Mean values of water pipe breakage rates around the world and in geographical areas

F. Dell’Orfano, V. Esposito, P. Gualtieri and G. Pulci Doria

ABSTRACT

One of the most frequently used parameters in water supply systems is the pipe breakage rate, which is often expressed by $\lambda$. Its definition can be deterministic or probabilistic, but both definitions lead to the same value under certain assumptions. This parameter can be defined per unit time only or per unit time and per unit length of the pipes (using a kilometre as the unit of measurement). The symbol $\lambda$ is used in both cases and can occasionally generate confusion. In a water supply system, most failures occur because of pipe breakage, and only those cases are considered in this work. For a pipe, $\lambda$ can be evaluated in terms of time (years) and length (km) and expressed by $\Lambda$. This paper aims to investigate the mechanical reliability of a water supply system with reference to pipeline breakages worldwide. A statistical approach is applied to a wide database of $\Lambda$ values relative to certain variables (e.g., the diameter, material, age, pressure, and chemical characteristics of the water) for water supply systems worldwide (over 3,500 data points were collected from approximately 200 papers). The pipe length $L$ and monitoring period $T$ are often reported in the database. For each water supply system, it is necessary to apply a statistical weight to each value of $\Lambda$ because the considered variables are notably different. The chosen weight is the product of $L \cdot T$; the weight is evaluated using statistical analysis if $L$ and/or $T$ are unknown. Finally, a particular treatment is applied to the obtained weights to eliminate distortions in the $\Lambda$ evaluation because of the different socioeconomic conditions of more or less developed countries. Four results are reported in this work: (1) the weighted average value of $\Lambda$ in the world ($\bar{\Lambda}_m$), (2) the average values of $\Lambda$ in different continents ($\bar{\Lambda}_{mc}$), (3) the average values of $\Lambda$ in different countries ($\bar{\Lambda}_{mn}$), and (4) the management correction factor ($f_{CE}$).

Key words | credibility, pipeline breakage rate, statistical approach, water supply systems

INTRODUCTION

The key elements to quantify the effectiveness of a water supply system are its reliability, resilience, and vulnerability (Hashimoto et al. 1982a, b). The reliability can be expressed as a function of four indicators: efficiency, adequacy, performance, and environmental compatibility (Alegre et al. 2012). The reliability is divided into environmental, hydraulic, mechanical, and sanitary components (Khomsi et al. 1996; Tanyimboh et al. 2001). Within the mechanical reliability, we refer to the following structural parameters: (1) supply interruptions and (2) pipeline breakages.

In the literature, the theoretical studies on mechanical reliability are often based on databases with reduced consistency (Bao & Mays 1990; Duan et al. 1990; Goulter & Bouchard 1990; Quimpo & Shamsi 1991; Bertola & Salandin 1992; Cullinane et al. 1992; Gupta & Bhave 1996).

In this context, the objective of this study is to investigate the mechanical reliability starting from an overall comprehensive analysis of water supply systems that have been described in the technical literature and whose data refer exclusively to pipeline breakages (Rajani et al. 1996; Stone et al. 2000; Boxall et al. 2007; Hu & Hubble 2007;
This study focuses on pipelines because they are the main components of water supply systems; therefore, their failures highly affect the statistical reliability of the system.

We only refer to pipe breakages, whereas the literature often refers to a broader concept of pipeline failure. In the literature, failure is often referred to as circumferential break, longitudinal break, explosion, hole, displacement joint, ovality, spiral break, split bell, reported breaks, and detected leakage. Breakages are the dominant phenomenon among all failures. In particular, a significant percentage of failures occur because of the type of breakage that occurs. Every breakage is also a failure but not vice versa.

Large, detailed datasets from several areas of the world have been examined. The analysis indicates that these datasets are often affected by temporal and geographical limitations and occasionally by the repetition of information, which has been included in previous calculations. Therefore, suitable weights have been introduced to eliminate the distortion resulting from the different numbers of monitored water supply systems in more or less developed countries. This study provides an average breakage rate, which is expected by many variables and can be used as a reference for water supply systems in the world, in a continent, and in a country. In addition to providing a deep theoretical background, such knowledge can improve the decision-making criteria for water supply system management and ensure standards of efficiency, effectiveness, and economy (Alegre et al. 2006, 2011; Alegre & Almeida 2009; ASCE 2009).

**PIPE BREAKAGE RATE**

Measurement is typically based on scientific analysis and included in every scientific approach. The management of water distribution systems requires accurate quantification of the network deterioration (AWWARF 2000; AWWA 2001; Kleiner & Rajani 2001; AWWSC 2002). However, the direct inspection of a distribution network is often excessively time consuming and expensive. Furthermore, the application of physical models to assess the structural resilience of each pipe is unrealistic and inaccurate in most cases because data are rarely available and expensive to obtain. In this regard, the use of statistical methods to identify breakages is an efficient and economical alternative to directly measure the deterioration of the pipes.

