

Statistical modelling of French drinking water pipe inventory at national level using demographic and geographical information

Eddy Renaud, Alain Husson, Alice Vacelet, Yves Le Gat, Anne-Emmanuelle Stricker

► **To cite this version:**

Eddy Renaud, Alain Husson, Alice Vacelet, Yves Le Gat, Anne-Emmanuelle Stricker. Statistical modelling of French drinking water pipe inventory at national level using demographic and geographical information. H2Open Journal, 2020, 3 (1), pp.89 - 101. 10.2166/h2oj.2020.028 . hal-03215000

HAL Id: hal-03215000

<https://hal.inrae.fr/hal-03215000>

Submitted on 3 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



ELECTRONIC OFFPRINT

Use of this pdf is subject to the terms described below





This paper was originally published by IWA Publishing. It is an Open Access work, and the terms of its use and distribution are defined by the Creative Commons licence selected by the author.

Full details can be found here: <http://iwaponline.com/content/rights-permissions>

Please direct any queries regarding use or permissions to h2open@iwap.co.uk

Statistical modelling of French drinking water pipe inventory at national level using demographic and geographical information

Eddy Renaud ^{a,*}, Alain Husson^a, Alice Vacelet^b, Yves Le Gat ^a
and Anne-Emmanuelle Stricker^a

^a INRAE, UR ETBX, 50 avenue de Verdun Gazinet, F-33612 Cestas cedex, France

^b PROLOG Ingénierie, F-33300 Bordeaux, France

*Corresponding author. E-mail: eddy.renaud@inrae.fr

Abstract

At the request of the French Ministry for Ecological and Inclusive Transition (MTES) and the French Biodiversity Agency (AFB), INRAE carried out a study focusing on the creation of a national asset knowledge base for drinking water networks, operating at Water Agency (WA) scale. Our study involved the creation of a number of statistical models. These models combine data relating to network characteristics (gathered from a sample of utilities) with geographical and demographical data collected from all municipalities in France. The SISPEA database, along with geographical information system (GIS) data from a sample covering around half of the total French drinking water network, were used to configure multilinear pipe models by diameter, installation period and type of material. On the basis of these models, the total length of drinking water pipes in mainland France was estimated to be 875,000 km. Networks are quite young (60% of the total length was laid after 1970), small diameter pipes are the majority (70% of the pipes have a diameter less than or equal to 100 mm) and the materials used are linked on geographic areas (pipes are mainly made of PVC in the west of France while cast iron pipes predominate in the other regions).

Key words: demographic and geographical data processing, drinking water networks, statistical modelling, technical asset knowledge

INTRODUCTION

In France, drinking water supply is overseen by some 13,000 Public Drinking Water Supply Services (WSSs). The majority of WSSs (around 56%) serve less than 1,000 subscribers. While WSSs run by private operators make up only 30% of the total number, they serve almost 60% of total subscribers (AFB 2018). It is clear that knowledge and management of French drinking water supply networks is somewhat fragmented and spread between a multitude of stakeholders. However, drinking water supply is a key element in the water cycle, and issues such as water conservation need to be dealt with at a scale larger than that of the WSS. For this reason, the French Government, through regulation, and the six French Water Agencies, through subsidies, have introduced a number of public policy initiatives aimed at encouraging local stakeholders to take a sustainable approach to asset management, with the more specific objective of limiting water loss. (The French Water Agencies (WAs) have for their territory the six major watersheds of the country: Adour-Garonne (AG), Artois-Picardie (AP), Loire-Bretagne (LB), Rhin-Meuse (RM), Rhône-Méditerranée-Corse (RMC), Seine-Normandie (SN).)

However, both the French Government and its WAs suffer from a lack of sufficient asset knowledge at their respective administrative scales. In 2009, the French Government created the Information System for Public Water and Sanitation Services (French: SISPEA), which records regulatory performance data from WSSs. However, the only asset data stored by this system are the total length of the network, and the data in question are not exhaustive (in 2015, the system only contained information about half of all WSSs). The only well-documented national study which includes pipe diameter, materials and installation period, dates back to 2002 (Cador 2002).

In view of this, the French Ministry for Ecological and Inclusive Transition (MTES) and the French Biodiversity Agency (AFB) asked INRAE to carry out a study focusing on the creation of a national asset knowledge base for drinking water networks at the France and WA scales. The study took place from 2016 to 2018 and included three key sections:

- Section 1 – Creation of a technical/statistical approach to assessing drinking water assets.
- Section 2 – Identifying the needs of WAs, the MTES and the AFB, and reviewing the suitability of any system developed.
- Section 3 – Economic and financial assessment of drinking water assets.

The present paper presents work carried out as part of Section 1. Our study focuses on the length, diameter, material and installation period of drinking water pipes which are networks' characteristics targeted by French regulations (French Law 2010-788 of 12/07/2010; Decree no. 2012-97 of 27/01/2012).

In view of the fragmented nature of French drinking water supply organisation, and given the limited knowledge some of the WSSs have about their assets, it is simply not realistic to try and achieve an exhaustive inventory of all water pipes in France. On this basis, missing data have to be estimated. Experience shows that network characteristics tend to be rather heterogeneous from one WSS to another, influenced by a number of physical, historical and social phenomena. To account for these specific local traits, it was decided in this study to use statistical models to link these characteristics to a number of explanatory variables. For estimations to take place, the values of these variables need to be known for the whole of France (Figure 1). Such data do exist, mainly at municipality scale, from publicly available sources. They provide exhaustive demographic and geographical information.

This paper will first detail the collection and processing of pipe networks' data and demographic and geographical explanatory variables. It will then present the calibration and validation of the statistical models built to estimate missing data. Finally, the main results obtained at France and WA scales will be highlighted.

