International Conference on Hydroinformatics

2014

Flow Analysis And Leak Detection With The CFPD Method In The Paris Drinking Water Distribution System

Peter van Thienen

Frank Montiel

FLOW ANALYSIS AND LEAK DETECTION WITH THE CFPD METHOD IN THE PARIS DRINKING WATER DISTRIBUTION SYSTEM

PETER VAN THIENEN (1), FRANK MONTIEL (2)
(1): Drinking water infrastructure group, KWR Watercycle Research Institute, Groningenhaven 7, 3433 PE Nieuwegein, Netherlands
(2): Eau de Paris, 19, rue neuve Tolbiac, 75013 Paris, France

In principle, DMA flow data can provide valuable information about new leaks, unregistered changes in valve status, etc. However, distinguishing one from the others is often difficult. The CFPD (Comparison of Flow Pattern Distributions) method is a flow time series data transformation which facilitates the identification, quantification, and interpretation of changes in the amounts of water supplied. In this way, it helps to distinguish e.g. new leaks from operational signals and demand changes. In the past years, it has been successfully applied at several Dutch drinking water companies. In this paper, we illustrate the application of the CFPD method by presenting selected results from CFPD analyses of flow data for 16 of 30 DMAs in the city of Paris. The findings are compared to a list of registered leaks.

INTRODUCTION

The Comparison of Flow Pattern Distributions (CFPD) method was introduced [1,2,3] as a new tool to look at flow data for a DMA or supply area in order to pinpoint, identify, and quantify changes in the amount of water supplied. It has since then been successfully applied in multiple projects with Dutch drinking water companies to identify for example leakages and incorrect valves statuses and network connections. Contrary to the often applied minimum night flow analysis, it uses all available flow data (24 hours per day, resolution of 1 measurement per hour or better) and recognizes different types of changes. In this way, it helps to distinguish e.g. new leaks from operational signals and demand changes.

The municipal drinking water company of the city of Paris, Eau de Paris, decided to test this method in order to improve its network efficiency. Paris is a city with 2 million drinking water consumers at night and 4 million during the day, with an average consumption of drinking water that amounts to 550 000 m^3/day. In order to achieve its objectives of quantity and quality of supply, Eau de Paris very swiftly equipped itself with a leading edge computerized system for purposes of network supervision and control.

In order to reduce water losses, Eau de Paris has, over the years, developed and set up tools of detection and analysis of real and differed time information of the central remote control and command system to identify and to locate water leaks, based on a subdivision of the city network into 30 District Metered Areas (DMAs) [4,5].

In this paper, we illustrate the application of the CFPD method by presenting selected results from CFPD analyses of flow data from the Paris SCADA system for 16 of 30 DMAs in the city of Paris. The findings are compared to a non-censored list of registered leaks, showing a very good performance of the method.

CFPD METHODOLOGY

A complete description of the CFPD methodology is presented in a paper by Van Thienen [1]. This section provides a brief overview of the method and a description of the CFPD block analysis, and is largely taken from a paper by Van Thienen *et al.* [2].

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 next year of the same length (Figure 1a,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 1c). 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 1d).
3. Determine a linear best fit with slope *a* and intercept *b*.

The slope or scaling factor *a* represents so called *consistent changes* in the supply volume; the y-axis intercept *b* (unit is the same as the flow rate unit used in the input data, e.g. m3/hour) represents the so called *inconsistent changes*. Both have distinctive interpretations, which will be discussed below. Note that consistent and inconsistent changes are purely numerical characteristics of the comparison of the two periods.

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 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 [1].

CFPD block analysis

Application of the CFPD procedure to long time series can be done using the CFPD block analysis, which performs a comparison of each period (which will be called block in the following) within this time series with each other period. For a more detailed description, the reader is referred to [1].

