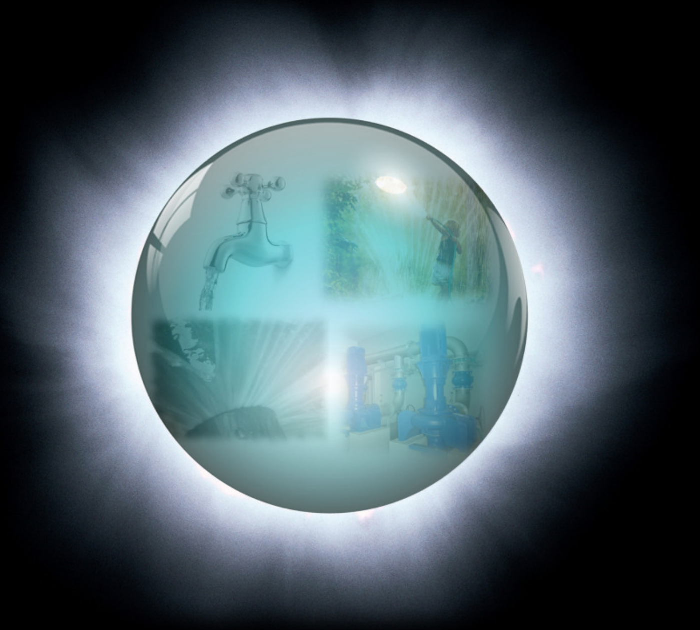
Optimised control and pipe burst detection by water demand forecasting



Martijn Bakker

Stellingen

Behorende bij het proefschrift
“Optimised control and pipe burst detection by water demand forecasting”
Martijn Bakker – 17 Oktober 2014

1. Hoe hoger het gemiddelde waterverbruik in het gebied waarvan het waterverbruik wordt voorspeld, hoe kleiner de procentuele voorspelfouten, maar hoe groter de absolute voorspelfouten (dit proefschrift).
2. In gebieden met een gematigd klimaat, kan het gebruik van informatie over het weer de prestaties van een model dat de drinkwatervraag voorspelt slechts beperkt verbeteren (dit proefschrift).
3. Geoptimaliseerde prognosebesturing leidt tot herhaalbare, significante verbetering van de waterkwaliteit en energiebesparingen (dit proefschrift).
4. De potentiële voordelen van dynamische druksturing in netwerken met een aanzienlijk lekverlies zijn groter dan de potentiële voordelen van prognosebesturing van de productie (dit proefschrift).
5. De introductie van slimme watermeters in huishoudens biedt de mogelijkheid om de sensordichtheid in water distributienetwerken sterk te verhogen, hetgeen zal leiden tot een revolutie op het gebied van leidingnet monitoring en besturing.
6. Het feit dat het nachtverbruik niet significant is veranderd in de afgelopen twintig jaar wijst erop dat de veronderstelde verdere toename van de 24-uurs economie het gedrag van mensen (nog) niet heeft beïnvloed.
7. Zelfs als de Nederlandse waterbedrijven het investeringsniveau voor leidingvervanging verdubbelen, zal het aantal ongeplande onderbrekingen in de watervoorziening in de komende decennia zeer sterk toenemen (Peter Horst, IWA Waterloss Conference 2014, Wenen).
8. Door social media nauwkeurig in de gaten te houden kunnen waterbedrijven ontdekken dat er een storing is in de watervoorziening, omdat veel mensen liever iets op internet zetten dan contact opnemen met het waterbedrijf.
9. Tijdens het doen van een promotieonderzoek, naast werken voor een ingenieursbureau, wordt men geconfronteerd met het spanningsveld tussen wetenschappelijke grondigheid en commerciële budget overwegingen, hetgeen zowel verrijkend als uitputtend is.
10. Arjen Robben heeft laten zien dat opstaan na te zijn gevallen of te zijn neergehaald, een effectieve strategie is om het doel te bereiken.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotoren
Prof.dr.ir. L.C. Rietveld and dr.ir. J.H.G. Vreeburg

Propositions

Accompanying the thesis
"Optimised control and pipe burst detection by water demand forecasting"
Martijn Bakker – 17 October 2014

1. The higher the average water demand in the area of which the demand is forecasted, the smaller the percentage forecasting errors but the larger the absolute forecasting errors (this thesis).
2. Weather input can only slightly improve the performance of a water demand forecasting model in areas with a moderate climate (this thesis).
3. Optimised control based on water demand forecasting leads to repeatable, significant water quality improvements and reductions of energy consumption (this thesis).
4. The potential benefits of dynamic pressure control in networks with considerable real losses are larger than the potential benefits of production flow control (this thesis).
5. The introduction of smart domestic water meters provides the opportunity to dramatically increase the sensor density in water distribution networks which will induce a revolution in the field of on-line network monitoring and control.
6. The fact that night flows in water demand patterns have not significantly changed in the past twenty years indicates that the perceived upcoming 24-hours economy did not change people's behavior.
7. Even if the water companies in the Netherlands double their investment level in pipe replacement, the number of unplanned interruptions of supply will increase largely in the next decades (Peter Horst, IWA Waterloss Conference 2014, Wenen).
8. Effectively monitoring social media can help water companies to detect that there is a failure in the water supply system, because many people rather post comments on internet than contact the water company directly.
9. When doing PhD research beside working for an engineering firm, one experiences the tension between scientific thoroughness and commercial budget limitations, which is both enriching and exhaustive.
10. Arjen Robben has shown that standing up after falling or after being attacked is an effective strategy to achieve the goal.

These propositions are considered opposable and defendable and as such have been opproved by the supervisors
Prof.dr.ir. L.C. Rietveld and dr.ir. J.H.G. Vreeburg

Optimised control and pipe burst detection by water demand forecasting

Martijn Bakker

Optimised control and pipe burst detection by water demand forecasting

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van Rector Magnificus Prof.ir. K.C.A.M. Luyben,
voorzitter van het College van Promoties,
in het openbaar te verdedigen
op 17 oktober 2014 om 12:30 uur

door Martijn BAKKER
civiel ingenieur
geboren te Niedorp

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr.ir. L.C. Rietveld

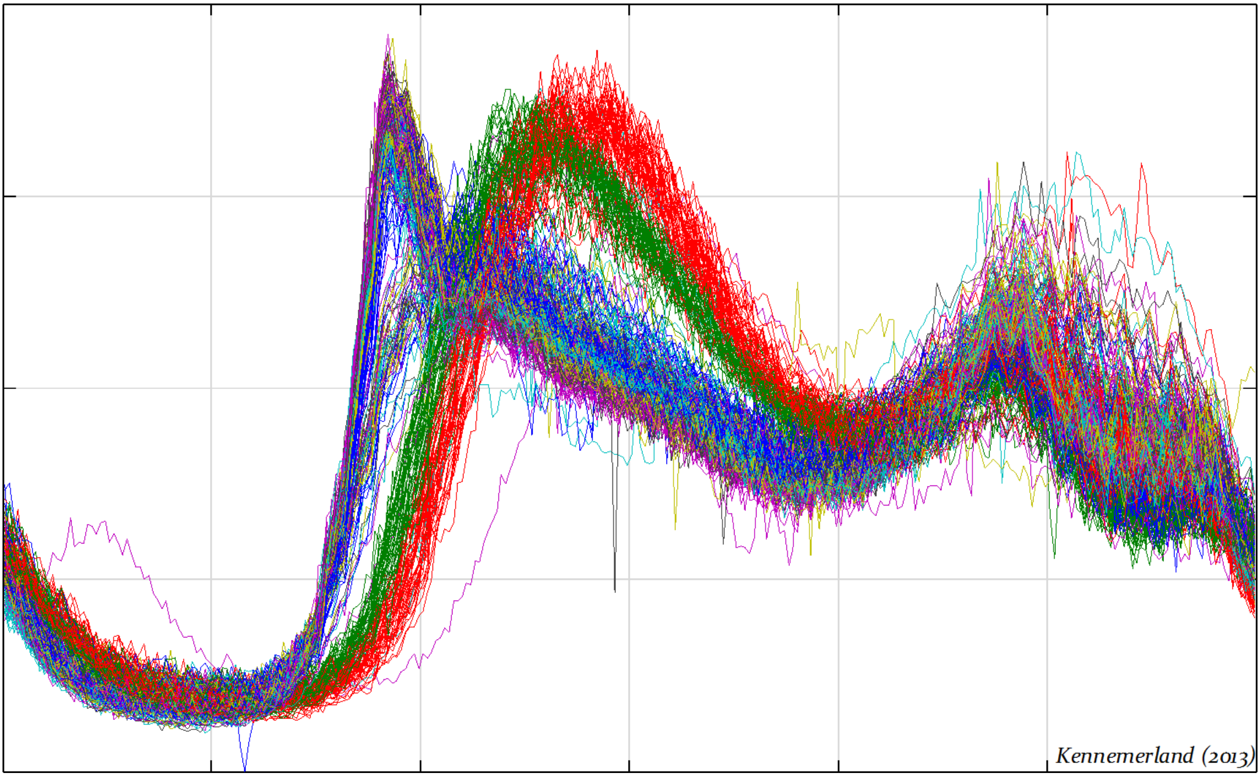
Samenstelling promotiecommissie:

Rector Magnificus, voorzitter
Prof.dr.ir. L.C. Rietveld, Technische Universiteit Delft, promotor
Dr.ir. J.H.G. Vreeburg, Wageningen University, copromotor
Prof.dr. J.E. Van Zyl, University of Cape Town
Prof.dr. K.E. Lansey, The University of Arizona
Prof.dr.ir. B. De Schutter, Technische Universiteit Delft
Prof.dr. D.P. Solomatine, UNESCO-IHE Institute for Water Education / Technische Universiteit Delft
Prof.dr. J.B. Boxall, The University of Sheffield
Prof.dr.ir. J.P. van der Hoek, Technische Universiteit Delft, reservelid

This work was performed in the DisConTO project (Distribution Control Training & Operation). The DisConTO project was executed by a collaboration of four Dutch water supply companies (PWN, Vitens, Dunea and Brabant Water), the Delft University of Technology, the National Institute for Public Health and the Environment (RIVM), Consulting and Engineering firm Royal HaskoningDHV and intelligent software provider UReason. The project was financially supported by the Dutch government through the "Innowator" programme.

Printed by Gildeprint Drukkerijen
ISBN 978-94-6186-323-2

To Mariska, Robin and Joey


Kennemerland (2013)

Summary

Water demand forecasting

The total water demand in an area is the sum of the water demands of all individual domestic and industrial consumers in that area. These consumers behave in repetitive daily, weekly and annual patterns, and the same repetitive patterns can be observed in the drinking water demand. The observations of the water demand were used to develop a fully adaptive forecasting model for short-term drinking water demand. This heuristic model automatically stores and updates water demand patterns and demand factors for all days of the week and for a configurable number of deviating days like national holidays, vacation periods, and individual deviating days. The model uses this information to adaptively learn the patterns and factors that are used when forecasting the water demand for the next 48 hours with 15-min. time steps (192 values). The model uses the measured water demand and static calendar information that appoints deviating days as only input. The model has functionality to identify extra water demand in the evening which is typical for peak water demand during good weather periods. Once this extra demand is identified, the model immediately changes its demand forecast to correctly forecast the deviating demand pattern. The model is easy to implement, fully adaptive and accurate, which makes it suitable for application in real-time control and pipe burst detection.

The model was tested on datasets containing six years of water demand data in six different areas in the central and southern part of Netherlands. The areas have all the same moderate weather conditions, and vary in size from very large (950,000 inhabitants) to small (2,400 inhabitants). The mean absolute percentage error (*MAPE*) for the 24-hours forecasts varied between 1.44-5.12%, and for the 15-min. time step forecasts between 3.35-10.44%. The analysis of the forecast errors showed a strong relation between the forecast error and the size of the area: the errors decreased linearly with a logarithmic increase of the average demand in the area.

Water demand is correlated to weather conditions: high temperatures and low rainfall lead to an increase of the water consumption. Therefore, it may be expected that using weather input will improve the performance of water demand forecasting models. The possible performance improvements when using weather input (temperature information) in the forecasting model and in two other models were studied. It was found that the average forecasting errors were 6.3% smaller, and the largest forecasting errors were 9.4% smaller when weather information was used in the models. This indicates that indeed the performance can be improved when using weather input, especially to reduce the largest overestimates and underestimates.

Optimised control
A first application of a short-term water demand forecasting model is using it for optimised control of water supply systems. The conventional automatic control of the production flow or the clear water pumps is often quite simple, resulting is highly varying production flows and pump flows. This basic control results in sub-optimal operation of the system, and the operation can be improved when forecasts of the water demands in the system are used. The short-term demand forecasts offer detailed information about the volume of water to be supplied to the consumers in the next 48 hours. The reservoirs in the water supply system provide the possibility to optimise the control, by balancing supply and demand, or to advance or postpose production or pump flows. When applying a control method that uses water demand forecasts, the optimal use of the reservoirs can be achieved.

In the Netherlands, the penetration of optimised control models based on a water demand forecast is quite high: more than half of the water supply systems are controlled by such models. To assess the differences between conventional basic control and optimised predictive control, five existing water supply systems in the Netherlands were examined. The operational results in one week with conventional control were compared to the results in one week with optimised control. The results showed that the variation in the production flow was 75% lower with optimised control. Due to the more constant operation of the production plants, the turbidity levels in the clear water were on average 17% lower. The optimised control also affected energy consumption and energy costs. Optimised control led to a decrease of 3.1% of the overall energy consumption and a 7.4% shift from high tariff to low tariff energy consumption. The resulting energy costs savings were 5.2% on average.

The effects of applying optimised control at a water supply system outside the Netherlands were studied as well. Therefore, the model was implemented for the control of part of the water supply system of Poznań, in the central western part of Poland. The model was applied for both the production flow control and the pump (pressure) control of the clear water pumps. The results of the production flow control were similar to the results observed at the Dutch water supply systems. The optimised pump pressure control at the Polish system proved to be especially valuable. The result of the optimised control was that the pump pressure was reduced by 29% and that the average pressure in the supply areas was reduced by 18%. As a result, the estimated background leakage was reduced by 18% as well. The combined effect of a more constant production flow, a lower pump pressure, and a reduction of the background leakage resulted in a reduction of the overall energy consumption of the system of 11.5%.

Pipe burst detection

A second application of a short-term water demand forecasting model is using it for pipe burst detection. The actual real-time value of the water demand forecast (the so-called now-cast) is a good estimator for the actual water demand under normal circumstances. By comparing the measured water demand to the forecasted water demand, anomalies like pipe bursts can be detected. Historic large pipe bursts were studied to understand the problems that pipe bursts cause to the water companies. Most bursts are not very problematic: the bursts cause only a small pressure drop in the water distribution system, and are reported and repaired shortly after they began. However, a small number of bursts are more problematic: this applies especially for smaller bursts that stay unnoticed for a long period and cause large water losses, or large bursts that begin in the night and are only noticed in the morning and cause damage to the urban environment.

A pipe burst detection method was developed that is based on a heuristic adaptive water demand forecasting model. The method also forecasts pressures in the system (by a data-driven model) and monitors the measured pressures as well. The monitoring threshold values that distinguish between normal forecasting inaccuracies and pipe bursts, are derived by an automatic procedure that evaluates the forecasting deviations in the year prior to the monitoring year. The threshold values equal the 5% exceedance probability of the forecasting deviations multiplied by C_{lim}. The C_{lim} factor is the tuning factor of the method and determines the trade-off between hit rates and false alarm rates. All measured and forecasted signals are transformed to moving average values over time frames of 2, 5, 10, 15, 30, 60, 120 and 240 minutes. The transformation to longer moving average time frames resulted in lower threshold values which enabled the detection of smaller pipe bursts.

The method was tested on different historic datasets with hydraulic data and pipe burst information in three areas in the western part of the Netherlands, and six areas in the Northern part of the Netherlands. When evaluating the method, a distinction was made between all pipe bursts and the large bursts. This was done because some of the areas were very large and many bursts in those areas caused no observable deviation in flow or pressure. The method proved to be ineffective for detecting all bursts, but effective for detecting the large bursts: 80-90% of the bursts could be detected within 20 minutes, while generating false alarms on 3% of the days without a burst. The data in the studied areas in combination with the proposed detection method was used to derive a relation between the size of the pipe burst that can be detected and the size of the area. Based on this relation and an analysis of the problematic bursts, it is recommended to apply the burst detection method to areas with an average demand of 150 m^3/h or less.

Samenvatting

Voorspelling water verbruik
Het totale waterverbruik in een gebied is gelijk aan de som van het waterverbruik van de individuele huishoudelijke en industriële consumenten in het gebied. Het gedrag van deze consumenten wordt gekenmerkt door dagelijkse, wekelijkse, en jaarlijkse herhalende patronen. Deze zich herhalende patronen zijn ook te zien in het drinkwaterverbruik. Op basis van deze waarnemingen, is een adaptief model voor het voorspellen van het waterverbruik voor de korte termijn ontwikkeld. Dit model vergaart automatisch verbruikspatronen en –factoren van alle dagen van de week en van een instelbaar aantal afwijkende dagtypen, zoals nationale feestdagen, vakantieperioden, en individuele afwijkende dagen. Het model gebruikt deze informatie bij het voorspellen van het waterverbruik voor de komende 48 uur met kwartierstappen (192 voorspelde waarden). Het model gebruikt als enige input het gemeten waterverbruik en kalender informatie waarin afwijkende dagen worden aangewezen. Het model heeft functionaliteit om extra waterverbruik tijdens de avondpiek op dagen met mooi weer te onderscheiden. Zodra dit extra verbruik geïdentificeerd is, dan wordt de voorspelling direct aangepast om het afwijkende verbruik zo goed mogelijk te voorspellen. Het model kan eenvoudig geïmplementeerd worden, is volledig adaptief en nauwkeurig, waardoor het geschikt is om toe te passen voor geavanceerde besturing en leidingbreuk detectie.

Het model is getest op datasets met zes jaar historisch waterverbruik in zes verschillende gebieden in centraal en zuid Nederland. Het klimaat in deze gebieden is ongeveer hetzelfde en de grootte van de gebieden varieerde van heel groot (950.000 inwoners) tot klein (2.400 inwoners). De gemiddelde absolute procentuele fout van de 24-uurs voorspelling lag tussen 1,44-5,12%, en van de kwartier-voorspelling tussen 3,35-10,44%. Uit een analyse van de voorspelfouten bleek dat deze sterk gerelateerd zijn aan het gemiddelde verbruik in het gebied: de voorspelfouten nemen lineair af met de log-waarde van het verbruik.

Waterverbruik is afhankelijk van de weersomstandigheden: hoge temperatuur en droogte leiden tot een toename van het waterverbruik. Op basis daarvan mag verwacht worden dat een model dat de watervraag voorspelt, beter zal presteren indien dit model weersinformatie gebruikt. Om deze verbetering te onderzoeken is het voorspellende model en twee andere modellen uitgebreid met temperatuurcorrectie. Hieruit bleek dat de voorspelfouten gemiddeld 6,3% kleiner waren, en de grootste voorspelfouten 9,4% kleiner. Dit laat zien dat een weercorrectie inderdaad leidt tot een nauwkeuriger voorspelling van de watervraag, en dat vooral de grootste onder- en overschattingen van het verbruik verkleind kunnen worden.

Geoptimaliseerde besturing

Een eerste toepassing van het vraagvoorspellingsmodel is geoptimaliseerde besturing. In conventioneel bestuurde drinkwatersystemen worden de productie en de reinwaterpompen nog gestuurd met een eenvoudige niveauregeling, hetgeen resulteert in grote fluctuaties van het productie- en pompdebiet. Deze suboptimale regeling kan verbeterd worden door het toepassen van een regeling gebaseerd op vraagvoorspelling. De korte termijn voorspelling van het verbruik biedt nauwkeurige informatie van de te leveren hoeveelheid drinkwater in de komende 48 uur. De buffers in het drinkwatersysteem bieden de mogelijkheid om de variatie in het verbruik af te vlaken, en om het produceren of verpompen van water te vervroegen of uit te stellen. Door in de regeling gebruik te maken van een vraagvoorspelling, kunnen de buffers van het drinkwatersysteem optimaal gebruikt worden.

De toepassing van besturing op basis van vraagvoorspelling in Nederland is behoorlijk ver doorgevoerd. Meer dan 50% van alle drinkwatersystem wordt op deze wijze gestuurd. Om de verschillen tussen conventionele besturing en geoptimaliseerde besturing te onderzoeken, is onderzoek gedaan bij vijf bestaande drinkwatersystemen in Nederland. Hierbij zijn de operationele resultaten vergeleken van een week met conventionele besturing met een week met geoptimaliseerde besturing. Uit dit onderzoek bleek dat de variatie in het zuiveringsdebiet 75% lager was met geoptimaliseerde besturing in vergelijking met conventionele besturing. Door het constantere debiet was de troebelheid in het reine water gemiddeld 17% lager. De geoptimaliseerde besturing had ook invloed op het energieverbruik. Geoptimaliseerde besturing leidde tot 3,1% minder energieverbruik, en een verschuiving van het verbruik in hoog tarief uren naar laag tarief uren van 7,4%. De totale energierekening was hierdoor gemiddeld 5,2% lager.

Tevens is onderzoek gedaan naar de toepassing van deze geoptimaliseerde besturing buiten Nederland. Hiervoor is de geoptimaliseerde besturing geïmplementeerd voor de besturing van een deel van het drinkwatersysteem van de Poolse stad Poznań. Hierbij werden zowel de productie als de distributiepompen (drukregeling) door de geoptimaliseerde besturing gestuurd. De resultaten met betrekking tot de productiesturing waren vergelijkbaar met de resultaten die bij de Nederlandse systemen werd waargenomen. Bij het Poolse systeem bleek vooral de distributiesturing zeer waardevol te zijn. Uit het onderzoek bleek dat de persdruk van de reinwaterpompen 29% lager was, en dat de gemiddelde druk in het leidingnet 18% lager was. Hierdoor nam het geschatte lekverlies ook met circa 18% af. Het gecombineerde effect van een constanter productiedebiet, een lagere distributiedruk, en een lager lekverlies resulteerde in een afname van het totale energieverbruik in het drinkwatersysteem van 11,5%.

Leidingbreuk detectie
Een tweede toepassing van het vraagvoorspellingsmodel is leidingbreuk detectie. De actueel voorspelde waarde is een goede schatting van het waterverbruik onder normale omstandigheden. Door het actueel gemeten verbruik te vergelijken met het actueel voorspelde verbruik kunnen anomalieën, zoals leidingbreuken, gedetecteerd worden. Historische grote leidingbreuken zijn onderzocht om na te gaan welke leidingbreuken problematisch zijn voor de waterbedrijven. Hieruit bleek dat de meeste niet erg problematisch zijn: de meeste breuken hebben weinig invloed op de druk, worden snel ontdekt en gemeld, en snel gerepareerd door het waterbedrijf. Slechts enkele breuken zijn problematischer: dit is in het bijzonder het geval voor kleinere breuken die pas na lange tijd (weken of maanden) ontdekt worden en waardoor veel water verloren gaat, of voor grotere breuken die ’s nachts ontstaan en pas ’s ochtends ontdekt worden en in de tussentijd veel schade aan de omgeving hebben toegebracht.

Een leidingbreuk detectie methode is ontwikkeld die is gebaseerd op een voorspelling van het waterverbruik. De detectiemethode voorspelt en monitort ook de drukken in het distributienet. De grenswaarden die het onderscheid maken tussen normale voorspelfouten en (mogelijke) leidingbreuken, worden vastgesteld op basis van een analyse van de voorspelfouten in het voorgaande jaar. De grenswaarde is gedefinieerd als de waarde van de 5% overschrijdingskans van de voorspelfouten vermenigvuldigd met C_{lim}. C_{lim} is de factor waarmee het functioneren van de detectiemethode getuned kan worden, en waarmee een evenwicht gevonden kan worden tussen trefkans en aantal valse alarmen. Alle gemeten en voorspelde waarden worden getransformeerd tot voortschrijdend gemiddelde waarden over verschillende tijdramen (2, 5, 10, 15, 30, 60, 120, 240 minuten). Langere tijdramen leidden tot lagere detectiegrenswaarden, zodat met deze transformatie ook kleinere leidingbreuken gedetecteerd konden worden.

Deze methode is getest op verschillende datasets met hydraulische data en leidingbreuk informatie van drie gebieden in het westen van Nederland, en zes gebieden in het noorden van Nederland. Bij het evalueren van de methode is onderscheid gemaakt tussen alle leidingbreuken, en de relatief grotere breuken. Dit is gedaan omdat sommige van de gebieden erg groot waren en veel breuken daardoor geen waarneembare invloed hadden op de gemeten debieten en drukken. De methode bleek ineffectief te zijn om alle breuken te detecteren, maar effectief te zijn voor het detecteren van de relatief grotere breuken: 80-90% van deze breuken kon binnen 20 minuten gedetecteerd worden, terwijl op gemiddeld 3% van de dagen zonder breuk een vals alarm gegenereerd werd. Bij de bestudeerde gebieden in combinatie met de voorgestelde methodiek, kon een relatie gelegd worden tussen de grootte van de breuk die ontdekt kon worden en de grootte van het gebied. Op basis van een analyse van de problematische breuken en bovengenoemde relatie, wordt aanbevolen om de leidingbreuk detectie methode toe te passen op gebieden met een gemiddeld verbruik van maximaal 150 m^3/h.

Table of Contents

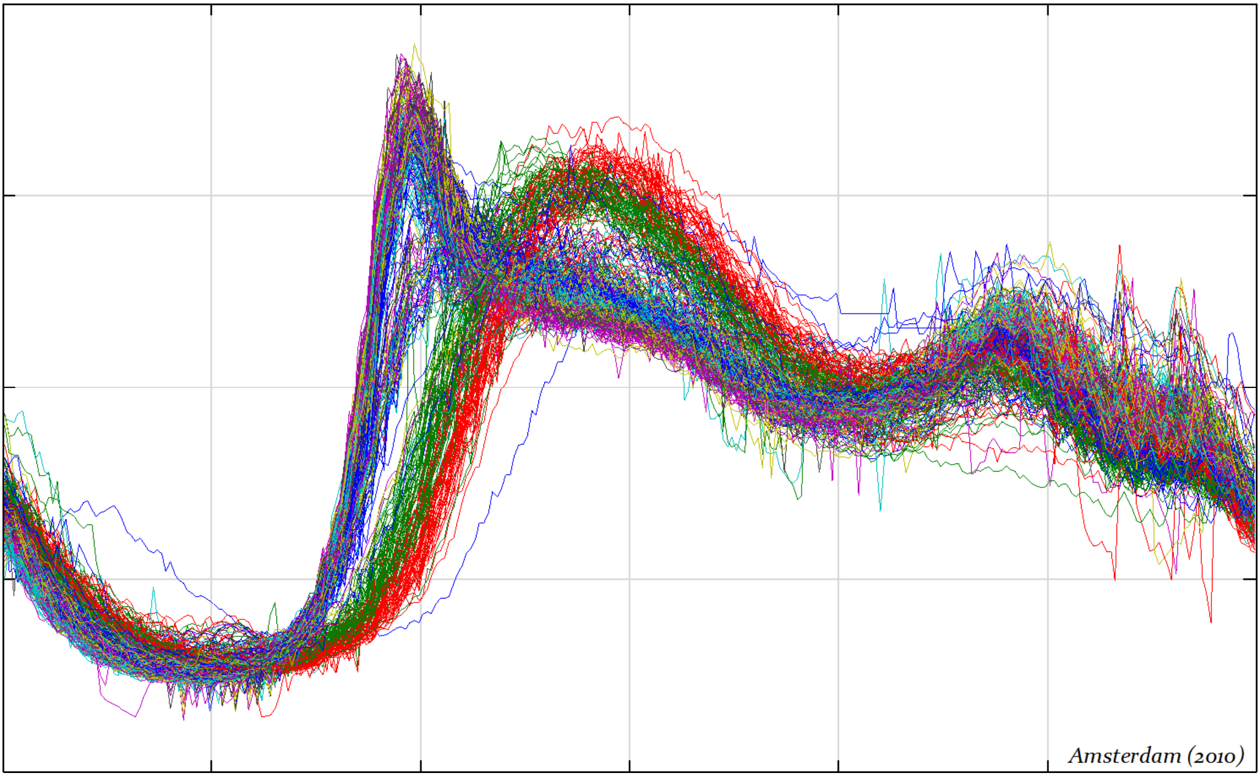


Table of Contents

Summary ... i

Samenvatting ... iv

Part I – Introduction ... 4

1 Introduction to optimised control, pipe burst detection and demand forecasting ... 5

Part II – Demand forecasting ... 18

2 A fully adaptive forecasting model for short-term drinking water demand ... 19

3 Improving the performance of water demand forecasting models by using weather input ... 49

Part III – Optimised control ... 64

4 Better water quality and higher energy efficiency by using model predictive flow control at water supply systems ... 65

5 Advanced control of a water supply system: a case study ... 87

Part IV – Pipe burst detection ... 104

6 Heuristic burst detection method using flow and pressure measurements ... 105

7 Analysis of historic bursts and burst detection in water supply areas of different size ... 131

Part V – Concluding remarks ... 146

8 Conclusions and recommendations ... 147

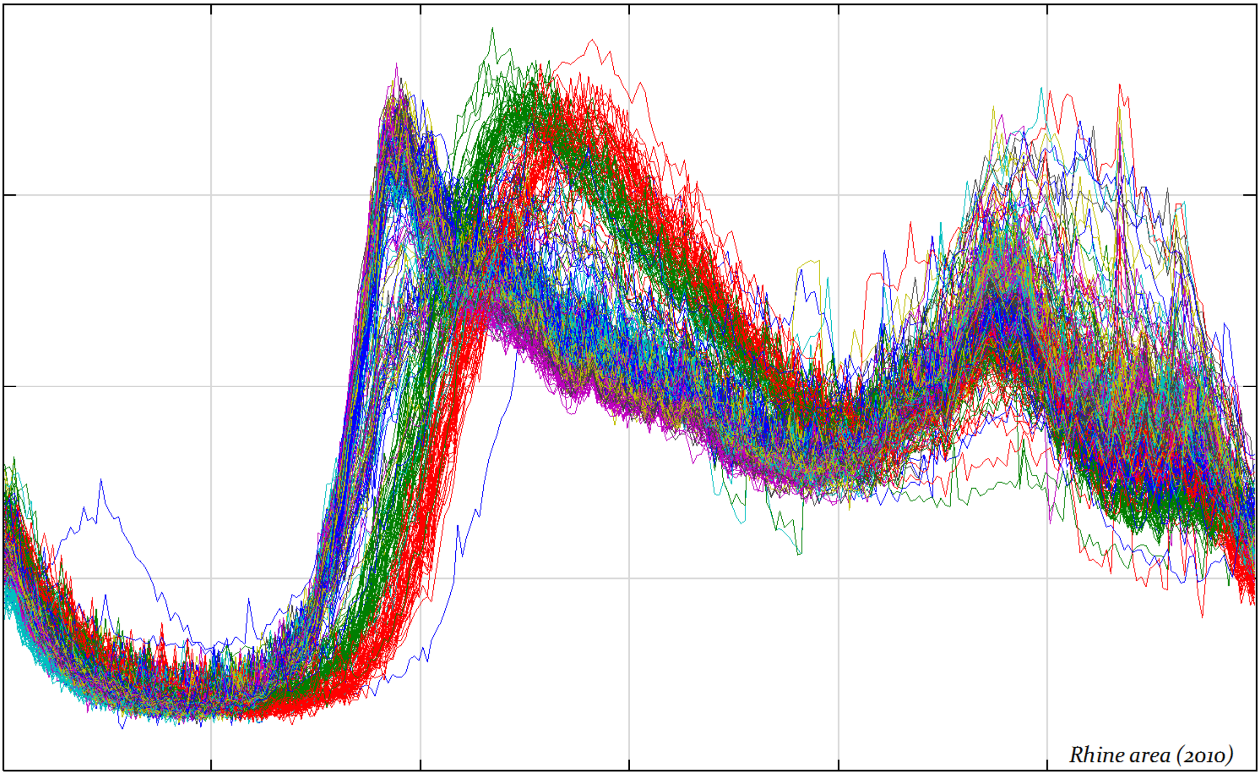
List of publications ... 151

List of abbreviations ... 154

Acknowledgements ... 156

Curriculum vitae ... 157

Part I – Introduction



1 Introduction to optimised control, pipe burst detection and demand forecasting

Based on

M. Bakker
K.M. van Schagen
J.L. Timmer

Flow control by prediction of water demand
Journal of Water Supply: Research and Technology – AQUA (2003). 52 (6): 417-424

And

M. Bakker
T. Lapikas
B.H. Tangena
J.H.G. Vreeburg

Monitoring water supply systems for anomaly detection and response
Proceedings New developments in IT & Water (2012), Amsterdam, the Netherlands

1.1 General

The automatic control and monitoring of the main elements of water supply systems is often based on a translation of empirical rules and 'historic' strategies into simple control loops (Cembrano et al., 2011). Because of computational limitations of the controllers and the wish of control engineers to limit the complexity of the control, many water supply systems are automated with simple and robust control loops. As a result, the operation of the systems is reactive and the operators are not fully in control of their systems. Examples of reactive operations are:

- The automatic controls of water supply systems react to real-time measured reservoir levels, pressures or flows, and only change the real-time operation when certain switching values are reached.
- Water supply companies are not directly aware of most pipe bursts, and only react when a burst is reported by consumers.

Operators become more in control of their systems when the operation is changed from reactive to pro-active. Pro-active operation means not only responding to the actual situation, but using short-term projections in the operation. The short-term projections provide information about the expected normal behaviour of the system, and this information can be used to control the system pro-actively. In this way unexpected situations and radical control actions as a result of these situations may be avoided. Moreover, anomalous behaviour of the system may be detected by comparing the actual behaviour of the system with projected behaviour. A second aspect of pro-active control is that it may improve the efficiency of the water supply systems. Efficiency improvements are important to the water sector, because the public expects that the companies do not only supply water continuously and safely, but also that the water supply systems are managed professionally and efficiently. In the Netherlands, the initiatives to increase the efficiency of the systems are intensified by benchmarks among water companies (Vewin, 2009).

1.2 Optimised flow control of water supply systems

Reservoirs in a water supply system are the key elements in the real-time flow control, because they provide the possibility to balance water demand and supply. The reservoirs are designed to balance the difference between water demand and (constant) production flow, and to guarantee the availability of sufficient water in case of equipment failure in the treatment plant or for fire-fighting (Twort et al., 2000). In most water supply systems in the Netherlands the total reservoir volume amounts approximately 25% of the peak day demand, which suffices to balance the fluctuations of normal domestic water demand. Conventional level based flow control methods are generally not capable of adequately balancing the variation in the water demand, and the pump and production flows vary considerably. This is caused by the fact that this conventional control method only responds to actual situations and does not take projections into account. We observed that it is not unusual that the conventional control method switches off the treatment plant completely during the night, and evokes production flows during the day that are significantly higher than the average demand.

A more pro-active way of controlling the flow in a water supply system is based on using forecasts of the water demand. These forecasts provide information about the amounts of water that need to be pumped (or flow by gravity) out of the reservoirs. Based on this information the required inflow into the reservoirs can be calculated. Note that this required inflow is not one single value, but an array (time series) of inflows with the same number of values as the array with forecasted outflows. The inflows must be such that all constraints (e.g. minimum and maximum allowed reservoirs levels) are met. Figure 1.1 shows an example of the user interface of a pro-active control method.

Figure 1.1 User interface of pro-active control method, named OPIR. The graph shows forecasted and actual measured data of the demand (green), the incoming flow (red), and the level in the reservoir (blue) on a 24-h time scale

In January 1996 a pro-active control method was implemented at the water treatment plant (WTP) Helden (12,000 m^3/day) in the Netherlands. The main driver for implementing the method was to improve the quality of the treated water. This improvement could be achieved by not switching on ground water wells with inferior water quality. This was made possible by running the treatment plant at a constant flow rate (Bakker et al., 1997). Therefore, the optimisation goal of the pro-active control method was defined as to maintain the production flow as constant as possible. Following the implementation at WTP Helden, the method was implemented at several other treatment plants, e.g. WTP Eindhoven. Figure 1.2 shows the change in production flow pattern (red line) when switching from conventional control to optimised control, as observed at WTP Eindhoven.

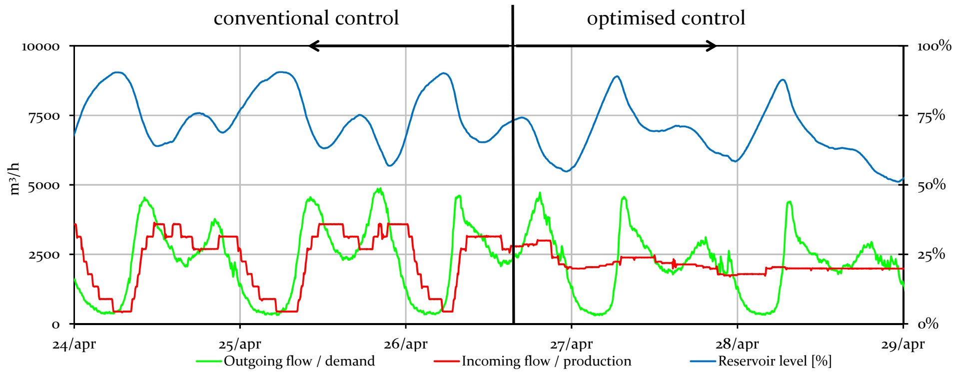


Figure 1.2 Switch to pro-active optimised control at WTP Eindhoven, the Netherlands

When being in control, the production and transportation flows in a water supply system can be stabilised. Stable flows have several advantages. For example, after the implementation of optimised control at WTP Reijerwaard, significant water quality improvements (hardness and turbidity) and a 50% decrease in the number of equipment failures (valves, filtration pumps, dosing pumps, blowers) were observed (Keuning et al., 1998). In addition, Bakker et al. (1998) showed that production flow changes lead to increased turbidity values, hence stabilizing production flow may lead to lower turbidity. Finally, the weighted average hydraulic head loss component of production and transportation flow is minimal at constant flows, so stabilizing flows should lead to a reduction of energy consumption in the water production and transportation. And more constant flows may also lead to relatively more energy consumption during low tariff periods, which may result in a lower energy bill.

1.3 Anomaly detection in water distribution systems

Water distribution systems consist of numerous elements exposed in public areas. This makes the systems inherently vulnerable for both unintended and intended damage caused by human actions (Perelman et al., 2012). Beside damage caused by human actions, the systems are vulnerable for spontaneous damage or water quality deterioration (Farley et al., 2010). Based on this high vulnerability of the water supply systems, high rates of failures of the system may be expected.

The number of outbreaks of waterborne diseases caused by a distribution deficiency is low. Smeets et al. (2009) and Van Lieverloo et al. (2007) reported that only three outbreaks occurred in the Netherlands since the end of World War II (65 year period). Although we cannot be sure that more (small scale) outbreaks have occurred, the low number indicates that outbreaks are quite rare. The number results in an average outbreak rate in the Netherlands of 0.0029 outbreaks /million people /year. In order to get a higher confidence in this number, a comparison was made to studies of outbreaks in other countries. Risebro et al. (2005) studied all outbreaks associated with drinking water in 10 countries in Western Europe (Finland, France, Germany, Greece, Italy, the Netherlands, Ireland, Spain, Sweden and the United Kingdom) in the period 1990 to 2004. The authors reported 19 outbreaks caused by a deficiency in the distribution system, resulting in an outbreak rate of 0.0036 outbreaks /million people /year. Craun et al. (2010) studied all outbreaks associated with drinking water in the United States in the period 1971 to 2006. The authors reported 49 outbreaks caused by distribution deficiencies in community water systems, resulting in 0.0046 outbreaks /million people /year. Assuming that the water distribution systems in The Netherlands, Western Europe and the United States are comparable, an outbreak rate between 0.05 and 0.07 outbreaks per year in the Netherlands may be expected (one event per 15 to 20 years).

The reported outbreaks in the Netherlands were caused by (coliform) bacteria (Smeets et al. (2009)), but both Risebro et al. (2005) and Craun et al. (2010) reported a wide range of contaminations, both microbiological and chemical. This diversity in possible contaminants puts researchers for a complex challenge, in finding the right (combination of) sensor(s) that is able to detect most of the contaminants. In addition, the number of sensors to install to protect people from water contamination is quite high. Cozzolino et al. (2011) presented a case study with a sensor density of 1 per 3,600 people. Such sensor density would mean that almost 5,000 sensors would need to be installed in the Netherlands, to detect an event with a probability of once per 15 to 20 years. It is likely that such number of sensors would evoke false alarms regularly. The probability of an outbreak due to water contamination is very low, and a sensor network for protecting people for contamination is likely to generate many false alarms and still cannot provide full protection against contamination.

The situation with respect to pipe bursts is quite the opposite. Pipe bursts are considered to be part of the normal operation of a water supply system, because of their daily occurrence. Trietsch and Vreeburg (2005) reported an average value of 0.07 failures per km of water main per year in the Netherlands. With a total length of all water mains of about 115,000 km, the number of pipe failures in the Netherlands amounts over 8,000 per year (22 per day). Most bursts have only limited effect on the water supply and cause an interruption of supply in a restricted area. However, every water company is regularly confronted with pipe burst events that have a larger impact on the water supply. Those events cause considerable damage to the surroundings and impose an interruption of supply to a larger number of consumers. Vreeburg and Boxall (2007) showed that 59% of all customer complaints at UK water companies are about interruption of supply or about low water pressures.

Figure 1.3 Example of a pipe burst event

The behaviour of most of the water supply companies with respect to pipe bursts is still reactive, despite the frequent occurrence of bursts. Most water companies rely on customers to report low pressures or water running in the street. A more pro-active behaviour of the water supply companies may lead to a reduction of the customer complaints and a reduction of the risk of a water contamination caused by a burst pipe. One of the first steps in more pro-active behaviour is using a pipe burst detection method. Actively monitoring the system shows this pro-active behaviour, although the water company still has to respond when a burst is detected by such system. A possible method for detecting pipe bursts is comparing the measured water demand in an area with the expected water demand. For generating an expected normal value of the water demand, a water demand forecasting model may be used.

1.4 Water demand forecasting

1.4.1 Time scales in water demand

Water demand in an area is the result of water consumption by individual people and industries in that area, reflecting their behaviour and habits. Water demands can be considered and forecasted on various time scales (House-Peters and Chang, 2011):

- Long-term: 5 to 20 years (unit: 1-1,000 million m^3 per year);
- Medium-term: 1 year (unit: 1,000-1000,000 m^3 per day);
- Short-term: 1 day to 1 week (unit: 10-100,000 m^3 per hour);
- Ultra short-term: real-time to 1 h (unit: 0.1-1,000 m^3 per second).

Long-term demand

Long-term demand forecasts are necessary for the planning and construction of new infrastructure. Long-term demand forecasts will identify the point in time that the forecasted demand will exceed the capacity of existing infrastructure. A number of years prior to this point in time, the design and construction of new infrastructure must be initiated (Frijns et al., 2013). Long-term demand forecasts are usually based on the planned construction of new living areas, or on sociological trends (e.g. the number of persons per household, the water use per person).

Medium-term demand

In the medium-term water demand, not only the total amount (m^3 per year) is considered, but also the distribution of the demand over the year and the forecasted daily peak demand (m^3 per day). The medium-term demand forecast is used to make operational plans for the existing infrastructure in the next year. In Figure 1.4, the daily demands over a one year period are shown for two areas. To compare the demands, the dimensionless demand factors (–) are shown rather than the absolute values (m^3 per day). The differences between a city (Amsterdam, 950,000 people) and a village (Helden, 39,000 people) are:

- The peak demand is higher in small areas.
- The day-to-day variation is larger in small areas.
- The weekly pattern is more distinct in small areas.
- Periods of high demand are not always at the same time.

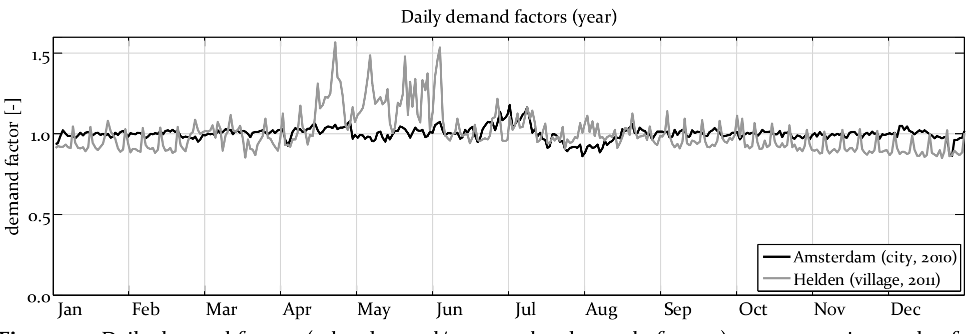


Figure 1.4 Daily demand factors (=day demand/average day demand of 1 year) on a 1-year time scale of a city (Amsterdam) and a village (Helden)

Short-term demand

On the short-term water demand scale, typically the daily demand patterns of hourly or 15-min. time steps are considered. In Figure 1.5 the daily demand patterns of one year (365 curves) are drawn on a 24-hours scale of both a city and a village. The demand patterns on Saturdays (green lines) and Sundays (red lines) are clearly different from the patterns on week days (other lines). The figure shows that the demand patterns on most days are quite similar. In particular, the variation in the demand in the night and early morning is low.

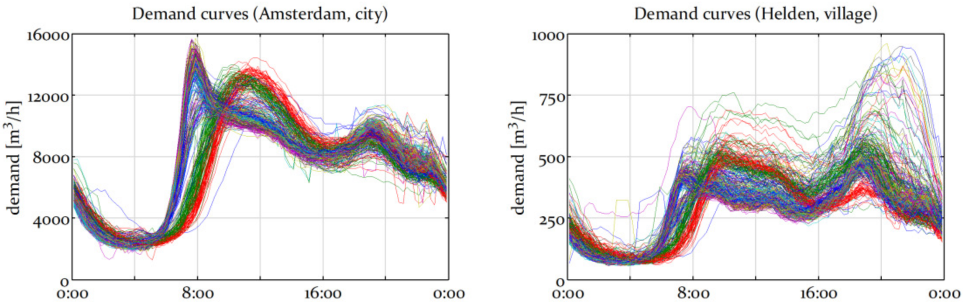


Figure 1.5 Daily demand curves of 1 year (365 curves) plotted on a 24-h scale of a city and a village

Ultra-short term demand

The water demand on the time scale of minutes to one hour is considered as the demand on the ultra-short term. When many consumers behave similarly as a group, extraordinary demand patterns can be observed on this time scale. Figure 1.6 shows the demand before, during and after an important soccer game during the world championship in 1998. Although the increases and decreases were very large, the average hourly demands before, during and after the game hardly differed from the average hourly demands under normal circumstances. This indicates that dramatic effects on the ultra-short term have little effect on the short-term demand.

