Co-UDlabs

BUILDING COLLABORATIVE URBAN DRAINAGE RESEARCH LABS COMMUNITIES

DATA VALIDATION with the Urban Drainage Metrology Toolbox

Webinar, 10 January 2025

Jean-Luc BERTRAND-KRAJEWSKI (INSA Lyon)
Francois CLEMENS-MEYER (Skillsinmotion)
Elodie BRELOT (GRAIE)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101008626

Lyon, 25 ans de la Fête des Lumières, 2024 © JLBK

Happy New Year 2025

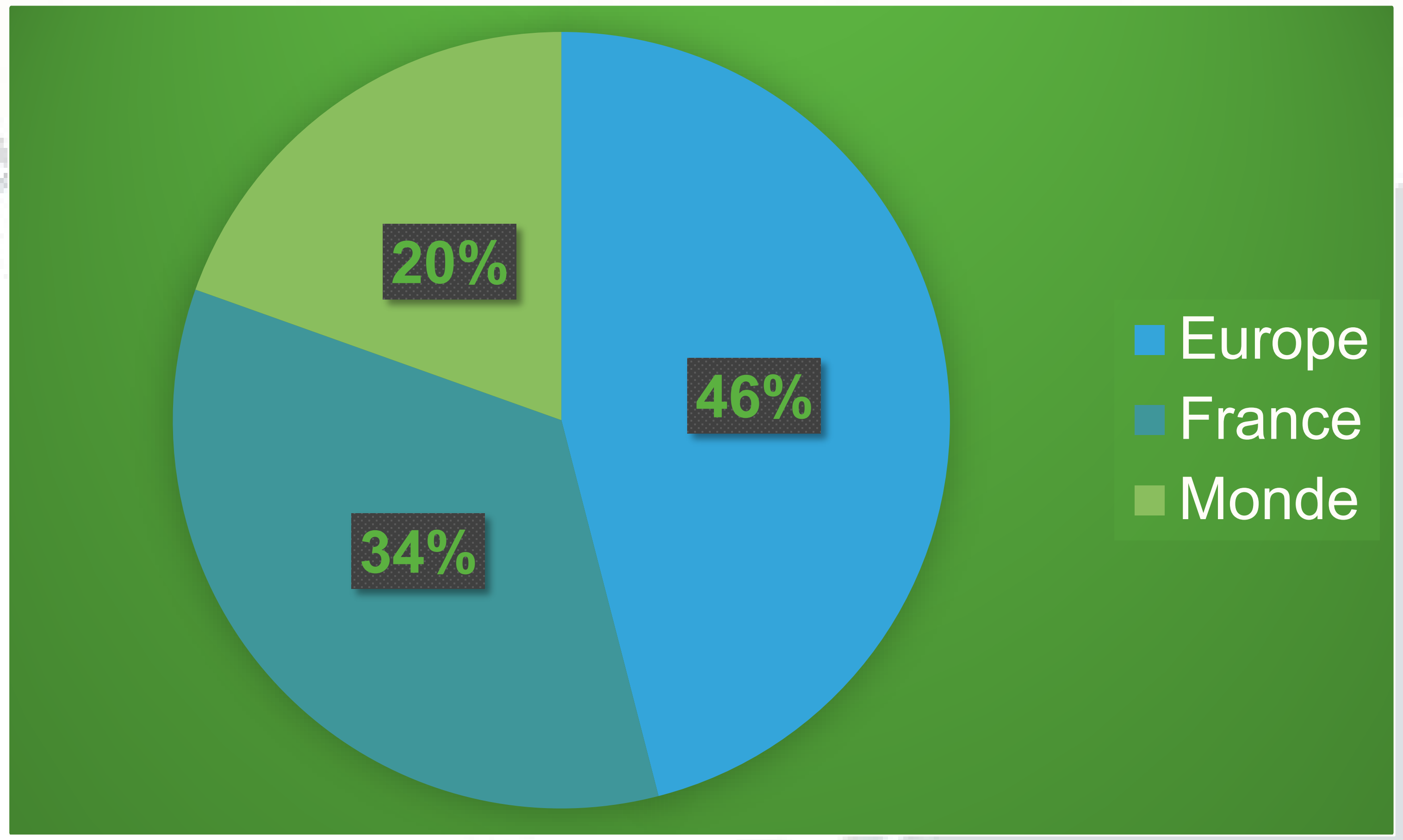
Bonne Année 2025

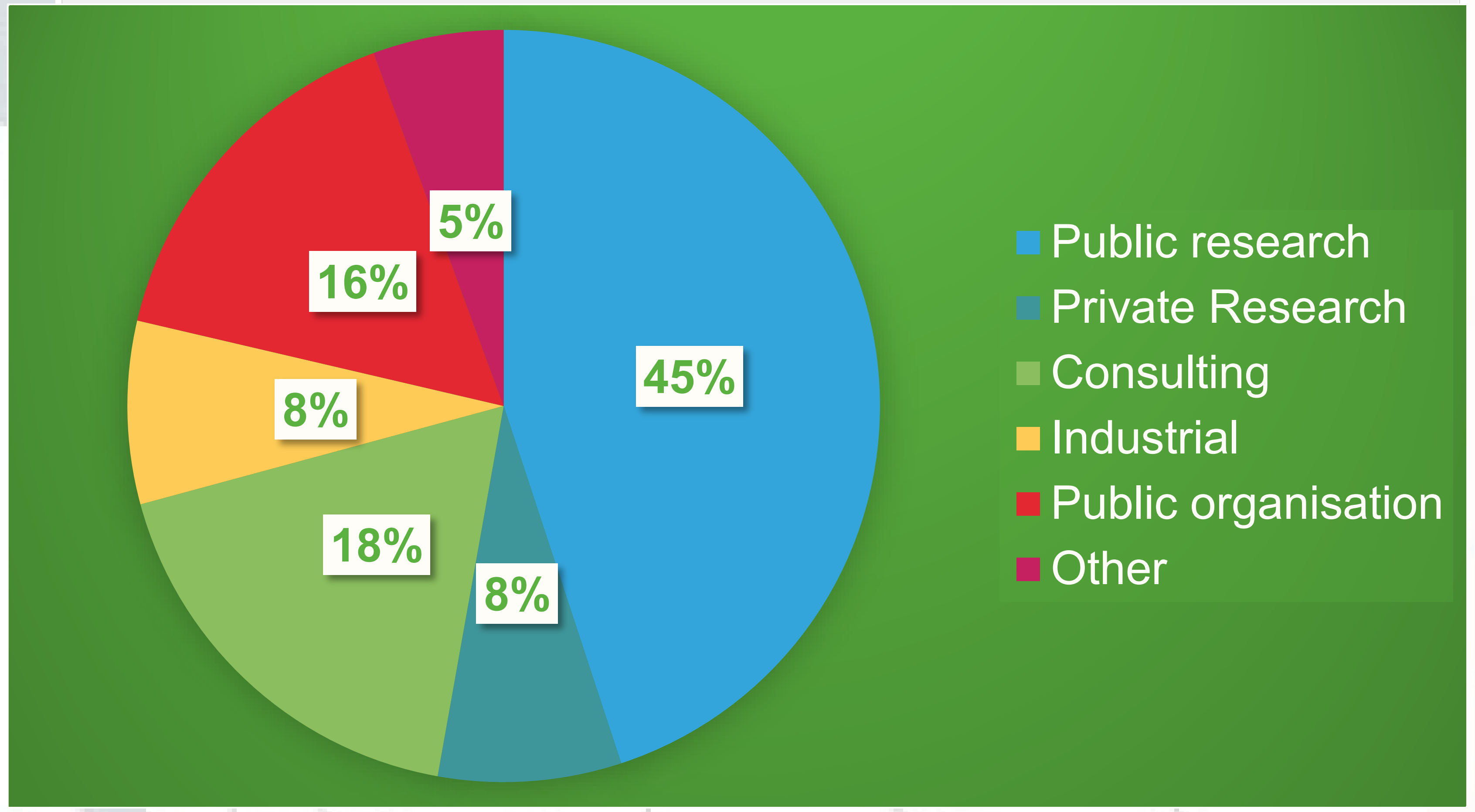
WEBINAR PROGRAMME

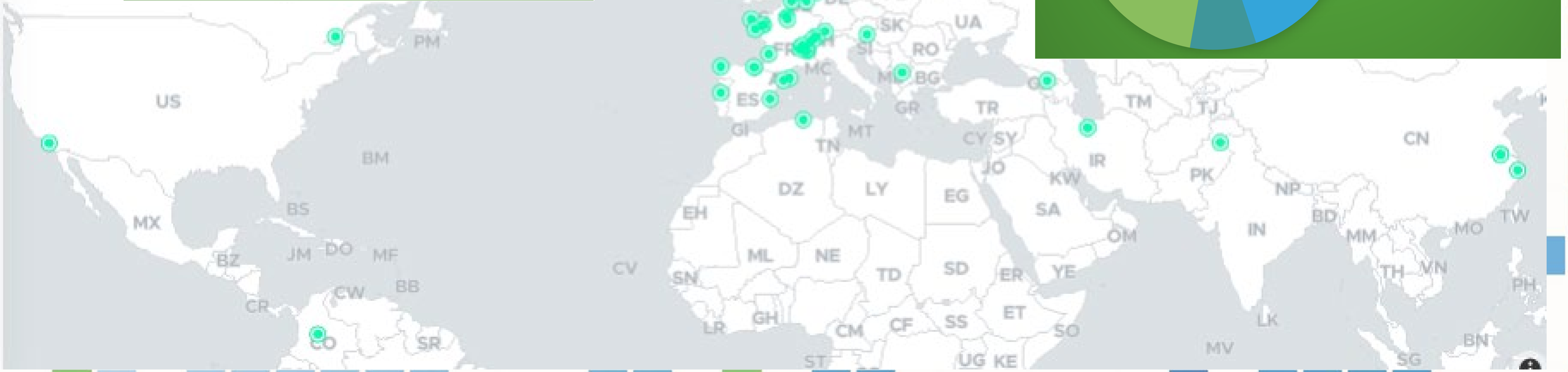
- 14:00 Welcome and introduction
  Elodie Brelot, GRAIE and Jean-Luc Bertrand-Krajewski, INSA Lyon
- 14:05 Importance of Data Validation in Urban Drainage monitoring and necessity to develop its systematic application
  François Clemens, Skillsinmotion
- 14:45 Coffee break
- 15:00 Presentation of the UDMT – Urban Drainage Metrology Toolbox developed in the Co-UDlabs project
  Jean-Luc Bertrand-Krajewski, INSA Lyon
- 15:15 Data validation with the UDMT (JLBK)
  - What is implemented in the UDMT?
  - Examples of application
- 15:50 Q&A, Concluding remarks,
- 16:00 End of the webinar

WELCOME!

✦ 99 registered participants / 83 organisations







Co-UDlabs

COLLABORATIVE URBAN DRAINAGE RESEARCH LABS COMMUNITIES

- 17 large scale research facilities
- 9 partners, from 7 countries
- coordinated by the university of Coruna

Activities in the Co-UDlabs project:

- Multidisciplinary research
- Innovation & transnational Access to platforms
- Training & transfer
- Networking

➔ Spring of 2025

Elodie Brelot, Graie

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101008626

AALBORG UNIVERSITY · euronovia · Deltares · INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON · graie · IKT · eawag aquatic research · The University Of Sheffield. · UNIVERSIDADE DA CORUÑA

Training & transfer - Next events

On **Wednesday, 28 January 2025** (12:00 - 13:00 CET), EAWAG will held a workshop on **FAIR (Findable, Accessible, Interoperable and Reusable) data**.

This workshop will explore the latest concepts, implementations and resources aimed at harmonizing and sharing data to enhance urban drainage water management strategies.

REGISTER HERE

On **Thursday, 30 January 2025** (10:00am-14:00pm CET), Deltares, IKT and The University of Sheffield will organise a webinar on **Techniques for monitoring underground infrastructures**.

The webinar will present some recent advances in inspection technologies that can be used to monitor various assets in sewer networks. The presentation will range from new approaches that are entering the market to techniques that are emerging from recent completed research.

AGENDA & REGISTRATION

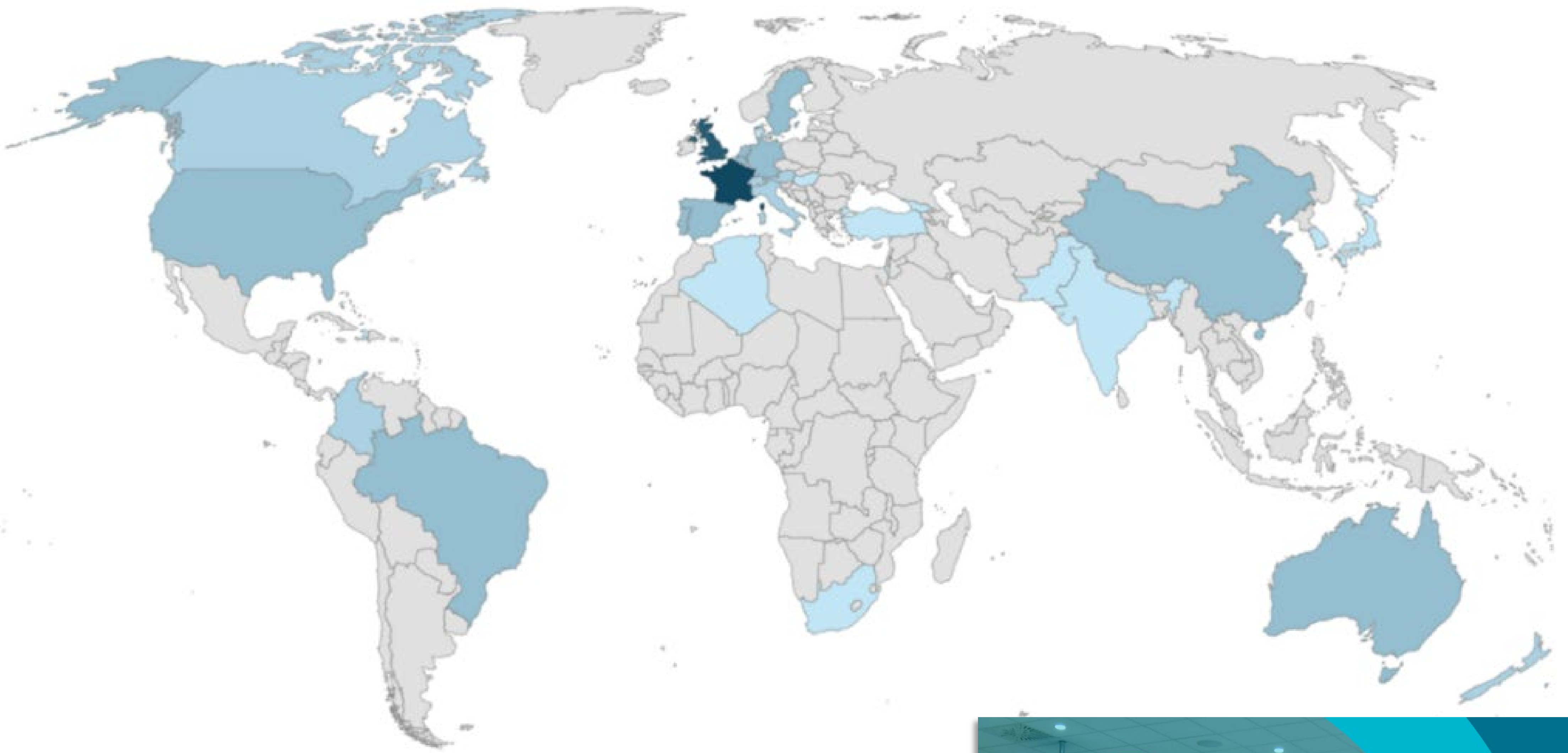
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101008626

Co-UDlabs

COLLABORATIVE URBAN DRAINAGE RESEARCH LABS COMMUNITIES

Networking - Next events

Join the UDRAIN Working Group on Large Research Infrastructure in UD!



FOLLOW THE Co-UDlabs PROJECT UPDATES 2020-2025

https://co-udlabs.eu/

Co-UDlabs
COLLABORATIVE URBAN DRAINAGE RESEARCH LABS COMMUNITIES

Sign for our news and newsletter
https://co-udlabs.eu/contact/

Social Media
https://www.linkedin.com/company/co-udlabs-project/
https://twitter.com/CoUDlabs

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101008626

AALBORG UNIVERSITY · euronovia · Deltares · INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON · graie · IKT · eawag aquatic research · The University Of Sheffield. · UNIVERSIDADE DA CORUÑA

USEFULL LINKS

- **The free UDMT Toolbox :** http://vps-7bc5cf87.vps.ovh.net:9988/webapps/home/session.html?app=coudlabs
- **Any comment :** mailto:UrbanDrainageMetrologyToolbox@gmail.com
- **The Co-Udlabs project :** https://co-udlabs.eu/
- **Metrology in Urban Drainage and Stormwater Management: Plug and Pray** https://iwaponline.com/ebooks/book/835/Metrology-in-Urban-Drainage-and-Stormwater
- **Asset Management of Urban Drainage Systems: If anything exciting happens, we've done it wrong!** https://iwaponline.com/ebooks/book/920/Asset-Management-of-Urban-Drainage-SystemsIf
- **Standardisation - the SWAN IUG group** https://swan-forum.com/interoperable-utility-group/
- **The Dutch project on data standardisation:** https://data.gwsw.nl/ It is in Dutch but I guess translations should be doable....

IMPORTANCE OF DATA VALIDATION IN URBAN DRAINAGE MONITORING AND NECESSITY TO DEVELOP ITS SYSTEMATIC APPLICATION

Francois CLEMENS-MEYER
(Skillsinmotion)

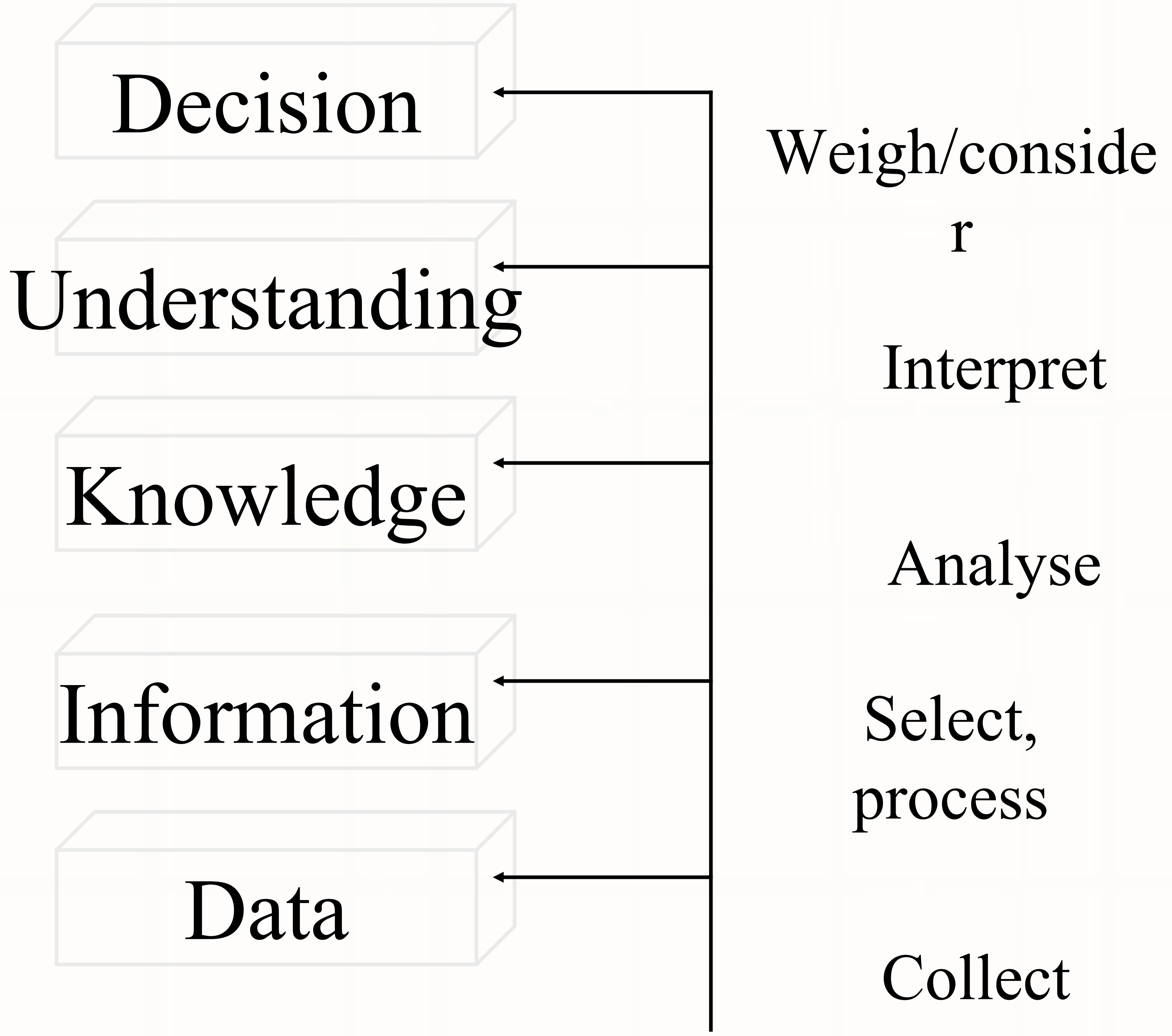
CONTENT

- Why monitor UD systems in the first place?
- Examples of the need for data validation/ consequences of not validating data
- Data Validation helps
- Concluding remarks
- Chapters 8 and 9 in IWA OA book: https://iwaponline.com/ebooks/book/835/Metrology-in-Urban-Drainage-and-Stormwater

WHY MONITOR UD SYSTEMS?

