Water in Celtic Countries: Quantity, Quality and Climate Variability (Proceedings of the Fourth InterCeltic Colloquium on Hydrology and Management of Water Resources, Guimarães, Portugal, July 2005).
IAHS Publ. 310, 2007.

Quality assurance for hydrometric network data as a basis for integrated river basin management

FRANK SCHLAEGER[1], MICHAEL NATSCHKE[1] & DANIEL WITHAM[2]

1 *Kisters AG, Charlottenburger Allee 5, 52068 Aachen, Germany*
frank.schlaeger@kisters.de
2 *Kisters AG, 28 Denton Road, Ilkley West Yorkshire LS29 0AA, UK*

Abstract Hydrological network data is increasingly demanded by hydrologists, engineers, decision makers and the general public. This means for hydrologists, that they have to process the raw data into quality controlled information in an efficient way. The hydrological information system WISKI offers a broad range of standardized quality assurance procedures to fulfil these requirements.
Key words data quality assurance; hydrometric network; information system

INTRODUCTION

The demand for information about different measured hydrometric parameters such as precipitation, water level or flow is increasing in quantity and quality significantly. This is regarding their use in hydrological modelling, water management techniques and public information. Today the level of information technology available represents an ideal base for hydrological information systems optimized for acquisition, data evaluation and dissemination of verified hydrological and ecological data sets. Consequently, the demand for tools for quality assurance during the whole process is on the increase.

For the successful quality assurance of hydrometric network data, work at different levels such as hydrometric standards, workflows, software and data security is needed. In addition to the data value itself, the user must be supplied with information about the data quality and the processing history of data values.

Most of these requirements can be complied by a hydrometric information system set up in a relational database environment. Typical elements such as time series plots, tables, reports, GIS functionality should be available as a user interface. For high acceptance, Windows-compatibility and office integration is desirable. All data quality information should always be available.

The WISKI software package is a modern tool for hydrological data management. WISKI is a Windows-based client/server system based on relational databases, designed in close cooperation with water agencies, authorities, engineers and hydrologists (Fig. 1). It combines modern standards of data management with advanced tools to collect, edit, store and present time series data to intranet. Internet and GIS users. The system allows the automatic and effortless flow of time series data from the measuring site into the database. This data is then reviewed using a powerful graphical or tabular graphical user interface.

Furthermore, a detailed quality flag sub-system has been integrated to present the data, together with its quality. This quality flag system deals with the data quality on two levels. A generic primary quality flag summarizes the data quality of data sets with standard flags, like good, suspect or missing. A more detailed secondary flag gives additional specialized information to the user. These secondary flags inform the user about the editing process, abnormalities during calculation or modelling. It is configurable for the different processing methods.

During the aggregation of instantaneous data to daily, monthly or annual data, not only are the values aggregated, but the quality flags are also aggregated. Consequently, for every aggregated value, information is available about the quality of the input data.

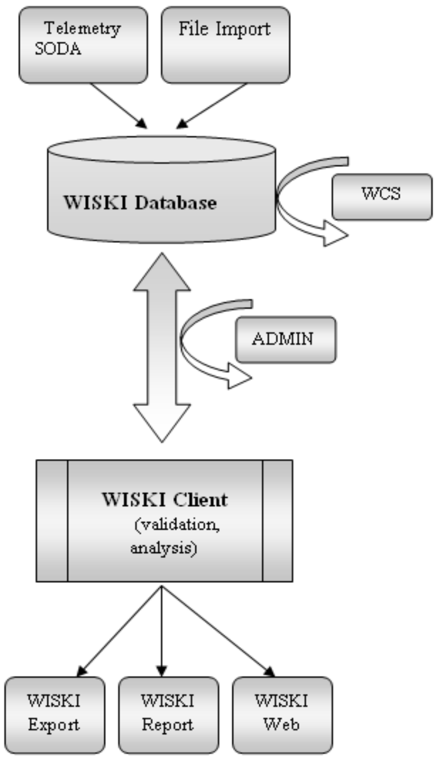

Fig. 1 Schema of the WISKI System.

