Waterbase – Water Quantity
Version 4

Quality control documentation

30 March 2007

Waterbase – Water Quantity

Data on water quantity are collected annually through the Eionet-Water process. Data and information obtained through the Eionet-Water process are primarily used to compile indicator factsheets, associated with the EEA's Core Set Indicators, upon which EEA assessment reports are based. Data collected through the Eionet-Water process are also published in Waterbase, a series of water topic-specific databases and web pages, publicly accessible via the EEA Data Service's web site.

Water Quantity dataset include physical characteristics of the water quantity stations, as well as the water quantity data.

QA/QC activities

This document briefly presents EEA activities focused on quality of Waterbase – Water Quantity dataset and results of these activities. In addition warning is given on the use of certain records for analytical purposes (see section 2).

Quality control tests have been performed on the Waterbase – Water Quantity database provided in December 2006 by ETC/WTR. This database is included in the EEA data service as version 4, and is publicly available. The database and metadata are available at the following URL:
http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=988

Waterbase – Water Quantity dataset contains two data tables:
- QUANTITY
- STATIONS

Two type of test have been performed on the data tables. Basic tests and Logical rules violation test.

1. Basic tests

1.1 Primary key tests

Primary key is a field or combination of fields with values which have to be unique in the data table. If primary key is duplicated it is an error.

List of data tables primary keys:
STATIONS: WaterbaseID
QUANTITY: WaterbaseID, Determinand, Year

Result:
No primary key error has been detected.

1.2 Table relations tests

Unique Waterbase identifier (WaterbaseID) is contained in each of the data tables. It can be used to link data from one table to another. Table relations tests detect identifiers which are not present in some of the tables.

1.2.1   Number of "QUANTITY" table records by country where WaterbaseID is not present in the "STATIONS" table

Country code	No. of records
TOTAL	0

1.2.2   Number of stations without any data in the "QUALITY" table by country

Country code	No. of stations	Percentage of total no. of stations
IE	1	0.28
TOTAL	1	0.04

1.2.3   Stations without any data in the "QUALITY" table

WaterbaseID	No. of stations
IE_QA_03056	1
TOTAL	1

2. Logical rule violation tests

Logical rules were tested in the "QUANTITY" data table. Certain logical relations can be defined between particular determinands contained in this table. Those relations can be mathematically transformed in a set of rules. Following rules have been detected and tested:

Rule	Basic validation rules
1	P >= Pmax
2	Q > Q95
3	Q < Qmax
4	Q95 < Qmax

Legend:
P   Mean annual precipitation at rain gauging stations
Pmax   Annual daily maximum precipitation
Q   Mean annual discharge at all gauging stations
Q95   Annual discharge exceeded 95% of the time at all gauging stations
Qmax   Annual maximum discharge at all gauging stations

A special QA field (QA_LRviolations) has been added to the data table. This field holds (as comma separated list) the numbers of the rules violated by the determinand in the record (see example).

Example:

WaterbaseID	Year	Determinand	Value	QA_LRviolations
XY_QA_1234	2005	Pmax	90	1
XY_QA_1234	2005	P	5	1
XY_QA_5678	2005	Q	50	2
XY_QA_5678	2005	Q95	1400	2, 4
XY_QA_5678	2005	Qmax	1350	4

It is recommended the records where QA_LRviolation field is not empty (2733 records), should not be used in further analysis. Detected data quality inconsistencies will be tried to be solved in the near future.