- Quantify actual performance vs design performance
- Operational purposes (pump operation, Real Time Control)
- Environmental impact (receiving surface water bodies, leakages)
- Legal and/or accountability reasons (permits, contracts)
- *Assess the technical condition (a.o. CCTV inspection)*
- *To know what is there, where it is and what dimensions it has.*

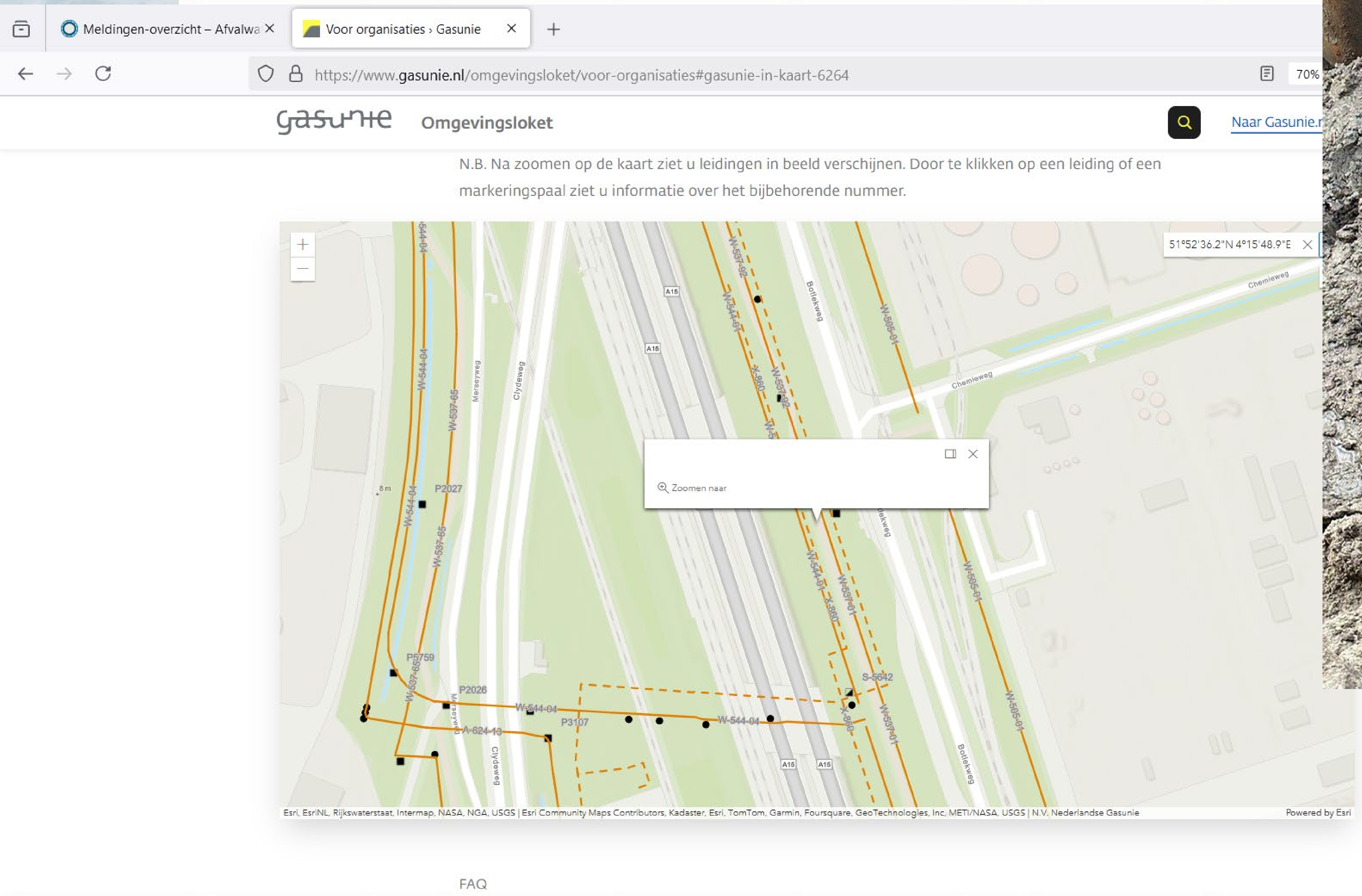
THE PATH BETWEEN DATA AND DECISION



DATA NEED IN URBAN DRAINAGE

- ***Asset Management in Urban Drainage: if anything exciting happens we've done it wrong!*** (IWA Publishing, open access) https://iwaponline.com/ebooks/book/920/Asset-Management-of-Urban-Drainage-SystemsIf

WHAT DO I MANAGE?



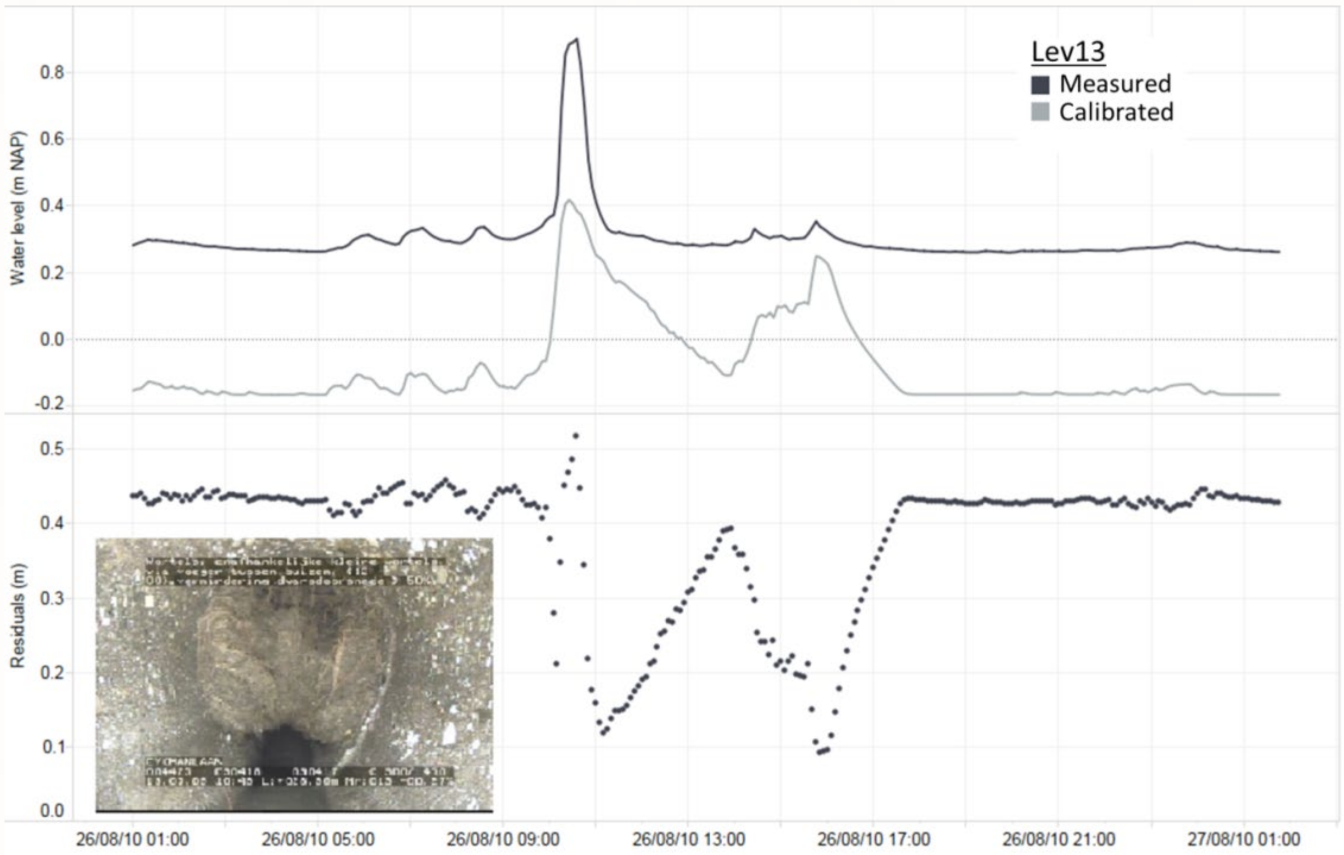
TECHNICAL CONDITION

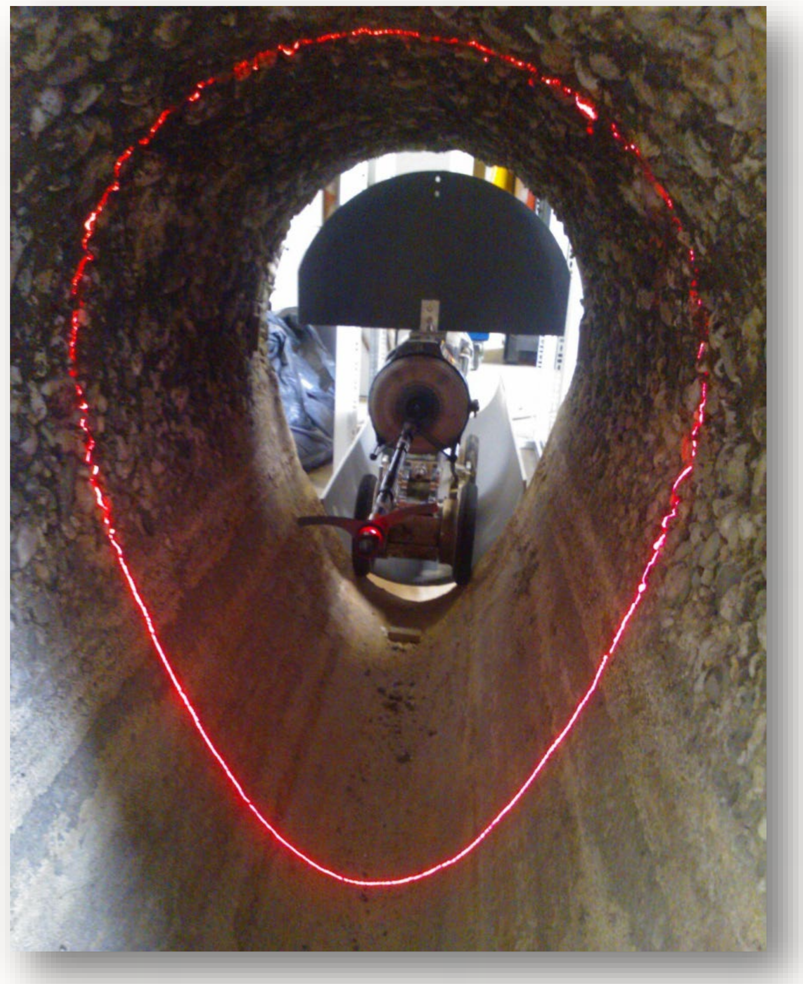
Figure 9.2 Ultra-high-pressure jetting tool with truck and operator controls. *Source*: vandervalk+degroot BV.

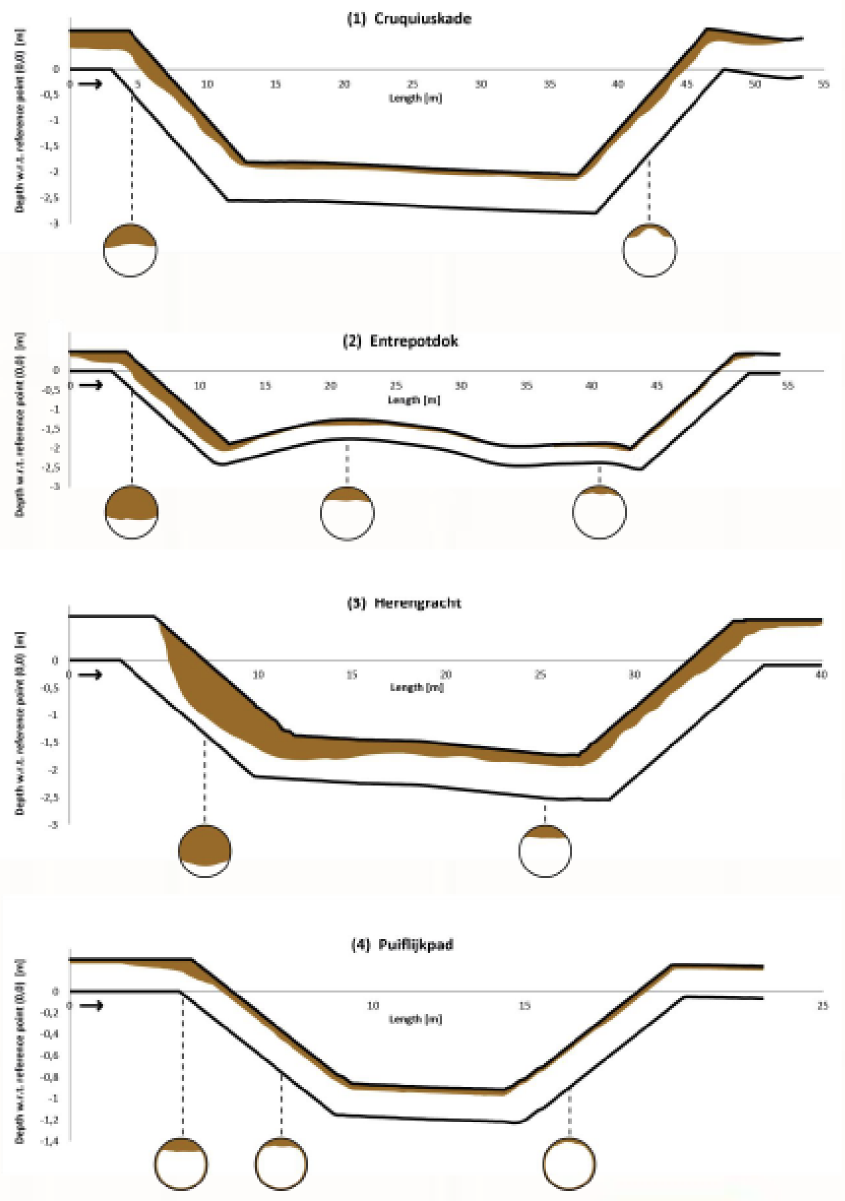
- Decisions for rehabilitation/replacement largely based on CCTV footage + (EU) standard
- On a scale from 0-5 deviations of ~2 classes are present, unless checked twice -> 1 class deviation. (Effect of primitive data validation)
- ~ 25% FP on occurrence of defects (as of ~ 2010)
- **Lessons learnt turned into new developments/insights:**
  - Application of AI image processing ~ 5% FP, no solution for the classification.
  - Standard is too detailed (too high a resolution): it suggests an accuracy that is not feasible (comp. Japanese standard identifying only 3 classes)
  - It measures the ‘wrong’ parameters (no physical quantity is addressed in the standard)

ASSESSING ACTUAL PERFORMANCE

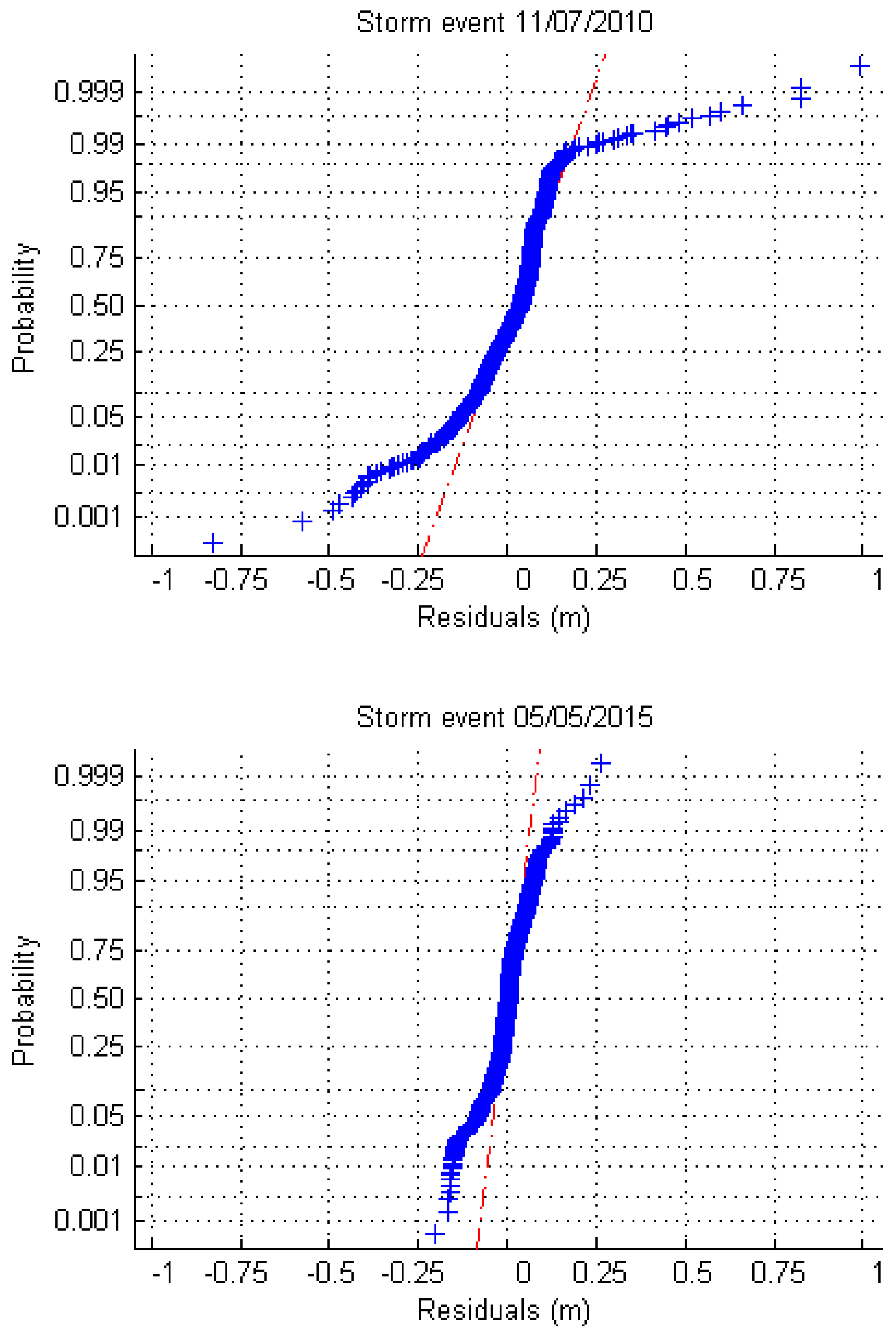
- Protection against flooding/minimising environmental impacts
- Often hydrodynamic models are applied without any validation (not to mention calibration)
- Unexpected issues may influence the actual hydraulic performance that are not taken into account when modelling:

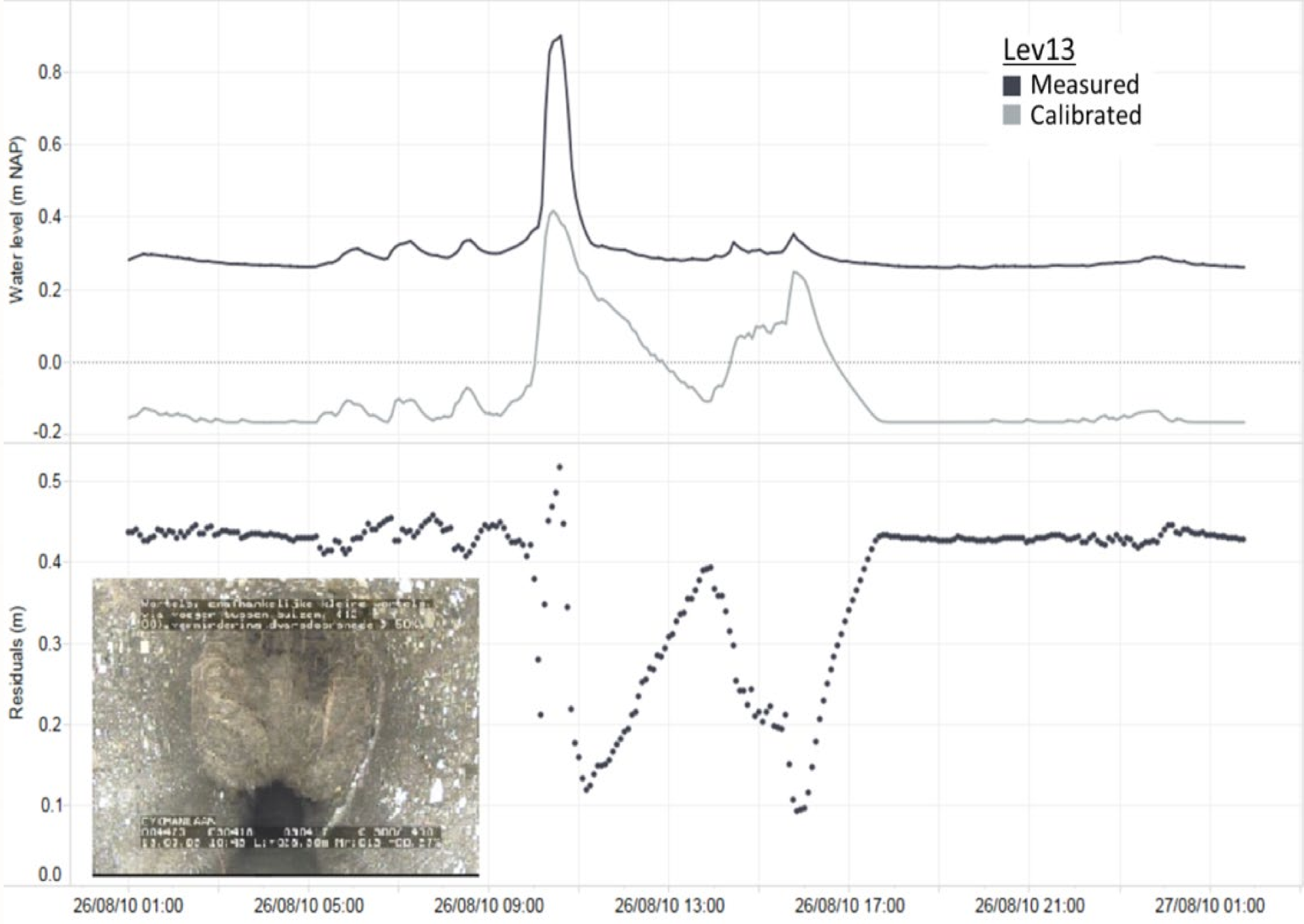




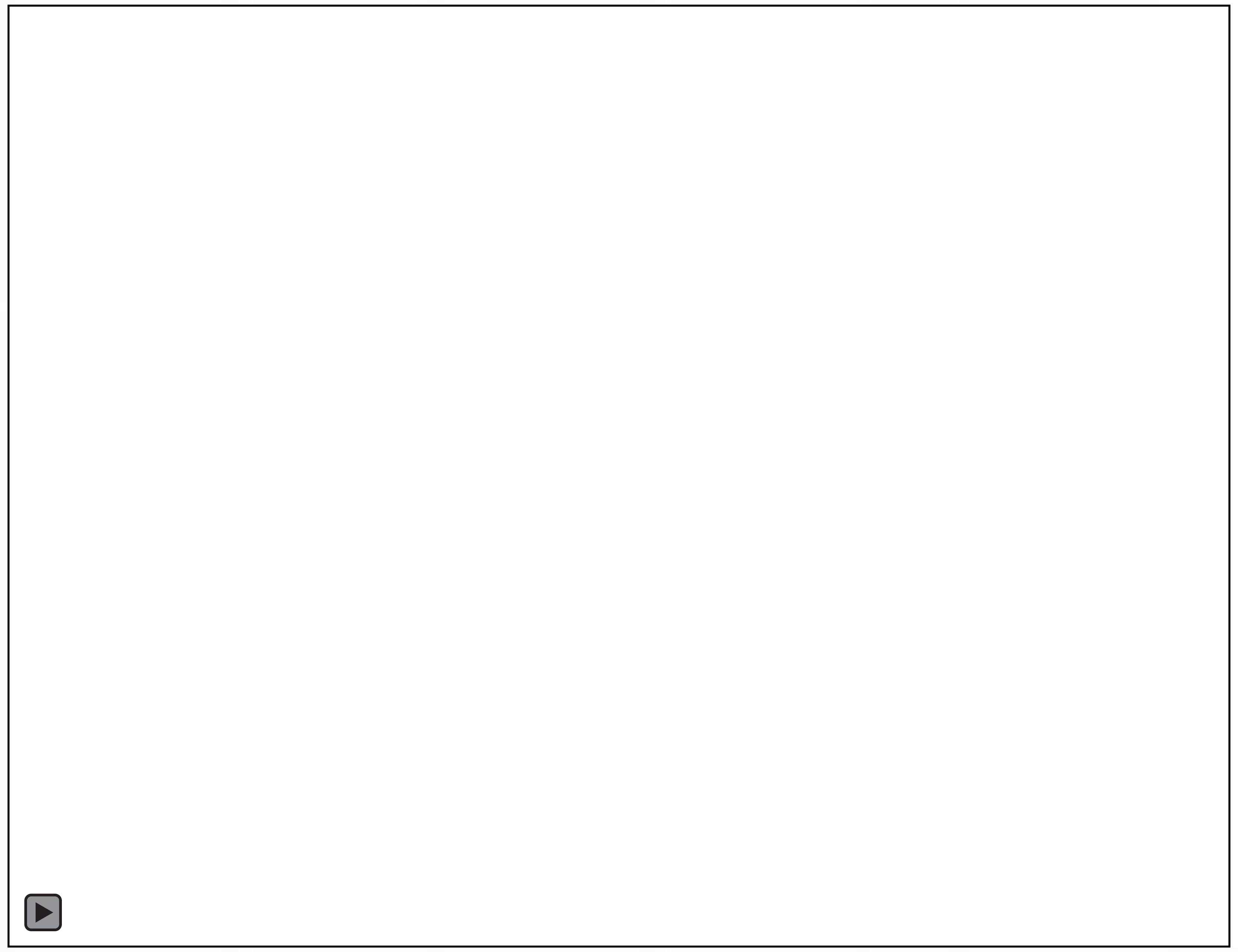


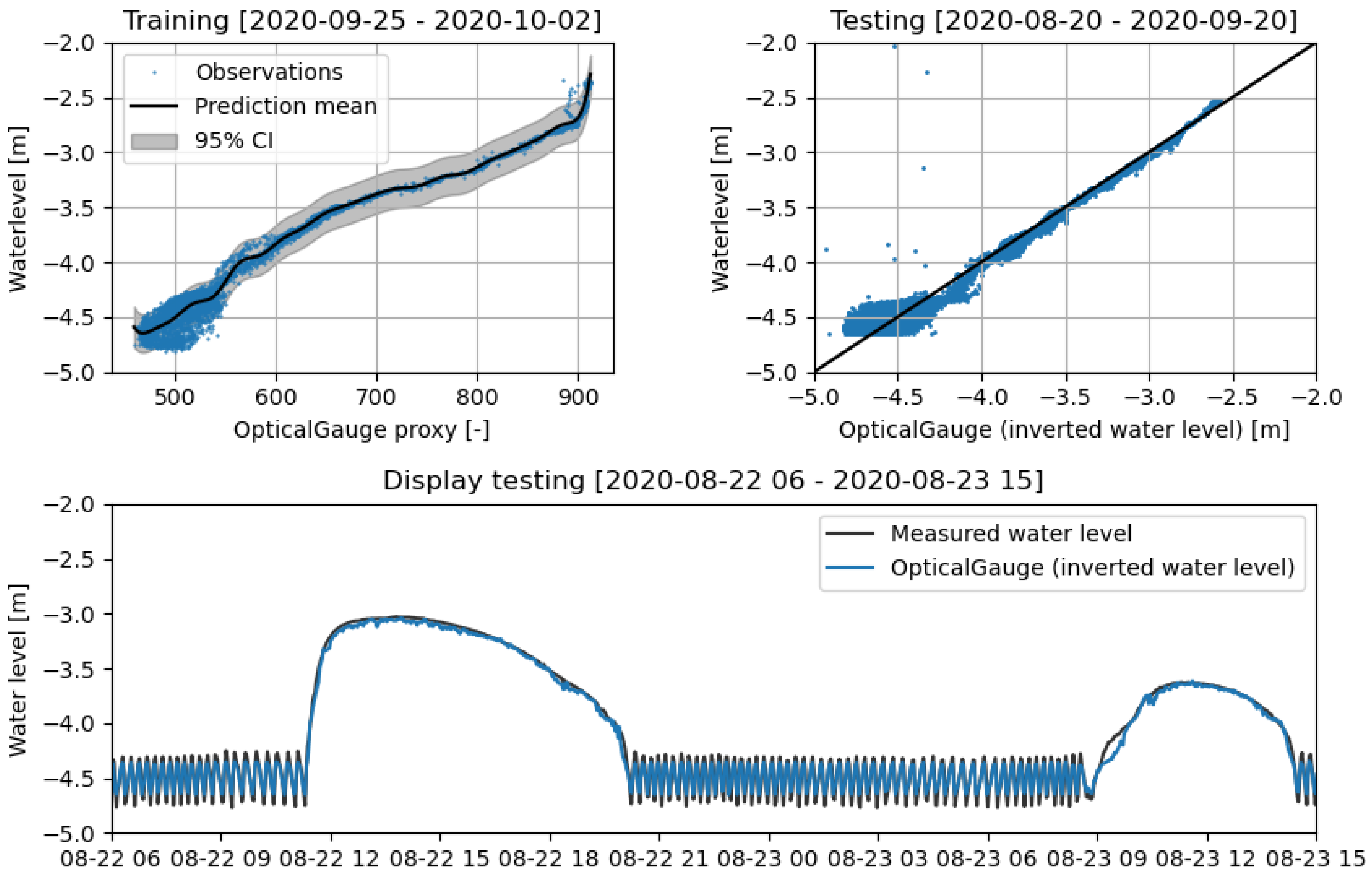
HYDRAULIC FINGERPRINTING





PUMP OPERATION, CHECK ON DESIGN PUMP SUMPS AND WATERLEVEL MEASUREMENT

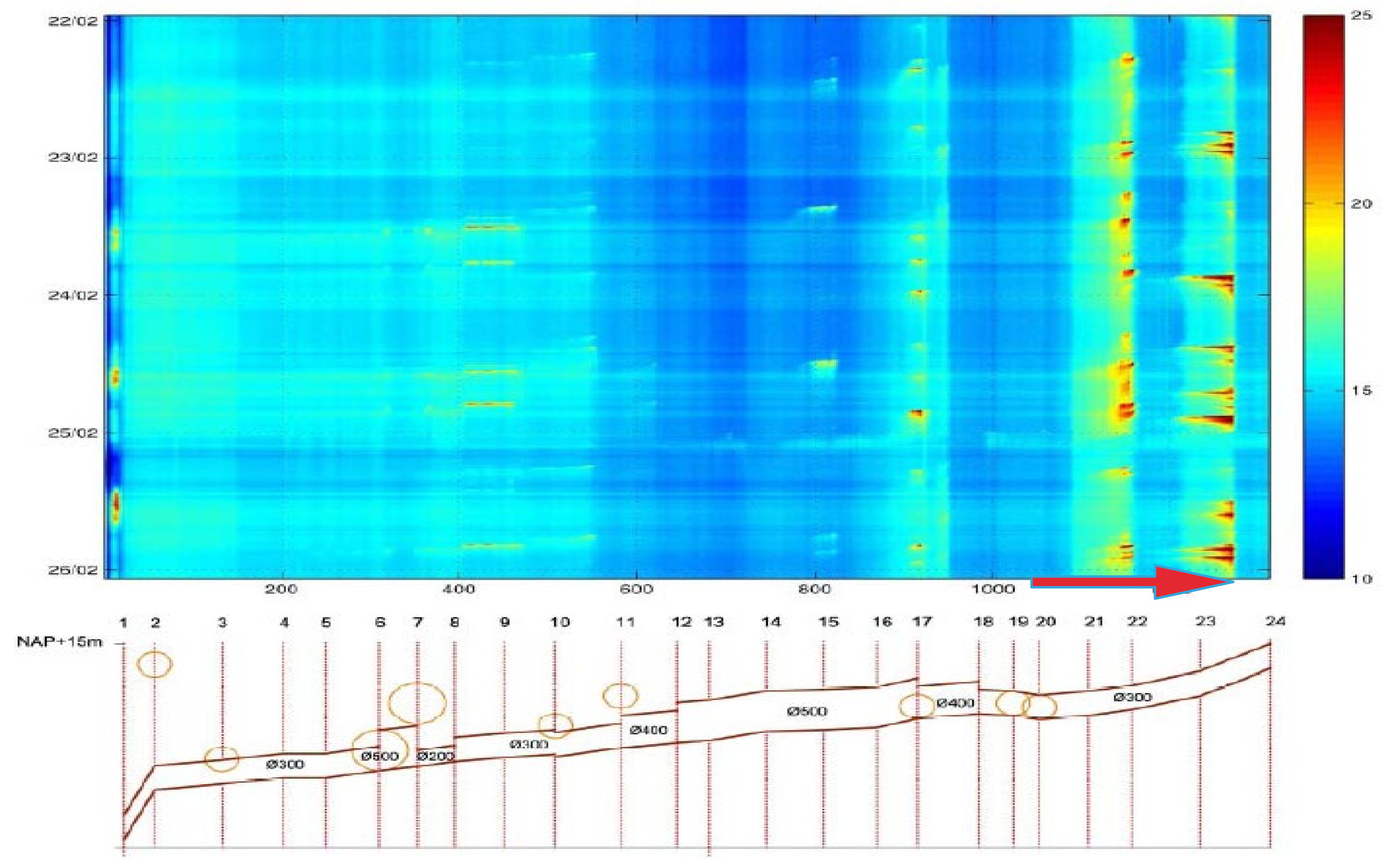




ENVIRONMENTAL IMPACT

- Obviously CSO events serve as **the** default example: monitoring their occurrence can help to identify (and potentially correct for):
  - Design flaws
  - Poor management (e.g. the occurrence of CSO events during dry periods)
  - When combined with observation on the receiving waterbodies the environmental impact can be quantified.
- Monitoring is often related to legislation -> demands put forward are not always realistic (e.g. 95% uncertainty of 5% in yearly CSO volume)

ENVIRONMENTAL IMPACT: MONITORING WRONG (ILLICIT) CONNECTIONS

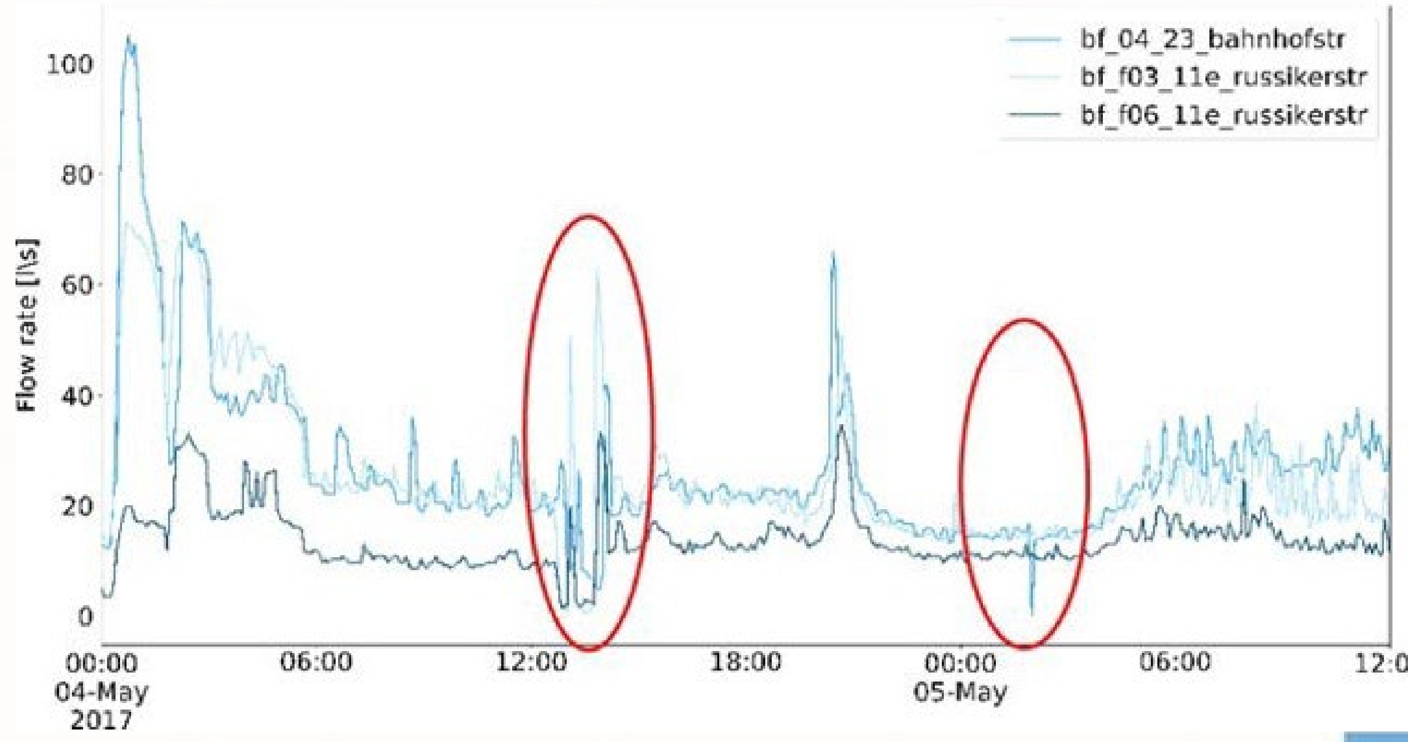


ORIGIN OF IMPERFECTIONS IN MEASUREMENTS

- Many monitoring ***systems*** are sensitive to variations in the conditions in which they operate (e.g. variations in temperature or the presence of EM fields)
- Not to be confused with SYSTEMATIC errors in measurements due to e.g. picking the wrong reference level in waterlevel measurements or setting a wrong scaling factor between e.g. mA and m, these can be avoided by deploying trained personnel
- Vandalism/theft, wrong installation, EM interference, no synchronisation, wrong data on system's dimension, etc. etc.