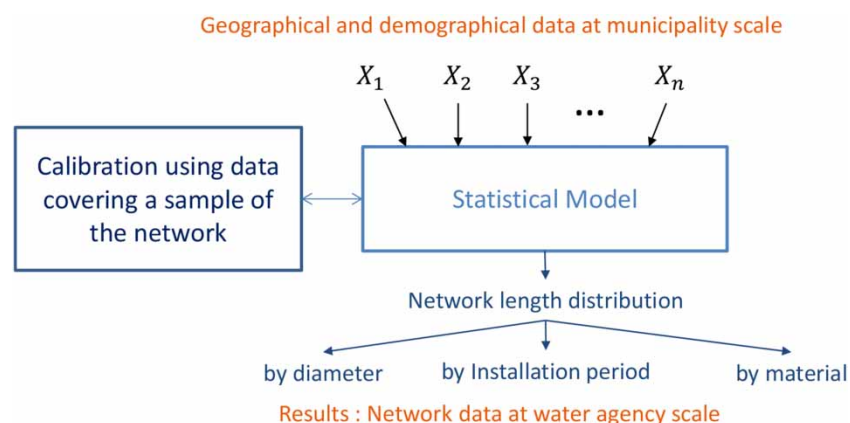


Figure 1 | Principle of our statistical model.

DATA COLLECTION AND PROCESSING

The raw data used in our modelling process were available at two spatial scales. Asset data for drinking water were available at the service area scale, while geographical and demographic data were available at the finer municipality scale. To link these two datasets together, they needed to be processed and presented at a common scale. We opted for the municipality scale, due to its stability and the volume of data available through different geographical layers.

Pipe networks' data

SISPEA data

The SISPEA database provides the means by which to link together data at different scales. SISPEA contains an exhaustive list of WSSs, each accompanied by a list of municipalities served. At the time of our study, the most recent set of data relating to the organisation of WSSs dated from 2015. However, one drawback of this database is that it does not provide detailed information on municipalities served by more than one WSS. As a consequence, at municipality scale, network length is only usable when WSS territory is a single municipality. As mentioned in the Introduction, SISPEA does not provide each year an exhaustive set of data. In view of this, it is better not to use only information from one year. In our study, we used the most recent network length values from the period 2009 to 2015.

GIS data preprocessing

To obtain information on the length of pipes according to their characteristics, we requested data from WSSs, including those in large cities and those serving more than 30 municipalities. These data were provided in the form of geographical information systems (GISs). Because the information systems used vary greatly from one WSS to another, it was necessary to homogenise the data taken from different sources. The steering committee for our study, made up of experts from across France, defined classes of data for each set of pipe characteristics. These classes were designed to take into account the varying contexts, regulatory environments and technical features of different assets (Table 1): material types according to installation periods (ASTEE *et al.* 2013); common interior diameter proposed by manufacturers; and installation period recommended by the regulation (MTES instruction 2015).

When material type was undetermined in the raw data available, some types were attributed based on installation period according to the rules presented in Figure 2.

Where necessary, the interior diameter (D_{int}) of pipes made from PVC, polyethylene and BPVC were deducted from nominal diameters (D_{nom}), which are external diameters. This calculation was carried out using Equations (1)–(3) shown below, based on data from the main suppliers of 16-bar nominal pressure pipes (Husson *et al.* 2018):

$$\text{PVC: } D_{int} = D_{nom} \times 0.89 - 2.47 \quad (1)$$

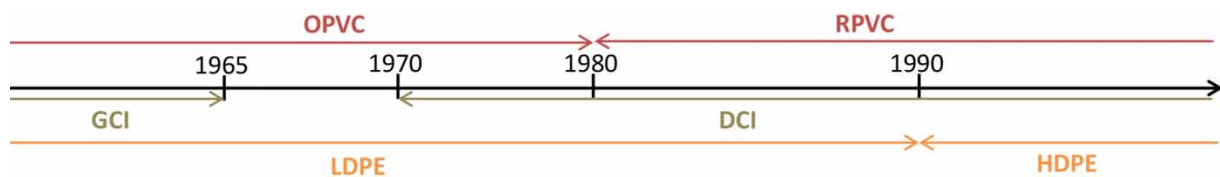
$$\text{PE: } D_{int} = D_{nom} \times 0.83 - 3.24 \quad (2)$$

$$\text{BPVC: } D_{int} = D_{nom} \times 0.95 \quad (3)$$

Data gathered were first combined within a single information system, and then broken down to municipality scale, using the borders between municipalities to define the relevant areas.

Table 1 | Classes defined for pipe characteristics

| Diameter | | Material | | Installation period Class |
|----------------------|-------|--------------------------------|-------|------------------------------|
| Interior D (mm) | Class | Names | Class | |
| Unknown | Unk | Unknown | Unk | Unk |
| $0 < D \leq 35$ | 39 | Steel | S | ≤ 1945 |
| $35 < D \leq 45$ | 40 | Asbestos cement | AC | [1945–1965] |
| $45 < D \leq 55$ | 50 | Steel core concrete | SCC | [1965–1970] |
| $55 < D \leq 72$ | 60 | Undetermined concrete | UC | [1970–1980] |
| $72 < D \leq 90$ | 80 | Grey cast iron | GCI | [1980–1990] |
| $90 < D \leq 113$ | 100 | Ductile cast iron | DCI | >1990 |
| $113 < D \leq 137.5$ | 125 | Centriflex cast iron | CCI | |
| $137.5 < D \leq 175$ | 150 | Undetermined cast iron | UCI | |
| $175 < D \leq 225$ | 200 | Low-density polyethylene | LDPE | |
| $225 < D \leq 275$ | 250 | High-density polyethylene | HDPE | |
| $275 < D \leq 325$ | 300 | Undetermined polyethylene | UPE | |
| $325 < D \leq 375$ | 350 | Old polyvinyl chloride | OPVC | |
| $375 < D \leq 425$ | 400 | Recent polyvinyl chloride | RPVC | |
| $425 < D \leq 475$ | 450 | Bi-oriented polyvinyl chloride | BPVC | |
| $475 < D \leq 550$ | 500 | Undetermined PVC | UPVC | |
| $550 < D \leq 650$ | 600 | Other materials | Other | |
| $650 < D \leq 750$ | 700 | | | |
| $750 < D$ | 701 | | | |

**Figure 2** | Key dates used for determining pipe material type.

Data collected

Collection of data through GISs and databases depended largely on the cooperation of staff of the WSSs. The cumulative length of the networks collected from 440 WSSs reaches 412,176 km.