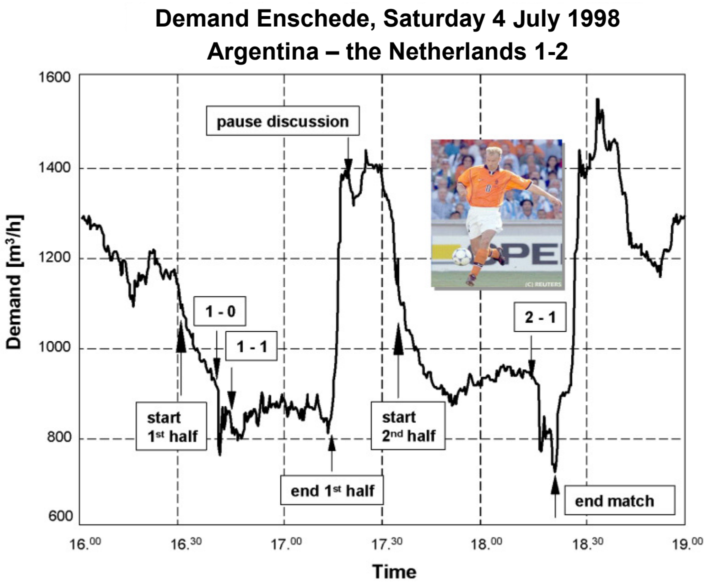


Figure 1.6 Demand during a soccer game at the world championships in 1998. During the game extreme demand variations occurred on the ultra-short time scale. For example, directly after the end of the first half, the demand increased from 800 to 1400 m^3/h within only 1-2 minutes

1.4.2 Short-term water demand forecasting

For optimised control of water supply systems a short-term water demand forecasting model may be used (Brdys and Ulanicki, 1994). A short-term water demand forecasting model generates forecasts for the next 1-2 days which is the same time frame as the optimisation horizon for the control of a water supply system. This horizon is restricted to 1-2 days, because the available clear water storage volume of most water supply systems is limited, which confines the time frame for which the control can be optimised. Forecasted hourly or 15 min. time step values of the demand need to be considered rather than day values, because the curve of the demand influences the level curve in the clear water reservoirs. As the level is a boundary condition that must be met continuously, one day time steps do no suffice.

For pipe burst detection, a forecast of the water demand for the actual point in time (a so-called now-cast) is necessary. By evaluating the deviation between the actual measured and actual forecasted values a pipe burst may be identified. A now-cast can be obtained from a short-term water demand forecasting model that forecasts the water demand with one hour time steps, or preferably 15 min. time steps. By interpolating between the hourly or 15 min. time steps forecasts, an actual forecasted value can be derived.

House-Peters and Chang (2011) and Donkor et al. (2014) presented overviews of available water demand forecasting models. Most of the available models are based on complex mathematical methods to generate the forecasts. A potential drawback of using (complex) mathematical methods is that such methods are hard to comprehend for the operational staff of water supply companies and are thus difficult to implement for day-to-day operation. However, it may be possible to translate a number of general observations about water demand into a comprehensible set of calculation rules to generate short-term water demand forecasts. Such heuristic forecasting model may be easier to understand and to implement and therefore be more acceptable to water supply operators.

Many of the currently available water demand forecasting models use weather information as input, because there is a correlation between the weather conditions and the water demand (Gato et al., 2007). A potential drawback of using weather information is that the complexity of the forecasting model increases, and that the costs of connecting real-time weather information to the model can be considerable. A possible motivation for not using weather input is that heuristic forecasting models may be sufficiently accurate for optimised control and pipe burst detection without using this input.

1.5 This thesis

1.5.1 Problem definition

The conventional operation of most water supply systems is reactive with respect to flow control and response to pipe bursts. This leads to inefficient operation and increased risk of high water losses, infrastructure damage and water quality deterioration after a pipe burst. Water demand forecasting models may be useful to transform the conventional reactive operation into more pro-active operation: Optimised control based on a water demand forecast may lead to operational efficiency improvements, like improved water quality and lower energy costs; Pipe burst may be detected at an early stage, by comparing real-time measured water demand with real-time forecasted water demand. A heuristic short-term water demand forecasting model may generate sufficiently accurate forecasts to be applied in optimised control models and in pipe burst detection models.

The research questions for this thesis are:

1. Can we develop a comprehensible, heuristic short-term water demand forecasting model, that is suitable for optimised control and pipe burst detection?
2. What is the influence of using weather data as additional input on the performance of water demand forecasting models?
3. What are the effects of optimised control of water supply systems with respect to water quality and operational costs compared to the current conventional control?
4. Can we detect pipe bursts by comparing real-time measured values to expected values generated by forecasting models?
5. On what scale should burst detection methods be applied to minimise water losses and damages caused by bursts?

1.5.2 Layout of this thesis

The heuristic water demand forecasting model, developed for optimised control and burst detection, is described in **Chapter 2** of this thesis. In this chapter the performance of the forecasting model is presented as well, when applied to forecast the demand in six different areas in a six years period. In **Chapter 3** the possible improvement in forecasting accuracy is presented when using weather input in the model. The performance improvement is not only considered when applied to the heuristic model, but also when applied to a Transfer/-noise and a Multiple Linear Regression (MLR) forecasting model.

Chapter 4 and **Chapter 5** describe the operational efficiency improvements that can be achieved when optimised control (based on a short-term demand forecast) is applied to water supply systems. In **Chapter 4**, differences in energy consumption and water quality at five full scale water supply systems in the Netherlands are presented. In **Chapter 5**, differences in energy consumption and leakage in the distribution network are presented for a Polish water supply system.

In **Chapter 6** the pipe burst detection method is described that is based on monitoring the deviation between expected (forecasted) and measured water demand. The chapter also presents the performance of the detection method when applied to historic datasets of water demand in three areas. In **Chapter 7** a thorough analysis of large historic pipe bursts in six different areas is presented. Based on this analysis, it could be determined which types of pipe bursts are problematic in the current operation, and where an early detection is critical to minimise water losses and damage to the urban environment. This chapter also presents the burst size that can be detected as a function of the size of the area.
Finally, in **Chapter 8** the general conclusions of this thesis are presented.

1.6 References

Bakker, M., Peters, E.G.J., and Verberne, A.J.P. (1997). "Productiebesturing van drinkwaterpompstations (in Dutch)". *H2O.* 30 (15): 466-469.

Bakker, M., Verberne, A.J.P., and van Schagen, K.M. (1998). "The benefits of demand forecasting and modelling". *Water Quality International.* (May-June): 20-22.

Brdys, M.A., and Ulanicki, B. (1994). "Operational control of water systems: Structures, algorithms and applications". United Kingdom: Prentice-Hall International.

Cembrano, G., Quevedo, J., Puig, V., Pérez, R., Figueras, J., Verdejo, J.M., Escaler, I., Ramón, G., Barnet, G., Rodríguez, P., and Casas, M. (2011). "PLIO: A generic tool for real-time operational predictive optimal control of water networks". *Water Science and Technology.* 64 (2): 448-459.

Cozzolino, L., Morte, R.D., Palumbo, A., and Pianese, D. (2011). "Stochastic approaches for sensors placement against intentional contaminations in water distribution systems". *Civil Engineering and Environmental Systems.* 28 (1): 75-98.

Craun, G.F., Brunkard, J.M., Yoder, J.S., Roberts, V.A., Carpenter, J., Wade, T., Calderon, R.L., Roberts, J.M., Beach, M.J., and Roy, S.L. (2010). "Causes of outbreaks associated with drinking water in the United States from 1971 to 2006". *Clinical Microbiology Reviews.* 23 (3): 507-528.

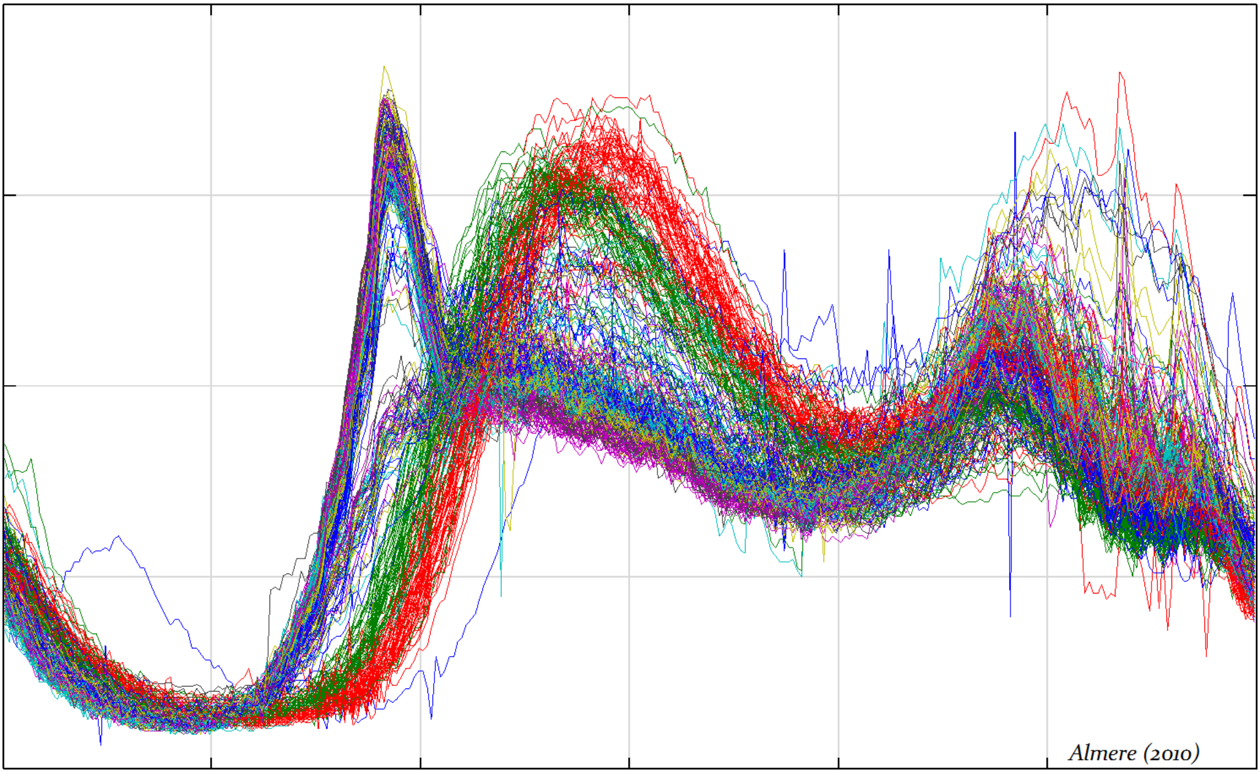
Donkor, E.A., Mazzuchi, T.A., Soyer, R., and Alan Roberson, J. (2014). "Urban water demand forecasting: Review of methods and models". *Journal of Water Resources Planning and Management.* 140 (2): 146-159.

Farley, B., Mounce, S.R., and Boxall, J.B. (2010). "Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network". *Urban Water Journal.* 7 (6): 345-356.

Frijns, J., Büscher, C., Segrave, A., and Van Der Zouwen, M. (2013). "Dealing with future

challenges: A social learning alliance in the Dutch water sector". *Water Policy.* 15 (2): 212-222.
Gato, S., Jayasuriya, N., and Roberts, P. (2007). "Temperature and rainfall thresholds for base use urban water demand modelling". *Journal of Hydrology.* 337 (3-4): 364-376.
House-Peters, L.A., and Chang, H. (2011). "Urban water demand modeling: Review of concepts, methods, and organizing principles". *Water Resources Research.* 47 (5).
Keuning, J., Mense, P., Van Schagen, K.M., and De Moel, P.J. (1998). "Hardheidsverlaging door middel van productiesturing (in Dutch)". *H2O.* 19: 73-75.
Perelman, L., Arad, J., Housh, M., and Ostfeld, A. (2012). "Event detection in water distribution systems from multivariate water quality time series". *Environmental Science and Technology.* 46 (15): 8212-8219.
Risebro, H., De Francia Doria, M., Yip, H., and Hunter, P.R. (2005). "Intestinal illness through drinking water in Europe. In: Report of the Microrisk Project to the European Commission (contract EVK1-CT-2002-00123)".
Smeets, P.W.M.H., Medema, G.J., and Van Dijk, J.C. (2009). "The Dutch secret: How to provide safe drinking water without chlorine in the Netherlands". *Drinking Water Engineering and Science.* 2 (1): 1-14.
Trietsch, E.A., and Vreeburg, J.H.G. (2005). "Reliability of valves and section isolation". *Water Science and Technology: Water Supply.* 5 (2): 47-51.
Twort, A.C., Ratnayaka, D.D., and Brandt, M.J. (2000). "Water supply (5th Edition)": IWA Publishing.
Van Lieverloo, J.H.M., Blokker, M.E.J., and Medema, G. (2007). "Quantitative microbial risk assessment of distributed drinking water using faecal indicator incidence and concentrations". *Journal of Water and Health.* 5 (SUPPL. 1): 131-149.
Vewin. (2009). "Reflections on performance 2009, Benchmarking in the Dutch drinking water industry".
Vreeburg, J.H.G., and Boxall, D.J.B. (2007). "Discolouration in potable water distribution systems: A review". *Water Research.* 41 (3): 519-529.

Part II – Demand forecasting



2 A fully adaptive forecasting model for short-term drinking water demand

Based on

M. Bakker
J.H.G. Vreeburg
K.M. van Schagen
L.C. Rietveld

A fully adaptive forecasting model for short-term drinking water demand
Environmental Modelling and Software (2013). 48 (1): 141-151

Abstract
For the optimal control of water supply systems and balancing demand and supply by reservoir level management, a short-term water demand forecast is necessary. In this chapter, a model that forecasts the water demand for the next 48 hours with 15-min. time steps is proposed and tested. The model uses measured water demands and static calendar data as only input. Based on this input, the model fully adaptively derives day factors and daily demand patterns for the seven days of the week, and for a configurable number of deviating day types. Although not using weather data as input, the model is able to identify occasional extra water demand in the evening during dry and sunny weather periods, and to adjust the forecast accordingly. The model was tested on datasets containing six years of water demand data in six different areas in the central and southern part of Netherlands. The areas have all the same moderate weather conditions, and vary in size from very large (950,000 inhabitants) to small (2,400 inhabitants). The mean absolute percentage error (*MAPE*) for the 24-hours forecasts varied between 1.44-5.12%, and for the 15-min. time step forecasts between 3.35-10.44%. The model is easy to implement, fully adaptive and accurate, which makes it suitable for application in real-time control.

Keywords
Demand forecasting; optimal control; short-term; water demand; water distribution

2.1 Introduction

The goal for a water supply company is to constantly supply water of good quality and under sufficient pressure. To achieve this, regular adjustments of pumps, valves and other controls of the water supply system are needed in order to balance supply and demand. The balancing of supply and demand is the normal daily operation of a water supply system. Initially the daily operation was done manually by operators, who intuitively made forecasts of the water demand. They made this forecasts based on their experience, taking information into account such as day of the week, hour of the day, water demand in previous days, weather, and special events like holidays.

Around the mid 1970's water utilities started automating their water supply systems by installing and operating telemetry and supervisory control and data acquisition (SCADA) systems (Bunn and Reynolds, 2009). The control loops of the first automated water supply systems were rather straightforward, resulting in inefficient operations with respect to energy consumption and costs, and fluctuations in the production flow. The operation of a system can be optimised by using forecasts in the control, which is effectively applied in different areas, like in the control of electricity grids (Manera and Marzullo, 2005), the control of open channels (Xu et al., 2013), and the control of the water quality reservoirs (Chen et al., 2012). Forecasts are also applied to increase the efficiency of the automatic control of water supply systems. Chapter 4 of this thesis (Bakker et al., 2013) will show that application of such optimal control software at water supply systems in the Netherlands, led to 3.1% reduction of energy consumption and 5.2% reduction of energy costs. Bunn and Reynolds (2009) reported 6%-9% reduction of energy consumption and 12% reduction of energy cost at water supply systems in the United States. Simulations with optimal control software showed that savings of 25% may be expected when applied at a real water supply system in Israel (Salomons et al., 2007), and savings of 17.6% when applied at a real system in Spain (Martínez et al., 2007).

All software applications for the optimal control of water supply systems contain a model that forecasts the water demand for the next 24 to 48 hours. This necessity for forecasting models has been one of the dominant reasons for researchers to develop such models. Both House-Peters and Chang (2011) and Donkor et al. (2014) present extensive overviews of water demand forecasting models. Although many researchers addressed water demand modelling, existing models may be improved with respect to adaptive functionality, forecasting time step and daily demand patterns, as will be explained in section 2.2.

We developed a model that forecasts the water demand for the next 48 hours with 15-min. time steps. The model we present in this thesis has enhanced functionalities: The model is fully adaptive, and can as a result be implemented and operated without manual (off-line) initial and interim data analysis for (re-)calibration; The model forecasts the water demand in

15-min. time steps for the next 48 hours; And the model discerns different demand patterns for the days of the week and a configurable number of deviating days (typically some 10 deviating day types are used). The forecasting model is an integral part of the advanced control software for water supply systems, named OPIR. This software is capable of generating optimised set-points for both production flow control of treatment plants, and detailed pump control in the water distribution network. The software was first implemented in 2006, and now forecasts and controls the water demand in 80 areas in the Netherlands, and in 20 areas in other countries (Belgium, Poland, Portugal and Canada). In these real water supply systems, the forecasting model has proved its effectiveness and its easy and reliable application.

In section 2.2, we describe some important issues in water demand forecasting, and where existing forecasting models may be improved. In section 2.3 we present the formulation of our model for forecasting the water demand, as well as for adaptively building up the database with the water demand characteristics. Section 2.4 presents the forecasting accuracy of the model when applied to six different datasets of water demand over a period of six years. In sections 2.5 and 2.6 we present the discussion and the conclusions of this chapter.

2.2 Water demand forecasting issues

2.2.1 Inputs water demand forecasting models

A number of one hour time step water demand forecasting models found in literature use measured water demand as only input (Jowitt and Chengchao, 1992; Shvartser et al., 1993; Homwongs et al., 1994; Alvisi et al., 2007). The papers show that it is possible to generate fairly accurate forecasts with measured water demand as only input. Other models use, beside measured water demand, weather information as input as well: The model of Zhou et al. (2002) uses the daily maximum temperature, the daily precipitation, the number of days since the last rainfall, and the pan evaporation; The Artificial Neural Network (ANN) model of Ghiassi et al. (2008) uses hourly values of temperature; the ANN model of Herrera et al. (2010) uses daily values of temperature, wind velocity, atmospheric pressure, and rain.

A drawback of using weather information is that the model needs an extra input. It is often difficult to make this extra input available in a production environment, because most automation networks are not connected to internet for ICT security reasons. And if weather input can be made available, it typically has a lower availability and reliability than measured water demand input, because it depends on a number of external systems. For this reason, the implementation of water demand forecasting models which use weather information is less easy and reliable, than the implementation of models that do not use weather information. So even though there is relation between the weather conditions and the water demand (as will be shown in chapter 3 of this thesis), there is an advantage from an implementation point of view not to use weather data as input.

2.2.2 Time scales

Water demand forecasting can be done at different time scales. The time scale for any water demand forecasting model is dictated by the purpose for which the model is to be used (Bakker et al., 2003). For the daily operation of treatment plants and pumping stations, a short-term forecasting model for the next 24-48 hours is needed. The output of the model can either be one day forecast for general production flow control of water treatment plants, or hourly forecasts for detailed distribution pump scheduling and operation of clear water reservoirs. Extensive research has been done to the forecast of the daily demand. To generate the daily demand forecast, various techniques can be used: Univariate time series models, which generate forecasts using observations as only input (Msiza and Nelwamondo, 2011); Time series regression models, which generate forecasts based on the relation between water demand and its determinants (Maidment and Miaou, 1986); Artificial neural network (ANN) models (Lertpalangsunti et al., 1999; Jain et al., 2001; Jentgen et al., 2007; Babel and Shinde, 2011); Composite or hybrid models in which two or more forecasting techniques are combined (Zhou et al., 2000; Aly and Wanakule, 2004; Gato et al., 2007; Alvisi et al., 2007; Bárdossy et al., 2009). In Ghiassi et al. (2008) and Adamowski et al. (2012) comparisons between several of the abovementioned techniques are presented. The forecast of water demand on an hourly basis has been studied by a smaller number of researchers. The applied techniques to generate hourly forecasts are identical to those to generate daily forecasts: Time series models (Jowitt and Chengchao, 1992; Shvartser et al., 1993; Homwongs et al., 1994), Time series regression models (Zhou et al., 2002) ANN models (Ghiassi et al., 2008; Herrera et al., 2010; Jentgen et al., 2007) and Composite models (Alvisi et al., 2007). Some forecasting models that are used for pipe burst detection, generate forecasts with smaller time steps (e.g. Eliades and Polycarpou (2012), Ye and Fenner (2011)).

We observed that a one hour time step that is used in most models, is too large to describe all the variations in water demand. We found that the dynamics in the water demand in the morning peak around 8:00 h, are not described properly with a one hour time step, see Figure 2.1. The 15-min. time step describes the water demand dynamics in more detail, which makes a 15-min. time step more suitable for application in water distribution control. Note that the more detailed 15-min. forecast has only added value when applied for detailed control, where the exact point in time to switch pumps is essential for the optimisation. Application of the smaller time step in a less critical time domain will be less valuable, and application in an area with a highly variable demand might result in less stable forecasts.

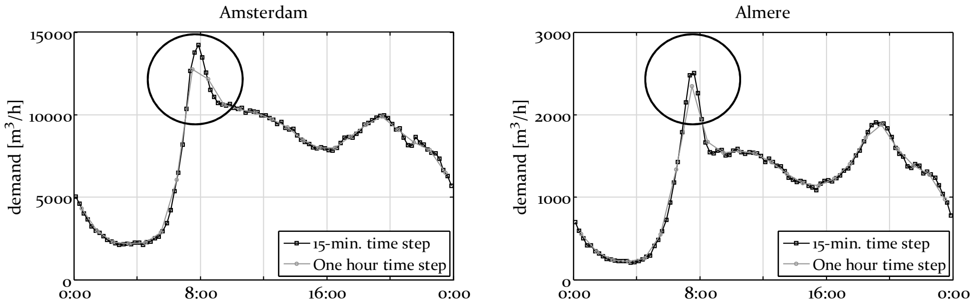


Figure 2.1 One hour versus 15-min. time step of water demand

2.2.3 Water demand patterns

Most water demand forecasting models found in literature use a limited number of demand patterns. The first models (Jowitt and Chengchao, 1992; Homwongs et al., 1994) use three different water demand patterns: one for weekdays, one for Saturdays and one for Sundays. The model described by Zhou et al. (2002) uses only two different patterns: one for weekdays and one for weekend days, including national holidays. A more recent model (Alvisi et al., 2007) uses demand patterns for each individual day of the week. In all four papers, it was observed that the patterns change with the seasons, and therefore the forecasting models use different patterns for each season. The models described by Ghiassi et al. (2008) and Herrera et al. (2010) use advanced mathematical modelling techniques (like Artificial Neural Networks), where the number of demand patterns is not explicitly discerned. To generate hourly forecasts, both models use information about the day of the week combined with weather information (Ghiassi et al. (2008) use hourly temperature readings at the end of each hour; Herrera et al. (2010) use daily values of temperature, wind velocity, rainfall and atmospheric pressure). These approaches imply that many different water demand patterns may be forecasted, depending on the combination of inputs in the mathematical models.

Observations of water demands in the Netherlands show that more deviating water demand patterns can be discerned. The first type of days with a deviating water demand pattern, as mentioned by Zhou et al. (2002), is national holidays (like Easter, Labour Day, Christmas, et cetera). On those days, the vast majority of the people and industries behave as on Sundays. The second type of days is the weekdays (Monday till Friday) in primary school holiday periods. In those periods, a substantial part of the population is not working and behaves differently than on normal weekdays. Typically the water demand peak in the morning is lower and smoothed out, compared to non-holiday water demand. For each holiday period (in the Netherlands: Summer holiday, Autumn holiday, Christmas holiday, Spring holiday, May holiday) a different pattern can be discerned, see the left graph in Figure 2.2.

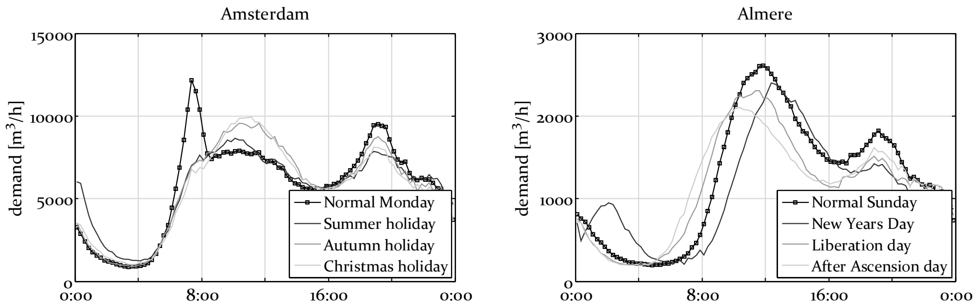


Figure 2.2 Deviating water demand patterns during primary school holiday periods (left graph) and during individual deviating days (right graph)

The third type of days is individual annual occurring days with a deviating demand pattern. Examples are, see the right graph in Figure 2.2: New Year's Day (very specific demand pattern), the day after Ascension Day (this is a Friday after a national holiday; many but not all people take an extra day off), Liberation Day (this is not a national holiday, but still many people take a day off to attend the festivities). The dates on which the three above mentioned types of deviating water demand patterns occur are known in advance. This information can be made available to a water demand forecasting model.

The fourth type of days with a deviating water demand pattern is related to the weather conditions, and is therefore not known in advance. On days with dry and sunny weather, the water demand resembles the normal pattern for the first part of the day, albeit somewhat higher than normal. In the (late) afternoon and evening the water demand is higher than the normal demand. This extra demand is presumably caused by people sprinkling their gardens. Figure 2.3 and Figure 2.7 (in paragraph 2.3.3) show examples of the water demand on a day with dry and sunny weather.

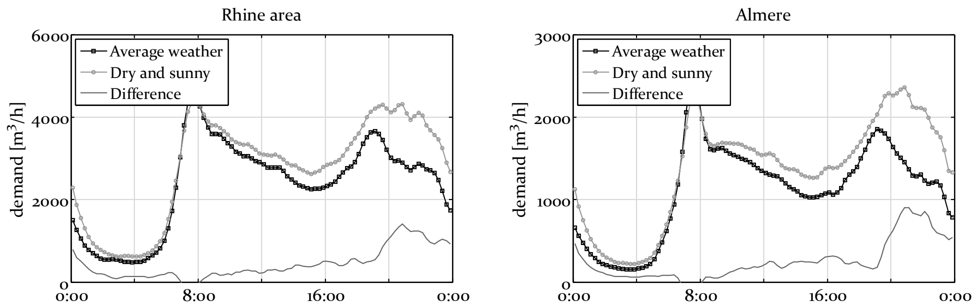


Figure 2.3 Deviating water demand pattern in two areas during dry and sunny weather. The graphs show a higher water demand during all day, but especially in the period between 18:00 and 0:00 h

2.2.4 Requirements for model (re-) calibration

All water demand forecasting models need a substantial initial dataset with historic water demands to calibrate the model. And when implemented in a practical situation, the models also need temporal re-calibration with new datasets, in order to keep up with the gradually changing water demand patterns. In the first models, this off-line data was used to derive the static seasonal demand curves and factors (Jowitt and Chengchao, 1992; Homwongs et al., 1994). The adaptive models described by Zhou et al. (2002) and Alvisi et al. (2007)) also need a dataset with historic demands to derive initial seasonal curves and factors. Data driven (ANN) models need data to train the mathematical model (Ghiassi et al., 2008; Homwongs et al., 1994), which is necessary for generating a forecast.

2.3 Materials and methods

2.3.1 Data and locations

We collected datasets of water demand in six different areas in the Netherlands in the period 2006-2011. We used the data we collected as input in simulations to assess the accuracy of the water demand forecasting model we developed. The weather conditions in the Netherlands in the whole country are more or less the same, and can be characterised as moderate with an average daily maximum temperature in summer (June to August) of around 19 ºC and in winter (December to February) of around 3 ºC. For each area, all water flows supplied to the area (from treatment plants, pumping stations and reservoirs) were summed in order to derive the net water demand in the area. Each number in the datasets represents the water consumption by all consumers in the area, plus all (occasional and planned) water losses in the area. Each dataset consisted of the water demand per 15-min. time step in m^3/h over a period of six years (210,336 values). The characteristics of the areas are shown in Table 2.1 and Figure 2.4.

Table 2.1 Characteristics of the six investigated areas (note the relatively low specific demand in the Almere area of 146 liter per capita per day (pcpd) which is caused by relatively low commercial demand in this area)

Area	Company	Average demand [m^3/day]	# consumers	Specific demand [liter pcpd]
1. Amsterdam	Waternet	179,800	950,000	189
2. Rhine area	Dunea	55,000	305,000	190
3. Almere	Vitens	28,200	193,000	146
4. Helden	WML	7,100	39,000	182
5. Valkenburg	WML	1,760	9,200	191
6. Hulsberg	WML	440	2,400	183

Figure 2.4 Location of the six investigated areas in the Netherlands

The water demands in the areas in a one year period are shown in the graphs at the beginning of the parts of the thesis (Amsterdam: page 2; Rhine area, page 4; Almere, page 18; Helden, page 64; Valkenburg, page 104; Hulsberg, page 146), and Figure 1.4 shows the daily demand factors in two areas. The graphs give an impression of the variability of the water demands in the different datasets.

2.3.2 Setup of the model

Considering the implementation of the model in real water supply systems, we focussed on three issues in the setup of the model: 1. easy and reliable implementation; 2. low maintenance costs, and; 3. high accuracy. For easy and reliable implementation of the model, we chose to use measured water demands and static calendar data as only input, and to use no other inputs like weather information. Although the relation between water demand and weather conditions seems evident, the model does not use weather input. Instead, the model uses a specific functionality to identify deviating weather related water demand, and to adjust the forecast accordingly (see section 2.3.3 of this thesis). To achieve low maintenance costs, we made the model fully adaptive. Fully adaptive means that the model automatically builds up a database with demand curves and patterns using its input (measured water demand), and while running, the model constantly renews these curves and factors (this process is described in section 2.3.5 of this thesis). In this way, the model automatically adapts to gradually changing water demand characteristics. This functionality makes manual (off-line) initial data analysis and temporal manual re-calibration unnecessary. It also enables to model to use one setup throughout the year, and not to use different setups for each season. For a high accuracy, the model forecasts the water demand with 15-min. time steps, using a relatively large number of demand patterns: The model discerns not only demand patterns for the seven days of the week, but also for a number of deviating day types like primary school holiday periods and individual deviating days (see section 2.3.5 of this thesis). National holidays are treated as Sundays. The flowchart of the model is shown in Figure 2.5.

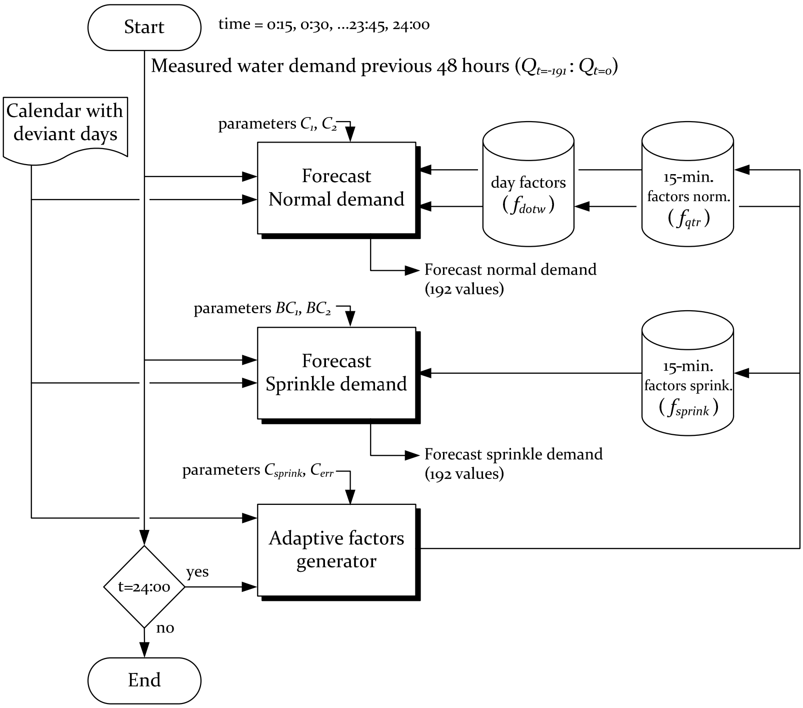


Figure 2.5 Flowchart of the water demand forecasting model

2.3.3 Description of the water demand forecasting model

The model forecasts the water demand for the next 48 hours with 15-min. time steps (192 forecasted values). Each time step, a new 48 hours forecast is calculated moving forward the array with forecasted water demands. The model forecasts the water demand in three main steps: in step one, the average water demand for the next 48 hours is forecasted; in step two, the normal water demands for the individual 15-min. time steps are forecasted; in step three, if applicable, extra sprinkle water demands for the individual 15-min. time steps are forecasted.

The forecast of the average water demand for the next 48 hours in step one is based on the measured water demand in the previous 48 hours. In order to correct for the influence of the day of the week, the measured water demand at 15-min. time step t (Q_t,) is divided by the typical day of the week factor for day type i ($f_{dotw,typ,i}$). The corrected water demand ($Q_{corr,t}$) is calculated with:

$$Q_{corr,t} = \frac{Q_t}{f_{dotw,typ,i}} \qquad [m^3 / h] \tag{2.1}$$

Section 2.3.5 will explain how $f_{dotw,typ,i}$ is adaptively derived from historical data. The forecasted average water demand ($Q_{forc,corr,avg}$) for the next 48 hours is based on the corrected water demands in the previous 48 hours (t=-95 to 0 and t=-191 to -96):

$$Q_{forc,corr,avg} = C_1 \cdot \left(\tfrac{1}{96} \sum_{t=-95}^{t=0} Q_{corr,t} \right) + C_2 \cdot \left(\tfrac{1}{96} \sum_{t=-191}^{t=-96} Q_{corr,t} \right) \qquad [m^3 / h] \tag{2.2}$$

The weighing constants C_1 and C_2 are by default set at 0.85 and 0.15 respectively, making the more recent measured water demands weigh heavier than the older demands. In section 2.5 of this thesis, a sensitivity analysis of these model parameters is presented. By using equation (2.2) the forecasted average water demand is based on a relative short period of measured demands (previous 48 hours, with emphasis on the previous 24 hours). This results in a rapid adjustment of the forecasted water demand, after a change of the measured water demand.

In step two, the normal water demands for the individual 15-min. time steps ($Q_{forc,norm,t}$) for the next 48 hours (t = 1 to 192) are calculated. This is done by multiplying the forecasted average demand by the typical day of the week factor ($f_{dotw,typ,i}$) and the typical 15-min. time step factor and ($f_{qtr,typ,i.j}$):

$$Q_{forc,norm,t} = Q_{forc,corr,avg} \cdot f_{dotw,typ,i} \cdot f_{qtr,typ,i,j} \qquad [m^3 / h] \tag{2.3}$$

Section 2.3.5 will explain how $f_{dotw,typ,i}$ and $f_{qtr,typ,i.j}$, are adaptively derived from historical data. Note that the model selects the proper factors ($f_{dotw,typ,i}$ and $f_{qtr,typ,i.j}$) that match with the period which is being forecasted. An example is shown in Figure 2.6.

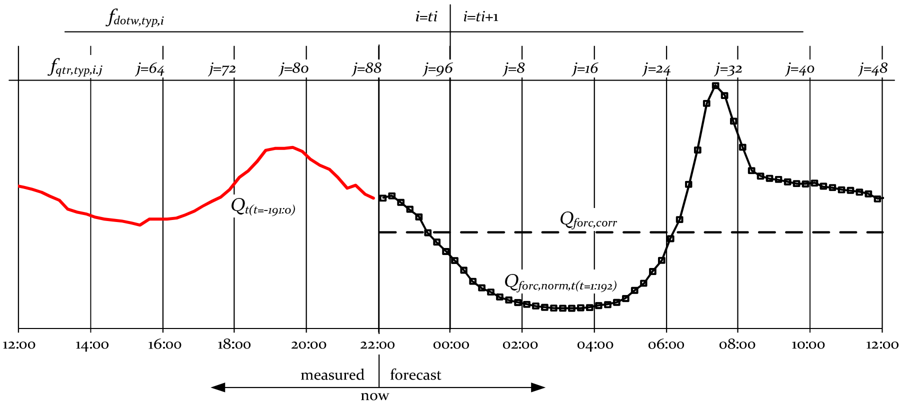


Figure 2.6 Selection of i and j in $f_{dotw,typ,i}$ and $f_{qtr,typ,i,j}$. Example of a forecast at 22:00 h: for the first 8 forecasted values $i=ti$, and j=89 to 96; for the next 96 forecasted values $i=ti+1$ and j=1 to 96; for the remaining 88 forecasted values $i=ti+2$ and j=1 to 88 (ti=day type today; $ti+1$=day type tomorrow; $ti+2$=day type day after tomorrow)

In step three, if applicable, extra sprinkle water demand is forecasted. We observed in water demand patterns in the Netherlands, that on a number of days the water demand in the late afternoon and evening (between 18:00 and 0:00 h) is much higher (Figure 2.7). In order to forecast this extra demand, the model identifies sprinkle demand in the measured demand. This identification is done by fitting the normal demand curve (consisting of 96 factors $f_{qtr,typ,i,j}$) on the measured water demand during the time frame between 0:00 and 18:00 h. This is the time frame where (in the Netherlands) no sprinkle water demand occurs. The fitting of the demand curve is done by multiplying the factors $f_{qtr,typ,i,j}$ with a value D, where:

$$\sum_{t=1}^{t=72} Q_t = D \cdot \sum_{j=1}^{j=72} f_{qtr,typ,i,j} \tag{2.4}$$

The D value is calculated at the end of 15-min. time step number 72 (at 18:00 h) and remains constant for the rest of the day. The model calculates the sprinkle demand at time t ($Q_{sprink,t}$) by taking the difference between the measured water demand at time t (Q_t) and the normal water demand according to the fitted demand curve ($D \cdot f_{qtr,typ,i,j}$):

$$Q_{sprink,t} = Q_t - \left(D \cdot f_{qtr,typ,i,j}\right) \qquad [m^3/h] \tag{2.5}$$

Figure 2.7 illustrates how the sprinkle demand is calculated.

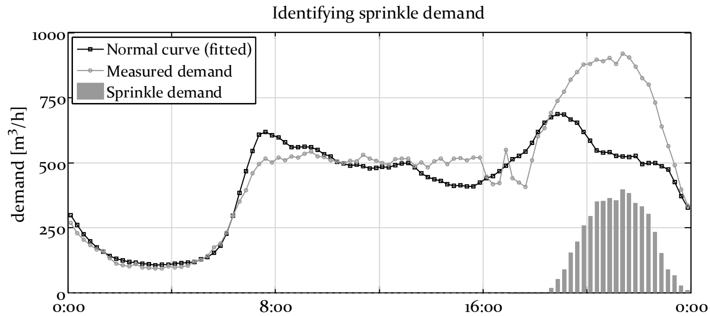


Figure 2.7 Identifying sprinkle demand by comparing measured demand with the normal expected demand according to the fitted normal demand curve

Sprinkle demand is only calculated in the time frame between 18:00 and 0:00 h (j = 73 to 96), at the end of each 15-min. time step j (j=73 at 18:15; j=74 at 18:30; to j=96 at 0:00 h). In the other time frames (j=1 to 72) the sprinkle demand is set at 0. The forecasted average sprinkle demand ($Q_{forc,sprink,avg}$) for the next 48 hours is based on the average of the observed sprinkle demands in the previous 48 hours ($Q_{sprink,t=-95\ to\ 0,}$ and $Q_{sprink,t=-191\ to\ -96}$):

$$Q_{sprink,forc,avg} = BC_1 \cdot \left(\tfrac{1}{24} \sum_{t=-95}^{t=0} Q_{sprink,t} \right) + BC_2 \cdot \left(\tfrac{1}{24} \sum_{t=-191}^{t=-96} Q_{sprink,t} \right) \ [m^3 / h] \qquad (2.6)$$

Note that by using equation (2.6) the forecast of the sprinkle demand does not change between 0:00 and 18:00 h ($Q_{sprink,t=-96\ to\ -1}$ and $Q_{sprink,t=-192\ to\ -97}$ do not change). The water demand between 0:00 and 18:00 h generally has no or very little predictive value for the sprinkle demand (see also Figure 2.7), and therefore the forecast of the sprinkle demand is not affected by the observations of the water demand in this time frame. The weighing constants BC_1 and BC_2 are by default set at 1.10 and 0.10 making the more recent observed sprinkle demands weigh heavier than the older demands. In section 2.5 of this thesis, a sensitivity analysis of these model parameters is presented. By using equation (2.6) the forecasted sprinkle water demand is rapidly adjusted once sprinkle water demand has been identified. Moreover, equation (2.6) implies that the forecasted sprinkle demand will be higher than the observed sprinkle demand: the sum of BC_1 and BC_2 = 1.20, which means that the forecast will be 20% higher than the observation. This is necessary because sprinkle water demand can increase rather quickly, quicker than the change in the normal water demand. In this way, the model performs well in a period with increasing sprinkle demand. However, in a period with constant or (sudden) decreasing sprinkle demand, the model will overestimate the sprinkle demand. We chose to design the model in this way, because an overestimate of the sprinkle demand is preferred over an underestimate when using the forecast for control.

Figure 2.8 shows an example of the observed and forecasted sprinkle demand per day as a percentage of the normal demand in a summer period. The figure shows that the sprinkle demand does not change from zero to its maximum value (around 12%) in one day, but that this takes some 5 to 7 days to rise to this maximum value. Although not using weather information as input, the model is still fairly able to forecast the sprinkle demand. The model is not able to forecast sudden changes in the sprinkle demand.

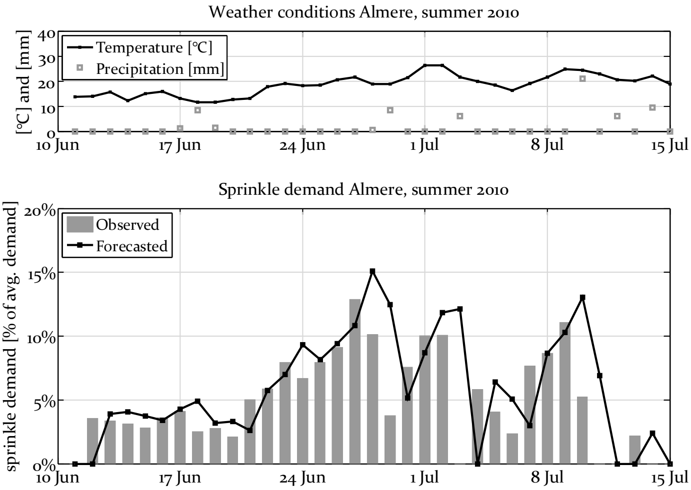


Figure 2.8 Weather conditions (average temperature and precipitation), and observed and forecasted sprinkle demand in the summer of 2010 in the Almere area. Note that the sprinkle demand is influenced by the weather conditions. The sprinkle demand it is underestimated when it (largely) increases (11, 21, 27, 4, 7 July) and overestimated when it (largely) decreases (18, 28, 29 June, 3, 10, 11 July)

Like the forecast of the normal demand, the forecasted average sprinkle demand must be transferred to the individual 15-min. time steps ($Q_{forcs,prink,t}$) for the next 48 hours. This is done by multiplying the forecasted average sprinkle demand ($Q_{forc,sprink,avg}$), by the typical 15-min. time step sprinkle factor ($f_{sprink,typ,i,j}$):

$$Q_{forc,sprink,t} = Q_{forc,sprink,avg} \cdot f_{sprink,typ,i,j} \qquad [m^3/h] \tag{2.7}$$

Section 2.3.5 will explain how $f_{sprink,typ,i,j}$ is adaptively derived from historical data. Finally, the model derives the total water demand forecast ($Q_{forc,tot,t}$) by summing the forecasts of the normal and the sprinkle demand:

$$Q_{forc,tot,t} = Q_{forc,norm,t} + Q_{forc,sprink,t} \qquad [m^3/h] \tag{2.8}$$

2.3.4 Day types

The model discerns the following different day types: seven types for the normal Monday till Sunday (including national holidays which are forecasted as if they were Sundays); five types for configured primary school holiday periods (Summer break, Autumn break, Christmas break, Spring break and May break); and four types for individual deviating days (New Year's Day, Good Friday, day after Ascension Day, and Liberation Day). For each day type, a day of the week factor ($f_{dotw,typ,i}$), and two arrays containing 96 factors for the normal demand pattern ($f_{qtr,typ,i,1\text{-}96}$) and for the sprinkle demand pattern ($f_{sprink,typ,i.1\text{-}96}$) are available. The dates when the non-normal day types occur, need to be selected in the calendar menu of the model. This enables the model to determine which day types have occurred in previous 48 hours and which will occur in the next 48 hours, in order to select the proper factors. The model uses the factors to normalise the measured demand (equation (2.1)), to forecast the normal water demand (equation (2.3)), to identify sprinkle demand (equations (2.4) and (2.5)), and to forecast the sprinkle demand (equation (2.7)). In contrast to the models of Zhou et al. (2002) and Alvisi et al. (2007), our model does not explicitly discerns different demand patterns for each season. However, the adaptive functionality of our model will result in different seasonal patterns, as will be explained in the next section.

