



Using Indico Validation

How to validate time series data



Airviro User's Reference

Issued by:

Subject:

Version

Volume 8

3.11

Approved:

Date:

1 Jan -2007

Using Indico Validation

How to validate time series data

Amendments

Version	Date changed	Cause of change	Signature
3.11	060910	V3.11, new Airvirofunctionality	GS

Contents

8.1	INTRODUCTION.....	5
8.1.1	What is Indico Validation?.....	6
8.2	GETTING STARTED	7
8.3	EDITING AND CHECKING THE VALUE DATABASE.....	9
8.3.1	Domain.....	9
8.3.2	Selecting time series	10
8.3.3	Period.....	12
8.3.4	Macro Load	13
8.3.5	Validate	13
8.3.5.1	Session.....	13
8.3.5.2	Interface	14
8.3.5.3	Mouse.....	15
8.3.5.4	Keys.....	15
8.3.5.5	Colour coding by status.....	17
8.3.5.6	Scalings.....	18
8.3.5.6.1	Scaling formula – Raw.....	19
8.3.5.6.2	Scaling formula – Scaled (by calibration)	19
8.3.5.6.3	Scaling formula – Linear	19
8.3.6	Ratify	20
8.3.7	Scaling.....	21
8.4	OUTPUT	23
8.5	DELETE.....	24

8.1 INTRODUCTION

Indico Validation is a JAVA-based user interface for validation of time series data. The Indico Validation client can be executed from a PC or other device running Internet Explorer 6, Netscape Navigator 7 (or higher), or a Mozilla based web browser like Firefox or Opera.

Indico Validation client requires no installation at the PC side. Anybody who knows the IAirviro web address and have privileges to do so may use the application. It is therefore important to change passwords regularly and to set privilege levels for the users with great concern to limit unauthorised access. Using an Intranet solution behind a firewall instead of the Internet prevents access for the World Wide Web, if properly set up.

Indico Validation is, together with Indico Presentation, Indico Administration and Indico Report part of the integrated Indico package.

The Indico package is one of the three main modules in Airviro; the others are EDB for storage of emission sources and Dispersion to run dispersion models. Indico is intended for acquiring, storing, editing, presenting, analysing, reporting and exporting time series data.

Indico Validation is one of the main tools in Airviro and it is used for validating time series data.

With the Indico Validation Module, one can view or edit the contents of the time series database. If un-calibrated data has been stored in the time series database, calibrations may be applied to scale the un-calibrated data.

8.1.1 What is Indico Validation?

Indico Validation is a powerful tool for validating data in the time series database. With Indico Validation, one can

- Select up to eight time series for validation
- Access values and status of stored data
- Select time series by using macros (saved in Indico Presentation)
- Scale data with validations values
- Apply different types of scaling for validated data (i.e. raw, linear and scaled)
- Delete time series data
- Edit time series data
- Adjust zero drift using ramps, and
- Save ratified time series as a new ratified instance in the time TSDB.

8.2 GETTING STARTED

Indico Validation has been designed to access, check and validate data stored in the Time Series Database.

It is possible to connect to an Airviro server using a browser over Internet/Intranet.

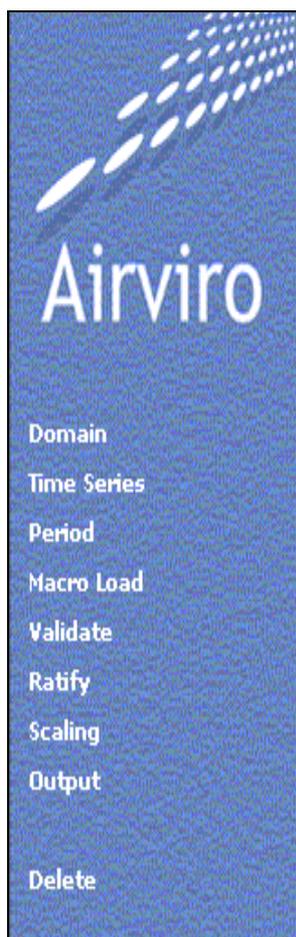
After logging in with user-ID and password, the user is presented with a list of available **Web Modules** and **Domains**, (See *Figure 8.2.1*).



Figure 8.2.1. Getting started.

After clicking on the Indico Validation module, the available menus for this module will be displayed.

The main window is split up into different sections:



- DOMAIN** Used to select the domain one wants to work with and to select the time resolution
- TIME SERIES** Provides an interface to the time series database, where you can select the station, parameters and instances you want to work with.
- PERIOD** Allows the user to select the period of time that will be used during the validation session.
- MACRO LOAD** Used to load an existing macro that has been previously created using the Indico Presentation module. This selects the time series in the macro and optionally the time period.
- VALIDATE** Loads the main Indico Validation page where one can view or edit the time series data. Data can also be scaled using calibration values.
- RATIFY** Is similar to VALIDATE but the validated data is stored as a new ratified instance.
- SCALING** Allows the user to scale time series data using calibration values. Using this interface calibration values for several parameters can be entered at once.
- OUTPUT** Displays the selected time series with the status in a new window with fixed headers.
- DELETE** Allows the user to delete time series.