Once data collected have been processed, standardised and fed into a single information system, the proportions of pipes whose material type, diameter or installation period are unknown can be calculated (Table 2).

As can be seen above, the diameters and material types of pipes are generally well known, given that the percentage of pipes for which they are not known is never greater than 10%. However, there is a much greater lack of knowledge of installation periods, with the unknown percentage over 30% for the whole country. There are also large fluctuations between WAs (from 16% to 40% in places).

Grouping together categories

Our data analysis shows that, for all WAs, certain categories of material type and diameter are fairly rare. To achieve statistically useful results, we decided to group together some of the initial categories shown in Table 1.

Table 2 | Network length collected by WA and shares of unknowns for each characteristic

| | Collected length (km) | Unknown percentage (%) | | |
|--------|-----------------------|------------------------|----------|---------------------|
| | | Material type | Diameter | Installation period |
| AG | 113,211 | 8 | 7 | 40 |
| AP | 21,433 | 2 | 2 | 17 |
| LB | 155,494 | 5 | 2 | 26 |
| RM | 12,840 | 2 | 2 | 16 |
| RMC | 65,118 | 8 | 6 | 35 |
| SN | 44,078 | 9 | 3 | 27 |
| France | 412,176 | 6 | 4 | 31 |

Material type:

- The categories UPE, LDPE, HDPE, UC, SCC and Other were grouped together into the category 'Various'.
- Pipes in the category BPVC were included in the RPVC category.
- Pipes in the category CCI were added to the category UCI.

Diameter:

- The class entitled '50 mm' includes the three classes '39', '40' and '50'.
- The class entitled '60–80 mm' includes the two classes '60' and '80'.
- The class entitled '100 mm' is merged with the class '100'.
- The class entitled '125–150 mm' includes the two classes '125' and '150'.
- The class entitled '200–250 mm' includes the two classes '200' and '250'.
- The class entitled ' ≥ 300 mm' includes the eight classes '300', '350', '400', '450', '500', '600', '700' and '701'.

Demographic and geographical municipality data**Gathering and preparing data**

Municipality data collected to create explanatory variables for our statistical models came primarily from the following public sources:

- INSEE (<https://insee.fr/>): Demographic and territorial data
- IGN (<http://www.ign.fr/>): Cartographical and urban development data
- CORINE LAND COVER (<https://land.copernicus.eu/pan-european/corine-land-cover/>), cartography and land cover
- OPENSTREETMAP (<https://www.openstreetmap.fr/>): road maps.

We used a number of processing techniques to extract municipality-level cartographical information from the raw data and to generate new data. The main methods used were cartographical processing (dividing and intersecting with municipality boundaries and grid squares), calculating ratios, combinations and classifications, i.e., ascending hierarchical classification (AHC) (Chessel *et al.* 2004).

The municipality database contains 97 variables, 24 of which come directly from public databases. The remaining 73 come from calculated data, or data created through geographical processing. Certain data are specifically constructed to show a highly probably link with length and other characteristics of drinking water networks.

Road length

Properties supplied by WSSs are generally accessible by a road, and pipes are often laid below or alongside those roads. On this basis, we assume that there is a correlation between road length and pipe length. OPENSTREETMAP provides data on a variety of roadways, not all of which are generally associated with water pipes (footpaths, bridleways, cycle lanes, etc.). In view of this, our first step was to filter out any irrelevant information. We then filtered the data for a second time, to remove roads running through large uninhabited areas. This was achieved by removing from the model any road located more than 250 m from the nearest residential property. The value of 250 m was calibrated by optimising the correlation between road length (after processing) and pipe length (Figure 3).

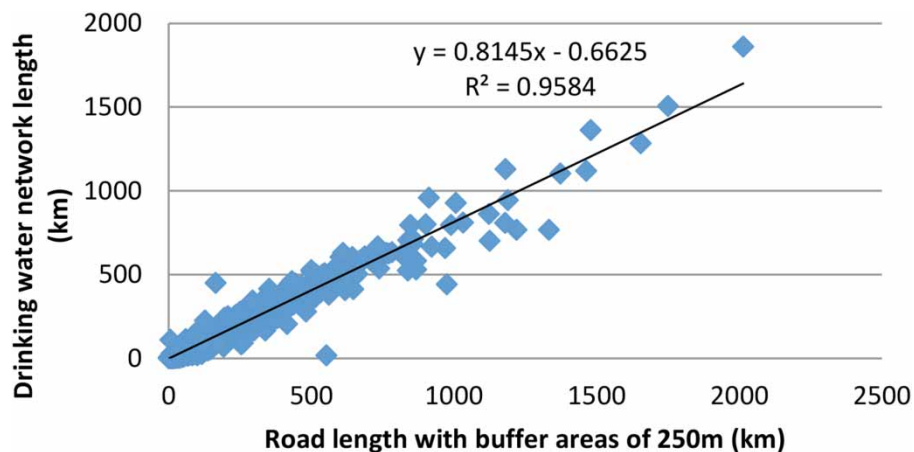


Figure 3 | Pipe length based on road length after processing (Han-Mui 2016).

