University of Thessaly

Advanced tools for urban water demand simulation and forecasting at multiple spatio-temporal scales

Dissertation for the degree Doctor of Philosophy in Civil Engineering

By

Dimitrios T. Kofinas

supervisor:

Professor Chrysi S. Laspidou

Volos, Greece

December 2019

Dissertation committee:

1. Prof. Chrysi Laspidou, department of Civil Engineering, UTH, member of the advisory committee
2. Prof. Vasilis Kanakoudis, department of Civil Engineering, UTH, member of the advisory committee
3. As. Prof. Christos Makropoulos, department of Water Resources and Environmental Engineering, School of Civil Engineering, NTUA, member of the advisory committee
4. Prof. Antonios Liakopoulos, department of Civil Engineering, UTH
5. Prof Nikitas Mylopoulos, department of Civil Engineering, UTH
6. Prof. Nikolaos Theodosiou, department of Civil Engineering, AUTH
7. Prof. Konstantinos Katsifarakis, department of Civil Engineering, AUTH

Contents

Acknowledgements

I feel grateful to my scientific supervisor Professor Chrysi Laspidou for opening a whole new world to me. Except for the supervision of my research work and conclusive scientific guidance, she always stood by me in all practical yet substantial matters, including finance and scientific network building. Most importantly, she instilled virtues that are fundamental to a scientist's mentality, such as interdisciplinarity, innovative thinking, silos breaking, a holistic view, a humanitarian approach to scientific implementation.

I would like to acknowledge the invaluable contribution of Professor Vasilis Kanakoudis and Associate Professor Christos Makropoulos with regards to their scientific guidance and insightful view to my research work.

I would like to express my sincere gratitude to Professors Antonis Liakopoulos, Nikitas Mylopoulos, Nikolaos Theodosiou and Konstantinos Katsifarakis for evaluating my thesis as committee members.

I feel obliged to acknowledge the contribution of Mr Bambos Charalambous concerning his generous sharing of knowledge in the field of Urban Water Supply.

I feel really indebted to Associate Professor Elpiniki Papageorgiou, Dr Konstantinos Kokkinos, Dr Alexandra Spyropoulou and Mr Rafał Ulańczyk for their support, where it came to their specific fields.

My feelings of thankfulness should also be expressed to Mr Nikolaos Mellios and Mr Ioannis Sarris for their support regarding the technical specifications of the case study.

I would also like to thank Ms Alexandra Ioannou and Mr Stylianos Mimis for their ethical support and excellent collaboration during the working hours under the same laboratory.

I thank my family and friends from the bottom of my heart, for being my backbone throughout my study.

Extended Abstract (in Greek)

Εκτενής Περίληψη

Εξελιγμένα εργαλεία για την προσομοίωση και την πρόβλεψη ζήτησης αστικού νερού σε πολλαπλές χωρο-χρονικές κλίμακες

Αντικείμενο της διατριβής είναι η διερεύνηση των δυνατοτήτων μαθηματικής προσομοίωσης διάφορων τμημάτων του υδρευτικού κύκλου, η εφαρμογή και η ανάπτυξη νέων εργαλείων, συγκεκριμένα της ζήτησης νερού στο επίπεδο των υδάτινων πόρων, του συστήματος διανομής και της οικιακής κατανάλωσης σε επίπεδο βρύσης.

Για την υλοποίηση των παραπάνω μαθηματικών προσομοιώσεων χρησιμοποιήθηκαν παραδοσιακές στατιστικές μέθοδοι, μέθοδοι που προέρχονται από το χώρο της τεχνητής νοημοσύνης και υβριδικές μέθοδοι, καθώς και διαδεδομένοι προσομοιωτές, ενώ αναπτύχθηκαν νέοι αλγόριθμοι για να καλύψουν περαιτέρω ανάγκες.

Οι μέθοδοι εξετάστηκαν και βαθμονομήθηκαν στην περίπτωση μελέτης της πόλης της Σκιάθου, όπου αποτελεί ένα παράδειγμα έντονα δυναμικού χαρακτήρα με περιοδικότητα στη ζήτηση που σχετίζεται με την τουριστική δραστηριότητα του νησιού και με την περιοδικότητα των μετεωρολογικών μεταβλητών. Τέλος, αναπτύχθηκαν δείκτες απόδοσης και χρησιμοποιήθηκαν τεχνικές οπτικοποίησης, ώστε να διευκολυνθεί η κατανόηση του συστήματος και να αναδειχθούν οι σχέσεις αιτίου- αιτιατού.

Η διατριβή διαρθρώνεται σε έξι κεφάλαια. Το πρώτο κεφάλαιο αφορά στα κίνητρα και στους αντικειμενικούς στόχους. Το δεύτερο κεφάλαιο αφορά στη συνεισφορά της διατριβής και στα πρωτότυπα σημεία. Το τρίτο κεφάλαιο αφορά στις μεθόδους πρόβλεψης της ζήτησης νερού, το τέταρτο στην ανάπτυξη ενός χωρο-χρονικού μοντέλου προσομοίωσης του υδρευτικού συστήματος, το πέμπτο στην ανάπτυξη ενός μοντέλου προσομοίωσης της οικιακής ζήτησης νερού σε επίπεδο βρύσης και το έκτο αφορά στα συμπεράσματα της διατριβής.

Πιο συγκεκριμένα, στο πρώτο κεφάλαιο γίνεται μια ανάλυση της σημερινής κατάστασης, ως προς τη διαχείριση του αστικού νερού, σε διάφορες κλίμακες, των προτεραιοτήτων που έχουν τεθεί σε επίπεδο στρατηγικού σχεδιασμού, καθώς και η τοποθέτηση της συγκεκριμένης διατριβής σε αυτό το πλαίσιο.

Στο δεύτερο κεφάλαιο απαριθμούνται τα βασικά βήματα της εργασίας και τονίζονται τα σημαντικά πρωτότυπα σημεία.

Στο τρίτο κεφάλαιο γίνεται μία ιστορική αναδρομή των μεθόδων προσομοίωσης και πρόβλεψης της ζήτησης νερού, οι οποίες κατηγοριοποιούνται ανάλογα με διάφορα κριτήρια, όπως η κλίμακα χρόνου, η θεώρηση ως προς την αβεβαιότητα (στοχαστικές ή αιτιοκρατικές προσεγγίσεις), ο αριθμός των εκτιμητριών (μονοπαραγοντικές ή πολυπαραγοντικές προσεγγίσεις) και ο ορίζοντας

(κοντοπρόθεσμος, μεσοπρόθεσμος, μακροπρόθεσμος). Στη συνέχεια χρησιμοποιούνται διάφορες μέθοδοι από την παραδοσιακή στατιστική, τις μεθόδους μηχανής εκμάθησης αλλά και υβριδικές για την μοντελοποίηση της ζήτησης του αστικού νερού της Πόλης της Σκιάθου. Η προσομοίωση γίνεται με διάφορα χρονικά βήματα (τριμήνου, μήνα, ημέρας), ενώ κάποιες από την μεθόδους που χρησιμοποιούνται είναι το Αυτοπαλίνδρομο Ολοκληρωμένο Υπόδειγμα Κινητού Μέσου (ARIMA), τα Τεχνητά Νευρωνικά Δίκτυα (ANN), το Προσαρμοστικό Σύστημα Νευροασαφούς Επαγωγής (ANFIS) και συνδυασμοί αυτών. Η ανάλυση γίνεται μονοπαραγοντικά και πολυπαραγοντικά, ενώ διερευνώνται διάφορες ενδεχόμενες εκτιμήτριες, με τη μέθοδο δοκιμής-λάθους, όπως μετεωρολογικές, η τουριστική δραστηριότητα και η τεχνική κατάσταση του δικτύου ύδρευσης. Στο κεφάλαιο αυτό αναδείχθηκε η μέθοδος ANFIS, ως ένα ισχυρό εργαλείο για την προσομοίωση της ζήτησης σε ημερήσιο βήμα, ακόμη ισχυρότερο από τα ANN, ενώ οι παραδοσιακές στατιστικές μέθοδοι αναδεικνύονται ιδανικές στην αντιμετώπιση της περιοδικότητας, της τάσης και του επιπέδου, αλλά όχι τόσο των μη γραμμικών σχέσεων. Η πολυπαραγοντική προσέγγιση υπερτερεί της μονοπαραγοντικής στην ευελιξία καθώς μπορεί να λειτουργήσει ως ένα εργαλείο διερεύνησης σεναρίων, αλλά προϋποθέτει μία καλή διερεύνηση των εκτιμητριών.

Στο τέταρτο κεφάλαιο αναπτύσσεται ένα χωρο-χρονικό μοντέλο προσομοίωσης του υδρευτικού συστήματος νερού. Η προσομοίωση αφορά σε μια σειρά μεταβλητών που σχετίζονται με το σύστημα όπως το ανταποδοτικό και μη ανταποδοτικό νερό, η κατανάλωση, οι απώλειες, τα σφάλματα μέτρησης, η κλοπή, η πίεση και οι αντιστοιχίες σε κόστος και ενέργεια. Η προσομοίωση υλοποιείται μέσω ενός συστήματος δύο επαναληπτικών διαδικασιών, μίας εσωτερικής και μίας εξωτερικής, που συγκλίνουν στο μέγεθος των διαρροών και στην πίεση του δικτύου αντίστοιχα. Η εκτίμηση των διαρροών βασίζεται στη μέθοδο της νυχτερινής παροχής, ενώ η σύγκλιση των πιέσεων γίνεται για τρία σημεία καταγραφής του δικτύου. Η όλη επαναληπτική διαδικασία βασίζεται στον σταδιακό διαχωρισμό του μη ανταποδοτικού νερού σε διαρροές και φαινόμενες απώλειες, οι οποίες προσομοιώνονται με διαφορετικές προσεγγίσεις ανάλογα με τη φύση των κατανομών τους, οι μεν πρώτες ως εξαρτώμενες από την πίεση, οι δε δεύτερες ακολουθώντας τις καμπύλες κατανάλωσης. Τέλος αναπτύσσεται ένα χωροχρονικό σύστημα δεικτών απόδοσης, ενώ αυτό εφαρμόζεται και ελέγχεται για ένα θεωρητικό σχήμα διαχείρισης μέσω ελέγχου πίεσης. Στο κεφάλαιο αυτό αναδεικνύεται η χρησιμότητα της χωροχρονικής διακριτοποίησης του υδατικού ισοζυγίου ενός δικτύου ύδρευσης, ειδικά σε περιπτώσεις έντονα δυναμικές, όπως οι τουριστικές περιοχές. Η χρησιμότητα αυτή έγκειται στην ανάδειξη περιοχών και περιόδων (ακόμη και ζωνών εντός του εικοσιτετραώρου) έντονης καταπόνησης του συστήματος, δηλαδή με υψηλές συγκριτικά και απόλυτα τιμές διάφορων κριτηρίων απόδοσης, όπως μη ανταποδοτικού νερού, διαρροών, φαινόμενων απωλειών και κατανάλωσης, ενεργειακής κατανάλωσης. Τέλος αναδεικνύεται η έμμεση χρήσιμότητα μιας τέτοιας προσέγγισης στην ταυτοποίηση των αστικών χρήσεων γης (όπως η τουριστική και μη τουριστική). Τα παραπάνω αποτελούν χρήσιμα

εργαλεία προς την πιο στοχευμένη και αποδοτικότερη διαχείριση του συστήματος ύδρευσης και αναδεικνύουν τη σύνδεση του υδρευτικού συστήματος με άλλα επίπεδα αστικών δραστηριοτήτων, όπως η κατανάλωση της ενέργειας και χρήση των χώρων.

Στο πέμπτο κεφάλαιο γίνεται αρχικά μία επισκόπηση της χρησιμότητας των μοντέλων προσομοίωσης οικιακού νερού σε επίπεδο βρύσης και μία σύντομη ιστορική αναδρομή των προσεγγίσεων. Στη συνέχεια, αναπτύσσεται ένα πρωτότυπο μοντέλο παλμού για την προσομοίωση της οικιακής κατανάλωσης νερού. Η ιδιαιτερότητα του μοντέλου αφορά στην εισαγωγή ενός μηχανισμού σχηματισμού των συμβάντων κατανάλωσης, μέσω της ένωσης μοναχικών συμβάντων με τη χρήση του νόμου του αντίστροφου τετραγώνου απόστασης. Το υπόδειγμα υλοποιείται για την περίπτωση μελέτης της πόλης της Σκιάθου και της πόλη του Σοσνόβιετς της Πολωνίας για έναν αριθμό πιλότων. Οι δύο περιπτώσεις μελέτης παρουσιάζουν διαφορετικά χαρακτηριστικά περιοδικότητας. Το υπόδειγμα αποδεικνύεται αποδοτικότερο για περισσότερα δεδομένα εκπαίδευσής του και για πυκνότερα συμβάντα κατανάλωσης. Στη συνέχεια, χρησιμοποιείται το μοντέλο σε συνδυασμό με το χωροχρονικό μοντέλο προσομοίωσης της ύδρευσης για τη μοντελοποίηση της μείωσης της κατανάλωσης μέσω ελέγχου πίεσης. Στο κεφάλαιο αυτό προτείνεται ένα αποδοτικό πρωτότυπο υπόδειγμα παλμού για την προσομοίωση της οικιακής χρήσης νερού, ενώ με τη χρήση του και τη σύνδεσή του με τη χωροχρονική προσομοίωση του δικτύου αναδεικνύεται η προοπτική χρήσης του για ακριβέστερη εκτίμηση της δυνατότητας εξοικονόμησης νερού όταν εφαρμόζεται διαχείριση του δικτύου με έλεγχο πίεσης.

Στο έκτο και τελευταίο κεφάλαιο γίνεται μία ανασκόπηση της εργασίας και τονίζονται τα σημαντικά σημεία.

Τα πρωτότυπα σημεία της διατριβής συνοψίζονται στα ακόλουθα:

- Η χρήση του Συστήματος Νευροασαφούς Επαγωγής (ANFIS) και υβριδικών μεθόδων που βασίζονται σε αυτό, στα Τεχνητά Νευρωνικά Δίκτυα και σε παραδοσιακες στατιστικές μεθόδους για την προσομοίωση της ζήτησης αστικού νερού.
- Η χρήση της τουριστικής εισροής και του ποσοστού Μη Ανταποδοτικού νερού, ως μεταβλητές πρόβλεψης της ζήτησης του νερού.
- Η εκτίμηση των πινάκων συνιστωσών του εισερχόμενου νερού, όπως προτείνεται από τον Παγκόσμιο Οργανισμό Νερού, σε κλίμακα ημέρας (ή και μικρότερης) και σε επίπεδο γειτονιάς.
- Η εισαγωγή του μεθοδολογικού πλαισίου για το σύστημα των δύο κύκλων επαναλήψεων για την προσομοίωση της ύδρευσης με κριτήριο την σύγκλιση εσωτερικά των διαρροών και εξωτερικά των πιέσεων.
- Η χωροχρονική εκτίμηση, εφαρμογή και οπτικοποίηση δεικτών απόδοσης (καθιερωμένων και καινοτόμων) του υδρευτικού συστήματος.
- Η ανάπτυξη ενός αλγόριθμου προσομοίωσης της οικιακής χρήσης νερού σε επίπεδο βρύσης

- Η χρήση του προσομοιωτή οικιακής χρήσης νερού για την εκτίμηση της δυνατότητας μείωσης της κατανάλωσης με διαχείριση ελέγχου πίεσης
- Η χρήση των χωροχρονικών διαφοροποιήσεων των συνιστωσών του πίνακα του Παγκόσμιου Οργανισμού Νερού για την διεξαγωγή συμπερασμάτων ως προς τις αστικές χρήσεις γης (τουριστική και μη).

Οι προοπτικές και τα θέματα προς περαιτέρω διερεύνηση συνοψίζονται στα παρακάτω σημεία:

- Η διερεύνηση της εφαρμογής διαφορετικών καμπυλών ζήτησης ανάλογα με τον χαρακτηρισμό της αστικής χρήσης της κάθε περιοχής.
- Η διερεύνηση της δυνατότητας εφαρμογής μεθόδων μηχανής εκμάθησης για την προσομοίωση των συνιστωσών του εισερχόμενου νερού, με εκπαίδευση των αλγορίθμων μηχανής εκμάθησης πάνω στις προσομοιωμένες τιμές. Αυτό θα μπορούσε να αποκαλύψει τη δυνατότητα διευκόλυνσης της προσομοίωσης ενός νέου υδρευτικού συστήματος μέσω τις αποθήκευσης προτύπων σε βιβλιοθήκες.
- Η χρήση του μοντέλου προσομοίωσης του υδρευτικού συστήματος για την αξιολόγηση μέσω των δεικτών απόδοσης διάφορων διαχειριστικών σχημάτων, όπως αυτό της διαμερισματοποίησης του δικτύου ύδρευσης.
- Η εμβάθυνση στην ανάλυση της σχέσης νερού ενέργειας καθ' όλο το σύστημα ύδρευσης.
- Η εξέλιξη του αλγόριθμου παλμού ως μοντέλο προσομοίωσης πολλών μεταβλητών.

1. Motivation and Objectives

The United Nations (UN) 2030 Agenda, which culminates in the Sustainable Development Goals (SDGs), sets the preservation of freshwater at the core of sustainable development. Discussions on SDG 6 that refers to clean water and sanitation focus on the threat of climate change to the availability of water, as well as on the need to integratedly manage water and energy and their interlinkages to climate change. The European Environmental Agency (EEA, 2018) states that water scarcity is a major threat for Europe, especially for the Southern regions. Among the reasons that cause this stress and have triggered drought incidents are the spatial and temporal variability in demand and availability, the overall increasing demand due to the rise in urbanization and changing lifestyles, the unevenly increased touristic water demand and the climate change effects. The Water Exploitation Index (WEI) indicates that quite a few European countries have surpassed the warning threshold of 20% (Raskin et al., 1997) for the overall exploitation of freshwater, or for solely ground or surface freshwater (Figure 1). Among the stressed countries, Kosovo has the overall highest WEI of 86.4%, followed by Cyprus with 67.4 % (reported values for 2017), while a few countries are below the 20% threshold when considering freshwater as a whole, but seem to perform poorly when groundwater or surface water are taken individually. For example Greece had an overall WEI of 13.8% but a groundwater WEI of 158.1% for 2015. These values both increased to 15.6% and 193.1%, respectively, for 2016 (Eurostat, 2019). Regarding the WEI at Adriatic area, for which Eurostat does not have any available data, Kanakoudis et al. (2016 and 2017) report that many areas, especially in eastern Italy, are also characterized with very high and high risk. These statistics highlight the stress on fresh water availability, which sometimes cannot be seen in the overall picture, but one needs to zoom in time- and space-wise, or to distinguish between surface and groundwater resources. The issue becomes more complicated, when water quality is considered, since water security is threatened by both water quantity and quality issues.

Aligned with the aforementioned concerns, the UN Agenda 21 (Sitarz, 1993) which later on evolved into the Agenda 2030 (Colglazier, 2015) encourages, among others, the development of interactive databases, the construction of forecasting models and the involvement of Information and Communication Technologies (ICT) to enhance water management. At the same time, Liemberger and Wyatt, (2019)estimated annual global non-revenue water (NRW) levels in urban Water Distribution Networks (WDNs) as high as 126 billion m^3 or 39 billion USD, with 9.8 billion m^3 or 3.4 billion USD corresponding to Europe. The International Water Association (IWA) Water Loss Task Force recommends four primary measures for optimized Real Losses Management, namely: (i) management of pipelines and assets, (ii) pressure management, (iii) speed and quality of repairs, and (iv) active leakage control. To enhance these measures, it sets at the core of such a multi-targeted task a detailed component analysis of the WDN inflow into revenue and NRW, real and apparent losses, etc, (Lambert, 2002).

