A RESEARCH PROJECT SUPPORTED BY THE EUROPEAN COMMISSION UNDER THE FIFTH FRAMEWORK PROGRAMMME
AND CONTRIBUTING TO THE IMPLEMENTATION OF THE KEY ACTION "SUSTAINABLE MANAGEMENT AND QUALITY OF WATER"
WITHIN THE ENERGY, ENVIRONMENT AND SUSTAINABLE DEVELOPMENT

EVK1-CT-2000-00053

REPORT

No. 2.1 - August 2002

WP2
Description and Validation of Technical Tools:

D3 – Models Description

Patrick Eisenbeis[1], Yves Le Gat[1], Katia Laffréchine[2], Pascal Le Gauffre[2], Axel König[3], Jon Røstum[4], Ladislav Tuhovcak[5], Pavel Valkovic[5]

[1]Cemagref Bordeaux, [2]INSA Lyon, [3]SINTEF, [4]NTNU, [5]Brno University of Technology

COMPUTER AIDED REHABILITATION OF WATER NETWORKS
RESEARCH AND TECHNOLOGICAL DEVELOPMENT PROJECT OF EUROPEAN COMMUNITY

COMPUTER AIDED REHABILITATION OF WATER NETWORKS
RESEARCH AND TECHNOLOGICAL DEVELOPMENT PROJECT OF EUROPEAN COMMUNITY
Hydraulics and Civil Engineering Unit, Cemagref Bordeaux, 50, avenue de Verdun, F-33612 CESTAS Cedex
Phone: +33-557.89.08.33 or 29 Fax: +33-557.89.08.01 Email: carew@bordeaux.cemagref.fr

CARE – W

Computer Aided REhabilitation of Water networks.
Decision Support Tools for Sustainable Water Network Management

WP2: Description and validation of Technical tools

WP2.1: Report on models description[1]

Deliverable D3

Authors : Patrick Eisenbeis, Yves Le Gat (Cemagref), Katia Laffréchine, Pascal Le Gauffre (INSA Lyon), Axel König (SINTEF), Jon Røstum (NTNU), Ladislav Tuhovcak, Pavel Valkovic (Brno University)

August 2002

---

[1] *Reference*:
EISENBEIS P., LE GAT Y., LAFFRÉCHINE K., LE GAUFFRE P., KÖNIG A., ROSTUM J., TUHOVCAK L., VALKOVIC P. (2002) *CARE-W: WP2 – Description and validation of Technical tools D3 – Report on models description.* CARE-W (Computer Aided Rehabilitation of Water networks), EU project under the 5th Framework Program, contract n°EVK1-CT-2000-00053. Cestas (F): Cemagref, August 2002, 116 p.

Table of contents

1 INTRODUCTION

CARE-W project aims to develop methods and software that will enable engineers of the water undertakings to establish and maintain an effective management of their water supply networks, rehabilitating the right pipelines at the right time. The results shall be disseminated as a manual on Best Management Practice (BMP) for water network rehabilitation.

This project is organised in the following Working Packages (WP):

- WP1: Construction of a control panel of performance indicators for rehabilitation;
- WP2: Description and validation of technical tools;
- WP3: Elaboration of a decision support system for annual rehabilitation programmes;
- WP4: Elaboration of long-term strategic planning and investment;
- WP5: Elaboration of CARE-W prototype;
- WP6: Testing and validation of CARE-W prototype;
- WP7: Dissemination;
- WP8: Project management.

Cemagref is responsible for WP2, which is divided in three Tasks. This report refers to this Task 2.1.

The objective of this task is to describe some existing scientific models, that can be technical tools useful and helpful in the framework of a rehabilitation policy. These tools are linked with some technical Performance Indicators proposed in the Work Package n°1, coordinated by LNEC, and also with technical criteria, that are, up to now, criteria used to choose the pipe to replace.

2 METHODOLOGY OF DESCRIPTION

2.1 Technical criteria to assess pipe degradation

Several technical indicators could be included in the study. They are all indicators that can give an idea of the state of a pipe or a group of pipe, or its functionality. That is to say:

- water quality,

  In rehabilitation policy, this criterion is used, regarding water colour ("red water") or regarding chemical rate about substances linked with internal wall pipe. These indicators can be assessed in current functioning of the network or after a flushing. Their occurrence define an area that can be suspected, but difficultly a particular pipe. It is however a reactive criteria, that is used only if the phenomenon occurs.

  Studies on methods to assess pipe degradation, in relation to the water quality are in progress, but not yet available. That's why this criterion is not used in this Work-Package, but it will be necessary to define it in the WP3 (annual programming rehabilitation), even in WP4 (Strategic …) if it can be predicted.

- water losses,

  Water losses are a Performance Indicator linked with the maintenance of drinking water networks. In most of the water utilities, values of water losses are used as objectives to reach. The Table 1, proposed by IWA, presents water balance, including revenue water and non revenue water (NRW) and gives a definition of different terms. NRW are including water losses.

Table 1 : components of water balance

<table>
<tr><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th></tr>
<tr><td rowspan="10">System input volume
[m^3/year]</td><td rowspan="4">Authorised consumption
[m^3/year]</td><td rowspan="2">Billed authorised consumption
[m^3/year]</td><td>Billed metered consumption (including water exported)
[m^3/year]</td><td rowspan="2">Revenue water
[m^3/year]</td></tr>
<tr><td>Billed unmetered consumption
[m^3/year]</td></tr>
<tr><td rowspan="2">Unbilled authorised consumption
[m^3/year]</td><td>Unbilled metered consumption
[m^3/year]</td><td rowspan="8">Non-revenue water
[m^3/year]</td></tr>
<tr><td>Unbilled unmetered consumption
[m^3/year]</td></tr>
<tr><td rowspan="6">Water losses
[m^3/year]</td><td rowspan="2">Apparent losses
[m^3/year]</td><td>Unauthorised consumption
[m^3/year]</td></tr>
<tr><td>Metering inaccuracies
[m^3/year]</td></tr>
<tr><td rowspan="4">Real losses
[m^3/year]</td><td>Real losses on raw water mains and at the treatment works (if applicable)
[m^3/year]</td></tr>
<tr><td>Leakage on transmission and/or distribution mains
[m^3/year]</td></tr>
<tr><td>Leakage and overflows at transmission and/or distribution storage tanks
[m^3/year]</td></tr>
<tr><td>Leakage on service connections up to the metering point.
[m^3/year]</td></tr>
</table>

These are assessed for the network or a group of pipes, but rarely at the pipe level. The assessment is still uncertain, in a lot of utilities. Indeed it depends:

* on the pressure existing on the network,

* on the accuracy of water meters,

* on the assessment of non-metered consumption water (used, for instance, for street cleaning or fire),

* on the method to assess leakages.

To conclude water losses can not be used directly to choose a pipe to replace, because they are not assessed to the pipe level. Studies, associating leak rates with particular pipe and its parameters (age, material, etc…),could maybe help to a better comprehension of the phenomenon, useful for annual programming as for strategic planning of rehabilitation. No results are now available, but this criteria will have to be taken into account in the concerned Work Packages.

- failures,

A failure is a leak, break or burst involving a repair on the pipe. This criterion is important in the decision of rehabilitation, regarding several aspects.

First, as it involves a repair, it causes a functioning costs. These costs can be direct or indirect, i.e. causing obstruction in the road, problem of traffic, risk of flood or dissatisfaction of water consumption.

Secondly a failure can cause water losses. Indeed the total water losses caused by failures could represent 25% of water losses caused by the pipes.

Thirdly one (or more) failure is representative of a defective state of the pipe. It can announce future failures.

Moreover some tools have been developed since 1990 in Europe by different Research centres or Universities. Some of them have already been applied in water utilities, but they need validation on a larger sample. They are based on maintenance historic records or physical data. They will be presented in this report.

- flow capacity of the pipe,

The decreasing of flow capacity can be representative of:

- an increasing of pipe roughness (or a decreasing of pipe diameter),
- an increasing of water consumption.

In the first case, a specific examination of the pipe makes it possible to note the problem of capacity. But it can only be made after a suspicion of defective capacity. This last one can be possible with an hydraulic model.

An hydraulic model, combined with failures probability, could also give hydraulic reliability for a pipe or a group of pipes, in term of impact on downstream nodes in case of failure

Models whose objectives are such have been developed by different European Research Centres. They will be studied in this Work Package.

2.2 Methodology

As presented in the previous part, two major kind of tools will be studied:

- tools (5 totally), whose objectives are to forecast failures on a pipe or a group of pipes, by statistical or physical mean,
- tools (3 totally), whose objectives are to assess hydraulic reliability of a pipe or a group of pipes.

Totally 8 models are studied : 5 forecast failures models and 3 hydraulic reliability models (Cf. Table 2).

Table 2: Models studied in the Work Package

	Forecast failure models					Hydraulic reliability models		
Original model acronym	Failnet-Stat	Asset-map1	Asset-map2	Utilnets	Winroc	Aquarel	Failnet-Reliab	Relnet
CARE-W acronym	PHM model	Markov model	Poisson analysis		NHPP model	Aquarel	Failnet-Reliab	Relnet
Research Centres / Universities	Cemagref	INSA-Lyon	INSA-Lyon	SINTEF	NTNU	SINTEF	Cemagref	Brno University

A model form has been filled by all the partners, responsible of one or more models. These form are presented in the appendices.

The form has been established in 6 parts:

1) A **general description**, useful to describe the models to the end-user (water utilities manager, municipalities),
2) A **theoretical framework overview**, to describe the scientific theory used in the models,
3) **Specifications of computational steps**, that describe operational functioning of the model (including input data, scheme and output),
4) **Possible improvements** of the model, that highlight default or not yet validated functions, which could be tested firstly,
5) **Software specifications,** to describe data format and software programs,
6) **References**.

3 DESCRIPTION

Up to the end of the report, the used names of the models are the "CARE-W" acronyms presented in the Table 2.

All the works concerning the elaboration of the models have begun between 1990 and 1995 and they have been going on up to now, with improvements and applications in different water utilities.

3.1 General description

3.1.1 Forecast failure models

The Table 3 gives a rapid description of the models. Regarding outputs and level approaches NHPP, PHM and Markov models are very close.

Markov and Poisson models are rather adapted to water utilities with "poor level" data, that means for Markov models incomplete data and for Poisson model short maintenance data. They could also be tested on "rich level" data utilities, to be validated.

On the other hand, NHPP and PHM models seem to be more efficient with long maintenance record and complete data, even if they have already been tested on networks with short maintenance records (5 to 10 years) and with good results.

Utilnets is different than the other ones because of the type of model, not based on historical maintenance but on the physical resistance of the pipes.

Figure 1 presents a comparison of the models, according to output and the type.

Table 3 : General description of failures forecasting models

	Markov	Poisson	PHM model	Utilnets	NHPP model
Output:	1) **Failure rate** $FR(C_i, t)$ for each section category C_i : 2) **Failure rates or predicted number of failures** $PNoF(C_i, Y_k)$ for year Y_k, according to the chosen MCG model, data on the asset stock.	1) **Rate Ratios** characterising influence of failures factors 2) PFR(Cj) **Predicted Failure Rate** for category of pipe	1) **Influence of failures** factors (Weibull parameters or ratio) 2) **Forecasted number of failures**/ pipe or group of pipes or failure probability in a time fixed horizon.	1) **Expected life-time** according to failure probability threshold 2) **Order of rehabilitation** (and cost of rehabilitation)	1) **Regression coefficients** 2) **Failure intensity** 3) **Expected number of failures**
Approach level:	group	group	pipe	pipe	pipe
Output level:	group	group	pipe, group or network	pipe	pipe, group or network
Model type	Probabilistic model	Statistical model	Statistical model	Physical model	Statistical model
Past application sites	Lyon	Lyon	Bordeaux, Lausanne, Charente-Maritime, Irrigation networks	North West Water, Trondheim	Trondheim

Table 4: Hydraulic reliability models

	Aquarel	Failnet-Reliab	Relnet
Output:	Water supply availability Frequency of degraded pressure Link importance (3 indices)	Hydraulic reliability indices at different level (node, pipe, group or network)	Expected life-time Reliability indices (node, pipe, group or network)
Approach level:	pipe	pipe	
Output level:	pipe, group, network	pipe, group, network	pipe, group, network
Model type	Mathematical and statistical model	Mathematical model	
Application site	Trondheim	Charente-Maritime	Brno

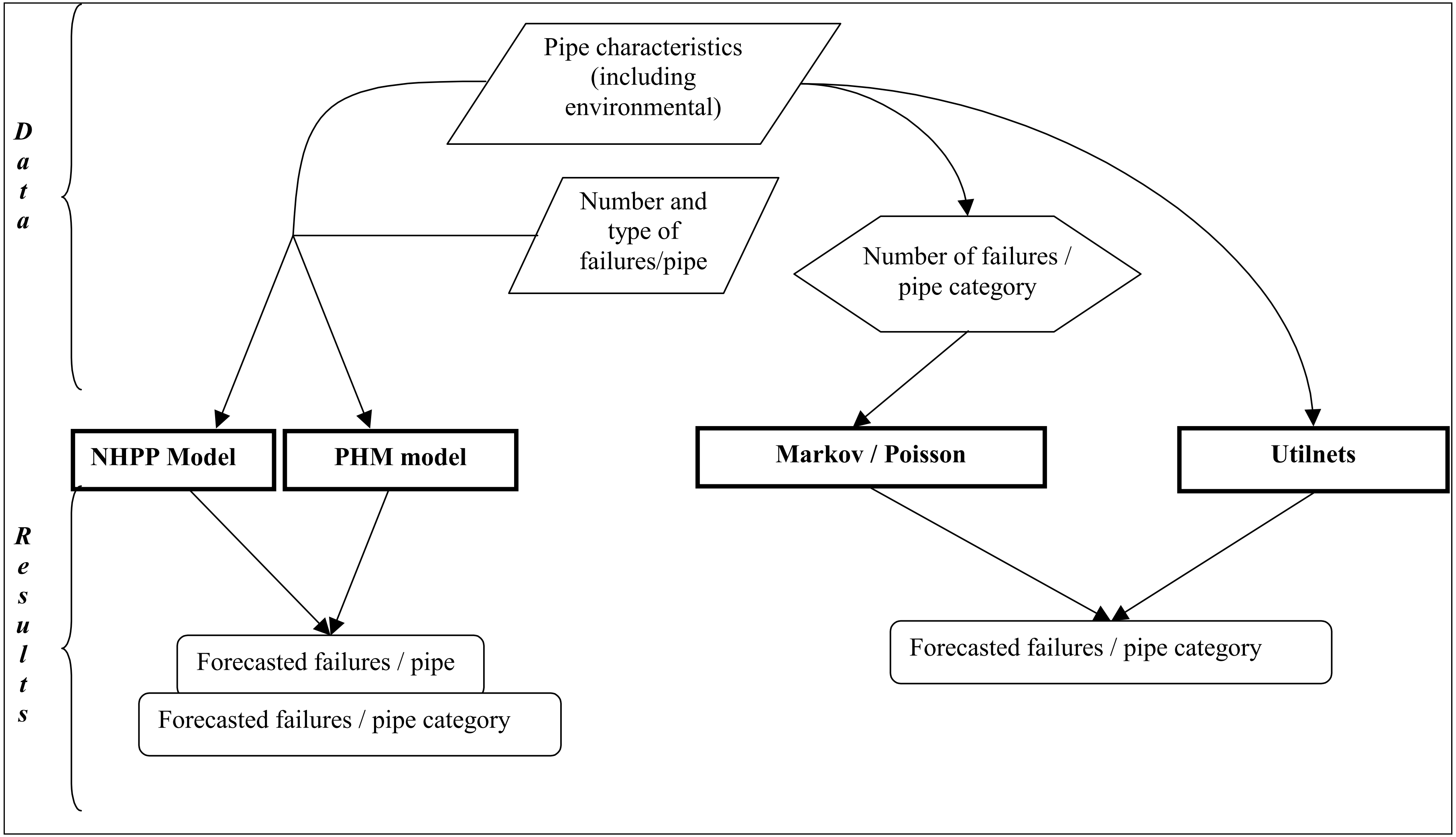


Figure 1: Differentiation of forecasting failures models according to output and type of model

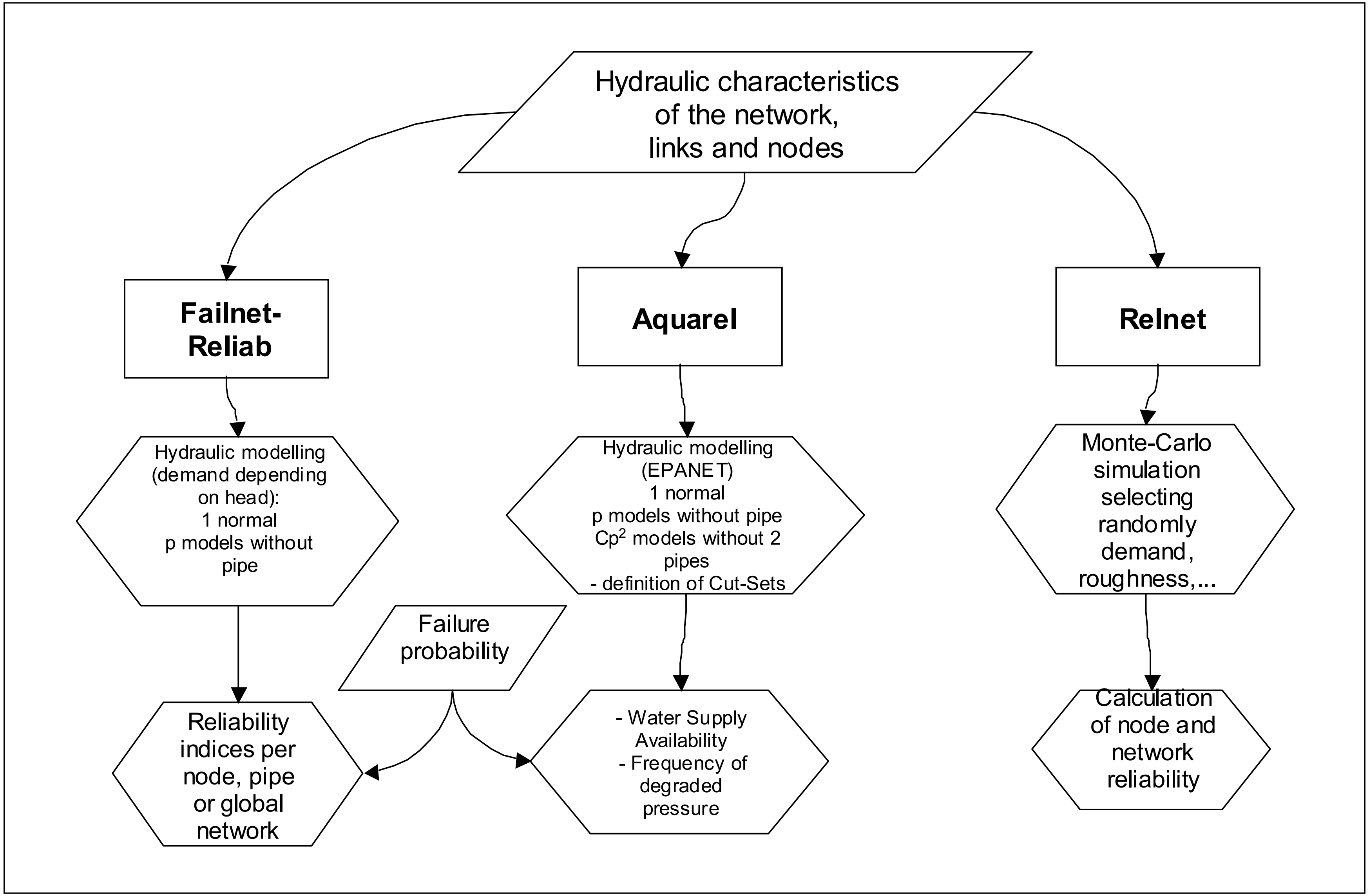


Figure 2 : Functional schemes of hydraulic reliability models

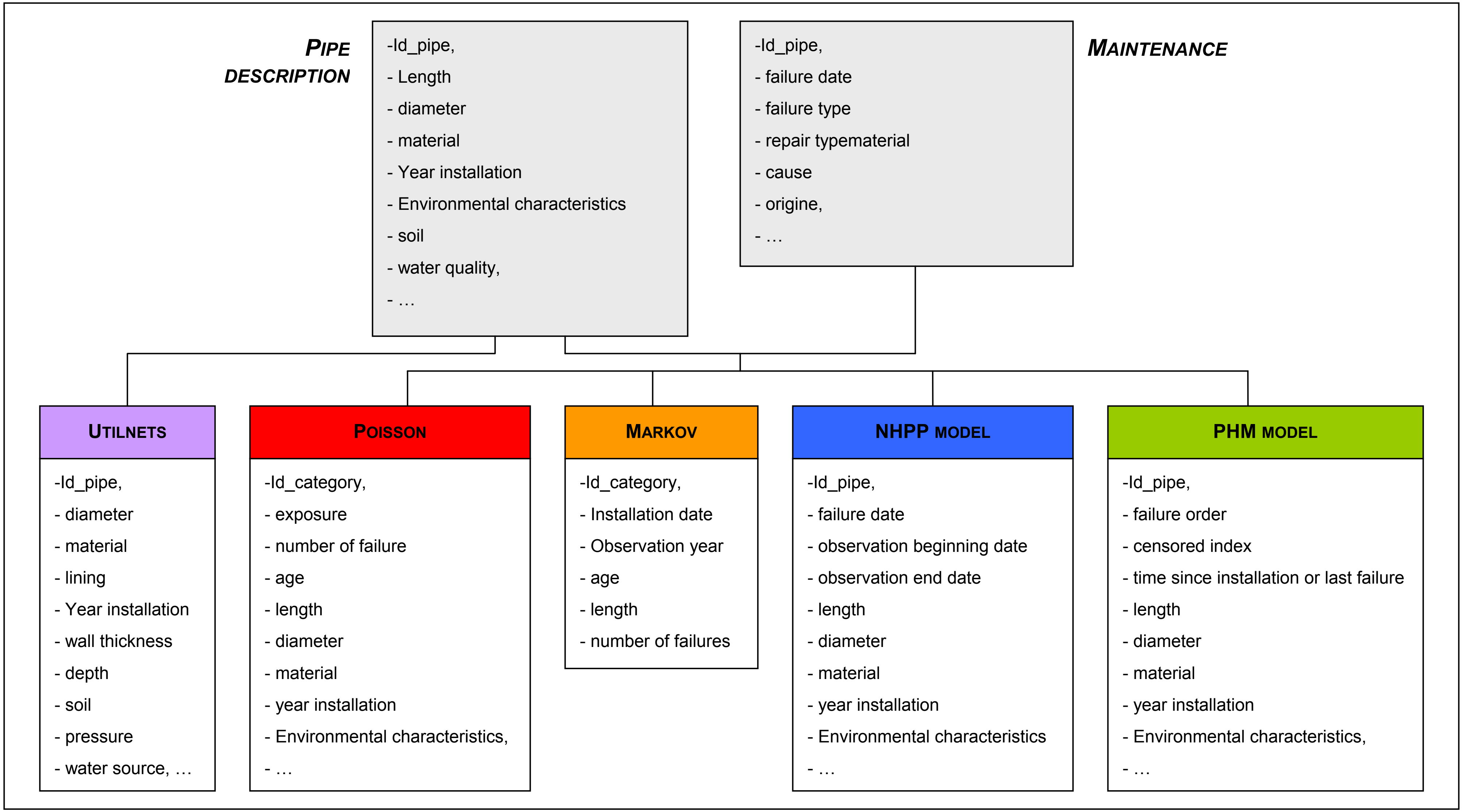


Figure 3 : Data description for failure forecasted models

3.1.2 Hydraulic reliability models

These models are presented in the Table 4. The three models are based on hydraulic model. Aquarel and Relnet use the model EPANET (more accurately Relnet uses a derived program of EPANET called ODULA). Failnet-Reliab uses a algorithm close to the algorithm used in Porteau software, elaborated by Cemagref.

Each model is established in two steps but the outputs and the progress are different.

Aquarel aims to define, for each closed link or for each closed links pair, a matrix of nodes and pressure class. It takes also into account the tank volume compared to the MTTF (Mean Time To Failure) and TTR (Time To Repair). These results allow the calculation of different values, such availability calculation or frequency of degraded pressure calculation.

Failnet-Reliab aims to define reliability indices. They are computed from the quotient between available consumption and the assumed demand at each node. The available consumption is computed with a function depending the head at each node, different of classical hydraulic model. The indices give an assessment of the influence of pipe break on the nodes, a zone or the whole network. Nodes vulnerability is also computed and a global index.

Relnet is based on stochastic principle using Monte-Carlo method.

Useful data are almost the same between the models. Failures probabilities can also be used to compute the different output. Failnet-Reliab and Aquarel could be linked, indifferently, with NHPP or PHM models.

Figure 2 presents a functional scheme of hydraulic reliability data.

3.2 The data

3.2.1 Failure forecast models

The Figure 3 presents useful data for these models.

Two types of data are used:

- Pipe description data:

  This concerns the description of the pipes existing in the network. The data are pipe characteristics (Identification n°, street, diameter, material, length, installation year) or environmental characteristics (soil type, water quality, pressure, traffic, location, average number of service connection…). All data, that could influence a priori failures occurrence, can be included.

- Maintenance data:

  This concerns the failures occurred on the networks and their characteristics (date, type of failure, Pipe concerned, …).

Then these two types are joint to be used, differently among the models.

For PHM model:

The two types of data are crossed to define only one file. One item correspond to a couple (pipe, failure) including pipe description, failure description, occurrence of the failure, censored value, time since the last failure or installation date.

For NHPP model:

Data are crossed to have a file, where items are failures or pipes without failure. In the first case failure date is given, as well as observation beginning and observation end.

For Markov models:

Data are transformed per pipe category. Per category each item corresponds to an age of pipe and a number of failures to be compared for the calibration.

For Poisson model:

Each category is defined according to the different covariates. To each category a number of failures is given.

For Utilnets:

Data used are only pipe description data. They are a little more accurate than for the other models. Indeed physical data, like wall thickness or depth, are useful.

3.2.2 Hydraulic reliability models

These models have two kinds of data (Cf. Table 5):

- data for hydraulic models, common to the three,
- data for reliability computing, different according to the models.

Table 5 : data useful for hydraulic reliability models

<table>
<tr><th></th><th>Aquarel</th><th>Failnet-Reliab</th><th>Rel-net</th></tr>
<tr><td>Hydraulic model</td><td colspan="3">- for the consumption nodes:
• the demand,
• the elevation,
• required pressure,
- for tank nodes:
• water level,
• tank volume,
- for links:
• the length,
• the material,
• the roughness.</td></tr>
<tr><td>Reliability</td><td>- failures rates on pump or pipe,
- Repair rates (MTTR, Mean Time To Repair or TTR_{99} that is the 99% percentile of the time to repair.</td><td>- Failure probability
- Weight (importance) of the node</td><td></td></tr>
</table>

3.3 Theoretical framework

3.3.1 Failure forecast models

Markov and Poisson models use data at a large level approach, that means that they are used for each group.

Markov process propose to transit from a condition state to another one, each state corresponding to a number of failures or a failure rate. By pipe category, each state and each transition are linked with probability functions or density functions. These functions are either exponential functions and/or Weibull functions. They are selected using least squares method. Then after the choice functions by pipe category it I possible to compute, in the next year(s), failure rate by pipe category.

Poisson model proposes to use directly Poisson regression for each pipe category. The parameters of these functions are assessed maximising log-likelihood function. Rate-ratio between two values of a parameters are given and failure rate is computed for each category.

PHM model uses survival data analysis. The time between laid date or previous failure and next failure or end of observation is analysed. Weibull function is used, with parameters assessed maximising a likelihood function. After determining the Weibull functions (according to the number of previous failures), a number of failure is computed using Monte-Carlo simulations.

Comparing to PHM model, NHPP model analyses failure process using a Non-Homogeneous Poisson Process (NHPP). That means that, on a determined time interval, the number of failures is taken into account. An intensity function is computed, that depends on a parameters linked to covariates. Thus a number of forecasted failures can be directly computed from this function, for each pipe or each pipe category.

Utilnets is designed for the failure prediction of cast iron pipes. The process is driven by corrosion, with an equation assessing max corrosion pit, according to time in service and coefficients. These coefficients are given according to soil types and water quality.

3.3.2 Hydraulic reliability models

Failnet-Reliab uses a hydraulic model, based on classical hydraulic laws: flow rate conservation at each node, energy conservation for each pipe, head-loss formula (Hazen-Williams). An equation defining consumption according to the demand and the head is given.

The different equations are solved using a minimisation method and Newton-Raphson algorithm. A special algorithm has been defined to take into account the elevated nodes that have downstream nodes. Then reliability indices can be computed, with failure probability.

Aquarel defines Cut-sets, that are matrix (nodes, pressure classes) per closed pipe or closed pipe pair. Hydraulic model EPANET is used. This model uses the same equations than failnet-reliab but with a fixed demand. Hydraulic models (as much as number of pipes) are computed with a closed pipe. 4 pressure class are defined. To each node a pressure class is assigned after modelling. Then same cut-sets are established with for each model, 2 pipes closed. When the cut-sets are established, water supply availability and frequency of degraded pressure can be calculated.