THE SOFTWARE PACKAGE

Data acquisition and integration

In order to support a more integrated approach to water management, a modern hydrological information system has to acquire and store all types of data from a wide range of parameters. Not only the commonly used parameters such as water stage and discharge, but also, for example, climate parameters, to allow analysis within the same system.

Typical data input sources are remote data acquisition from field data loggers, import of third party data via input files in different formats, read out of field devices, digitization from graphical charts, or manual input using the interactive time series editor.

The import process should be automated as much as possible to free staff resources from manually entering data. The WISKI data acquisition server allows the automatic collection of data from field data loggers. The Import Server can scan directories and import data files, which are stored directly into the database. When the user digitizes graphical charts, the digitized data is imported automatically afterwards, without user intervention.

Data storage

Due to the complexity of data modelling, WISKI stores measured data in relational database systems (RDBMS). As historical data plays a fundamental role in hydrology, one of the greatest challenges is to achieve a good performance when dealing with large amounts of measured data; this is a key objective of the developers of WISKI. Another fundamental aspect is to allow multiple users to work on the same system simultaneously. They can benefit from working on the same data, but also preventing users from editing the same data at the same time. The data model of WISKI was developed to deal with large amounts of measured data. Therefore it has all the necessary functionality to manage time series on such a level. Another important topic is time series classification. One of the major differences found is the distinction between equidistant data (e.g. 15 min values) and non-equidistant data (e.g. rainfall event data).

Data processing

The data stored in the WISKI database can be accessed through the Graphical User Interface (GUI) of the WISKI application. The WISKI application is based on 32-bit technology and runs on Windows platforms. Therefore, WISKI's toolbox has a modern interactive graphical time series editor, to allow graphical editing of time series.

Following the data flow through the system, WISKI delivers the following processes and functionalities:

Data validation

Identifying abnormalities in the data is normally the first step of the validation process. Due to malfunctioning of field sensors and other devices, or due to the maintenance of those devices, measured data is likely to have a range of different quality that has to be corrected. To help the user locate data of poor quality, WISKI has auto validation routines. These routines validate the data using predefined criteria and place remarks for user notification each time one of the criteria is violated. Examples of these quality controls are identification of gaps, maximum/minimum exceeding of thresholds, differences to neighboured stations, or other user defined formulas and rules.

Data manipulation

For editing purposes, WISKI offers efficient data editing functionality using an interactive graph or table to allow graphical editing using the computer mouse and keyboard. Examples of functions and methods to correct errors in time series are the filling of gaps with interactive linear or spline interpolation, adjusted data from neighbouring stations or regressions, or e.g. vertical or horizontal stretching of the trace.

Thereby WISKI allows data quality flags to be assigned to each data value. This will identify the primary quality of *Good, Suspect, Estimated, Unchecked* or *Missing*; a secondary flag will provide additional information about the quality, for example edited. Standard comments from a pick list and free text comments added by users can be used to provide additional information about the data (Fig. 2).

WISKI keeps a record of who changed the data, for which period, and writes the name of the applied method automatically for the predefined editing functions listed above. The original data is not edited; instead a "production" time series is defined, which stores a copy of the original values, which can be validated.

Working within a multi-user environment also implies being able to regulate access to the data by defining roles with corresponding access rights. This allows the hierarchy of the organization to be represented by the information system.

