A method for quantitative discrimination in flow pattern evolution of water distribution supply areas with interpretation in terms of demand and leakage

Peter van Thienen

ABSTRACT

A method, the comparison of flow pattern distributions (CFPD), is described in which the specific representation of flow measurements for two different time periods allows a direct, quantitative interpretation of changes in the pattern. Two types of changes can be distinguished. The first is changes from one period to the next in demand consistent with the existing pattern, e.g. due to changing weather or changes in the population size. The second type is inconsistent changes which may be due to increased leakage. The method is successfully applied to drinking water distribution systems of different sizes and characteristics. Being data driven, it is independent of model assumptions and therefore insensitive to uncertainties therein which may hinder some other leakage determination methods. Because it is simple to implement and apply but nevertheless powerful in distinguishing between consistent and inconsistent changes in water demand, the method provides water companies with a way to constantly monitor their networks for possible changes in customer demand and the possible occurrence of new leakages and also check archived data for similar changes. This could render additional information about customer behavior and the evolving condition of the network from data which is usually readily available at water companies.

Key words | demand patterns, leakage, numerical methods

Peter van Thienen
KWR Watercycle Research Institute,
Post Box 1072,
3430 BB Nieuwegein,
The Netherlands
E-mail: *peter.vanthienen@kwrwater.nl*

INTRODUCTION

Knowledge and understanding of flow patterns into drinking water supply areas are important for the proper operation, maintenance and rehabilitation of existing drinking water distribution systems and for the design of new networks. Flow patterns may also provide valuable information about the occurrence of leaks and bursts, but this information is not always easy to distill from the data.

Leakage continues to be a problem for water companies around the world, with numbers ranging from 3% to more than 50% (Lambert 2002; Beuken *et al.* 2006). The water which is lost in this way represents a financial value, but its disappearance is also undesirable from a sustainability point of view. Classically, the two main methods to determine the amount of non-revenue water (NRW; water losses including, in addition to leakage, unauthorized consumption and unbilled authorized consumption) in a supply area are the top-down and the bottom-up methods (Farley & Trow 2003; Wu 2011). The top-down method consists of a water balance in which the registered amount of water delivered to a supply area over the period of a year is compared to the billed amount of water. The bottom-up method essentially compares the minimum flow rate during the quiet night hours into a district metered area (DMA) or demand zone or the integrated flow of a 24-h period to an estimate for the demand in this DMA or demand zone based on the number of connections (Puust *et al.* 2010).

Because the former method is rather labor-intensive and has large error margins (Farley & Trow 2003) and the latter method is only applicable to small supply areas

doi: 10.2166/hydro.2012.171

with predictable demands, several researchers have tried to develop alternative methods to determine the amount of NRW and/or leakage. Also, methods to determine the location of leakages have been the focus of research. Several methods which combine both are based on the optimization of a hydraulic model including leaks to match measurements of flow rate and/or pressure. These include inverse transient analysis (Liggett & Chen 1994; Savic *et al*. 2005; Vítkovský *et al*. 2007), probabilistic leak detection (Poulakis *et al*. 2003; Puust *et al*. 2006), and pressure dependent leak detection (Wu *et al*. 2010). Applications of these methods to real water distribution networks are few (e.g. Saldarriaga *et al*. 2006; Wu *et al*. 2010). These model optimization based methods generally require a hydraulic model of (or information about the system in) the supply area and can be computationally demanding (Vítkovský *et al*. 2007; Colombo *et al*. 2009). Developments in other transient test-based techniques for the detection of leaks and illegal connections are promising (Menicone *et al*. 2011).

A statistical approach to leak detection from night flow patterns was presented by Buchberger & Nadimpalli (2003). However, the literature does not contain any reports on application of the method in practice, and our own investigation of night flow data (see Figure 1) shows that, at least in the cases which were considered, the assumption in their method that the statistics of the night flow pattern follow a truncated normal distribution does not hold.

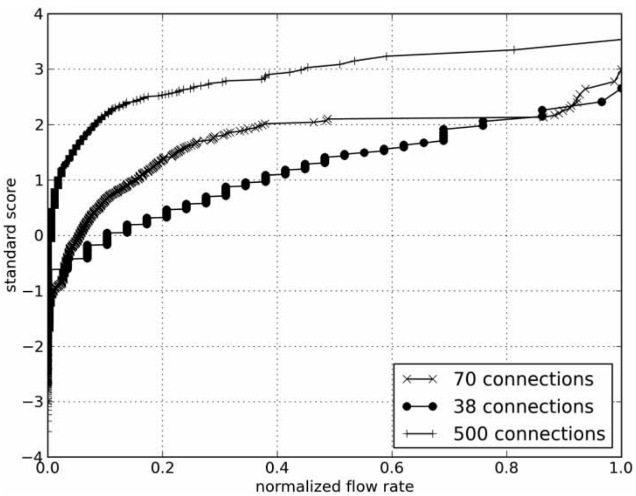

Figure 1 | Normal probability plots (with scaled flow rates) for three data sets from the Netherlands (38 and 500 connections) and Belgium (70 connections), each of which show significant periods of zero demand during the night. None of these curves is a straight line, which would indicate a (truncated) normal distribution of the data.

In addition to these mostly retrospective methods for determining leakage rates, several sensor based field methods for locating leaks and on-line monitoring techniques for burst detection are described in the literature. For an overview, the reader is referred to, for example, Wu (2011).

In this paper, a method called the comparison of flow pattern distributions (CFPD) is presented which allows its user to compare flow patterns of arbitrary duration for an arbitrarily sized supply area and distinguish consistent from inconsistent changes in the pattern. Consistent and inconsistent changes will be defined below. The former can be interpreted in terms of changes in demand due to changes in the population characteristics (growth or shrink on longer term, holiday periods on shorter term). The latter can be interpreted in terms of new large volume customers, new types of water use or a change in leakage. As water companies have (access to) information about the first two (and provided their interpretations are correct), the method allows quantitative statements about the third: leakage.

The method presented here is relatively simple, not computationally intensive, independent of any model assumptions and easily implemented. Nevertheless, it provides water companies with a new tool to monitor their distribution systems on arbitrary time scale for possible changes in customer demand and the possible occurrence of new leaks. Also, archived flow data can be checked for the occurrence of new leaks which may still be present. When looking for leaks, the CFPD method provides an alternative for the classical top-down and bottom-up methods, rendering more information with fewer assumptions. More generally, the method provides additional information about both customer behavior and the evolving condition of the network from data which is usually readily available at water companies. This paper aims to describe the basic principles of the CFPD method, illustrate its operation and the interpretation of its results with a number of simple field data sets, test its sensitivity, and discuss directions for further research and development and suggested application at water companies.

METHODS

Procedure

Consider a supply area for which the flow rate into the area (accounting for all inflow, outflow and storage) is registered for a period of time (e.g. a day, a week, a month or an entire year) and again for a comparable (in terms of parameters affecting water use, such as average air temperature, rainfall, holiday periods, etc.) period in the next year, not necessarily of the same length of time. Examples include comparing patterns for January (winter in Northern hemisphere temperate zones, no garden watering) or August (holiday periods in Northern hemisphere temperate zones, significantly fewer people in urban areas but possibly significant garden watering). The registered patterns are likely to be similar in shape but not exactly the same (Figure 2(a), showing flow rates φ_1 and φ_2 into a single supply area for two different periods as a function of time). The simple procedure which is presented here allows a quantitative comparison of these patterns, taking the following steps:

1. Sort both data sets from small to large magnitude (Figure 2(b)). Sorted measurement ordinal numbers are on the horizontal axis, flow rates are on the vertical axis. Note that the ordinal numbers have been scaled to have the same ordinal range [0,1] for both series in spite of their different lengths.
2. When the sets are not of equal length, resample the sorted curve of one of them so that the number of data points is equal for both (Figure 2(c)). If, for example, set 1 has 5 data points and set 2 has 10 data points, interpolation of the sorted curve of set 2 is done at $x = [0, 0.25, 0.5, 0.75, 1]$. Note that it is preferable to resample the longest of the two datasets in order to reduce the amount of data. In the example shown in Figure 2, interpolation of the dashed curve for the second period on the scaled ordinal points of the solid curve for the first period is performed.
3. Plot one data set against the other in a CFPD plot (Figure 2(d)).
4. Determine a linear best fit with slope a and intercept b (Figure 2(e)).

