Improving data for the asset management of the water supply network of the Walloon Water Company

N. Rodriguez*^, A. Mirebeau**, A. Husson*, M. Collet*, E. Renaud*, and Y. Le Gat*

*INRAE Nouvelle-Aquitaine Bordeaux, UR ETTIS, Cestas-Gazinet, France
(E-mail: *nicolas.rodriguez@inrae.fr)*

** Centre National de Recherches Météorologiques (CNRM, UMR 3589), Toulouse, France
(E-mail: *aurelien.mirebeau@meteo.fr*)

^Corresponding author

Keywords
Infrastructure Asset Management, Database, Formatting, Imputation, Modeling, Water Supply Network.

ABSTRACT / EXTENDED ABSTRACT

The Infrastructure Asset Management (IAM) of water supply networks (WSNs) relies on vast amounts of data. Ideally, these data should be stored and organized in one single database (DB). Water utilities may however need to collect data in separate files or systems, for various reasons. In the case of the Walloon Water Company (SWDE, see Figure 1), separate files describe the pipes, their environmental attributes, their maintenance history, the network structure, and the daily to yearly operations. This stems from the adaptation of their information system to a historical expansion of the managed pipe network and from rapidly increasing data needs. In circumstances like these, managing the assets can imply data manipulation and formatting efforts, which may be goal-specific. For instance, using pipe failure prediction models (such as the one embedded in the Casses software) typically requires the complete knowledge of each pipe's characteristics and maintenance history (repairs, renewals). This may involve data manipulation (Rodriguez et al., 2022), and sometimes data imputation. In this article, we show examples of data manipulations that were designed (using the programming language R) to help SWDE further develop its IAM. Three problems are addressed in the case of SWDE:

- Detecting the underlying DB structure in the various files (see Figure 1) and making use of DB concepts to maximize the use and interoperability of the data for IAM ;
- Detecting and correcting data which are not consistent with "field constraints" (list of pre-defined values), especially manually input text such as locations ;
- Imputing missing data.

Particular emphasis is put on the imputation of missing pipe data (e.g., installation date, material), for which up to 4 imputation methods are developed and compared. The quantity (completeness) and the quality (accuracy) of the imputed data relative to the original data are first compared. The compromise found between completeness and accuracy lies in successfully implementing business rules within the imputation method. The impact of imputing (or not) missing pipe characteristics on the LEYP (Linearly Extended Yule Process) pipe failure prediction model (Le Gat, 2014) is then measured. The effect of the imputation methods on model coefficients and on predicted pipe failure rates (per km and per year) relative to the original data is specifically compared. This confirms the hypothesis that an accurate data imputation improves the ranking of pipe renewals priorities with respect to predicted failure rates (Figure 2). Random effects are accounted for by using repeated random experiments (of deletions and imputations).

The processed (imputed) and formatted data is used in several models (e.g., failure model, regressions, long-term simulations) that will be a basis in the IAM of SWDE. Eventually, this work allows us to provide guidelines for the design of DBs in the context of the IAM of WSNs.

