Modelling daily water consumption through potential curves. Disaggregating apparent and real losses.

Sergi Grau[a]* and Ramon Pérez[b]

[a]*Aigües de Manresa, S.A, Plaça Plana de l'Om 6, Manresa 08241, Manresa, Spain;*
sgrau@aiguesmanresa.cat

 [b]*Research Center for Supervision, Safety and Automatic Control (CS2AC), Universitat Politècnica de Catalunya (UPC), Terrassa Campus, Gaia Building, Rambla Sant Nebridi 22, 08222 Terrassa, Barcelona, Spain; ramon.perez@upc.edu*

Modelling daily water consumption through potential curves. Disaggregating apparent and real losses.

This paper presents a model, based on potential curves, that describes the behaviour of the inverse of the daily cumulated frequency of the flows provided to a District Metered Area (DMA). The model has two terms, the first corresponds to the variable consumption due to the aggregation of demand patterns of consumers. The evolution of this term presents periodic behaviours with annual and weekly frequency. An extreme drought episode that affected Catalunya, reduced this parameter 19%. A second term presents exponential behaviour in its evolution and includes the real leakage. The leakage disaggregation together with the billing information allows the estimation of the apparent loses, 14.89% in the case study. The difficulty of estimating the parameters in a potential model, a complex problem of optimization, is simplified by applying mathematical moments. Hence, daily parameters become a linear relation of the daily moments that allows their algebraic operation.

Keywords: water demand, potential curves, model calibration, real losses, apparent losses, mathematical moment.

Motivation

There are two main questions for a utility that provides drinking water through a Water Distribution Network (WDN), which is the current leakage of the system and which is the future demand. In order to study the demand, its balance has to be understood. Lambert (2003) classifies the components of the water balance. These components have different behaviour depending on their origin. Recognising these behaviours from the available data requires models with parameters that pick them up.

(1) The components linked with the authorised customers' demand and apparent leakage are voluntary behaviours. They are brief, with determined start and end, generally with constant flow produced by the action of a mechanical device as a tap. The superposition of all these consumptions may present a daily, weekly or

annual periodicity. In the literature these components are extensively studied (Blokker 2010, Al-Humoud 2003).

(2) The components linked with the real leakage are involuntary behaviours. They are not periodic and with a flow that lasts days or months. Between the origin and the repair, they often grow exponentially (Rogerson 2017).

As the companies currently control strictly the consumptions, the unmetered authorized ones are reduced to zero. Thus, the study will be oriented to the metered authorised consumptions, apparent loses and actual leakage. The tools used for determining the actual leakage are of two types. First the analysis of the minimal night flow (MNF) as an indicator of leakage flow (Fox 2016), applied to Capinzal in South Brazil (Cheung 2010) or in Perak State in Malaysia (Alkasseh 2013). A second approach is based on the study of the variation of daily consumption or certain coefficients as indicators of the variation of leakage flow. In (Mazzolani 2016) a method based on the relation between flows at different time slots is presented. The comparison of Flow Pattern Distribution (CFPD) allows identifying consistent or inconsistent changes in the provided flow in order to detect a variation in leakage flow (Irons 2015, Van Thienen 2014). The novelty of the present work is to identify and evaluate all the components of the demand including real and apparent losses. Actual leakage flow determination, together with the registration of authorised consumptions, enables the apparent losses estimation, an economically crucial term for the water supply companies.

Besides the actual leakage there are other priority question regarding the resources control, the pumping costs and the sanitary guarantees. It is the forecast of provided consumption in a District Metered Area (DMA) in short time, generally 24 or 48 hours. The tools used for this short-term forecasting are based on the analysis of

historical data (daily, hourly or higher frequencies) as indicator of future flows due to the cyclic repetition of the subscriber's consumptions. Examples of short time demand forecast are those based on chaos theory (Nobuo Oshima 2015, Wei Xizhu 2006), temporal series (Puig 2017) or fully adaptive forecasting model (Bakker 2013, Shang 2006).

Data used, both for the actual leakage flow and the demand forecasting, are series of measurements of flowmeters. These temporal series are characterised by their sampling frequency, resolution and sensibility, understood this last as the relation between the resolution of the equipment and the data value. From these temporal series data, the MNF data is obtained with daily periodicity as the day minimum value. Thus, the provided daily flow can be expressed by the addition of a variable term and a constant term, the MNF. The variable term has defined origin, the subscriber's consumption, nevertheless the constant flow hasn't. This MNF or constant flow is the addition of two components: the actual leakage and some consumptions that are time sustained and overlapping with the low consumption time slots (watering gardens, filling pools, industrial activities). In larger DMA there is an effect of overlapping that may give a constant behaviour to consumptions that aren't. Therefore, it is not possible to estimate straightforward the different components of the water balance.

Data used in this work is provided by the company Aigües de Manresa that supplies water to the town of Manresa (76.000 inhabitants) in the centre of Catalunya and the surrounding villages. Water comes from river Llobregat through a channel of 27km. Since 2001 there is a telecontrol system managed by a SCADA WIZCON. 20 flowmeters have been selected, located at the outflow of a tank and the inflow of a DMA. These flowmeters have a minimal resolution of 0.1m3/h. The 20 DMA they supply have different characteristics, but none contains tanks (except pools) or pumping

groups. No pressure variation has been taken into account due to the large dimensions of the pipes given the flows. The flow range is between 10m3/day till 750m3/day. 18 temporal series of 28 months, one of 6 years and one of 17 years have been used. Sampling time is one minute. The acquisition algorithms are programmed in C++ and the validation and analysis algorithms are programmed in R.

Objectives and scope

Potential law is one of the most usual behaviours observed, both in nature (magnitude of earthquakes, diameter of moon craters, intensity of solar flares) and in social phenomena (word frequency, citations of scientific papers, web hits, copies of books sold, telephone calls, intensity of wars, wealth of the richest people, frequencies of family names and population of cities) (Newman 2005). However, these models have only a potential function that adjusts to their growth. In some series of real values, a second potential behaviour is observed for small values. Therefore, though a function of growing type can provide a very good adjustment for the largest values, a decreasing potential is required in the opposite end for a good fit in all the range of data. Most data of the flowmeters for water supply ordered from highest to lowest values present these two behaviours.

The objectives of this paper are:

(1) To create, identify and validate a model of potential curves that adjusts the inverse cumulated frequency (the values ordered in decreasing order against the normalised samples of the daily flows provided by a DMA). The potential curves should be estimated by a methodology with low computational and time costs.

(2) To use the parameters of potential curves to estimate the components of the water balance: registered and authorised consumptions, apparent losses and actual leakage. These parameters describe de behaviour of the subscribers and the leakages.

(3) Demand forecast is not the objective of this model. Nevertheless, demand forecasting based on estimated parameters using the state-of-the-art algorithms is used to validate the model.

This paper is organised as follows. Next section presents the model, its identification and validation on real data. Section 4 presents a simplified identification process that allows an online parameter estimation, it is validated too. Disaggregation of the demand components using the model's parameters is described in section 5. In section 6 the physical interpretation of the parameters is discussed while section 7 focusses on the linear manipulation of the models. Section 8 compares the results of demand forecast using the linear approximation proposed with those obtained on real flow data in order to validate the model. Finally, discussion and conclusions are presented.

