SMEAR database

Introduction

Processed data from continuous measurements is stored in MySQL database hosted by CSC. All data is located in database 'smear' which contains several tables from each station.

Web user interface in AVAA portal provides easy access and visualization of the data. The UI does not yet show all database tables, you must use other tools to retrieve data directly from the database, see instructions and metadata descriptions further below.

Search page allows visualization and downloading of shorter (time period of <6 months) datasets.

Download larger datasets via the download page.

AVAA portal also provides API for scripted queries. This is the recommended way to retrieve data if you want to set up automated visualization or processing of data or just download the data "on the fly" without saving intermediate files. See instructions on the API page and check some sample scripts to get started.

General questions on data availability, terms of use etc. contact atm-data@helsinki.fi. It's basically a mailing list with couple of "data-aware" people. More detailed list of contact persons

The old databases on wisp.atm.helsinki.fi are not in use any more.

Data from FMI weather stations in Värriö, Hyytiälä and Kumpula can be downloaded from FMI database.

Software configuration for direct database access and SQL queries

Database structure and metadata

All data is located in database 'smear' which contains several tables from each stations, e.g. HYY_META for Hyytiälä SMEAR II meteorological and gas data, SII1_EDDY for Siikaneva1 eddy fluxes.

An easy way to find out the names and locations of the variables in the database (that is, table and column names) is using AVAA:

1. Tooltips on the variable lists show COLUMN:[column name]:[table name] for each variable.
2. Output file headers identify the variables as [table name].[column name]

Table 'VariableMetadata' contains the column names ('variable'), locations ('tableID'), descriptions, units, source instruments etc. for (almost) all variables in the database. 'TableMetadata' is the key to table ID's in VariableMetadata, it also contains basic station information of the corresponding tables.

Download variable metadata in .csv format

Download table metadata in .csv format

Note that the attached files are UTF-8 encoded, MS Excel will not show the contents correct. Open for instance with Notepad++ using some Unicode font.

You can also access the metadata via the API.

If you have database connection open, you can take a look at metadata contents using MySQL command

```
SELECT * FROM smear.VariableMetadata;
```

Example of more focused MySQL queries for table and column names:

```
SELECT tableID, variable FROM smear.VariableMetadata WHERE title LIKE `Air temperatur%`;
```

Field 'variable' is the column name. Table names must be parsed from 'tableID' values, for example tableID 2:

```
SELECT name FROM smear.TableMetadata WHERE tableID=2;
```

Table 'Events' contains descriptions of changes in measurement and data processing. Events are linked to the actual data variables in 'variableEvents' table.

'Tags' are standard (e.g. NetCDF CF) variable names and linked with actual variables with 'variableTags' table.

More detailed description of metadata tables.

Data quality levels

Most variables in the database are calculated and inserted to the database near real time with only rudimentary automated quality check. Later they are updated with data processed and checked by the responsible researchers. The quality level is indicated as [variable name]_EMEP column. Level 1 refers to online processed data and 2 for quality checked data.

Miscellaneous

Basic time step in all tables except flux data (*_EDDY* tables) is one minute. One record can be instantaneous observation (most met & soil measurements) or average or accumulation over 1 min (precipitation, runoff) or 30 min (fluxes).

Not all variables are measured every minute, in such case there is empty value field (NULL) in the table.

You can apply basic math, for instance time-averaging, to the data within the database engine using SQL commands but in some cases you better first download 1-min data and do the math yourself:

– wind direction: vector mean requires fairly complex SQL scripting
– is it reasonable to calculate time average that contains just one 1 min observation and the rest is NULL?

Time step of one minute means that the amount of data in the database is huge! Don't try to select everything at once!

Data descriptions

Värriö

Hyytiälä

SMEAR II

Siikaneva 1 & 2

Lake Kuivajärvi

Helsinki

Kumpula

Hotel Torni

Erottaja Fire station