Land use

Whether or not there is a drinking water network in a given area depends on the potential uses of water, which, in turn, depend on land use. CORINE LAND COVER (CLC) divides territories up into 44 types of land use. To associate these with water use, we grouped them into four categories: C1 (inhabited area), C2 (artificial spaces outside inhabited areas), C3 (farming areas) and C4 (natural area). To make best use of this information, we used an AHC to create a typology of municipalities. The typology is based on the respective proportions of these types of areas within a given municipality. This leads to the qualitative variable *Label*, which is divided into six classes: Label1: natural municipalities, Label2: natural municipalities with large farming areas and an urban area, Label3: farming municipalities, Label4: urban municipalities, Label5: joint natural and farming municipalities, and Label6: farming municipalities with large natural areas and an urban area.

Buildings

The number, size and height of the buildings within a zone are generally indicative of the number of users of its water network. This, in turn, provides an indication of the length and diameter of the pipes within that network. To make best use of the information from the IGN database relating to buildings, and to complement the variables created by cross-referencing with the zones in the four CLC categories, we generated a 500 m × 500 m grid system for the whole of France. Within each grid square, we calculated ratios of buildings in terms of number and size. Using this approach, we were able to divide France into two sets of five classes. For each class, the proportion of municipal territory within it is a new variable.

STATISTICAL MODELLING

Modelling method

The aim of our statistical model was to estimate, at municipality scale, the total length of the network and the length of pipes in each class of characteristics. Pipe characteristics are not mutually exclusive: material type is linked to installation period (we know that certain materials were only used during certain periods) and diameter (plastic pipes tend to be narrower in diameter). Installation period is, in turn, linked to diameter (wide-diameter pipe networks are generally among the oldest). In each class, pipe length is linked to total network length. To take into account these interdependencies, our models were constructed in the following order: total length, diameter, installation period and material type. Having tested a number of approaches, including the Dirichlet model (Maier 2014), without a great deal of success, we decided to make use of multilinear models with interaction effects. Where the variable to be explained is Y and with explanatory variables X_1 to X_n which interact with variable X_0 , our model uses Equation (4), where $\beta_i(X_0)$ is the value of the parameter of variable X_i , associated with the value acquired by X_0 :

$$Y = \beta_0 + \beta_1(X_0)X_1 \dots + \beta_n(X_0)X_n \quad (4)$$

Our models were constructed with R software, using the following process:

1. Selection of the initial set of variables through a process based on Akaike's information criterion, using the 'stepAIC' function in R (Venables & Ripley 2002).
2. Calibration of a model based on the initial set of variables, using the 'lm' function in R. Non-significant variables (with Student's t -test p -value of more than 5%) were excluded (Student 1908).
3. Certain variables which appeared to be beneficial to the model were reintroduced. We used the R^2 determination variable (among others) to decide which were beneficial.
4. Cross validation of the model was carried out using the Monte Carlo method. Where results were unsatisfactory, we started again at step 3, or at step 1, using new variables.

The cross validation mentioned in step 4 involves taking 1,000 random samples and splitting data into two subsets. The first subset, containing 75% of the data, is used to calibrate the model, while the second, containing the remaining 25%, is used to compare predicted values with actual values. The 1,000 results obtained through this process provide a distribution of relative deviation between actual and predicted values. This distribution is represented on a boxplot graph. The more the distribution is tightened and centred on 0, the better the quality of the model.

When a satisfactory model is obtained, a value for the relevant variable is attributed to each municipality. This is the actual measured value (where known), or the modelled value otherwise. Estimated and measured values at municipality scale are then agglomerated to provide results at WA and national scales.

Modelling total network length

To create the sample used to calibrate our model of total municipality pipe length, data from SISPEA (4,526 single-municipality WSSs with a total length of 128,787 km) were combined with data from GISs for municipalities not present in SISPEA (12,926 municipalities with a total length of 381,160 km). The total calibration sample was made up of 17,452 municipalities, with a total length of 509,947 km. Based on the previously described method, the parameters for the multiple linear model selected are provided in Table 3.

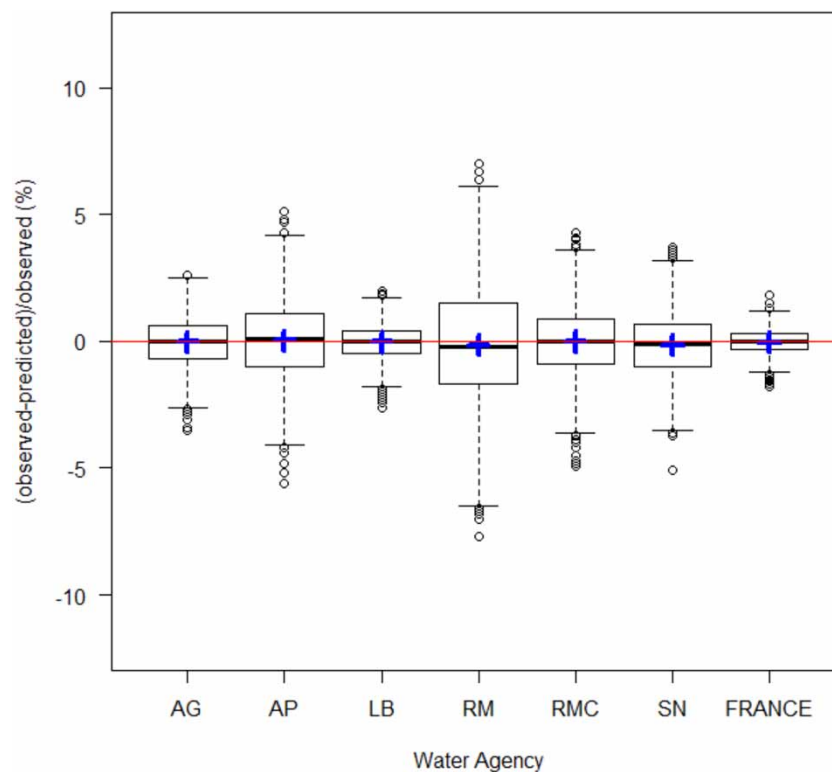
The model uses most of the variables relating to buildings, including number and size (BN and BS), as well as number depending on CLC zone ($BNCn$) and proportionate size depending on grid square

