

# **GEOmon Harmonized data set of in-situ ground based data of O3, NO2, CO**

Stephan Henne ([stephan.henne@empa.ch](mailto:stephan.henne@empa.ch))

Zoë Fleming ([zf5@leicester.ac.uk](mailto:zf5@leicester.ac.uk))

Version 0

2008-07-21

<b>GEOmon Harmonized data set of in-situ ground based data of O3, NO2, CO</b> .....	<b>1</b>
<b>1 Introduction</b> .....	<b>2</b>
1.1 Version 0.....	2
<b>2 Data product</b> .....	<b>2</b>
<b>3 Data sources</b> .....	<b>7</b>
3.1 EMEP.....	7
3.2 WDCGG.....	7
3.3 Individual data providers.....	7
3.3.1 PAES network (France).....	7
3.3.2 Cabauw.....	7
3.3.3 Weybourne .....	8
3.3.4 Monte Velho .....	8
<b>4 Data treatment</b> .....	<b>8</b>
4.1 EMEP.....	8
4.2 WDCGG.....	8
4.3 Individual data providers.....	9
<b>5 Merging data</b> .....	<b>9</b>

# 1 Introduction

The purpose of this dataset is to provide a set of observations that

- Cover the European continent
- Are reported with uniform time stamps and units
- Are formatted uniformly and in an easy to read manner
- Are checked by a uniform quality control procedure
- Are uniformly quality flagged

The current (preliminary) version of the dataset focuses on uniform data format and reporting times and units and data flagging information that is already available while the final version will include the results of additional quality control procedures.

For publications using data from any of the selected sites the authors are asked to contact the data providers about acknowledgements and give proper reference to the data source.

## 1.1 Version 0

The following steps for 1-hourly and daily data harmonization were performed for version 0:

- Alignment of time stamps to UTC, including information of time stamp relative to observation interval
- Conversion of all data to units volume mixing ratio for all selected parameters
- Merging data from WDCGG, EMEP and local PIs:
  - French PAES network (3 additional non-EMEP sites)
  - Cabauw, NL
  - Weybourne, GB
  - Monte Velho, PT
- Plausibility tests for data range
- Manual QC for data received from additional sites
- Uniform flagging of data
- Uniform ASCII format and R object file for internal use

# 2 Data product

A short description of the format of the harmonized dataset and the data availability is given here before the procedures to derive the harmonized dataset are discussed below.

For each site one file containing hourly and one file containing daily data (not yet finished) were produced.

The data is formatted in a simple comma separated ASCII layout with the columns:

All times are given as UTC and all trace gases are quantified as volume mixing ratios with units ppb.

dtm.start beginning of measurement period  
 dtm.end end of measurement period  
 O3 ozone volume mixing ratio (ppb)  
 O3flag 0: valid, >0: invalid, questionable, etc.  
 O3origin Where the data originated from, currently only "EMEP", "WDCGG", "Other"  
 .... The same for NO2 and CO

Figure 2 and Figure 3 give an overview of the data availability of the whole dataset. Station abbreviations are defined in Table 1 and the location of the sites within Europe is shown in Figure 1.