With Relnet, random selection are established. Demand, roughness, network topology are randomly selected, defining for each case a load-state. For each load state a hydraulic analysis is established. These Monte Carlo simulations can result to the computing of reliability index for each node and pipe.

4 POSSIBLE IMPROVEMENTS / PROPOSITION OF VALIDATION

Possible improvements of the models concern:

- utilisation easiness,
- theoretical framework,
- testing and validation,

4.1 Utilisation easiness

Some of this model are still programmed with specific software (PHM Model, Poisson, Markov) like Access, SAS, Stata or Matlab. For a best and easier use, it could be interesting to create a specific program, from the beginning to the end.

It is also the case for Hydraulic reliability model that use hydraulic modelling software like EPANET, MIKENET or PORTEAU. Indeed these models progress in two steps, one more hydraulic and a second more statistical.

4.2 Theoretical framework

Some models could be improved and more detailed on the theoretical level. These improvements are linked to the assumptions or the method that could be done in the different models. They concern, for instance :

- values to specify in Utilnets, about physical data,
- weights given in Failnet-reliab to determine node importance,
- improving assessment of parameters in NHPP model (problem of significance matrix),
- possibility to calibrate a set of models with Poisson Regression.

Of course these improvements will be made thanks to the tests and validation.

4.3 Tests and validation

Most of the models have been tested on a limited number (1 to around 5) and will need more tests to be validated and improved. This is one of the objectives of CARE-W. These tests will be useful:

- for theoretical frame work (Cf. previous paragraph),
- to compare the models with the same objectives especially regarding the need of data,
- to define confidence intervals of results and parameters,
- to compare forecasted failures to real failures,
- to study the influence of data missing, uncompleted or uncertain data on the results,
- to improve the interface and the utilisation.

The comparison of forecasted and real failure will need to define an index to measure efficiency of the models. An index is proposed in PHM model description.

4.4 Utilisation of the models in the rehabilitation policies (annual programming or strategy planning)

The models give directly outputs that are not always directly useable for rehabilitation. A way to transform the output in global or specific indicators will be necessary, as well as a method.

APPENDIX 1 : MARKOV MODELS

1 GENERAL DESCRIPTION

1.1 Name and/or Acronym of the Model

Markov models

1.2 Company/Research Center/University

INSA Lyon

1.3 Objectives

The objective is to define ageing models for pipe categories in using incomplete failure records.

Markov processes, i.e. continuous in time Multi-state models are used to represent the ageing and failure rates of homogeneous pipe categories.

Applied to a set of categories representing the asset stock, these ageing functions make it possible to forecast the evolution of the number of failures per km and per year.

This enables to test different renewal hypotheses on these categories.

1.4 Functional description

Several MCG (multi condition grades) models are proposed.

These models correspond to continuous in time multi-state Markov Processes: ageing is represented by successive transitions from one state to another and each state corresponds to a particular failure rate.

→ transition from state i to state $(i+1)$ is represented by a rate (or hazard function): $h_i(.)$

→ at a given time t, the probability that an individual is in state i is $P_i(t)$

→ failure rate for state i is $FR_i(.)$

The overall behaviour (failure rate) is given by: $\boxed{FR(t) = \sum_{i=1}^{n} P_i(t) FR_i(.)}$

These models can be calibrated in using observed failure rates based on incomplete failure records.

1.5 Brief Historical Overview of the Model

As regards the appearance of breaks on water mains, one of the most significant results of previous researches is the distinction between different stages of deterioration during the operating life of a water main.

Using a Cox Proportional Hazard Model Andreou (1987) and Eisenbeis (1994) have shown that the number of previous failures is a very important explanatory variable to model failure rates.

(Andreou *et al*., 1987; Karaa and Marks, 1990; Eisenbeis, 1994 and Lochbaum, 1994 in the case of gas mains) show clearly that:

- During the early stage of deterioration (0 to first break for Eisenbeis or 0 to second break for Andreou), the failure risk increases with age (Weibull model).
- Once a pipe was experiencing breaks, the time to the following failure started to decrease.
- During the late stages of deterioration ("fast-breaking stage"), the break-rate shows no particular trend towards increase or decrease with time. Failure thus becomes independent of time with a constant return period (Exponential model).

These results have been obtained with epidemiological approaches (survival analysis) based on rich historical records. The times to failure used in the analysis have to be observed or right-censored failure time (pipes still in place at the end of the observation period). In other words, for each pipe segment, we are supposed to know the time to the first failure, time to the second failure, etc.

For left censored data (i.e. for pipe segments with a laying date preceding the beginning of the observation) failures may have occurred before the beginning of the observation. In this case Le Gat & Eisenbeis (2000) proposed to consider the first observed failure as the first failure. Time to the first failure is thus calculated from the date of the beginning of the observation.

Another approach has been proposed by (Malandain, 1999) (Le Gauffre *et al*.,2000) to deal with incomplete data (left-censored data).

Stochastic multi-state models are used to represent the ageing of an homogeneous pipe category: the ageing of a group of pipe segments is represented by the evolution of the distribution of pipe segments in several condition grades.

At age *t* one part $P_1(t)$ of the cohort is supposed to be in condition grade 1, one part $P_2(t)$ is supposed to be in condition grade 2, etc.. Each condition grade is represented by a specific failure rate function $FR_i(.)$, and the overall failure rate $FR(t)$ is the result of these different behaviours. For *n* possible states we have

$$\boxed{FR(t) = \sum_{i=1}^{n} P_i(t) FR_i(.)}$$

Le Gauffre P., Malandain J., Miramond M.(2000) Modélisation du vieillissement et maintenance des réseaux d'eau potable. Revue Française de Génie Civil, Volume 4 – n°2-3/2000, p. 397-410.

Malandain J., (1999) Modélisation de l'état de santé des réseaux de distribution d'eau pour l'organisation de la maintenance. Étude du patrimoine de l'agglomération de Lyon. Thèse de Doctorat n° 99 ISAL 0040 de l'INSA de Lyon, URGC / Hydrologie Urbaine, 206 p.

Andreou S.A., Marks D.H. and Clark R.M. (1987), 'A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory', Journal of Advances in Water Resources, vol. 10, March 1987, pp. 2-10.

Andreou S.A., Marks D.H. and Clark R.M. (1987), 'A new methodology for modelling break failure patterns in deteriorating water distribution systems: Applications', Journal of Advances in Water Resources, vol. 10, March 1987, pp. 11-20.

Eisenbeis P. (1994), 'Modélisation statistique de la prévision des défaillances sur les conduites d'eau potable', Ph.D. thesis, Université Louis Pasteur Strasbourg, 1994, 190 p.

Le Gat Y., Eisenbeis P. (2000) Using maintenance data to forecast failures in water networks. Urban Water. Vol. 2 173-181

KARAA F.A. and Marks D.H. (1990), 'Performance of water distribution networks : integrated approach', Journal of performance of constructed facilities, vol. 4 n°1, February 1990, pp.51-67.

LOCHBAUM B.S. (1994), 'Cast-iron main break predictive models guide maintenance plans', Journal Pipeline Industry, April 1994, pp.29-32.

2 THEORETICAL FRAMEWORK OVERVIEW/PAST STUDIES

2.1 Scientific background

Markov processes discrete in space and continuous in time

Denote by $\{E_1, E_2, ..., E_n\}$ the state space associated with the process $\{X(t), t \geq 0\}$ that is the set of all the possible values which the random variable X(t) can assume.

A Markov Process is completely determined by its transition probabilities:

$P\{X(t'') = j / X(\tau), \tau \leq t'\} = P\{X(t'') = j / X(t') = i\} = P_{ij}(t', t'')$

For the time-homogeneous or stationary case we have:

$P_{ij}(t', t'') = P_{ij}(t = t'' - t')$

For time-homogeneous Markov Processes that are continuous in time, transitions can be studied with two functions $q_i(t), Q_{ij}(t)$

$P_{ii}\{[t, t + \Delta t]\}$	Probability that no transition occurs during the interval $[t, t + \Delta t]$
$\lim\limits_{\Delta t \to 0}\left[\dfrac{1 - P_{ii}\left([t, t + \Delta t]\right)}{\Delta t}\right] = q_i(t)$	Intensity function
$\lim\limits_{\Delta t \to 0}\left[\dfrac{P_{ij}\left([t, t + \Delta t]\right)}{\Delta t}\right] = q_i(t) \times Q_{ij}(t)$	

$Q_{ij}(t)$ is the conditional probability that $X(t + \Delta t) = j$ given that $X(t) = i$ and a random change has taken place in the interval $[t, t + \Delta t]$

For ageing processes, where $E_i \;\; _{i=1,...,n}$ represent n successive condition grades (or deterioration levels) we can assume that: $Q_{i,i+1}(t) = 1$, et $Q_{i, j \neq i+1}(t) = 0$

Transition intensity function & survival time:

If we denote by T_i the survival time in state i and by $h_i(t)$ the risk function associated with the random variable T_i we have:

$$\lim_{\Delta t \to 0}\left[\frac{1 - P_{ii}\left([t, t + \Delta t]\right)}{\Delta t}\right] = \lim_{\Delta t \to 0}\left[\frac{P\left(\mathrm{t} \leq \mathrm{T_i} < t + \Delta t / T_i \geq t\right)}{\Delta t}\right]$$

$$q_i(t) = h_i(t)$$

2.2 Nature of the model

Application of Markov processes in order to study failure rates can be represented with the table below (example with four possible states):

Condition grade & transition function	Probability	Failure rate	Expected Number of events for each condition grade
CG1 $\downarrow$ h1(t1)	P1(t)	FR1(t1)	P1(t) x FR1(t1)
CG2 $\downarrow$ h2(t2)	P2(t)	FR2(t2)	P2(t) x FR2(t2)
CG3 $\downarrow$ h3(t3)	P3(t)	FR3(t3)	P3(t) x FR3(t3)
CG4	P4(t) = 1- P1(t) - P2(t) - P3(t)	FR4(t4)	P4(t) x FR4(t4)

with

t the age of the element

t_i time spent in condition grade i

$h_i(t_i)$ intensity function for condition grade i

$P_i(t)$ probability that the element is in condition grade i at age t

$FR_i(t_i)$ failure rate associated with condition grade i

The expected failure rate can thus be defined with the following expression:

$$\boxed{FR(t) = \sum_{i=1}^{n} P_i(t).FR_i(.)}$$

This general expression can be developed according to assumptions (see 2.3) relative to transition functions $h_i(t_i)$ and failure rate functions $FR_i(t_i)$

This approach is applied to homogeneous pipe categories.

Homogeneous pipe categories can be defined with Poisson regression (*see INSA model*_1).

2.3 Underlying assumptions

Several models are proposed which correspond to various possible assumptions:

Model Name	No of condition grades	Functions for CG1 & parameters	Functions for CG2 & parameters	Functions for CG3 & parameters	Functions for CG4 & parameters
MCG_WE	2	h1: Weibull (λ, p) FR1: Weibull (λ, p)	h2 = 0 FR2: Expo. (λ_2)	/	/
MCG_W2E	3	h1: Weibull (λ, p) FR1: Weibull (λ, p)	h2: Exponential. (λ_2) FR2: Exponential. (λ_2)	h3 = 0 FR3: Expo. (λ_3)	/
MCG_4E	4	h1: Exponential (λ_1) FR1 = 0	h2: Exponential. (λ_2) FR2: Exponential. (λ_2)	h3: Expo. (λ_3) FR3: Expo. (λ_3)	h4 = 0 FR4: Expo. (λ_4)
MCG_nE	n	…	…	…	…

2.4 Algorithm

Equations:

MCG_WE Model:

Using a Weibull function and an Exponential function (Le Gauffre *et al.*, 2000) we have:

State	Probability	Transition and Failure rate functions
CG1	$P_1(t) = S_{weibull}(t) = P(T_1 > t) = \exp(-(\lambda t)^p)$	$h_1(t) = p\lambda^p t^{p-1}$
CG2	$P_2(t) = 1 - P_1(t)$	$h_2 = \lambda_2$

This gives:

$\boxed{FR(t) = \exp(-(\lambda t)^p\ p\lambda^p t^{p-1} + (1 - \exp(-(\lambda t)^p)\lambda_2}$ Expected Failure Rate at age t

MCG_W2E Model:

Using a Weibull function and two Exponential functions (Le Gauffre *et al.*, 2000) we have:

State	Probability	Transition and Failure rate functions
CG1	$P_1(t) = S_{weibull}(t) = P(T_1 > t) = \exp(-(\lambda t)^p)$	$h_1(t) = p\lambda^p t^{p-1}$
CG2	$P_2(t) = 1 - P_1(t) - P_3(t)$	$h_2 = \lambda_2$
CG3	$P_3(t) = \Pr(T_1 + T_2 < t) = \int_{t_1=0}^{t} \left(f_1(t_1) \int_{t_2=0}^{t-t_1} f_2(t_2)dt_2 \right) dt_1$ $= \int_{t_1=0}^{t} f_1(t_1)F_2(t-t_1)dt_1$ $= \int_{t_1=0}^{t} p\lambda(\lambda t_1)^{p-1} \exp(-(\lambda t_1)^p)\left(1 - \exp(-\lambda_2(t-t_1))\right)dt_1$ (*numerical integration*)	$h_3 = \lambda_3$

MCG_4E Model:

Using four Exponential functions we have (Le Gauffre *et al.,* 2001) :

State	Probability	Transition and Failure rate functions
CG1	$P_1(t) = \Pr(T_1 > t) = \exp(-\lambda_1 t)$	$h_1 = \lambda_1$ & $FR_1(t) = 0$
CG2	$P_2(t) = \frac{\lambda_1}{(\lambda_2 - \lambda_1)} \exp(-\lambda_1 t) + \frac{\lambda_1}{(\lambda_1 - \lambda_2)} \exp(-\lambda_2 t)$	$h_2 = \lambda_2$
CG3	$P_3(t) = \frac{\lambda_1\lambda_2 \exp(-\lambda_1 t)}{(\lambda_2 - \lambda_1).(\lambda_3 - \lambda_1)} + \frac{\lambda_1\lambda_2 \exp(-\lambda_2 t)}{(\lambda_1 - \lambda_2).(\lambda_3 - \lambda_2)} + \frac{\lambda_1\lambda_2 \exp(-\lambda_3 t)}{(\lambda_1 - \lambda_3).(\lambda_2 - \lambda_3)}$	$h_3 = \lambda_3$
CG4	$P_4(t) = 1 - P_1(t) - P_2(t) - P_3(t)$	$h_4 = \lambda_4$

$\boxed{FR(t) = \sum_{i=2}^{4} P_i(t)\lambda_i}$ Expected Failure Rate at age t

2.5 Past studies and conclusions

Malandain (1999) has calibrated ageing models (MCG_W2E and MCG_WE) on a sample of grey cast iron sections (231 km in Villeurbanne, France) laid between 1954 and 1968 and observed from 1982 to 1997.

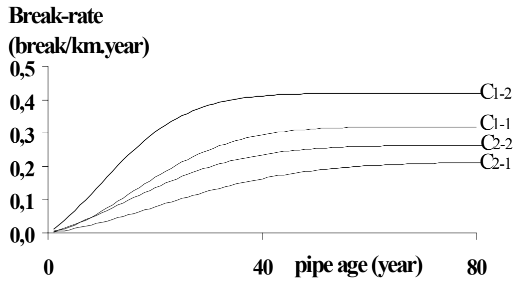


fig. 1

Four categories of grey cast iron mains have been studied:

C_{1-1} : 60 or 80 mm diameter, under footpath

C_{1-2} : 60 or 80 mm diameter, under roadway

C_{2-1} : >80 mm diameter, under footpath

C_{2-2} : >80 mm diameter, under roadway

Using these failure rate functions, we can compare observed and modelled yearly failure rates.

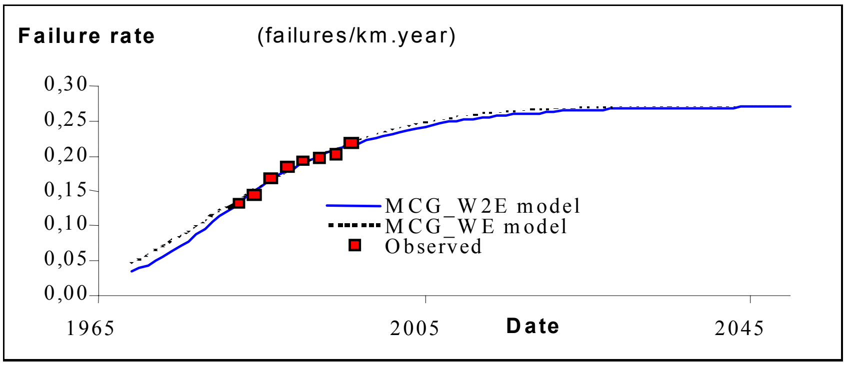


fig. 2

Introducing hypothesis on rehabilitation rates, it is thus possible to forecast failures rates or failures numbers for the next years.

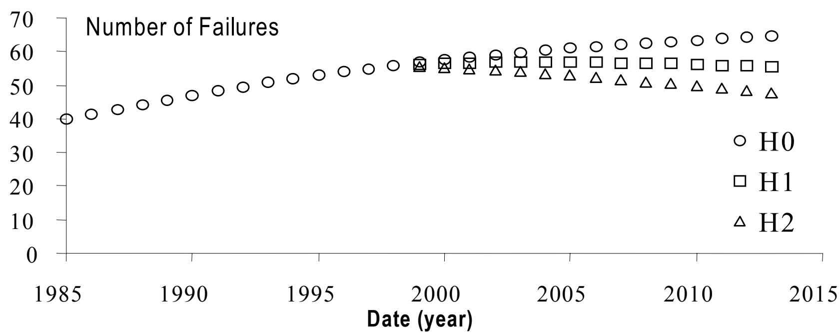


fig. 3

H0, H1 and H2 are 3 hypothesis about renewal rates and criteria.

H0: no renewal,

H1: 1% per year random renewal, and

H2: 1% per year renewal of oldest sections in category C_{1-2}

3 SPECIFICATIONS OF COMPUTATIONAL STEPS

3.1 Functional Scheme

1) Definition of pipe categories (*a priori*, or after Poisson regression)

↓

2) Raw data formatting and selection of data that will be used for calibration

↓

3) Calibration and comparison of MCG models

↓

4) Failure rate forecasting with various hypotheses on renewal rates

3.2 Raw Data Formatting

Describe how raw data are formatted to be analyzed. Raw data, originating from maintenance records or hydraulic modeling for instance, have to be formatted before being used as input of the model. This step, very often crucial and time consuming, assures the validity of the input data, as well as their compliance with the requirements of the model computation procedure. Inter alia, it will be a question of explaining what to do in case of missing data, partially or not.

FILE A-1 : Data relative to pipe sections

(1)	(2)	(3)	(4)	(5)	(6)	(7)	..		..
Id	**Length (m)**	**Start_Obs** first year of failure records	**End_Obs** last year of failure records	**Missing Data** No of years ; years	**Category**	**Laying date**	Z2 Material	..	Zi Traffic
1253	50	1982	2000	*3;1990;1991;1992*	4	1965	GCI	..	H
8502	100	1993	2000	*0*	12	1972	DI		VH

FILE A-2 : Maintenance data

Id	**Laying_date**	**Obs_Year**	*Age*	**Failure_type#1**	**Failure_type#2**	Reliability $\in [high, medium, low]$
1253	1965	1997	32	1	0	

↓

Table B – Raw data for a given pipe category

Y_L Laying Date	**Y_Obs** Year of Observation	Age	**Total Length** (km)	**NoF** Number of Failures
…	…	…	…	…
1955	1995	40	0.550	2
…	…	…	…	…

↓

Table C.1 – NoF(Y_L,Y_Obs) Y_obs: Year of Observation, Y_L: Laying Date

Selection of data used for calibration (or for cross validation)

	Selected	…	☐	☐	☐	☑	☑	☑	☑	☑	☑	☑	☑	
	Y_obs	…	…	…	1992	1993	1994	1995	1996	1997	1998	1999	2000	…
Selected	Y_L													
☑	…													
☑	1955							2						
☑	…													
☑	1968													
☑	1969													
No of Failures for Y_Obs														

Table C.2 – Length(Y_L,Y_Obs)

	Selected (from C.1)		☐	☐	☐	☑	☑	☑	☑	☑	☑	☑	☑	
	Y_obs		…	…	…	…	1994	1995	1996	1997	1998	1999	2000	…
Selected	Y_L													
☑	…													
☑	1955							550						
☑	…													
☑	1967													
☑	1969													
Length(Y_Obs)														

For selected data:

Table C – Failure rates & total length of pipes observed at age t

t (Age of pipes)	FR(t)	Length(t)
…	…	…
…	…	…

3.3 List and Definition of Explanatory Factors

Definition of section categories can be done *a priori* or after statistical analysis.

Poisson Regression (*see Poisson model*) provides a way of defining homogeneous categories.

3.4 Model Parameters Estimation or Assignation

Denote by

$L(t',t'')$: the total length of sections laid in year t' and observed in year t''(for selected data)

$NoF_{obs}(t',t'')$: the total number of failures for sections laid in year t' and observed in year t''.

$L(t)$ the total length of sections observed at age t, and $NoF_{obs}(t)$ the number of observed failures at age t are given by:

$$L(t) = \sum_{t'} \sum_{t''=t'+t} L(t',t'')$$

$$NoF_{obs}(t) = \sum_{t'} \sum_{t''=t'+t} NoF_{obs}(t',t'')$$

The observed failure rate at age t is thus:

$$r_{obs}(t) = \frac{NoF_{obs}(t)}{L(t)}$$

Each MCG model (modelling $r(t)$) is calibrated by a Least Squares Minimisation, with:

$$\boxed{E^2 = \sum_t (r(t) - r_{obs}(t))^2 .(L(t))^K} \quad \text{to minimise}$$

Differences between observed failure rates and calculated failure rates are weighed according to the total length of pipes observed at age t.

3.5 Output

We can distinguish two levels of results.

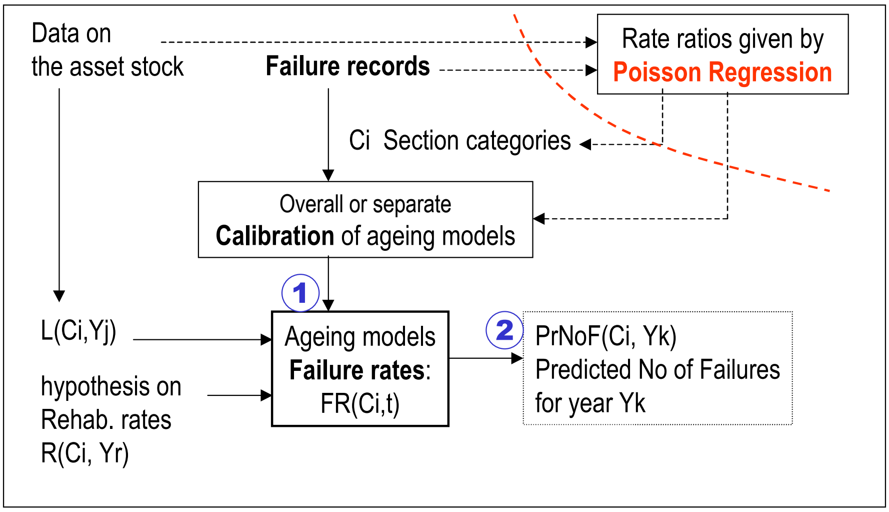


fig. 4

1) Failure rate $FR(C_i, t)$ for each section category C_i :

$FR(.)$ is given by each MCG model after calibration with failure records.

2) Failure rates or predicted number of failures $PNoF(C_i, Y_k)$ for year Y_k , according to

- the chosen MCG model
- data on the asset stock: $L(C_i, Y_j)$ is the total length of pipes in category C_i laid in year Y_j
- and hypothesis on renewal rates: $R(C_i, Y_r)$ is the renewal rate in category C_i for year Y_r

3.6 Model Validation or calibration

(Statistical model)

External Validation or Cross Validation – Statistical test(s)

We can propose two ways to define data-set 1 (for calibration) and data-set 2 (for validation)

Fist method: available data are divided into two data-sets according to the laying dates of pipe sections.

		Data-set 1	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	
		Data-set 2	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	
		Y_obs	…	…	1992	1993	1994	1995	1996	1997	1998	1999	2000	…
	☑	1954												
☑	☐	1955												
…	…	…												
☐	☑	…												
☑	☐	…												
☐	☑	1967												
☑	☐	1968												
☐	☑	1969												

Second method: available data are divided into two data-sets according to the observation dates.

This method allows testing the influence of failure records.

		Data-set 1	☑	☑	☑	☑	☑	☑	☐	☐	☐	☐	☐	
		Data-set 2	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	
		Y_obs	…	…	1992	1993	1994	1995	1996	1997	1998	1999	2000	…
Data-set 1	**Data-set 2**	Y_L												
☑	☑	1954												
☑	☑	1955												
☑	☑	…												
☑	☑	…												
☑	☑	…												
☑	☑	1967												
☑	☑	1968												
☑	☑	1969												

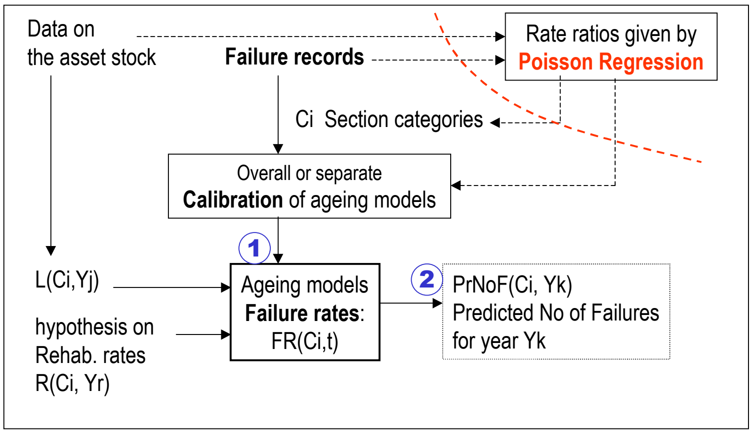
3.7 PI(s) Estimation Method

Give the precise definition of the PI(s) to be estimated. Explicit the procedure for computing PI(s) as function(s) of the dependent variable(s) (in some cases PI and dependent variable may be the same).

3.8 PI(s) Forecasting Method

Give the computational procedure for PI(s) forecasting.

Output: Cf. 3.5.



This approach provides a forecast of the failure rate for a given pipe category.

A forecast can be made for the asset stock in dividing this stock into a set of pipe categories.

This forecast can take into account hypothesis on rehabilitation rates applied to these categories, and could be used for the simulation of strategies.

A) Connection between a Poisson regression analysis & the Markov models:

MCG Models can be calibrated for a single pipe category defined a priori or after Poisson regression.

One possible improvement of these models consists in calibrating a set of ageing models in using results from Poisson Regression.

Typically, Poisson Regression gives the following results:

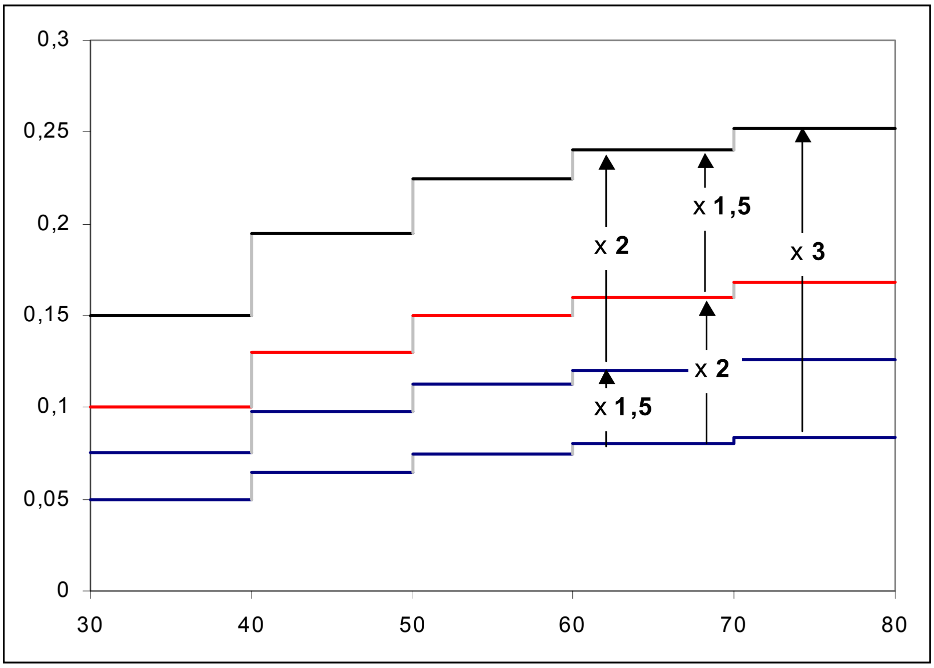


fig. 5

	X2 = 0	X2 = 1	Rate Ratio = RR2 = 2.0
X1 = 0	Category C0	Category C2	RR (C0 → C2) = 2.0
X1 = 1	Category C1	Category C3	RR (C1 → C3) = 2.0
Rate Ratio = RR1= 1.5	RR (C0 → C1) = 1.5	RR (C2 → C3) = 1.5	RR (C0 → C3) = 1.5 x 2 = 3

In the above example, four categories have been defined with two binary variables, and for each of these categories 5 sub-categories correspond to age groups.