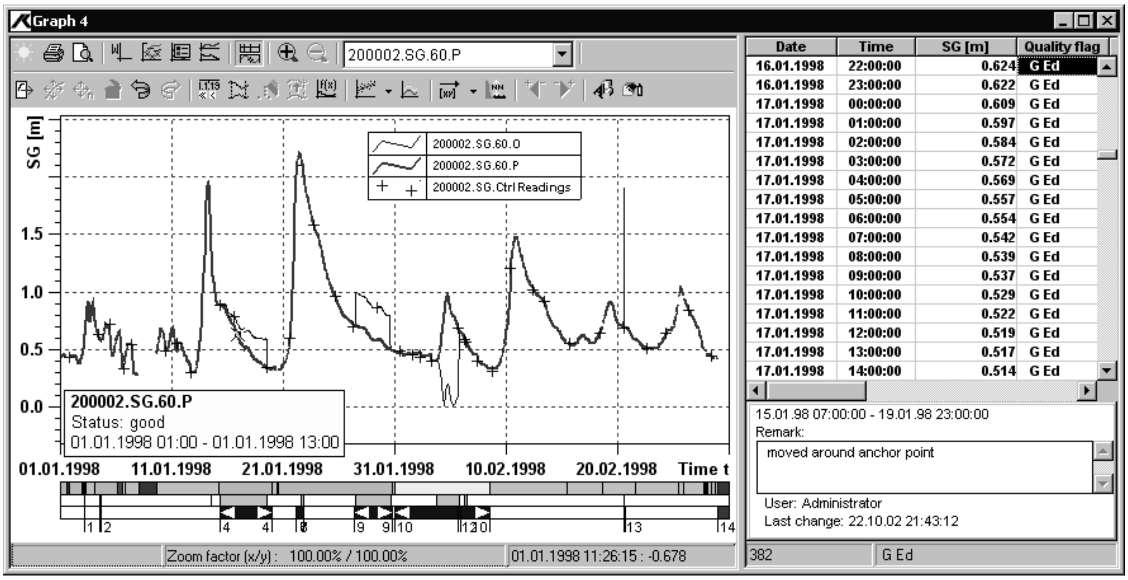

Fig. 2 Interactive data editing in the WISKI graph.

Rating curves

WISKI has a fully integrated rating curve editor. Based on discharge measurements, rating curves can be managed without having to leave the main application. After validation of the stage data all necessary functionality is available to perform the complex task of defining the stage/discharge relationship of a river. The discharge is

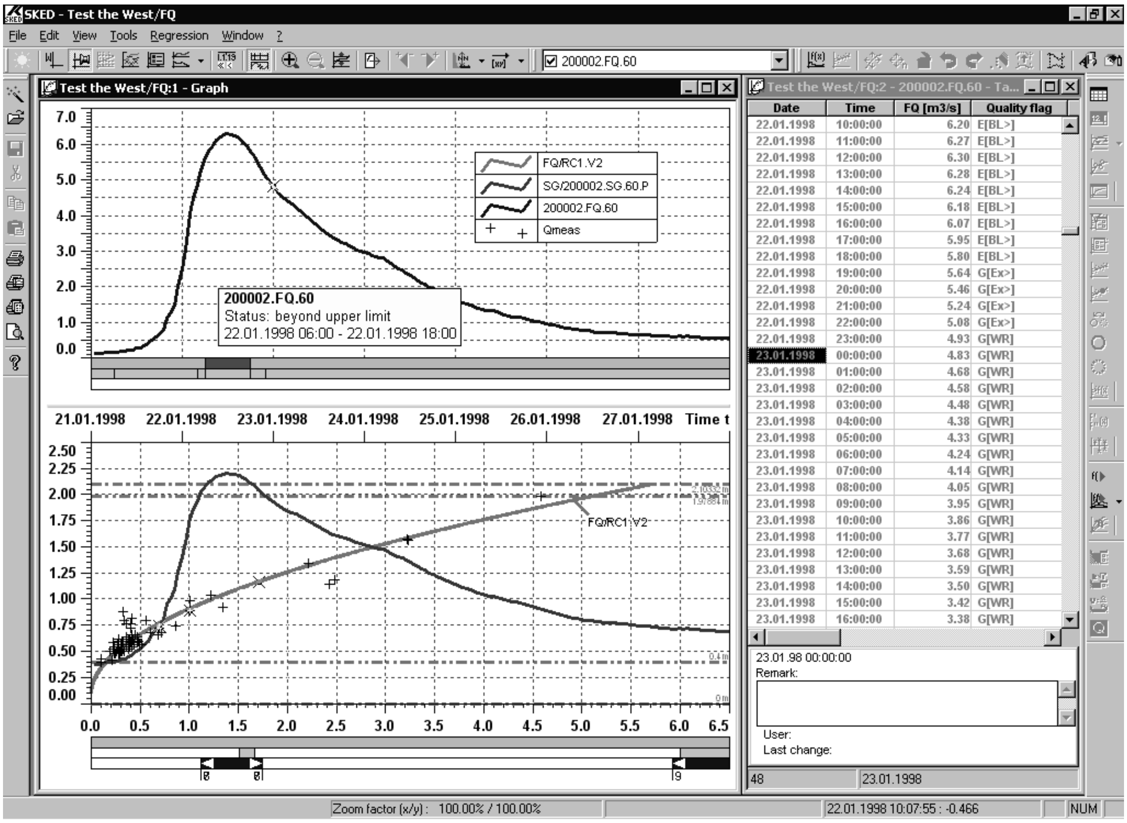

Fig. 3 SKED, the rating curve editor.

calculated automatically, or can be triggered by opening the corresponding time series in a graph or table.