**Table 1: List of sites included in harmonized dataset.**

Station	ID	Longitude (°E)	Latitude (°N)	Altitude (m) a.s.l.
Bialystok	BIA	22.750	53.200	120
Birkenes	NO01	8.250	58.383	190
Cabauw	NL11	4.933	51.967	60
Campisabalos	ES09	-3.150	41.283	1360
Donon	FR08	7.133	48.500	775
Finokalia	GR02	25.667	35.317	150
Harwell	GB36	-1.317	51.567	137
Hegyhatsal	HNG	16.650	46.950	344
Hohenpeissenberg	HPB	11.017	47.800	985
Ispra	IT04	8.633	45.800	209
Jungfrauoch	CH01	7.983	46.550	3580
K-pusztta	HU02	19.583	46.967	125
Kollumerwaard	NL09	6.283	53.333	0
Kosetice	CZ03	15.083	49.583	534
Lampedusa	LMP	12.633	35.517	60
Lough Navar	GB06	-7.900	54.433	126
Mace Head	IE31	-9.900	53.333	25
Mahón	ES06	4.250	39.900	10
Monte Cimone	CMN	10.683	44.167	2165
Monte Velho	PT04	-8.800	38.083	43
Neuglobsow	DE07	13.033	53.150	62
Observatoire de Haute-Provence	OHP	5.700	43.917	650
Pic du Midi	PDM	0.167	43.067	2860
Preila	LT15	21.067	55.350	5
Puy de Dome	PUY	3.000	45.750	1465
Roquetas	ES03	0.500	40.817	50
Schauinsland	DE03	7.900	47.917	1205
Schmücke	DE08	10.767	50.650	937
Sniezka	PL03	15.733	50.733	1604
Sonnblick	AT34	12.967	47.050	3106
Weybourne	WEY	1.333	52.950	16
Zavizan	HR04	14.983	44.817	1594

Station	ID	Longitude (°E)	Latitude (°N)	Altitude (m) a.s.l.
Zingst	DE09	12.733	54.433	1
Zugspitze	ZUG	10.983	47.417	2950