2.3.5 Adaptive factors and curves

The model uses different factors in transforming measured water demand in a forecast (see equations (2.1), (2.3), (2.4), (2.5), and (2.7)). Each water supply area has its own characteristic water demand, and therefore these factors are unique for each area. The factors are adaptively learned by the model based on the measured demand. In this way, the model re-calibrates continuously and automatically all the factors, using recent observations of water demand for the day type. Initially the model starts with default values for all factors, and once the model is running, it starts updating the factors. Each day at 0:00 h the model stores the water demand information of the previous day. Along with the type of day, the average water demand ($Q_{avg,i}$) [m^3/h], as well as an array of 96 dimensionless 15-min. time step demand factors ($f_{qtr,i,j}$ or $f_{sprink,i,j}$) are stored. The model stores either the factors for the normal demand or the factors for the sprinkle demand, depending on whether a substantial amount of sprinkle demand was identified. The model uses the following filter to store the factors for normal demand:

$$\frac{\sum_{t=-95}^{t=0} Q_{sprink,t}}{\sum_{t=-95}^{t=0} Q_t} < C_{sprink} \tag{2.9}$$

where C_{sprink} is the discriminating factor, with a default value of 0.02. This means that if the identified sprinkle demand is less than 2% of the total measured demand, the normal demand factors will be stored; otherwise, the sprinkle factors will be stored. In this way, the normal pattern will not be disturbed by sprinkle demand, and vice versa. The dimensionless factors for the normal demand ($f_{qtr,i,j}$) or for the sprinkle demand ($f_{sprink,i,j}$) are calculated for each t by:

$$for\ each\ t\ (t = -95 \text{ to } 0;\ j = 1 \text{ to } 96)$$

$$f_{qtr,i,j} = \frac{Q_t}{\frac{1}{96}\sum_{t=-95}^{t=0} Q_t} \qquad [-] \tag{2.10}$$

$$f_{sprink,i,j} = \frac{Q_{sprink,t}}{\frac{1}{24}\sum_{t=-24}^{t=0} Q_{sprink,t}} \qquad [-] \tag{2.11}$$

After calculating and storing the new array with dimensionless factors, the model validates this array. This validation is necessary to avoid that an error in the measured demand or a very deviating demand, disturbs the typical factors which are used in the forecast. An array of demand factors of a day is disapproved if any of the factors differs more than the maximum allowed error:

$$Max\left(\left|f_{i,j} - f_{typ,i,j}\right|\right)_{j=1\ to\ 96} > C_{err} \tag{2.12}$$

where C_{err} is the validation factor with a default value of 0.50. This value provides a balance between an allowable normal variation in a demand curves, and a disturbing large deviation. The model uses the stored data to derive the typical factors that are used in the forecast. The array of 15-min. time step factors for the normal demand ($f_{qtr,typ,i,j}$) and for the sprinkle demand ($f_{sprink,typ,i,j}$) for day type *ti* are calculated by using equation (2.13) and (2.14). In both formulas *n* is the number of previous stored arrays of factors, that are used to calculate the typical factors:

$$f_{qtr,typ,ti,j} = \frac{1}{n}\sum_{i=1}^{i=n} f_{qtr,\{typ=ti\},j} \qquad [-] \tag{2.13}$$

$$f_{sprink,typ,ti,j} = \frac{1}{n}\sum_{i=1}^{i=n} f_{sprink,\{typ=ti\},j} \qquad [-] \tag{2.14}$$

The model selects only the approved arrays of curves when applying equation (2.13) and (2.14). When less than *n* (approved) arrays of factors are available, which is typically the situation if the model has just been implemented with only default factors, *n* is temporarily set to the number of available arrays. With this functionality, the model can forecast the water demand using proper arrays of normal demand factors, already one week after implementation.

The typical day of the week factors ($f_{dotw,typ,i}$) for each day type ti are calculated by using equation (2.15) over a time frame of m previous observations of that type of water demand.

$$f_{dotw,typ,ti} = \frac{\frac{1}{m}\sum_{i=1}^{i=m} Q_{avg,\{typ=ti\},i}}{\frac{1}{m\cdot 7}\sum_{i=1}^{i=m\cdot 7} Q_{avg,all,i}} \qquad [-] \tag{2.15}$$

$$f_{dotw,typ,ti} = \frac{\frac{1}{m}\sum_{i=1}^{i=m} Q_{avg,\{typ=ti\},i}}{\frac{1}{m\cdot 7}\sum_{i=1}^{i=m\cdot 7} Q_{avg,all,i}} \qquad [-] \tag{2.16}$$

The number of previous observations n and m from which the model calculates the demand factors is 5 and 10 respectively by default. In section 2.5 of this thesis, a sensitivity analysis of these model parameters is presented. By default, the number is chosen small enough to rapidly adjust the factors during the seasons, but large enough to average out random variations in the water demand. By continuously updating the factors, it is not necessary to discern different patterns for the seasons. Moreover, the functionality to forecast the sprinkle demand, results in a very quick adjustment of the forecasted demand pattern when applicable in the summer. With this functionality, the model will forecast the water demand in the summer more accurately, than when using a (static) summer demand pattern. A (static) summer demand pattern will be an averaged pattern of both days with and without substantial sprinkle demand. With such average demand pattern, the model will not be able to describe the dynamics in the water demand in summer properly.

2.3.6 Simulations with the water demand forecasting model

The datasets with water demands in the years 2006-2011 in six different areas were used as input in simulations to test the water demand forecasting model. The first year (2006) of the datasets was used by the forecasting model to adapt the typical demand factors to the demand characteristics of the simulated area; the other years (2007-2011) were used to evaluate the accuracy of the model. When evaluating the accuracy, the forecasted values were compared to the measured values. In Bennett et al. (2013) a wide range of measures is suggested to characterise the performance of environmental models. We selected four of those measures (generating dimensionless outputs), which are also commonly used by other researchers addressing water demand forecasting (e.g. Alvisi et al. (2007) and Adamowski et al. (2012)): The Relative Error for each individual value (RE_i); The Mean Absolute Percentage Error (*MAPE*, noted as Relative Volume Error *RVE* in Bennett et al. (2013)); The Relative Root Mean Square Error (*RRMSE*); The Nash-Sutcliffe Model Efficiency (R^2, noted as *NSE* in Bennett et al. (2013)):

$$RE_i = \frac{\left(\hat{y}_i - y_i\right)}{\overline{y}} \cdot 100\% \tag{2.17}$$

$$MAPE = \frac{\frac{1}{n}\sum_{i=1}^{i=n}\left|y_i - \hat{y}_i\right|}{\overline{y}} \cdot 100\% \tag{2.18}$$

$$RRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{i=n}\left(y_i - \hat{y}_i\right)^2}}{\overline{y}} \cdot 100\% \tag{2.19}$$

$$R^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{i=n}\left(y_i - \hat{y}_i\right)^2}{\frac{1}{n}\sum_{i=1}^{i=n}\left(y_i - \overline{y}_i\right)^2} \cdot 100\% \tag{2.20}$$

where y_i is the measured value, $\hat{y}_i$ is the forecasted value, and $\bar{y}$ is the mean of the measured values. The accuracy of the forecasts of both the water demand per 24-hours and per 15-min. time step were investigated. For the 24-hours evaluation, the 15-min. time step values of each 24 hours period (96 values) were transformed to 24-hours averages. For the 15-min. time step evaluation, each measured value is compared to 96 forecasted values corresponding to the forecasts of the first 24-hours period (note that the model produces a forecast for the next 48 hours, where each demand is forecasted 192 times, of which only the first 96 values are considered in the evaluation).

The simulations were executed on a 64-bit Windows7 HP EliteBook 8470p, with an Intel® Core™ i5-320M 2.60 GHz processor and 8 GB installed memory. The model was built in a Matlab R2012b simulation environment. The simulation time to generate the forecasts for one area for six years amounted 43.1 seconds.

2.4 Results

2.4.1 Overall accuracy of the model

Table 2.2 shows the *MAPE*, the *RRMSE*, the R^2, and the 0.5%, 25%, 75% en 95.5% confidence intervals of *RE* for the 24-hours forecast, and Table 2.3 shows the values for the 15-min. time step forecasts.

Table 2.2 Performance of the model per 24-hours in the period 2007-2011

Area	$\bar{y}$	***MAPE***	***RRMSE***	R^2	**Confidence intervals *RE*** [%]			
	[m^3/h]	[%]	[%]	[-]	**0.5%**	**25%**	**75%**	**95.5%**
1. Amsterdam	7,540	1.44	2.01	0.785	-6.5	-0.9	1.2	7.2
2. Rhine area	2,295	1.86	2.78	0.710	-8.2	-1.0	1.4	11.2
3. Almere	1,160	2.12	3.12	0.740	-9.7	-1.3	1.6	12.7
4. Helden	291	3.40	5.17	0.803	-16.8	-1.8	2.4	21.7
5. Valkenburg	73	3.49	4.83	0.802	-15.0	-2.3	2.8	17.5
6. Hulsberg	18	5.12	8.21	0.658	-26.5	-2.9	3.6	29.0

Table 2.3 Performance of the model per 15-min. time step in the period 2007-2011

Area	$\bar{y}$	***MAPE***	***RRMSE***	R^2	**Confidence intervals *RE*** [%]			
	[m^3/h]	[%]	[%]	[-]	**0.5%**	**25%**	**75%**	**95.5%**
1. Amsterdam	7,540	3.35	4.85	0.987	-16.4	-2.1	2.5	16.8
2. Rhine area	2,295	4.64	7.22	0.978	-24.1	-2.8	3.2	28.9
3. Almere	1,160	5.28	8.68	0.972	-29.9	-2.9	3.4	35.6
4. Helden	291	6.55	10.32	0.952	-35.8	-3.8	4.5	40.2
5. Valkenburg	73	6.90	10.00	0.949	-33.0	-4.5	5.3	31.9
6. Hulsberg	18	10.44	16.71	0.905	-57.9	-6.0	7.5	53.1

For the 24-hours forecast, the values of the *MAPE*, the *RRMSE*, and the R^2 varied between 1.44%-5.12%, 2.01-8.21, and 0.658-0.803 respectively. The errors expressed as *MAPE* and *RRMSE* were smallest in the largest areas and increased where the size of the areas decreased. However, the R^2 values were smallest for the relative smaller areas Helden and Valkenburg. For the 15-min. time step forecast, the values of the *MAPE*, the *RRMSE*, and the R^2 varied between 3.35%-10.44%, 4.85-16.71, and 0.905-0.987 respectively. The errors expressed as *MAPE* and *RRMSE* were smallest in the largest areas and increased where the size of the areas decreased. The *MAPE* and *RRMSE* of the 15-min. time step were on average a factor 2.2 larger than the *MAPE* of 24-hours forecast. For both the 24-hours forecast and the 15-min. time step forecast, the largest forecast errors which occur during 1% of the time (the 0.5% and 95.5% intervals of *RE*) were on average a factor 7.7 larger than the errors which occur during 50% of the time (the 25% and 75% intervals of *RE*).

Figure 2.9 shows the *MAPE* for the individual years. The figure shows the accuracy in both the years that were evaluated (2007-2011), and the initial year (2006) that was used to adapt the typical demand factors of the area. The graph shows that the *MAPE* varied in the years. This indicates that the unexplained variability of the water demand is not constant in all years. A possible explanation is that the variability in water demand is related to the weather conditions which vary among years. In order to test this assumption, additional analysis should be done to relate the forecasting errors to available weather data. Figure 2.9 does not show a gradually increasing or decreasing *MAPE* over the years. This indicates that (after adapting the typical factors in the initial year) the *MAPE* was rather constant, and the accuracy is not improving further after the first year. This can be explained by the fact that the model derives the typical factors from the last five measured demand patterns per day type (for the normal days this implies a “memory” of five weeks). It also indicates that the adaptive functionality was able to keep up with changes in the water demand.

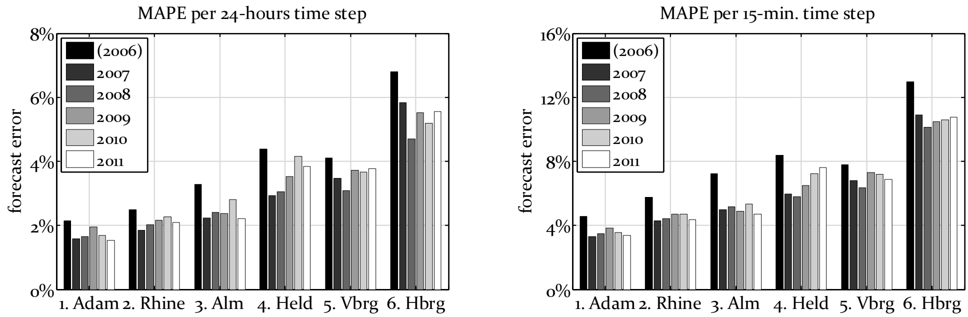


Figure 2.9 MAPE for the 24-hours forecast and the 15-min. time step forecast

Figure 2.9 also shows that on average the *MAPE* in the initial year was 20% higher than in the other years. The reason for this is that the model needs to observe each day type (at least) one time, before it is able to generate an accurate forecast. We observed that after approximately 30 days the *MAPE* for the 15-min. time step in the first year had the same magnitude as in subsequent years, and each time a day type occurred for the first time, we observed a peak in the error (see Figure 2.10).

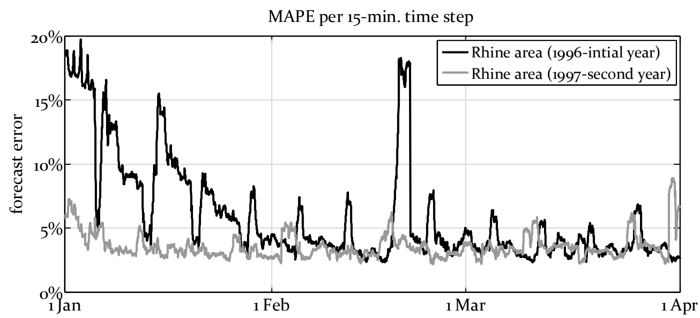


Figure 2.10 Trend of the MAPE for the 15-min. time step forecast in the initial year and a second year. Note that in the initial year on day nr. 47 (20 February 2006) a new day type (Spring break) for the first time occurred. This explains the peak in the error trend

2.4.2 Accuracy in relation to the water demand in the area

The forecast errors showed a relation with the average demand in the area (see Table 2.2, Table 2.3 and Figure 2.9): The percentage errors (*MAPE*) were smaller when the average demand in the area was higher. The relation between the water demand and the forecast error was further investigated by plotting the relation between the demand and mean absolute errors (*MAE*) in a log log diagram, and the relation between the demand and the percentage errors (*MAPE*) in logX-Y diagram (Figure 2.11).

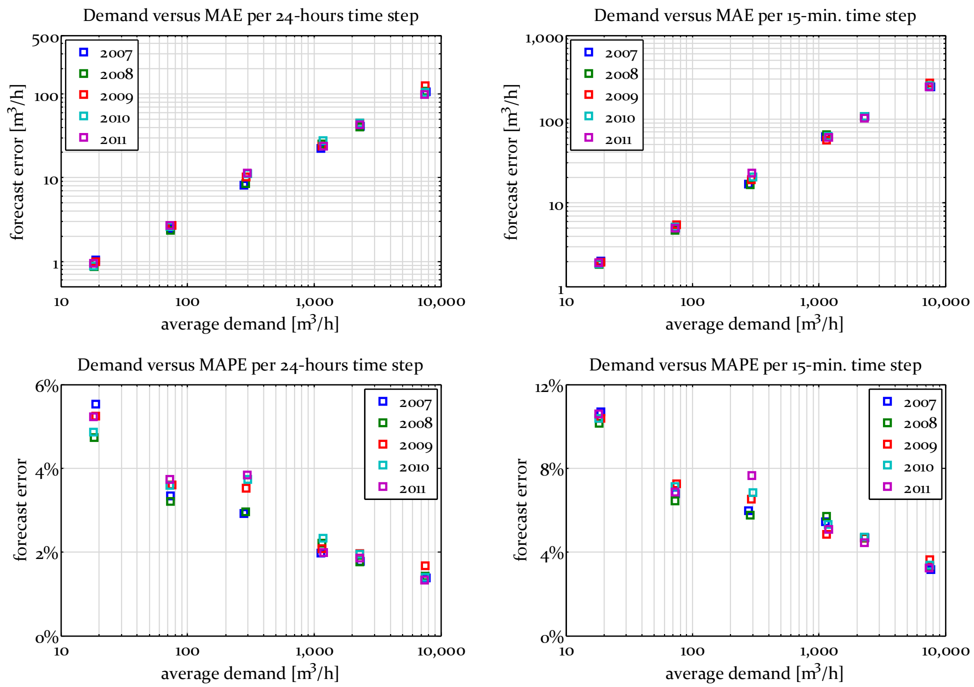


Figure 2.11 Relation between the water demand in the area and the *MAE* (upper graphs), and the MAPE (lower graphs). Left graphs who 24-hours forecast, right graphs show 15-min. time step forecast

Figure 2.11 shows that the higher the average water demand in the area of which the demand is forecasted, the larger the absolute forecasting errors but the smaller the percentage forecasting errors. The relation between the *MAPE* and the log value of the average water demand in the area (Q_{avg} [m^3/h]) can be approximated with:

$$MAPE_{day} = 6.5 - 0.59 \ln(Q_{avg}) \qquad R^2=0.92$$

$$MAPE_{qtr} = 12.6 - 1.05 \ln(Q_{avg}) \qquad R^2=0.90$$

The correlation (R^2) between the forecast error and the average demand in the array is rather good. Based on this correlation, we can estimate the average accuracy of the model when implementing it in a new area.

2.4.3 Accuracy in relation to the hour of forecast and the number of hours ahead

The previous sections showed the average accuracy of the model. This section will show how the accuracy varied in time. The accuracy of the model was not constant for each point in time on a day. In order to assess the differences in accuracy, the errors were calculated and plotted in Figure 2.12. The figure shows that in all areas the errors were smaller during the night, especially between 1:00 and 5:00 h. This indicates that the water demand in the night varied little over time, and that the predictability was high. The patterns of the errors somewhat resembled the patterns of the water demand, albeit that the amplitude of the error patterns was smaller. This indicates that the *MAPE* was related to forecasted water demand: the *MAPE* increased as the forecasted water demand increased.

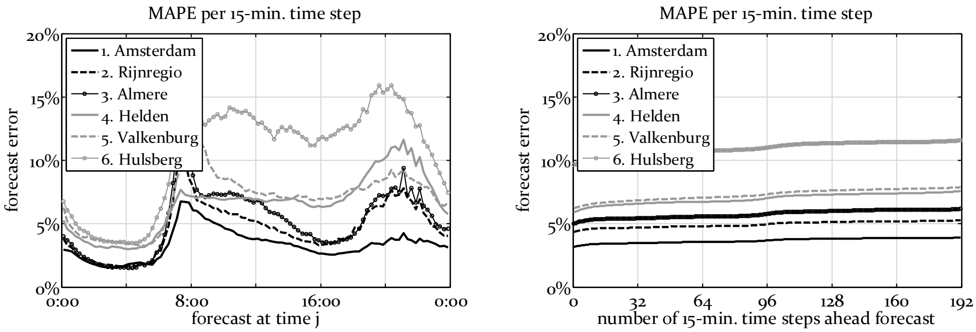


Figure 2.12 *MAPE* in relation to the forecasted point in time on the day (left graph), and in relation to the number of 15-min. time steps ahead forecast (right graph)

The model forecasts the water demand for 48 hours ahead (192 15-min. time step values). The first forecasted value is mainly based on the measured water demand right before this time step, as can observed in equation (2.2). The last forecasted value (for 48 hours ahead) is mainly based on the same information, but this information is measured some 48 hours before the time step actually occurs. This implies that a lower accuracy may be expected as the number of time steps ahead forecast increases. Figure 2.12 shows the *MAPE* as a function of the number of 15-min. time steps ahead for the 192 forecasted values. The figure shows that the *MAPE* gradually increased as the number of time steps ahead increased. Compared to the first forecasted value, the *MAPE* of the 96th value (24 hours ahead) was on average 15.7% higher, and of the 192th value (48 hours ahead) 23.7% higher.

2.5 Discussion

Influence model parameters

When assessing the accuracy of the model at different datasets, the model parameters were kept the same and were not optimised for the specific dataset. By doing this, the results resemble most the case when the model is implemented in a practical situation. However, the model parameters will affect the performance. In order to assess the influence of the model parameters, a sensitivity analysis was carried out. The model parameters which influence the forecast per 24 hours are: the weighing factors C_1 and C_2 (equation (2.2)) for forecasting the total demand per 24 hours; and the number of m previous observations to determine the day of the week factors, $f_{dotw,typ,i}$ (equation (2.15)). Figure 2.13 shows the sensitivity analysis for the C_1 / C_2 parameters, where C_1 was varied between 0.6 and 1.0, and $C_2 = 1 - C_1$. The graph shows optimal values for C_1 between 0.80 and 0.90. The graph also shows little variation in *MAPE* (per 24 hours) value when changing the values of C_1 / C_2: for values of C_1 between 0.80 and 0.90 the variation was on average 0.014%. This indicates that using the information about the water demand of 48 to 24 hours ago ($Q_{t=-191}$:$Q_{t=-96}$) only marginally improves the performance of the model. Simplification of the model by leaving out C_1 and C_2 factors and only use the information about the water demand in the previous 24 hours ago ($Q_{t=-95}$:$Q_{t=0}$) could therefore be considered.

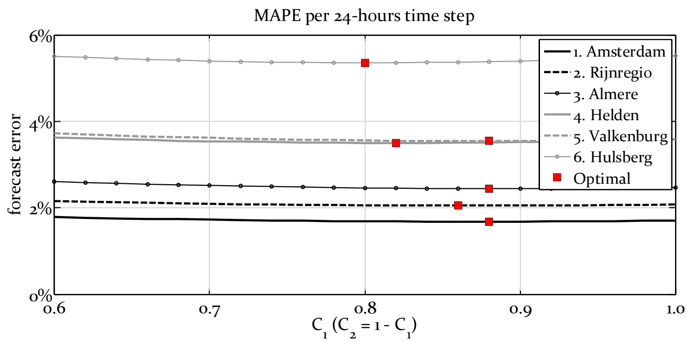


Figure 2.13 *MAPE* in relation to values of C_1 and C_2 (default 0.85 and 0.15)

In the sensitivity analysis, we also varied the m parameter between 2 and 22, and found minimal errors for m between 8 and 12. Like the C_1 / C_2 parameters, we found little variation in the *MAPE* (per 24 hours) value when changing the values of m: for values of m between 8 and 12 the variation was on average 0.011%.

The parameters which influence the forecast per 15-min. time step are: the number of n previous observations to determine the 15-min. time step factors, $f_{qtr,typ,i,j}$ and $f_{sprink,typ,i,j}$ (equations (2.13) and (2.14)); and the weighing factors BC_1 and BC_2 (equation (2.6)) for forecasting extra sprinkle demand. In the sensitivity analysis, we varied the n parameter between 1 and 7 and found minimal errors for n between 2.5 and 6.5. Like the parameters influencing the demand per 24-hours, we found little variation in MAPE value when changing the values of n: for values of n between 4 and 6 the variation was on average 0.07%. Figure 2.14 shows the sensitivity analysis for the BC_1 and BC_2 parameters. In order to illustrate the differences, the figure shows the frequency distribution of the *MAPE* (in summer of the period between 18:00 and 0:00 h) of one area for three different sets of BC_1 and BC_2. The figure shows that large underestimates occur when BC_1 and BC_2 are 0. This illustrates that sprinkle demand forecast mechanism is able to improve the forecast. The differences between the other sets of BC_1 and BC_2 are small. The largest underestimates are smaller when the sum of BC_1 and BC_2 exceeds 1. Because large underestimates need to be avoided when using the forecast for control, the default values for BC_1 and BC_2 were chosen at 1.1 and 0.1.

The sensitivity analysis showed that the model performance is affected only to a small degree by the model parameters. Therefore we conclude that calibrating the model parameters for a specific dataset instead of using the default values, will result in a relatively small improvement of the performance of the model. In most implementations the better performance will not justify the extra effort for calibrating the model. This justifies using the default parameters in most situations.

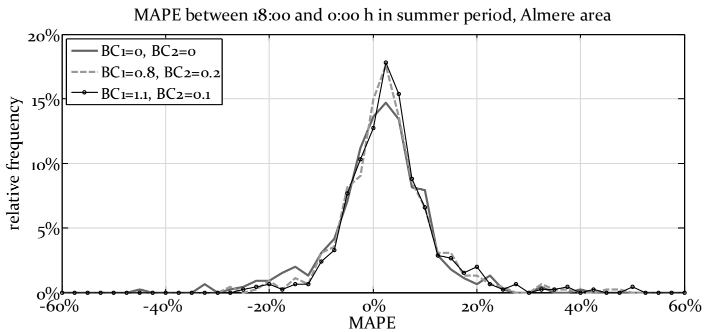


Figure 2.14 *MAPE* in the hours between 18:00 and 0:00 h (sprinkle period) in the summer period in the Almere area, in relation to parameters BC1 and BC2

Over-parameterisation

The forecasting model uses a large number of factors that describe the water demand characteristics: for each normal (7) and deviating (10) day type a normal pattern (96 values), a sprinkle pattern (96 values), and a typical day of the week factor (1 value). These factors are adaptively learned by the model by using the demand data. The implementation of the model in practice and the application of the model to six historic datasets in this chapter, have shown that the patterns and factors differ significantly from each other which justifies using the chosen number of patterns. Furthermore, the model uses eight parameters that are applied when forecasting the demand and storing the data. This section has shown (for six parameters) that the accuracy of the model is not affected much when the values of the parameters are varied. This justifies that default values are used for these parameters when implementing the model, rather than optimising the values by analysing historical data. However, enhancing the model by adaptively tuning the parameters based on historical data may be considered.

Weather conditions

This chapter shows that the model we developed is able to generate fairly accurate forecasts without using weather data as input, in the researched areas. All researched areas are situated in the Netherlands and have an equal moderate climate. The average daily maximum temperature in summer is around 19 ºC and in winter around 3 ºC. The moderate climate probably induces moderately changing water demands. In general, the peak day demand is 30–50% higher and the minimum day demand 20–30% lower than the average demand. The moderately changing water demand might be the reason for the good performance of the model without using weather data as input. The model has not been tested in areas with more variation in the weather conditions. Implementing the model in such areas may show a limitation of the model's performance.

Model accuracy in smaller areas

Table 2.2 and Table 2.3 show that the forecast errors expressed as *MAPE* and *RRMSE* increase as the size of the area decreases. This is the result of the fact that in larger areas random deviating water uses by individual consumers (or groups of consumers) are equalised by a large group of consumers who consume water normally. This results in an increasing unexplained variability in the water demand in smaller areas, which cannot be forecasted by the forecasting model. This is especially true for the smallest area (“Hulsberg”), which does not only have the highest values for *MAPE* and *RRMSE*, but also a considerable lower value for R^2 compared to all other areas. This limits the applicability of the model, and when implementing the model in small areas, the expected errors need to be considered in relation to the purpose the forecasting model.

Comparison to other water demand forecasting models

We think that it is not possible to compare the performance of our model with the performance of other models found in literature. The reason for this is that the forecast accuracy highly depends on the variability (or rather the forecastibility) of the water demand in the datasets. Table 2.2 and Table 2.3 show that the *MAPE* and the *RRMSE* can vary by a factor 3.5 and the R^2-values by 0.15 when we applied our model on different datasets. This indicates that the accuracy of forecasting models obtained by simulations on different datasets cannot be compared objectively. For a good comparison, different models should be tested on the same dataset(s). Therefore, we recommend that researchers in the field of water demand modelling share their datasets to allow objective model performance comparisons.

Performance evaluation

The performance evaluation of the model presented in this chapter only shows the average and the largest forecasting errors of the complete dataset. This evaluation does not provide information about the variation of the performance of the model in time, like differences between days of the week, differences between seasons, the variation of the performance on the defined deviating days, et cetera. Also the datasets on which the model was tested, were not analysed in detail. These evaluations and analyses would provide more insight in the data and the model, and would support in understanding how the model works. Therefore, we recommend future work for an in depth analyses of the datasets and forecasting residuals.

Influence of water loss and fire-fighting on the forecasting model

The model forecasts the water demand based on the measured water flow supplied to the area. This flow consists of the water consumption by all consumers in the area plus all water losses in the area plus eventual water use for fire-fighting. Water loss can be divided in background leakage and water loss as a result of pipe bursts. Background leakage is rather constant and forms a consistent value in the measured flow. Therefore, it will be captured by the water demand forecasting model in the adaptive typical demand curves. As a result, background leakage will not disturb the performance of the model. Only in case when new (small) leaks arise or when existing leaks are repaired, a subtle change in the demand pattern will occur. The model will gradually adapt the demand curves to this new situation, and after some five weeks the curves are fully adapted. The effect of background leakage on the forecast accuracy is therefore rather small, for both low levels (e.g. 5%) and high levels (e.g. 30%) of background leakage.

A catastrophic pipe burst and water use for fire-fighting (generating a flow exceeding e.g. 20% of the average flow) results in a sudden, large, but temporal change in the water supplied to an area. In contrast to background leakage, the effect of a pipe burst or water use for fire-fighting to the forecast accuracy can be large. The sudden extra water flow is completely random, and cannot be forecasted by the model. And also the point in time that the leak is isolated or that the water use for fire-fighting stops (which results in a sudden decrease of the flow) cannot be forecasted. The forecast error on a day with a (large) pipe burst or fire is likely to be one of the largest errors in a year. However, large pipe bursts and fires do not occur very frequently, which limits the effect of these events on the average forecast accuracy.

The total water losses in the Netherlands are relatively low (3%-7% for all water supply companies, (Beuken et al., 2007)) and therefore, the influence of water losses on the water demand is small on average in the researched datasets. However, the datasets will most likely contain measured flows during pipe burst events or fire events. Some of the largest forecast errors (see Table 2.2 and Table 2.3) may have been the result of such events, but because of a lack of information about bursts and fires, we were not able to confirm this hypothesis.

2.6 Conclusions

The model presented in this chapter, forecasts the drinking water demand for the next 48 hours with 15-min. time steps. The model uses measured water demand and static calendar data as only input to calculate the forecast. The performance of the model was examined by doing simulations with six datasets of water demands. The results showed that *MAPE* of the model varied between 1.44-5.12% for the 24-hours forecast, and between 3.35-10.44% for the 15-min. time step forecasts. The error had a strong relation with the average water demand in the area: The error increased as the average water demand in the area decreased.

The model has enhanced functionality compared to existing models with respect to the small time step (15-min.), the fully adaptive functionality, the number of discerned water demand patterns, and the mechanism to detect and forecast sprinkle demand without using weather input. The 15-min. time step is especially valuable, when the forecast is used for detailed optimisation where one hour time steps are too coarse. Because the model does not use weather input, it can be implemented reliably and easily. Still, with the mechanism to detect and forecast sprinkle demand, the model is able to generate fairly accurate forecast under different weather conditions. However, under highly changing weather conditions, the forecast errors increase. Inclusion of weather input in the model will likely reduce the errors under such circumstances.

The setup of the model is such, that it can be implemented and operated at low cost, and still generate accurate forecast. The characteristics make the model very suitable for implementation in practical situations. Experiences with implementations, showed that the model performs well, and that its forecast can be used for optimal control of the water supply systems.

2.7 References

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

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

Aly, A.H., and Wanakule, N. (2004). "Short-term forecasting for urban water consumption". *Journal of Water Resources Planning and Management.* 130 (5): 405-410.

Babel, M.S., and Shinde, V.R. (2011). "Identifying Prominent Explanatory Variables for Water Demand Prediction Using Artificial Neural Networks: A Case Study of Bangkok". *Water Resources Management.* 25 (6): 1653-1676.

Bakker, M., Van Schagen, K., and Timmer, J. (2003). "Flow control by prediction of water demand". *Journal of Water Supply: Research and Technology - AQUA*. 52 (6): 417-424.

Bakker, M., Vreeburg, J.H.G., Palmen, L.J., Sperber, V., Bakker, G., and Rietveld, L.C. (2013). "Better water quality and higher energy efficiency by using model predictive flow control at water supply systems". *Journal of Water Supply: Research and Technology - AQUA*. 62 (1): 1-13.

Bárdossy, G., Halász, G., and Winter, J. (2009). "Prognosis of urban water consumption using hybrid fuzzy algorithms". *Journal of Water Supply: Research and Technology - AQUA*. 58 (3): 203-211.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., and Andreassian, V. (2013). "Characterising performance of environmental models". *Environmental Modelling and Software*. 40: 1-20.

Beuken, R.H.S., Lavooij, C.S.W., Bosch, A., and Schaap, P.G. (2007). "Low leakage in the Netherlands confirmed", *Proceedings 8th Annual Water Distribution Systems Analysis Symposium 2006*, Cincinnati (OH), USA.

Bunn, S.M., and Reynolds, L. (2009). "The energy-efficiency benefits of pumpscheduling optimization for potable water supplies". *IBM Journal of Research and Development*. 53 (3).

Chen, L., Kao, S.J., and Traore, S. (2012). "Predicting and managing reservoir total phosphorus by using modified grammatical evolution coupled with a macro-genetic algorithm". *Environmental Modelling and Software*. 38: 89-100.

Donkor, E.A., Mazzuchi, T.A., Soyer, R., and Alan Roberson, J. (2014). "Urban water demand forecasting: Review of methods and models". *Journal of Water Resources Planning and Management*. 140 (2): 146-159.

Eliades, D.G., and Polycarpou, M.M. (2012). "Leakage fault detection in district metered areas of water distribution systems". *Journal of Hydroinformatics*. 14 (4): 992-1005.

Gato, S., Jayasuriya, N., and Roberts, P. (2007). "Forecasting residential water demand: Case study". *Journal of Water Resources Planning and Management*. 133 (4): 309-319.

Ghiassi, G., Zimbra, D.K., and Saidane, H. (2008). "Urban water demand forecasting with a dynamic artificial neural network model". *Journal of Water Resources Planning and Management*. 134 (2): 138-146.

Herrera, M., Torgo, L., Izquierdo, J., and Pérez-García, R. (2010). "Predictive models for forecasting hourly urban water demand". *Journal of Hydrology*. 387 (1-2): 141-150.

Homwongs, C., Sastri, T., and Foster, J.W. (1994). "Adaptive forecasting of hourly municipal water consumption". *Journal of Water Resources Planning & Management - ASCE*. 120 (6): 888-905.

House-Peters, L.A., and Chang, H. (2011). "Urban water demand modeling: Review of concepts, methods, and organizing principles". *Water Resources Research*. 47 (5).

Jain, A., Varshney, A.K., and Joshi, U.C. (2001). "Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks". *Water Resources Management*. 15 (5): 299-321.

Jentgen, L., Kiddler, H., Hill, R., and Conrad, S. (2007). "Energy management strategies use short-term water consumption forecasting to minimize cost of pumping operations". *Journal American Water Works Association*. 99 (6): 86-94.

Jowitt, P.W., and Chengchao, X. (1992). "Demand forecasting for water distribution systems". *Civil Engineering Systems*. 9 (2): 105-121.

Lertpalangsunti, N., Chan, C.W., Mason, R., and Tontiwachwuthikul, P. (1999). "Toolset for construction of hybrid intelligent forecasting systems: Application for water demand prediction". *Artificial Intelligence in Engineering.* 13 (1): 21-42.

Maidment, D.R., and Miaou, S.-P. (1986). "Daily water use in nine cities". *Water Resources Research.* 22 (6): 845-851.

Manera, M., and Marzullo, A. (2005). "Modelling the load curve of aggregate electricity consumption using principal components". *Environmental Modelling and Software.* 20 (11): 1389-1400.

Martínez, F., Hernández, V., Alonso, J.M., Rao, Z., and Alvisi, S. (2007). "Optimizing the operation of the Valencia water-distribution network". *Journal of Hydroinformatics.* 9 (1): 65-78.

Msiza, I.S., and Nelwamondo, F.V. (2011). "Predictive analysis of a water demand time series using a numerical method", *Proceedings IEEE Control and System Graduate Research Colloquium*, Shah Alam, Malaysia.

Salomons, E., Goryashko, A., Shamir, U., Rao, Z., and Alvisi, S. (2007). "Optimizing the operation of the Haifa-A water-distribution network". *Journal of Hydroinformatics.* 9 (1): 51-64.

Shvartser, L., Shamir, U., and Feldman, M. (1993). "Forecasting hourly water demands by pattern recognition approach". *Journal of Water Resources Planning & Management - ASCE.* 119 (6): 611-627.

Xu, M., van Overloop, P.J., and van de Giesen, N.C. (2013). "Model reduction in model predictive control of combined water quantity and quality in open channels". *Environmental Modelling and Software.* 42: 72-87.

Ye, G., and Fenner, R.A. (2011). "Kalman filtering of hydraulic measurements for burst detection in water distribution systems". *Journal of Pipeline Systems Engineering and Practice.* 2 (1): 14-22.

Zhou, S.L., McMahon, T.A., Walton, A., and Lewis, J. (2000). "Forecasting daily urban water demand: A case study of Melbourne". *Journal of Hydrology.* 236 (3-4): 153-164.

Zhou, S.L., McMahon, T.A., Walton, A., and Lewis, J. (2002). "Forecasting operational demand for an urban water supply zone". *Journal of Hydrology.* 259 (1-4): 189-202.

3 Improving the performance of water demand forecasting models by using weather input

Based on

M. Bakker
H. van Duist
J.H.G. Vreeburg
K.M. van Schagen
L.C. Rietveld

Improving the performance of water demand forecasting models by using weather input
Proceedings 12th International Conference on Computing and Control for the Water Industry, CCWI (2013), Perugia, Italy.

Abstract
Literature shows that short-term water demand forecasting models that use water demand as only input, are capable of generating a fairly accurate forecast. However, at changing weather conditions the forecasting errors are quite large. In this chapter, three different forecasting models are tested: an Adaptive Heuristic model, a Transfer/-noise model, and a Multiple Linear Regression model. The performance of the models both with and without using weather input (temperature information) is presented, in order to assess the possible performance improvement due to using weather input. Simulations with the models show that on average when using temperature information the largest forecasting errors can be reduced from 14.3% to 13.0% (9.4% reduction), and the average errors from 3.07% to 2.88% (6.3% reduction). This reduction may be important for the application of the forecasting model for the control of water supply systems and for anomaly detection.

Keywords
Demand forecasting; short-term; weather input; transfer/-noise model; MLR model

3.1 Introduction

There is an on-going trend towards a fully automated centralised operation of water supply systems (Worm et al., 2010; PWN, 2006). When utilities implement centralised automatic control, they aim to reduce costs and at the same time improve the quality of the operations or at least keep the same quality of the operations executed by motivated operators. This goal can be achieved by implementing models for the operational control of the systems. Short-term water demand forecasting models are often part of those kinds of control models. The forecast of the demand can be used for optimal flow control or for optimal pump scheduling, which leads to an improved water quality and a reduction of the energy consumption and costs (see chapter 4 of this thesis (Bakker et al., 2013a)). These models are used by a number of utilities around the world. In the Netherlands, for instance, in 2012 57% of all supplied water was controlled based on short-term water demand forecasts (Bakker et al., 2013a), while the penetration of forecasting models was expected to rise to over 90% in 2016. Other examples of implementations of control based on a short-term water demand forecast are at four large utilities in the United States (Bunn and Reynolds, 2009).

The accuracy of a water demand forecasting model is important to avoid sub-optimal control of the system, and to prevent that operators overrule the control settings in order to meet all operational constraints (e.g. to avoid a reservoir to overflow or to run empty). Not the average forecast errors but the largest errors –underestimates and overestimates– of the forecasting model play the most important role in the operational control. The reason for this is that large forecast errors may lead to undesired or unacceptable adjustments in control, or to violating operational constraints in case of limited (over)capacity of the water supply system. Despite the importance of the largest errors, many papers describing water demand forecasting models only report the average performance, expressed as the average error or as the coefficient of determination (R^2). Lertpalangsunti et al. (1999) and Jain et al. (2001) mention the largest errors explicitly, but both do not indicate whether the largest errors are underestimates or overestimates.

To forecast the water demand on the next day, various models have been developed and tested. House-Peters and Chang (2011) and Donkor et al. (2014) present extensive overviews of existing models. One of the first described models was based on linear regression of observed values of the daily water demand combined with transfer functions for rainfall and air temperature as independent variables (Maidment et al., 1985). Other papers describe forecasting models, based on the assumption that water demand is made up of base, seasonal, and weather dependent consumption (Zhou et al., 2000, 2002; Gato et al., 2007a, 2007b; Alvisi et al., 2007). In the models, a persistence component of the observed water demand is combined with different methods to transfer the independent observations (like temperature and rainfall) to forecasted water demand. The application of artificial neural networks (ANN) is most popular to forecast water demand. Initially, conventional ANN models were applied (Joo et al., 2002; Jain et al., 2001), but, as the development of ANN models proceeded, more complex and dynamic ANN models were applied (Ghiassi et al., 2008). Hybrid models which combine ANN models with other models like Fuzzy Logic and Fourier Transformations are described in several other papers (Lertpalangsunti et al., 1999; Bárdossy et al., 2009; Odan and Reis, 2012; Adamowski et al., 2012). All water demand forecasting models found in literature are based on mathematical techniques, either conventional regression / transfer functions, or more advanced data driven techniques. The models are often complex and abstract, and hard to understand for operators. This may be a disadvantage, because many operators only gain confidence in a model when they understand the model. However, we found no papers that describe heuristic (and understandable for operators) forecasting models based on observations of water demands.

Different inputs may be used to generate water demand forecasts. A limited number of papers describe forecasting models that use measured water demand as only input (Jowitt and Chengchao, 1992; Alvisi et al., 2007; Cutore et al., 2008; Caiado, 2010; Bakker et al., 2013b). Most models also use weather information as input, like temperature (Lertpalangsunti et al., 1999; Ghiassi et al., 2008), temperature and precipitation (Maidment et al., 1985; Jain et al., 2001; Bárdossy et al., 2009; Adamowski et al., 2012), and temperature, precipitation, and evaporation, wind speed and/or humidity (Zhou et al., 2002; Joo et al., 2002; Babel and Shinde, 2011). Although the relation between water demand and weather conditions seems obvious, some papers report that the performance of the forecasting model did not improve when using weather inputs (Ghiassi et al., 2008; Odan and Reis, 2012). In some papers, the difference in forecasting accuracy of models with and without using weather input is not reported, which limits the determination of the performance improvement. Chapter 2 of this thesis (Bakker et al., 2013b) stated that for an easy and reliable implementation of forecasting models, it is preferred to use measured water demand as only input, because connecting weather data as input to the model results in extra costs and possible risks of missing input for the model (depending on how the input of weather data has been implemented). In order to make a decision whether to use weather input or not, both the performance improvement of the model and the extra costs and risks need to be considered.

The performance of a forecasting model can be assessed by testing it on a historic dataset of water demands. Generally, the dataset is split into two parts: a part for developing the model and setting the parameters of the model ("training" or "calibration" set); and a part for assessing the model's performance on an independent dataset ("testing" or "validation" set). The data for testing the model should describe all normal variations in the water demand, in order to obtain a reliable performance evaluation of the model. Water demand varies over the year, and also the variability (or rather the forecastability) varies, and therefore we suggest to use a dataset for testing a model of at least one year and preferably three years. However, the data for testing a model was shorter than four months in several papers (Zhou et al., 2002;

Jain et al., 2001; Ghiassi et al., 2008; Bárdossy et al., 2009; Babel and Shinde, 2011; Caiado, 2010).

In this chapter we compare the performance of an adaptive heuristic model (this is a simplified version of the model presented in chapter 2 of this thesis), a transfer- /noise model and a MLR model for forecasting the water demand on the next day. The central research question in this chapter is, what performance improvement can be achieved by using weather data, and therefore all models are evaluated both with and without using weather inputs (temperature information). We selected the three models for this comparison, because those were readily available for the authors. The performance of the models is evaluated by applying the models to water demand data in six different water supply zones in the Netherlands for a period of six years.

3.2 Methods and materials

3.2.1 Study areas and data

The same data is used as in chapter 2 (see Table 2.1 and Figure 2.4). The data contains the water demand in six different areas in the Netherlands in the period 2006-2011. Each dataset consists of the measured water demand per day in m^3 over a period of six years (2,192 values). The models were trained with three years of data (2006-2008), and tested with a subsequent dataset of three years (2009-2011).

3.2.2 Adaptive heuristic model

Based on the experience of operators observing water demands, we developed a heuristic adaptive water demand forecasting model, as described in chapter 2 of this thesis. The model for forecasting the daily water demand was set up using the following three main observations: 1. The variation in the daily water demand from one day to the next is limited; 2. Subsequent daily water demands describe a weekly pattern; 3. Changes in daily average temperature, result in changes in the water demand. The heuristic model forecasts the water demand for the next day ($Q^*_{D,i}$), primarily based on the measured water demand in the previous two days. In order to correct for the day of the week, the measured water demand on day i ($Q_{D,i}$) is divided by the typical day of the week factor of day i ($f_{dotw,typ,i}$). The water demand is forecasted by applying the day of the week factor to the corrected water demand of the previous two days:

$$Q^*_{D,i} = f_{dotw,typ,i} \cdot \left(C_1 \cdot \frac{Q_{D,i-1}}{f_{dotw,typ,i-1}} + C_2 \cdot \frac{Q_{D,i-2}}{f_{dotw,typ,i-2}} \right) \qquad [m^3/day] \tag{3.1}$$

Based on a sensitivity analysis presented in section 2.5, the model parameters C_1 and C_2 are by default set at 0.85 and 0.15 respectively for the best forecasting results. When using these values, the more recent measured water demands weigh heavier than the older demands. By using equation (3.1) the forecasted average water demand is based on a relative short period of measured demands (previous two days, with emphasis on the last day). This results in a quick adjustment of the forecasted water demand, after a change of the measured water demand. The day of the week factor for day type *ti* ($f_{dotw,typ,ti}$) is adaptively learned by the model, using measurements of the daily water demand in the previous *m* (default 10) weeks:

$$f_{dotw,typ,ti} = \frac{\frac{1}{m}\sum_{i=1}^{i=m} Q_{D,\{typ=ti\}}}{\frac{1}{m\cdot 7}\sum_{i=1}^{i=m\cdot 7} Q_{D,i}} \qquad [-] \tag{3.2}$$

By applying equation (3.1) in combination with the day of the week factor derived with equation (3.2), the model automatically corrects for the variation in the daily demands which occur in a weekly pattern.

Changing weather conditions affect the water consumption. To improve the performance of the forecasting model during changing weather conditions, we added a temperature correction factor. When a change in the temperature occurs, the model calculates the corrected forecast ($Q^{**}_{D,i}$) by multiplying the original forecast with the temperature correction factor (f_T) and the change in temperature between the forecasted and previous day (T_i – T_{i-1}, note that for T_i a forecasted temperature from a weather bureau needs to be used):

$$Q^{**}_{D,i} = Q^{*}_{D,i} \cdot f_T \cdot \left(T_i - T_{i-1}\right) \qquad [m^3/day] \tag{3.3}$$

The temperature correction factor is calculated by deriving a relation using a least squares fit between the forecast errors (of the uncorrected forecast) and the change in temperature. The change in temperature is used rather than the absolute temperature, because a change in temperature will result in a change in the water demand; a constant temperature results in a rather constant demand (relatively high or low, depending on the temperature), which is forecasted well by the model. When deriving the relation, only those days were selected on which the temperature exceeded 10 °C, the temperature change exceeded 0.5 °C, and a positive relation between temperature change and the change in water demand was observed (see Figure 3.1). We used this heuristic approach rather than a mathematical approach (e.g. formulate it as a univariate optimisation problem), because this allows to show the historic data and derived relation to the operator, which is helpful in understanding and verifying the model.