Note that when comparing individual day patterns to each other, stochastic effects may have a significant influence on the resulting CFPD plot, which may deviate from a linear pattern. Using a longer comparison period (e.g. a week or longer) will dampen these effects. The user is free to choose which period is put on the horizontal axis and which is put on the vertical in the CFPD plot, but this choice should be taken into account when interpreting the

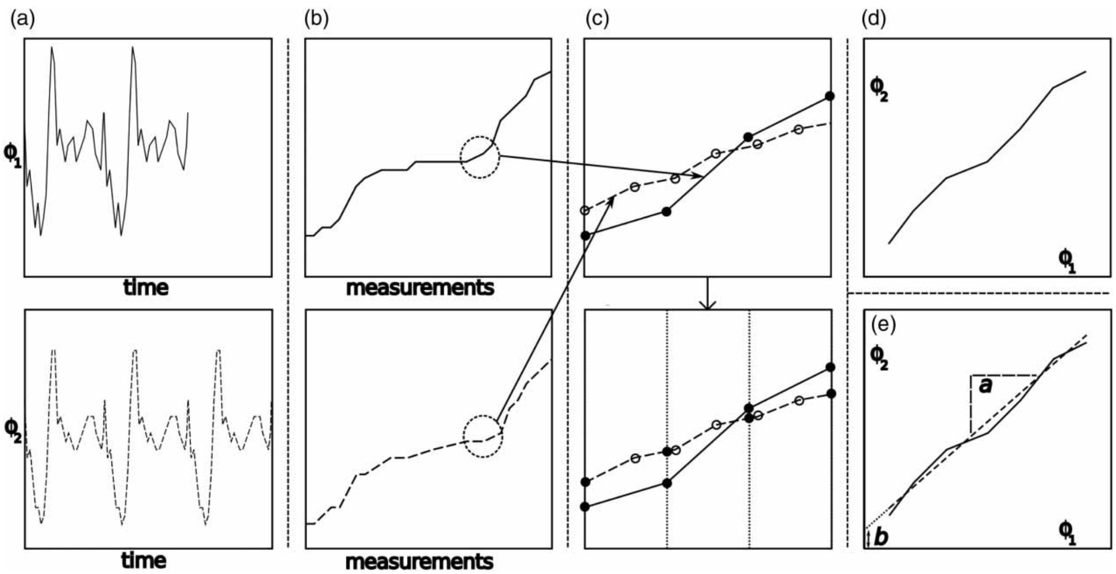

Figure 2 | Comparison procedure for flow patterns. (a) Flow patterns, (b) sorted flow patterns, (c) zoom-ins on small part of sorted flow patterns, (d) CFPD diagram, (e) CFDP diagram with linear fit. Note that these images merely illustrate the procedure but do not include real measured data.

results. When the horizontally plotted period precedes the vertically plotted period, $a > 1$ and/or $b > 0$ corresponds to an increase in flow rate. In general, this is preferable. However, in some cases it may be preferable to plot the more recent of the two patterns on the horizontal axis, in case the latter can be considered a reference pattern for some reason and the former is deviating from it. When two datasets of the same length are compared, step 2 can be skipped and the procedure can easily be performed in an ordinary spreadsheet program.

Properties

The characteristics of the curve which is thus produced depend in a simple way on the differences between the flow patterns on which it is based. If the shape of the flow patterns is identical, the resulting curve will be a straight line (Figure 3(a)). If the shapes are the same, but one is offset relative to the other by a constant value (e.g. a constant amount of leakage), the slope of the curve will be one but it will be offset (Figure 3(b)). This change is defined here as an inconsistent change, since it does not follow the existing flow rate distribution but is uniform. The corresponding offset b (unit is the same as the flow rate unit used in the input data, e.g. m^3/h) in the CFPD-plot is equal to the offset value of the pattern, so an inconsistent change of $+40\ m^3/h$ in the second flow pattern compared to the first results in a factor b of $+40\ m^3/h$. If the shape is the same but the pattern has been scaled by a certain value, this scale factor will be reflected in the slope a (dimensionless) of the curve (Figure 3(c)), which continues to cross the origin of the plot. This change is defined here as a consistent change, since it does follow the existing flow rate distribution. If, for example, the water demand increases by 10% during all parts of the day, the value of a will be 1.1. Note that consistent and inconsistent changes are purely numerical characteristics of the comparison of the two periods.

If the second pattern differs from the first only during a part of the day, a deviation from the ideal line is observed for a part of the curve (Figure 3(d)). The interpretation of consistent and inconsistent changes is summarized in Table 1.

It can be easily verified that when a measured flow pattern is compared to a uniformly perturbed (scaled and/or shifted) version of itself, the scaling and shifting factors are retrieved with high accuracy. In fact, only rounding errors in the numerical operations of the procedure might cause a slight deviation.

Comparison of periods of different length requires some caution. For example, if one would compare a 5 day

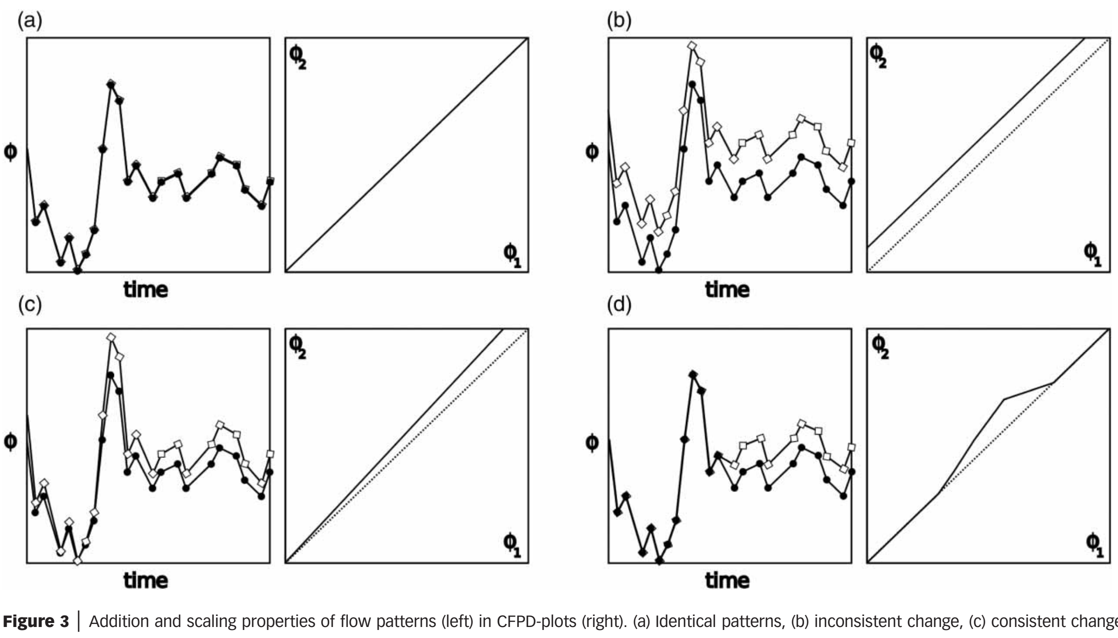

Figure 3 | Addition and scaling properties of flow patterns (left) in CFPD-plots (right). (a) Identical patterns, (b) inconsistent change, (c) consistent change, (d) time-windowed change causing a local deviation in the CFPD diagram.

Table 1 | Interpretation of consistent and inconsistent changes (factors *a* and *b*)

		No inconsistent change $b = 0$	**Inconsistent change** $b \neq 0$
No consistent change	$a = 1$	No change in demand	Increase in leakage or increase in demand not following established pattern
Consistent change	$a \neq 1$	Increase in demand following established pattern	Combination of increase in demand following established pattern and increase not following the established pattern and/or increase in leakage

workweek period to a 7 day full week period, the former pattern would not include the weekend days, which have a significantly different pattern from week days, whereas the latter would. These differences are sure to show up in the *a* and *b* factors of the CFPD analysis, likely drowning any other changes in *a* and/or *b*. Therefore, periods should be longer than the time scale of natural variability in the signal one is not interested in (such as weekly variations) and/or include proportionate numbers of days of different types.

Consistent and inconsistent changes in the water demand can be caused by several factors. Table 2 lists a range of demand change scenarios and the corresponding consistent and inconsistent demand pattern changes. These need to be kept in mind when interpreting a CFPD-plot. In general, significant inconsistent changes are likely attributable to either large volume costumers or leaks. Note that changes in the network configuration (e.g. opening and/or closing of valves) or flow meters may also induce consistent and/or inconsistent changes.

The power of this method stems from the fact that the time factor is removed from the comparison. Any stochastic variation is only seen in terms of its amplitude and can be expected to occur at some time in each of the compared sets of measurements. Also, the somewhat different characteristics of different weekdays are caught in the same way.

