Applications of discriminative flow pattern analysis using the CFPD method

Peter van Thienen, Ilse Pieterse-Quirijns, Jan Vreeburg, Karel Vangeel and Zoran Kapelan

ABSTRACT

Several applications of a new method for flow pattern analysis called the Comparison of Flow Pattern Distributions (CFPD) are presented. This method allows the user to compare flow patterns of a supply area and distinguish *consistent* from *inconsistent* changes in them. The so-called *consistent* changes are mainly caused by modifications in the characteristics of the population. The so-called *inconsistent* changes are generally related to new large volume customers, new types of water use and/or changes in leakage. Detailed knowledge of the supply area allows quantitative statements to be made about leakage. The method presented here is simple, not computationally intensive, independent of any model assumptions and easily implemented. The automated application of this method on long time series, called CFPD block analysis, is applied on data from three different areas. These applications illustrate, respectively, the identification and pinpointing (in time) of a small leak, the independent quantification of concurrent different types of changes with opposite signs in a supply pattern, and the difficulties of interpretation in cases in which climate has an overwhelming influence on demand patterns. In each case, discriminative quantification and visualization by the CFPD method result in features and trends in complicated time series becoming apparent at a glance.

Key words | demand patterns, flow pattern analysis, leakage, numerical methods

Peter van Thienen (corresponding author)
Ilse Pieterse-Quirijns
Jan Vreeburg
KWR Watercycle Research Institute,
Post Box 1072,
3430 BB Nieuwegein,
The Netherlands
E-mail: *peter.vanthienen@kwrwater.nl*

Karel Vangeel
Pidpa,
Desguinlei 246,
B-2018 Antwerpen,
Belgium

Zoran Kapelan
College of Engineering,
Mathematics and Physical Sciences,
University of Exeter,
Harrison Building,
North Park Road,
Exeter EX4 4QF,
United Kingdom

INTRODUCTION

The proper operation, maintenance and rehabilitation of existing drinking water distribution systems and the design of new networks require knowledge and understanding of flow patterns into drinking water supply areas. These contain information about both demand and leakage, but this information is often difficult to retrieve and interpret.

Leakage continues to be a problem for water companies around the world, with water losses ranging from 3% to more than 50% of distribution input (Lambert 2002; Beuken *et al*. 2006). The water lost represents a financial loss, but is also undesirable from the environmental, social and sustainability points of view. The two principal methods to determine the amount of Non-Revenue Water (NRW) in a supply area are the top-down and the bottom-up methods (Farley & Trow 2003; Wu 2011), both with a considerable degree of uncertainty. Several new, alternative methods have been developed which make use of optimization-based inverse type models, including 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). These model optimization-based methods generally require a hydraulic model of, and/or information about the system in, the supply area and can be computationally demanding (Vítkovský *et al*. 2007; Colombo *et al*. 2009). Applications of these methods to real water distribution networks are few (e.g. Saldarriaga *et al*. 2006; Wu *et al*. 2010). Developments in other transient test-based techniques for the detection of leaks and illegal connections are promising (Meniconi *et al*. 2011).

doi: 10.2166/ws.2013.080

The methods mentioned above are mostly retrospective, as is the method applied in this paper. For an overview of real-time and field methods for identifying and/or finding leaks, the reader is referred to e.g. Wu (2011).

Van Thienen (2013) presented a new method, the Comparison of Flow Pattern Distributions (CFPD) method, to look at flow data in a different way in order to identify and quantify two different types of changes in a supply pattern, so-called *consistent* and *inconsistent* changes. These different types have different, distinctive interpretations and only one of them can be attributed to changes in leakage. More generally, the method provides additional information about both customer behavior and the evolving condition of the network from data which are usually readily available at water companies.

The methodology section presented below provides a summary of Van Thienen (2013), which gives a more elaborate explanation of the methods used. The focus of this paper is on its applications. The corresponding section provides a number of CFPD analyses of flow data sets from several different supply areas with varying characteristics.

