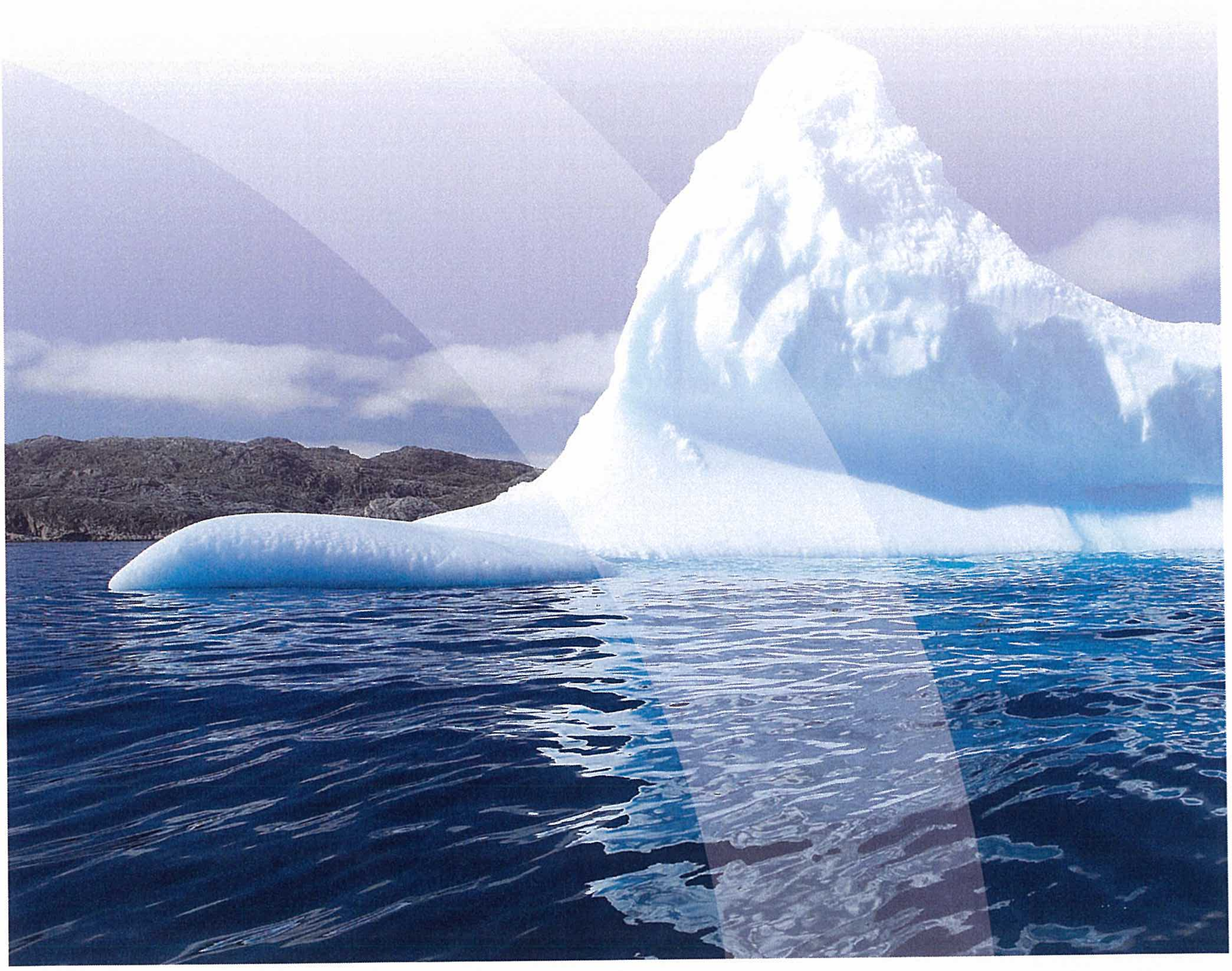
TR 1/2011

Norsk institutt for luftforskning
Norwegian Institute for Air Research

# HTAP-EBAS

N° 070307/2007/481644/MAR/C5
Final Report, Version 1.0

Kjetil Tørseth, Paul Eckhardt, Aasmund Fahre Vik, Michael Schulz



Technical report

TR 1/2011

# HTAP-EBAS

## N° 070307/2007/481644/MAR/C5
## Final Report, Version 1.0

Kjetil Tørseth, Paul Eckhardt, Aasmund Fahre Vik (NILU)
Michael Schulz (met.no)

**Technical report**

# 1 Introduction

The Convention on Long-Range Transboundary Air Pollution has set up a Task Force on Hemispheric Transport of Air Pollution which aims at obtaining a fuller understanding of the role of hemispheric transport of air pollution and specifically to have tools for access to integrated observations and a data set for the evaluation of chemical transport models. A specific service contract (this project) was designed to provide data sets for evaluation of models and more generally create a global system of ground based observations of transboundary air pollution.

The objective of the contract was specifically to

- Define **a set of generic data standards** for observations of air pollution (air quality, deposition and vertical profiles and linked meta information) that is compatible with the requirements of evaluation of the performance of chemical transport models. This set of standards should build on existing efforts of naming conventions and be flexible to accommodate the main standards in use for air pollution information and to be adapted to modern information and communication data technologies.
- Provide a **core data set** of observations suitable for evaluation of chemical transport models in the assessment of hemispheric transport of air pollution. This core set should include the information on ozone and its precursors and Particulate Matter and its precursors. The core data set has to be relevant for the years 2000/01 and 2005/06 although also data for other years are of relevance and thereafter having the potential of being annually updated through automatic schemes.
- Provide **recommendations** for further work to fully streamline the access to observational data.

## 1.1 Approach

To support long term operations and availability of the core observational data set, it was decided to utilize the existing operational system EBAS (EMEP database) for archival and dissemination of the data. Focus was put on the creation of conversion routines to convert existing data from different regional monitoring networks into a common data format suitable as input to EBAS. This enabled harmonization of observations and export of data with the regular EBAS tools and interfaces into a common format with common metadata and flagging routines.

## 1.2 Deliverables

The contract specifies seven deliverables:

D1: Inception report. Proposed a detailed work plan and formed a basis for the Contract kick-off meeting (held 21 February 2008 (Participants were Zuber, Kobe, Dentener, Schulz, Schultz, Fahre Vik, Eckhardt, Tørseth).

D2: Updated final work plan (Deadline 20 March, delivered 15 March 2008).

D3: Interim report for progress of work under WP1 and WP2 (delivered March 2009)

D4: Final report (this document) (delivered December 2010)

D5: A preliminary core dataset made available through the NADIR database (September 2008)(delivered)

D6: A draft core dataset available through the TFHTAP database (fully operational) (delivered September 2009)

D7: An updated final set of observations available at the end of project (delivered November 2010). This final set of observations is termed the HTAP Observations Database … (HTAP-Obs).

## 1.3 Definition of terms

- EBAS: database system operated by NILU used to archive ground-based in-situ observations. Originally designed as database for EMEP observations, it serves as a database system for various observational frameworks today. Direct user access to the EBAS database is maintained through a web application (http://ebas.nilu.no).
- HTAP-Obs: HTAP Observations Database. A dataset of ground-based in-situ measurements for evaluation of chemical transport models in the assessment of hemispheric transport of air pollution. The compilation of this database was one of the main objectives of this project. This database is technically implemented using the EBAS database system.
- HTAP-EBAS refers to the project behind the HTAP-Obs – of which the current document is the final report.

## 1.4 Major achievements

The major outcome of the project is the compilation of the HTAP-Obs database and the availability of these data through the EBAS web interface and from an ftp-server. Starting from the existing EBAS database (including data from projects and programmes such as CREATE, EMEP, HELCOM, EUSAAR and GAW-WDCA), several relevant data collections (IMPROVE, NatChem, EANET) were added in order to create a homogenous database of observation data covering the northern hemisphere.

Three major steps were necessary to achieve this:

1) Negotiations with the contributing measurement networks and development of a data policy that could be accepted by all networks.
2) Research about the different data standards used by the networks (understanding the data standards require sometimes the understanding of different field and laboratory routines as well).
3) Converting and ingesting the new data into the EBAS database.

Due to the fact that EBAS has been used as primary database for several European measurement networks, the coverage for Europe was already very good before this project started. The coverage outside Europe was very sparse, and the efforts for extension of the database therefore concentrated on the northern American and Asian regions (see Figures 1-6).

For a more detailed description of the database extension please refer to chapter 2.2.3.1.