In the literature, the frequency of breakage is expressed by a discrete (number of breakages) or continuous perspective (breakage rate). The definition of the breakage rate can be (a) deterministic or (b) probabilistic (i.e., expressed in terms of determination or probability):

(a) The deterministic definition is the average number of breakages per unit time (one year), whereas the probabilistic definition is the probability that the breakage occurs in the next unit time (1 year). Under certain assumptions, the two definitions lead to the same quantity. Thus, breakage is occasionally defined per unit time but often both per unit time and per unit pipe length (using a kilometre as the unit of measurement). In both cases, the symbol \( \lambda \) is used, which often generates confusion. In this work, the pipe breakage rate is assessed per unit time (year) and unit length (km) and expressed by the symbol \( \Lambda \). The equation to calculate a deterministic value of \( \Lambda \) is

\[
\Lambda = \frac{r}{L \cdot T}
\]

where \( r \) is the number of breakages, \( L \) is the length of the network or the pipe, and \( T \) is the monitoring period. A review of the literature demonstrates that the breakage rate is used in many models that were proposed to optimise water pipe rehabilitation/replacement (Kaara 1984; Smith 1994; Kleiner & Rajani 1999). The breakage rates that were obtained from different studies are quantified as follows. The maximum \( \Lambda \) value in the United States is \( 16.7 \times 10^{-2} \text{ km}^{-1} \text{ year}^{-1} \), which is significantly lower than the rates reported in Australia \((40 \times 10^{-2} \text{ km}^{-1} \text{ year}^{-1})\) and the United Kingdom \((18.8 \times 10^{-2} \text{ km}^{-1} \text{ year}^{-1})\). According to McDonald et al. (1994), reports of 40 breakages per \( 10^2 \text{ km} \) per year are considered high and indicate a poor network condition. Networks with ratios between 20 and 39 are considered in acceptable condition, whereas ratios of less than 20 indicate that the network is in good condition. The published values of \( \Lambda \) in the literature range from a minimum of \( 2.24 \times 10^{-2} \text{ km}^{-1} \text{ year}^{-1} \) to a maximum of \( 4.70 \times 10^{-1} \text{ km}^{-1} \text{ year}^{-1} \) (O’Day 1982; Su et al. 1987;
Cullinane et al. 1989; Guercio et al. 1995). According to Pelletier et al. (2003), it can be assumed that a network is in good condition when $\Lambda \leq 0.2$, in acceptable condition when $0.2 < \Lambda < 0.4$, and in poor condition when $\Lambda \geq 0.4$. According to Sundahl (1997), the breakage rate is low when $\Lambda \leq 0.08$, normal when $0.08 < \Lambda < 0.16$, high when $0.16 < \Lambda < 0.28$, and extremely high when $\Lambda \geq 0.28$.

(b) The probabilistic definition refers to the concept of probability of breakage $p$. This concept considers a pipe without breakages at the instant $t = 0$. In particular, $p_1$ is the probability that a breakage occurs in the time between 0 and 1, and $p(t)$ is the probability that a breakage occurs in the time between 0 and $t$. For comparison purposes, we consider one year as the unit of time.

An analysis based on the theory of probability (Brunone et al. 2009) indicates that $p(t)$ can be derived from the previously defined breakage rate, which specifically results in

$$p(t) = 1 - e^{-\beta t}$$  \hspace{1cm} (2)

$$p_1 = 1 - e^{-\beta}$$  \hspace{1cm} (3)

$$\beta = \Lambda \cdot L$$  \hspace{1cm} (4)

where for the considered pipe, $L$ is the length, $\beta$ is the expected number of breakages per year, and $\Lambda$ is the expected number of breakages per year and per unit length.

There are differences in the practical uses of these two approaches.

The quantity $\Lambda$ can be applied with greater generality to an extremely short or extremely long pipe that is homogeneous along its length. Two pipes can be considered equal if they have identical $\Lambda$ values. $\Lambda$ can also be applied to a network even if it is heterogeneous, in which case it becomes the average $\Lambda$ but continues to indicate the expected average number of breakages during a year if it is multiplied by the total length of the network.

In any case, $\Lambda$ is a powerful and reliable parameter that an operator can use to perform technical and economic planning for future managerial actions. The quantity $p(t)$ can be applied only to a pipe that is not overly long and has certain characteristics. If the pipe is extremely long, $p$ ($t$) rapidly approaches unity (but always remains lower) and loses all predictive significance.

When referring to a network and its average $\Lambda$ value, the probabilistic approach is faster than the deterministic approach. The management of a water supply network can be planned by applying Equation (2) to each pipe. As a result, both approaches are discussed in the technical literature and are implemented in the management of water supply systems when appropriate.

THE DATABASE AND ITS MAIN CHARACTERISTICS

In this section, the logic, characteristics, and peculiarities of the collected data are described. They refer to different water supply systems in the world (≈200 publications consulted) and are collected in a database that consists of 3,656 lines and 35 columns with spatial and temporal information, which has never been proposed in the literature before. The output parameter of the database is the breakage rate $\Lambda$.

The complexity of the database is linked to a level of aggregation/disaggregation of data from a spatial and temporal perspective. More than a century of observations has been examined (1875–2011). This database is the first to link different worldwide water supply systems and thus allow for an in-depth analysis of these systems (Table 1).

The study of $\Lambda$ values for several water supply systems worldwide allows us to obtain some preliminary results.

First, the most important variables for describing the deterioration of a water supply system can be grouped into four categories (Aslani 2003): structural conditions of pipelines (1), environmental external conditions (2), hydraulic internal conditions (3), and operational maintenance conditions (4).