Derived and aggregated values

While validation takes place on high resolution time series, WISKI calculates the derived aggregated time series for long-term analysis of the state of the environment. The daily, monthly and yearly values are derived from the high resolution time series and are calculated automatically. In combination with built-in reports, the user can rapidly generate these reports without exporting data to external publishing software. An internal mechanism makes sure that each time the underlying high resolution time series are modified, the derived aggregated time series are updated automatically. This mechanism prevents a user from publishing or exporting out-of-date data, and prevents them from having to keep a record of when to update derived time series.

Data analysis

For advanced analysis of measured data, WISKI offers statistical analysis tools such as linear and non-linear regression analysis, statistical analysis of durations, floods and low floods, double mass plots, rain storm frequency analysis (Fig. 4). The WISKI

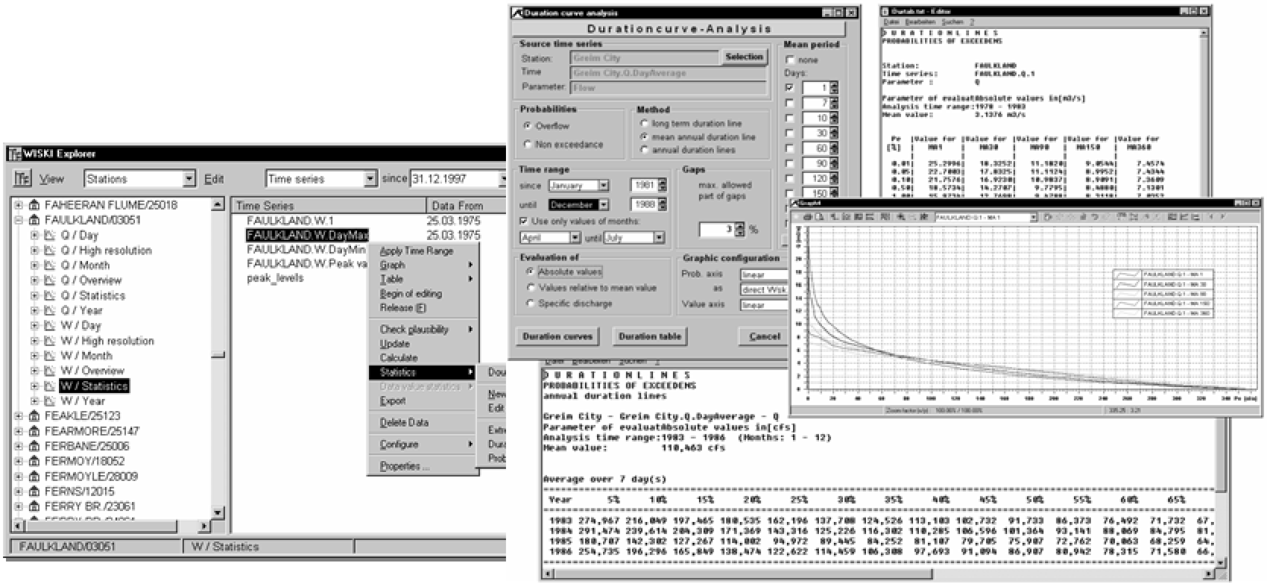

Fig. 4 WISKI duration curve analysis.

statistical analysis tools were developed in close collaboration with hydrologists, and are based on national and international standards, such as the United States Geological Survey, the World Meteorological Organization guidelines, and ISO standards.