Poisson regression provides Rate Ratios that can be used to calibrate an overall multi-state model associated with n categories.

Let us denote by:

$FR_i(t), i \in \{0,1,2,...,n\}$ the modelled failure rates

$FR_{i,obs}(t)$ the observed failure rates

A baseline ageing function is defined

$FR_0(t)$ failure rate for category C_0 chosen as reference

$FR_i(t) = \alpha_i FR_0(t)$ with:

$\alpha_i = RR(C_0 \to C_i)$ the rate ratio calculated by Poisson regression

The calibration of an overall ageing model can be done in minimising:

$$\boxed{E^2 = \sum_{i=0}^{n}\left[\sum_t \left(FR_i(t) - FR_{i,obs}(t)\right)^2 .\left(L_i(t)\right)^K\right] = \sum_{i=0}^{n}\left[\sum_t \left(\alpha_i FR_0(t) - FR_{i,obs}(t)\right)^2 .\left(L_i(t)\right)^K\right]}$$

with

$L_i(t)$ the total length of pipes belonging to category C_i observed at age t

With the MCG_WE model, the expression becomes:

$$\boxed{E^2 = \sum_{i=0}^{n}\left[\sum_t \left(\alpha_i \left(\exp(-(\lambda t)^p\, p\lambda^p t^{p-1} + (1-\exp(-(\lambda t)^p)\lambda_2\right) - FR_{i,obs}(t)\right)^2 .\left(L_i(t)\right)^K\right]}$$

λ, p, λ_2 are Weibull and Exponential parameters that have to be determined

α_i are rate ratios that can be given by Poisson Regression, or parameters that have to be determined.

B) **<u>Markov models taking into account the effects of past rehabilitation programmes:</u>**

Application of Markov processes in order to study failure rates can be represented with the table below (example with four possible states):

Condition grade & transition function	Failure rate	Probabilities		
CG1 ↓ h1(t1)	FR1(t1)	P1(t)	P'1(t) = P1(t)	P*1(t) = P'1(t) / (1-R(t))
CG2 ↓ h2(t2)	FR2(t2)	P2(t)	P'2(t) = P2(t)	P*2(t) = P'2(t) / (1-R(t))
CG3 ↓ h3(t3)	FR3(t3)	P3(t)	P'3(t) = P3(t)	P*3(t) = P'3(t) / (1-R(t))
CG4		P4(t) = 1- P1(t) - P2(t) - P3(t)		
CG4 – not rehabilitated	FR4(t4)		P'4(t) = P4(t) – R(t)	P*4(t) = P'4(t) / (1-R(t))
CG4 – rehabilitated	/		R(t)	

with

t	the age of the element
t_i	Time spent in condition grade i
$h_i(t_i)$	Intensity function for condition grade i
$P_i(t)$	Probability that the element is in condition grade i at age t, with no rehabilitation
$R(t)$	Probability that the element has been rehabilitated before age t
$P'_i(t)$	Probability that the element is in condition grade i at age t, and is not rehabilitated
$P_i^*(t)$	Probability that the element is in condition grade i at age t, given that the element has not been rehabilitated (conditional probability)
$FR_i(t_i)$	Failure rate associated with condition grade i

The expected failure rate, due to the ageing process and with no rehabilitation programmes, can be defined with the following expression:

$$FR(t) = \sum_{i=1}^{n} P_i(t).FR_i(.)$$

The expected failure rate, due to the ageing process and rehabilitation programmes, can be defined with the following expression:

$$FR^*(t) = \sum_{i=1}^{n} P_i^*(t).FR_i(.)$$

This type of model can be calibrated by a Least Squares Minimisation, with:

$$E^2 = \sum_t \left(FR^*(t) - FR_{obs}(t)\right)^2.\left(L(t)\right)^K$$ to minimise

Some examples are given in fig. 6. For these examples, r is a constant rehabilitation rate applied to non rehabilitated elements (1-R(t)) after age t = 20 years (hypothesis).

Fig. 6a corresponds to the hypothesis: r = 0 (no previous rehabilitation for these studied pipes).

Fig. 6b, 6c, 6d, correspond to the hypotheses r = 0.2%, 0.5%, 1% respectively.

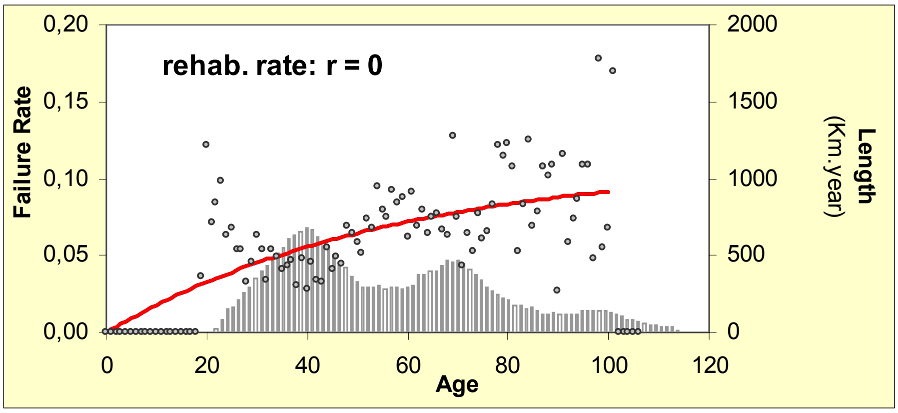


fig. 6a

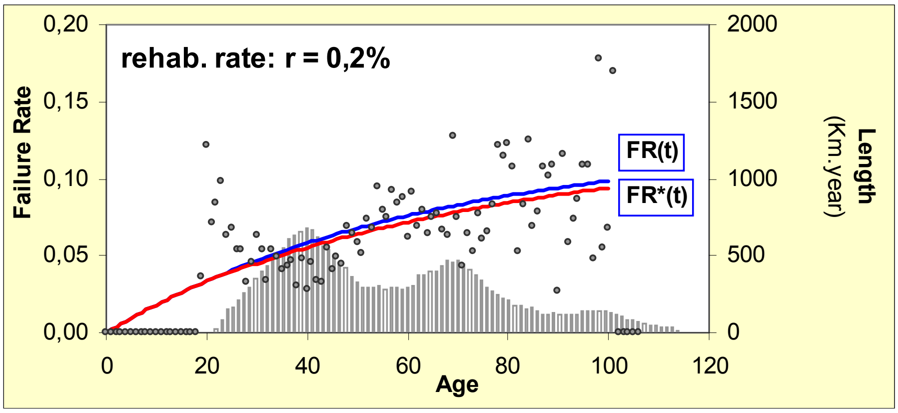


fig 6b

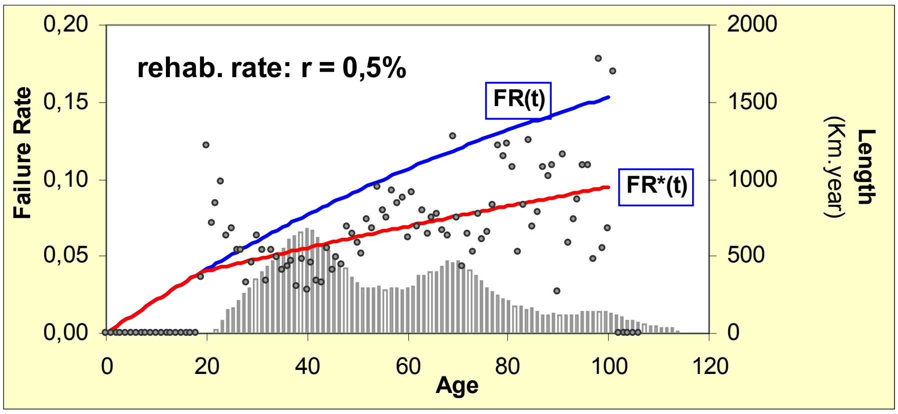


fig. 6c

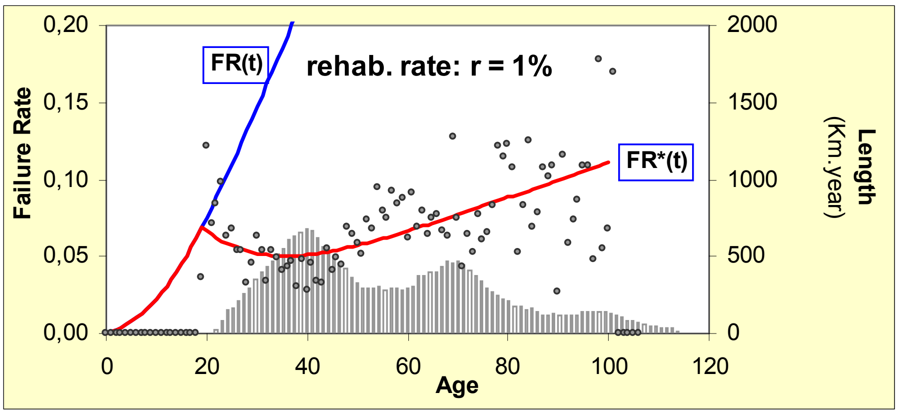


fig 6d

4 SOFTWARE SPECIFICATIONS

4.1 Programming Language(s) or Mathematical-Statistical Software(s)

Microsoft Access

Microsoft Excel

4.2 Theoretical Framework References

BHARUCHA-REID A.T. (1960) *Elements of theory of Markov Processes and their applications*. New York : McGraw-Hill, 468 p.

4.3 Practical Use and Results References

LE GAUFFRE *et al*. (2001) « *Projet CAPTUR : Consolidation d'un cadre d'Analyse des Patrimoines Techniques Urbains de type Réseau – Cas des réseaux d'eau et d'assainissement ».* Rapport de fin de contrat n°99 V 0492 Action Concertée Incitative Ville – Ministère de l'Éducation Nationale, de la Recherche et de la Technologie. Lyon (F) : INSA Lyon / URGC Hydrologie Urbaine.204 p.

LE GAUFFRE P., MALANDAIN J., MIRAMOND M.(2000) Modélisation du vieillissement et maintenance des réseaux d'eau potable. *Revue Française de Génie Civil,* Volume 4 – n°2-3/2000, p. 397-410.

MALANDAIN J., (1999) *Modélisation de l'état de santé des réseaux de distribution d'eau pour l'organisation de la maintenance. Étude du patrimoine de l'agglomération de Lyon*. Thèse de Doctorat n° 99 ISAL 0040 de l'INSA de Lyon, URGC / Hydrologie Urbaine, 206 p.

APPENDIX 2 : POISSON MODEL

1 GENERAL DESCRIPTION

1.1 Name and/or Acronym of the Model

POISSON REGRESSION

1.2 Company/Research Center/University

INSA Lyon

1.3 Objectives

Output from a Poisson Regression are:

→ The influence of failure factors can be characterised with Rate Ratios calculated by the Poisson Regression:

e.g.: RR(Under Roadway / Under Footpath) = 2.0 means that the failure rate for sections situated under roadway is estimated to be 2 times higher than the failure rate for sections situated under footpath.

Statistical tests and confidence intervals for Rate Ratios allow selecting a set of significant variables.

→ The asset can be divided into Pipe categories defined in combining the statistical variables that are significant according to the Poisson Regression Analysis

→ PFR(Ci) is the predicted Failure Rate for category Ci.

This predicted failure rate can be applied to a single section in addition to information available for this section.

→ Each category can be characterised with a Rate Ratio: $RR(C_i) = PFR(C_i) / Reference\_Failure\_Rate$

→ A set of indices can be calculated in order to evaluate the efficiency of the dividing of the asset into categories: e.g. $\%TL(80) = \%$ of the total length corresponding to 80% of the failures

1.4 Functional description

- Data available that are supposed to correspond to explanatory factors for failure (bursts or leaks) rates

↓

- Pipe categories defined *a priori*: data on failures (bursts or leaks)

Poisson Regression:

↓ *Analysis of the relationship between a count (No of failures) with a Poisson distribution and a set of explanatory variables.*

- Variables, which are significant according to the Poisson Regression: Rate Ratios

- Pipe categories defined with variables selected by Poisson Regression: Predicted Failure Rates, Rate Ratios

1.5 Brief Historical Overview of the Model

A Poisson regression is used to estimate models of the number of occurrences (counts) of an event.

It provides an analysis of the relationship between a count (No of failures) with a Poisson distribution and a set of explanatory variables

A Poisson regression analysis has been applied by INSA during the Ph-D Thesis by (Malandain, 1999) dedicated to the water mains of the Water Supply System owned by the Urban Community of Lyon and operated by Vivendi Water (Générale des Eaux).

Limited historical data on the whole asset (5 years, 1993 → 1997) and incomplete data concerning laying dates make it impossible to apply models based on rich historical data. However, the important length of the asset (2980 km – about 40 000 sections) and the availability of urban data (about the infrastructure, traffic, soil, etc.) allowed to study failure factors (and to define pipe categories) in using Poisson regression.

Predicted failure rates are determined for pipe categories but can be used for a single section in addition to the limited information available at this scale (No of previous failures since 1993).

For failures on pipes, 80 categories have been defined after Poisson regression.

For failures on joints, 32 categories have been defined after Poisson regression.

2 THEORETICAL FRAMEWORK OVERVIEW/PAST STUDIES

2.1 Scientific background

A Poisson regression analysis is a particular form of regression modelling.

A Poisson regression model provides an analysis of the relationship between a count (No of failures) with a Poisson distribution and a set of explanatory variables.

Statistical tests or confidence intervals make it possible to define statistical variables that are significant.

2.2 Nature of the model

→ Statistical model (providing a function *g* reflecting the relationship between a dependant variable y and the explanatory variables $x_1, x_2, ...x_n$

→ Approach level: Pipe Categories

→ Dependent variable: No of failures, for a given exposure

→ Exposure (km.years): $e_j = \sum_i Length(i)histo(i)$

Each category C_j corresponds to a set of sections. *Length(i)* is the length of each section, *histo(i)* is the duration of the failure records corresponding to this category.

2.3 Underlying assumptions

1) There is a quantity called the *incidence rate* (failures/km.year) that is the rate at which events occur.

2) The incidence rate can be multiplied by exposure to obtain the expected number of observed events.

3) Over very small exposures ε , the probability of finding more than one event is small compared with ε

4) Non overlapping exposures are mutually independent.

2.4 Algorithm

Equations:

In the Poisson Regression model the incidence rate for the *j*th observation (pipe category) is assumed to be given by

$$\boxed{\ln(r_j) = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,j}} \quad (1)$$

$$\boxed{r_j = \exp(\beta_0)\prod_{i=1}^{n} \exp(\beta_i x_{i,j})} \text{ incidence rate} \quad (2)$$

If e_j is the exposure, the expected number of events is

$$\boxed{\lambda_j = e_j r_j} \text{ expected number of events} \quad (3)$$

2.5 Past studies and conclusions

A Poisson regression analysis has been applied by INSA during the Ph-D Thesis by (Malandain, 1999) dedicated to the water mains of the Water Supply System owned by the Urban Community of Lyon and operated by Vivendi Water (Générale des Eaux).

Limited historical data on the whole asset (5 years, 1993 → 1997) and incomplete data concerning laying dates make it impossible to apply models based on rich historical data. However, the important length of the asset (2980 km) and the availability of urban data (about the infrastructure, traffic, soil, etc.) allowed to study failure factors (and to define pipe categories) in using Poisson regression.

Predicted failure rates are determined for pipe categories but can be used for a single section in addition to the limited information available at this scale (No of previous failure since 1993).

For failures on joints, 32 categories have been defined after Poisson regression.

For failures on pipes, 80 categories have been defined after Poisson regression.

Variables	Risk (or Rate) Ratio	95 % confidence Intervals	
Geol_0 « Urban Soil »	**1**	/	/
Geol_1 « Nappe Alluviale Fluvio-Glaciaire »	0.59	**0.51**	**0.68**
Geol_2 « Alluvions Fluviatiles Wurmiennes »	0.69	**0.56**	**0.85**
Geol_3 « Moraines Argileuses »	**2.98**	0.46	19.21
Geotech_0 « No Geotech. Risk »	**1**	/	/
Geotech_1 « Geotech. Risk (soil movements)»	**1.12**	0.77	1.62
∅ = 60-80 mm	**1**	/	/
∅ = 100-135 mm	**0.76**	**0.69**	**0.83**
∅ = 150-175 mm	**0.35**	**0.30**	**0.42**
∅ = 200-350 mm	**0.15**	**0.12**	**0.19**
∅ ≥ 400 mm	**0.01**	**0.001**	**.11**
Mat_0 « Ductile Iron »	/	/	/
Mat_1 « Grey Cast Iron »	**12.49**	**10.34**	**15.08**
TR_0 « Under Footpath »	/	/	/
TR_1 « Under roadway – light traffic or Mun. Road)	**1.29**	**1.14**	**1.45**
TR_2 « Under roadway – heavy traffic or Dep. Road)	**1.39**	**1.26**	**1.53**
TR_3 « Under roadway – very heavy traffic or Nat. Road)	**2.06**	**1.44**	**2.97**

The above table shows preliminary results with 320 categories. Statistical tests on these results lead to group two variables (Geol_1 and Geol_2 → "Alluvium") and to eliminate two variables: Geol_3 and Geotech_1 that cannot be considered as significant (confidence interval including 1). Finally 80 categories have been defined with the remaining variables.

These results display low differences between light and heavy or very heavy traffic, and can be considered an underestimation of relative risks. Actually it has been shown, (Malandain *et al*., 1999), that the position of pipes must be considered as uncertain data. In studying this uncertainty on a sample (a 211km network in Villeurbanne) with a Bayesian approach, the point estimate of the relative risk increased from 1.6 to 4.4. These results demonstrate that improving accuracy and reliability of databases appears to be highly profitable for failure modelling.

Pipe Location	Rate Ratio: Point estimate and 95% Conf. Interv., without considering uncertainty of data	RR: Point estimate and 95% Conf. Interv, in considering uncertainty of data (Bayesian approach)
L0: Under Footpath	1	1
L1: Under Roadway	**1.64** ; [1.06 ; 2.54]	**4.4** ; [2.68 ; 5.96]

3 SPECIFICATIONS OF COMPUTATIONAL STEPS

3.1 Functional Scheme

- Data available that are supposed to correspond to explanatory factors for failure (bursts or leaks) rates

↓

- Pipe categories defined *a priori*: data on failures (bursts or leaks)

Poisson Regression:

↓ *Analysis of the relationship between a count (No of failures) with a Poisson distribution and a set of explanatory variables.*

- Variables that are significant according to the Poisson Regression: Rate Ratios
- Pipe categories defined with variables selected by Poisson Regression: Predicted Failure Rates, Rate Ratios

3.2 Raw Data Formatting

FILE A-1 : Data relative to pipe sections

(1)	(2)	(3)	(4)	(5)			..		..
Id	**Length (m)**	**Start_Obs** first year of failure records	**End_Obs** last year of failure records	**Missing Data** No of years ; years	Z1 Laying date	Z2 Material	..	Zi Traffic	..
1253	50	1982	2000	*3;1990;1991;1992*	1965	GCI	..	H	..
8502	100	1993	2000	*0*	1972	DI		VH	

FILE A-2 : Maintenance data

Id	Laying_date	Obs_Year	*Age*	Failure_type#1	Failure_type#2	...
1253	1965	1997	32	1	0	

↓

FILE B: Data relative to categories

Category	Exposure (km.years)	No of Failures Type #1	No of Failures Type #2	Z1 Age group	Z2 Material	Z3 Diameter	Z4 Traffic and Position	..
..								
4	580	42	15	30-40	GCI	60	URW_HT	

↓

FILE C: Data relative to categories – Binary variables

Category	Exposure (km.years)	No of Failures Type #1	No of Failures Type #2	*R1*	X1	X2	X3	X4	*R2*	X5	....
4	580	42	15	0	0	0	1	0	..	..	..

R_i are binary variables which are not used in the regression. These variables correspond to modalities of variables Z_i which are taken as reference.

X_i are binary variables (or design variables) used as explanatory variables in the Poisson regression.

Each design variable X_i corresponds to a given possible value of a variable Z_j or corresponds to a sub-set or interval.

Example for Z_3 (Diameter):

Z_3 Diameter	R_3 *60-80*	X_9 **100-135**	X_{10} **150-175**	X_{11} **200-350**	X_{12} **>400**
if $Z_3 \in [60,80]$	*1*	0	0	0	0
if $Z_3 \in [100,135]$	*0*	**1**	0	0	0
if $Z_3 \in [150,175]$	*0*	0	**1**	0	0
if $Z_3 \in [200,300]$	*0*	0	0	**1**	0
if $Z_3 \geq 400$	*0*	0	0	0	**1**

Example for Z_4 (Traffic x Position):

Z_4 Traffic x Position	R_4 *UFP*	X_{13} **URW_LT**	X_{14} **URW_HT**	X_{15} **URW_VHT**
Under Footpath	*1*	0	0	0
Under Roadway & Light Traffic	*0*	**1**	0	0
Under Roadway & Heavy Traffic	*0*	0	**1**	0
Under Roadway & Very Heavy Traffic	*0*	0	0	**1**

3.3 List and Definition of Explanatory Factors

3.3.1 Required data

FILE A.1 (data on pipe sections)

→ Length (m)

→ First year of failure records

→ Last year of failure records

→ Missing data (list of years)

→ Material

FILE B (data on categories)

→ Total length (km)

→ Exposure (km.years)

→ Material

3.3.2 Highly recommended factors

FILE A.1 (data on pipe sections)

→ Laying date

→ Diameter

→ Class of Traffic

→ Position (under roadway / footpath)

→ Class of Soil

→ Water pressure

FILE B (data on categories)

→ Age group

→ Diameter

→ Class of Traffic

→ Position (under roadway / footpath)

→ Class of Soil

→ Class of Pressure

3.3.3 Possibly Useful factors

→ type of joint

→ depth

→ type of joint

→ class of depth

3.4 Model Parameters Estimation or Assignation

For each parameter β_i associated to the explanatory variable X_i statistical software provide a point estimate and a confidence interval.

Point estimates (b_i) are calculated in maximising the log-likelihood function L which is the log-transform of the joint probability of the observations:

$$\boxed{L = \ln\left\{\prod_{j=1}^{k} \Pr(Y_j = y_j)\right\} = \sum_{j=1}^{k} \ln(\Pr(Y_j = y_j))} \quad (1)$$

Using the Poisson distribution we have:

$$\Pr(Y_j = y_j) = \frac{\exp(-\lambda_j)\lambda_j^{y_j}}{y_j!}$$

with

Y_j Number of events observed for the *j*th category

$\lambda_j = r_j \times e_j = \exp(\xi_j)$ Expected number of events in the *j*th category

$\xi_j = \ln(r_j) + \ln(e_j)$

$\ln(r_j) = X_j^t \beta = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,j}$

e_j Exposure for the *j*th category

$r_j = \exp(\beta_0) \prod_{i=1}^{n} \exp(\beta_i x_{i,j})$ Expected rate for the *j*th category

In replacing terms in expression (1) function L can be expressed by (2)

$$\boxed{L = \sum_{j=1}^{k} \left\{ -\exp(\xi_j) + \xi_j y_j - \ln(y_j!) \right\}} \quad (2)$$

3.5 Output

Final output from a Poisson Regression are:

→ The influence of failure factors can be characterised with Rate Ratios calculated by Poisson Regression (point estimates and confidence intervals).

e.g. RR(URW_HT / UFP) = 2.4 (the failure rate for sections situated under roadway with heavy traffic is estimated to be 2.4 times higher than the failure rate for sections situated under footpath).

→ With these results, the asset can be divided into Pipe categories defined in combining the statistical variables that are significant according to the Poisson Regression Analysis

PFR(Cj) Predicted Failure Rate for category Cj is given by:

$$PFR(C_j) = r_j = \exp(b_0) \prod_{i=1}^{n} \exp(b_i x_{i,j})$$

with: b_i the point estimate for coefficient β_i

On this basis, each category can be characterised with a Rate Ratio:

$$RR(C_j) = PFR(C_j) / \text{AverageFailureRate}$$

A set of indices can be calculated in order to evaluate the efficiency of the dividing of the asset into categories:
e.g. $\%TL(80) = \%$ of the total length corresponding to 80% of the failures

These indices are quite similar to the Gini Index.

3.6 Model Validation or calibration

(Statistical model)

3.6.1 Check of Parameters Significance – Internal Validation (Statistical model)

a) Parameters significance

Poisson regression analysis will lead to select and group explanatory variables according to the confidence interval calculated for each regression coefficient, or for each IRR (Incidence Rate Ratio).

The incidence rate ratio for one-unit change in x_i is $\exp(\beta_i)$

Intermediate outputs are:

Variable	Rate Ratio (Point estimate)	Std. Err.	z	P>|z|	[95% confidence interval]	
..	..					
X_{12} Soil_2	1.46	0.40	1.37	0.171	**0.84**	**2.52**
X_{16} Pressure	2.82	0.79	3.71	0.000	1.63	4.88

In the example above we can see that the 95% confidence interval for $\exp(\beta_{12})$ includes 1. This brings to consider that X_{12} is not a significant variable that can be used to define pipe categories.

There is no evidence that allows distinguishing "Soil_2" from "Soil_1" which has been taken as a reference.

In this case, pipe categories will be redefined in grouping these two types of soil.

b) Sensitivity to the value of explanatory factors

In (Malandain *et al.*, 1999) we have shown that uncertain data (e.g. position under roadway or under footpath) could lead to underestimate the incidence rate ratios.

A Bayesian approach has been proposed in order to quantify the effect of an uncertain data (Malandain *et al.*, 1999).

Another way to study the effect of uncertain data can be tested during the CARE-W project. This can be done in using a reliable database in which we can introduce a certain proportion of false data.

c) Goodness of fit

Statistical software provide two different ways to assess the accuracy of a Poisson Model.

- the differences between the estimated values and observed values can be evaluated by a Pearson chi-square statistic.
- Likelihood statistics measure the goodness-of-fit of a Poisson model and can be used to compare two nested models.

3.6.2 External Validation or Cross Validation – Statistical test(s)

External validation can be done in dividing the maintenance data into two data sets. For example if the available maintenance data cover a period from 1990 to 2000 (11 years) events from 1990 to 1996 (data set #1) can be used to calibrate the Poisson model and events from 1997 to 2000 (data set #2) can be used to compare observed failure rates and counts with predicted failure rates and counts.

Data set #1

For each category, Poisson regression gives:

$\boxed{PFR(C_j) = r_j = \exp(b_0)\prod_{i=1}^{n}\exp(b_i x_{i,j})}$ Expected failure rate for the *j*th category

We can calculate $RR(C_j) = {}^{r_j}\!/_{r}$ which is the estimated rate ratio for the *j*th category (r being the average failure rate)

Data set #2

Given a certain exposure e_j for the *j*th category in the second data set we obtain:

$\boxed{ExNoF(C_j) = \lambda_j = r_j e_j}$ Expected number of failures in the *j*th category

This expected number can be compared to $NoF(C_j)$ number of failures observed in the *j*th category

However, in order to take into account the possible effects of climatic conditions we suggest to compare

$RR(C_j / C_0) = {}^{r_j}\!/_{r_0}$ Rate ratio calculated with data set #1

and

$Obs\_RR(C_j / C_0) = {}^{obs\_r_j}\!/_{obs\_r_0}$ Rate ratio observed with data set #2

with C_0 a category chosen as reference.

3.7 PI(s) Estimation Method

PI: Expected number of failures for a given period

For a given category C_j, this estimate is directly determined from Poisson regression.

For a given section $L_k \in C_j$ the expected failure rate is $r_k = r_j$ and the expected number of events during one year is $ExNoF(L_k) = r_k.length(k)$

4 SOFTWARE SPECIFICATIONS

4.1 Programming Language(s) or Mathematical-Statistical Software(s)

Preparation of data: **Access**

Statistical analysis: **Stata Statistical Software** Release 6.0. College station Texas (USA) Stata Corp.

5 REFERENCES

5.1 Theoretical Framework References

SELVIN, S. (1995) *Practical Biostatistical Methods*, Duxbury Press ITP, 503 p., ISBN 0-534-23802-5.

StataCorp. (1999) *Stata Reference manual - Release 6.0.* College Station, Texas (USA) Stata Press.

5.2 Practical Use and Results References

MALANDAIN J., LE GAUFFRE P., MIRAMOND M. (1999) Organizing a decision support system for infrastructure maintenance: application to water supply systems. *Journal of Decision*, Vol. 8 – N°2/1999, pp.203-222.

LE GAUFFRE P., MALANDAIN J., MIRAMOND M.(2000) Modélisation du vieillissement et maintenance des réseaux d'eau potable. *Revue Française de Génie Civil,* Volume 4 – n°2-3/2000, p. 397-410.

MALANDAIN J., (1999) *Modélisation de l'état de santé des réseaux de distribution d'eau pour l'organisation de la maintenance. Etude du patrimoine de l'agglomération de Lyon*. Thèse de Doctorat n° 99 ISAL 0040 de l'INSA de Lyon, URGC / Hydrologie Urbaine, 206 p.