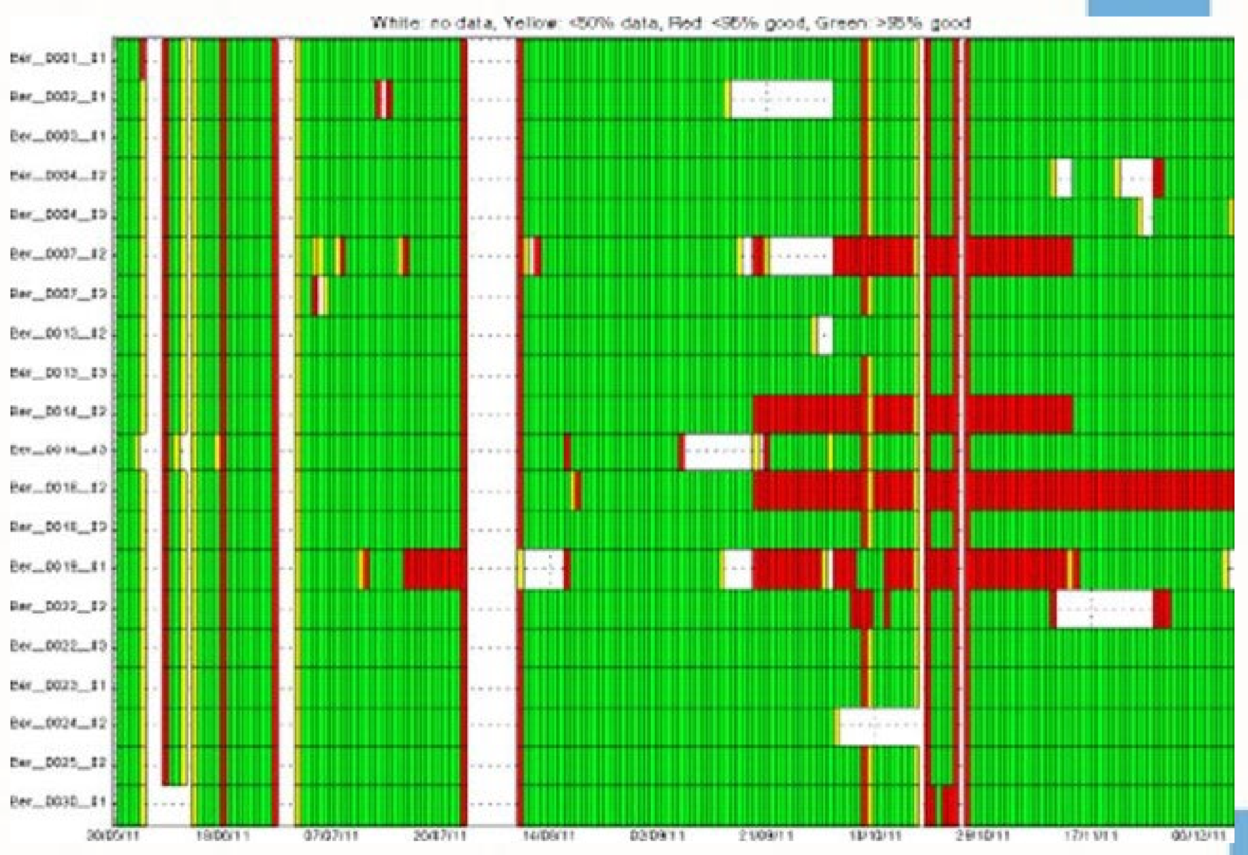
DATA VALIDATION

- Check if obtained (monitoring) data are correct (which is strictly speaking impossible, as this would imply the ‘real’ values are known while they are not.)
- Validation encompasses **the whole monitoring system** (not just the sensor, the data storage, - transmission and possible post processing steps have to be taken into account as well, as each step may introduce some deviation)
  - So retreat to a obtaining a level of plausibility
    - Physical realistic (waterlevel < lowest invert level ?)
    - When compared to other datasources the results are plausible



DATA VALIDATION

- Gives important information on the effectiveness of a monitoring system (e.g. in the 1990’s ~ 50% data yield of waterlevel data was achieved, while at present ~ 99% uptime and ~ 95% correct yield data is feasible)
- A important activity is finding the cause of incorrect/missing data -> ‘detective’ talents come into play!!!!!
- To be taken into account preferably during the design of a monitoring network:
  - Choice of locations so as to allow for cross referencing
  - How to synchronise all measuring stations
  - Choice of applied measuring principles



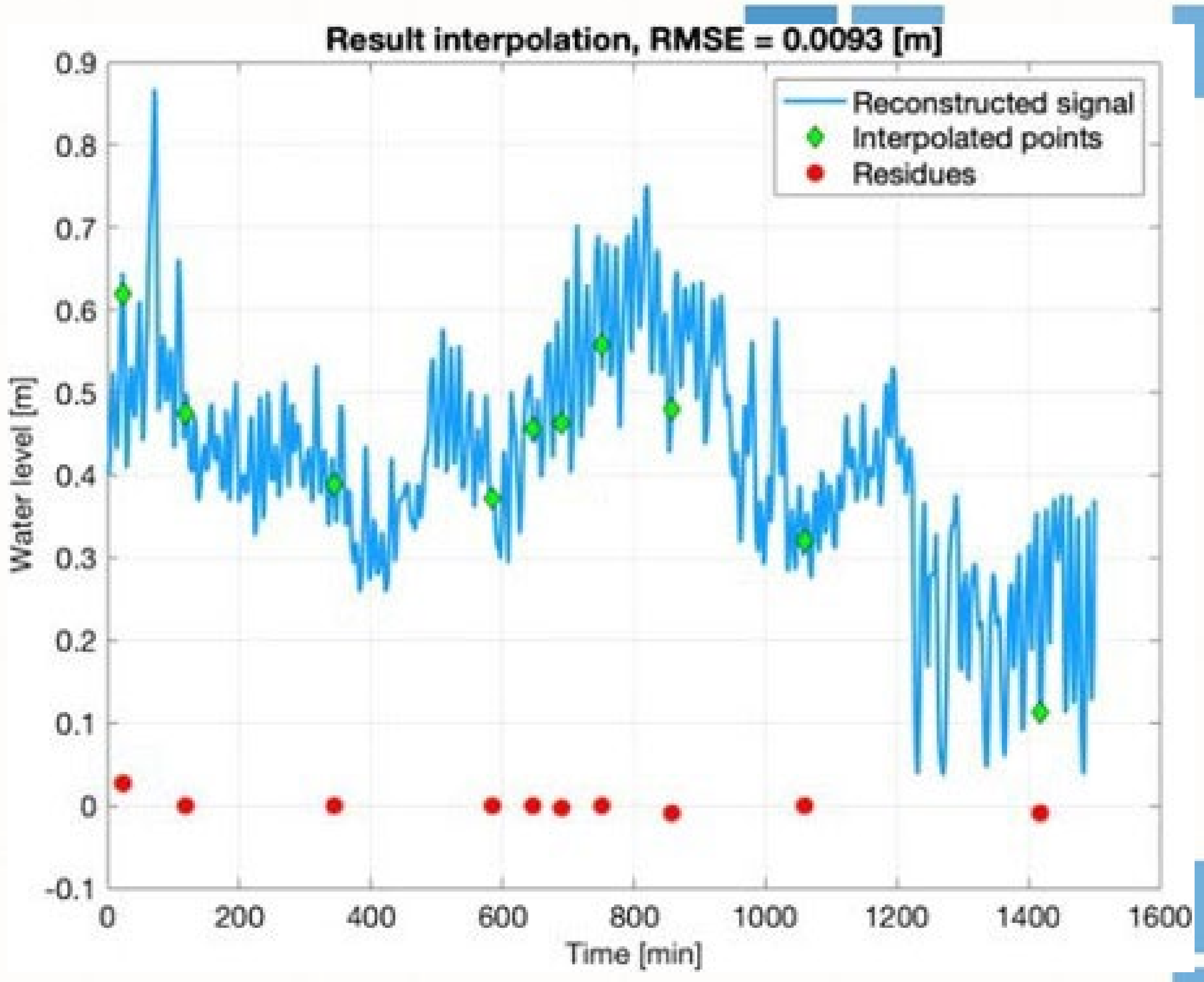
DATA VALIDATION

- There is no formal standard for data validation in UD, benefits of having a would be:
  - Standardised tools for validation -> improved quality on the long term
  - Improved accessibility of monitoring results by third parties
  - A reduction of workload, now ‘the wheel' is being invented over and over again by companies, practitioners and academics.

DATA VALIDATION

- A more or less systematic approach would encompass:
  - Define levels of data quality and data yield strived for depending on the purpose of use
  - When deviations occur to the demanded data quality and - yield, find the causes and cure them
  - Ensure the validation process (and tools) is in place and the personnel is trained to use them well before data start to be produced as data validation is an important tool to finetune the monitoring set-up
  - Regular validation (e.g. every 2-4 weeks for basic checks, but preferably continuously)
  - Ensure to store the original data to avoid effects of software bugs in validation software cannot be repaired afterwards.

CURATING DATA



- What if data points are missing?
  - Straight Interpolation
  - Use a model (either purely mathematical or physically based)
  - Or don’t use the data set……….
  - The choice will depend on the use of the data, in e.g RTC there is no time for a detailed scrutiny, so missing data is normally dealt with in a very practical manner: interpolate (for small data gaps) or fall back on a default (safe) setting of the system
  - **In any case: label the data quality**, as they may be fit for use for one purpose but not for another (see chapter 9)

CONCLUDING REMARKS

- Large scale monitoring has become a possibility over the last decade, due to robust sensor technology, IT technology and the availability of cheap components making it 'look easy' which may be a deceiving perspective.

- The need for monitoring is increasing as we are faced with (rapidly) changing conditions (climate, increase in urban population), combined with heavier demands on functionalities this asks for careful and often costly redesign of existing systems and the evaluation of new concepts that are being deployed.

- Using the notion of the presence of uncertainty/errors in monitoring data is essential to avoid making wrong decisions in (re)design, operation and law enforcement pertaining to UD systems.

CONCLUDING REMARKS

- Data validation asks for:
  - Awareness of and acknowledging the limitations of measuring systems and the occurrence of (human) error, communication issues, software bugs.
  - Knowledge of the system under scrutiny.
  - A sound understanding of the relevant processes influencing the monitored parameters.
- Datavalidation serves to:
  - Enhance the quality of **all data** involved in UD management
  - Reduce damage/hindrance due to malfunctioning, operational mishaps, damage due to digging/tunnelling activities

CONCLUDING REMARKS

- Organisations managing UD systems will need to incorporate the knowledge and the means (personnel) to perform monitoring on the required level. *Monitoring is a specialism that asks for attention.*
- Adapting a systematic approach as e.g. described in 'Metrology in Urban drainage and stormwater management: plug & pray,' may be beneficial for a plethora of reasons.
- Developments in AI show promising progress (e.g. image processing has largely solved the human interpretation issues in visual inspection), still human supervision when performing data validation and/or -curation will be needed (I guess).

Co-UDlabs

BUILDING COLLABORATIVE URBAN DRAINAGE RESEARCH LABS COMMUNITIES

DATA VALIDATION
with the Urban Drainage Metrology Toolbox

Webinar, 10 January 2025

Jean-Luc BERTRAND-KRAJEWSKI (INSA Lyon)
Francois CLEMENS-MEYER (Skillsinmotion)

ENJOY

PRESENTATION OF THE UDMT
- URBAN DRAINAGE METROLOGY TOOLBOX -
DEVELOPED IN THE Co-UDlabs PROJECT

Jean-Luc BERTRAND-KRAJEWSKI
(INSA Lyon)

PRESENTATION OF THE UDMT

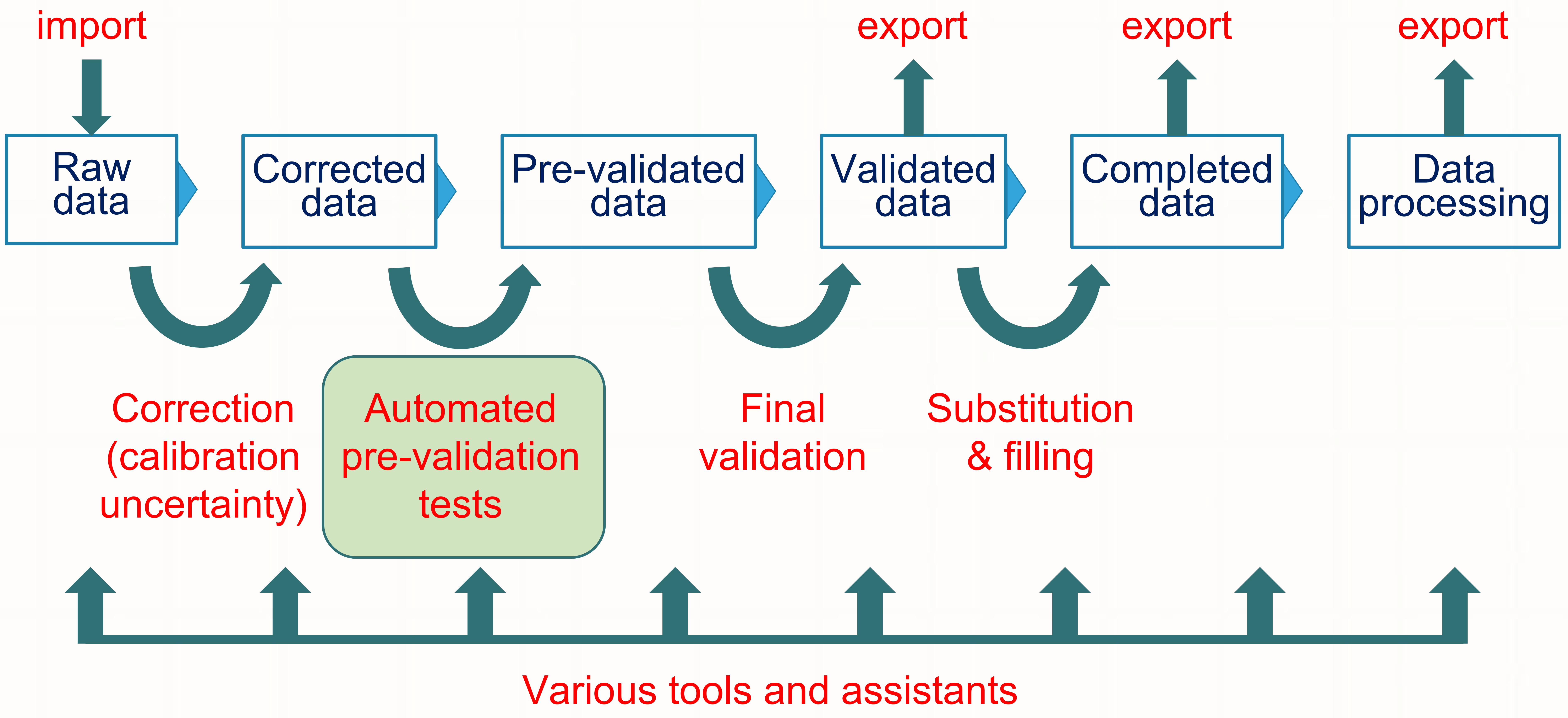
A FREE TOOLBOX FOR EVERYONE

- A free WebApp accessible by everyone (on- and off-line)
- Adress : www.coudlabs.alisonen.com
- User manual, training files and off-line version available
- For the moment : in English, French and Spanish languages (German in Spring 2025)
- Both versions updated until end of the project (April 2025)
- CSV files required, according to templates described in the user manual
- Comments, remarks and suggestions/ to be sent to UrbanDrainageMetrologyToolbox@gmail.com

UDMT – URBAN DRAINAGE METROLOGY TOOLBOX

- User interface as simple as possible, and as complete as necessary
- For both researchers and practitioners
- Free access, no registration (anonymous users)
- No user data stored on the virtual machine (up- and download)
- All results accessible by the user (csv files)
- Access on the cloud to
  - UDMT User manual
  - Examples data files
  - Repository of source Matlab codes (in prep., end of Co-UDlabs by April 2025)

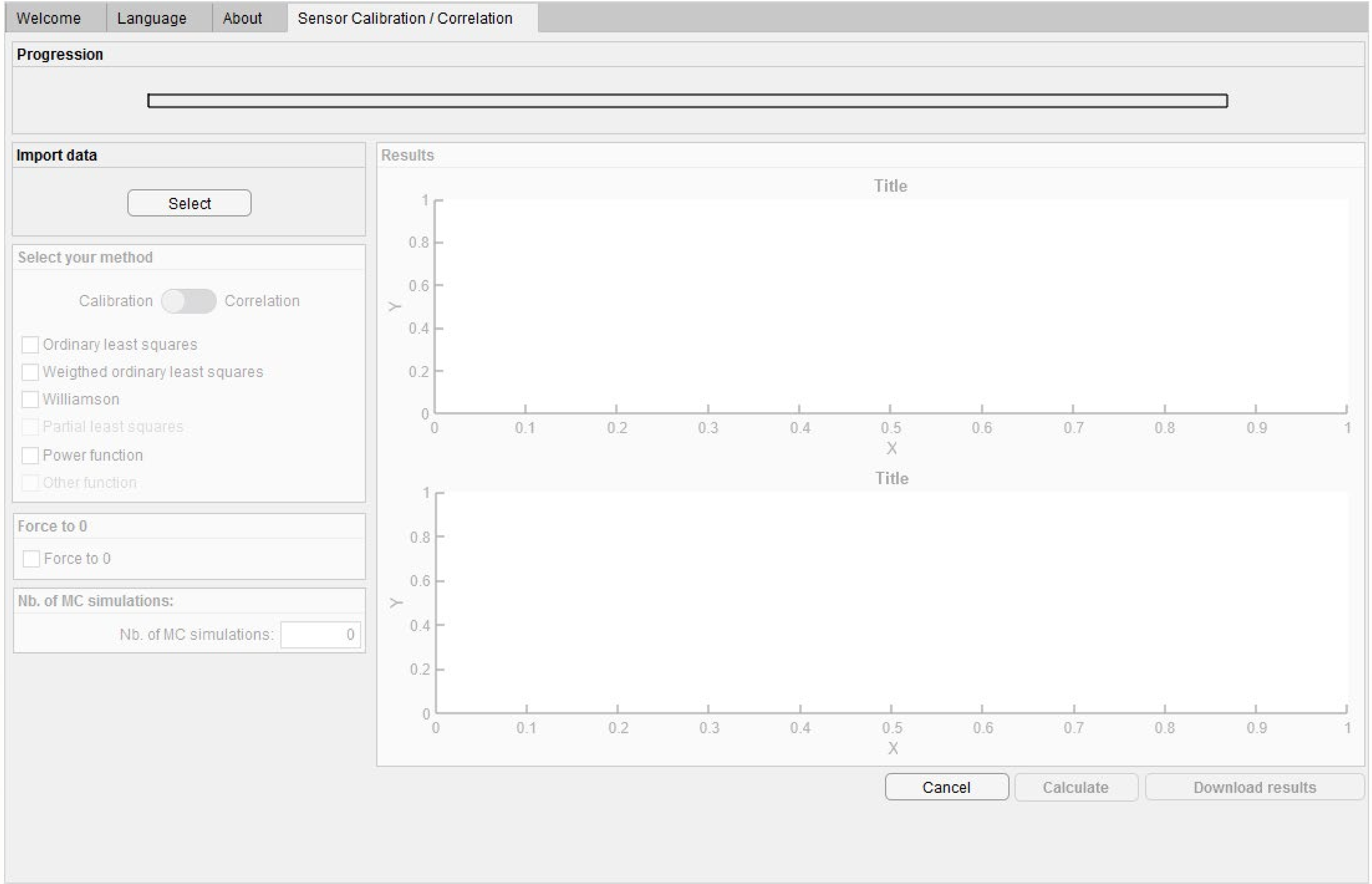
RELATION TO METROLOGY BEST PRACTICE



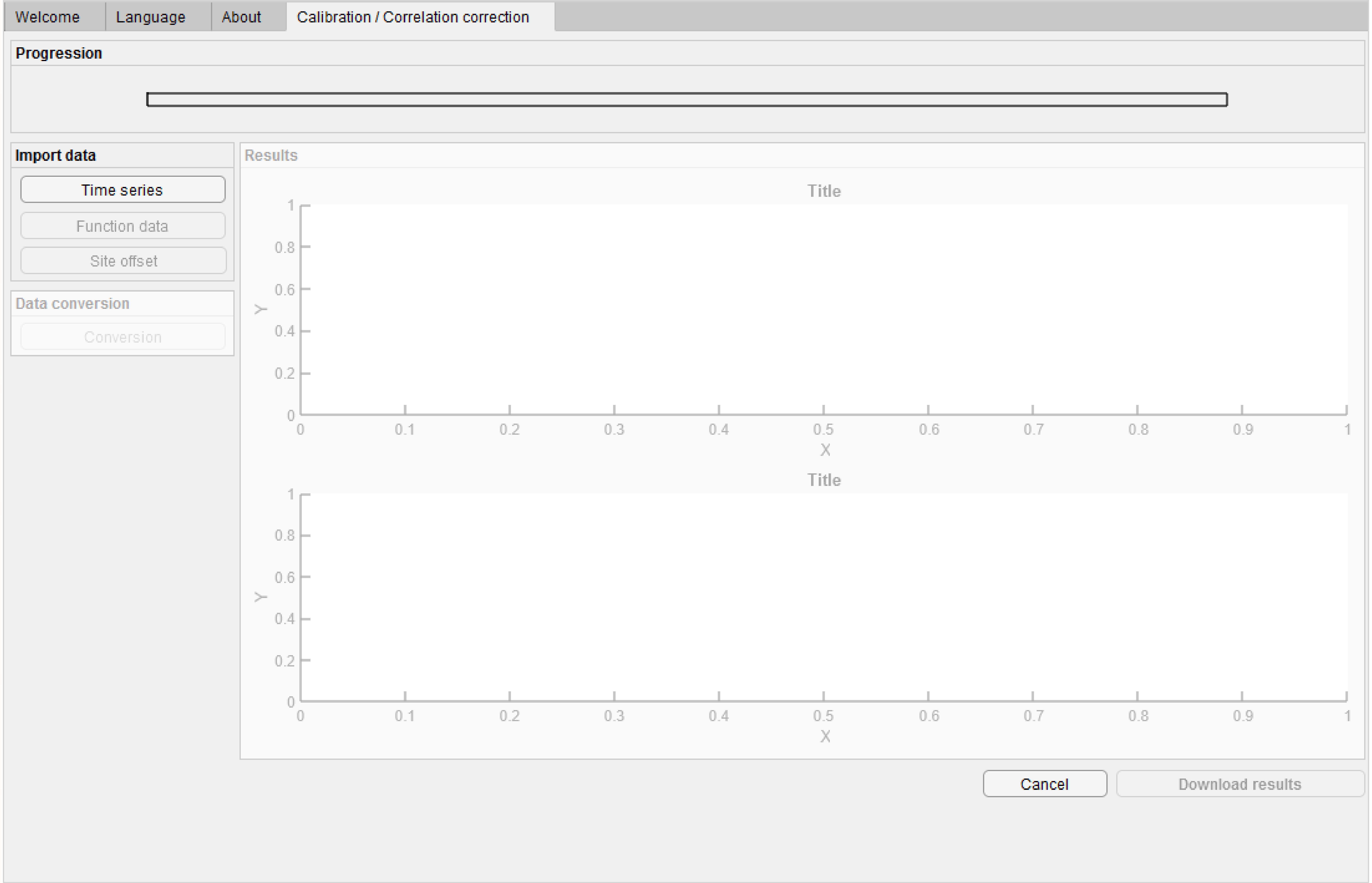
UDMT – URBAN DRAINAGE METROLOGY TOOLBOX

- Facilitating the application of metrology best practice in UD
  - Sensor calibration (with standards)
  - Sensor correlation (with traditional analyses)
  - Data correction from calibration or correlation functions
  - Uncertainty assessment
  - Data validation
  - Tracing experiments for flowmeter qualification