8.3 EDITING AND CHECKING THE TIME SERIES DATABASE

For a certain domain a time series database may contain a large number of measuring stations and parameters. The parameters can be related to mass concentrations of pollutants or meteorological data, traffic intensities, instrument readings of other kinds or quality control data from data loggers. For each parameter there is also a quality flag (status).

On the other hand, data may arrive occasionally, for example, once per year from some other sources. This data can be imported into the Airviro time series database with Waved®. Time series data can also be generated by the postprocessor in the Dispersion module or by statistical forecasting in the Aircast module or from some meteorological agency. Also postprocessors in Indico can generate averaged values from lower time resolutions to be stored in the TSDB.

When Indico Validation has been initialised, a list of available menus is displayed on the left-hand side. The user must first choose a **Domain & Time resolution**. After that, time series must be selected by choosing a combination of parameters and instances for the selected stations. The time period is selected using **Period**. Also **Macro Load** can be used to load predefined macros containing time series and optionally a time period.

Validate can be used to view or edit the time series data. Data can also be scaled using calibration values. With **Scaling** one can scale time series data using calibration values. Finally one can ratify the changes with **Ratify**, which is used less frequently, typically each year or half a year.

8.3.1 Domain

Domain shows the Domain & Time resolution page. Here one selects the domain and one of the available time resolutions for the domain. Typically only the lowest resolution of data is validated (*Figure 8.3.1*).

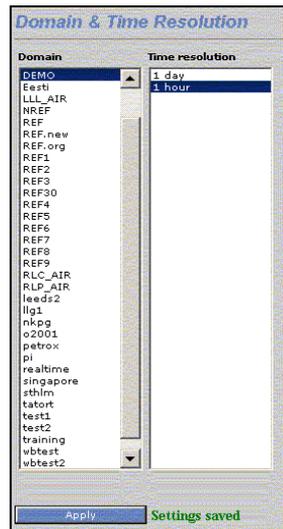


Figure 8.3.1. Domain and Time Resolution.

It is possible to work simultaneously with many instances of Indico Validation. Many users can work on the same domain and one user can work on different domains or time resolutions without any substantial risk for interference.

8.3.2 Selecting time series

In the **Time Series** page, you will see a list of all stations - active or inactive, in the station database and all parameters that are stored in the time series database. If one selects a station by clicking on its name in the station list, only the parameters that the chosen station has data for will be shown. Clicking **Re-scan** will update the list.

Clicking **Clear** removes selections and shows all stations and parameters. If you are interested in all stations that measure a certain parameter, start by clicking on the parameter in the parameter list box and all the stations that have data for this parameter will be shown in the station list (*Figure 8.3.2.*).

It is possible to **sort** the stations or the parameters in the list box alphabetically or by station key or parameter key, by selecting sort key in the associated drop-down list box. Checking/un-checking **Reverse** reverses the sort order accordingly. It is also possible to promote active stations by moving them to the top of the station list box by checking **Active first**. Sorting stations also by reverse death time creates a list of increasingly older stations. Click **Clear** to get a full list of stations. You may notice that some stations are preceded by an asterisk (*). This is to show that they are operational stations, i.e. they collect data automatically (*Figure 8.3.2.*).

When you have selected both station and parameters, you will get a list of available instances for the actual selection. The instance is used to differentiate between

simultaneous measurements of the same parameter at the same site. For meteorological data, the instance is normally the height of the instrument above ground.

During the Data Collection (managed by cold) collected data is stored in the Time Series Database according to the settings defined in the Indico Administration module for each station. One can only enter scalings for the instances with the parameter type "v" (volt value). A status flag is also assigned in the quality control in Indico Administration. This status can be changed if you edit this data with **Validate**. Instances with type M cannot be scaled (*Figure 8.3.3.*).

When you have clicked on station, parameter, instance and attribute, the time series is uniquely identified (for the current time resolution). Click **New** to select the time series for further processing (Also time series can be selected by pressing the space bar or by double clicking on the selected items in the list). One can remove a highlighted selected variable from the list box with the **Remove selected** button or replace it by identifying another time series and clicking **Replace**. If you click **Clear all**, all time series will be removed from the "Selected" list box.

The same restrictions for selecting the time series as in the Indico Presentation are valid. The user can select up to 15 time series. Up to 8 variables at a time can be presented in the selection page. When one is satisfied with the time series selections, click **Apply** to save the settings (*Figure 8.3.2.*).