CFPD METHODOLOGY

A complete description of the CFPD methodology is presented in Van Thienen (2013). This section provides a brief overview of the method and a description of the CFPD block analysis.

CFPD 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 period in the following year of the same length (Figures 1(a) and 1(b)). The registered patterns are likely to be similar in shape but not exactly the same. The simple CFPD procedure allows a quantitative comparison of these patterns, taking the following steps:

1. Sort both data sets from small to large magnitude (Figure 1(c)). Sorted measurement ranks, scaled to a 0–1 range, are on the horizontal axis, flow rates are on the vertical axis.
2. Plot one data set against the other in a CFPD plot (Figure 1(d)).
3. Determine a linear best fit with slope a and intercept b.

Note that the word *pattern* is used here in the sense of a time series which is generally repetitive to a significant degree with some variations. The supply area can be quite small (Van Thienen 2013, shows a case with just 70 connections), as long as there is still a more or less regular daily pattern. Theoretically, there is no upper limit to the supply area size, although interpretation of the results may be more difficult in larger areas. Note that it is preferable to use a comparison period of several days or longer to dampen effects of variations in the patterns due to the stochastic nature of water demand. In general, it is preferable to construct the CFPD plot with the first period on the horizontal axis and the second on the vertical. In this case $a > 1$ and/or $b > 0$ corresponds to an increase in flow rate. The procedure for comparison periods of equal length, described here, can easily be performed in an ordinary spreadsheet program. Note that comparison of periods of different length is also possible but requires an expanded procedure and special software, see Van Thienen (2013).

Interpretation

The CFPD diagram which is thus constructed reflects a number of characteristics of the relation between the two

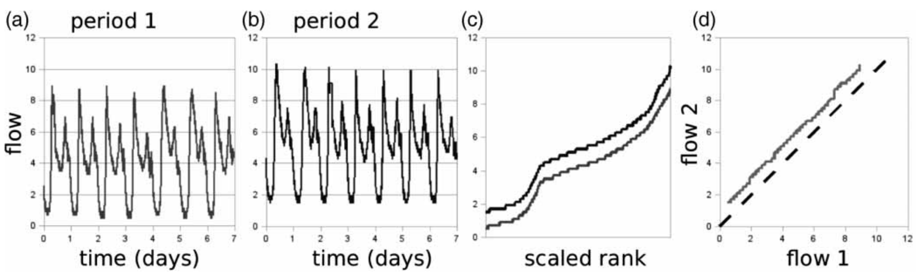

Figure 1 | The CFPD analysis procedure for patterns of equal length. The unit of flow can be any volume over time unit.

flow patterns which are compared in a very simple way. These are illustrated in a number of hypothetical cases, for which perfect data are assumed. If the shape of the flow patterns which are compared is identical, the resulting CFPD curve will be a straight line. 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 1 but it will be offset. This change is defined here as an *inconsistent* change, as it does not follow the existing pattern of variations throughout the day but is uniform. The y intercept b (unit is the same as the flow rate unit used in the input data, e.g. m^3/hour) in the CFPD-plot is equal to the offset value of the pattern. 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. This change is defined here as a *consistent* change, since it does follow the existing flow rate variation. 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. Ideally, a fit should then be made of the part of the CFPD curve which does not deviate if information is sought on processes operating throughout the day, such as leakage.

In general, several processes may be operating simultaneously, and therefore they may be obscured in the flow pattern. The CFPD analysis allows the consistent and inconsistent changes to be isolated, facilitating quantification and interpretation. Note that processes such as the pressure dependence of leakage rates, events with a similar or shorter duration than the comparison time window, and noise will affect the quality of the CFPD fit. These issues are discussed in more detail in Van Thienen (2013).

