



SAVONIA

THESIS - MASTER'S DEGREE PROGRAMME
TECHNOLOGY, COMMUNICATION AND TRANSPORT

AUTOMATION OF WATER DISTRIBUTION SYSTEM MODEL BUILD-UP

AUTHOR:

Janne Väyrynen

Field of Study Technology, Communication and Transport	
Degree Programme Master's Degree Programme in Environmental Engineering	
Author Janne Väyrynen	
Title of Thesis Automation of Water Distribution System Model Build-up	
Date 8 June 2021	Pages/Appendices 42/1
Client Organisation /Partners Fluidit Ltd.	
<p>Abstract</p> <p>International Water Association (IWA) has listed smart and intelligent network technologies as one of the major trends in the water industry in the next decade (IWA, 2020). Hydraulic models help water utilities become more sustainable by reducing leaks, optimizing the network investments, and the system's energy use. This thesis, commissioned by Fluidit Ltd., had the objectives of documenting the current Water Distribution System (WDS) model build-up used by Fluidit Ltd. and automating the process by implementing scripts and pseudo-codes. The aim was to improve network processing, which has usually been the most time-consuming part when constructing hydraulic network models. The validation methods created in this thesis aim at easily accessible, better-quality hydraulic models at a global scale.</p> <p>In this thesis, the methods for documenting and improving the model build-up were gradually improved. At first, the current model build-up process was documented. Manual processing phases were replaced step by step with scripts and the pseudo-codes implemented in this thesis. The processes include validating network connectivity and the essential pipe attributes in the WDS models. Finally, the two hydraulic model building methods, manual and import methods, were compared in terms of quality and speed.</p> <p>The improved model build-up process was documented in this thesis, and it will be used when training modelers using Fluidit software. The modeling process includes the requirements of input data and how to deliver the data to the model. The network data was put under validation using scripts and the pseudo-codes created in this thesis, and a method for the automatic WDS network model update was introduced. Implementing the automatic build-up tool was not fully completed during the thesis; thus, the performance could not be thoroughly evaluated. The comparison of the manual building with a semi-automatic import process showed that the manual build-up could be efficient when modeling small networks in terms of speed, but importing the network is beneficial in terms of model quality, improving the quality of the original network data, and reducing data loss when creating a hydraulic model. The developed automatic validation method has shown promising results with actual network data. Visual, map-based network validation reports made it easier to correct the network in network information system (NIS), as the flaws in data became more visible. For example, a test case using data from a Finnish water utility showed that the topological connectivity of the modeled network had significant improvement just after two iterations of network validation. The data validators and the model building processes will be further improved after this thesis, adding implementations in sewer, storm, and district energy networks. Some input data sources that limit fully utilizing the hydraulic model build-up in WDS networks are listed in Appendix 1.</p>	
<p>Keywords</p> <p>District metering area, DMA, EPANET, Fluidit, hydraulic model, model, network, network information system, NIS, optimization, pump, SCADA, topology, validation, valve, water, water distribution system, WDS</p>	

PREFACE

I want to thank Pasi Pajula and Juha-Matti Aalto for guiding me in this thesis. I express gratitude to Fluidit Ltd. and Timo Ranta-Pere for allowing this work to happen. Special thanks to Dr. Markus Sunela for reviewing my work and answering my endless questions throughout the years. Thank you, Kalervo Aho, for your guidance and support in completing this thesis. I also want to thank my wife for the endless support during my studies.

Kuopio 3.6.2021

Janne Väyrynen

CONTENTS

1	INTRODUCTION	7
2	WATER SUPPLY SYSTEMS	9
2.1	Water supply system structure	9
2.2	Water distribution system model (WDS model)	10
2.3	Fluidit Water	11
2.4	EPANET	11
3	WDS MODEL COMPONENTS AND THE PROPERTIES OF INPUT DATA.....	13
3.1	Pipe network.....	13
3.2	Water consumers	15
3.3	Flow and pressure history data	15
3.4	Control device settings	16
3.5	Pumps.....	16
3.6	Tanks	17
3.7	Digital elevation model (DEM)	18
4	WDS MODEL BUILD-UP PROCESSES.....	19
4.1	Processing digital elevation model (DEM).....	19
4.2	Processing network	20
4.2.1	Creating network model manually.....	20
4.2.2	Creating a network with import tools	20
4.2.3	Pipe material library.....	22
4.3	Processing gate valves.....	23
4.4	Processing stations.....	24
4.4.1	Processing tanks and reservoirs.....	25
4.4.2	Processing pumps and valves	25
4.4.3	Processing controls settings	26
4.5	Processing water consumers	27
4.6	Processing emitters and patterns.....	27
5	IMPROVED DATA PROCESSING TOOLS FOR NETWORK DATA	30
5.1	Base model and scenario structure	30
5.2	Pseudo-code for checking network topology	30
5.3	Pseudo-code for checking diameter and material.....	31

5.4	Pseudo-code for checking pipe categories.....	31
5.5	Pseudo-code for valve connections checking	32
5.6	Pseudo-code for generating a network model	32
5.7	Pseudo-code for updating the network model	33
6	DATA PROCESSING PERFORMANCE	35
7	CONCLUSIONS & RECOMMENDATIONS.....	38
	REFERENCES.....	40
	APPENDIX 1: COMMON ISSUES WHEN ANALYZING AND GATHERING WDS MODEL INPUT	
	DATA.....	42

LIST OF FIGURES

FIGURE 1.	Schematic drawing of a water supply system with multiple water sources, water towers, and pressure zones (Sunela, 2017)	9
FIGURE 2.	Hydraulic model components (Fluidit, 2021)	13
FIGURE 3.	Example of pump control methods in mock-up picture of a SCADA system. The pressure booster controls outgoing pressure with a constant pressure setting alternating a bit during the day. Pump stations flow limited. Two rightmost control types are available but not in use.....	16
FIGURE 4.	Pressure booster station with two pumps. A pump plate is used to find pump curves using online catalogs (Väyrynen 2015)	17
FIGURE 5.	Data processing workflow from each data source to final model component	19
FIGURE 6.	Adding of digital elevation model (DEM) to the hydraulic model	19
FIGURE 7.	Creating network by manually adding pipes on top of the network map	20
FIGURE 8.	Import components to the Model tool, input of field values.....	21
FIGURE 9.	Data Processing workflow for network data to create a pipe network model	22
FIGURE 10.	Material library in Fluidit Water and selecting a pipe material from the library to a pipe	23
FIGURE 11.	Data processing workflow for gate valve data	24
FIGURE 12.	Data processing workflow for control devices and tanks.....	24
FIGURE 13.	Station objects with zones and tags included. Control device sizes and lengths are exaggerated for visual purposes.	25
FIGURE 14.	Modeling of control SCADA control settings.....	26
FIGURE 15.	Data processing workflow for water consumer data.....	27
FIGURE 16.	Demands connecting to the nearest junctions in Fluidit Water	27
FIGURE 17.	Data processing workflow for flow history data	28
FIGURE 18.	District metering areas in Fluidit Water model. Control device sizes and lengths are exaggerated for visual purposes.	28

FIGURE 19. Example of water usage in a DMA	29
FIGURE 20. Simplified process to report topology errors	30
FIGURE 21. The process to generate a WDS model network	32
FIGURE 22. The process to update the network and demands of a WDS model.....	34
FIGURE 23. Comparison of network modeling speed using import methods and manual model build-up	37

GLOSSARY:

Term	Definition
CRS	Coordinate Reference System
DEM	Digital Elevation Model
DMA	District Meter Area
Geocoding	Process of converting addresses into geographic coordinates
NIS	Network Information System
SCADA -control system	Supervisory control and data acquisition control system
Topology	The study of geometrical properties and spatial relations unaffected by the continuous change of shape or size of figures
UI	User Interface
WDS	Water Distribution System

1 INTRODUCTION

After the author finished his bachelor's thesis back in 2015, his position as a design engineer has been chiefly building up hydraulic network models and analyzing networks. Modeling tools, methods, and input data sources have been improved throughout the years, but the whole modeling process itself is not well documented. In this thesis, the author combines his current understanding of the Water Distribution System (WDS) model building into this one document. This thesis is written with a practical approach, explaining the different input data sources and the typical errors found in each of them. Most of the author's modeling projects have been in Finland, but the issues found are universal, and the same problems can be found across the globe. In this case, the data conversion process in the hydraulic modeling software, Fluidit Water, is explained and improved.

Still, in the 2010s, models were often built manually on top of the network map. As the modeling tools improved, the author could visualize model topology dynamically within the modeling software. Programming languages, such as Python, Java, and JavaScript, enhance network data processing to a whole new level. Python scripts for automatizing the building process were implemented in this thesis. Although these scripts themselves are not included in this document, explanations of what the scripts do are described as pseudo-codes.

Data validators created in this thesis have been created using the data provided by Finnish water utilities. The problems with network documentation may differ with different NIS software providers, which should be considered when using these validators. Even though some products' user interface (UI) may prevent the user from easily exporting the data for modeling, restrictions added may have been implemented for a good reason. From the modeler's point of view, the author wants to improve all data source quality and availability, and open dialogue between many players in the field. The list of the most common problems found with these data sources is listed in Appendix 1. These data sources need more transparency and standardization to smoothen the modeling process, provide more valuable information for the utilities, and facilitate broader use of data at the water utilities.

From a water utility's point of view, computational models are relatively new compared to managing the network and storing the network information. Data has been collected for a long time before any models had been created, and this is one reason why the quality of network data does not necessarily meet the demands for hydraulic models. Although the pipes documented in the network information system (NIS) were based on field measurements or plans, there was no need to keep an eye on the network continuity, add properties for all the line elements or to avoid creating duplicate lines. Line continuity was used for visual purposes only: to show where the pipes were and which size and material they had. Gaps in the topology and data caused no problems. Modern modeling tools allow using of utility data directly, accelerating the model build-up. Modeling tools and processes can help the water utilities to improve their data, so that models created using the data will be more reliable and easier to maintain.

Fixing the network data requires many resources from the water utility, and even finding out what to fix may be an overwhelming task. When hydraulic models are built, much data goes through check-ups. Errors found in the data should be addressed in the original data storage, and not only fixed in

the model. The author has recognized the need to prioritize what should be fixed in the model and estimate missing information. Also, the number of gaps in the network data does influence the cost of the modeling project. When the network model is updated roughly every year, the input network data should not contain all the same errors as the last time. That is why it is logical to send the network continuity report also to the water utility. As the data processing tools estimate some data gaps in the network data, utilities can prioritize the manual fixing only to the critical locations, where automatic fixing cannot be applied. With the advanced import functionalities, the network information system can import the results of the automatic fixing into the original data source as well.

This thesis focuses on the water supply systems and especially the water distribution networks. Model build-up functions can also be applied to the sanitary sewer, stormwater and district heating networks, although they may require extra processing steps to get the correct results as the number of variables increases. For example, in gravitational sewer systems, elevations add another layer of complexity.

Being able to generate standardized network models aims for easily accessible, better-quality models. More time should be left for the analysis and design tasks. Here are the main objectives for this work. The author will evaluate the accomplishment of these points at the end of the thesis:

1. Document the model building process: what data should be added to the model and how.
2. Write scripts and pseudo-code to validate the network continuity errors.
3. Highlight the continuity errors in a format, that can be imported back into the NIS.
4. Update models easily when new data is available.
5. Compare the speed and the quality of the automatic network building process to a manual model build-up.
6. Learn programming skills while doing the tasks above.

The overall aim of the thesis is to improve water utility resilience, as the most up-to-date information can prevent and solve a possible water crisis. Models can help water utilities to become more sustainable by reducing leaks, optimizing the network investments and the energy use of the system.

In 2020, the International Water Association (IWA) listed smart and intelligent network technologies as one of five major trends impacting the water industry in the next decade. Infrastructure management is enhanced, and non-revenue water losses are reduced using improved data analytics and Internet of Things (IoT) devices. (IWA, 2020)

2 WATER SUPPLY SYSTEMS

2.1 Water supply system structure

Water supply system is an infrastructure for collecting, transmitting, treating, storing, and distributing water for homes, commercial establishments, industry, and irrigation. Water can also be delivered to public needs such as firefighting and street flushing. Of all the municipal services, providing potable water is perhaps the most vital. Water is delivered for drinking, cooking, washing, carrying away wastes and other domestic needs. Water must fulfill both quality and quantity requirements for public, commercial and industrial activities. (Nathanson, Jerry A., 2021) A schematic representation of a water supply system is shown in Figure 1.

A water supply system consists of potable water source(s), water consumers, and transport facilities. As a water source, treated surface water or disinfected groundwater are used. In the water source, there is usually a clear well to increase the contact time of disinfectants, buffer water demand difference between the treatment process and the distribution system, and to serve as a source for back-wash water for plant filter cleaning. Water consumer demand varies over time, both in the long-term (seasonally) and the short-term (daily); this is the driving mechanism of how the water distribution system behaves. Moving water from the source to the consumers requires a network of pipes, pumps, and valves. Water storages help to smoothen out the fluctuations in demand. Piping, storage and supporting infrastructure are referred to as the water distribution system (WDS). (Walski et al., 2003).