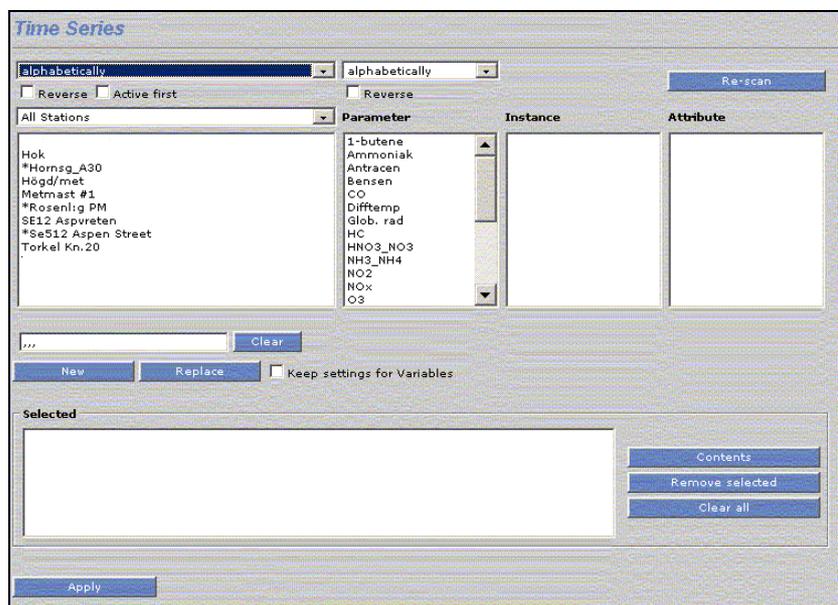


Figure 8.3.2. Time Series Windows.

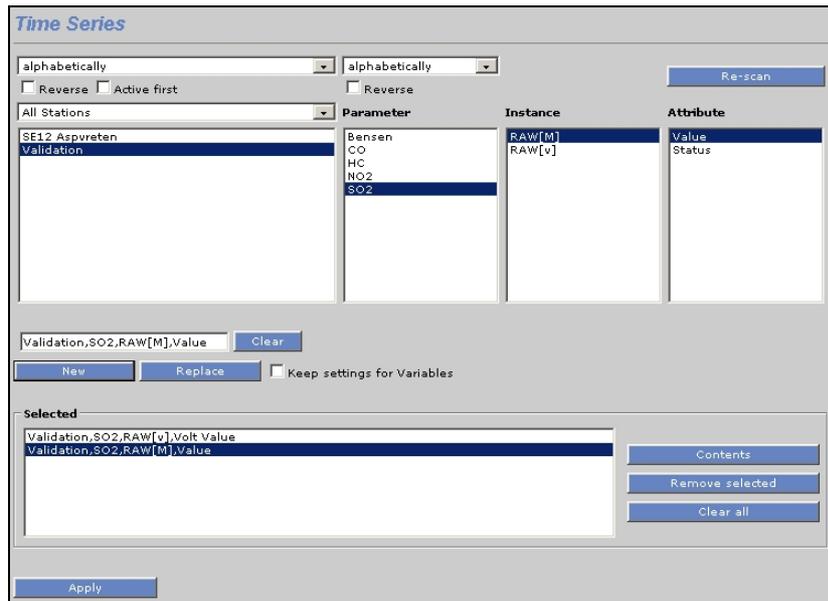


Figure 8.3.3. Example of selected time series.

8.3.3 Period

By default, Indico Validation selects the last week of data. In the **Period** page, one can change the time period. Date and time can be presented in European, UK or US date format. If one want to set another start date, one can write the date in the **From** box. Alternatively, use the double arrows between the From and To boxes to transfer a date between the boxes. One can also use the double arrows adjacent to Year, Month, Week, Day or Hour to step forwards or backwards with one time step. One can always reset the **To** box to present time by clicking **Present**.

When one has selected a period, click **Apply** (Figure 8.3.4.).

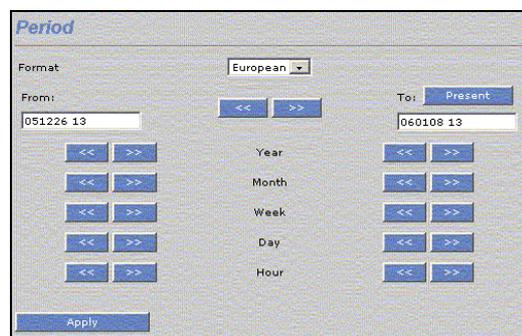


Figure 8.3.4. Period.

The hour starting at 00:00 and ending at 01:00 is named 01. The hours are numbered from 01 to 24. This means that the hour starting 23:00 and ending 00:00 is named 24.

The date and time given is inclusive in the From box and exclusive in the To box.