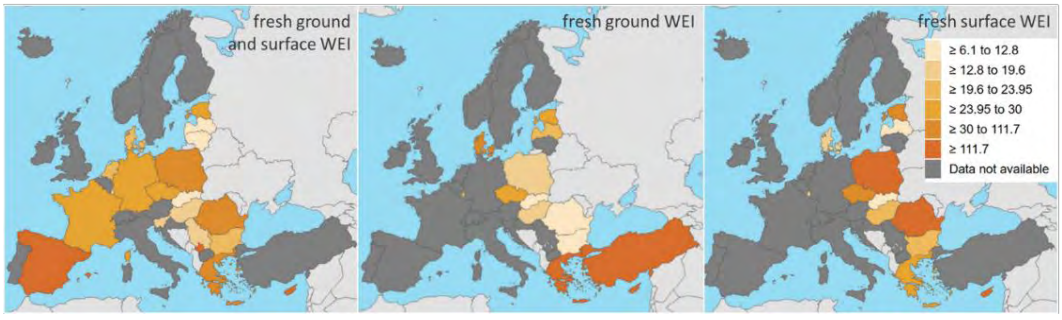


Figure 1. Europe Water Exploitation Index for overall fresh groundwater and surface water, fresh ground and fresh surface water for 2016 (Source: ec.europa.eu/eurostat)

In 2012, the Smart Water Networks Forum—SWAN Forum—conducted a Global Utility Survey (Sensus and SWAN, 2012), in order to identify the challenges and opportunities that utilities face around the world. Water supply network leakage tops the list of utility challenges for over half of the utilities surveyed; at the same time, leakage reduction is seen as the key opportunity for improving network efficiency. It is estimated that, on average, water distribution networks lose 20% of the transported water in Europe. The problem of water loss through network leakage becomes really serious, especially when coupled with increased urbanization, climate change and extreme weather phenomena such as extensive drought and water scarcity. Information and Communication Technology (ICT) tools offer a great potential for early leakage detection, while smart metering may lead to substantial reductions in water resource losses (Laspidou, 2015).

There are many challenges to the efficient and effective operation of the water supply network, especially when leakage levels are high. In particular, excessive energy consumption used for pumping water and chemicals used for treating it are related to leakages. With an ageing infrastructure, burst rates are rising, while replacing affected network sections requires large capital investments. Pressure management and optimization is really a central issue in addressing the challenges of leakage, bursts and high operational costs. Since pressure in the network is essential in the context of operation and maintenance for achieving high customer satisfaction and meeting water demand, even before simulating and optimizing flow and pressure in the network, a good grasp on forecasting demand is of critical importance (Alvisi et al, 2007; Jentgen et al, 2007).

Contemporary approaches on Water Distribution System (WDS) management need to deal with the temporal and spatial pressure variation within the networks. Meeting water demand needs includes the development of network pressure profiles that vary in time and space. Such pressure management schemes are at the core of smart water networks, making possible energy savings due to controlled

pumping, leakage minimization and overall reduction of Non-Revenue Water. Smart water networks are most commonly implemented by dividing the network in District Metered Areas (DMAs) and employing a Decision Support System (DSS) that is intended to match water supply with water demand, usually exploiting Information and Communication Technologies—ICT. Real-time monitoring and water demand forecasting methodologies have emerged as important components of numerous multi-objective schemes, such as the reduction of leakage, pumping costs, pipe bursts, green-house gas emission, chemical use, etc. Successful smart water network deployment can contribute to better management of water resources, to better informed consumers and to cities with a smaller water and carbon footprint overall, an area that is becoming increasingly important for water-scarce regions under climate change pressures.

A smart water network and generally optimised and informed urban water management would constitute one of the main achievements of what has been characterised in the literature as a "smart city"(Nam and Pardo, 2011; Hollands, 2008). Though the definition of a smart city is rather diverse and vague, often relevant to the ideological label of its context, the idea has always implied the involvement of new technologies or innovation in general to facilitate or improve economy, people, governance, mobility, environment and living within a city (Giffinger et al., 2007). Key to a smart city concept is the idea of measurement, of instrumenting the urban landscape and associated activity and monitoring their state and behavior in a way that leads to technological, governmental and societal advances. Peter Ferdinand Drucker (1909-2005), a thinker considered to be the founder of modern management, has said that "you can't manage what you can't measure", which greatly applies to a city of the future, in which near real-time measurements enable stakeholder awareness, engagement and quick response to new conditions, thus leading to a new model of civic behavior and involvement. This new paradigm is based on almost individualized planning on one hand, and near real-time information on another (Lim et al., 2010). A recent study (Cominola et al., 2015) reviews water smart metering projects taking place in the last decades worldwide. According to this work, these projects, which focus on real-time water use monitoring at high spatial and temporal granularity, stimulate modeling approaches and behavior adaptive urban water management strategies, investigating the potential of building aware consumers who will be more considerate of their water use. In parallel, consumer awareness campaigns—supported by smart meters that provide feedback—have been documented in the literature in the last decade (Novak et al., 2016; Perren and Yang, 2015; Shan et al., 2015; Russell and Fielding, 2010), while latest advances include the development of gaming platforms (Wang and Capiluppi, 2015) for water management and the involvement of social media for citizen engagement in water saving practices. The European Commission has funded a series of research projects that developed a series of diverse case studies that all showed how

building consumer awareness could limit water consumption (all these projects are grouped under the ICT4WATER cluster—http://ict4water.eu). A number of water utilities increasingly attempt to influence the behavior of consumers towards improving water consumption, by using communication tools to give information back to users and display their consumption or customized feedbacks or water-saving tips. At the same time, various companies have been established lately that specialize solely on transforming the way customers think about their household water consumption, as well as the way utilities engage with their customers. Such companies combine Machine Learning (ML) and other data science tools with cloud computing and behavioral science to develop Software-as-a-Service solution to customer engagement and efficiency issues faced by utilities.

Subsequently, the need for the collection and management of large quantities of temporal and spatial high-resolution data emerges as a useful tool for urban planning, while at the same time, the radical evolution in the technological sector of sensors, Information and Communication Technologies (ICTs) used in the whole water supply chain, social network data analysis and Data Mining (DM) techniques reveals new potentials for more efficient planning (Kanakoudis and Tsitsifli, 2019; Yang et al., 2017; Laspidou et al., 2015b; Laspidou, 2014). In the urban water domain, due to fast urbanization, increasing demands, climate change and high pressure on water resources, research activity increasingly focuses on monitoring, understanding and better managing urban water activities. Detailed monitoring of household water consumption can reveal useful information about citizen behavioral patterns, not only related to their water use *per se*, but also concerning a range of socio-economic factors, directly or indirectly related to water, such as circadian rhythms, working hours, daily habits, house amenities, familial structure and profile, etc. (Koiv and Toode, 2006). Furthermore, the spatiotemporal analysis of household water use can help make water consumption a key indicator of human behavior, thus helping authorities and relevant stakeholders identify changes in city-living conditions, such as local development, migration (Sadat, 2012), epidemics (Kleczkowski and Maharaj, 2010), or it can disclose population shifts due to events, such as terrorist attacks (Khan et al., 2001), natural disasters (Liu et al., 2015), large-scale organized meetings or tournaments (Mol, 2010), etc. Besides the wealth of information potentially extracted by monitoring water consumption, channeling this data back to the consumers will contribute to an increased awareness that will lead to a smaller household water footprint (Al-Hoqani and Yang, 2015; Perren and Yang, 2015; Lanzarone and Zanzi, 2010). The effectiveness of similar schemes regarding energy consumption through energy metering, billing and direct display methodologies has already been documented (Darby, 2006), concluding that feedback to consumers is an important element of an energy savings scheme for consumers. Numerous relative examples are reported in Ehrhardt-martinez and Donnelly (2010) and (Fischer, 2008) works. Indicatively, in the (Staats et al., 2004) study energy savings increase of

approximately 3 % in 16 months in Netherland is reported and in Wilhite and Ling (1995) study an increase of 2.4 % in energy saving in one year for a Norway case study is presented.

Optimization of the operation of Water Distribution Systems (WDSs) is inextricably linked to Urban Water Management (UWM). Quite often, urban water supply infrastructure is ageing and deteriorating with increasing bursts and extended leakage. Pressure in Water Distribution Networks (WDNs) is often kept constant and is not adjusted temporally and spatially according to water demand; District Metered Areas (DMAs) are not always in place, thus reinforcing even more bursts due to excessive stress and leading to energy overconsumption.

In their review on urban hydroinformatics, Makropoulos and Savic (2019) present a scheme according to which the modeling is facilitated by the increasingly fluent flows of data and information in the context of the developments in Information and Communication Technologies, cloud based information platforms, and remote monitoring. Such a technological landscape in combination to well-advanced methodological frameworks could lead to the next step in water distribution networks modeling, what has been referred as Water Distribution Network Digital Twins (Sun et al., 2020).

This dissertation is an attempt to develop a series of simulation and management tools incorporated in a methodological framework to facilitate detailed supervision of a WDS via a number of key parameters across the WDN from source (groundwater) to tap, such as urban water demand, real losses, apparent losses, pressure, and water consumption at the tap, while also developing and assessing a number of key performance indicators, such as the energy-water nexus, leakage reduction potential, pressure driven demand reduction potential and economic savings.

The aim of this dissertation is to create tools for the mathematical modelling, analysis, forecast and performance indexing of various components of a water distribution system which is intensely seasonal regarding demands and stresses. Specifically the presented work answers the following objectives:

1. The investigation among traditional statistical, ML and hybrid approaches, for appropriate modelling and forecasting of the water demand in multiple periodicities (trimester, months and days) and the analysis of the drivers that force intense seasonality, such as the weather, the tourism and the technical status of the Water Distribution Network.
2. The construction of a spatio-temporal simulation model of a Water Distribution Network, which will facilitate the assessment of the iconic International Water Association Water Balance tables in small time and

spatial resolution (hourly and neighborhood) and the localization of stresses such as leakage, theft, and billed water consumption.

3. The proposition of appropriate key performance indicators that can indicate the performance of the Water Distribution Network regarding real losses, revenue, energy consumption, as well as the relation of the water balance profiles to the urban land uses (touristic or residential).
4. The proposition of key performance indicators for the spatiotemporal assessment of the performance of alternative management schemes, such as Pressure Control Management, regarding the potential reduction of leakage, energy consumption and non-revenue.
5. The construction of a simulation tool for the tap water consumption mimicking the consumption incidents (time slot, duration, and flowrate) based on the pulse models paradigm.
6. The investigation of the potential of reducing pressure driven demand based on the Water Distribution Network Simulation and the tap water simulation and the proposition of relevant indicators for this potential.

Parts of the content of chapter 1 is included in the following published articles in journals:

Kofinas, D. T., Spyropoulou, A., Laspidou, C. S. (2018). A methodology for synthetic household water consumption data generation. *Environmental modelling & software*, *100*, 48-66.

- The contribution of Mr. Kofinas, D. involves the conceptualization, the methodology, the validation, the formal analysis, the investigation, the writing, and the visualization.
- The contribution of Dr Spyropoulou, A. involves the programming of the involved code.
- The contribution of Professor Laspidou, C. involves the scientific supervision

This work was supported by the project ISS EWATUS—Integrated Support System for Efficient Water Usage and Resources Managementdwhich is implemented in the framework of the EU 7th Framework Programme, Specific programme Cooperation In- formation and Communication Technologies; Grant Agreement Number 619228.

Kofinas, D., Ulanczyk, R., & Laspidou, C. S. (2020). Simulation of a Water Distribution Network with Key Performance Indicators for Spatio-Temporal Analysis and Operation of Highly Stressed Water Infrastructure. *Water*, *12*(4), 1149.

- The contribution of Mr. Kofinas, D. involves the conceptualization, the methodology, the review of the software, part of the data curation, the validation, the formal analysis, the investigation, the writing, and the visualization.
- The contribution of Mr. Ulancyk, R. involves the software, the data curation, part of the investigation (relevant to the programming), and the review.
- The contribution of Professor Laspidou, C. involves the scientific supervision and review.

The work described in this paper has been conducted within the project Water4Cities. This project has received funding from the European Union's Horizon 2020 Research and Innovation Staff Exchange programme under grant agreement number 734409. This paper and the content included in it do not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

2. Original Contributions

1 The current work explores the drivers for water demand. It investigates forecasting approaches, including univariate and multivariate algorithms, traditional statistics, Machine Learning (ML) and hybrid approaches in different periodicities, from daily to trimester. The investigation concludes that the properties of ML algorithms tested, such as Artificial Neural Networks (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS), can be exploited to overcome the collinearity and non-linearity issues in water demand forecasting. Hybrid approaches of ML and statistical algorithms can exploit the beneficial properties of each.

original contribution: the use of ANFIS and hybrid (ANFIS-traditional statistics) approaches for water demand forecasting

2 Throughout the testing of different multivariate forecasting approaches, the investigation identifies two variables as predictors, except for the well-established meteorological variables: The first is touristic activity, as described by the total incoming arrivals into the case study island and the second is a composite index of Revenue and Non-Revenue Water.

original contribution: the use of total incoming arrivals and percentage of Non Revenue water as water demand drivers

3 The research work creates a simulation model of the Water Distribution Network. The simulation is spatio-temporal, in neighborhood (or even household) granularity and hourly time step.

4 For the construction of the simulation model, the well-known Water Balance table, which is recommended by the International Water Association as a prerequisite to any urban water supply management strategy is assessed in neighborhood granularity and hourly time step. The Water Balance components assessed include System Input Volume, Billed Consumption, Non-Revenue Water, Unauthorized Consumption, Real Losses, and Metering Inaccuracies. All the components are identified locally and instantly offering a useful tool to a water utility for creating dynamic hotspot maps that can contribute to the optimization of the Water Distribution Network management through:

- the localization of leakage in neighborhood resolution and the identification of drivers of leakage throughout the network;
- the localization of theft in neighborhood resolution;
- the assessment of metering inaccuracies;
- the quantification of tourism impact on water demand;
- the application of a Pressure Control Management scheme and the quantification of its beneficial effect.

original contribution: the increase of temporal and spatial granulariy in the simulation of the water balance components in a water distribution network

5 The assessment of the Water Balance table is implemented by distinguishing the simulation of the components into pressure-driven and demand driven and an application of a nested system of two loops, an inner and an outer, which through iterative runs, close when the simulated pressure best fits the actual network pressure. The aforementioned process suggest a methodological approach for a spatiotemporal assessment of the IWA WB tables, facilitating well informed and detailed supervision and management of a WDN.

original contribution: the introduction of the two-loops methodological framework for the simulation of the water balance components according to their nature, pressure driven or demand driven

6 A theoretical Pressure Control Management scheme is applied to simulate and quantify all the benefits spatio-temporally. All differences in the Water Balance Components are assessed in neighborhood granularity and hourly time step.

7 Spatio-temporal critical Key Performance Indicators (KPIs) are introduced to facilitate a detailed supervision of the Water Distribution Network. The Key Performance Indicators include the energy consumption in the water components (water-energy nexus), a pressure-driven demand indicator, and various expressions of water components such as ‘‘per-connection’’, ‘‘per-customer,’’ and ‘‘per-network-length’’ indicators

original contribution: the spatio-temporal assessement of KPIs that have been used in an aggregate manner for the whole Water Distribution Network in the past and for bimonthly time scale

8 The aforementioned KPIs are also used for the assessment of the Pressure Control Management scheme performance.

9 The localization of the water components is used to conduct rough conclusions regarding the urban land uses, specifically the residential and the touristic ones.

10 A synthetic household consumption data generator algorithm has been built and used to simulate household tap water consumption. The algorithm has been adjusted and tested for different types of seasonality according to the specifications of two case studies. A monthly seasonality for a Greek touristic case study and a week/weekend seasonality for a Polish industrial case study.

11 The suggested algorithm belongs to the group of pulse models but introduces a structural modification based on the well-known Newton gravity law. Through this modification, the model prioritizes the household consumption profile throughout the day rather than other attributes that are typically prioritized by the other pulse models.

original contribution: the development of a prototype household consumption algorithm and the use of the Newton gravity law for the creation of water consumption incident clusters

12 The household consumption model is used to better assess the pressure driven demand reduction. The two approaches for pressure driven demand assessment are compared.

original contribution: the use of the household consumption model prototype to simulate pressure driven consumption decreace thanks to the application of a theoretical pr;essure management scheme

13 The current thesis also explores, throughout the different aforementioned modelling components, the implications of intensely dynamic demand drivers due to tourism and weather, such as the ones that occur in a Mediterranean touristic resort.

original contribution: the use of water balance components spatio-temporal differences to extract implications for urban land uses dynamics related to tourism

3. Urban Water Demand Forecasting

3.1 The need to develop forecasting methodologies for water demand

3.1.1. The different planning tasks that need the forecasting of water demand

Urban water demand is a key parameter for the operation of urban water supply and distribution systems. Water demand forecasting plays a significant role in managing and planning water supply operations and water conservation strategies. A series of operational, tactical and strategic planning decisions set water demand as a core designing aspect. The variable appears in different forms, depending on the context, such as integrated water demand, including or excluding losses, per capita water demand, peak demand, etc. Among the various planning tasks that require water demand estimation, literature refers to the optimization of pumping scheduling, the operation of wells, the operation of treatment plants, the potential need to develop new water sources, the risk of water shortages assessment, the revenue risk assessment, the optimization of operational and investment decisions, and pricing purposes (Donkor et al., 2014; Billings and Jones, 2008; Hazen and Sawyer and PMCL, 2004). Reuse management and public awareness are also two systems that emerge to correlate with water demand forecasting, both in multiple scales (Hajeeh, 2010). In the context of operational control, the water demand, from the perspective of water resources, comprises, not only that required for customers, but also the losses of the water supply system—from drilling to household tap—since it is the combined amount which is put into supply (Alvisi et al., 2007)).

3.1.2 Horizon and periodicity

Many of the involved system objectives, such as the minimization of cost, the minimization of water losses, the optimization of water quality and increase of consumer satisfaction, require a dynamic approach towards water demand assessment, so as to facilitate frequent adjustments in response to variations. These dynamic schemes vary in terms of horizon and periodicity. Decision Support Systems (DSSs) often require forecasting approaches of short-term horizon and fine periodicity. Contemporary DSSs tend to include real-time monitoring schemes in order to increase their ability for short-term and or even near real-time decision that can adjust dynamically (daily, hourly) to the dynamic nature of the critical system input variables (Figure 2). Dynamic properties of a DSS may also apply to bigger periodicities and horizons such as weekly, monthly, seasonally and annually. Perception of seasonal variation of designing parameters, may often be adequate enough taking into account that uncertainty of forecasts forces planners to reduce planning accuracy and preferably work within ranges.