APPENDIX 3 : PHM MODEL

1 GENERAL DESCRIPTION

1.1 Name and/or Acronym of the Model

PHM Model

1.2 Company/Research Center/University

Cemagref / ORH Unit

1.3 Objectives

The main objective of PHM Model is to portray approximately the distribution of the random variable consisting in the number of future failures a given section of drinking water network is likely to be subjected to in a given time horizon. The main output PI of the model is the future failure rate (number of future breaks per km) of each section of the network This PI can finally be also aggregated at the level of a category of pipes (e.g. of same material and diameter), or a sub network, or the whole network.

1.4 Functional description

PHM Model is based on the statistical survival analysis of the past failures dates (maintenance data over at least 5 years) observed for each section of the network (pipeline homogeneous in material, diameter, road location and installation date). These occurrences are probabilistically explained by a set of covariates either proper to the sections of the network or related to their environmental conditions, influencing proportionally the failure risk. The key analysis variable is the inter failure time, which distribution is modeled by a Weibull distribution function which depends on a scale parameter and a position parameter designed as a linear combination of the covariates. The analysis is stratified by material and number of observed previous failures. The parameter estimates are computed via the maximization of the log likelihood function of the observed inter failure times, including those right censored by the observation stopping date or the removal date of the sections. The number of future failures for each section is estimated by Monte Carlo simulations.

1.5 Brief Historical Overview of the Model

The underlying Stochastic Process Model of PHM Model belongs to the category of the Renewal Processes. More specifically, it derives from the so called by D. R. Cox & V. Isham (1980) “Time Dependent Renewal Process”. The first application to pipe breaks modeling, using Weibull inter-failure time distributions, was proposed by Andreou (1986). The model was then formalized and validated for French drinking water networks by the Ph D thesis work of P. Eisenbeis (1994). Further validations and improvements, especially in the case of short failure records, were thereafter achieved between 1995 and 2000 for other water and irrigation systems: Société du Canal de Provence, Service des Eaux de Charente Maritime, villes de Lausanne (Swiss), de Roubaix-Tourcoing and Dijon (in collaboration with Lyonnaise des Eaux for these two last).

2 THEORETICAL FRAMEWORK OVERVIEW/PAST STUDIES

2.1 Scientific background

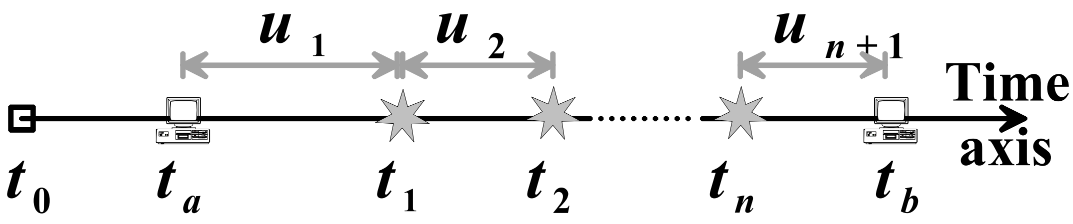


Figure 1 – Observational window, failure dates and inter-arrival times.

PHM Model is a statistical modeling tool derived from the Renewal Theory. This theory is designed to study recurrent events experienced by the components of a system. As it is shown on the figure 1, a given component installed at date t_0 is observed during the time interval $[t_a, t_b]$, called the observational window, within which it experiences n events occurring at dates t_1, t_2, …, t_n. The event dates t_1, t_2, …, t_n are considered as the realizations of the random variables T_r , where $r \in \{1,2,...,n\}$. This leads to define the inter-arrival times as the non-negative random variables $U_r = T_r - T_{r-1}$. Each U_r is assumed to have a two parameters Weibull distribution, denoted by $W(\mu_r, \sigma_r)$, with Probability Density Function (PDF), Cumulative Distribution Function (CDF), Survivor Function (SF), and Hazard Function (HF) respectively given by the following formulae:

$$f_r(u) = \lim_{h \to 0+} \frac{P\{u \le U_r < u+h \mid \mu_r, \sigma_r\}}{h} = \exp\left[-\exp\left(\frac{\ln u - \mu_r}{\sigma_r}\right)\right] \exp\left(\frac{\ln u - \mu_r}{\sigma_r}\right) / (u\sigma_r) \quad (1)$$

$$F_r(u) = P\{U_r \le u \mid \mu_r, \sigma_r\} = \int_0^u f_r(s)\,ds = 1 - \exp\left[-\exp\left(\frac{\ln u - \mu_r}{\sigma_r}\right)\right] \quad (2)$$

$$S_r(u) = P\{U_r > u \mid \mu_r, \sigma_r\} = \int_u^{+\infty} f_r(s)\,ds = \exp\left[-\exp\left(\frac{\ln u - \mu_r}{\sigma_r}\right)\right] \quad (3)$$

$$h_r(u) = \lim_{h \to 0+} \frac{P\{u \le U_r < u+h \mid U_r > u, \mu_r, \sigma_r\}}{h} = \frac{f_r(u)}{S_r(u)} = \exp\left(\frac{\ln u - \mu_r}{\sigma_r}\right) / (u\sigma_r) = \delta_r u^{\delta_r - 1} \exp(-\delta_r \mu_r) \quad (4)$$

where: $\delta_r = 1/\sigma_r$

The parameters μ_r and σ_r are positive, and respectively called the location parameter and the scale parameter, consistently with the above formulae (1) to (4).

As these parameters vary according to the rank r of the observed event, chaining the successive random inter-arrival times defines an Event Dependent Renewal Process (EDRP).

An important feature of this model consists in modulating the inter-arrival times according to a set of explanatory variables, also named covariates, either proper to the components, or characterizing their environment. The set of covariates values form a vector denoted by *x*. The covariates are assumed to be either continuous, or indicator variables (*i.e.* taking the value 0 or 1) of the levels of categorical explanatory variables.

In the Reliability Theory, the generic event of interest is called a failure, and *x* a stress vector.

The location parameter is assumed to be a linear combination of the covariates values:

$$\mu_r = x'\beta_r \tag{5}$$

where x' is the transpose of *x*.

The first component of *x* is defined to be a constant equal to 1, and the corresponding component of vector β_r is denoted β_{0r} and commonly called the intercept parameter.

When no covariate is taken into account, μ_r reduces to β_{0r}, which defines the Baseline Distribution.

2.2 Nature of the model

Whereas the nature of the model appears to be essentially statistical, it is able to account for stress factors, which effects on system components are to some extent deterministic. It is thus possible to consider the location parameter μ_r as being composed of a purely statistical part β_{0r} corrected by the deterministic effect of covariates. However, the estimates of the deterministic components of β_r are statistical, in the sense where their values depend on the sample of inter-arrival times observed on an actual system.

2.3 Underlying assumptions

As it will be explained below the model parameters are estimated by the maximum likelihood method. When performing such maximum likelihood estimation, three major assumptions are made:

- The events of a given rank are mutually independent ;
- The inter-arrival times are Weibull distributed ;
- The logarithm of the inter-arrival times depend linearly on the covariates.

The validity of the first assumption is considered to be assured as long as the vector *x* contains all main stress factors susceptible to produce events correlated in time or space. For example, some system components may show similar events patterns due to the same environmental condition which is assumed to be well described by a given level of a covariate ; introducing this covariate into the stress vector removes the correlation between residuals.

The Weibull assumption is easily checked graphically by plotting against the logarithm of the inter-arrival time the $\ln(-\ln(.))$ transform of the empirical Kaplan-Meyer estimate of the survivor function, and observing an at least rough linear relationship.

The covariate linear effect assumption do not pose problem in the case of categorical covariates. The validity of this assumption for continuous covariates can thus be easily asserted by splitting their range into classes and verifying the linear relationship between the β estimates and the central value of the corresponding classes.

2.4 Algorithm

The components of the vector β_r and the scalar σ_r are estimated by maximizing the likelihood function L_r, which is now to be defined for the events of rank *r* occurring on the *c* system components under observation. The notations used until now have thus to be enhanced by introducing a new index *i* referring to the system components: $i \in \{1,2,...,c\}$. The i^{th} system component has so been installed at date t_{i0}, has experienced n_i events recorded inside the observational window $[t_{ia}, t_{ib}]$, and is characterized by the stress vector x_i. This leads to the n_i+1 inter-arrival times u_{ir} with $r \in \{1,2,...,n_i+1\}$, the first n_i of which are not censored and hence characterized by the indicator variable $c_{ir}=0$, whereas the $(n_i+1)^{th}$ takes the value $u_{in_i+1}=t_{ib}-t_{in_i}$ and is associated with the censor variable $c_{in_i+1}=1$. If moreover L_r is to be computed for $r>n_i+1$, u_{ir} is conventionally considered as taking the value 0 (involving $S_r(u_{ir})=1$), and $c_{ir}=1$. The Weibull location parameter depends on the rank of the event and on the stress vector proper to the system component: $\mu_{ir}=x'_i\beta_r$, whereas the scale parameter σ_r depends only on the rank of the event. The likelihood function L_r is defined as the joint probability of the *c* observed inter-arrival times. The actual events contribute to L_r with the values taken by the PDF for the corresponding inter-arrival times, whereas the events censored by the stopping of observations, for the system components for which $n_i+1=r$, contribute with $P\{U_r>u_{in_i+1} \mid \mu_{ir}, \sigma_r\}$.

The likelihood function L_r can so be written as follows:

$$L_r=\prod_{i=1}^{c} f_r(u_{ir})^{1-c_{ir}} S_r(u_{ir})^{c_{ir}} \tag{6}$$

The maximization of the likelihood deals with the natural logarithm of L_r, rather than with L_r itself, and is usually performed using a Newton-Raphson algorithm. A rough initial guess of the solution can be obtained by ignoring the censoring information and solving the linear system $\ln u_{ir}=x'_i\beta_r+\varepsilon_{ir}$ by the classical GLM (General Linear Model) technique (the first guess of σ_r is simply the estimate of the standard deviation of ε_{ir}).

3 SPECIFICATIONS OF COMPUTATIONAL STEPS

3.1 Functional Scheme

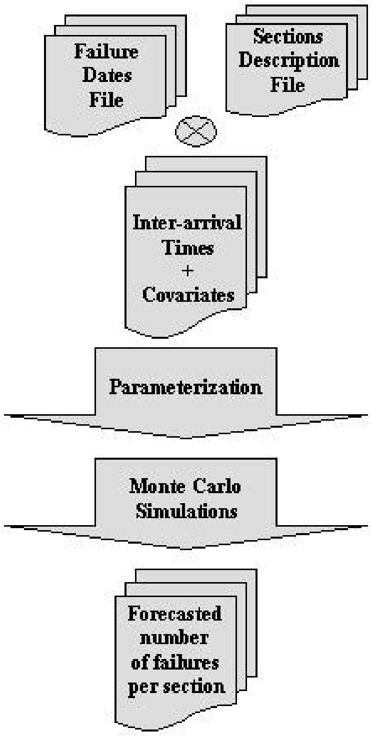


Figure 2 – Main Computational Steps

3.2 Raw Data Formatting

Up to now the PHM Model has been described in rather general terms, such as system, system components and events. These generic notions are now to be transposed to the technical context of pipe failures forecasting in drinking water networks.

What was previously called a system is now a drinking water supply system.

The events under consideration are pipe bursts leading to repair operations recorded in the digital database of the maintenance service. Such computerized records are rarely available for a statistical study since the date the water system began to be installed, but rather since the years 1980 or 1990. That is why an observational window has to be considered in stead of the whole operational life of the network components.

The definition of the system components depends on the nature of the maintenance records, which are still seldom linked to a GIS, but rather mainly refer to the notion of homogeneous network section. The term section shall thus be used in the sequel and refers to a pipeline serving, between 2 valves (operating homogeneity), the totality or a part of a road (spatial and hence environmental homogeneity) and constituted by pipes made of the same material, in the same diameter (pipe category homogeneity) and installed at the same date (age homogeneity).

Raw data generally consist in two databases:

- Description of the sections,

- Dates of failures.

Both files contain some common variables that serve for section identification, and make it possible to merge the two databases.

The description of the sections is generally issued from the hydraulic modeling database and/or selected GIS databases. In order to lead to operational results, the description of the network must be exhaustive. The structure of this file is illustrated at table 2 below.

The failures dates file comes from the Maintenance Service database. Its structure is illustrated at table 1 below. It must concern all sections of the network or a representative sample, provided that, in this case, the list of the sections having experienced not any failure is available. The date unit is generally the day and it is practical that the software used for the modeling task be able to represent calendar dates, formatted *e.g.* as “dd/mm/yyyy”, as a number of days elapsed since a given date (*e.g.* 01/01/1960 for the SAS System).

The files of failure dates and sections description are firstly sorted out by alpha-numerical value(s) of the section identification variable(s) and failure date (for the first one). These datasets are then horizontally merged in order that failure date be accompanied with the covariate values related to the concerned section.

The main formatting operation consists lastly in transforming the list of dates into a list of inter-arrival times accompanied with their censoring variable. This operation is illustrated at table 3 below. As already explained above, the $n_i - 1$ failures recorded for the i^{th} section give rise by difference to $n_i + 1$ inter-arrival times, the last of which is censored whereas the first ones are not. If no failure is recorded for a section (it is fortunately the most frequent case) a single censored inter-arrival time is generated. It s worth noticing that the bounds t_a and t_b of the observational window are the same for all sections except those:

- laid out at $t_{i0} > t_a$,
- or replaced at $t_{in_i} < t_b$.

For computing the inter-arrival times, the starting and stopping dates t_{ia} and t_{ib} are then defined by:

$$t_{ia} = Max(t_{i0}, t_a) \text{ and } t_{ib} = Min(t_{in_i}, t_b) .$$

Id.	**Fail.**
…	**…**
***i*-1**	***t*(*i*-1,*n*$_{i-1}$)**
i	***t*(*i*,1)**
…	**…**
i	***t*(*i*,*n*$_i$)**
***i*+1**	***t*(*i*+1,1)**
…	**…**

Table 1 – Typical Format of Failure Dates File

Id.=Section Identification Variable – Fail.=Failure Date

Id.	t_a	t_b	Instal.	Mat.	Diam.	Length	X
…	…	…	…	…	…	…	…
i-1	$t_a(i$-1)	$t_b(i$-1)	$t(i$-1,$0)$	M(i-1)	D(i-1)	$L(i$-1)	$X(i$-1)
i	$t_a(i)$	$t_b(i)$	$t(i,0)$	M(i)	D(i)	$L(i)$	$X(i)$
i+1	$t_a(i$+1)	$t_b(i$+1)	$t(i$+1,$0)$	M(i+1)	D(i+1)	$L(i$+1)	$X(i$+1)
…	…	…	…	…	…	…	…

Table 2 – Typical Format of Sections Description File

Id.=Section Identification Variable – t_a and t_b =Bounds of the Observational Window

Instal.=Section Installation Date – Mat.=Pipeline material – Diam.=Pipes diameter

Length=Section Length – X=One Generic Covariate

Id.	U	C	R	Age_1	Mat.	Diam.	Length	X
…	…	…		…	…	…	…	…
i-1	$t(i$-1,n_{i-1}) -$t(i$-1,n_{i-1}-1)	0	n_{i-1}	$t(i$-1,n_{i-1}-1) -$t(i$-1,0)	M(i-1)	D(i-1)	$L(i$-1)	$X(i$-1)
i-1	$t_b(i$-1) -$t(i$-1,n_i)	1	n_{i-1}+1	$t(i$-1,n_{i-1}) -$t(i$-1,0)	M(i-1)	D(i-1)	$L(i$-1)	$X(i$-1)
i	$t(i$,1) -$t_a(i)$	0	1	$t_a(i)$ -$t(i$,0)	M(i)	D(i)	$L(i)$	$X(i)$
…	…	…	…	…	…	…	…	…
i	$t(i,n_i)$ -$t(i,n_i$-1)	0	n_i	$t(i,n_i$-1) -$t(i$,0)	M(i)	D(i)	$L(i)$	$X(i)$
i	$t_b(i)$ -$t(i,n_i)$	1	n_i+1	$t(i,n_i)$ -$t(i$,0)	M(i)	D(i)	$L(i)$	$X(i)$
i+1	$t(i$+1,1) -$t_a(i$+1)	0	1	$t_a(i$+1) -$t(i$+1,0)	M(i+1)	D(i+1)	$L(i$+1)	$X(i$+1)
…	…	…		…	…	…	…	…

Table 3 – Merging by Id. Tables 1 and 2

U=Inter-arrival time – C=Censoring variable – R=Failure rank

Age_1=Age of the section at the preceding failure date

3.3 List and Definition of Explanatory Factors

3.3.1 Required factors

Among the numerous covariates that can be introduced into the stress vector, two appear to be indispensable:

- The internal pipe diameter,
- The length of the section.

The inter-arrival time generally shows to decrease with the pipe diameter, which can be used as a either continuous or categorized covariate. When the diameter is small a pipe tends to mechanically behave like a beam, whereas the mechanical behavior of a large diameter pipe tends to that of a ring. The ratio of the diameter to the pipe unit length can alternatively be used as a covariate.

The inter-arrival time generally tends to increase like the square root of the section length. This fact has often be observed without being yet well understood ; one possible explanation is that short sections are preferentially encountered in town centers, where human presence and activities are more concentrated, and thus environmental stresses are likely to be at a higher level than in less dense peripheral areas.

3.3.2 Possibly Useful factors

The following covariates may sometimes significantly improve the model fit but are more seldom available:

- installation date,
- hydraulic conditions,
- connections density,
- soil conditions,
- level of the ground water table,
- road traffic level and quality (especially trucks, buses, tramway *etc.*),
- pipe location under sidewalk or roadway.

The knowledge of the section installation date makes it possible to compute the logarithm of the age of the section at the previous failure date $\ln(t_{ir-1} - t_{i0})$,. which may partially compensate for the lack of failure information before t_{ia}. The knowledge of the section installation date may also help to infer some technical pieces of information like:

- distinction between different qualities of the same material due to well dated technological innovation (*e.g.* replacement of molded cast iron by centrifuged cast iron in the thirties, of gray cast iron by ductile cast iron in the sixties, or of stuck PVC by sealed PVC in the early seventies) ;
- similar distinction between different qualities of laying bed or embankment.

Sometimes missing information about soil quality or level of water table can be usefully replaced by indicator variables of geographical zones which are assumed to be almost homogeneous with respect to these characteristics.

It is worth mentioning that the model can only be calibrated, and then used for forecasting purpose, for the sole sections completely documented with respect to the entire set of covariates. When a covariate value lack for a notable proportion of the sections (*i.e.* more than 10 or 15 %) it is better to neglect it than parameterizing the model on the sole documented sections and reducing thus drastically the sample size.

3.4 Model Parameters Estimation or Assignation

Practically, the calibration operations begin by splitting the whole maintenance dataset into strata. This means that all failure data corresponding to a given stratum are processed together to maximize the same likelihood function. The stratification is made by crossing the two following variables:

- The pipe material,
- The rank of the observed failure.

The first level of stratification is motivated by the huge differences in the aging process observed among the usual pipe materials, due to their variety of mechanical and chemical resistance to environmental stresses ; for example, the β parameters corresponding to traffic load or soil corrosivity are a priori expected to take quite different values for concrete, cast iron or PVC pipes.

The second level of stratification directly results from the EDRP theory presented above. However, statistical reasons of sample size lead often to consider two main categories of failure ranks:

- The first for the random variable U_1, i.e. corresponding to the time elapsed between the beginning of the observational window and the first observed failure date,
- The second for the random variables U_r with $r \geq 2$.

In the second category a single scale parameter value is estimated: $\forall r \geq 2, \sigma_r = \sigma_2$. However different values may be obtained for the location parameter by introducing the natural logarithm of the failure rank $\ln r$ into the stress vector ; It is worth mentioning that the Wald test generally reveals this covariate as being highly significant.

3.5 Output

Describe precisely the different outputs of the model

The major output of the model is provided by MCS forecasting step and consists in a portray of the distribution of the number of future failures each section is likely to be subjected to.

A less important but possibly technically interesting output is the list of the significant covariates which may lead to useful reflections concerning the technical management of the network.

3.6 Model Validation

(Statistical model)

3.6.1 Check of Parameters Significance – Internal Validation (Statistical model)

An important property of the maximum likelihood estimates is their asymptotic normality. Let l_r denote the natural logarithm of the likelihood function L_r defined in (6): $l_r = \ln L_r$. Let also θ_r be the vector of parameters composed of σ_r and the β_{jr}'s. If there are *p* covariates, then: $\theta_r{}'=(\theta_{jr})=(\sigma_r \;\; \beta_{0r} \;\; \beta_{1r}...\beta_{pr}),\; j\in\{1,...,p+2\}$. The Hessian matrix H_r of the maximization problem is thus: $H_r=(H_{jkr})$ with: $H_{jkr}=\dfrac{\partial^2 l_r}{\partial\theta_{jr}\partial\theta_{kr}}\;\; j,k\in\{1,...,p+2\}$. The covariance matrix V_r of the θ_r components is estimated by: $\hat{V}_r=(\hat{V}_{jkr})=(-H_r)^{-1}$. Hence, for a sufficiently large number of observed events of rank *r*, the estimate of θ_{jr} has a gaussian distribution $N(\hat{\theta}_{jr},\hat{V}_{jjr})$. This theoretical point makes it possible to test the null hypothesis $\beta_{kr}=0$, or equivalently $\theta_{jr}=0$ with $j=k+2$, which means that the corresponding covariate do not have any influence on the inter-arrival time U_r. The usual test, called Wald test, is based on the Chi-Square distribution of a squared gaussian $N(0,1)$ random variable, and consists in calculating the statistic $K=\hat{\theta}_{jr}^2/\hat{V}_{jjr}$ and then the probability that a Chi-Square distributed random variable with 1 degree of freedom exceeds this statistic: $P\{\chi_1^2>K\}$; this probability of exceedence is often called p-value. The risk of wrongly rejecting the null hypothesis is commonly considered as sufficiently low when the p-value is less than 0.05, and the effect of the covariate is thus ascertained ; otherwise, the null hypothesis is accepted and the covariate has to be removed from the model.

3.6.2 External Validation or Cross Validation – Statistical test(s)

Once the model has been calibrated, by finding optimal estimates of the parameters, its forecasting ability has to be validated.

The validation is performed by artificially stopping the series of observations at date t_v so as $t_v<t_b$. A new calibration of the model is then performed, by using the sole event dates $t_{ir}\leq t_v$. The number of events which are likely to occur in $[t_v,t_b]$ is computed by a Monte Carlo simulation technique to be described thereafter.

The expected and actually observed numbers of events in $[t_v,t_b]$ are finally compared by building a contingency table which rows and columns are defined as follows:

- The system components are ranked in descending order according to their expected numbers of events and grouped in quantiles, i.e. classes of nearly equal frequencies, which constitute the rows of the table, and are called risk quantiles ;
- The table contains two columns, one for the expected numbers of failures, and the other for the observed ones ;
- The cell frequencies are the expected and observed total numbers of events per risk quantile.

The model is finally validated by examining how well expected and observed numbers of events match across the risk quantiles ; this examination can be helped by graphical transformation of the contingency table in histograms. Unfortunately, no inferential test has been yet proposed in the statistical literature to quantify the departure between expected and observed events frequencies.

After the validation phase and the definitive calibration on the complete time interval $[t_a,t_b]$, the model is run to forecast the distribution of the number of future events that are likely to occur on each system component in a given time horizon $]t_b,t_h]$.

Theoretically, a rigorous computation should use the convolution of the distributions $W(\mu_i, \sigma_i)$, the calculation of which seems tremendous to achieve. A Monte Carlo simulation method (MCS) is thus proposed to circumvent this difficulty.

MCS is an iterative method consisting in performing successively 1000 elementary random simulations (or more according to the available computation power). One elementary simulation consists in computing the following steps, for each system component successively:

- Choose a first random number s_1 uniformly distributed in $[0,1]$ and transform it in a random $W(\mu_{n+1}, \sigma_{n+1})$ distributed inter-arrival time w_1 by applying the formula $w_1 = \exp[\sigma_{n+1} \ln(-\ln r_1) + \mu_{n+1}]$ where $r_1 = s_1 \exp\left[-\exp\left(\frac{\ln u_{n+1} - \mu_{n+1}}{\sigma_{n+1}}\right)\right]$ (the intermediate calculation of r_1 is justified by the absence of observed event in $[t_n, t_b]$ which involves the use of a conditional survivor function after t_b) ;
- Add the inter-arrival time w_1 to t_b and obtain $t_{n+1} = t_b + w_1$;
- If $t_{n+1} < t_h$, choose a second random number s_2 uniformly distributed in $[0,1]$ and transform it in a random $W(\mu_{n+2}, \sigma_{n+2})$ distributed inter-arrival time w_2 by applying the formula $w_2 = \exp[\sigma_{n+2} \ln(-\ln r_2) + \mu_{n+2}]$, else stop ;
- Add the inter-arrival time w_2 to t_{n+1} and obtain $t_{n+2} = t_{n+1} + w_2$;
- Repeat the two preceding steps until t_{n+k} overpasses t_h ;
- Record finally the number *k*-1 of possible future events for the considered system component.

After this iterative process has been completed, one disposes of 1000 possible numbers of future events for each system component. It makes it then possible to portray the distribution of the random number of future events for each system component, and more synthetically for homogeneous groups of components or for the whole system.

3.7 PI(s) Forecasting Method

As already mentioned in the MCS method description, the raw output of this modeling step is constituted by a list of 1000 (or more) possible numbers of failure in the time interval $[t_b, t_h]$ for each section. From the end-user point of view, the Performance Indicator (PI) that summarizes these results in the most interesting way is the estimate of the expected number of failures per time and length units, generally expressed in km and year. This ratio is simply computed as the arithmetic mean of the above mentioned 1000 possible numbers of failure, divided by the length of the section in km and by the duration $t_h - t_b$ in years. More synthetic PI's can be computed for a part or the totality of the network as the arithmetic mean of the PI's of the concerned sections weighted by their length.

4 POSSIBLE IMPROVEMENTS OF THE MODEL

One interesting problem, that has not yet been addressed neither in the statistical literature nor in the practical Cemagref studies, consists in building a tool to measure the usefulness of a given covariate, from the point of view of the forecasting ability of the model.

It is here proposed to build first an index to characterize the efficiency of the model to forecast the failure risk. In the sequel, this index is called the Failure Risk Forecasting Efficiency (FRFE) and denoted Φ, and can be computed in the model validation phase previously described. The model validation phase can then be performed twice:

- a first time with the complete model, *i.e.* using all covariates found as being significant in the calibration phase, including the covariate which contribution to the forecasting ability of the model is to be assessed,
- and a second time with the reduced model, *i.e.* deprived of the given covariate.

If the FRFE obtained with the complete and the reduced model are respectively denoted Φ_+ and Φ_-, the difference $\Phi_+ - \Phi_-$ measures the contribution of the given covariate to the forecasting ability of the model.

It is proposed to compute the FRFE as follows. Let first the random variable N_i stand for the number of failures the i^{th} section may be subjected to in the time interval $[t_v, t_{ib}]$. The number of failures actually observed in this interval is denoted by $\tilde{N}_i$, and the expected value (*i.e.* forecasted by the model) by $\hat{N}_i$. The *c* sections are then ranked in three ways:

- the first ranking consists in sorting out the sections by descending values of expected numbers of failures $\hat{N}_i$ and the resulting ranks are denoted by $\hat{R}_i$; this means that:

  $$\hat{N}_k = \underset{i=1\ldots c}{Max}\left(\hat{N}_i\right) \Rightarrow \hat{R}_k = c \text{ and } \hat{N}_k = \underset{i=1\ldots c}{Min}\left(\hat{N}_i\right) \Rightarrow \hat{R}_k = 1 \text{ ;}$$

- the second ranking consists in sorting out the sections by descending values of observed numbers of failures $\tilde{N}_i$ and the resulting ranks are denoted by $\tilde{R}_i^+$; this means that:

  $$\tilde{N}_k = \underset{i=1\ldots c}{Max}\left(\tilde{N}_i\right) \Rightarrow \tilde{R}_k^+ = c \text{ and } \tilde{N}_k = \underset{i=1\ldots c}{Min}\left(\tilde{N}_i\right) \Rightarrow \tilde{R}_k^+ = 1 \text{ .}$$

- the third ranking consists in sorting out the sections by ascending values of observed numbers of failures $\tilde{N}_i$ and the resulting ranks are denoted by $\tilde{R}_i^-$; this means that:

  $$\tilde{N}_k = \underset{i=1\ldots c}{Max}\left(\tilde{N}_i\right) \Rightarrow \tilde{R}_k^- = 1 \text{ and } \tilde{N}_k = \underset{i=1\ldots c}{Min}\left(\tilde{N}_i\right) \Rightarrow \tilde{R}_k^- = c \text{ .}$$

The quantity $\sum_{i=1}^{c} \hat{R}_i \tilde{N}_i$ can take any integral value between $\sum_{i=1}^{c} \tilde{R}_k^+ \tilde{N}_i$ and $\sum_{i=1}^{c} \tilde{R}_k^- \tilde{N}_i$.