8.3.4 Macro Load

Under **Macro Load** one can select the available macros. The macros have been previously created using the Indico Presentation module.

Macros are stored in groups, where there is a group for each user in a domain and a common group. The system administrator decides who is allowed to store macros in the common group (using `Indico.WriteGroup.user` in `priv.rf`). Users can always save macros in their own group, but usually not in other groups, although it is possible to load macros from other groups (*Figure 8.3.5*).

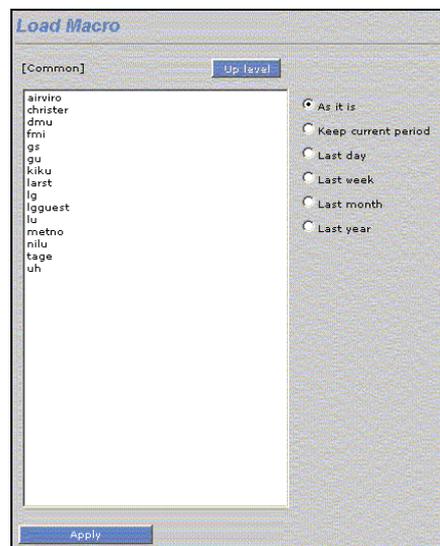


Figure 8.3.5. Macro.

8.3.5 Validate

In this sub window the user can view or edit the selected time series. The user must select time series before starting a validation session.

The **Validation** page is shown in *Figure 8.3.6*.

8.3.5.1 Session

A session consists of 1 to 8 time series of data for a limited time period.

If there are Scalings available for a selected time period, the first active scaling before the selected period will be included.

Both raw and scaled data will be loaded for each time series if available. The raw data is never modified in any way, but the scaled data may be edited or changed by adding, changing or deleting scaling.

If no raw data is available, no scaling can be applied and it is only possible to edit data. No changes are actually saved to the database until the **Apply** button is pressed (*Figure 8.3.6*).

Data is stored as raw data if it is stored as type **v**. The other options are to store data as type **M** or **O** (Opsis time series). In those cases scalings cannot be used.

8.3.5.2 Interface

Spreadsheet	The spreadsheet contains an active time window, which usually is a subset of the whole session. The active window may also contain one value that has keyboard focus (marked with a black square).
Graph	<p>If no time series in the spreadsheet has focus, all time series in the session will be shown. Graphs will be colour coded per time series in the same way as the time series labels in the spreadsheet.</p> <p>If a time series has focus, only that time series will be shown. The graph will be colour coded by status in the same way as the data in the spreadsheet.</p> <p>The active window is marked with a light grey hollow rectangle in the graph.</p> <p>Any Scalings are marked with triangles at the top of the graph area. Active Scalings have a dark triangle and inactive ones have a gray triangle. If a scaling is selected in the Scalings list, the corresponding marker in the graph will be slightly lower than the others.</p> <p>When a shorter time period is shown, each value is marked with a diamond. This disappears when there are too many values to be shown.</p>
Scaling	If there is raw data available, the Scalings interface will be shown to the right of the graph. Here it is possible to add, edit and delete Scalings.
Axis	If auto scaling of the x-axis is enabled, the application will try to find the best possible scaling. It is also possible to enter manual limits to the x-axis. To enter manually the Y-axis scale, just type the min and max value for it in the text boxes and press enter after that.

Fonts & Columns	The size of the text in the spreadsheet and the number of columns can be selected. These settings will be saved in a cookie between sessions.
-----------------	---

8.3.5.3 Mouse

Clicking on a spreadsheet in a cell with a value will set focus to that value. Clicking on the time series labels to the left will remove focus from an individual value. Clicking on the graph will set the focus on the nearest value and move the active window so that it is centred on the value. Pressing and holding down the left mouse button will allow interactive dragging of the active window.

8.3.5.4 Keys

Page Up	Scroll active window back one window size.
Page Down	Scroll active window forward one window size.
Arrow Up	Move focus to cell above.
Arrow Down	Move focus to cell below.
Arrow Left	Move focus to cell to the left.
Arrow Right	Move focus to cell to the right.
Home	Move active window to beginning of session.
End	Move active window to beginning of session.
Esc	If editing a value, this will abort the editing without changes; otherwise it will remove focus from the spreadsheet.
Enter	If editing a value, this will confirm the changes you have made and return to spreadsheet.
SHIFT+ALT+Arrow Up	Copy value to cell above and move focus.
SHIFT+ALT+Arrow Down	Copy value to cell below and move focus.
SHIFT+ALT+Arrow Left	Copy value to cell to the left and move focus.

