EMEP ASSESSMENT 1980 - 2000

Comparison and control of Danish EMEP data at DMU and EMEP.

Niels Z. Heidam
DMU - National Environmental Research Institute, Denmark

1. Extent of data and assessment

The Danish EMEP network comprises the stations *Tange, Keldsnor, and Anholt (with acronyms TANG, KELD, ANHO).* The first two stations have been in operation since 1978 whereas the latter started in 1988-1989 and was appointed EMEP station in 1991 (air) and 1996 (precipitation). In the EMEP terminology these stations are named *DK03 Tange, DK05 Keldsnor, and DK08 Anholt*.
Daily measurements of airborne sulphur- and nitrogen compounds in gaseous and particulate form are performed at all stations and in precipitation concentrations of these compounds are determined from wet-only samples. Wet-only samples have been collected on a daily basis at *Tange* and *Keldsnor* and on a weekly basis at *Anholt*. From 1999 the sampling frequency for precipitation was changed to ½-monthly samples at all stations.
The objective for the data comparison is to establish the extent of agreement between the original data held in the database at DMU (the DMU-data) and the corresponding data reported annually to the EMEP/CCC at NILU in Norway. These data are referred to as EMEP-data and are available at the EMEP website[1].
Whereas the two datasets should be identical this assessment shows that this ideal has not quite been achieved. There can be many reasons for the disagreements, *e.g.* oversights or omissions by DMU or CCC, data misplacements, and misunderstandings in the communication between DMU and CCC or typing errors (in the early years).
In this context it should also be mentioned that the DMU data files were first organised in a relational database in 1987 and that the data were transferred to a more sophisticated version of the database in 1998.
Whatever the reasons for the deviations the original DMU-data must *à priori* be considered the correct ones.

The assessment concerns the following selection of pollutants:

A. Air concentrations: SO_2, SO_4^{2-}/S, NO_2, $TNH_4 = NH_3 + NH_4^+$, $TNO_3 = HNO_3 + NO_3^-$. It should be noted that the results for the two composite compounds are not stored in the DMU database. Instead the concentrations are calculated as the sums shown.
B. Precipitation concentrations: SO_4^{2-}, NH_4^+, NO_3^-, pH, Na, Mg, K, Ca.

The EMEP Assessment covers officially the period 1980 – 2000, but here the period since the start of the measurements 1978 is also included. All data for the above components are included, also results that have failed in the quality control, maybe fatally so, and have been flagged accordingly.

[1] www.emep.int (/ccc/measurements)

2. Matches and mismatches between DMU and EMEP data

Data extracted from the DMU database and data downloaded from the EMEP website were combined to construct aggregate data records.
The essentials of such an aggregate data record looks as follows:

station	date	compound	unit	DMU	EMEP

Here DMU and EMEP denotes the concentration values recorded in the DMU-database and at the EMEP website, respectively, for the specified compound at the given station and date.
Since the DMU- and EMEP-values can either agree or disagree and can either be present or missing the following cases arise:

A. Matches: DMU= EMEP[2].

1. Non-missing entries: The ideal situation.
2. Both DMU and EMEP values are missing: A quite normal occurrence where the two databases agree that no data are available; possible reasons are absence of precipitation or equipment failure. These 'no-data' entries have consequently been removed from the assessment.

No further action required.

B. Mismatches

1. Non-missing but deviating entries DMU ≠ EMEP: The EMEP-data must *à priori* be considered as erroneous, misplaced or possibly outdated.
2. The DMU value is present but the EMEP value is missing: A large part of these mismatches occur because some Danish stations have not always have been operating as EMEP stations.
3. The EMEP value is present but the DMU value is missing: The EMEP-data must *à priori* be considered as erroneous or misplaced with respect to date and/or station.

The extent of the mismatches should be mapped and the reasons for the mismatches should - if possible - be identified to avoid future mistakes.
The data at EMEP/CCC may need to be corrected, in particular if the mismatches occur frequently within a limited period or the deviation is large. Correction is certainly required if the general concentration level is affected. Updates e.g. to fill empty EMEP-periods may also be called for. Corrections at EMEP/CCC normally entail renewed transmission of correct data.