Table 2 | Different scenarios causing changes in the water demand in a supply area and the corresponding differentiation between consistent and inconsistent effects

Scenario	**Aspect**	**Change**	**Note**
Population size changes	New neighborhood is built and populated	Increase consistent with household part of existing demand pattern	Assuming behavior is comparable to that of existing population
	Vacation period	Decrease consistent with household part of existing demand pattern	Fewer people present, but with the same demand pattern
	Seasonally visiting tourists	Part consistent, part inconsistent	Tourists may have comparable habits of water use, but probably not identical
Warm season (temperate climate)	Showering/ bathroom	Increase	Combined effect is an increase which has a large consistent component and a small inconsistent component in household demand
	Laundry	Increase	
	Toilet use	Same/decrease?	
	Kitchen	Increase	
	Outside tap[a]	Inconsistent increase	Typical pattern of garden watering
	Evaporative coolers[a]	Inconsistent increase	Probably a typical pattern
Cold season (temperate and cold climates)	Prevention of pipe freezing[a]	Inconsistent increase	
Large volume customer sets up or leaves		Inconsistent change	Costumer specific demand pattern
Change in network configuration		Consistent and inconsistent changes expected	
New leak		Inconsistent change	
Repair of a leak		Inconsistent change	Sign opposite to new leak

[a]Billings & Jones (2008).

Block analysis

The procedure described above allows the user to compare two preselected periods from a dataset to each other. However, when it is unknown at what time specific events occur, as is generally the case in practice, manual preselection of comparison periods is undesirable. Instead, a comparison of all periods of prescribed characteristics (to make the number of combinations less than infinite) would circumvent this issue. Figure 4 illustrates the procedure and results of such a block analysis. A CFPD analysis is made (Figure 4(a)) of all possible combinations of time blocks of a preselected length of the comparison frame ℓ (unit of time, in this example 1 week) within the complete dataset of length L (unit of time, in this example 4 weeks). For a preselected comparison frame length ℓ, the dataset is split into L/ℓ frames (rounded up, in this case four frames), the first starting at the starting time of the dataset t_0 and each consecutive frame starting at $t_{i+1} = t_i + \ell$. Two matrices A (Figure 4(b)) and B (Figure 4(c)) are made, in which row i and column j represent frames i and j, respectively, and entries A_{ij} and B_{ij} are the factors a and b, respectively, resulting from a CFPD comparison of frame i with frame j. Since both matrices are antisymmetric about the main diagonal, only the upper triangle is shown. As the main diagonals show the results of a comparison of each frame with itself, these consist of ones (A) and zeros (B) exclusively. The entries in the upper triangle are gray-toned as a function of their deviation from 1 (A) and 0 (B), respectively, with small deviation having a light tone close to white and larger deviations having a darker tone. The sign of the deviations is indicated as well (+/–/=). Note that it is important to perform a comparison of all possible combinations, since beforehand it is usually not clear which time block is suitable as a reference time block.

Changes in a or b which remain in the signal longer than the frame length will show up in the block analysis as blocks of similar gray tone and sign, allowing direct pinpointing of events which cause these changes. This will be illustrated below.

Data quality

The effectiveness of the CFPD method will depend on the quality of the input data. If flows into and out of tanks are derived from level meters rather than flow meters, significant errors may be introduced when transported volumes are relatively small and the surface area of the reservoir is large, resulting in very small level changes. Also, when the registered flow data are instantaneous values rather than integrated values, errors may be introduced when large variations occur between sensor readings, e.g. related to the cycling on or off of pumps. Therefore, storing and using integrated rather than instantaneous flow data is preferable.

Systematic errors in flow meters will affect coefficient b. Random errors in flow meter readings or related to the

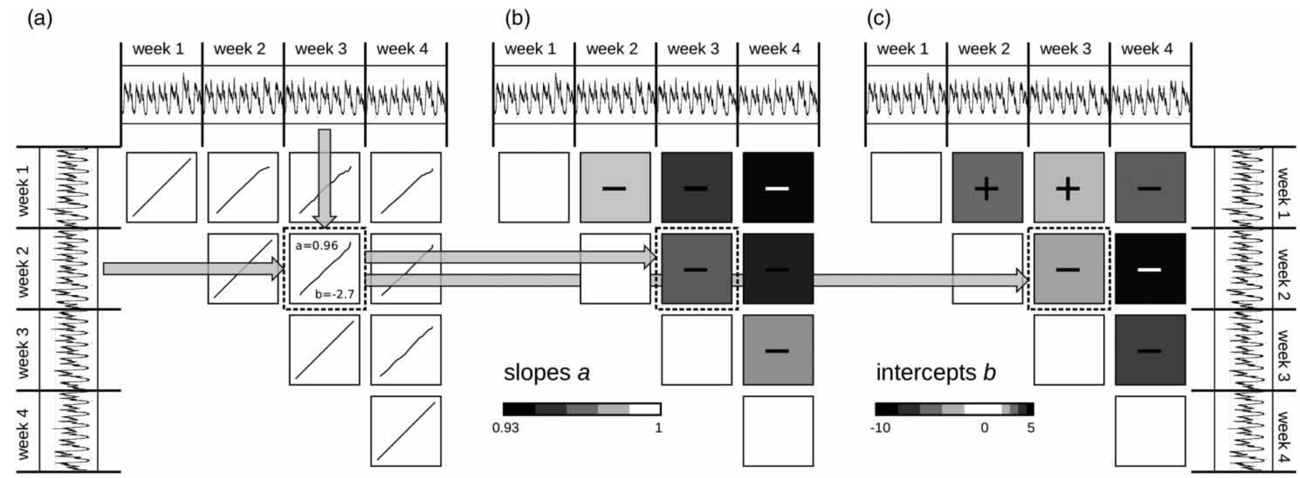

Figure 4 | Procedure of the CFPD block analysis. (a) CFPD analysis of combinations of individual periods within a long time series; (b) gray tone representation of a coefficients for all combinations; (c) gray tone representation of b coefficients for all combinations. Arrows indicate the comparison of week 3 to week 2 (reference).

issues described above are expected to average out to some degree in the CFPD analysis, since the mechanisms introducing the errors behave similarly under similar conditions (e.g. larger errors in flow measurements at low flow rates, larger errors in tank flow determinations for small level changes) and are sampled with each reading. Below, the sensitivity of the method to Gaussian noise is studied.

Testing data

Several Dutch and Flemish water companies have supplied recent flow rate measurement data for selected supply areas with measurement intervals of 1–15 min. These supply areas are individually discussed in the next section.

RESULTS AND DISCUSSION

The method presented above is applied to a number of real life cases. In addition to this, a number of synthetic tests are performed to determine the sensitivity of the method.

Case 1: Fixing of a leak

From top-down method results, Vitens water supply company (The Netherlands) suspected that one or more significant leaks were present in one of their supply areas (about 13,000 connections). Eventually a leak of about 40–50 m^3/h was found and fixed. Figure 5 shows the flow patterns for a period of 3 weeks before and about 6 weeks after the fixing of the leak. The shape of the pattern is quite similar but a significant shift has been caused by the fixing of the leak. This can be readily seen in the CFPD-plot of Figure 6. The good fit of the curve to a straight line demonstrates the similarity of the flow pattern between the two periods. However, a shift is clearly visible. Quantitatively, the slope of the curve is 1.0008, showing almost no consistent change in the pattern. The intercept is located at –56.5 m^3/h. The estimate of the size of the leak made here is somewhat higher than the estimate of the water company. However, their estimate was based on observations at a single pumping station, whereas the present estimate is based on the complete patterns of several weeks and is therefore

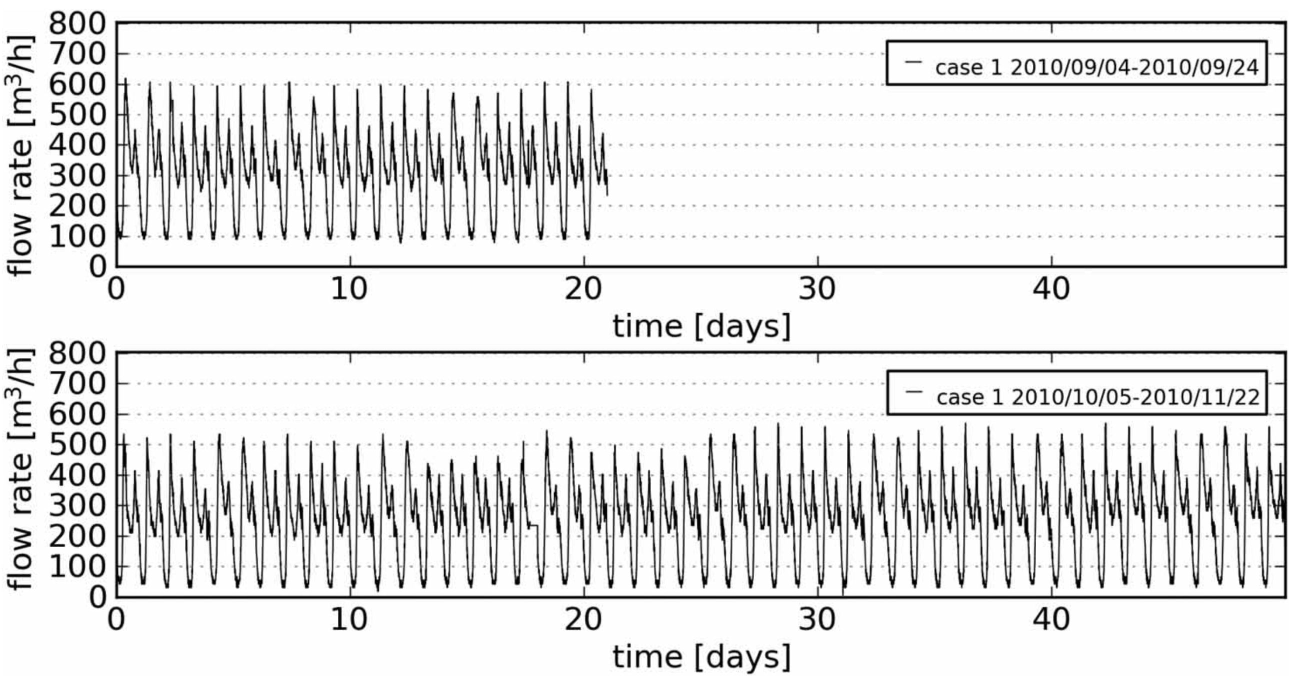

Figure 5 | Flow patterns for case 1 in the weeks before (top) and after (bottom) the fixing of a big leak.