It is then proposed to define the FRFE as:

$$\Phi = \frac{\sum_{i=1}^{c} \hat{R}_i \tilde{N}_i - \sum_{i=1}^{c} \tilde{R}_k^- \tilde{N}_i}{\sum_{i=1}^{c} \tilde{R}_k^+ \tilde{N}_i - \sum_{i=1}^{c} \tilde{R}_k^- \tilde{N}_i} \qquad (7)$$

The FRFE index Φ has the property: $\Phi \in [0,1]$. If the model produced a perfect forecast, the ranks $\hat{R}_i$ and $\tilde{R}_i^+$ should be equal for all sections, and thus $\Phi = 1$, which means that the perfect model has a forecasting efficiency of 100 %. It is important to notice that FRFE does not measure the exactness of the forecasted numbers of failures, but rather the ability to correctly rank the sections according to their actual risk of failure.

It remains to carry out the theoretical investigation of the distribution of the FRFE considered as a random variable Φ_0 under the null hypothesis H_0 of independence between $\hat{R}_i$ and $\widetilde{R}_i^+$. This would then make it possible to compute the risk $P\{\Phi_0 > \Phi \mid H_0\}$ to reject wrongly the null hypothesis when asserting the forecasting efficiency of the model.

5 SOFTWARE SPECIFICATIONS

Length not limited

5.1 Programming Language(s) or Mathematical-Statistical Software(s)

All above described modeling steps are carried out under SAS System. The version presently used is the 6.11 in Microsoft Windows 98 environment. The SAS System provides all language and macro-language statements, functions, and procedures that are needed to manage, format and process the raw data, estimate the model parameters and carry out the forecasts.

5.2 Possible Input File(s) Formats

Up to now all the studies performed by Cemagref/ORH necessitated input files in either ASCII (.txt) or Excel (.xls) format. It is possible to develop programs accepting various other input formats but this would necessitate to contract with the SAS Institute (Software provider) an enhanced license including the SAS/Access module.

5.3 Possible Output File(s) Formats

The same comment as the one made above for input format holds for output files.

6 REFERENCES

6.1 Theoretical Framework References

Cox, D.R., & Isham, V. (1980). Point Processes, Chapman and Hall, London, pp. 188.

Eisenbeis, P. (1994) « Modélisation statistique de la prévision des défaillances sur les conduites d'eau potable ». *Thèse*, Spécialité Sciences de l'eau, Université Louis Pasteur, Strasbourg, France, 156 p.

Kalbfleisch, J.D., & Prentice, R.L. (1980). The statistical analysis of failure time data. John Wiley & Sons.

6.2 Practical Use and Results References

SAS Institute Inc. (1989). SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 2, Cary, NC: SAS Institute Inc., pp. 997-1025.

SAS Institute Inc. (1990). SAS® Procedures Guide, Version 6, Third Edition, Cary, NC: SAS Institute Inc., pp. 483-502.

Eisenbeis, P., & Le Gat, Y. (2000), « Using Maintenance Records to Forecast Failures in Water Networks », Urban Water.

APPENDIX 4 : UTILNETS

1 GENERAL DESCRIPTION

1.1 Name and/or Acronym of the Model

UTILNETS - A decision support system for water mains rehabilitation

1.2 Company/Research Center/University

The project is funded by the EC DG XIII, Innovation Programme

Project Coordinator during the Upswing phase: North West Water (UK)

Partners: ACEA S.p.A. (I)

Computer Technology Institute (GR)

SINTEF (NO)

TECNIC S.p.A. (I)

Trondheim Kommune (NO)

UBIS (DE)

1.3 Objectives

UtilNets is a prototype predictive decision support system for the rehabilitation of water distribution pipes. It determines the prospective life expectancy of pipe segments and supports the prioritisation of rehabilitation measures. The possible failure causes are represented by safety factors for each pipe segment. UtilNets analyses all important environmental influences and loads that have affected or will affect the individual pipe during its whole lifetime. The deterministic-probabilistic approach of UtilNets requires a huge amount of physical and environmental input parameters. For the rehabilitation order, further information about the importance of a pipe is needed. UtilNets has strong import facilities and a GIS interface. The application of UtilNets is menu driven with a Multi Document Interface (MDI).

1.4 Functional description

For the prediction of pipe life and the calculation of pipe deterioration, UtilNets takes each factor into account influencing this process. It applies physical models based on engineering well known equations for the calculation of structural pipe condition. Loads on and resistance of the pipe are derived from input data, which must be given for each individual pipe. Each pipe can be further divided by segments if changes in load make it necessary. The tool has been designed for the failure prediction of cast iron pipes. The deterioration process is driven by corrosion, which depends on soil conditions and water type on one hand and pipe material and protection on the other hand. The UtilNets tool consists of several subtools handling data import, data editing, calculation processing and result presentation. The user is able to keep an overview over input data and intermediate calculation results all along the application.

1.5 Brief Historical Overview of the Model

UtilNets has started in 1994 as a project under the Brite/Euram Programme, funded by the European Union. The model has been developed by a consortium of structural and reliability engineers, as well as IT experts in GIS, relational databases and expert systems, assisted by a large water utility. The former consortium consisted of CTI (Greece), TECNIC S.p.A. (Italy), UBIS (Germany) and NWW Ltd. (UK). Associated partner was PDL (UK) and subcontractor was SEPTE Ltd. and the University of Thessaloniki (Greece).

During the first three years of tool development, the complete model has been established, containing the analysis modules structural, hydraulic and water quality reliability and the optimisation modules for rehabilitation costs, consequences of failure and rehabilitation prioritisation. These modules are further described in chapter 2.1.

This first version was based on the database of North West Water, the only user at this time. Most of the necessary calculation data was hardcoded. The results of the test evaluation were indifferent. It has been found that many calculation procedures where insensitive to the results and some input parameters were redundant.

In a second phase, emphasis has been laid on a strong import facility to make UtilNets applicable for any utility. Further on, the tool database has been explored regarding redundant variables and calculation procedures. Since UtilNets had no interface towards an hydraulic network calculator, the hydraulic module has been comprehensively reduced to a few rules and the water quality module has been removed. Also the integration of another module that had been developed independendly from UtilNets during the first phase, network reliability, has been postponed.

SINTEF has joined the group during the second phase to contribute with its experience in water network rehabilitation. To improve user-friendliness, two further utilities have joined the consortium during this phase, ACEA in Rome and the municipality of Trondheim.

The project has been officially ended in January 2001. There were extended problems of data-technical nature to overcome during the last two years. A re-implementation became necessary due to the loss of support by Intellicorp for PowerModel that has been used for UtilNets. The new import modules had to be cleaned to assure a smooth application. The tool development has so far been finished. Yet, there is still some work to be done regarding the refinement of default rules and values in general and for the single test cases specifically. Only when the adaptation of values that were missing in the imported data, cannot any longer improve the calibration results, a trustworthy statement can be made on UtilNets.

2 THEORETICAL FRAMEWORK OVERVIEW/PAST STUDIES

2.1 Scientific background

The following paper has been published at the NODIG conference in Orlando, USA, in 1999. It represents the status reached in the UtilNets project after the first phase, finished in 1997. The scientific background remained the same since then, but changes to modules and extensions have been made.

Erreur ! Liaison incorrecte.

2.2 Nature of the model

Model approach

A common understanding for the form of a survival function for a water main is represented by the so-called "bath tub" curve as shown in Figure 1. The higher failure probability in the beginning is due to initial factors as manufacturing faults or bad workmanship on the construction site. During a long life-span only occasional failures appear. The length of this period depends on many factors and can differ widely even for the same pipe material and strength. At the right hand side of the "bath tub" curve the probability for failure starts to rise more or less significantly. Again, there are lot of factors influencing the rise and shape of this part of the curve. The factors leading to a pipe break must not necessarily be the same ones, which contributed to a continuous weakening of the pipe. Therefore, a statistical model can hardly be applied on the single pipe. In the best case it is valid for a certain group of pipes.

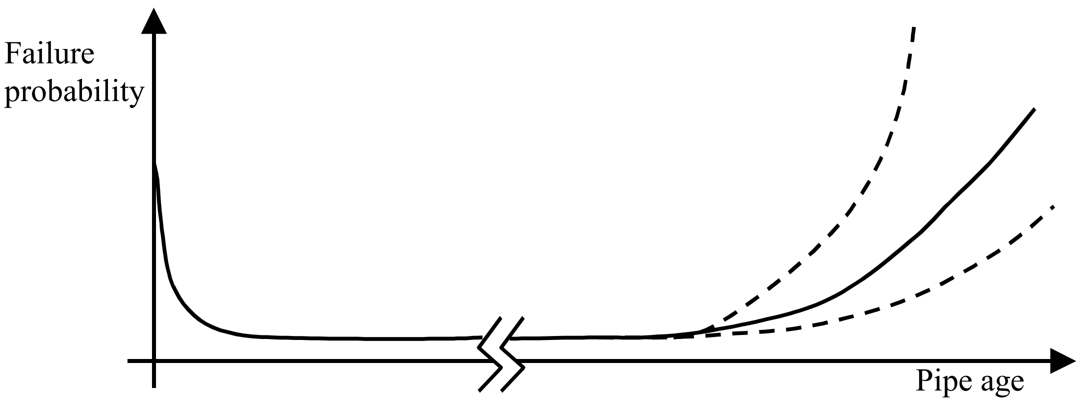


Fig. 1 Illustration of the typical failure course for a water main, the so-called "bath tub" curve

For the prediction of pipe life and the calculation of pipe deterioration, UtilNets takes each factor into account influencing this process. It applies physical models based on engineering well known equations for the calculation of structural pipe condition. Loads on and resistance of the pipe are derived from input data, which must be given for each individual pipe. Each pipe can be further divided by segments if changes in load make it necessary. Figure 2 gives an overview over the most important loads influencing pipe life.

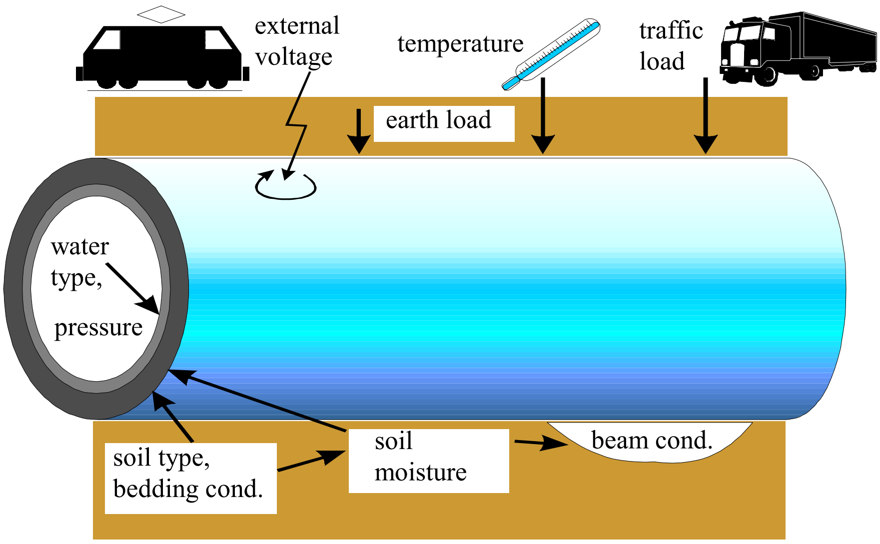


Fig. 2 Parameters influencing pipe life used by UtilNets

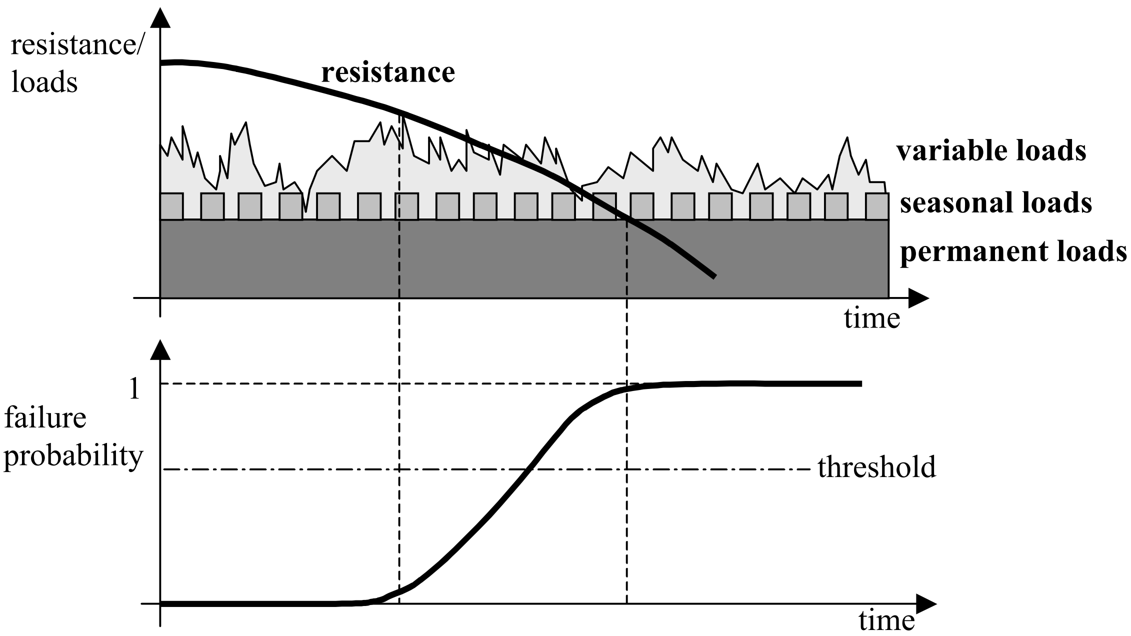


Fig. 3 Structural analysis for pipe failure probability in UtilNets

The counterpart of the loads is the pipe resistance, which is weakened over time until it reaches the level of loads, as shown in Figure 3. The loads can be permanent, e.g. earth load and water load, seasonal, e.g. frost load and temperature stress, or variable as traffic load. Variable loads are calculated in a probabilistic process.

UtilNets has been designed for the failure prediction of cast iron pipes. The deterioration process is driven by corrosion, which depends on soil conditions and water type on one hand and pipe material and protection on the other hand. The corrosion process is calculated by equation (1):

$$d = a \cdot t^{n} \qquad (1)$$

with d: max corrosion pit [mm]

a,n: corrosion coefficients

t: time in service

UtilNets has implemented default values for the external and internal corrosion coefficients. The user can choose from a variety of soil types and water types, but can as well adapt or define values of his own if he has data available.

When the pipe structure is weakened so far that the resistance gets as low as the maximal possible loads, the failure probability curve starts to rise. The user can set a threshold for the failure probability, which is giving the year of failure. Additionally, the user can define the importance of each single pipe regarding supply and consequences of failure. Together with an annual budget information UtilNets proposes a succession of pipe rehabilitation. Depending on pipe material, corrosion, diameter, pressure and importance, different methods of rehabilitation, e.g. PE-lining or replacement, are suggested.

For each segment UtilNets distinguishes the reason for failure which is represented by a calculated safety factor. For the reliability of results a table is provided which gives a confidential factor for each link. This factor depends on the amount of input data provided by the user and not filled by the default manager. The different input parameters are weighted, depending on their importance. A trace window offers further possibilities to track the calculations. With all these features, UtilNets is not just a black box, but offers several control options especially for an experienced user.

2.3 Underlying assumptions

Since UtilNets is a physical model, the only assumption that has to be made is the reliability of input data. This can only be taken granted to a certain extend. General definitions and overall assumptions have to be made for many parameters where homogeneousness can not be assured, like in soil conditions. There might additionally exist some loads and degradation processes, like a special type of corrosion, which are not taken into acount by UtilNets.

2.4 Algorithm

The algorithm back UtilNets can be split up into the several modules. Since the hydraulic, water quality and reliability module is not longer implemented, the following documents contain only the description of the remaining modules.

Module 1 contains the description of the calculation of structural reliability:

Module 5,6 and 7 contain the optimisation modules for the prioritisation of rehabilitation:

2.5 Past studies and conclusions

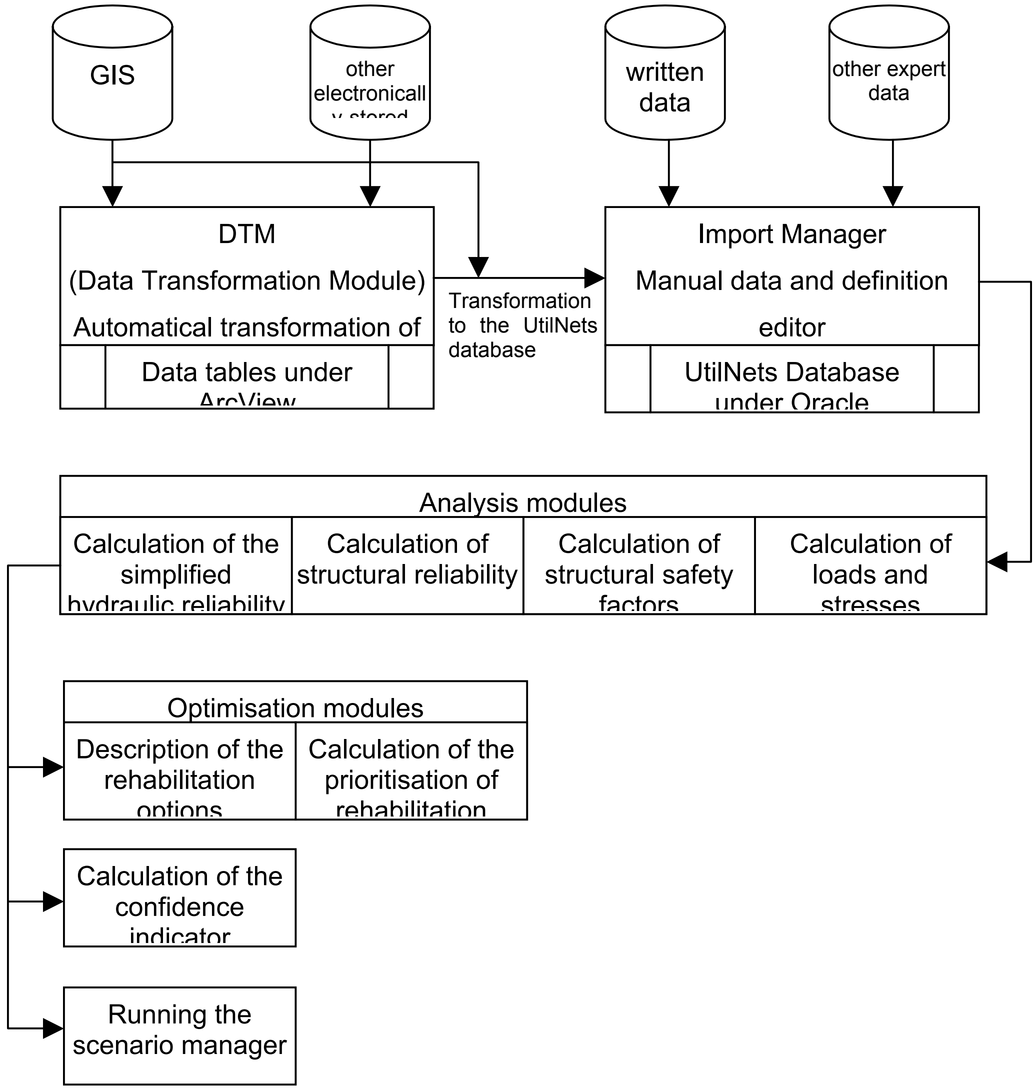
North West Water has written a comment on the first version of UtilNets in 1997. It includes user experiences and conclusions from a first application. These intermediate test results are not valid for the current version but give an impression for the development of the tool.

After finalising the second phase (Upswing) in the UtilNets project a final report is going to be delivered to the EU, which will be opened to public after its approval. The test evaluation from SINTEF for the Trondheim municipality is given in advance.

The tool development of UtilNets is not finished yet. The major outstanding work is the refinement of default rules and values in general and for each single test case specifically. A final conclusion can therefore not be drawn now.

3 SPECIFICATIONS OF COMPUTATIONAL STEPS

3.1 Functional Scheme



Raw Data Formatting

One of the main emphasisis for the development of UtilNets during the last years has been laid on the Import of data. Two separated tools have been developed, handling electronically available mass data as well as oral information. There are preliminary user manuals available for the handling with Import Data. These manuals will be available after acceptance of the final report by the EU comission. The following document contains a description of the import facilities, taken from the 18-month periodic report. Some changes had been made since then. The document contains as well the hierarchical data structure of all parameters used by UtilNets.


UtilnetRef4.pdf

3.2 List and Definition of Explanatory Factors

3.2.1 Required factors

None of the parameters in UtilNets is actually required. Some basic informations must be given to make an application reasonable. For most of the data fields suggests UtilNets some default values. The user can addionally define its own rules for the filling with fields. If for example the material, which is a basic information, is unknown and the conatruction year is known, a rule helps, which says that pipes laid between year x and y are most probable made of material z. The more information the user can give, the more reliable are the results. A confidence indicator calculated by UtilNets represents the amount and importance of data that is given by the user.

One differentiation can be made between data used for the calculation of failure and data used for the prioritisation of rehabilitation. The following list gives an overview over all parameters that can be imported into UtilNets:

Variable	Unit	Short description
Link_ID	[-]	Code of single pipe
A_End	[-]	Code of node A connected to a Link ID
B_End	[-]	Code of node B connected to a Link ID
Length	[m]	Length of Link ID
Max_Flow_Capacity	[l/s]	
People	[-]	number of people supplied be the Link
DMA	[-]	Code of district meter area the link belongs to
Supply_Zone	[-]	Code of supply zone the link belongs to
Compliance_Zone	[-]	Code of compliance zone the link belongs to
Is_Made_Of	[-]	Code of pipe material
Supply	[-]	Category of importance for link regarding supply
Danger	[-]	Category of potential danger in case of pipe burst
Damage	[-]	Category of potential damage in case of pipe burst
Nominal_Diameter	[mm]	
Orig_Int_Diameter	[mm]	Original internal diameter
Orig_Ext_Diameter	[mm]	Original external diameter
Orig_Wall_Thickness	[mm]	Original pipe wall thickness
Year_Laid	[-]	
Trench_Width	[m]	Width of the trench
Burst_Rate	[times/yr]	Estimated failure rate of the single link
Internal_Protection_Date	[-]	Year when link was protected internal
External_Protection_Date	[-]	Year when link was protected external
Bedding_Description	[-]	Type of soil the pipe is bedded on
Backfill_Description	[-]	Type of soil the trench is filled with
Joint_Type	[-]	Type of connection between two links (rigid, flexible,…)
Internal_Lining	[-]	Type of internal lining
External_Lining	[-]	Type of external lining
Has_Subsistance	[-]	Is the pipe likely fully supported? [yes/no]
Working_Pressure	[N/m*m]	Average working pressure
Surge_Pressure	[N/m*m]	Expected pressure in surge conditions
Surge_Pressure_Occur_Rate	[times/year]	Expected number of surge conditions per year
Water_Source_type	[-]	Name of water type defined in an extra table
Hydraulic_Failure	[-]	Has the pipe likely a leakage? [yes/no]
Cathodic_Protection	[-]	Has the pipe a cathodic protection? [yes/no]
Other_Utilities	[-]	Are other utilities affected when working on this pipe? {yes/no]
Segment_ID	[-]	Code of a segment (part of a link, identically with Link ID when a link consists only of one segment)
Depth_At_Crown	[m]	Depth of the segment (Depth of the link)
Length	[m]	Length of the segment (part of a link)
Truck_Load	[-]	Refers to the Road Class (expected traffic load)
Temperature_Zone	[-]	Code of temperature zone defined in an extra table
Pavement_Condition	[-]	Type of pavement (for traffic load distribution)
Soil_Class_Zone	[-]	Type of surrounding soil, defined in an extra table
Node_ID	[-]	Code of Node between to links
Depth	[m]	Depth of node

Further tables, containing specific parameters and values for the calculation of loads and resistance, partly filled with default values:	Soil type parameters
	Water type parameters
	Pipe Material specifications
	Rehabilitation methods and costs
	Air temperature and frost parameters
	Customer specifications
	Road class specifications

3.2.2 Highly recommended factors

The following data can be seen as important for gaining a reliable result when running UtilNets:

Variable	Unit	Short description
Link_ID	[-]	Code of single pipe
Is_Made_Of	[-]	Code of pipe material
Nominal_Diameter	[mm]	
Orig_Wall_Thickness	[mm]	Original pipe wall thickness
Year_Laid	[-]	
Internal_Lining	[-]	Type of internal lining
External_Lining	[-]	Type of external lining
Has_Subsistance	[-]	Is the pipe likely fully supported? [yes/no]
Water_Source_type	[-]	Name of water type defined in an extra table
Depth_At_Crown	[m]	Depth of the segment (Depth of the link)
Soil_Class_Zone	[-]	Type of surrounding soil, defined in an extra table
Working_Pressure	[N/m*m]	Average working pressure
Surge_Pressure	[N/m*m]	Expected pressure in surge conditions
Tables		
Soil type parameters		
Water type parameters		
Pipe Material specifications		

3.2.3 Possibly Useful factors

These are the other data given in chapter 3.2.1

3.3 Output

The following parameters are the results of the UtilNets application.

3.4 Model Validation or calibration

(Statistical model)

Variable	Unit	Short description
Expected life-time	years	Prediction of the pipe lifetime (50% probability)
Expected life-time	year of failure	Prediction of the pipe lifetime for each single pipe (threshold of failure probability can be chosen), plus survival curve
Expected failures		Prediction of specific Link and Segment failures in a chosen time horizon
Order of rehabilition		Recommended order of rehabilition considering year of failure, costs, budget, methods, customer specifications and more
Costs of rehabilitation	[valuta/year]	Expected cost for each rehabilitation
Reliability factor		A factor representing to what extension the input variables are filled by the user (weighted regarding sensitivity)

4 POSSIBLE IMPROVEMENTS OF THE MODEL

As mentioned in the previous chapters, UtilNets is not finalised yet. The main development of import tools and main modules has been done. The program is running satisfyingly stable. What has yet to be done is the refinement of default rules which are filling data gaps with pre-assumptions. Additionally, definitions for general data have to be adjusted. These are for example the values for corrosion factors, the assumed mean unsupported length of the pipes and material resistance specifications.

An evaluation of the results given by UtilNets has either not be done yet.

5 SOFTWARE SPECIFICATIONS

5.1 Programming Language(s) or Mathematical-Statistical Software(s)

The current version of UtilNets is written in Visual Basic 6.0. The interface is a Multi Document Interface (MDI), Thus, several forms can be displayed within the main form. Also the import facilities are implemented in Visual Basic 6.0 and provide a graphical environment with which the end-user can access the database through calls to the Data Manager using a common mechanism. This mechanism is the Dynamic-Link Library (DLL) written in C++, which permits data retrieval from and storage in the *UtilNets* database. The DLL uses the ODBC standard to access any database.

The Utilnets database runs under Oracle. It seems to be sensitive which version is used as the users experienced during testing. The Oracle version 7.3.3.0.0 for Windows NT should be installed. Sintef uses Personal Oracle 7.3.4.0.0 which works also fine.

As Utilnets has direct GIS import and export facilities it needs additionally the installation of ArcView. Here the versions 3.1 and 3.2 are applicable. Sintef used for testing ArcView 3.2.

UtilNets demands a number of system requirements and preinstalled tools. It should be run on a PC with at least a Pentium III processor with 128 MB memory. For the operating system Windows NT 4 + service pack is recommended.

5.2 Possible Input File(s) Formats

The DTM (Data Transformation Module) has three possible import facilities, via tables, complete ArcView projects and SQL scripts. When selecting the import via tables, the user has the possibility to choose between ArcView tables with the extension ".dbf" or ordinary text files. The latter can be prepared by e.g. Excel sheets.

Any other information can be manually edited under the Import Manager. UtilNets has an enormous flexibility regarding data import and any common database can be read or at least transformed.

5.3 Possible Output File(s) Formats

During the UtilNets calculations each input or result table can be printed, saved or exported to an Excel file. The results can also be exported again to a GIS system and the single pipe failure can thus be located.

6 REFERENCES

6.1 Theoretical Framework References

Some of the relevant documents have already been included in the respective chapters. The complete final report and the final user manuals to each tool part will be delivered as soon as the EU Commission has approved the finalising of the project.

6.2 Practical Use and Results References

The development and intermediate results of the UtilNets project have been presented at various conferences and scientific journals. A complete list will be delivered after the approval of the EU Commission. Two relevant articles are attached below, a third one (NODIG conference) is already included in chapter 2.1.

UtilnetRef5.pdf

UtilnetRef6.pdf

APPENDIX 5 : NHPP MODEL

1 GENERAL DESCRIPTION

1.1 Name and/or Acronym of the Model

NHPP model

1.2 Company/Research Center/University

Norwegian University of Science and Technology, NTNU.

Department of Hydraulic and Environmental Engineering

1.3 Objectives

The main objective of NHPP MODEL is to predict failures for each individual pipe in a water distribution network based on historical failure data. The relative importance of different explanatory variables is reported by its regression coefficients.

1.4 Functional description

NHPP MODEL models the failure-process in water supply networks as a Non Homogeneous Poisson Process (NHPP) which also takes into account the factors influencing (e.g. material, diameter, length) the failure history. The relative importance of the explanatory variables is reported and future failures for each pipe in the network in predicted.