SHIFT+ALT+Arrow Right	Copy value to cell to the right and move focus.
d	Delete current value, i.e, change the status to 2.
D	Delete a number of values. The time period is entered in a dialog.
u	Undelete current value (change status 2 to 15).
U	Undelete a number of values. Time period entered in a dialog.
e	Edit current value. Changes the status to 15.
f	Apply ramp on a series of values. Time period, offsets and multiplication factor entered in dialog. The offset is applied before the multiplication factor. The status of the changed values is 15.
r	Restore current value and status to the original value at the start of this session.
R	Restore a number of values. Time period entered in dialog.
i	Recover current value from raw value with applied scaling. Overrides status 2 and 15.
I	Recovers a number of values. Time period entered in a dialog.
s	Search for values of a certain status and sets the value to a specified value. Time period, status and new value entered in dialog. Note: Applies to all loaded time series, not only the active time series. The status will be set to 15 (Manually changed).
t	Apply a lower threshold on a series of values. Time period and the lower threshold entered in dialog. Status is set to 2 (Deleted).

z	Toggle graph between zoomed mode and whole mode. Only the data in the active window will be shown when in zoomed mode.
----------	---

8.3.5.5 Colour coding by status

The following colour codes are used for statuses:

1	Blue	Unchecked (should normally not appear)
2	Dark Red	Manually marked invalid.
3	Red	Error from logger.
4	Red	Value below configured minimum.
5	Red	Value above configured maximum.
6	Red	Exceeded maximum configured gradient.
7	Red	Variations less than configured minimum.
8	Red	Too large standard deviation (OPIS only)
9	Red	< not used >
10	Orange	Logger warning.
11	Orange	< not used >
12	Orange	< not used >
13	Orange	Suspect
14	Blue	Value checked ok.
15	Green	Manually changed.

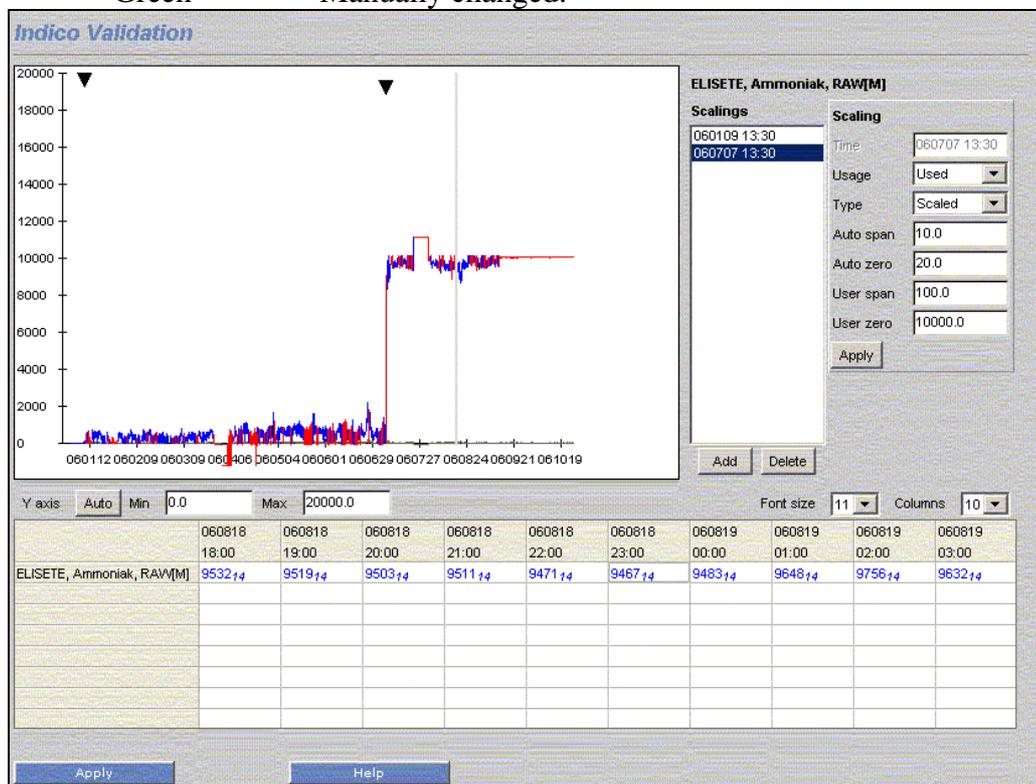


Figure 8.3.6. Example of the Validation.

8.3.5.6 Scalings

In the **Scaling** sub window the user can enter zero and span voltage values obtained during calibration of the actual analysers.

In this sub window one can specify scalings for the parameters measured at different stations. **Apply** must be pressed to save changes (*Figure 8.3.7.*).

To add a new Scaling, click on **Add** and then input: Time, Usage, Type and their characteristics. In the *Usage* combo box, different options can be selected (inactive, used or discarded.) In the combo box *Type*, three different options can be selected (raw, scaled or linear) and for each of them their characteristics must be set up according to this detail:

- Raw: Cylinder concentration, user span and user zero must be entered.
- Scaled: Auto span, auto zero, user span and user zero must be entered.
- Linear: Line slope and line offset must be entered.