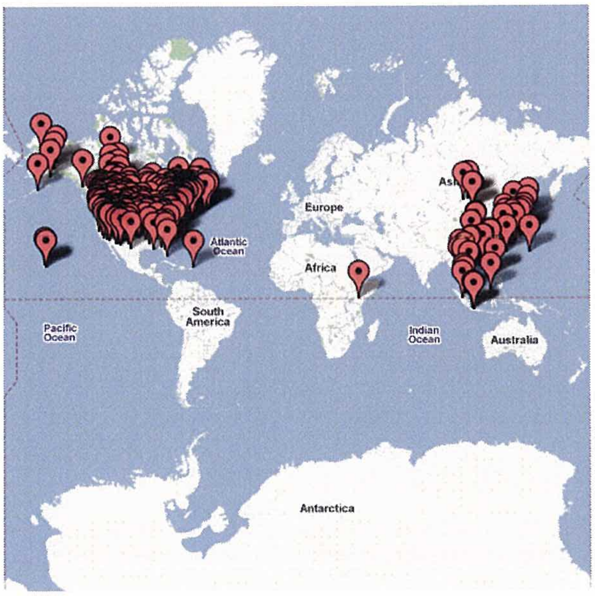



***Figure 1:*** *Distribution of measurement locations imported for HTAP-Obs*


***Figure 2:*** *Distribution of all measurement locations available for HTAP users*

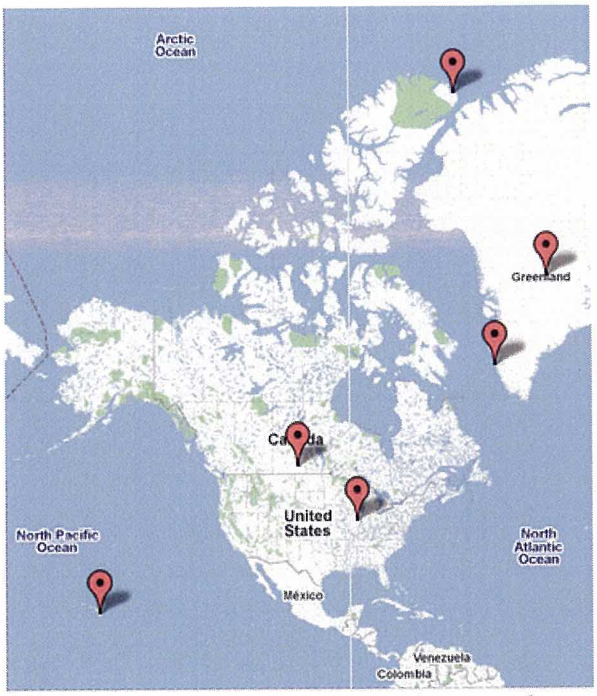



*Figure 3: EBAS coverage for northern America prior to HTAP-Obs*


*Figure 4: EBAS coverage for northern America after implementation of HTAP-Obs*


*Figure 5: EBAS coverage for south-eastern Asia prior to HTAP-Obs*


*Figure 6: EBAS coverage for south-eastern Asia after implementation of HTAP-Obs*

# 2 Efforts and results from the work packages

The current chapter follows the structure of the final work plan which describes the activities to be undertaken as part of the contract. The current chapter describes how the different sub tasks and planned achievements have been approached and solved. Specific emphasis is given to goals that were either not achieved or tasks that were solved in a different way than planned.

## 2.1 Work package 1 (data standards)

Work package 1 (WP1) were mainly addressing data standards, and a review of the data and meta data standards used under various monitoring frameworks were to be undertaken through WP1.1. The overall goal was to get an overview of the most commonly used data formats, meta-data standards and quality specifications. WP1.2 addressed how a draft set of standards for data sharing could be established. This formed the basis for the data flow and the harmonised observational dataset to be established.

### 2.1.1 Work package 1.1 (review of standards)

The efforts in WP1.1 and WP2.1 were closely harmonized already from the beginning in order to ensure that the review of standards included the datasets of interest for the users. NILU and LSCE participated in a number of meetings and workshops where the potential sources of data were discussed (see additional comments under WP2.2 below). Major focus was, as originally planned, on the various regional monitoring programmes. The standards used by the monitoring frameworks were reviewed using web-search as well as direct interaction with contact persons.

Together with the conclusions from WP2.1, the following networks were selected for further analysis: CAPMoN, EANET, IMPROVE, NADP, CASTNET, DEBITS and CAAD.

While analysis the metadata and reporting standards of the different monitoring networks is was early seen that networks had good overviews of the applied meta-data and data transfer standards internally, but they often did not have routines to accommodate batch mode export of data with sufficient metadata. Several networks had only simple meta-data exports (i.e. only providing units, time indication and site name, while no information about instrument techniques etc. were available without consulting off-line written documentation).

Based on these findings it was decided that metadata had to be added from external data sources in parallel with the import into EBAS. The meta-data was then to be provided from information made available through the original data repositories. All regional monitoring programmes were approached individually to query their preferred way of making data available.

There was also at an early stage a meeting between NILU and LSCE (at NILU during summer 2008) to discuss how export from the final HTAP-Obs could be made compatible with the CF-convention[1]. Work started on creating look-up tables for converting names and units from the EBAS standard into the CF-standard, but this work was never completed.

The CF standard originates from the atmospheric modelling community and is strongly related to formatting of model data in the netCDF format[1]. Much of the CF-standard deals with rules and guidelines for how multidimensional model data should be structured in this binary format. An extensive list of component/parameter names have been developed, but it is clear that this list is mainly created for model data and not for observational data. The CF standard is suitable for identifying the origin of chemical constituents for model calculations, e.g. "$NO_x$ originating from lighting" or "CO originating from biomass burning". Monitoring stations on the other hand generally measure the overall quantity of a chemical or physical property (e.g. concentration of $NO_x$ or CO) and cannot directly determine the origin. The CF convention is not well suited to accommodate essential information for observations such as measurement techniques, observational site classification, observational flags, statistical outliers, etc. After initial attempts, it was decided not to pursue the work on describing observational metadata through the CF convention any further.

It is worth noticing the recent developments of an international standard for formatting of observational data for the purpose of satellite validation. This effort is built upon previous developments done by NILU through their Envisat Validation Data Centre and by NASA through the Aura Validation Data Centre. Through the ESA-led initiative GECA (Generic Environment for Calibration and validation Analysis), previously developed standards (by NILU and NASA) have been brought together, harmonized and further developed into a new standard named GEOMS (Generic Earth Observation Metadata Standard). More information on the standard is available at: http://avdc.gsfc.nasa.gov/index.php?site=1178067684. The standard is currently released in version 0.7 (18th of November 2010) and only final comments from authors with minor contributions are pending before version 1.0 will be released. The GEOMS standard is purely developed for observational data and may be useful for exchanging data also for EBAS in the future – especially for the purpose of satellite validation. NILU (including members of the current project) has been actively involved in the development of GEOMS and is co-author of the documentation soon to be published.

Another important standard or standardization work rather, is the ongoing processes associated to the EU INSPIRE[2] directive. The directive specifies how geospatial data are to be made available within EU and externally. Important prototyping work (e.g. the SEIS-CAFE project led by JRC) has been undertaken in recent years in order to specify how observational data may be reported from countries and new data models and formats have been proposed. Definition of metadata standards related to Air Quality is currently being worked out (during 2010 and 2011) and it is expected that these standards will influence both the input to and the output from EBAS in the near future.

ANNEX I presents the data formats according to the preferred ways of export from the contributing monitoring networks. It also lists how data were received (url/ftp/CD), file formats, naming of parameters, conversion of units, flags etc.

---

1 Conventions for Climate and Forecast (CF): Metadata conventions to promote the exchange of NetCDF files within the climate modeling community [http://cf-pcmdi.llnl.gov/]

2 Infrastructure for Spatial Information in Europe [http://inspire.jrc.ec.europa.eu/]

### 2.1.2 Work package 1.2 (establishment of standards).

#### 2.1.2.1 Input data standard

As introduced above, import of data originating in different format into EBAS was chosen as a method to bring the datasets into a common standard. The contributing networks had different preferences on how this should be done. In some cases NILU actually extracted the datasets directly from public web services (US data using the Views interface), others made their files available at an FTP-site (Canada) and some provided their data in excel-files on a CD (EANET). For this reason, it was decided to focus on making generic routines for converting data to EBAS compatible files for import (NASA-Ames). Due to the very large number of sites and parameters, this effort became challenging and more time consuming than anticipated. The details about developed data conversion utilities are described in ANNEX I. These utilities are adapted to the data exchange formats and metadata content of the individual monitoring networks. While the original concept of this work package was to deliver a set of formats and methods for data exchange, the utilities themselves represent the deliverable on data standards. These software tools developed will be delivered on CD-ROM together with this report.