Data dissemination

WISKI has an open data model. This means that it is possible to access the data through special views so that users can execute queries using other commercial software such as Crystal Reports and Business Objects, or even their own software to generate reports. WISKI provides standard reports that can be adapted to an organization’s requirements. Today, the Internet has become one of the key gateways for exchanging information. For this purpose a web module has been developed for WISKI for disseminating data on the Internet/Intranet using a web browser as a Graphical User Interface. Moreover, WISKI supports fully automated services such as the sending of data files via FTP (file transfer protocol) or email.

QUALITY ASSURANCE

Today’s demand on hydrological network data by hydrologists, engineers, decision makers and the public is increasing in quantity, quality and time to delivery. For us, this means that we have to process the raw data to value added products and information in an efficient way. Standardized quality assurance procedures can help to fulfill this requirement.

For best results such tools are needed in the whole chain:
measurement → data acquisition → validation and correction → dissemination → use

The hydrological workbench WISKI supports this process with different tools and functions. The quality assurance on the following levels is available:

– data level;
– workflow level;
– system level;
– IT level.

The quality assurance (QA) on these levels is now described.

QA on data level

The quality assurance on data level can be ensured by:

– securing the raw data before processing;
– standardized procedures for data validation, correction and evaluation;
– use of national and international standards or agreed organization wide standards;
– continuous quality marking of the data by users or automatically by the system;
– use of audit trail tools for logging and evaluating data changes.

The raw data is secured in WISKI by saving it in a so-called "original" time series. These time series are write-protected. They can only be changed by importing new data. To correct any errors in the raw data the "original" time series is copied into a so-called "production" time series. The "production" time series is the source for any other derived time series (e.g. daily mean data).

To validate, correct and evaluate data a broad range of different tools are provided. A set of so-called plausibility checks can be assigned to each time series. These plausibility checks can be defined for a particular time range or season. The following list shows a selection of supported plausibility checks:

– boundary check;
– completeness check;
– gradient check;
– distance check, etc.

Furthermore, specific checks for tidal or rainfall data are available.

For defining very specific checks, which depend on particular boundary conditions, the user can define his own plausibility checks by applying the WISKI macro scripting language KiBasic.

The plausibility check results are shown in the WISKI graph and in the WISKI table and will be summarized in a plausibility check table.

The calculation methods, which are applied for time series derivation and statistical analysis, are based on national and international standards. For example the German *Pegelvorschrift,* the English *Flood Estimation Handbook* (FEH) or EC ISO standards are supported.

An important step to assure the data quality is the WISKI flagging and remark concept. Each time series value gets a quality flag assigned to. Three levels of quality flags are supported, primary, secondary and tertiary quality flags (Table 1). Additionally remarks and tasks can be assigned to time series data. The quality flags, listed in Table 1, are available.

Table 1 Summary of the quality flags.

Quality flag level	Quality flag	Abbreviation	Description
Primary	Good	G	These flags are assigned automatically or manually to high resolution data.
	Missing	M	
	Unchecked	U	
	Estimated	E	
	Suspect	S	
	etc.		Further primary flags are supported.
Secondary	Complete	C	These flags are assigned automatically to derived data on daily and upper time level.
	Incomplete	I	
	Missing	M	
	Edited	Ed	This flag is automatically assigned to data which has been edited manually or automatically.
	Within rating	WR	These flags are assigned to derived high resolution data, e.g. flow data, which is derived by applying rating curves.
	Beyond upper limit	BL>	
	Beyond lower limit	BL<	
	No rating	NR	
	etc.		Further secondary flags are supported.
Tertiary	Snow	d	This flag is assigned to rainfall data as additional information.
	Trace	t	
	Apportioned	a	This flag is automatically assigned.
	etc.		Further tertiary flags are supported.

Derived data on daily and upper time level generally gets the primary CIM flag. Additionally this data gets secondary flags which are derived directly from the primary flags of the source values (including their percentages). A typical flag of a daily mean value would be I [30% G, 55% U, 10% S, 5% M].

By analysing the quality flags data can be included or excluded automatically from calculations and statistical analysis. The quality flags can also be exported into files and reports and can be viewed or edited in the WISKI graph and WISKI table (see Figs 2 and 3).

In order to trace all modifications of basic data and time series data, the corresponding transactions are logged. By logging the transaction, the user name, the time of modification as well as the server name an extended audit trail is available. In the figure below the audit trail log dialog as well as the audit trail filter are depicted.