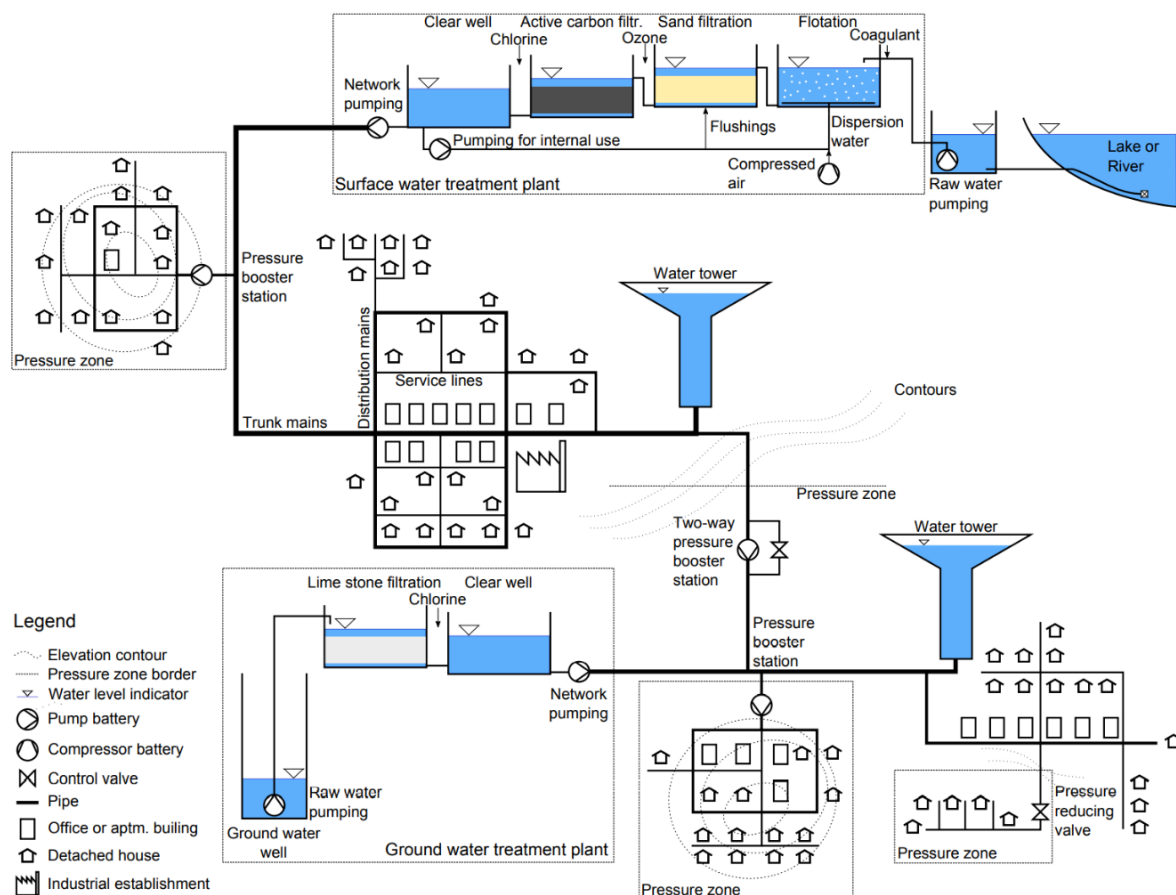


FIGURE 1. Schematic drawing of a water supply system with multiple water sources, water towers, and pressure zones (Sunela, 2017)

WDS piping is often categorized into transmission/trunk mains and distribution mains. Transmission mains are designed to convey large amounts of water over great distances, and usually individual customers are not served from the transmission mains. The distribution mains are smaller in diameter than the transmission mains and usually follow the topology and alignment of the city streets. In the distribution lines, fire hydrants, isolation valves, control valves, blow-offs, and other maintenance and operational appurtenances are frequently connected. Finally, service lines provide the water from the distribution mains to the end customers. (Walski et al., 2003) This thesis focuses on the networks and how the network data is processed to be suitable for hydraulic calculations.

2.2 Water distribution system model (WDS model)

Simulation refers to imitating the behavior of one system through the functions of another. Models are mathematical representations, replicating the dynamics of an existing or proposed system. (Walski et al., 2003). In WDS models, pipes, consumers, leaks and even water itself are imitations of the real counterpart. Network capacity is calculated using a mix of data about mapped location, pipe manufacturer-provided diameters, and commonly used roughness values for each pipe material. Water consumers use water with district-based patterns instead of treating them individually. The pipes do not leak in the models, as the water goes out from the "Junctions" instead. Model simulations are usually done using deterministic approach, always resulting in the same results with unchanged input values.

Simulating the water distribution systems is useful when it is not practical to directly conduct experiments in the real system, as testing in the actual network can be expensive and risky for public health. Simulations can also be used to evaluate a system before it is built. (Walski et al., 2003) WDS models are a tool for understanding the behavior of the water distribution network. A model can serve as a design tool for numerous what-if -scenarios, and it can provide up-to-date information on the current state of a network. With the use of models, information gathered from metering stations is enhanced to the whole network-wide, from the reservoirs to each consumer. There can be multiple hydraulic models in a single water utility, each serving the different design, operation and management levels.

Most of the models the author has been involved with have been made from designers to designers. Design models can be used as a tool to dimension new pipes and pumps. The design models are excellent to test how the hydraulic system behaves in different circumstances. The pipe network of these design models is usually created from a snapshot of NIS data, which has been imported or manually added to the model. Although these models lack a direct link to the original NIS data, modeling results can still be imported back to the NIS via unique ID or background layers. As the NIS and the WDS model are separate systems, all future and what-if -scenarios are easily distinguished from the system's current state. These design models are updated as needed, usually when significant modifications have been made to the real system. With proper network documentation and finalized model build-up process, the author estimates that in the future, the networks for hydraulic models can be easily updated at any time as needed.

Some utilities have the option to create hydraulic models directly from NIS using hydraulic model integration. With an integrated model, the user can make edits to the network and simulate it directly

from a single UI. Users can check the modeling results directly from NIS and can make minor adjustments to the system behavior. In this case, models always include the up-to-date network. Although integrated modeling solutions provide easily accessible data for all NIS users, they may lack tools to validate modeling results and perform more complex simulations. On a critical note, analyzing and validating modeling results require special expertise. Relying on incorrect information can lead to catastrophic results and danger to the water quality for real water users.

Hydraulic models can also be integrated with the supervisory control and data acquisition system (SCADA). With SCADA integration, all the latest control device settings and water balances can be automatically calculated for each district metering area (DMA). SCADA integrations provide up-to-date information in real-time of the system and can be used, for example, to show water quality in all parts of the network. An example of a SCADA integrated model is "Vellamo", which shows water quality in real-time in Tampere, Finland (Tampere Water, 2021). SCADA integrated models can be built on top of design models and NIS-integrated models.

Regardless of the model creation method, modeling requires proper tools and expertise to evaluate the model and modeling results. The model should behave identically to the real system to create valid results. The model should have control systems inbuilt to react to changes in the system correctly. Expert users should validate the modeling results to avoid decisions made based on misleading data. The more automatic the model build-up process will be, the more critical it is to fix flaws in the initial data. Automated processes will increase the quality requirements for all input data. With incorrect data, the model gives incorrect results.

2.3 Fluidit Water

Fluidit Ltd. is a Finnish start-up company founded in 2017. Fluidit develops and distributes hydraulic modeling software to analyze water distribution, sewerage, stormwater, and district energy systems. Fluidit also provides consulting services related to these systems.

Fluidit Water is a modeling software created by Fluidit Ltd., providing tools to model, analyze and optimize water supply systems from the deepest of wells to the furthest of water consumers. Fluidit Water provides a model-based analysis solution for water supply system networks and operations design. (Fluidit, 2020) In this thesis, the network is imported to Fluidit Water, and the program's Jython console is used to import scripts created in this thesis. Fluidit Water uses an enhanced version of the EPANET simulation engine to simulate the network (Fluidit, 2021).

2.4 EPANET

EPANET is a computer program created by the U. S. Environmental Protection Agency that performs simulations of hydraulic and water quality behavior within pressurized pipe networks. During a simulation, EPANET tracks water in each pipe, pressure at each node, and water level in each tank. It also tracks the concentration of a chemical species, water age, and source tracing throughout the network. EPANET computes friction headloss using the Hazen-Williams, Darcy-Weisbach or Chezy-Manning formulas. EPANET is designed to be a research tool for improving understanding of the movement of drinking water constituents within distribution systems. (EPA, 2000)

EPANET application is used throughout the world to model water distribution systems. EPANET is public domain software that can be freely copied, enhanced and distributed. (EPA, 2021) In this thesis, the water utility data is processed to a format accepted by the simulator. For example, the simulator requires that water demands connect to a water source via a pipe network, and that all of the pipes must have diameter and roughness values.

3 WDS MODEL COMPONENTS AND THE PROPERTIES OF INPUT DATA

The various layers of the water distribution system model are shown in Figure 2. The inputs to the models are network data, water storage dimensions, water consumer data, flow and pressure measurement history data, control station settings, pump characteristics, and digital elevation model (DEM). In Finland, all input data are provided by the water utility, except the DEM, which the National Land Survey of Finland provides.

The demand model (Figure 2, orange color) consists of water consumers, leaks, and hydrants. Water demand sets requirements for pipe network model (blue) and hydraulic equipment (purple). Control system (yellow) controls hydraulic equipment: for example when to increase or decrease pressure and where the flow is directed in the system. The ground elevation model (dark blue) provides elevations to the pipes, as the elevation data is seldom included in the water network data.

MODEL AND PHYSICAL LAYERS

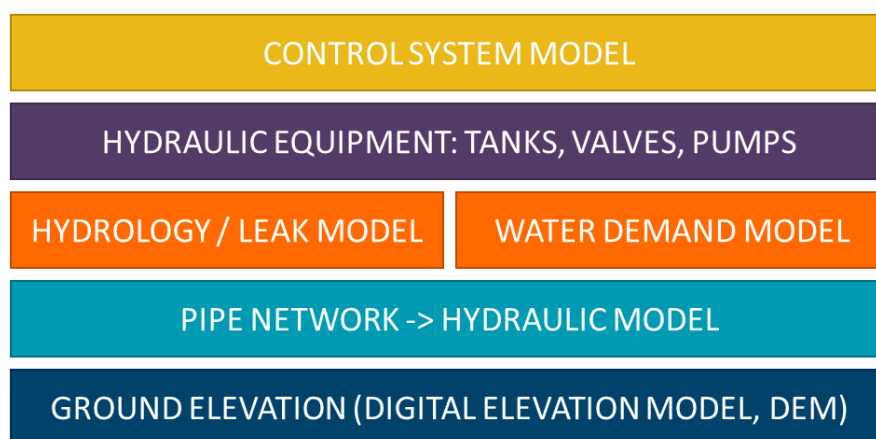


FIGURE 2. Hydraulic model components (Fluidit, 2021)

3.1 Pipe network

Water utility may have hundreds of kilometers of pipes constructed from various materials, ranging from new plastic pipes with thin wall thickness to asbestos cement pipes that are more than 50 years old. Network information contains pipe and gate valve data and the locations and names of pressure booster stations, measurement stations, tanks, and reservoirs. Network data is stored in a network information system (NIS) in a local coordinate system. In Finland, the regulation requires that utilities have converted their network datasets into a digital format by the end of 2016 (Finlex, 2014).

For WDS models, the critical properties of pipe network data are geographical location, pipe diameter and material (roughness). Usually, the pipe sizes are stored as nominal diameter, not as inner diameter, and thus may require further processing. Pipe material information may not use material standards, and pipes may have multiple different values for the same material, for example, "plastic", "M", "Muovi". Secondary pipe attributes are pipe construction year, id, category (known as a classid or type name) and comments, all of which are not used in the simulation but can be used as reference information. The most important of the secondary attributes is the id, which can be used for transferring the simulation results back to the NIS and for updating the model. Different pipelines can be

identified with a pipe category, such as water mains and water cooperative networks. Comment information may reveal some critical information of individual pipes, for example, "crossing pipes are not connected".

In this thesis, attention is paid to the correctness of the network topology. Network topology errors are discontinuities in the pipe network. The NIS may not include a data model which would connect pipes, for example, to create route selections between locations. Errors in the network topology may not be visible from the NIS user's perspective. From the hydraulic model point of view, leaving gaps in the network model can compromise the accuracy of the modeling results.

Topology errors can occur in the NIS, for example, when field measurements of pipe location points (or lines) are imported to the NIS, but line continuity is not validated. Pipelines can also be drawn by hand with too distant zooming and without snapping to existing endpoints. NIS UI may prevent a user from zooming in enough, leaving almost invisible gaps between the pipes, almost impossible to find.

In addition to the pipe network, NIS also stores gate valves (isolation valves). Modeling gate valves are optional, as every pipe in the model can be closed. However, as some pipes cannot be closed in the actual network, adding the gate valves to the model is recommended. Gate valve input data include location, status whether the valve is open or closed, and usually an id, which identifies to which pipe the valve is connected. Gate valve's physical location may differ from the pipe to which it is connected if either the pipes or valves do not have measured locations. In NIS, the gate valves should be located at the pipe, or the pipe moved to the valve's location.

In the NIS, the modeler can usually find physical locations (and possibly names) of pressure boosters, control valves, measurement stations, hydrants, tanks, and reservoirs. In most cases, the essential information about these stations is stored elsewhere, like in the SCADA system or water utility's network drive. However, knowing where the stations are is crucial when calculating water balances for each district metering area and when modeling the stations themselves.