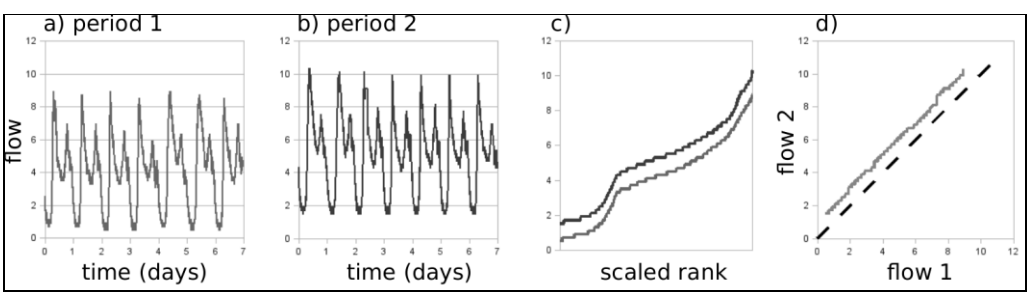

Figure 1: The CFPD analysis procedure for patterns of equal length. The unit of flow can be any volume over time unit.

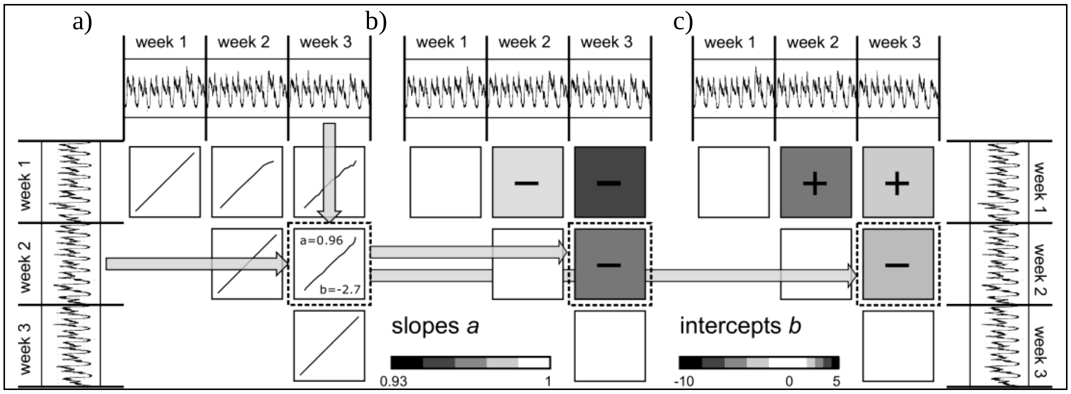

Figure 2: Illustration of the CFPD block analysis. a) CFPD analysis for each combination of blocks, b) visualization of slope values (matrix *A*), c) visualization of intercept values (matrix *B*). Copied from [2].

Figure 2 illustrates the procedure and results of such a block analysis. A CFPD analysis is made (Figure 2a) of all possible combinations of time blocks of a preselected length of the comparison frame within the complete dataset. Two matrices *A* (Figure 2b) and *B* (Figure 2c) are made, in which row *i* and column j represent blocks *i* 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 deviation having a light tone close to white and larger deviations having a darker tone and a sign (-/=/+) indicating the direction of the deviation (or, as in the following, by using different color scales for positive and negative deviations). 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.

Interpretation

In general, several processes may be operating simultaneously, and therefore they may be obscured in the flow pattern. The CFPD analyses 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 [1].

The interpretation of consistent and inconsistent changes is summarized in Figure 3. 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