Model for daily water consumption

Daily data of a flowmeter are organized in time series of work data (Y), with constant sampling. Each Y includes a constant number of samples (n_d) of consumption $\{Y_1 \dots Y_{nd}\}$ provided by a flowmeter. These values are ordered from highest to lowest. The frequency (f_i) for each of the work data (Y_i), defined by eq.1 (Mijares 1989), expresses the probability of having a value of flow rate greater than or equal to Y_i.

$$f_i = \frac{i}{n_d+1} \text{ where } 0 < f_i < 1 \text{ and } i \in \{1, \dots, n_d\} \tag{1}$$

We define $\mathbf{Y}^{sym}$ as the time series of n_d data $\{Y_{nd} \ldots Y_1\}$ where $f_{i_{sym}} = 1 - f_i$. We call $N \geq 0$ the minimal daily flow measured by the flowmeter.

$$N \approx Y_{nd} \tag{2}$$

Y_i flows adjust potential law where $a \geq 0\ and\ k \geq 0$ are the parameters that expresses the growth.

$$Y_i \approx k\ f_i^{-a} \text{ when } f_i \to 0 \tag{3}$$

The lowest flows present a potential behaviour with two possible options, concave or convex. These behaviours can be modelled using an inverse potential where $a' \geq 0\ and\ k \geq 0$ are the parameters that express the decrease degree. For $a' > 1$ the values present a concave behaviour for $a' < 1$ presents a convex behaviour.

$$Y_i \approx k\ (1 - f_i)^{a'} \text{ when } f_i \to 1 \tag{4}$$

Merging equations 3 and 4 the variable component can be modelled by:

$$Y_i \approx k\ f_i^{-a}\ (1 - f_i)^{a'} \tag{5}$$

The provided flow for a certain frequency (f_i) of the new model are calculated using equations 2 and 5:

$$\hat{Y}_i = k\ f_i^{-a}\ (1 - f_i)^{a'} + N \tag{6}$$

Or for the symmetric curve:

$$\hat{Y}_i^{sym} = k\ f_i^{a'}\ (1 - f_i)^{-a} + N \tag{7}$$

In Figure 1 ($\widehat{Y}$), the daily series of the flowmeter QX00111 has been adjusted to equation 6.

One of the key information that the daily time series contain is the daily measured volume:

$$V = \frac{\sum_{i=1}^{n_d} Y_i}{n_d} * 24 = \bar{Y} * 24 \tag{8}$$

Where:

$$\bar{Y} = \frac{\sum_{i=1}^{n_d} Y_i}{n_d} \tag{9}$$

Applying Taylor series, we get an expression of the model that enhances its analytical usage.

$$(1 - f)^{a'} = \sum_{n=0}^{\infty} f^n \, (-1)^n \binom{a'}{n}$$

Replacing in $\widehat{Y}_i$:

$$\widehat{Y}_i = k \, f^{-a} \, (1 - f)^{a'} + N = k \left(\sum_{n=0}^{\infty} f^{n-a} \, (-1)^n \binom{a'}{n} \right) + N$$

Hence:

$$\bar{\widehat{Y}} = k \left(\sum_{j=0}^{\infty} \frac{1}{(j-a+1)} \, (-1)^n \binom{a'}{j} \right) + N \tag{10}$$

Integrating, we find the expression of the estimated daily volume:

$$\widehat{V} = k \left(\sum_{j=0}^{\infty} \frac{1}{(j-a+1)} \, (-1)^j \binom{a'}{j} \right) * 24 + N * 24 = \bar{\widehat{Y}} * 24 \tag{11}$$

The parameters k, N, a, a' of $\hat{Y}_i$ are estimated by means of minimizing a combined error (Err_{com}) that includes the point error (Err_p) evaluation and the global shape error (Err_s) evaluation:

$$Err_{com} = Err_p + Err_s \tag{12}$$

Point error (Err_p) evaluates the difference between the pairs of values $Y_i - \hat{Y}_i$. It is usual in the literature to use the mean square error (MSE, equation 13) so that the highest errors are most weighted. Otherwise the absolute mean error (MAE, equation 14) can be used as well.

$$MSE = \frac{\sum_{i=1}^{n_d}(Y_i - \hat{Y}_i)^2}{n_d} \tag{13}$$

$$MAE = \frac{\sum_{i=1}^{n_d}|Y_i - \hat{Y}_i|}{n_d} \tag{14}$$

In our case the weight in the cost function has another aim. There are three parameters (a, a' and N) that are relevant from a qualitatively point of view and for an ulterior behaviour analysis but can be easily masked due to the few numbers of observations that are affected by them. Thus, it is essential to weight the extremes of the curves through the expression:

$$Err_p = \sum_{i=1}^{n_d}\left(1 + |Y_i - \hat{Y}_i|\right)^{n_e} \tag{15}$$

Where $n_e = 2 - \frac{P(Y_i)}{P_{max}}$, the normalized probability density function P() assures that n_e takes values between 1 and 2; for frequencies with high density of values, n_e tends to 1 (equivalent to MAE) and for frequencies with low density values, mainly the extremes, n_e tends to 2 (equivalent to MSE). We combine both effects of equation 15 for avoiding the distortion produced for the several orders of magnitude in the probability

density. Within the potential, a 1 is added to avoid the distorting effect that occurs when $\left|Y_i - \hat{Y}_i\right| < 1$.

Shape error (Err_s) is introduced for avoiding local minima induced by discontinuities in data or extreme values. This error measures the difference in the shape of measured time series ($\boldsymbol{Y}$ and $\boldsymbol{Y^{sym}}$) and their model counterparts ($\boldsymbol{\hat{Y}}$ and $\boldsymbol{\hat{Y}^{sym}}$) using mathematical moments with respect $f=0$ and $f=1$. These moments, that will appear to be very handy in the management of the models, are measurements of the shape of a mathematical function and are intensively used in mechanics and statistics.

Moment of degree n respect to $f=0$:

$$\mu_n = \sum_{i=1}^{n_d} f_i^n Y_i \tag{16}$$

Moment of degree n respect to $f=1$:

$$\mu_n^{sym} = \sum_{i=1}^{n_d}(1 - f_i)^n Y_i \tag{17}$$

Once again, Taylor series produce analytical expressions for the moments:

$$\hat{\mu}_n = k\left(\sum_{j=0}^{\infty} \frac{1}{(j-a+n+1)} (-1)^j \binom{a'}{j}\right) + \frac{N}{n+1} \tag{18}$$

$$\hat{\mu}_n^{sym} = k\left(\sum_{j=0}^{\infty} \frac{1}{(j+a'+n+1)} (-1)^j \binom{-a}{j}\right) + \frac{N}{n+1} \tag{19}$$

The shape error (Err_s) includes the moments till fourth degree.

$$Err_s = \sum_{n=0}^{n=4}(\mu_n - \hat{\mu}_n)^2 + \left(\mu_n^{sym} - \hat{\mu}_n^{sym}\right)^2 \tag{20}$$

Once the parameters are estimated. The model is validated with real data. In (Benett 2013) a set of indicators are used, we choose three of them used for the demand models (Bakker 2013, Alvisi 2007 and Adamowski 2012): The Mean Absolute

Percentage Error (MAPE); the Relative Root Mean Square Error (RRMSE); The Nash-Sutcliffe Model Efficiency (NSE). Table 1 values of these indicators when the model is used to estimate the daily volume and sampled flow for the 20 flowmeters, temporal series of 28 months from January 2016 to April 2018.