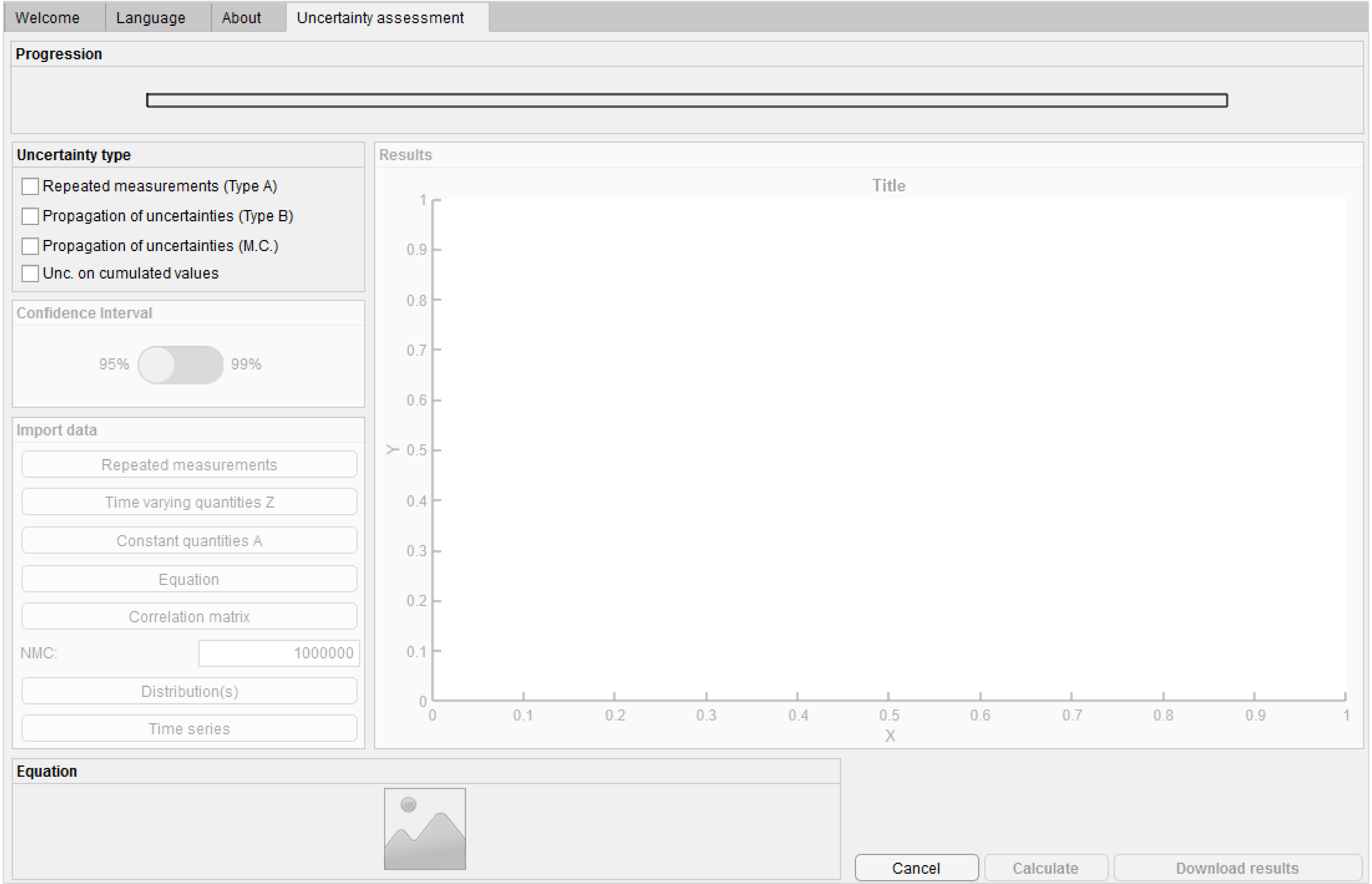
SENSOR CALIBRATION / CORRELATION



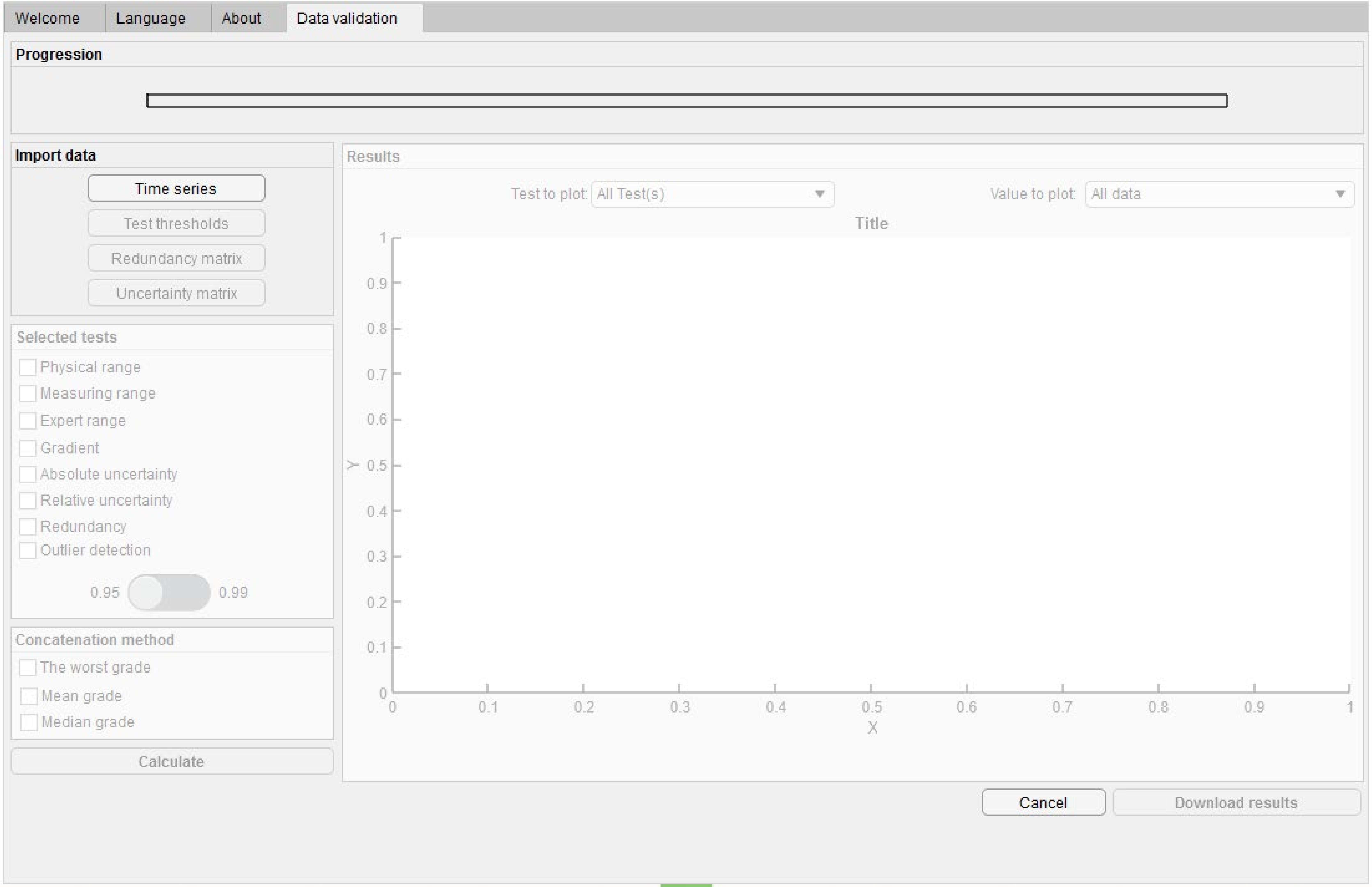
CALIBRATION / CORRELATION CORRECTION



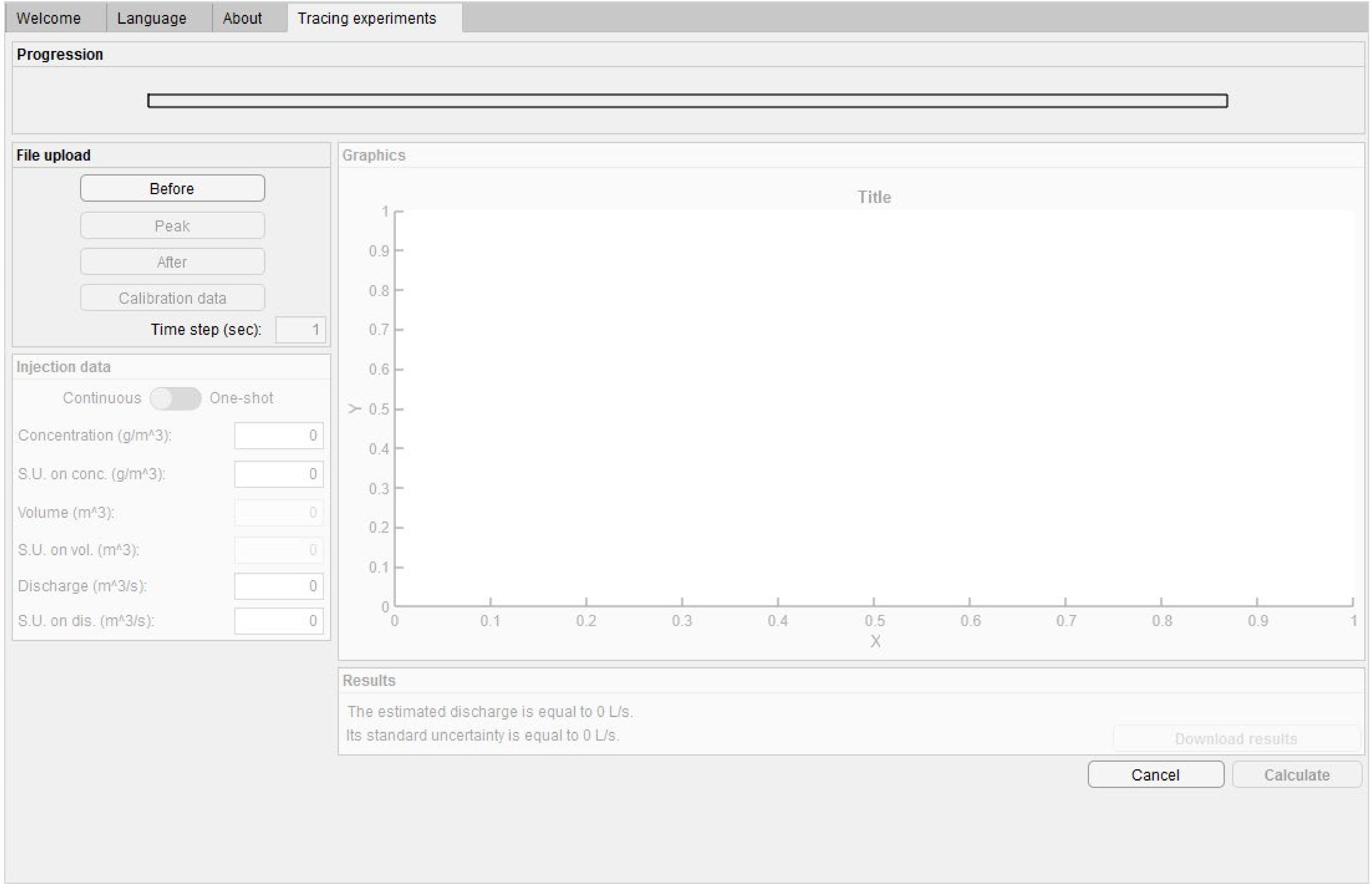
UNCERTAINTY ASSESSMENT



DATA VALIDATION



TRACING EXPERIMENTS



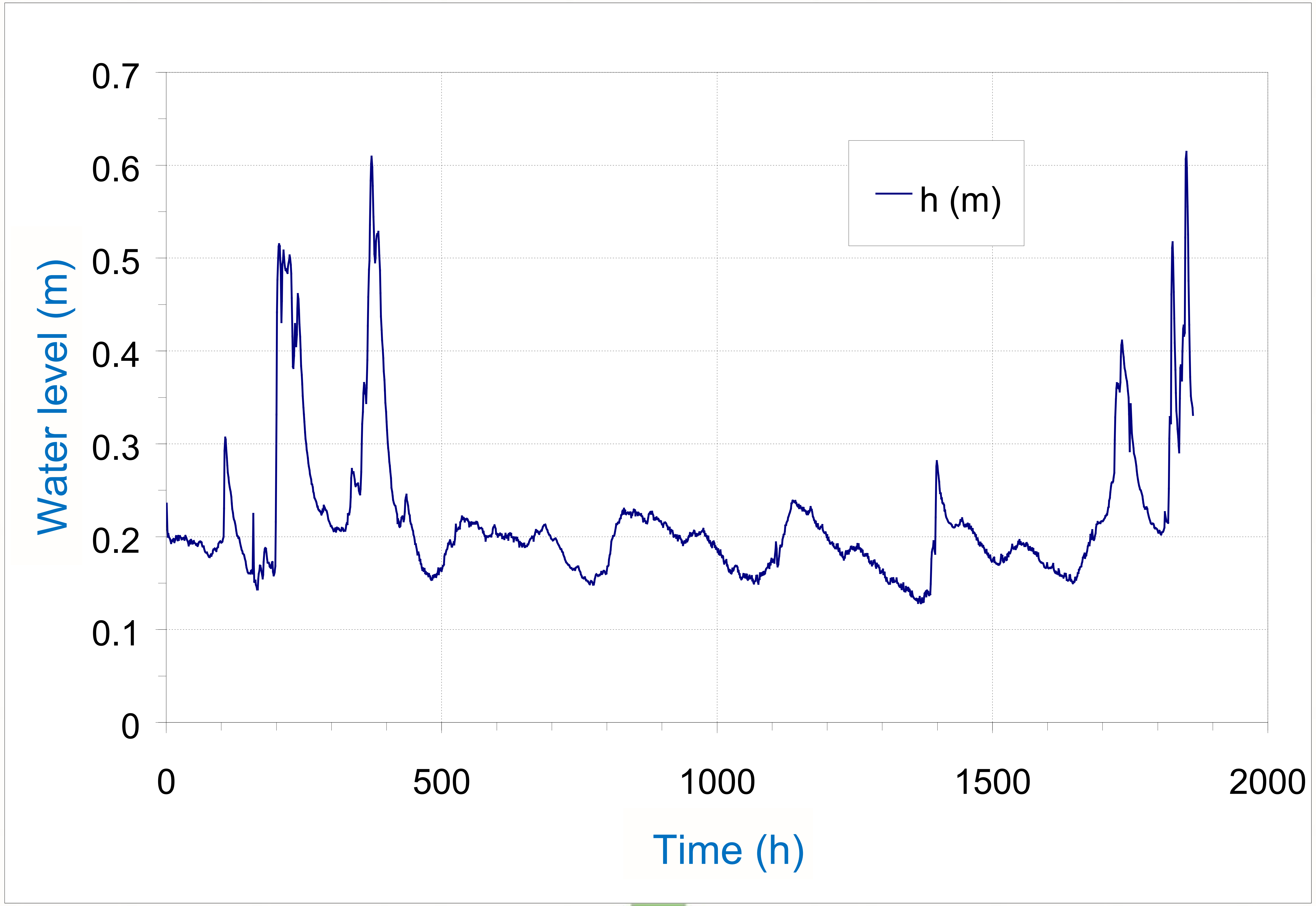
DATA VALIDATION WITH THE UDMT

Jean-Luc BERTRAND-KRAJEWSKI
(INSA Lyon)

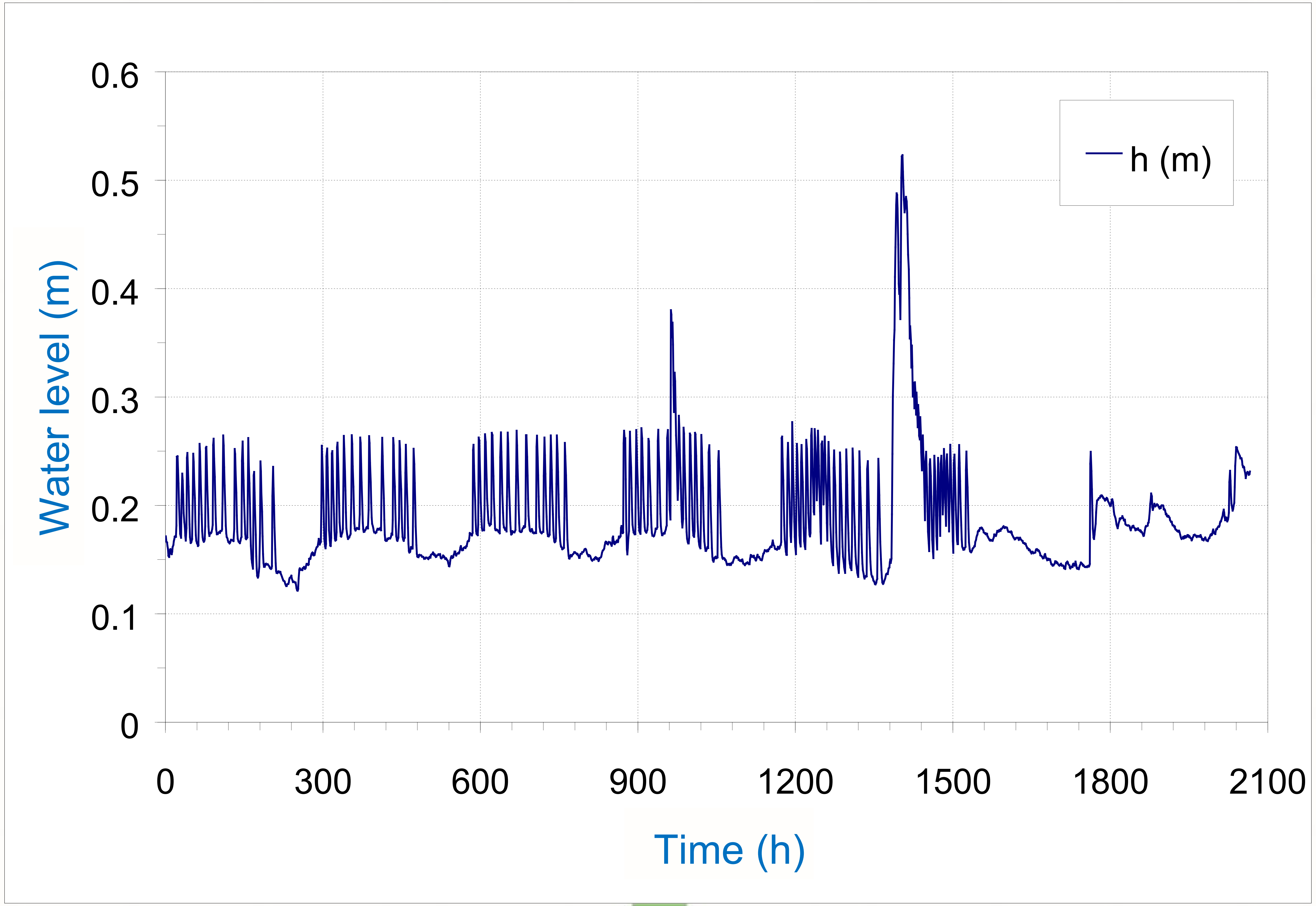
ASSISTED DATA VALIDATION

- Necessary
  - Large amount of data
  - Avoid operator's subjectivity
- Main steps
  - Pre-validation: automatic detection of abnormal / doubtful data
  - Final validation: confirmation or correction of pre-validation outputs