#### 2.1.2.2 Output data standard

To enable semi-automatic evaluation of models it was necessary to support a standardised format for export of data from EBAS. The format had to be strictly implemented in order to allow modellers to develop reading routines for automatic ingestion of data into their own systems. Furthermore, the format needed to contain sufficient amounts of metadata to describe the context in which the measurement was performed (provided that this was available in the database).

Through discussions with potential users of the HTAP database and through the experiences gained with delivering the first core HTAP dataset to the LSCE (a subset of the EBAS database in the form of custom formatted ASCII files), it became apparent that an ASCII output format of EBAS would work fine for the purpose of ingesting larger subsets of the database into e.g. model evaluation tools of modellers. The export in form of ASCII files required accompanying metadata tables that had to be provided in addition to the data files. In addition to the custom formatted ASCII files, it was also necessary to support data export in a self-descriptive file format such as NASA-Ames.

#### 2.1.2.3 Data policy

In order to clearly state the circumstances under which the measurement networks provide their data and the users access the data, common data use protocol had to be developed. The challenge was to harmonize the data policy needs of all contributing networks.

This effort was significantly more challenging than what was originally anticipated and the final version of the data policy document was therefore delivered later than planned. After comments and discussions with the individual networks, all regional monitoring programmes finally present in the database approved the data usage policy. The data policy document is presented in ANNEX II to this report.

## 2.2 Work package 2 (database implementation and population)

Work package 2 focused on developing an actual database and to populate it with observations. It was divided into three sub-WPs; WP2.1 addressed user needs and observations to be included, WP2.2 dealt with the actual development of a data centre and WP2.3 focused on its population.

### 2.2.1 Work Package 2.1 (user needs).

There were close interaction with data users and the concepts of the HTAP database as well as the progress of work were presented at all meetings held under HTAP. It was also presented at several other relevant meeting including:

- HTAP modelling workshop (Jülich Oct 07)
- HTAP TF-meeting (Rome May 08)
- HTAP modelling workshop (Washington June 08)
- IITAP meeting (IIanoi, Oct 08)
- EANET meeting (Hanoi, Oct 08)
- WMO Expert group on World Data Centres (St. Petersburg, Oct 08)

Further, some specific items were discussed with Martin Schultz and Frank Dentener by e-mail. JRC-Ispra, FZ- Jülich and LSCE had from an early stage access to the datasets both through "batch mode exports" as well as through the password protected web-interface.

#### 2.2.1.1 Selection of datasets

Selection of Sites

In order to assess the impact on atmospheric composition from hemispheric transport, it was necessary to focus on rural background and/or global type stations with little local influence on air pollution levels. The regional monitoring networks all include an exhaustive list of stations providing high quality observations – in general, observations from more stations than what was needed for the TFHTAP model evaluation effort. The networks were asked if they could recommend some specific stations of areas of sites that were of particularly high quality (e.g. stations with little and known local influence from pollution sources), but they were generally reluctant to providing such screening of their data. All stations were of equal importance to them and it was therefore chosen to include all stations from all contributing networks in the database.

Selection of parameters

This issue was discussed in some detail at the kick-off meeting on 21st of February 2008 and it was stated that the main parameters to be delivered included those of relevance to acidification, eutrophication, photo-oxidants and particulate matter (air and precipitation data). These parameters were provided by the regional monitoring networks. In addition to the mentioned focus, it was expressed a clear wish from the modelling community to also include measurements of CO, either at the surface or in the free troposphere.

#### 2.2.1.2 Evaluation of the database

The database in its final form was probably most intensely tested by the LSCE and the evaluation of the database is mainly based on the experience of this remote-from-NILU user.

After consultation between the LSCE and NILU it was decided to propose three ways to access the database, a) the EBAS web-interface; b) a bulk custom format ASCII extract fetched by LSCE through ftp and c) a bulk NASA-AMES extract accessible by ftp. These database access functionalities and extracts were tested by LSCE throughout the year 2009. An overview of experiences is presented in the following including recommendations for future development with respect to a) data content and completeness, b) data quality and c) the data accessibility and format. Note that the following remarks summarize a subjective user experience.

## Background and Data Usage Procedure

The specific motivation for LSCE as a user of the HTAP database was to compare global models to a large number of aerosol, deposition and ozone data and to use these to achieve a fast automatic benchmarking of the HTAP and Aerocom models via the Aerocom analysis and visualization tools. The Aerocom tools are built on IDL and are basically a set of modular subroutines which: (a) read in model results of different netCDF formats (AeroCom, HTAP, CMIP5 and generic LMDZ-INCA), (b) read in different observational datasets, such as satellite data, Aeronet data, LIDAR data and surface site data such as the EBAS-HTAP extracts and (c) merge the model and observation data so that temporal and spatial structures are matched and can be compared in a coherent way. Finally, the merged data structures allow for statistical analysis and visualization of the comparison as fields, profiles, time series, scatter, histograms and scores.

The procedure developed in the frame of this study was to read in EBAS-HTAP data into the AeroCom tools without reformatting of the observations. This way, new and updated data could easily be used, once available. The extraction was based on the available key fields of the EBAS database, the matrices and components. The following extract shows the formalized details of the request made to NILU to obtain an appropriate database large extract:

| **Projects:** CREATE, EMEP, EUSAAR, NILU, WMO-PCSAG, HTAP | |
|---|---|
| **Matrix** | **Components** |
| aerosol | sulphate corrected, sulphate total, sodium, magnesium, potassium, calcium, nitrate, lead, nickel, iron, phosphorus, ammonium, cloride<br>aerosol light scattering coefficient<br>organic carbon corrected, total carbon, total carbon corrected, organic mass |
| pm10 | sulphate corrected, sulphate total, sodium, magnesium, potassium, calcium, nitrate, lead, nickel, iron, phosphorus, ammonium, chloride<br>pm10 mass<br>organic carbon corrected, total carbon, total carbon corrected, organic mass |

| | |
|---|---|
| pm25 | sulphate corrected, sulphate total, sodium, magnesium, potassium, calcium, nitrate, lead, nickel, iron, phosphorus, ammonium, chloride<br>pm 2.5 mass<br>organic carbon corrected, total carbon, total carbon corrected, organic mass |
| pm10+pm25 | sulphate corrected, sulphate total, sodium, magnesium, potassium, calcium, nitrate, lead, nickel, iron, phosphorus, ammonium, chloride<br>pm10+pm2.5 mass<br>organic carbon corrected, total carbon, total carbon corrected, organic mass |
| air | sulphur dioxide, ozone, nitric acid, ammonia, hydrochloric acid |
| air+aerosol | sum nitric acid and nitrate |
| precipitation | sulphate corrected, sulphate total, sodium, magnesium, potassium, calcium, nitrate, lead, nickel, iron, phosphorus, ammonium, chloride<br>pH, precipitation amount |

***Table 1:*** *Extraction criteria for defining the set of data to be extracted for the bulk data export for HTAP users*

The extraction resulted in an ensemble of individual files which separated data from different stations, parameters, instruments, periods and matrices. This extraction was done on several occasions in 2009 to update the data and resolve inconsistencies found. For the period 1980-2007 the extraction resulted in ca 14000 files with 2.7 Gbyte data.

The AeroCom tools were selecting and eventually combining the parameters into an ensemble of observational data. Time series were constructed at all sites with valid data of daily and/or monthly frequency for the time period under consideration. Investigation of wet deposition required for instance the combination of precipitation and rain concentrations (e.g. wet deposition of sulphate required combining the corresponding instruments at a given station for precipitation amount and sulphate concentration in rain). If necessary, precipitation had to be obtained from a different instrument than rain concentration. Any change in station instrumentation, change in measurement frequencies or data gaps required specific algorithms in order to reconstruct unique time series at a given site.

Observational data were obtained at very different frequencies from hourly to weekly. In a first step, daily data were thus constructed from all datasets. If the measurement frequency was lower than daily, the daily re-sampling was done to facilitate matching of model and observations. This was especially useful for cases when e.g. a particular sampling went on from one month into another month or when data gaps existed in a series of weekly sampling. From these daily time series monthly model and data averages were constructed at all measurements sites (excluding from the daily model times series those days where no observations existed). All statistics and comparisons were based on monthly time series from model and data. It should be noted that months with fewer observations (some months have lesser number of days) or incomplete coverage, had equal weight when an annual average was created based on the monthly means. An annual average constructed from daily data would therefore be different. Although there are small differences between the annual averages based on days and months, the aggregation first to monthly values has important advantages; important seasonal anomalies and overall variability can be documented from the same dataset, even if some seasons have low data coverage. The monthly dataset is less subject to daily extremes and is more suited to characterize a global model of aerosol transport in the first place, when the overall quality of this model is in question.