For the daily volume, the errors (*MAPE, RRMSE)* as a result of applying adjustment of the model are below 0.03% and 0.2% respectively while the efficiency (*NSE*) is over 0.999. For the ten-minute data the errors vary between 1.55%-8.49% and 3.84-35.17% respectively and the efficiency between 0.874-0.988. The worst performance obtained has not to do with the size of the DMA, but with the volume of extreme values. The frequency of extreme values is independent of the size of the DMA, in some cases it may depend on a large consumer or an industrial activity. Extreme values have an important weight in the error, and when their number is high, it can significantly increase the error.

Model for daily consumptions expressed in terms of its moments.

The time series of the n-moments for each flowmeter appear to be highly correlated (Figure 2). The Pearson's correlation coefficient for these time series are close to 1. The moments present scale invariance.

Empirically a relation between two consecutive moments has been found:

$$\frac{\mu_{n-1}}{\mu_n} \cong 1.68 \, n^{-1.145} + 1 \tag{21}$$

In the potential adjustment of previous section, the values of *a* and *a'* present a random behaviour with median $\tilde{a} = 0.086$ and $\widetilde{a'} = 1.079$. Selecting *a*=0.09 and *a'*=1, the expression of momenta (equation 18) simplifies:

$$\hat{\mu}_n = k \left[\frac{1}{n+0.91} - \frac{1}{n+1.91} \right] + \frac{N}{(n+1)} \tag{22}$$

In equation 21 we state that increasing the order n the new moment (n+1) tends to the previous one, furthermore figure 2 allowed a visual confirmation of this fact. For the purposes of calculation, only moments up to fourth order are considered to introduce new significant information. Expressing the simplified (eq. 22) form of the first 4 momenta in matrix form:

$$\begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \\ \hat{\mu}_4 \end{pmatrix} = \begin{pmatrix} 0.1799 & 0.5000 \\ 0.0879 & 0.3333 \\ 0.0521 & 0.2500 \\ 0.0345 & 0.2000 \end{pmatrix} \begin{pmatrix} k_\mu \\ N_\mu \end{pmatrix} \tag{23}$$

The estimation of k_μ and N_μ minimising $(\mu_n - \hat{\mu}_n)^2$ becomes a linear regression problem (Ljung 1987). The result of solving this problem are equation 24 and 25.

$$k_\mu = 13.44\,\mu_1 - 5.49\,\mu_2 - 10.35\,\mu_3 - 11.53\,\mu_4 \tag{24}$$

$$N_\mu = -2.96\,\mu_1 + 2.37\,\mu_2 + 3.65\,\mu_3 + 3.89\,\mu_4 \tag{25}$$

The total daily volume is calculated using them momentum $\hat{\mu}_0$ (eq.22):

$$V_\mu = 24\left(0.575\,k_\mu + N_\mu\right) \tag{26}$$

We take the potential model of equation 6 and simplify it with the results obtained in this section:

$$\hat{Y}_{\mu_i} = k_\mu\left(f_i^{-0.09} - f_i^{0.91}\right) + N_\mu \tag{27}$$

Demand disaggregation and apparent leakage estimation.

The relation of the flows for different time slots is used for determining the real leakage variation (Mazzolani 2016). Applying this concept to N_μ/k_μ, using the weekly mean for

both parameters, the time series obtained includes exponential curves with a sharp interruption coinciding with a leak repair (Figure 3). These kind of curves, exponential growth with sharp end, can be related to real leaks (Rogerson 2017). The signal presents little noise and high stability. The calculation through the momenta filters it.

A proportionality is observed between k_μ, associated with the variable fraction of consumption, and N_μ, associated with the constant fraction of daily consumption (Figure 3). In addition, the time series of N_μ/k_μ does not present oscillation with annual period like those observed in Y_{nd} or N_μ, in other words, the normalisation of N_μ by k_μ eliminates the fraction related with the consumers' demand. Thus, elevated values in N_μ/k_μ correspond with elevated volumes of real leakage.

N_μ is the addition of two signals, $N_{\mu s}$ and $N_{\mu l}$. The first one follows the periodicity of k_μ and thus it can be associated with the consumers' demand while the second one follows a pattern of consecutive exponentials associated with the real leaks:

$$N_\mu = N_{\mu s} + N_{\mu l} \qquad (28)$$

From N_μ/k_μ we can obtain the value of $N_{\mu l}$ multiplying by a factor related with the provided volume, this factor is defined by $k_{\mu_{year}}$, the annual mean of the daily variable fraction of V_μ (eq.26), and a coefficient S_μ characteristic of the sector.

$$N_{\mu l} = (N_\mu/k_\mu) \, k_{\mu_{year}} S_\mu \qquad (29)$$

S_μ is obtained constraining that, throughout its time series, $N_\mu - N_{\mu l} > 0$. For the flowmeters used in this work, S_μ has values between 0.5 and 0.8.

We obtain $N_{\mu s}$ using eq.28:

$$N_{\mu s} = N_\mu - N_{\mu l} \qquad (30)$$

The total daily volume V_μ is expressed (eq.26 and eq.28):

$$V_\mu = 24 \left(0.575\, k_\mu + N_{\mu\,s} + N_{\mu\,l} \right) \tag{31}$$

Where the daily volume associated with the consumer's demand is:

$$V_{\mu_s} = 24 \left(0.575\, k_\mu + N_{\mu\,s} \right) \tag{32}$$

And the daily volume associated with real leakage is:

$$V_{\mu_l} = 24 \left(N_{\mu\,l} \right) \tag{33}$$

Figure 4 presents the example of flowmeter QX00111 located at the input of a residential area. The first graphic shows the daily volume V for 17 years. It can be separated clearly in two components. On the one hand there is V_{μ_s}, volume with periodic behaviour with annual oscillation, maximum in summer and minimum in winter, coherent with the consumers' demand in residential areas. On the other there is V_{μ_l}, volume where different exponentials can be observed, they are coherent with the real leaks. The annual minima of V follow the behaviour of V_{μ_l} while the annual minima of V_{μ_s} are constants, this reinforces the relationship of V_{μ_l} with the real leakage, and V_{μ_s} with consumptions not affected by repairs.

Following the classification of the first section, the flow components are three. The authorised metered consumptions (mc), apparent losses (al) and real losses (V_{nd}). The MNF is an indicator for real losses (Fox 2016), in our case it is represented by Y_{nd}. Therefore, the water balance can be adjusted to the expression:

$$V = mc + al + V_{nd} \tag{34}$$

$$V_{nd} = \sum_{days} (Y_{nd} * 24)$$

Where:

$$al = V - V_{nd} - mc \qquad (35)$$

Using V_{μ_l}, we obtain an alternative expression for the apparent losses$_\mu$ (al_μ):

$$al_\mu = V - V_{\mu_l} - mc \qquad (36)$$

From equation 35 and equation 36 we calculate the al and al_μ for the sector provided by flowmeter QX09698 throughout 22 trimesters with $S_\mu = 0,69$, where V_T is the trimestral volume measured at the input of the sector. We estimate the real trimestral leakage from V_{nd} and V_{μ_l} as the sum of the daily values (Table 2).