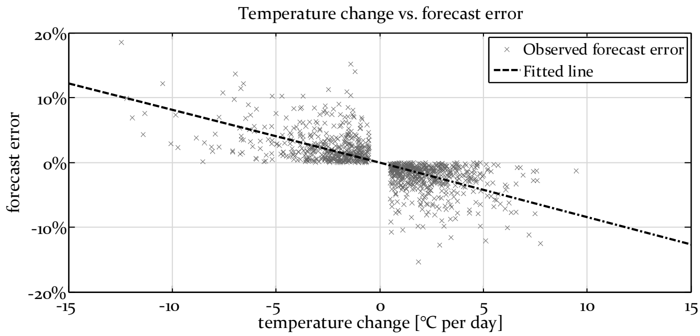


Figure 3.1 Forecast error as function of temperature change. The derivative of the fitted line is the f_T factor, which is derived for both increases and decreases of the temperature

3.2.3 Transfer- /noise model

In modelling a time series, Box and Jenkins (1976) combined the benefits of both an auto regressive integrated moving average (ARIMA) model and a transfer model, in a so-called transfer-/noise model. The basic assumption in this model is that the output signal of a dynamical system is driven by several signals, including white noise (Figure 3.2). These kinds of models have successfully been applied in many scientific areas, including hydrology (Castellano-Méndez et al., 2004) and economics (Grillenzoni, 2000).

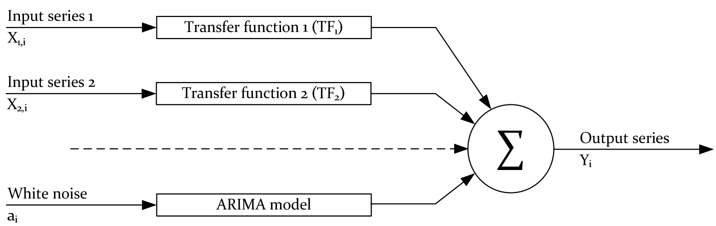


Figure 3.2 Setup of transfer-/noise model

In case the output series y_i cover the same time interval as the input series x_i, the dynamical relation between two series can be modelled with a transfer function of order (r,s):

$$\left(1-\delta_1 B-\delta_2 B^2-\ldots-\delta_r B^r\right)\cdot y_i=\left(\omega_0-\omega_1 B-\omega_2 B^2-\ldots-\omega_s B^s\right)\cdot x_i \qquad (3.4)$$

where δ_1 to δ_r are the autoregressive parameters, ω_0 to ω_s the moving-average parameters and B the backshift or lag operator ($B\cdot x_i = x_{i-1}$), or simplified $y_i = \omega(B)/\delta(B)\cdot x_i = \mathrm{TF}(B,x_i)$.

The ARIMA model (noise model) is a special form of the transfer model. The output signal y_i is modelled as a linear function of white noise only. The general form is:

$$\left(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p\right)\cdot\left(\nabla^d y_i-c\right)=\left(1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q\right)\cdot a_i \tag{3.5}$$

where ϕ_1 to ϕ_r are the autoregressive parameters, θ_1 to θ_q the moving-average parameters, a the white noise, ∇ the difference operator ($\nabla y_i = y_i - y_{i-1} = (1-B)\cdot y_i$), d the number of differences ($\nabla^d y_i = y_i - y_{i-d} = (1-B^d)\cdot y_i$) and c a constant. The transfer-/noise model is generated by linear superposition of the transfer function(s) and the noise function. The role of the latter is to account for any influences on the system output, which are not covered by any of the input series $x_{n,i}$. In case of two inputs, the relation between output y_i and the inputs $x_{n,i}$ becomes:

$$y_i = TF_1(x_{1,i}) + TF_2(x_{2,i}) + N_i \tag{3.6}$$

where TF_n is the Transfer Function and N_i is the white noise component. When using the Box-Jenkins methodology to forecast water demand, a number of modifications were applied. Firstly, only the Moving-Average parameters turned out to be valuable (no significant δ parameters were found during the model building phase, and therefore not used in the model). Secondly, water demand is not a stationary series but usually shows a weekly modulation. Therefore, we differentiated the original demand series ($Q_{D,i}$) twice, and used the resulting output series y_i as the output of the transfer-/noise model:

$$y_i = \nabla\nabla^7 Q_{D,i} = \left(Q_{D,i} - Q_{D,i-1}\right) - (Q_{D,i-7} - Q_{D,i-8}) \tag{3.7}$$

Like in the heuristic model, the daily average temperature is used as an independent variable to forecast the water demand. Temperature only affects water demand during the spring and summer (average temperature higher than 10 ºC), and the effect increases as temperature increases. We found optimal forecasting results when using a transformed value (z_i) of the original temperature values (T_i) with:

$$z_i = 0.001\cdot\left(T_i - 10\right)^3 \;\; if\; T_i > 10 \qquad Else\; z_i = 0 \tag{3.8}$$

Note that in equation (3.8) a factor of 0.001 is applied to ensure that the z_i term has the same magnitude as the other terms in equation (3.10), which enabled efficiently finding the unknown parameters. To compensate for special days with deviating water demand (like national holidays), a dummy input series F_i was created with values of -1 on special days with expected lower demand and +1 for expected higher demands. Both abovementioned input series (transformed temperature z_i and the intervention series for special days F_i) were differentiated in the same way as the demand series (equation (3.7). The noise model (N_i) was kept simple, by constructing it as an additive model with 3 moving average parameters:

$$N_i = \left(a_i - \theta_1 \cdot a_{i-1}\right) - \left(\theta_7 \cdot a_{i-7} - \theta_8 \cdot a_{i-8}\right) \tag{3.9}$$

Assuming the white noise elements *a* are 0, the resulting forecast model is formulated as:

$$\begin{aligned} Q^*_{D,i} = {} & Q_{D,i-1} + \left(Q_{D,i-7} - Q_{D,i-8}\right) + \\ & \omega_{1,0}\left(z_i^* - z_{i-1} - z_{i-7} + z_{i-8}\right) + \omega_{1,1}\left(z_{i-1} - z_{i-2} - z_{i-8} + z_{i-9}\right) + \\ & \omega_{2,0}\left(F_i - F_{i-1} - F_{i-7} + F_{i-8}\right) + \omega_{2,1}\left(F_{i-1} - F_{i-2} - F_{i-8} + F_{i-9}\right) + \\ & -\theta_1 a_{i-1} - \theta_7 a_{i-7} + \theta_8 a_{i-8} \end{aligned} \tag{3.10}$$

where $Q^*_{D,i}$ is the forecasted demand and z^*_i is the transferred forecasted average temperature, obtained from a weather bureau. The unknown parameters (θ_1, θ_7, θ_8, $\omega_{1,0}$, $\omega_{1,1}$, $\omega_{2,0}$, $\omega_{2,1}$) were estimated using the training dataset of Q_D, T and F values.

3.2.4 Multiple Linear Regression (MLR) model

When omitting the noise component, the transfer-/noise model is reduced to a simplified multiple linear regression model. We executed a sensitivity analysis and found that the regression parameters z_{i-7} and z_{i-8} were not significant and we therefore adopted the following forecast model:

$$Q^*_{D,i} = Q_{D,i-1} + Q_{D,i-7} - Q_{D,i-8} + \beta_1 z_i + \beta_2 z_{i-1} + \beta_3 F_i + \beta_4 F_{i-1} + \beta_5 F_{i-7} + \beta_6 F_{i-8} \tag{3.11}$$

where the Q_{Di}, z_i and F_i are the same as in the transfer-/noise model. The β_n were estimated using the training dataset with the Least Squares Algorithm.

3.2.5 Model performance evaluation

The performance of the different models and the performance improvement by using weather input (temperature information) were assessed by comparing forecasted and measured values. As stated before, not only the average errors are important, but also the largest errors (underestimates and overestimates). For evaluating the models we therefore chose the following evaluation parameters: The 99% confidence intervals of the relative error RE_i (the values of $RE_{0.5\%}$ and $RE_{95.5\%}$, see equation (2.17)); The mean absolute percentage error (*MAPE*, see equation (2.18)), and The Nash-Sutcliffe model efficiency (R^2, see equation (2.20)).

The $RE_{0.5\%}$ and $RE_{95.5\%}$ are the extremes of the 99% confidence interval, which represent the underestimates and overestimates which are statistically exceeded approximately two days in a year. We chose this measure rather than the absolute largest underestimates and overestimates, because those might be related to isolated anomalies (e.g. caused by pipe bursts or fire-fighting) that are not representative for the performance of the model. By calculating the abovementioned confidence intervals, the evaluation is not influenced by single observations, and still the largest errors that may be expected are indicated properly.

3.3 Results

Table 3.1 to Table 3.3 show the results of the three investigated forecasting models, both without and with using weather input, and Figure 3.3 shows the average values.

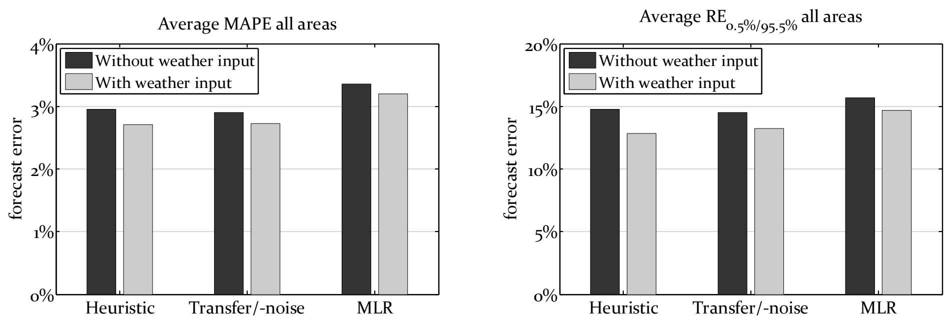


Figure 3.3 Average forecasting errors in all areas, expressed as *MAPE* (left graph) and as the mean of the *RE*0.5% and *RE*95.5% values (right graph)

The forecast errors of both the heuristic and the transfer/-noise model were 10-15% smaller than the MLR model. This indicates that water demand cannot be described very accurately with regression formulas only, which was also observed by Adamowski et al. (2012). The tables show a comparable performance of the heuristic model and the transfer/-noise model (see Table 3.1 and Table 3.2), with, on average, a slightly better performance of the transfer/-noise model. When using weather input, the forecast errors were smaller for all models in all areas. The average performance improvement of all models in all areas by using weather input was 6.3% for the *MAPE*, and 9.4% for the maximum errors (calculated as average of the $RE_{0.5\%}$ and the $RE_{99.5\%}$ improvements). The largest improvements were observed at the heuristic model (8.2% and 13.0%), next the transfer/-noise model (6.1% and 8.7%), and the smallest improvements at the MLR model (4.6% and 6.5%). This indicates that the performance improvements due to using weather input were most significant for the heuristic model and for reducing the largest forecast errors.

Table 3.1 Results heuristic water demand forecasting model (testing period: 2009-2011)

Area	**Without weather input**				**With weather input**			
	MAPE	***R^2***	***$RE_{0.5\%}$***	***$RE_{95.5\%}$***	***MAPE***	***R^2***	***$RE_{0.5\%}$***	***$RE_{95.5\%}$***
1. Amsterdam	1.39%	0.756	-6.5%	7.0%	1.32%	0.790	-5.7%	6.3%
2. Rhine area	1.88%	0.694	-8.9%	9.8%	1.73%	0.751	-8.9%	7.9%
3. Almere	2.08%	0.718	-11.0%	10.8%	1.97%	0.764	-9.4%	8.4%
4. Helden	3.72%	0.794	-24.5%	17.3%	3.33%	0.848	-19.5%	14.1%
5. Valkenburg	3.55%	0.788	-16.5%	15.1%	3.38%	0.808	-16.1%	14.5%
6. Hulsberg	5.08%	0.688	-28.7%	21.0%	4.48%	0.755	-22.5%	20.7%
Average	*2.95%*	*0.738*	*-16.0%*	*13.5%*	*2.71%*	*0.784*	*-13.7%*	*12.0%*

Table 3.2 Results Transfer-/noise water demand forecasting model (testing period: 2009-2011)

Area	**Without weather input**				**With weather input**			
	MAPE	***R^2***	***$RE_{0.5\%}$***	***$RE_{95.5\%}$***	***MAPE***	***R^2***	***$RE_{0.5\%}$***	***$RE_{95.5\%}$***
1. Amsterdam	1.36%	0.766	-6.3%	5.9%	1.25%	0.812	-5.7%	5.2%
2. Rhine area	1.77%	0.740	-8.6%	8.7%	1.65%	0.777	-7.4%	8.2%
3. Almere	2.03%	0.733	-9.6%	10.9%	1.87%	0.793	-8.2%	9.6%
4. Helden	3.74%	0.791	-21.6%	20.4%	3.47%	0.832	-19.4%	18.6%
5. Valkenburg	3.44%	0.807	-15.8%	14.7%	3.32%	0.818	-16.3%	14.2%
6. Hulsberg	5.04%	0.696	-26.1%	25.9%	4.77%	0.727	-22.7%	23.2%
Average	*2.90%*	*0.754*	*-14.6%*	*14.4%*	*2.73%*	*0.792*	*-13.3%*	*13.2%*

Table 3.3 Results MLR water demand forecasting model (testing period: 2009-2011)

Area	**Without weather input**				**With weather input**			
	MAPE	***R^2***	***$RE_{0.5\%}$***	***$RE_{95.5\%}$***	***MAPE***	***R^2***	***$RE_{0.5\%}$***	***$RE_{95.5\%}$***
1. Amsterdam	1.54%	0.709	-7.4%	6.3%	1.47%	0.738	-6.4%	6.0%
2. Rhine area	1.99%	0.681	-8.3%	10.5%	1.87%	0.731	-7.4%	9.5%
3. Almere	2.37%	0.681	-11.0%	10.7%	2.26%	0.724	-10.3%	8.7%
4. Helden	4.25%	0.747	-23.2%	20.4%	4.01%	0.780	-22.4%	18.7%
5. Valkenburg	3.97%	0.756	-20.3%	15.0%	3.81%	0.772	-18.3%	14.6%
6. Hulsberg	5.97%	0.619	-29.2%	26.2%	5.73%	0.643	-28.8%	24.7%
Average	*3.35%*	*0.697*	*-16.6%*	*14.8%*	*3.20%*	*0.729*	*-15.6%*	*13.8%*

3.4 Discussion

Performance in areas of different size

Table 3.1 to Table 3.3 show increasing forecast errors (expressed as *MAPE*) with decreasing water demand in the areas. The errors in the smallest area (6. Hulsberg) were approximately a factor 3.5 larger than the errors in the largest area (1. Amsterdam). The R^2 values indicate how well the model is able to describe the variation in the data. The R^2 values for the different areas were in the same range, varying between 0.7 and 0.85 for the heuristic and transfer/-noise model. This shows that the performance of the forecasting models is comparably in the different areas. The higher *MAPE* values in combination with the comparable R^2 values in the smaller areas, indicate that the variability of the water demands in those areas was larger, but that the models were able to describe those variations to a comparable degree. The larger percentage errors in the smaller areas are thus in particular explained by the higher variation in the demand rather than by a worse performance of the forecasting models. This indicates that larger forecast errors may be expected when the model is applied to forecast the demand in smaller areas (see also chapter 2 (Bakker et al., 2013b), which limits the applicability of the model in such small areas.

Comprehensible forecasting model

The heuristic model is easier to understand for operators than the mathematical transfer/-noise and MLR models. In the heuristic model, the adaptively learned factors that influence the forecast can be presented to the operators, which allows them to verify the validity of the factors and to understand how future demands will be forecasted. This may be important when implementing a forecasting model in a real water supply system. It allows an open discussion between the operator and the control engineer about the current performance and possible improvements of the model. This will result in an early detection of sub-optimal performance of the model and timely response. However, a more complex mathematical model may also be accepted by operators, when the operators have been experiencing for a while that the model is performing well, although they do not quite understand how it works.

Significance performance improvement

The average performance improvement in all areas due to using weather input (temperature information) was around 6.3% (reduction from 3.07% to 2.88%) for the average errors and 9.4% (reduction from 14.3% to 13.0%) for the largest errors. To achieve this improvement in practice, a more complex model must be used and additional effort must be put in connecting the weather information to the forecasting model. It is questionable whether the extra effort and costs weigh up to the forecasting performance improvement. The answer to this question will be different in each individual case. When using the forecast for the control of a small scale water supply system with robust treatment plants and some installed overcapacity, more robust, simple and low-cost solutions may be preferred, and thus a forecasting model without weather input. However, when the forecast is used for the control of large scale systems with more complex treatment plants, the forecast errors must be minimised, and thus a forecasting model with weather input may be preferred.

When the forecast is used for detection of anomalies (e.g. pipe burst detection), the normal forecast errors determine the accuracy of the detection method. Chapter 6 of this thesis (Bakker et al., 2014) will show that the detection threshold values in a pipe burst detection method can be directly related to forecast errors of the forecasting model. As a result, the size of the pipe burst that can be detected has a direct relation with the forecast errors, and reducing the forecast errors will enable the detection of smaller bursts. Again, case-by-case must be evaluated if the increased effort and costs of connecting weather input weighs up against the increased accuracy of forecasts and detection performance.

Other weather parameters
In this chapter, we only studied the potential performance improvement by using temperature information, and no other weather parameters like precipitation, humidity, wind speed, et cetera were evaluated. Other papers (e.g. Adamowski et al. (2012); Babel and Shinde (2011)) showed that temperature is the most significant weather parameter for water demand forecasting. This indicates that further performance improvements by using other weather parameters may be possible, but the significance of improvements will be limited.

3.5 Conclusions

Simulations with six different sets of water demands, showed that a heuristic model and a transfer/-noise model outperformed a multiple linear regression model when forecasting the water demand for the next day. The transfer/-noise model performed somewhat better than the heuristic model, but the heuristic model is easier to understand which is favourable for implementation in practice. The average and maximum errors showed a correlation with the size of the area: the errors increased as the size of areas decreased. Because the R^2 values were comparable in all areas, we conclude that the larger errors in smaller areas were in particular caused by a larger variation in the water demand rather than by a worse performance of the forecasting models.

When using weather input, the performance of the forecasting models could be improved by 6.3% with respect the average errors, and 9.4% with respect to the largest errors. The largest performance improvements were observed at the heuristic model, and the smallest improvements at het MLR model. This improvement can be relevant when higher forecasting accuracies are necessary when using the forecasts for optimal control or for anomaly detection. A case-by-case evaluation must show whether the extra effort and costs of connecting weather information to the forecasting model weighs up against the improved performance of the model.

3.6 References

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

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

Babel, M.S., and Shinde, V.R. (2011). "Identifying Prominent Explanatory Variables for Water Demand Prediction Using Artificial Neural Networks: A Case Study of Bangkok". *Water Resources Management.* 25 (6): 1653-1676.

Bakker, M., Vreeburg, J.H.G., Palmen, L.J., Sperber, V., Bakker, G., and Rietveld, L.C. (2013a). "Better water quality and higher energy efficiency by using model predictive flow control at water supply systems". *Journal of Water Supply: Research and Technology - AQUA.* 62 (1): 1-13.

Bakker, M., Vreeburg, J.H.G., van Schagen, K.M., and Rietveld, L.C. (2013b). "A fully adaptive forecasting model for short-term drinking water demand". *Environmental Modelling and Software.* 48 (1): 141-151.

Bakker, M., Vreeburg, J.H.G., Van de Roer, M., and Rietveld, L.C. (2014). "Heuristic burst detection method using flow and pressure measurements". *Journal of Hydroinformatics.* (in press).

Bárdossy, G., Halász, G., and Winter, J. (2009). "Prognosis of urban water consumption using hybrid fuzzy algorithms". *Journal of Water Supply: Research and Technology - AQUA.* 58 (3): 203-211.

Box, G.E.P., and Jenkins, G.M. (1976). "Time Series Analysis: Forecasting and Control". 2 nd ed. San Francisco: Holden-Day.

Bunn, S.M., and Reynolds, L. (2009). "The energy-efficiency benefits of pumpscheduling optimization for potable water supplies". *IBM Journal of Research and Development.* 53 (3).

Caiado, J. (2010). "Performance of combined double seasonal univariate time series models for forecasting water demand". *Journal of Hydrologic Engineering.* 15 (3): 215-222.

Castellano-Méndez, M., González-Manteiga, W., Febrero-Bande, M., Manuel Prada-Sánchez, J., and Lozano-Calderón, R. (2004). "Modelling of the monthly and daily behaviour of the runoff of the Xallas river using Box-Jenkins and neural networks methods". *Journal of Hydrology.* 296 (1-4): 38-58.

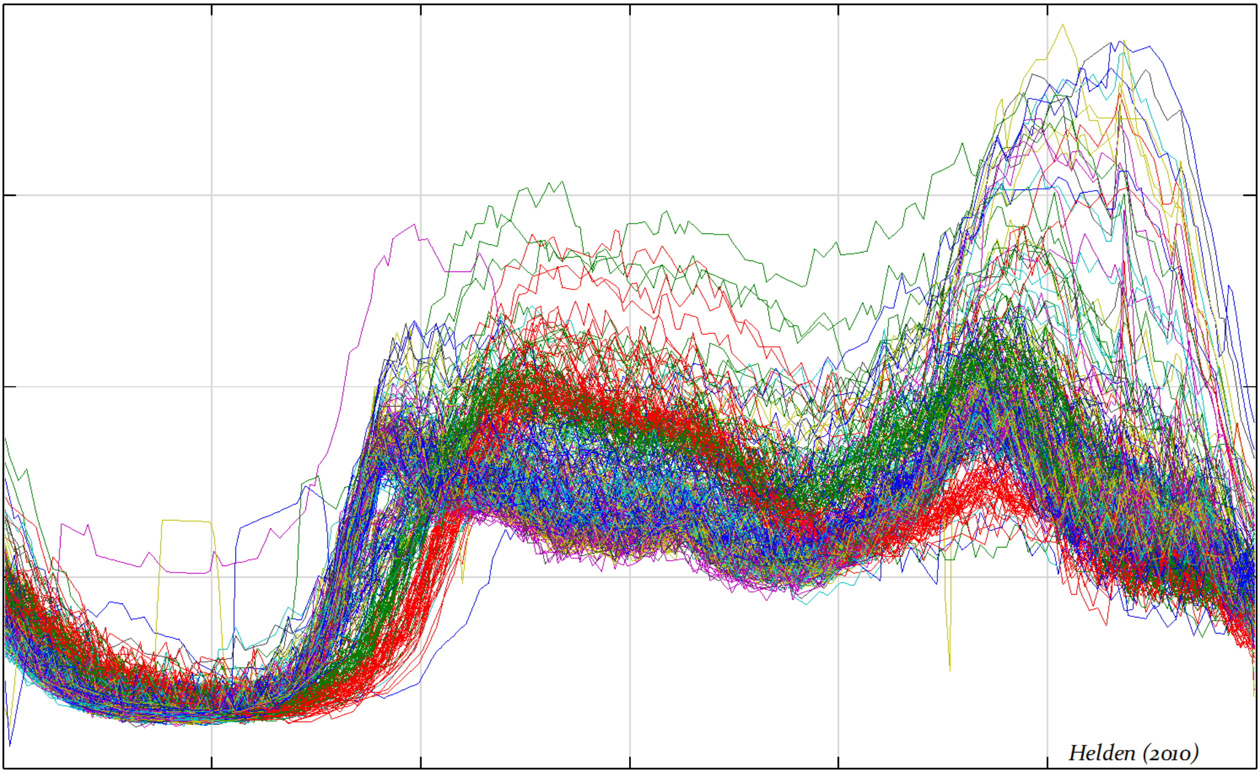
Cutore, P., Campisano, A., Kapelan, Z., Modica, C., and Savic, D. (2008). "Probabilistic prediction of urban water consumption using the SCEM-UA algorithm". *Urban Water Journal.* 5 (2): 125-132.

Donkor, E.A., Mazzuchi, T.A., Soyer, R., and Alan Roberson, J. (2014). "Urban water demand forecasting: Review of methods and models". *Journal of Water Resources Planning and Management.* 140 (2): 146-159.

Gato, S., Jayasuriya, N., and Roberts, P. (2007a). "Temperature and rainfall thresholds for base

use urban water demand modelling". *Journal of Hydrology.* 337 (3-4): 364-376.
Gato, S., Jayasuriya, N., and Roberts, P. (2007b). "Forecasting residential water demand: Case study". *Journal of Water Resources Planning and Management.* 133 (4): 309-319.
Ghiassi, G., Zimbra, D.K., and Saidane, H. (2008). "Urban water demand forecasting with a dynamic artificial neural network model". *Journal of Water Resources Planning and Management.* 134 (2): 138-146.
Grillenzoni, C. (2000). "Time-varying parameters prediction". *Annals of the Institute of Statistical Mathematics.* 52 (1): 108-122.
House-Peters, L.A., and Chang, H. (2011). "Urban water demand modeling: Review of concepts, methods, and organizing principles". *Water Resources Research.* 47 (5).
Jain, A., Varshney, A.K., and Joshi, U.C. (2001). "Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks". *Water Resources Management.* 15 (5): 299-321.
Joo, C.N., Koo, J.Y., and Yu, M.J. (2002). "Application of short-term water demand prediction model to Seoul". *Water Science and Technology* 46 (6-7): 255-261.
Jowitt, P.W., and Chengchao, X. (1992). "Demand forecasting for water distribution systems". *Civil Engineering Systems.* 9 (2): 105-121.
Lertpalangsunti, N., Chan, C.W., Mason, R., and Tontiwachwuthikul, P. (1999). "Toolset for construction of hybrid intelligent forecasting systems: Application for water demand prediction". *Artificial Intelligence in Engineering.* 13 (1): 21-42.
Maidment, D.R., Miaou, S.-P., and Crawford, M.M. (1985). "Transfer function models of daily urban water use". *Water Resources Research.* 21 (4): 425-432.
Odan, F.K., and Reis, L.F.R. (2012). "Hybrid Water Demand Forecasting Model Associating Artificial Neural Network with Fourier Series". *Journal of Water Resources Planning and Management.* 138 (3): 245-256.
PWN. (2006). "Automated Water System Control". *Control Engineering.* 53 (10).
Worm, G.I.M., van der Helm, A.W.C., Lapikas, T., van Schagen, K.M., and Rietveld, L.C. (2010). "Integration of models, data management, interfaces and training support in a drinking water treatment plant simulator". *Environmental Modelling and Software.* 25 (5): 677-683.
Zhou, S.L., McMahon, T.A., Walton, A., and Lewis, J. (2000). "Forecasting daily urban water demand: A case study of Melbourne". *Journal of Hydrology.* 236 (3-4): 153-164.
Zhou, S.L., McMahon, T.A., Walton, A., and Lewis, J. (2002). "Forecasting operational demand for an urban water supply zone". *Journal of Hydrology.* 259 (1-4): 189-202.

Part III – Optimised control



4 Better water quality and higher energy efficiency by using model predictive flow control at water supply systems

Based on

M. Bakker
J.H.G. Vreeburg
L.J. Palmen
V. Sperber
G. Bakker
L.C. Rietveld

Better water quality and higher energy efficiency by using model predictive flow control at water supply systems
Journal of Water Supply: Research and Technology – AQUA (2013). 62 (1): 1-13

Abstract

57% of all water supply systems in the Netherlands are controlled by model predictive flow control; the other 43% are controlled by conventional level based flow control. This chapter shows the differences between conventional level based flow control and model predictive control, based on experiments at five full scale water supply systems executed in the first half of 2011. Quality parameters of the drinking water and energy consumption of the treatment and distribution processes are presented. The experiments show 12-28% lower turbidity values, and 12-42% lower particle volume values when the systems are controlled by model predictive flow control. The experiments also show that predictive flow control leads to a 1.0-5.3% lower overall energy consumption and a 1.7-7.4% lower overall energy costs.

Keywords

Demand prediction; drinking water; energy reduction; optimal control, water quality

4.1 Introduction

4.1.1 Automation of water supply systems

Water utilities around the developed world started automating their water supply systems around the mid 1970's by installing and operating telemetry and supervisory control and data acquisition (SCADA) systems (Bunn, 2007; Bunn and Reynolds, 2009). Prior to this period the treatment plants and pumping facilities were mainly operated manually. In the Netherlands most small scale water treatment plants (typically groundwater treatment plants serving on average 50,000 people, producing and distributing 6,000 m^3 per day) were automated extensively at that time, enabling unmanned operation. Unmanned operation in this perspective means that under normal operational conditions no manual actions of operators are necessary for pressure and flow control. Therefore both pressure control of the pumping stations and flow control of the water treatment plants were fully automated. At night and during the weekends no personnel is present at the small scale facilities since the automation systems were implemented. Small scale water supply systems were automated extensively at an early stage because of two aspects. The first aspect is that manually operating small scale systems is labour intensive and therefore rather expensive. The second aspect is that small scale systems are rather simple systems: the treatment plants consist mainly of robust aeration and filtration steps, and also the distribution systems are rather straightforward. Initially, the small scale water supply systems were automated with relative simple control loops: the set-point for the production flow was derived directly from the level in the reservoir. This level based production flow control is simple and robust. However this way of control results in variations in the production flow, which causes variations in the water quality. In the 1990's the desire for more advanced control loops grew to achieve a more constant production flow.

4.1.2 Level based versus model predictive flow control

In the current practise, two methods are commonly applied for the automatic production flow control: level based control and model predictive control. In level based control loops (see Figure 4.1) the production flow set-point is directly related to the level in the reservoir. The production flow set-point increases at a decreasing level in the reservoir, the set-point decreases at an increasing level. This set-point can be given as discrete commands to start or stop pumps or filters (based on fixed switch levels), or a continuous value for variable speed pumps. In general the production flow set-point more or less follows the outgoing flow, with a time lag of 2-4 hours. The reservoir is merely used as a switching buffer, rather than a buffer to balance the variation in the distribution flow. The production flow varies and the maximum and minimum flow values of the production flow are comparable with the maximum and minimum flow values of the distribution flow (Bakker et al., 2003).

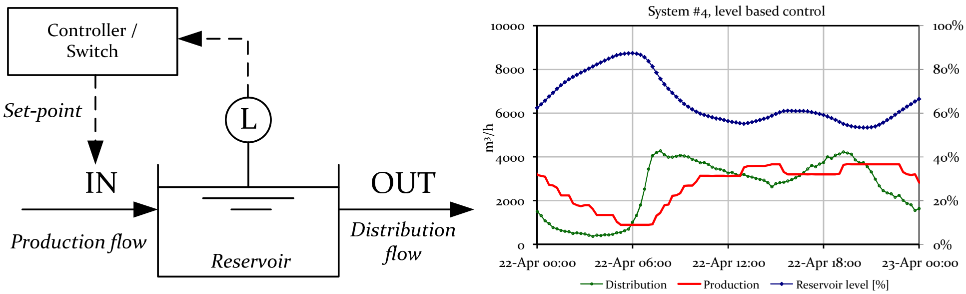


Figure 4.1 Principles of level based flow control and trends of production flow, distribution flow, and level in the clear water reservoir on a day with level based flow control

A model predictive flow control algorithm (see Figure 4.2) consists of a short-term water demand forecasting algorithm and a control algorithm. For production flow control, the forecasting horizon is typically 24 to 48 hours (Bakker et al., 2003). The control algorithm calculates production flow set-points that match the forecasted demand, under the condition that the level in the reservoir stays between a chosen upper and lower limit. In general, the upper and lower limits in the reservoir are set as static values, but these may also be optimised to e.g. aiming to maximise the chlorine residual in the clear water. The control algorithm can be configured to optimise various optimisation goals, such as minimal changes in the production flow, minimal energy use, minimal energy costs or a combination of those. In most cases mathematical optimisation techniques are applied in the control algorithm in order to find an optimum. However, if the optimisation goal is formulated strictly (e.g. constant production flow), the optimum can be calculated directly by a deterministic model (Bakker et al., 2003). Like level based flow control, the set-point can be either discrete switching commands for wells, filters or pumps, or a continuous value. Predictive flow control may be applied to achieve a constant production flow, where the reservoir is used to balance the fluctuations in the outgoing distribution flow (see Figure 4.2).

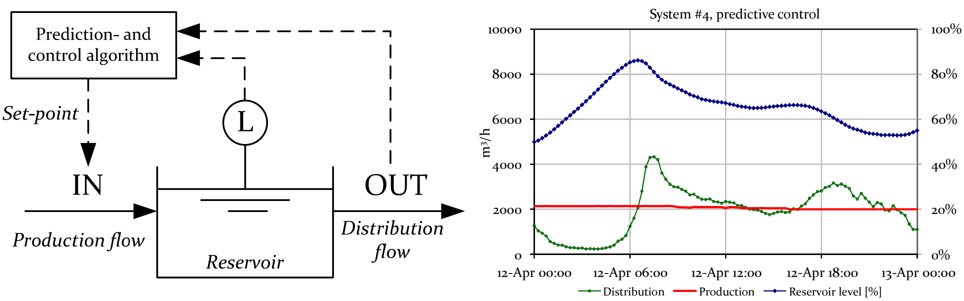


Figure 4.2 Principles of model predictive flow control and trends of production flow, distribution flow, and level in the clear water reservoir on a day with model predictive flow control

4.1.3 Application of predictive flow control in the Netherlands

All ten Dutch water supply companies were interviewed in 2011 to determine the penetration of predictive flow control in the Netherlands. The result of the interviews is that at present 57% of the total production flow is controlled by predictive flow control. The production flow of the other 43% of the systems is controlled by level based flow control, see Table 4.1.

Table 4.1 Application of predictive control (MPC) for the production flow control at water supply companies in the Netherlands, based on interviews at the water supply companies (summer 2012)

Company	Total production [million m^3 in 2009]	MPC controlled production [million m^3 per year]	Fraction
Brabant Water	167	27	16%
Dunea	69	69	100%
Evides	176	176	100%
Oasen	46	43	93%
PWN	100	95	95%
Vitens	329	68	21%
Waternet	65	65	100%
WBG	42	20	48%
WMD	28	0	0%
WML	72	62	86%
Total	1,094	625	57%

4.1.4 Objectives for predictive flow control

There are two main reasons to apply predictive flow control rather than level based flow control. The first reason is that drinking water treatment plants perform better at a constant production flow rate. Keuning et al. (1998) reported 20% lower values for turbidity and total hardness at a drinking water treatment plant, after the production flow control was changed from level based control to predictive control. In studies of Gauthier et al. (2001), Vreeburg et al. (2004) and Vreeburg et al. (2008) it was observed that the major part of the particle load in drinking water occurs as a result of peak flow and start-up procedures at the treatment plant. The occurrence of particles in the distribution network leads to discolouration events, which in England and Wales account for 80% of all customer complaints about drinking water quality (Husband and Boxall, 2011). Particles in drinking water are therefore a dominant factor in the customer satisfaction regarding water supply.

The second reason to apply predictive flow control is the reduction of energy consumption and energy costs. Since the late 1980's the near optimal control of water supply systems (pump scheduling of distribution pumps) to reduce energy consumption and costs, is a topic that has been studied by many researchers. Ormsbee and Lansey (1994) and Brdys and

Ulanicki (1994) give good overviews of algorithms developed until 1994 and report several successful implementations in Europe and Israel. More recent publications present newly developed optimisation algorithms with new mathematical optimizing techniques (Rao et al., 2007; Martínez et al., 2007; Salomons et al., 2007; Jamieson et al., 2007; Ulanicki et al., 2007; Shamir and Salomons, 2008; Cembrano et al., 2011; Savić et al., 2011).

The dominant reason for most water supply companies in the Netherlands to implement predictive control, was the wish for a more constant operation of the water treatment plants in order to get lower turbidity values. The implemented predictive control algorithms in the Netherlands therefore focus mainly on constant production flow set-points. Energy savings and energy cost savings are an important, but less dominant, second reason for the implementation of predictive control. Note that a constant production flow is generally not the optimal strategy to minimise the energy costs. Although predictive flow control is widely applied in the Netherlands, the effectiveness of this way of control has never been studied in detail. This chapter describes the results of research that was carried out to quantify the differences between level based flow control and predictive flow control at five full scale water supply systems in the Netherlands. This research considers both water quality aspects, and energy consumption and costs.

4.2 Materials and methods

4.2.1 Experiments at five full scale water supply systems

Five full scale water supply systems were examined in the first half of 2011. Under normal operating circumstances, the selected systems are controlled with the predictive flow control model OPIR (Bakker et al., 2003). This control software runs on a Windows platform and connects to the existing SCADA system to read real-time measurements and send set-points. In the experiments the predictive flow control was switched off for one week, in which the systems were controlled with level based flow control loops. The predictive flow control model aims at constant production set-points, and controls both the production flow of the treatment plants, and the intake and distribution flows of the service reservoirs of the water supply system. The characteristics and the configurations of the systems are shown in Figure 4.3. The research comprised of examining the behaviour of the systems, during:

- One week with predictive flow control.
- One week with level based flow control.

During both periods the same (static) process constraints, like minimum/maximum allowed reservoir levels, minimum/maximum flows applied.

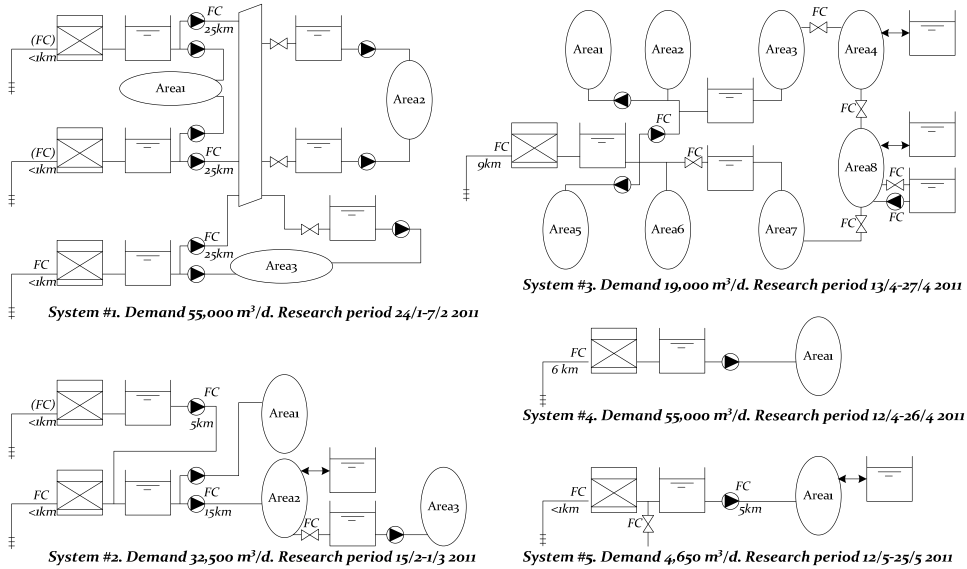


Figure 4.3 Configurations of the five examined water supply systems. The flow controlled elements are indicated by “FC”. The elements indicated with “(FC)” are always level based flow controlled (also when predictive flow control was active).

4.2.2 Evaluated control methods

Level based flow control

The conventional level based control method was initially installed for the normal operation at all five systems. The parameters for this control method (switching levels and production capacities) were chosen by the operators to ensure that the reservoir levels would not drop undesirably low during all demand situations. Figure 4.5 gives an impression of operational results (reservoir levels and production flow) with the chosen settings.

Model predictive flow control

The predictive control method (named OPIR) was installed later at all five systems to enhance the production and transportation flow control. The control method uses the adaptive short-term water demand forecasting model that was described in chapter 2 of this thesis (Bakker et al., 2013). The water demand in each supply area of the systems is forecasted, and translated into the total outflow from each reservoir. OPIR’s control method calculates the set-point for the inflow into the reservoir. The (optimisation) goal is to keep this inflow as constant as possible, for achieving the optimal drinking water quality. The constraints for the control method are the minimum and maximum allowed reservoir levels that can be set by the operators.

The method executes the following calculations to derive the inflow (production) set-point:

1. The forecasted total outflow from the clear water reservoir is derived from the forecasted water demands in the areas.
2. This forecasted outflow is drawn as two cumulative series, where:
    a. The upper series is the cumulative forecast plus the buffer volume between the actual level and the maximum allowed level. This line depicts the maximum allowed level in the buffer.
    b. The lower series is the cumulative forecast minus the buffer volume between the actual level and the minimum allowed level. This line depicts the minimum allowed level in the buffer.
3. The possible inflows are drawn as cumulative series (the method always calculates with discrete flows; if the inflow can be controlled continuously, the method internally calculates with small discrete flow steps).
4. The intersection of a possible inflow with the upper / lower cumulative forecast series, depicts the time frame that the inflow can be kept constant without violating the maximum / minimum allowed levels.
5. The inflow that does not violate the maximum / minimum allowed levels for the longest time frame is selected as the set-point; if the current inflow does not violate the levels at all, then the current inflow is maintained as set-point.

Figure 4.4 shows the forecasted outflow (upper graph), the transformation to the minimum and maximum allowed levels, and the evaluation of the possible inflows (lower graph). The selected inflow (production flow), is the inflow that results in a minimum number of flow changes.

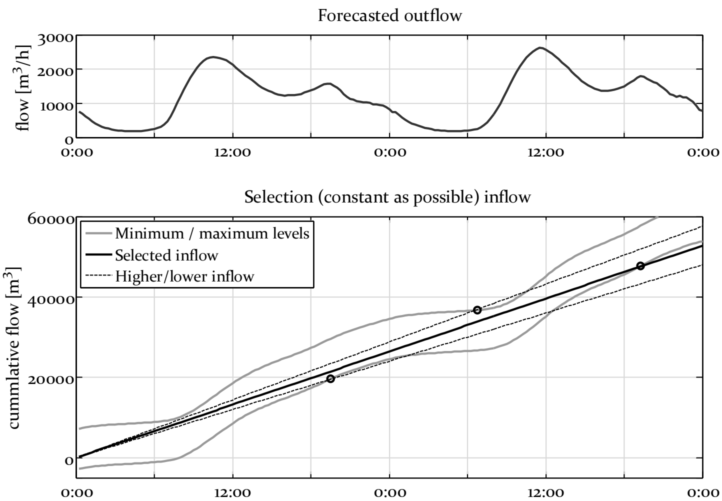


Figure 4.4 Selection of the desired inflow by transforming the forecasted outflow

4.2.3 Water quality and production flow variation

The following water quality parameters were measured (sensors in the clear water main at the treatment plants):

- Turbidity (all systems), measured by Hach Lange 1720 turbidimeter / Endress+Hauser CUR22 turbidimeter.
- Particles (systems #1 and #4), measured by Pamas Waterviewer system.

At all five systems turbidity measuring devices were installed. The turbidity rate gives a good indication of the load of suspended solids in the water (Low et al., 2011). At systems #1 and #4 particle measuring devices were installed. Measurements of particle numbers and sizes give insight in the particle load of the clear water. The measured numbers and sizes of the particles are transformed in a “particle volume concentration” in the units parts per billion (10^{-9} m^3/m^3) as described by Vreeburg et al. (2008).

To assess the variability of the production flow, the Production Variation per day (PV_d) is defined as the sum of the (absolute values of) the difference between subsequent hourly average production flow values ($Q_{prod,d,h}$) divided by the total daily production:

$$PV_d = \frac{\sum_{h=1}^{h=24} \left| Q_{prod,d,h} - Q_{prod,d,h-1} \right|}{\sum_{h=1}^{h=24} Q_{prod,d,h}} \cdot 100\% \qquad (4.1)$$

A value of 10% indicates that on average the production flow changes on each hour with 10% of the average production flow of that day (d).

4.2.4 Energy consumption and costs

The specific energy consumption (kWh/m^3) and the percentage of the energy consumption during high tariff hours were analysed. In the Netherlands, the high tariff applies to each weekday from 7 AM to 11 PM, the low tariff applies to all the other hours and in the weekends. At all researched systems one continuously measuring electricity meter was available at each water treatment plant and each service reservoir. Using measurements of flow, pressure and reservoir level, the measured energy consumption was divided in three main components: 1. Abstraction/ treatment (including base consumption), 2. Transportation / distribution (clear water pumped in a transport main or towards a high reservoir) 3. Direct boosting (clear water pumped to consumers in an area without a high reservoir):

$$E_{abst} = \frac{1}{\eta_{abst}} \cdot \frac{Q_{abst}}{3600} \cdot \rho \cdot g \cdot \left(dH_{stat,abst} + C_{dyn,abst} \cdot Q_{abst}^2 \right) \cdot \frac{1}{1000} \quad [kW] \tag{4.2}$$

$$E_{pump} = \frac{1}{\eta_{pump}} \cdot \frac{Q_{pump}}{3600} \cdot \rho \cdot g \cdot (H_{pump} - L_{res}) \cdot \frac{1}{1000} \quad [kW] \tag{4.3}$$

$$E_{boost} = \frac{1}{\eta_{boost}} \cdot \frac{Q_{boost}}{3600} \cdot \rho \cdot g \cdot (H_{boost} - L_{res}) \cdot \frac{1}{1000} \quad [kW] \tag{4.4}$$

E_{abst}, E_{pump} and E_{boost} are the calculated values of the energy consumption [kW] for abstraction / treatment, transportation / distribution, and direct boosting respectively. The flows (Q_{abst}, Q_{pump}, Q_{boost} [m^3/h]), pump head (H_{pump}, H_{boost} [m]) and the reservoir level (L_{res} [m]) are measured at the treatment plants and service reservoirs. The values for efficiency (η_{abst}, η_{pump}, η_{boost} [-]), the static and dynamic head loss parameters ($dH_{stat,abst}$ [m] and $C_{dyn,abst}$ [m / $(m^3/h)^2$] respectively), and the constant base energy consumption for heating/cooling, controllers, and other non-process related use (E_{base} [kW]) were estimated.

4.3 Results

4.3.1 Comparison

The differences between level based control and predictive control were quantified by comparing average values of the measured parameters of both researched periods. Examples of the differences are shown in Figure 4.5, and the results for all water supply systems are summarised in Table 4.2. Note that at some of the graphs, the reservoir level at the end of the day did not return to the original value at the beginning of the day. However, the graphs show only one day of the evaluated seven days, and the level differences on one day will be averaged out over the evaluated days in the seven days period. The level differences will therefore not influence the results.

The graphs show that predictive flow leads to a lower variation in the production flow and a higher production flow rate at night (during low energy tariff) compared to level based control. The observed differences in flow patterns and the use of reservoirs during level based control and predictive control were distinct for each of the five water supply systems.