Table 3 | Parameters of the model created for municipality network lengths

| $R^2 = 0.9613$ | $\beta_0 = 4.094$ | AG | AP | LB | RM | RMC | SN |
|----------------|-------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| <i>BN</i> | β_1 | -6.23×10^{-4} | 5.33×10^{-4} | -2.47×10^{-4} | -1.97×10^{-3} | -4.82×10^{-3} | 2.00×10^{-3} |
| <i>BNC2</i> | β_2 | -4.24×10^{-3} | 5.68×10^{-5} | 1.16×10^{-2} | 1.96×10^{-2} | 1.30×10^{-2} | 6.47×10^{-3} |
| <i>BNC3</i> | β_3 | 1.03×10^{-3} | 5.59×10^{-3} | 2.11×10^{-3} | 7.25×10^{-3} | 1.32×10^{-2} | 6.16×10^{-3} |
| <i>RC1</i> | β_4 | -6.20×10^{-4} | -3.24×10^{-4} | -6.27×10^{-4} | 3.11×10^{-4} | 9.18×10^{-5} | -2.67×10^{-4} |
| <i>RC2</i> | β_5 | -8.12×10^{-4} | -6.40×10^{-4} | -1.00×10^{-3} | -4.31×10^{-4} | -4.13×10^{-4} | -5.66×10^{-4} |
| <i>RC3</i> | β_6 | -2.66×10^{-4} | -1.15×10^{-4} | -4.42×10^{-4} | -3.65×10^{-5} | -1.91×10^{-5} | -1.71×10^{-4} |
| <i>RC4</i> | β_7 | -3.08×10^{-4} | -1.66×10^{-5} | -5.88×10^{-4} | 1.12×10^{-5} | 3.54×10^{-5} | -3.74×10^{-4} |
| <i>R250</i> | β_8 | 8.88×10^{-4} | 6.26×10^{-4} | 9.57×10^{-4} | 2.97×10^{-4} | 2.16×10^{-4} | 7.43×10^{-4} |
| <i>GS</i> | β_9 | 3.28×10^{-7} | 1.06×10^{-7} | 6.16×10^{-7} | 9.94×10^{-9} | 1.25×10^{-7} | 2.96×10^{-7} |
| <i>DBSCAR1</i> | β_{10} | -7.822 | -6.123 | -8.531 | -3.034 | -6.289 | -6.321 |
| <i>DBSCAR2</i> | β_{11} | -1.874 | -2.793 | 0.247 | -3.463 | 1.577 | -3.320 |
| <i>DBSCAR3</i> | β_{12} | -7.668 | -8.156 | -8.104 | 0.845 | -4.236 | -4.046 |
| <i>DBSCAR5</i> | β_{13} | -17.664 | 0.168 | -2.749 | -9.446 | -28.209 | -3.739 |
| <i>BS</i> | β_{14} | 5.82×10^{-5} | 4.52×10^{-5} | 3.95×10^{-5} | 4.69×10^{-5} | 8.27×10^{-5} | 2.87×10^{-5} |
| <i>SU</i> | β_{15} | 2.66×10^{-6} | 3.58×10^{-6} | 3.39×10^{-6} | 1.80×10^{-6} | 3.68×10^{-6} | 7.50×10^{-7} |

class (*DBSCARn*). Several of the variables also relate to road data after processing, such as total length (*R250*) and roads included in a CLC zone (*RCn*). Other variables in the model are total surface area (*GS*) and total built-up surface area (*SU*).

The boxplots shown in [Figure 4](#) relate to the distribution of relative deviation between actual and predicted lengths, by WA and for the whole country. Despite a number of striking disparities between WAs, the distributions are all tightened and centred on 0.

**Figure 4** | Cross validation of model to estimate total length.

Using this model, the lengths of networks from municipalities not included in the sample can be estimated. These values can then, in turn, be used to estimate total network length for the whole of France and by WA (Table 4). On this basis, using a confidence interval of 95%, we were able to estimate total pipe length in France at around 875,000 km, give or take 2,000 km.

Table 4 | Network lengths and collection rates by WA

| Length in km | Estimated | Collected | Collection rate (%) |
|--------------|-----------|-----------|---------------------|
| AG | 195,236 | 113,211 | 58 |
| AP | 39,512 | 21,433 | 54 |
| LB | 288,679 | 155,494 | 54 |
| RM | 38,308 | 12,840 | 34 |
| RMC | 170,922 | 65,118 | 38 |
| SN | 142,307 | 44,078 | 31 |
| France | 874,963 | 412,176 | 47 |

To compare collected and estimated lengths we calculated collection rates. The results show an average collection rate close to 50% and significant disparities between WAs, with the collection rate in some WAs almost twice that of others.

Modelling of pipe lengths by diameter, installation period and material type

As explained previously, to allow data from a model to be used as variables in subsequent models, we modelled characteristics in the following order: (1) diameter, (2) installation period, (3) material type.