The pipe elevations can be imported to the model. In practice, the pipe elevation data has many variations in quality, resulting in inconsistent long section profiles. In the WDS models the author has created in Fluidit, the measured pipe elevations are seldom imported to the model. The measured elevations are only applied to the pressure booster stations, control valve stations, and tanks. The pipe network is mostly modeled to the terrain elevation, showing a slight pressure offset between modeled results and reality, depending on how deep or above the modeled pipe is. The assumption is valid, as the focus is on street level pressures at the consumer locations.

High-quality network data includes the correct topology, pipe sizes, and materials. The pipe data should not contain duplicate pipes, and the pipes which are not in use should not exist in the data, or they can be identified with a category identifier. Valve states should be documented in the NIS, especially for the ones which are closed. If the modeling process highlights errors in network data, those errors should be corrected directly into the NIS, not (only) in the model.

3.2 Water consumers

Water utilities document the amount of billed water consumed by each customer. Typically, the water usage has been measured using a mechanical meter, and the data is stored as yearly estimates.

Finnish law requires that after the entry into force of the legislation on 23 November 2020, remotely read water meters must be used in new buildings for which a building permit is applied. Existing housing companies adapt gradually to the apartment level billing based on actual consumption, as remotely read water meters should be installed to each apartment during water and sewage system renovations. (Ministry of Economic Affairs and Employment, 2020) Some water utilities in Finland have decided to switch to using only remotely read meters by systematically replacing the meters for every client during the next few years.

For modeling purposes, the water consumer data must include the water usage and a location field: coordinates, address, or a property identifier. A field for recognizing industrial demands is helpful from the modeler's point of view. Consumer IDs can be used to update existing water demands. In Fluidit Water, the demands are usually added to the model as individual users, and the demands are connected to the nearest junction.

Water demand data may include sensitive GDPR data, which could prevent adding water users to the model. Name information from individual customers is not required or utilized in modeling. Water demand locations can also be obscured: With the list of all model junctions, the water utility could join the consumer data to the nearest junctions. Knowing where the demands connect in the model, the demands could be imported without geocoding but still recognized as individual consumers (adding a custom pattern is possible). From the model results point of view, it does not matter if the water demands are individual users or a pre-grouped set of demands: after all, the water consumption is taken from the junctions, not from the demands. Exact list of users may help when analyzing and reporting the results properly, as the elevation difference between the consumers and the network can be considered.

Continuous measurements from individual water consumers are becoming increasingly popular in Finland. This new data source may provide great insight for model calibration purposes. The new data source could allow improved leak localization modeling and the creation of user-specific patterns for the most demanding water consumers.

3.3 Flow and pressure history data

To understand the water distribution system behavior, measurements made in the system are crucial. The most informative variables for hydraulic modeling are flow, pressure, and water level. Flow and pressure variables are commonly measured before and after all the pumping stations, control valves, and measurement points. Water level data is gathered from water towers and water tanks. All history of measured data is stored in the SCADA systems history database. One year of hourly data is usually exported from the SCADA system for further processing for the model.

Later, when hydraulic model components (shown previously in Figure 2) have been modeled, the history data is used for model calibration. The modeling results are validated in the calibration by comparing them to the flow and pressure history data. This phase should not be skipped even if all model components could be imported directly to the model. If differences are found between modeled and measured flow/pressure, the model does not match the real system, and fixes must be made. Most commonly, the difference has been identified as a non-documented closed gate valve, error in model continuity, or unmatched control setting. Pipe capacity can also decrease over time, which may not have been considered in the model.

3.4 Control device settings

The SCADA system includes how the water distribution system is controlled. Control can be automatic by user-defined control settings or manually controlled by a system operator. Settings of pumps and control valves are displayed in SCADA UI (example in Figure 3). For WDS models, the current control system settings are replicated to the model. SCADA systems rarely have functions to export current settings easily. The current control settings are exported as screenshots or notes for further processing. This processing is done manually and requires expertise: the modeler must know where the station is, how the station is controlled and how the stations would control if the system changes, for example, during a pipe burst.

EPANET requires tanks to operate between a minimum and maximum level (EPA, 2000). The model control system should respect these boundaries, and the model tanks should not run empty or overflow during the simulation.

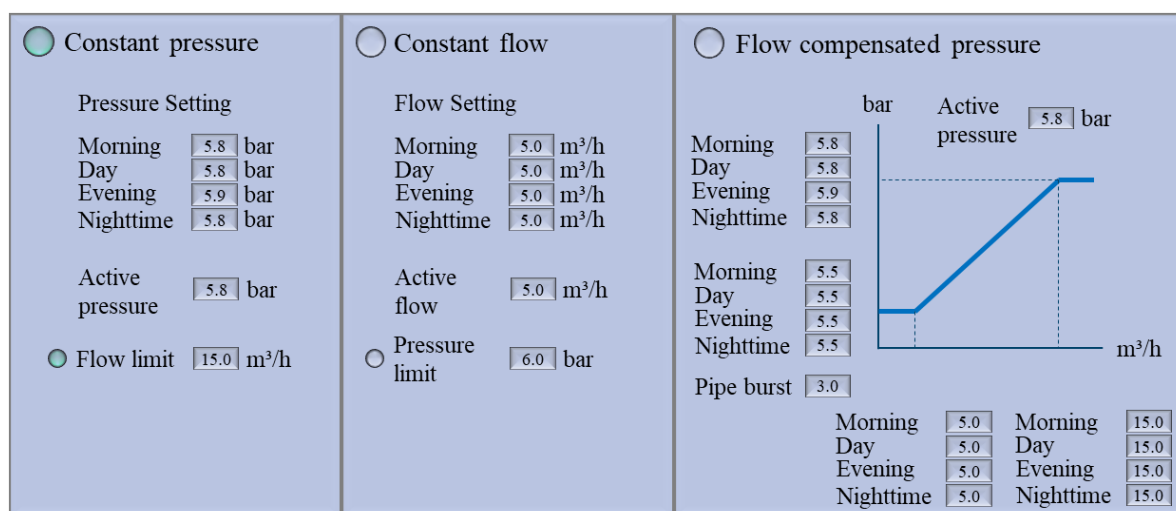


FIGURE 3. Example of pump control methods in mock-up picture of a SCADA system. The pressure booster controls outgoing pressure with a constant pressure setting alternating a bit during the day. Pump stations flow limited. Two rightmost control types are available but not in use.

3.5 Pumps

In water distribution systems, pumps deliver potable water into the network (Figure 1). Elevation and network capacity may require pressure booster stations (Figure 4) to increase the pressure in some parts of the network. From a hydraulic model point of view, pump performance is replicated using

pump performance and efficiency curves. For modeling a pump, motor performance, nominal motor power, efficiency at full and $\frac{3}{4}$ loads, and IE-class are used. Water utilities usually store pump models and pump curve information in a spreadsheet on a network drive or paper format. If the pump curves are not available, some of the pump characteristics can be found in the online pump catalogs. Network information system usually includes the locations of pressure booster station which is crucial for understanding how the system behaves overall.

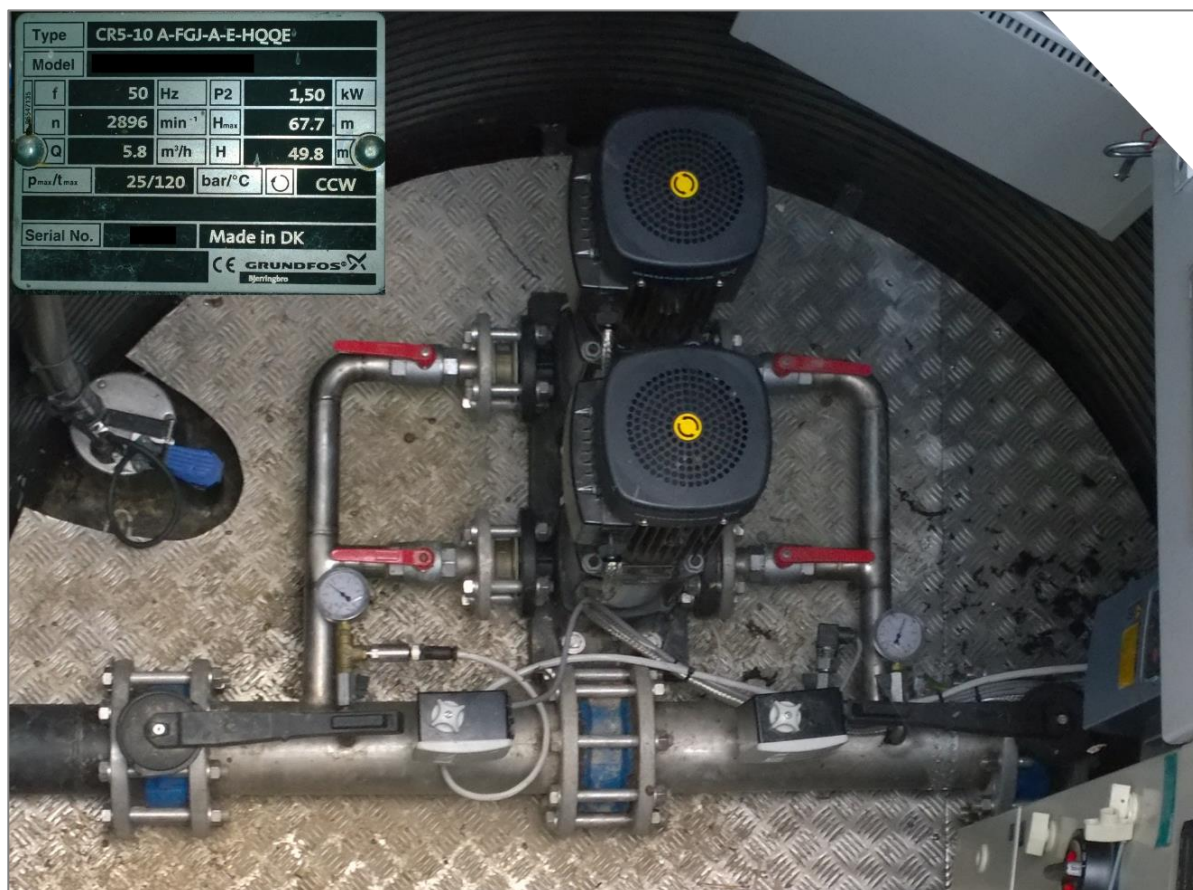


FIGURE 4. Pressure booster station with two pumps. A pump plate is used to find pump curves using online catalogs (Väyrynen 2015)

Modeling of pump characteristics is still mostly manual work. Although pump performance and efficiency curves may be available in online catalogs, the database is rarely exportable in numerical format. Allowing proper export functions should be in pump manufacturers' interest as well. For example, using a pre-created pump library, hydraulic modeling software could recommend a pump model based on different simulated scenarios (for example, typical day use, fire flow, future forecast).

3.6 Tanks

Tanks are used to provide storage capacity to meet fluctuations in demand, provide reserves for fire-fighting use and other emergencies, and equalize pressure in the distribution system. The storage facility can be elevated or in-ground. (EPA, 2005)

For WDS models, tanks are modeled based on structural drawings. Tank volume curves are created manually if the tank cross-sectional area changes in different water levels. As tanks can be many decades old, all elevation markings should be checked and converted to the current elevation system. As

a part of the modeling process, the time used in the modeling of tanks is insignificant. Therefore, there is no practical reason to improve the tank import process, but the tanks existing in the model should be preserved when updating the network model.

3.7 Digital elevation model (DEM)

Terrain elevation models are used in the model for adding elevation for junctions and water consumers. For WDS models, an elevation grid of 10x10 meter is sufficient. In Finland, data can be accessed freely from National Land Survey of Finland. Abroad, custom grid DEM models (for example, TIFF files) can be used. As adding a DEM happens almost instantly, there is no need to speed up this part of the model build-up.

4 WDS MODEL BUILD-UP PROCESSES

Figure 5 summarises how models are created in Fluidit Water. Phases below the blue line indicate which part can be done using the modeling software at its current state. The figure is read one column at a time from top to bottom. The bottom cell contains the component name found in the model.