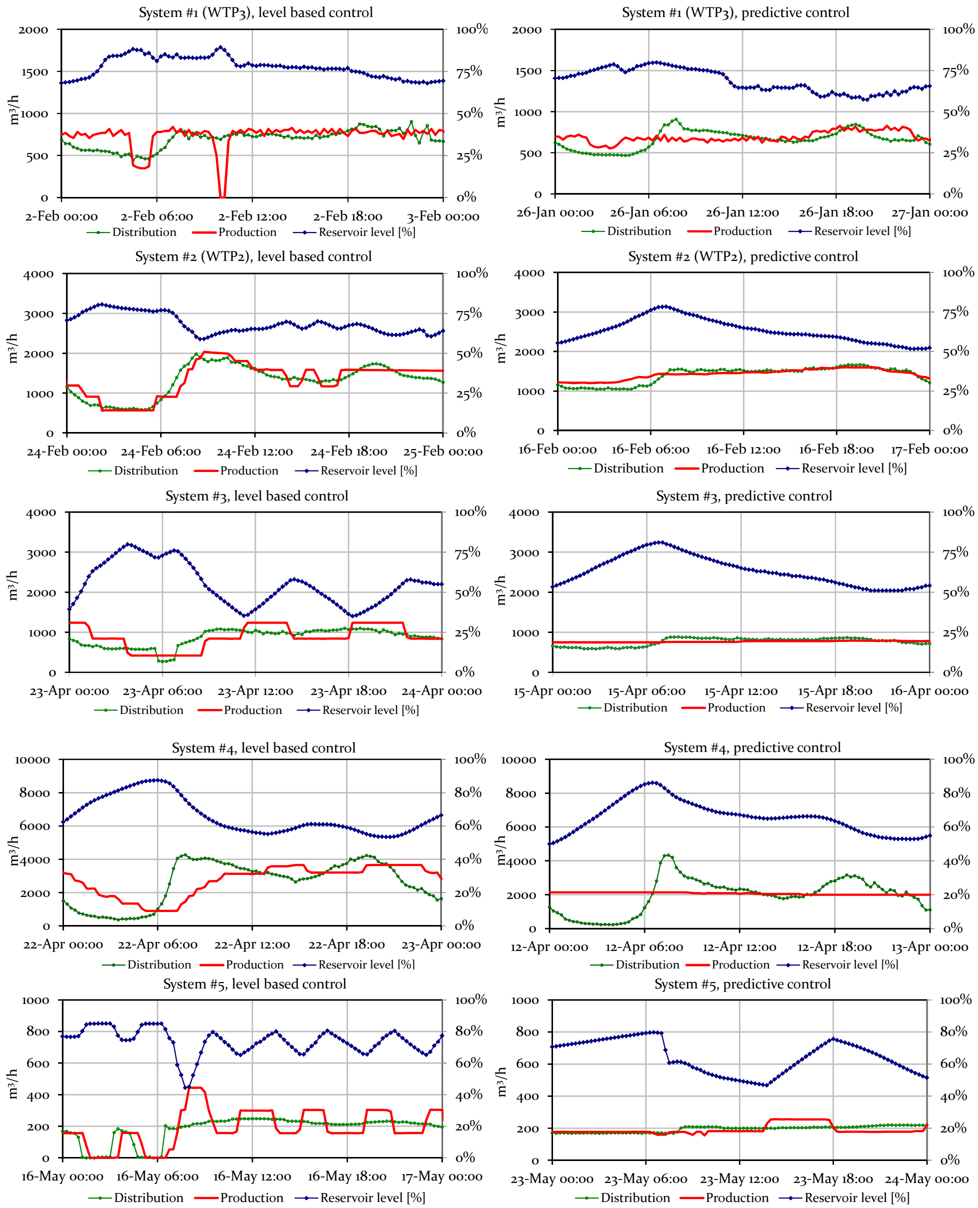


Figure 4.5 One-day example trends of level based control (left) and model predictive control (right)

Table 4.2 Differences between level based control and predictive control

	Level based control	Predictive control	Difference
System #1			
Production Variation [%]	11.7%	4.1%	-65%
Min. / max production flow [m^3/h]	0 / 840	408 / 808	-52%
Turbidity [NTU]	0.398	0.345	-13%
Particle load [ppb]	30.8	17.9	-42%
Specific energy consumption [kWh/m^3]	0.340	0.336	-1.0%
Energy use at high tariff [%]	51.7%	51.0%	-1.4%
Energy costs [€ per 1,000 m^3]	€ 22.47	€ 22.09	-1.7%
System #2			
Production Variation [%]	14.2%	2.9%	-80%
Min. / max production flow [m^3/h]	186 / 921	356 / 681	-56%
Turbidity [NTU]	0.809	0.654	-19%
Specific energy consumption [kWh/m^3]	0.733	0.694	-5.3%
Energy use at high tariff [%]	55.5%	51.2%	-7.7%
Energy costs [€ per 1,000 m^3]	€ 49.30	€ 45.65	-7.4%
System #3			
Production Variation [%]	12.5%	1.2%	-90%
Min. / max production flow [m^3/h]	250 / 1260	430 / 933	-50%
Turbidity [NTU]	0.078	0.056	-28%
Specific energy consumption [kWh/m^3]	0.605	0.587	-3.0%
Energy use at high tariff [%]	51.0%	48.6%	-4.8%
Energy costs [€ per 1,000 m^3]	€ 39.77	€ 38.10	-4.2%
System #4			
Production Variation [%]	11.7%	2.8%	-76%
Min. / max production flow [m^3/h]	208 / 1729	755 / 1560	-47%
Turbidity [NTU]	0.044	0.039	-12%
Particle load [ppb]	5.43	4.78	-12%
Specific energy consumption [kWh/m^3]	0.329	0.324	-1.4%
Energy use at high tariff [%]	58.3%	54.5%	-6.5%
Energy costs [€ per 1,000 m^3]	€ 22.38	€ 21.67	-3.2%
System #5			
Production Variation [%]	37.7%	7.6%	-80%
Min. / max production flow [m^3/h]	0 / 473	53 / 264	-55%
Turbidity [NTU]	0.313	0.273	-13%
Specific energy consumption [kWh/m^3]	0.399	0.389	-2.5%
Energy use at high tariff [%]	57.3%	48.2%	-15.9%
Energy costs [€ per 1,000 m^3]	€ 27.19	€ 25.29	-7.0%

4.3.2 Water quality

Figure 4.6 shows the relation between production flow changes and turbidity at all locations, and the relation between production flow changes and particle volume at systems #1 and #4.

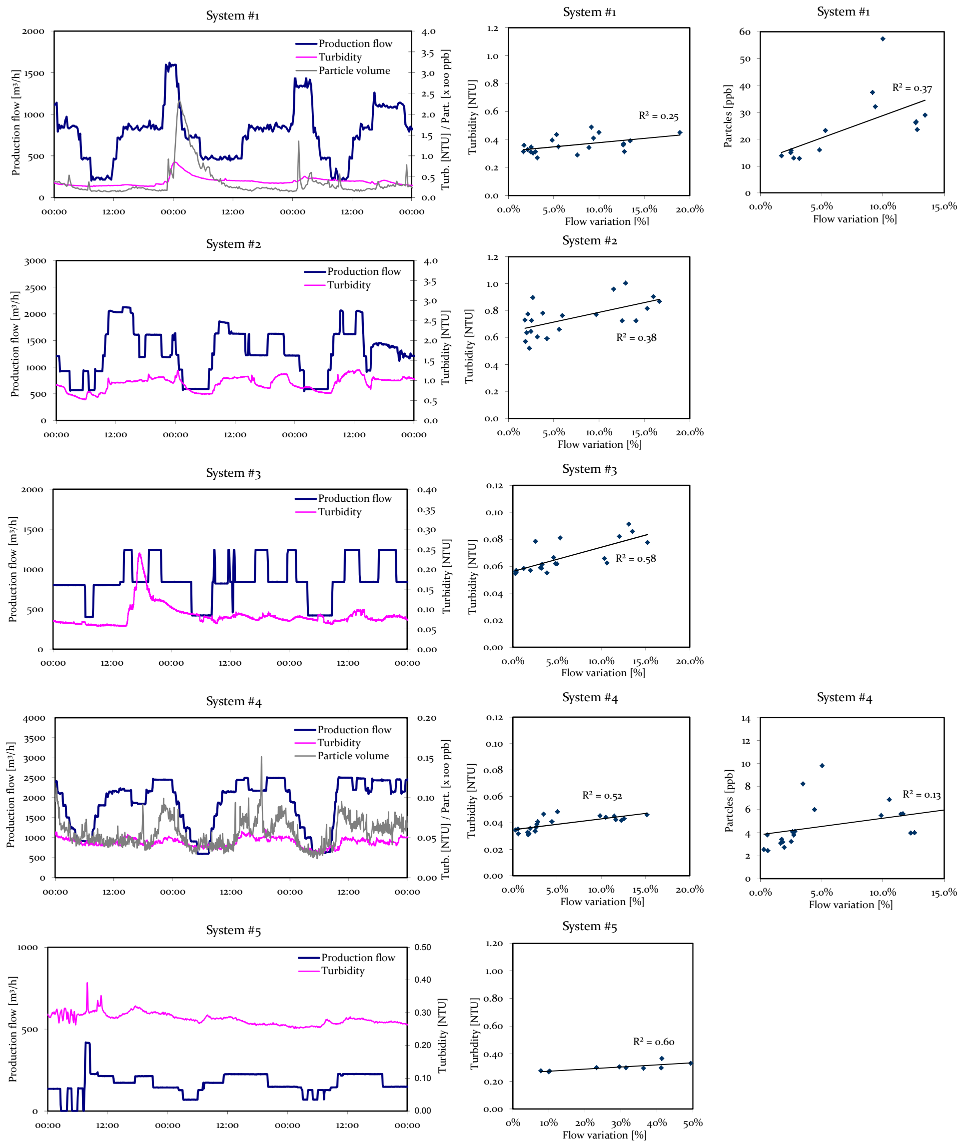


Figure 4.6 Relation between production flow variations and turbidity / particle volume in clear water. The graphs at the left show the turbidity and the flow variations in time. The graphs at the right show the relation between average daily values of flow variations and average daily values of turbidity (middle graphs) and particle volume (right graphs).

At all systems a relation was found between flow changes and turbidity, although the correlation was weak (R^2 values between 0.25 and 0.6). Especially at systems #1 and #2 and #5 lower turbidity rates are valuable, because the rates were quite high at those systems. The experiments at the systems with particle counters installed (system #1 and #4), showed that the values of particle volumes were 12% to 42% lower with predictive control compared to level based control. The graphs of system #1 and system #4 in Figure 4.6 show a strong response of particle volume to production flow variations (high increase of particle volume after flow increase). However the correlation of average values per day between production flow variation and particle volume is weak (R^2 values 0.13 and 0.37).

Table 4.2 shows the average values of the experiments. The table shows that the Production Variation with predictive control (1.2%-7.6%) was lower than with level based control (11.7%-37.7%). This resulted in turbidity values that were 12%-28% lower (average 17%) at all five systems.

4.3.3 Energy consumption

The variables of equations (4.2), (4.3), and (4.4) were estimated in a way that that the calculated energy consumption best fitted the measured energy consumption. By doing this, the total measured energy consumption was assigned to the individual "components" of energy consumption, making it possible to evaluate the effect of the control on these components. Figure 4.7 gives an example of a trend with both the measured energy consumption and the calculated energy consumption. The figure shows that measured and calculated values resemble well (R^2 = 0.92), which indicates that the parameters were chosen well.

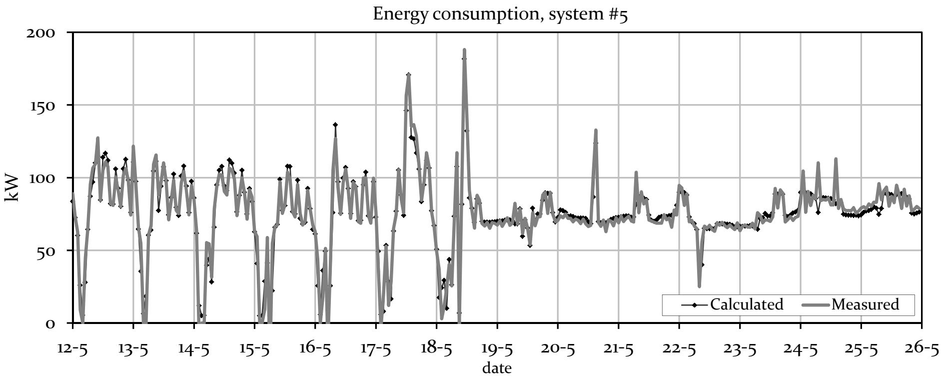


Figure 4.7 Example of trend with calculated energy consumption and measured energy consumption

Figure 4.8 shows the average division of the total energy consumption over the three components as observed during the experiments. Energy costs were calculated by multiplying the calculated energy consumption with the average energy costs per kWh. The tariffs were 0.08229 €/kWh during high tariff hours, and 0.04849 €/kWh during low tariff hours.

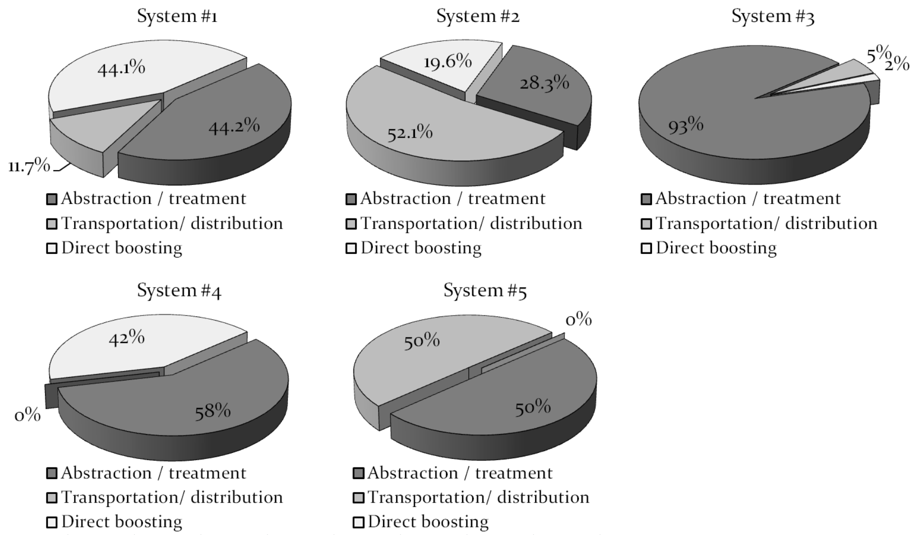


Figure 4.8 Division of total energy consumption over the three components of energy consumption

Table 4.2 shows that the average specific energy consumption is 1.0% to 5.3% lower for predictive control compared to level based control. This is the result of the fact that with level based control the difference between minimum and maximum production flows is larger (see Table 4.2). The specific energy consumption [kWh/m^3] at high flows is relatively higher, because of the dynamic head loss components in the abstraction and treatment process, and in the transportation and distribution process. As a result the average specific energy consumption is higher at varying flow rates. The hydraulic head loss occurs predominantly if the water is pumped over longer distances between abstraction and treatment plant, or for transportation. The distances in the examined water supply systems are indicated in Figure 4.3. Figure 4.9 shows the difference in specific energy consumption for the distinguished components of energy consumption.

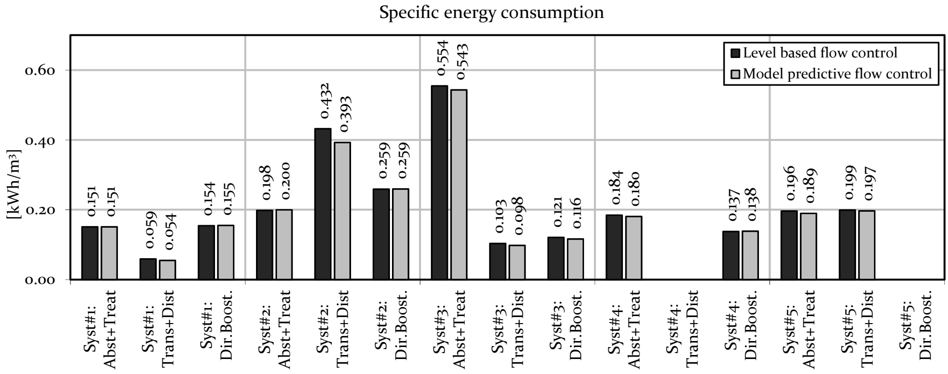


Figure 4.9 Specific energy consumption per component of energy consumption for each system

A second aspect is a shift of energy consumption from high tariff to low tariff hours at predictive control compared to level based control (shift varying from 1.4% to 15.9%). This is caused by the fact that with level based control the reservoirs are filled with too high flow rates in the evening and night. As a result the reservoir level becomes high early in the night, and the level based control decreases the production flow. As a consequence less water is produced or transported and therefore less energy is consumed in the period with low energy tariff (see also Figure 4.1 and Figure 4.5). Figure 4.10 shows the shift in energy consumption for the different components of energy consumption for all five examined systems. The figure shows that the shift occurs especially in the Transportation and Distribution component, and to a lesser extent in Abstraction and Treatment. No shift is observed in the Direct Boosting component, which is the consequence of the fact that this component is pressure controlled and not influenced by the researched flow control.

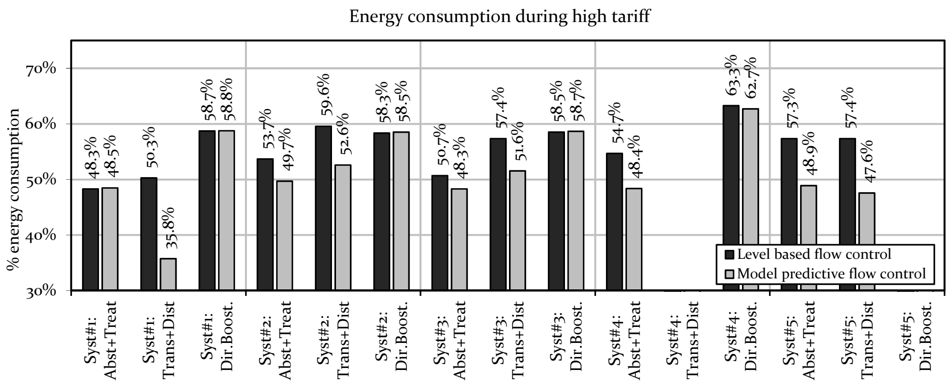


Figure 4.10 Percentage of energy consumption during high tariff per component of energy consumption for each system

The combination of the lower specific energy consumption and the shift from high tariff to low tariff, results in lower energy costs of 1.7% to 7.4% for predictive control compared to level based control. The results show a relatively large variation between the five examined water supply systems. This variation was caused by the fact that the five systems have quite different configurations, as can be seen in Figure 4.3. As observed in Figure 4.8 there are also large differences in the way the total energy consumption is divided over the components. The presented differences in Table 4.2 are related to the energy consumption of all three components, including the energy consumed for (pressure controlled) direct boosting. Direct boosting is not influenced by the control, and therefore the numbers of Table 4.2 (of specific energy consumption, energy use during high tariff, and energy costs) are not fully representative for the real difference between the control of the systems. Table 4.3 shows for each of the water supply systems which part of the energy consumption is influenced by the way of control, and what the differences in energy costs are related to the influenced energy consumption. The table shows that the observed differences between predictive control and level based control are relatively smaller when the differences are related to the influenced energy consumption.

Table 4.3 Energy consumption influenced by the type of control, and differences in energy costs. (Note that the influenced energy consumption in system #1 is rather low, because the abstraction and treatment flow control of 2 of the 3 treatment plants are always level based controlled).

	Energy consumption influenced by control	**Difference in energy costs between predictive control compared to level based control**	
		related to total energy consumption	**related to influenced energy consumption**
System #1	26.3%	-1.7%	-6.4 %
System #2	80.4%	-7.4%	-9.2 %
System #3	98.0%	-4.2%	-4.3 %
System #4	57.9%	-3.2%	-5.5 %
System #5	100.0%	-7.0%	-7.0 %

4.4 Discussion

4.4.1 Relation Production Variation – Turbidity

In comparison to the period with level based control, in the period with predictive control at all five systems lower turbidity values (and particle volumes), and lower Production Variation values were observed. However, the relation between Production Variation and turbidity is weak (see Figure 4.6, R^2 values between 0.25 and 0.6). This indicates that the Production Variation is not the only factor which influences the turbidity. Other disturbances in the treatment process, (like filter backwashing, see Vreeburg et al. (2004), Bakker et al. (1998)) might also be responsible for variations in the turbidity. However, backwashing and other disturbing events were not monitored in this study.

4.4.2 Limited energy savings

The observed energy cost savings in this study (5.2% on average) were smaller than reported energy cost savings in other studies, like Bunn and Reynolds (2009): 12% measured savings; or Martínez et al. (2007): 17% simulated savings. This difference can be explained by a number of aspects. The effectiveness of changing the control of a water supply system depends on four main factors: 1- the effectiveness of the optimal control in relation to the existing control; 2- flexibility in the system to change the control; 3- influence of control on operational costs; 4- the optimisation goal.

In the first aspect the existing control plays a dominant role: if the existing control is rather bad, than the potential savings can be quite large. However, if the existing control is already rather good, the potential savings are limited. The second aspect relates to the flexibility of the system. If there are large buffers in the network and pump and treatment capacity largely exceed the average demands, a large range of operational control strategies is possible without violating the constraints. Among the very different control strategies the most efficient can be selected by the optimal control algorithm. In systems with less flexibility, the range of possible control strategies is smaller, and therefore the potential savings are smaller. The third aspect relates to how much the energy consumption and costs are influenced by the control. If pump efficiencies or pump heads change dramatically depending on the control of the system, or if there are large differences in energy tariffs, potential savings can be very high. In systems where the control hardly influences energy consumption and where tariff differences are small, the potential savings are much smaller. The final aspect relates to the explicit optimisation goal: if one aspect is targeted as the main goal, the results at other aspects will be less compelling.

In this study there was quite a large difference between the examined controls (aspect 1). The flexibility of the systems was limited due to relative small buffers in the system (aspect 2), which also limited the energy savings. The dominant factor that the energy savings in this study were limited, is that the control did not influence the energy to a large extent (aspect 3). This is quite different from other studies, where the water supply systems have large high reservoirs (which can be filled during low energy tariff) and where pump efficiencies are highly influenced by the control strategy. Bunn and Reynolds (2009) show examples where the pump efficiencies vary up to 20%, which offers the possibility to achieve large reductions of energy consumption. The optimisation goal (aspect 4.) of the applied optimising control method was stabilizing the production flow to obtain an optimal water quality. The energy cost savings were a valuable but less important side effect and not an explicit target of the optimisation. This also explains why the energy cost savings observed in this chapter were smaller than the savings reported in other papers.

4.4.3 Other aspects

The experiments were carried out in a relative short time and a limited number of parameters were studied. Therefore this study describes not all differences between the level based control and the predictive control of a water supply system. Some difference can only be measured over a longer period of time. Two aspects which were not studied in this chapter are wear of pumps and valves, and the occurrence of process alarms and alerts. Both aspects are reported by Keuning et al. (1998), who studied the effects of the implementation of predictive control at one location over a longer period. Keuning reported that predictive control resulted in less wear, because less variation in the operation, and therefore less starts and stops of pumps occurred. The value of the Production Variation in Table 4.2 is a measure for the number of starts/stops of pumps. As can be observed in the table, the values for level based control were some 3-8 times higher than for predictive control. Keuning also reported that failures and alarms occurred less frequent when using predictive control, because processes were switched on and off less often. At each switch there is a small risk of a failure in the installation, resulting in a process alert or alarm.

Another aspect is the ease of operation. Process operators of system #1 and system #3 stated that the most valuable aspect of predictive control for them was the ease of operation. With level based control the operation required much more attention, especially during high demand periods. The predictive control was better able to cope with changing demand situations and to adjust the control accordingly.

4.5 Conclusions

Experiments at five full scale water supply systems showed that predictive control leads to a better water quality and a more energy efficient water supply compared to level based control:

- The Production Variation was 3-8 times lower.
- Turbidity values were 12-28% lower.
- Particle volume values were 12-42% lower.
- The overall energy consumption was 1.0-5.3% lower.
- The overall energy costs were 1.7-7.4% lower.

The quality improvements were the result of the fact that the variations in the production flow were on average some 3-8 times lower with predictive control compared to level based control. Variations in production flow resulted in peaks in the turbidity values of the clear water. The observed higher energy efficiency was the result of more constant production flows which led to a lower average energy consumption. Moreover relatively less energy was consumed during high tariff, resulting in a lower energy bill. Both quality improvements and higher energy efficiency make predictive control a valuable asset for water supply companies.

Acknowledgements
We thank the water companies WML, Vitens and Brabant Water for changing the operation for one week, and providing us the data that was used in this research.

4.6 References

Bakker, M., Verberne, A.J.P., and van Schagen, K.M. (1998). "The benefits of demand forecasting and modelling". *Water Quality International.* (May-June): 20-22.
Bakker, M., Van Schagen, K., and Timmer, J. (2003). "Flow control by prediction of water demand". *Journal of Water Supply: Research and Technology - AQUA.* 52 (6): 417-424.
Bakker, M., Vreeburg, J.H.G., van Schagen, K.M., and Rietveld, L.C. (2013). "A fully adaptive forecasting model for short-term drinking water demand". *Environmental Modelling and Software.* 48 (1): 141-151.
Brdys, M.A., and Ulanicki, B. (1994). "Operational control of water systems: Structures, algorithms and applications". United Kingdom: Prentice-Hall International.
Bunn, S. (2007). "Closing the loop in water supply optimisation", *Proceedings IET Water Event*, Coventry, United Kingdom.
Bunn, S.M., and Reynolds, L. (2009). "The energy-efficiency benefits of pumpscheduling optimization for potable water supplies". *IBM Journal of Research and Development.* 53 (3).

Cembrano, G., Quevedo, J., Puig, V., Pérez, R., Figueras, J., Verdejo, J.M., Escaler, I., Ramón, G., Barnet, G., Rodríguez, P., and Casas, M. (2011). "PLIO: A generic tool for real-time operational predictive optimal control of water networks". *Water Science and Technology*. 64 (2): 448-459.

Gauthier, V., Barbeau, B., Millette, R., Block, J.C., and Prévost, M. (2001). "Suspended particles in the drinking water of two distribution systems". *Water Science and Technology: Water Supply*. 1 (4): 237-245.

Husband, P.S., and Boxall, J.B. (2011). "Asset deterioration and discolouration in water distribution systems". *Water Research*. 45 (1): 113-124.

Jamieson, D.G., Shamir, U., Martinez, F., and Franchini, M. (2007). "Conceptual design of a generic, real-time, near-optimal control system for water-distribution networks". *Journal of Hydroinformatics*. 9 (1): 3-14.

Keuning, J., Mense, P., Van Schagen, K.M., and De Moel, P.J. (1998). "Hardheidsverlaging door middel van productiesturing (in Dutch)". *H2O*. 19: 73-75.

Low, H.X.D., Handojo, D.U., and Lim, Z.H.K. (2011). "Correlation between Turbidity and Total Suspended Solids in Singapore Rivers". *Journal of Water Sustainability*. 1 (3): 313-322.

Martínez, F., Hernández, V., Alonso, J.M., Rao, Z., and Alvisi, S. (2007). "Optimizing the operation of the Valencia water-distribution network". *Journal of Hydroinformatics*. 9 (1): 65-78.

Ormsbee, L.E., and Lansey, K.E. (1994). "Optimal control of water-supply pumping systems". *Journal of Water Resources Planning & Management - ASCE*. 120 (2): 237-252.

Rao, Z.F., Wicks, J., and West, S. (2007). "Optimising water supply and distribution operations". *Proceedings of the Institution of Civil Engineers: Water Management*. 160 (2): 95-101.

Salomons, E., Goryashko, A., Shamir, U., Rao, Z., and Alvisi, S. (2007). "Optimizing the operation of the Haifa-A water-distribution network". *Journal of Hydroinformatics*. 9 (1): 51-64.

Savić, D.A., Bicik, J., and Morley, M.S. (2011). "A DSS generator for multiobjective optimisation of spreadsheet-based models". *Environmental Modelling and Software*. 26 (5): 551-561.

Shamir, U., and Salomons, E. (2008). "Optimal real-time operation of urban water distribution systems using reduced models". *Journal of Water Resources Planning and Management*. 134 (2): 181-185.

Ulanicki, B., Kahler, J., and See, H. (2007). "Dynamic optimization approach for solving an optimal scheduling problem in water distribution systems". *Journal of Water Resources Planning and Management*. 133 (1): 23-32.

Vreeburg, J.H.G., Schaap, P.G., and Van Dijk, J.C. (2004). "Particles in the drinking water system: From source to discolouration". *Water Science and Technology: Water Supply*. 4 (5-6): 431-438.

Vreeburg, J.H.G., Schippers, D., Verberk, J.Q.J.C., and van Dijk, J.C. (2008). "Impact of particles on sediment accumulation in a drinking water distribution system". *Water Research*. 42 (16): 4233-4242.

5 Advanced control of a water supply system: a case study

Based on

M. Bakker
T. Rajewicz
H. Kien
J.H.G. Vreeburg
L.C. Rietveld

Advanced control of a water supply system: a case study
Water Practice & Technology (2014). 9 (2): 264-276

Abstract
Conventional automatic production flow control and pump pressure control of water supply systems are robust and simple: production flow is controlled based on the level in the clear water reservoir and pump pressure is controlled on a static set-point. Recently, more advanced computer-based control methods were developed in which production flow is controlled by using a short-term water demand forecasting model and pressure is controlled by a dynamic pressure control module. To show the differences between conventional and advanced control, this chapter presents operational data of water treatment plant (WTP) Gruszczyn that supplies drinking water to a part of the city of Poznań, Poland. A three weeks period of conventional is compared with a three weeks period of advanced control. The comparison shows that with advanced control the variation in the production flow is 83% lower, and the pump pressure of the clear water pumps is 29% lower. The lower pressure results in 18% less background leakage and the overall system's energy costs are 11.5% lower.

Keywords
Water demand forecasting; optimal control; pressure management

5.1 Introduction

A water supply system is designed to produce drinking water of good quality, and to supply this water to the consumers under sufficient pressure. The goal in the operation of the system is to produce and supply the drinking water with a high reliability at the lowest operational costs. Initially, the water supply systems were operated manually by operators, but since the mid 1970's water utilities started automating the systems (Bunn, 2007). The control loops were rather straightforward because of the limited computational force of the automation systems. This simple and robust automation is sub-optimal with respect to the performance of the treatment plant and energy efficiency (Bakker et al., 2003). In the meantime, there is an on-going trend towards the fully automated (centralised) operation of water supply systems (Worm et al., 2010; PWN, 2006): local control rooms are closed, and one central, highly automated, control room is created. When utilities implement modern centralised automatic control, they aim to reduce costs and, at the same time, improving the quality of the operations.

A possibility to achieve this goal is to apply a short-term water demand forecasting model for optimal production flow control or optimal pump scheduling. Forecasting models are used by utilities around the world: In the Netherlands in 2012, 57% of all supplied water was controlled with predictive control models (see chapter 4 of this thesis (Bakker et al., 2013)), leading to 5.2% lower energy costs and 17% lower turbidity of the clear water. Other examples of the implementation of predictive control are at four large utilities in the United States (Bunn and Reynolds, 2009), where a reduction of energy costs of 12% was achieved. Another possibility to improve the operation of water supply systems is the implementation of pressure management. In most cases, implementing pressure management includes both creating smaller pressure zones, called district metered areas (DMAs), and installing pressure reducing valves (PRVs) in the distribution network. Pressure management can lead to a reduction of the water losses. Girard and Stewart (2007) showed in their case study a reduction of the water losses of 21% as a result of reducing the pressure in the DMAs. Pressure management, including the reduction of water losses, also leads to a reduction of the system's energy consumption (Colombo and Karney, 2005).

In this chapter, we present a case study of the implementation of advanced control software that aims to combine the two control strategies mentioned above. The software controls both the production flow based on an adaptive water demand forecasting model, and the pump pressure by applying flow depending (dynamic) pressure control.

5.2 Materials and methods

5.2.1 Case study

Water company Aquanet S.A. is responsible for the water supply in the city of Poznań (550,000 inhabitants), in the central western part of Poland. Like most water supply companies in Poland, Aquanet manually operated the water treatment and pumping facilities by operators. In 2011 Aquanet decided to fully automate the control of the system of water treatment plant (WTP) Gruszczyn and run this system unmanned. The aim was to optimise the control of the system, and to achieve a better water quality and a reduction of the operational costs by minimizing the energy consumption. The lay-out of the system is shown in Figure 5.1 and Figure 5.2.

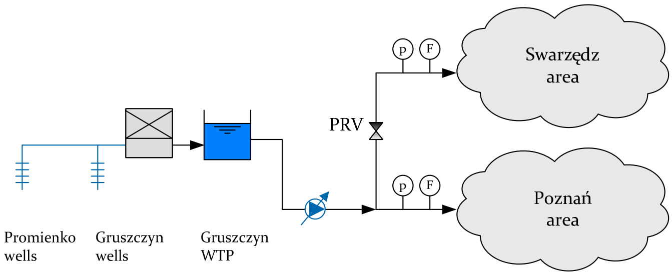


Figure 5.1 Schematic drawing of the water supply system of Gruszczyn

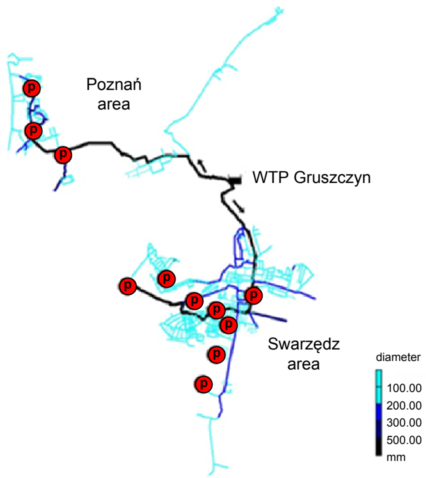


Figure 5.2 Distribution system of Gruszczyn, including nine new and two existing pressure measuring points

5.2.2 Production flow control

At first, a relatively simple level based production flow control loop was programmed in the programmable logic controller (PLC). In this control loop the production flow set-point was directly derived from the level in the clear water reservoir. An increasing level resulted in a decreasing set-point, and vice versa. This level based production flow control was capable of controlling the production unmanned. However, many production flow changes occurred, which resulted in energy inefficiency and sub-optimal water quality.

Influence of production flow control

Production flow control influences the variability of the production flow, which can be expressed in the production variation per day (PV_d, see equation (4.1) in the previous chapter). Production flow control also influences the energy consumption for the production of the drinking water. Because real-time energy measurements were not available, the energy consumption for abstraction and treatment (P_{prod}) was estimated with:

$$P_{prod} = P_{base} + C_1 \cdot Q_{prod} + C_2 \cdot Q^3{}_{prod} \qquad [kW] \tag{5.1}$$

where P_{base} [kW] is the constant, flow independent, energy consumption. Q_{prod} [m^3/h] is the production flow. C_1 is the parameter for the energy consumption for static head loss in the abstraction and treatment process (linearly depending on the flow). C_2 is the parameter for the dynamic head loss (the head loss has a quadratic relation with the production flow, and must be multiplied with the flow to calculate the power consumption). We estimated the values for P_{base}, C_1 and C_2 in such a way that the calculated yearly energy consumption matched with the total energy consumption that could be derived from the energy bill of 2011. The specific energy consumption for abstraction and treatment was 0.456 kWh/m^3.

5.2.3 Pressure control

The distribution pumping station of WTP Gruszczyn consisted of five identical pumps all equipped with variable speed drives (VSD). The pumps were operated as one group at a fixed pump pressure. The clear water was pumped in two directions towards the individual supply areas. Initially the two areas were separated by a PRV in order to reduce the pressure in one zone while keeping a higher pressure in the other. The installed PRV was a medium driven automatic control valve Cla-val NGE9001 (DN250). The operators chose a relatively high pressure set-point for the clear water pumping station, because information about the pressure in the entire network during flow variations was lacking.

In the automation and optimisation process, nine new pressure measuring points were installed in the distribution network (see Figure 5.2). After examining the measurements, it was decided to remove the PRV, as will be explained in section 5.3. Removing the PRV resulted in equalised pressures in both zones. For the control of the clear water pumping station, a dynamic pressure control module (DPCM) was installed. In the conventional control loop, the pressure set-point was a static value chosen by the operator. The DPCM is a feedback control model, which dynamically calculates a pressure set-point for the pumping station by comparing the measured pressures at the measuring points with their individual lower bound values. The measuring point with the lowest pressure in relation to its lower bound value is the master in the control loop. The DPCM uses a manually tuned proportional integral derivative (PID) control mechanism to derive a pressure set-point for the pumping station, based on the desired (lower bound) and measured pressure value of this master pressure measuring point. Figure 5.3 shows the user interface of the DPCM.

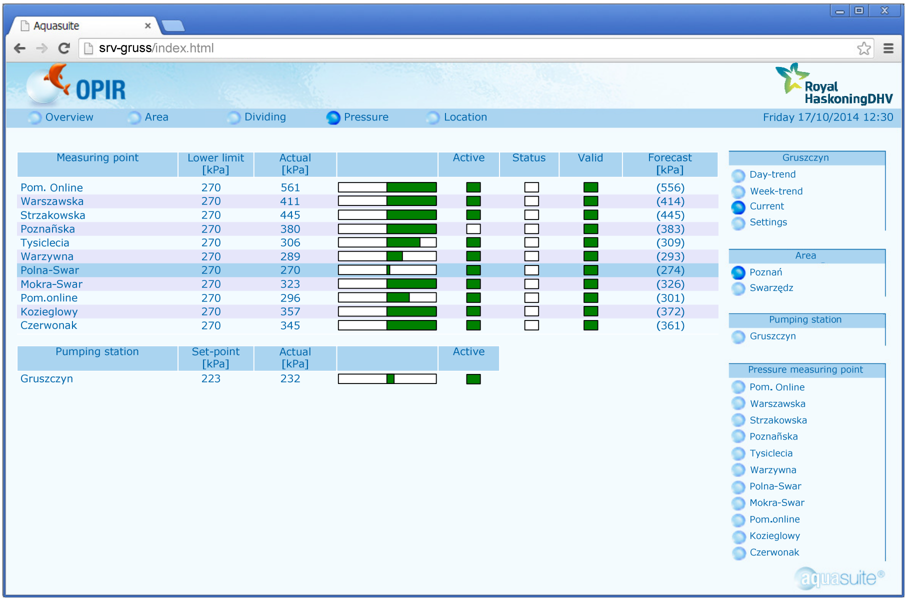


Figure 5.3 Interface of the dynamic pressure control module (DPCM), showing all measured pressures and their lower bound values, and highlighting which measuring point is the master in the control loop

Using off-line pressure measuring points

The nine installed pressure measuring points were equipped with a local logger and GSM modem. The measured pressures were not available in real-time, but stored locally and sent to the SCADA system of WTP Gruszczyn once per day. The DPCM estimated the real-time pressure p_i for each pressure measuring point i as a function of the real-time measured pressure at the pumping station p_{ps} and distribution flow to the area Q_{dist}, with:

$$p_i = p_{ps} + a + b \cdot Q^2_{dist} \quad [kPa] \tag{5.2}$$

The values for a and b in equation (5.2) were derived by the DPCM using the least-squares method on the data of the previous 72 hours (see Figure 5.4). The R^2 values of the fits were

0.84 in the Poznań area and 0.95 in the Swarzędz area. The fits in the Poznań area were relatively worse, because a large industrial user in the area caused that the demands were not always consistently distributed. With this functionality, the DPCM is a feedback control model that uses a predicted value as input, and can therefore be qualified to be a hybrid form of a predictive controller and a feedback controller as described by Ulanicki et al. (2000).

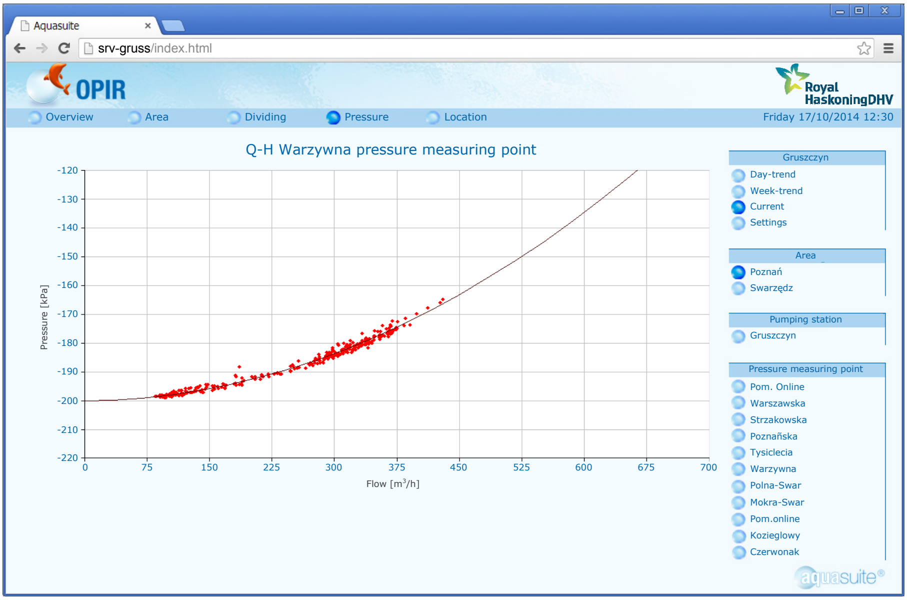


Figure 5.4 Least squares fit of measured pressure drop between pumping station and pressure measuring point as a relation of the flow to the area

Influence of pressure control

Pressure control influences the average pressure in the water distribution network and as a result also the leakage in the distribution network. The background leakage Q_{leak} can be described by (Gomes et al., 2011; Araujo et al., 2006; Vairavamoorthy and Lumbers, 1998):

$$Q_{leak} = K_f \cdot p^{\beta} \quad [m^3 / h] \tag{5.3}$$

where K_f is a leakage coefficient for the area, p is the average pressure in the area, and β is pressure exponent. In general, the pressure component β varies between 0.5 and 1.5 (Van Zyl and Cassa, 2014), but in certain networks values up to 2.5 were observed (Gomes et al., 2011). When no specific information about the network is known, a value of 1.0 is likely to be the best approximation (Van Zyl and Cassa, 2014; Gomes et al., 2011), and therefore we use 1.0 in this chapter.

If the pressure in the area changes, the background leakage will change:

$$\frac{q_{leak,1}}{q_{leak,2}} = \left(\frac{p_1}{p_2}\right)^{\beta} \tag{5.4}$$

A reduction of the leakage will lead to a reduction of the amount of water to be pumped. Therefore, also the energy consumption will be reduced. This reduction dE_{loss} can be estimated with:

$$dE_{loss} = dV_{loss} \cdot E_{spec,tot} \qquad [kWh] \tag{5.5}$$

where dV_{loss} [m^3] is the difference in water loss in the water distribution system and $E_{spec,tot}$ [kWh/m^3] is the total specific energy consumption for abstraction, treatment and distribution (0.600 kWh/m^3).

Changing the pressure will also affect the energy consumption by the clear water pumps. The difference in energy consumption dE_{pump} is calculated with:

$$dE_{pump} = \frac{\rho \cdot V \cdot dp}{1000 \cdot 3600 \cdot \eta} \quad [kWh] \tag{5.6}$$

where ρ is the specific mass of water (1,000 kg/m^3), V [m^3] is the total pumped volume of water, dp [kPa] is the difference in pump pressure, and η is the total efficiency of pump, motor plus VSD of the clear water pumps (estimated to be constant at 0.69 as observed by Hydratek (2013)).

5.3 Results

Comparison of operational periods

To evaluate the results of the improvements in control, the operational data (flows, pressures, water levels) of a period with conventional control were compared to a period of advanced automatic control. The implementation of the advanced control software was done in several phases, and after initial implementation a period of tuning followed. Therefore, a contiguous period with a sharp transition from conventional control to advanced control was not available. In this chapter, we compared a three weeks period with conventional control in November 2011 to a three weeks period with advanced control in November 2012.

Production flow control

Figure 5.5 shows trends of the total water demand, production flow and reservoir level of both examined periods.

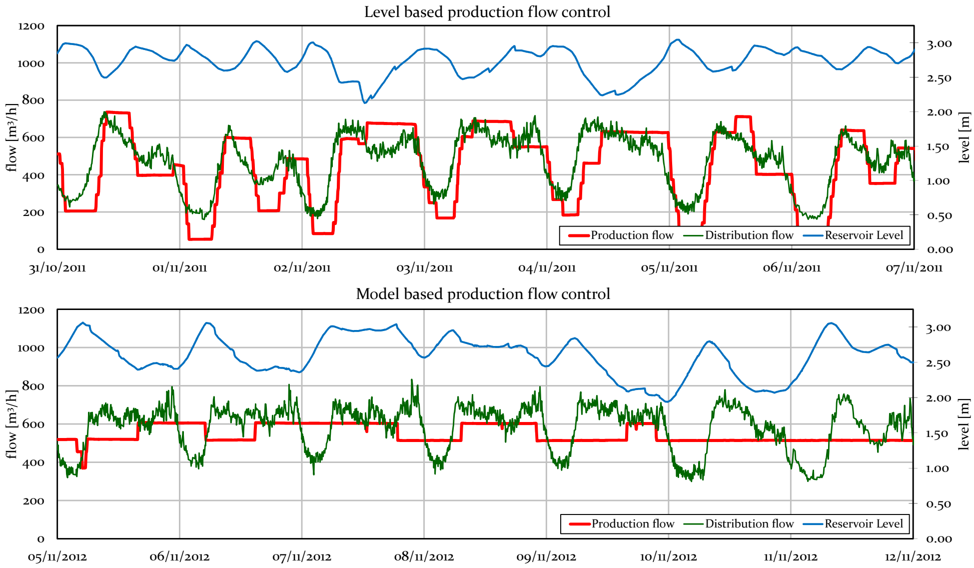


Figure 5.5 Level based (upper graph) and model based (lower graph) production flow control

The trend with level based control shows that the production flow was switched up and down every day, and that the minimum and maximum flows were almost equal to the minimum and maximum distribution flows. The trend with model based control shows a more stable production flow, with a smaller difference between maximum and minimum flows. Table 5.1 shows that model based control led to an 83% lower value for the production variation calculated with equation (4.1). Chapter 4 of this thesis (Bakker et al., 2013) showed that treatment performance was better at lower values of production variation, resulting in lower values of the turbidity of the clear water. Table 5.1 also shows that the difference between the maximum and minimum production flows was 67% lower with model based control. A third aspect shown in Table 5.1 is the energy consumption and energy costs, which were calculated with equation (5.1) for both periods. With predictive control, the energy consumption [in kWh/m^3] was 1.9% lower, and relatively more energy was consumed during low tariff hours resulting in 2.7% lower energy costs.