Contrary to V_{nd} the calculation of V_{μ_l} does not overestimate leakage. al_μ present a more clearly periodical behaviour winter/summer, typical in residential areas, than al do. It should be highlighted that, in this particular case, the mean, for the 22 trimesters, of al was of 1.3%, too low, while the mean of al_μ becomes 14,9%, a more realistic estimation and with no negatives values.

The behaviour of V_{μ_l} and V_{nd} are similar, nevertheless the values are lower in V_{μ_l} which avoids some of the peaks observed in V_{nd} that coincide with periods of higher consumption. These peaks are transmitted to al_μ.

Physical interpretation of the parameter k_μ

k_μ is the principal parameter of the variable part of the daily volume. It is a good indicator of the consumers' demands. As it does not include the real leakage flow it

shows clearly the induced behaviours. As an example, we take the flowmeter QX18909, located in a touristic town of the Pyrenees (Figure 5).

k_μ, in red, shows clearly the users' pattern of consumption. There are peaks in the weekends, higher peaks for bank holidays, Eastern and holiday period in August. There are annual patterns that are repeated clearly (Eastern, New Year Eve, and August), there is a weekly oscillation where the value in the working days depend on the snow and the year period.

Another example is flowmeter QX00111, it provides water to a residential area where $k_{\mu_{year}}$ (yearly mean for k_μ) presents a sharp change between summer of 2007 and summer of 2009 coinciding with an extreme drought episode that affected Catalunya. The Catalan government published a decree that lasted from April 2007 till January 2009 that promoted a water saving campaign and increased the taxes. This action changed the habits of the highest and luxury consumers (Figure 6). New habits persisted after the drought. However, it didn't affect the basic consumes which kept inelastic.

Till 2007 $k_{\mu_{year}}$ value was allways over 7, with peaks that got 9 and a mean of 8.0; after 2009 it has been always bellow 7 with a mean of 6.5. The minimum value, below 6, is found in summer 2008 in full drought.

Addition of frequency distributions.

The model for daily consumptions can be expressed linearly in terms of its moments. The linear properties of this model can be used for studying global systems composed by some sectors, which allows modelling the joint flow of various flowmeters by means of the arithmetic sum of the parameters of the model.

Given the linear relation of parameters k_μ, N_μ (equation 27) the sum of m daily series may be expressed in terms of $k_{\mu_{sum}}$ and $N_{\mu_{sum}}$ calculated by summing their m parameters k_μ, N_μ.

$$k_{\mu_{sum}} = \sum_{1<i<m} k_{\mu_i} \tag{37}$$

$$N_{\mu_{sum}} = \sum_{1<i<m} N_{\mu_i} \tag{38}$$

The parameters $k_{\mu_{sum}}$ and $N_{\mu_{sum}}$ can be calculated using the sum of the m momenta due to their linear form (equations 22, 24 and 25).

$$\hat\mu_{n\,sum} = \sum_{1<i<m} \hat\mu_{n_i} \tag{39}$$

From equations 26 and 27 we can estimate the total volume of the global system $V_{\mu_{sum}}$ and its distribution model $\widehat{Y}_{\mu_{sum}}$. Both calculus can be done directly adding the flows of the m sensors (V_{sum} and $\mathbf{Y}_{sum}$) while N_μ can be estimated taking the lowest flow value in the time series $Y_{nd_{sum}}$. The comparison of both results appears in table 3. It presents mean values for 1 year and the errors associated. The System is the aggregation of the 10 sensors of table 1, the subset of sensors in table 1 that are complete enough in data.

Demand forecast

The linear approximation induces inaccuracy in the information contained by the model. In order to validate this approximation, its parameters (both the original model and its linear approximation) have been used to forecast the consumed volumes and compared with forecast based on real flow data.

We apply the fully adaptive forecasting model for short-term (Bakker 2013) to the 10 selected flowmeters of previous section. We apply this methodology to predict in a horizon of 24 and 48 hours the parameter k_μ each 15 minutes and with the last N_μ (its dynamic is much slower so in these horizons it can be assumed constant) the daily volume $V_{\mu\,forecast}$ is predicted(both to the measured volumes and to the volumes obtained by means of the simplification (Y_μ)). Table 4 shows the results obtained by this method compared with the actual daily volume V measured by means of the error indicators *MAPE*, *RRMSE* and *NSE* used in (Bakker 2013).

Comparing the errors in our 10 systems with those obtained from literature it is obvious that this error decreases with the size of the sector (evaluated for the mean flow). In order to have a sector of a comparable size with those used in the previous work the 10 flowmeters are aggregated. The methodology proposed in previous section eases this process. Table 5 shows the result of applying the forecast using the time series replicating the results of Bakker with a similar forecast error. The last row presents the error obtained using the potential model generated by the moments.

The result obtained with the sum of the 10 flowmeters is in the range of the results obtained by Bakker & others (2013); Values are similar to those of Valkenburg in terms of daily flow as well as fifteen-minute flow rate. In addition, we note that the error of the sum of the 10 flowmeters *MAPE*, *RRMSE*, and *NSE* is lower than the 10 flowmeters calculated individually (Table 4 and 5).

The values of *MAPE* show a logarithmic behaviour depending on the average flow (Figure 7). The sectors larger of the 10 work flowmeters, and the result of the sum of all, have similar values to the smaller sectors simulated by Bakker & others (2013) (Valkenburg and Hulsberg)

Conclusions

The inverse of the cumulative frequency of the daily flows sampled each 10 minutes of a DMA can be modelled using a two-term potential law and a constant value. The modelling error after the adjusting process is small both for the daily flow and the 10 minutes samples. The adjusting methodology requires a sophisticated error that includes form factors. Thus, the form indicators (moments) are explored in order to enlighten this behaviour.

A relation between the consecutive order momenta is observed. Moreover, two of the parameters of the model present stochastic distribution and can be taken constants as their variation is very low. Thus, the potential model can be obtained directly from a linear combination of the moments avoiding the optimisation process involved in the parameter estimation.

The interest of this potential model resides in the physical interpretation of its parameters. The first parameter (k_μ) is associated with the daily variable fraction of consumption. It is a good indicator of patterns and behaviours of the customers where the induced flows are highlighted. The second (N_μ) associated with the constant daily fraction can be subdivided in a potential behaviour ($N_{\mu l}$) associated with real leaks and a second one ($N_{\mu s}$) with a periodic behaviour with higher period associated with flows induced by the customers. This last term is valuable for the apparent losses estimation improving the state of the art based on MNF.

The linear properties of momenta and the model itself ease the aggregation of flowmeters for the analysis of bigger systems. The forecast methods can be applied to the parameters of the model in order to estimate the short term volume required by the network. The evolution of these parameters enhances the capacity of studying the evolution of customers' behaviours in the long term.

The results of this work encourage the effort in obtaining, storing and processing great amount of data further than 10 or 15-year horizon in order to better manage the assets and the control of such a complex system as the WDS are.

In order to improve the adjustment of the potential model, the frequency and resolution of the time series data of the flow meters should be increased. Another improvement would come with the installation of telemetering counters, even if it was partially in a DMA, since the analysis of the data would improve the adjustment between real losses and apparent losses.