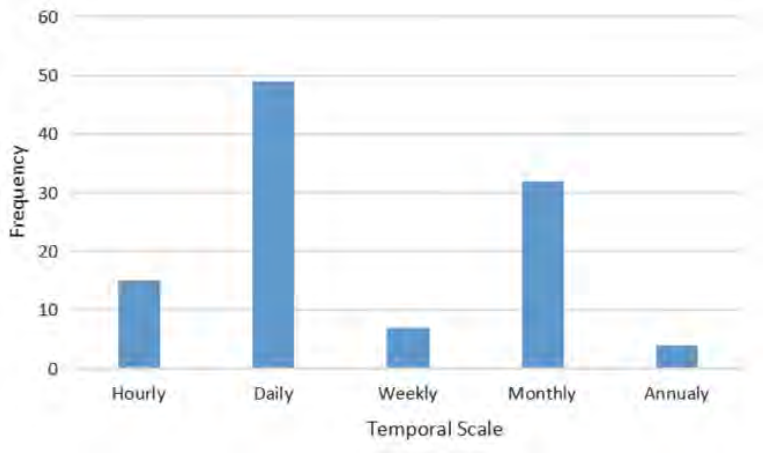


Figure 2. Frequency of scaling applied in the reviewed papers in a recent review chapter of over 100 papers that suggest or compare ML urban water demand forecasting approaches. The figure is taken by Singleton and Liu (2017)

Future projections of the demand are often required for future planning of the aforementioned systems. The nature of the target sets the projection aspects, namely the forecast variable form, the periodicity or time step, and the horizon. Short-term, mid-term and long-term forecasts may apply to different exercises such as pumping scheduling, pricing purposes and investment decisions, respectively. The three time horizons, however, are not sharply defined, since short-term may imply a horizon of one to six months, mid-term a horizon of two to ten years and long-term a horizon of longer than two to ten years (Billings and Jones, 2008; Gardiner and Herrington, 1990; Ghiassi et al., 2008), as presented in Figure 3. For the efficient operation of water supply and distribution networks, the type of short-term demand forecasting is the most appropriate in order to program the pumping arrangements and thus to supply water with a more efficient way (Alvisi et al., 2007).

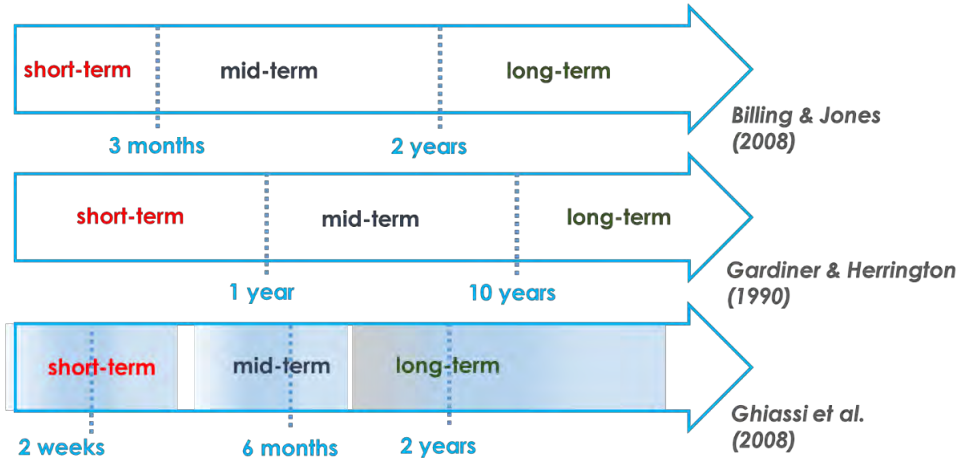


Figure 3. Different definitions on forecast horizons

3.1.3 Categorization of modeling approaches

Indirectly, the approach of forecast is indicated through the efficiency achieved in the different contexts. For example, the definition of different target variables, such as peak day, daily total system demand, monthly total system demand, annual demand per capita, annual demand by customer class and revenue, shall imply a different forecast architecture.

Forecasting methods have been developed mainly to control water distribution operation system nearly in real time, based on time series data that are collected sequentially overtime period. In the literature there are several time series forecasting models that forecast future water demand on the basis of past observations and associated error terms (Billings and Jones, 2008). In general, the time series forecasting models of water demand can be divided into deterministic and stochastic models. Deterministic models are used to model time series by a series of seasonal, weekly and daily patterns considering physical nature. Stochastic models usually adopt a numerical approach since they are usually formulated by using statistical and probabilistic models that are built on historical data (Box and Jenkins, 1976; Shahin et al., 1993).

Demand forecasting approaches vary from the simplest historical extrapolation to sophisticated analytical models; therefore the choice of an appropriate modeling ought to address the purpose of the forecasting needed by water utility companies, and the quality and quantity of data (Mamo et al. 2013). Univariate and multi-variate approaches have been used to address the different target variables. Specifically for multivariate approaches, a series of socio-economic, demographic, weather related and technological predictors have been used (Donkor et al., 2014). Income, Gross National Product, price of water, Consumer Price Index, population, temperature, precipitation, wind, availability, attitude towards water, lifestyle are some of the predictors that are met in different multi-variate approaches (Rockaway et al., 2011; Arbués et al., 2003; Brekke et al., 2002; Burney et al., 2001). All of the aforementioned variables are proven to affect water demand, however, not necessarily in the same context, horizon or periodicity. For example Burney et al. (2001) have used the demographic and economic variables to implement a long-term and annual periodicity forecasting scheme, having excluded the weather variables that mostly affect water demand in smaller periodicity. On the contrary for short-term approaches the mild socio-economic trends might not prove to have a significant impact.

Other approach selections include unit consumption models, smoothing exponential and moving average, stochastic processes, time series regression, scenario based models, Artificial Neural Networks (ANN) (Bennett et al., 2013; Adamowski and Chan, 2011; Adamowski, 2008; Chau, 2006; Bougadis et al., 2005; Baxter et al., 2001) and composite or hybrid approaches (Lima et al., 2017; Kang et al., 2015; Mohammed and Ibrahim, 2012; Li and Huicheng, 2010), with respect to the criterion of the utility

(or user) requirements. A recent very extensive review work among ML approaches on urban water demand forecasting by Singleton and Liu (2017), showcases the increasing interest in ML techniques, which is fairly justified taking into account that such techniques overcome and simplify issues such as the non-linear relations and the collinearity between the system variables. The review work discerns the ML techniques into eleven groups including Wavelet ANN, ANFIS, ARIMA-ML hybrid schemes, and Multiple Linear Regression-ML hybrid schemes.

3.1.4. Methodological framework for water demand forecasting

Figure 4 depicts the sequence of steps for developing a water demand model and forecasting methodology. Prerequisite to any forecasting approach is the definition of objectives. The definition of objectives are usually set after stakeholders' involvement process and analysis of their requirements. Water utilities are the typical primary stakeholders, while secondary stakeholders may include municipalities, policy makers, non-governmental organizations and scientists (Edelenbos et al., 2011; Tillman et al., 2001). The definition of the objectives will imply the system boundaries. For example, if the objective is the optimization of water sources selection, this would imply the involvement of a broader system compared to the one that needs to be analyzed for Non-Revenue Water minimization. The context and the objectives, in turn, implies the selection of the water demand form, for example including or excluding water losses, the periodicity and the horizon. A good understanding of the system in a specific study case and the involved specifications can lead to a preselection of the variables that need to be introduced as drivers of water demand. The following phase involves data collection from various sources and data preparation. The periodicity and horizon define the datasets length and time step that preferably need to be common for all attributes. Data preparation, also includes tasks, such as removing outliers, imputation of missing values and normalizations.

Next, the forecasting approach needs to be selected. The approach also needs to agree to the requirements and the involved variables. For example, future scenarios testing or parameter analysis implies a multivariate approach rather than a univariate, while detection of complex interrelations, non-linearities and co-linearities suggest ML algorithms, such as Agent Based Models (Koutiva and Makropoulos, 2019), rather than traditional statistical approaches. The analysis of the involved data may approve or disprove the variables preselected. When the data are analyzed according to the requirements of each methodology, the construction of the model is implemented by calibrating the functions at the data on a selected calibration sample of the data and later on validating on the validation period. The validation is implemented with use of selected accuracy metrics. Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Coefficient of Determination (CoD), and Coefficient of Correlation (CoC) are the most popular metrics of accuracy and evaluation criteria according to Singleton

and Liu (2017). After the evaluation of the overall process and the potential iterations to improve the accuracy of the model, the conclusions, results, implications, parameter analysis or scenario runs are implemented with use of the constructed model.

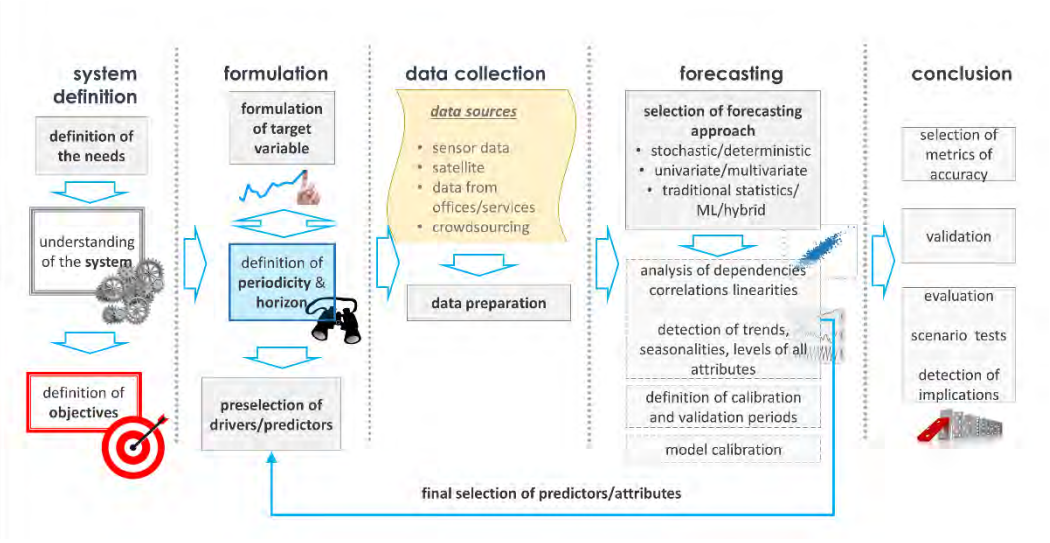


Figure 4. Schema of the sequence of steps for water demand forecasting

The increasing demand for near real-time forecasts of water demand, and for finest time scale set the requirements for increasing the accuracy of water demand forecasting, even if the simulation of a taget variable at a single temporal scale does not necessarily mean similarly consistent performance at higher time scales (Kossieris et al., 2019; Tsoukalas et al., 2019). New algorithms need to be tested and methodologies and approaches that are already tested in other fields such as econometrics and energy consumption need to be brought to the water demand field. Special focus needs to be given to the proper selection and further involvement of precise predictor attributes that at the same time are easy for utilities to collect.

3.2 Implications of tourism in the seasonality of water demand

3.2.1. The Mediterranean touristic resorts

Mediterranean touristic resorts often share a common profile of stresses that is shaped up by the subtropical climate, the intense touristic activity and the inadequate infrastructure that cannot efficiently serve such a wide range of population which grows at summer and decreases in winter. These three main stresses lead to a series of aftereffects, such as the spatial and temporal intense variability in resources demand and availability, the overexploitation of resources, the deterioration of the natural environment, etc. (Spilanis et al., 2009; Spilianis and Vayanni, 2004). These implications are related through a series of interlinkages, and often lead to a vicious cycle of trade-offs. Water resources are typically at the core of any implications and usually suffer both quantity and quality issues. The Mediterranean area has been noted to be a highly vulnerable region regarding water resources, taking into account the expected 2050 scenario of a 30-505 decrease in available freshwater, due to climate change (Figure 5), while at the same time the current efficiencies in irrigation and urban water supply networks are reported inadequate in average, ranging from 35% (Libanon) to 85 % (Cyprus) for irrigation and from 34 % (Albania) to 88 % (Israel) for urban water supply (Milano et al., 2013).

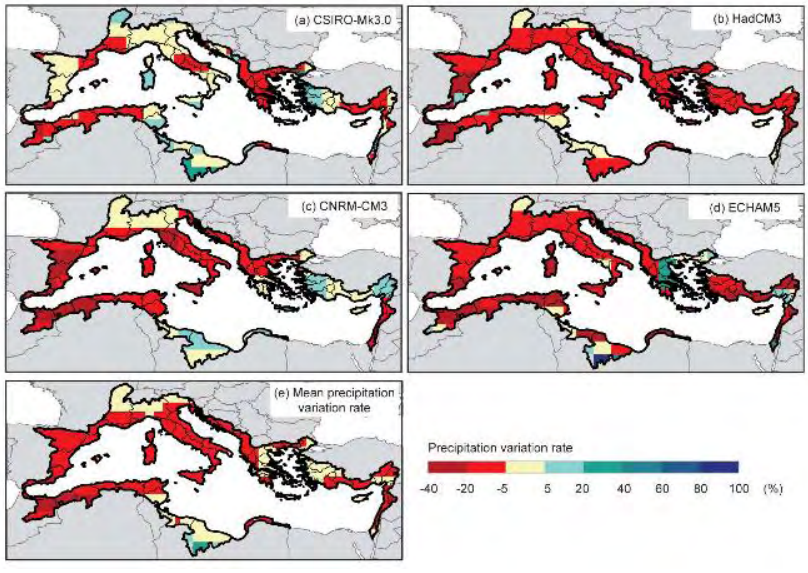


Figure 5. Mean precipitation variation rate over the Mediterranean basin by the 2050 horizon as compared to the1971–1990 period; projections based on: (a) the CSIRO-MK3.0 model; (b) the HadCM3 model; (c) the CNRM-CM3 model; (d) the ECHAM5 model; and (e) an average of the four GCMs. The figure is taken from Milano et al. (2013)

Greek islands are often supplied with water through old and poorly designed Water Distribution Networks (WDNs). Aging infrastructure naturally leads to massive leakage, while the bold relief combined with the absence of DMAs, forces utilities to maintain high pressure to the whole network in order to satisfy water supply at remote and elevated households. Indicative of the aging infrastructure of many Greek islands is the recent, 2018, ministry announcement for renovation of the WDNs of nine islands that reached exceptionally high leakage rates (Minister of Maritime Affairs Islands and Fisheries- Greek Republic, 2018).

The arid and hot summers that coincide with large touristic influxes, multiplying water demand are followed by wet winters and low water consumption by the limited number of locals; these conditions shape an apparently seasonal regime of water demand. Taking into account the shortage of water resources—usually groundwater in diminishing quantity and quality—one can understand why water demand forecasting emerges as a basic step in dealing with water supply issues in such special conditions.

3.2.2. The Skiathos paradigm

Most Greek islands are inhabited by small communities with old and inadequate water distribution networks. Their economic development is usually based on tourism (Table 1 and Figure 6), which results in high seasonal water demand variance, with summer water demand surpassing by far winter demand. Moreover, the networks typically serve local households and some large hotel units as well, while a lot of the large hotels have their own drillings and water supply systems.

The island of Skiathos has a small water distribution network with highly variable and seasonal water demand and Non-Revenue Water as high as 50% and even almost 70 % at times. Any significant improvement in network water loss prevention would require an analysis of historical water demand data in order to capture the stochastic nature of the data. In the context of operation and maintenance management, short-term forecasting of water demand or consumption plays a critical role for water utility companies that do not only try to identify optimal ways to supply water and minimize pumping energy costs, but are also interested in controlling water losses due to leakage from the distribution network and minimizing non-revenue water.

Table 1. Touristic enterprises as a percentage of total number of enterprises at indicative Greek islandic prefectures (Spilianis and Kizos, 2015)

	touristic enterprises as a % of total
Iraklion	47.2
Chios	48.8
Evia	50.8
Chania	51.3
Rethimno	51.4
Lesvos	55.3
Lasithi	57.5
Kefallinia	59.0
Samos	59.2
Dodekanisa	60.4
Cyclades	64.0
Lefkada	64.2
Corfu	64.8
Zante	66.5

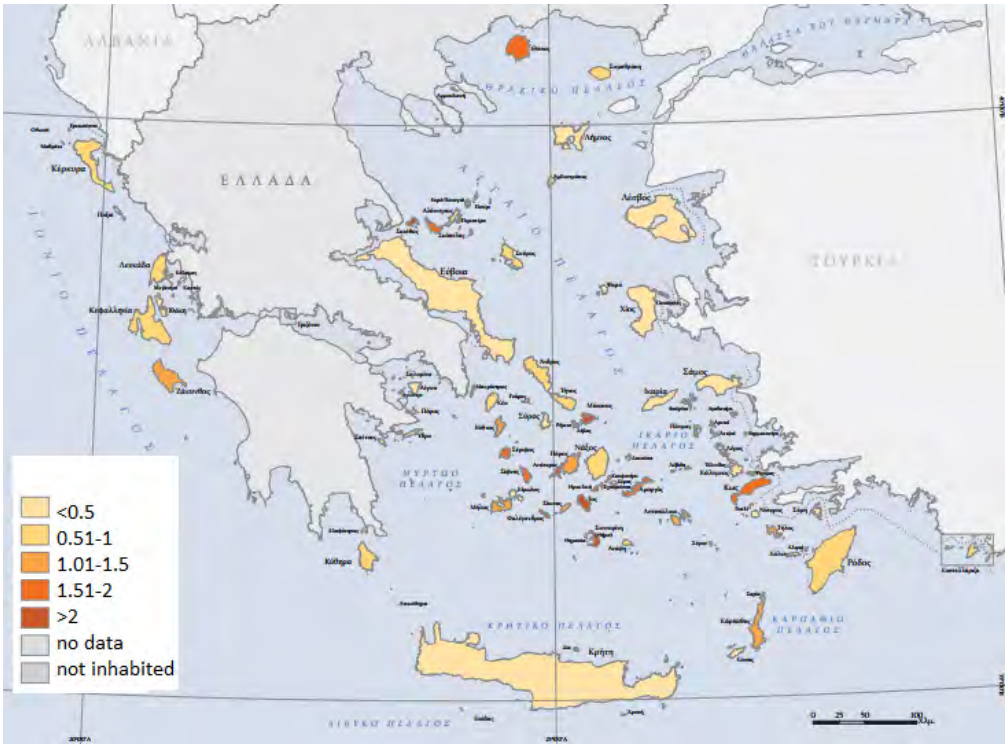


Figure 6. Tourist beds per citizen for 2013 (Spilianis and Kizos, 2015)

Skiathos (Figure 7) is an interesting case study and unique in many ways, when compared with other European water companies. Other than the high water distribution network leakage rate, it has the peculiarities of an island with limited fresh water resources. The economy in Skiathos is mainly driven by tourism, especially in the summer months, while in the winter there are agricultural and urban activities. The weather in Skiathos is typical Mediterranean, with long, hot and dry summers and cool and wet winters. The water distribution network serves the small town of Skiathos and some major hotels that reach 100% occupancy throughout the summer months. Water is abstracted by the Skiathos water utility company from a single drilling and includes a single tank and approximately 3,500 water meters. In addition, most of the big hotels of the island are not connected to the network but are supplied by private drillings. Groundwater over-exploitation causes the water level in the aquifer to drop quickly and unsustainably, resulting in seawater intrusion and aquifer salinization effects. This, in turn, has been suspected to be the reason for mercury release in the aquifer, which is later on transferred to the water supply (Spyropoulou et al., 2018). Skiathos Water Company has declared the supply water as non-potable for this reason. The monthly water demand profile follows an ascending slope reaching a high peak and then follows a descending slope, from July to October, down to the low winter demand level. The average annual water demand trend is slightly positive (Figure 8).