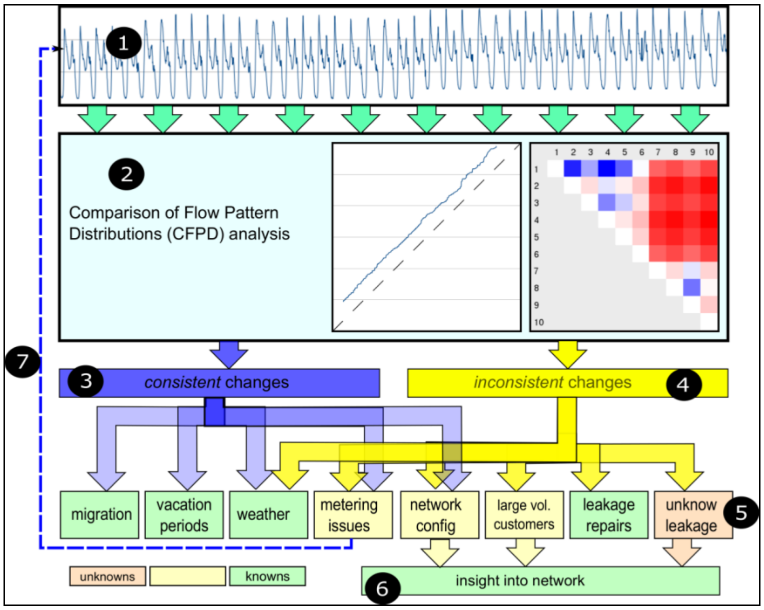

Figure 3: CFPD analysis procedure and interpretation. 1) flow time series; 2) CFPD analysis; 3,4) identification of consistent and inconsistent changes; 5) interpretation of these in terms of known and unknown mechanisms; 6) discarding changes by known mechanisms results in a reduced list of unknown events which a limited set of possible interpretations; 7) any data quality issues which are found may initiate improvement measures (modified from [3]).

(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.). By elimination of these known factors, remaining observed inconsistent changes ($b \neq 0$) can tentatively be ascribed to one of a limited number of factors, most likely increased leakage or unregistered changes in the network configuration (boundary valves) .

Parisian distribution network and flow data

The Parisian distribution network has been described in detail by Montiel *et al.* [4,5]; a summary of their description is provided here.

Paris is a densely populated city of about 2 million inhabitants with again as many workers coming to the city during the day, built on somewhat hilly terrain. To ensure the security and the reliability of water distribution in Paris, the pipe network follows a meshed design philosophy. The city covers an area of 100 km^2 and is equipped with a 1800 km long drinking water pipe network. A specific water distribution regulation is defined for Paris: the pressure on the network is maintained constant. This regulation presents several important advantages, including energy savings, protection of the infrastructure from pressure fluctuations, and offering consumers a stable pressure at their tap.

Given the topography of the city, the water distribution network has been divided into sub-networks depending on the ground elevation, and for each sub-network the pressure is maintained constant. The water at the entry point of the city is distributed into three different isolated groups of sub-networks. If necessary, it is possible to connect two sub-networks for water transfer. To allow a reliable control of the water distribution in Paris, a 15 000 information points SCADA system is used. It controls more than 200 flow meters and 120 pressure meters dispersed throughout the network. This real time information is used for leakage detection [4,5]. In order to facilitate this, the network has been naturally or virtually separated sub-networks. The natural sub-networks are topographically based isolated from each other by valves which are generally closed. The virtual sub-networks are controlled by flow-meters. Each sub-network is divided with flow-meters into several district metering areas (DMAs).

The dataset used in this paper is an extract from Paris real time SCADA system historical records (with data since 1998). The resolution of the dataset depends on the considered flow meters; it ranges from 2.3 minutes to 15 minutes time resolution (in the real time SCADA system information are available on a second basis).

In order to calculate the distributed volume in the 30 areas, the system takes into account around 130 flow meters with a precision ranging from 0.5 to 5%. For some areas including tanks, the variation of volume of the tank is taken into account, considering the tank capacities.