Notation

The following symbols are used in this paper:

DMA District Metered Area

MSE Mean squared error

MAE Mean absolute error

$MAPE$ Mean Absolute Percentage Error

$RRMSE$ Relative Root Mean Square Error

NSE Nash-Sutcliffe Model Efficiency

Y Series of work data, a single series per day and flowmeter, with n_d values $\{Y_1 \ldots \ldots Y_i \ldots \ldots Y_{nd}\}$

Y_{sum} Series of work daily data adding the flows of m flowmeters, with n_d values $\{Y_{sum_1} \dots \dots Y_{sum_i} \dots \dots Y_{sum_{nd}}\}$

f_i Frequency associated with the order number of Y_i in $\boldsymbol{Y}$

Y_{nd} Minimal daily flow measured

$Y_{nd_{sum}}$ Minimal daily flow measured adding the flows of m flowmeters

$\bar{Y}$ Arithmetic mean of work daily data Y_i

$\bar{\bar{Y}}$ Arithmetic mean of the daily arithmetic mean

$\hat{Y}$ Potential adjust data of $\boldsymbol{Y}$ series

$\hat{Y}_\mu$ Potential adjust data of $\boldsymbol{Y}$ series calculated through moments

$\hat{Y}_{\mu\,sum}$ Potential adjust data of $\boldsymbol{Y}_{sum}$ series calculated through moments

$\bar{\hat{Y}}$ Arithmetic mean of the daily potential adjust data

$\bar{\bar{\hat{Y}}}$ Arithmetic mean of the daily arithmetic mean

$\boldsymbol{Y^{sym}}$ Series with the order of the inverted data, a single series per day and flowmeter, with n_d values $\{Y_{nd} \dots \dots Y_i \dots \dots Y_1\}$

$\hat{Y}^{sym}$ Potential adjust data of $\boldsymbol{Y^{sym}}$ series

V Daily flow volume calculated by $\boldsymbol{Y}$ series

V_{sum} Sum m daily flow volume V

V_T Trimestral volume measured

$\bar{V}$ Arithmetic mean of V

$\hat{V}$ Daily flow volume, calculated with $\hat{Y}$

$\bar{\hat{V}}$ Arithmetic mean of $\hat{V}$

V_μ Daily flow volume, calculated with the adjust potential daily data through moments

$V_{\mu\,s}$ Daily flow volume V_μ associated with the consumer's demand

$V_{\mu\,l}$ Daily flow volume V_μ associated with real leakage

$V_{\mu\,sum}$ Sum m daily flow volume V_μ

a Exponent of the potential function, daily calculated

a' Exponent of the potential function, daily calculated

k Parameter of the potential function, daily calculated

k_μ Parameter of the potential function, daily calculated through moments

$k_{\mu\,year}$ Annual daily mean k_μ

$k_{\mu\,sum}$ Sum m daily parameters k_μ

N Offset of the potential function, daily calculated

N_μ Offset of the potential function, daily calculated through moments

$N_{\mu\,s}$ Fraction of N_μ, associated to consumers

$N_{\mu\,l}$ Fraction of N_μ, associated to real leaks

S_μ Coefficient characteristic of the sector, to calculate $N_{\mu\,l}$

$N_{\mu\,sum}$ Sum m daily parameters N_μ

μ_n Moment of degree n respect to f=0, calculated with the work daily data

μ_n^{sym} Moment of degree n respect to f=1, calculated with the work daily data

$\hat\mu_n$ Moment of degree n respect to f=0, calculated with the adjust potential daily

data

$\hat\mu_{n\,sum}$ Sum m daily moments $\hat\mu_n$

Err_{com} Combined error

Err_p Point error

Err_s Shape error

mc Metered Consumption

al Apparent Losses

al_μ Apparent Losses calculated by adjust potential data through moments

$P()$ Normalized probability density function

References:

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., and Sliusarieva, A. (2012). "Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada". Water Resources Research. 48 (1).

Al-Humoud, J. M., & Al-Ghusain, I. (2003). Household demand for water: A case study in Kuwait. Kuwait Journal of Science and Engineering, 30(1), 197–211.

Alkasseh, J. M. a, Adlan, M. N., Abustan, I., Aziz, H. A., & Hanif, A. B. M. (2013). Applying Minimum Night Flow to Estimate Water Loss Using Statistical Modeling: A Case Study in Kinta Valley, Malaysia. Water Resources Management, 27(5), 1439–1455.

Alvisi, S., Franchini, M., and Marinelli, A. (2007). "A short-term, pattern-based model for water-demand forecasting". Journal of Hydroinformatics. 9 (1): 39-50.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., and Andreassian, V. (2013). "Characterising performance of environmental models". Environmental Modelling and Software. 40: 1-20. Blokker, E. J. M., Vreeburg, J. H. G., & Van Dijk, J. C. (2010). "Simulating Residential Water Demand with a Stochastic End-Use Model". *Journal of Water Resources Planning and Management*, *136*(1), 19–26.

Cheung, P. B., Guilherme, V., Abe, N., & Propato, M. (2010). Night flow analysis and modeling for leakage estimation in a water distribution system, (August), 509–513.

Feng Shang, James G. Uber, Bart G. van Bloemen Waanders, Dominic Boccelli and Robert Janke (2006) Real Time Water Demand Estimation in Water Distribution System.

Fox, S., Collins, R., & Boxall, J. (2015). Traditional leakage models for leakage modelling: Effective or not? Procedia Engineering, 119(1), 35–42.

Irons, L. M., Boxall, J., Speight, V., Holden, B., & Tam, B. (2015). Data driven analysis of customer flow meter data. Procedia Engineering, 119(1), 834–843.

Lambert, M., Simpson, A., Vítkovský, J., & Wang, X. (2003). "A Review of Leading-edge Leak Detection Techniques for Water Distribution Systems". Awa Convention Australia.

Ljung, L. "System Identification, Theory for the user". Prentice Hall. New Jersey 1987.

M. Bakker, H. van Duist, K. van Schagen, J. Vreeburg, L. Rietveld (2013). "Improving the performance of water demand forecasting models by using weather input"

Mazzolani, G., Berardi, L., Laucelli, D., Martino, R., Simone, A., & Giustolisi, O. (2016). A Methodology to Estimate Leakages in Water Distribution Networks Based on Inlet Flow Data Analysis. Procedia Engineering, 162, 411–418.

Mijares, F. J. A. (1989). Fundamentos de Hidrología de Superficie. MEXICO: EDITORIAL LIMUSA, S.A.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. Power Laws, Pareto Distributions and Zipf's Law. Contemporary Physics, 46(5), 323–351. http://doi.org/10.1016/j.cities.2012.03.001

Oshima, N. (2015). Information Integration Type Chaos Theory-Based Demand Forecasting for Predictive Control of Waterworks. Water Purification Technologies.

Puig, V., Ocampo-Martínez, C., Pérez, R., Cembrano, G., Quevedo, J., & Escobet, T. (Eds.). (2017). *Real-time Monitoring and Operational Control of Drinking-Water Systems*. Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-50751-4

Sarah Rogerson (2017) Tactical leakage targeting, a modelling and control system to AID efficient leakage strategies

Thien, P. van, & Montiel, F. (2014). Flow Analysis and Leak Detection With the Cfpd Method in the Paris Drinking Water Distribution System. 11 International Conference on Hydroinformatics HIC 2014, New York City, USA.