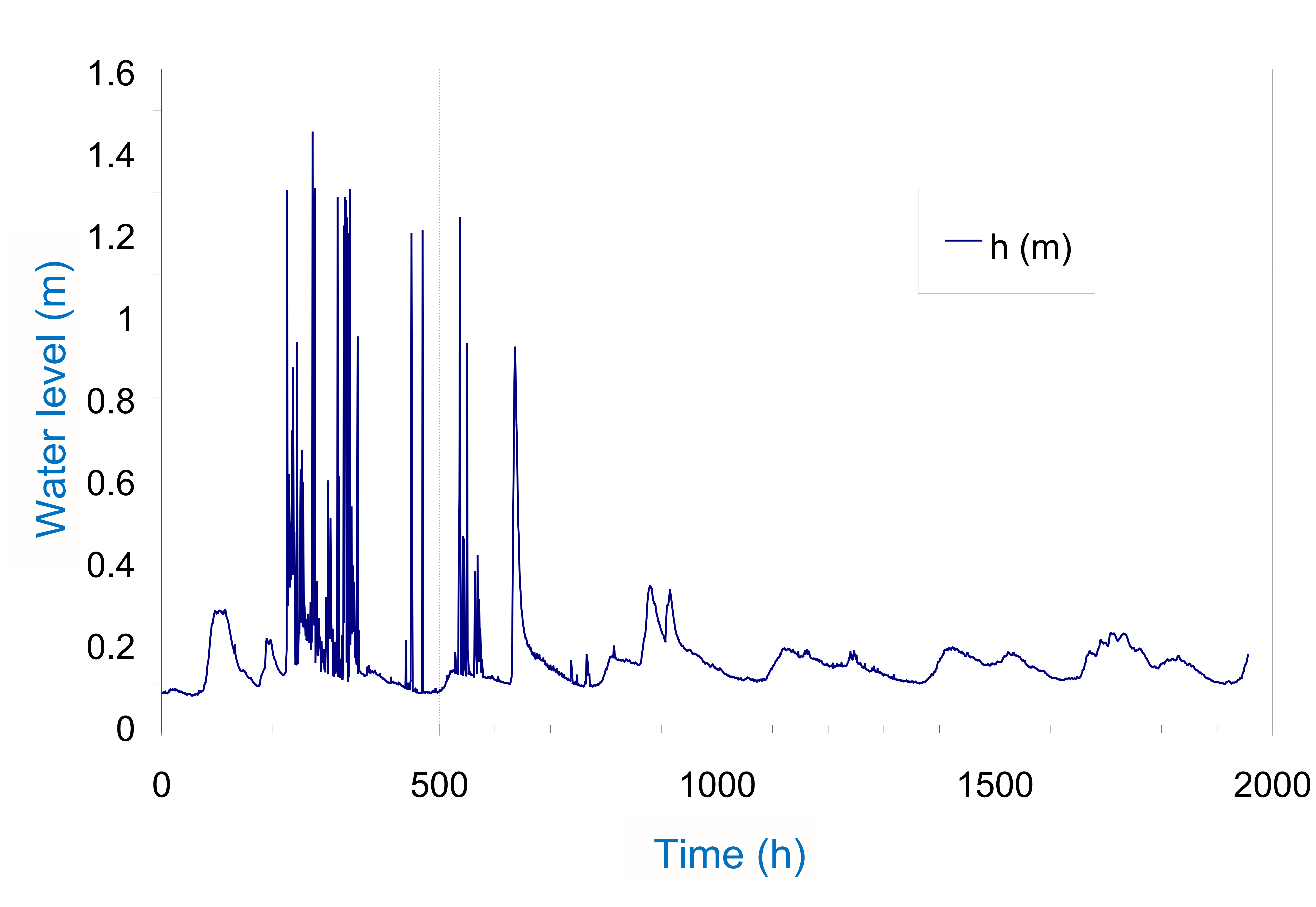
NORMAL VALUES



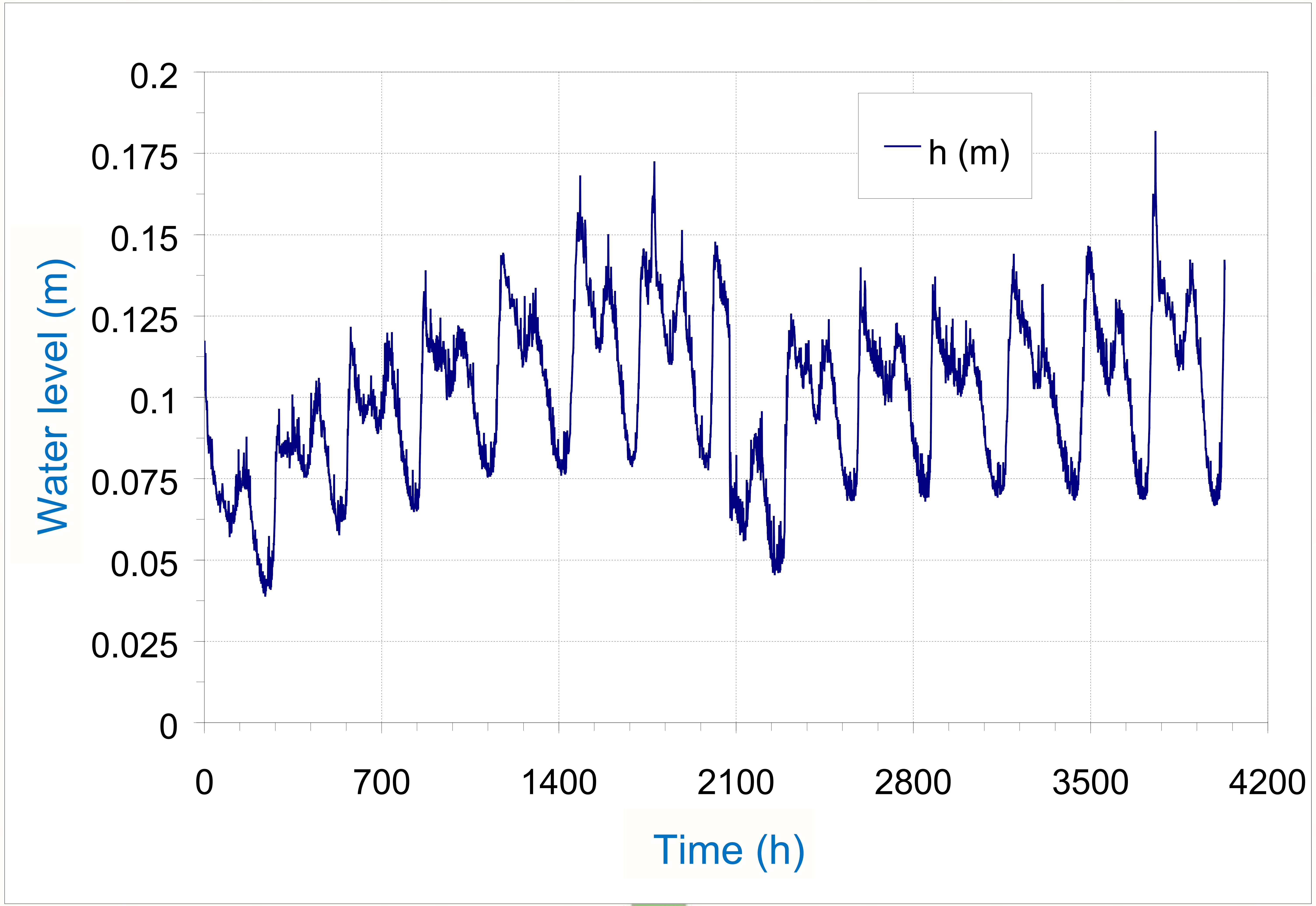
ABNORMAL BEHAVIOUR



ERRATIC BEHAVIOUR



SENSOR DRIFT



AUTOMATIC PRE-VALIDATION

- Pre-validation mark
  - 1 : satisfactory value
  - 2 : doubtful value
  - 3 : false or outlier (rejected) value

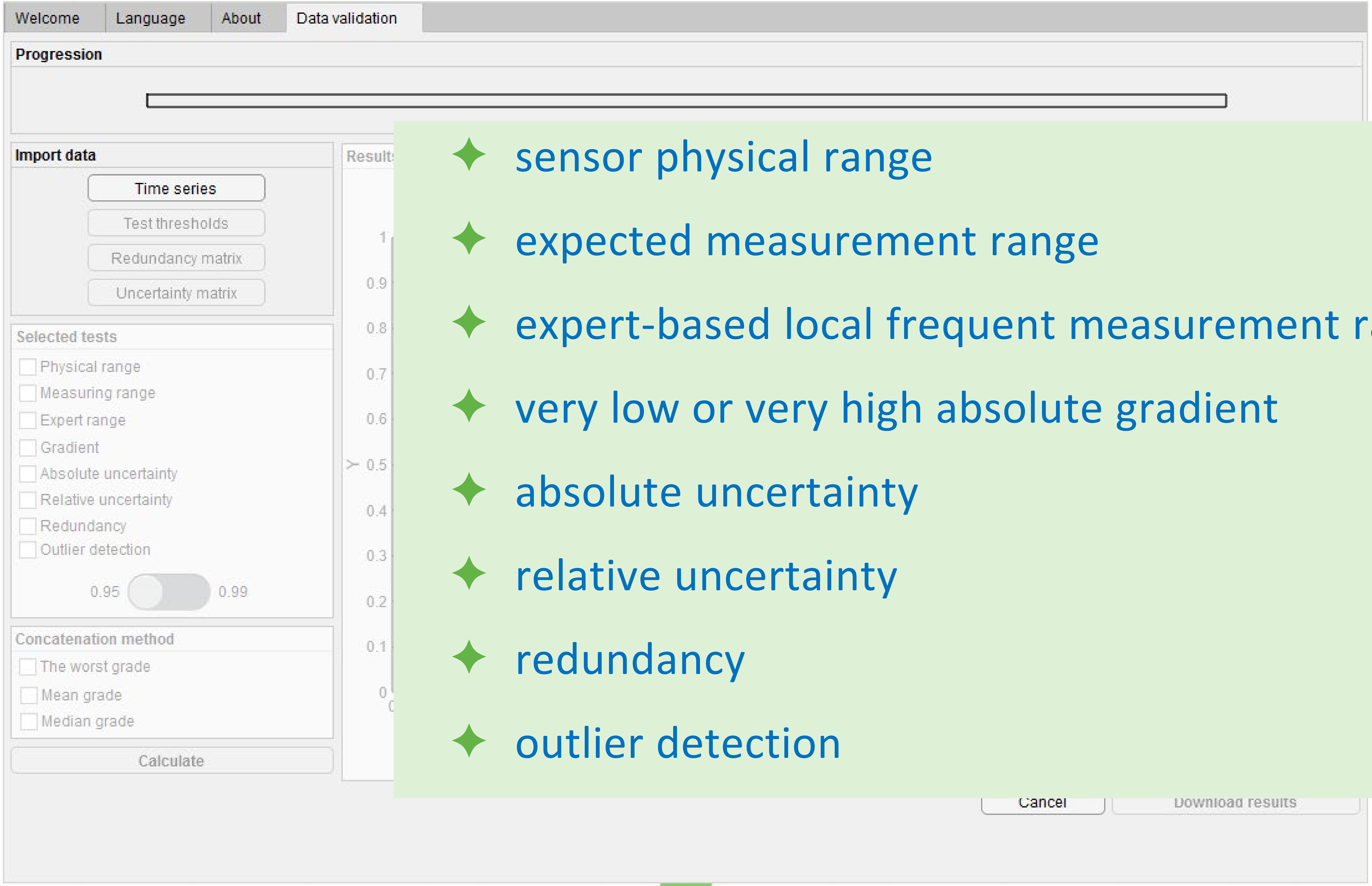
- Final validation mark
  - 1 : valid value
  - 3 : non-valid (rejected) value

TRACEABILITY

IS OF

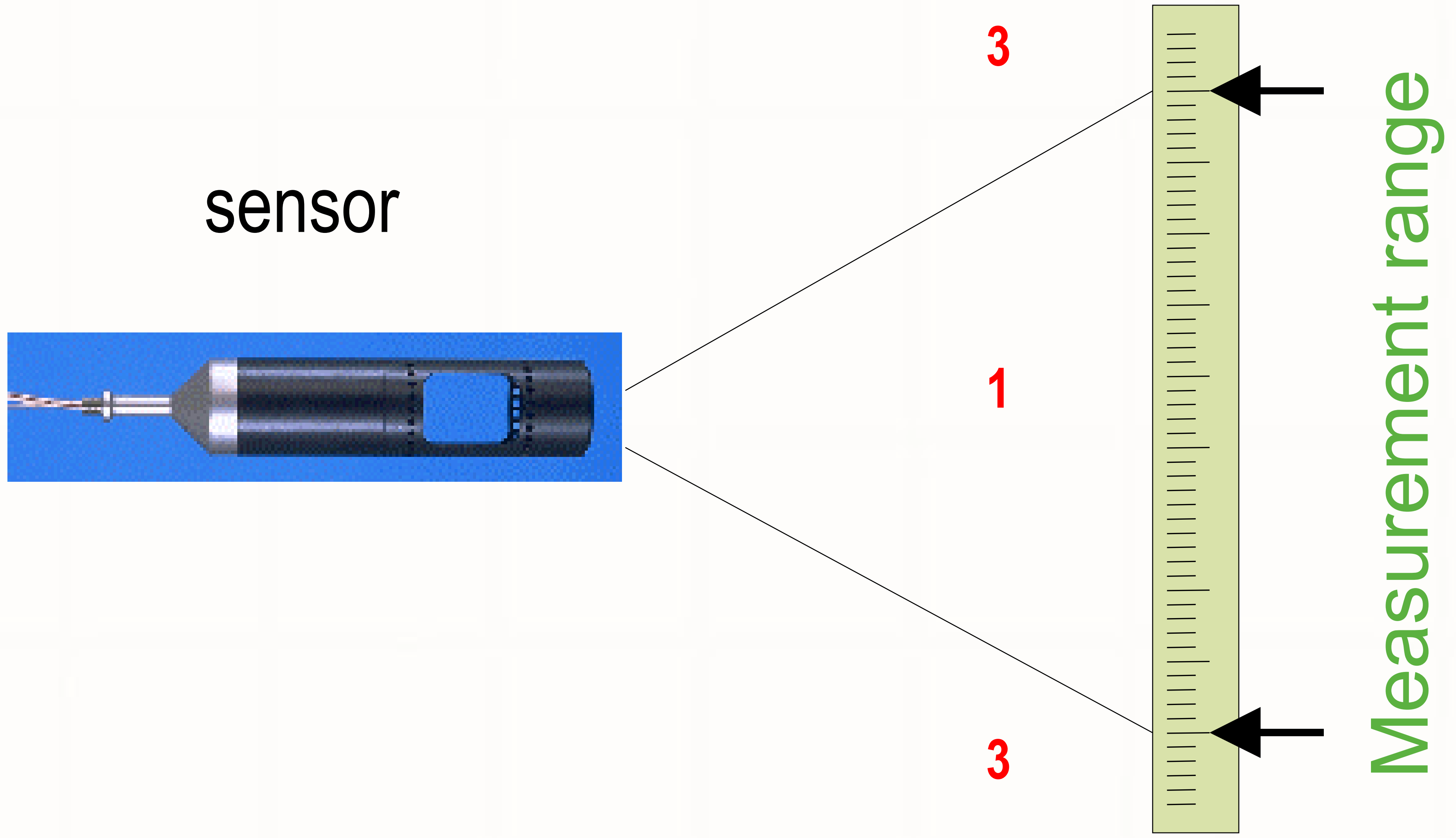
PARAMOUNT IMPORTANCE

8 BASIC PRE-VALIDATION TESTS

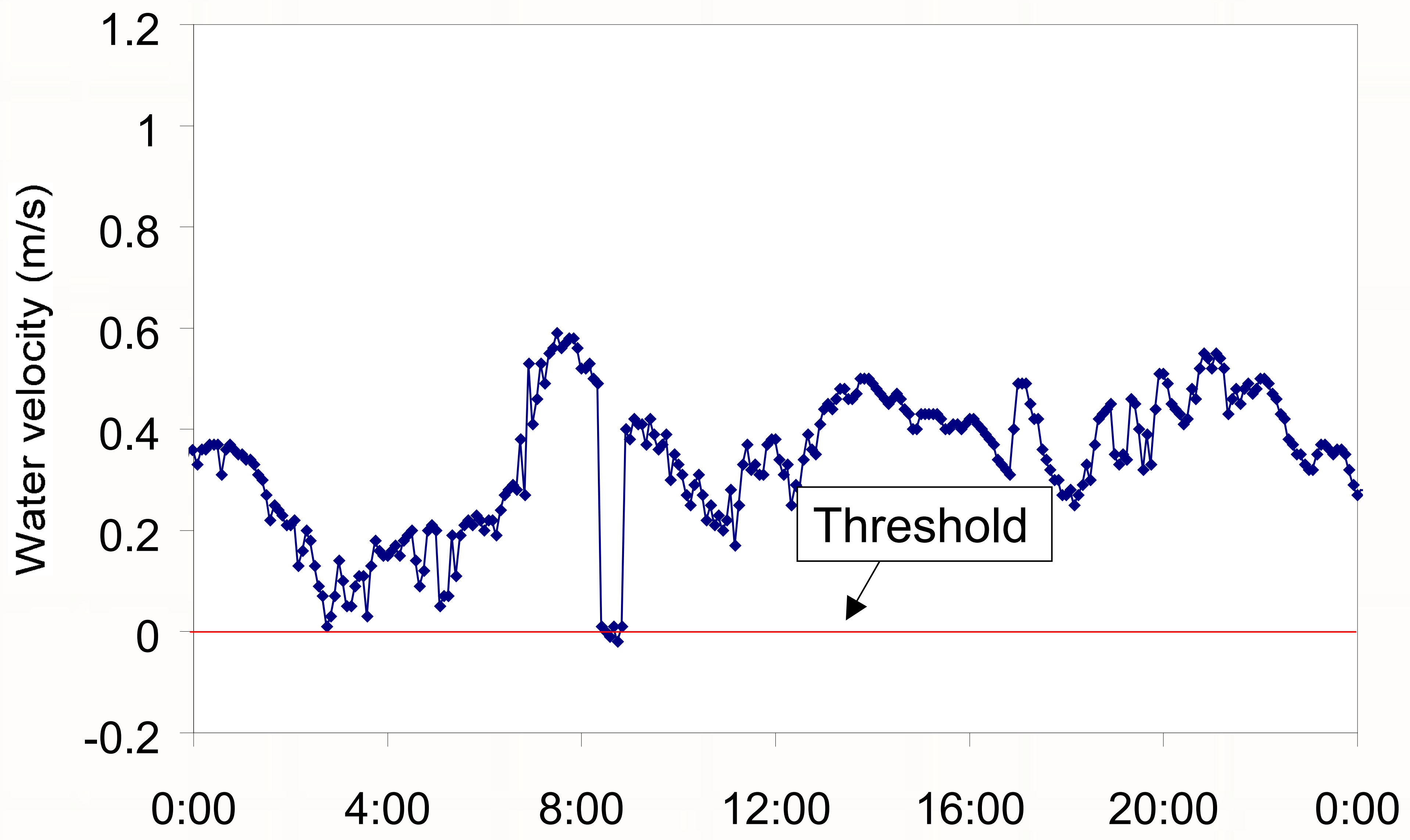


- sensor physical range
- expected measurement range
- expert-based local frequent measurement range
- very low or very high absolute gradient
- absolute uncertainty
- relative uncertainty
- redundancy
- outlier detection