Skiathos inhabited by a small community throughout winter and flooded by a large number of tourists during the summer period, which lasts for approximately 5 to 6 months and shows a high peak in August. It has an ageing water distribution network that is small, with a total of about 3,500 water meters; it was built without prior planning and was expanding throughout the years as needed. The island's water demand presents high variability and seasonality, while its network has reportedly some of the highest Non-Revenue Water in Greece—as high as 50% or more (even 70 % in some cases, Figure 8). Skiathos faces serious water shortage issues, with aquifer salinization and deteriorating groundwater quality; since all urban water needs are covered by groundwater, it is important that network leakage is reduced to a minimum. In the summer of 2014, August, the town of Skiathos faced a phenomenal water scarcity incident—reportedly there was no water flow in the distribution system for more than three days—causing a series of problems to all related aspects including healthcare and tourism. Any significant improvement in leakage prevention would involve a series of strategic decisions and actions on behalf of the local water utility. At the bottom of these actions lies a reliable water demand forecasting routine that will be based on historical data and will be able to capture its stochastic nature.

The network is not divided into DMAs and no pressure regulation to meet demand profiles has ever been implemented until today.

Non-Revenue Water—leakage being its main component— seems to increase with time, when comparing winters or summers of subsequent years, with the highest NRW percentages being recorded in winter months, when network pressure is high due to reduced demand. This can lead to a primary assumption that NRW increases due to leakage increase, since the latter one is related to the network pressure (Germanopoulos, 1985). Among typical water uses in the town of Skiathos, except for household use, there are a variety of touristic-related businesses such as small hotels, taverns, restaurants, pubs and cafes. Thus, water demand follows strongly the touristic inflow regime. The touristic season lasts from April to September with a sharp peak in August (Figure 9), which drives to a water demand increase, up to more than 170% of the winter demand.

The climate of the island is typical Mediterranean (Figure 10). Observing the profiles of the meteorological variables it can be qualitatively conducted that they are related to Skiathos water demand. Specifically mean and high temperature are two variables with a significant positive correlation, while precipitation has a negative one (Mellios et al., 2015). Past research has shown though that weather variables drive water demand in a rather non-linear way. The linear effect of precipitation has been questioned; on the contrary, it has been suggested that the response of consumers to water consumption is psychological and concerns the incident of rain more than the amount of precipitation (Martinez-Espiñeira, 2002; Miaou, 1990; Maidment and Miaou, 1986).

Figure 7. Skiathos Island

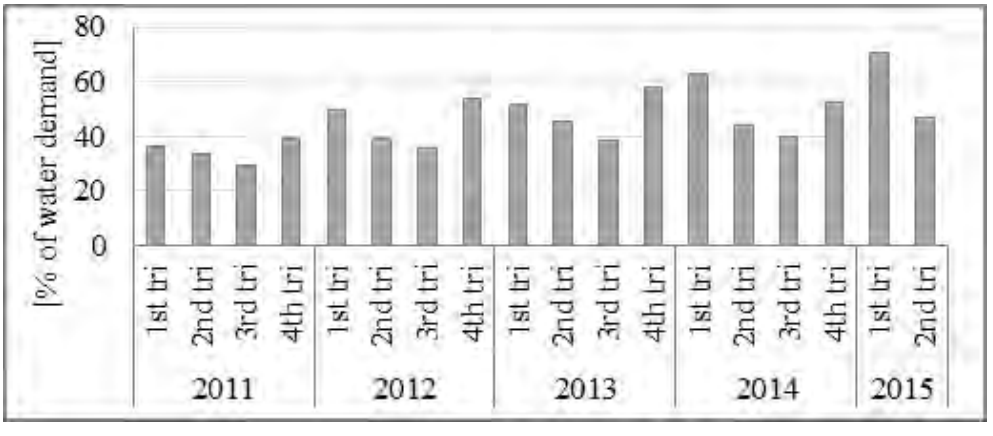


Figure 8. Trimester Non Revenue Water as a percentage of water demand for 2011-2015

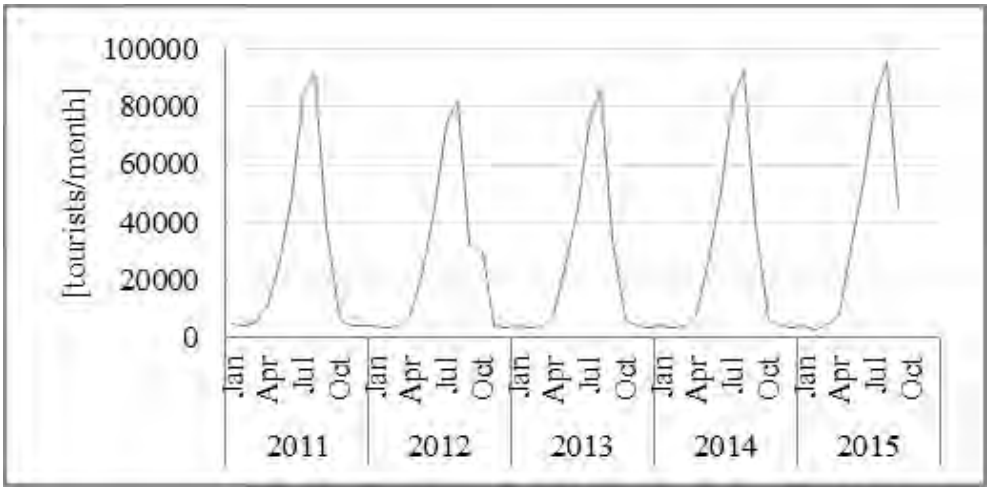


Figure 9. Fluctuation of touristic activity 2011-2015 in Skiathos Island. The variable of the diagram adds up the total arrivals in the island of Skiathos by all means: air and sea. Total arrivals is perceived as an indicator for touristic activity.

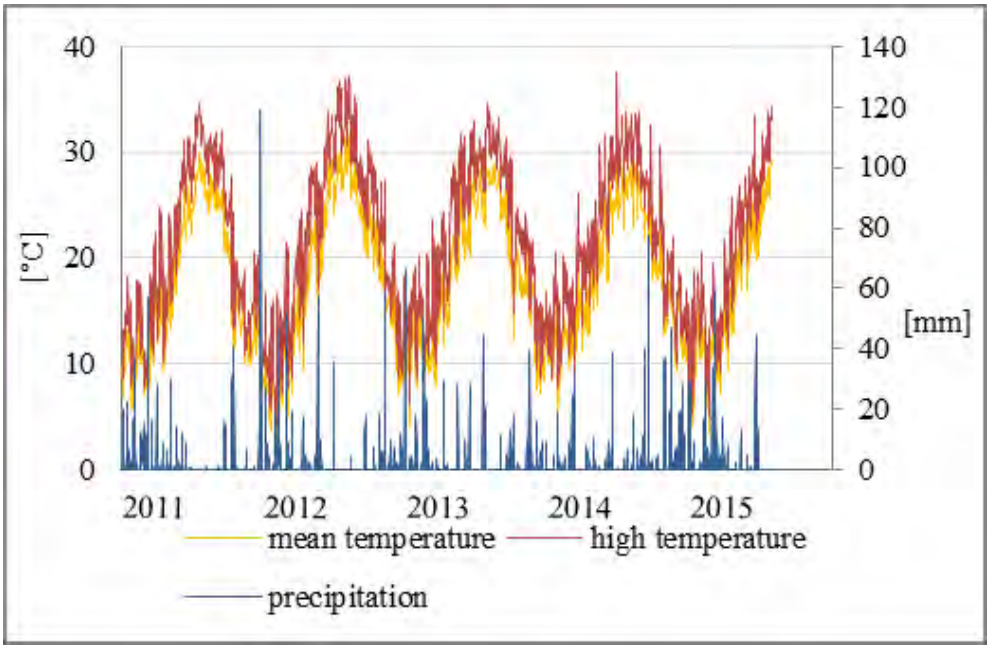


Figure 10. Weather observations for 2011-2015 in Skiathos Island

3.3 Univariate approach in trimester and monthly **time scale: ARIMA, Winters'** additive, ANN, hybrid approach

Foreword: In this subchapter, an analysis of historical water demand data from the water utility of Skiathos is presented and demand-forecasting tools using the stochastic nature of short term (or mid term) historical water demand and supply data is demonstrated. The approaches tested are univariate and are using as predictors the components of the target variable itself. The goal here is to identify the method and model useful for specific water-utility decision making problems, to apply this method to the analysis of water demand forecasting data and use it for short-term forecasting. The forecasting approaches tested in this chapter are the ARIMA, the Winter's additive, the ANN and an ARIMA-ANN hybrid approach.

3.3.1. Introduction in water demand univariate forecasting

In water demand time series modeling, common stochastic models are: pure random (or white noise) model, autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) model, autoregressive integrated moving average (ARIMA) model and seasonal autoregressive integrated moving average (SARIMA) models (Box et al., 2015; Machiwal and Jha, 2012). These models—known as traditional statistical models—are linear, which means that predictions of the future values are constrained to be linear functions of past observations. Because of their relative simplicity in understanding and implementation, linear models have been the main research focus providing applied tools during the past few decades.

Although ARIMA models are quite flexible, having the ability to represent several different types of time series, i.e., pure autoregressive (AR), pure moving average (MA) and combined AR and MA (ARMA) series, their major limitation is the pre-assumed linear form of the model. That is, a linear correlation structure is assumed among the time series values and therefore, no nonlinear patterns can be captured by the ARIMA model.

To overcome the restriction of ARIMA model and to account for certain nonlinear patterns observed in real problems, ANNs have been proposed in the literature. ANNs have been suggested as an alternative to time series forecasting to deal with linear and nonlinear relationships. The major advantage of neural networks is their flexible nonlinear modeling capability. With ANNs, there is no need to specify a particular model form. Instead, the model is adaptively formed based on the features presented from the data. This data-driven approach is suitable for many empirical data sets where no theoretical guidance is available to suggest an appropriate data generating process.

Donkor et al. (2014) recent study reviews the literature on urban water demand forecasting published from 2000 to 2010 in order to identify methods and models useful for specific water utility decision making problems. This study presents an annotated reference list of the methods and models for water demand forecasting identifying the forecast variable and periodicity, the determinants used and the forecast horizon. Results show that although a wide variety of methods and models have attracted attention, applications of these models differ, depending on the forecast variable, its periodicity and the forecast horizon. The popular models used for short-term forecasting are ARIMA and ANNs. They gained popularity due to the advantageous characteristics of each one.

Alvisi et al. (2007) proposed a short-term pattern-based forecasting approach using an ARIMA model based on statistical analysis of historical water demand data. Mamo et al. (2013) in their recent study demonstrated the usefulness of the stochastic nature of short term historical water demand and supply pattern models to study and forecast short term water demand for water utility companies.

Maier and Dandy (2000) reviewed recent papers dealing with the use of neural network models for the prediction and forecasting of water resources variables. Rojas (1996), Hassoun (1995), and Kartam (1994) have used feedforward networks with sigmoidal-type transfer functions for the prediction and forecasting of water resources variables. Chau (2006) has reviewed the development and current progress of the integration of artificial intelligence into water quality modeling. Hatzikos et al. (2005) utilized neural networks with active neurons as a modeling tool for the prediction of seawater quality indicators like water temperature, pH, dissolved oxygen (DO) and turbidity. Palani et al. (2008) demonstrated the application of ANNs to model the values of selected seawater quality variables, having the dynamic and complex processes hidden in the monitored data itself. Most of the studies reported above were simple applications of using traditional time series approaches and univariate ANNs.

Also, hybrid models have been suggested, combining the ARIMA model and neural networks, in order to overcome the deficiencies of single models. Su et al. (1997)used a hybrid model to predict a time series of reliability data with growth trend. Their results showed that the hybrid model produced better predictions than either the ARIMA model or the neural network by itself. Zhang (2003) proposed a hybrid ARIMA and ANN model to take advantage of the two techniques and applied the proposed hybrid model to some real data sets. He concluded that the combined model can be an effective way to improving predictions achieved by either of the models used separately. Ömer Faruk (2010) proposed a hybrid neural network and ARIMA model, developed for the Buyuk Menderes river, for water quality time series prediction. He indicated that the approach of combining the strengths of the conventional and ANN techniques provides a robust modeling framework capable of capturing the nonlinear nature of the complex water quality time series, thus producing more accurate forecasts.

In general, research activities in water demand forecasting with ANNs suggest that ANNs can be a promising alternative to the traditional linear methods. Jentgen et al. (2007) exploited the performance of ARIMA models and ANNs for forecasting, which are often compared with mixed conclusions in terms of the superiority in forecasting performance. The ARIMA model cannot deal with nonlinear relationships while the neural network model alone is not able to handle both linear and nonlinear patterns equally well. Thus, hybrid models were investigated that are capable of exploiting the strengths of traditional time series approaches and ANNs (Ömer Faruk, 2010). Due to the present complexity in real-life time series efficient approaches are needed.

The forecasting of short-term water demand is the one playing a critical role for water utility companies which are trying to find more efficient ways to supply water (Alvisi et al., 2007). In this subchapter, the use of the two most popular time series forecasting methods for water demand, ARIMA and ANN are investigated. The aim is to forecast the water demand in urban water supply network of Skiathos using the methods of ARIMA, ANN, Winters' Additive Exponential Smoothing and a hybrid ARIMA-ANN approach to obtain more reliable and accurate short- term forecasting.

3.3.2. Materials and Methods for the univariate approach

- *Available data*

The period on which the four models are calibrated and validated is approximately three years, from January 2011 to November 2013. The daily time series is normalized, after the outliers are removed. Monthly and annual-quarterly averaged time series are produced, so as to investigate the trends and seasonality of water demand in micro and macro scale. The forecasting models applied are linear and nonlinear and specifically the seasonal ARIMA, the Winters' Additive Exponential Smoothing and ANN among others.

- *Methodology*

ARIMA

Autoregressive integrated moving average (ARIMA) is one of the most important and widely used linear models in time series forecasting during the past three decades (Zhang et al., 1998). The popularity of the ARIMA model is due to its statistical properties as well as the well-known Box–Jenkins methodology (Box et al., 2015) in the model building process. In addition, various exponential smoothing models can be implemented by ARIMA models (McKenzie, 1984). In an ARIMA model, the future value of a variable is assumed to be a linear function of several past observations and random errors. The linear function is based upon three parametric linear components: autoregression (AR), integration (I), and moving average (MA) method (Box et al., 2015; DeLurgio, 1998).

An ARIMA model can be explained as ARIMA (p, d, q) (P, D, Q)s, where (p, d, q) is the non-seasonal part of the model and (P, D, Q)s is the seasonal part of the model (Box et al., 2015). The p is the order of non-seasonal autoregression, d is the number of regular differencing, q is the order of non-seasonal MA. The ARIMA model order is identified by the trial and error method, with the criteria of the optimal combination of four statistical amounts, the R square, the Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE) and the Mean Absolute Error (MAE).

Winters' Additive Exponential Smoothing

The Winters' Additive Exponential Smoothing technique forecasts a time series that has a linear trend and additive seasonal variation. The initial estimates of the parameters that are updated are usually obtained from the additive decomposition model. However, the initial estimates can be calculated using a multiple regression analysis on the data and employing dummy variables as a measure of the seasonal components (Winters, 1960).

Artificial Neural Networks

An ANN model is a massively parallel-distributed processor that has a natural propensity for storing experiential knowledge and making it available for later use (Allende et al., 2002). It resembles the brain in two respects: The ANN models can recognize trends, patterns, and learn from their interactions with the environment. The firstly most extensively studied and used ANN models are the multilayer feed forward networks (Rumelhart et al., 1986), which allow information transfer only from an earlier layer to the next consecutive layers. Hence, the ANN model, in fact, performs a nonlinear functional mapping from the past observations to the future value y_t. Thus, the neural network is equivalent to a nonlinear autoregressive model.

A neural network must be trained to determine the values of the weights that will produce the correct outputs. In a training step, a set of input data is used for training and presented to the network many times. The performance of the network is tested after the training step is stopped. The backpropagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient) (Govindaraju and Rao, 2000).

Hybrid approach

The combination of the ARIMA and ANN models was performed to use each model capability to capture different patterns in the data. An ARIMA model is not sufficient if there are still linear correlation structures left in the residuals. Therefore, the residuals can be modeled by using ANNs to discover nonlinear relationships. The methodology consisted of two steps: In the first step, the ARIMA model is developed to forecast water demand; and in the second step, the ANN model presented above is used to describe the residuals from the ARIMA model. The hybrid model steps are

implemented using the IBM SPSS Statistics 20 and Zaitun Time Series softwares, respectively.

3.3.3. Results and Discussion

Through the quarterly averaged data processing, it emerges that there is a peak in summer, while spring and autumn periods are ascending and descending, respectively. The dramatic summer increase can be explained by the temperature increase and the intense summer touristic activity, yet this needs to be justified through a multivariate approach. The distribution is seasonal and faintly follows an ascending trend through the years (Figure 11). The ascending trend could be more justified if longer time series data were available; however, it corresponds to the observed annual population and/or tourism increase. The investigation through the adequacy of different models is performed by comparing four statistical amounts, the R square, the Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE) and the Mean Absolute Error (MAE). Winters' Additive Exponential Smoothing is a well performing model for the annual-quarterly demand forecast (Table 2).

Observing the monthly averaged water demand, it can be obtained that the spring ascending slopes are smoother than the autumn descending, which could be related to the tourism distribution (Figure 12). The water demand distribution in winter is more rippled, possibly due to the more intense weather disturbances, causing the models' relative difficulty to fit. Although the ANN model is the most fitting one, it does not obtain the generally ascending trend as the linear models ARIMA (2.0.2)(1.1.0) and Winters Additive Exponential Smoothing do. The Hybrid model seems to fit almost as well as the ANN, additionally it can catch the general trend due to the linear models' properties. The hybrid model seems to be a very adequate forecasting tool (Table 3).