Eighteen variables have been identified and classified in the literature. They are reported in Table 2.

Moreover, the output parameter $\Lambda$ can be divided into four types:

Type 1 (dominant): The $\Lambda$ value with the corresponding $L$ and $T$ values is correlated to one or more of the considered variables.

Type 2: The $\Lambda$ value without the corresponding $L$ and/or $T$ values is correlated to one or more of the considered variables.
variables and can be used with those from Type 1 after an appropriate statistical evaluation of $L$ and/or $T$.

Type 3: The $\Lambda$ value with the corresponding $L$ and/or $T$ values is not correlated to any considered variables and can be used with those of Types 1, 2, and 4 to build Model <zero> to obtain an average worldwide value of $\Lambda$.

Type 4: The $L$ and/or $T$ values are never correlated to the corresponding $\Lambda$ values and are occasionally correlated to one or more of the considered variables.

This research provides, for the first time, an exhaustive list of variables that affect $\Lambda$ and a robust and significant database of $\Lambda$ values for different water supply systems and monitoring periods throughout the world.

**ASSIGNMENT OF WEIGHTS AND ESTIMATION PROBLEMS**

The $\Lambda$ value for an extremely large water supply system that is monitored over many years is more valuable than a $\Lambda$ value that is evaluated for a small water supply system that is monitored over only a few years. The product $L \cdot T$ is used as the weight $w$.

The use of an unweighted $\Lambda$ value determines significantly different results. For example, the $\Lambda$ values can become an order of magnitude higher than the weighted value (Dell'Orfano 2013a).

In some cases, the required elements to evaluate the weights are entirely or partially unknown.

---

**Table 1 | Geographical areas in the database**

<table>
<thead>
<tr>
<th>Continent</th>
<th>Nation/State</th>
<th>Town</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oceania</td>
<td>Australia.</td>
<td>Canberra, Melbourne, Sydney.</td>
</tr>
<tr>
<td>Asia</td>
<td>South Korea, U.A.E., Iran, Kurdistan, Russia, Taiwan, Georgia, Kazakhstan, Republic of Kyrgyz, Malaysia, Ukraine, Moldova, Tajikistan, Republic of Azerbaijan.</td>
<td>Seoul, Abu Dhabi, Sananday, Behshahr, Sari, Ramsar, Moscow, Luodong.</td>
</tr>
</tbody>
</table>
In these cases, the weight evaluation can be divided into two sub-phases: (1) calculation or estimation of the product $L \cdot T$ at the network scale, which considers a network whose $L$ and $T$ values are partially or not at all known, and (2) calculation or estimation of the product $L \cdot T$ at the pipe scale, which occurs when the product is known (or has been attributed via statistical analysis) at the network scale, but the breakage rates are disaggregated into functions with different variables.

In the first case, it is necessary to have at least a statistical knowledge of the monitored $L$ and $T$ in relation to some variables, specifically the diameter $D$, material $M$, and age $A$. In the database, data without $\Lambda$ are used to estimate $w$ if information on $L$, $A$, or $D$ is provided.

Three cases can arise in the estimation of $L \cdot T$ at either the network scale or pipe scale: (1) estimation of $L$ when $T$ is known, (2) estimation of $T$ when $L$ is known, and (3) estimation of both. For the estimation of $L$, it is necessary to split the calculation into two phases that correspond to the network scale and/or pipe scale. For the estimation of $T$, the calculation is identical for both scales, which implies that $T$ is independent of the characteristics of the considered system (network or pipe). Now, we can provide the

<table>
<thead>
<tr>
<th>Number</th>
<th>Category</th>
<th>Variable</th>
<th>Symbol</th>
<th>Unit of measurement</th>
<th>Investigated range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1)</td>
<td>Diameter</td>
<td>D</td>
<td>mm</td>
<td>[20 : 1,350]</td>
</tr>
<tr>
<td>2</td>
<td>(1)</td>
<td>Age</td>
<td>A</td>
<td>years</td>
<td>[0 : 147.5]</td>
</tr>
<tr>
<td>3</td>
<td>(1)</td>
<td>Material</td>
<td>M (I)</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>(1)</td>
<td>Pressure(a)</td>
<td>P</td>
<td>m</td>
<td>[12.5 : 95]</td>
</tr>
<tr>
<td>5</td>
<td>(2)</td>
<td>Soil type</td>
<td>T (II)</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>(2)</td>
<td>DIPRA</td>
<td>S (III)</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>(2)</td>
<td>Rainfall level</td>
<td>h</td>
<td>mm</td>
<td>[0 : 270]</td>
</tr>
<tr>
<td>8</td>
<td>(2)</td>
<td>Air temperature</td>
<td>$t_a$</td>
<td>°C</td>
<td>[0.5 : 27]</td>
</tr>
<tr>
<td>9</td>
<td>(2)</td>
<td>Freezing index$_2$</td>
<td>$F_2$</td>
<td>(IV)</td>
<td>[0 : 28]</td>
</tr>
<tr>
<td>10</td>
<td>(2)</td>
<td>Traffic</td>
<td>$T_r$</td>
<td>(V)</td>
<td>–</td>
</tr>
<tr>
<td>11</td>
<td>(2)</td>
<td>Humidity(b)</td>
<td>U</td>
<td>mm</td>
<td>[–20 : 325]</td>
</tr>
<tr>
<td>12</td>
<td>(2)</td>
<td>Freezing index$_1$</td>
<td>$F_1$</td>
<td>°C</td>
<td>[1,100 : 2,200]</td>
</tr>
<tr>
<td>13</td>
<td>(3)</td>
<td>Index of aggression</td>
<td>$I_A$</td>
<td>(VI)</td>
<td>[8.3 : 12.6]</td>
</tr>
<tr>
<td>14</td>
<td>(3)</td>
<td>Sulphate content</td>
<td>$S_r$</td>
<td>mg/l</td>
<td>[0 : 255]</td>
</tr>
<tr>
<td>15</td>
<td>(3)</td>
<td>Hardness</td>
<td>H</td>
<td>mg/l</td>
<td>[5 : 295]</td>
</tr>
<tr>
<td>16</td>
<td>(3)</td>
<td>Alkalinity</td>
<td>$Al_c$</td>
<td>mg/l</td>
<td>[5 : 138]</td>
</tr>
<tr>
<td>17</td>
<td>(4)</td>
<td>Depth laying</td>
<td>d</td>
<td>m</td>
<td>[0.54 : 2.37]</td>
</tr>
<tr>
<td>18</td>
<td>(4)</td>
<td>Bed laying</td>
<td>$L_p$</td>
<td>(VII)</td>
<td>–</td>
</tr>
</tbody>
</table>