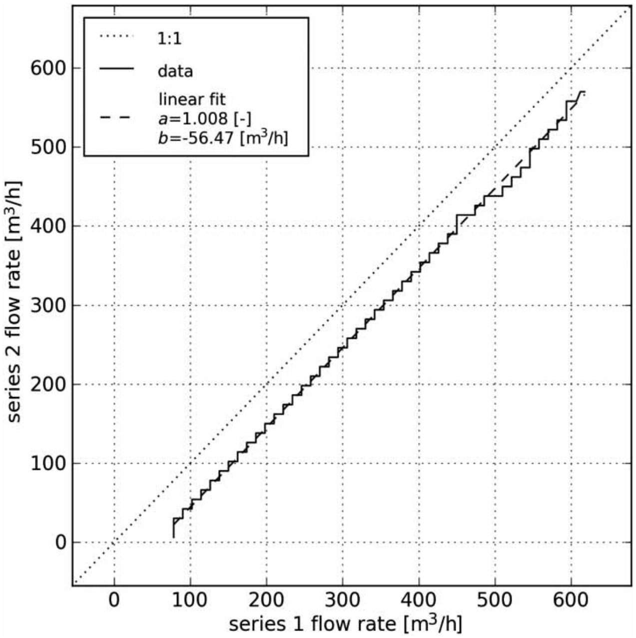

Figure 6 | CFPD-plot combining flow patterns for case 1.

more representative. The intercept value is also quite close to the difference of the mean flow rates over the two periods, which is –54.0 m^3/h. Note that the stepped shape of the curve in Figure 6 is related to the limited flow rate resolution of the flow patterns supplied. However, this poses no problem for applying the method.

Case 2: Spring versus fall

One of the supply areas of Evides water company (The Netherlands) has a number of small towns, some agriculture and some camping and holiday houses (in total less than 10,000 connections). Measurements of the flow pattern for the period from the end of April through to the end of May 2010 and for October 2010 were supplied to us. These are shown in Figure 7. A larger variability in the peak flow rates can be seen in the flow pattern for spring compared to fall. The CFPD-plot which was made with these data (Figure 8) shows again a good linear fit of the curve, with a slope $a = 0.874$ and an intercept $b = 12.6\ m^3/h$ (with the minimum night flow increasing from 23.3 to 32.0 m^3/h). This can be interpreted as a reduction in the actual water demand in fall (cool) compared to spring (warmer, more tourists) by about $1.0 - a = 13\%$ combined with an inconsistent increase $b = 12.6\ m^3/h$, possibly due to increased leakage. Note that this

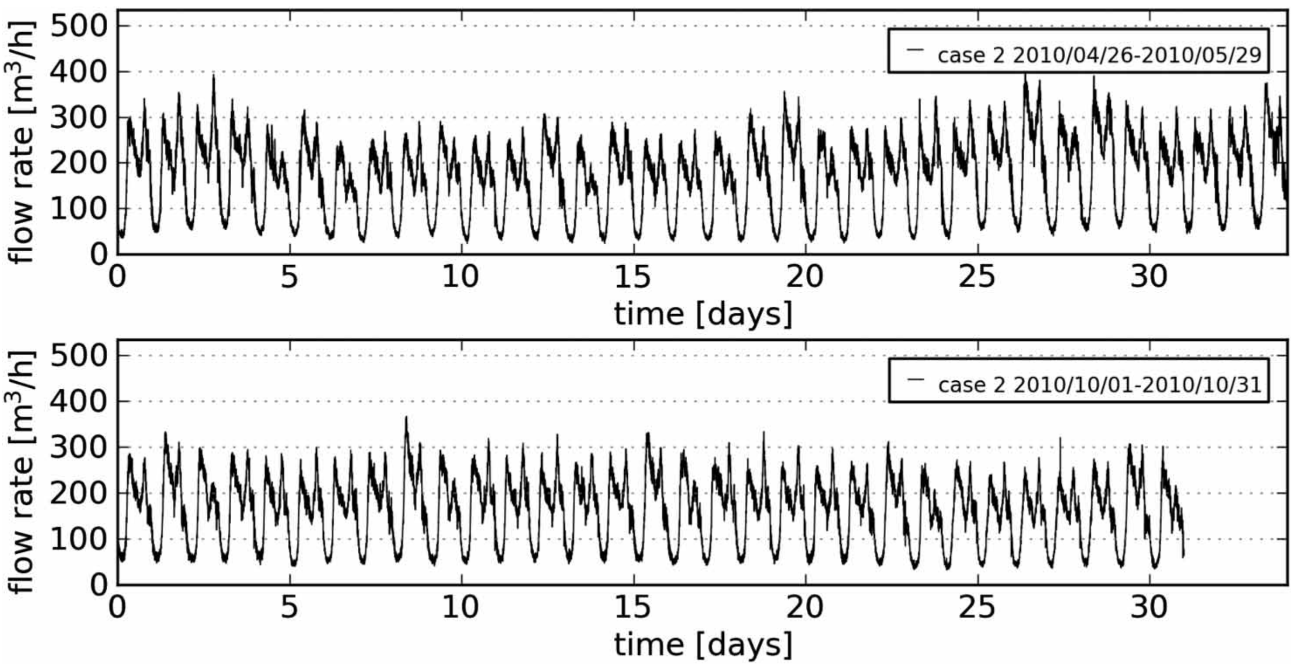

Figure 7 | Flow patterns for case 2 in spring (top) and fall (bottom).

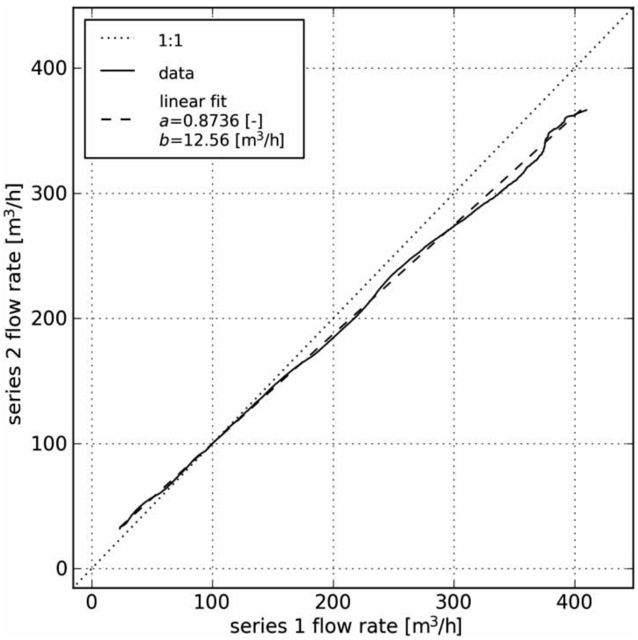

Figure 8 | CFPD-plot combining flow patterns for case 2.

inconsistent increase is in line with but somewhat larger than the increase in the minimum observed flow rate of 8.7 m^3/h.