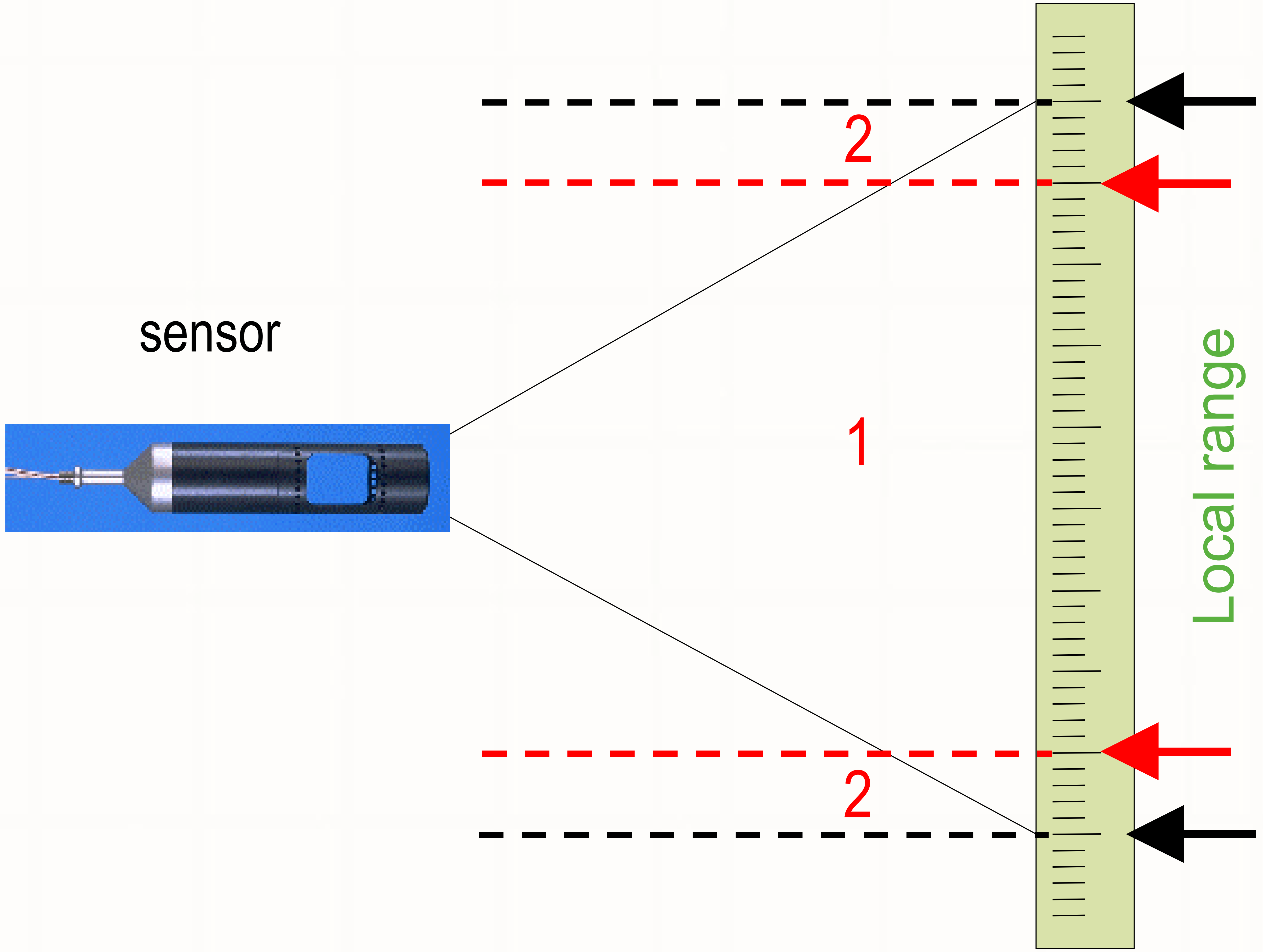
SENSOR PHYSICAL RANGE



Examples:
- water level 0-2 m
- pH from 0 to 14



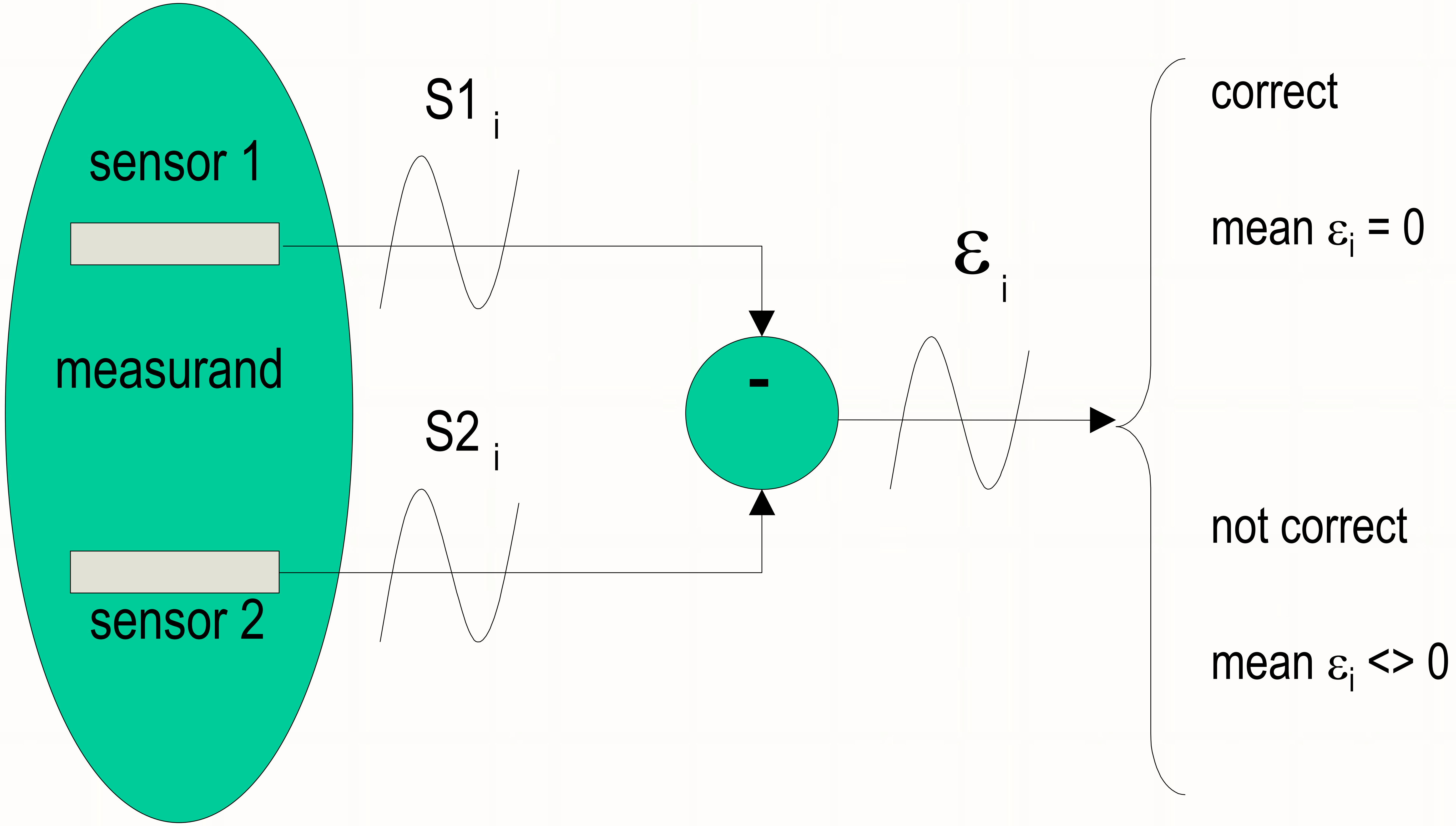
LOCAL FREQUENT MEASUREMENT RANGE



Example:

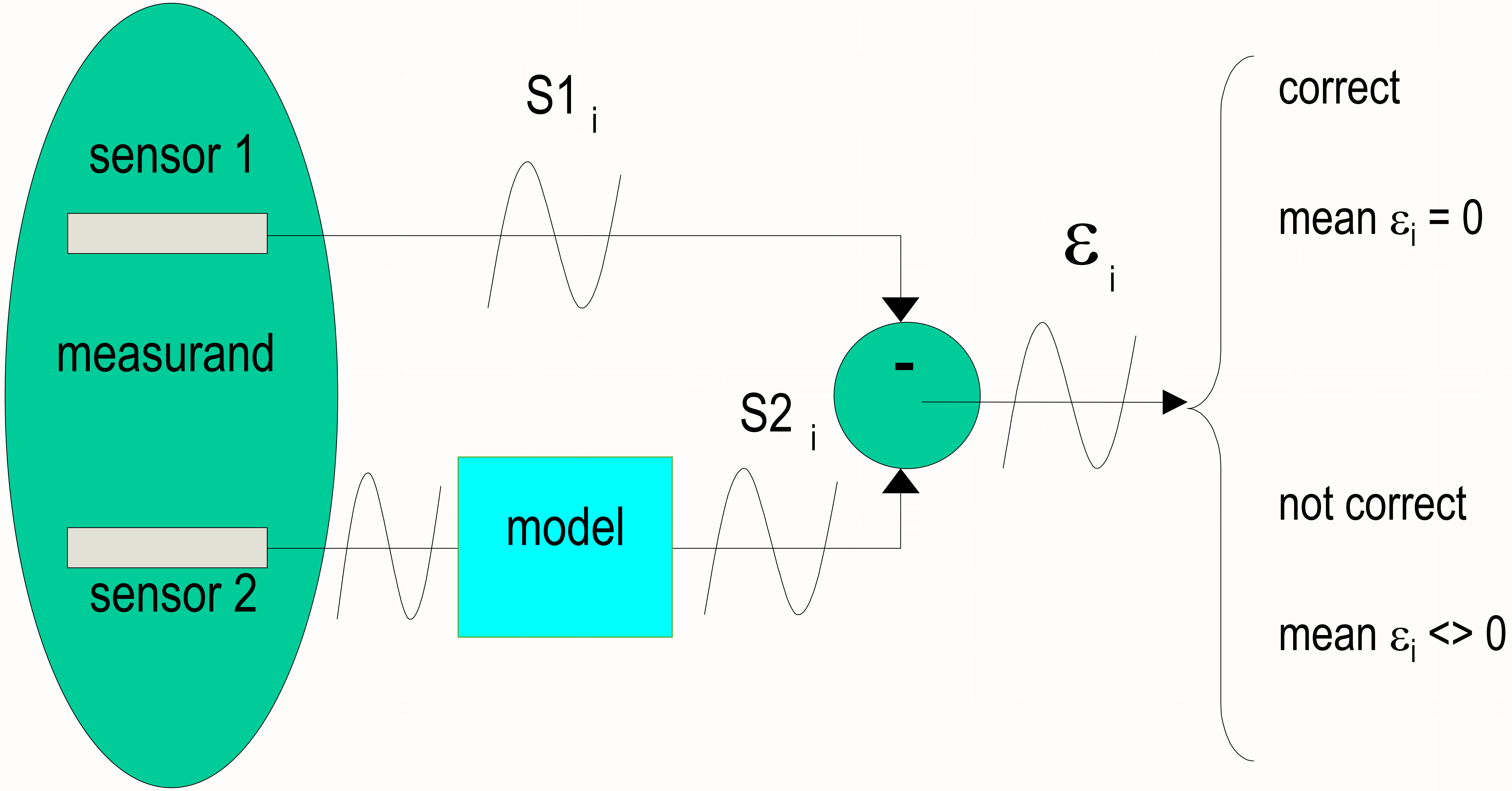
95 % of pH values in this location are expected between 6 and 8

SENSOR REDUNDANCY



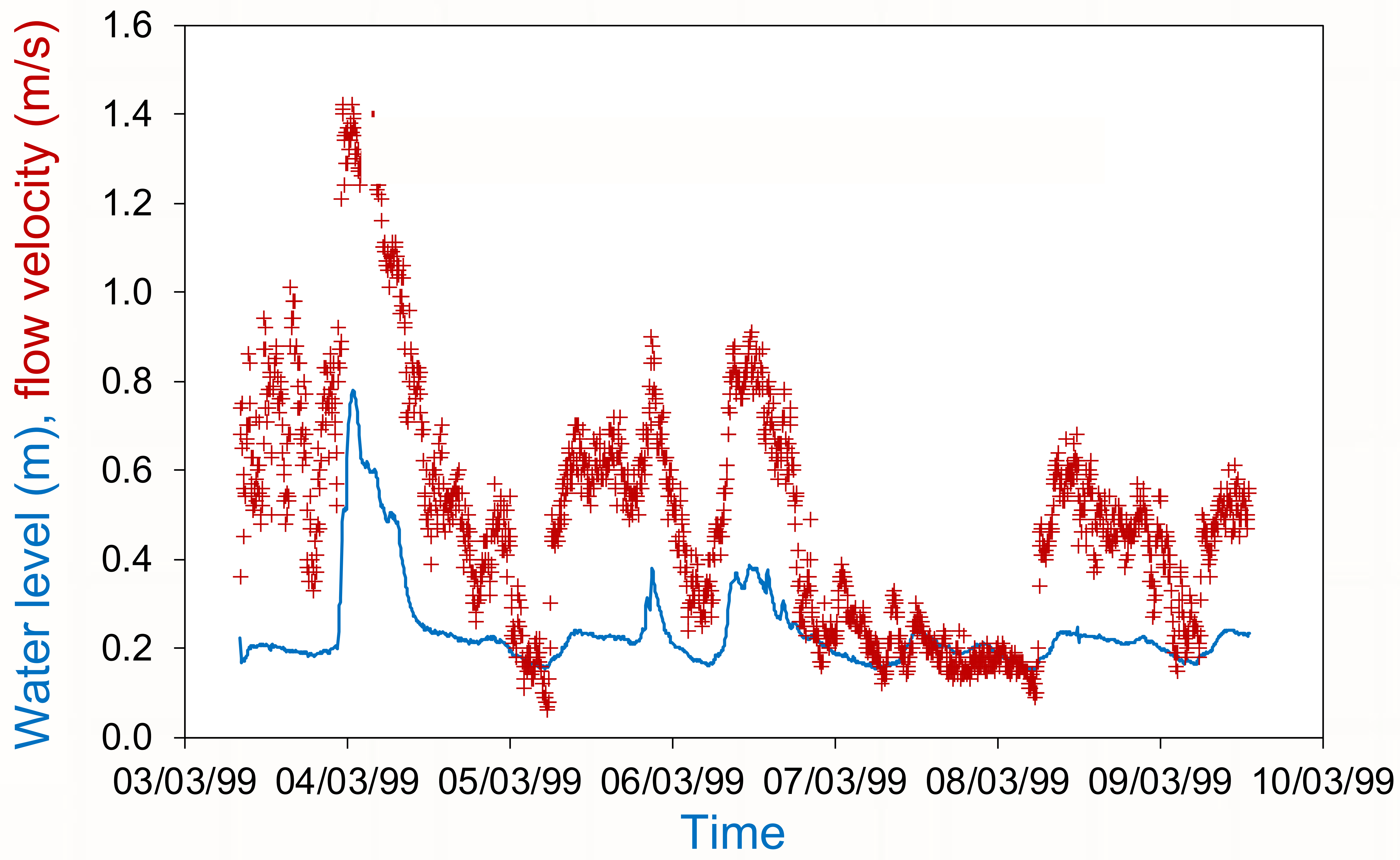
Page-Hinkley algorithm

ANALYTICAL REDUNDANCY



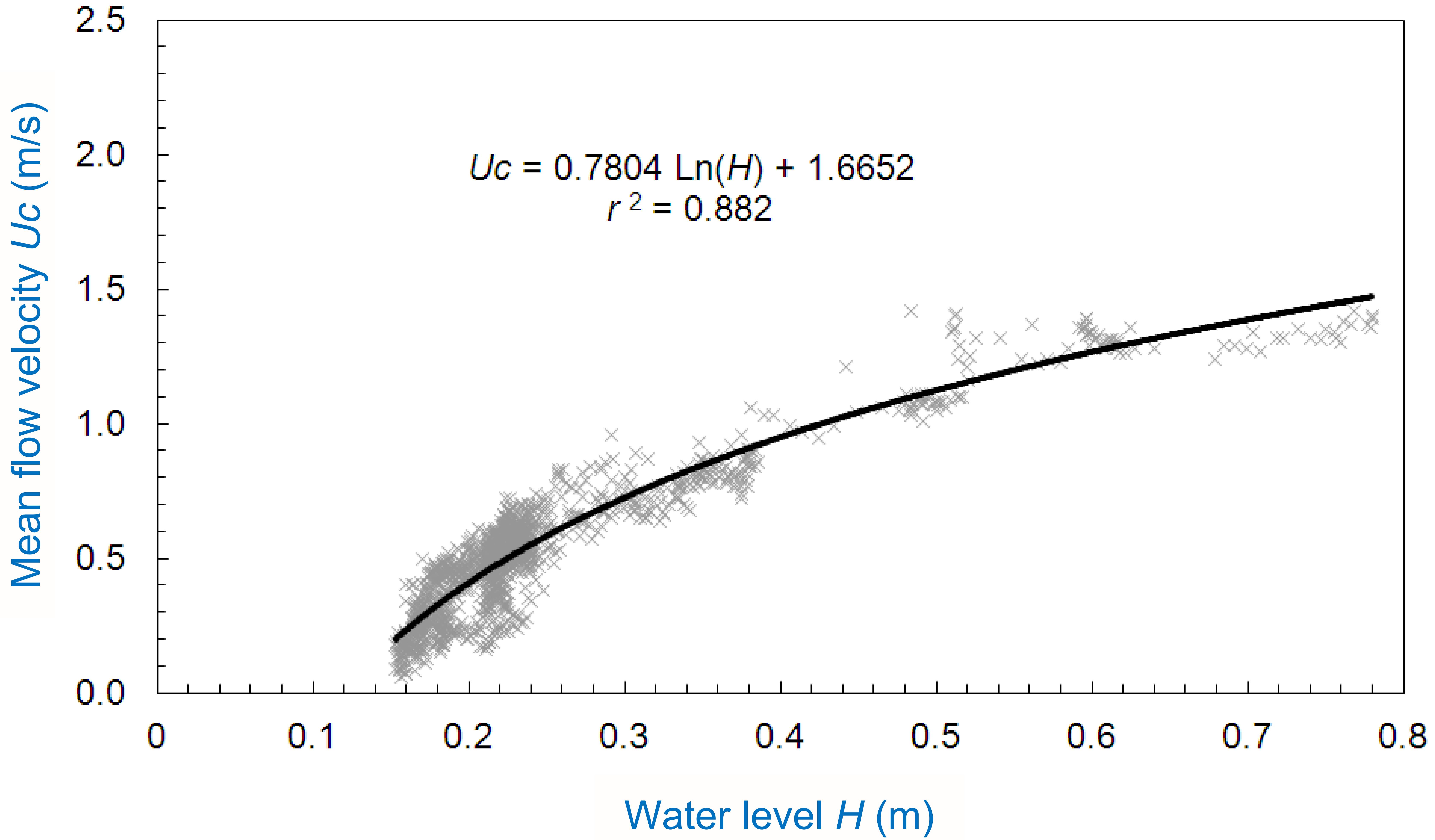
Page-Hinkley algorithm

EXAMPLE



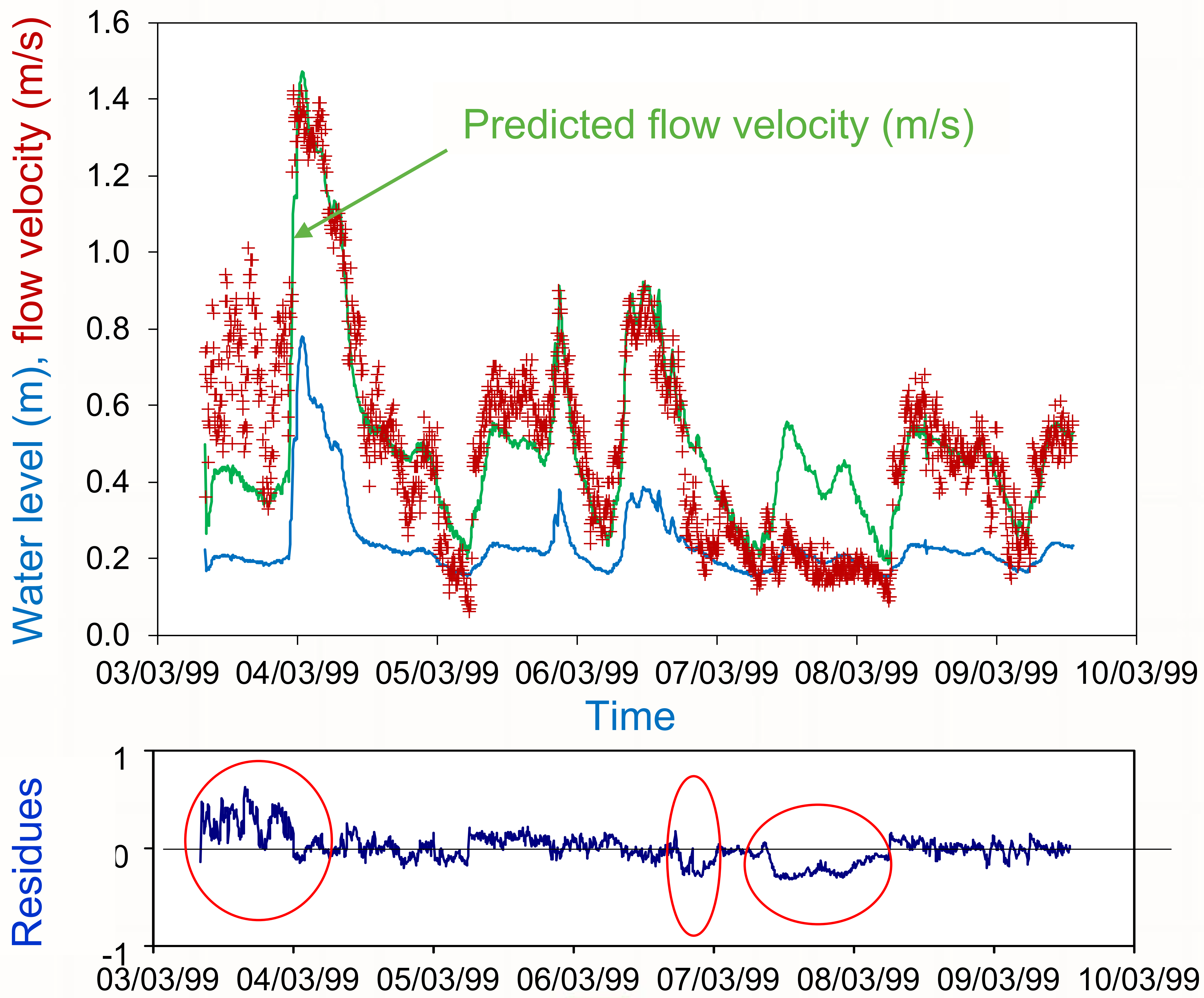
Water level (m), flow velocity (m/s)
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
03/03/99
04/03/99
05/03/99
06/03/99
07/03/99
08/03/99
09/03/99
10/03/99
Time