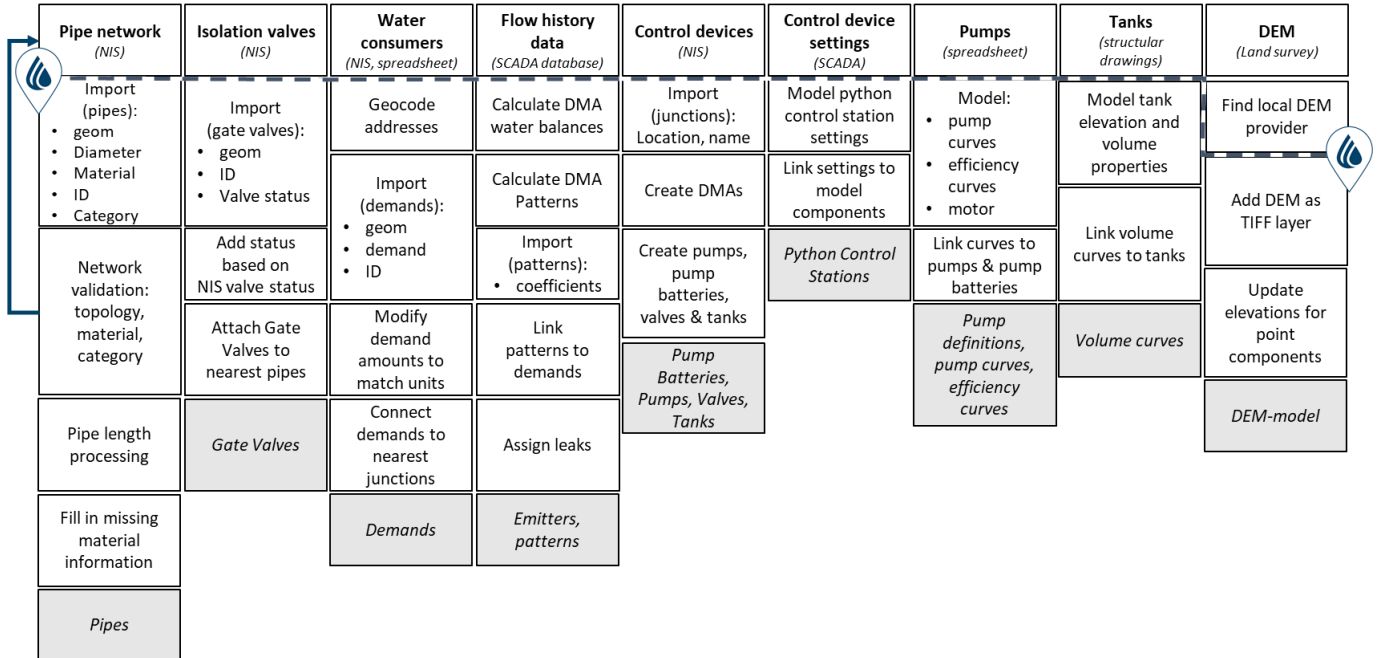


FIGURE 5. Data processing workflow from each data source to final model component

4.1 Processing digital elevation model (DEM)

When starting a modeling project, adding a digital elevation model (DEM) is recommended (Figure 6). Elevation can be defined as ground elevation or the user's defined offset of the ground. The DEM may use a different coordinate reference system than the model components themselves. When using a DEM, the elevations are automatically added when new components are added to the model.

Apart from the DEM, exact measured elevations are used in storage tanks, reservoirs, and pressure measurement nodes. Using exact elevations in pipe networks can also be done, although variation in data quality may result in strange-looking hydraulic profiles. Measured pipe elevations can also be updated to finished model junctions later if preferred.

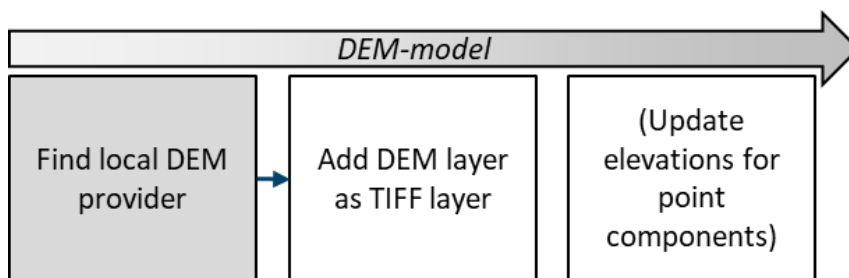


FIGURE 6. Adding of digital elevation model (DEM) to the hydraulic model

4.2 Processing network

4.2.1 Creating network model manually

Some years ago, models were built on top of the background map (from NIS or scanned paper maps). The method was good topology-wise, but it lacked precise location information. Approximated locations complicated model updates, as the model topology was different from the background network. With the manual build-up, only the critical attributes and features were added to the model. Other useful NIS information like identifiers, class information, construction years and comments, were not added into the model and were not used when reports about the modeling results were written.

A small test was conducted to measure how much network can be modeled manually using Fluidit Water software within five-minute intervals. Pipes were added on top of the network map in the demonstration area shown in Figure 7. As created, pipes were added material and diameter information using Fluidit material library functionalities. Other fields were not added (like year, id, pipe category). In the four 5 min intervals, manual modeling speed was measured from 3.1 km/5 min to 4.6 km/5 min of a network in each of the sessions. Based on this small test, the average manual network modeling speed was around 44 km/h if this intensive pace could be kept continuously. This result can be used as one way to validate the model build-up process in terms of speed.

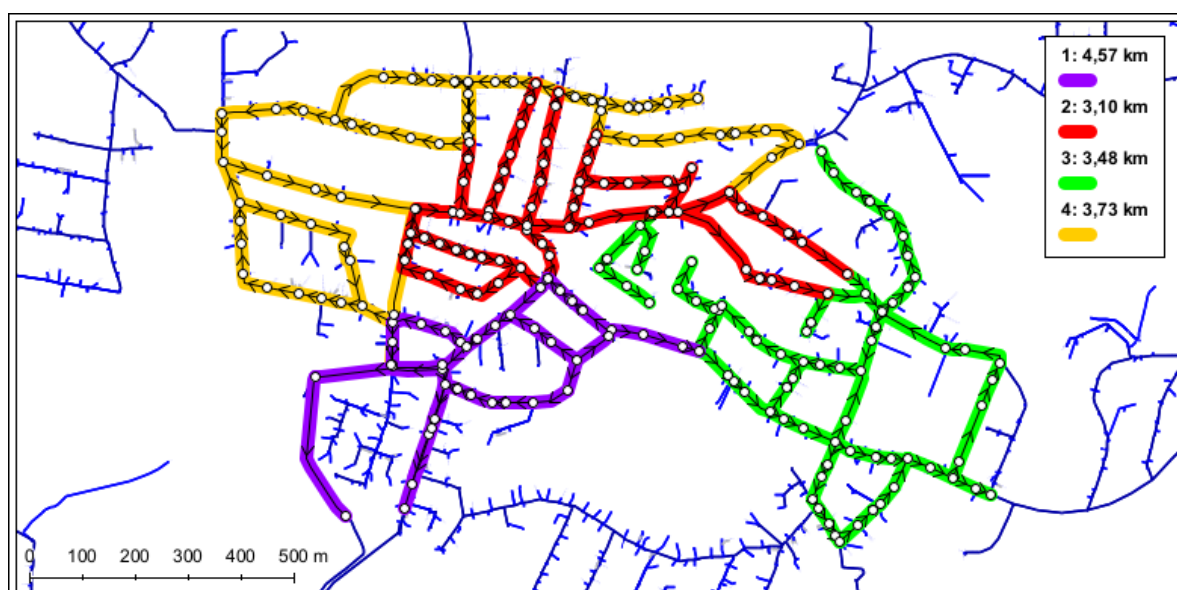


FIGURE 7. Creating network by manually adding pipes on top of the network map

4.2.2 Creating a network with import tools

Import Components to Model is a tool in Fluidit Water to import and update model physical and non-physical components. The tool allows importing different external data sources (like .xlsx, .geojson, shapefile) one file at a time.

After a file is selected, the import dialogue opens (Figure 8). A model component type from the top-left dropdown menu is selected from a list of all possible model components. File Coordinate Reference System (CRS) can be defined to apply coordinate system transformation when importing. If the

user wants to split the imported links with existing model nodes, he/she can do it with the "Allow Splitting Links" option. The "Match Type" setting includes importing features as new components or updating existing components based on a common attribute. At the bottom of the import dialogue, all columns of the original data source are displayed, and the user can pick which fields are brought into the model. After the Import button is pressed, model components are generated into the model.

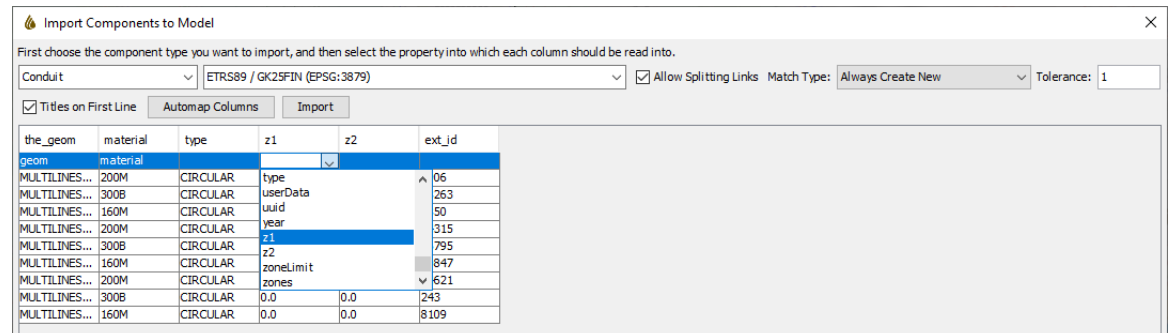


FIGURE 8. Import components to the Model tool, input of field values

The import itself, in its current form, does not include processing of the imported data. For example, it does not check network connectivity when importing pipes, and it does not validate flow measurements when data is imported as time series. Respecting the original data is a priority to keep the results of the import consistent.

As the import tools have been created and improved throughout the years, now it is more efficient to import the network data to the modeling software and check its topology than to do the modeling manually. As attributes from the input network data can be added at ease, there is no point in leaving useful data out of the import. For example, the identifiers can be imported and used as a reference point when writing modeling reports. The comments added by the water utility's personnel can help understanding some gaps in the data. Later, the identifiers can be used for updating the model and transferring the modeling results back to the NIS.

The current workflow for creating a pipe network model is shown in Figure 9. In the leftmost part is the input data (export from network information system). The input data is imported into the model for topology validation. After the topology validation is finished, the results are reported back to the NIS to improve the original data. Ideally, network data would be iterated through the process multiple times until there won't be need to fix any errors in the model. After the connections are fixed, the pipe materials and category identifiers are validated. The results of this validation are reported to the NIS for fixing. If the model must be built up fast, fixes are made directly to the model. At the last step, pipe lengths are processed. To minimize unnecessary pipes in the model, ensure that pipes include elevation at the highest and the lowest point in terrain elevation (requires Digital Elevation Model).

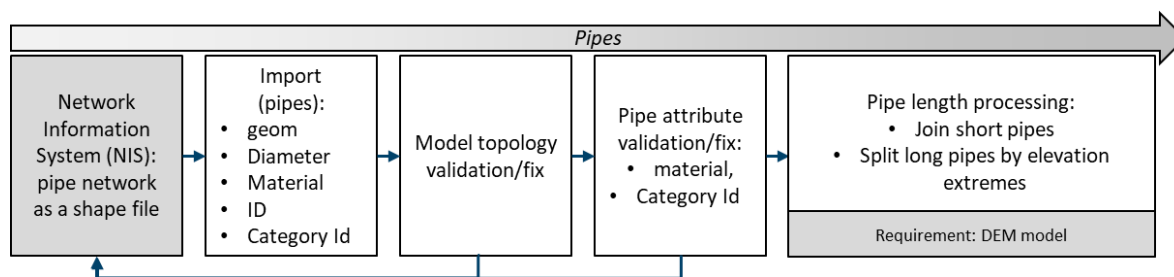


FIGURE 9. Data Processing workflow for network data to create a pipe network model

The model build-up process should avoid any manual preprocessing of input data. The current process requires pipe network data and gate valve data preprocessing. In preprocessing, excess networks, household pipes, and abandoned pipes are removed from the network data. In some cases, diameters and material are stored in a non-human readable format, and in these cases, those codes are transformed into human-readable form with a correlation table. Preprocessing phase should not be manual work, and this phase can be automated, for example, with custom templates made for different water utilities and NIS providers.

After preprocessing, the network is imported into modeling software, and a network continuity check should be made. WDS models require a topologically connected network model. Network topology errors are discontinuities between pipes and leaving those errors into the model can lead to incorrect modeling results. Pipes documented in wrong categories can create gaps in the model, as some of the classes may have been removed in the model preprocessing phase. Network with duplicate pipe geometries may result in too much capacity in some parts of the model.

Network topology can be corrected with the tools found in modeling software, GIS software, or NIS. It is recommended to correct topology flaws in the original data source (NIS) and use the corrected data to re-create the model. If the model must be created quickly, topology errors are fixed directly to the model, but the list of topology errors should still be reported to the utility.

In addition to the topology flaws, network data may include pipes lacking diameter or material information. Ignoring these would affect the capacity of the hydraulic network model. Missing pipe sizes and materials may be approximated to some extent if the pipe is continuing without branching and at least one of the nearest connections has these attributes. Note that recognizing branches require correct topology! Pipes without diameter and material and pipes which use estimated attributes should be reported and fixed in NIS. Typing errors made in diameter and material fields should be recognized and later fixed by the water utility. The modeling process should list all the incorrect values and report which pipes use incorrect values to correct the original data. For NIS providers, open fields should be avoided in diameter and material input fields and the use of a pre-made list of pipes should be encouraged to avoid variations in the documentation.

4.2.3 Pipe material library

Fluidit pipe material library, as shown in Figure 10, is a feature to combine pipe characteristics into a readily usable form. Using the material library is optional, and pipes can also be modeled with individual pipe diameters and roughnesses. Material library objects have names (usually concatenated from

pipe nominal diameter and material), roughness (Chezy-Manning, Darcy-Weisbach, and Hazen-Williams equations), and inner diameter. Also, a cost-estimate value can be assigned to each material. In the pipe component, the material library is referred to by the material name. When the material is selected for the pipe, pipe diameter and roughness values are updated to the edited pipe(s).