Case 3: A small supply area

The third case presented is a special one in that it consists of only ~70 connections, including houses, agricultural customers and an abbey which is a tourist attraction. Figure 9 shows flow patterns for the period of August–September 2009 and November–December 2010, respectively. The latter pattern shows a peak demand which is generally around 0.6 m^3/h and many nights in which the demand drops to 0. This means that total leakage in this supply area is less than the smallest measurable flow rate. The pattern in summer, however, shows a significantly higher demand, both during the day (peak demand generally around 2 m^3/h) and during the night (consistently around 0.8 m^3/h). Figure 10 shows a CFPD-plot combining these two time series. A very convincing linear fit can be made, with a few small deviations at the peak demand end. The fit parameters show a consistent doubling of the demand ($a = 1.98$) and in addition to this an inconsistent increase of 0.31 m^3/h. The consistent change can be attributed to the increased presence of tourists in summer. In this case, we

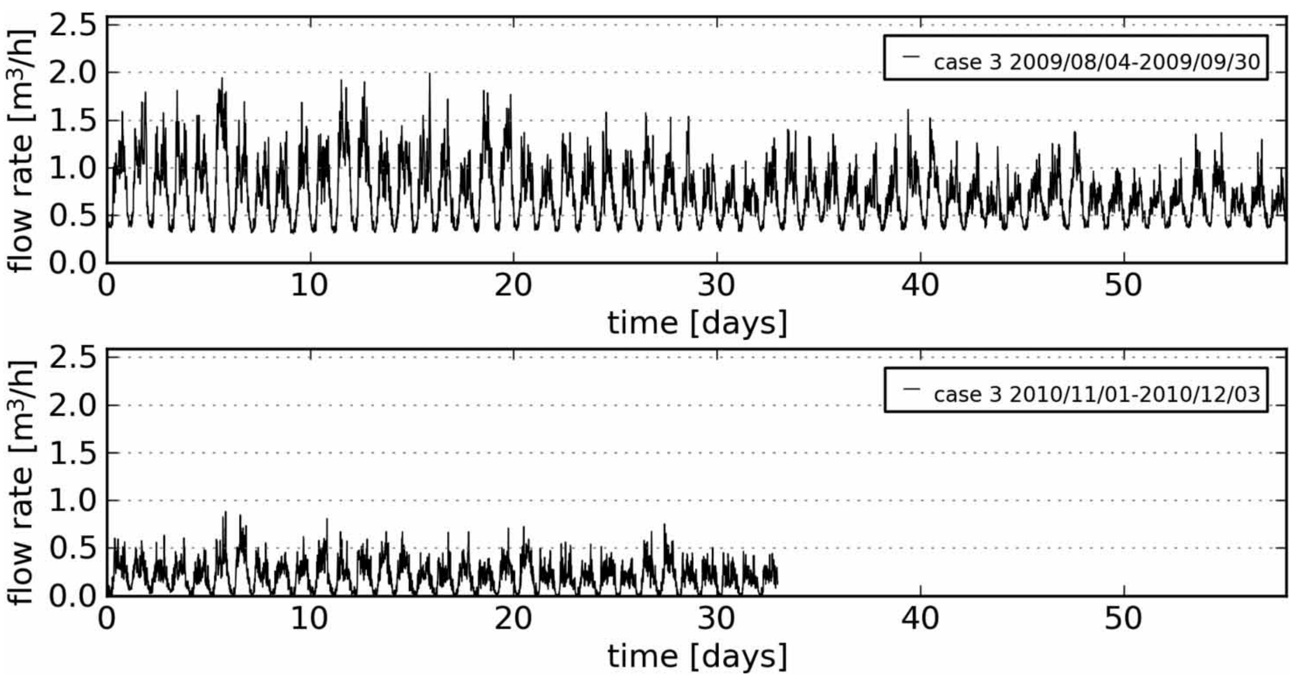

Figure 9 | Flow patterns for case 3.

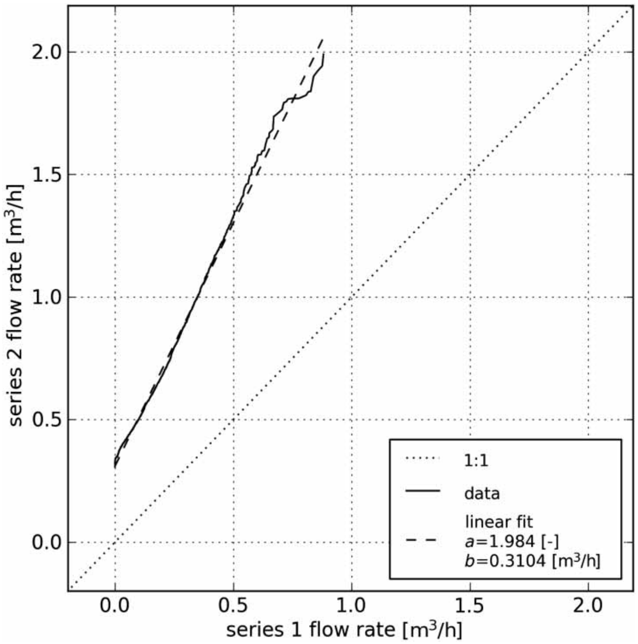

Figure 10 | CFPD-plot combining two flow patterns for case 3.

know there is no leakage which could explain the inconsistent increase. The water company suspects (on different grounds) some illegitimate water use during spring and summer in this area, to which the inconsistent increase may be attributed. This case shows that the method can also be successfully applied to very small supply areas.

Case 4 (theoretical): Varying leakage rate due to pressure variations

Real life leaks are thought to have a pressure dependent leakage rate. This is generally modeled (for an individual leak) using an equation of the following form:

$$Q_l(t) = K \cdot P(t)^{\alpha}$$

in which $Q_l\ (t)$ is the leakage rate as a function of time t, K a constant which is defined for an individual leak, $P(t)$ the local pressure as a function of time and α an exponent, the value of which is usually assumed to be close to 0.5 but may be higher in some cases (Greyvenstein & van Zyl 2007; Wu 2011). In order to test its effect on CFPD-curves, a pressure dependent leakage is added to a measured data set. This measured set is a 1 month pattern with a 1 min measuring interval and minimum and maximum observed flow rates of 32 and 366 m^3/h, respectively (fall period of case 2). Two different pressure scenarios are applied (Figure 11). The first assumes that during the morning and evening water demand peaks, pressure drops to 50 and 60%, respectively, of the reference level. In the second scenario, it is assumed that the water company actively manages the pressure and increases its value by a factor of 2 during the peak hours.

Figure 12 shows the original and perturbed flow patterns and Figure 13 shows the corresponding CFPD-plots

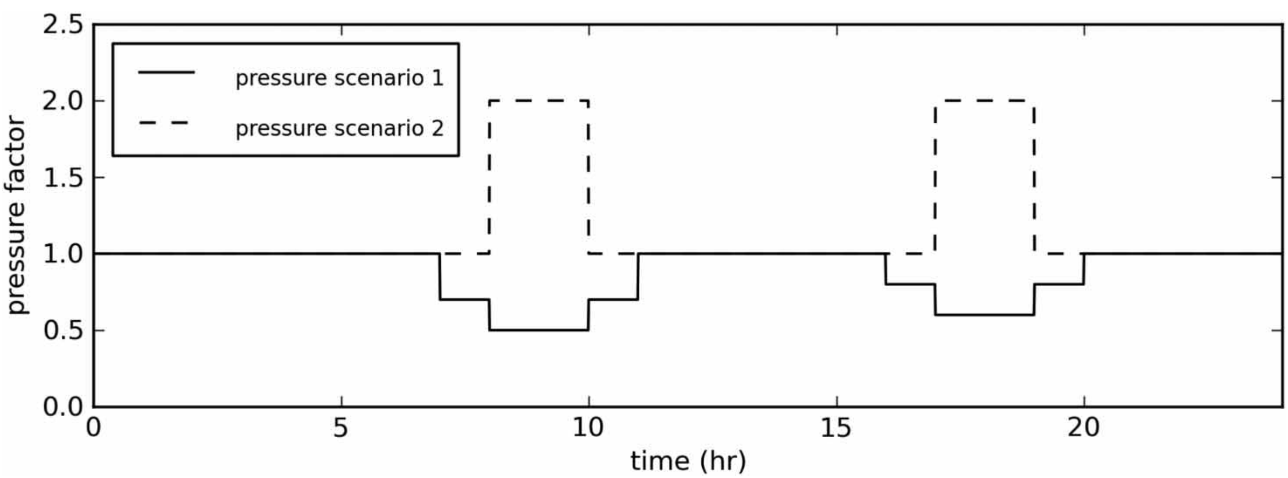

Figure 11 | Synthetic pressure curves for the determination of the effect of pressure dependent leakage on CFPD-plots.