RESULTS

Error! Reference source not found. gives an overview of registered leaks in the period of available data. The flow rate of these registered leaks has been determined using the average night flow increase, combined with the decreased flow when the broken pipe has been isolated using the real time information of the SCADA system. Table 1 also shows CFPD block analysis results. The corresponding block diagrams are shown in Figure 4. All but one of the registered leaks can be recognized in the data, although the clarity with which they can discerned varies. Different types of leakage and other features seen in the data are:

Bursts lasting less than a day	These show up as a simultaneous anomaly in the a and b tables, with opposite signs (which either combination possible: a burst during low consumption hours will results in $a<1$ and $b>0$ and a burst during high consumption hours results in $a>1$ and $b<0$. An example is shown in Figure 4a,b.
Multiple day leaks with constant flow rate	These show a very clear, constant positive anomaly in the block diagrams (factor *b*). An example is shown in Figure 4c.
Expanding leaks	These leaks start out small, and grow over the course of several days to reach a plateau. An example is shown in Figure 4d.
Leak repairs	Leak repairs on their own (i.e. without the leak initiation in the same diagram) show up as more or less constant negative anomalies in block diagrams (factor *b*). An example is shown in Figure 4e.
Holiday periods	Incidental or structural changes in population size clearly show up in the factor a block diagrams as positive or negative anomalies. A clear example is the end of the 2012 summer holidays in Figure 4f, with consistent increases in two subsequent weeks (the first is marked).
Small leaks	When the flow rate of a leak is comparable to natural variation caused e.g. by consumption pattern differences between week days and weekends, a leak is more difficult to discern. An example is shown in Figure 4g, which shows a leak superimposed on a very clear and regular weekday-weekend pattern.
Superposition of	When multiple processes with a significant impact on the flow rate

multiple effects occur simultaneously in a DMA, CFPD block diagrams become more difficult to read and interpret. Examples are shown in Figure 4h,i.

Note that some of the leaks listed in **Error! Reference source not found.** show up at slightly different (starting and/or ending) dates in the CFPD analyses. The can be due either to incorrect registration or misinterpretation of the diagrams.

Table 1: Overview of registered leaks and their detection in CFPD block analyses.

DMA	**Reported**			**CFPD analysis**				
	Start	**End**	**Flow**	**Start**	**End**	**Flow**	**Fig. 4**	**Cla-rity**
Belleville	27-4-11 00:00	5-5-11 00:00	80	28-4-11	5-5-11	60	h	low
Belleville Réservoir	09-12-11 4:00	09-12-11 12:00	peak 3500	9-12-11	9-12-11	unk.		High
Chapelle	12-3-12 00:00	23-3-12 10:30	300	6-3-12	23-3-12	200		High
Cité Universitaire	5-10-12 10:00	24-10-12 11:00	35	Slow increase 1 through 7 Oct, burst on 8-10-12	24-10-12	20		Med
Convention	11-6-11 03:00	15-6-11 14:00	40	11-6-11	14-6-11	200		Low
Vaugirard	12-6-11 12:00	17-6-11 01:00	110	12-6-11	16-6-11	140	c	High
Courcelles	11-2-12 21:30	16-2-12 15:00	700	12-2-12	16-2-12	300		High
Courcelles	27-4-12 00:00	2-5-12 00:00	300	27-4-12	2-5-12	400		High
Courcelles	26-9-12 05:15	27-9-12 12:00	1100	26-9-12	27-9-12	unk.	a,b	Low
Daumesnil	11-9-12 03:00	8-10-12 14:00	100	12-9-12	7-10-11	100	i	Low
Fabien	15-1-11	20-1-2011	700	16-1-11	19-1-11	550		Low
Fabien	8-7-11 00:00	3-8-11 00:00	640	11-7-11	3-8-11	500		Low
Maine	24-5-12 05:10	26-5-12 20:50	250	25-5-12	28-5-12	150		high
Menilmontant	13-5-11 13:00	15-5-11 09:00	2200	14-5-11	15-5-11	2500		High
Nation	25-12-11 04:30	27-12-11 00:00	1100	25-12-11	27-12-11	1100		High
Olympiades	28-8-12 14:00	18-10-12 09:00	50	28-8-13	?	35	g	Low
Olympiades	24-9-12 10:50	24-9-12 17:30	700	24-9-13	24-9-13	unk.		High
Plaine Vaugirard	29-11-2010		260	29-11-13	29-11-13	unk.		High
Plaine Vaugirard	30-3-11 00:00	27-4-11 00:00	60	2-4-11	25-4-11	15-55[1]	d	High
Rivoli	20-12-10 00:00	25-01-11 12:00	300	15-12-10	24-1-11	150-300 [1]	e	Med
Rivoli	27-2-12 01:30	27-2-12 18:45	670	27-2-12	27-2-12	unk.[2]		Low
Sorbonne		17-6-11 12:00	50					None