The interpretation of consistent and inconsistent changes is summarized in Figure 2. Structural changes in population size result in structural consistent changes. Holiday periods may result in either increases (holiday areas: in) or decreases (people leaving to spend their holidays elsewhere: out) in effective population size, which translate into consistent changes in water demand. Warm periods in temperate climates will scale up parts of the water demand (washing, showering) but also add additional demand types, such as garden watering. Large volume customers typically have patterns very different from the average demand pattern of a supply area. Several aspects of the operation and configuration of the network may affect CFPD analyses (reservoirs, valves, connections, etc.). A more elaborate description is available in Van Thienen (2013). By elimination of these known factors, remaining observed *inconsistent* changes ($b \neq 0$) can tentatively be ascribed to increased leakage. *Consistent* changes ($a \neq 1$) do not include a significant leakage signal.

CFPD block analysis

This paragraph describes the application of the CFPD procedure on a long time series, resulting in a comparison of each period (which will be called a *block* in the following) within this time series with each other period. For a more detailed description, the reader is referred to Van Thienen (2013).

Figure 3 illustrates the procedure and results of such a block analysis. A CFPD analysis is made (Figure 3(a)) of all possible combinations of time blocks of a preselected length of the comparison frame within the complete dataset. Two matrices A (Figure 3(b)) and B (Figure 3(c)) are made, in which row i and column j represent blocks i

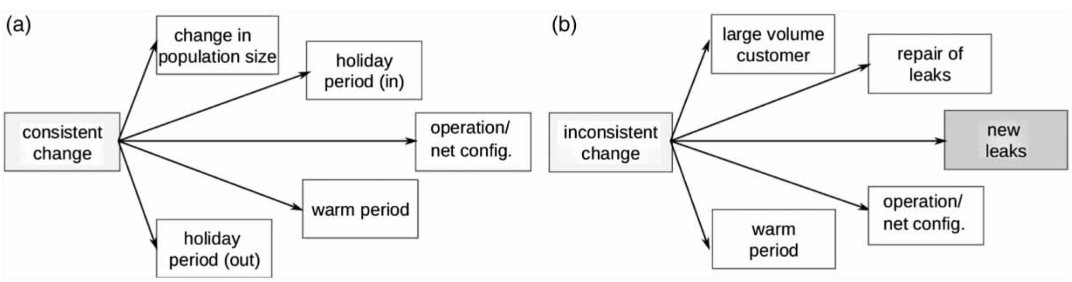

Figure 2 | Interpretation by elimination of consistent (a), and inconsistent changes (b), found in a CFPD analysis.

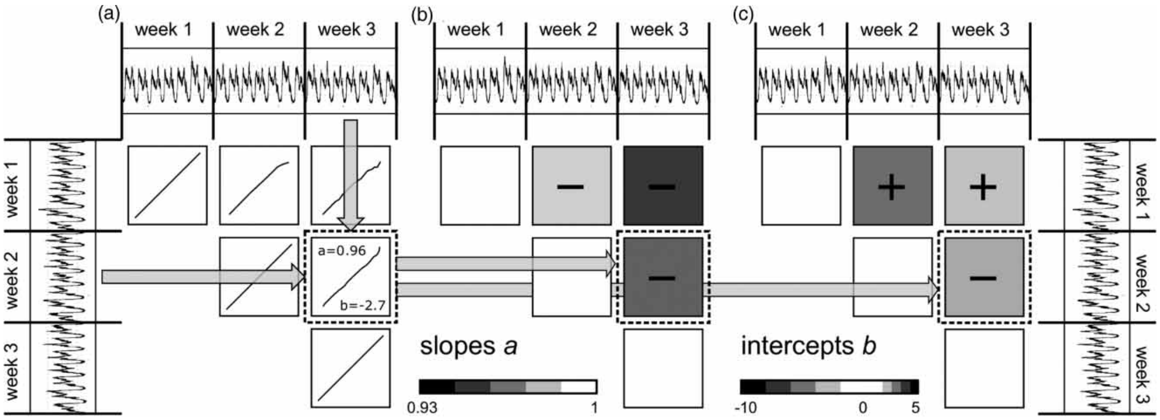

Figure 3 | Illustration of the CFPD block analysis. (a) CFPD analysis for each combination of blocks; (b) visualization of slope values (matrix A); (c) visualization of intercept values (matrix B).