We calibrated a model for each category of each characteristic. Explanatory variables differ depending on the categories modelled. That being said, all models contain variables relating to buildings (number, size and weighted average height, cross-referenced with CLC layers, or processed using grid squares),

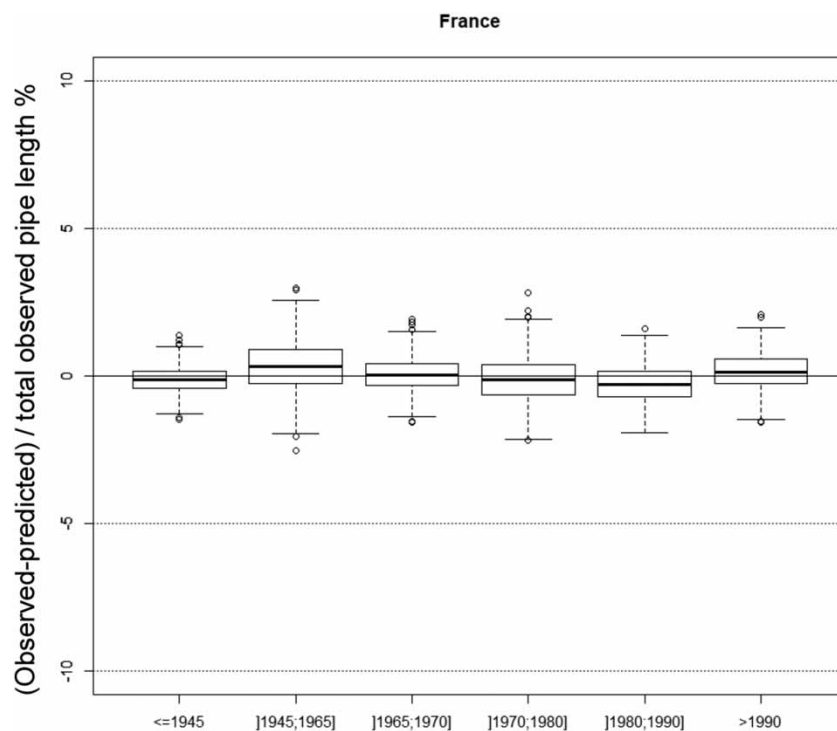


Figure 5 | Cross validation of models to estimate installation period.

roads (processed length values cross-referenced with CLC layers), surface areas (total and built-up areas), land use (*Label* variable), population (often attached to another variable) and total pipe length. The ‘installation period’ models use explanatory variables based on the results of the ‘diameter’ models (proportion of a diameter class within total length). The ‘material type’ models also include explanatory variables constructed using ‘installation period’ categories. For categories relating to cast iron and PVC, models were constructed using samples in which the proportion of undetermined cast iron (UCI) or undetermined PVC (UPVC) pipes was low (less than 10%). On this basis, it was possible to obtain estimations without using these categories, i.e., all cast iron pipes are assigned to GCI or DCI and all PVC pipes are assigned to OPVC (PVC installed prior to 1980) or RPVC (PVC installed after 1980).

For each characteristic, the sum of the lengths modelled per categories is slightly different from total length. In view of this, we corrected the lengths for each category based on a pro rata of total length. As shown in Figure 5 for installation periods, cross validation of our models showed relative deviations between predicted and actual values that were very close to zero.

RESULTS

Estimation of lengths by diameter

Estimations of length by diameter show that small diameter pipes are by far the most represented since 70% of the pipes have a diameter less than or equal to 100 mm. On the other hand, wider pipes (300 mm and greater) account for only 4% of total national pipe length, or around 35,500 km (Figure 6).

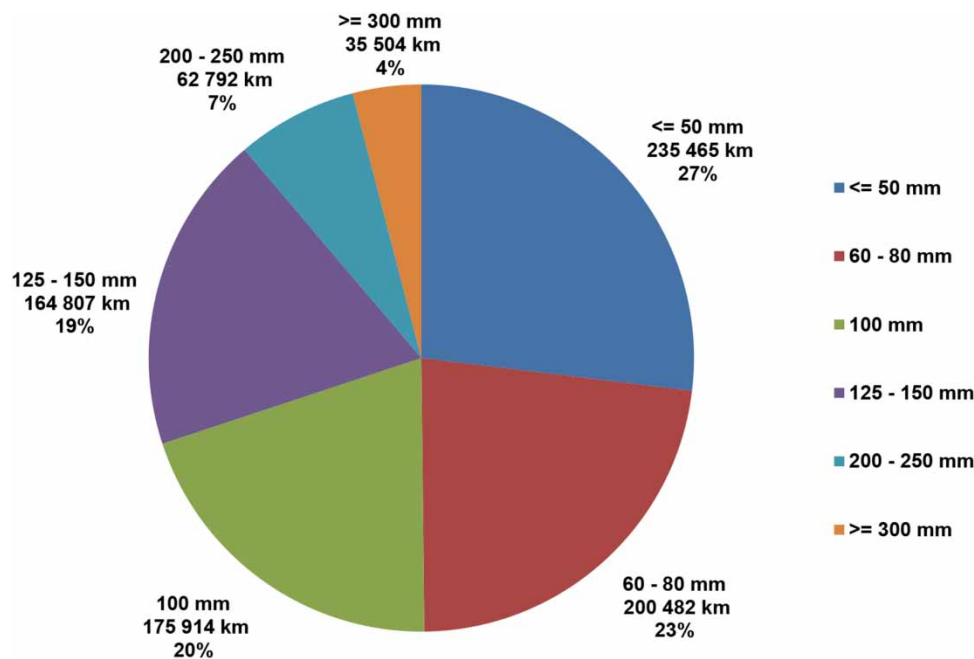


Figure 6 | Spread of pipe length by diameter.