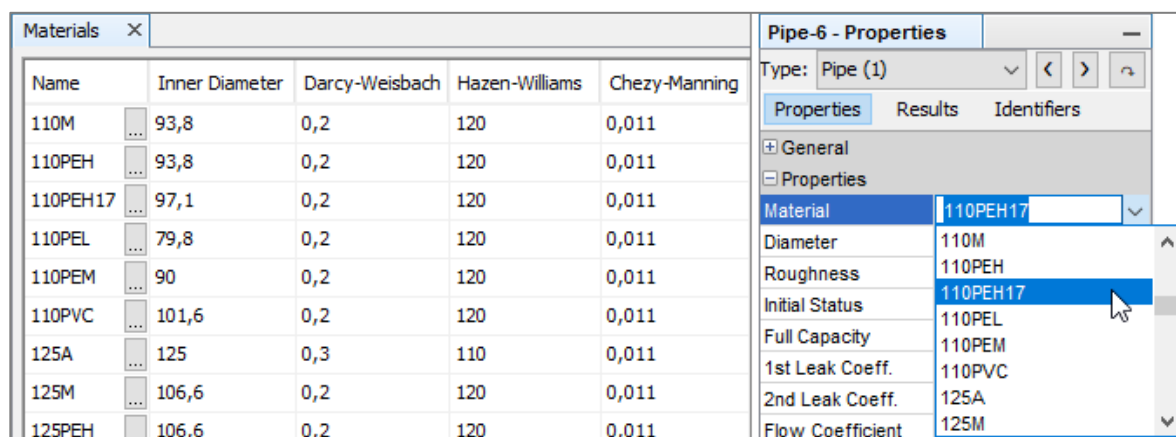


FIGURE 10. Material library in Fluidit Water and selecting a pipe material from the library to a pipe

The current Fluidit Water material library, referred to as Finnish Material Pack, includes the commonly used plastic, cast iron, and asbestos cement pipes. Pipe materials are the most used pipe materials in Finland. Pipe diameters are standard. For example, 160PEH pipe in material library corresponds to the dimensions described in the SFS-EN 12201-2 standard for 160PEH-pipe with SDR class 13.6. SDR is the ratio between nominal outside diameter and wall thickness. SDR 13.6 pipes include the nominal pressure of PN10 (maximum pressure of 10 bars) and PN12.5 pipes. SDR17 pipes are also included in the material library and expressed in postfix, for example, 160PEH17. According to one pipe manufacturer in Finland, the most used SDR classes in Finland are 11 (PN12.5/PN16), 17 (PN8/PN10), and 26 (PN5/PN6) (Muoviteollisuus RY). In practice, pipe pressure classes are rarely documented in the NIS, but the water utility may prefer one SDR class when installing new pipes to their system.

If the network material does not match with the material library, the material name should be edited to correspond to the material library, or a new material should be added to the library. The material library can be extended with new materials, and the existing materials can also be edited. Edits made in the material library automatically update the pipes' diameter and roughness, which refer to the edited material.

4.3 Processing gate valves

The processing of gate valve input data is shown in Figure 11. Gate valves are point data sources exported from the network information system. Gate valves can be imported directly to the model. Gate valve data has an attribute to define its status (is the valve closed or open), which may not correspond to the model attributes (true or false). In the manual process, Gate Valve status is imported into Description or External system fields, which are then used to identify and close the correct gate valves in the model. If the model build-up process is extended to include importing gate valves, valve statuses could automatically be transformed with custom templates. The templates could be tailored for each of the water utility and network information system providers.

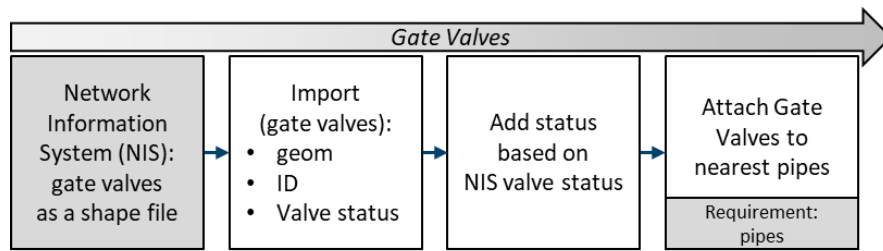


FIGURE 11. Data processing workflow for gate valve data

4.4 Processing stations

The station model includes the hydraulic equipment model of pumps, valves, tanks, and reservoirs. The overall process of creating a station model is shown in Figure 12. Locations of hydraulic equipment are exported from the NIS and imported to the model as junctions with station names. Stations are modeled with the corresponding model component and attached to the pipe network.

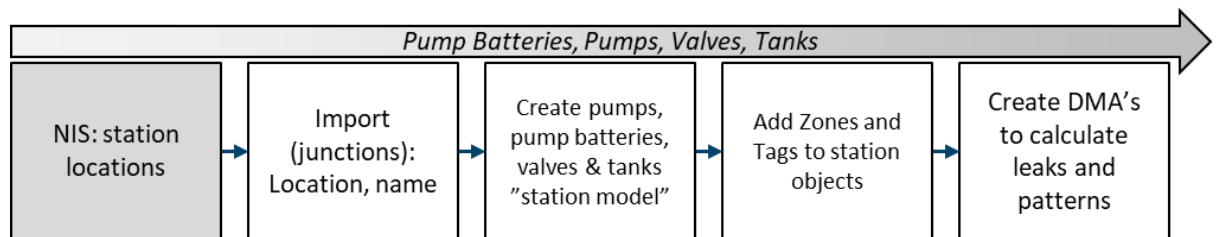


FIGURE 12. Data processing workflow for control devices and tanks

Stations include location, control device properties (such as pump curve, tank elevation), and how the station is controlled. In Figure 13 map view, model stations are highlighted as yellow. Stations are connected to the network model pipes (blue). DEM is not required to create a station model, but it can be used if not all elevations are known.

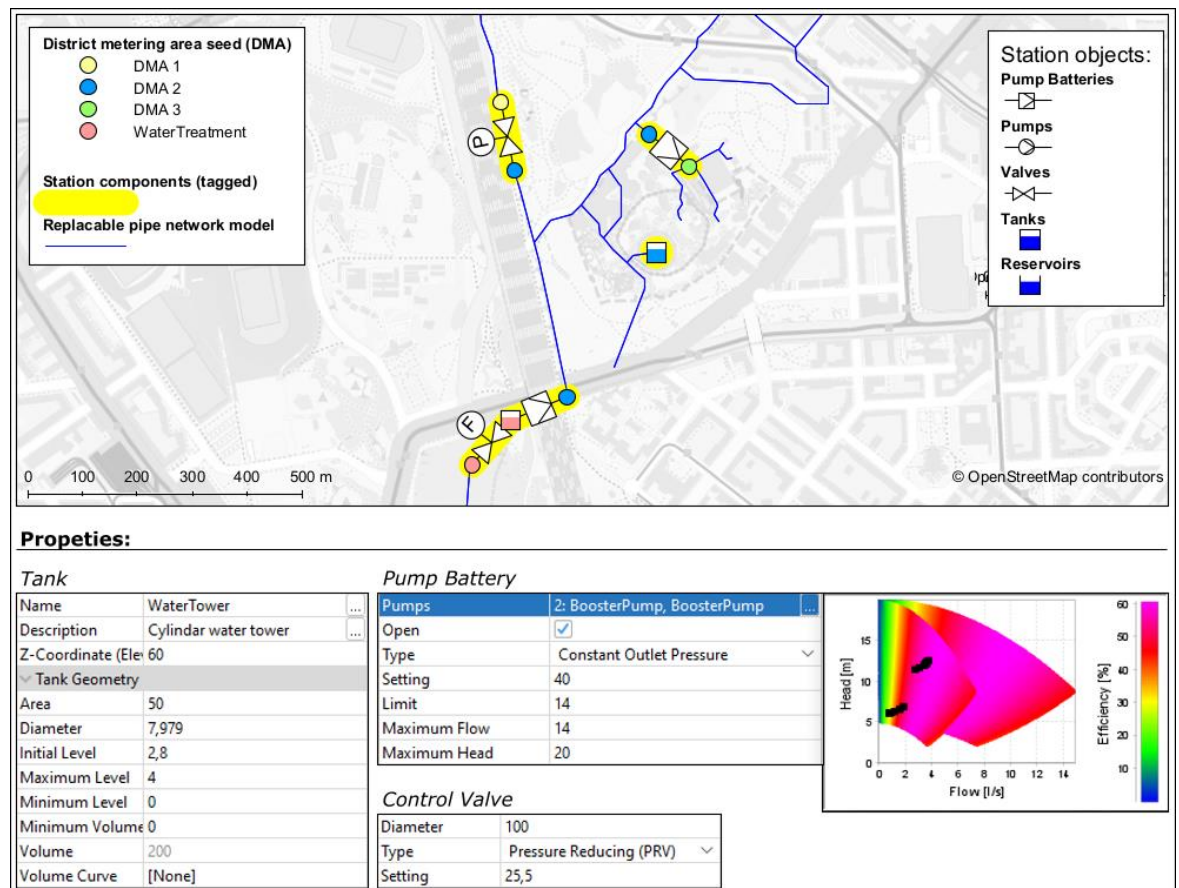


FIGURE 13. Station objects with zones and tags included. Control device sizes and lengths are exaggerated for visual purposes.

Creating a station model requires some manual work. Usually, the created station is still valid after many years of modeling work: station locations, elevations, and working principles do not change, although the control setting may be altered. Stations are preserved when the network model is updated. To easily distinguish the stations from the pipe network model, it is recommended to tag stations and the points connecting to station objects. DMA names are added as zones to tanks, reservoirs, and points before and after a pump or a valve. Zones can be programmatically added to a connecting network of pipes and junctions, helping to assign leaks to each DMA and adding patterns to model demands based on the parent node's zone name.

4.4.1 Processing tanks and reservoirs

Modeling of storage units includes adding tank bottom elevation, volume, and maximum level. If the tank area changes in different water levels, a volume curve is created and linked to the tank.

Reservoirs are modeled accordingly to the water head in groundwater, raw water, or clear well (if the treatment process is not modeled). Water sources can be added to each reservoir as quality parameters.

4.4.2 Processing pumps and valves

Pumps are modeled either as pump or pump battery components based on the control type. The pump battery component allows driving the pump based on the control setting; thus, the pump curve

is not mandatory. Pump component is used if the pump is controlled with on/off setting, freely based on a pump curve.

The pressure booster station's pumps are modeled by adding pump performance curves and efficiency curves to the model (as in EPANET). Fluidit Water's "Pump Definition" combines performance and efficiency curves, as well as motor information (nominal motor power, motor efficiency in full and $\frac{3}{4}$ loads, and the motor IE-class) and variable-speed drive. By adding the pump performance curves to the model, the flow and head of the pump can be limited by the curves. Efficiency curves and motor characteristics allow calculating the energy efficiency of pumping to identify pumps running in a non-cost-effective way.

Valves are modeled based on their control type. For example, valves can control tank filling or pressure in the pressure zone. Like pumps and pump batteries, the valves can have control based on SCADA settings.

4.4.3 Processing controls settings

As in the real system, the WDS model is also controlled with pumps and valves. The water utility's present (or the most frequently used) controls settings are replicated to the model from the SCADA controls (Figure 14). There are multiple options to control stations in the model. The most straightforward pump or valve controls can be directly modeled using model component properties, allowing simple constant pressure/flow control. More complex control can be modeled using EPANET control settings or with a Python script. The latter one is recommended, as the Python control stations can be directly linked to the model stations, allowing easy access for editing the station controls later if the station's controls are changed.

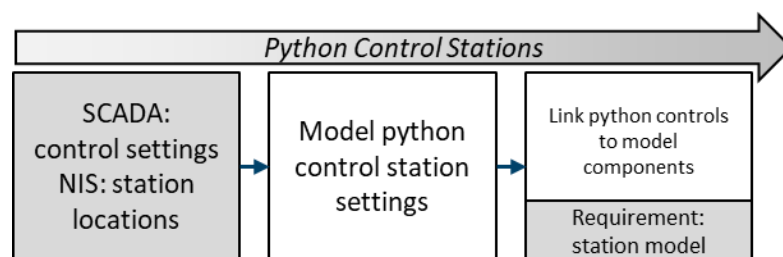


FIGURE 14. Modeling of control SCADA control settings

In order to update the control station settings automatically to the model, the SCADA control system should have proper export functions available. As the SCADA systems cannot be directly accessed externally due to security reasons, the best available export method is screenshots of the control settings and updating the model station settings once a year or more. More advanced update methods have been developed with few clients, allowing the update of the control settings to an online model as they are changed. For design models, one option would be to identify control settings changes from the measurement history data and using this data to update the model controls. However, this method would not result in actual control settings, and it may be hard to identify the control type, for example, if the control is based on flow, pressure, water level, or water consumption estimate.

4.5 Processing water consumers

The data processing workflow of water consumer data is shown in Figure 15. Water consumer data is provided as yearly measured or estimates in an excel sheet. If the input data lacks coordinates, data is geocoded based on address or a property id. Geocoding is possible, for example, with Bing or Google APIs, freely with Open Street Map Nominatim (coverage may vary) or using a national data-base for streets, for example, "avoindata.fi".