and j (within the time series), respectively, and entries A_{ij} and B_{ij} are the factors a and b, respectively, resulting from a CFPD comparison of block i with period j. The entries in the upper triangle (the lower triangle is not shown, as the matrices are antisymmetric) are grey toned as a function of their deviation from 1 (A) and 0 (B), respectively, with small deviations having a light tone close to white and larger deviations having a darker tone and a sign ($-/=/+$) indicating the direction of the deviation. The complete matrices are constructed because it is usually not clear beforehand 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 (in time) of events which cause these changes.

RESULTS AND DISCUSSION

This section describes applications of the CFPD methodology to several datasets from water distribution networks with varying characteristics.

Area 1 (Amsterdam area – Waternet)

Case description

A predominantly residential area in the Amsterdam area (about 2,850 connections, average daily consumption in the Netherlands is approximately 125 L) has occasional small leaks. A CFPD analysis of flow data for this area was performed and the occurrence of a small leak was identified easily.

CFPD analysis

Figure 4 shows some CFPD analysis results for this area. In Figure 4(a), a CFPD diagram for the comparison of early 2009 to late 2009 is shown (one month periods excluding the first and final days of the year, respectively, which have a deviating pattern). The similarity of the pattern between the periods is illustrated by the absence of significant deviations from the linear best fit line. Its slope shows an 11% consistent increase in water demand, and in addition to this a 3.3 m^3/hour inconsistent increase. A more detailed analysis of the supplied volumes through the year using a block analysis (Figures 4(b) and 4(c)) shows two striking features. The first is the strong consistent drop in water demand in the month of July (Figure 4(b)), corresponding to the school summer holidays. The second is the inconsistent increase which can be seen for the months of July through December relative to the preceeding months. This increase has a maximum amplitude of over 3 m^3/hour. Even though a secondary signal appears to be superimposed on this inconsistent change (the b value is not constant in the period of July–December), the appearance of a new leak of about 3 m^3/hour is quite apparent. When zooming in on the period