QA on workflow level

The quality assurance on the workflow level must be ensured by:

– clear definitions and policies;
– adoption to the requirements of the individual organization;
– regular training of the users.

Clear definitions and policies naturally must be communicated by the institution itself, but it is helpful if they are supported by the software, for example, by online guidelines

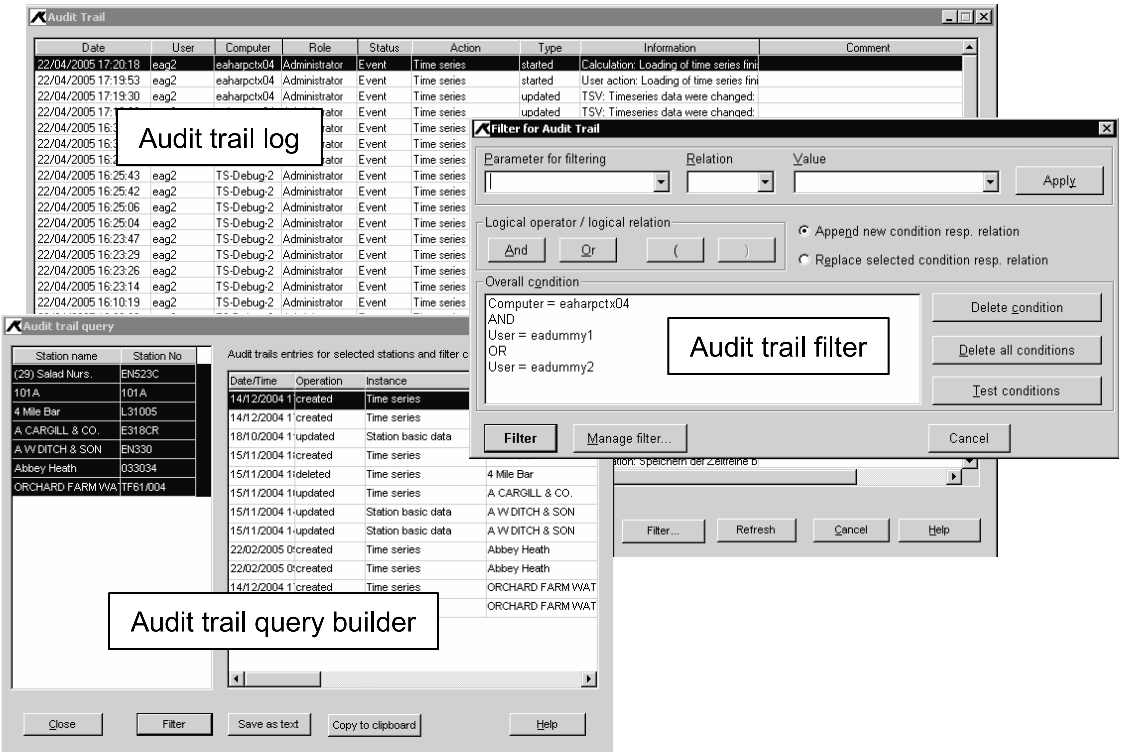

Fig. 5 Audit trail log and filter.

or use cases. For the WISKI package many hundred of use cases are available which describe workflows in a user friendly manner.

Furthermore, the workflow must be supported by adapting to the requirements of individual organizations. This is supported by different wizards for exporting data, creating reports and so on. These wizards lead through the steps of fulfilling the required tasks.

QA on system level

The quality assurance on the system level, which is related to the hydrological network, must be ensured by:

– hydrological cross-checking;
– spatial analysis in GIS environment;
– optimization of number and location for the gauges.

A hydrological cross-checking could for example be the validation of rating curves in a catchment area by calculation of the annual water balance. This calculation can be carried out by applying the formula editor and the macro scripting language KiBasic. By using these tools almost every calculation is possible. This tool can also be applied to create new virtual time series.