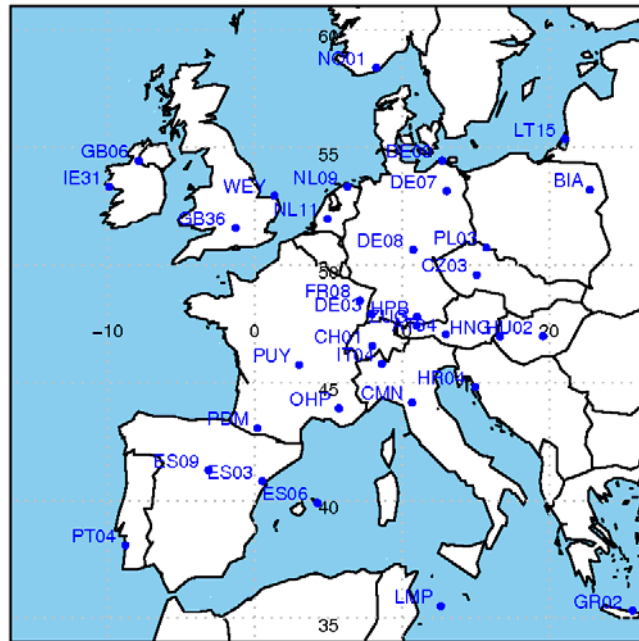
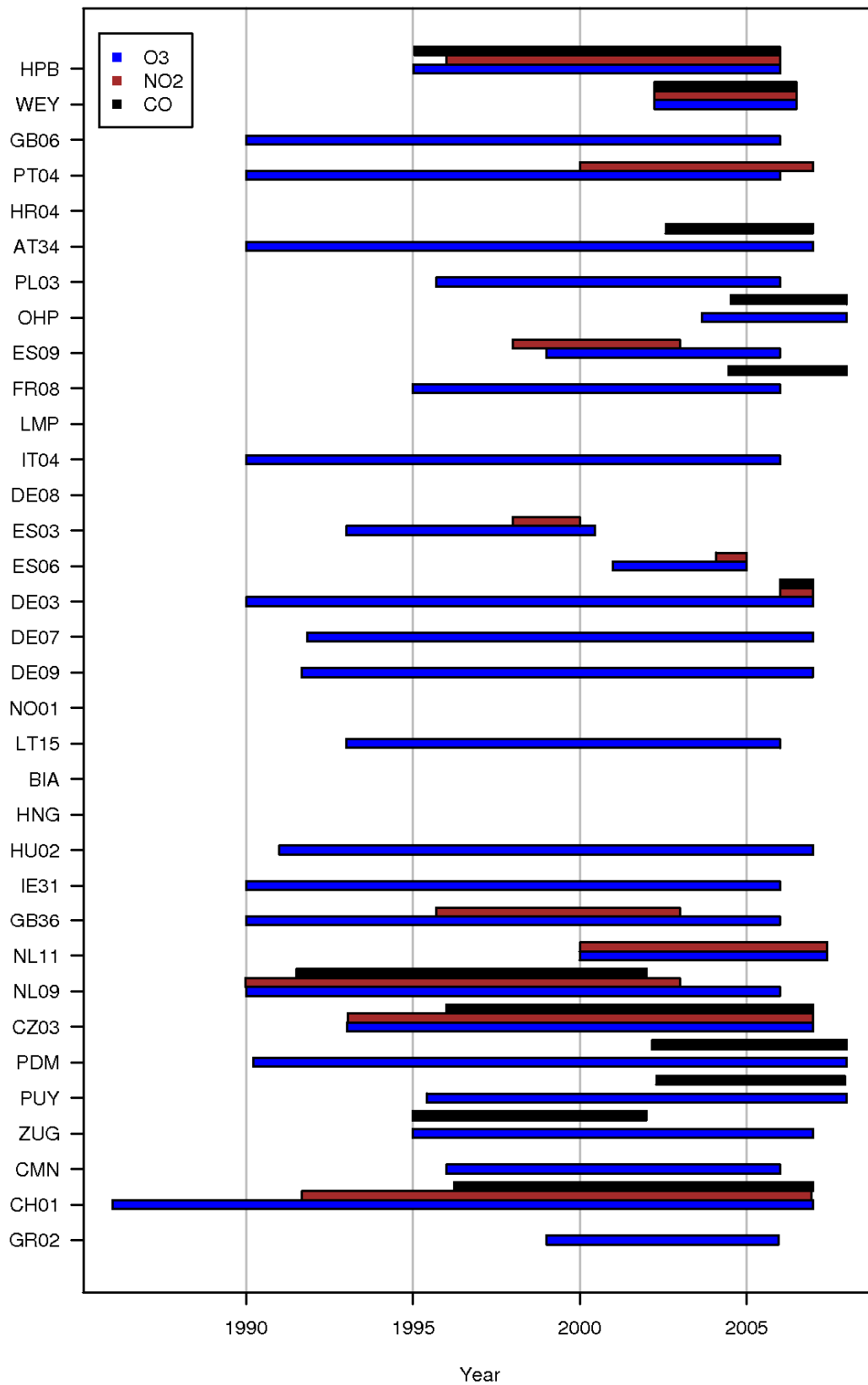
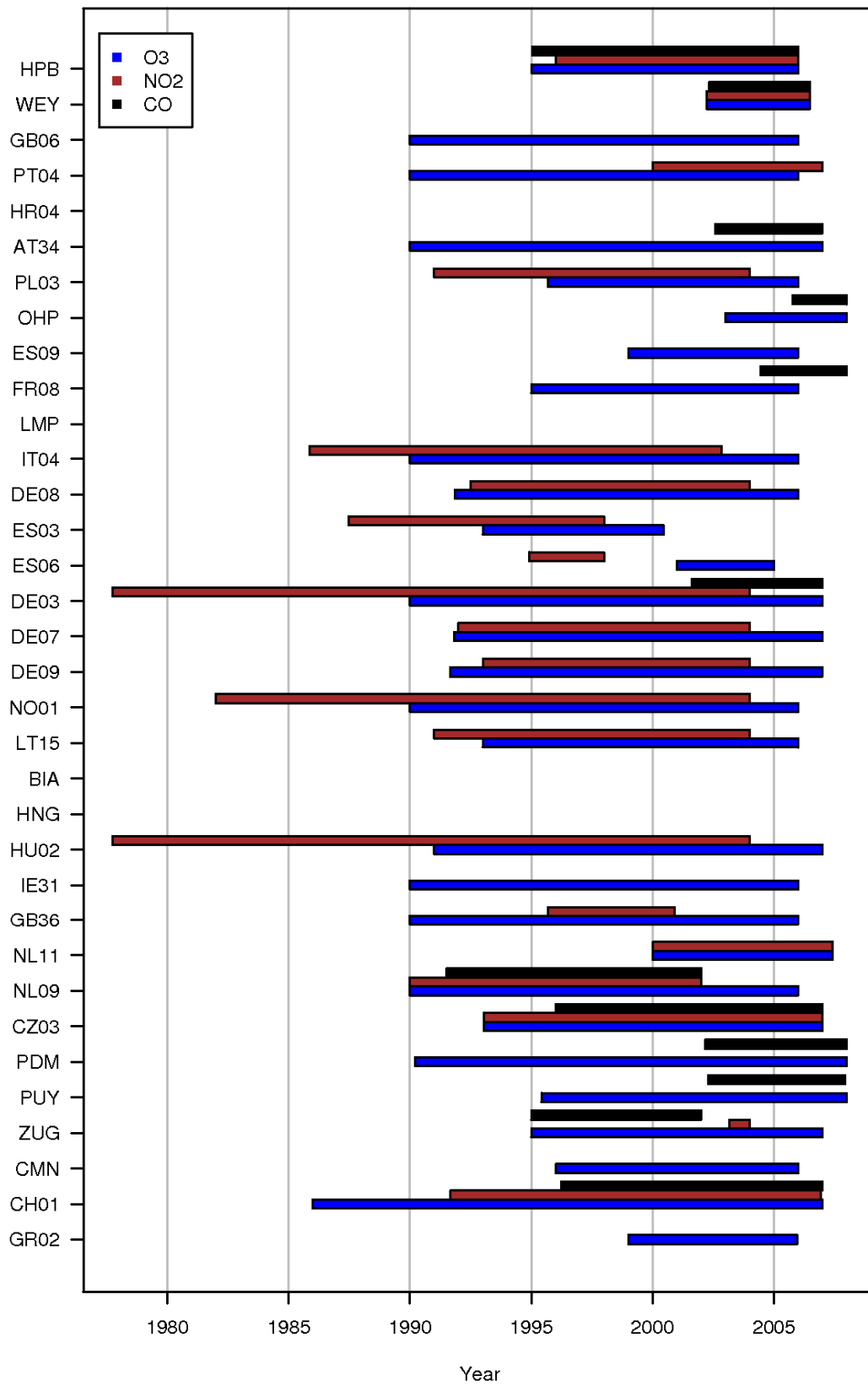