**Table 2 | Considered variables**

Predictor Variable column:

- **(a)** $P$ is the mean pressure of the water supply system.
- **(b)** $U$ is an expression of the moisture.
- **(c)** $F_1$ is the accumulated average daily temperature below 0 °C for 5 months (November to March).

Unit of measurement:

- **(I)** There are nine types of material $M$ (steel, cast iron, ductile iron, PVC, polyethylene, pead (a high density polyethylene), prestressed concrete, fibre cement, and lead). There are also intermediate types of aggregate materials (e.g., grey cast-ductile iron) and unknown types.
- **(II)** There are 10 types/definitions of soil $T$ (clay and sand, clay, expanded, slightly expanded, silt and clay, very expanded, sandy gravel and clay, sandy-loam, stable, and urban terrain).
- **(III)** DIPRA is a group of four variables: soil pH, redox potential (mV), resistivity (Ω•cm), and sulfides (mV). The four parameters are translated into scores according to an international reference (ANSI/AWWA C-105/A21.5 1999).
- **(IV)** $F_2$ is the number of days in a month when the soil temperature is less than or equal to – 1 °C.
- **(V)** There are two descriptions of the traffic, which leads to 6 types of $T_r$: (low load, high load, nil, moderate, ordinary, and intense).
- **(VI)** AWWA (2003a, b).
- **(VII)** There are two types of bed laying (sand and yellow sand).
In this method, there are 3 sub-cases: (1) lack of data, (2) lack of T data, and (3) lack of L·T product data.

Method 1: sub-case 1. A distinction should be noted between a water supply system and utility. A water supply system refers to a country or a city, whereas a utility refers to a set of water supply systems with a single management. In the database, the average value of the length of a water supply system \( L^a \) is calculated to be 1,542.48 km. When the length of a water supply system is unknown, \( L^a \) is considered as the reference value; when the length of the utility is unknown, \( L^a \) multiplied by the number of water supply systems of the utility is considered as the reference value. In either case, the final expression is

\[
L_S = (L_{sv})^{(N_{LSV})(\mu_{LSV} \cdot c_{LSV})^{(N_{LS})}(\mu_{c} \cdot c_{c})^{(N_{c})}(\mu_{U} \cdot c_{U})^{(N_{U})}}
\]

where \( L_S \) is the length (true or estimated) of the water supply system; \( L_{sv} \) is the true length of the system (if known); \( N_{LSV} \) is a factor of binary control of the true lengths, which is 1 when the network length is known and 0 when it is not; \( \mu_{LSV} = L^a; \) \( c_{LSV} \) is the credibility associated with the \( \mu_{LSV} \), which depends on the greater or lesser dispersion of the true lengths around their average value and is taken as 0.2670; \( N_{LS} \) is a factor of binary control of the estimated lengths, which is 1 when the network length is known and 0 when it is not; \( \mu_{c} \) is the arithmetic average of the ratios of the lengths with respect to the total length of the water supply system, which becomes relevant if only a part of it has been investigated, and its obtained value from a statistical analysis is 0.5750; \( c_{c} \) is the credibility associated with \( \mu_{c} \), which depends on the dispersion of the data used for the calculation of \( \mu_{c} \) and is equal to 0.5428; \( N_{c} \) is a factor of binary control because of a partial analysis of the water supply system, which is 1 if only a part of it has been investigated and 0 otherwise; \( \mu_{U} \) is the arithmetic average of the number of water supply systems of a utility, which becomes relevant if the real number is unknown and is equal to 16.25; \( c_{U} \) is the credibility associated with \( \mu_{U} \), which depends on the greater or lesser dispersion of data that contribute to the calculation of \( \mu_{U} \) and is equal to 0.3640; \( N_{U} \) is a factor of binary control of the presence of a utility, which is 1 if the number of water supply systems is not provided and 0 otherwise.