C. Mapping matches and mismatches

The data can be considered 3-dimensional with respect to station (space), date (time) and pollutant and are far too numerous to be considered on an individual basis. The aggregate records have therefore been subjected to mass processing in order to find on an annual basis the numbers of

[2] A note on the equality of data. Most data are produced digitally in the laboratory and that is reflected in the many (superfluous) decimals recorded in the DMU database. The EMEP-data are on the other hand presented on the website with a small number of decimals. This resolution is a reflection of the uncertainty of the measurement. Any deviation less than or equal to this resolution cannot be considered real. Therefore, the difference between entry such as EMEP= 0.73 and DMU= 0.7264438, which mathematically is *dev*= 0.0035562, must be considered as zero.
In view of the differing resolutions of the data we have therefore defined

$$DEV = EMEP - DMU \quad \text{and} \quad DEV = 0 \quad \text{if} \quad |dev| \leq 10^{-dec}$$

where *dec* is the smallest number of decimals occurring in the two results.

different cases described above. This is still a quite large body of data and the results have therefore been mapped on to the 3 'axes' of the data space, supplemented by a grand total.
In the mappings below the column OBS lists the total number of observations. It has been split into the number of simultaneous observations, termed DMU_EMEP, and the number of cases, MISSING where one of the results is missing. The simultaneous results can either be *Equal* (cf. note) or *Differ*. In the MISSING cases either the *DMU*-result or the *EMEP*-result is the *missing* one. Mismatches of rejected data, *i.e.* data with a fatal flag, are not included in the discussion and have been removed from the mapping.

D. Assessment of mismatches
The mismatches have been evaluated from the hypothesis that they do not have any influence on annual mean values.
Deviating results
Only annual mean deviations that differ from zero at a significance level of 5% will be considered here.
Missing results
These cases are evaluated from the total annual means of DMU- and EMEP-data with the exclusion of deviating results. These total annual mean thus contain, respectively, both results that are unique to either one of them and those that are equal in the two sets. If the two annual means differ from zero at a significance level of 5% then the missing results are of importance.

3. Atmospheric data

Overview

The summary covers 3 stations, 23 years, and 5 components with a total of almost 70 000 records. On the average there are thus approximately

- 14 000 records per component;
- 3 000 records per year;
- 600 records per year and component;
- 23 000 records per station.
- 4 600 records per station and component;
- 1 000 records per station-year;
- 200 records per station-year and component.

Data modifications

For the comparison some modifications to the aggregate data records had to be introduced. Thus it has been necessary to take account of the different reference conditions used in the two data systems. Whereas the EMEP-concentrations are given at 20^0 C, DMU-concentrations are referred to 0^0 C. A common reference temperature of 20^0 C was selected for the aggregate assessment records, and a 6 % reduction (20/293) was consequently applied to the DMU-results.

3.1 Mapping of atmospheric data

The 3 dimensional mapping of the atmospheric records is shown in Table 1a and Table 1b below. As mentioned above this is a censored mapping where all fatally flagged data that either differ or are missing in one of the databases have been removed from the tables.

Table 1a. Aggregate atmospheric records mapped by components and by stations.

ATMOSPERE			OBS	DMU_EMEP	Equal	Differ	MISSING	Miss_DMU	Miss_EMEP
STATION	Year	Comp							
ALL	ALL	ALL	69387	65620	65588	32	3767	165	3602
		NO2-N	3760	3380	3380	0	380	13	367
		SO2-S	20014	19196	19195	1	818	4	814
		SO4-S	20360	19529	19526	3	831	6	825
		TNH4-N	12598	11732	11732	0	866	70	796
		TNO3-N	12655	11783	11755	28	872	72	800
ANHO	ALL	ALL	20449	16860	16859	1	3589	13	3576
KELD			24202	24047	24018	29	155	147	8
TANG			24736	24713	24711	2	23	5	18

As expected the great majority of data registered by DMU and EMEP, almost 95 % are not only simultaneous but also equal. The remaining 5 % of the records are incomplete and hold only single results.
It is a gratifying surprise that only 32 cases of valid data contain discrepancies (in the sense defined above). As can be seen the differing results pertain mainly to TNO_3 at *Keldsnor*, supplemented with a few cases for SO_2 and SO_4.