Model-data comparison documentation is constantly updated, and includes comparisons to new model versions. These are accessible via the HTAP/AeroCom surfobs web interface (*http://nansen.ipsl.jussieu.fr/cgi-bin/AEROCOM/aerocom/surfobs_annualrs.pl?MODELLIST=HTAP*)

## Data Content and Completeness

The extension of EBAS to include HTAP data resulted in the availability of a larger than ever dataset from one database covering N-America, Europe and East Asia targeted at multi-parameter and multi-station comparisons to model results.

It should be noted that the data amounts available for the HTAP work is not fully identified as the HTAP database. E.g. the extraction of data for LSCE builds on obtaining data from additional (and publicly available) projects. If only the data added through the HTAP project would have been used, the dataset would have been considerably smaller.

## Data Quality

The data extracted for LSCE were fully readable by the Aerocom tools and only minor issues on units and formats needed to be resolved in subsequent extractions from the HTAP database. Some problems were nevertheless encountered when developing the analysis algorithms. E.g. a closer inspection revealed that longer time series (1990-2005) of sulphate deposition were not consistent over time for all sites. Sulphur units suffered from confusion between S and $SO_4$ masses and flags for identifying dry days in the precipitation networks were not harmonized. The frequency of measurements varied largely. Precipitation originated, in some networks, from different samplers then those used for establishing rain concentrations. Parallel sampling by different samplers during campaigns or when replacing an old instrument generation created ambiguities for deciding on which data to be used at a given site. It thus became clear that the use of large data extracts required more work on the data analysis side to recommend a set of consolidated data for benchmark testing of models. Already the algorithms used in the AeroCom tools to work-up the HTAP-Obs large data extract became too complicated to be documented in all detail here. This may not be a problem if different models are compared within the same analysis framework. However, data usage will depend on the interpretation of the data base, especially when combining data from different networks and instruments.

## Data accessibility and format

Three data access pathways were realized and the evaluation is presented here:

### *1. Via the web interface ebas.nilu.no*

The web interface was used in this project mainly to check the consistency and understand the large data extracts that was processed with the Aerocom tool. It was appreciated that the web interface allowed for a flexible check on the data availability, coverage in space and time, on data origins, and on the units used for a given parameter. The plotting facility allowed inspecting the variability of the parameter.

However, the evaluation by LSCE of the web interface lead to some suggestions to be implemented and a new version of the EBAS-web interface was developed during spring and summer 2010. It was released October 22nd 2010 and implemented the following suggestions from LSCE:

- Improvement in performance speed to help the user find the data wanted
- Search filters are not to be reset when going back after viewing datasets and plots, filters applied when sub-selecting data are more visible in the menus, so that the choices done are better understood
- Long names of parameters are now fully showing up in the menus
- Display of detailed metadata linked from datasets/measurements
- Graphical plots for non-continuous data (wet deposition, grab samples etc.)

The following recommendations from LSCE have not yet been implemented:

- Addition of a frequently asked section would help users find documentation
- Facilitation of repeated large extractions by storing user specific search profiles and an ftp download facility where datasets are prepared for downloads
- Grouping of components (e.g. group POPs, group VOCs, group Aerosol physical properties etc.) to avoid being diverted by too much speciation
- Possibility to plot and download aggregates (monthly, annual means etc.)
- Download option for metadata lists (station list, component list, flag list)

2. *Via a bulk data request to NILU, so that multiple data are selected based on user defined criteria. The data are put out in a simplified ASCII format.*

The procedure has been described above and is found to satisfy the needs for a large data extract. In general LSCE could receive a new data extract within 1-2 weeks after formulating the request. The access to the database was done by experts at NILU and earlier extraction procedures were reused.

3. *Via a bulk data request to NILU, as under 2), but files are provided in Nasa-Ames format.*

This procedure was developed in the extension period of the project and it allowed for export of data in multicolumn format (e.g. export of all three wavelengths + associated metadata from a nephelometer measurement into a single NASA-Ames format). Since these procedures were developed late in the project, they were not extensively tested by the modelling groups.

The evaluation efforts by LSCE have (as already described) discovered some minor inconsistencies in the existing datasets, which had to be corrected. The effort has thus led a general improvement of the data quality.

### 2.2.2 Work package 2.2 (development needs)

#### 2.2.2.1 Data access

While the EBAS database has been fully operational since the mid-1990ies, the web-interface for external users was at an early stage of development at the time of the project initiation. The search and extract functionalities were only partly developed which allowed the finalization of system requirements to be done jointly with the scoping work of the HTAP database. Establishing a system and routines to preserve data ownership (at Principal Investigator (PI, e.g. the owner of the data from one station) and network level) was achieved (partly) because of this and work was also initiated on an automatic system for management of protocol signature by users and granting of corresponding data access. This system has, however, not yet been finalized.

The web-interface is mainly to be seen as a "search tool" for identifying available datasets, and to review/plot and extract a limited number of datasets. Modelling groups will thus receive larger extracts of data in defined formats (see "Export routines" below).

#### 2.2.2.2 Export routines

As described in WP1.2, batch mode export routines were developed to deliver a full extract of the observation data in two different data formats (NILU NASA-Ames and a simple ASCII format with accompanying metadata files).

#### 2.2.2.3 Input routines

As described above, the data ingestion was highly dependent on conversion from various data formats. These data formats varied in many ways: File formats, metadata standards and data models. Different data models appeared for example in measurement periods, and where some data formats provided start and end time for each sample, others provided just the sample day as time reference.

The general approach was to develop one converter for each file format, which was to be kept generic and mainly deal with the syntactical layer of the data format (i.e. the file format). The different metadata- and data-models were implemented in data format specific configuration settings.

NARSTO

The approach described above proved useful for the ingestion of CAPMoN precipitation and filter-pack measurements. Both were provided by CAPMoN in the NARSTO file format, but different data exchange standards (DES) were used - both according to NAtChem specifications.

When NatChem delivered updated data in 2009, which included previously unreported data from the NatChem regional networks (ABPM, NFPM, NBPN, REPQ) in addition to revised data for CAPMoN, the same converter could be used to reprocess all the import data.

The developed conversion utility used the same program code for both input formats, but depending on the DES and the reporting organization stated in the file header, a different configuration and thus the applicable metadata model were used. There were no hardcoded dependencies on the DES or organization within the converter's program code.

VIEWS

The IMPROVE data used for ingestion originated from the network's VIEWS database system which allowed download through a web portal. The download format could be customized in many ways and a specific setting of output options had to be specified. The "custom designed" output file format that was used for the data export is further called the VIEWS file format.

The software design of the converters is similar to the NARSTO converter.

Manual conversion

For some data formats, the development of automatic conversion routines did not seem appropriate.

EANET: The delivered data were in human readable csv format and in MS-Excel files and the format definitions were not strictly implemented by the network. Automatic processing would therefore have been required a high percentage of manual interventions during conversion. It was concluded that manual conversion was far more efficient is this case.

EMPA: The original data format was according to WDCGG file format. The data format was strictly defined and would be a good candidate for automatic conversion. The development of an automatic converter would, however, have been far too time consuming since only 12 datasets were to be inserted in EBAS.

### 2.2.3 Work package 2.3 (population)

#### 2.2.3.1 Overview of the database

Data from following Monitoring frameworks have been ingested into the Database: IMPROVE, NATChem, EANET and EMPA.

NATChem provided data from CAPMoN and 4 regional networks (ABPM, NFPM, NBPN, REPQ).

EMPA provided hourly monitoring data for ozone and carbon monoxide from Mt. Kenya (Kenya) and ozone from Kototabang (Indonesia). Auxiliary meteorological data were imported as well.