Method 1: sub-case 2. The applicable considerations are similar to those in sub-case 1, but the situation itself is considerably simpler. The final expression is

\[
t_{LS} = (t)^{(N_{TV})(\mu_{TV} \cdot c_{TV})^{(N_{TV})}}
\]

where \( t_{LS} \) is the monitoring period to be introduced in the final value of \( w; \) \( t \) is the monitoring period; \( N_{TV} \) is a factor of binary control that is 1 if the actual period is known and 0 if it is unknown; \( \mu_{TV} \) is the arithmetic mean of the true monitored times, which is involved when the true monitoring period is unknown (9.20 years); \( c_{TV} \) is the credibility associated with \( \mu_{TV} \), which depends on the greater or lower dispersion of data that contribute to the calculation of \( \mu_{TV} \) and is equal to 0.5237; and \( N_{ts} \) is a factor of binary control relative to the period estimation that is 1 if the period is estimated and 0 if it is known.

Method 1: sub-case 3. In this case, expressions (6) and (7) can be used with one explanation.

The products \( \mu_{LSV} \cdot \mu_{U} \) and \( c_{LSV} \cdot c_{U} \) are 14,559 and 0.1398, respectively, because of a deeper analysis of the water supply system.

Method 2. The evaluated \( L_a \) can pass at the pipe scale using statistical methods in relation to the variables \( D, A, \) and \( M. \) Thus, a statistical survey of the data is performed, and a probability density function that
correlates \( L \) and \( D \) is obtained and presented in Figure 1

\[ f(D) = a \cdot D \cdot e^{-kD} \] (8)

This expression is chosen because it tends toward 0 for \( D = 0 \) and \( D \rightarrow \infty \). Thus, because this curve represents a probability density, its integral from 0 to \( \infty \) is unity, and the result is \( a = k^2 \). Finally, the remaining parameter \( k \) is chosen using the least mean square (l.m.s.) method, which results in

\[ f(D) = 9.51^2 \times 10^{-6} \cdot D \cdot e^{-9.51 \times 10^{-3} \cdot D} \] (9)

This methodology can be similarly performed for the distribution according to \( A \) to obtain

\[ f(A) = 3.07^2 \times 10^{-4} \cdot A \cdot e^{-3.07 \times 10^{-2} \cdot A} \] (10)

Finally, it is impossible to obtain a probability density function for \( M \). However, a simple probability function is used instead because \( M \) is a discrete variable.

The direct analysis of the data, which results in a weighted average of \( M \) with the associated \( L \) value, leads to the final percentages shown in Table 3 [Dell’Orfano 2013a].

The statistics related to other variables (e.g., pressure, abundant rainfall, and soil type) cannot be obtained because of data scarcity. The previous statistics can be used to evaluate \( L \) and thus \( w \). Three possible cases are determined: (1) the data for \( L \) are known; (2) the data for \( L \) are unknown, but data are available for \( D, A, \) or \( M \); (3) the data for \( L \) are unknown, and the data for \( D, A, \) and \( M \) are unavailable (Dell’Orfano 2013a).

Method 2 may be separately or simultaneously performed. When a water supply system that is composed of pipes corresponds to a finite number of \( D, A, \) and \( M \) values, it is necessary to pass from a continuous concept to a discrete probability distribution and use a probabilistic process to regain the unitary realignment of probability. According to \( D, A, \) and \( M \), the density probability function \( g \) can be obtained from the overlapping of the various multiplicative factors, which are separately investigated. Then

\[ g(D, A, M) = f(D) \cdot f(A) \cdot i_{LM} \] (11)

which indicates that

\[ \int_{0}^{\infty} \int_{0}^{\infty} f(D) \cdot f(A) \cdot d(D, A) \cdot \sum_{i} i_{LM} = 1 \] (12)

![Figure 1](https://iwaponline.com/ws/article-pdf/14/5/766/415759/766.pdf) | Graphical representation of \( f(D) \) and \( f(A) \). Continuous: interpolated curve; dashed: extrapolated curve.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Distribution of the rates of materials in the network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grey cast iron</td>
<td>0.3065</td>
</tr>
<tr>
<td>Ductile iron</td>
<td>0.1713</td>
</tr>
<tr>
<td>Asbestos cement</td>
<td>0.1709</td>
</tr>
</tbody>
</table>
In the discrete field, it is necessary to pass from the integral sum to the discrete sum and apply a probabilistic process as follows:

\[
\sum_{i=1}^{n_A} [f(D) \cdot f(A) \cdot i_{LM}]_i = 1; \quad \sum_{i=1}^{n_A} |f(D) \cdot f(A) \cdot i_{LM}|_i = 1
\]

\[
1 = \sum_{i=1}^{n_A} g_i (D, A, M) = 1
\]

It is possible to calculate the distribution lengths at the pipe scale if the system lengths \( L_{SV} \) or \( \mu_{LSV} \) are known (the second quantity is corrected with the term \( \mu_U \) if necessary). To unify the above descriptions, we analytically report the synthesis of the process as follows:

\[
L_{ci} = \left[ (L_{ew})^{(N_{ew})}, \langle (\mu_{ew})^{(N_{ew})} \rangle \cdot \langle (\mu_U) \rangle^{(N_U)} \right]
\]

\[
\cdot \left[ (f(l)_D)^{\beta} \cdot (f(l)_A)^{\gamma} \cdot (i_{LM})^{\alpha} \right] \cdot \left( \frac{1}{\sum_{i=1}^{N_M}} \right)^x
\]