Table 1b. Aggregate atmospheric records mapped by years.

ATMOSPERE			OBS	DMU_EMEP	Equal	Differ	MISSING	Miss_DMU	Miss_EMEP
STATION	Year	Comp							
ALL	ALL	ALL	69387	65620	65588	32	3767	165	3602
	1978		990	988	988	0	2	0	2
	1979		1305	1304	1304	0	1	0	1
	1980		1376	1376	1376	0	0	0	0
	1981		1424	1424	1424	0	0	0	0
	1982		1416	1416	1413	3	0	0	0
	1983		1360	1360	1360	0	0	0	0
	1984		1312	1312	1312	0	0	0	0
	1985		1578	1438	1438	0	140	140	0
	1986		1286	1286	1286	0	0	0	0
	1987		1454	1454	1454	0	0	0	0
	1988		1861	1491	1491	0	370	0	370
	1989		4089	2594	2594	0	1495	3	1492
	1990		5050	3318	3290	28	1732	1	1731
	1991		4401	4400	4400	0	1	1	0
	1992		4181	4178	4178	0	3	0	3
	1993		4340	4335	4335	0	5	5	0
	1994		4583	4583	4583	0	0	0	0
	1995		4544	4544	4544	0	0	0	0
	1996		4644	4644	4643	1	0	0	0
	1997		4452	4452	4452	0	0	0	0
	1998		4682	4680	4680	0	2	2	0
	1999		4393	4380	4380	0	13	13	0
	2000		4666	4663	4663	0	3	0	3

Table 1b shows that the differing TNO_3 results are all from 1990 at *Keldsnor* and a closer look reveals that they occur in August and September and that in all cases the EMEP results are substantially smaller than the DMU results. Since the EMEP results are neither resembling HNO_3 nor NO_3 there is no obvious reason for the discrepancies and the EMEP results must be considered erroneous.
It turns out that the differing results for SO_4 all occur early in 1982 with one case at *Keldsnor* and two at *Tange*. For one case at each of these stations a typing error seems to be the cause.

Tables 1 also shows that the vast majority of the incomplete records hold only DMU results but no EMEP-data. More details on missing data can be seen in Table 2, which is a mapping by station and year.

Table 2. Missing atmospheric results mapped by stations and years.

ATMOSPERE	Miss_DMU				Miss_EMEP			
	ALL	ANHO	KELD	TANG	ALL	ANHO	KELD	TANG
ALL YEARS	**165**	**13**	**147**	**5**	**3602**	**3576**	**8**	**18**
1978	0	.	0	0	2	.	1	1
1979	0	.	0	0	1	.	0	1
1980	0	.	0	0	0	.	0	0
1981	0	.	0	0	0	.	0	0
1982	0	.	0	0	0	.	0	0
1983	0	.	0	0	0	.	0	0
1984	0	.	0	0	0	.	0	0
1985	140	.	140	0	0	.	0	0
1986	0	.	0	0	0	.	0	0
1987	0	.	0	0	0	.	0	0
1988	0	0	0	0	370	367	1	2
1989	3	0	0	3	1492	1482	5	5
1990	1	0	1	0	1731	1725	0	6
1991	1	0	1	0	0	0	0	0
1992	0	0	0	0	3	2	0	1
1993	5	1	3	1	0	0	0	0
1994	0	0	0	0	0	0	0	0
1995	0	0	0	0	0	0	0	0
1996	0	0	0	0	0	0	0	0
1997	0	0	0	0	0	0	0	0
1998	2	0	2	0	0	0	0	0

ATMOSPERE	Miss_DMU				Miss_EMEP			
	ALL	ANHO	KELD	TANG	ALL	ANHO	KELD	TANG
1999	13	12	0	1	0	0	0	0
2000	0	0	0	0	3	0	1	2

A large majority of 3574 missing EMEP-data is from *Anholt* in 1988 – 1990, *i.e.* from the pre-EMEP period of this station. Accordingly the missing EMEP-data are distributed evenly over the compounds except NO_2 where measurements did not start until late in 1989.