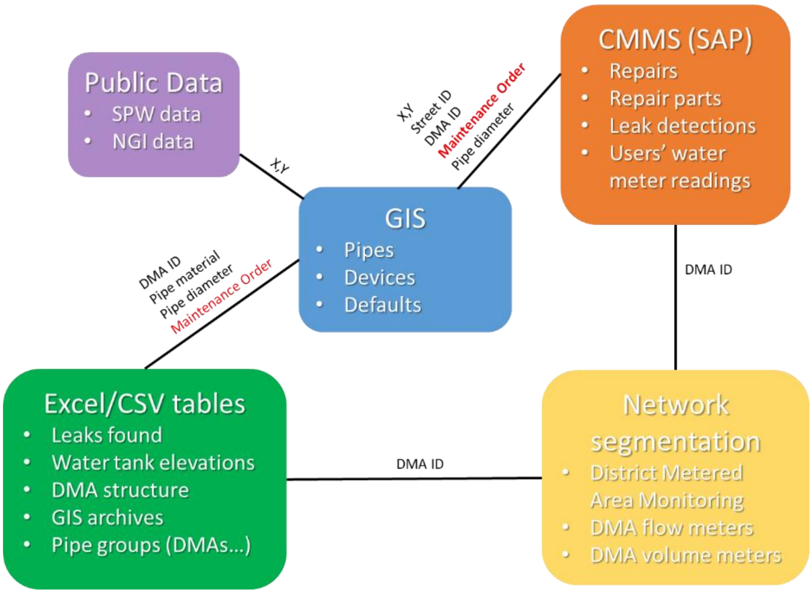

Figure 1: Overview of the data files and of the keys linking those in the information system of the Walloon Water Company (SWDE). CMMS: Computerized Maintenance Management System. GIS: geographic information system. CSV: comma separated values. SPW (Service Public de Wallonie) and NGI (National Geographic Institute): public institutes in Belgium.

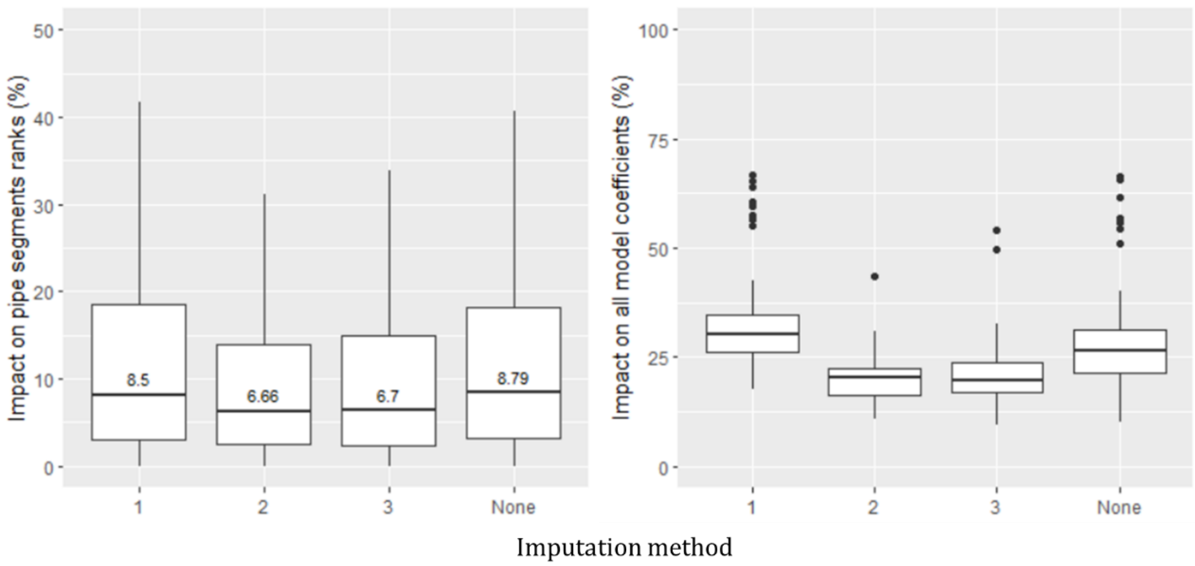

Figure 2: Impacts of the imputation methods for missing pipe materials on the pipe segment ranks and on all model coefficients from the LEYP model. 1: mode. 2: nearest neighbor method. 3: nearest neighbor method constrained by pipe diameter equality.

REFERENCES

Le Gat, Y. (2014) Extending the Yule process to model recurrent pipe failures in water supply networks. *Urban Water J.* **11**, 617–630. https://doi.org/10.1080/1573062X.2013.783088

Rodriguez, N., Husson, A., Le Gat, Y., Renaud, E. (2022) Automatized data preparation methods for the asset management of water supply networks. *Techniques Sciences Méthodes* **10**, 71-87. https://doi.org/10.36904/tsm/202210071