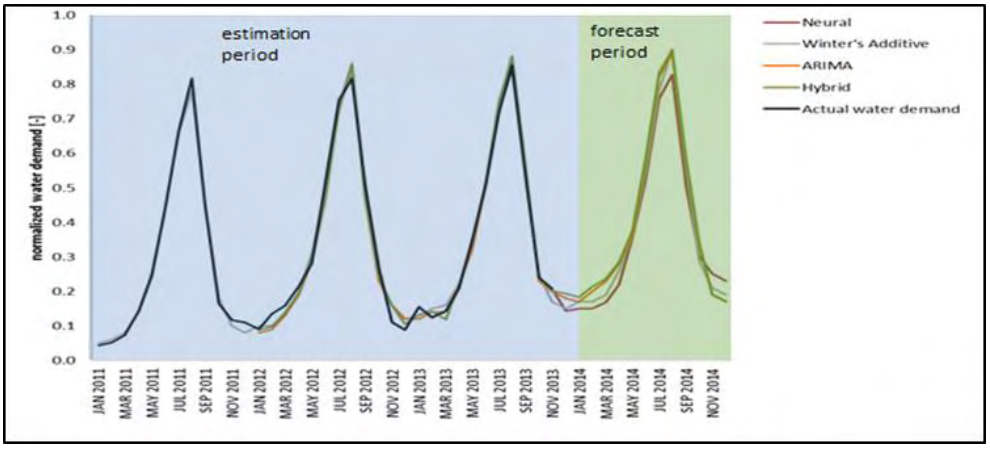


Figure 11. Fit and forecast of the normalized quarterly water demand

Table 2. Statistical amounts for the adequacy of the Winters' Additive Smoothing Exponential model for the quarterly water demand

R squared	RMSE	MAPE	MAE
0.997	0.014	5.009	0.010

Table 3. Statistical amounts for the adequacy of four models for the monthly water demand

Model	R squared	RMSE	MAPE	MAE
Neural	0.999	0.053	0.649	0.003
ARIMA	0.987	0.158	10.493	0.025
Winters' Additive	0.940	0.149	10.797	0.022
Hybrid	0.990	0.125	6.323	0.016

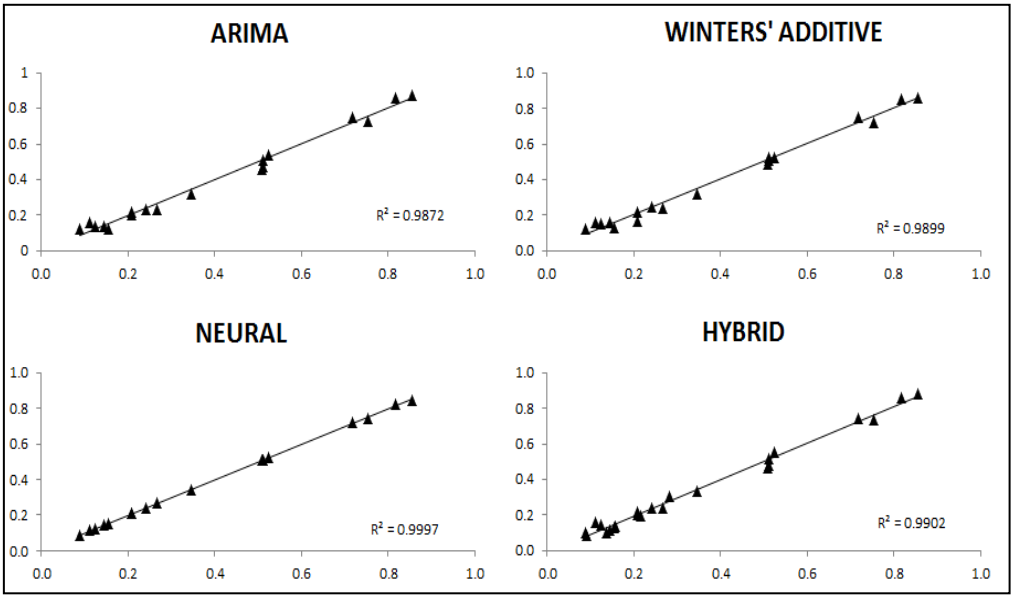


Figure 12. Scatter plots of actual and simulated monthly water demand for the ARIMA, Winters' Additive, Neural and Hybrid models

Afterword: Summarizing, in this subchapter, the strengths of three different popular approaches for urban water demand forecasting via univariate time series analysis are exploited. The first approach is devoted to statistical time series modeling using ARIMA, the second one to Winters' Additive Exponential Smoothing, and the third one to artificial intelligent modeling using neural networks. Each one of the investigated approaches presents the advantageous characteristics of linear and nonlinear modeling respectively. Four common measures of accuracy are applied to assess model performance. Performance of individual time series models is compared to decide the best model so as to ensure appropriate simulation of the water demand time series. Experimental results with monthly water demand data sets indicate that all models can be effective tools to improve the forecasting accuracy obtained by each one of the models.

3.4 Multivariate forecasting of water demand in monthly time scale: ARIMAX, multiple regression, ANN, hybrid approaches

Foreword: In this subchapter a multivariate analysis of Skiathos historical water demand data and five different water demand forecasting methodologies are demonstrated and compared. Linear and nonlinear forecasting methods to a three-year time series water demand are applied. The analysis is held for five variables, mean and high temperature, precipitation, wind speed, describing the meteorology of the island, and arrivals to the island by air and sea, a variable which is used as an indicator for the touristic activity. Time series related to the considered variables are subjected to stationarity test, as the latter plays a key role in the selection of a proper modeling procedure. Five multivariate models are applied, the multiple regression model, multivariate Autoregressive Integrated Moving Average (ARIMAX), Artificial Neural Networks (ANNs) and two hybrid approaches multiple regression-ANNs and ARIMAX-ANNs. The best fit for the monthly averaged data is observed for the multiple regression method –simple and hybrid- and the ANNs.

3.4.1. Introduction on multivariate water demand forecasting

This subchapter deals with the issue of multivariate forecasting water demand. Water demand forecasting can be conducted through a univariate analysis of the water demand fluctuation and a multivariate analysis. The univariate analysis, held out in the previous subchapter for quarterly and monthly data indicates as well fitting forecasting tools the ANNs, the seasonal Autoregressive Integrated Moving Average model (sARIMA), the Winters' Additive method and a hybrid method combining the capability of sARIMA to catch the linear part of the water demand curve and the capability of ANN's to catch any non-linear parts. A multivariate analysis would benefit in terms of depicting the importance of the touristic activity related and meteorological variables to water demand. Such a tool would be quite useful in the case of a dramatic change in tourism or climate. Under such circumstances a univariate model would be inadequate to predict the effect of these changes on water demand levels (Alvisi et al., 2007).

In this subchapter, the analysis focuses on a multivariate approach that uses several characteristic linear and nonlinear forecasting methods; namely, the multiple regression model, the multivariate Auto-Regressive Integrated Moving Average model (ARIMAX), the ANNs and two hybrid approaches, a multiple regression- ANNs and an ARIMAX- ANNs one. The methodologies are tested and applied for the specifics and peculiarities of the Greek island of Skiathos. Skiathos presents a special interest, since it has one of the highest water distribution network Non-Revenue Water levels in Greece—as high as 50% or more, even 70% at times—it is highly touristic, and faces serious water scarcity issues especially during the summer

months. The models are validated and compared through widely used statistical metrics for a single-year forecast.

3.4.2. Materials and Methods for multivariate forecasting

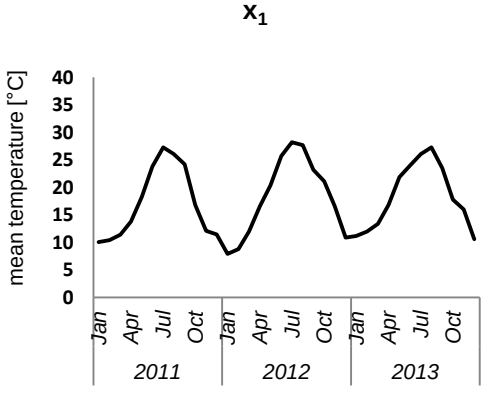
- *Available data*

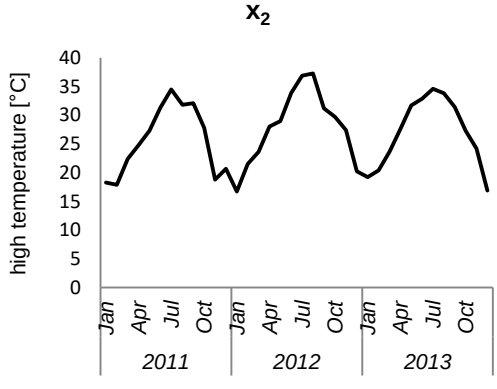
The analysis explores whether meteorological variables and tourist influx are strong predictors for water demand, while evidence of this correlation can be found in the literature (Donkor et al., 2014; Mamo et al., 2013; Billings and Jones, 2008; Arbués et al., 2003; Stevens et al., 1993; Griffin and Chang, 1990; Rumelhart et al., 1986; Agthe and Billings, 1989; Foster and Beattie, 1979; 1981). Although tourist influx may not be explicitly mentioned as a predictor for water demand, population is definitely a strong predictor. Meteorological variables include daily mean and high temperature (variables x_1 and x_2, respectively), precipitation (variable x_3) and daily average wind speed (x_4), while touristic activity is expressed as numbers of tourists arriving at the island by all possible transportation means (ferry and airplane) (variable x_5). Relevant meteorological data are obtained from the private weather station in Skiathos Island, Metar, which operates in collaboration with the National Observatory of Athens (http://penteli.meteo.gr/stations/skiathos/). Touristic activity data are provided by the Touristic Department of the Municipality of Skiathos Island and the Research Institute for Tourism of the Hellenic Chamber of Hoteliers. Water demand (dependent variable y) is a time series of daily groundwater pumping to cover the island's water supply needs. It should be noted that y variable includes water distribution network leakage and other Non-Revenue Water components, but expresses the demand exerted on local water resources. The analysis was based on three-year time series (2011 to 2013) for all variables. Graphs of all variables time-series are shown in Figure 13.

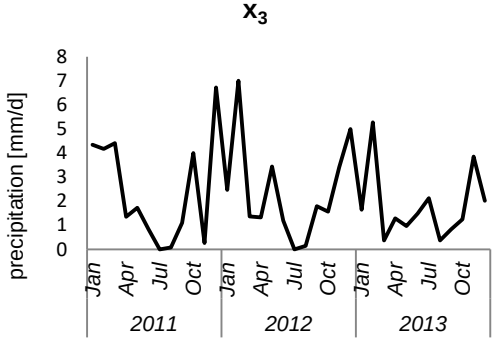
All daily data is averaged to monthly values. The forecast concerns the monthly water demand, y as function of the five independent variables. Corresponding units for all variables are shown in Table 4. A preliminary qualitative analysis shows that the pattern of variability exhibited by independent variables x_1, x_2 and x_5 follows the pattern of the dependent variable y, while x_3 and x_4 are not that consistent with the fluctuation of y. Having said that, we keep in mind that the more noisy variables x_3 and x_4 could play a key role on the nonlinear and/or non-predictable character of winter fluctuations of water demand. The pattern of water demand variability is quite typical for a touristic resort, with summer consumption surpassing by far winter levels. The y variable follows a rapidly ascending and a rapidly descending slope, in spring and autumn respectively, creating a sharp, high summer peak. Winter consumption is generally noisier, indicating that the consumption is possibly more closely correlated to the less seasonal variables, such as x_3 and x_4.

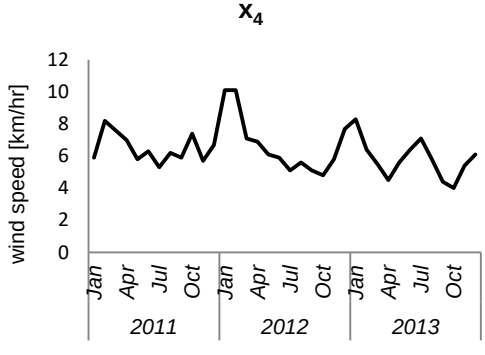
Table 4. Independent and dependent variables

	description of variable	variable symbol	measurement ur
independent variables	mean daily temperature	x_1	°C
	high daily temperature	x_2	°C
	precipitation	x_3	mm/d
	wind speed	x_4	km/h
	total touristic arrivals	x_5	-
dependent variable	normalized water demand	y	-









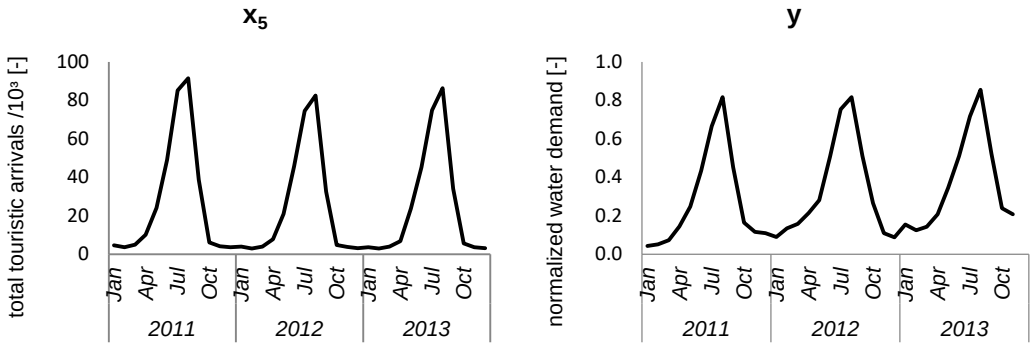


Figure 13. Diagrams of the monthly fluctuation of the five independent and one dependent variables for the three-year period 2011-2013.

❖ *Methodology*

Multivariate analysis is carried out using linear and non-linear models. Two linear models are used: multiple regression and ARIMAX, as well as the multivariate ANNs model (non-linear). Additionally, two hybrid approaches are conducted, a multiple regression-ANN and an ARIMAX-ANN that are both non-linear. Hybrid approaches have been used in different research works in the past, in an effort to combine advantages of two forecasting methods (Zhou et al., 2014; Machiwal and Jha, 2012). Linear regression is used to examine the variables' correlations; while at the same time, stationarity tests are conducted, in order to select appropriate methods for forecasting.

The monthly data of years 2011, 2012 are considered for training the multiple-regression model. To check the quality of fitting of the produced model to historical time series the Akaike Information Criterion -AIC (Akaike, 1974) is used. A stepwise selection of predictor variables is performed that enables to select the best fitted model. For the training of the model the R software package is used (http://www.r-project.org). The initial pool of factors involves all predictor variables: x_1, x_2, x_3, x_4, and x_5. Also, in order to check for possible trends and examine the polynomial regression, first and higher orders of the time index are included: t, t_2 and t_3. In addition, in order to better capture the seasonality, the following harmonic factors are added to the pool: a1 sin (2πt/b1), a2 cos (2πt/b2), where, constants a_1, a_2, b_1, b_2 are tested and determined on a trial and error basis. After running the stepwise selection algorithm, the finally obtained model is shown in equation 1:

$$y = 1.127 \cdot *10^{-2}\, x_1 + 5.615 \cdot *10^{-6}\, x_5 + 4.691 \cdot *10^{-3}\, t + 3.622 \cdot \cdot *10^{-2} \cdot \sin t - 9.955 \cdot *10^{-2} \quad \text{(equation 1)}$$

with AIC equal to -85.992. From equation 1 it is shown that the stepwise selection algorithm identifies only the variables mean daily temperature (x_1), touristic influx (x_5), time (t), and the harmonic function sin(t) as relevant.

The ARIMAX model is defined, in terms of the p, d, q values - parts of the Autoregressive (p), Integrated (d) Moving Average (q) model- by trial and error, with the criterion of an optimal R squared value, as close to 1 as possible (Box et al., 2015; Zhou et al., 2014; Contreras et al., 2003; DeLurgio, 1998). Trials are conducted with the statistical software package SPSS. The model is finally defined as shown in Table 5.

Table 5. Definition, after trial and error procedure, of the p, d and q values of the ARIMAX model

	y	x_1	x_2	x_3	x_4	x_5
p	1	0	0	0	0	1
d	0	0	0	0	0	0
q	1	0	0	0	0	1

ANNs is a widely-used modeling methodology applied with beneficial results in describing non-linear parts of a variable's dynamic function (Bennett et al., 2013; Adamowski and Karapataki, 2010; Adamowski, 2008; Chang and Liu, 2009; Bougadis et al., 2005). The ANN in this sub-chapter is formed in a two-hidden-layer architecture, which was found as most appropriate for the number of independent variables involved in the analysis.

For the two hybrid approaches, firstly, a multiple regression or an ARIMAX model is applied for water demand forecasting and secondly, ANNs are applied for fitting the residuals of forecasted values. This way, it is possible to examine whether the two linear models are missing some important non-linearity, while full advantage of their capability is taken to perform adequately for the linear parts of the variables, such as level, trend or seasonality (Ömer Faruk, 2010; Zhang, 2003). To check the predictive capabilities of the models for the out-of-sample data, year 2013 is considered as a testing period. To quantify the prediction errors, several standard accuracy measures are used such as ME, RMSE, MAE, MPE and MAPE. To evaluate and compare the goodness of fit and suitability of models, the R square criterion is also used.

3.4.3. Results and Discussion

- *Pre-processing: stationarity test and linear correlations*

Stationarity is a key distinctive feature of time series determining the selection of methods for forecasting. An augmented Dickey Fuller (ADF) test (Dickey and Fuller, 1979) is usually used for that purpose. The p-value close to 1 confirms the null hypothesis of non-stationarity. The lag order indicates the order of the autoregressive processes fitted to data with the highest p-value achieved.

The results in Table 6 show that, since p-values are low, there is no evidence for non-stationarity of all univariate components of the considered multivariate time series. Even the largest p-value of 0.341 that is observed for variable x_4 (average wind speed) is quite low, so we conclude that the ADF does not confirm the hypothesis of non-stationarity.

The linear correlation coefficients are shown in table 7 and Figure 14. The strongest positive linear correlation is detected between the total arrivals of tourists (x_5) and water demand (y). Besides the positive influence of temperature (x_1 and x_2), also quite strong negative correlation between precipitation (x_3) and water demand is recognized. The strong correlation between touristic activity and water demand is verified by the ANNs, as well: for almost every run, x_5 turns out to be the main predictor.

Table 6. ADF test of non-stationarity

variable	lag order	p-value
water demand (y)	3	0.023
mean daily temperature (x_1)	3	0.010
high daily temperature (x_2)	3	0.010
daily precipitation (x_3)	3	0.072
average wind speed (x_4)	3	0.341
total touristic arrivals (x_4)	3	0.023

Table 7. Linear correlation coefficients

cor (x_1, y)	cor (x_2, y)	cor (x_3, y)	cor (x_4, y)	cor (x_5, y)
0.918	0.858	-0.606	-0.377	0.965

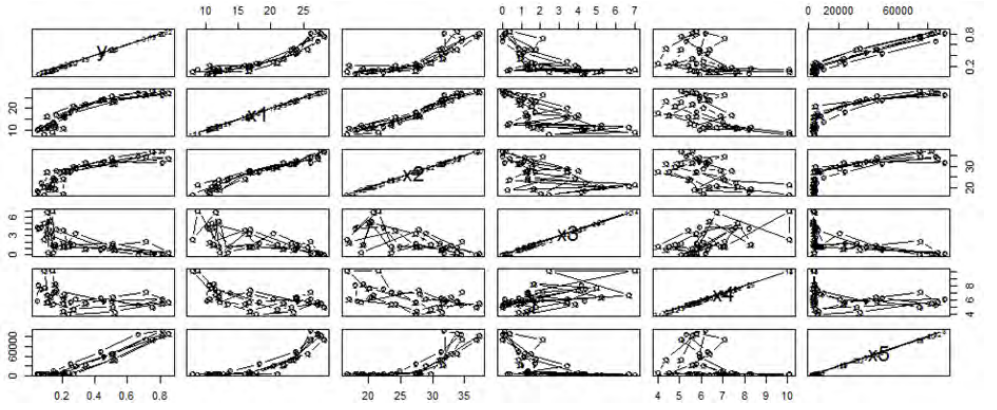


Figure 14. Scatterplot matrix for the dependency between water demand and predictor variables

- *Model Performance*

The fitting of all five models is shown quantitatively in table 5, with five statistical quantities that act as metrics and is visualized in Figure 15, through a comparative diagram of the actual normalized water demand and the models' output fitting time series. The two linear models show adequate fitting capability, which is improved distinctively in their hybrid forms with use of ANNs for modeling the respective residuals. The ANN model also fits quite adequately the actual time series. Improvement, in the case of the hybrid approaches, although not great, is due to the ANNs' advantageous ability to capture non-linear parts of water demand. However, the ANNs cannot capture linearity quickly on their own; thus, they cannot capture the overall slightly annual increasing demand that is observed in the data set for the Skiathos island. In Figure 15, it can be seen that each consecutive peak is higher than the previous one, depicting a definite rise in water demand over the years. All five methods seem to adjust much better in the second training year, which probably means that they have a good training skill. It is also noticeable, that it is harder for the models to fit the actual demand in the winter months than in the ascending and descending summer slopes. This is due to the water demand's high correlation to touristic activity: during the winter months that touristic activity is low, the other independent variables take over and guide the predictions giving a fit that is less accurate. During the summer months, the increase in touristic activity is so dramatic and its correlation with water demand so strong that excellent fits are obtained with most models.