The few cases (only about 0.25 % of all data) where results are missing from DMU but exist at EMEP are more difficult to understand and need to be checked.
The majority of the missing DMU-results are dominated by the composite compounds TNH_4 and TNO_3, mainly at *Keldsnor* in August - October 1985 (2·69 cases). Since measurements of the constituting components that are used for calculating these phase sums, NH_3, HNO_3, and NO_3, did not start at the Danish EMEP stations until late in 1988 these EMEP-data must be erroneous. Furthermore it turns out that most of the remaining mismatches are also caused by erroneous entries in the EMEP-database. These EMEP-values appear to have arisen as a false-memory effect when there is a gap in the DMU-values reported to EMEP. The EMEP-values in these composite records with no corresponding DMU-data are in fact all identical to the common value in the most recent case of simultaneous (and equal) EMEP-and DMU-values.
This false-memory effect explains the mismatches for the 12 cases of missing DMU-data for NO_2 at *Anholt* in 1999 (Tables 1 and 2, Fig. 1c). This also applies to one case of TNO_3 and the sole remaining case of missing DMU-data on TNH_4 at *Keldsnor* in March 1998 and for one case of TNO_3 at Tange in August 1999. The corresponding EMEP-data are artefacts and should be deleted. Finally for 5 of the 6 cases of missing DMU-results for SO_4 the corresponding EMEP-values are zero. These EMEP-data are probably also artefacts.

The remaining incidents of missing DMU data are then reduced to only 7 solitary cases for which there seems to be no explanation. In line with the general view adopted at the outset these EMEP-data must be considered erroneous.

Some examples of the cases described are illustrated graphically in Fig. 1.

3.2 Assessment of atmospheric mismatches

Deviating results

Only the deviations found for the 28 records of TNO_3 from 1990 at *Keldsnor* lead to annual deviation means that differ significantly from zero.

Missing results

No significant effects on annual mean values were found.