Table 5.1 Difference between level based and model based production flow control

	Level based	Model based	Difference
PV [%]	9.3	1.6	-83%
Production flow			
Min. flow [m^3/h]	52	367	+600%
Max. flow [m^3/h]	875	636	-27%
Max.-Min. [m^3/h]	823	269	-67%
Energy (estimation)			
Cons. [kWh/m^3]	0.456	0.447	-1.9%
Cost [€/1,000 m^3]	27.43	26.68	-2.7%

Pressure control

Figure 5.6 and Figure 5.7 show trends of the water demand, the outlet pressure at the pumping station and the average pressure in the area of both examined periods.

The measurements showed that initially the minimum pressures in the Poznań area were some 80-100 kPa higher than the desired minimum pressure. Therefore the pump pressure could be reduced significantly while still maintaining sufficient pressure in the network. In the initial setup of the system, a PRV was reducing the pressure to the other supply area (Swarzędz area, see Figure 5.1). A reduction of the pump pressure would eliminate the need of the PRV to reduce the pressure in the Swarzędz area. Based on the analysis of the pressure data, it was decided to remove the PRV.

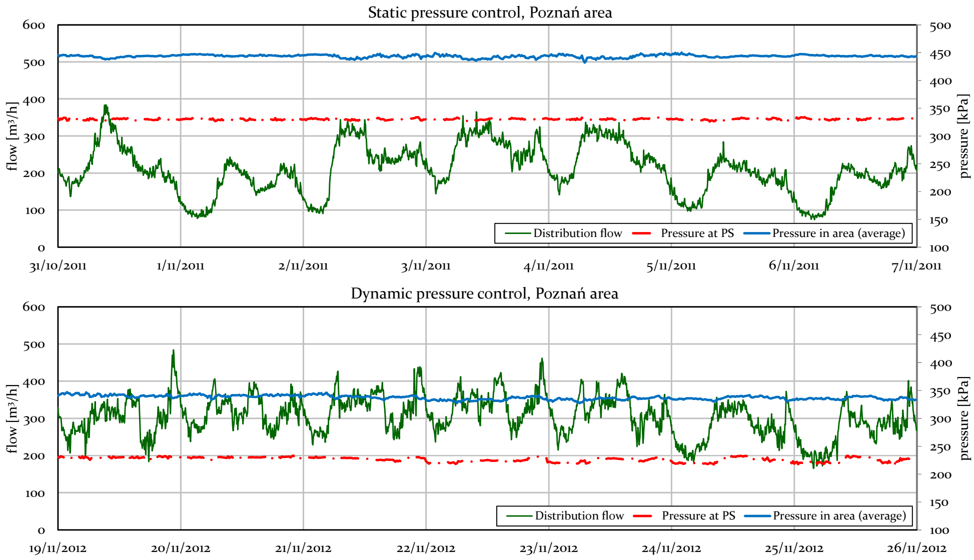


Figure 5.6 Static (upper graph) and dynamic (lower graph) pressure control, Poznań area

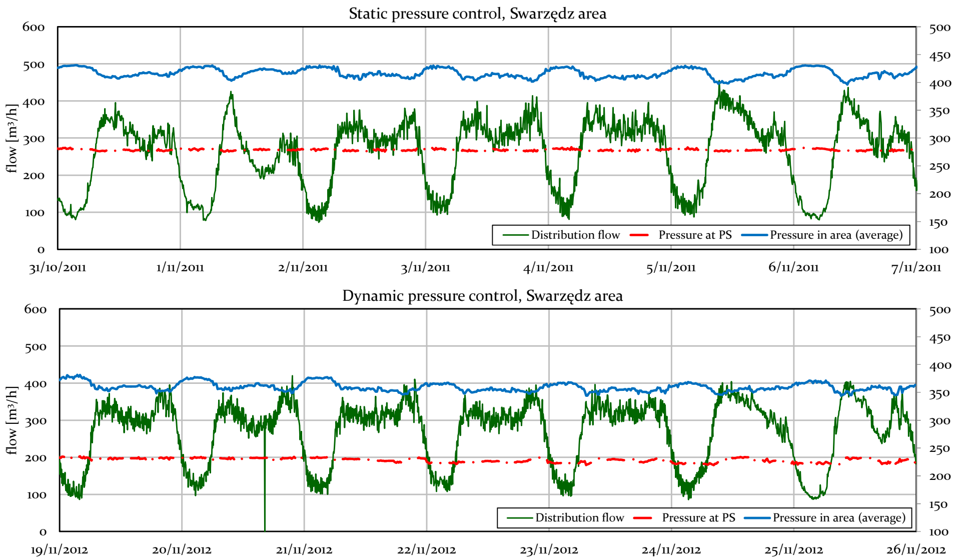


Figure 5.7 Static (upper graph) and dynamic (lower graph) pressure control, Swarzędz area

Initially, the pumps were operated at a fixed pressure (330 kPa), and the fixed outlet PRV was set to reduce the pressure to the Swarzędz area to 280 kPa. The outlet pressure of the PRV was not constant, but was inversely related to the flow. The behaviour of the PRV is shown in the left graphs of Figure 5.8.

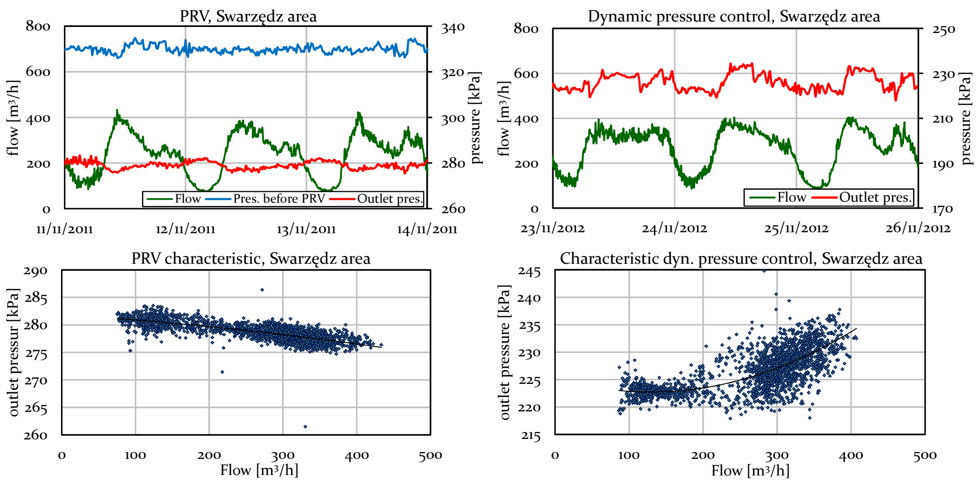


Figure 5.8 Outlet pressure in the water main to Swarzędz area with conventional pressure control including PRV (left graphs), and with dynamic pressure control (right graphs)

The right graphs of Figure 5.8 show the outlet pressure at the clear water pumping station when it was controlled by the DPCM. The trend shows that the outlet pressure was lower during low flows and higher during high flows. This was caused by the fact that the pressure set-point for the pumping station was based on the (calculated) minimum pressure in the distribution network. Using the DPCM resulted, as expected, in a quadratic relation between the pressure and the flow, as can be seen in the lower right graph of Figure 5.8.

The pump pressure of the clear water pumping station was 97 kPa lower (29%) when controlled by the DPCM, see Table 5.2. The average pressure in the distribution network was 100 kPa (23%) lower in the Poznań area and 50 kPa (12%) in the Swarzędz area. The difference in pressure in the Swarzędz area was smaller, because in the initial setup of the system the pressure was already reduced in this area with a PRV (reduction by 50 kPa on average). The water flow to the Poznań area was in the period with dynamic pressure control 92 m^3/h (43%) higher than in the period with static control. This increase was caused by a large industrial customer in the area, who increased its water consumption. The water flow to the Swarzędz area was almost equal in the two periods.

Table 5.2 Difference between static and dynamic pressure control

	Level based	**Model based**	**Difference pressure**	**Difference leakage**
Poznań area				
Flow [m^3/h]	214	306	+43%	
PS [kPa]	330	233	-29%	
MP1 [kPa]	395	298	-25%	
MP2 [kPa]	475	373	-22%	
MP3 [kPa]	462	361	-22%	
Area avg. [kPa]	*444*	*344*	-23%	-23%
Swarzędz area				
Flow [m^3/h]	261	267	+2%	
Outlet [kPa]	279	233	-16%	
MP1 [kPa]	470	422	-10%	
MP2 [kPa]	516	452	-12%	
MP3 [kPa]	439	390	-11%	
MP4 [kPa]	365	318	-13%	
MP5 [kPa]	329	281	-14%	
MP6 [kPa]	381	337	-12%	
Area avg. [kPa]	*417*	*367*	*-12%*	*-12%*
Total area avg. [kPa]	431	335	-17%	-18%

The lower pressure in the area when controlled by the DPCM resulted in lower water losses in the distribution network. By applying equation (5.4), the background leakage was calculated to be 23% lower in the Poznań area, and 12% lower in the Swarzędz area, resulting in 18% lower in the entire Gruszczyn system. Aquanet's total water losses were 5.30 million m^3 per year (11.3%) in 2011 (Aquanet, 2012). With the above, we estimated the water losses in the Gruszczyn system at 565,000 m^3 per year in 2011 (with static conventional control) and at 465,000 m^3 per year in 2012 (with dynamic pressure control).

Reduction of energy consumption

The reduction of energy consumption due to the implementation of advanced control software consisted of three elements:

1. Savings due to production flow control.
2. Savings due to lower pump pressure of clear water pumps.
3. Saving due to reduced water losses.

Table 5.1 showed that the energy costs were 2.7% lower with model based production flow control. With a total annual production of WTP Gruszczyn of 5 million m^3 per year, the implementation of model based production flow control led to a reduction in energy consumption of 43,000 kWh per year. The energy reduction due to lower pump pressure was calculated with equation (5.6). With a pumped volume (V) of 5 million m^3 per year, and a difference of the pump head (dp) of 97 kPa, this results in a dE_{pump} of 225,000 kWh per year. The energy reduction due to reduced water losses was calculated with equation (5.5). Based on a reduced water loss of 100,000 m^3 per year (see above), the reduction in energy consumption was calculated at 60,000 kWh per year. The overall reduction of the energy consumption was 328,000 kWh (€ 21,000) per year which corresponds to 11.5% of the total energy costs. The reduction of the pump pressure is the main contributor to the reduction of the overall energy consumption.

5.4 Discussion

Increased water consumption

Surprisingly, in the year with advanced control the water flows from the pumping station to the areas were higher than in the year with conventional control (Poznań area +48%; Swarzędz area +2%). A lower rather than a higher flow was expected, because the pressure in the areas was reduced significantly, and therefore significant reduction in background leakages were expected. The higher pump flows to the both areas can be explained by an increase of the water consumption. Especially, the higher water flow to the Poznań area can be linked to the increased water consumption of a large industrial customer in that area.

Availability of pressure measurements

An important difference between the periods of conventional control and advanced control, was the availability of pressure measurements in the distribution area. Initially, only two (unreliable) measurements were available. The operators had no confidence in the measurements and therefore chose a set-point based on their experience to operate the pumping station. The installation of new (more reliable) measurements showed that the pressure in the network was unnecessarily high. By using the new measurements, the advanced controller was able to reduce the pump pressure by 29%. However, if only new pressure measurements were installed and no advanced controller, the operators would have reduced the pumping pressure as well based on the observation of unnecessarily high pressures in the network. In this case, the availability of reliable measurements was presumably more important than the introduction of an advanced controller. Still, the combination of installing new sensors and introducing advanced control is an important success factor to achieve efficiency improvements, because the sensors and models help to understand the possible improvements, and the advanced control helps to achieve the best operational results continuously.

Static and flow modulated PRVs versus the DPCM

The installed PRV to reduce the pressure in the Swarzędz area was a classic medium controlled valve. Prescott and Ulanicki (2003) used this type of valve to develop their dynamic model of PRVs. According to this model, this type of PRV is somewhat flow dependent: during low flows the outlet pressure is slightly higher than during high flows. This behaviour was also observed in the trends of flow and pressure to the Swarzędz area (see Figure 5.8). However, this behaviour is undesirable. During low flows the dynamic head loss between the pumping station and the distribution area was lower, and therefore, a lower, instead of a higher, outlet pressure was desired.

The DPCM worked as a flow modulated PRV, which is also available as integrated medium controlled device, like the AQUAI-MOD® hydraulic controller (Abdelmeguid et al., 2011). Both the DPCM and medium controlled valves are able to reduce the outlet pressure during low flow, and increase the outlet pressure during higher flows. The advantage of the DPCM over a flow modulated PRV, is that the DPCM can be tuned easier, and that the DPCM adapts automatically to changing hydraulic or demand characteristics. A potential drawback of the DPCM is that it needs power and functioning communication infrastructure, which is not necessary for a medium controlled device. In the considered case study, this was not an issue because the pressure was controlled at the treatment facility which had a permanent and failsafe power supply and communication infrastructure.

Reduction of pipe breaks
Girard and Stewart (2007) and Gomes et al. (2011) showed that a reduction of the pressure in a water distribution system also leads to a reduction of the number of main breaks and service breaks. Based on the lower pressure in the area when controlled by the DPCM, a lower number of breaks may be expected in this case study as well. However, the number of breaks in the concerning areas were not registered separately, making it impossible to confirm the expected reduction of breaks.

5.5 Conclusion

Many water supply systems are still controlled by basic conventional control loops. Although this control is reliable and easy to understand, it is inefficient with respect to energy consumption and leakage. More advanced controllers can help to improve the efficiency of the water supply systems. The introduction of an advanced production flow control model at WTP Gruszczyn resulted in an 80% reduction of the production flow variation; the introduction of a dynamic pressure control model resulted in a 29% reduction of the pump pressure and an 18% reduction of the background leakage compared to conventional control. The dynamic pressure control module could achieve these results by using the data from nine newly installed pressure measurements. The resulting estimated reduction of the total energy consumption was 11.5% (328,000 kWh per year, or € 21,000 per year). Based on these results, we conclude that extra information from the distribution network (nine reliable pressure measuring points) in combination with advanced control software led to a more efficient water supply system.

Acknowledgements
The project is financially supported by the Dutch government through the "Partners for Water" programme. The project was implemented by Royal HaskoningDHV the Netherlands, supported by water utility Oasen (the Netherlands), water utility Aquanet (Poland) and Royal HaskoningDHV Poland.

5.6 References

Abdelmeguid, H., Skworcow, P., and Ulanicki, B. (2011). "Mathematical modelling of a hydraulic controller for PRV flow modulation". *Journal of Hydroinformatics.* 13 (3): 374-389.

Aquanet. 2012. Annual report 2011.

Araujo, L.S., Ramos, H., and Coelho, S.T. (2006). "Pressure control for leakage minimisation in water distribution systems management". *Water Resources Management.* 20 (1): 133-149.

Bakker, M., Van Schagen, K., and Timmer, J. (2003). "Flow control by prediction of water demand". *Journal of Water Supply: Research and Technology - AQUA*. 52 (6): 417-424.

Bakker, M., Vreeburg, J.H.G., Palmen, L.J., Sperber, V., Bakker, G., and Rietveld, L.C. (2013). "Better water quality and higher energy efficiency by using model predictive flow control at water supply systems". *Journal of Water Supply: Research and Technology - AQUA*. 62 (1): 1-13.

Bunn, S. (2007). "Closing the loop in water supply optimisation", *Proceedings IET Water Event*, Coventry, United Kingdom.

Bunn, S.M., and Reynolds, L. (2009). "The energy-efficiency benefits of pumpscheduling optimization for potable water supplies". *IBM Journal of Research and Development*. 53 (3).

Colombo, A.F., and Karney, B.W. (2005). "Impacts of leaks on energy consumption in pumped systems with storage". *Journal of Water Resources Planning and Management*. 131 (2): 146-155.

Girard, M., and Stewart, R.A. (2007). "Implementation of pressure and leakage management strategies on the gold coast, Australia: Case study". *Journal of Water Resources Planning and Management*. 133 (3): 210-217.

Gomes, R., Marques, A.S., and Sousa, J. (2011). "Estimation of the benefits yielded by pressure management in water distribution systems". *Urban Water Journal*. 8 (2): 65-77.

Hydratek. 2013. Toward Municipal Sector Conservation: A Pump Efficiency Assessment and Awareness Pilot Study. Toronto, Ontario, Canada.

Prescott, S.L., and Ulanicki, B. (2003). "Dynamic modeling of pressure reducing valves". *Journal of Hydraulic Engineering*. 129 (10): 804-812.

PWN. (2006). "Automated Water System Control". *Control Engineering*. 53 (10).

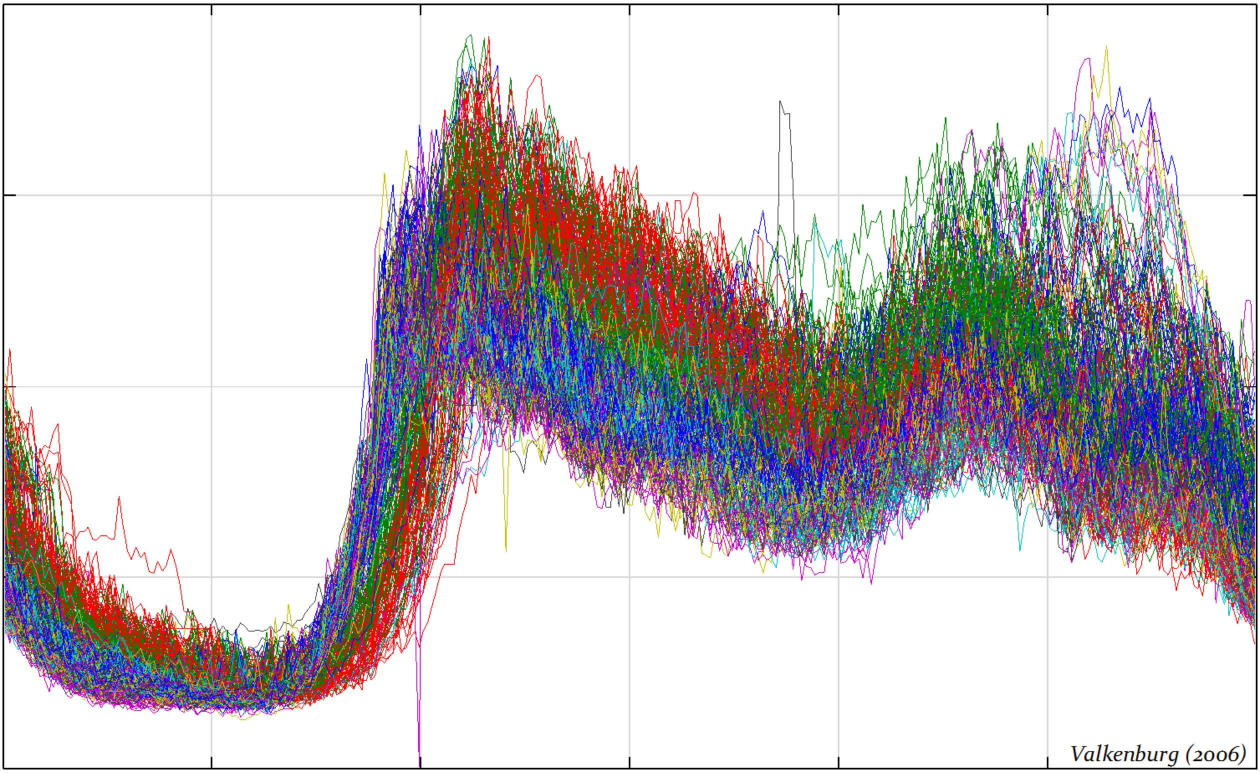
Ulanicki, B., Bounds, P.L.M., Rance, J.P., and Reynolds, L. (2000). "Open and closed loop pressure control for leakage reduction". *Urban Water*. 2 (2): 105-114.

Vairavamoorthy, K., and Lumbers, J. (1998). "Leakage reduction in water distribution systems: Optimal valve control". *Journal of Hydraulic Engineering*. 124 (11): 1146-1154.

Van Zyl, J.E., and Cassa, A.M. (2014). "Modeling elastically deforming leaks in water distribution pipes". *Journal of Hydraulic Engineering*. 140 (2): 182-189.

Worm, G.I.M., van der Helm, A.W.C., Lapikas, T., van Schagen, K.M., and Rietveld, L.C. (2010). "Integration of models, data management, interfaces and training support in a drinking water treatment plant simulator". *Environmental Modelling and Software*. 25 (5): 677-683.

Part IV – Pipe burst detection



6 Heuristic burst detection method using flow and pressure measurements

Based on

M. Bakker
J.H.G. Vreeburg
M. Van De Roer
L.C. Rietveld

Heuristic burst detection method using flow and pressure measurements
Journal of Hydroinformatics (2014, in press)

Abstract
Pipe bursts in a drinking water distribution system lead to water losses, interruption of supply, and damage to streets and houses due to the uncontrolled water flow. To minimise the negative consequences of pipe bursts, an early detection is necessary. This chapter describes a heuristic burst detection method, which continuously compares measured and expected values of water demands and pressures. The expected values of the water demand are generated by an adaptive water demand forecasting model, and the expected values of the pressures are generated by a dynamic pressure drop – demand relation estimator. The method is tested off-line on a historic dataset of five years of water flow and pressure data in three supply areas (with 650, 11,180 and 130,920 connections) in the western part of the Netherlands. In the period 274 bursts were reported of which, based on the definition we propose in this chapter, 38 were considered as (relative) large bursts. The proposed method was able to detect 50%, 25.9% and 7.8% in the considered areas related to all bursts, and around 80% in all three areas related to the subset of large bursts. The method generated false alarms on 3% of the evaluated days on average.

Keywords
Water distribution networks; burst detection; demand forecasting; data-driven pressure model

6.1 Introduction

6.1.1 Pipe bursts in Dutch water distribution systems

Response to pipe bursts and leakages are part of the daily operation of water companies. The burst frequency on water mains in the Netherlands is 7 bursts/ 100 km/year (Trietsch and Vreeburg, 2005; Vreeburg et al., 2013). This number is low compared to other countries: e.g. United States, 17 bursts/100 km/year (AWWA, 2007; Srirangarajan et al., 2013); The city of Trondheim, Norway, 30 bursts/100 km/year (Røstum, 2000); Three Canadian water companies: 34 bursts/100 km/year (Pelletier et al., 2003). The International Water Association (IWA) base level of burst frequencies for well-maintained systems is 13 bursts/100 km/year (Lambert and Thornton, 2011). Possible explanations of the low burst rate in the Netherlands are the relatively young age of the networks, and the low pressure in the networks because the country is rather flat and densely populated. Based on observations of 112 systems in 10 different countries, Thornton and Lambert (2006) showed that the burst frequency is reduced by 51% when reducing the maximum pressures by 37%. Other indicators showing a good performance of the Dutch distribution networks are the low rate of physical losses of 5% on average (Beuken et al., 2007) and a low infrastructure leakage index (ILI) of 0.7 (Lambert et al., 1999). The ground conditions in the country favour leaks showing rapidly at the surface (Lambert et al., 1999) which results in a quick detection and repair of bursts.

As a result, the issue of pipe bursts is not a top priority for the water companies, and budgets are limited to minimise the number and the impacts of bursts. Still, a few large bursts occur occasionally which have disruptive consequences for the urban environment and people. In order to improve the service level to consumers and to act more proactively in case of a pipe failure, the Dutch water companies wish to detect and locate bursts at an early stage. The main objective in this respect is the timely detection of large (catastrophic) burst, rather than effectively detecting small burst.

6.1.2 Unmanned operation of water systems

The operations of water supply systems is becoming centralised and unmanned (Worm et al., 2010). Operators who are continuously controlling a single location are replaced by supervisors who are supervising a number of locations in a region only during office hours. This increasing distance between the human operator and the water production and distribution processes, results in an increasing risk that failures in the system remain unnoticed. Distribution networks are often only monitored by a simple “flat-line” alerting system that raises an alarm when flow or pressure exceeds a static threshold value. Mounce et al. (2010) showed the limitations of such a system in detecting pipe bursts, and as a result many bursts stay unnoticed. Often, the water companies only take action after consumers’ complaints of low pressure or consumers reporting flooding caused by a burst pipe.

6.1.3 Life cycle of pipe bursts

Thornton et al. (2008) classified leakages in water distribution networks into 'background' (small continuous running leakages), 'unreported' (slightly bigger leakages, that tend to increase and need attention) and 'reported' (big leaks that need to be repaired as soon as possible). After the beginning of a burst, some time elapses before it is reported and the water company is aware of the situation. In the time frame between the beginning and the isolation of the burst, the broken pipe causes negative consequences like interruption of supply, water loss, and damage to streets and houses. The aim of a burst detection method is to minimise the time frame between the beginning and the moment that the water company is aware of the burst. This is the unawareness period in the life cycle of a burst, as shown in Figure 6.1 (based on WRc (1994), and expanded by Bakker et al. (2012)). The other periods, the awareness period, the location period, the isolation period and the repair period, are not affected by a burst detection method. This indicates that such method will only be valuable for bursts that have a relatively long unawareness period. Typically, this is the case for smaller bursts where only a small amount or no water surfaces, or for larger bursts that occur at night and surfacing water remains unnoticed.

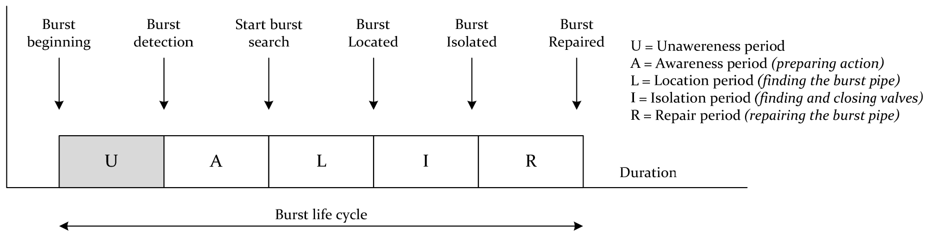


Figure 6.1 Life cycle of a burst (note that the duration of the periods can differ largely)

6.1.4 Previous work

Pipe burst detection can be considered as the application of anomaly detection techniques to a specific, narrow defined phenomenon. Anomaly detection is defined as "the problem of finding patterns in data that do not conform to expected behaviour" (Chandola et al., 2009). Anomaly detection can be divided in two sub-problems: 1) generating "expected behaviour" of the phenomenon; and 2) evaluating the "non-conformity" of observed and expected behaviour. For the detection of pipe bursts, various techniques can be used (Puust et al., 2010).

Monitoring hydraulic parameters (flow, pressure)

Flow and pressure are commonly measured in water distribution networks, and therefore used in most burst detection methods. The sampling interval of flow and pressure sensors installed in the network is usually 15 minutes (Romano et al., 2013), but for burst detection both longer intervals (1 hour, e.g. Palau et al. (2011)) and shorter intervals are applied (5 min., e.g. Eliades and Polycarpou (2012); 1 minute, e.g. Misiunas et al. (2006)). Mounce et al. (2012) studied the relation between the sampling interval and the performance of detection methods. The paper shows that shorter sampling intervals result in earlier detection, but questions if the earlier detection compensates the extra costs for communication and data handling.

Different techniques are used to generate “expected” values of the flows and pressures, like artificial neural networks (ANN) (Mounce et al., 2002; Romano et al., 2014), support vector machines (SVM) (Mounce et al., 2011), Fourier transformations (Eliades and Polycarpou, 2012) and Kalman filtering (Ye and Fenner, 2011). But the “expected” value can also be defined as the mean of observations in a previous period, differentiated in week days and weekend days. When applying this approach, different models need to be constructed for each season or periodicity, because of the (seasonal) variation in the water demand. This simplified approach is especially used in combination with statistical detection methods, like cumulative sum method (CUSUM) (Jung et al., 2013) or principle component analysis (PCA) (Palau et al., 2011). As the evaluation of (non-) conformity is based on a comparison of observed and expected values, the accuracy of the expected value under normal conditions plays a key role in the performance of the detection method. Surprisingly, in the abovementioned papers, little attention is paid to the analysis of the performance of the applied models that generate the expected values.

For the detection of events, the deviation between expected and observed behaviour is evaluated. Different techniques are applied for this evaluation, like an ANN combined with a rule based system (Mounce et al., 2003), fuzzy logic (Mounce et al., 2010), Bayesian inference systems (BISs) (Romano et al., 2013), a self-organizing map (SOM) (Aksela et al., 2009), or the CUSUM method (Misiunas et al., 2006; Jung et al., 2013).

Monitoring pressure transients

Pressure transients are evoked by a sudden failure (rupture) of a pipe, and can be monitored to detect pipe bursts. Transient monitoring consists of measuring pressure at different locations at a high sampling rates (250 Hz (Srirangarajan et al., 2013) up to 2,000 Hz (Misiunas et al., 2005a)). By analysing these measurements, a pipe burst can be detected and the burst location can be estimated. Colombo et al. (2009) presented a literature overview of transient monitoring techniques. Brunone and Ferrante (2001) and Gong et al. (2013) studied transients in a single water pipe. Misiunas et al. (2005b) and Srirangarajan et al. (2013) studied transients in distribution networks, and Allen et al. (2012) tested this approach on a test bed in Singapore.

Although the abovementioned papers report promising results, monitoring pressure transients to identify pipe bursts has some important disadvantages. The method is expensive because of the high sampling rates, which cannot be obtained with existing sensors and communication equipment. Furthermore, a large number of sensors need to be installed because pressure transients will only travel a few hundred meters (Srirangarajan et al., 2013). A second disadvantage is that transients will only arise, and thus can only be observed at pipe failures that happen (almost) instantaneously. Pipe failures that develop more gradually will not induce a pressure transient, and will therefore not be detected by this technique.

Monitoring other parameters
Other parameters besides flow and pressure can be monitored to identify pipe bursts. Khan et al. (2005) applied opacity and temperature sensors for burst detection, and Mounce et al. (2002) applied opacity sensors in combination with flow and pressure sensors. Also multi-probe devices were applied, that measure, in addition to the hydraulic parameters, conductivity, pH, oxidation reduction potential (ORP), and acoustics (Allen et al., 2012).

Performance evaluation
To develop and test detection methods, researchers have used different approaches. Some researchers only described the theoretical platform (Palau et al., 2011; Poulakis et al., 2003) or used simulated data (Misiunas et al., 2006). In most papers, the methods were tested on data from a real network. In some papers this was done with data of a rather short period of several weeks to one month (Mounce et al., 2002), sometimes including engineered burst events (Mounce et al., 2003; Mounce and Machell, 2006; Ye and Fenner, 2011; Romano et al., 2013). In some other papers, the performance of the detection methods was evaluated over longer periods: 2-3 months (Mounce et al., 2010; Palau et al., 2011); 6 months (Aksela et al., 2009; Mounce et al., 2011); 12 months (Eliades and Polycarpou, 2012; Mounce and Boxall, 2010; Romano et al., 2014). Most papers present results obtained in off-line simulation; only the results presented in Mounce and Boxall (2010) and Mounce et al. (2010) were obtained by an implemented on-line system.

6.1.5 Development of heuristic burst detection method

In this chapter, we research whether bursts can be detected by using only existing flow and pressure measurements in one very large and two smaller water supply areas. For the detection, we developed a heuristic burst detection method that monitors flow and pressure data. The key element of the method is the water demand forecasting model that was developed for control (see chapter 2), but is now applied for the detection of anomalies.

In section 6.2 we describe the area and data we used to test the method, and we describe the method itself. In section 6.3 we present the accuracy of the forecasting models, and the performance of the detection method expressed in detection probability (*DP*), rate of false alarms (*RF*) and detection time (*DT*). In section 6.4 we discuss the results and in section 6.5 we present the conclusions of this chapter.

6.2 Materials and methods

6.2.1 Study area and dataset

To develop and test the burst detection method, we collected a dataset with historic flow and pressure data. We collected all measured flows and pressures of three supply areas of the water company Dunea in the western part of the Netherlands. Data was available at 5 min. intervals for the period 2007-2012 (630,296 values per time series). The flows and pressures were measured at the permanent assets of the water company (treatments plant, reservoirs, boosters and permanent measuring points) and were stored in a central database system. Flow and pressure were only measured at these (12) locations; no further sensor data was available. The three researched areas are shown in Figure 6.2.

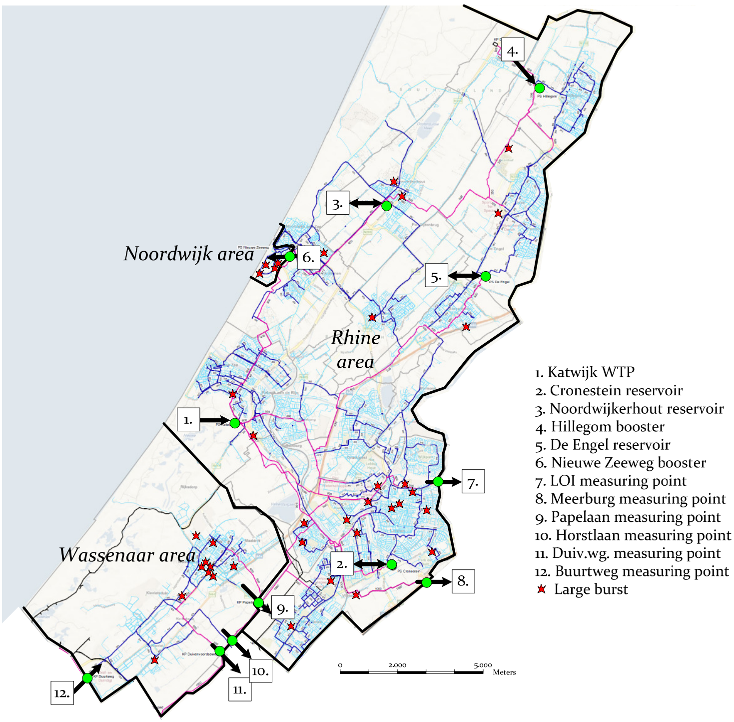


Figure 6.2 Study areas, including the treatment plant, reservoirs, boosters, measuring points, and bursts

The water supplied in the Rhine area is mainly produced at the Katwijk water treatment plant (1.), and buffered in the clear water reservoirs Cronestein (2.), Noordwijkerhout (3.), and De Engel (5.). These are ground level reservoirs that are filled during low demand (at night) with water from the network, and water is pumped back to the network during high demand. The Wassenaar area receives water from adjacent areas through four measured connections (measuring points 9. to 12.). The Noordwijk area receives water from the Rhine area through the Nieuwe Zeeweg booster (6.). The Rhine area contains 130,920 connections. This is a factor 100 larger than district metered areas (DMAs) in other countries, which generally contain 500-3,000 connections (Thornton et al., 2008; Machell et al., 2014). The Wassenaar area contains 11,180 connections (a factor 10 larger than a DMA); The Noordwijk area contains 650 connections (DMA size). When testing a burst detection method in very large areas with a low sensor density (Rhine area and Wassenaar area), a poor performance may be expected. The effect on flow and pressure of small bursts in such large areas is expected to be minimal, and therefore these bursts are likely to be missed by a detection method. As explained in section 6.1.1, the objective in this chapter is mainly the detection of large bursts. For this reason, we paid especially attention to the large bursts when evaluating the method.

Main repair (MR) records from the period 2008-2012 were available that contained information about the repairs carried out in the three areas. In these reported events, we distinguished between the total number of bursts and the number of large bursts. We subjectively defined large bursts as those where the burst flow exceeded the root mean square error (*RMSE*) of the demand forecast error (which will be presented in section 6.3, Table 6.2). The characteristics and reported incidents are summed in Table 6.1.

Table 6.1 Characteristics of the three researched areas (2008-2012)

	# connections	**Average demand (m^3/h)**	**# bursts (total / rel. large)**
1. Rhine area	130,920	2,290 m^3/h	242 / 24
2. Wassenaar area	11,180	212 m^3/h	26 / 10
3. Noordwijk area	650	31 m^3/h	6 / 4

6.2.2 Heuristic burst detection method

We developed a heuristic burst detection method that aims to detect bursts in real-time. In the case-study presented in this chapter, we applied the method off-line on the historic data from section 6.2.1. The method consists of five main steps: 1. Generating expected values of water demands and pressures; 2. Validity check of the signals and forecasts; 3. Transformation of the measured and expected values; 4. Analysis of the deviations between the measured and expected values, for A: generating threshold values based on historic data, and B: generating alarms by comparing the actual deviations with the threshold values; 5. Classifying detected bursts. The detection method ran in parallel for all three monitored areas, where the results from one area were used when monitoring the other areas. The setup of the method is shown in in Figure 6.3.

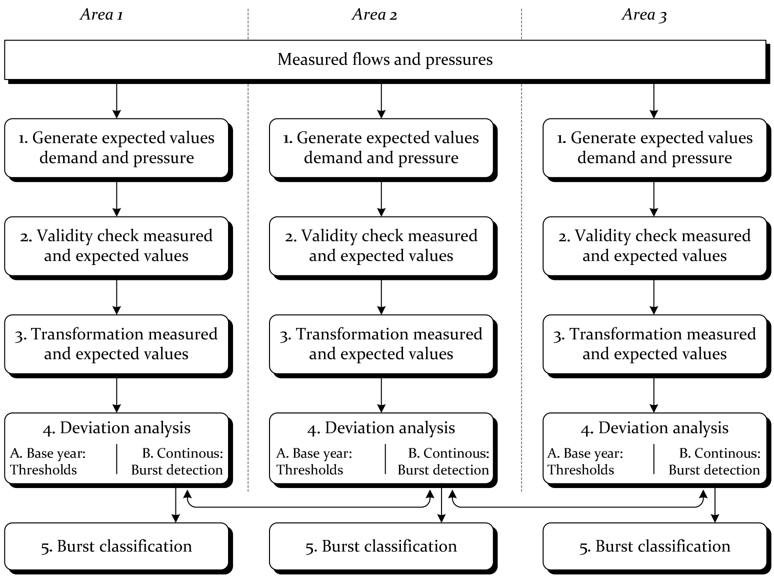


Figure 6.3 Setup of the heuristic burst detection method

1. Generate expected values for demand and pressure

For each monitored area, the net water demand was determined by performing a water balance calculation. The net water demand in the area was the input for the (data-driven) adaptive water demand forecasting model described in chapter 2 of this thesis (Bakker et al., 2013b). This model generates a water demand forecast for the next 48 hours with 15-min. time steps. This forecasting model has been implemented in real-time at a number of water supply systems for optimal control, see chapters 4 and 5 (Bakker et al., 2013a; Bakker et al., 2014a). For detection of pipe bursts, only the actual forecasted value (the so-called now-cast) was used. This now-cast was calculated by interpolating between the previous and the next 15-min. time step forecast.

In the Rhine area, pressure was measured at the entry point and at eight other locations. The method generated expected values for these pressures by fitting a relation between the pressure at the entry point and the pressure at a location based on measured data. The intake flow or pump flow at a location plays an important role when calculating the expected pressure at that location. Therefore different relations needed to be made for situations with intake flows (equation (6.1)), situations with (approximately) no flows (equation (6.2)), and situations with pump flows (equation (6.3)). The relations between the pressure at a location (p_{loc}) and the pressure at the entry point (p_{entry}) were formulated as:

$$Q_{loc} < -Q_{loc,min}$$
$$\rightarrow p_{loc} = p_{entry} + C_{1,1} + C_{1,2} \cdot Q^2{}_{area} + C_{1,3} \cdot Q^2{}_{entry} + C_{1,4} \cdot Q^2{}_{loc} \quad [kPa] \tag{6.1}$$

$$-Q_{loc,min} \leq Q_{loc} \leq Q_{loc,min}$$
$$\rightarrow p_{loc} = p_{entry} + C_{2,1} + C_{2,2} \cdot Q^2{}_{area} + C_{2,3} \cdot Q^2{}_{entry} \quad [kPa] \tag{6.2}$$

$$Q_{loc} > Q_{loc,min}$$
$$\rightarrow p_{loc} = p_{entry} + C_{3,1} + C_{3,2} \cdot Q^2{}_{area} + C_{3,3} \cdot Q^2{}_{entry} + C_{3,4} \cdot Q^2{}_{loc} \quad [kPa] \tag{6.3}$$

where Q_{area} is the net water demand in area, Q_{entry} is the water flow at the entry point, Q_{loc} is the water flow at the location, and $Q_{loc,min}$ is the threshold value to distinguish between zero flows, equation (6.2), and non-zero flows, equations (6.1) and (6.3). An automatic procedure derives the three sets of parameters $C_{n,1}$ to $C_{n,4}$ from the previous 72 hours of data using the least squares method. We chose to use only quadratic terms in the equations, because the pressure drop between two points is approximated by a quadratic function of the flow (Mays, 2000). This data-driven model can be implemented with very limited effort, and because the parameters are searched automatically, no manual tuning is required. However, this approach may have limitations because the pressure regime in the network changes as demand distribution (and flows) through the network change (see discussion in section 6.4.2). Figure 6.4 shows example trends of measured and expected demand and pressure.

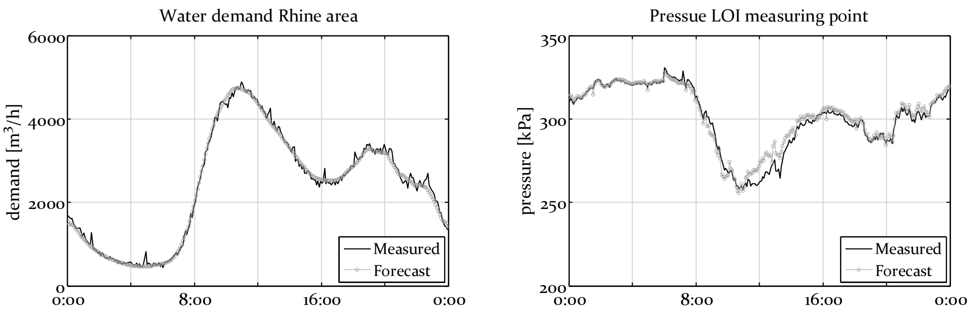


Figure 6.4 Examples of trends of measured and expected values, with the water demand in the Rhine area (left graph) and the pressure at LOI measuring point (right graph)

2. Validity check measured and expected values

The measured signals can be invalid due to measurement errors, sensor failures, and communication failures. To prevent false alarms caused by these failures, the method contains a rather basic validity check of the signals. A signal was considered invalid if: 1. The sensors' validity bit indicated a sensor or communication failure; 2. The value was outside a static upper / lower band that was configured for each signal individually (e.g. if a pressure measurement was <10 kPa or >500 kPa); 3. The value was exactly zero (only for pressure measurements); 4. The signal was "dead" (constant value). In addition to the validity check of the signals, the validity of the forecasts were checked as well. A forecast was considered invalid if the average forecast error of the previous four hours exceeded a configurable threshold value. The invalid status of the forecast was reset if the invalid condition was not true for 24 hours. When a signal or forecast was considered invalid, alarms were suppressed of the monitoring module that used the signal or forecast.

3. Transformation of measured and expected values

The measured time series of water demand and pressure showed unexplained variations due to random temporal and spatial variation in the water demand. These variations in the signal can be reduced by transforming the signal to a moving average, where the variations decrease as the moving average time frame is increased. Reducing the unexplained variation enables closer monitoring without increasing the number of false alarms. However, when calculating the moving average, the abnormal values after the burst will be equalised by normal values prior to the burst. This means that by taking the moving average value, smaller bursts can be detected but only some time after the beginning of the burst. In the burst detection method, the moving average signals over time frames of 5, 10, 15, 30, 60, 120 and 240 min. were monitored. Figure 6.5 shows that the errors of the expected water demand were more centred around zero when the time frame was increased from 5 minutes to 1 hour.

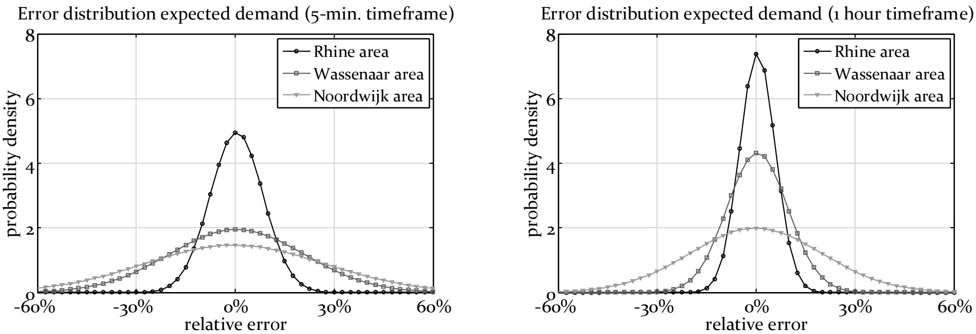


Figure 6.5 Error distribution of the expected water demand in the three areas (left: 5-min. time frame; right: 1 hour time frame)

4A. Deviation analysis: setting threshold values

Deviations between measured and expected values indicate a pipe burst. For monitoring purposes, threshold values need to be set to distinguish between normal forecasting inaccuracies versus possible burst events. All (transformed) forecasted and measured values are stored, and an automatic procedure derives the threshold parameters based on the stored data of previous one year period. The procedure splits the stored data in five classes (from low to high values), and calculates the 5% exceedance probability value of the percentage deviations for each class (see left graph of Figure 6.6). When monitoring, the actual threshold value is derived from the multiplication of the expected value and the (interpolated) 5% exceedance probability value of the percentage deviation. The right graph of Figure 6.6 shows both the percentage values and the absolute values as a function of the expected demand.

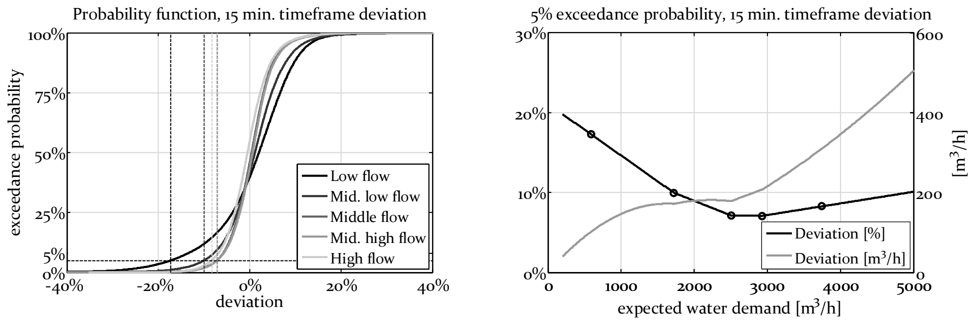


Figure 6.6 Example of the deviation analysis of the 15-min. time frame water demand forecast (Rhine area). The left graph shows the deviation probability functions of the five classes of expected water demand; the right graph shows the 5% exceedance probability values

We used the 5% exceedence probability deviations values (rather than 1% or 0.5% values), to ensure that these deviations were part of the normal forecasting inaccuracies and not caused by burst events (we made the implicit assumption that deviating flows and pressures due to pipe bursts occurred during less than 5% of the time). By using this value, there was no need to have detailed burst repair records available in order to derive the threshold value.