[1] slowly increasing; [2] only in standard deviations

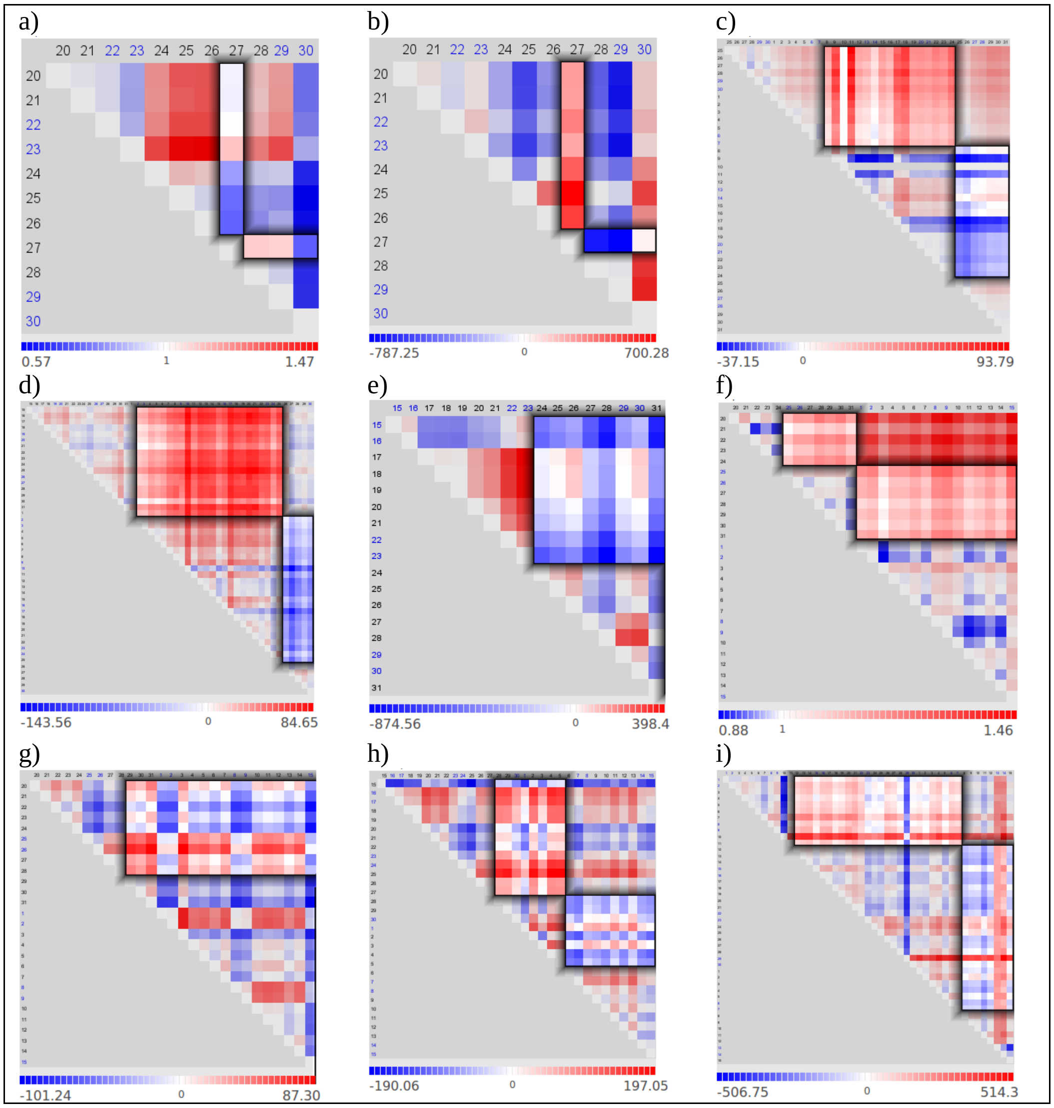

Figure 4: a) Courcelles, September 2012, factor *a*; b) Courcelles, September 2012, factor *b*; c) Vaugirard, June 2011, factor *b*; d) Plaine Vaugirard, March-April 2011, factor *b*; e) Rivoli, January 2011, factor *b*; f) Olympiades, August-September 2012, factor *a*: end of the summer holidays; g) Olympiades, August-September 2012, factor *b*; h) Belleville, April-May 2011, factor *b*; i) Daumesnil, September-October 2012, factor *b*.

DISCUSSION

Performance

Of all 22 leaks registered in the period of the available data, 21 of varying types could be recognized, many quite easily and some with more difficulty. This provides a very good illustration of the CFPD method and its ability to show leaks in an easy and accessible way in complex flow data. This does not provide a validation, however. There are many features in the provided data set which cannot be linked to any registered leak. These may be related to unknown leaks, but also to any of the other factors shown in Figure 3. Note that leak registration is far from complete for this period, as is the registration of other important factors, such as valve manipulations. Complete and accurate registration of everything that happens in a distribution network (leakage repairs, valve manipulations, changes in pressure regime, etc.) is not common practice yet with drinking water companies. However, in order to get the most information out of flow data using the CFPD or any other method, this information is essential. It allows a drinking water company to separate the wheat from the chaff in flow data, allowing it to focus on the real issues, such as unknown new leaks.

Outlook