Estimating length by installation period

The spread of network length by installation period is slightly different from that observed in our sample of collected data (Figure 7). This proves – if proof was ever necessary – the advantages of using multi-variable modelling rather than simple extrapolation, when correcting any bias in collected data. For example, as expected, it was confirmed that it is much less common for the installation date

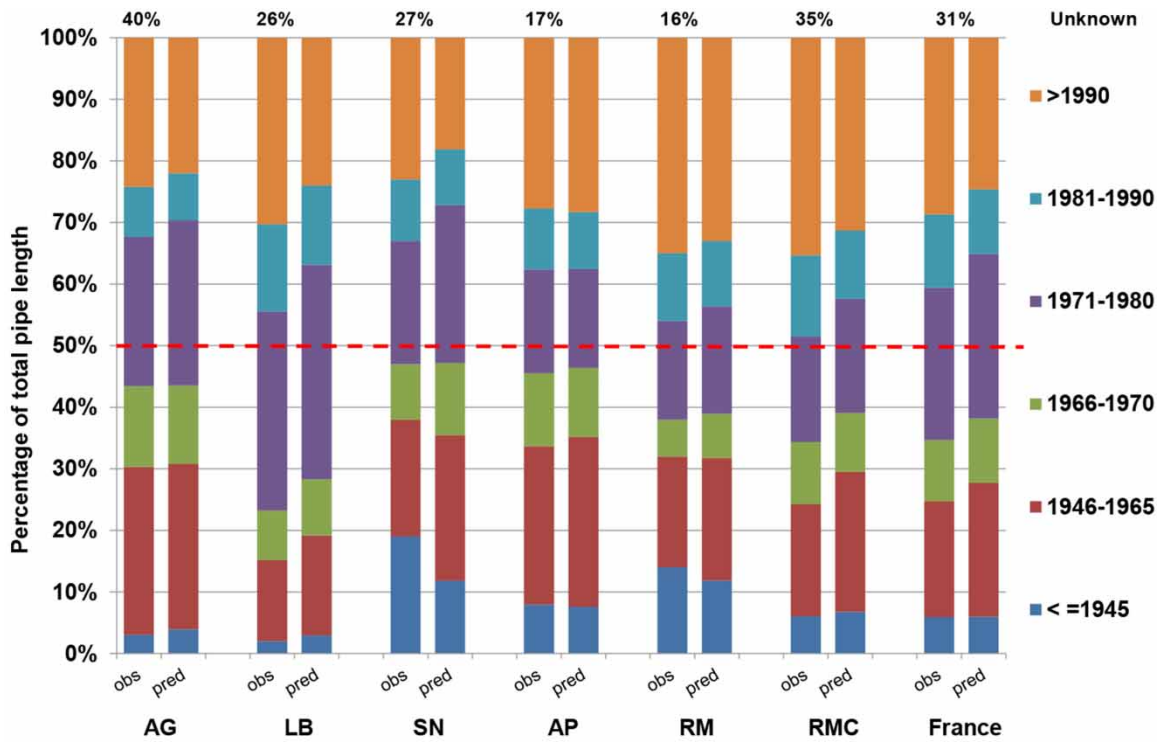


Figure 7 | Combined proportions of length by installation period.

of recently installed pipes to be unknown. This is reflected by a 4% drop in the proportion of the class ‘ > 1990’ (29% of the known length of the sample compared with 25% of estimated length).

Networks are quite young (60% of the total length was laid after 1970), and this is even more marked for WA LB for which only 30% of the pipes were installed before 1970. Over a quarter of total water pipes in France (27%) was laid during the ten years between 1971 and 1980. This shows how water supply was rolled out to French rural areas in a relatively short period of time; a phenomenon which will need to be taken into account for asset management strategies.

Estimating length by material type

The vast majority of French drinking water pipes are made of cast iron and PVC (Table 5), with slightly more of the latter (47% PVC and 41% cast iron). However, this national average is not necessarily indicative of the situation within different WAs; there are in fact significant differences from one WA to another. For example, in four of the six WAs, there are more cast iron pipes than PVC pipes (Figure 8). Conversely, the two WAs located in the west of France (LB and AG), which account for over half of the total French network (Table 5), have over 60% of their pipes made from PVC (68% and 63%, respectively).

PVC pipes installed before 1980 (OPVC) appear almost twice as much as those installed after 1980 (RPVC). For cast iron pipes, the most common variant is ductile cast iron (DCI). Asbestos cement

Table 5 | Total pipe length by material

| | GCI | DCI | OPVC | RPVC | AC | A | Various | Total |
|--------|---------|---------|---------|---------|--------|--------|---------|---------|
| Length | 147,815 | 205,896 | 275,014 | 141,826 | 35,440 | 25,869 | 43,103 | 874,963 |
| % | 17% | 24% | 31% | 16% | 4% | 3% | 5% | 100% |

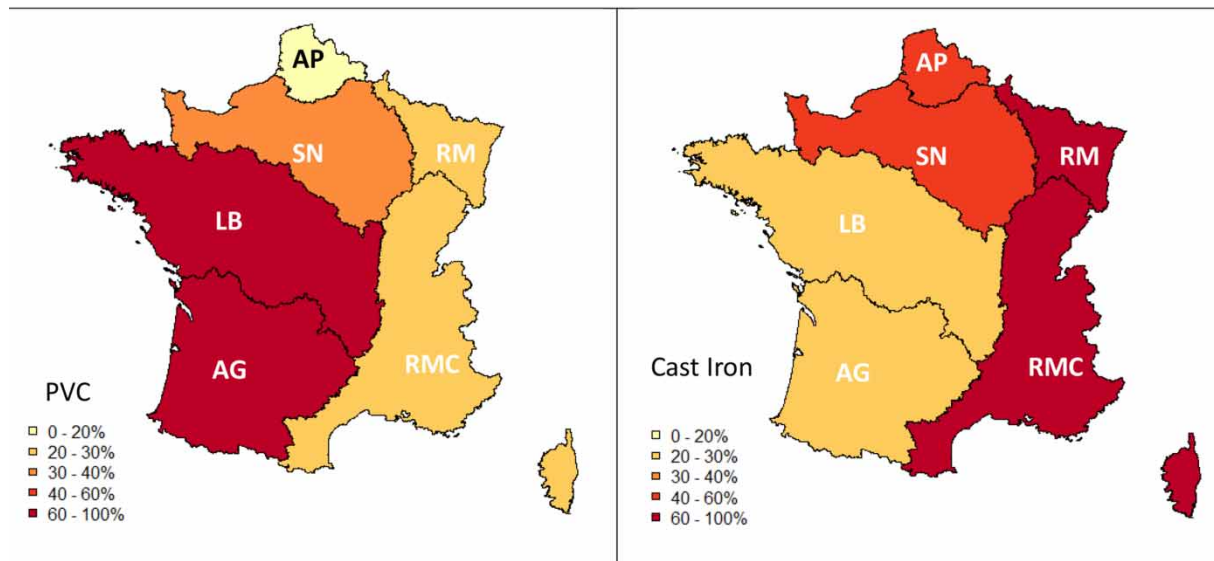


Figure 8 | Proportions of PVC and cast iron pipes by WA.