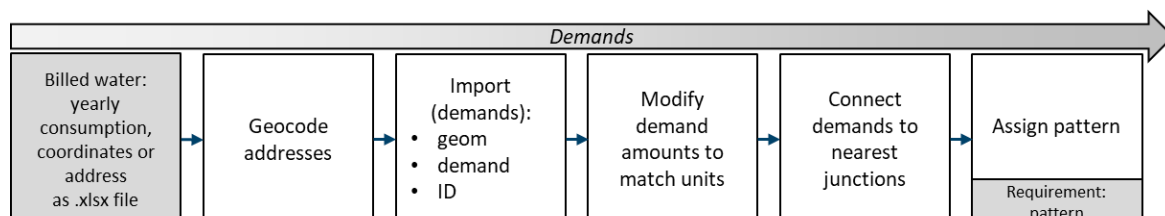


FIGURE 15. Data processing workflow for water consumer data

Geocoding results should always be checked in case of wrong locations. At a minimum, check if there are multiple demands at the same coordinates. Online geocoding may provide a rough location, like the town center, if the exact address is not found. Leaving this unnoticed may decrease model quality significantly.

After geocoding, locations and yearly estimates are imported to the model. In Figure 16, the demands (purple and red) have been successfully connected to the nearest junctions (black points) using an inbuilt tool in Fluidit Water. When connected, water demand is taken from the connecting junction without any hydraulic losses in the connector line (light grey line in the picture). Demand elevations and floors can be utilized when analyzing the model results. Demand patterns are set to each DMA after patterns are calculated.



FIGURE 16. Demands connecting to the nearest junctions in Fluidit Water

4.6 Processing emitters and patterns

The data processing workflow of flow history data is shown in Figure 17. Calculating leakage amount and how water is consumed in each of the DMA's requires general information on how the system

works overall. The modeler must know where each station is located (usually found in network information export) and which direction water is flown (using SCADA control settings). Flow history data is used to calculate median water demand and hourly demand in each district metering area (DMA).

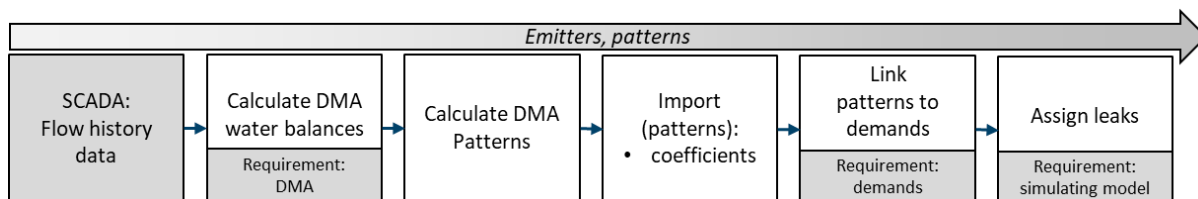


FIGURE 17. Data processing workflow for flow history data

In Figure 18, a schematic of DMAs and (continuous) measurement stations are shown. For example, to calculate water usage and leakage in DMA 2 (pipes with grey background), the modeler must know how much water is pumped to the system (from the southmost pumping station) and how much water is taken from the system to other district metering areas (DMA 1: yellow and DMA 2: green). As the DMA 2 also has a water tank, flow to the tank should be subtracted from the water usage and flow from the tank added to the water usage. Tanks may include flow meters for incoming/outgoing water. If this information is not available, the water flow and direction are calculated using water level measurements.

District measurement areas can be grouped to form larger DMAs. For example, if the flow data of DMA 1 is not usable or a more simplified model is preferred, the area can be joined to the DMA 2. In that case, the flow data of the control valve station is ignored when calculating DMA leaks and patterns.

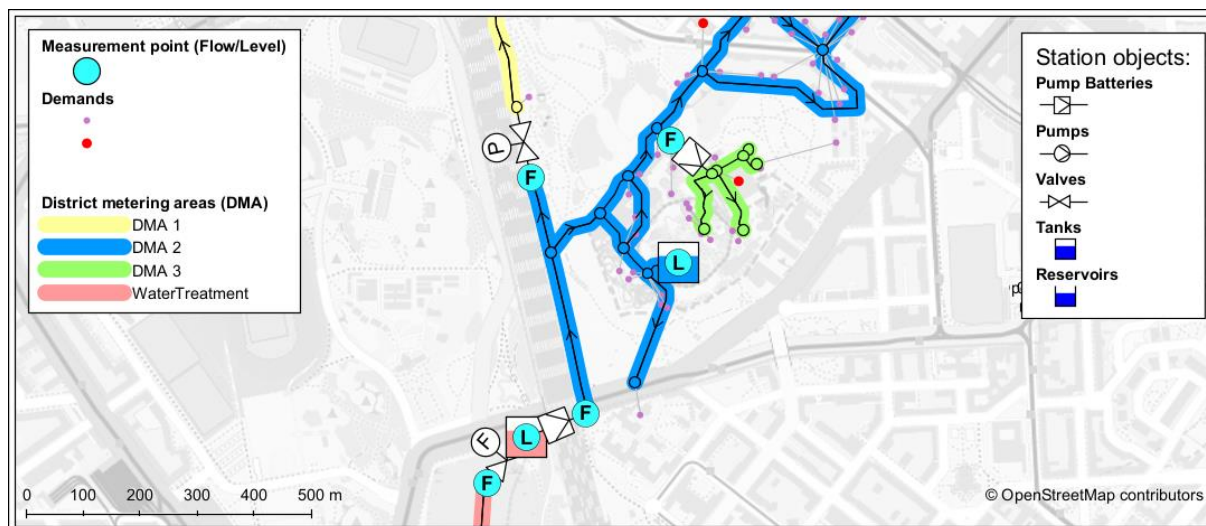


FIGURE 18. District metering areas in Fluidit Water model. Control device sizes and lengths are exaggerated for visual purposes.

In Figure 19, an example of a water demand analysis for one DMA is shown. In the dark green line, the median is used to assign leakage and pattern coefficients in this DMA and validate model results in the average consumption day scenario. Within the light green color area, 90 % of all recorded hours are shown. This area represents the variance in day-to-day water usage. The maximum values of this area can be used for validating the model results in high consumption day scenarios. Absolute

maximum hours in each of the hours are shown as red. Absolute maximum values help to identify, for example, industrial water usage and anomalies in the data. For example, due to measurement errors or pipe burst, flow measurements can show too much or too little calculated water demand in the system, as Figure 19 at three o'clock. If data errors are infrequent, they can be ignored automatically using the median to calculate the water usage instead of average. If the data is thoroughly misleading, it should be left out, and the DMA should be included in a larger DMA.

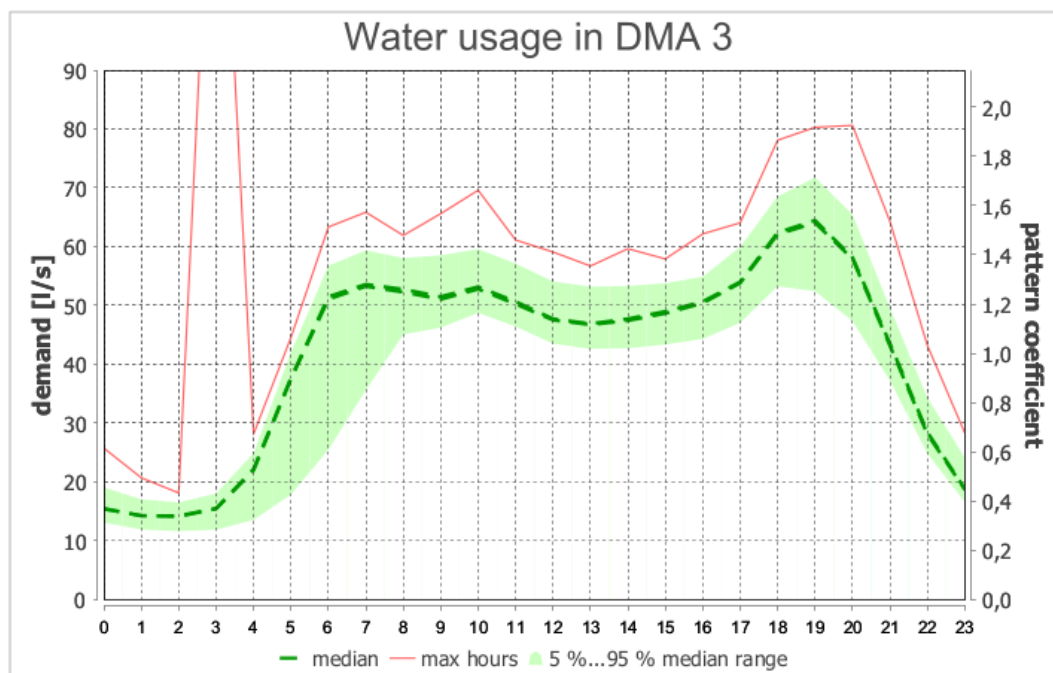


FIGURE 19. Example of water usage in a DMA

DMA's may also "leak" to other DMA's due to valves left open between the pressure zones. These cases may become apparent when the DMA's water usage is compared to the water billing of the same area. A second option is to examine DMA water usage in the long term and compare the usage to each other DMA's.

The consumptions are converted into hourly patterns, which multiply water demand in the model. Adding patterns to model demands requires that demands have been added to the model. Demand patterns can be set manually using zone-selection tools or programmatically if model components have attributes (like zones or tags) to identify each of the DMA's.

If the modeling area includes major water users with highly fluctuating demands, patterns for these demands should also be made. In these cases, using the hourly data may not be sufficient for creating accurate enough patterns and data with smaller resolution should be used.

Knowledge of daily water demand in each DMA is used for assigning leaks and unbilled water based on pressure in each of the junctions. Model leaks are added to the model with the assign leaks tool. The leak to each junction is calculated with average pressure in the junction and connecting pipe lengths and diameters. As this phase requires that the model simulates and the model pressures are on the correct scale, this phase is one of the last steps before the model calibration.

5 IMPROVED DATA PROCESSING TOOLS FOR NETWORK DATA

5.1 Base model and scenario structure

Before adding model components to the model, the model coordinate reference system is changed to correspond to the local coordinate system used by the input data. Finnish material pack is loaded to the model. Network data is imported to the model using 0.01 m tolerance. Import pipe attributes as model external systems ids, using the same names as they are documented in NIS export, for example, id, category, diameter, material, build year, remark.

Network data is imported in the model base scenario. The base scenario represents the original imported data without any processing made in the network. Further, each of the pseudo-codes should be run in child scenarios in order to see the differences between the original and the corrected data. Results created in child scenarios (notifications of gaps and missing attributes) can be copied from child scenario to base scenario to check how the error exists in the original data.

5.2 Pseudo-code for checking network topology

Checking the network topology can be one of the most time-consuming tasks when building a network model, and therefore this simple topology check can save much time, as most topology errors are pinpointed on the map. An example of a topology checking process is shown in Figure 20. The purpose of the topology checking is to find out network continuity errors, regardless of pipe attributes. Only the locations of the pipes are used. As a result, one or two files (shape files) are created, which can be used as a background map to correct NIS data. Optionally, a list of pipe identifiers can also be created if it further helps to locate the problematic cases. This process is repeated until network data quality is sufficient.

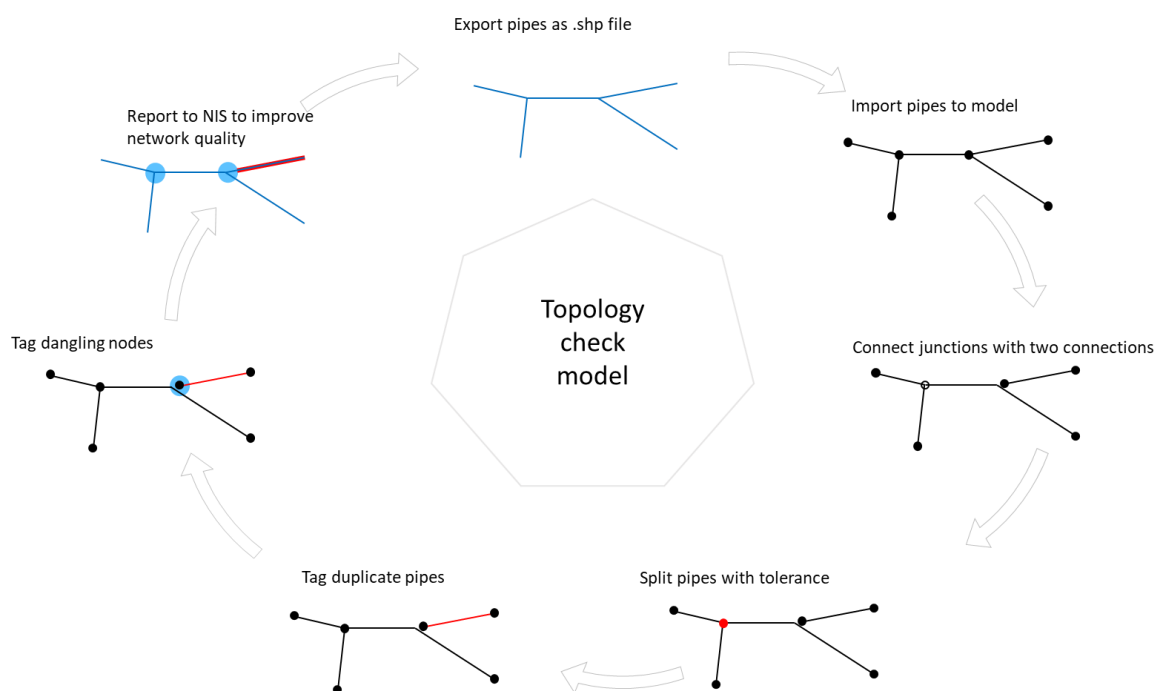


FIGURE 20. Simplified process to report topology errors

1. Import pipe geometry with a small tolerance, such as 0.01 m.
2. Join nodes with two connections.
3. From every junction point of view, if nodes or pipes are close to the node and not connected, mark those locations with point components.
4. For every pipe's point of view, search for other pipes connected to the same start and end nodes. If other pipes are sharing the same junctions, mark those pipes as duplicates.
5. As a result, create GIS-export from all disconnection points and duplicate lines for error report back to the NIS.