(14)

where \( L_{ci} \) is the final length of a pipeline; \( f(l)_D \) is the density of the probability function of \( L \) as a function of \( D \); \( f(l)_A \) is the density of the probability function of \( L \) as a function of \( A \); \( i_{LM} \) is the probability of the presence of \( L \) as a function of \( M \); \( \alpha \) is a factor of binary control of the presence of \( D \) data that is 1 if the \( D \) data are provided and 0 if they are not; \( \beta \) is a factor of binary control of the presence of \( A \) data that is 1 if the \( A \) data are provided and 0 if they are not; \( \gamma \) is a factor of binary control of the presence of \( M \) data that is 1 if the \( M \) data are provided and 0 if they are not; \( x \) is a factor of binary control relative to the absence of \( D, A, \) and \( M \) data that is 1 if no information on \( D, A, \) or \( M \) are provided and 0 otherwise. Similarly, it is possible to evaluate \( T \) of the pipes according to the reported observations in the previous paragraphs

\[
L_{ci} = \left( f_{ci} \right)^{(N_{ew})} \cdot \left( \mu_{ci} \cdot c_{ci} \right)^{(N_{ew})}
\]

(15)

where \( f_{ci} \) is the monitoring period to be introduced in the final weight. After calculating \( L_{ci} \), the subsequent corrections can be made. Thus, the final \( L_{ci} \) value is

\[
L_{ci} = L_{ci} \cdot \left( c_{ci} \right)^{(N_{ew})} \cdot \langle \mu_{ci} \cdot c_{ci} \rangle^{(N_{ew})} \cdot \langle c_{ci} \rangle^{(N_{ew})}
\]

(16)

Social realignment of weights: Economic, social, and management aspects are very different among the countries that we considered in our study. These differences are not clearly distinguishable through physical parameters but certainly influence the evaluation of the weight \( w \) and breakage rate \( \Lambda \).

In fact, in more developed countries, there are many long water supply systems, and their management is often strongly oriented toward a careful measurement of the variables that are involved in the systems. Thus, many studies have been conducted on these systems. In contrast, the presence of short water supply systems, which are geographically located in less developed areas of the world and often poorly managed, decisively shows problems relative to the difficulty of collecting measurement data and publicising them in the literature.

Thus, the information about water supply systems of developed countries tends to nullify the information relative to water supply systems in countries that are characterised by a low gross domestic product ratio.

To accurately estimate the global weighted arithmetic mean of \( \Lambda \), which is expressed in Equation (5), it is necessary to reconstitute a balance among the weights of different countries.

The starting point is Equation (5). This average consists of the ratio between the sum of the products of \( \Lambda \) multiplied by the respective weights and the sum of the simple weights. In the products of the numerators, \( \Lambda \) can be divided by their average value and added to their variation from the mean value.

Thus, the influence of each data point compared to the general average breakage rate is related to the product of its deviation for the corresponding weight. We assume that the so-weighted influences of poor and rich countries similarly contribute to the final evaluation of \( \Lambda \).

With the previous assignment of the weights, if the maximum deviations with a positive sign (relative to poor countries with high \( \Lambda \) values and minimum weights) are compared with the maximum deviations with a negative sign (relative to rich countries with low \( \Lambda \) values and maximum weights), the influence of the second case clearly exceeds the influence of the first case.

Then, a correction exponent \( n \) is defined to balance the influence of the two extreme factors. The resulting value \( n \)
from this reasoning is 0.4711, which is adopted for the socio-economic realignment of the weights.

**Technical weight:** In some cases, data relative to a water supply system are inserted by the authors several times in the same publication, with reference to different variables, such as $D$, $M$, or $A$. Then, the same information is provided several times, and the weight is overestimated if the proposed methodology to evaluate the weight is applied to any information. In these conditions, it is necessary to divide the obtained weights by the number of presentations of the data. Therefore, a technical weight is considered as the inverse of the number of times that the same data point is presented.

**PIPE BREAKAGE RATE ON CONTINENTAL AND COUNTRY SCALES**

The worldwide value of $\Lambda$, which is denoted as $\Lambda_{mc}$, is calculated using the methods described above.

This value is 0.3615, with a total weight of 3,020.4389. This value is higher than that reported in Dell’Orfano et al. (2013a) (0.3352) because of the deeper observations in Dell’Orfano et al. (2013b).

The final results of $\Lambda_{mc}$ at a continental scale are given in Table 4, and the results of $\Lambda_{mn}$ at a country scale are reported in Table 5.