4B. Deviation analysis: Continuous burst detection

The method raised an alarm when the deviation of any of the signals exceeded its threshold value. Figure 6.7 shows an example of the method, where both the forecasted water demand (and related threshold values), and the actual measured demand are shown.

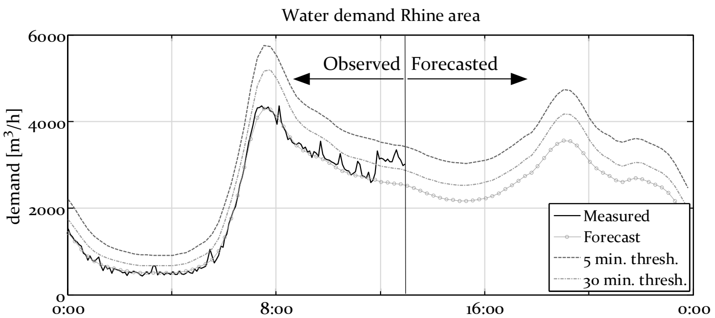


Figure 6.7 Monitoring the water demand by the detection method. The measured value was continuously compared with the threshold value. At the actual point in time (14:25 h) a burst was detected that began 30-min. earlier

The threshold value was defined as the 5% exceedance probability value multiplied by C_{lim}. The C_{lim} value controls the performance of the method: a low value results in precise monitoring and many false alarms; a high value results in less precise monitoring and fewer false alarms. We evaluated the performance of the method with a default C_{lim} value of 2.5. This value was set (based on the analysis of the data in the year prior to the monitoring years) to limit the number of false alarms to approximately 1 per month per area. In the discussion section of this chapter, we present a sensitivity analysis of the C_{lim} factor.

Certain exceptional water demands are not forecasted properly by the forecasting model. Exceptional water demands are often spatial correlated (Buchberger and Li, 2009) and occur in multiple areas at the same time. An example of such water demand evoked by collective human behaviour is the sudden peaks and drops in the water demand during and after important sports games, see Figure 1.6 (Bakker et al., 2003). This non-forecasted demand pattern may result in deviations that exceed the threshold value, resulting in false alarms. To eliminate those kinds of false alarms, the heuristic burst detection method simultaneously monitored the demand deviations in all areas. An alarm was suppressed in one area if the deviation in a neighbouring area exceeded its 5% exceedance probability value multiplied by C_{supp}. The default value of C_{supp} was 1.0, based on the analysis of the data in the year prior to the monitoring years. This alarm suppressing mechanism may wrongly retain alarms when real bursts occur at the same time in neighbouring areas. However, the statistical chance of simultaneous bursts is very small, and therefore this mechanism will not limit the applicability of the method for the case study described here, but could be an issue for systems with densely interconnected, smaller DMAs. In section 6.4.1 a sensitivity analysis of the C_{supp} factor is presented. The C_{lim} and the C_{supp} factor are the method's main parameters that can/may be tuned manually when implementing the method. All forecasts and threshold parameters are derived by the method automatically by data-driven procedures.

5. Burst classification
When a burst was detected, the method classified the burst by estimating the burst flow and the certainty whether the detected event is really a burst. The classification procedure will further be explained in section 7.2.2.

6.2.3 Performance evaluation

To assess the added value of a burst detection method, the detection time (*DT*), and both the detection probability (*DP*, also noted as True Positive Rate) and the rate of false alarms (*RF*, also noted as False Positive Rate) need to be evaluated, which can be expressed as (Jung et al., 2013; Metz, 1978):

$$DT = T_{detection} - T_{beginning} \qquad [minutes] \tag{6.4}$$

$$DP = \frac{number\ of\ detected\ bursts}{number\ of\ days\ with\ burst} \cdot 100\% = \frac{True\ Positives}{Positives} \tag{6.5}$$

$$RF = \frac{number\ of\ false\ alarms}{number\ of\ days\ without\ burst} \cdot 100\% = \frac{False\ Positives}{Negatives} \tag{6.6}$$

where $T_{detection}$ is the point in time of detection by the method, and $T_{beginning}$ is the point in time of beginning of the burst. Note that the latter was not measured but estimated by us by examining the hydraulic data.

The area under curve (AUC) (Hanley and McNeil, 1982) of the receiver operating characteristics (ROC) graph (Egan, 1975) represents the effectiveness of a detection method: the closer the AUC value to one, the more effective. The ROC graph depicts the trade-off between the hit rate (true positive rate) and false alarm (false positive rate) of a detection method. We derived the ROC-curve and AUC-value for the burst detection method, by varying the C_{lim} value. Note that we applied the *DP* as true positive rate and the *RF* as false positive rate. This means that we evaluated “days” (on which a burst event occurred or not) rather than all individual 5-min. time-steps (in which water was flowing from a burst pipe or not) monitored by the method.

6.3 Results

6.3.1 Analysis of the reported large burst events

Detailed information about the exact point in time of beginning, detection, location, and isolation of the considered 38 large burst events was lacking. However, from the flow data, information about the running time of the burst could be extracted. The flow patterns of the examined large burst events showed a sudden increase at the beginning and a sudden decrease at the isolation point in time. In the intermediate time frame the water ran freely from the burst pipe. This time frame covers the unawareness period + awareness period + location period + isolation period of the life cycle of a burst (Figure 6.1). Figure 6.8 shows the relation between the running time of the burst and the burst flow (left graph) and the beginning point in time (right graph).

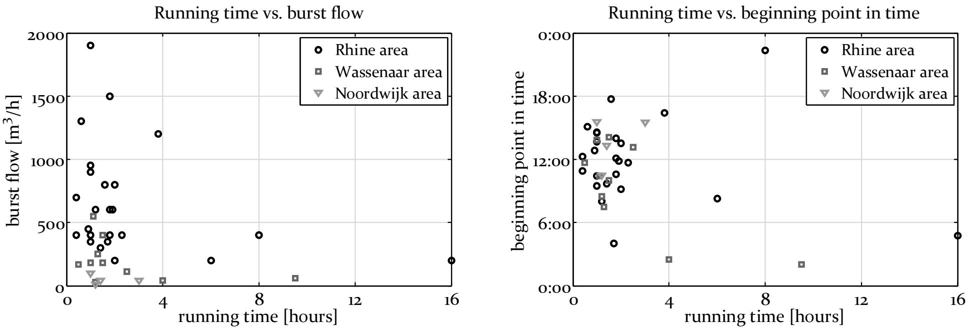


Figure 6.8 Relation between the running time of all observed bursts versus the burst flow (left graph) and the running time versus the beginning point in time of the bursts (right graph)

Figure 6.8 shows that the running time of most bursts was short: 31% were less than one hour, and 45% were between one and two hours. The isolation period (to identify the proper valves in the GIS system, and to locate and close the valves in the field) takes 30-45 min. according to servicemen of Dunea. Closing the valves of the concerning (large diameter) pipe requires a long closing time to prevent water hammer. This indicates that most bursts were discovered and reported shortly after they began, resulting in a rather short unawareness period. Figure 6.8 shows that four bursts had a running time of six hours or more. These were all bursts that had a low burst flow compared to other bursts, and three out of four began in the late evening or night (between 22:00 and 5:00). Assuming the same isolation, awareness and location period, the unawareness period was considerably longer for these bursts. This indicates that bursts that began in the night were not noticed by consumers and not promptly reported to the water company. Note that the vast majority of the examined bursts began during the day (between 7:00 and 17:00). This may be caused by the fact that some of the large bursts are caused by excavation activities that are mainly executed during the day (see chapter 7). However, because we did not study the cause of the bursts, we were not able to verify this assumption in this case study.

6.3.2 Deviations analysis and detection threshold values

Deviations expected water demand

Table 6.2 shows the percentage of valid values, the root mean square error of the deviations (*RMSE*, expressed as relative and absolute value), and the 5% exceedence probability values of the five classes of the expected water demands. Table 6.2 shows that the smallest percentage deviations occurred in the largest area (Rhine area) and the largest percentage deviations in the smallest area (Noordwijk). This is in accordance with the observations in chapter 2 of this thesis (Bakker et al., 2013b). Meanwhile, the absolute deviations in m^3/h were the largest in the largest area and the smallest in the smallest area. This indicates that despite the worse performance of the forecasting model in the smaller areas, still the absolute values of the deviations were smaller which enabled the detection of smaller bursts in the smaller areas.

Table 6.2 Percentage of valid values, and deviations of the expected water demands, related to 15-min. moving average value (L=Low; ML=Middle Low; M=Middle; MH=Middle High; H=High values)

	Valid	*RMSE*		5% exceedence prob. deviation				
				L	ML	M	MH	H
		[%]	[m^3/h]	[%]	[%]	[%]	[%]	[%]
1. Rhine area	100%	6.7	154	17.3	9.5	6.8	6.4	7.9
2. Wassenaar area	100%	11.3	24	32.1	15.9	11.2	11.5	12.1
3. Noordwijk area	100%	30.6	9	80.3	64.8	41.8	31.7	24.4

The method's automatic procedure executed the signal transformation and deviation analysis to derive alarm threshold values, as explained in section 6.2.2. Figure 6.9 shows the resulting threshold values for the water demand in the Rhine area.

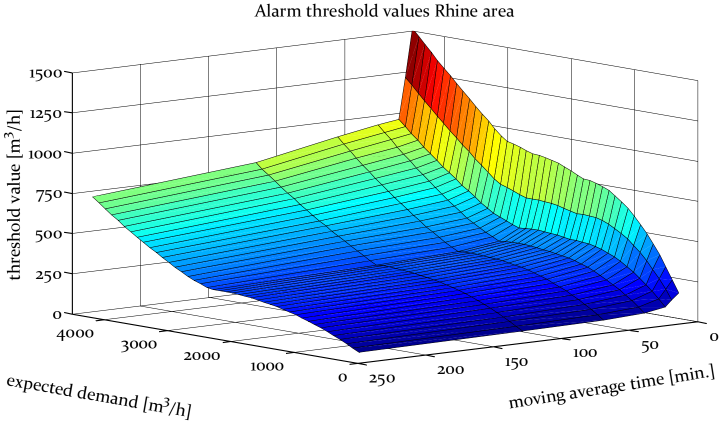


Figure 6.9 Alarm threshold values for the expected water demand in the Rhine area

The figure clearly shows high threshold values for short moving average time frames and high values of the expected demand. A high threshold value means that only a large deviation between measured and expected flow (which is the case for bursts with high burst flows) will raise an alarm. Lower threshold values occurred when expected demand was lower and/or when the moving average time frames were longer. This means that bursts with a smaller burst flow, only raised an alarm when demand was lower (e.g. at night) and/or somewhat longer after the beginning of the burst.

Deviations expected pressure

For eight measured pressures, the method generated expected values by applying equations (6.1) to (6.3). Table 6.3 shows the percentage of valid values, the *RMSE* of the deviations (expressed as relative and absolute value), and the 5% exceedence probability values of the five classes of the expected water pressures. The table shows that the percentage deviations of the expected pressures were lower than those of the expected demands (Table 6.2). This indicates that the pressure could be estimated more accurately.

Table 6.3 Percentage of valid values, and deviations of the expected pressures related to 15-min. moving average value (L=Low; ML=Middle Low; M=Middle; MH=Middle High; H=High values)

	Valid	***RMSE***		**5% exceedence prob. deviation**				
				L	**ML**	**M**	**MH**	**H**
		[%]	**[kPa]**	**[%]**	**[%]**	**[%]**	**[%]**	**[%]**
1. loc. Cronestein	96.2%	0.7	2.1	1.0	0.9	0.5	0.5	0.6
2. loc. de Engel	98.4%	3.1	9.2	3.0	2.8	2.4	2.2	2.0
3. loc. Hillegom	97.2%	11.0	31.7	5.2	4.2	4.0	4.5	3.3
4. loc. LOI	86.0%	0.6	1.9	1.0	0.8	0.7	0.7	0.6
5. loc. Meerburg	96.9%	0.9	3.0	2.7	0.5	0.5	0.5	0.5
6. loc. Noordwijk	93.9%	2.5	7.4	4.2	3.2	3.3	3.1	3.5
7. loc. Nw'hout	96.8%	5.3	16.4	3.7	2.9	2.7	2.4	2.2
8. loc. Papelaan	91.8%	4.0	10.9	5.7	4.2	4.4	5.0	10.5

6.3.3 Performance burst detection method

The initial detection threshold parameters were derived by the method's automatic procedure from the 2007 data. Next, the burst detection method was applied to evaluate the 2008-2012 data. We analysed the performance of the method by evaluating the *DP* and the *RF* (both related to all observed burst: DP_{All} and RF_{All}; and to the selected subsets of large bursts: DP_{Large} and RF_{large}) and the *DT* see Table 6.4. The table shows that the DP_{All} values were quite low, especially in the larger areas. The DP_{Large} values were on average around 80%, the RF values around 3% and the *DT* values around 20 min. The *DT* was 5-10 min. for most bursts, but a number of bursts was detected much later (the highest *DT* was 75 min.). Figure 6.10 shows the ROC graph of the burst detection method, applied to the Rhine area. The AUC-value of the curve related to all bursts was 0.535, and related to the large bursts 0.972. This indicates that in the Rhine area the method was very ineffective related to all bursts, but effective related to the large bursts.

Table 6.4 Performance of the burst detection method (*DP*=Detection Probability; *RF*=Rate of False (related to all burst or to the subset of large bursts); *DT*= Detection Time)

	DP_{All} [%]	RF_{All} [%]	DP_{Large} [%]	RF_{Large} [%]	***DT*** **[min.]**
Rhine area (24 bursts)	7.8	4.2	79.2	3.4	12
Wassenaar area (10 bursts)	25.9	2.3	90.0	2.1	24
Noordwijk area (6 bursts)	50.0	2.1	75.0	2.0	20

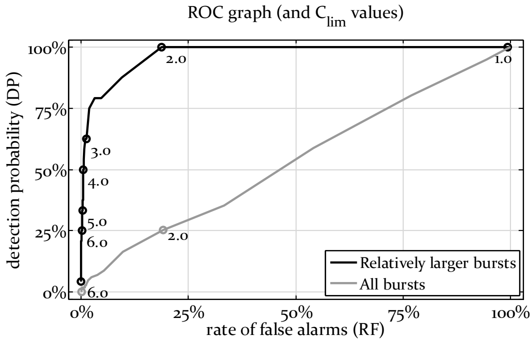


Figure 6.10 ROC graphs of the burst detection method in the Rhine area of the large bursts only, and of all bursts

An analysis of the results showed that all alarms were raised by a deviation of the water demand. During four burst events, a deviation of the pressure raised an alarm as well. During the other events, the deviation of any of the pressures did not exceed the alarm threshold value. This indicates that monitoring the pressures did not provide additional information for detecting the bursts. This limited sensitivity of pressure sensors (that were at some distance from the burst location), is in accordance with observations in Mounce et al. (2011) and Farley et al. (2013).

6.4 Discussion

6.4.1 Sensitivity analysis model parameters

Detection probability (C_{lim})

The C_{lim} factor directly influences the alarm threshold value, and as a result, the factor determines the trade-off between hit rates and false alarm rates. The acceptable number of false alarms can be determined by the water company that uses the method. We aimed at false alarm rate of 3% and analysed the data prior to the monitoring years (2007) to find the proper C_{lim} (and C_{supp}) values. A C_{lim} value of 2.5 appeared to be applicable for all areas, which indicates that this factor is not very sensitive to the size of the area. To assess its influence, we tested different C_{lim} values at the data of the Rhine area and evaluating the large bursts only. The left graph of Figure 6.11 shows *DP* and *RF* as a function of C_{lim}. The graph shows that a *DP* of 100% was achieved with a C_{lim} value of 1.8 or smaller. With this value however, a *RF* of at least 25% occurred. The default C_{lim} value of 2.5 resulted in acceptable *DP* and *RF* values.

Alarm suppressing (C_{supp})

The C_{supp} factor directly influences the threshold value to suppress alarms. The right graph of Figure 6.11 shows *DP* and *RF* as a function of C_{supp}. With smaller values of C_{supp}, more potential alarms are suppressed and the chance of wrongly suppressed alarms increases. The graph shows that the *DP* decreased at a C_{supp} value smaller than 0.5 which indicates that alarms were suppressed wrongly. The default C_{supp} value of 1.0 resulted in acceptable *DP* and *RF* values.

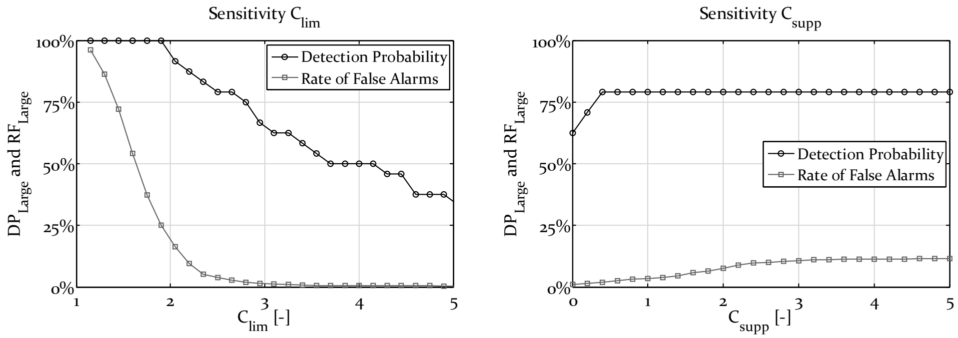


Figure 6.11 Sensitivity analysis of the C_{lim} (default 2.5) and C_{supp} (default 1.0) parameters when monitoring the Rhine area water demand (related to the large bursts)

With C_{supp} values of 4.5 and larger, the alarm suppressing functionality was virtually turned off, resulting in a constant *RF* value of 11.4%. The applied C_{supp} value of 1.0 resulted in a *RF* value of 3.4%. Hence, the alarm suppressing functionality was able to reduce het *RF* value of 11.4% to 3.4%, which means that 70% of all false alarms were effectively suppressed.

6.4.2 Data-driven pressure estimation

We used a grey-box data-driven model (equations (6.1) to (6.3)) instead of a hydraulic network model or ANN model to estimate the pressures. We chose to do so, to enable easy and low-cost implementation and to reduce calculation time. However, the data-driven model may have limitations because the pressure regime in the network changes as demand distribution (and flows) through the network change. This might explain why monitoring the pressure proved not to be very valuable for burst detection in our research. However, the pressure estimation was rather accurate: Table 6.3 shows that the *RMSE* was 10 kPa (3.5%) on average. This was more accurate than the demand forecast: the *RMSE* of the demand forecast in the Rhine area was 6.7%, see Table 6.2. A more detailed analysis showed that the maximum pressure deviation at any of the pressure measuring points during the bursts was around 12 kPa for most bursts (4.1% of the average pressure). The average flow deviation caused by bursts was 660 m^3/h (22.8% of the average demand). This indicates that the main reason that pressure monitoring was not very sensitive in our research, was that the effect of bursts on pressure at the measuring points was much less profound than the effect on the flow. The cause of the low sensitivity of the pressure measurements to bursts, is that most bursts did not occur in the direct vicinity of the sensor location (see Figure 6.2). The burst only caused a local pressure drop, but did not affect the pressure at the sensor locations much.

6.4.3 Applicability of the method

Availability of measurements

We tested the burst detection method on one very large (100 DMAs), one large (10 DMA) and one average (DMA size) network. At the very large network, flow and pressure were measured at the main entry point and at eight other locations, and all these measurements were used in the burst detection method. In many other water distribution networks in the Netherlands, there are not that many measurements available in a network. The large and average networks resemble more normal networks in the Netherlands, where flow and pressure are only measured at one or two locations. Compared to international standards, this sensor density is quite low. The burst detection method proved to be equally effective in all areas for detecting the large bursts. The reason for this is that monitoring the water demand in the area is the key element for burst detection (as explained in section 6.4.2), and the extra information provided by monitoring the pressure is limited. Therefore, the method can be applied to any network where a water balance can be made.

Size of the area

The size of the area that is monitored very much influences the size of the bursts that can be detected. In the large area (Rhine area) only bursts were detected where the burst flow exceeded 150 m^3/h; in the smallest area (Noordwijk area) bursts starting at 7 m^3/h were detected. This is an important limitation of the burst detection method, and must be borne in mind when implementing the method in larger areas. The minimum burst size that can be detected in an area, can be approximated by calculating the *RMSE* of the water demand forecast error. This line of reasoning can also be reversed: if the minimum burst size that must be detected is set at a certain value, the (maximum) size of the areas can be determined. If the existing areas are larger, additional flow meters must be installed to created smaller areas. In such case, additional investments are needed for installing and operating new sensors when implementing the detection method. The further analysis of the relation between the size of the area and the minimum detectable burst size is reported in chapter 7 of this thesis (Bakker et al., 2014b).

Changes in operation or topology

Changes in the operation or in the topology of the distribution system can result in changes in the water demand in the area or in the hydraulic behaviour of the network. Because the detection method uses data-driven models to generate expected values of demands and pressures, this deviating behaviour will not be forecasted properly. As a result, false alarms may be expected when the changes in demand and hydraulic behaviour are large. To avoid undesirable false alarms, the burst detection method should be put off-line temporarily in an area, if the operations or topology are changed in this area. The data-driven methods are mainly based on the behaviour in the previous 72 hours and therefore expected values will be accurate again only days after the change. As a result, the detection can be put on-line shortly after the change in operation or topology of the network.

Testing with off-line data

We tested the burst detection method only with historic off-line data. One year of off-line data was used for determining the monitoring threshold parameters. Next, the rest of the historic data was processed by the method to find the bursts. When developing the method, we took into account the aim of implementing the method in an on-line matter. The next step is the transformation of the off-line method into an on-line method, and test it when monitoring a water distribution network in real-time.

6.4.4 Added value burst detection method

Low detection probability related to all bursts

The detection probability related to all bursts was very low in the very large area (7.8%) and the large area (25.9%). This indicates that when applying our method at areas of this scale, the (vast) majority of all bursts will be missed. The added value of applying the method at this scale is therefore questionable. As the major part of the physical water loss in water distribution systems is caused by small bursts and leaks, applying the burst detection method will not be valuable for reducing water loss. However, because the water losses in the Dutch networks are quite low, reducing background leakage is not the main driver for the application of a burst detection method.

Minimise damage caused by large bursts

The running time of most large bursts (76%) was two hours or less, and an earlier detection could not have prevented any damage. Four bursts occurred with longer running times of 6 hours or more. The burst detection method was able to identify all four events and to raise an alarm. Thus, for these bursts the detection method could have been valuable for the water company, and damage caused by the bursts could have been minimised. As the damage of any single burst can be large, the costs for implementing this burst detection method may compensate the avoided damage.

Need for burst localisation

When the detection method detects a burst, the burst pipe needs to be located before it can be isolated and repaired. This can be difficult and time consuming, especially in case of a burst in a large area (like the Rhine area). Therefore, the detection method should be combined with a localisation method to enable effective response to a burst event. Possible methods to locate a burst are described by e.g. Misiunas et al. (2005a), Farley et al. (2013) and Romano et al. (2013).

False alarms caused by deviating water use

Occasional deviating water use of large customers (e.g. tanker filling, cleaning activities, fire-fighting, et cetera) will result in a deviating water flow and possible local pressure drop. If the effect of these activities on flow or pressure exceeds the monitoring threshold value, the method will raise a false alarm. The method does not have a functionality to distinguish between pipe bursts and other events when observing deviating hydraulic parameters.

Comparison to other burst detection methods
A variety of approaches and techniques has been published for the application area of burst detection (see section 6.1.4). We think that it is not possible to compare the performance of our method with the performance of other methods found in literature. The reason for this is that the performance highly depends on the size of the area and the sensor density. The results presented in this chapter are mainly based on the analysis of data from a very large (100 DMAs) and large (10 DMAs) area with few sensors, where other methods presented in literature typically evaluated data from areas of normal DMA size. It is obvious that most of the bursts (small bursts that only locally affect flow and pressure) cannot be detected if the area is very large and only few measurements are available. As far as we know, no other researchers studied burst detection techniques at this large scale. The main difference between our method is and other methods, is that our method uses understandable heuristic models and procedures to derive expected values and threshold values of the monitored signals, and it has an effective alarm suppressing functionality. Other methods have more sophisticated (combined) procedures for comparing expected and measured values for detecting bursts.

6.5 Conclusions

The main assets (treatment plants, reservoirs, boosters, and measuring points) of most water distribution systems are equipped with permanent flow and pressure sensors. Most water companies only use the sensor data for real-time control and management of the water distribution networks. However, the sensors can provide valuable information for detecting abnormal events like pipe bursts. Analysis of large burst events showed that most (large) pipe bursts were reported by consumers shortly after they began, and as a result 76% of all bursts were isolated within two hours after the beginning. However, some bursts had a longer running time, and the water flows of those bursts could cause considerable damage to the urban environment. To minimise this potential damage and other negative aspects of pipe burst, an early detection is required.

We developed and tested a heuristic burst detection method off-line on a historic dataset, containing five years of hydraulic data in three distribution areas in the western part of the Netherlands. The three areas varied largely in size: 650, 11,180 and 130,920 connections. The data of the latter area contained most burst events, and therefore, this area was dominant in developing and evaluating the method. Because of the large size of this area, only large pipe bursts could be detected, and small size bursts were not detected by the method.

The heuristic burst detection method we propose in this chapter is based on monitoring water demands and pressures in the water distribution network. The method uses adaptive data-driven models to generate expected values of the water demands and pressures. An automatic procedure analysed historic deviations between measured and expected values to set the value of threshold parameters, and (continuous) observed deviations were evaluated to raise alarms. The method monitored multiple areas in parallel, which enabled the suppressing of alarms in case abnormal water demands occurred simultaneously in different areas. The functionality reduced the rate of false alarms (*RF*) from 11.4% to 3.4% (70% of the false alarm were suppressed). The deviation between measured and expected demand was the key element to detect pipe bursts; deviations in pressure appeared to be less valuable for burst detection. Because the detection method uses only existing measurements, and is comprised of adaptive data-driven models, it can be implemented and operated at low cost.

When evaluating the method, we considered both all reported bursts and a subset of large bursts which were selected by applying our (subjective) definition. When all reported bursts were considered, the method detected 7.8% of the bursts in the large area, 25.9% in the medium area, and 50% in the small area. When the subset of large bursts was considered, the method detected around 80% of the bursts in all three areas. The method generated an acceptable number of false alarms, and the average detection time was 20 min. The detection time was 5-10 min. for most bursts, but a number of bursts was detected much later (up to 75 min.). The method was able to detect the critical bursts which had a long running time, at an early stage. This shows that the burst detection method can shorten the “unawareness period” of a burst, and therewith deliver a contribution to minimise the negative aspects of large pipe bursts. A further reduction of the negative aspects of bursts can be achieved if the “location period” can be shortened as well. To achieve this, the burst detection method should be extended with a burst localisation method.

6.6 References

Aksela, K., Aksela, M., and Vahala, R. (2009). "Leakage detection in a real distribution network using a SOM". *Urban Water Journal.* 6 (4): 279-289.

Allen, M., Preis, A., Iqbal, M., and Whittle, A.J. (2012). "Case study: A smart water grid in Singapore". *Water Practice and Technology.* 7 (4).

AWWA. (2007). "Distribution System Inventory, Integrity and Water Quality". Edited by Office of Groundwater and Drinking Water Prepared for the Environmental Protection Agency. Washington.

Bakker, M., Van Schagen, K., and Timmer, J. (2003). "Flow control by prediction of water demand". *Journal of Water Supply: Research and Technology - AQUA.* 52 (6): 417-424.

Bakker, M., Vreeburg, J.H.G., Rietveld, L.C., and Van De Roer, M. (2012). "Reducing customer minutes lost by anomaly detection?", *Proceedings 14th Water Distribution Systems Analysis Conference 2012, WDSA 2012*, Adelaide (SA), Australia.

Bakker, M., Vreeburg, J.H.G., Palmen, L.J., Sperber, V., Bakker, G., and Rietveld, L.C. (2013a). "Better water quality and higher energy efficiency by using model predictive flow control at water supply systems". *Journal of Water Supply: Research and Technology - AQUA*. 62 (1): 1-13.

Bakker, M., Vreeburg, J.H.G., van Schagen, K.M., and Rietveld, L.C. (2013b). "A fully adaptive forecasting model for short-term drinking water demand". *Environmental Modelling and Software*. 48 (1): 141-151.

Bakker, M., Rajewicz, T., Kien, H., Vreeburg, J.H.G., and Rietveld, L.C. (2014a). "Advanced control of a water supply system: A case study". *Water Practice and Technology*. 9 (2): 264-276.

Bakker, M., Trietsch, E.A., Vreeburg, J.H.G., and Rietveld, L.C. (2014b). "Analysis of historic bursts and burst detection in water supply areas of different size". *Water Science and Technology: Water Supply*. (in press).

Beuken, R.H.S., Lavooij, C.S.W., Bosch, A., and Schaap, P.G. (2007). "Low leakage in the Netherlands confirmed", *Proceedings 8th Annual Water Distribution Systems Analysis Symposium 2006*, Cincinnati (OH), USA.

Brunone, B., and Ferrante, M. (2001). "Detecting leaks in pressurised pipes by means of transients". *Journal of Hydraulic Research*. 39 (5): 539-547.

Buchberger, S.G., and Li, Z. (2009). "Preserving spatial correlation in network water demands", *Proceedings 10th Annual Water Distribution Systems Analysis Conference, WDSA 2008*, Kruger National Park.

Chandola, V., Banerjee, A., and Kumar, V. (2009). "Anomaly detection: A survey". *ACM Computing Surveys*. 41 (3): 15:1-58.

Colombo, A.F., Lee, P., and Karney, B.W. (2009). "A selective literature review of transient-based leak detection methods". *Journal of Hydro-Environment Research*. 2 (4): 212-227.

Egan, J.P. (1975). "Signal detection theory and ROC Analysis, Series in Cognition and Perception": Academic Press, New York.

Eliades, D.G., and Polycarpou, M.M. (2012). "Leakage fault detection in district metered areas of water distribution systems". *Journal of Hydroinformatics*. 14 (4): 992-1005.

Farley, B., Mounce, S.R., and Boxall, J.B. (2013). "Development and field validation of a burst localization methodology". *Journal of water resources planning and management*. 139 (6): 604-613.

Gong, J., Lambert, M.F., Simpson, A.R., and Zecchin, A.C. (2013). "Single-event leak detection in pipeline using first three resonant responses". *Journal of Hydraulic Engineering*. 139 (6): 645-655.

Hanley, J.A., and McNeil, B.J. (1982). "The meaning and use of the area under a receiver operating characteristic (ROC) curve". *Radiology*. 143: 29-36.

Jung, D., Kang, D., Liu, J., and Lansey, K.E. (2013). "Improving resilience of water distribution system through burst detection", *Proceedings 2013 ASCE Environment and Water Resources Institue Annual Conference*, Cincinnati (OH), USA.

Khan, A., Widdop, P.D., Day, A.J., Wood, A.S., Mounce, S.R., and Machell, J. (2005). "Performance assessment of leak detection failure sensors used in a water distribution system". *Journal of Water Supply: Research and Technology - AQUA*. 54 (1): 25-36.

Lambert, A., and Thornton, J. (2011). "The relationships between pressure and bursts – a 'state-of-the-art' update". *Water 21*. (April): 37-38.

Lambert, A.O., Brown, T.G., Takizawa, M., and Weimer, D. (1999). "A review of performance indicators for real losses from water supply systems". *Aqua*. 48 (6): 227-237.

Machell, J., Mounce, S.R., Farley, B., and Boxall, J.B. (2014). "Online data processing for proactive UK water distribution network operation". *Drinking Water Engineering and Science.* 7 (1): 23-33.

Mays, L.W. (2000). "Water Distribution System Handbook". Edited by McGraw-Hill. New York (NY), USA.

Metz, C.E. (1978). "Basic principles of ROC analysis". *Seminars in Nuclear Medicine.* 8 (4): 283-298.

Misiunas, D., Lambert, M., Simpson, A., and Olsson, G. (2005a). "Burst detection and location in water distribution networks". *Water Science and Technology: Water Supply.* 5 (3-4): 71-80.

Misiunas, D., Vítkovský, J., Olsson, G., Simpson, A., and Lambert, M. (2005b). "Pipeline break detection using pressure transient monitoring". *Journal of Water Resources Planning and Management.* 131 (4): 316-325.

Misiunas, D., Vitkovský, J., Olsson, G., Lambert, M., and Simpson, A. (2006). "Failure monitoring in water distribution networks". *Water Science and Technology.* 53 (4-5): 503-511.

Mounce, S.R., Day, A.J., Wood, A.S., Khan, A., Widdop, P.D., and Machell, J. (2002). "A neural network approach to burst detection". *Water Science and Technology.* 45 (4-5): 237-246.

Mounce, S.R., Khan, A., Wood, A.S., Day, A.J., Widdop, P.D., and Machell, J. (2003). "Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system". *Information Fusion.* 4 (3): 217-229.

Mounce, S.R., and Machell, J. (2006). "Burst detection using hydraulic data from water distribution systems with artificial neural networks". *Urban Water Journal.* 3 (1): 21-31.

Mounce, S.R., and Boxall, J.B. (2010). "Implementation of an on-line artificial intelligence district meter area flow meter data analysis system for abnormality detection: A case study". *Water Science and Technology: Water Supply.* 10 (3): 437-444.

Mounce, S.R., Boxall, J.B., and Machell, J. (2010). "Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows". *Journal of Water Resources Planning and Management.* 136 (3): 309-318.

Mounce, S.R., Mounce, R.B., and Boxall, J.B. (2011). "Novelty detection for time series data analysis in water distribution systems using support vector machines". *Journal of Hydroinformatics.* 13 (4): 672-686.

Mounce, S.R., Mounce, R.B., and Boxall, J.B. (2012). "Identifying sampling interval for event detection in water distribution networks". *Journal of Water Resources Planning and Management.* 138 (2): 187-191.

Palau, C.V., Arregui, F.J., and Carlos, M. (2011). "Burst detection in water networks using principal component analysis". *Journal of Water Resources Planning and Management.* 138 (1): 47-54.

Pelletier, G., Milhot, A., and Villeneuve, J.P. (2003). "Modeling water pipe breaks - three case studies". *Journal of Water Resources Planning and Management.* 129 (2): 115-123.

Poulakis, Z., Valougeorgis, D., and Papadimitriou, C. (2003). "Leakage detection in water pipe networks using a Bayesian probabilistic framework". *Probabilistic Engineering Mechanics.* 18 (4): 315-327.

Puust, R., Kapelan, Z., Savic, D.A., and Koppel, T. (2010). "A review of methods for leakage management in pipe networks". *Urban Water Journal.* 7 (1): 25-45.

Romano, M., Kapelan, Z., and Savić, D.A. (2013). "Geostatistical techniques for approximate location of pipe burst events in water distribution systems". *Journal of*

Hydroinformatics. 15 (3): 634-651.

Romano, M., Kapelan, Z., and Savić, D.A. (2014). "Automated detection of pipe bursts and other events in water distribution systems". *Journal of Water Resources Planning and Management*. 140 (4): 457-467.

Røstum, J. (2000). "Statistical modelling of pipe failures in water networks", *Dr. thesis, Norwegian University of Science and Technology, Department of Hydraulic and Environmental Engineering, Trondheim, Norway*.

Srirangarajan, S., Allen, M., Preis, A., Iqbal, M., Lim, H.B., and Whittle, A.J. (2013). "Wavelet-based burst event detection and localization in water distribution systems". *Journal of Signal Processing Systems*. 72 (1): 1-16.

Thornton, J., and Lambert, A. (2006). "Managing pressures to reduce new breaks". *Water 21*. (DEC.): 24-26.

Thornton, J., Sturm, R., and Kunkel, G. (2008). "Water Loss Control", *2nd edition, McGraw-Hill, New York, USA*.

Trietsch, E.A., and Vreeburg, J.H.G. (2005). "Reliability of valves and section isolation". *Water Science and Technology: Water Supply*. 5 (2): 47-51.

Vreeburg, J.H.G., Vloerbergh, I.N., Van Thienen, P., and De Bont, R. (2013). "Shared failure data for strategic asset management". *Water Science and Technology: Water Supply* 13 (4): 1154-1160.

Worm, G.I.M., van der Helm, A.W.C., Lapikas, T., van Schagen, K.M., and Rietveld, L.C. (2010). "Integration of models, data management, interfaces and training support in a drinking water treatment plant simulator". *Environmental Modelling and Software*. 25 (5): 677-683.

WRc. (1994). "Managing leakage", *U.K. Water Industry Research Ltd/WRc, Wiltshire, England*.

Ye, G., and Fenner, R.A. (2011). "Kalman filtering of hydraulic measurements for burst detection in water distribution systems". *Journal of Pipeline Systems Engineering and Practice*. 2 (1): 14-22.

7 Analysis of historic bursts and burst detection in water supply areas of different size

Based on

M. Bakker
E.A. Trietsch
J.H.G. Vreeburg
L.C. Rietveld

Application of heuristic burst detection method in water supply areas of different size
Water Science and Technology: Water supply (in press)

Abstract
Pipe bursts in water distribution networks lead to water losses and a risk of damaging the urban environment. This chapter presents information for a better understanding of pipe bursts, based on an analysis of hydraulic data and customer contact records of 44 real (large) bursts. The information shows that most bursts are reported to the water company shortly after the beginning, and the negative consequences of the bursts are limited. However, smaller bursts that stayed unnoticed for a longer time period or larger bursts that began in the late evening or in the night are problematic to the water company that has no burst detection method installed. Detection of those bursts is critical to minimise the negative consequences, and a burst detection method can perform this task. This chapter presents the relation between the size of supply area and the size of the bursts that can be detected. For finding this relation, we applied a heuristic burst detection method on historic datasets of eight areas varying in size between 1,500 and 48,300 connections. We found a correlation between the size of the area and the minimum detectable burst size and quickly detectable burst size, for the areas we studied in combination with the method we applied. To reduce the risk of substantial water losses or damage to the urban environment, the burst detection method can effectively be applied to areas with an average demand of 150 m^3/h or less.

Keywords
Pipe bursts; water distribution; demand forecasting; detection system

7.1 Introduction

The risk to exhaust water resources is increasing rapidly around the world, due to population growth, urbanisation, pollution and climate change (Bates et al., 2008). This fact urges water companies to act responsibly and to enhance the efficiency of their systems. Important efficiency improvements are possible in the water distribution systems (WDS), which often have high rates of physical water losses. The amount of water loss in the UK systems for instance, is estimated between 15% (Wu et al., 2010) and 25% (Farley et al., 2013). These water losses are primarily caused by small background leakages or larger reported or unreported leakages (also indicated as bursts) due to pipe failures (Thornton et al., 2008). When trying to reduce the water loss, many water companies aim to achieve the economical level of leakage (ELL), where the costs of leakage control and the benefits of lower water losses are balanced (Trow and Farley, 2004). However, the main problem of pipe burst in the Dutch networks, does not relate to the background losses caused by small bursts, but to the damage caused by large bursts (see chapter 6). Therefore, this chapter will continue to mainly consider the large bursts in the network.

A basic instrument that may be used to detect pipe bursts is to monitor hydraulic data, recorded by the supervisory control and data acquisition (SCADA) system, with a simple “flat-line” alerting system (Mounce et al., 2010). However, such system has important limitations and detects only a small fraction of all bursts. A review of more advanced techniques that may be used for the detection of pipe bursts is presented in Puust et al. (2010). An example of such technique that has been studied in several papers, is analysing hydraulic parameters (flow and pressure) with complex statistical methods and machine learning techniques (e.g. Mounce et al. (2010) and Romano et al. (2014)). Another example is monitoring pressure transients that occur after a failure of a pipe in the WDS. By measuring pressure at high sampling rates, the propagation of the transient in the WDS can be observed.

Papers addressing the topic of pipe burst detection are mainly focussing on a detailed description of the detection method, and assessing the performance of the method in a selected area by using simulated or real data (e.g. Palau et al. (2011); Wu et al. (2010); Ye and Fenner (2011); Eliades and Polycarpou (2012); Aksela et al. (2009)). This research has led to the availability of very valuable detection methods and tools that can help the water industry with their pipe burst problem. However, an examination of the phenomena of pipe bursts itself has generally received less attention in the papers, as well as an assessment of the size of the bursts that can be detected in relation to the area where the method is applied. In order to gain a better understanding of bursts, we analysed and characterised a number of real (large) bursts in different areas. We also applied a burst detection method to these (differently sized) areas to analyse in depth the relation between the size of the area and the size of the detectable bursts. This knowledge is important when water companies consider introducing a burst detection method and create district metered areas (DMAs) or larger areas for this purpose. And finally, we analysed which bursts can be detected quickly (within 10-20 min.) and which bursts can only be detected with some delay (4-20 hours).

7.2 Materials and methods

7.2.1 Pipe failure data and hydraulic data

To obtain a better understanding of the phenomena of pipe bursts, we examined historic large burst events. Therefore, we collected pipe failure data and hydraulic data in the period 2012-2013 of five differently sized areas. Water company (Vitens) registered each pipe repair they executed in the main repair (MR) database, and the hydraulic data was stored in a central database system with 10 seconds intervals. The hydraulic data was measured by permanent flow and pressure sensors installed at the assets (treatment plants, reservoirs, boosters) in the area. We estimated the beginning and isolation point in time and the burst flow, by closely examining the detailed hydraulic data on days when bursts were reported in the MR database. For research purposes, water company Vitens constructed a number of virtual DMAs in one of the areas by installing flow (and pressure) sensors in the network. The data of three DMAs was used in this chapter. Figure 7.1 shows one of the sensor locations and the characteristics of the areas that were used in this chapter are shown in Table 7.1.

Figure 7.1 Installing sensors in the pipe and accompanied cabinet for the electronic equipment

Table 7.1 Characteristics of the researched areas (the # sensors column indicates how many flow sensors had to be used to make the water balance in that area)

Area	Available data	Avg. demand [m³/h]	# connections	# flow/pres. sensors
1. Leeuwarden	2012-2013	730	48,300	4
2. Buitenpost	2012-2013	380	26,600	3
3. Drachten	2012-2013	270	21,100	1
4. Burgum	2012-2013	150	11,100	2
5. Wolvega	2012-2013	85	6,800	1
6. DMA-Camminghaburen	Apr-Dec 2013	51	4,900	5
7. DMA-Westeinde	Apr-Dec 2013	16	1,500	3
8. DMA-Hemrik	Apr-Dec 2013	16	400	4

7.2.2 Burst detection and classification

In this chapter, we applied the method for burst detection that was first introduced in Bakker et al. (2013a) and described in depth in chapter 6 of this thesis (Bakker et al., 2014). The burst classification procedure of the method classifies an detected event by estimating the burst flow ($Q_{burst,estimated}$) and calculating the event confidence factor (*ECF*), which are defined as:

$$Q_{burst,estimated} = Q_{deviation,MA15} = \left(Q_{measured,MA15} - Q_{expected,MA15}\right) \qquad [m^3/h] \tag{7.1}$$

$$ECF = \max_{i=\{2,5,10,15,30,60,120,240\}}\left(\frac{Q_{deviation,MAi}}{Q_{threshold,MAi}}\right)\cdot 100\% \qquad [\%] \tag{7.2}$$

where $Q_{deviation,MA15}$ is the deviation between the measured ($Q_{measured,MA15}$) and expected ($Q_{expected,MA15}$) water demand of the 15-min. moving average timeframe. We chose the 15-min. moving average value, because this timeframe is sufficiently long to provide a rather stable value and sufficiently short to provide a good value shortly after the beginning of the burst. $Q_{deviation,MAi}$ and $Q_{threshold,MAi}$ are the deviation and the detection threshold values of the *i*-min. moving average timeframe. The method also calculates the deviations between expected and measured pressures in the area, and it designates which pressure measurement is affected most by the event. For monitoring a real WDS, we designed an informative and easy to understand user interface (see Figure 7.2).

Figure 7.2 User interface of the burst detection method prototype

The interface provides information about the burst and about the confidence level that the detected event is really a burst. It also shows the estimated burst flow, the pressure that is affected most by the burst, and on a geographical map how all pressures in the network are affected by the event.

7.2.3 Performance evaluation burst detection

We analysed the performance of the method by testing it on historic data of five areas of different size (see Table 7.1). The method was only tested at areas 1-5, because no bursts were reported in areas 6-8 in the period of which data was available. Although testing a method on live data might be more convincing, we used historic data which enabled us to assess the performance in a condensed time. When evaluating the performance, we calculated the detection probability (*DP*), the rate of false alarms (*RF*), and the burst detection time (*DT*) see equations (6.4) to (6.6).

7.3 Results

7.3.1 Analyses of pipe bursts

In Vitens' MR database 142 bursts were recorded in the selected areas in 2012-2013. All bursts occurred in areas 1-5; no bursts were recorded in the smaller (DMA scale) areas 6-8, see Table 7.1. After studying this data, we concluded that for 98 bursts the effect on the hydraulic parameters was smaller than the normal variation of the measured values. The trends of flow and pressure showed no abnormal values or changes in the values that could be linked to a burst. This can be explained by the fact that the areas 1-5 were quite large. The relative flow deviations caused by smaller bursts in a large area are too small to be observable. For 44 bursts, we observed abnormal hydraulic trends, like a sudden increase and decrease of the water demand. Of these 44 bursts we collected and analysed the parameters: evolution timeframe (the timeframe in which the bursts evolved from zero to maximum burst flow), beginning point in time, duration, burst flow, and total water loss. Note that real bursts were considered, and therefore these parameters were not measured but estimated by closely examining the detailed historic flow data. We also analysed how bursts were notified to the water company by studying the customer contact database. The results are shown in Figure 7.3.