Figure 1: Location of selected sites presented in Table 1.



**Figure 2: Availability of hourly data.**



**Figure 3: Availability of daily data.**

## 3 Data sources

This section describes the originators of the data. These are mainly from EMEP and WDCGG but also through personal communication with network or station PIs.

### 3.1 EMEP

EMEP data were download on: 2008-05-09. EMEP data start in 1990 and currently include data up to end of 2005. Main and ozone data were downloaded from: <http://www.nilu.no/projects/ccc/emepdata.html> (web download).

Additional information about time stamps received from Anne-Gunn Hjellbrekke ([annehj@nilu.no](mailto:annehj@nilu.no)). O<sub>3</sub> data contain time zone. NO<sub>2</sub> data are given in UTC. O<sub>3</sub> time stamp is at the beginning of the period. NO<sub>2</sub> data contain two time stamps for the beginning and end of the measurement interval.

EMEP data are provided as  $\mu\text{g}/\text{m}^3$  (for O<sub>3</sub>) or  $\mu\text{g N}/\text{m}^3$  (for NO<sub>2</sub>) at standard atmospheric conditions (293.15 K and 1013.25 hPa).

### 3.2 WDCGG

WDCGG data were downloaded 2007-11-01

Time stamp information is given as start and end of measurement interval. Time zone UTC.

WDCGG data files contain information about the units the data are provided in. Units are not uniform for each species.

### 3.3 Individual data providers

#### 3.3.1 PAES network (France)

Data were kindly provided by Francois Gheusi ([francois.gheusi@aero.obs-mip.fr](mailto:francois.gheusi@aero.obs-mip.fr)) and Yves Meyerfeld.

Time stamps correspond to UTC and the end of the 1-h measurement interval.

All provided data were given as volume mixing ratios (ppbv) independent of species.