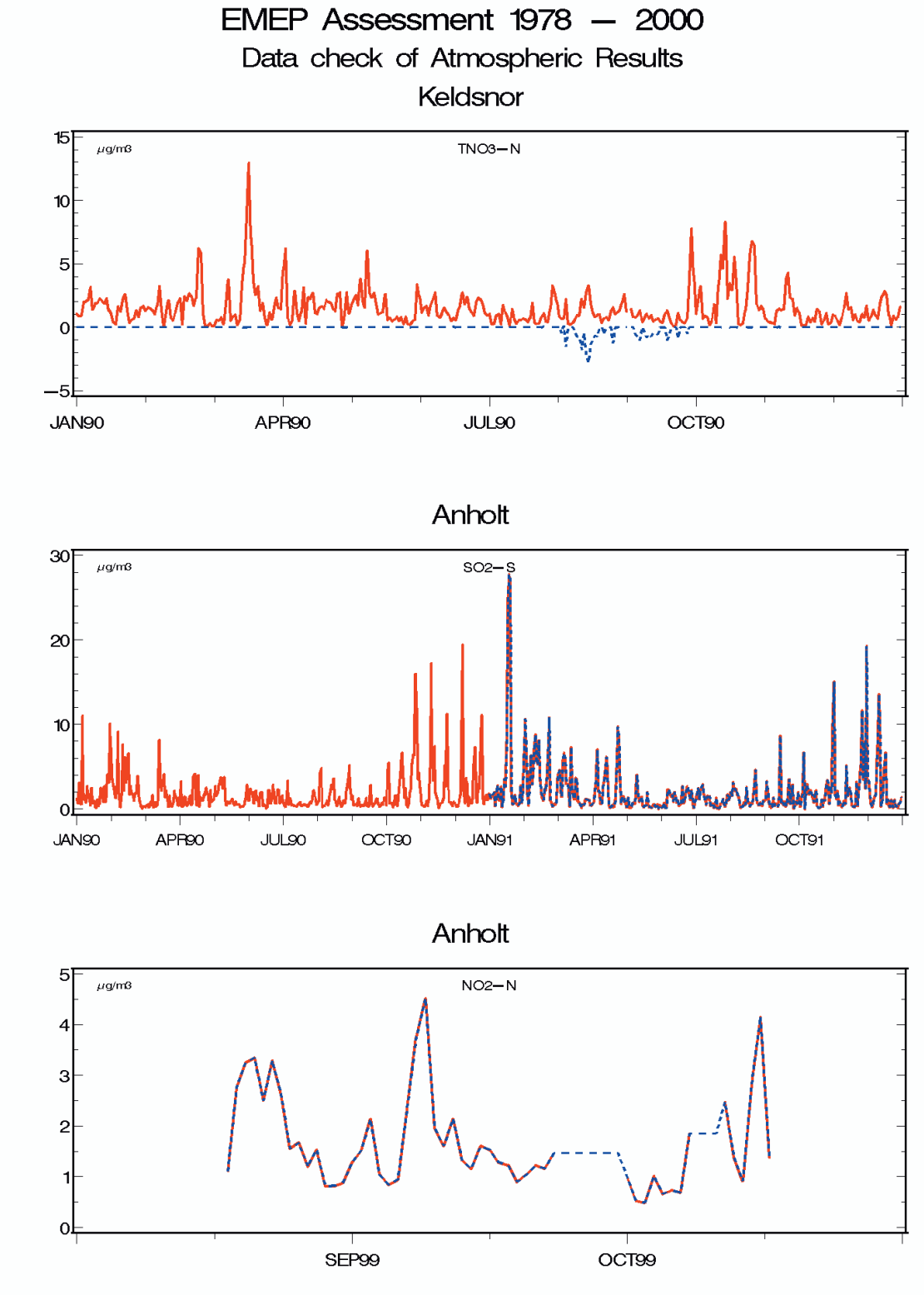

Figure 1. Time series of DMU-data (full red) and EMEP-data or Deviations (broken blue).

a. Differing data in 1990, Deviation= EMEP – DMU.
b. Example of data mismatches and matches in the transition period when the purely national station Anholt also became an EMEP station.
c. Example of false-memory effect in EMEP data in October 1999.

3.3 Conclusions and recommendations for atmospheric data

On the basis of the findings and assessments of the mismatches it is it is clear that some results will have to be corrected and retransmitted.

1. **All**[3] results from *Anholt* prior to 1991 will be reported to EMEP;
2. The results for TNO_3 from *Keldsnor* in August – September 1990 will be retransmitted to EMEP (cf. 5. below);
3. The results for NO_2 at *DK08 Anholt* from October 1999 will be retransmitted to EMEP;
4. The 3 cases of deviating and erroneous SO_4 results from 1982 at *Keldsnor* on 24. January and *Tange* on 19. January and 24. February will be corrected and retransmitted to EMEP;
5. The phase sums of nitrogen, TNH_4 and TNO_3 will in the future be calculated and reported as the sum of the individual constituents, if - and only if - both the gaseous and the particulate result is non-missing. Otherwise the total result will be defined as missing. For non-missing sums the 'worst' flag of the constituents will be attached to the result.
6. There are at present no plans to correct the remaining mismatches. They are very few and seem to occur randomly as solitary concentration values for individual pollutants and have furthermore been found to be of no significance for annual means.

It is recommended to EMEP-CCC that:

1. The procedures for handling gaps in the reported data be scrutinized in order to avoid the production of false-memory data or artificial zero data;
2. The results in the EMEP-database for the phase sums of nitrogen, TNH_4 and TNO_3, prior to 1988 – *i.e.* at *DK05 Keldsnor* in August – October 1985 be deleted;

[3] **All** results should in this context be interpreted as all the official EMEP-components measured in Denmark, including also those that are not considered here, such as the elements and heavy metals detected in aerosols by PIXE.

4. Precipitation data

Overview

The summary covers 3 stations, 23 years, and 8 components with a total of almost 37 000 records. On the average there are thus approximately

- 4 500 records per component;
- 1 600 records per year;
- 200 records per year and component;
- 12 000 records per station.
- 1 500 records per station and component;
- 500 records per station-year;
- 65 records per station-year and component.