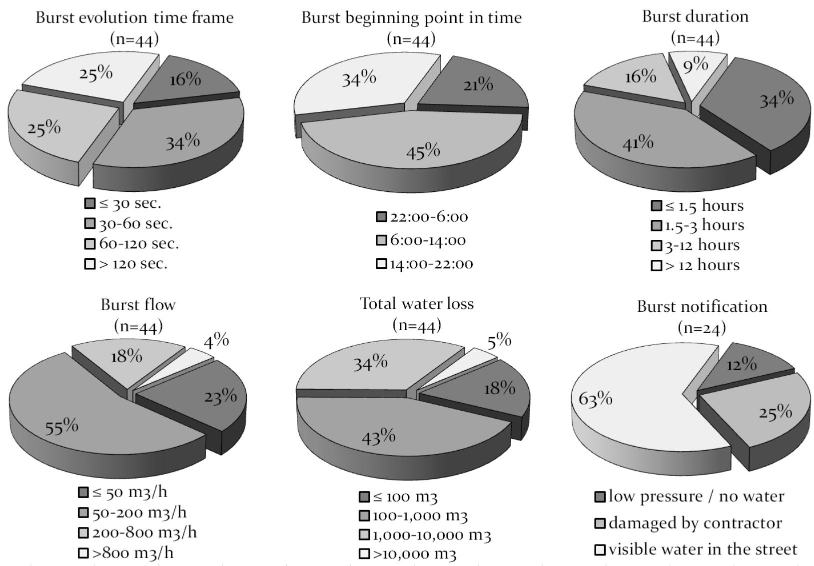


Figure 7.3 Characteristics of the examined bursts

We estimated the burst evolution timeframe by counting how many 10 seconds samples of flow data were present between the beginning of the burst (last sample with normal flow), and the burst being fully evolved (sample that maximum burst flow was observed). The upper graph of Figure 7.4 shows an example of a burst event including the detailed observation of the evolution timeframe. We found that 75% of the 44 examined large bursts evolved in two minutes or less. This shows that a sudden change in the measured water demand is a good indication of a pipe burst. Therefore, the change in water demand may be monitored to identify bursts. We found that the majority of the 44 bursts (80%) began during day time. That more bursts began during day time can be explained by the fact that a part of the bursts were caused by excavation activities by contractors or other third parties, which are only executed during day time.

We studied the duration of the bursts, which we defined as the timeframe in which the water ran freely from the burst pipe. This is the timeframe between the beginning and the isolation points in time. In this timeframe, the burst must be reported, found, and isolated by the water company. We found that the burst duration of 75% of the 44 large bursts was shorter than three hours. This indicates that most bursts were reported rather quickly to the water company, and the water company was able to stabilise the situation quickly. The burst duration of 16% of the bursts was between three and twelve hours, and of 9% of the bursts lasted longer than twelve hours. The burst in upper graph of Figure 7.4 had a short duration (two hours), and burst in the lower graph had a long duration: the burst was only repaired after 20 days.

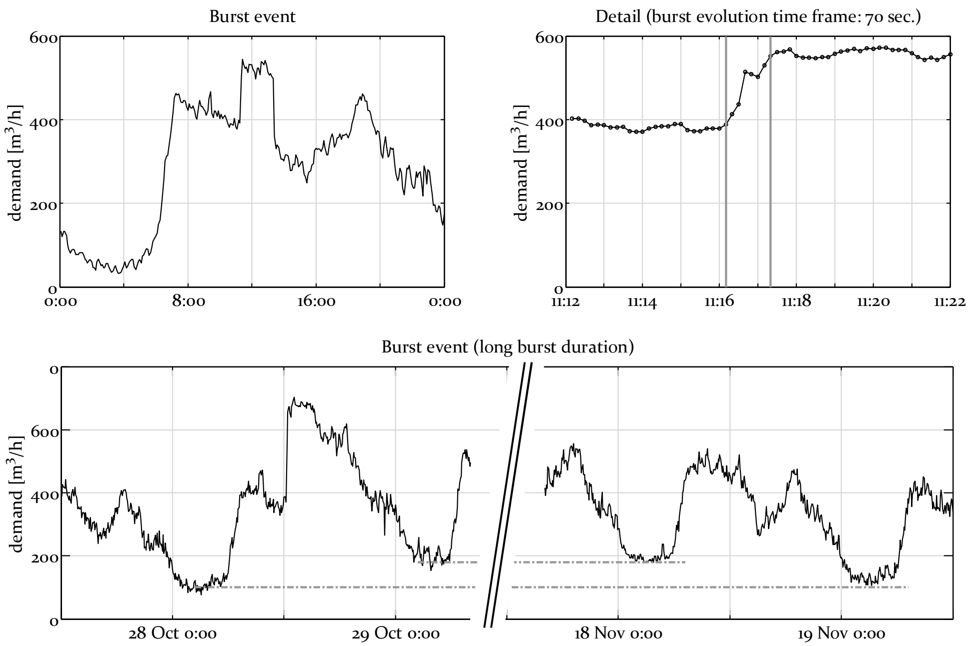


Figure 7.4 Calculation of the burst evolution timeframe (upper graph), and example of burst with long duration. The burst started on October 28th at 12:10, and was (fully) repaired on November 18th at 15:05

The burst flow of the examined bursts showed a large variation: the lowest flow was 20 m^3/h and the highest flow was over 2,800 m^3/h. This large variation was a result of the variation of the diameters of the burst pipes and the variation in local pressure. Most burst flows were between 50-200 m^3/h. We estimated the total water loss of the bursts by multiplying the estimated burst duration by the estimated burst flow. We found a large variation of the total water loss: the smallest total loss was 15 m^3 and the highest total loss was over 100,000 m^3. The highest water loss was caused by a burst that had a duration of 240 days and a burst flow of 20 m^3/h. Most of the total water losses were between 100-1,000 m^3. Finally, we studied how bursts were reported to the water company by tracking down the notifications in the customer contact database: 62% of the bursts were reported by people who observed water on street level; 25% were reported by contractor's staff who damaged a water main during excavation activities; The remaining 12% were reported by customers who experienced low pressure or no water at the tap.

7.3.2 Forecasting errors and burst detection performance

The water demand forecast appeared to be the dominant factor for the detection of pipe bursts; the pressure forecast was only supportive in the detection, which corresponds to the observations in chapter 6 of this thesis (Bakker et al., 2014). Therefore, we studied the errors of the water demand forecasts in the eight examined areas, see Table 7.2. The table shows both the percentage errors [%] and the absolute errors [m^3/h], expressed as root mean square error (*RMSE*) and mean absolute error (*MAE*).

Table 7.2 Forecasting errors water demand forecasts (15 min. time step forecasts)

Area	Demand [m^3/h]	*RMSE* [%]	*RMSE* [m^3/h]	MAE [%]	MAE [m^3/h]
1. Leeuwarden	730	8.2	59.2	5.4	38.9
2. Buitenpost	380	7.9	29.5	5.7	21.4
3. Drachten	270	9.2	24.2	5.9	15.4
4. Burgum	150	7.8	12.7	5.4	8.8
5. Wolvega	85	8.8	7.3	5.9	4.9
6. DMA-Camminghaburen	51	11.2	5.8	8.3	4.3
7. DMA-Westeinde	16	12.5	2.6	9.4	2.0
8. DMA-Hemrik (ind.)	16	48.0	8.1	35.1	6.0

We found that in smaller areas (areas with lower water demand), the percentage errors were larger but the absolute errors were smaller, which was in line with the observations in chapter 2 (Bakker et al., 2013b). Figure 7.5 shows examples of the forecasted and measured demand in two areas.

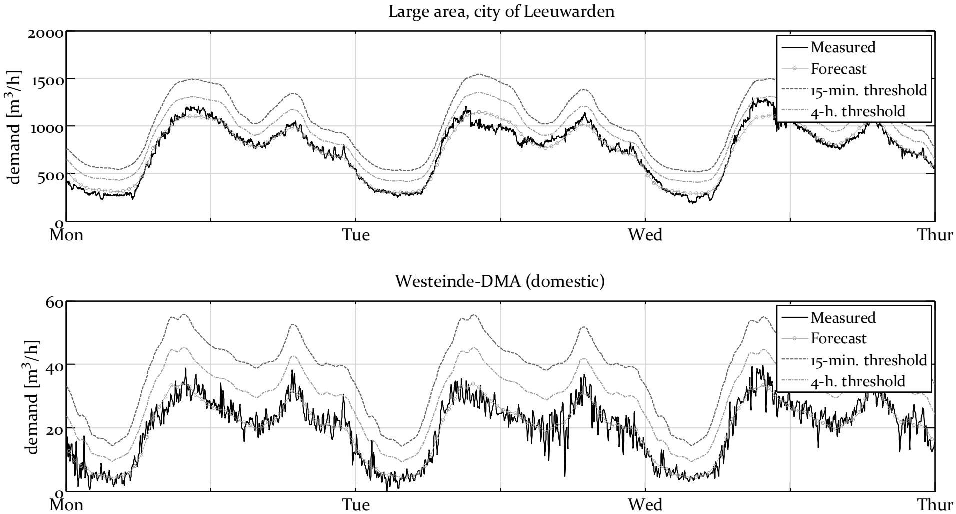


Figure 7.5 Trends of measured and forecasted demand in two different areas, including the 15-min. and 4-h. moving average detection threshold values

The graph shows small percentage differences between forecasted and measured values in the upper graph (large area), and larger percentage differences in the smaller area. Because the burst detection threshold value is closely related to the absolute demand forecast error, smaller bursts can be detected in areas with lower demand. This can be observed in Figure 7.5, where the difference between the forecasted demand and the threshold values were smaller (in absolute values [m^3/h]) in the smaller area. The graph also shows that the detection threshold values were lower when the moving average timeframe was larger, which is a known phenomenon (Montgomery, 2009). This indicates that smaller bursts could be detected when a longer moving average timeframe was applied.

The method's automatic procedure processed the 2011 data to derive the detection threshold parameters, and examined the 2012-2013 data to detect the bursts. We analysed whether the method was able to identify the bursts in the areas and we expressed the performance of the method in the performance parameters as defined in the previous chapter (section 6.2.3). We calculated the detection probability (*DP*) both related to the total number of bursts (142) and to the subset of burst that were visible in the flow data (44), see Table 7.3. The table shows that the method was able to detect 31.7% related to all bursts, and 88.3% related to the subset of visible bursts. The bursts were detected within 19 minutes after the beginning, while generating false alarms on 3.3% of the days. We considered any alarm that was raised on a day with no reported burst as a false alarm. Hence, we did not relate alarms to other possible causes / explanations (like operational activities or customer contacts), because we had no such information available. The performance of the method could not be evaluated in the DMAs, because no burst events occurred in the evaluated period.

Table 7.3 Performance of the burst detection method (DP_{All} is the detection probability, DP_{Large} is related to the burst that were visible in the flow data, *RF* is rate of false alarms, *DT* is detection time). Note that in area 6-8 no bursts were reported in the period, and therefore these areas are not included in the table.

Area	**# bursts all / large**	DP_{All} **[%]**	DP_{Large} **[%]**	***RF*** **[%]**	***DT*** **[min.]**
1. Leeuwarden	33 / 8	27.3	90.0	1.8	16.2
2. Buitenpost	32 / 8	25.0	80.0	2.2	16.2
3. Drachten	42 / 12	26.2	91.7	5.2	14.2
4. Burgum	15 / 6	40.0	100.0	3.6	14.2
5. Wolvega	20 / 10	40.0	80.0	3.9	32.6
Average	*142 / 44*	*31.7*	*88.3*	*3.3*	*18.7*

Table 7.3 shows that the detection probability increases as the size of the area decreases (see also Table 1, area size decreases from area 1. to 5.). This indicates that the detection method performs better in smaller areas. Table 7.3 also shows that the rate of false alarm (*RF*) varies considerably among the different areas. The high value in area 3. is quite remarkable, especially because the area is monitored by only one sensor (see Table 7.1). After examining the data, we observed quite deviating demand patterns during several periods, resulting in relative frequent false alarms in this area. We did not find any other specific variable that might explain the variability in the rate of false alarm.

7.3.3 Area size and size of detectable bursts

The sizes of the areas that we studied varied largely (see Table 7.1). To assess the detectable burst size, we studied the method's dynamic detection threshold values that were applied when monitoring the water demand in the different areas. Note that the results presented in this section are specific to the researched areas, the applied detection method, and the chosen parameters that determine the trade-off between hit rates and false alarms (*DP-RF*).

As explained in section 6.2.2 and illustrated in Figure 7.5, the threshold values of the method are (dynamically) related to the forecasted water demand and the moving average timeframe. By simultaneously monitoring transformed signals, the method was able to detect both large bursts very quickly, and smaller bursts with some delay. When applying the detection method to monitor an area, we were interested in knowing:

1. The quickly detectable burst size. We defined this burst size as the detection threshold value of the 10-min. moving average timeframe of the water demand with a 90% exceedance probability. This means that bursts of this size could be detected within 10 minutes during 90% of the time.
2. The minimum detectable burst size. This burst size is equal to the detection threshold value that occurred when demand was low (at night) and when the longest moving average timeframe was applied (4 hours).

We determined the quickly detectable and minimum detectable burst sizes in all researched areas, and plotted both the absolute values [m³/h] and the percentages against the average demand in the area (Figure 7.6).

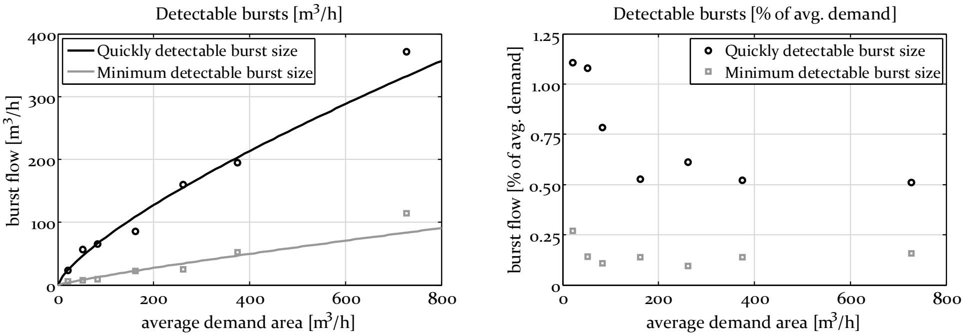


Figure 7.6 Detectable burst size as a function of the water demand in the area

The left graph of Figure 7.6 shows a correlation between the burst flow of on the one hand the quickly detectable bursts ($Q_{burst,quick}$) and minimum detectable bursts ($Q_{burst,min}$), and on the other hand the average demand in the area (Q_{area}). The correlation is approximated by:

$$Q_{burst,quick} = 2.48 \cdot Q_{area}^{0.74} \qquad R^2 = 0.98 \tag{7.3}$$

$$Q_{burst,min} = 0.27 \cdot Q_{area}^{0.87} \qquad R^2 = 0.93 \tag{7.4}$$

7.4 Discussion

The 44 historic bursts we examined in this chapter were a subset of all 142 reported bursts. Therefore, the observations presented in section 7.3.1. only apply to these large bursts that are visible in series of flow data. Moreover, all parameters –except the “burst notification”– of the 44 examined bursts were estimated by closely studying the detailed hydraulic data. We had to estimate those variables because they are not measured at real bursts. This means that the observations presented in section 7.3.1. have a degree of uncertainty. However, we are confident that our estimations are quite accurate, because we used very detailed data (10 sec. intervals). We approximate that our point in time estimation had an accuracy of ±30 sec. and our burst flow estimation had an accuracy of ±10-15%.

Based on the analysis of the historic bursts in the selected areas, we conclude that the added value of a burst detection method is limited for the majority of the bursts. Most bursts (98 of 142) were so small that they did not affect the hydraulic parameters, and were therefore undetectable. Of the bursts that were visible in the flow data (44 of 142), the burst duration of 75% was shorter than three hours. It is unlikely that a burst detection method could have shortened this duration. Furthermore, the total water loss of 61% of the bursts was less than 1,000 m^3, which indicates that financial loss due to lost water was limited for a large number of bursts. However, for five of the observed bursts, the consequences for total water loss, damage to the environment, or customer minutes lost (CML) were large. This was the case for the following three types of bursts:

1. Long running small sized bursts. We observed two of those (burst flow 20 and 40 m^3/h) that stayed unnoticed for a long period (320 days and 28 days). Only when the bursts developed to a larger size, they were noticed and repaired. The major part of the total water loss was caused by those bursts. An early detection would have reduced the total water loss in the system.
2. Medium sized bursts. We observed two of those (burst flow 140 and 150 m^3/h) that began in the late evening. Because a rather high flow ran from the burst pipes, they caused damage to the urban environment. The bursts stayed unnoticed until the next morning when people experienced a lower water pressure at the tap or observed the water in the street. An early detection of this type of bursts could have reduced the damage caused by the burst pipe.
3. Very large bursts. We observed one (burst flow 2,800 m^3/h) that caused a dramatic pressure drop in the entire network and as a result approximately 100,000 people had no water for three hours. One of the problems was that the burst could not be located quickly, because the broken pipe was located close to large open water. An early location of this burst would have helped to reduce the customer minutes lost (CML) caused by the burst pipe.

In order to detect the problematic small bursts (± 20 m^3/h, where a delay in the detection is acceptable) and the problematic medium sized bursts (± 100 m^3/h, within 10-20 min.), the size of the monitored area is restricted. From Figure 7.6 and equations (7.3) and (7.4) it can be derived that the average demand in the area should not exceed 150 m^3/h. If the sizes of the existing water supply areas are larger, new sensors need to be installed to created smaller areas to enable effective burst monitoring.

In addition, the results of Figure 7.6 and equations (7.3) and (7.4) are valid for the areas we examined in combination with the burst detection method we applied. We think the results are generally valid for other areas and other detection methods that are mainly based on monitoring water demand, although the constants in equations (7.3) and (7.4) may vary depending on the effectiveness of the detection method. This is based on the observation that the absolute water demand forecast errors increase as the size of the area increases, see chapter 2 of this thesis. Therefore, only larger deviations in measured water demand are significant to identify a burst in larger areas, and as a result the pipe burst that can be detected increases as the size of the area increases. However, if a detection method is used that not only monitors the total water demand in the area but also uses other sensors (e.g. (a dense network of) pressure sensors, noise sensors, et cetera), different relationships between the water demand in the area and the detectable burst sizes may be found.

Finally, we must stress that not only detecting but also locating of bursts is necessary to avoid the negative consequences of bursts. Therefore, the detection method should be accompanied by a burst location method or staff to perform active leakage detection activities (e.g. use of leak noise correlators) in order to locate the burst and enable quick repair. In several papers, e.g. Romano et al. (2013) and Farley et al. (2013), methods are proposed to estimate the burst location.

7.5 Conclusions

Pipe bursts are unavoidable and part of the normal operation of water distribution systems. To understand the phenomena of pipe burst better, we studied historic real pipe bursts that happened in networks in the Northern part of the Netherlands. We focussed on large bursts, and closely studied the 44 large (of 142) bursts. We found that the majority were not problematic: the events were reported shortly after the beginning, and because the water company responded adequately little water was lost and little damage was done to the urban environment. However, a small fraction of the large bursts was more problematic and an early detection could reduce the negative consequences. This was the case for small bursts that stayed unnoticed very long (multiple weeks or even months), and for medium sized bursts that began in the late evening or night and caused substantial damage to the environment. The financial consequences of the bursts can be quite large, and highly depend on the marginal costs of the lost water or on the damage that was caused.

We also studied the relation between the area size and the size of the detectable burst, and therefore we tested a heuristic burst detection method on a historic dataset containing hydraulic data and pipe burst date in eight differently sized areas. In the data that we analysed, we found that the detectable burst size was a function of the average area demand to the power n, where n was 0.74-0.87. When applying this correlation in the inversed way, the (maximum) size of the area can be determined in order to meet a target for detectable bursts. The analysis of 44 historic burst events showed that bursts of 100 m^3/h and larger should be detected quickly (with 10-20 min.) to avoid damage to the urban environment. Bursts of 20 m^3/h and larger should be detected (where a delay in the detection is acceptable) to avoid large water losses. The conclusion for the examined areas in combination with the heuristic burst detection method, is that the burst detection method should be implemented in areas with a maximum average demand of 150 m^3/h.

7.6 References

Aksela, K., Aksela, M., and Vahala, R. (2009). "Leakage detection in a real distribution network using a SOM". *Urban Water Journal.* 6 (4): 279-289.

Bakker, M., Jung, D., Lansey, K.E., Vreeburg, J.H.G., and Rietveld, L.C. (2013a). "Detecting pipe bursts using Heuristic and CUSUM methods", *Proceedings 12th International Conference on Computing and Control for the Water Industry, CCWI2013,* Perugia, Italy.

Bakker, M., Vreeburg, J.H.G., van Schagen, K.M., and Rietveld, L.C. (2013b). "A fully adaptive forecasting model for short-term drinking water demand". *Environmental Modelling and Software.* 48 (1): 141-151.

Bakker, M., Vreeburg, J.H.G., Van de Roer, M., and Rietveld, L.C. (2014). "Heuristic burst detection method using flow and pressure measurements". *Journal of Hydroinformatics.* (in press).

Bates, B.C., Kundzewicz, Z.W., Wu, S., and Palutikof, J.P. (2008). "Climate change and water", *Technical paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva*.

Eliades, D.G., and Polycarpou, M.M. (2012). "Leakage fault detection in district metered areas of water distribution systems". *Journal of Hydroinformatics.* 14 (4): 992-1005.

Farley, B., Mounce, S.R., and Boxall, J.B. (2013). "Development and field validation of a burst localization methodology". *Journal of water resources planning and management.* 139 (6): 604-613.

Montgomery, D.C. (2009). "Statistical Quality Control: A Modern Introduction". 6th edition ed: John Wiley & Sons, Inc.

Mounce, S.R., Boxall, J.B., and Machell, J. (2010). "Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows". *Journal of Water Resources Planning and Management.* 136 (3): 309-318.

Palau, C.V., Arregui, F.J., and Carlos, M. (2011). "Burst detection in water networks using principal component analysis". *Journal of Water Resources Planning and Management.* 138 (1): 47-54.

Puust, R., Kapelan, Z., Savic, D.A., and Koppel, T. (2010). "A review of methods for leakage management in pipe networks". *Urban Water Journal.* 7 (1): 25-45.

Romano, M., Kapelan, Z., and Savić, D.A. (2013). "Geostatistical techniques for approximate location of pipe burst events in water distribution systems". *Journal of Hydroinformatics.* 15 (3): 634-651.

Romano, M., Kapelan, Z., and Savić, D.A. (2014). "Automated detection of pipe bursts and other events in water distribution systems". *Journal of Water Resources Planning and Management.* 140 (4): 457-467.

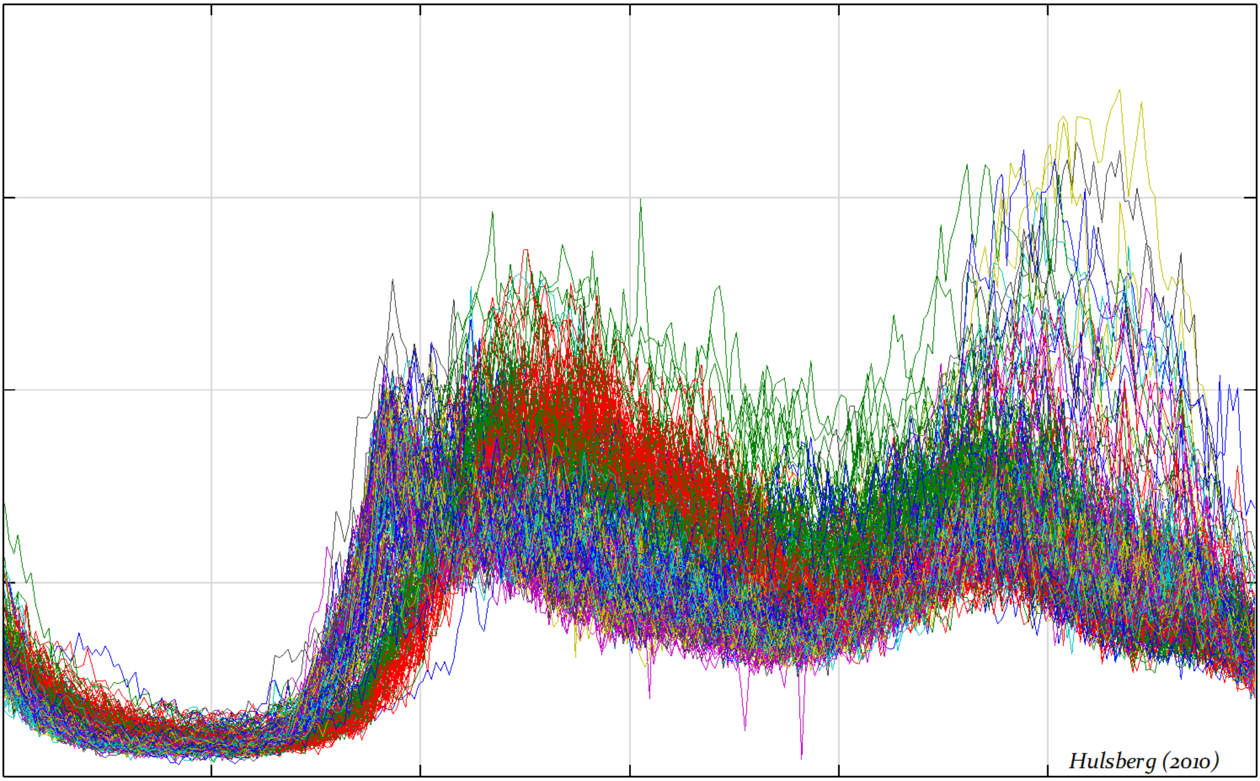
Thornton, J., Sturm, R., and Kunkel, G. (2008). "Water Loss Control", *2nd edition, McGraw-Hill, New York, USA*.

Trow, S., and Farley, M. 2004. Developing a strategy for leakage management in water distribution systems.

Wu, Z.Y., Sage, P., and 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.

Ye, G., and Fenner, R.A. (2011). "Kalman filtering of hydraulic measurements for burst detection in water distribution systems". *Journal of Pipeline Systems Engineering and Practice.* 2 (1): 14-22.

Part V – Concluding remarks



8 Conclusions and recommendations

8.1 General conclusions

In order to become in control of the water supply systems, there is a need to transform the conventional reactive control and monitoring to a more pro-active control and monitoring. Therefore, a heuristic water demand forecasting model was developed that can be applied to achieve this goal. This heuristic forecasting model is more comprehensible and may be more acceptable for the water companies' operating staff, compared to existing (more or less) complex mathematical models. The heuristic model is more comprehensible because it consists of a conversion of general observations of water demand into a series of straightforward calculation rules. The model is accepted as a leading tool in the Netherlands and controls approximately 50% of all water supply facilities in this country (www.aquasuite.eu). The performances of the heuristic model and a complex mathematical (transfer-/noise) model were almost equal: The average error of day forecasts was 2.81% of the mathematical model and 2.83% of the heuristic model. This indicates that the examined more complex model performed only slightly better than the heuristic model. The performance (expressed as forecasting errors) of the heuristic forecasting model highly depended on the area where it was applied: The average forecast errors were low in large (urban) areas, like the city of Amsterdam, and rather high in small (rural) areas. The percentage errors in small areas were approximately 350% higher than the percentage errors in large areas. Part of the water consumption is influenced by weather conditions, like water consumption for garden sprinkling, but also for washing and showering. Therefore, using weather information as (extra) input can improve the performance of water demand forecasting models. The average forecasting errors were reduced by 6.3% and the largest errors by 9.4% when using weather forecasts. However, using weather input will increase the complexity of the model, and will increase the risk of malfunctioning when the weather input is false or not available.

The conventional automatic operation of water supply systems often consists of simple rule-based methods (e.g. level based flow control). This reactive way of operations results in an ineffective way of using the reservoirs in the system, which causes unstable (production and transportation) flows with many flow changes. The possible enhancements of the flow control when using water demand forecasts were studied. The demand forecasts provide a projection of the amount of water that needs to be pumped or supplied out of the reservoirs in the next 48 hours. The control algorithm can find the optimal (production or transportation) flows to fill the reservoir by effectively using the reservoirs volume. The optimisation can be focussed on the water quality (by aiming at a constant production flow), energy costs (by aiming at a minimal energy consumption or optimally using the different energy tariffs), reduction of water losses (by minimizing the pressure in the network), or a combination. A heuristic water demand forecasting model in combination with an analytical control method was tested at five water supply systems in the Netherlands and at one system in Poland. A period with optimised control was compared with an equal period with conventional control. It was observed at the systems in the Netherlands that with optimised control the average turbidity in the clear water was 17% lower, the average energy consumption 3.1% lower and the energy costs 5.2% lower. At the Polish system, it was observed that optimised control led to a 18% lower back-ground leakage and 11.5% lower energy costs. The research at these six full scale water supply systems proved the benefits of optimised control, and resulted in tangible savings in the day to day operation of the water supply systems.

A water demand forecasting model is not only valuable with respect to the real-time control, but can also be used for pipe burst detection in water distribution systems. In the current practice, most water companies in the Netherlands rely on customers to report a pipe burst when they experience low water pressure or observe water running in the street. As a result, the companies are only aware of a burst a considerable time after the beginning, which causes high water losses and an increased risk of damage to the urban environment due to the uncontrolled water flow. Forecasting models are capable of generating fairly accurate estimates of water flows and pressures under normal operational circumstances. A pipe burst causes abnormal flows and pressures in the water distribution network, and by comparing expected (forecasted) with measured flows and pressures, pipe bursts can be detected. A burst detection method was developed that uses heuristic water demand and pressure forecasting models, and raises an alarm when the deviation between the measured and forecasted value exceeds a dynamic threshold value. By taking the moving average value of signals that are monitored over different time frames, the method was both able to detect large bursts quickly and smaller bursts with some delay.

The method was tested on a number of historic datasets with hydraulic and pipe burst data of different areas varying in size from very large (Rhine area (Dunea): 130,920 connections) to small (Westeinde-DMA (Vitens): 1,500 connections). The forecasting model performed relatively worse in smaller areas (generating higher percentage forecast errors), and therefore the percentage pipe burst detection threshold values were higher as well. However, because the average demand is lower in smaller areas, the *absolute* values of the detection threshold values (in m^3/h) are lower in smaller areas. As a result, the size of pipe burst that can be detected decreases as the size of the area decreases. The examination of 44 historic large pipe burst in the Northern part of the Netherlands, showed that medium sized bursts (around 100 m^3/h) should be detected quickly within 10-20 minutes to prevent damage to the urban environment. Small size bursts (around 20 m^3/h) should also be detected to avoid large water losses, but some delay in the detection is allowed. Both types of bursts can be detected if the proposed burst detection method is applied in those areas with an average demand of 150 m^3/h or less. The costs for implementing the method at this scale are low if the flow and pressure measurements are already installed and real-time available. If this is the case, the reduced water losses and reduced risk of damage to the environment may pay-back the investment and operational costs of having the method. If new flow and pressure measurements need to be installed in the network to create areas of 150 m^3/h, the costs can be considerable. In such case, the costs of the pipe bursts must be weighed against the costs of installing the measurements to determine the economic feasibility of implementing the burst detection method.

From this thesis, it can be concluded that the heuristic water demand forecasting model is valuable in the operation of water supply systems. The benefits of the forecasting model when applied for real-time control have explicitly been proven in experiments at six full scale water supply systems in Poland and the Netherlands. The possible benefits of the forecasting model when applied for pipe burst detection have been studied with off-line data. This analysis showed that the forecasting model in combination with the proposed detection method can only detect large (catastrophic) bursts in the large water distribution networks that are common in the Netherlands. This limits the benefits of the forecasting model for burst detection.

8.2 Recommendations and future outlook

A pipe burst detection method by itself will not reduce the negative consequences of a burst: the water loss and damage to the environment only stops when the burst pipe is isolated in the field by closing the right valves. Therefore, water companies need tools that help them to pinpoint the location of the burst pipe. The detection method we developed already shows which pressure measurements are influenced most by the burst pipe which may be helpful in locating the burst. However, this is not sufficient to effectively direct the repair team to the right location, especially if the area is large. Promising results in the development of burst location methods have already been published (e.g. Ye and Fenner (2011)), but further research is recommended to develop methods with increased accuracy and practical applicability. A possible burst location method that not yet (or very little) been has studied, is based on the statistical analysis of pressure measurements from a very high density network of pressure sensors. Such very high density sensor network may become available when the conventional (domestic) water meters are replaced with smart meters that are equipped with a pressure sensor and data transfer module. When fully deployed, each house connection would be a sensor location. With such high sensor density and sensors sending the data in (near) real-time, it is expected that burst can be detected and located more accurately. Research is necessary in the field of handling large quantities of data, interpretation of the data, and privacy considerations related to using information from smart (domestic) meters.

This thesis focusses on the detection of pipe burst by analysing hydraulic data. Hydraulic data may also be used to detect contamination events in case the contamination is caused by an illegal or accidental cross connection of the water supply network with another pressurised system. In such case, the inflow of non-drinking water in the water supply network will (locally) influence the hydraulic parameters in the network. When the inflow is relatively large, the deviations of the hydraulic parameters may be sufficient to identify the contamination event. However, the burst detection method develop in this thesis only responds to flow increases and pressure decreases, and will therefore not detect possible anomalous water inflows in the network. Future research is recommended to study the phenomena of unwanted inflow of non-drinking water in the water supply network, and the possibilities to detect such events by using forecasting models.

The approach of using heuristic, adaptive models may be valuable in the control of other processes in the water supply system as well. For example, the conventional operation of the treatment processes in a water treatment plant shows the same characteristics as the conventional control of the water distribution networks: control and monitoring are based on static switching and alerting levels. Simplified models (continuously calibrated with real-time measurements) that describe the water treatment processes can be valuable in the control and monitoring of these processes. Therefore, future research is recommended to develop and apply heuristic water quality models to enhance the performance of drinking water treatment plants. This means a continuation of the development of process models in the Stimela modelling environment (www.stimela.com) of the Delft University of Technology, and further research to applying the models for control.

Finally, this thesis has not been able to compare the performance of the proposed forecasting and detection methods with existing methods. An objective comparison appeared not to be possible because the performance is very site- and dataset specific. To allow an objective comparison among methods, the methods should be tested on the same datasets. Therefore it is recommended that researchers share their data, and make their datasets available to other researchers.

List of publications

Peer reviewed journals

Bakker, M., Van Schagen, K., and Timmer, J. (2003). "Flow control by prediction of water demand". *Journal of Water Supply: Research and Technology - AQUA.* 52 (6): 417-424.

Bakker, M., Vreeburg, J.H.G., Palmen, L.J., Sperber, V., Bakker, G., and Rietveld, L.C. (2013). "Better water quality and higher energy efficiency by using model predictive flow control at water supply systems". *Journal of Water Supply: Research and Technology - AQUA.* 62 (1): 1-13.

Bakker, M., Vreeburg, J.H.G., van Schagen, K.M., and Rietveld, L.C. (2013). "A fully adaptive forecasting model for short-term drinking water demand". *Environmental Modelling and Software.* 48 (1): 141-151.

Bakker, M., Vreeburg, J.H.G., Van de Roer, M., and Rietveld, L.C. (2014). "Heuristic burst detection method using flow and pressure measurements". *Journal of Hydroinformatics.* (in press).

Bakker, M., Trietsch, E.A., Vreeburg, J.H.G., and Rietveld, L.C. (2014). "Application of heuristic burst detection method on water supply areas of different size". *Water Science and Technology: Water Supply.* (in press).

Bakker, M., Rajewicz, T., Kien, H., Vreeburg, J.H.G., and Rietveld, L.C. (2014a). "Advanced control of a water supply system: A case study". Water Practice and Technology. 9 (2): 264-276.

Other journals and conference proceedings

Bakker, M., Verberne, A.J.P., and van Schagen, K.M. (1998). "The benefits of demand forecasting and modelling". *Water Quality International.* (May-June): 20-22.

Bakker, M., and Rietveld, L.C. (1998). "Combining quantity control and quality modelling", *Proceedings IWSA workshop: Trends in on-line monitoring and control of water supply*, Amsterdam, Netherlands.

Bakker, M., and Schaap, M.E. (2010). "Intelligent control of the water supply system in the province of Groningen". *Arab Water World (AWW).* XXXIV (1): 61-63.

Bakker, M., Verberk, J.Q.J.C., Palmen, L.J., Sperber, V., and Bakker, G. (2011). "Higher energy efficiency and better water quality by using model predictive flow control at water supply systems", *Proceedings Water and Innovation: Water Technology, International Water Week, 2011*, Amsterdam, The Netherlands.

Bakker, M., Verberk, J.Q.J.C., Palmen, L.J., Sperber, V., and Bakker, G. (2011). "Quantifying the benefits of model predictive flow control at different water supply systems in the Netherlands ", *Proceedings Water Convention Conference, Singapore International Water Week, SIWW 2011*, Singapore, Singapore.

Bakker, M., Lapikas, T., Tangena, B., and Vreeburg, J.H.G. (2012). "Monitoring water supply systems for anomaly detection and response", *Proceedings New developments in IT & Water*, Amsterdam, The Netherlands.

Bakker, M., Vreeburg, J.H.G., Rietveld, L.C., Blom, T., and Van De Roer, M. (2012). "The use of an adaptive water demand prediction model", *Proceedings 14th Water Distribution Systems Analysis Conference 2012, WDSA 2012*, Adelaide (SA), Australia.

Bakker, M., Vreeburg, J.H.G., Rietveld, L.C., and Van De Roer, M. (2012). "Reducing customer minutes lost by anomaly detection?", *Proceedings 14th Water Distribution Systems Analysis Conference 2012, WDSA 2012*, Adelaide (SA), Australia.

Bakker, M., Vreeburg, J.H.G., Van de Roer, M., and Sperber, V. (2012). "Detecting pipe bursts by monitoring water demand", *Proceedings Waterloss Europe 2012*, Ferrara, Italy.

Bakker, M. (2013). "Monitoring van het distributienet om afwijkingen te detecteren (in Dutch)", *Proceedings 65e Vakantiecursus 2013: 'De Watercyclus; daar zit wat in!'*, Delft, The Netherlands.

Bakker, M., Jung, D., Lansey, K.E., Vreeburg, J.H.G., and Rietveld, L.C. (2013). "Detecting pipe bursts using Heuristic and CUSUM methods", *Proceedings 12th International Conference on Computing and Control for the Water Industry, CCWI2013*, Perugia, Italy.

Bakker, M., Rajewicz, T., Kien, H., Vreeburg, J.H.G., and Rietveld, L.C. (2013). "Reducing energy consumption and leakage by active pressure control in a water supply system", *Proceedings IWA conference on Asset Management for Enhancing Energy Efficiency in Water and Wastewater Systems, 2013*, Marbella, Spain.

Bakker, M., Rajewicz, T., Kien, H., Vreeburg, J.H.G., and Rietveld, L.C. (2013). "Advanced control of a water supply system: A case study", *Proceedings OZ Water, 2013*, Perth, WA.

Bakker, M., Van Duist, H., Van Schagen, K.M., Vreeburg, J.H.G., and Rietveld, L.C. (2013). "Improving the performance of water demand forecasting models by using weather input", *Proceedings 12th International Conference on Computing and Control for the Water Industry, CCWI2013*, Perugia, Italy.

Bakker, M., Vreeburg, J.H.G., Van de Roer, M., and Rietveld, L.C. (2014). "Burst detection using forecasting models", *Proceedings IWA Water Loss 2014 Conference*, Vienna, Austria.

Tangena, B.H., Schijven, J.F., Foret, J.M., and Bakker, M. (2012). "Development and use of a decision support system for the response to drinking water contamination events", *Proceedings 5th Water Contamination Emergencies: managing the threats, WCEC5*, Mülheim-an-der-Ruhr, Germany.

Van Schagen, K.M., Bakker, M., Rietveld, L.C., Veersma, A.M.J., and Babuska, R. (2006). "Using on-line quality measurements in drinking water process control", *Proceedings Water Quality Technology Conference and Exposition 2006: Taking Water Quality to New Heights*, Denver (CO), USA.

National publications

Bakker, M., Peters, E.G.J., and Verberne, A.J.P. (1997). "Productiebesturing van drinkwaterpompstations (in Dutch)". *H2O.* 30 (15): 466-469.

Bakker, M., and Van Schagen, K.M. (1998). "Optimalisatie van de inzet van win- en productiemiddelen. Van kwantitatief naar kwalitatief (in Dutch)", *Proceedings Euroseminar Procesautomatisering in de Drinkwaterproductie en Afvalwaterzuivering, 1998*, Amsterdam, The Netherlands.

Bakker, M., Van Schagen, K.M., Geusen, R., and Keuning, J. (2000). "Sproeien in beeld gebracht (in Dutch)". *H2O.* 33 (18): 33-35.

Bakker, M., Van Schagen, K.M., and Backus, N. (2002). "Prognosebesturing in een complex voorzieningsgebied (in Dutch)". *H2O.* 35 (20): 27-30.

Bakker, M., Brunklaus, O., and Swildens, B. (2010a). "Leidingbreuk eerder opmerken door bewaking uitgaande hoeveelheid drinkwater (in Dutch)". *H2O.* 43 (16): 10-11.

Bakker, M., Hartel, O.H., Munneke, J., and Aagten, E. (2010b). "Automatische distributiesturing in Groningen (in Dutch)". *H2O.* 43 (4): 26-29.

Siarkiewicz, A., Maciołek, A., Bakker, M., Karolczak, P., and Rajewicz, T. (2014a). "The optimisation of the control system of the Water Treatment Plant (WTP) Gruszczyn - Part I (in Polish)". *Gaz, Woda I Technika Sanitarna.* (February 2014): 53-56.

Siarkiewicz, A., Maciołek, A., Bakker, M., Karolczak, P., and Rajewicz, T. (2014b). "The optimisation of the control system of the Water Treatment Plant (WTP) Gruszczyn - Part II (in Polish)". *Gaz, Woda I Technika Sanitarna.* (March 2014): 89-93.

Verheijden, S., Zwamborn, M., and Bakker, M. (1999). "OS-PIM: On-line advisering voor optimale bedrijfsvoering (in Dutch)". *H2O.* 32 (22): 32-33.

List of abbreviations

ANN: Artificial neural network
ARIMA: Auto regressive integrated moving average
AUC: Area under curve
BIS: Bayesian inference system
CML: Customer minutes lost
CUSUM: Cumulative sum method
DisConTO: Distribution Control Training & Operation
DMA: District metered areas
DP: Detection probability
DPCM: Dynamic pressure control module
DT: Detection time
ECF: Event confidence factor
ELL: Economical level of leakage
FC: Flow controlled
ILI: Infrastructure leakage index
IWA: International Water Association
MA: Moving average
MAE: Mean absolute error
MAPE: Mean absolute percentage error
MLR: Multiple linear regression
MPC: Model predictive control
MR: Main repair
NSE: Nash-Sutcliffe model efficiency
OPIR: Optimal Production through Intelligent control
ORP: Oxidation reduction potential
PPB: Parts per billion
PV: Production variation
R^2: Coefficient of determination
RE: Relative error
RF: Rate of false alarms
PCA: Principle component analysis
PID: Proportional integral derivative
PLC: Programmable logic controller
PRV: Pressure reducing valves
PS: Pumping station
ROC: Receiver operating characteristics
RMSE: Root mean square error
RRMSE: Relative root mean square error

SCADA: Supervisory control and data acquisition
SD: Standard deviation
SOM: Self-organizing map
SVM: Support vector machines
VSD: Variable speed drive
WDS: Water distribution systems
WTP: Water treatment plant

Acknowledgements

Doing PhD research and writing a thesis is impossible without the help and support of many people, and I thank all people who were somehow involved in this process. First of all I thank my father, Jaap Bakker, and Professor Hans van Dijk who both made me interested in water supply and inspired me to become a water supply engineer. Next I thank everyone who enabled the development and implementation of OPIR, the control software for water supply systems. I thank not only my colleagues, especially Kim van Schagen and Hans van der Kolk, who helped me creating and expending the functionality, but also all the water companies that showed their confidence and decided to use the software for the control of their precious infrastructure. Being a water supply engineer, and having knowledge about water demand forecasting, made me decide to start working on this research project.

The work was carried out in the DisConTO project, and I thank all parties that were involved in this project. In the first place, I thank Igasz Worm and Jasper Verberk who wrote the winning project plan, that was awarded with an Innowator grant that made the project possible. Next, I thank the members of my work package within the DisConTO project: Ben Tangena (RIVM) for covering all the organisational parts; Eelco Trietsch, Ton Blom and Geo Bakker (Vitens), Veerle Sperber (Brabant Water), Maurice van de Roer (Dunea), and Henk van Duist and Martin Klein-Arfman (PWN) for supplying data and/or being co-author of one of my papers. I also thank Edwin Blaauwgeers who did a great job as the overall project leader of the DisConTO project. Next, I thank Gerard van Houwelingen, who drew my attention to the vacant position of researcher at the Delft University of Technology within the DisConTO project. And finally, I thank Eric Zandbergen, who provided me the support from Royal HaskoningDHV to work on this project.

All scientific papers and conference proceedings on which this thesis is based, were written in close cooperation with my supervisors Luuk Rietveld (Delft University of Technology) and Jan Vreeburg (Wageningen University). I thank you both for the inspiring discussions about the project and for your valuable comments on the draft papers. Your contributions definitely improved the quality of the papers.

Although the support of all professionals mentioned above is very important, a stable and supportive environment at home is invaluable. My terrific kids, Robin and Joey, (almost) never complained that I spent many private hours working on this thesis instead of playing with them. And Mariska, thank you so much for your support, and creating the conditions in which I could concentrate on this work. But also for taking me out once in a while to dance all night and show me that there is more in life than this research work. *Oh my baby, baby, I love you more than I can tell...*

Curriculum vitae

Martijn Bakker was born in Niedorp, the Netherlands, on 24 December 1970. In 1989 he graduated from the Rijks Scholen Gemeenschap (R.S.G.) Schagen, and in 1990 he started his studies at the Delft University of Technology, Faculty of Civil Engineering. For writing his master thesis, Martijn did an internship at engineering and consultancy firm DHV, and studied water demand forecasting and advanced production flow control. Directly after his graduation in 1995, the internship at DHV was converted in employment and Martijn started his career as a water supply specialist. One of his first professional assignments was further developing the forecasting and controlling model for the automatic control of a single water treatment plant. At the end of the project in January 1996 the model -named OPIR- was implemented, and until now the implemented model has controlled the plant continuously. A substantial part of his following professional life, Martijn worked on enhancing the model's functionality and implementing the model at other water supply systems. In 2011, Martijn Bakker started a part-time PhD research at the Delft University of Technology. The research project provided the option to further enhance the forecasting and controlling model and to scientifically prove its effectiveness. It also provided the option to develop a new field of application of the model, namely pipe burst detection.

If you know quite precisely what will happen in the next one to two days, you can execute your activities systematically and effectively. You are in control and the chance of unpleasant surprises is reduced to a minimum. This is the essence of this thesis, but then applied to water supply systems. The repetitive nature of water demand makes the water demand very well forecastable by an adaptive model. The water demand forecasts can be used to optimise the operation of the systems and to detect deviating behaviour caused by a pipe burst. This thesis explores the possibilities of forecasting the water demand, and describes the efficiency improvements and detection capabilities that can be achieved by using these forecasts.