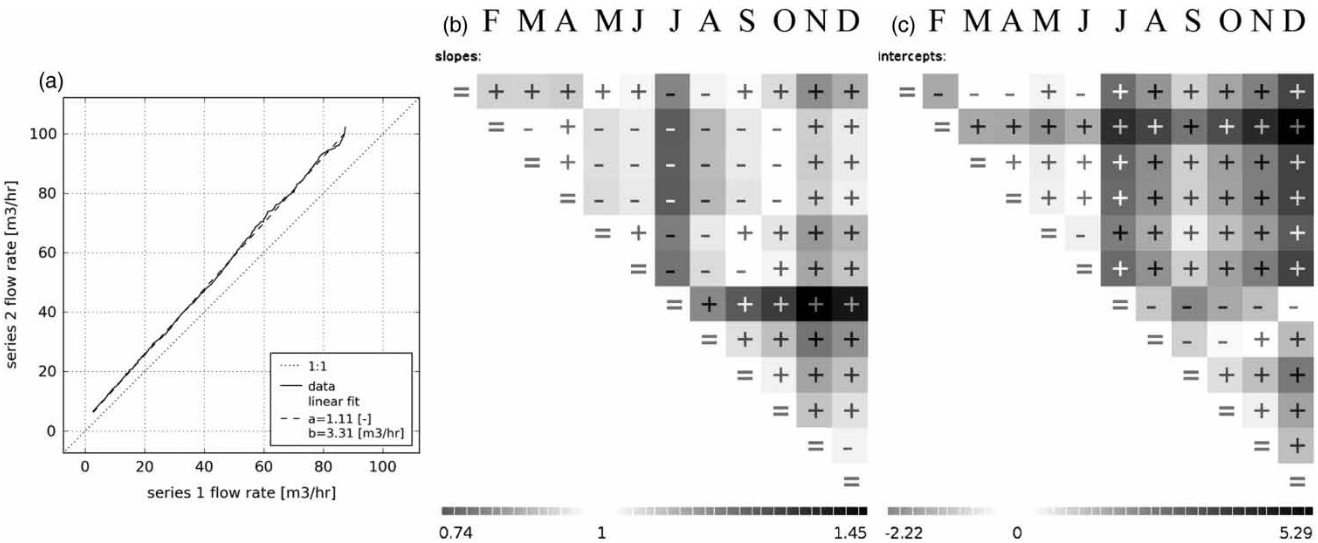

Figure 4 | (a) CFPD diagram comparing the pattern of January 8, 2009 – February 7, 2009 to the pattern of November 22, 2009 – December 22, 2009 for Area 1. (b) Slope matrix for CFPD block analysis of Area 1 data for 2009. (c) Intercept matrix for CFPD block analysis of Area 1 data for 2009.

in which it appears, late June or early July (see Figure 5), its actual first occurrence can be seen on Friday, June 26 in Figure 5(b). Its first occurrence is somewhat obscured by the typical low *b* values which are seen on all Sundays, but when this effect is taken into account, the initiation is unmistakable. Equally prominent is the start of the school summer holidays (Figure 5(a)).

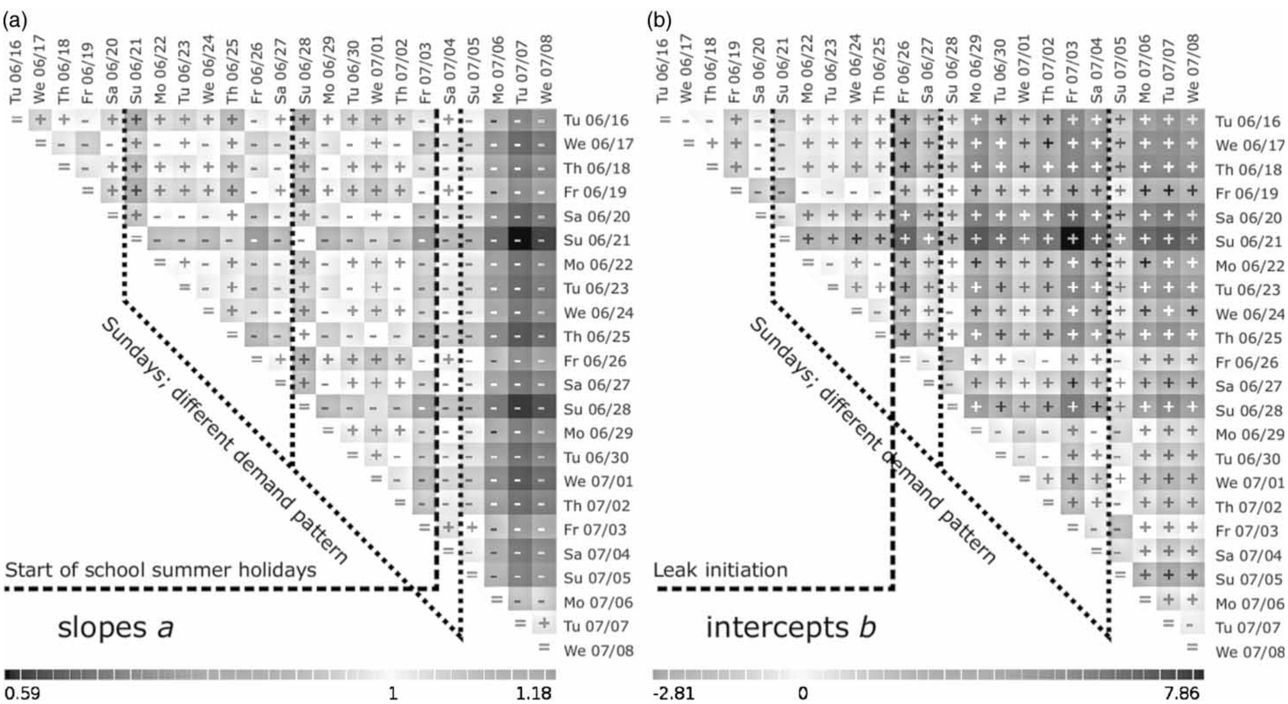

Figure 5 | CFPD block analysis of Area 1 data on a single day basis for the period in which the leak was thought to have initiated.