The total number of datasets imported to EBAS is 3793:

| Dataset Type | Region | Years | Time Res. | # Sites | # Dataset |
|---|---|---|---|---|---|
| IMPROVE filter | US[3] | 2000-2006 | every 3rd day / 24h or Wed+Sat / 24h | 187 | 2488 |
| NAtChem filter | CA[4] | 2000-2007 | daily / 24h | 15 | 165 |
| NAtChem prec. chem. | CA | 2000-2007 | daily / 24h [5] | 40 (83) | 473[6] (935) |
| EANET filter | SE Asia | 2001-2005 | weekly | 30 | 358 |
| EANET precip. chem.[7] | SE Asia | 2001-2005 | daily or weekly | 51 | 626 |
| EMPA | Mt. Kenya, Kototabang | 2002-2006<br>1996-2007 | hourly | 2 | 12 |
| Total | | | | 288[8] | 3981[8] |

***Table 2:*** *Overview of number of datasets imported for HTAP-Obs*

EBAS is the primary database for the frameworks CREATE, EMEP, HELCOM, EUSAAR and GAW-WDCA. These data are also fully available for HTAP users. Furthermore EBAS holds a compilation of global precipitation chemistry monthly mean values for WMO-PC-SAG which is also available for HTAP users.

Overall, HTAP users have currently access to 31666 atasets from 933 stations.

---

[3] Includes one Canadian station (Egbert)

[4] Includes one US Station (Penn State)

[5] NBPN data have weekly resolution

[6] 462 datasets could not be converted due to data file problems – request for clarification/correction pending. For the time being, these datasets are contained in the database as monthly means

[7] Import currently in progress, counters to be updated, not included in the total numbers yet

[8] Due to overlaps, the total number of sites / datasets does not equal the according column sum

| Network | Parameters |
|---|---|
| IMPROVE | pm 10: pm10 mass<br>pm 2.5: pm 2.5 mass, ammonium, iron, lead, magnesium, nickel, nitrate, organic carbon, organic mass, phosphorus, potassium, sulphate total<br>pm 2.5- pm 10: pm2.5-pm10 mass (calculated) |
| NAtChem filter | aerosol: ammonium, calcium, chloride, magnesium, nitrate, potassium, sodium, sulphate total,<br>air: nitric acid, sulphur dioxide<br>air+aerosol: nitric acid + nitrate |
| NAtChem prec. chem. | precipitation amount<br>precipitation: ammonium, calcium, chloride magnesium, nitrate, pH, potassium, sodium, sulphate total, sulphate corrected |
| EANET filter | aerosol: ammonium, calcium, chloride, magnesium, potassium, sodium, sulphate total, nitrate<br>air: ammonia, hydrochloric acid, nitric acid, sulphur dioxide |
| EANET prec. chem. | precipitation_amount<br>precipitation: ammonium, calcium, chloride, conductivity, magnesium, nitrate, pH, potassium, sodium, sulphate total |
| EMPA | ozone (both stations), carbon monoxide (Mt. Kenya) as well as auxiliary meteorological data |

*Table 3: Components imported from different frameworks*

#### 2.2.3.2 Data conversion and import

NATChem

As described in WP 2.2 (chapter 2.2.2.3, section NARSTO), NatChem data have been converted using an automated conversion routine.

A number of problems were discovered during the first ingestion. Some of these were based on lack of information about laboratory routines and data conventions in NatChem, and could be clarified through interaction with the data provider. Others problems revealed data inconsistencies and could not readily be solved (such data were excluded in the data ingestion).

A second data revision was received in 2009 where most of the identified problems had been fixed and the temporal coverage had been extended to 2006. The spatial and contextual coverage was also widened by including the data from the NatChem regional networks (ABPM, NFPM, NBPN, REPQ) and the number of stations and datasets therefore increased.

This data revision were reprocessed with an improved version of the converter utility (considered feedback regarding previous questions to the data provider).

IMPROVE

After first consultations with network representatives it was decided to use the networks web access system (Views) to download the data. Creating a subset of sites, in order to reduce the number of datasets to the most representative sites, was not attempted since the IMPROVE representatives made clear, that all the sites are likewise representative depending on what the data user is interested in. Creating any subset of sites would rather create a bias. After this clear recommendation from the network it was agreed to include all the regional and background site's data, but to exclude urban sites.

As described in WP 2.2 (chapter 2.2.2.3, section VIEWS), IMPROVE data were converted using an automated conversion routine. There were no major problems detected and no reviewed import was necessary.

EANET

We received EANET data in non-standardized data formats. As explained in WP 2.2 (chapter 2.2.2.3, section Manual conversion), manual conversion of EANET data was chosen.

Filter pack measurements and precipitation chemistry were made available – last datasets were added in December 2010.

EMPA

A total of 12 datasets were converted manually (see also WP 2.2, chapter 2.2.2.3, section Manual conversion).

## 2.3 Work package 3 (recommendations for further work)

The HTAP-EBAS project has been a challenging task, both technical through the import of data in heterogeneous data formats, but also through interaction with the contributing networks working towards a common data policy. The resulting database and the final data policy document has, however, been a valuable contribution to the HTAP work and will likely also be used in the future to assess intercontinental transport issues. Due to the reusability of the import routines developed for this project, it is expected to involve minimal work to continue updating the database whenever new years of data become available from contributing networks.

The Call, as issued by the commission, specified linking of the database to other international data services and initiatives. This was toned down in the proposal from NILU and LSCE and did not become a specific topic in the final workplan of the project. This has, however, become an issue towards the end of the project. An exchange of metadata has been established between EBAS and the GEOmon distributed data centre (both operated by NILU) and the whole HTAP observational dataset is now therefore fully searchable at http://geomon.nilu.no together with observational and model data from a number of networks and databases. The data access and acceptance of the data policy is still handled by the EBAS system, but the possibility of searching through these (meta)data without having to provide a username/password is a good chance to promote wider use of the efforts laid down in the current project. Users with a registered username and who have signed the data protocol may, if they wish, also download the HTAP-Obs data (together with data from all the other databases) through the GEOmon portal.

Another key issue is the development of an interface to the American-led DataFed initiative. This brings together data-streams from a series of observational monitoring networks and model groups using a common Web Content Service (WCS). A dedicated meeting between Rudolf Hussar, Aasmund Fahre Vik and Paul Eckhardt was arranged at NILU to discuss this issue on 17th and 18th of June 2010. The discussions focused on practical aspects on how the EBAS data system could be made directly searchable and accessible though DataFed and an agreement was made that WCS-code from previous installations were to be made accessible to NILU for easy implementation. This code was sent to NILU on September 14th 2010, and a proof of concept installation was successfully demonstrated by October 27th. The installation is available at http://knulp.nilu.no:8080/NILU and provides access to publicly available EBAS data for two components. It is uncertain if the HTAP observational dataset can be made available to DataFed since there seem to be a lack of proper access control with the WCS standard. This needs to be further investigated though.

# 3 Conclusions

The HTAP-EBAS project has resulted in an updated database of observational data from regional background stations covering essential parts of the northern hemisphere. Especially areas susceptible for receiving pollution from intercontinental transport are covered. The work has proven the usefulness of the concept of ingesting heterogeneously formatted data into a common data model (EBAS). The work on exporting data from EBAS and the evaluation of the system by modellers, have proven that data export routines are currently sufficiently well developed. The HTAP observational dataset is currently well harmonized, both internally between the contributing networks and externally to the other datasets being handled by EBAS. The data will be kept updated and made available to users in the foreseeable future.

# 4 ANNEX I - Data conversion Issues

The following section provides detailed information about the conversion of data for import of different datasets from external networks to EBAS.

## 4.1 General

### 4.1.1 Station Codes

For converting external data formats to EBAS, the station code has to be translated to a unique EBAS station code. This procedure must first seek to find an existing station in the database (for the case that there are already existing data in the database), if no matching station can be found, a new station has to be created and the EBAS station code of this newly created station has to be used for conversion of the input data. The process of matching stations is not always straight forward and cannot be easily automated. Reasons for this are, that stations might have slightly different positions or station names in different organizations metadata. In order to avoid the multiple creation of stations, a semiautomatic approach has been chosen. A program takes the metadata information (station table) of the input data's network as input (simple csv format providing station name, network station code, lat, lon, alt), and performs the following checks for each station in the list:

If both, the station name and the position match exactly an existing station, the station is chosen automatically without user interaction

Otherwise, a list of best name matches (similar station names) and a list of nearest geographical matches is presented to a user. Following information is available: existing and new station name, existing EBAS station code, projected distance, triangular distance and altitude difference. Based on this information the user decides (in unclear cases after further investigations) if one of the existing stations should be used, or a new station should be created.

The output of this procedure (again a csv formatted file) can easily be used as station translation configuration for the converter program.

### 4.1.2 Instrument names

In order to create a unique instrument entity in EBAS, an instrument name was generated automatically. In default cases the instrument name "<instrument_type>_<EBAS_stationcode>" was used. In exceptional cases, an extension was necessary to create a unique name. These exceptions are described in the following network specific sections.