Xizhu, W. (2006). Forecasting of Urban Water Demand Based on Chaos Theory. 2007 Chinese Control Conference, 441–444.

Table 1. Error indicators arithmetic mean of the daily volume between $\bar{V}$ and $\bar{\hat{V}}$, and arithmetic mean of the daily arithmetic mean $\bar{\bar{Y}}$ and $\bar{\bar{\hat{Y}}}$.

	$\bar{V}$ [m3/day]	$\bar{\hat{V}}$ [m3/day]	MAPE [%]	RRMSE [%]	NSE [-]	$\bar{\bar{Y}}$ [m3/h]	$\bar{\bar{\hat{Y}}}$ [m3/h]	MAPE [%]	RRMSE [%]	NSE [-]
QX00111	148.09	148.07	0.0162	0.0376	0.999999	6.16	6.16	3.33	6.82	0.988
QX00546	56.76	56.74	0.0377	0.0738	0.999997	2.38	2.35	6.38	13.24	0.955
QX00798	45.63	45.63	0.0071	0.0276	1.000000	1.90	1.89	5.39	12.10	0.977
QX01451	27.49	27.49	0.0052	0.0144	1.000000	1.15	1.15	7.80	10.92	0.975
QX01674	69.74	69.72	0.0204	0.0422	0.999998	2.90	2.89	3.73	6.83	0.986
QX01722	19.07	19.07	0.0074	0.0296	1.000000	0.79	0.79	4.07	8.03	0.988
QX02811	117.02	117.01	0.0081	0.0471	1.000000	4.92	4.94	8.48	20.14	0.983
QX04654	520.70	520.63	0.0171	0.0293	0.999994	21.72	21.66	3.21	4.75	0.977
QX05456	142.10	142.08	0.0128	0.0246	0.999999	5.91	5.90	4.09	7.42	0.986
QX09698	50.58	50.57	0.0266	0.1242	0.999988	2.11	2.09	5.60	18.71	0.926
QX10922	8.41	8.41	0.0107	0.0430	0.999999	0.35	0.35	7.70	18.06	0.972
QX11016	98.26	98.24	0.0143	0.0368	0.999999	4.11	4.11	3.80	7.50	0.983
QX11164	224.01	223.97	0.0204	0.0438	0.999999	9.41	9.38	8.49	14.35	0.970
QX13281	379.74	379.69	0.0155	0.0398	0.999999	15.96	15.83	8.04	11.28	0.970
QX15509	54.64	54.64	0.0063	0.0164	0.999999	2.27	2.27	1.55	3.84	0.972
QX15088	58.71	58.70	0.0209	0.0549	0.999997	2.45	2.44	8.19	12.93	0.946
QX15737	733.93	733.80	0.0220	0.0439	0.999991	30.73	30.54	6.26	8.12	0.963
QX16687	63.42	63.41	0.0087	0.0439	0.999998	2.41	2.40	5.59	11.92	0.965
QX16790	50.44	50.43	0.0071	0.0341	1.000000	2.11	2.09	6.22	35.17	0.874
QX18909	566.32	566.21	0.0208	0.0379	0.999999	23.65	23.53	5.30	7.32	0.979

Table 2. Trimestral balance, in m^3, for the flowmeter *QX09698*.

Data reading	V_T	mc	V_{nd}	$V_{\mu l}$	al	al_μ
09/01/2012						
10/04/2012	3,602.4	2,115.0	1,648.6	1,253.9	-161.3	233.5
09/07/2012	3,614.0	2,403.0	1,046.4	629.9	164.6	581.1
09/10/2012	5,216.3	3,277.0	2,085.1	948.1	-145.7	991.2
04/01/2013	3,726.9	1,846.0	2,130.8	1,858.5	-249.9	22.4
08/04/2013	3,575.2	1,768.0	1,866.1	1,562.0	-58.9	245.2
05/07/2013	3,687.8	2,177.0	1,303.2	712.7	207.6	798.1
07/10/2013	5,876.7	2,913.0	2,637.6	1,186.1	326.1	1,777.6
07/01/2014	3,535.5	2,223.0	1,377.6	989.1	-65.1	323.4
08/04/2014	3,404.6	2,122.0	1,504.3	1,126.6	-221.6	156.0
04/07/2014	4,155.4	2,585.0	1,399.2	655.3	171.2	915.1
07/10/2014	5,070.0	3,139.0	1,924.8	1,034.1	6.2	896.9
08/01/2015	3,312.0	2,091.0	1,198.5	902.0	22.5	319.0
08/04/2015	3,340.8	2,039.0	1,264.8	1,105.8	37.0	196.0

07/07/2015	4,094.2	2,845.0	609.6	210.9	639.6	1,038.4
07/10/2015	4,450.5	3,448.0	816.0	328.3	186.5	674.3
07/01/2016	3,539.6	2,255.0	1,123.4	704.1	161.2	580.5
07/04/2016	3,905.9	2,602.0	1,557.6	1,110.6	-253.7	193.3
08/07/2016	3,958.0	2,631.0	880.8	426.5	446.2	900.5
07/10/2016	5,513.7	3,631.0	1,977.6	861.7	-94.9	1,021.0
10/01/2017	4,721.7	2,193.0	2,557.0	1,942.4	-28.3	586.3
09/04/2017	3,427.1	2,155.0	1,454.4	1,258.3	-182.3	13.7
10/07/2017	4,330.8	2,964.0	1,070.4	417.9	296.4	948.9
TOTAL %	90,059.2	61.54%	37.12%	23.57%	1.34%	14.89%

Table 3. *MAPE, RRMSE and NSE between* Y_s *and* $\widehat{Y}_{\mu_s}$; V_s *and* $\widehat{V}_{\mu_s}$; Y_{nd_s} *and* $\widehat{N}_{\mu_s}$.

	Mean	MAPE [%]	RRMSE [%]	NSE [-]
Y_s	98.83	10.26	14.80	0.834
$\widehat{Y}_{\mu_s}$	100.81			
V_s	2,353.23	2.99	3.28	0.969
$\widehat{V}_{\mu_s}$	2,421.86			
Y_{nd_s}	40.79	5.23	6.76	0.917
N_{μ_s}	41.51			

Table 4. *Errors between* V *and* $V_{\mu forecast}$.