Data modifications

For the comparison some modifications to the aggregate data records had to be introduced. Thus it was necessary to take account of retrospective corrections that DMU applied in 1991 to all pH-results prior to 1990. The corrections take on one of two constant values, -0.10 until the end of 1985 and -0.12 until the end of 1989. These corrected results were apparently never reported to EMEP. The EMEP-values for pH in this assessment therefore had to be similarly amended.

Also the dating peculiarities of the two data systems had to be taken into consideration. Whereas both the EMEP- and the DMU-data for 1-day samples are marked with the date of the sampling start, DMU samples taken over more than 1 day[4] are, according to meteorological tradition, marked with the end-date of the sampling period. Since the DMU records also hold the start date this change to the DMU-data was introduced for the aggregate assessment records. A closer study revealed, however, that these dates were not always correct – or at least not equal to the starting dates in the EMEP-data, once reported by DMU. However, in many instances agreement could be obtained from neighbouring DMU-samples. Clearly the dating control procedures at DMU for weekly, now half-monthly, samples have not been given sufficient priority.

The mapping described below was based on aggregate records amended according to these findings. But it may contain a few artificial mismatches since it has not been possible to reproduce completely all sample start-dates reported to EMEP.

4.1 Mapping of precipitation data

The 3 dimensional mapping of the precipitation records is shown in Table 4a and Table 4b below. As mentioned above this is a censored mapping where all fatally flagged data that either differ or are missing have been removed from the tables.

As expected the great majority of data registered by DMU and EMEP, more than 97 % are not only simultaneous but also equal. The remaining 3 % of the records contain incomplete data with only one result where the majority of 93 % are missing from the EMEP records for Anholt, which as explained in the Introduction was initially operating as a non-EMEP station for 7 years.

Table 4a. Aggregate precipitation records mapped by components and by stations.

PRECIPITATION			OBS	DMU_EMEP	Equal	Differ	MISSING	Miss_DMU	Miss_EMEP
STATION	Year	Comp							
ALL	ALL	ALL	36815	35912	35900	12	903	52	851
		Na	3328	3222	3218	4	106	5	101
		Mg	3320	3218	3218	0	102	3	99
		K	3287	3185	3185	0	102	3	99
		Ca	3563	3462	3462	0	101	1	100
		NH4-N	5752	5622	5622	0	130	10	120
		NO3-N	6004	5885	5882	3	119	12	107
		SO4-S	6034	5917	5914	3	117	10	107
		pH	5527	5401	5399	2	126	8	118
ANHO	ALL	ALL	2243	1445	1444	1	798	8	790
KELD			15706	15661	15657	4	45	23	22
TANG			18866	18806	18799	7	60	21	39

It is a gratifying surprise that only 12 cases of valid data contain discrepancies. These differing results pertain mainly to scattered results from *Keldsnor* and *Tange* in 1988. But there is no obvious explanation for these deviations.
The samples in this period were all on a daily basis and so they were originally marked – presumably quite correctly - with the starting date. As the values as explained above nevertheless have been tentatively – and unsuccessfully - matched with the neighbouring samples it is unlikely that the deviations can be caused by a dating error. It must therefore be assumed that the EMEP-data are erroneous.

Table 4b. Aggregate precipitation records mapped by years.