## 4.2 IMPROVE

### 4.2.1 Data source

http://vista.cira.colostate.edu/views/Web/Data/DataWizard.aspx

### 4.2.2 Data Format

The data format for download at the Views system is a highly customizable ASCII format. In order to achieve a reproducible and consistent data format we have defined a specific set of configurations for download:

- selected fields: Network, Site Code, Date, Parameter code, parameter occurrence code, data value, status flag, sampling duration
- output format:
    - format: ASCII Text
    - column format: ',' delimited
    - row format: standard (wide format),
    - content options: data & metadata
    - headers & title: display column header, display section titles
    - string quotes: double quotes
    - missing value: -999
    - date format: YYYY/MM/DD

### 4.2.3 Standard Flag Translation

| IMPROVE | | EBAS | |
|---|---|---|---|
| **flag** | **description** | **flag** | **description** |
| V0 | Valid value | 000 | unflagged |
| V2 | Valid esimated value | 798 | Measurement missing (unspecified reason), data element contains estimated value. Considered valid. |
| V4 | Valid value despite failing to meet some QC or statistical criteria | 100 | Checked by data originator. Valid measurement |
| V5 | Valid value but qualified because of possible contamination | 559 | Unspecified contamination or local influence, but considered valid |
| V6 | Valid value but qualified due to non-standard sampling conditions | 659 | Unspecified sampling anomaly |
| V7 | Valid value set equal to the detection limit (DL) since the value was below the DL | 781 | Value below detection limit, data element contains detection limit |
| M1 | Missing value because no value is available | 999 | Missing measurement, unspecified reason |
| M2 | Missing value because invalidated by data originator | 456 | Invalidated by data originator |
| M3 | Missing value due to clogged filter | 699 | Mechanical problem, unspecified reason |
| I0 | Invalid value - unknown reason | 899 | Measurement undefined, unspecified reason |
| I1 | Invalid value - known reason | 899 | Measurement undefined, unspecified reason |
| I2 | Invalid value (-999), though sample-level flag seems valid (SEM) | 999 | Missing measurement, unspecified reason |
| NA | [no description, only appeared with missing values] | 999 | Missing measurement, unspecified reason |

### 4.2.4 Additional Flags

The duration field (no equivalent in EBAS) was used to generate an applicable EBAS flag:

duration < 75% of nominal sampling duration:
EBAS flag 653 (Sampling period shorter than normal, observed values reported)[9]

9 EBAS flag 653 is usually used in cases of considerable deviations (<75%), in order to preserve at least some of the source formats accuracy, the finer grained data completeness flag 394 (usually used for statistical aggregates) has been used to flag values with sampling durations < 90%

75% of nominal sampling duration<= duration < 90% of nominal sampling duration:
EBAS flag 394 (Data completeness less than 90%)Error: Reference source not found

duration > nominal sampling duration:
EBAS flag 654 (Sampling period longer than normal, observed values reported)

### 4.2.5 Parameters

| IMPROVE | | | EBAS | | | Unit conversion factor |
|---|---|---|---|---|---|---|
| **Code** | **Parameter Name** | **Unit** | **Component Name** | **Matrix** | **Unit** | |
| **Parameter Description** | | | | | | |
| MF | Mass, PM2.5 (Fine) | ug/m$^3$ | pm25_mass | pm25 | ug/m$^3$ | |
| Gravimetric fine mass | | | | | | |
| MT | Mass, PM10 (Total) | ug/m$^3$ | pm10_mass | pm10 | ug/m$^3$ | |
| Gravimetric mass < 10 um in diameter | | | | | | |
| CM_calcul | Mass, PM2.5 - PM10 (Coarse) | ug/m$^3$ | pm10_pm25_mass | pm10_pm25 | ug/m$^3$ | |
| Calculated coarse mass | | | | | | |
| NO3f | Nitrate (Fine) | ug/m$^3$ | nitrate | pm25 | ug N/m$^3$ | 0.22590114 |
| Mass of nitrate particles < 2.5 um in diameter | | | | | | |
| SO4f | Sulfate (Fine) | ug/m$^3$ | sulphate_total | pm25 | ug S/m$^3$ | 0.33380525 |
| Mass of sulfate particles < 2.5 um in diameter | | | | | | |
| NH4f | Ammonium Ion (Fine) | ug/m$^3$ | ammonium | pm25 | ug N/m$^3$ | 0.77648428 |
| Mass of ammonium particles < 2.5 um in diameter | | | | | | |
| SO2 | Sulfur Dioxide | ug/m$^3$ | sulphur_dioxide | air | ug S/m$^3$ | 0.50048385 |
| FEf | Iron (Fine) | ug/m$^3$ | iron | pm25 | ng/m$^3$ | 0.001 |
| Kf | Potassium (Fine) | ug/m$^3$ | potassium | pm25 | ug/m$^3$ | |
| Mass of potassium particles < 2.5 um in diameter | | | | | | |
| MGf | Magnesium (Fine) | ug/m$^3$ | magnesium | pm25 | ug/m$^3$ | |
| Mass of magnesium particles < 2.5 um in diameter | | | | | | |
| NIf | Nickel (Fine) | ug/m$^3$ | nickel | pm25 | ng/m$^3$ | 0.001 |
| Pf | Phosphorus (Fine) | ug/m$^3$ | phosphorus | pm25 | ng/m$^3$ | 0.001 |
| PBf | Lead (Fine) | ug/m$^3$ | lead | pm25 | ng/m$^3$ | 0.001 |
| OMCf | Carbon Mass (Fine) (organic) | ug/m$^3$ | organic_mass | pm25 | ug/m$^3$ | |
| 1.4 * OC | | | | | | |
| OCf | Carbon, Organic Total (Fine) | ug/m$^3$ | organic_carbon | pm25 | ug/m$^3$ | |

#### 4.2.6 Other Metadata Attributes

The parameter occurrence code (POC) in IMPROVE is used to distinguish between different series of the same parameters measured. This code has a legal range from 1 to 9. In order to transfer this information to EBAS, different instrument identities have been created for different POC values. The instrument name was modified with a "_<POC>" extension in cases of POC $\neq$ 1 (POC=1 is interpreted as standard instrument name, without any extension).

Station codes: The IMPROVE sites MALO1, MALO2 and YELL1, YELL2 have been merged to one station in EBAS respectively, and converted to different instrument identities instead. In these cases, the generated EBAS instrument name was extended with "_<IMPROVE_sitecode>" (e.g. "_MALO2")

### 4.3 NAtChem Precipitation Chemistry

#### 4.3.1 Data source

The data for 2000 to 2004 have been provided by CAPMoN for ftp download upon request by NILU. In 2009 we received a data update. In addition to revised CAPMoN data, the coverage has been widened to the whole NAtChem network. The temporal extent was extended to 2006.

#### 4.3.2 Data Format

The data format is a NARSTO ASCII format following NAtChem specifications (DES specification in the file's header is given as "NATCHEM PRECIP 2003/01/17 (1.01)"). The source files were composed of one file per year, each file containing data of all the networks sites.

#### 4.3.3 Standard Flag Translation

| NAtChem precip | | EBAS | |
|---|---|---|---|
| flag | description | flag | description |
| V0 | Valid Value | 000 | unflagged |
| V1 | Valid value is below detection limit. | 781 | Value below detection limit, data element contains detection limit |
| V2 | Valid estimated value | 798 | Measurement missing (unspecified reason), data element contains estimated value. Considered valid. |
| V3 | Valid but qualified because of possible contamination | 559 | Unspecified contamination or local influence, but considered valid |
| V4 | Valid but qualified due to non-standard sampling conditions | 659 | Unspecified sampling anomaly |
| M1 | Missing value because no value is available | 999 | Missing measurement, unspecified reason |
| M2 | Invalidated by Network or missing value. | 999 | Missing measurement, unspecified reason |

### 4.3.4 Parameters

| NAtChem | | | EBAS | | | Unit conversion factor |
|---|---|---|---|---|---|---|
| **Parameter Name** | **CAS** | **Unit** | **Component Name** | **Matrix** | **Unit** | |
| pH | None | pH units | pH | precip | pH units | |
| Nitrate | C14797-55-8 | mg/L | nitrate | precip | mg N/l | 0.22590114 |
| Ammonium ion (NH4) | C14798-03-9 | mg/L | ammonium | precip | mg N/l | 0.77648428 |
| Sulfate | C14808-79-8 | mg/L | sulphate_total | precip | mg S/l | 0.33380525 |
| Sulfate: non-sea salt | C14808-79-8 | mg/L | sulphate_corrected | precip | mg S/l | 0.33380525 |
| Chloride | C16887-00-6 | mg/L | cloride | precip | mg/l | |
| Sodium, ion (Na1+) | C17341-25-2 | mg/L | sodium | precip | mg/l | |
| Calcium, ion (Ca2+) | C14127-61-8 | mg/L | calcium | precip | mg/l | |
| Magnesium | C22537-22-0 | mg/L | magnesium | precip | mg/l | |
| Potassium, ion (K1+) | C24203-36-9 | mg/L | potassium | precip | mg/l | |
| Precipitation amount | None | mm | precipitation_amount_off | precip | mm | |

### 4.3.5 Other Metadata Attributes

Station codes: The CAPMoN sites CAPMCAON1EGB and CAPMCAON2EGB (same coordinates) have been merged to one station in EBAS and converted to different instrument identities instead. In these cases, the generated EBAS instrument name was extended with "_1" and "_2" respectively.