The forecasting strength of all models, which is tested for 2013, is presented quantitatively in table 8, through the same five statistical metrics previously presented and is also visualized in the blue-shaded part of the comparative diagram in Figure 15. Indicative are also the R square values that are presented in Figure 16. All five models have exceptionally high R square values, ranging from 0.984 for the hybrid ARIMAX-ANNs to 0.990 for the multiple regression model. The latter also gives the optimal mean prediction errors, while the ANNs give relatively small errors. The ARIMAX model and the hybrid ARIMAX-ANNs approach give slightly worse results, numerically; however, once the forecast is visualized.

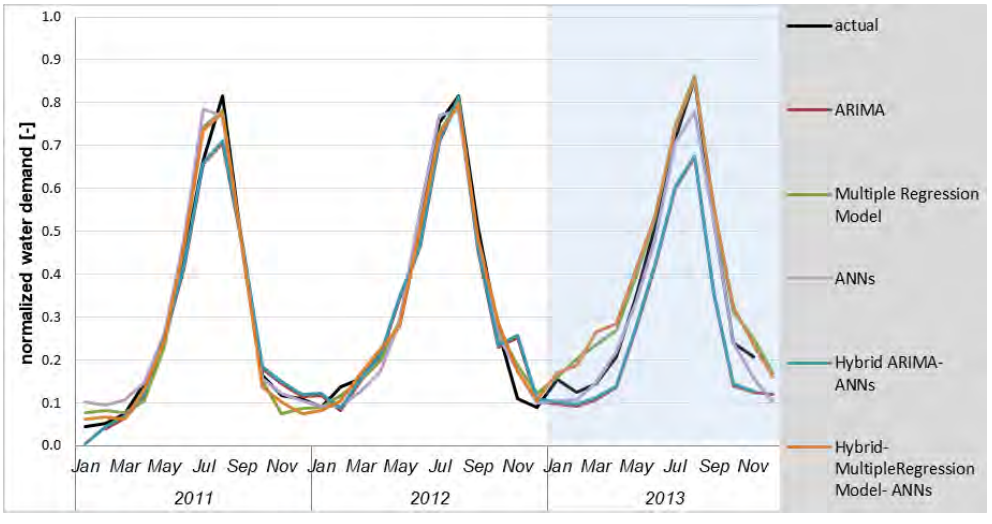
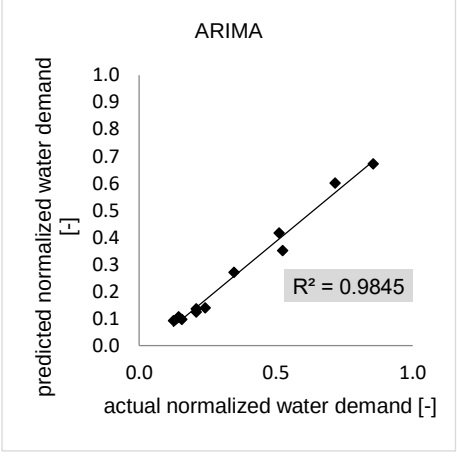


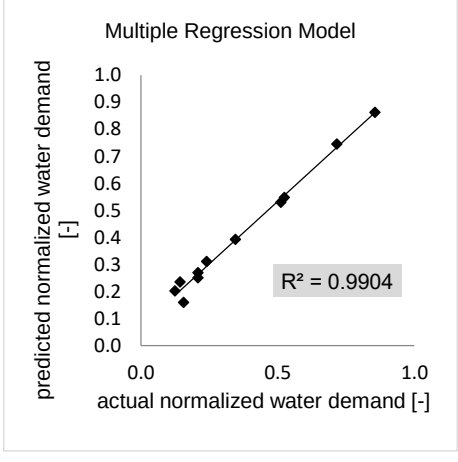
Figure 15. Comparison of actual normalized water demand with the models' fitting and forecasting values, for the training period 2011-2012 and testing - blue shaded-period 2012, respectively

In Figure 15, it is obvious that the two ARIMAX models—simple and hybrid—are incapable of capturing the 2013 summer peak. The seasonal ARIMAX would be expected to be more efficient than the non-seasonal ARIMAX, but needs longer time series than a two-year, to be trained. The ANNs capture better the summer peak, than the ARIMAX, but still not at as well as the multiple regression model. Both hybrid approaches do not contribute anything to the whole forecasting procedure, compared to the simple linear models, probably due to the shortness of the time series, combining with the complexity of the non-linear relationships in the data. A longer training period could give them the benefit of capturing the yet undescribed non-linearity. The multiple regression model seems to generally give the best fit in the output.

Table 8. Accuracy metrics for the five models

Model	ME	RMSE	MAE	MPE	MAPE
ARIMAX	0.006	0.047	0.032	-0.031	0.173
Multiple regression models	0.001	0.031	0.024	-0.020	0.143
ANNs	-0.006	0.040	0.029	-0.063	0.154
Hybrid ARIMAX-ANNs	0.002	0.047	0.032	-0.062	0.171
Hybrid Multiple regression-ANNs	0.001	0.028	0.022	-0.002	0.121





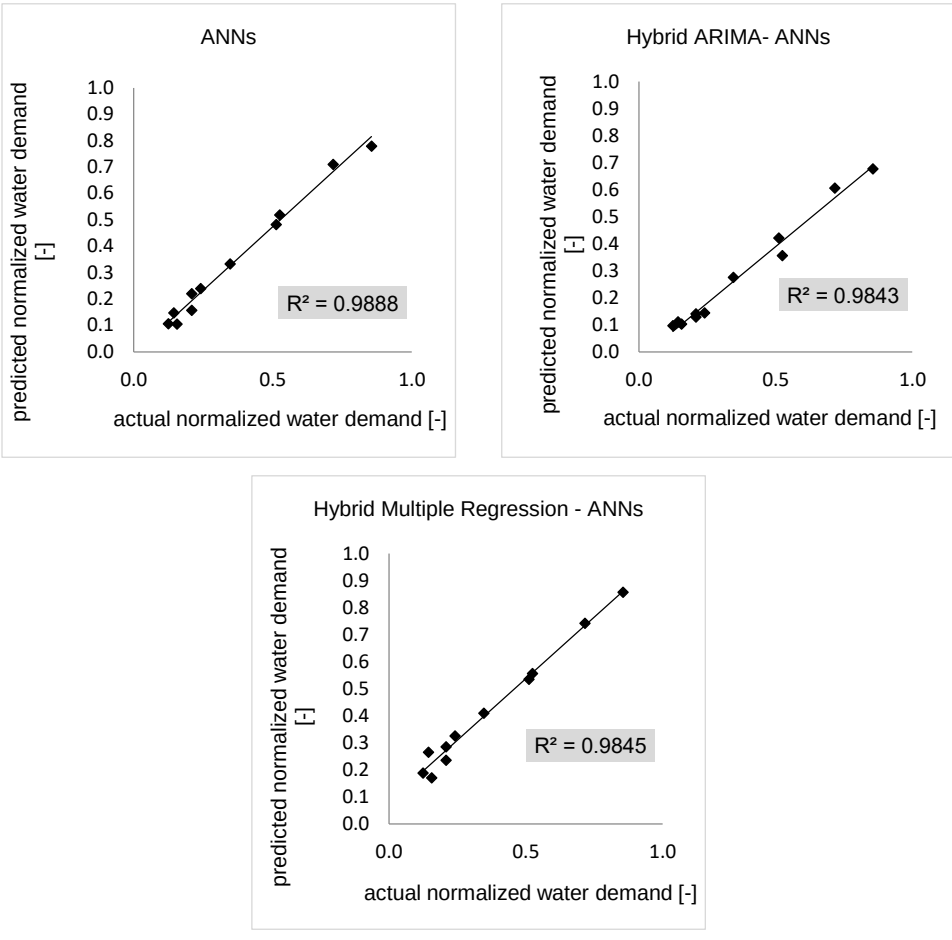


Figure 16. Scatter plots of actual and simulated monthly water demand for the five models

Afterword: In this subchapter, a multivariate analysis of the monthly water demand of the island of Skiathos is presented. The analysis, which is held for a three-year time period, explores the involvement of five variables, meteorological and touristic, in the prediction of the water demand function. An ADF test proves the involved variables to be stationary. Among variables, touristic influx to the island is identified as the most important predictor of water demand and is found highly positively correlated. The temperature is also positively correlated to the demand, unlike precipitation, which is found negatively correlated. The five forecasting methods applied, namely, the multiple regression model, ARIMAX, ANNs, the hybrid ARIMAX-ANNs and multiple regression-ANNs approaches, prove to have adequate fitting and forecasting capacity, especially the multiple regression and the ANNs models. ANNs seem to overcome the difficulty of the winter time more unpredictable fluctuations in water demand, while the multiple regression model captures better the summer peak and the general ascending trend. Longer time series could possibly give the hybrid approaches or the seasonal ARIMAX a chance to fit better.

3.5 Multivariate forecasting of water demand in daily time scale: ANFIS

Foreword: Considering the increasing demand for optimizing water distribution networks in terms of leakage detection and pressure management, as well as the need to reduce urban water consumption, a lot of effort has been invested in the past decade in order to define accurate, long-term and short-term water demand forecasting methods in decreasing time scale. Linear regression models, such as ARIMA, and ANN have been used, as well as different hybrid approaches. In this subchapter, a multivariate analysis of daily water demand of Skiathos Island, Greece and an investigation on the benefits of the ANFIS forecasting method are presented. The applied methodology considers how touristic activity and meteorological and hydraulic variables influence water demand. The applied method benefits when water demand includes non-linear parts and performs adequately; it also provides a Fuzzy Rule Base, giving researchers and water managers a handy tool for interpreting the physical aspects of the inter-relationships.

3.5.1. Introduction

Short-term demand forecasting is needed in order to increase the stability of urban freshwater supply by adjusting water supply to actual demand and consumption, thus resulting in the optimal and timely use of water resources. Demand forecasting is also critical for optimal pump scheduling and thus, for supplying water in a more energetically efficient manner (Skworcow and Ulanicki, 2011). The development of ML algorithms facilitates approaches with finest periodicities such as daily, or even hourly or less allowing for management schemes that would operate towards a near real-time horizon. ANFIS is a promising approach, already used successfully in other fields, such as the energy domain. The adoption of this approach in the urban water management field for demand forecasting is a natural next step.

3.5.2. Materials and Methods

- *Available data*

A multivariate analysis of Skiathos monthly water demand has been presented elsewhere in the previous subchapter. It has been shown that the dominating predictors for water demand are found among meteorological and touristic variables (Donkor et al., 2014; Mamo et al., 2013; Billings and Jones, 2008; Arbués et al., 2003; Rumelhart et al., 1986). Meteorological variables include high daily and mean temperature, rain and wind, while in order to describe arithmetically the touristic activity in the island, the arrivals into the island by any popular transportation means were used as a representative indicator.

The data on which the current analysis is based includes time series of daily water pumping, daily mean and high temperature, daily precipitation, daily wind speed and monthly arrivals by air and sea for a three-year period. Meteorological data are obtained from the private weather station in Skiathos Island, Metar, which operates in collaboration with the National Observatory of Athens[a]. The touristic activity data is taken from the Touristic Department of the Municipality of Skiathos Island and the Research Institute for Tourism[b]. For touristic arrival data, an assumption is made that the monthly arrivals are normally distributed in each month.

The forecast concerns the daily water demand, y as a function of the five independent variables, x_i (Table 9). The variability of the six variables in time is presented in Figure 17. Initially, we note that the independent variables x_1, x_2 and x_5 follow the dependent variable y, while x_3 and x_4 are not consistent to the fluctuation of y. The more noisy variables x_3 and x_4 are expected to play a key role on the nonlinear relations that seem to form the less seasonal-related fluctuations of water demand. The shape of the water demand curve is quite typical for a touristic resort, with summer consumption surpassing by far—almost six times—the winter level. The y variable follows a rapidly ascending and a rapidly descending slope, in spring and autumn respectively, creating a sharp, high summer peak. Winter consumption is generally noisier, showing that the consumption is then mostly correlated to less seasonal variables, such as x_3 and x_4.

Table 9. Independent and dependent variables

	description of variable	variable symbol	measurement unit
independent variables	daily mean temperature	x_1	°C
	daily high temperature	x_2	°C
	daily precipitation	x_3	mm/d
	daily wind speed	x_4	km/h
	monthly total arrivals (assumed to follow a uniform distribution through the days of the month)	x_5	arrivals/month
dependent variable	water demand	y	m^3

[a] http://penteli.meteo.gr/stations/skiathos/

[b] http://www.grhotels.gr/

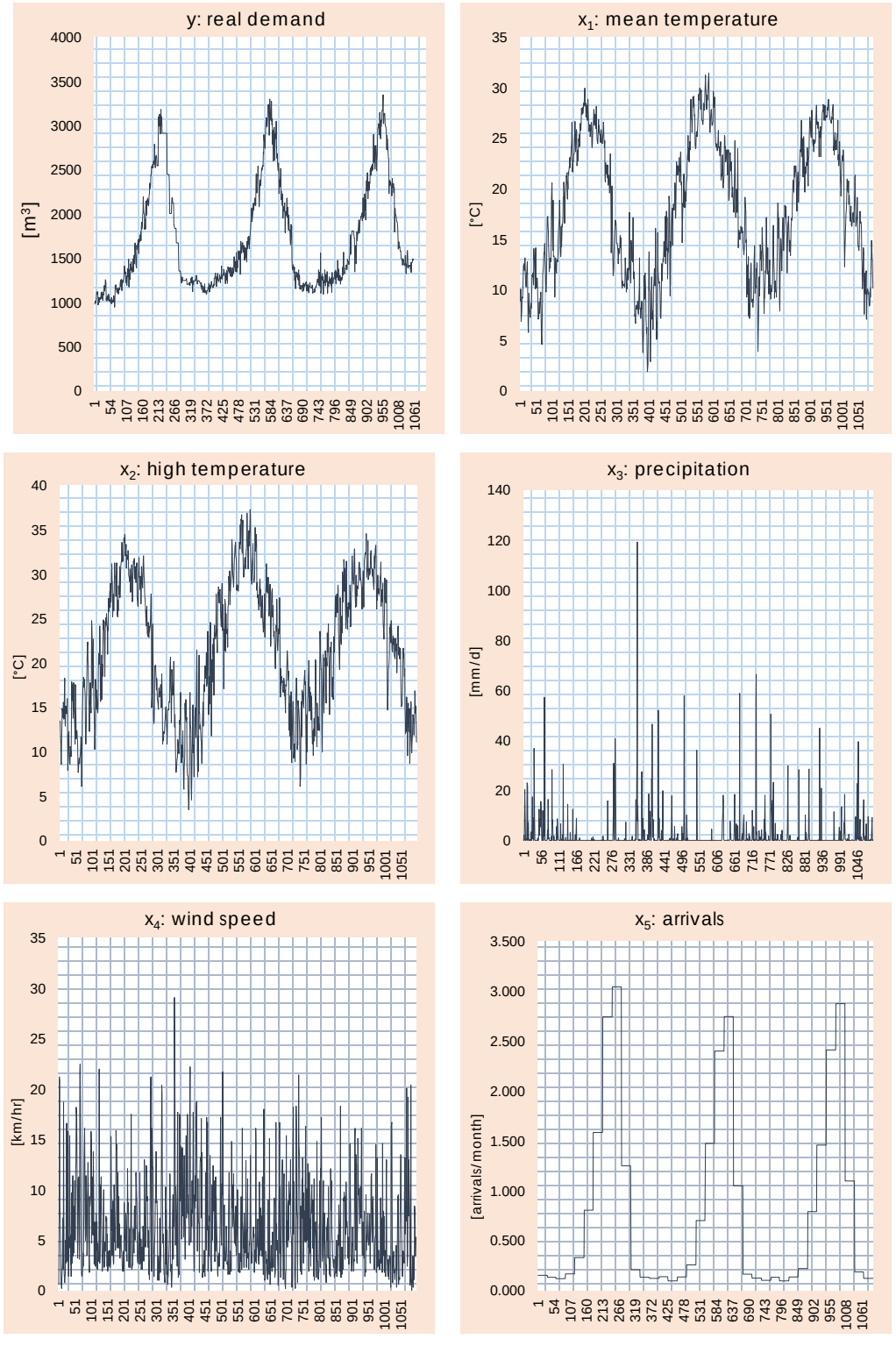


Figure 17. The daily fluctuation of the dependent and independent variables for a three-year period (1096 days)

❖ *Methodology*

The Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS (Jang, 1993) is considered to be an adaptive network, which is very similar to neural networks (NN) (Jang et al., 2005). Adaptive network has no synaptic weights, but it has so-called adaptive and non-adaptive nodes. It should be noted that adaptive networks can be easily transformed to neural networks with classical feed-forward topology. ANFIS integrates the best features of fuzzy systems and neural networks. From fuzzy systems, it inherits the representation of prior knowledge into a set of constraints to reduce the optimization search space and from NN it inherits the adaptation of back-propagation to a structured network in order to automate fuzzy control parametric unit (Jang et al., 2005)

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$.

Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$.

The node in the i-th position of the k-th layer is denoted as Ok,i, and the node functions in the same layer are of the same function family as described below:

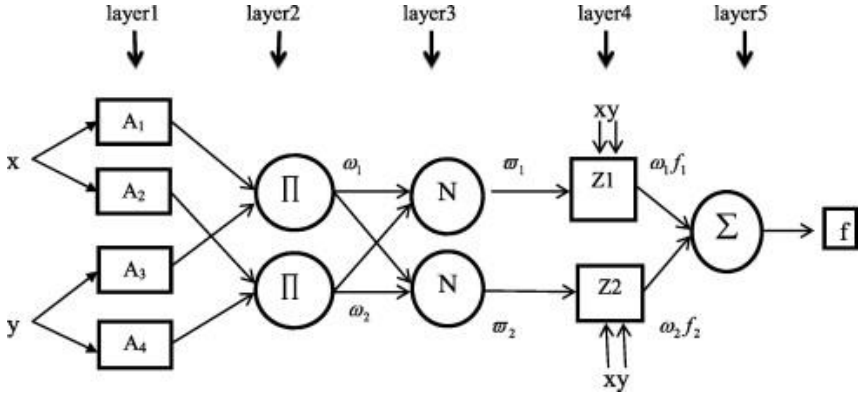


Figure 18. ANFIS structure (Khoshnevisan et al. 2014)

The typical ANFIS structure as shown in Figure 18, includes 5 layers. These five network layers are used to perform the following fuzzy inference steps:

Layer 1: Input fuzzification.