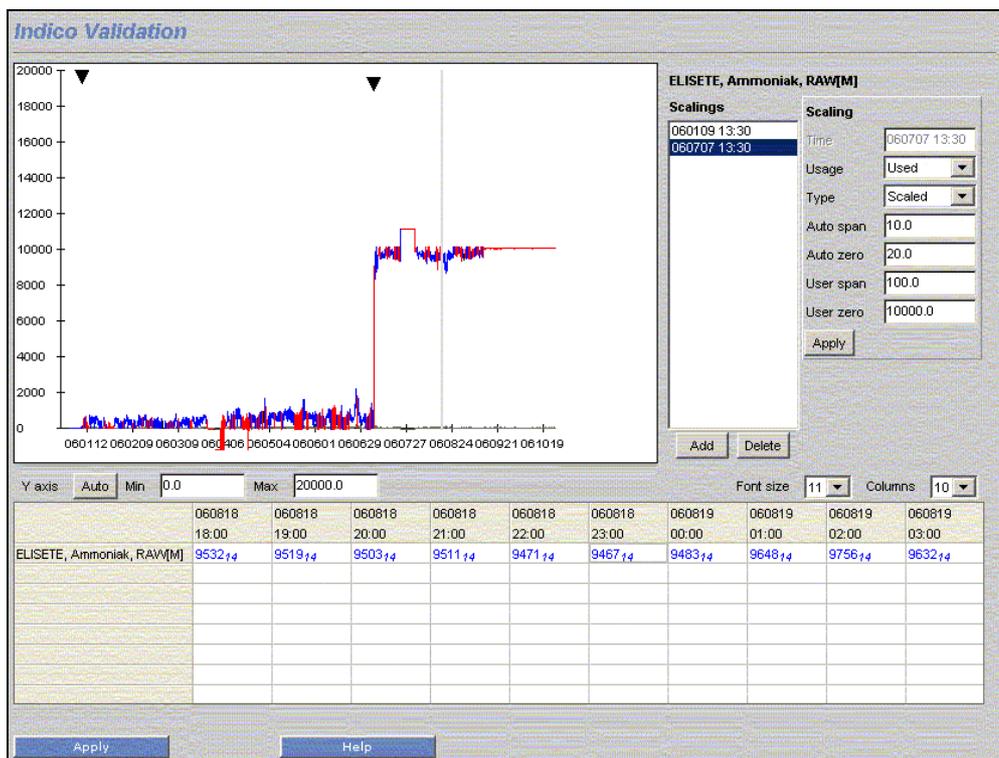


Figure 8.3.7. Example of the Scalings

A scaling is applied from the period defined, until next available scaling. When a scaling is performed the raw data is always left untouched. The scaled value is stored in the instance M. The raw in instance v. The two instances exist in parallel in the database.

Scalings can be deleted using the Scaling menu. The first scaling cannot be deleted but by changing it to a neutral scaling, i.e., the scaled value is the same as the raw, the scaled data will have the same value as the raw.

8.3.5.6.1 Scaling formula – Raw

In this case the values received from the logger/instrument are unscaled, e.g. mV values.

$$V_{scaled} = (V_{raw} - Z_u) \frac{CC}{(S_u - Z_u)}$$

Where:

V_{scaled}	The scaled value calculated in Airviro.
V_{raw}	The value received from the logger/instrument.
Z_u	The zero point entered.
S_u	The span point entered.
CC	Span gas cylinder concentration.

8.3.5.6.2 Scaling formula – Scaled (by calibration)

In this case the values received from the logger/instrument are scaled using the latest valid calibration.

$$V_{scaled} = (V_{raw} - Z_m) \frac{S_u}{(S_m - Z_m)} + Z_u$$

Where:

V_{scaled}	The scaled value calculated in Airviro.
V_{raw}	The value received from the logger/instrument.
Z_m	The zero point measured in calibration.
S_m	The span point measured in calibration.
Z_u	The zero point entered.
S_u	The span point entered.

8.3.5.6.3 Scaling formula – Linear

In this case the values received from the logger/instrument are scaled using a linear transformation – useful for entering multipoint calibrations.

$$V_{scaled} = V_{raw} \cdot L_k + L_m$$

Where:

V_{scaled}	The scaled value calculated in Airviro.
V_{raw}	The value received from the logger/instrument.

L_m Offset.
 L_k Slope.

Multipoint calibrations are normally made in a laboratory and produce a best fit of a linear equation.

8.3.6 Ratify

Ratify is used to ratify data. Validation of data is made rather frequently while the ratification of data is less frequent, each half year or year, and it spans over a much larger time period.

The "Ratify" function is more or less the same as "Validate". The difference is that the ratified values are stored as a different instance that is selectable.