The quality assurance of spatial data can be carried out by applying the WISKI Arc View extension, which is directly linked to the WISKI database. This tool

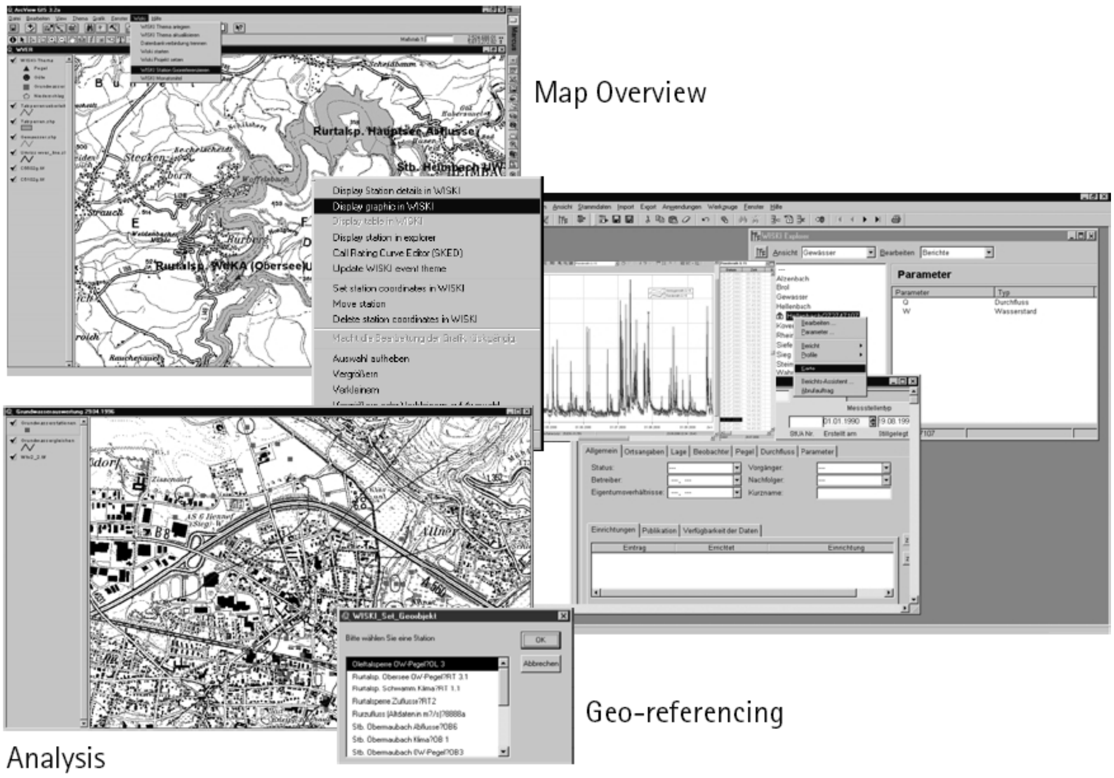

Fig. 6 Spatial analysis in the WISKI Arc View extension.

provides a cartographic overview of the station network and allows by its bi-directional communication with WISKI the correction and analysis of spatial data (see Fig. 6).

The quality assurance on the system level, which is related to the data archive, must be ensured by:

– overall analysis about trends in number and length of gaps in the records;
– overall analysis about trends in data quality;
– overall analysis about trends in time to publish the data.

Comparisons between these indicators between different regions, directorates or departments.

In order to fulfil these important tasks several areas of WISKI functionalities (statistical tools, reporting etc.) are available. This is a part of data quality assurance, which has still to be extended. This extension is planned for a next release within the Kisters R&D budget.

QA on IT level

Last, but not least, the quality assurance on the IT level should not be omitted. Generally these tasks have to be carried out by the authority itself. The quality assurance on the IT level must be ensured by:

– standardized methods and policies for data base back-up and maintenance;
– back-up strategy for data base and software servers, multiple network lines;
– regular data base maintenance (statistics, etc.)
– regular checks of hardware and backup systems, etc.

CONCLUSION

The WISKI system should provide a reliable and flexible archive enabling a central, consolidated source of information. It can provide a modern software platform with sufficient flexibility to encompass future business change and data demands. Its implementation in different authorities reduces costs for support and maintenance, and frees up staff resources to improve data quality, undertake more comprehensive analysis and to convert data into information for managing the environment.

On all levels of data quality assurance WISKI provides ideas, strategies or tools to increase the quality of the data and information. However, due to the never ending process of increasing quality, continuous discussion on this topic is needed.