1.5 Brief Historical Overview of the Model

NHPP MODEL results from a strategic university program at NTNU carried out during 1996-2000. The work resulted in a *Doktor ingeniør* thesis (Røstum, 2000) with the main focus on statistical modelling of pipe failures in water networks. The work introduces the Non Homogeneous Poisson Process (NHPP) with covariates (i.e. explanatory variables) as an appropriate method for modelling pipe failures in water networks. NHPP is well known in the fields of reliability analysis and medicine. The NHPP can be used to model minimal repair processes, i.e. processes where intensity of failures remains the same after a repair. This is the normal situation for pipe repairs in water distribution systems. Most water pipes are repaired by replacing a very small segment of the pipe, or by using a repair sleeve.

2 THEORETICAL FRAMEWORK OVERVIEW/PAST STUDIES

In NHPP MODEL the failure process for each pipe is modelled as a Non-Homogeneous Poisson Process (NHPP) which also includes covariates (i.e. explanatory variables). The NHPP has shown to be capable to model the complex failure process existing in water networks (Røstum, 2000).

In a non-homogeneous Poisson process each pipe is studied within the time interval (a_i,b_i), i.e. time interval where observations are available. Each pipe has its own covariate vector $\mathbf{z_i}$ and a number n of recorded failures, with the time of their occurrence: $T_1<T_2<...<T_n$. The components of the covariate vector are all independent variables that have a significant influence on the pipe's service-life. The effect of the covariates on the rate of occurrence of these failures (ROCOF) is of interest.

The intensity of the NHPP is described with the following model when also covariates are included:

$$\lambda(t,\boldsymbol{\beta},\mathbf{z}_i) = \lambda\delta t^{\delta-1}\exp(\mathbf{z}_i{}'\boldsymbol{\beta}) \qquad (1)$$

To estimate the unknown parameters (λ, δ and β) in the chosen NHPP, the principle of maximum likelihood is used.

In Trondheim the following covariates have been used to describe the intensity function for each pipe in the network: *pipe diameter*, *pipe length, soil condition* and *age of pipe when observation starts*.

2.1 Scientific background/nature of the model

Figure 4 shows an example of the failure data typically available for water distribution networks. The failure events are marked with an "o" on the time axes. The time window reflects the period where failure data is available.

The failure data on the left side of the time window is not known. Failures may have occurred in this period, but are unrecorded. We call this left-censored failure data. The right side of the time window corresponds to an upper bound of time for which failure data is available. Failure data will be recorded in the future, but these data are not included in the analyses. This means that the data is also right censored.

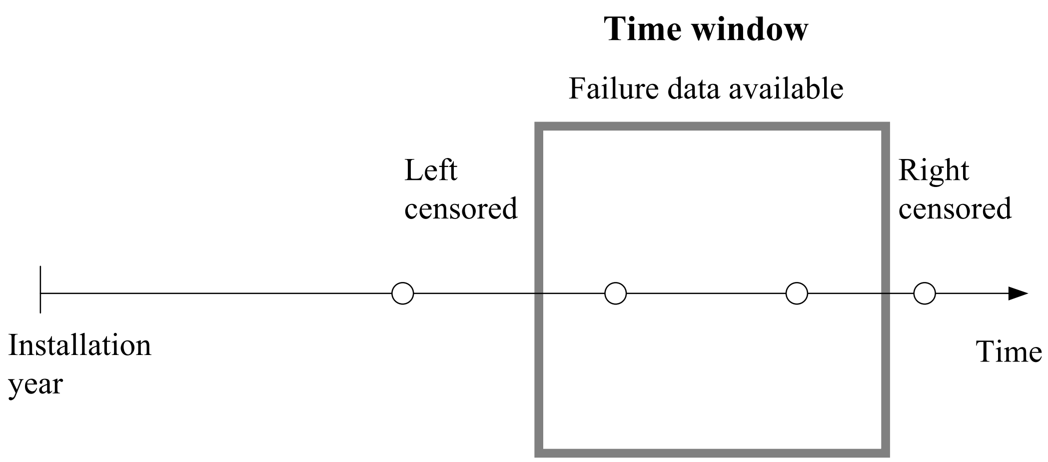


Figure 4. Typically availability of failure data in water networks.

A data set might also consist of some wrong/false data such as impossible inventory data, typing errors, etc. Before the data can be analysed, these *false* data must be detected and then discarded or corrected. Otherwise the results can be distorted with the presence of false data.

2.2 Underlying assumptions

As for other statistical models, NHPP MODEL uses *historical* failure data to predict future failures in a network. It is then assumed that the history will be repeated and that the factors behave ii the same way in the future as they did in the past.

In the NHPP it is assumed that the pipe is not restored to a ‘good-as-new’ state after the repair, and the intensity of failures for the repaired pipe is unchanged (i.e. minimal repair process). This is the normal situation for pipe repairs in water distribution systems.

2.3 Algorithm

In a non-homogeneous Poisson process each pipe is studied within the time interval (a_i,b_i), i.e. time interval where observations are available. Time 0 corresponds to the laying year of the pipe. Each pipe has its own covariate vector $\mathbf{z}_i$ and a number *n* of recorded failures, with the time of their occurrence: $T_1<T_2<...<T_n$. The components of the covariate vector are all independent variables that have a significant influence on the pipe's service-life. The effect of the covariates on the rate of occurrence of these failures (ROCOF) is of interest.

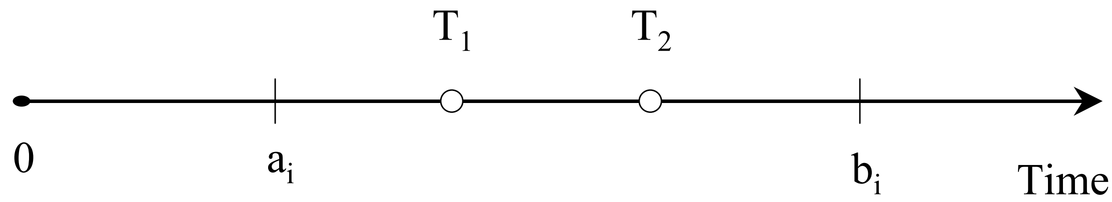


Figure 5. Definition of terms used in NHPP.

The intensity function applied in NHPP MODEL , when covariates are included is:

$$\lambda(t,\boldsymbol{\beta},\mathbf{z}_i) = \lambda\delta t^{\delta-1}\exp(\mathbf{z}_i'\boldsymbol{\beta}) \tag{2}$$

The cumulative or integrated intensity function is

$$E(N(t)) = \Lambda(t,\boldsymbol{\beta},\mathbf{z}) = \int_0^t \lambda(u,\boldsymbol{\beta},\mathbf{z})du \tag{3}$$

where *N(t)*= number of failures in (0,*t*]. The integrated intensity function for the interval (a_i,b_i), corresponding to the expected number of failures in the interval (a_i,b_i) is given by:

$$E(N(b_i)-N(a_i)) = \int_{a_i}^{b_i} \lambda(u,\boldsymbol{\beta},\mathbf{z})du = \lambda\left(b_i^{\delta} - a_i^{\delta}\right)\exp(\mathbf{z}'\boldsymbol{\beta}) \tag{4}$$

An illustration of the intensity function for the NHPP is shown in Figure 6. The area under the curve is equivalent to the expected number of failures for the time interval. For the NHPP model, this curve can be integrated using Eq. (4).

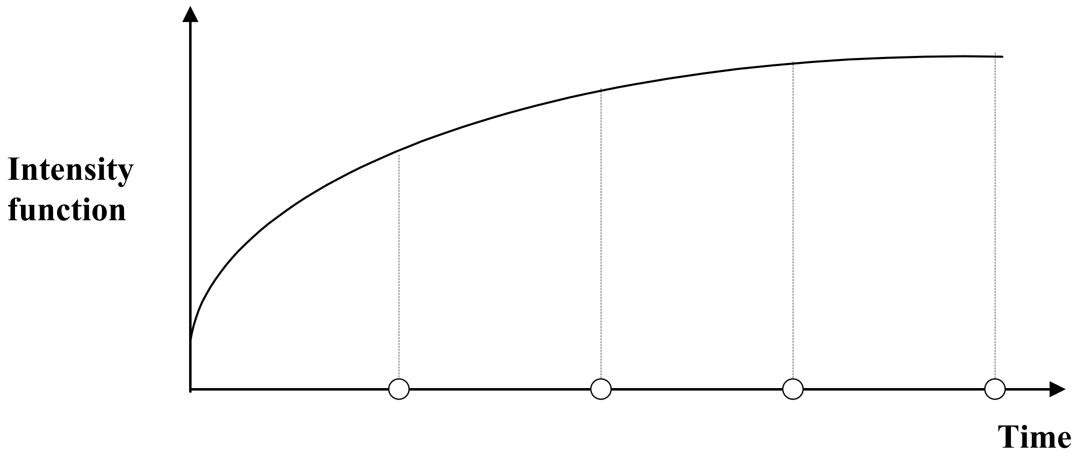


Figure 6. An illustration of the pattern of the intensity function for NHPP.

To estimate the unknown parameters (λ, δ and β) in the chosen NHPP, the principle of *maximum likelihood* is used. The likelihood function when covariates are present is denoted as $L(\theta;\mathbf{z},t)$. We might think of the likelihood function as a measure of how "likely" θ is to have produced the observed T values.

Information about m independent observations with identical intensity function $\lambda(t)$ is available (i.e. inventory and failure data). Individual (e.g. pipe) i is observed over the time interval (a_i, b_i) and n_i events are registered at the times t_{ij}, where j=1,2,…n_i and i=1,2,…m.

The likelihood function for the power law model for all m processes is given by:

$$L(\theta;t)=\prod_{i=1}^{m}\left[\cdot\prod_{j=1}^{n_i}\left[\lambda\left(t_{ij}\right)\right]\cdot e^{-\int_{a_i}^{b_i}[\lambda(u)du]}\right] \tag{5}$$

The maximisation of Eq. 5 is achieved taking the logarithm of L and maximising the new function (l=ln$L$). The log- likelihood function (l) for the power law model is given by:

$$l(\theta;\mathbf{z},t)=\sum_{i=1}^{m}\left[n\mathbf{z}_i\beta+n\ln\lambda+n\ln\delta+(\delta-1)\sum_{j=1}^{n}\ln t_{ij}-e^{\mathbf{z}_i\beta}\lambda\left(b^{\delta}-a^{\delta}\right)\right] \tag{6}$$

The maximisation of the log-likelihood function is performed in the program by a special optimisation algorithm, which only requires the following formulas for the first derivative of $l(\theta;\mathbf{z},t)$:

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^{m} \left[\frac{n}{\lambda} - e^{\mathbf{z}_i \beta} \left(b^{\delta} - a^{\delta} \right) \right] \tag{7}$$

$$\frac{\partial l}{\partial \delta} = \sum_{i=1}^{m} \left[\frac{n}{\delta} + \sum_{j=1}^{n} \ln t_{ij} - e^{\mathbf{z}_i \beta} \lambda \left(b^{\delta} \ln b - a^{\delta} \ln a \right) \right] \tag{8}$$

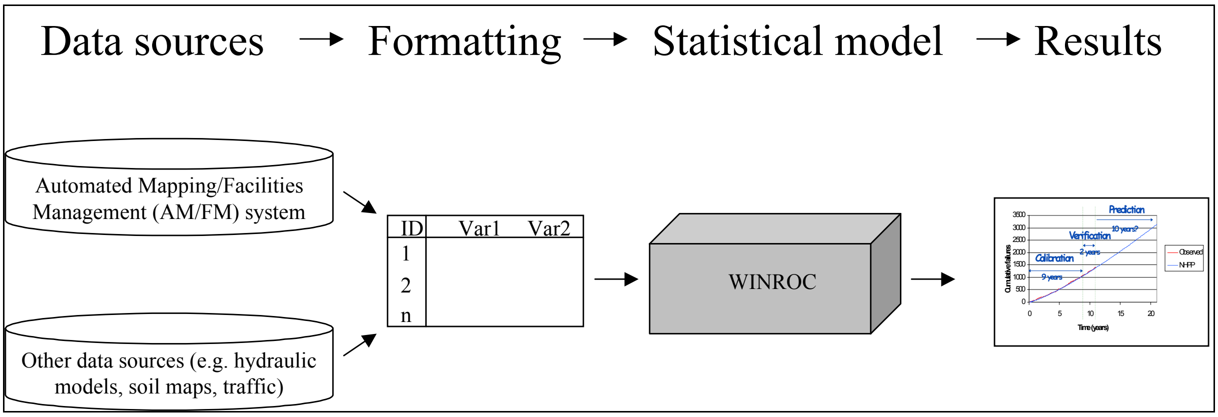
$$\frac{\partial l}{\partial \beta_r} = \sum_{i=1}^{m} \left[n\mathbf{z}_i - e^{\mathbf{z}_i \beta} \lambda \mathbf{z}_i \left(b^{\delta} - a^{\delta} \right) \right] \tag{9}$$

2.4 Past studies and conclusions

The NHPP has shown to be capable to model the complex failure process existing in water networks. So far it has been applied in a case study in Trondheim Norway (Røstum, 2000).

3 SPECIFICATIONS OF COMPUTATIONAL STEPS

3.1 Functional Scheme



3.2 Raw Data Formatting

The raw data, originating from maintenance records or hydraulic modelling for instance, have to be formatted before being used as input of the model. This step can be carried out within spreadsheet, statistical software or in some cases also generated by the program storing the maintenance data.

3.3 List and Definition of Explanatory Factors

Explanatory factors are variables specific of the individuals, and that can be used to explain the variation in the observed values of the dependent variable(s).

The most important variables describing the structural deterioration of water networks can be grouped into four (4) categories; structural or physical variables, external or environmental variables, internal or hydraulic variables and maintenance variables (Røstum et al., 1997).

Table 1. Factors affecting structural deterioration of water distribution pipes.

Structural variables	External/environmental variables	Internal variables	Maintenance variables
Location of pipe	Soil type	Water velocity	Date of failure
Diameter	Loading	Water pressure	Date of repair
Length	Groundwater	Water quality	Location of failure
Year of construction	Direct stray current	Water hammer	Type of failure
Pipe material	Bedding condition	Internal corrosion	Previous failure history
Joint method	Leakage rate		
Internal protection	Other networks		
External protection	Salt for de-icing of roads		
Pressure class	Temperature		
Wall thickness	External corrosion		
Laying depth			
Bedding condition			

The variables that have and an influence on the failure intensity will vary from case to case, depending on local conditions. All variables that you think might have some kind of influence on the rate of occurrence of these failures are of interest. For each pipe in the network the values of the variables has to be found. You seldom have all the information given in Table 1, so the answer is to use the data, which is available.

3.3.1 Required factors

The following variables for each pipe are normally available via the inventory database of the network and can be seen as a minimum of required data:

Variable	Unit	Short description
Pipe_ID		Code of single pipe
Failure times		Date of failures
Pipe material	-	Code of the pipe material according to a convention
Pipe diameter	mm	
Pipe length	m	Length of a pipe
Pipe age	years	Time from installation of pipe
Type of soil	-	Soil classification system
Water pressure	m	Static water pressure in each pipe
Water velocity	m/s	Water velocity in each pipe
No of previous breaks		Number of previous failures for each pipe
Other variables		The examples of the explanatory variable explained above are just examples. Other variables might be of interest in other cases, depending on the local conditions.

3.4 Output

NHPP MODEL gives the following results:

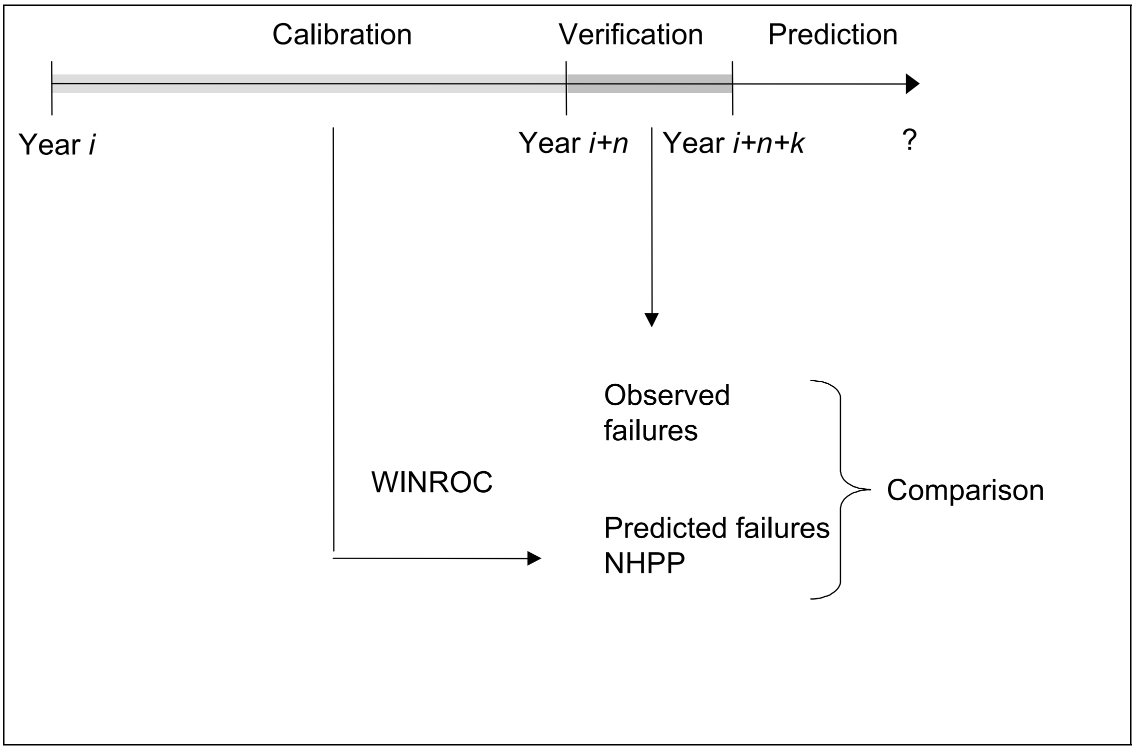
Variable	Unit	Short description
Regression coefficients	-	Relative importance of the significant variables
Failure intensity	-	Time derivate of the expected cumulative number of failures
Expected number of failures	-	Expected number of failures for each pipe within a given time interval

3.5 Model Validation or calibration

Following techniques are useful for evaluating the model:

- Cumulative plots for observed versus predicted failures (Nelson-Aalen plot)
- Annual plots for observed failures versus predicted failures for each year
- Plotting the pairs (observed failures, predicted failures) for each pipe

The procedure for calibration and validation of the results is shown bellow:



3.6 PI(s) Estimation Method/forecasting method

Two performance indicators are calculated, namely the failure intensity function (i.e. λ) and the expected number of failures in an interval (i.e. N(t)).

Intensity of failures:

The failure intensity function for a NHPP, when covariates are included is given by:

$$\lambda(t, \boldsymbol{\beta}, \mathbf{z}_i) = \lambda \delta t^{\delta - 1} \exp(\mathbf{z}_i{}' \boldsymbol{\beta})$$

N(t):

The integrated intensity function for the interval (a_i,b_i), corresponding to the expected number of failures in the interval (a_i,b_i) is given by:

$$E(N(b_i) - N(a_i)) = \int_{a_i}^{b_i} \lambda(u, \boldsymbol{\beta}, \mathbf{z}) du = \lambda\left(b_i^{\delta} - a_i^{\delta}\right) \exp(\mathbf{z}' \boldsymbol{\beta})$$

Both the intensity function and the expected number of failures are time dependent and might be used for prediction of future conditions.

4 POSSIBLE IMPROVEMENTS OF THE MODEL

The inverse matrix from the calculations of the regression parameters is normally used for checking the statistical significance of the parameters. Due to programming practice in *NHPP MODEL* the inverse matrix is not available, and therefore it is not possible to evaluate the significance of the parameters. In order to carry out this type of test, some reprogramming has to be carried out.

When analysing some particular types of dataset, NHPP MODEL sometimes does not find a solution. In order to evade this “problem” it has shown useful to change the order of covariates or to remove one covariate from the dataset.

5 SOFTWARE SPECIFICATIONS

5.1 Programming Language(s) or Mathematical-Statistical Software(s)

The programming language for NHPP MODEL is FORTRAN. No statistical software is required for running the analysis.

5.2 Possible Input File(s) Formats

The NHPP MODEL require the following input file format (Tabulator separated text file):

No. of pipes (p)

No. of failures (f)

No. of strata (*)

No. of covariates (c)

*

Name $Covariate_1$

Name $Covariate_2$

…

Name $Covariate_{\#}$

a_i	b_i	$Pipe_i$	*	$Covariate_{i1}$	$Covariate_{i2}$	$Covariate_{i\,c}$	
a_j	b_j	$Pipe_j$	*	$Covariate_{j1}$	$Covariate_{j2}$	$Covariate_{j\,c}$	Inventory data
…							
a_p	b_p	$Pipe_p$	*	$Covariate_{p1}$	$Covariate_{p2}$	$Covariate_{p\,c}$	

T_1	$Pipe_i$	
T_2	$Pipe_l$	
…		Failure data
T_f	$Pipe_p$	

Where * is an option for stratification and (a_i,b_i) represents the time window where $pipe_i$ is observed.

The program is executed by typing *winroc def* in DOS mode.

5.3 Possible Output File(s) Formats

The output file is automatically generated by the program. The user designates the name of the output file in a definition file (def). The output file is a ".txt" file which easily can be imported to other programmes like Excel for further analysis, presentations etc. The output file includes:

- descriptive statistics for the input file
- estimated values for the parameters
- cumulative plots
- predicted failures for the observed period

6 REFERENCES

6.1 Theoretical Framework References

Røstum, J. (2000). Statistical modelling of pipe failures in water networks. PhD thesis, 2000:12, Norwegian University of Science and Technology, Trondheim, Norway. ISBN 82-7984-033-8.

6.2 Practical Use and Results References

Røstum, J. (2000). Statistical modelling of pipe failures in water networks. PhD thesis, 2000:12, Norwegian University of Science and Technology, Trondheim, Norway. ISBN 82-7984-033-8.

Røstum, J. Schilling, W. (1999). Predictive service life models for water network management. In: *Proceedings of the 13th EJSW*, 8 September – 12 September, Dresden University of Technology, ISBN: 3-86005-238-1.

APPENDIX 6 : AQUAREL

1 GENERAL DESCRIPTION

1.1 Name and/or Acronym of the Model

AQUAREL

1.2 Company/Research Center/University

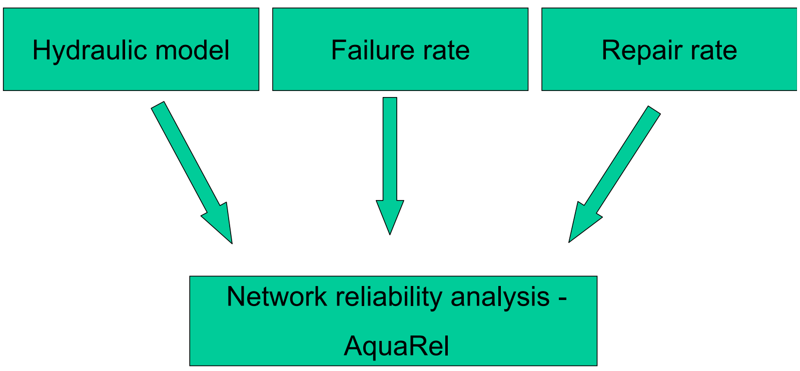
SINTEF, Department of Water and Waste Water, Norway

1.3 Objectives

AQUAREL calculates reliability measures for the reliability of a water distribution network allowing simultaneous failure of equipment. The approach is based on hydrostatic simulations of the conditions in the network (EPANET 2.0) combined with standard reliability calculation techniques. The idea is to close the links in the network and examining the effect on the supply nodes using EPANET. The model also takes into account the volume-effect of the elevated reservoirs (tanks). As input data AQUAREL requires the failure intensity ("failure rate") for all links (i.e. pipes and pumps) in the network. AQUAREL calculates several reliability measures at pipe (i.e. node) level (i.e. water supply availability, frequency of degraded pressure, link importance_ B, link importance_ U and link importance_ F).

1.4 Functional description

The system reliability is dependent on the hydraulics in the network, failure rate and the repair rate. Failure rates and repair rates vary from link to link. The integration of these elements leads to a water network reliability analysis as shown bellow.



1.5 Brief Historical Overview of the Model

Water distribution networks are traditionally designed to be completely reliable. However, the increasing scarcity of public money for construction and maintenance and the advanced age of many water supply systems are causing system operators to focus on reliability analysis. Reliability models have been adopted in other fields, but what's special for water supply is that the analysis also must include a hydraulic analysis. AQUAREL is developed for this purpose.

The method is based on the hydraulic simulation program EPANET (version 2.0), a failure time model and standard reliability theory. The idea is to close the links in the network and examining the effect on the supply nodes using EPANET. The results obtained from AQUAREL are suited for aiding operators in design, verification and vulnerability assessments of water distribution networks and for establishing maintenance and preparedness strategies. The cut sets are also valuable information in these analyses.

The Research Council of Norway has financed the development of AQUAREL. Lately, AQUAREL has been further developed to also include the effect of increased reliability caused by including elevated reservoirs (tanks) in the network.

2 THEORETICAL FRAMEWORK OVERVIEW/PAST STUDIES

2.1 Scientific background

The AQUAREL procedure combines hydraulic modelling with failure rate modelling and links these to elements together with standard reliability theory.

2.2 Underlying assumptions

Since AQUAREL requires a hydraulic network model (i.e. EPANET) for the hydraulic simulations, it is important the hydraulic model actually represents the real life conditions (i.e. the model has to be calibrated). The AQUAREL approach is based on closing pipes and evaluating the resulting pressure in nodes. Closing crucial pipes might result in steep gradients. The hydraulic models are normally not calibrated for these situations. However, the results so far show that the AQUAREL model is robust with respect to this.

2.3 Algorithm for the AQUAREL procedure

In the following the AQUAREL procedure for calculating reliability of water networks are outlined:

Notation

N	=	Number of nodes in the network
M	=	Number of pipes
P_i	=	Required pressure at node i, i = 1,…, N
R_i	=	Actual water pressure at node i, i = 1,…, N
I_j	=	The pressure is divided into 4 intervals to give the availability distribution, I_j denotes interval j, j = 1,…, 4
x_l	=	$\begin{cases} 0 \text{ if pipe } l \text{ is in a fault state, i.e broken, } l = 1,\ldots, M \\ 1 \text{ otherwise} \end{cases}$
$\boldsymbol{x}$	=	$[x_1, x_2, \ldots, x_M]$ = state vector
1_l, 0_l	=	The notation 1_l is used to explicit state that link l is functioning, and 0_l is used to explicit state that link l is in a fault state.
λ_l	=	Failure frequency for link l, l = 1, …, M
q_l	=	Unavailability for link l, l = 1, …, M
K_{1i}^j	=	All cut set of order 1 wrt. I_j, i.e $\{ l \mid x_l = 1 \cap x_m = 0, l \neq m \Rightarrow R_i \in I_j \}$
K_{2i}^j	=	All cut set of order 2 wrt. I_j, i,e $\{ l,m \mid x_l = 1 \cap x_m = 1 \cap x_n = 0, n \neq l,m \Rightarrow R_i \in I_j \}$ (Non-minimal cut sets are excluded from the cut set list)
p_i^j	=	Probability that pressure at node i is in the interval I_j (availability distribution)

		distribution)
F_i^j	=	Frequency of transition into interval I_j with respect to the pressure at node i
I_l^B	=	Basic importance measure of link l, l = 1, …, M
I_l^U	=	Criticality importance of link l, with respect to *unavailability* of link l
I_l^F	=	Criticality importance of link l, with respect to *frequency* of link l failures

Algorithm for Evaluating Hydraulic Cut Sets Using EPANET

The EPANET simulator (version 2.0) from the US EPA Drinking Water Research Division is used as the hydraulic engine in this approach. The performance of the network at the end users may be described by water quality, pressure, flow rate and availability. In this paper we will only focus on pressure and availability for the end user.

Failures of equipment such as pipes, pumps and valves are simulated by closure and the effects of these failures are examined by EPANET to give the hydraulic cut sets for each node. In the presentation we will use the general term "link" to represent pipes, pumps and valves. Failures of all this equipment is treated in the same manner, and is represented by a link failure in the following. The algorithm for utilising EPANET is now:

1. Start with a definition of the entire network in EPANET format, this network is denoted NW
2. The cut sets list is initially empty, i.e. $K_{1i}^j = \emptyset$, $K_{2i}^j = \emptyset$
3. Repeat for all links l:
   Let NW|0_l denote the network when link l is closed (i.e. is in a fault state). Use EPANET with NW|0_l as input, and calculate the actual pressure R_i for each node i = 1, …, N. Determine which interval R_i belongs to, i.e. find j such that $R_i \in I_j$. {l} is now said to represent a cut set of order 1 with respect to level j and node i. {l} is added to the cut set list of order 1, K_{1i}^j.
4. Repeat for all combinations of links l and m:
   Let NW|0_l,0_m denote the network when links l and m are closed (i.e. is in a fault state). Use EPANET with NW|0_l,0_m as input, and calculate the actual pressure R_i for each node i = 1, …, N. Determine which interval R_i belongs to, i.e. find j such that $R_i \in I_j$. {l,m} is now said to represent a cut set of order 2 with respect to level j and node i. Link l is added to the cut set list of order 2, K_{2i}^j.
5. Analogous to standard fault tree analysis, we exclude non-minimal cut sets. If {l} or {m} belongs to K_{1i}^j, then {l,m}is not minimal, and is excluded from K_{2i}^j.