## 4.4 NAtChem Filter Measurements

### 4.4.1 Data source

The data for 2000 to 2004 have been provided by CAPMoN for ftp download upon request by NILU. In 2009 we received a data update. In addition to revised CAPMoN data, the coverage has been widened to the whole NAtChem network. The temporal extent was extended to 2006.

### 4.4.2 Data Format

The data format is a NARSTO ASCII format following NAtChem specifications (DES specification in the file's header is given as "NARSTO 2002/05/28 (2.301)", but in fact, a NAtChem specific DES has been used). The source files were composed of one file per year, each file containing data of all the networks stations.

### 4.4.3 Standard Flag Translation

| NAtChem filter | | EBAS | |
|---|---|---|---|
| **flag** | **description** | **flag** | **description** |
| V0 | Valid Value | 000 | unflagged |
| V2 | Valid estimated value | 798 | Measurement missing (unspecified reason), data element contains estimated value. Considered valid. |
| V6 | Non-conforming sampling period | 653 | Sampling period shorter than normal, observed values reported |
| | | 653 | Sampling period longer than normal, observed values reported |
| V7 | Valid value but set equal to the detection limit (DL) | 781 | Value below detection limit, data element contains detection limit |
| M1 | Missing value because no value is available | 999 | Missing measurement, unspecified reason |
| M2 | Missing value because invalidated by data originator | 999 | Missing measurement, unspecified reason |
| M3 | Invalidated by NAtChem/PM | 999 | Missing measurement, unspecified reason |

The NAtChem flag V6 does not provide any information whether the period was shorter or longer than usual. Therefore both EBAS flags (653 and 654) have been applied, in order to indicate that either of both is the case.

### 4.4.4 Parameters

| NAtChem | | | EBAS | | | Unit con-version factor |
|---|---|---|---|---|---|---|
| **Parameter Name** | **CAS** | **Unit** | **Component Name** | **Matrix** | **Unit** | |
| Nitrate | C14797-55-8 | ug/m$^3$ | nitrate | aerosol | ug N/m$^3$ | 0.2259011 |
| Nitric acid | C7697-37-2 | ug/m$^3$ | nitric_acid | air | ug N/m$^3$ | 0.222291 |
| Nitrate ion + Nitric acid | C14797-55-8+ C7697-37-2 | ug/m$^3$ | sum_nitric_acid_and_nitrate | air+ aerosol | ug N/m$^3$ | 0.2259011 |
| Ammonium ion (NH4) | C14798-03-9 | ug/m$^3$ | ammonium | aerosol | ug N/m$^3$ | 0.7764843 |
| Sulfur dioxide | C7446-09-5 | ug/m$^3$ | sulphur_dioxide | air | ug S/m$^3$ | 0.5004838 |
| Sulfate | C14808-79-8 | ug/m$^3$ | sulphate_total | aerosol | ug S/m$^3$ | 0.3338053 |
| Calcium, ion (Ca2+) | C14127-61-8 | ug/m$^3$ | calcium | aerosol | ug/m$^3$ | |
| Chloride | C16887-00-6 | ug/m$^3$ | cloride | aerosol | ug/m$^3$ | |
| Potassium, ion (K1+) | C24203-36-9 | ug/m$^3$ | potassium | aerosol | ug/m$^3$ | |
| Magnesium | C22537-22-0 | ug/m$^3$ | magnesium | aerosol | ug/m$^3$ | |
| Sodium, ion (Na1+) | C17341-25-2 | ug/m$^3$ | sodium | aerosol | ug/m$^3$ | |

## 4.5 EANET Precipitation Chemistry

### 4.5.1 Data source

The data for 2001 to 2005 have been provided by EANET as MS-Excel documents on Compact Disk.

### 4.5.2 Data Format

The data format is MS-Excel document, according to the EANET data reporting procedures. The documents are created by the data originators for each measurement site. The format human readable, and details divert slightly between files.

As a support utility, a spreadsheet document has been created. This support utility helps with routine conversion steps and warns about most common data inconsistencies. However, due to the differences in the datafiles between laboratories, the manual effort was still very high. Most common inconsistencies were incompatible value/flag combinations and different reporting of overflow samples.

### 4.5.3 Standard Flag Translation[10]

| EANET precip | | EBAS | |
|---|---|---|---|
| **flag** | **description** | **flag** | **description** |
| | Valid Value | 000 | unflagged |
| 477 | Inconsistency between measured and estimated conductivity. | 477 | Invalid due to inconsistency between measured and estimated conductivity |
| 599 | Contamination not specified. | 599 | Unspecified contamination or local influence |
| 699 | Mechanical problem, reason not specified. | 699 | Mechanical problem, unspecified reason |
| 701 | Less accurate than usual, reason not specified. | 701 | Less accurate than usual, unspecified reason. (historical flag) |
| 781 | Below detection limit. | 781 | Value below detection limit, data element contains detection limit |
| 782 | Low precipitation, value is obtained from diluted sample (EANET original flags) | 782 | Low precipitation, concentration estimated |
| 783 | Low precipitation, concentration unknown. | 783 | Low precipitation, concentration unknown |
| 899 | Measurement not defined, reason not specified. | 899 | Measurement undefined, unspecified reason |
| 999 | Missing measurement, reason not specified. | 999 | Missing measurement, unspecified reason |

[10] "Quality Assurance/Quality Control (QA/QC) Program for Wet Deposition Monitoring in East Asia", Acid Deposition Monotoring Network in East Asia, March 2002, Chapter 7, pp. 9ff

### 4.5.4 Parameters

| EANET | | EBAS | | | Unit conversion factor |
|---|---|---|---|---|---|
| Parameter Name | Unit | Component Name | Matrix | Unit | |
| pH | pH units | pH | precip | pH units | |
| EC | mS/m | conductivity | precip | uS/cm | 10 |
| $NO_3^-$ | umol/l | nitrate | precip | mg N/l | 0.0140067 |
| $NH_4^+$ | umol/l | ammonium | precip | mg N/l | 0.0140067 |
| $SO_4^{2-}$ | umol/l | sulphate_total | precip | mg S/l | 0.032065 |
| $Cl^-$ | umol/l | cloride | precip | mg/l | 0.035453 |
| $Na^+$ | umol/l | sodium | precip | mg/l | 0.02298977 |
| $Ca^{2+}$ | umol/l | calcium | precip | mg/l | 0.040078 |
| $Mg^{2+}$ | umol/l | magnesium | precip | mg/l | 0.024305 |
| $K^+$ | umol/l | potassium | precip | mg/l | 0.0390983 |
| Amount of precipitation | mm | precipitation_amount | precip | mm | |
| | | precipitation_amount_off | precip | mm | |

### 4.5.5 Other Metadata Attributes

The method for measuring precipitation amount is coded in the datafiles.

Information about the analytical methods for each parameter, year and laboratory have been derived from the annual reports[11,12,13,14,15].

Sufficient station metadata where available for download as pdf document[16]. Most stations could be mapped to existing stations in EBAS.

---

[11] "Data Report on the Acid Deposition in the East Asian Region 2001", Network Center for EANET, November 2002, Table 3.3., p. 10

[12] "Data Report on the Acid Deposition in the East Asian Region 2002", Network Center for EANET, November 2003, Table 3.3., p. 12

[13] "Data Report on the Acid Deposition in the East Asian Region 2003", Network Center for EANET, November 2004, Table 3.3., p. 12

[14] "Data Report on the Acid Deposition in the East Asian Region 2004", Network Center for EANET, September 2005, Table 3.3., p. 12

[15] "Data Report on the Acid Deposition in the East Asian Region 2005", Network Center for EANET, November 2006, Table 3.3., p. 12

[16] "EANET Site Information", November 2009, retrieved from http://www.eanet.cc/site/site_p/all.pdf

## 4.6 EANET Filter Measurements

### 4.6.1 Data source

The data for 2001 to 2005 have been provided by EANET as CSV files on Compact Disk.