	$\overline{V}$ [m3/day]	$\overline{V}_{\mu\,forecasting}$ [m3/day]	MAPE [%]	RRMSE [%]	NSE [-]
QX00111	148.09	151.45	9.83	13.61	0.900
QX00798	45.63	47.43	18.23	28.94	0.661
QX01451	27.49	28.08	12.09	17.21	0.627
QX01674	69.74	70.80	8.43	11.92	0.863
QX01722	19.07	19.63	15.90	23.79	0.819
QX04654	520.70	520.83	4.03	5.71	0.776
QX05456	142.10	145.91	9.65	14.16	0.818
QX011016	98.26	102.33	12.40	17.98	0.803
QX015737	733.92	747.78	5.17	7.75	0.702
QX018909	566.31	576.33	8.75	14.35	0.721
SUM	2,353.23	2,327.06	3.83	5.08	0.911

Table 5. *Forecast errors for daily Volume for 6 sectors in Netherlands and Manresa.*

Area	$\bar{y}$ [m³/h]	MAPE [%]	RRMSE [%]	NSE [-]
1. Amsterdam	7,540	1.44	2.01	0.785
2. Rhine area	2,295	1.86	2.78	0.710
3. Almere	1,160	2.12	3.12	0.740
4. Helden	291	3.40	5.17	0.803
5. Valkenburg	73	3.49	4.83	0.802
6. Hulsberg	18	5.12	8.21	0.658
Manresa	98	3.83	5.08	0.911
Manresa $V_{\mu\,forecasting}$	98	5.13	6.80	0.869

Figure 1. Inverse cumulated frequency of data for flowmeter QX00111 on 26/09/2011 adjusted by two potential curves. The data are ordered in decreasing order and the number of samples is normalised in X-axis.
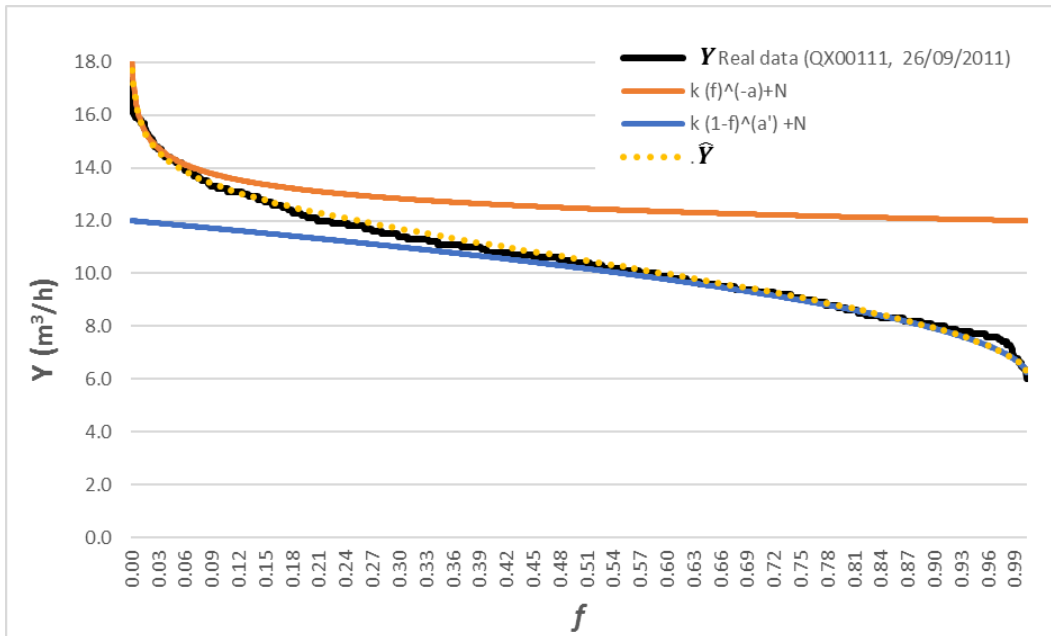
Figure 2. Time series of μ_0 μ_1 μ_2 μ_3 μ_4 for the flowmeter QX00111
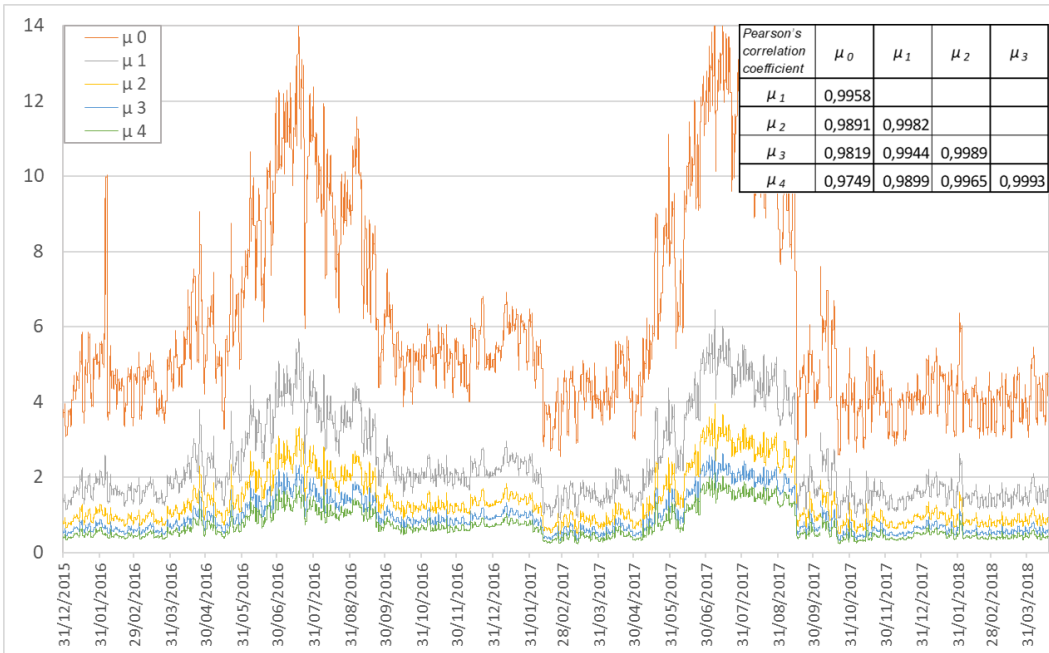
Figure 3. Time series of k_μ, N_μ and N_μ/k_μ for flowmeter QX00111.
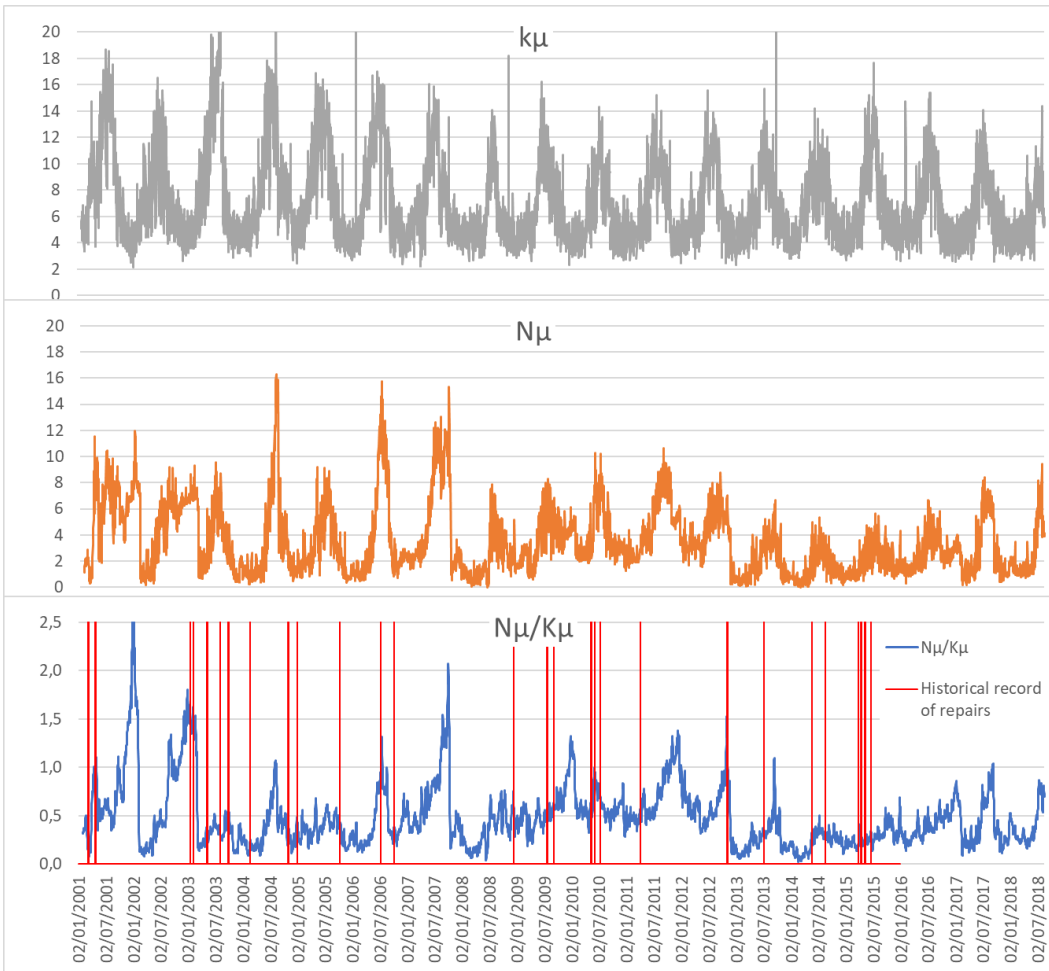
Figure 4. Temporal series of V, V_{μ_s} and V_{μ_l} for flowmeter QX00111 with S_μ=0.57.
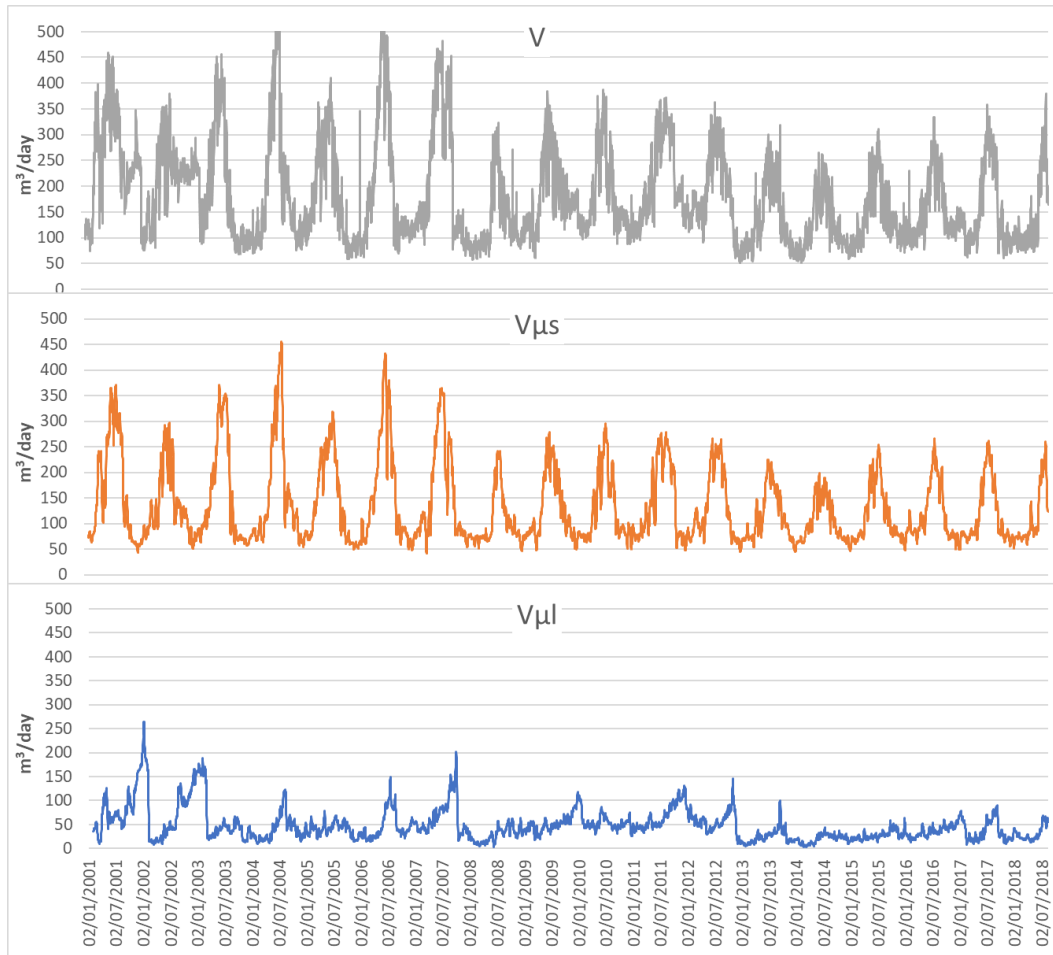