The result from the EPANET analysis is then for each node i the minimal cut sets {K_{1i}^j, K_{2i}^j} for level j, j = 1,…,4. Further we also save the actual pressure for each node i, given link failure of link l, and given link failure of both link l and m.

Algorithm for including the effect of increased safety due to installation of tanks

Installing tanks in water networks has a positive effect on the water supply reliability. However, the size of a tank is limited and normally it can only feed water for a limited time (i.e. hours or days). Cutting out different pipes/set of pipes influence the emptying of the tanks in different ways. The time to repair a pipe to repair a pipe also varies from pipe type to pipe type. For minor pipe failures the repair might be is carried out before the tank is emptied, but for some critical pipes the tank might be emptied before the pipe is repaired.

The log- normal distribution is normally used for modelling repair times.

$$f(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\tau} \frac{1}{t} e^{-(\ln t - \nu)^2 / 2\tau^2}$$

Different categories of pipes have follows different repair times distributions. For example, larger pipe dimensions need more time for repair than smaller pipes. This will be the case for both mean time values (MTTR, mean time to repair) and for maximal repair times.

For each group of pipes the repair time distribution estimated is estimated based on two point on the distribution curve, namely the mean time to failure (MTTF) and the 99% percentile of the time to repair (TTR_{99}). Table 1 shows typically repairs data for Norwegians conditions. However, the repair times will vary from city to city and often the municipalities have defined some minimum required standards. TTR_{99} can normally be assumed to be in the order of 2-3 times MTTR.

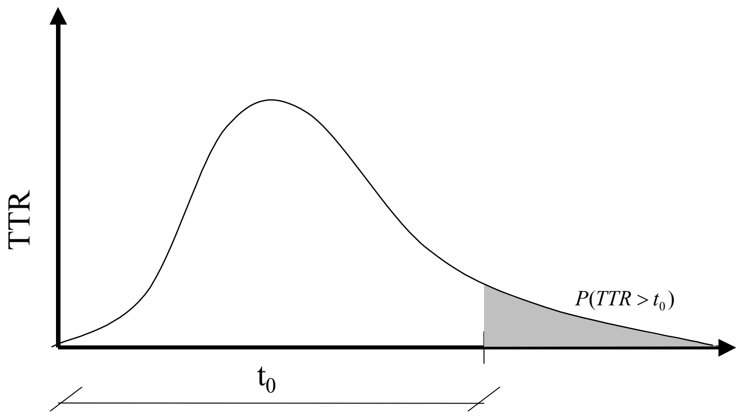


Figure 1. Illustration of the probability that time to repair (TTR) is larger than the time to empty a tank (t_o).

In order to analyse the effect of tanks the following algorithm is included:

For each node in the network:

1. For each pipe/pair of pipes
2. Run AQUAREL in normal mode (steady state)
3. If pressure in node < *critical pressure*: register pipe as a cut set
4. Run AQUAREL with empty tanks. Tanks are emptied by closing pipes in/out of the tanks and adding a by-pass pipe manually
5. If pressure in node < *critical pressure*: If pipe is already registered as a cut set and the pressure reduction is larger than with FILLED tanks, the cut sets are updated with information about pressure reduction with EMPTY tanks, included $t_0 = V/Q$, where Q is water flow out of tank and V is the total tank volume. If the cutset is so far not included a new cut set is recorded.
6. Repeat step 1 –5 until all pipes/pair of pipes have been applied
7. The cut set might be sorted according to pressure difference or probability ($P(TTR < t_0)$).
8. For the first 100 cut sets a *multiple extended period simulation* is carried out. The time, t_0 , where *all* tanks are emptied is recorded. If there is still more water again in the tanks after simulation or in case of some kind of errors arise in EPANET the values for t_0 from the steady state simulations are used. The 100 most important pipes (sorted) have now received updated values for t_0.

For each node the reliability is calculated based on all cut set (also those with no updated value for t_0). We define:

I_i = pressure interval i, i = 1, 2, 3, 4
PFB_K = Pressure in node for cut set K, with filled tanks

PTB_K = Pressure in node for cut set K with empty tanks

q_i = P(failure at $link_i$)
$r_i = P(TTR < t_0)$, t_0 = time to empty the tanks

We assume that K = {$link_i$, $link_j$} and $PFB_K = I_k$, $PTB_K = I_l$. The availability is then updated as follows:

$$P(I_k) = P(I_k) * (1 – (q_i * r_i) * (q_j * r_j))$$
$$P(I_l) = P(I_l) * (1 – (q_i * (1 - r_i)) * (q_j * (1 - r_j)))$$

2.4 Past studies and conclusions

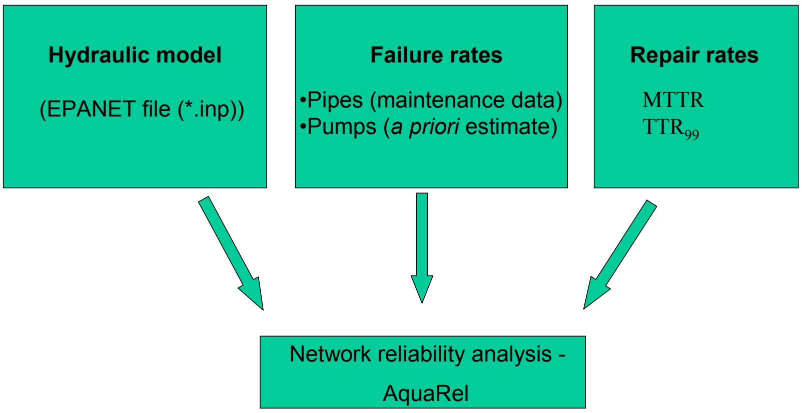
So far AQUAREL has been applied in two different case studies in Norway (i.e. Trondheim and Narvik). In Trondheim the network is about 700 km, supplies more than 150 000 persons and consist of more than 9000 pipes. The required input data for failure rates for each pipe in the network was estimated by using the statistical failure prediction model NHPP MODEL (Røstum, 2000). Data for pumps was not available and therefore as a priori estimate, data collected for similar, offshore installations has been used to estimate failure and repair rates for pumps. The results from the AQUAREL procedure was used as input for the new master plan in Trondheim.

In Narvik the network is smaller (20000 persons) and the available failure records where very limited. Therefore, rougher estimates for the failure rates were used. Nevertheless, the AQUAREL procedure showed to be robust and very useful as input for rehabilitation plans and defining candidates for rehabilitation.

3 SPECIFICATIONS OF COMPUTATIONAL STEPS

3.1 Functional Scheme

The system reliability is dependent on the hydraulics in the network, failure rate and the repair rate. Failure rates and repair rates vary from link to link. The integration of these elements leads to a water network reliability analysis as shown bellow.



Figur 2. Analysis procedure with AQUAREL

3.2 List and Definition of required input data

The AQUAREL procedure requires the following input data:

Variable	**Unit**	**Short description**
Pipe_ID	[-]	Code of single pipe
Failure rate	failures/year	Failure rate for each pipe
MTTR	hours	Mean time to repair
TTR_99	hours	99 % percentile of the time to repair
Hydraulic model		An EPANET compatible hydraulic model is required

3.3 Output

AQUAREL calculated the following reliability measures:

Variable	Unit	Short description
Water supply availability	%	Portion of time in a given state
Frequency of degraded pressure	times/year	Number of times per year with degraded pressure
Link importance_ B		Birnbaum's importance measure of each link
Link importance_ U		Importance measure of each link wrt unavailability
Link importance_ F		Importance measure of each link wrt frequency

The water supply availability is averaged over time and might be interpreted as the probability distribution of the pressure at a given node. The Link importance_ B measure is similar to Birnbaum's measure of reliability importance in reliability theory.

In addition the method gives the hydraulic cut sets of a given node, i.e. the links that – if they fail simultaneously – causes the pressure at the node to drop below the predefined pressure.

3.4 PI(s) Estimation Method

AQUAREL calculates several reliability measures (i.e. performance indicators, PI(s)) at pipe (i.e. node) level). In this session the computational procedure for the PIs is given.

The calculations of the different reliability measures are based on the minimal hydraulic cut-sets cut sets { K_{1i}^j, K_{2i}^j } identified by the EPANET program. These cut sets form the basis for the reliability calculations. We will perform an availability calculation, a frequency calculation and a calculation to evaluate importance measures of links.

Water supply availability calculation

A hydraulic cut set at level *j* for node *i* was defined as a set of links, such that if all links in the set fail, the pressure at node *i* is in interval *j*. Let L_k^j denote the event that cut set *k* is in a fault state. Since L_k^j are positively dependent for different cut sets *k*, the probability that the pressure at node *i* is in interval *j* is now given by:

$$p_i^j = P(R_i \in I_j) = P\Big(\bigcup_{k \in \{K_{1i}^j, K_{2i}^j\}} L_k^j\Big) = 1 - P\Big(\bigcap_{k \in \{K_{1i}^j, K_{2i}^j\}} L_k^{j*}\Big) \leq 1 - \prod_{k \in \{K_{1i}^j, K_{2i}^j\}} (1 - P(L_k^j)) \qquad (2)$$

where

$$P(L_k^j) = \begin{cases} q_l \, for \; k = \{l\} \in K_{1i}^j \\ q_l \times q_m \, for \; k = \{l, m\} \in K_{2i}^j \end{cases} \qquad (3)$$

Equation (7) is now used as an upper bound approximation for the availability of node *i,* for the various levels *j*, *j* = 1,…,4.

Frequency of degraded pressure calculation

We now define F_i^j as the transition frequency into interval I_j for node *i*. This frequency is the total frequency into that interval from all intervals above. If we limited us to the cut sets up to order two, there are two contributions to the frequency:

- All links are functioning, and then link *l* fails with frequency λ_l
- Link *l* is in a fault state, and the link *m* (in the same cut set as *l*) fails with frequency λ_m

Thus, the total frequency is given by equation (7):

$$F_i^j = \sum_{\{l\}\in K_{1i}^j} \lambda_l + \sum_{\{l,m\}\in K_{2i}^j} (q_l \lambda_m + q_m \lambda_l) \tag{4}$$

Importance measures of each link

As a basis we define the importance measure of a link to be the total reduction in water pressure at one or more end users as a consequence of the link failure. We define the importance measure in relation to all nodes in the network (i.e. we examine the effect of the link failure on each node). Furthermore, to be flexible, we introduce the weight w_i that indicates the importance of node *i*. E.g., a hospital may be considered more important than regular resident areas.

When cut sets up to order two are considered, there are basically two situations that will cause a reduction in water pressure at node *i*. Either all links are functioning, and then link *l* fails and causes a reduction from $R_i(1_l)$ to $R_i(0_l)$. Further there is a probability q_m that another link *m* is in a fault state, and a failure of link *l* will then cause a reduction from $R_i(1_l,0_m)$ to $R_i(0_l,0_m)$. In total we therefore define the basic importance measure, I_l^B as:

$$I_l^B = (1 - \sum_{m=1,m\neq l}^{M} q_m) \cdot \sum_{i=1}^{N} w_i \left[R_i(1_l) - R_i(0_l)\right] + \sum_{m=1,m\neq l}^{M} q_m \sum_{i=1}^{N} w_i \left[R_i(1_l,0_m) - R_i(0_l,0_m)\right] \tag{5}$$

The importance measure given by equation (5) states the importance of link *l* is independent of the probability that link *l* is in a fault state, and may be viewed as an analogue to Birnbaum's measure of importance in standard reliability analysis. Multiplying I_l^B with the unavailability of link *l* will give a measure analogue to the criticality importance measure in standard reliability analysis. To emphasise the *unavailability*, we superscript with index *U,* i.e:

$$I_l^U = q_l \times I_l^B \tag{6}$$

The criticality importance measure in equation (6) takes the unavailability of link *l* into account, i.e. it is some average loss of pressure due to link *l*. It will also be interesting to investigate how often this happens, and we define a *frequency* version of the criticality importance measure:

$$I_l^F = \lambda_l \times I_l^B \tag{7}$$

3.5 PI(s) Forecasting Method

The PI(s) defined in session 3.4 might all be predictable as a function of *time*. Due to the deterioration processes in water networks, the failure rates might increase with time. This will also affect the reliability of the water network. The roughness coefficients applied in the hydraulic analysis might also be expressed as a function of time (increased friction with time). This will also affect the future reliability of the network. It is thus possible to calculate future reliability measures with AQUAREL by running different scenarios.

4 POSSIBLE IMPROVEMENTS OF THE MODEL

There is a need for improving the possibilities for visualizing the results (e.g. GIS). So far AQUAREL require EPANET format for the input file. In the future AQUAREL should be more flexible with respect to input format for the hydraulic models.

So far AQUAREL has been applied in two (2) case studies in Norway. By running more case studies future possible developments can be found. The new feature with also taking into account the improved reliability caused by the volume-effect of the elevated reservoirs (tanks) has so far not been applied in real case studies.

The hydraulic simulations are based EPANET. In EPANET the demand in each node is constant and not a function of the available water head. According to information received from EPA, the hydraulic engine is planned to be improved in future versions of the program.

5 SOFTWARE SPECIFICATIONS

5.1 Programming Language(s) or Mathematical-Statistical Software(s)

The program AQUAREL is programmed in Visual Basic and Visual C++. Visual Basic is used for programming the user interface and Visual C++ for the reliability calculations. The dynamic link library (DLL) of functions provided by EPANET Programmer's Toolkit are used for the hydraulic calculations.

The AQUAREL user interface is shown in Figure 7. The EPANET input file, the reliability data and the corresponding results are shown in separate windows. The user might choose a specific node for the analysis should be focused. There is also an option for ranking the cut-sets according to probability or pressure difference.

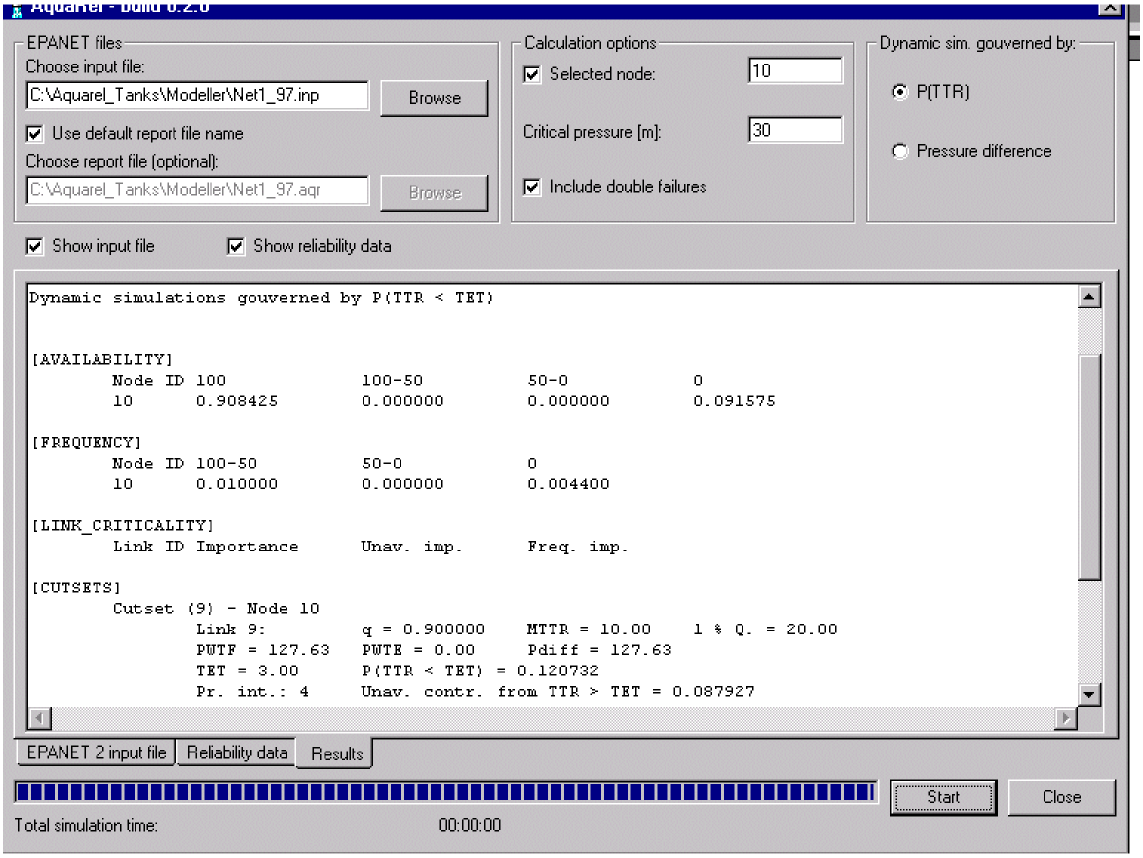


Figure 7. AQUAREL user- interface

In the result window the following abbreviations are used:

PWTF = Pressure with tanks full

PWTE = Pressure with tanks empty

Pdiff = PWTF - PWTE

TET = Time to Empty Tanks (t0)

Pr. int. = Pressure interval (Ij)

Unav. contr. from TTR > TET = bidraget til Ij når TTR > TET

Unav. contr. from TTR <= TET = bidraget til Ij når TTR <= TET

5.2 Possible Input File(s) Formats

AQUAREL requires an EPANET input file (*.inp) for the reliability calculations. If other hydraulic water network simulation programs are available these input files can relatively easily be converted. AQUAREL also requires reliability data for each link (i.e. pipes and pumps) in the network. The data format for the reliability data is Microsoft Access database (*.mdb). However, as long as the data is within a table the table can easily be imported to Access.

Format for input reliability data for pipes

ID	Availability	Frequency	MTTR	99 % percentile TTR
10	0.990000009536	9.99999977648258	5	15
11	0.980000019073	1.99999995529652	2	6
…	…	…	…	…

Format for input reliability data for pumps

ID	Availability	Frequency	MTTR	99 % percentile TTR
9	0.990000009536	9.99999977648258	10	30
…	…	…	…	…

The availability, *A(t)* at time *t* is the probability that an object (e.g. pipe) is functioning at time *t*. The average availability $A_{av}(t)$ denotes the mean proportion of time the object is functioning. If we have an object that is repaired to an "as good as new" condition every time it fails, the average availability is

$$A_{av} = \frac{MTTF}{MTTF + MTTR}$$

where MTTF (mean time to failure) denotes the mean functioning time and the MTTR (mean time to repair) denotes the repair time of the object. MTTF is the inverse of the intensity of failures ($MTTF = 1/\lambda$) and might result from the statistical failure models.

For each group of pipes the repair time distribution estimated is estimated based on two point on the distribution curve, namely the mean time to failure (MTTF) and the 99% percentile of the time to repair (TTR_{99}). Table 1 shows typically repairs data for Norwegians conditions. However, the repair times will vary from city to city and often the municipalities have defined some minimum required standards. TTR_{99} can normally be assumed to be in the order of 2-3 times MTTR.

Table 1. Example of repair data for pipes

	Repair time (hours)	
Pipe diameter	**MTTR**	**TTR_{99}**
< 300	8	24
300-400	16	48
> 400	24	72
Water tunnels	168	504

5.3 Possible Output File(s) Formats

AQUAREL reports the results as a text file (*.txt). The text file can later be exported to other programs (e.g. spreadsheet, GIS) for further presentations.

6 REFERENCES

6.1 Theoretical Framework References

Røstum, J., Sægrov, S. Vatn, J. and Hansen, G. K. AQUAREL- a computer program for water network reliability analysis combining hydraulic, reliability and failure time models. *In Proceedings*: CWS2000 Water Network Modelling for Optimal Design and Management 11 - 12 September 2000 Woodbury Park, Exeter, UK

Røstum, J. (2000). Statistical modelling of pipe failures in water networks. PhD thesis, 2000:12, Norwegian University of Science and Technology, Trondheim, Norway. ISBN 82-7984-033-8.

6.2 Practical Use and Results References

Røstum, J. (2000). Analyse av leveringssikkerhet i vannforsyningen til Trondheim kommune. SINTEF rapport STF22 A00313 (in Norwegian).

APPENDIX 7 : FAILNET-RELIAB

1 GENERAL DESCRIPTION

1.1 Name and/or Acronym of the Model

Failnet-Reliab

1.2 Company/Research Center/University

Cemagref

1.3 Objectives

This tool aims to assess the **reliability of drinking water networks**. Reliability is defined in the sense of water demand satisfaction, and, summarily, it is the quotient between the available consumption and the water demand.

After a specific hydraulic modeling, where available consumption is computed according to the head at each node, several reliability indices are assessed and could be used as performance indicators (PI). The different scales of assessment are:

- pipes: it is the impact of a break in the pipe on all the nodes of the network,
- nodes: it is the reliability of supply at the node in relation with all the links,
- global network (or a sector): it is overall reliability of the network.

1.4 Functional description

The model is elaborated in two steps.

First an hydraulic model is computed. This model is different than classical hydraulic models, because water consumptions are not fixed and depend on computed heads and water demands. Newton-Raphson method is used to solve hydraulic equation and compute the outputs.

Secondly reliability indices are assessed. They depend on results of hydraulic models (with or without pipe breaks), on weight of each nodes (quantity, vulnerability) and on pipe failure probabilities (assessed or not with forecast probability models).

Necessary data are classical hydraulic data (**node**: altitude, water demand, kind of water use, **pipe**: roughness, length, diameter, **tank**: volume, altitude, **pumps**…) and, optionally, failures probability.

1.5 Brief Historical Overview of the Model

Main dates:

1980-…: Elaboration of hydraulic models:

Zomayet: hydraulic simulation of drinking water networks in 24h.

Opointe: hydraulic simulation of drinking water networks at instantaneous demand pick time.

1994: new algorithm for Zomayet in thr framework of Olivier Piller Thesis.

1995: Elaboration of Porteau, graphical software, that allow a more “user-friendly” use of the computation program.

1995: Study of Stéphane Berthin, on reliability indices and hydraulic model with consumption depending on the head.

1991- 2001 : Elaboration of forecasting failure model to assess failure probability (P. Eisenbeis thesis, work of Y. Le Gat on Monte-Carlo simulation, work on Lausanne, Charente-maritime, Canal de Provence networks).

2000-2001: Elaboration of reliability model, application on Charente-Maritime networks

2001: Elaboration of water quality model (time of water in the network, computation of chlorine rate in each node) implemented in Porteau

Comments:

For now 25 years “Hydraulics and civil engineering” Unit of Cemagref Bordeaux has been working on water networks modeling. A deterministic model (Zomayet) has been firstly elaborated, then a probabilistic model (Opointe), that assesses water pick demand by the way of binomial law.

First destined to research work, these programs has been used up to now by water services and engineer consultants, for projects elaboration. This large diffusion has been allowing to the Unit to be more close than the “End-users” and to know their problems and requirements.

In the theme of diagnosis, it was obvious that classical hydraulic models were not sufficient to assess service quality of a network, because too fare from the reality. Then close to a lot of studies about reliability, the Unit worked in 1995 on this theme, and using the past knowledge (Berthin).

On another side, the Unit has been working on statistical failure models since 1994, study which was able to be used and useful for reliability studies. Thus Failnet has been created, which is the tool unifying the results of the two studies. The two parts of the tool can be used independently.

2 THEORETICAL FRAMEWORK OVERVIEW/PAST STUDIES

2.1 Scientific background

Conventional models satisfy the demand required at nodes. The flows in pipes match necessarily the demand. Conventional models have been developed for many years. Hardy Cross (1936) wrote the first set of equations leading to different ways of resolution. Collins et al (1978) introduced first the content model, where balancing equations appear as optimality conditions of an optimization problem. Piller (1995) wrote explicitly the derivatives of variable head as a function of demand. Wagner et al (1988a), Jowitt and Xu (1993), and Gupta and Bhave (1994) introduce a hydraulic model where the effective demand is a discontinuous function of pressure; consumer demand is assumed to be met if the pressure is greater than a fixed pressure and nil if the pressure is lower. Wagner et al (1988b) correct the effective demand after calculating the head corresponding to the total demand.

Generally speaking, network reliability problems have been examined from the perspective of stochastic performance (connexity). Wagner et al. (1988a) define two indices: reachability, which indicates the probability of water reaching a given node, and connectivity, which indicates the probability of all the nodes being connected to at least one source node. Connectivity is a stricter index than reachability. The problem of connexity is qualified as being NP complete, i.e. it requires algorithms with calculation times that are non-polynomial and, more often than not, exponential. To solve the problem, the solution consists in examining particular cases, such as series-parallel networks (Rosenthal, 1981, Wagner et al., 1988a), establishing simplifying hypotheses such as limiting the maximum number of simultaneous link failures (Fujiwara and De Silva, 1990, Jacobs and Goulter, 1991), or confining ourselves to the study of the probability that at least one of the links connected to a node is not defective (Goulter and Coals, 1986, Goulter and Bouchart, 1990).

2.2 Nature of the model

Deterministic model (hydraulic)

2.3 Underlying assumptions

Several underlying assumptions have been used:

HYDRAULIC MODELLING

- Head-losses formula

The head-losses formula is the Hazen-Williams formula:

$$J(q)=10.69\times\frac{q^{0.852}}{Chw^{1,852}D^{4.871}}$$

With *J* linear head-losses in the link (m/m),

q the flow (m^3/s),

D the diameter in m,

Chw Hazen-Williams roughness coefficient.

- Energy conservation on a link

This equation is relates to the equality of the difference of head between two nodes of a same link and the head-losses on this link:

$$J(q_{AB}) = h_B - h_A \,,$$

where $J(q_{AB})$ is head-loss existing on the link *AB* according to the flow q_{AB} and h_B and h_A are the heads at the nodes *B* and *A.*

- Flow conservation at a node

This equation is:

$$\sum_i q_i + c = 0$$

where q_i are the flows coming from or out the node and *c* the consumption at the node.

- Formula of available consumption according to the head

$$c_i(h_i) = \begin{cases} 0 \text{ if } h_i < h_i^m \\ d_i \cdot \sqrt{\dfrac{h_i - h_i^m}{h_i^s - h_i^m}} \text{ if } h_i \in \left[h_i^m, h_i^s\right] \\ d_i \text{ if } h_i > h_i^s \end{cases} \tag{1}$$

where h_i^s is the desired head to satisfy the demand and h_i^m the minimum required head at non-tank node i. This function is represented graphically in figure 1.

INDICES BUILDING

- Maximal number of simultaneous bursts

The maximal number of simultaneous bursts is assumed to be 1. The probability of 2 simultaneous bursts is assumed to be negligible.

- Breaks probabilities

Breaks probabilities can be fixed by different ways:

* pipe by pipe, coming from a failure forecasting program

* by pipes group or sector, according to the area, the material, the diameter, from a global failure forecasting program

* assumed, if there is no forecasting program.

- Weight of the nodes

The weight of the nodes are assumed according:

- the demand at the node
- the type of consumer (domestic, industrial, school, …)
- the vulnerability of the consumer (hospital, dialyzed person,…)

2.4 Algorithm

Newton Raphson iterative method

This method allows to solve some equations systems, by an iterative approach. The solution are approximated.

Algorithms

Three algorithms are elaborated according to the assumptions on energy conservation, flow conservation, head-loss formula and available consumption.

Written in mathematical language, two systems are established:

- HQ':

$$\begin{cases} Aq + c(h) = 0_n \\ \xi - {}^tAh - {}^tA^fh^f = 0_a \\ \xi = \xi(q) \end{cases}$$

used in a first approach and without taking into account elevated nodes (node with a pressure <0).

- HQ'':

$$\begin{cases} Aq + c(h_\lambda) = 0_n & \text{(mass balancing)} \\ \xi(q) - {}^tAh_\lambda - {}^tA^fh^f - {}^tA.{}^tS.\lambda = 0_a & \text{(energy balancing)} \\ H.h_\lambda - H.{}^tS.\lambda = H.h^m & \text{(pressure constraint at elevated nodes)} \end{cases} \qquad (2)$$

used when elevated nodes are existing. In this case a fictive head-loss is affected to each node, existing downstream to the elevated node. These head-losses permits to have a pressure at the elevated node equal to zero. It is thus possible to compute available consumption at each isolated node.