Layer 2: Fuzzy set database construction

Layer 3: Fuzzy rule base construction

Layer 4: Decision making

Layer 5: Output defuzzification

In Layer 1, every node *i* is an adaptive node with a node function, given in equation 2:

$O_{1,i} = \mu Ai(x)$ *(equation 2)*

where x is the input to node i, A_i represents the linguistic label associated with this node function, and $O_{1,i}$ is the membership function of A_i that specifies the degree to which the given x satisfies A_i. To the input y, the node functions in the same layer are of the same function family as x. The most common MFs encompass triangular, trapezoidal and bell-shaped. Bell-shaped MF with a maximum equal to 1 and a minimum equal to 0 are calculated as follows:

$$\mu A(x) = \frac{1}{1 + \left[\frac{(x-c)}{a}\right]^{2b}} \quad \textit{(equation 3)}$$

In Layer 2, every node is a fixed node, and acts as a simple multiplier. The outputs of these nodes are given by eq. (3)

$O_{2,i} = \omega_i = \mu Ai(x) \times \mu Bi(y), i = 1,2, \ldots$ (equation 4)

which are the so-called firing strengths of the rules.

In Layer 3, each node is an adaptive one labeled as N. The i-th node calculates the ratio of the i-th rule's firing strength to the sum of all rules' firing strengths,

$$O_{3,i} = \overline{\omega i} = \frac{\omega_i}{\omega_1 + \omega_2}, for\ i = 1,2 \quad \textit{(equation 5)}$$

In Layer 4, every node is an adaptive one with a function, equation (5)

$O_{4,i} = \bar{\omega}_i f_i = \bar{\omega}_i(p_i x + q_i y + r_i), for\ i = 1,2, \ldots$ *(equation 6)*

where ϖ is the output of layer 3, and p_i, q_i, r_i are referred to as consequent parameters.

In the final Layer 5, the single node is a fixed node labeled as $\sum$ that computes the overall output as the summation of all incoming signals, i.e. equation 7.

$$O_{5,i} = \sum_{i=1}^{2} \bar{\omega}_i f_i = \frac{\sum_{i=1}^{2} \omega_i f_i}{\sum_{i=1}^{2} \omega_i} = overall\ output \quad \textit{(equation 7)}$$

The functions of the various layers are explained in Jang et al. (2005). The number of rules defined in the Sugeno-Type model is a product of the number of membership functions in each input. The output signals from nodes in the previous layers will be accepted as the input signals in the present layer. After manipulation by the node function in the present layer, the output will be served as input signals for the next layer (Cheng and Wei, 2010).

This network is trained in a supervised learning fashion. So some data are used as training data and our goal is to train the adaptive network to enable it to approximate unknown functions given by training data. Training data on water demand were gathered from the water utility company of Skiathos. Our goal is to adapt and determine the right value of the above parameters for function approximation of the mean square error.

ANFIS development for water demand prediction

The ANFIS method is selected, due to its efficiency in decision making, classification and prediction. To find the most effective architecture of ANFIS model, five necessary modifications can be made to increase the accuracy of the network and decrease the errors. These settings include the number of membership functions (MFs), types of MFs (triangular, trapezoidal, bell-shaped, Gaussian and sigmoid), types of output MFs (constant or linear), optimization methods (hybrid or back propagation) and the number of epochs (Bonissone et al., 1995) To develop ANFIS models, MATLAB M-file environment version 7.14.0.739 (R2012a) was used to program ANFIS networks.

The main objective of this study is to develop an ANFIS model based on historical data for predicting the water demand in Skiathos Island. The dataset collected from the previous subsequent years (2011-2013) was used to structure and formulate the ANFIS (Jang, 1993). From this dataset, the first two years were used for learning and the last year for testing (2013).

The efficiency of the model for the different runs-structures is validated through the Root Mean Square Error (RMSE) that corresponds to the testing period and a qualitative look on the produced diagrams of forecasted water demand for the same period. This way, the five most adequate runs are selected and the respective models are validated with the use of Mean Error (ME), Mean Absolute Error (MAE) Mean Absolute Percentage Error (MAPE) and R^2.

3.5.3. Results and Discussion

In Table 10 the RMSE of several different structures of ANFIS and types of MFs are presented for the training and testing periods. Types of MFs are explicitely defined under Table 10. All of the ANFISs are finally realised for the number of epochs that give the smallest average testing error, after it converge. The noteworthy variance

in testing errors proves the importance of selecting the most fitting structure, especially concerning the number and type of input MFs. The testing error varries from 193.0 to 736.6, while the training error varries from 148.1 to 216.8. The five ANFISs that give the best fit based on the training and the average testing errors (shown in bold in Table 10 and having a number in the "Run" column) are further examined. In Figure 19 their scatter plots and a table of more statistical measurements are presented. In Figure 19, comparative diagrams for the testing periods are presented for the 5 ANFISs. The five ANFISs give five well-fitting models, with R^2 up to 0.916 and no lower than 0.910, shown in each diagram. The same conclusion is also conducted from a qualitative observation of the comparative diagrams, as shown in Figure 20, in which the actual and forecasted water demand for a period of about one year are plotted.

The best fitting ANFIS (ANFIS1) has trapezoidal type of MFs and has a 3-2-3-2-3 structure respectively for variables x_1, x_2, x_3, x_4, and x_5. In other words, the best fitting ANFIS divides variables x_2 (high temperature) and x_4 (wind speed) into two MFs each and variables x_1 (mean temperature), x_3 (precipitation) and x_5 (tourist arrivals) into three MFs each. This seems to be an indirect way to understand how extensively each variable needs to be fuzzyfied. To understand what the division in two or three MFs means, the following example is given: Wind speed is classified into two categories—windy and not-windy, for example—while mean temperature is classified in three categories—hot, normal and cold for example. This type of classification in MFs is something that is ultimately defined by an iteration process, in which the number and type of MFs is specified *a priori* and the simulation error that corresponds to each combination is implemented. The optimum combination of MFs is concluded for the one that minimizes the error.

The type of ANFIS architecture that emerges from this analysis is logical, since it can be seen that it requires a more extensive classification in more classes for variables that are highly relevant to the independent variable, water demand (y). In other words, water demand in an island with hot summers is highly correlated with mean temperature; thus, the ANFIS divides this variable (x_1) into three MFs to allow it to better describe its variability; the same is true for precipitation—although here the relationship is reverse. For other variables that are less correlated with water use, such as wind speed, the classification requires only two MFs.

Table 10. Training and testing errors for some of the different tested structures of ANFIS

MFs-Input	Type of MFs	Epochs	Training error	Average testing error	*Run*
3-2-3-2-3	trimf	100	170.5	315.6	
3-2-3-2-3	trapmf	60	172.4	193.0	*1*
3-2-3-2-3	gbellmf	100	164.4	430.4	
3-2-3-2-3	gaussmf	80	165.2	203.2	*2*

3-2-3-2-3	gauss2mf	100	168.8	736.6	
3-2-3-2-3	pimf	100	174.1	295.5	
3-2-3-2-3	dsigmf	37	171.0	516.8	
3-2-3-2-3	psigmf	37	170.9	556.7	
3-2-2-2-3	trimf	20	171.2	245.4	
3-2-2-2-3	trapmf	60	172.8	194.0	*3*
3-2-2-2-3	gbellmf	120	164.5	248.6	
3-2-2-2-3	gaussmf	80	166.2	356.1	
3-2-2-2-3	gauss2mf	40	171.2	200.7	*4*
3-2-2-2-3	pimf	100	174.8	227.1	
3-2-2-2-3	dsigmf	120	167.5	634.2	
3-2-2-2-3	psigmf	120	173.2	487.0	
2-2-2-2-2	trimf	150	181.1	218.0	
2-2-2-2-2	trapmf	100	216.8	252.9	
2-2-2-2-2	gbellmf	200	175.5	274.5	
2-2-2-2-2	gaussmf	100	177.1	227.0	
2-3-2-2-2	trimf	100	176.3	223.2	
2-3-2-2-2	trapmf	100	195.7	238.6	
2-3-2-2-2	gbellmf	100	176.6	315.2	
2-3-2-2-2	gaussmf	100	174.8	247.5	
2-3-2-2-3	trimf	100	173.7	207.1	
2-3-2-2-3	trapmf	100	194.8	237.2	
2-3-2-2-3	gbellmf	250	167.6	217.5	
2-3-2-2-3	gaussmf	150	168.6	214.4	*5*
3-3-3-3-3	trimf	50	162.9	234.1	
3-3-3-3-3	trapmf	100	184.3	217.5	
3-3-3-3-3	gbellmf	100	155.4	325.5	
3-3-3-3-3	gaussmf	100	148.1	265.8	

The types of MFs shown in the table are the following:

trimf: triangular MF

trapmf: trapezoidal MF

gauss2mf: Gaussian2 MF

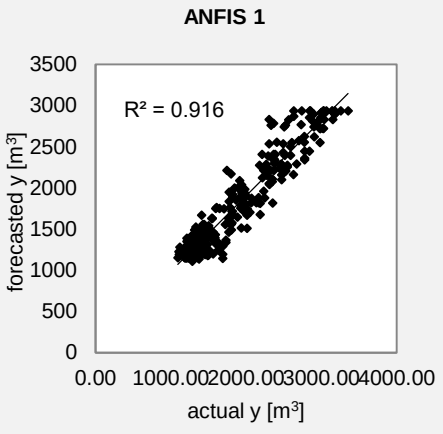
pimf: pi MF

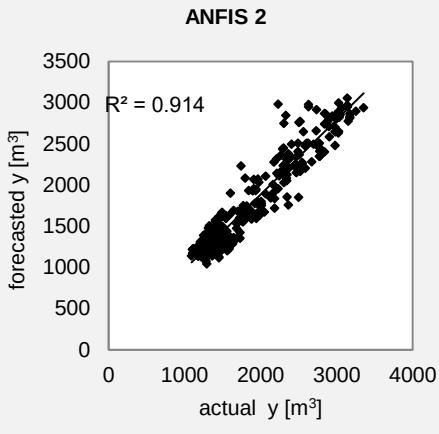
gbellmf: generalized bell MF

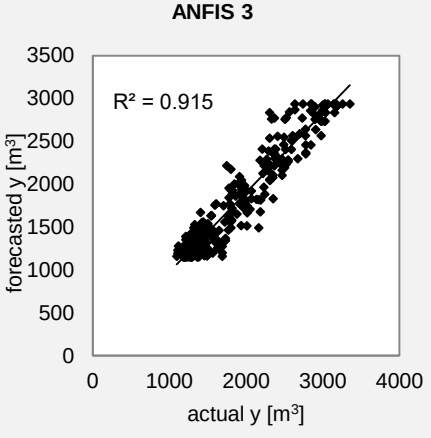
gaussmf: Gaussian MF

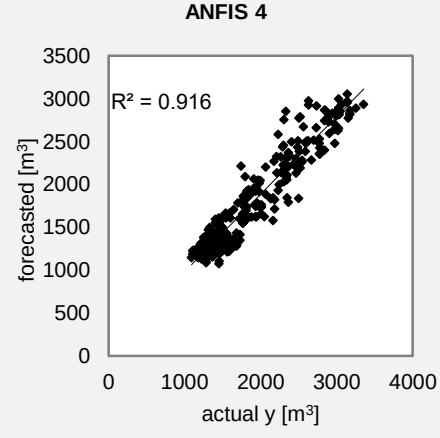
dsig: difference of two sigmoidal MFs

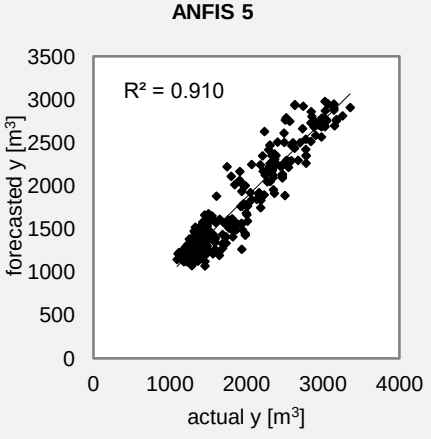
psig: product of two sigmoidal MFs





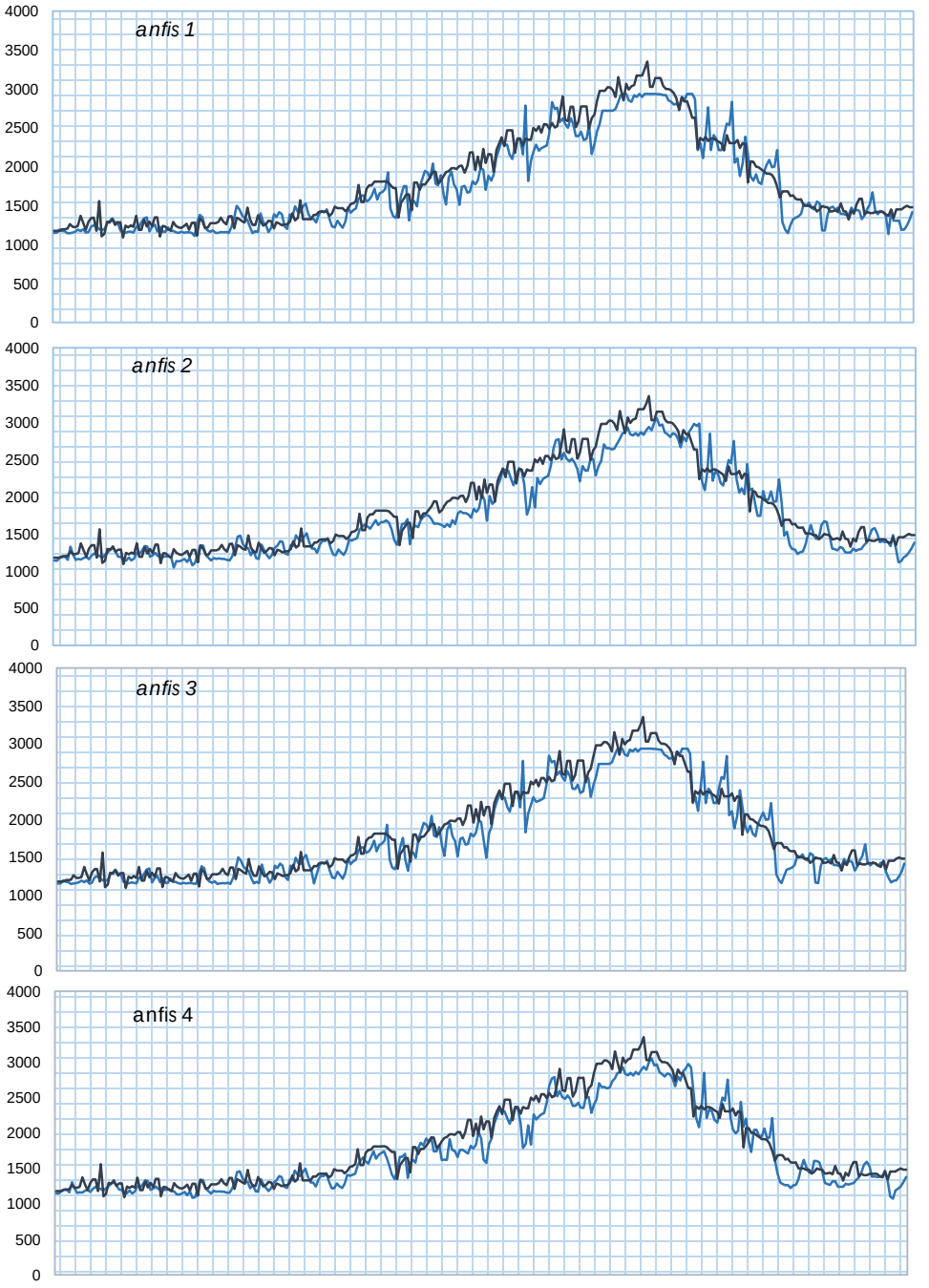






ANFIS	ME	RMSE	MAE	MAPE
1	82.86	192.99	151.23	0.081
2	100.28	203.16	159.91	0.085
3	82.91	194.04	151.26	0.082
4	99.75	200.66	156.37	0.083
5	114.6	214.4	170.7	0.090

Figure 19. Comparative diagrams of forecasted water demand versus actual water demand for the five ANFIS structures for the testing period



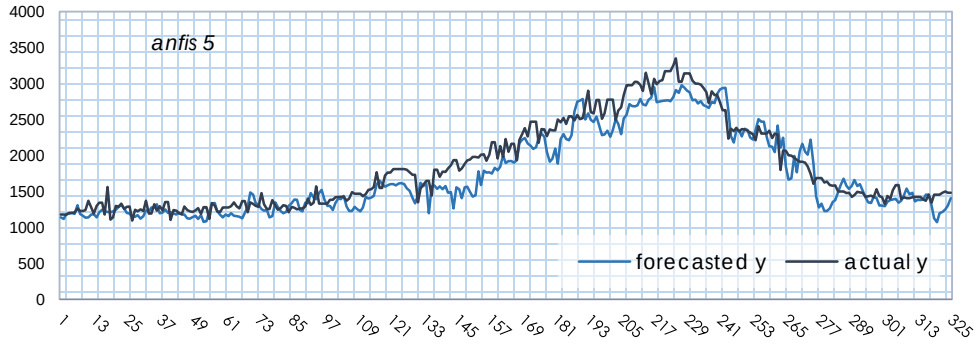


Figure 20. Comparative diagrams of forecasted and actual daily water demand (m^3) for the days of the testing period, for the 5 best ANFISs

In Figure 21, the stucture of the best fitting ANFIS, ANFIS 1, as described in Table 10, is depicted through its nodes and rules. The figure shows the complexity added to the system for just a single membership function or even more for an extra variable. In the figure, the black dots under the "input" heading represent the variables, while the white dots under the "inputmf" heading represent the MFs. The number of rules define the number of black lines leaving the white dots, while the number of rules depends on the number of variables and MFs. It is obvious that the addtion of an extra black or white dot (variable or MF, respectively) will add in a multiplicative fashion to the complexity of this structure and to the calculation load.

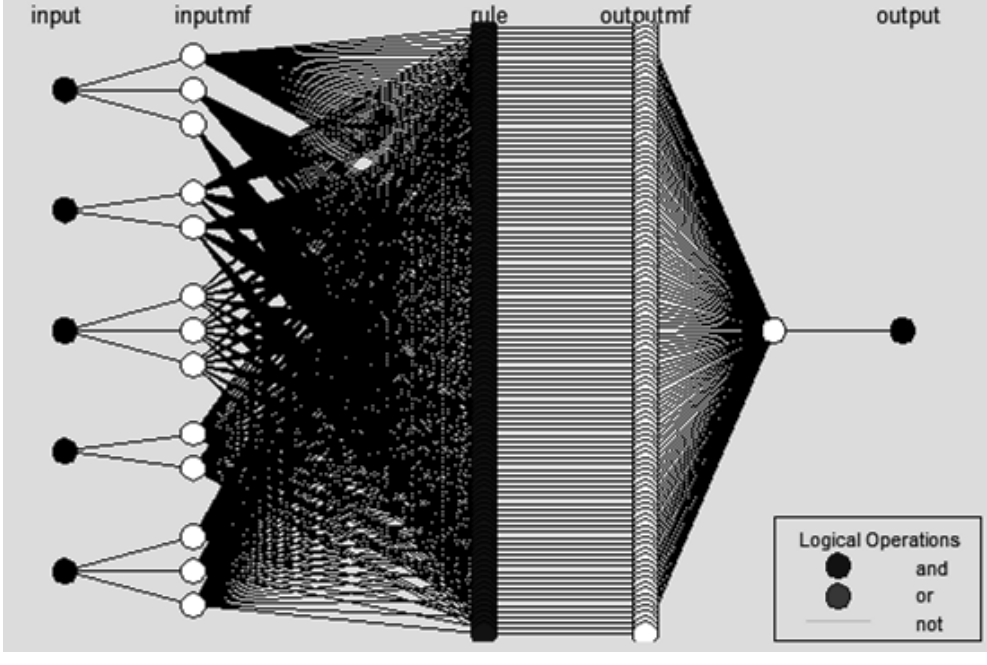


Figure 21. ANFIS 3-2-3-2-3 structure

Afterword: A daily multivariate forecasting approach for urban water demand is tested. The ANFIS methodology is implemented. The predictors tested are

meteorological variables and a total arrivals into the island by air and sea that performs as an indicator of the touristic influx. The algorithm is trained on two years time series and evaluated on approximately one year daily time series. Several architectures have been tested and the methodology performs sufficiently with R^2 up to 0.916.