Area 2 (Vlietregio – Dunea)

Case description

For an area in the west of the Netherlands (about 140,000 connections) with both urban and agricultural land use, and also a number of large volume industrial customers, a CFPD block analysis for a 2-year period was made.

CFPD analysis

A simple CFPD analysis comparing quarter 1 of 2008 data to quarter 1 of 2010 data reveals a small consistent increase in demand of 1.7% and a small inconsistent increase of 33 m^3/hour (compared with a maximum flow of more than 6,000 m^3/hour). Block analysis of the flow data for the years 2008 and 2009 (Figure 6) reveals a number of interesting features. The summer holidays are clearly revealed by consistent decreases in July and August of both years (distinctive negative columns in Figure 6(a)). However, the decrease is much stronger in 2008 than in 2009. This is possibly related to the wet and cloudy summer of 2008, with a below average amount of sunshine and more rainfall which may have made more people decide to spend their holidays abroad, compared with the relatively sunny summer of 2009 (KNMI 2012). The number of holidays taken by Dutch people slightly decreased in 2009 compared with 2008 (CBS Statline 2012), although not by a sufficient amount to explain the observed difference in summer demand dips. However, these numbers apply to the entire year, so people may have taken more holidays outside the summer period.

A second interesting feature is the inconsistent increase which is observed in May–August 2008 and June–August 2009, indicated by dark blocks of pluses in Figure 6(b). These increases are probably related to garden watering. The months of April, May and June 2008 had a below average amount of rainfall. In 2009, August and September were unusually dry, whereas July had more rainfall than the many year average. Note that these inconsistent increases partially coincide with the consistent decreases related to summer holidays described above. The CFPD method has no problem distinguishing these opposite effects.

The third interesting feature is that most of the long-term consistent increase which was observed in the first quarter comparison of 2008–2010 (described above) appears to have taken place in 2008. The upper half of Figure 6(a), comparing 2008 and 2009 data with 2008 data, shows a predominance of pluses (apart from the holiday dips), suggesting accumulating increases throughout the year. On the other hand, 2009 data compared with 2009 data (lower