The values used in the equations are:

A	incidence matrix at non-tank nodes	Δ_k	Diagonal matrix of the derivative of c
A_f	incidence matrix at tank nodes	H	Matrix of the elevated nodes locations
${}^tA$	Transpose matrix A	S	Matrix of the simple nodes locations which are separed from the rest of the network by the elevated nodes
D_k	Diagonal matrix of the derivative of ξ at q^k		
D_j	Link diameter vector	h_m	Minimum heads at the nodes
ξ	Head losses vector	h_s	Desired head to satisfy the demand
q	Flows in the links	C	Node consumption
h	Heads at the simple nodes	λ	Virtual head losses when elevated nodes are taken into account
q_k	Flows in the links at the k-th iteration	λ_k	Virtual head losses when elevated nodes are taken into account at the k-th iteration
h_k	Heads at the simple nodes at the k-th iteration		
a	Number of links	n_{eg}	Number of elevated nodes
n	Number of simple nodes		

2.5 Past studies and conclusions

Study of Stéphane Berthin

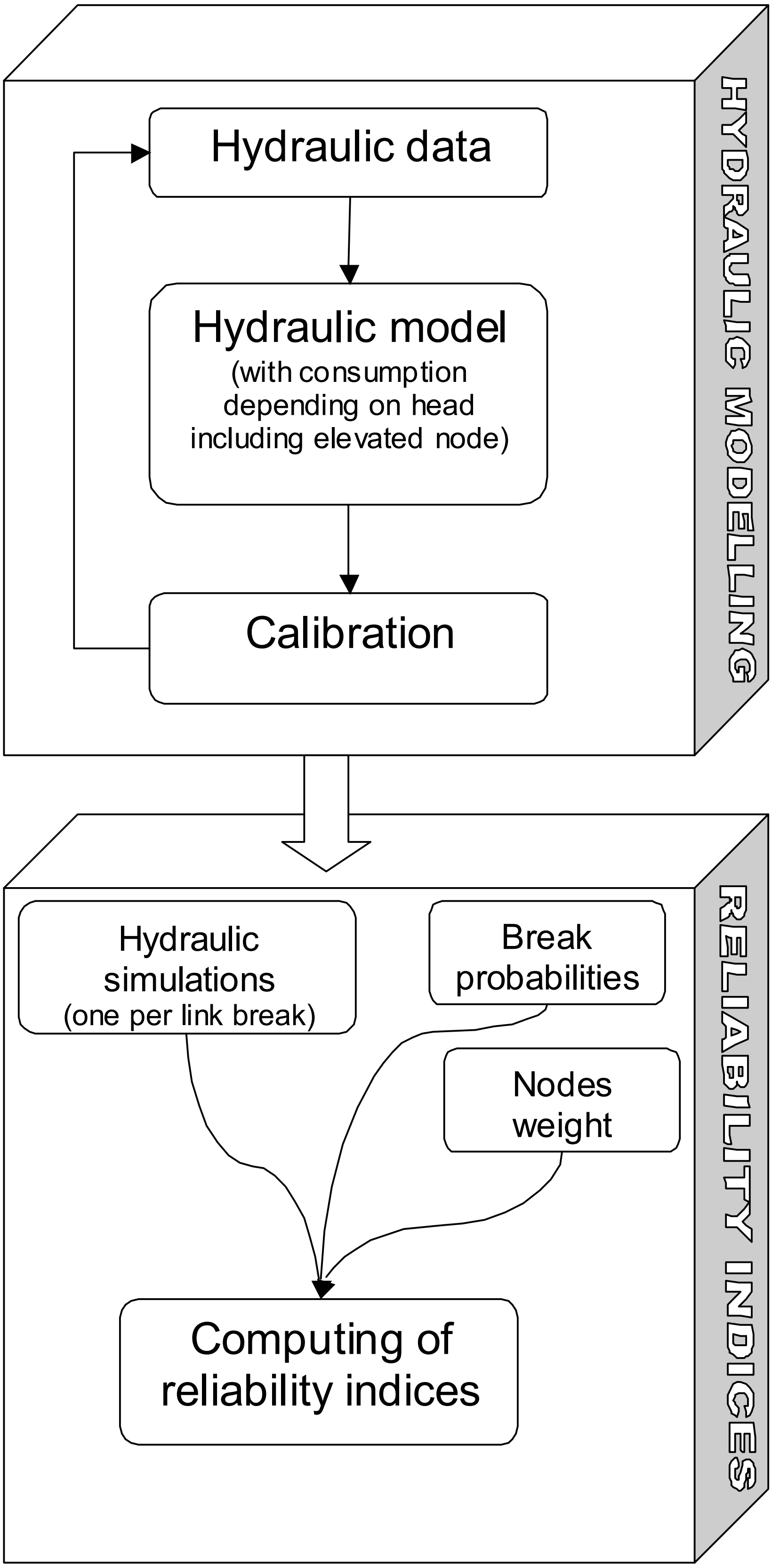
This study permits the elaboration of the algorithm proposed in the model. The problem of the elevated nodes has been evocated, but not formulated. The reliability index has been constructed. The method has been applied on a little city (30 000 inhabitants) in the West of the France.

Study of Aurélien Le Goff

This study will finish in July 2001. Its objectives are to write again the program, including the problem of elevated nodes. It will be applied of some rural and little urban networks.

3 SPECIFICATIONS OF COMPUTATIONAL STEPS

3.1 Functional Scheme



3.2 Raw Data Formatting

3.3 List and Definition of Explanatory Factors

Describe here the data and factors necessary for the model. These factors are either factors to be studied and analysed (statistical models) or data necessary for instance for a deterministic model. Explanatory factors are variables specific of the individuals, and that can be used to explain the variation in the observed values of the dependent variable(s). According to past studies, three level of "necessity" are proposed: Required, highly recommended, possibly useful. Each factors will be presented (Unit, qualitative or not, way to assess it…)

3.3.1 Required factors

Name	**Dimension**	**Description**	**Assessment**
Link ID	Alpha-numeric	Identification variable of the links	given
Node ID	Alpha-numeric	Identification variable of the nodes	given
Length	m	Link length	known in GIS or measured
Material	Alpha-numeric	Codes the pipe material.	GIS or personal or map
Diameter	mm	Internal diameter of the pipe	GIS, maps or personal
Roughness	mm or Hazen-Williams Unit	This value is used to compute head-losses with Hazen-William's Formula	assumed and calibrated
Height	m	Node height above sea level.	GIS or map or measured
Type of node	Alpha-numeric	Demand node, Tank, Water source	known
Desired pressure	m	Pressure desired by the consumer.	given or assessed
d_i	l/s	Maximum demand of the consumers at node i	assessed
Minimum pressure	m	Pressure below which the actual consumption vanishes (c=0)	
Maximum pressure	m	Pressure above which the actual demand is satisfied (c=d)	
Water level	m	Level of water in a tank or a water source	
w_i	Dimension-less	Weight of node i representing the qualitative and quantitative importance of the demand at this node.	
Unavailability Probability pe_j	Dimension-less	Probability of link j to be under repair, all other links being operational. Takes into account the probability of	

		failure and the mean repair duration of a break on link j.
pe_0	Dimension-less	Probability of all links to be simultaneously operational.

3.4 Output

Pressure at a node	m	Value of the pressure computed at a given node.
Actual consumption	l/s	Actual water quantity consumed at a given node with respect to the available pressure
Flow	l/s	Flow of the link
Satisfaction Rate	Dimension-less	Actual supply (l/s) divided by the demand
SR_{ij}	Dimension-less	Satisfaction Rate at node i when link j is under repair (event of probability pe_j).
SR_{i0}	Dimension-less	Satisfaction Rate at node i when none of the links is under repair.
SRN_i	Dimension-less	Weighted (with weights pe_j) Mean Satisfaction Rate at node i when one or none of the links of the network is under repair.
SRP_j	Dimension-less	Weighted (with weights w_i*d_i) Mean Satisfaction Rate over all nodes when link j is under repair.
SRP_0	Dimension-less	Weighted (with weights w_i*d_i) Mean Satisfaction Rate over all nodes when none of the links is under repair.
Global Satisfaction Rate of the system GSR	Dimension-less	Overall reliability of the network defined as the weighted mean of SRP_j (with weights pe_j), or equivalently of SRN_i (with weights w_i*d_i).

3.5 PI(s) Estimation Method

3.6 PI(s) Forecasting Method

4 POSSIBLE IMPROVEMENTS OF THE MODEL

Possible improvements can be defined at different viewpoints:

- programming level:

The model exists now in a mathematical language linked to the software "Matlab". When it will applied and validated, it will be implemented in the software "Porteau" that models the hydraulic functioning of drinking water networks.

It shorter run, it will be written to be used more automatically than now.

- Hydraulic level:

The hydraulic modeling has to be validated on different networks. It is now made on French networks.

- Reliability level:

About the definition of the indices, some improvements can be done:

- the assumption of "1 simultaneous break" is tested,
- the definition of nodes weight has to be more accurate. The influence of the weights has to be tested, especially for the assessment of "pipe reliability index" and global index"
- the influence of breaks probabilities will be tested. A method of determination of these probabilities will be proposed, according to data availability and existence of failures forecasting models.

5 SOFTWARE SPECIFICATIONS

5.1 Programming Language(s) or Mathematical-Statistical Software(s)

The software is MATLAB, that is a mathematical software with a specific language. This language is really close to the language C.

6 REFERENCES

6.1 Theoretical Framework References

Andreou, S. A. (1996). "Predictive models for pipe break failures and their applications on maintenance planning strategies", *Massachusetts Institute of Technology,* 291 p.

Collins M., Cooper L., Helgason L., Kennington J. and Leblanc L. (1978). “Solving the pipe network analysis problem using optimization techniques”, *Management Science*, Vol. 24, n° 7, p. 747-760, March.

Cross Hardy (1936), “Analysis of Flow in Networks of Conduits or Conductors”, *Bulletin n° 286*, Univ. of Illinois Engr. Experiment Station.

Damelin, E., Shamir, U., Arad, N. (1972). "Engineering and Economic Evaluation of the Reliability of Water Supply", *Water Resources Research*, 4(8), 861-877.

Eisenbeis, P. (1994) "Modélisation statistique de la prévision des défaillances sur les conduites d’eau potable". *Thèse*, Spécialité Sciences de l’eau, Université Louis Pasteur, Strasbourg, France, 156 p.

Fujiwara, O., De Silva, A. U. (1990). "Algorithm for Reliability-Based Optimal Design Of Water Networks", *Journal of Environmental Engineering*, 116(3), 575-587.

Goulter, I. C.and Bouchart, F. (1990). "Reliability-Constrained Pipe Network Model". *Journal of Hydraulic Engineering*, 116(2), 211-229.

Goulter, I. C.and Coals, A. V. (1986). "Quantitative Approaches to Reliability Assessment in Pipe Networks", *Journal of Transportation Engineering*, 112(3), 297-301.

Gupta, R.and Bhave, P. R. (1994). "Reliability Analysis of Water-Distribution Systems". *Journal of environnmental engineering*, 120(2), 447-460.

Hobbs, B. F.and Beim, G. K. (1988) " Analytical Simulation of Water System Capacity Reliability : 2. A Markov Chain Approach and Verification of the Models". *Water Resources Research*, 24(9), 1445-1458.

Hobbs, B. F.and Beim, G. K. (1988) "Analytical Simulation of Water System Capacity Reliability : 1. Modified Frequency-Duration Analysis". *Water Resources Research*, 24(9), 1431-1444.

Jacobs, P.and Goulter, I. (1991). "Estimation of Maximum Cut-Set Size For Water Network Failure". *Journal of Water Resources Planning and Management*, 117(5), p.588-605.

Jowitt, P. W., Xu, C. (1993)."Predicting Pipe Failure Effects in Water Distribution Networks". *Journal of Water Resources Planning and Management*, 119(1), 18-31.

Kalungi P., Tanyimboh T.T., (2000)"An efficient algorithm for the determination of head-dependent consumption for water distribution systems", *2nd international conference on decision making in urban and civil engineering, Lyon, 20-22 November 2000*

Le Gat, Y., Eisenbeis, P., (2000) "Using maintenance records to forecast failures in water networks", *Urban water,* 3(2), pp. 173-181.

Mays, L. W. (1989). *Reliability Analysis of Water Distribution Systems*. New York (U.S.A.) : American Society of Civil Engineers, 532 p.

Piller O. (1995). “Modeling the behavior of a network - hydraulic Analysis and a sampling procedure for estimating the parameters”, *thesis from the University of Bordeaux,* defended on 03 February 1995, 288 pages (in French).

Rosenthal, A., (1981). "Series-Parallel Reduction for Difficult Measures of Network Reliability", *Networks*, 11, 323-334.

Wagner, J. M., Shamir, U., Marks, D. H. (1988a). "Water Distribution Reliability : Analytical Methods". *Journal of Water Resources Planning and Management*, 114(3), 253-275.

Wagner, J. M., Shamir, U., Marks, D. H. (1988b). "Water Distribution Reliability: Simulations Methods". *Journal of Water Resources Planning and Management*, 114(3), p. 276-294.

6.2 Practical Use and Results References

Berthin, S. (1994). "Water distribution system reliability indices", DEA Report, Water Science, Louis Pasteur University, Strasbourg, France.

Brémond B, Berthin S (2001). "Reliability of a drinking water supply system", *Iwa Conference "System approach to leakage control and water distribution systems management" Brno, 16-18 May 2001*

Brémond, B. Piller, O., Eisenbeis, P. (2001). " water supply network modeling including head dependent demand and elevated nodes", *IWA conference, 16-18 May, Brno*.

APPENDIX 8 : RELNET

1 GENERAL DESCRIPTION

1.1 Name and/or Acronym of the Model

RelNet

1.2 Company/Research Center/University

Brno University of Technology

Faculty of Civil Engineering

Institute of Municipal Water Management

Žižkova 17, 602 00, Brno, Czech Republic

1.3 Objectives

The aim of this model is to assess service reliability of each node and consequently the total reliability of the network. Reliability of the water distribution network depends on reliability of network elements (nodes and pipe sections). The model outputs are :

1) node reliability
2) total reliability of the network (pressure zone)
3) impact of each pipe section on total reliability of the network (pressure zone)

1.4 *1.4* Functional description

The model is based on theory of reliability. In this case, the reliability means immediate reliability in particular time step. Reliability is based on required pressure in each node of the network and the model simulates random network load state (topology, demand, selected physical parameters – roughness etc.. For each generated random status of the network, the hydraulic analysis of flows and pressure for each node is realized. After statistical data processing, we receive following curves for each node :

- probability density function
- cumulative distribution function

- probability function

1.5 Brief Historical Overview of the Model

The first theoretical steps of **RelNet** has been published in [1] (see chapter 6.1.). This approach was tested and developed also in the diploma projects in years 1986 and 1993. Mr.Pavel Viščor started development of current model in 1997 as a part of his PhD thesis. He tested the model on the pressure zone Lesná (node reliability, impact of each pipe section on total reliability of pressure zone). The results of his case study were presented at the 13th Junior scientist workshop in Dresden [4] (see chapter 6.1) . Mr. Pavel Dvořák collected failure data of pipe sections in another pressure zone of the Brno water distribution network and realized statistical analysis of failures including the impact of each pipe section on total reliability of pressure zone, as a part of his PhD thesis.

We have just started to develop the compact software package of the model. We use the ODULA software package (MIKE NET) for hydraulic analysis and Excel for statistical processing. We plan to cooperate with DHI Hydroinform, developer of ODULA software, on implementation of our model into the ODULA software package.

2 THEORETICAL FRAMEWORK OVERVIEW/PAST STUDIES

2.1 Scientific background

The RelNet model is based on stochastic principle using the Monte-Carlo method. Selected parameters of the network are generated pursuant to the adequate probability density functions. Pipe sections are eliminated from the network structure according to the failure rate of each pipe section. The load states and the topology of the network are prepared as outputs of the random data processing. Afterwards the hydraulic analysis of each state is realized and the results are recorded. As a correctly supplied node of the network is considered the node, in which the calculated pressure is greater than the required pressure. The probability density function, the cumulative distribution function and the node reliability is evaluated for each node of the network.

2.2 Algorithm of the model

A) Total reliability of the network

1. Data collection and processing

(water network topology - GIS, statistical data processing – water consumption, failure rates etc.)

2. Generation of pseudo-random load states

3. Hydraulic analysis of generated load states (repeating in loop)

4. Statistical processing of hydraulic analysis results

5. Calculation of the node reliability for each node

6. Calculation of the total reliability of the network

B) Impact of each pipe section on total reliability of the network

1. Data preparation

(water network topology - GIS, statistical data processing – water consumption)

2. Generation of network states. Each state corresponds to one eliminated pipe section out of service.

3. Hydraulic analysis of generated load states

4. Processing of hydraulic analysis results

5. Evaluation of the impact of each pipe section on total reliability of the network

2.3 Past studies and conclusions

The **RelNet** model has been tested on the pressure zone Lesná, which is a part of the Brno water distribution network. This pressure zone includes 165 pipe sections and 155 nodes. One thousand of pseudo random load states of the pressure zone has been realized. The values of the node reliability and the impact of each pipe section on the total reliability of the pressure zone has been calculated.

The impact of each pipe section on the total reliability of the network has been tested also in the other pressure zone of the Brno water distribution network. This pressure zone includes 167 pipe sections and 127 nodes.

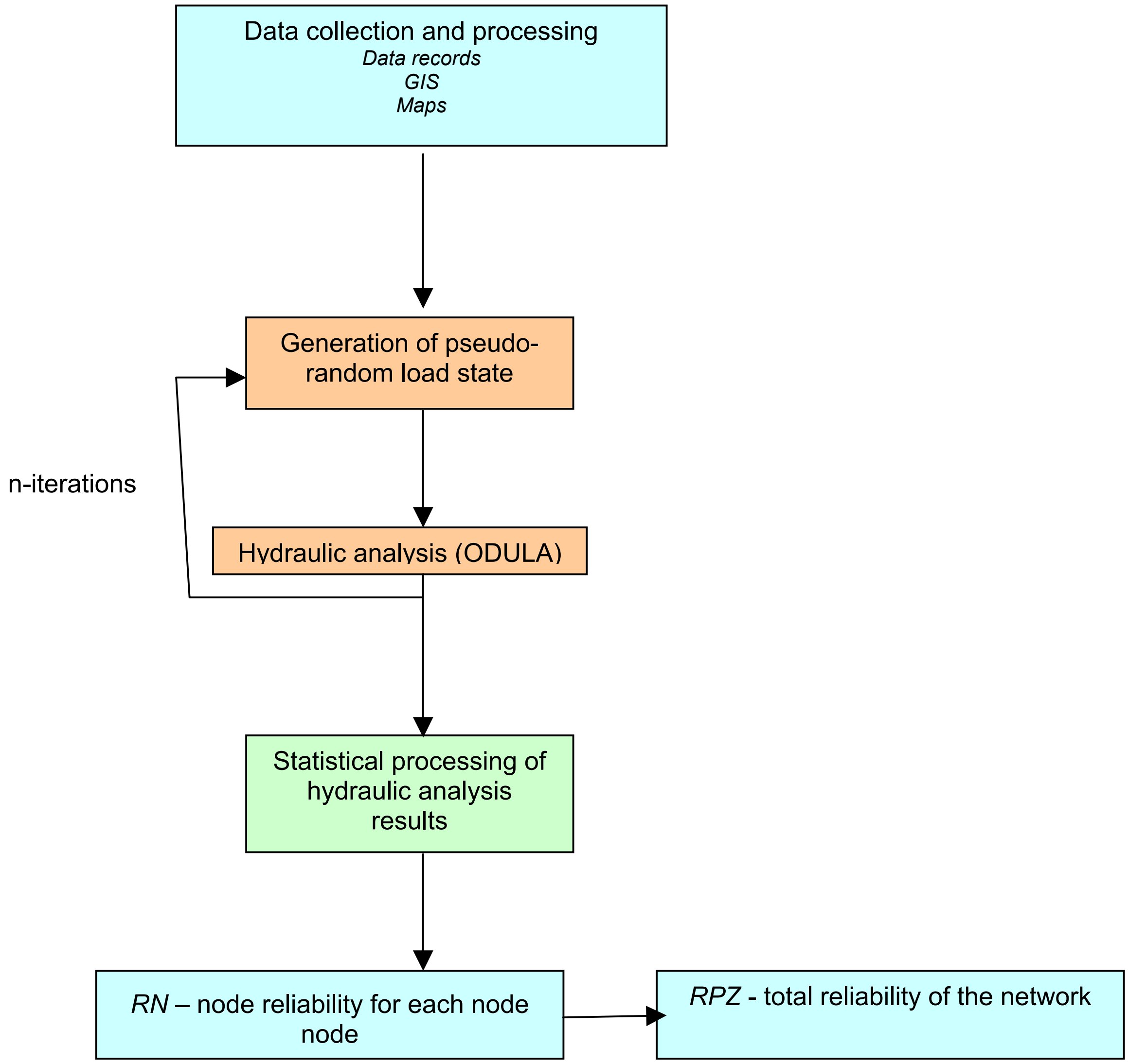
Both case studies has proved that the **RelNet** model is applicable for evaluation of the reliability of water networks.

The theory of the model is still developed. We would like to change the criteria of the node reliability. The number of supplied consumers or the number of supplied properties (flats) will be tested instead of pressure. We don’t suppose to apply the water quality approach for the calculation of the node reliability at this moment.

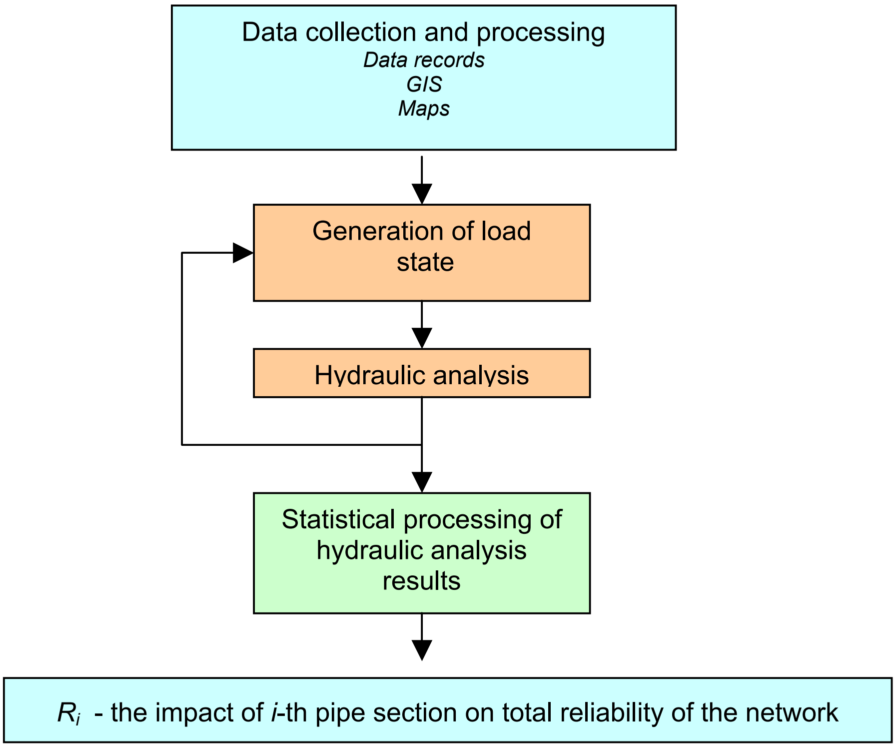
3 SPECIFICATIONS OF COMPUTATIONAL STEPS

3.1 Functional Scheme

A) Total reliability of the network



B) Impact of each pipe section on total reliability of the network



3.2 Raw Data Formatting

The crucial problem is the data unavailability for failure prediction. Hydraulics software ODULA is based on EPANET computing core. ODULA reads ASCII formatted files. ODULA can export data using ODBC. The basic data structure for hydraulic analysis including pipe section availability (output from failure forecasting model) is needed for the model.

3.3 List and Definition of Explanatory Factors

3.3.1 Required factor

Input variables

Variable	Unit	Short description
Pipe material	-	Code of the pipe material according to a convention
Link_ID	-	Code of single pipe section
NB	-	Code of node 1 connected to the Link_ID
NE	-	Code of node 2 connected to the Link_ID
Lenght	m	Length of Link_ID
DN	mm	Nominal diameter of the Link_ID
k	mm	Roughness coefficient of the material used in the Link_ID
c1	-	Number of units supplied (e.g. flats, persons, etc.)
c2	-	Specific consumption per unit per day
PZ	-	Pressure zone
TC	l/s	Total max. consumption in the evelauated pressure zone
NT	-	Code of tank node
TNL	m	Tank node water level
NL	m	Node elevation
PN	m	Requested min. hydrodynamical pressure in the node
APS	-	The Link_ID availability (probability that the pipe section is in service function)
TPC1	-	Type of probability curve for total water consumption
TPC2	-	Type of probability curve for roughness
TPC3	-	Type of probability curve for tank water level
TPC4	-	Type of probability curve for node consumption
n	-	Number of hydraulic analysis
And other parameters necessary for hydraulic modelling (pumps, valves, PRV, etc.)		

Output variables

Variable	Unit	Short description
Expected life-time	years	Prediction of the pipe lifetime (50% probability)
RN	-	Node realibility
RPZ	-	Total realibility of the pressure zone
Ri	-	Impact of i-pipe section on total realibility of the pressure zone

3.3.2 Highly recommended factors

3.3.3 Possibly Useful factor

- number of failures per year for each pipe section

- average duration of the failure

3.4 Model Parameters Estimation or Assignation

1) The calibration of the hydraulic model must be realized by measurement of the pressure (flows) directly on the network.

2) The calculated pressure is distributed to frequency rates for each node.

3) The probability density curve for each node is the result of this process.

3.5 Output

3.6 Model Validation or calibration

(Statistical model)

3.6.1 Check of Parameters Significance – Internal Validation (Statistical model)

The verification of the impact of each pipe section on total reliability of the network is difficult. It's practically impossible to close the most important pipe sections on the real network and realize the measurement of the pressure.

The other problem is the number of generated load states (iterations). This number depends on the size of the pressure zone (number of nodes). There is not still exact method to evaluate the number of iterations. We tried 1000 iterations on tested pressure zones.

3.6.2 External Validation or Cross Validation – Statistical test(s)

(Deterministic/hydraulic model and physical model)

3.6.1 Method for best calibration of the model

3.7 PI(s) Estimation Method

Depends on estimation of all 3 types outputs considering particular time horizon.

3.8 PI(s) Forecasting Method

Depends on estimation of future water consumption and estimation of pipe section availability (output of failure forecasting model).

4 POSSIBLE IMPROVEMENTS OF THE MODEL

We develop the complex software package for this model. The main task is the direct link and data processing among hydraulic software ODULA and statistical software Excel. We plan to cooperate with DHI Hydroinform, the producer of ODULA software, on implementing our model to their software package. We consider different approaches to estimate node reliability (number of supplied consumers). We also consider different approaches in estimation of the total reliability of the network.

5 SOFTWARE SPECIFICATIONS

5.1 Programming Language(s) or Mathematical-Statistical Software(s)

Hydraulics modeling – ODULA by DHI Hydroinform (www.dhi.cz) based on EPANET computing core

Statistical software – MS Excel (Office 2000 package) by Microsoft

5.2 Possible Input File(s) Formats

ASCII files as file input of ODULA , otherwise ODULA can exchange data between other programs under MS Windows by ODBC ,all programs capable of ODBC can exchange data between ODLULA and other software running under MS Windows.

.XLS and .TXT files as main file input of Excel and all other format of which is Excel capable.

5.3 Possible Output File(s) Formats

.XLS as output of Excel and all other format of which is Excel capable

ODULA - data exchange between other programs under MS Windows is provided by ODBC, all programs capable of ODBC can exchange data between ODULA and other software running under MS Windows.

6 REFERENCES

6.1 Theoretical Framework References

[1] Votruba, L., Heřman, J. and others: *Spolehlivost vodohospodářských děl*, Česká matice technická, Praha, SNTL, Praha, 1993 (in Czech)

[2] Bao, Y., Mays, L. W.: *Model for Water Distribution System Reliability*, Journal of Hydraulic Engineering, September 1990

[3] Viščor, P., Tuhovčák, L.: *Spolehlivost vodovodních sítí*, Proceedings of the II. International conference VODA Zlín 98, Zlin, p. 97-102 (in Czech)

[4] Viščor, P.: *Management of Water Distribution Systems and Leakage Control*, Proceedings of the 13th European Junior Scientist Workshop on Service-Life Management

6.2 Practical Use and Results References

[5] Viščor, P.: *Spolehlivost vodovodních sítí*, Diploma project, Brno University of Technology, Brno 1997, (in Czech)

[6] Tuhovčák, L., Viščor, P., Valkovič, P., Raclavský, J.: *Expertní systém jako podpůrný prostředek pro plánování rekonstrukcí a sanací vodovodních řadů*, Proceedings from 6. International conference of CzSTT, Brno, 2000 (in Czech)

[7] Dvořák, P.: *Vliv poruch rozváděcích řadů na spolehlivost vodovodních sítí*, PhD Thesis, Brno University of Technology, Brno, 2000 (in Czech)

[8] Tuhovčák, L., Viščor, P., Valkovič, P.: *Analysis of failures and reliability of the water distribution network as criteria for water mains rehabilitation planning*, Proceedings of the IWA specialised conference - System approach to leakage control and water distribution systems management, Brno, 2001 (in English)

SINTEF The Foundation of Scientific and Industrial Research, Trondheim, Norway
Norwegian University of Science and Technology, Trondheim, Norway
Brno University of Technology, Brno, Czech Republic
Cemagref, Bordeaux, France
Dresden University of Technology, Dresden, Germany
INSA, Lyon, France
LNEC, Lisbon, Portugal
WRc, Swindon, UK
University of Bologna, Bologna, Italy
University of Ferrara, Ferrara, Italy
AGAC, Reggio Emilia, Italy

Institute name: *Cemagref*
Contact person: Patrick Eisenbeis
E-mail: carew@bordeaux.cemagref.fr
www.cemagref.fr/English/more-infos/water-waste.htm

Address: 50, avenue de Verdun
33612 CESTAS Cedex, France
Phone: (+33) 557 89 08 33
Fax: (+33) 557 89 08 01
Work Package: 2 Task no.: 2.1