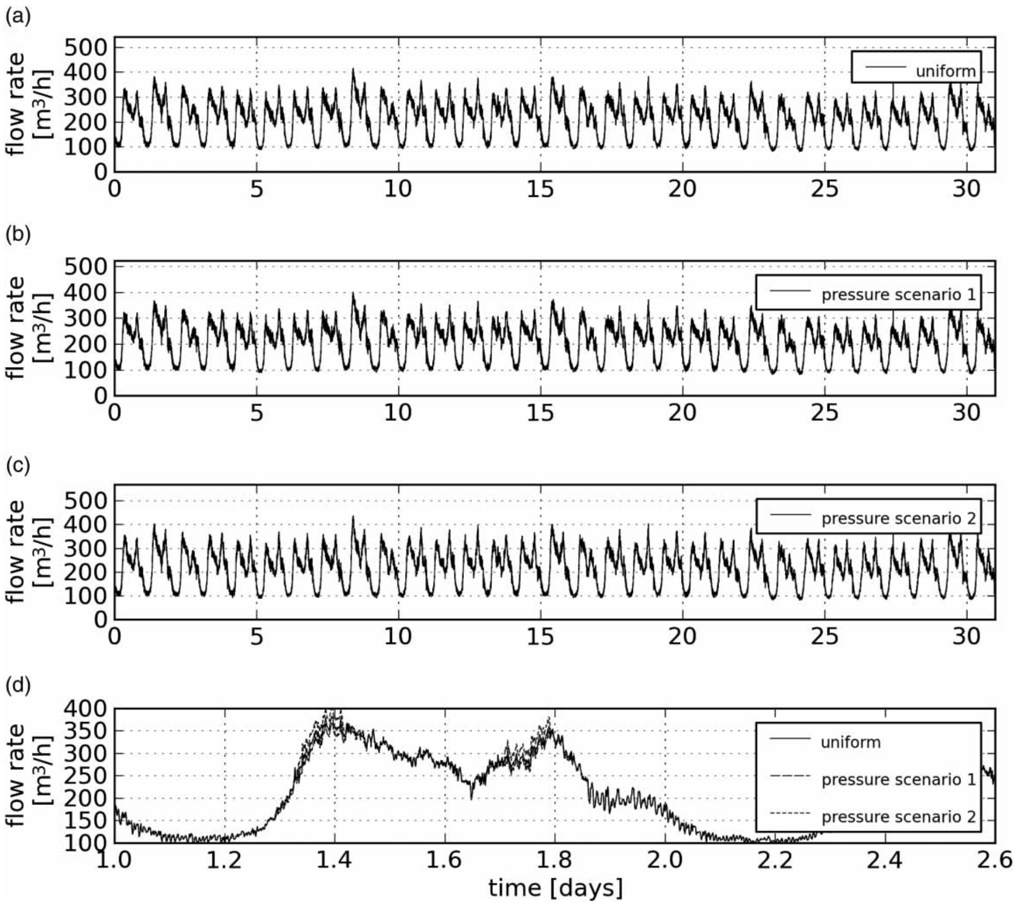

Figure 12 | Flow patterns for uniformly perturbed case (a), pressure scenario 1 (b), and pressure scenario 2 (c), with a zoom-in on a single day of all three (d).

comparing the original pattern to the same pattern with a uniform shift (a), to the pattern perturbed according to the first scenario (b), and to the pattern perturbed according to the second scenario (c). For easier comparison, these curves are shown together in Figure 13(d). In each case, the magnitude of the perturbation is 50 m^3/h (for a pressure factor of 1 in the two pressure dependent cases). A pressure exponent of 0.5 is applied, representative of round holes and circumferential cracks (Greyvenstein & van Zyl 2007).

It can be clearly seen from Figure 13 that the significant variations in pressure prescribed in the two scenarios have a quite small effect on the CFPD-curves. The corresponding fit parameters are listed in Table 3 and they support this conclusion. The nominal added leakage rate of 50 m^3/h is recovered to a large degree. It should be noted that it is overestimated in the first scenario, which has a pressure reduction during peak hours, and underestimated in the second, which has a pressure increase during peak hours.

This test is in fact a simplification of the actual situation in a distribution network. Variations in the demand during the day and variations in the input pressure at the pumping station cause the pressure variations at individual leaks to depend on their location in the network. For example, in the case where pressure is increased during peak hours, a leak close to the pumping station may have a higher leakage rate during peak hours due to the higher pressure at the pumping station, whereas a leak far from the pumping station may have a lower leakage rate at the same time due to the lower local pressure related with the high demand at that time.

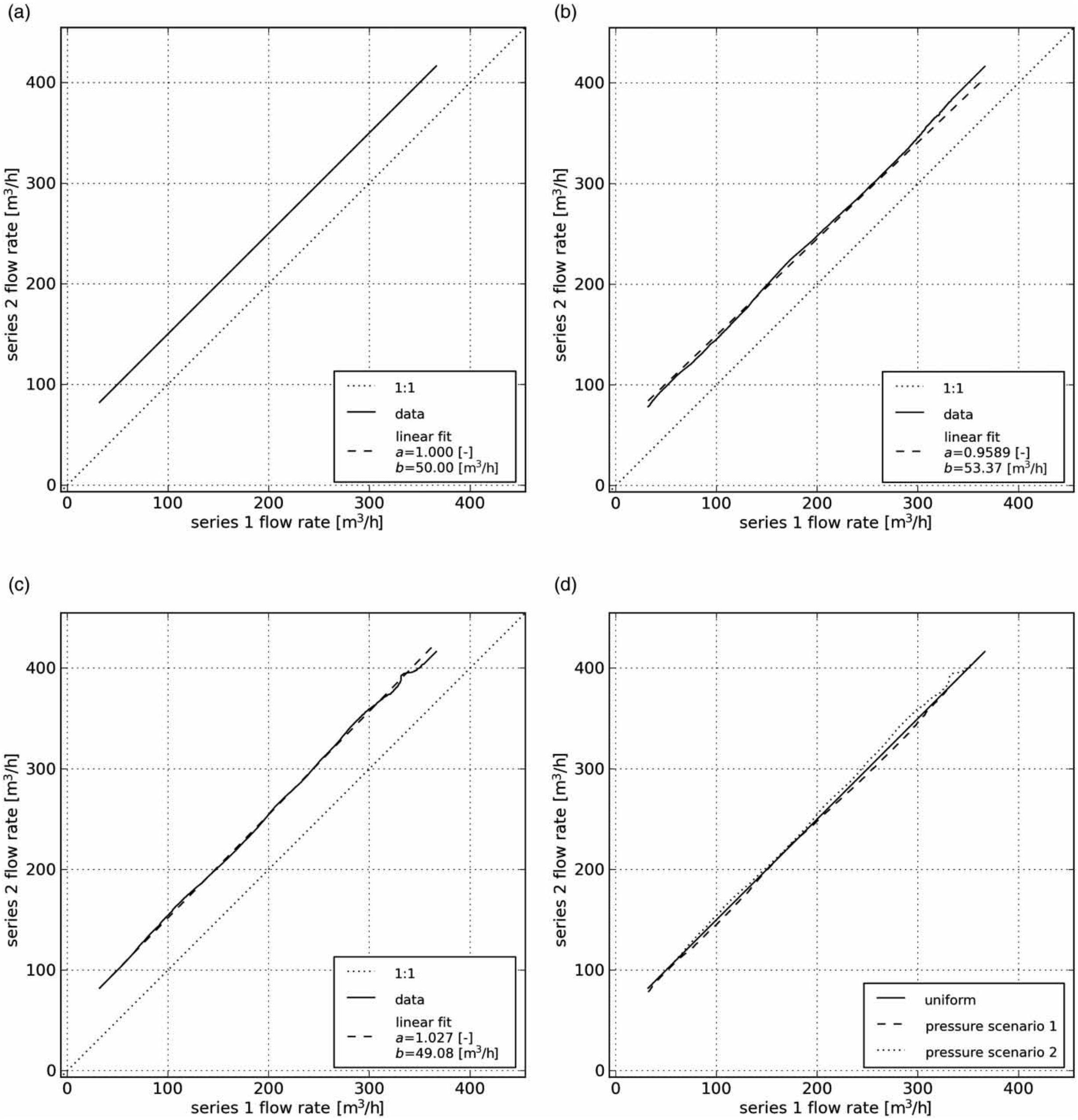

Figure 13 | CFPD-plot for a uniform perturbation of $+50$ m^3/h (a), a pressure dependent leakage case with lower pressures at peak hours (b), and a pressure dependent leakage case with increased pressures at peak hours (c), with all combined in a single graph (d).

Case 5 (theoretical): Measurement interval

The measurement interval of the flow data should be sufficient to characterize the shape of the flow pattern. The question is, however, what sufficient means, or which minimum measurement interval of the flow data is required for a successful application of the CFPD methodology. In order to establish this, the measured data of case 2 have been aggregated in increasingly large time blocks in order to simulate the data registration at larger time intervals. Table 3 lists the fit parameters of the aggregated data at intervals of 15 min, 1 h, 4 h, 8 h and 24 h. It can be seen that at 1 h

Table 3 | Scale factors and intercept values from CFPD-plots for different real life cases, pressure dependent leakage scenarios and measurement intervals

Case	Scenario	Scale factor *a* (–)	Intercept *b* (m^3/hour)	Figure
1	Data	1.00	–56.5	Figure 6
2	Data	0.874	12.6	Figure 8
3	Data	1.98	0.310	Figure 10
4	Uniform	1.00	50.0	Figure 13(a)
4	Low pressure during peak hours	0.988	48.7	Figure 13(b)
4	High pressure during peak hours	1.03	49.1	Figure 13(c)
5	1 min interval	0.874	12.6	
5	15 min interval	0.873	12.8	
5	1 h interval	0.870	13.2	
5	4 h interval	0.863	14.3	
5	8 h interval	0.844	17.5	
5	24 h interval	0.716	38.1	

intervals, the resulting fit is still quite close to the fit of the original data, with deviations only in the third decimals. Longer measurement intervals result in an underestimation of the scale factor *a* and an overestimation of the intercept factor *b*.