(AC), which makes up 4% of France's total network length, accounts for only a small fraction of most WAs' networks, with the exception of AP, where 22% of pipes (8,500 km) are made of this material.

CONCLUSION

The work carried out in this study presents a statistics-based method to estimate network length by diameter, installation period and material type. This method uses demonstrated links between drinking water networks and demographic and geographical elements of their environment, such as roads, urban planning, population and land use. When applied with data gathered and processed for a sample of WSSs and for all municipalities in mainland France, our method provided an overview of drinking water pipes both at national scale and at the scale of French WAs. It provides updated and improved asset knowledge, which can be used in the future for public policy management at these specific scales (*Assises de l'eau 2018*).

The technical results obtained through our method were used to perform an initial economic and financial assessment of assets within the French drinking water network as a whole. The replacement value of the networks has, as a first approach, been estimated at 135 billion euros. However, given the complexity of the issues determining cost, it will be necessary to carry out additional work on obtaining and processing economic and financial data to ensure effective results.

In terms of future update of the present study, there should be no problem reproducing this approach at a later stage, as all our modelling techniques are fully documented. However, it will be necessary to carry out a new data gathering campaign and a new comprehensive data processing.

To go further in the study of asset management strategies for drinking water networks on a large scale, INRAE and AFB are currently working together on a project aimed at identifying the links between asset characteristics and network performance.

ACKNOWLEDGEMENTS

The authors would like to thank all the WSS staff who provided us with their asset data, without which this project would have been impossible. We would also like to thank all the members of

the project committee for their assistance in collecting data and for their help in piloting the study throughout.

REFERENCES

- AFB 2018 *Observatoire des services publics d'eau et d'assainissement – Panorama des services et de leur performance en 2015 (Observatory of Public Water and Sanitation Services – Overview of Services and Their Performance in 2015)*. Les rapports Eaufrance, Vincennes, France, 81 pp.
- Assises de l'eau 2018 *Des réseaux renouvelés pour une eau de qualité (Renovated Networks for Quality Water)*. Communiqué de presse du gouvernement français (Press release from the French Government), Paris.
- ASTEE, AITF, ONEMA 2013 *Gestion patrimoniale des réseaux d'eau potable – Elaboration du descriptif détaillé des ouvrages de transport et de distribution d'eau, Coordination Charrière S (Drinking Water Networks Asset Management – Detailed Description of Water Transport and Distribution Networks, Coordination Charrière S)*. Onema, Vincennes, France, 48 pp.
- Cador, J. M. 2002 *Le patrimoine en canalisations d'AEP en France - Bilan des huit enquêtes départementales et estimation nationale (French Drinking Water Pipes Networks – Assessment of Eight Departmental Surveys and National Estimate)*. GEOPHEN, Université de Caen Basse, Normandie, France, 192 pp.
- Chessel, D., Thioulouse, J. & Dufour, A. B. 2004 *Introduction à la classification hiérarchique (Introduction to Hierarchical Cluster Analysis)*. Université de Lyon, Lyon, France, 56 pp. <https://pbil.univ-lyon1.fr/R/pdf/stage7.pdf>.
- Han-Mui, K. 2016 *Etude des déterminants et modélisation de la consistance du patrimoine des réseaux d'eau potable (Modelling the Consistency of Drinking Water Network Assets)*. ENGEES, Strasbourg, France, 77 pp.
- Husson, A., Vacelet, A., Stricker, A.-E., Brejoux, E. & Renaud, E. 2018 *Améliorer la connaissance des réseaux d'eau potable à l'échelle nationale: analyse des données disponibles (Improving Asset Knowledge of French Drinking Water Networks: Analysis of Available Data)*. TSM 10/2018, pp. 41–54.
- Maier, M. J. 2014 *DirichletReg: Dirichlet Regression for Compositional Data in R*. Institute for Statistics and Mathematics. Vienna University of Economic and Business, Vienna, Austria, 25 pp.
- Student 1908 The probable error of a mean. *Biometrika* **6** (1), 1–25.
- Venables, W. N. & Ripley, B. 2002 *Modern Applied Statistics with S*, 4th edn. Springer, New York, USA, 495 pp.

French regulation

- Décret n°2012-97 du 27/01/2012 relatif à la définition d'un descriptif détaillé des réseaux des services publics de l'eau et de l'assainissement et d'un plan d'actions pour la réduction des pertes d'eau du réseau de distribution d'eau potable (Definition of a detailed description of public water and sanitation networks and of a water loss reduction action plan)*, Journal officiel, n° 0024 du 28 janvier 2012.
- Instruction du 16/06/2015 relative au doublement de la redevance pour prélèvement sur la ressource en eau pour l'usage 'alimentation en eau potable' (Doubling of the water resource withdrawal fee for drinking water supply)*. Ministère de l'écologie, du développement durable et de l'énergie. Bulletin officiel, n° 2015-13 du 25 juillet 2015.
- Loi n°2010-788 du 12/07/2010 portant engagement national pour l'environnement (Grenelle II) (National commitment for environment)*, Journal officiel, n° 0160 du 13 juillet 2010.