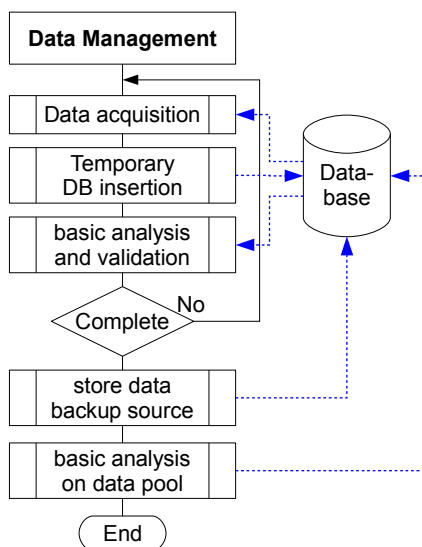


PROPOSITION FOR INTERNSHIP

Time series analysis, on-line data validation and feature reporting

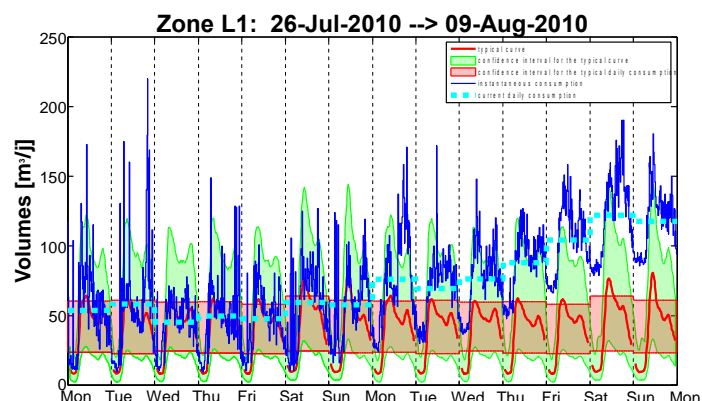


In the framework of the research project MoGREPo [1], focusing on the management and control of drinking water distribution networks, a longterm data acquisition is done. Different variables of the network reflecting the water consumption are gathered over time (15min sampling) from distributed locations of the network. After a basic manual validation process the data is stored in a centralized relational database. Thus most of the data in the database represent time series [2], in this case information on water consumption changing over time.

The available and basically validated data exists now for 27 month for about 26 variables. Currently a weekly report on detected exceptions is send to the project partners based on a manual data analysis. Some statistical pattern are extracted

like daily or weekly mean values or typical day curves together with there confidence intervals to give some examples. Also existing redundant information are analyzed for a possible use of data validation. But, currently this information is not directly used in order to produce the weekly report in an automatic way.

The framework for the data management is realized using the python [3] programming language. Most of the analysis is done using python and Matlab [4], but an option could be the statistical framework R [5].



The internship will be done at the modeling and simulation unit (MODSI) of AMS under the responsibility of Dr. Georges Schutz.

BRIEF DESCRIPTION.

The main goal of this internship project is to automate the current weekly reporting of exceptional behavior of the consumptions in the network. For this different analysis, that are currently done off-line, should be systematically integrated into the existing data-management schema and results stored in the centralized database. In order to automatically detect most of the exceptional behavior additional analysis especially on night time consumptions have to be realized. Based on indicators combining this information it will be possible to generate the weekly report in an automatic way.

THE INTERNSHIP WORK-FLOW

The internship is done in the framework of a research project of the CRP Henri Tudor (MoGREPo). The existing analysis methods and the existing historical network data as well as the expertise of the research team of this project will support the student in order to bring this project to a positive end.

After a brief introduction and an analysis of the internship project topic the student will have to become familiar with the tools used in the research project, mainly python, Matlab and eventually R, and with the drinking water network specificities and the available measurement data.

In a next step the database schema will be adapted in order to store results of the different off-line analysis and for the later storage of detected exceptions. The existing off-line analysis should be adapted in that way that it is able to directly interact with the database and store or update stored results if needed.

Finally, indicators combining the different analysis results will be specified in order to detect exceptional behavior and produce a weekly report of detected exceptions in an automatic way.

THE OBJECTIVES OF THE INTERNSHIP

1. Initiation to research and development work,
2. Working on projects in a team,
3. Relational database development using object relational managers,
4. Combine data analysis results into indicators for exception detection,
5. Use mathematical methods and programming environments for solving real word problems.

THE DEFENSE

The internship will produce a report of the realized work and a presentation of the most important outcomes will be done in-front of the department members.

REFERENCES

- [1] http://www.crte.lu/cms/crte/content.nsf/id/newsletter_January10
- [2] http://en.wikipedia.org/wiki/Time_series
- [3] <http://www.python.org>
- [4] <http://www.mathworks.com/>
- [5] <http://www.r-project.org/>