Case 6 (theoretical): Sensitivity to noise and special events

Varying types of noise, including stochastic noise, measurement errors, etc., can conceivably affect the functioning of the presented method. In order to determine the sensitivity of the method to noise, the fall data of case 2 (minimum and maximum observed flow rates of 32 and 366 m^3/h, respectively) have been perturbed (scaled and/or shifted) with prescribed values of *a* (1.0 and 1.1) and *b* (0, 5, and 50 m^3/h). Varying amounts of Gaussian noise (standard deviation being 0, 5 and 10% of individual values, respectively) have been added as well. The resulting data sets have been compared to the original data in CFPD-plots. The corresponding CFPD fit parameters are listed in Table 4.

When no noise is added, recovery of the parameters *a* and *b* is perfect. When 2% noise is added, both recovered parameters are still quite close to their prescribed values. For 5% noise, slope values (*a*) still match quite well, but the relative deviation for small prescribed values of *b* starts to become large. For larger amounts of noise, *a* values are consistently too high and *b* values consistently too low, the former apparently seeping into the latter.

Special events, such as flushing campaigns, significantly affect flow rates and may therefore also affect the results of the CFPD method. An additional flow rate of 50 m^3/h has been added to the fall data set of case 2 on a single day between 10:00 and 14:00 to represent a series of flushings in this supply area, again including varying amounts of noise. As can be seen in Table 4 (test 7), the effect on the fit parameters is minimal and rapidly drowned by added noise.

Note that in all cases, the coefficient of determination R^2 of the fit is above 0.99.

Table 4 | CFPD fit parameters for a range of test scenarios with varying amounts of Gaussian noise added (percentage indicates percentage of sample value, used as standard deviation for Gaussian noise)

	No noise		2% noise		5% noise		10% noise	
Test	*a* [–]	*b* [m^3/h]	*a* [–]	*b* [m^3/h]	*a* [–]	*b* [m^3/h]	*a* [–]	*b* [m^3/h]
1: Pattern	1.000	0.000	1.001	−0.2084	1.008	−1.223	1.030	−4.696
2: Leak 5 m^3/h	1.000	5.000	1.001	4.805	1.008	3.694	1.032	−0.09075
3: Leakage 50 m^3/h	1.000	50.00	1.003	49.61	1.012	48.03	1.045	42.71
4: Increase 10%	1.100	0.000	1.101	−0.2252	1.109	−1.397	1.135	−5.504
5: lkg 5 m^3/h + inc 10%	1.100	5.000	1.101	4.776	1.109	3.545	1.133	−0.4520
6: lkg 50 m^3/h + inc 10%	1.100	50.00	1.102	49.63	1.112	48.08	1.148	42.22
7: Single day increase 50 m^3/hr	1.004	−0.3673	1.005	−0.5648	1.011	−1.578	1.034	−5.203

Processing and stability of long time series

A block analysis has been performed for a selected period of 8 weeks of the case 1 data (see Figure 14; data have been aggregated to one measurement per 15 min to reduce computation times). Figure 15 shows cross tables in which (as described in the Methods section) the magnitude of parameters a (Figures 15(a) and (c)) and b (Figures 15(b) and (d)) in CFPD analyses of combinations of time blocks within this 8 week period are indicated by gray tones. Signs indicate whether the values are above, below or equal to the neutral values of $a = 1$ and $b = 0$, respectively. The rows and columns of the tables represent the individual time blocks, with rows representing reference time blocks and columns representing the time blocks which are compared to individual reference time blocks. Figure 15 shows these cross tables for selected time block durations of 1 day (Figures 15(a), (b)) and 7 days (Figures 15(c), (d)).

When applying a 1-day comparison period, a strikingly regular pattern emerges in the slope table (Figure 15(a)). Weekdays are relatively similar to each other, showing light tones, and weekend days as well, but they are quite different from each other, as indicated by dark tones. When looking at the intercept table (Figure 15(b)), the same pattern is visible, but it is drowned by a dark block in the upper right quadrant of the table, which represents the significant decrease in demand by the fixing of the leak described above. As can be observed from the actual values of a and b shown in the tables, there is significant variability, both stochastic (which may be important when using such short comparison windows) and related to the deviating flow patterns associated with the fixing of the leak. By applying a longer comparison window (7 days in Figures 15(c) and (d)), the magnitude of the (stochastic) variability strongly decreases (note the changing extrema of the scale bars) and the general picture stabilizes to values which are close to those found for case 1.

It is easy to determine the time of occurrence of events in a block analysis diagram. Figure 15(b) shows that the dark block, corresponding to the decreased leakage due to the repair, starts on the second day of week 5. The coarser time scale image of Figure 15(d) similarly shows that the event took place close to the boundary between week 4 and week 5. Another interesting observation on Figure 15(b) is that the significant deviation of b can already be seen in an analysis on a single day basis. One can speculate that in a less clean, more noisy dataset, a single day might not be sufficient.

Interpretation of consistent and inconsistent changes

Interpretation of the results of the pattern distribution comparison can only be done with some knowledge of the operation of the supply area. More specifically, for any change, consistent or inconsistent, an explanation must first be found in terms of things that are known to have happened in the supply area (see Table 2). These range from a holiday period resulting in a temporary shift and decrease of the morning demand peak (see the slight deviation in the upper right part of Figure 6, which is related to the autumn school holidays) to changes in the operation of a large industrial client. Also, it is important to know the condition and status of all valves relevant for isolating a supply area. For example, when a boundary valve is open which should not be and water flows from the supply area under investigation into a neighboring supply area, flow

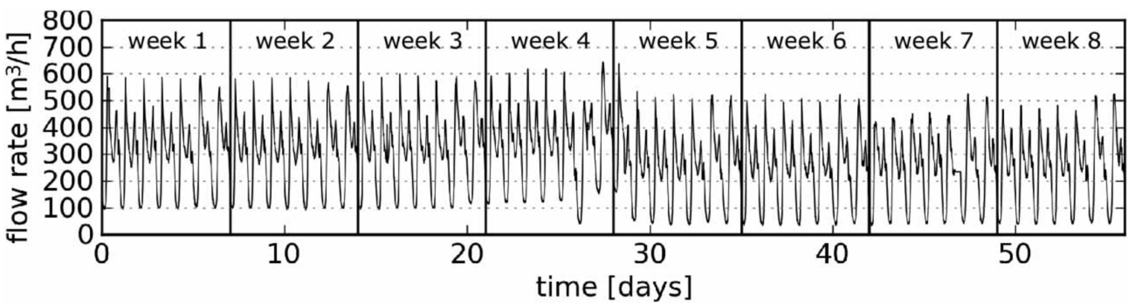

Figure 14 | Eight week set of flow data for case 1 on which Figure 15 is based.