The ratified data is thus its own time series and stored in parallel with raw and scaled data.

When you have finished ratifying click on **Apply** to save your data (See *Figure 8.3.8.*).

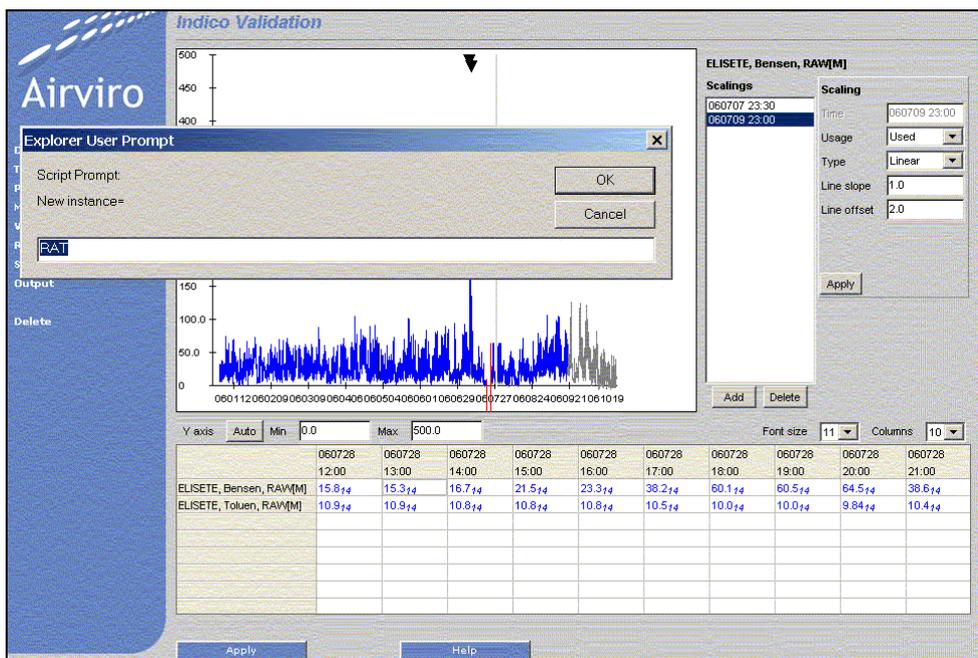


Figure 8.3.8.Ratify

8.3.7 Scaling

In this sub window, one can define several scalings at once for a station. This should allow scalings to be input faster than in the *Validate* interface. It is possible to add, edit and delete Scalings.

In the **Scaling** window, select a station from the stations list and another list box will appear showing the scalings defined for the selected station. In the first column description of the scalings for each calibration type is shown. In this window, one can define a new calibration type for each parameter and date period.

In this raw case (see formula, 8.3.5.6.1.) the values received from the logger/instrument are un-scaled. The user must enter the following:

- The zero point entered (Z_u)
- The span point entered (S_u)
- Span gas cylinder concentration (CC)
- Status: inactive, used and discarded
- And ZS; Z is Z_u and S is $(cc/su-zu)*1000$

Note: the column ZS is a result of calculations.

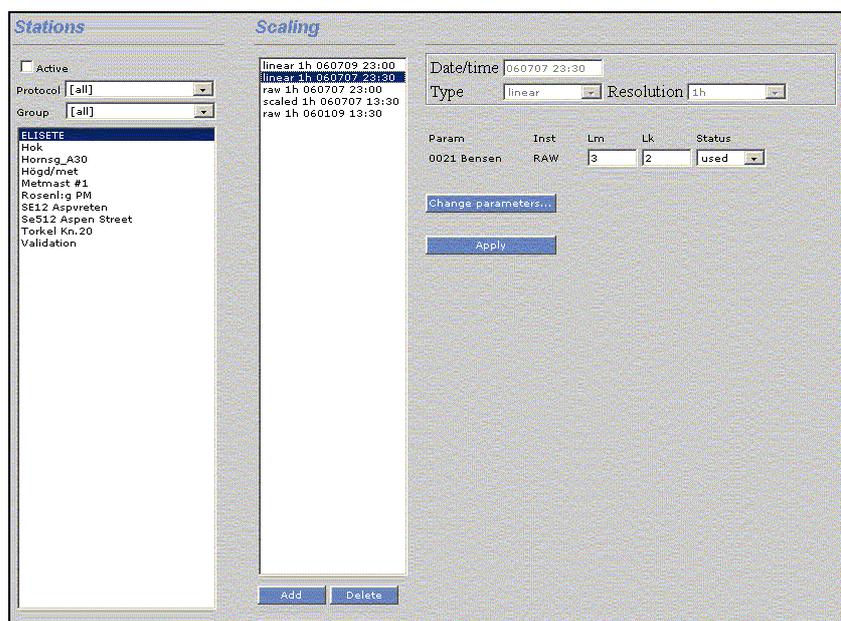


Figure 8.3.9. Scaling.