The visual identification and interpretation of the CFPD analysis results as presented in this paper still requires human intervention. Ongoing research focuses on the automated recognition of features in CFPD block diagrams, which has several advantages, including reduction of analysis time, more objective analysis, and the possibility of automatized alarms.

This automated feature recognition, combined with anomaly comparison of neighboring DMAs as already implemented at Eau de Paris [4,5], can be implemented on top of a SCADA system giving automatic alarms when anomalies are observed and classified as potentially important.

CONCLUSIONS

The successful recognition of 21 out of 22 registered leaks of varying types provides a very good illustration of the CFPD method and its ability to show leaks in an easy and accessible way in complex flow data. The simplicity of the method facilitates its implementation in a real time process.

ACKNOWLEDGMENTS

The authors would like to acknowledge W-SMART and the utilities participating in and sponsoring its INCOM project (Eau de Paris and Eaux du Nord), and also the European Water Innovation fund of the Dutch water companies for additional funding.

REFERENCES

[1] Van Thienen, P., "*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, doi: 10.2166/hydro.2012.171 (2013).

[2] Van Thienen, P., Pieterse-Quirijns I., Vreeburg J.H.G., Vangeel K. and Kapelan Z., "*Applications of discriminative flow pattern analysis using the CFPD method*", Water Science and Technology: Water Supply, Vol. 13, No. 4, (2013), pp. 906-913.

[3] Van Thienen, P., Vreeburg J. and De Kater H. , "*Water flow data key to pinpointing change*", Water21, June, (2013), p. 36.

[4] Montiel, F. and Nguyen B., *"Efficient real and differed time tools and method for leakage detection in the city of Paris*", 6th IWA Specialist Conference on Efficient Use and Management of Water, (2011).

[5] Montiel F., Nguyen B., Juran I. and Shahrour I., "*Real-time water leak detection and analysis tools*", 7th IWA International Conference on Efficient Use and Management of Water, (2013).