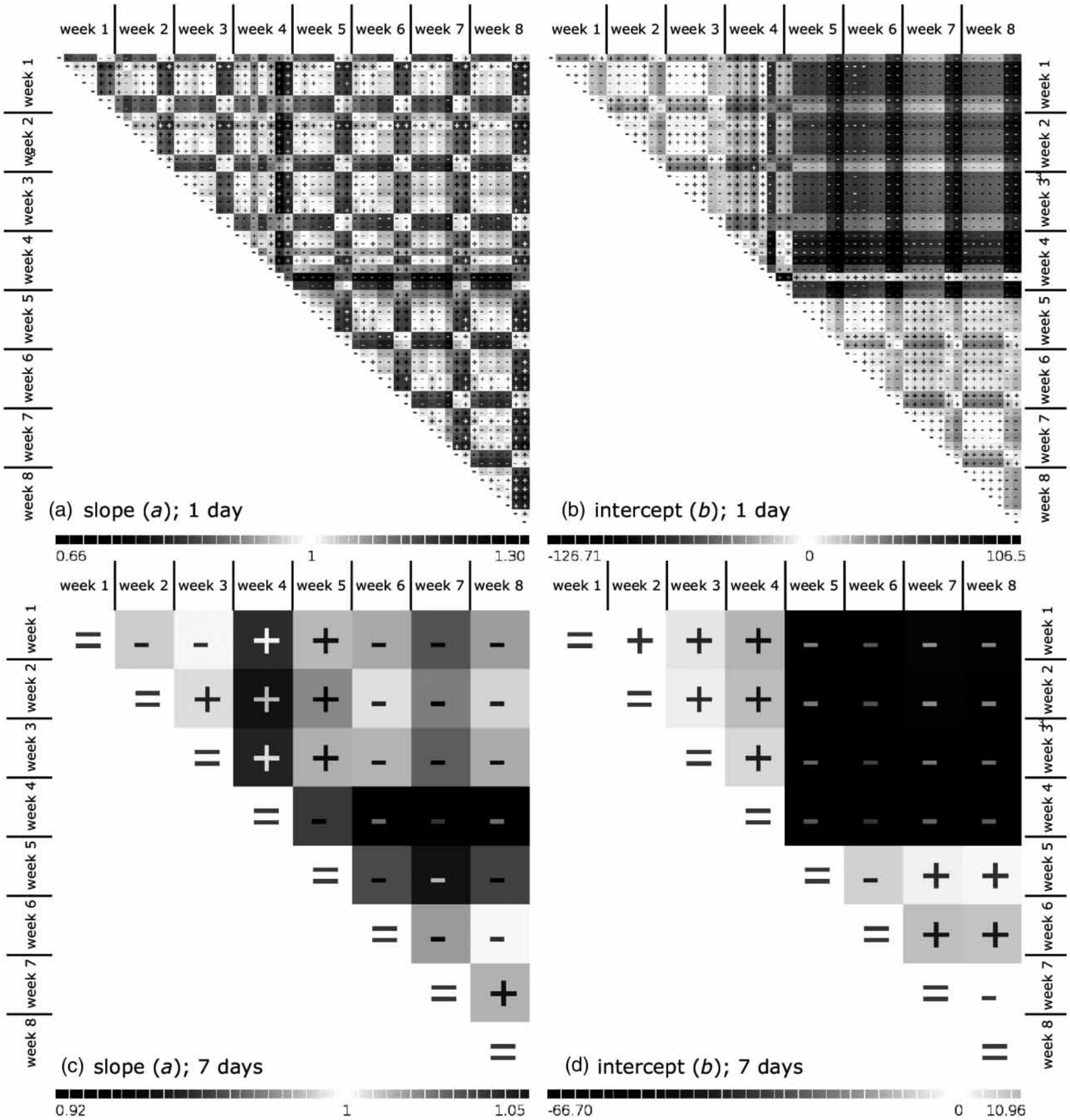

Figure 15 | Cross tables comparing periods of 1 (a, b) and 7 (c, d) days, respectively, of the 8 week set of flow data for case 1. Gray scale values indicate the magnitude of the slope a (a, c) and intercept b (b, d) for combinations of periods, signs in the cross tables indicate whether the values are positive, negative or neutral. Note that each subfigure has a different scale bar.

measurements may represent a much larger supply area than thought. If the demand characteristics of the unintentionally coupled supply area are similar to those of the area under investigation, the open valve will result in a consistent demand increase.

Consistent and inconsistent changes which cannot be explained from the knowledge of the supply area and its operation can possibly be ascribed to new leakages and require further investigation, e.g. in the field using noise correlators.

Suggestions for further research and development

A more detailed study on the sensitivity of the presented method to high exponent pressure dependencies of specific types of leaks, as well as the sensitivity of the method to specific sources of errors in flow data (such as the level meter issue mentioned above) would widen the applicability and strengthen the basis of the CFPD method. Automated coherent feature detection in CFPD block analyses would simplify the analysis of long historic time series of flow data. When performing the CFPD analysis in a moving time frame instead of a fixed time frame, i.e. comparing data for the past 24 h to the same 24 h a week before or a reference pattern of 24 h, on a minute to minute basis, application of the method on real time leakage monitoring may be possible. Further research and development on these issues are recommended and being pursued.

Suggested implementation at water companies

In its current form, the CFPD method and CFPD block analysis can be applied by water companies to study their archived data for changes in demand and possible leakage. This is expected increase their knowledge on customer behavior and the evolution of the condition of their networks. This gives water companies an (additional) tool for prioritizing field leakage detection campaigns. As shown in the processing and stability of long time series paragraph, significant changes in the *b* factor can show up in day to day comparisons, so a daily analysis (near real time) may be implemented as an additional leakage monitoring tool. With the developments suggested in the previous section, it is expected that the possibilities of the method as a monitoring tool for both demand and leakage may be enhanced.

CONCLUSIONS

The comparison of flow pattern distributions (CFPD) method was introduced as a new tool for supply pattern analysis with possible applications in leakage detection and demand management. Because it is simple to implement and apply but nevertheless powerful in distinguishing between consistent and inconsistent changes in water demand, the method provides water companies with a way to review historic data and monitor current data with a small time lag (up to some days) for possible changes in customer demand, increasing leakages and illegal connections. Additional advantages are that the method is independent of any model assumptions and that it scales very well (70 to >10,000 connections, no upper limit in view). Automated processing of long time series in a CFPD block analysis allows easy pinpointing and identification of events and gives insight into the stochastic variability of the flow patterns of their supply area, allowing an appropriate choice of the comparison time window (i.e. long enough) for the single period CFPD method. More generally, the method could provide additional information about both customer behavior and the evolving condition of the network from data which is usually readily available at water companies.

ACKNOWLEDGEMENTS

The author gratefully acknowledges data supplied by Henk de Kater of Evides water company, Karel Vangeel of Pidpa water company, and Johan Duifhuizen of Vitens water supply company, as well as constructive and stimulating reviews by Tom Walski and two anonymous reviewers that have helped to significantly improve the paper.

REFERENCES

Beuken, R. H. S., Lavooij, C. S. W., Bosch, A. & Schaap, P. G. 2006 Low leakage in the Netherlands confirmed. Water Distribution Systems Analysis symposium. Cincinnati.

Billings, R. B. & Jones, C. V. 2008 *Forecasting Urban Water Demand*. American Water Works Association, Denver, Colorado.

Buchberger, S. G. & Nadimpalli, G. 2003 Leak estimation in water distribution systems by statistical analysis of flow readings. *J. Water Resour. Plan. Manage.* **130** (4), 321–329.

Colombo, A. F., Lee, P. & Karney, B. W. 2009 A selective literature review of transient-based leak detection methods. *J. Hydro-Environ. Res.* **2**, 212–227.

Farley, M. & Trow, S. 2003 *Losses in Water Distribution Networks*. IWA Publishing, London.

Greyvenstein, B. & van Zyl, J. E. 2007 An experimental investigation into the pressure – Leakage relationship of some failed water pipes. *J. Water Supply: Res. Technol. AQUA* **56** (2), 117–124.

Lambert, A. O. 2002 Water losses management and techniques. *Water Supply* **2** (4), 1–20.

Liggett, J. A. & Chen, L.-C. 1994 Inverse transient analysis in pipe networks. *J. Hydraul. Eng.* **120** (8), 934–955.

Menicone, S., Brunone, B., Ferrante, M. & Massari, C. 2011 Potential of transient tests to diagnose real supply pipe systems: what can be done with a single extemporary test. *J. Water Resour. Plan. Manage.* **137** (2), 238–241.

Poulakis, Z., Valougeorgis, D. & Papadimitriou, C. 2003 Leakage detection in water pipe networks using a Bayesian probabilistic framework. *Prob. Eng. Mech.* **18**, 315–327.

Puust, R., Kapelan, Z., Savic, D. & Koppel, T. 2006 Probabilistic Leak Detection in Pipe Networks using the SCEM-UE Algorithm. Water Distribution Systems Analysis Symposium, Cincinnati.

Puust, R., Kapelan, Z., Savic, D. A. & Koppel, T. 2010 A review of methods for leakage management in pipe networks. *Urban Water J.* **7** (1), 25–45.

Saldarriaga, J. G., Araque Fuentes, D. A. & Castañeda Galvis, L. F. 2006 Implementation of the hydraulic transient and steady oscillatory flow with genetic algorithms for leakage detection in real water distribution networks. Water Distribution Systems Analysis Symposium, Cincinnati.

Savic, D., Lambert, A. & Kapelan, Z. 2005 Water losses management and leakage detection techniques for water distribution systems. *Water Sewer. J.* **2**, 25–27.

Vítkovský, J. P., Lambert, M. F., Simpson, A. R. & Liggett, J. A. 2007 Experimental observation and analysis of inverse transients for pipeline leak detection. *J. Water Resour. Plan. Manage.* **133** (6), 519–530.

Wu, Z. Y., Sage, P. & Turtle, D. 2010 Pressure-dependent leak detection model and its application to a district water system. *J. Water Resour. Plan. Manage.* **136** (1) 116–128.

Wu, Z. Y. (ed.) 2011 *Water Loss Reduction*. Bentley Institute Press, Exton, Pennsylvania.

First received 2 November 2011; accepted in revised form 29 March 2012. Available online 9 July 2012