Continental macroregions are also identified. In particular, the American continent is divided into North and South America to highlight their differences, whereas Russia and the Republic of Azerbaijan are specifically considered in

<table>
<thead>
<tr>
<th>Continent</th>
<th>$w$</th>
<th>$\Lambda_{mc}$</th>
<th>$f_{CE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>31.4987</td>
<td>0.1721</td>
<td>2.1000</td>
</tr>
<tr>
<td>North America</td>
<td>1,248.7768</td>
<td>0.2379</td>
<td>1.5195</td>
</tr>
<tr>
<td>Eurasia</td>
<td>15.0152</td>
<td>0.2379</td>
<td>1.5193</td>
</tr>
<tr>
<td>Europe</td>
<td>1,325.1017</td>
<td>0.5077</td>
<td>1.1748</td>
</tr>
<tr>
<td>Oceania</td>
<td>274.4186</td>
<td>0.7303</td>
<td>0.4949</td>
</tr>
<tr>
<td>Asia</td>
<td>70.9053</td>
<td>1.2583</td>
<td>0.2873</td>
</tr>
<tr>
<td>Africa</td>
<td>51.3230</td>
<td>1.4583</td>
<td>0.2479</td>
</tr>
<tr>
<td>South America</td>
<td>3.6011</td>
<td>1.7100</td>
<td>0.2114</td>
</tr>
</tbody>
</table>

Table 5 | Weighted mean values of $\Lambda_{mn}$ by country

<table>
<thead>
<tr>
<th>Nation</th>
<th>$w$</th>
<th>$\Lambda_{mn}$</th>
<th>$f_{CE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>43.7375</td>
<td>0.0609</td>
<td>5.9336</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>128.8128</td>
<td>0.0945</td>
<td>3.8253</td>
</tr>
<tr>
<td>Sweden</td>
<td>143.7401</td>
<td>0.0966</td>
<td>3.7399</td>
</tr>
<tr>
<td>Austria</td>
<td>64.5826</td>
<td>0.1300</td>
<td>2.7806</td>
</tr>
<tr>
<td>USA</td>
<td>693.6570</td>
<td>0.1788</td>
<td>2.0218</td>
</tr>
<tr>
<td>Russia</td>
<td>12.8329</td>
<td>0.2281</td>
<td>1.5846</td>
</tr>
<tr>
<td>Norway</td>
<td>30.5774</td>
<td>0.2746</td>
<td>1.3163</td>
</tr>
<tr>
<td>Germany</td>
<td>37.1209</td>
<td>0.3136</td>
<td>1.1525</td>
</tr>
<tr>
<td>Cyprus</td>
<td>47.2878</td>
<td>0.3220</td>
<td>1.1224</td>
</tr>
<tr>
<td>Canada</td>
<td>459.6290</td>
<td>0.3482</td>
<td>1.0381</td>
</tr>
<tr>
<td>South Africa</td>
<td>23.4068</td>
<td>0.3775</td>
<td>0.9575</td>
</tr>
<tr>
<td>Taiwan</td>
<td>17.8006</td>
<td>0.4131</td>
<td>0.8749</td>
</tr>
<tr>
<td>Italy</td>
<td>637.2996</td>
<td>0.4569</td>
<td>0.7911</td>
</tr>
<tr>
<td>Australia</td>
<td>314.5645</td>
<td>0.6663</td>
<td>0.5425</td>
</tr>
<tr>
<td>Libya</td>
<td>13.7693</td>
<td>0.6927</td>
<td>0.5218</td>
</tr>
<tr>
<td>Morocco</td>
<td>8.3961</td>
<td>0.8613</td>
<td>0.4196</td>
</tr>
<tr>
<td>Spain</td>
<td>8.0643</td>
<td>0.9031</td>
<td>0.4002</td>
</tr>
<tr>
<td>Poland</td>
<td>10.2776</td>
<td>0.9241</td>
<td>0.3911</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>0.1805</td>
<td>0.9367</td>
<td>0.3859</td>
</tr>
<tr>
<td>Kyrgyzstan</td>
<td>1.8005</td>
<td>1.0000</td>
<td>0.3615</td>
</tr>
<tr>
<td>Argentina</td>
<td>1.8005</td>
<td>1.1900</td>
<td>0.3037</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>1.8005</td>
<td>1.3400</td>
<td>0.2697</td>
</tr>
<tr>
<td>Iran</td>
<td>43.7331</td>
<td>1.4617</td>
<td>0.2473</td>
</tr>
<tr>
<td>Tajikistan</td>
<td>1.8005</td>
<td>1.5900</td>
<td>0.2273</td>
</tr>
<tr>
<td>South Korea</td>
<td>0.3690</td>
<td>1.9200</td>
<td>0.1883</td>
</tr>
<tr>
<td>Peru</td>
<td>1.8005</td>
<td>2.2300</td>
<td>0.1621</td>
</tr>
<tr>
<td>Vietnam</td>
<td>1.8005</td>
<td>2.4100</td>
<td>0.1500</td>
</tr>
<tr>
<td>Ukraine</td>
<td>1.8005</td>
<td>3.0500</td>
<td>0.1185</td>
</tr>
<tr>
<td>Georgia</td>
<td>1.8005</td>
<td>3.2100</td>
<td>0.1126</td>
</tr>
<tr>
<td>Malaysia</td>
<td>1.8005</td>
<td>3.2500</td>
<td>0.1119</td>
</tr>
<tr>
<td>Moldova</td>
<td>1.8005</td>
<td>4.3600</td>
<td>0.0829</td>
</tr>
<tr>
<td>Uganda</td>
<td>5.7507</td>
<td>8.5620</td>
<td>0.0422</td>
</tr>
</tbody>
</table>

The continent of Eurasia, Asia and Europe are presented separately, Oceania is largely represented by Australia, and the African continent is also shown.