Figure 5. Parameter k_μ for flowmeter QX18909 (19/10/2016-5/10/2018)
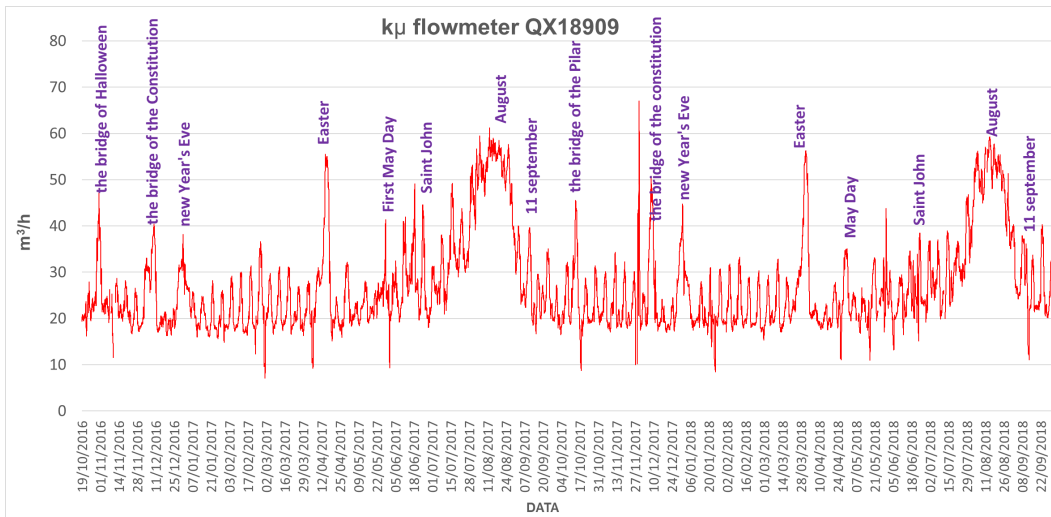
Figure 6. Parameter $k_{\mu_{year}}$ for the flowmeter QX00111 between 2001and 2018

Figure 7. MAPE depending on the logarithm of mean daily flow.