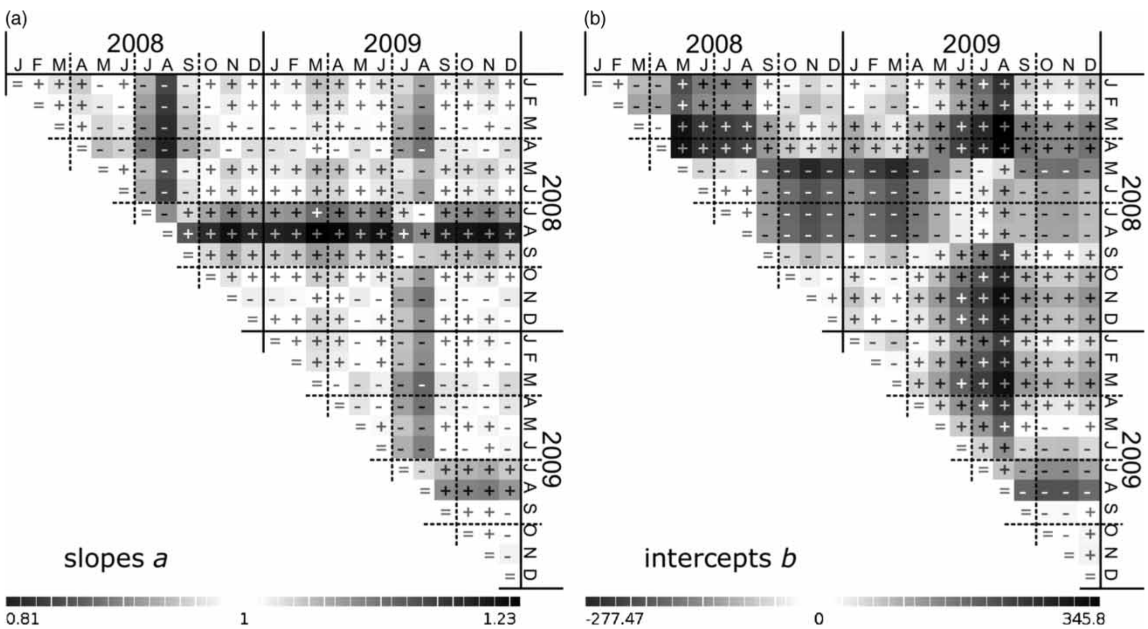

Figure 6 | CFPD block analysis for Area 2 data.

half of Figure 6(a)) shows pluses and minuses to be more balanced and generally having a smaller amplitude.

Obviously these effects have been known for a long time. However, they provide a good illustration of how these different effects which may occur simultaneously and have an opposite sign can be distinguished qualitatively and quantitatively by the CFPD method. It should be noted that for an area of this size (140,000 connections), many (applications of) processes which affect net water demand may be operating simultaneously, which makes interpretation of CFPD block analysis diagrams more difficult for these large areas than for smaller ones.

Area 3 (DMA in a Vietnamese city)

Case description

This city in Vietnam has slightly under 1 million inhabitants. It is divided into six supply areas, two of which are isolated. A CFPD block analysis was performed on 18 months of data for one of these DMAs (district metering area – approximately 137,000 inhabitants, average daily demand of 110 L per person).

CFPD analysis

Figure 7 shows a CFPD block analysis for parts of 2010 and 2011 with monthly average high temperatures (solid curves) and monthly average rainfall (dashed curves, both many year averages) indicated on top. In this case, both the consistent and the inconsistent changes (Figures 7(a) and 7(b), respectively) show a very significant change through the year and a strong correlation to climate parameters, but with a lag. These variations are presumably related to the use of evaporative coolers, garden watering, etc. These strong variations through the year make an interpretation of parts of the observed changes in terms of non-climate parameters (such as increased leakage and demand) more challenging. An interesting exercise is to do a month-by-month comparison of subsequent years. The associated cells have been marked with dashed squares in Figure 7. These show both positive and negative consistent changes for the period of May–October 2011 compared with the same period in the preceding year (Figure 7(a)), and varying inconsistent increases for the period of June–October 2011 compared with the same months of the preceding year. A