At the continental scale, the corresponding weight $w$ is shown for each continent with the average continental breakage rate $\Lambda_{mc}$ and a corresponding management correction factor $f_{CE}$, which is obtained as the ratio between
which is represented by the term related to the management of the water supply system, considered variables and stochastic systematic errors, i.e., physical errors, because of random credibility. Table 5 provides the country-scale result.

The management correction factor $f_{CE}$ is not known a priori: it is considered unitary in the case of the world but must be regarded as substantially unknown in other cases. If we set the value of $\Lambda_m$ in Equation (17), it becomes an equation in two unknowns: $\Lambda_{mc}$ and $f_{CE}$. After $f_{CE}$ is known, it is possible to estimate $\Lambda_{mc}$, or vice versa, if $\Lambda_{mc}$ is known it is possible to evaluate the management correction factor $f_{CE}$. A high level of management ($f_{CE}$ high) indicates a low breakage rate, on the contrary a poor level of management ($f_{CE}$ low) indicates a high breakage rate.

A second observation, which is directly connected to the previous one, is that for a given water supply system, under equal conditions of the deteriorating infrastructure, efficient management can reduce $\Lambda_{mc}$, whereas poor management causes it to increase.

The predicted value $\Lambda_{mc}$ is influenced by stochastic physical errors, i.e., because of random fluctuations of the considered variables and stochastic systematic errors, i.e., related to the management of the water supply system, which is represented by the term $f_{CE}$.

The above influences can be minimised. The first influence can be minimised by expanding the research in other unexplored geographical areas; the knowledge of $f_{CE}$ may improve the management.

Ultimately, expression (17) strongly reflects the power of the proposed model and illustrates the importance of the considered variables in the definition of the breakage rate.

Table 4 illustrates that North America has a management correction factor of 1.5195, whereas South America has a management correction factor of 0.2114. In addition, Europe has a high management correction factor of 1.1748. Although the $f_{CE}$ values reveal the management quality of the considered area, the $w$ values provide the corresponding credibility. Table 5 provides the country-scale result.

Table 5 illustrates that the European water supply systems perform better than the corresponding American systems. The South American and African water supply systems appear more distant in the ranking because of an economic bias factor, which should be investigated in further detail.

The evaluated weights $w$ direct further research about failure mechanisms in water supply systems in the world, particularly in countries that are not present in Table 5 but are characterised by rapidly growing economies, such as Brazil and China, or by strategic geographical location, such as Russia, Japan, and New Zealand.

There is no linear dependence between the weight and the breakage rate. It is not obvious that a heavy weight corresponds to a low breakage rate or vice versa.

A high breakage rate may lead one to believe that the water supply systems of that country are in extremely bad condition, but if the breakage rate is not supported by a significant weight, it is necessary to investigate the condition further.

In contrast, a reasonable breakage rate and a heavy weight lead one to believe that the information about the water supply systems in that country is in excellent condition.

CONCLUSIONS

A key element to quantify the effectiveness of a water supply system is its mechanical reliability. Within it, we refer to the following structural parameters: (1) supply interruptions and (2) pipeline breakages. In this paper, the mechanical reliability of water supply systems is investigated with reference to pipeline breakages. This study focuses on pipelines because they are the main components of water supply systems; therefore, their failures highly affect the statistical reliability of the system.

In the literature, the theoretical studies on the mechanical reliability are often based on databases with reduced consistency. In contrast, in this study, large, detailed datasets from several areas of the world have been examined, and 18 variables for describing the deterioration of a water supply network have been identified.

The analysis indicates that these datasets are often affected by temporal and geographical limitations and occasionally by the repetition of information, which has been included in previous calculations. Therefore, suitable weights have been introduced to eliminate the distortion resulting from the different numbers of monitored water
supply systems in more or less developed countries, and the weights evaluation is described.

The use of an unweighted pipe breakage value \( \Lambda \) determines significantly different results. For example, the \( \Lambda \) values can become an order of magnitude higher than the weighted value.

Weighted mean values of the pipe breakage rate on continental and country scales are evaluated and a management correction factor \( f_{CE} \) is calculated for different continents and countries and compared to a reference value (unitary).

The management correction factor \( f_{CE} \) provides, in a compact and synthetic manner, indications about the management of a water supply system and is certainly an element of correction and improvement for the overall planning of a water supply system.

**ACKNOWLEDGEMENT**

This study was funded by PON_01596 WATERGRID_01 (CUP B61C11000880005), ‘Activity support development of a hydraulic simulation model (in terms of quantity and quality) for water distribution systems’.

**REFERENCES**


Brunone, B., Ferrante, M. & Meniconi, S. 2009 Ricerca e controllo delle perdite nelle reti di condotte, 1st edn. Research and Monitoring of Leaks in Pipe Networks, Città Studi, Italy.


Dell’Orfano, F. 2013a Modello fisico-matematico per la rappresentazione del deterioramento delle condotte in un acquedotto. Tesi di Dottorato XXIV ciclo, aprile (Physical-mathematical model for representing the deterioration of the water pipelines, PhD Thesis).

Dell’Orfano, F., Esposito, V., Qualtieri, P. & Pulci Doria, G. 2013b Mean values of water pipe break rate in the world and in geographical areas. In: Seventh IWA International Conference...