PRECIPITATION			OBS	DMU_EMEP	Equal	Differ	MISSING	Miss_DMU	Miss_EMEP
STATION	Year	Comp							
ALL	ALL	ALL	36815	35912	35900	12	903	52	851
	1978		1238	1208	1207	1	30	12	18
	1979		1163	1160	1160	0	3	0	3
	1980		1169	1166	1166	0	3	1	2
	1981		1060	1050	1050	0	10	0	10
	1982		1073	1071	1071	0	2	1	1
	1983		1024	1024	1024	0	0	0	0
	1984		1039	1039	1039	0	0	0	0
	1985		1114	1114	1114	0	0	0	0
	1986		1163	1158	1158	0	5	0	5
	1987		1833	1833	1833	0	0	0	0
	1988		2184	2142	2132	10	42	26	16
	1989		1828	1689	1689	0	139	0	139
	1990		2083	1897	1897	0	186	1	185
	1991		1656	1488	1488	0	168	0	168
	1992		1987	1819	1819	0	168	0	168
	1993		2279	2255	2255	0	24	0	24
	1994		2278	2278	2278	0	0	0	0
	1995		1986	1881	1881	0	105	0	105
	1996		1800	1797	1797	0	3	0	3
	1997		2068	2064	2064	0	4	0	4
	1998		2889	2889	2889	0	0	0	0
	1999		1526	1521	1520	1	5	5	0
	2000		375	369	369	0	6	6	0

The missing results are distributed rather inhomogenously over the years, as shown in Table 4b. The missing EMEP-data are most numerous in 1989-1995 and apart from that period it is observed that the remaining few missing data from both databases are most numerous in 1978, 1981, and 1988.

A more detailed mapping of missing results by station and year is shown in Table 5.

Table 5. Missing precipitation results mapped by stations and years.

PRECIPITATION	Miss_DMU				Miss_EMEP			
	ALL	ANHO	KELD	TANG	ALL	ANHO	KELD	TANG
ALL YEARS	**52**	**8**	**23**	**21**	**851**	**790**	**22**	**39**
1978	12	.	12	0	18	.	7	11
1979	0	.	0	0	3	.	1	2
1980	1	.	0	1	2	.	0	2
1981	0	.	0	0	10	.	2	8
1982	1	.	1	0	1	.	1	0
1983	0	.	0	0	0	.	0	0
1984	0	.	0	0	0	.	0	0
1985	0	.	0	0	0	.	0	0
1986	0	.	0	0	5	.	5	0
1987	0	.	0	0	0	.	0	0
1988	26	.	7	19	16	.	6	10
1989	0	0	0	0	139	139	0	0
1990	1	0	0	1	185	184	0	1
1991	0	0	0	0	168	168	0	0
1992	0	0	0	0	168	168	0	0
1993	0	0	0	0	24	24	0	0
1994	0	.	0	0	0	.	0	0
1995	0	0	0	0	105	104	0	1
1996	0	0	0	0	3	3	0	0
1997	0	0	0	0	4	0	0	4
1998	0	0	0	0	0	0	0	0
1999	5	5	0	0	0	0	0	0
2000	6	3	3	.	0	0	0	.

A great majority of 787 missing EMEP-data is from the pre-EMEP period 1989 – 1995 at *Anholt*. As can be seen from Table 4a the missing EMEP-data are accordingly distributed evenly over the compounds. The remaining missing EMEP-data, most numerous at *Keldsnor* and *Tange* in 1978, 1981 and 1988, are scattered in an apparent random order over both dates and components, not consistent with simply missing samples. The samples in this period were all on a daily basis and so they were originally marked – presumably quite correctly - with the starting date. The absence from the EMEP-database of these few results therefore remains unexplained.

This random scatter also applies to the very few missing DMU-data. They are most numerous at *Keldsnor* in the first few days of 1978 when the monitoring programme started, and in 1988 where the reason might be one or two misplaced samples, although the dating of the 1-day samples of this period presumably is correct. Otherwise there is no common feature among the absence of these data and their solitary appearance in the EMEP-database remains unexplained.
In line with the general view adopted at the outset these EMEP-data must be considered erroneous.

4.2 Assessment of precipitation mismatches

Deviating and missing results

There are no annual deviation means and no deviations between annual means that - at a significance level of 5% - differ from zero because of mismatches between EMEP- and DMU-results.

4.3 Conclusions for precipitation data

1. **All**[5] results from *Anholt* prior to 1996 will be reported to EMEP;
2. All pH-results prior to 1990 will be retransmitted to EMEP;
3. The quality control procedures for dating the samples have been tightened.
4. There are at present no plans to correct the remaining mismatches. They are very few and seem to occur randomly as solitary concentration values for individual pollutants and have furthermore been found to be of no significance for annual means.