### 4.6.2 Data Format

The data format is a CSV format according to the EANET data standards. The files are condensed to one datafile per year containing all measurements at all stations.

### 4.6.3 Standard Flag Translation

The original datafiles contained no flag information. All missing values have been translated to EBAS flag 999 (missing measurement, unspecified reason), all reported measurements have been flagged 000 (valid measurement).

### 4.6.4 Parameters

| EANET | | EBAS | | | Unit conversion factor |
|---|---|---|---|---|---|
| **Parameter Name** | **Unit** | **Component Name** | **Matrix** | **Unit** | |
| SO2 | ppb | sulphur_diaxide | air | ug/m3 | 1.33 |
| HNO3 | ppb | nitric_acid | air | ug N/m3 | 0.58 |
| HCl | ppb | hydrochloric_acid | air | ug Cl/m3 | 1.47 |
| NH3 | ppb | ammonia | air | ug N/m3 | 0.58 |
| SO4 | ug/m3 | sulphate_total | aerosol | ug S/m3 | 3 |
| NO3 | ug/m3 | nitrate | aerosol | ug N/m3 | 0.2257 |
| Cl | ug/m3 | chloride | aerosol | ug/m3 | |
| NH4 | ug/m3 | ammonium | aerosol | ug N/m3 | 0.7778 |
| Na | ug/m3 | sodium | aerosol | ug/m3 | |
| $K^+$ | ug/m3 | potassium | aerosol | ug/m3 | |
| Ca | ug/m3 | calcium | aerosol | ug/m3 | |
| Mg | ug/m3 | magnesium | aerosol | ug/m3 | |

## 4.7 EMPA CO and $O_3$ datasets

### 4.7.1 Data source

The data have been downloaded from the World Datacenter for Greenhouse Gasses (WDCGG) with permission from EMPA.

### 4.7.2 Data Format

The source data were in WDCGG data format[17]. Thus the data format was well defined and no interaction with the data originators for clarification was necessary.

17 "World Datacenter for Greenhouse Gases Data Submission and Dissemination Guide", GAW Report No. 174, June 2007, Annex 2

### 4.7.3 Standard Flag Translation

The input files provides no flag information. All missing values have been translated to EBAS flag 999 (missing measurement, unspecified reason), all reported measurements have been flagged 000 (valid measurement).

### 4.7.4 Parameters

| EMPA | | EBAS | | | Unit conversion factor |
|---|---|---|---|---|---|
| **Parameter Name** | **Unit** | **Component Name** | **Matrix** | **Unit** | |
| O3 | ppb | ozone | air | ug/m3 | 2 |
| CO | ppb | carbon_monoxide | air | ppb | |

# 1 ANNEX II - Data policy document

**Database for evaluation of Hemispheric Transport of Air Pollutants (database project name "HTAP")**

Version 1, valid from 26. February 2009, Author: Kjetil Tørseth

In support of the EMEP Task Force on Hemispheric Transport of Air Pollutants (www.htap.org), the European Commission has issued a service contract (contract 070307/2007/481644/MAR/C5) with the Norwegian Institute for Air Research. The objective of the contract is to integrate observational data in a common database to support the understanding of Hemispheric Transport of Air Pollutants.

This document outlines the data policy associated with the use of data made available through this effort (a general description of how data policy and access control is handled is described in the web-interface to the database (http://ebas.nilu.no)).

The main focus of the integration effort has been to import data from various international monitoring programmes together with sufficient meta-data to allow for the interpretation of model results. This has the advantage that users can download data having common standards for nomenclature, formats etc.

The main source of data is 'public domain' established under the different regional monitoring programs. These original data repositories are to be considered as the official data archives (primary archives). The datasets associated with this "HTAP" project are considered as secondary copies. This means that more recent updates may be available in the primary archives. The original project/program from where data are made available can be found as part of the meta-data description.

We request that users consider an acknowledgement or citation to the primary data base upon use of the data for publication purposes. The name of individual data providers is normally available in the meta-data within the data files. For scientific studies where substantial use of data sets is made, users are expected to consider an offer of co-authorship to the data originators. Please also consult the data policy for "public data" which can be found in the login section of http://ebas.nilu.no.

EANET requests data users to provide a copy of all papers and publications originating from use of their data. Contact details for correspondence with EANET network centre can be found at www.eanet.cc

Access to the HTAP database is granted upon signing this data policy according to the "restricted but simplified" procedure described at the web interface. The NILU data centre keeps a record of all data policy agreements made.

Data access is limited to those undersigning the data exchange protocol, and data must not be redistributed to any third parties.

I approve the data policy of the HTAP-project:

Full name/affiliation: ______________________________

E-mail: ____________________ Phone: ____________________

Date: ____________________ Sign: ____________________

**NILU** Norwegian Institute for Air Research

Norwegian Institute for Air Research
P.O. Box 100, N-2027 Kjeller, Norway
***Associated with CIENS and the Environmental Research Alliance of Norway***
***ISO certified according to NS-EN ISO 9001***

| | | |
|---|---|---|
| REPORT SERIES<br>TECHNICAL REPORT | REPORT NO. TR 1/2011 | ISBN: 978-82-425-2359-4 (print)<br>978-82-425-2360-0 (electronic)<br>ISSN: 0807-7185 |
| DATE 20/1-11. | SIGN. | NO. OF PAGES 35 / PRICE NOK 150.- |
| TITLE<br>HTAP-EBAS<br>No 070307/2007/481644/MAR/C5<br>Final Report, Version 1.0 | | PROJECT LEADER<br>Paul Eckhardt<br>NILU PROJECT NO.<br>O-1080114 |
| AUTHOR(S)<br>Kjetil Tørseth, Paul Eckhardt, Aasmund Fahre Vik, Michael Schulz | | CLASSIFICATION *<br>A<br>CONTRACT REF.<br>No 070307/2007/481644/MAR/C5 |

REPORT PREPARED FOR
European Commission
DG Environment

ABSTRACT
The Convention on Long-Range Transboundary Air Pollution has set up a Task Force on Hemispheric Transport of Air Pollution which aims at obtaining a fuller understanding of the role of hemispheric transport of air pollution and specifically to have tools for access to integrated observations and a data set for the evaluation of chemical transport models. A specific service contract (this project) was designed to provide data sets for evaluation of models and more generally create a global system of ground based observations of transboundary air pollution.
The objective of the contract was specifically to

- Define **a set of generic data standards** for observations of air pollution (air quality, deposition and vertical profiles and linked meta information) that is compatible with the requirements of evaluation of the performance of chemical transport models. This set of standards should build on existing efforts of naming conventions and be flexible to accommodate the main standards in use for air pollution information and to be adapted to modern information and communication data technologies.
- Provide a **core data set** of observations suitable for evaluation of chemical transport models in the assessment of hemispheric transport of air pollution. This core set should include the information on ozone and its precursors and Particulate Matter and its precursors. The core data set has to be relevant for the years 2000/01 and 2005/06 although also data for other years are of relevance and thereafter having the potential of being annually updated through automatic schemes.
- Provide **recommendations** for further work to fully streamline the access to observational data.

NORWEGIAN TITLE
HTAP-EBAS. No. 070307/2007/481644/MAR/C5. Sluttrapport, versjon 1.0

| KEYWORDS | | |
|---|---|---|
| TF-HTAP | Database | Observations |

ABSTRACT (in Norwegian)

| | | |
|---|---|---|
| ** Classification* | *A* | *Unclassified (can be ordered from NILU)* |
| | *B* | *Restricted distribution* |
| | *C* | *Classified (not to be distributed)* |

REFERENCE: O-1080114
DATE: JANUARY 2011
ISBN: 978-82-425-2359-4 (print)
978-82-425-2360-0 (electronic)

NILU is an independent, nonprofit institution established in 1969. Through its research NILU increases the understanding of climate change, of the composition of the atmosphere, of air quality and of hazardous substances. Based on its research, NILU markets integrated services and products within analyzing, monitoring and consulting. NILU is concerned with increasing public awareness about climate change and environmental pollution.

NILU
Norsk institutt for luftforskning
Norwegian Institute for Air Research