No NO<sub>2</sub> data but only NO<sub>x</sub> data were provided.

#### 3.3.2 Cabauw

Data were kindly provided by the Dutch National Institute for Public Health and the Environment (RIVM) and Alex Vermeulen ([a.vermeulen@ecm.nl](mailto:a.vermeulen@ecm.nl)).

Time stamps correspond to Central European Time and the end of the 1-h measurement interval.

All provided data were given as mass concentrations ( $\mu\text{g}/\text{m}^3$ ) at standard atmospheric conditions (293.15 K and 1013.25 hPa).

### 3.3.3 Weybourne

Data were provided by the British Atmospheric Data Centre (BADC) and Brian Bandy ([B.Bandy@uea.ac.uk](mailto:B.Bandy@uea.ac.uk)).

Time stamps correspond to UTC and the beginning of the 1-h measurement interval.

All provided data were given as volume mixing ratios (ppb).

### 3.3.4 Monte Velho

NO<sub>2</sub> data were kindly provided by Paulo Beliche ([paulo.beliche@ccdr-a.gov.pt](mailto:paulo.beliche@ccdr-a.gov.pt)).

Time stamps correspond to UTC and the beginning of the 1-h measurement interval.

All provided data were given as mass concentrations (µg/m<sup>3</sup>) at standard atmospheric conditions (293.15 K and 1013.25 hPa).

## 4 Data treatment

### 4.1 EMEP

All EMEP data were converted to volume mixing ratios (ppbv) assuming standard conditions (20°C, 1013.25 hPa).

Some time stamps were formatted incorrectly (e.g. 01.13.1990 instead of 01.01.1991) and had to be modified.

No flag information on O<sub>3</sub> and hourly NO<sub>2</sub> were available. Flags for daily NO<sub>2</sub> measurements were available. These were kept for the harmonized data set. 0 corresponds to an unquestionable valid measurement, while other flags indicate a problem with the measurement, however, EMEP does not consider all flags that are not 0 to be invalid. Combinations of different flags are also found in the data. Flags are given in three groups of three digit identifiers. The definitions can be found under: <http://www.nilu.no/projects/CCC/flags/flags.html>.

Where necessary daily aggregates were created from hourly data. Averages of days with less than 75 % valid data available were flagged as invalid (flag: 1).

### 4.2 WDCGG

All WDCGG data were converted to volume mixing ratios (ppbv). Usually the reporting units were provided with the data files. The following units were discovered in the global WDCGG dataset:

"ppm"

"ug/m3-20C" (Assuming 293.15 K and 1013.25 hPa)

"mg/m3 at 20" (Assuming 293.15 K and 1013.25 hPa)

"mg/m3-25C" (Assuming 298.15 K and 1013.25 hPa)

"ug/m3" (Assuming standard conditions 293.15 K and 1013.25 hPa)



"ug/m3-25C" (Assuming 298.15 K and 1013.25 hPa)  
"ugN/m3-20C" (Assuming 293.15 K and 1013.25 hPa)  
"ugN/m3-25C" (Assuming 298.15 K and 1013.25 hPa)  
"µg/m<sup>3</sup> - 20 °C" (Assuming 293.15 K and 1013.25 hPa)  
"ugN/m3" (Assuming 298.15 K and 1013.25 hPa)  
"ppbv"  
"ppb"

Where flagging information was available it was converted to a valid/invalid flag.

Daily aggregates were created from hourly data. Averages of days with less than 75 % valid data available were flagged as invalid (flag: 1).

### **4.3 Individual data providers**

Next to the unit and time stamp conversion the data were plausibility checked and a brief manual inspection and subsequent flagging of obviously erroneous data was performed.

Daily aggregates were created from hourly data. Averages of days with less than 75 % valid data available were flagged as invalid (flag: 1).

## **5 Merging data**

From all data sources individual data files by station were created. In a final merge from all data sources one individual file per site was created.

The merge procedure takes into account multiple data sources for an individual parameter. Currently only the longest available (number of valid data points) time series is considered in the merge process to avoid mixing of data sources within the same parameter column.