Filters may be used to search for a station to which scalings want to be added. The active stations are visualised by checking the **Active** check box (Figure 8.3.9.).

One can also filter stations according the protocols used or the station group they belong to. These filters are useful if you have a system with a large number of stations.

When you select a station from the list, all the scaling associated to this station will be shown in a new list box.

If you click on a scaling, it will then be shown, and you can check it or change it by editing an existing parameter or by adding a new parameter to be scaled.

If you click on **Add**, you will be requested to complete date/time, type of the scaling, parameter and values for each scale variables.

In the scaled case (see formula, 8.3.5.6.2) the user must enter the following:

- the zero point measured in calibration (Z_m)
- the span point measured in calibration (S_m)
- the zero point entered (Z_u)
- and the span point entered (S_u)

In the linear case (see formula, 8.3.5.6.3) the user must enter the following:

- offset (L_m)
- and slope (L_k)

Click on **Apply** to save your changes (*Figure 8.3.9*).

8.4 OUTPUT

It is possible to create reports showing data stored in the TSDB.

The output is easy to read with fixed headers above the columns of data. It should be used to scan through large sets of data. The following is a typical data report output:

	Validation CO RAW[v] Volt Value	Validation CO RAW[M] Value	Validation CO RAT[M] Value		
# Header					
	Validation, CO, RAW[v], Volt Value				
	Validation, CO, RAW[M], Value				
	Validation, CO, RAT[M], Value				
# Data					
051215	1100	2.7 (4)	2.7 (14)	-	(0)
051215	1200	4.8 (4)	4.8 (14)	-	(0)
051215	1300	7.7 (4)	7.7 (14)	-	(0)
051215	1400	10 (4)	10 (14)	-	(0)
051215	1500	2.5 (4)	2.5 (14)	-	(0)
051215	1600	1.8 (4)	1.8 (14)	-	(0)
051215	1700	1.7 (4)	1.7 (14)	-	(0)
051215	1800	1.7 (4)	1.7 (14)	-	(0)
051215	1900	1.8 (4)	1.8 (14)	-	(0)
051215	2000	1.7 (4)	1.7 (14)	-	(0)
051215	2100	1.9 (4)	1.9 (14)	-	(0)
051215	2200	4.3 (4)	4.3 (14)	-	(0)
051215	2300	3.8 (4)	3.8 (14)	-	(0)
051216	0000	5.3 (4)	5.3 (14)	-	(0)
051216	0100	5.3 (4)	5.3 (14)	-	(0)
051216	0200	11 (4)	11 (14)	-	(0)
051216	0300	14.3 (4)	14.3 (14)	-	(0)
051216	0400	14.1 (4)	14.1 (14)	-	(0)
051216	0500	15.5 (4)	15.5 (14)	-	(0)
051216	0600	22.2 (4)	22.2 (14)	-	(0)
051216	0700	20.3 (4)	20.3 (14)	-	(0)
051216	0800	18.5 (4)	18.5 (14)	-	(0)
051216	0900	15.3 (4)	15.3 (14)	-	(0)
051216	1000	12.5 (4)	12.5 (14)	-	(0)
051216	1100	12.9 (4)	12.9 (14)	-	(0)
051216	1200	11.6 (4)	11.6 (14)	-	(0)
051216	1300	11.2 (4)	11.2 (14)	-	(0)

Figure 8.4.1. Output.

8.5 DELETE

In the **Delete** sub window, the user can delete time series. For each station using the **Update** button, the user can select using the available filters, which time series one wants to delete (*Figure 8.5.1*).

By selecting, for example, a station in the station list and pressing *Update*, the *Selected* list will be filled with all time series of that station. By selecting time series of that station in the *Selected* list and then by pressing *Delete*, the selected time series will be deleted.

NOTE: All the data from a time series will be deleted, i.e., any time period specified will not be taken into account.

It is not possible to undelete deleted data.

This functionality is enabled for specific users and roles according to the Airviro privileges settings carried out by the system administrator.

Station	Parameter	Resolution	Type
SE12 Aspvreten	SO2	1 day	000[M]
Högd/met	NO	1 hour	020[M]
Validation	NO2		026[M]
SeS12 Aspen Street	CO		036[M]
ELISETE	HCl		GJ8[M]
Hornsg_A30	O3		N_n[M]
Torkel Kn.20	Kol		N_s[M]
Rosenl:g PM	Ammoniak		PER[M]
Metmast #1	Hg		RAT[M]
Hok	Bensen		RAW[M]
	Toluen		S_n[M]
	P-Xylen		S_s[M]

None None None None

Update Max. results: 1000

Selected

Delete

Figure 8.5.1. Delete.