Note that topology errors can also be fixed directly in the model to some extent. However, automatic fixing may cause unintended errors in the network, especially if multiple pipes are close to each other. The automatic process does not know how the piping connects in the real system, and some gaps can exist in real life.

5.3 Pseudo-code for checking diameter and material

The purpose of the diameter and material check model is to go through pipe materials and find out the pipes missing these attributes. Network continuity is used for supplementing the data gaps in information; hence network topology should be fixed before this phase. If a pipe is missing diameter or material, it may be supplemented from other connecting pipes if there are no branches between them. As a result, a list of pipes without material is written, and more importantly: the list of pipes for which the material information cannot be automatically determined, is written. In the model build-up's perspective, the data gaps that cannot be automatically filled should be prioritized when fixing the data in the NIS.

1. Import pipes with geometry, diameter, and material to the model.
2. Join nodes with two connections if diameters and materials of both pipes match.
3. If pipe material is missing, try to fill the material based on links connected to the start and end nodes. Update the material if either the start or the end node has two connections, and the next connected pipe has diameter and material. If both nodes have two connections and the next pipe's materials differ, use material from the longer pipe.
4. As a result, create a list of pipes with materials that are still missing and a second list with pipes that have automatically determined diameter and material.

5.4 Pseudo-code for checking pipe categories

This pseudo-code is developed for finding out pipes documented as service lines, which usually are not included in the model. Service pipes in between distribution or transmission pipes should be included in the model, as they increase the network capacity. These pipes may have been documented incorrectly. Thus they are also reported to a NIS maintainer. Pipe categories could also be used exclusively so that only the category would be checked, for example, service pipes.

1. Import pipe geometry with categories.
2. Join nodes with two connections if the category ids are the same in both connecting pipes.

3. Search for pipes with two or more connections at both ends (ignore dead-end pipes). If the pipe category id does not match the next and the previous pipe category id's, mark the pipe to be checked.

5.5 Pseudo-code for valve connections checking

This pseudo-code is to process gate valves imported to the model. Gate valve status is updated to the model based on input data. The gate valves are attached to the nearest pipes within a tolerance. Valve connections to the pipes are tested if the pipe father id matches the connecting pipe id, as the gate valve may have been incorrectly documented to connect to a faraway pipe. Pseudo-code checks whether there are many valves documented in duplicate locations. Duplicate gate valves cannot be connected to the model pipes.

1. Import pipe geometry and id.
2. Import gate valves. Include id and possible father id, which links valve to correct link with link id. Include valve status.
3. Close the gate valves based on valve status.
4. Connect gate valves with a small tolerance, such as 0.50 m.
5. As a result, visualize the gate valves in which the father id does not match the connected pipe id. Visualize:
 - o closed gate valves
 - o detached gate valves
 - o gate valves with duplicate locations.

5.6 Pseudo-code for generating a network model

This pseudo-code (Figure 21) creates the actual network model. Pseudo-code assumes that network topology and pipe categories have been previously fixed; thus, gaps are not checked again. This process aims to provide a network model, which includes all main pipes and gate valves. The pipe network is processed to optimize the number of junctions and pipes in the model (joining of short pipes and splitting too long pipes).

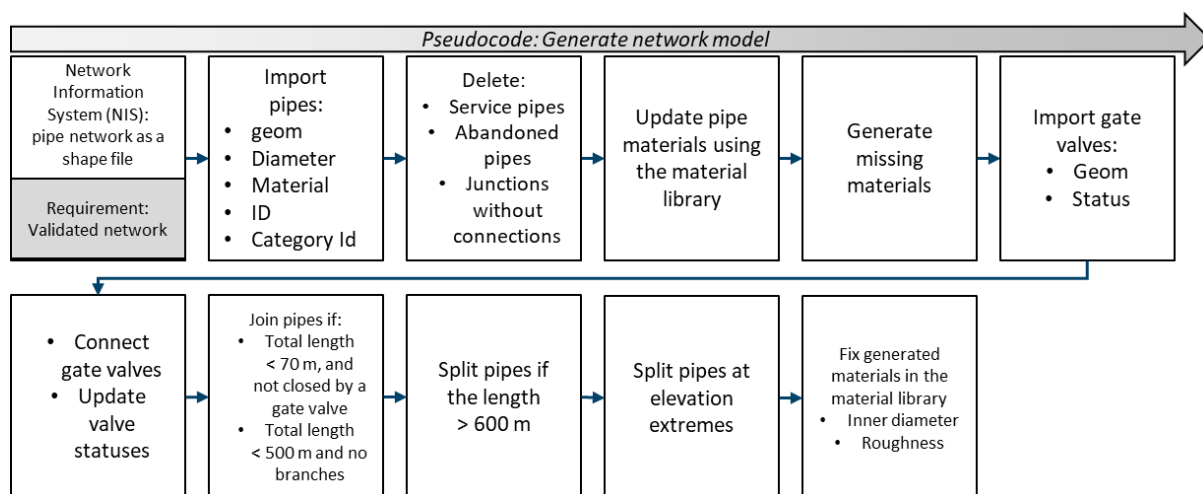


FIGURE 21. The process to generate a WDS model network

1. Import pipe data into a model. Pipe id, category, diameter, and material are imported in respectable external system ids created previously in the base model.
2. Using pipe category, remove service pipes and abandoned pipes from the model. Delete junctions without connections.

3. Update pipe material fields from the external system identifier fields.
 - a. If external system diameter and material match the material library, update pipe material according to external system data.
 - b. If the external system diameter and the material do not match existing materials in the material library, create a new material. The new material name is joined by diameter and material. Material inner diameter and roughness are left blank.
4. Import valve components. Mind the status (closed/open). Connect valves within 0.50 m tolerance to the nearest pipe.
5. Join junctions with two connections with the same materials with these exceptions:
 - a. Do not join pipes if the length exceeds a threshold value, such as 70 m.
 - b. Do not join junctions if a pipe is closed by a gate valve.
 - c. Detect long transmission mains with two connections. For long transmission lines, join junctions with a 500-meter threshold value.
6. Split extended pipe features if pipe length is more than 600 m.
7. Split pipes, if the terrain is fluctuating, for example, if elevation along pipe differs from start and end node elevations by 2 m, split pipe by elevation extremes.
8. Update new materials created in step 3 to the model material library.

As the material library may not match entirely with input data pipe materials and thus new duplicate materials could be created, the pseudo-code includes a correlation table for converting the most frequently non-matched material names (for example, 110Plastic → 110M). Another way would be to add all materials as custom materials (and not using the default material library). This method would require editing of all new materials to include the inner diameter and roughness values, but the created material library would contain the same look and feel as the NIS material library.

5.7 Pseudo-code for updating the network model

Pseudo-code replaces the model network with the new imported network (Figure 22). The process does not save any old model components outside stations; thus, it will remove any customizations added in the model components like tags and schematics referencing the old network.

1. Save the amount of leakage and pipe length in the model in each of the zones.
2. Remove existing pipe/node network without a station model tag. Demands with custom patterns (pattern name is different from zone name) are preserved. All control stations should have a zone named precisely as the connecting demand pattern.
3. Create a polygon that surrounds each of the stations.
4. Cut the new network with the polygons.
5. Import clipped network, new water consumers, and gate valves to the model (pseudo-code to generate a network model).
6. Connect demands to the nearest junctions.
7. Add zones to all the new networks based on connecting stations. Fetch demand pattern based on parent junction's zone name, excluding demands with custom patterns.
8. Simulate the model. Assign leaks to the model zones with the amount of leakage saved in the first step. Create a report for each zone: pipe length and water usage before and after the update.

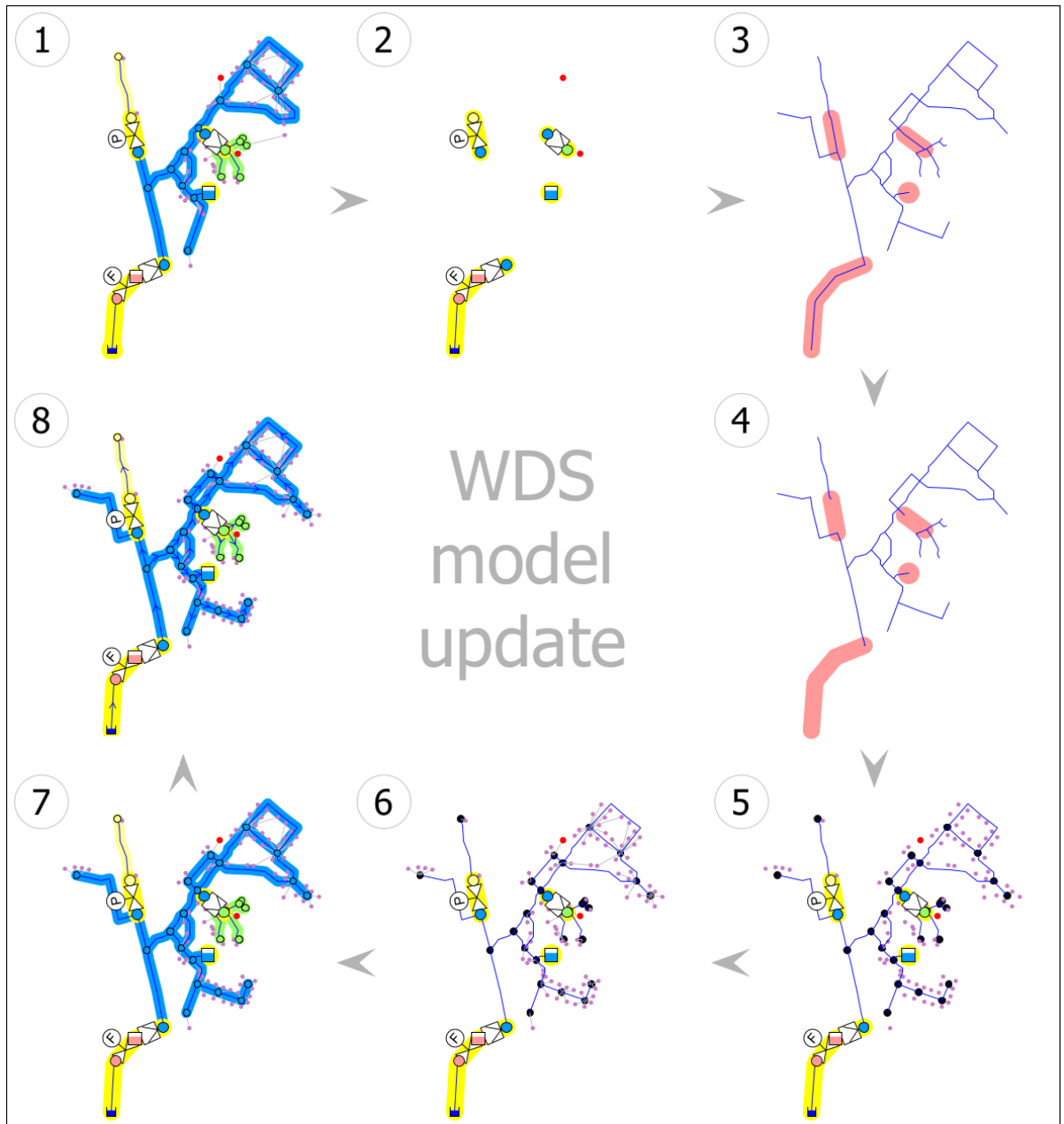


FIGURE 22. The process to update the network and demands of a WDS model

6 DATA PROCESSING PERFORMANCE

A comparison chart for two model-building approaches is presented in Table 1. When the networks are build using the automatic tools, the result will be reproducible, meaning that the created model will be the same if the input network does not change. With a manual approach, the result is different in every iteration. Compared to model manual build-up, the automatically created network model results in low data loss in location and missing attributes. With the automatic build, the number of added attributes can be increased without adding them individually. Attributes can give more information about the network, for example, original diameter and material, build year, condition, renovation history, pressure class, and criticality. Creating models by hand is error-prone, especially when multiple numerical attributes must be added to model components.

TABLE 1. Comparison chart for two model building approaches

Aspect	Model build-up using import and automatic processing of network data	Manual build-up
Reproducible results	Yes	No
Improvements to NIS data	Network gaps, pipes with missing or wrong attributes (diameter, material, category id) can be exported to NIS for data improvements.	The complete analysis cannot be done even with a high quality of work.
Loss of data: location	Exact location	Approximate location based on zoom level and modelers quality of work
Loss of data: Attributes	No restrictions on how many NIS attributes are added (it takes the same amount of time).	Adding more attributes increases the amount of manual work. Manual adding of attributes increases the risk of human-made errors.
Missing materials	Some of the data gaps can be filled automatically.	The modeler may not necessarily even notice that the pipe attribute is missing.
Wrong pipe category	Pipes documented as service pipes can be identified if they are located between distribution lines.	The modeler may not necessarily even notice that the pipe attribute is missing.
Network update	It can be automatized, requires specific documentation of model stations. Attributes can be later added to the model via pipe identifiers.	Manual work: going through pipes that are abandoned, new, and moved after the previous build. If pipe identifiers were not added, updating of pipe properties is more uncertain.