EXAMPLE



Uc = 0.7804 Ln(H) + 1.6652
r 2 = 0.882
2.5
2.0
1.5
1.0
0.5
0.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Mean flow velocity Uc (m/s)
Water level H (m)

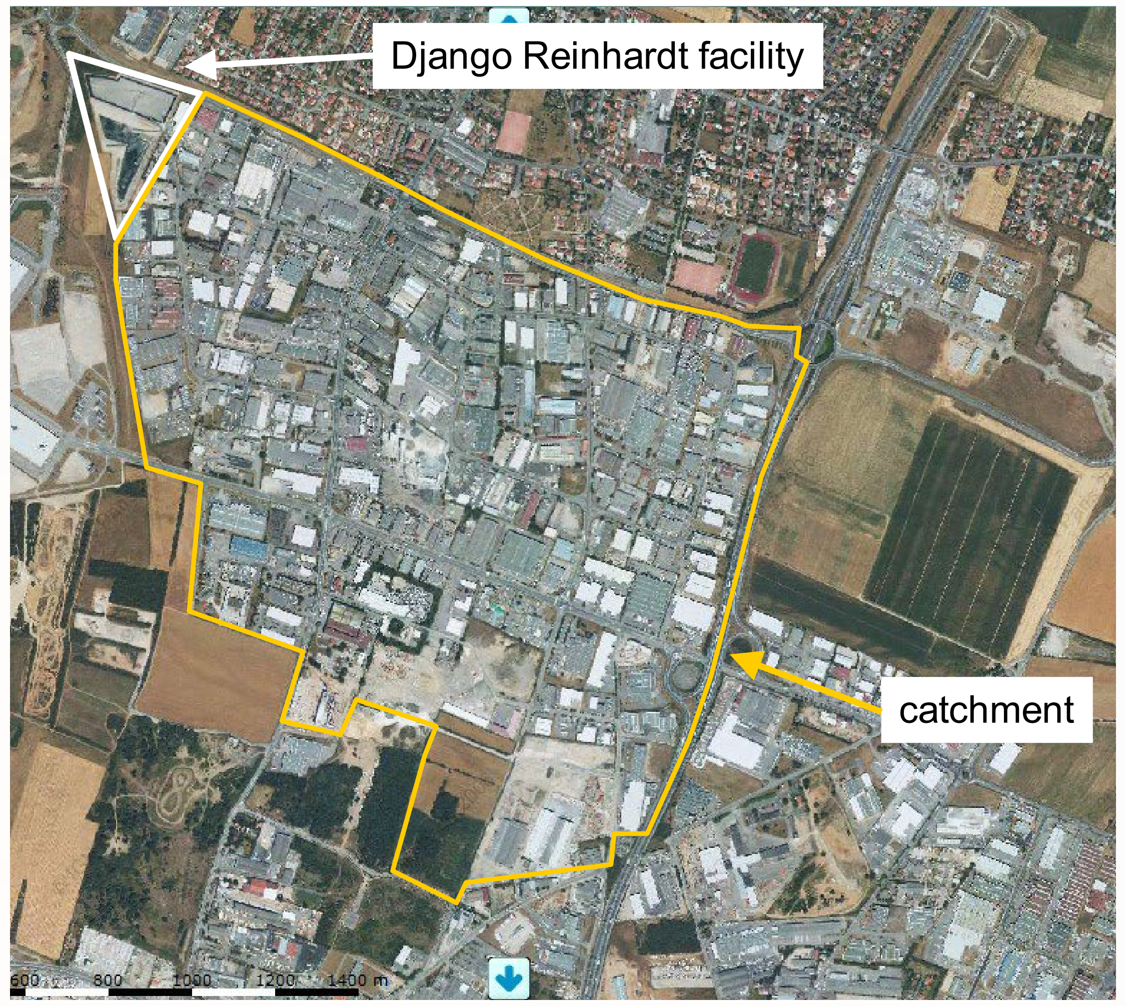
EXAMPLE



Water level (m), flow velocity (m/s)
Predicted flow velocity (m/s)
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
03/03/99 04/03/99 05/03/99 06/03/99 07/03/99 08/03/99 09/03/99 10/03/99
Time
Residues
1
0
-1
03/03/99 04/03/99 05/03/99 06/03/99 07/03/99 08/03/99 09/03/99 10/03/99

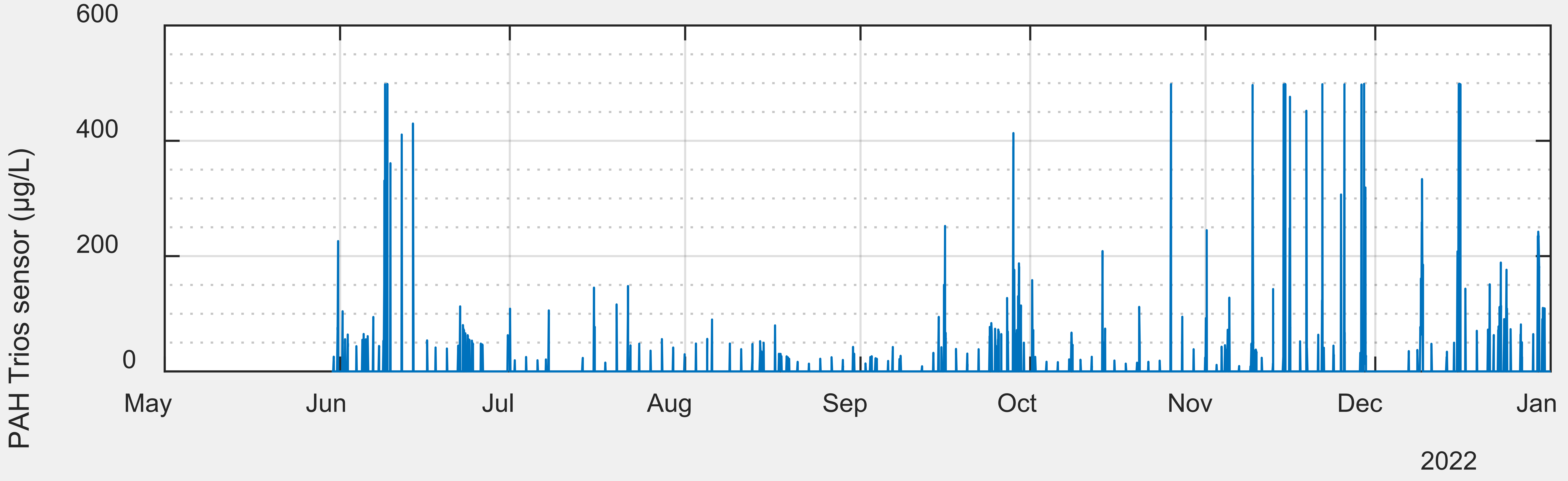
EXAMPLE OF UDMT APPLICATION 1/4

- Co-UDlabs sensor testing in WP 6 : case of a PAHs probe
- *In situ* testing in the Chassieu OTHU monitoring station from 30 May 2022 at 12:10 until 31 December 2022 at 00:00



EXAMPLE OF APPLICATION 2/4

- Co-UDlabs sensor testing in WP 6 : case of a PAHs probe
- *In situ* testing in the Chassieu OTHU monitoring station from 30 May 2022 at 12:10 until 31 December 2022 at 00:00

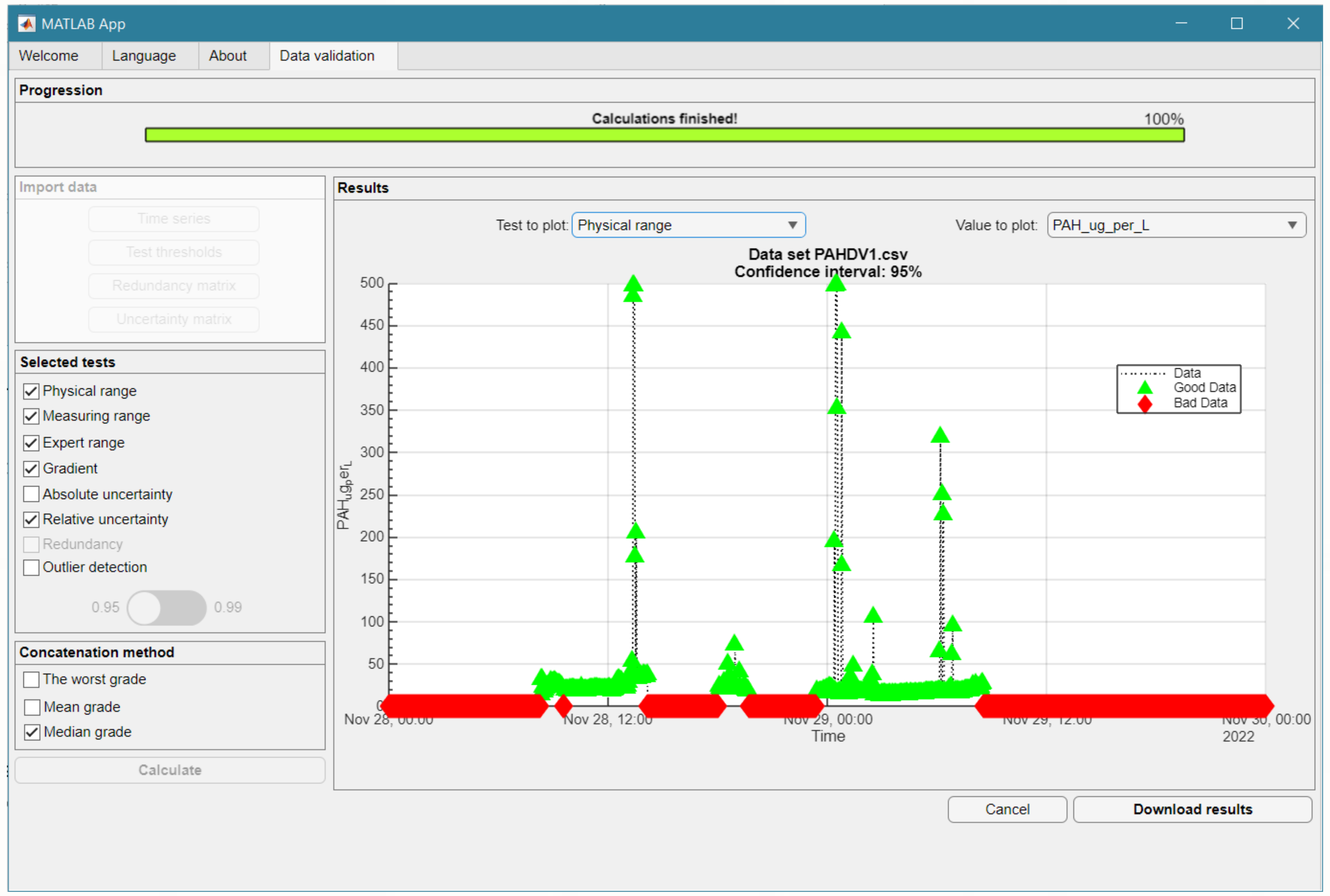


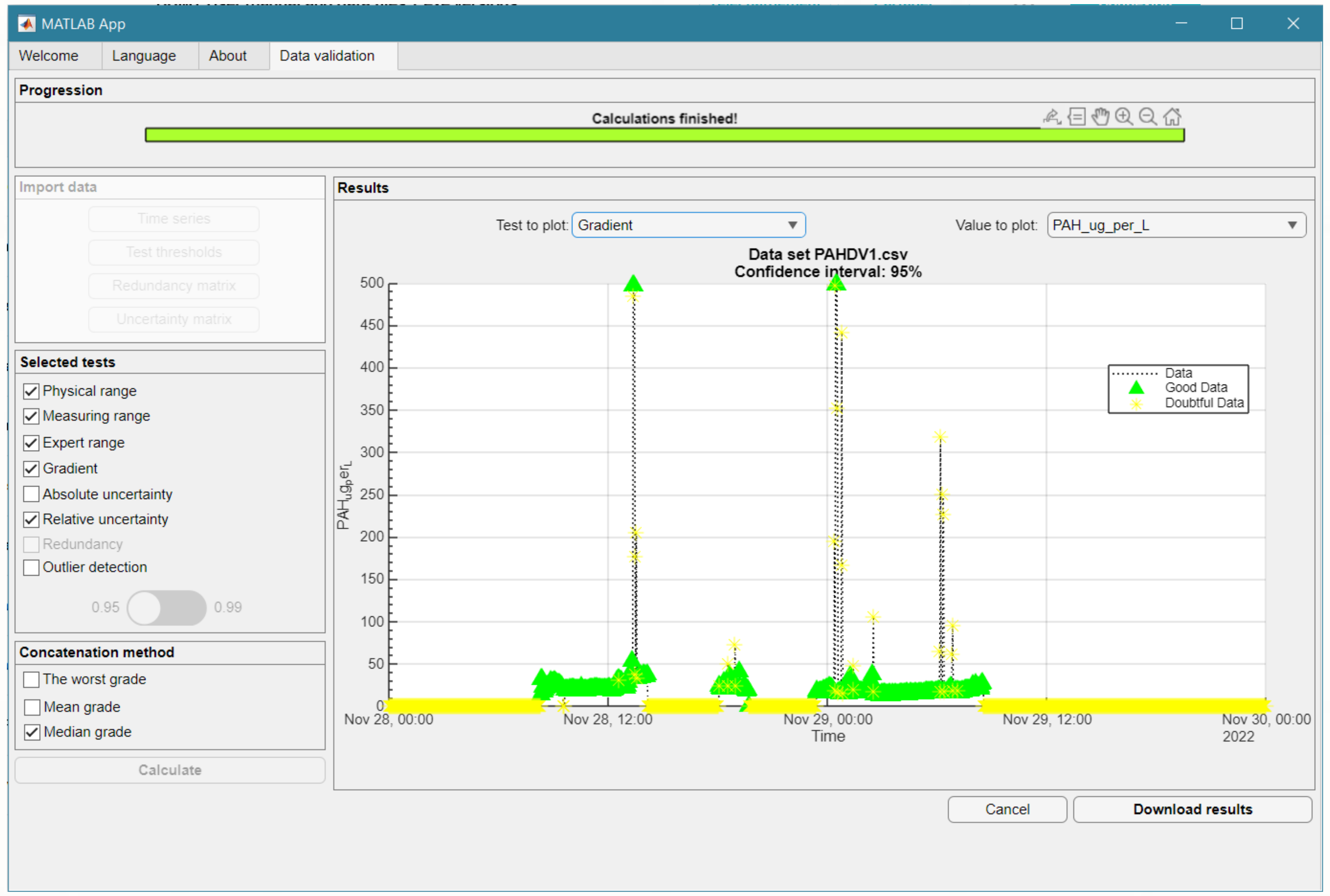
EXAMPLE OF APPLICATION 3/4

- Time series file PAHDV.csv, from 30 May 2022 10:12 to 31 Dec 2022 23:58

Time	PAH_ug_per_L	uPAH_ug_per_L
29-Nov-2022 05:16:00	15.9958	12.5
29-Nov-2022 05:18:00	16.3591	12.5
29-Nov-2022 05:20:00	15.3127	12.5
29-Nov-2022 05:22:00	17.305	12.5
29-Nov-2022 05:24:00	17.2083	12.5
29-Nov-2022 05:26:00	17.2306	12.5
29-Nov-2022 05:28:00	16.6133	12.5
29-Nov-2022 05:30:00	16.8899	12.5
29-Nov-2022 05:32:00	16.4885	12.5
29-Nov-2022 05:34:00	17.431	12.5
29-Nov-2022 05:36:00	16.8762	12.5

ON-LINE DEMO…


MATLAB App
Welcome
Language
About
Data validation
Progression
Calculations finished!
100%
Import data
Time series
Test thresholds
Redundancy matrix
Uncertainty matrix
Results
Test to plot: Physical range
Value to plot: PAH_ug_per_L
Data set PAHDV1.csv
Confidence interval: 95%
Data
Good Data
Bad Data
Selected tests
Physical range
Measuring range
Expert range
Gradient
Absolute uncertainty
Relative uncertainty
Redundancy
Outlier detection
0.95
0.99
Concatenation method
The worst grade
Mean grade
Median grade
Calculate
Time
Cancel
Download results


MATLAB App
Welcome
Language
About
Data validation
Progression
Calculations finished!
Import data
Results
Test to plot: Gradient
Value to plot: PAH_ug_per_L
Data set PAHDV1.csv
Confidence interval: 95%
Data
Good Data
Doubtful Data
Selected tests
Physical range
Measuring range
Expert range
Gradient
Absolute uncertainty
Relative uncertainty
Redundancy
Outlier detection
0.95
0.99
Concatenation method
The worst grade
Mean grade
Median grade
Calculate
Nov 28, 00:00
Nov 28, 12:00
Nov 29, 00:00
Nov 29, 12:00
Nov 30, 00:00
2022
Time
Cancel
Download results

EXAMPLE OF APPLICATION 3/4

T	PAH_ug_per_L	uPAH_ug_per_L
PR_Min	1	NaN
PR_Max	499	NaN
MR_Min	5	NaN
MR_Max	250	NaN
ER_Min	10	NaN
ER_Max	150	NaN
GR_Min	0.001	NaN
GR_Max	10	NaN
AU	NaN	NaN
RU	NaN	0.25

✦ Test thresholds file PAHTR.csv contains the following values:

- physical range: from MR_Min = 1 to MR_Max = 499 µg/L
(to detect extreme values 0 and 500 µg/L and beyond)
- sensor measuring range: values from PR_Min = 5 to PR_Max = 250 µg/L are considered as valid for this site and this sensor (assumed to be a realistic range in this example)
- the expert range: values from ER_Min = 10 to ER_Max = 150 µg/L are considered as valid for this site and this sensor, from previous knowledge (expected range of most frequent values)
- gradient: gradients below GR_Min = 0.001 µg/L/min (flat signal) or above GR_Max = 20 µg/L/min (abrupt changes) in one time step Δt = 2 min are detected for final manual validation
- absolute uncertainty: AU = NaN (not applied)
- relative uncertainty: values with relative standard uncertainty RU higher than 0.25 (i.e. 25%) are considered as not enough reliable for validation

USEFULL LINKS

- **The free UDMT Toolbox :** http://vps-7bc5cf87.vps.ovh.net:9988/webapps/home/session.html?app=coudlabs
- **Any comment :** mailto:UrbanDrainageMetrologyToolbox@gmail.com
- **The Co-Udlabs project :** https://co-udlabs.eu/
- **Metrology in Urban Drainage and Stormwater Management: Plug and Pray** https://iwaponline.com/ebooks/book/835/Metrology-in-Urban-Drainage-and-Stormwater
- **Asset Management of Urban Drainage Systems: If anything exciting happens, we've done it wrong!** https://iwaponline.com/ebooks/book/920/Asset-Management-of-Urban-Drainage-SystemsIf
- **Standardisation - the SWAN IUG group** https://swan-forum.com/interoperable-utility-group/
- **The Dutch project on data standardisation:** https://data.gwsw.nl/ It is in Dutch but I guess translations should be doable….