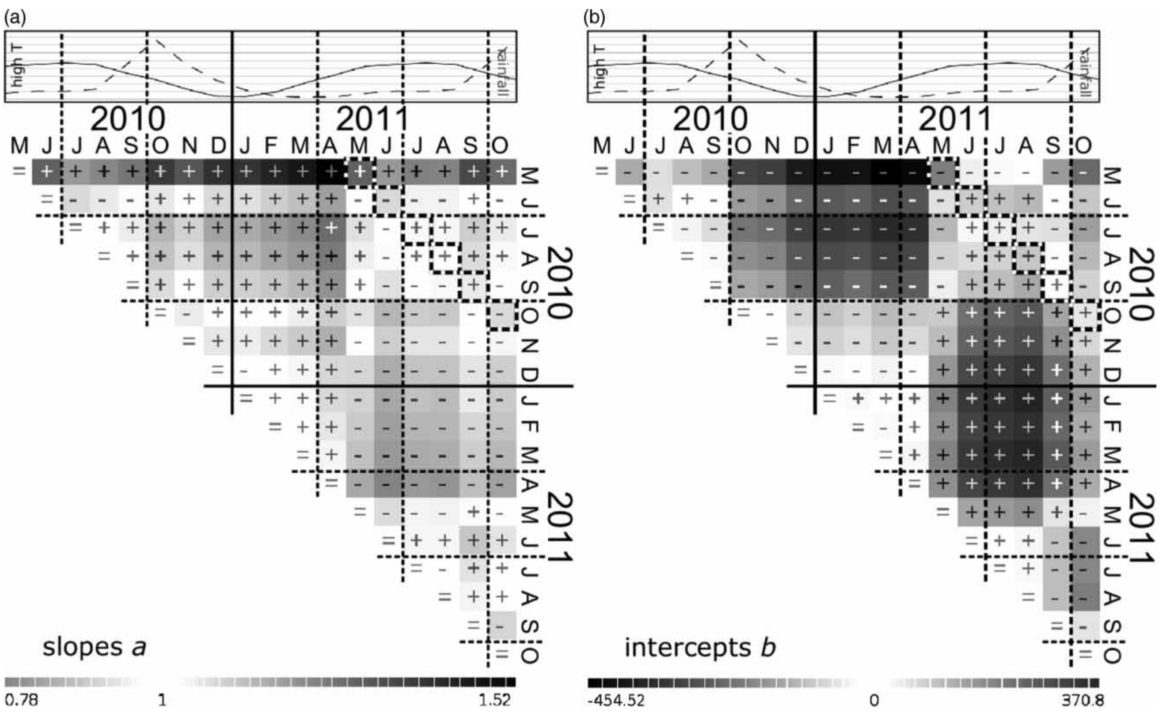

Figure 7 | CFPD block analysis of Area 3, with local climate trends (solid curves: high temperature, dashed curves: rainfall; many year averages, World Weather Information Service (2012)).

multi-year analysis should give a better understanding of the climate-driven patterns in the CFPD block diagram. This may allow these patterns to be eliminated (by subtraction) to get a clearer picture of other factors such as increases in demand and leakage. This is recommended for future work.

CONCLUSIONS

The cases which are described in this paper provide example applications of the CFPD method:

- simple identification and quantification and exact timing of a small new leak (3 m^3/hour) in flow data for a small supply area (2,850 connections);
- interpretative analysis of multi-year flow changes connected to consumer behavior and weather conditions, with coinciding opposing effects being separated, identified and quantified independently;
- difficulties of interpretation in supply patterns which are strongly influenced by climate.

The CFPD block analysis provides a simple tool to analyze flow changes, distinguishing between different types which aids interpretation and quantification, and to visualize them such that features and trends in complicated time series become apparent at a glance.

ACKNOWLEDGEMENTS

The authors would like to thank Arne Bosch of Waternet, Maurice van de Roer of Dunea, and Duc Hanh Nguyen for supplying the flow data analyzed in this paper. Also, the constructive comments of three anonymous reviewers are gratefully acknowledged.

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, USA.

CBS Statline 2012 http://statline.cbs.nl, data retrieved on April 4, 2012.

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

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

KNMI 2012 http://www.knmi.nl/klimatologie/maand_en_seizoensoverzichten/, data retrieved on April 4, 2012.

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. *Journal of Hydraulic Engineering* **120** (8), 934–955.

Meniconi, 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. *Journal of Water Resources Planning and Management* **137** (2), 238–241.

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

Puust, R., Kapelan, Z., Savic, D. & Koppel, T. 2006 Probabilistic leak detection in pipe networks using the SCEM-UE algorithm. *Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium*, Cincinnati, USA.

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, USA.

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

Van Thienen, P. 2013 A method for quantitative discrimination in flow pattern evolution of water distribution supply areas with interpretation in terms of demand and leakage. *Journal of Hydroinformatics* **15** (1), 86–102.

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

World Weather Information Service 2012 http://worldweather.wmo.int, World Meteorological Organization, data retrieved on April 6, 2012.

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

Wu, Z. Y., Sage, P. & Turtle, D. 2010 Pressure-dependent leak detection model and its application to a district water system. *Journal of Water Resources Planning and Management* **136** (1), 116–128.

First received 1 October 2012; accepted in revised form 13 December 2012