Quick response time for network quality analysis	Yes	No
Model results to NIS	Possible using pipe ids and as a background map.	Results can be exported as a background map.
Modeling speed	The process is not fully completed and tested.	Rough estimate: 44 km/h (WDS-network by adding one attribute)

In this thesis, network information input data was validated using automatic tools. Validations of network location (gaps, duplicate pipes), diameter, material, and category ids can be sent to the water utility for data improvements. If the model would be constructed manually, the reports about missing data may not be necessarily sent since the data gaps may not be visible, and that the reporting increases the amount of manual work.

When updating the water distribution model network, using a manual approach requires much work, as the modeler must go through all network updates done after the last model build: what pipes have been added, removed, and moved. With the automatic update, all network is replaced with a new one. Using model build-up processes allows utilizing new fields added to the network data, such as pipe maintenance and failure information to hydraulic models. With manually created networks, this would require adding all pipe identifiers individually to each pipe.

Network validations are a responsible way when creating models automatically. Input data should be validated before using the data blindly to analyze the network. Although the network would be once validated and corrected, the errors can still emerge in the future if the documentation of new networks is not valid.

Network validation processes provide fast ways to validate network data, allowing new services to be developed. For example, the process provides cues for improving the NIS data, and creating a model is optional. This kind of service could not be provided as manual work.

Validation of validating model speed should be further studied. In this thesis, manual modeling speed was measured using only one modeler's network build-up speed at 20 min period, with an average speed of 44 km/h. In Figure 23, there is a comparison between modeling speed with manual build-up (red) and networks that have been imported to the model and then processed (a total of 7 utility networks). The networks in the chart have been real modeling projects with the hours logged to the project. As projects with fully implemented automated model build-ups are not yet completed, the data analysis does not tell how fast the build-up could be. The chart does not include the rest of the modeling project (modeling of stations, demands, leaks and control settings, network analysis, and the report). The figure analysis shows that it is faster to create the network by hand when creating small networks, as it takes time to process the network data before it can be imported, for example, removing scrap pipes and household pipes from the network concatenating pipe materials. According to this data, the automatic build-up would be less efficient than the manual build-up until total network

length of 1800 km, if only network modelling speed is considered. However, the automatic approach provides other benefits, such as improvements to NIS data quality. Regards to the modelling speed, the automatic method can be beneficial for the smaller networks in the medium-term, as the models become outdated without maintenance, and new network updates are necessary.

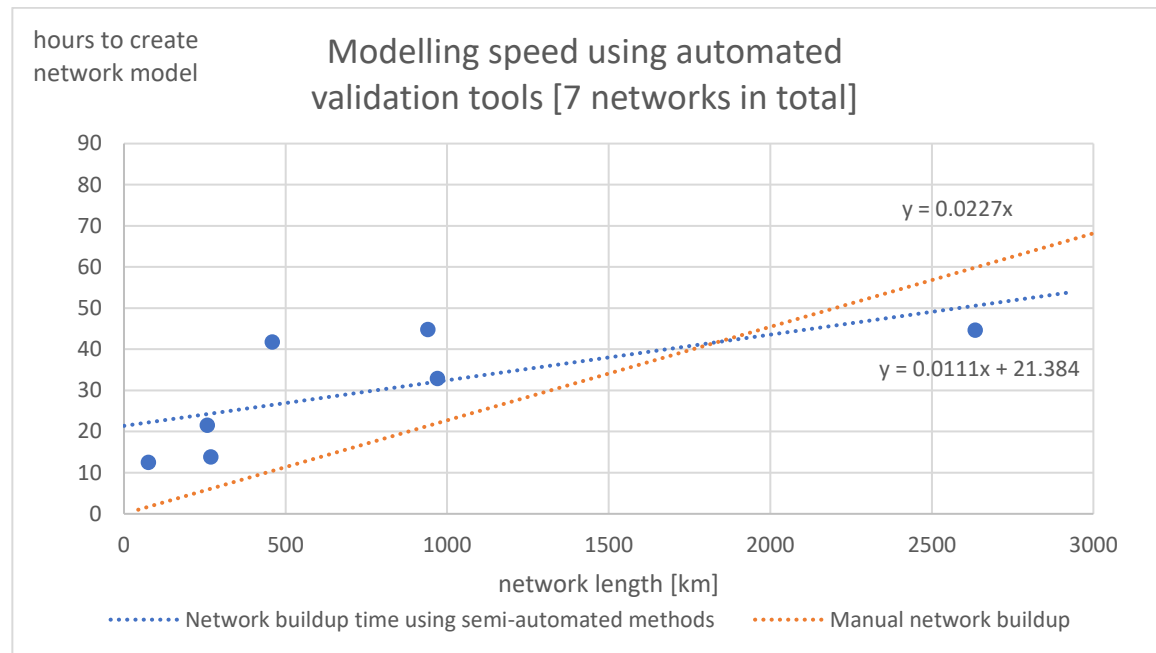


FIGURE 23. Comparison of network modeling speed using import methods and manual model build-up

The manual process could be further studied using multiple modelers with different skill levels. Also, tests could be conducted with sewer, storm, and district energy modeling software, calculating how much time to add additional attributes to the model components. In the test conducted in this thesis, only one attribute was added to the model pipes, but the process could be much slower when adding more attributes (like elevations). With this additional testing, one could evaluate the benefits of implementing a similar process, for example, to a sewer network.

The automatic processes should be further studied after it has been fully implemented and used. For example, the process for network topology checking could be tested using different tolerances and their effect on resulting reports. Addition to marking the gaps of the network data, the automatic fixing tools could be improved. Also, the gaps that could not be founded automatically, should be considered as ways to improve the process. New approaches, such as machine learning, could be utilized to network validation as well.

7 CONCLUSIONS & RECOMMENDATIONS

In this thesis, a total of six objectives were set. The model building process developed in Fluidit Ltd. was documented. This documentation will be helpful when training new water distribution system modelers in the field. With the documentation of the input data requirements and how the data is processed, the new modelers can anticipate what should be considered when building up a hydraulic model.

The network information was validated using the scripts developed in this thesis. The scripts include validating and fixing the network connectivity and the essential attributes, like pipe diameters, materials and pipe category identifiers, and estimating and filling in the missing properties. The scripts were written in the Python programming language, but only the language-independent pseudo-code was released in this thesis. The scripts were tested in practice, and they worked well with the real network data. Validation reports were successfully created and used for improving network data. In one test case with data from a water utility in Finland, the author got promising results, as the input data continuity showed significant improvement after just two iterations of the network validation. The continuity reports, imported into the NIS, made it easier to correct the network as the data flaws became more visible to the users maintaining the network information.

Automatic update of the network model using new network data was almost finished during the thesis. The process of replacing the network (Figure 22) cannot dynamically clip the imported network. However, replacing the network in the model has already been implemented with the models integrated with the NIS. With the NIS integrated models, the hydraulic model network is automatically updated and available when simulating the network. In the future, the model update process will be enhanced to apply design models as well, and the models could always use the latest network data if available.

The automatic hydraulic model build-up was compared to a manual approach regarding speed and quality (Table 1). The network data validation tools and the automatic hydraulic model build-up offer much value for creating models and improving the water utility's network asset management. Compared to the manual modeling, the automated tools can provide reproducible results with no loss of data. The model network can be updated at any time, and the model results can be more easily transferred to the NIS. In this thesis, a comparison of model build-up times for the semi-automated build-up and the manual build-up was performed. The manual modelling process had average network model building speed of 44 km/h, as the semi-automatic process had the average speed of 90 km/h after the data pre-processing had been completed. The modeling speed of a fully automated process could not be yet measured, as the process is not fully finished.

While documenting, developing, and testing the process in this thesis, the author learned programming and new ways of thinking about problems from a broader perspective. The problems highlighted in the thesis are universal, and the problems should be handled with a standardized approach. Now, as the roadmap to automated model build-up is documented, it will be easier to discuss and tackle the problems, providing easily accessible solutions to all the modelers out there. It has been a good

learning process to build the manually at first, and to gradually improve the process by utilizing programming to develop scripts for automatizing the process. Now, these validators will be used for cleaning the data of many different systems.

The input data validators and the model builder will be improved further after this thesis. For example, it would be beneficial to add templates for each network information system provider, helping to create the model from different input data systems. The data validator functions could be added as Fluidit software plugins, making the network errors more visible for many modelers to improve hydraulic models and for water utilities to improve their network assets.

Appendix 1 lists the input data problems that limit fully utilizing the automatic hydraulic model build-up.

REFERENCES

- Nathanson, Jerry A. "Water supply system", 2021. *Encyclopedia Britannica*. <https://www.britannica.com/technology/water-supply-system>. Accessed 18 February 2021.
- EPA, 2000. EPANET 2 USERS MANUAL. Lewis A. Rossman, Water Supply and Water Resources Division, National Risk Management Research Laboratory, Cincinnati, OH 45268. U. S. Environmental Protection Agency. Available from: <https://www.epa.gov/water-research/epanet>
- EPA, 2005. Water Distribution System Analysis: Field Studies, Modeling and Management - A Reference Guide for Utilities. U. S. Environmental Protection Agency. Available from: https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NRMRL&dirEntryId=150585
- EPA, 2021. EPANET - Application for Modeling Drinking Water Distribution Systems. Referenced in 11.3.2021. Available from: <https://www.epa.gov/water-research/epanet>
- Finlex, 2014. Laki vesihuoltolain muuttamisesta. 681/2014 15 §. Referenced in 17.4.2021. Available from: <https://finlex.fi/fi/laki/alkup/2014/20140681>
- Fluidit, 2020. Fluidit Water product page. Referenced in 3.11.2020. Available from: <https://www.fluidit.com/products/fluidit-water/>
- Fluidit, 2021. Fluidit training material. Referred in 3.12.2019.
- Fluidit, 2021. Fluidit home page. R&D. Available from: <https://www.fluidit.com/rd/>. Referred in 16.4.2021.
- IWA, 2020. 5 major trends impacting the Water industry in the next decade. Online-post. Referred in 20.5.2021. Available from: <https://iwa-network.org/five-major-challenges-and-emerging-trends-impacting-the-water-industry-in-the-next-decade/>
- Ministry of Economic Affairs and Employment, 2020. Water bills to be based on actual consumption in each apartment – Remotely read water meters to be installed in old housing companies during water and sewage system renovations. Press release. Referred in 18.4.2021. Available from: https://tem.fi/-/huoneistokohtainen-vesilaskutus-todellisen-kulutuksen-mukaan-vanhoihin-taloyhtiöihin-etaluettavat-vesimittarit-putkiremontin-yhteydessä?languageId=en_US
- Muoviteollisuus RY, 2012. Paineputkijärjestelmät polyeteenistä. Available from: https://www.plastics.fi/fin/organisaatio/putkijaosto/putkijaoston_julkaisut/
- Sunela, 2017. Real-time control optimization of water distribution system with storage. Available from: https://tutcris.tut.fi/portal/files/16756236/Real_Time_Control_Optimization_of_Water_Distribution_System_with_Storage.pdf
- Tampere Water, 2021. Vellamo - Water quality in the water supply system in Tampere. Available from: <https://vellamo.tampere.fi/en>. Referred in 16.4.2021.

Thomas M. Walski, Donald V. Chase, Dragan A. Savic, Walter Grayman, Stephen Beckwith, and Edmundo Koelle, 2003. *Advanced Water Distribution Modeling and Management*. Haestad Press, 1st edition. ISBN 0-9714141-2-2

Väyrynen, Janne 2015. Pressure booster station. Image. Location: Image library of Janne Väyrynen.

APPENDIX 1: COMMON ISSUES WHEN ANALYZING AND GATHERING WDS MODEL INPUT DATA

Water utility's documentation issues:

- Network:
 - Network gaps and duplicate pipes
 - Inconsistent pipe materials show duplicate attributes for the same material and non-existent pipe materials, like 100Plastic → 110Plastic.
 - Missing pipe SDR classes, inner diameter is not available from the network data. The utility may have preferences for certain SDR and PN classes', but they are not added to the NIS.
 - The isolation valve does not have status documented, even when the valve is closed.
 - The isolation valve's connection to the pipe does not match the valve's location.
 - Station and control device names are not documented in GIS format.
 - Water consumers lack coordinates and addresses.
 - Billed water includes other data, like sewer flow measurements, or the billed water contains double measurements (individual customers and DMA flow measurement points), resulting in too much billed water consumption.
- Tanks: Old tank drawings do not include elevations in the current elevation system.
- Pumps: pump models are unknown and not documented.

Software issues:

- **NIS:** The user cannot export network data with attributes included.
- **SCADA:** The user cannot export flow/pressure history data.
- **SCADA:** User cannot export history data from a year or more period at hourly or more frequent intervals.
- **SCADA:** Control device settings are not exportable in tabular format. The user must manually go through all stations and their settings to export control settings, taking notes or screenshots.
- **Pump manufacturer:** Pump curves are unavailable from the online catalogs.
- **Pump manufacturer:** Pump curve points are not exportable in tabular format (or API). Adding API for online pump catalogs would help create models and analyze and optimize pumping.

DEM: Accurate DEM is not available at the target location.