3.6 Multivariate Forecasting of water demand in daily time scale, introducing the NRW predictor: ANN and ANFIS

Foreword*:* Advanced techniques have the ability to overcome the non-linearity issues commonly met when investigating the complex relationship of water demand and weather, socioeconomic and other variables. In this subchapter two approaches, an ANN and an ANFIS are presented, for forecasting a Mediterranean touristic resort daily water demand based on weather variables, tourism and infrastructure efficiency. For the two later indicative metrics are used, namely the total arrivals into the island by sea and air and the Non-Revenue Water percentage, respectively. Both models seem to have an adequate response, though ANFIS can more smoothly catch winter non-touristic water demand profile.

3.6.1. Introduction

The ANFIS approach is increasingly adopted by various fields as a target variable forecasting algorithm. Among other fields, the water domain has used ANFIS to predict damn levels, groundwater levels, water quality parameters etc. Urban water demand forecasting can for sure make use of a promising forecasting tool that can operate in multivariate analysis. While ML algorithms such ANN have been used in water demand forecasting, a comparison of the performances of ANN and ANFIS needs to be done to showcase the best choice for planners.

3.6.2. Materials and methods

- *Available data*

In this subchapter, five predictors are used to forecast daily water demand of the town, including leakages, namely daily mean temperature (x_1), daily high temperature (x_2), daily precipitation (x_3), monthly touristic inflows through all possible transportation means (x_4) and trimester leakage level (x_5) as a percentage of water demand (Table 11). The forecasted daily water demand (y), represents customer consumption and leakage and is estimated through the daily water amount pumped from groundwater. The meteorological data sets are obtained from the weather station in Skiathos, Metar, which operates in collaboration with the National Observatory of Athens (penteli.meteo.gr). The touristic data, for which the assumption of uniform monthly distribution has been made, were provided by the Touristic Department of the Municipality of Skiathos Island and the Research Institute for Tourism of the Hellenic Chamber of Hoteliers. The NRW level has been estimated as the relative difference of trimester water bills and total pumped water. Network and billing data are provided by the Skiathos water company, DEYAS

(www.deyaskiathos.gr). The rough assumption of uniformity is made to surpass the problem of a constant NRW rate for each trimester (Figure 8 and Figure 10).

Two forecast methodologies are applied, ANN and ANFIS. Both methods are applied for numerous possible architectures, in order to find the most fitting ANN and ANFIS structure. All data sets refer to the time period from January of 2011 to July 2015 and are divided to 70% training sets and 30% testing sets.

Table 11. The variables used in the two models

Variables	*Description*	*Units*
x_1	daily mean temperature	°C
x_2	daily high temperature	°C
x_3	daily precipitation	mm
x_4	daily arrivals (monthly arrivals are uniformly distributed)	arrivals/d
x_5	trimester NRW as a percentage of water demand (uniformly distributed)	% water demand
y	water demand including NRW	m^3/d

- *Methodology*

Artificial Neural Network

ANN is a system of interconnected units known as neurons interacting across weighted connections. Inspired by the architecture of the human brain, these neurons can compute output values from inputs by learning complex patterns of information and generalizing the learned information. ANNs can be classified into several categories based on supervised and unsupervised learning methods and feedforward and feedback recall architectures. Multilayered feedforward networks use a supervised learning method and feedforward architecture. A backpropagation neural network is one of the most frequently utilized neural network techniques for classification and prediction (Deyfus, 2005).

The proper selection of ANN modeling parameters contributes to the success of ANN in prediction and classification tasks. The number of nodes (neurons) and layers, the nonlinear function used in the nodes, the learning algorithm, the initial weights of the inputs and layers, and the number of epochs for which the model is iterated, are the most important parameters that need to be defined for ANN performance. In general, the ANN model has a structure of an input layer, a hidden layer, and an output layer. In ANN approach, the input data are divided into 2 main subsamples, which are called learning and testing sets (Haykin, 2005). The neural network uses initially the learning sets to learn the relationship between the output and input criteria, while the test set is used to assess the performance of the model during the

testing process. ANNs have been suggested as an alternative to time series forecasting to deal with linear and nonlinear relationships. The major advantage of NNs is their flexible nonlinear modeling capability where the data are fitted for prediction purposes. With ANNs, there is no need to specify a particular model form. Rather, the model is adaptively formed based on the features presented from the data. This data-driven approach is suitable for many empirical data sets where no theoretical guidance is available to suggest an appropriate data generating process.

The multilayered feedforward ANN models were implemented in Matlab R2012a, using the functions provided by ANN toolbox. The two most used training functions, the Levenberg-Marquardt (LM) and gradient decent (GD) were used. The Levenberg-Marquardt backpropagation algorithm worked more efficiently to train the ANN model for prediction tasks, providing less mean square errors and correlation coefficients. The optimization algorithm was selected as a conjugate gradient algorithm

Adaptive Neuro-Fuzzy Inference Systems

A fuzzy inference system (FIS) is an inference mapping that provides an intuition for the relationship between a series of input and output sets. This mapping from a given input to an output using fuzzy logic is called Fuzzy Inference (Adriaenssens et al., 2004). These systems have proved to work better when the input and output sets are time series data of the same time step. The FIS uses fuzzy logic principles to establish the input-output relationship through a rule based inference engine that consists of: (a) a rule-base, containing fuzzy if–then rules, (b) a data-base, defining the membership functions (MF) and (c) an inference system, combining the fuzzy rules and producing the system results. There are two types of popular FIS, the Takagi–Sugeno FIS, (Takagi and Sugeno, 1985) and the Mamdani FIS (Jang et al., 2005). The difference between the two approaches is the definition of the consequent parameters in the network. The FIS used in this study is a Takagi and Sugeno type FIS in which the rule base is constructed from the input–output pairs and it consists of five layers as seen in Figure 22: (L1) Input fuzzification, (L2) Fuzzy set database construction, (L3) Fuzzy rule base construction, (L4) Decision making and (L5) Output defuzzification.

In Layer 1, every node is an adaptive node with a node function, given in equation 8:

$$O_{1,i} = \mu_{A_i}(x) \quad for\ i = 1,2 \qquad \textit{(equation 8)}$$

Where x indicates the input to node I, A_i represents the linguistic label associated with this node function, and $O_{1,i}$ is the membership function of A_i that specifies the degree to which the given x satisfies A_i. Regarding all other input y, the node functions have exactly the same behavior with the function family as x, with the condition that they belong to the same layer. In Layer 2, every node is a fixed node and acts as a simple

multiplier. The outputs of these nodes, which are the so-called firing strengths of the rules, are given by equation 9.

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y) \quad for\ i = 1,2 \qquad \textit{(equation 9)}$$

Each node, in Layer 4, is an adaptive node with a function given by equation 11,

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2} \qquad \textit{(equation 10)}$$

$$O_{4,i} = \overline{w_i} f_i = w_i(p_i x + q_i y + r_i) \qquad \textit{(equation 11)}$$

Where $\overline{w_i}$ is the output of layer 3, and p_i, q_i, r_i are referred to as consequent parameters. Finally, the single node, in Layer 5, is a fixed node indicated as ∑ (sum) that computes the overall output as the sum of all incoming inputs:

$$O_{5,i} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \qquad \text{(equation 12)}$$

To construct an ANFIS from a given input/output data set, we first construct the FIS whose membership function parameters are tuned (adjusted) using either a back propagation algorithm alone or in combination with a least squares type of method (Singh et al., 2012). Learning using the neuro-adaptive method works similarly to that of neural networks as for the procedure to learn information about a data set. In other words, ANFIS, which is a combination of ANN and FIS, has the benefits of the two models (Azadeh et al., 2011). Propagation and hybrid are two learning methods, which are generally applied in ANFIS to clearly describe the relationship between input and output (Khoshnevisan et al., 2014). Hybrid learning, which is a combination of gradient decent method and least squares approach, can decrease the complexity of the algorithm and simultaneously increase the learning efficiency. The parameters associated with membership functions will change through the learning process using a gradient vector that facilitates in this recalculation. So every time the gradient vector is obtained, an optimization procedure can be performed to adjust parameters in order to reduce errors.

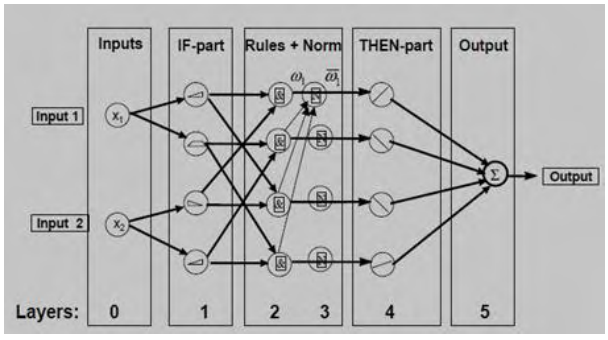


Figure 22. Adaptive neuro-fuzzy inference system structure (Khoshnevisan et al., 2014)

3.6.3 Results and Discussion

Daily forecast of water demand seems to be achievable at satisfactory accuracy as tested for numerous ANN and ANFIS structures. Several popular metrics are assessed to evaluate the effectiveness of the two methods and select the best fitting architecture of each approach. The metrics of accuracy used are the root-mean-square error (RMSE), the mean absolute error (MAE), the mean percentage error (MPE), the mean absolute percentage error (MAPE), the Nash–Sutcliffe model efficiency coefficient (E), used to assess the predictive power of water demand models (equations 13-17) and the R^2 values produced by scatter plots of actual versus forecasted values (Figure 25).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - f_i)^2}{n}} \qquad \textit{(equation 13)}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - f_i|}{n} \qquad \textit{(equation 14)}$$

$$MPE = \frac{100\%}{n} * \sum_{i=1}^{n}\frac{y_i - f_i}{y_i} \qquad \textit{(equation 15)}$$

$$MAPE = \frac{1}{n} * \sum_{i=1}^{n}\left|\frac{y_i - f_i}{y_i}\right| \qquad \textit{(equation 16)}$$

$$E = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \qquad \textit{(equation 17)}$$

where, y_i is the actual value and f_i is the forecasted value.

Regarding ANFIS, the Matlab 2014(b)-ANFIS tool named ANFIS-Editor is used. The tool is designed to utilize different variables including a normalization method, trial step quantity and various data classification methods to achieve the minimum error between predicted values and real data. The number and type of membership functions, (MF), the type of output MF, the optimization method (hybrid or back propagation) and the number of epochs are five important adjustments in ANFIS to reach the most effective model with minimum errors. Figure 23 summarizes most of the types of membership functions used in our simulation. The primary goal is to find the effect of these adjustments and their subdivisions in different combinations in order to develop these ANFIS models and compare the results. For this purpose, all possible combinations of numbers of membership functions for each predictor, from 2 to 4 and types of membership functions [triangular (trimf), generalized bell-shaped (gbelmf), Gaussian (gaussmf), Gaussian combination (gauss2mf), trapezoidal (trapmf), Π-shaped (pimf) and sigmoidal (dsigmf)] are implemented.

The structure that gives the best fitting would fuzzificate variables x_1-x_5 into 3,2,2,2 and 3 membership functions respectively and the fuzzification would be implemented with the use of a triangular shaped function. The partitioning is implemented through the grid partitioning method.

The success of ANN models depends on properly selected parameters such as the number of nodes (neurons) and layers, the nonlinear function used in the nodes, the learning algorithm, the parameters of learning, the initial weights of the inputs and layers, and the number of epochs for which the model is iterated (Maier and Dandy, 2000; Bishop, 1995). A large number of experiments with ANN multilayered feed-forward architecture is accomplished, considering different number of hidden layers, different learning rate and momentum parameters, two different (the most efficient) learning algorithms of backpropagation ANN technique (LM and GD) and number of epochs. Through the experiments, the best results of back propagation ANNs have been received for the architecture of one hidden layer, with 10 neurons, learning rate = 0.3 and momentum = 0.1, random values of initial weights, and the Levenberg-Marquardt back-propagation algorithm as the learning algorithm. The optimization algorithm is selected as a conjugate gradient algorithm. The hyperbolic tangent transfer function was used in the hidden layer, and a linear transfer function is used in the output layer. The number of epochs for best configuration is 100.

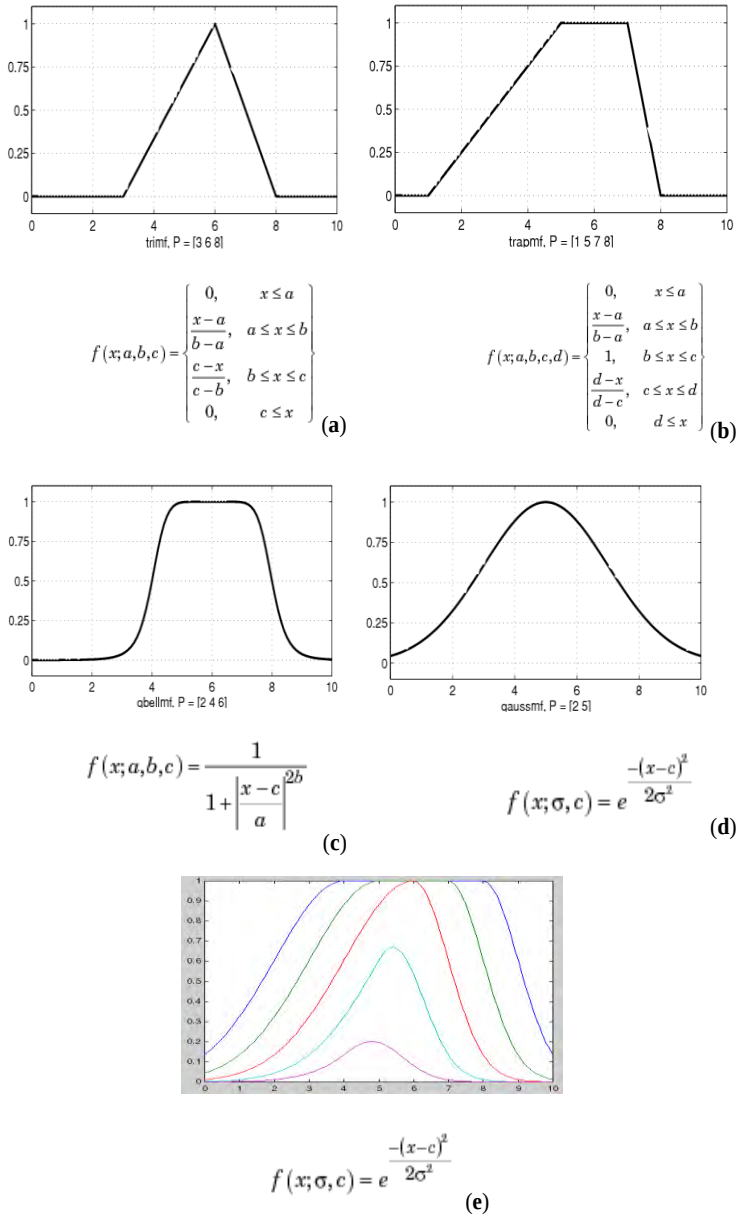


Figure 23. Indicative Membership functions used in the ANFIS: (a) trimf, (b) trapmf, (c) gbelmf, (d) gaussmf and (e) gauss2mf

Table 12 and Figure 24 and Figure 25 show that both methods are adequate and can produce quite accurate water demand predictions. Comparing the two methods, ANFIS gives better results in all estimated metrics of accuracy. In Figure 24, it is showcased that the weakest—yet acceptable—fitting is observed during the period from October to December 2014. This is a result of using a single NRW rate throughout the trimester (October through December), which is realistic overall, but performs sub-optimally, when compared with water demand that is reported on a daily basis.

Naturally, the metrics and overall predictions will significantly improve if NRW percentage data of greater time-scale are included. For investigating if the two methods perform differently in warm and cold periods, the evaluation period is split into April to October and January to March plus November to December. The estimated performance metrics prove that ANFIS performs better than ANN in both sub periods, as shown in table 12.

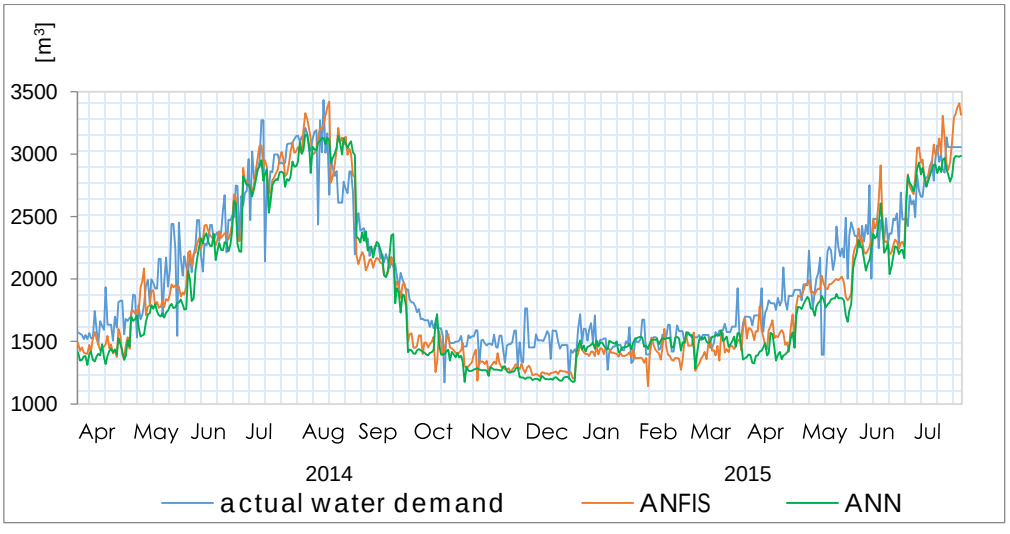


Figure 24. Comparative plots of actual water demand, ANFIS and ANN forecast versus time for the testing period (April 2014-July 2015)

Table 12. Metrics of accuracy for the ANFIS and ANN models for the whole evaluation period and for split warm and cold periods

	Method	*RMSE*	*MAE*	*MPE*	*MAPE*	*E*
the whole of the year	ANFIS	220	177	5.34	0.09	0.84
	ANN	249	200	7.36	0.10	0.79
Jan-Mar & Nov-Dec	ANFIS	218	193	11.74	0.13	-2.60
	ANN	254	231	14.39	0.15	-3.85
Apr-Oct	ANFIS	212	176	5.09	0.09	0.66
	ANN	270	232	9.89	0.116	0.45

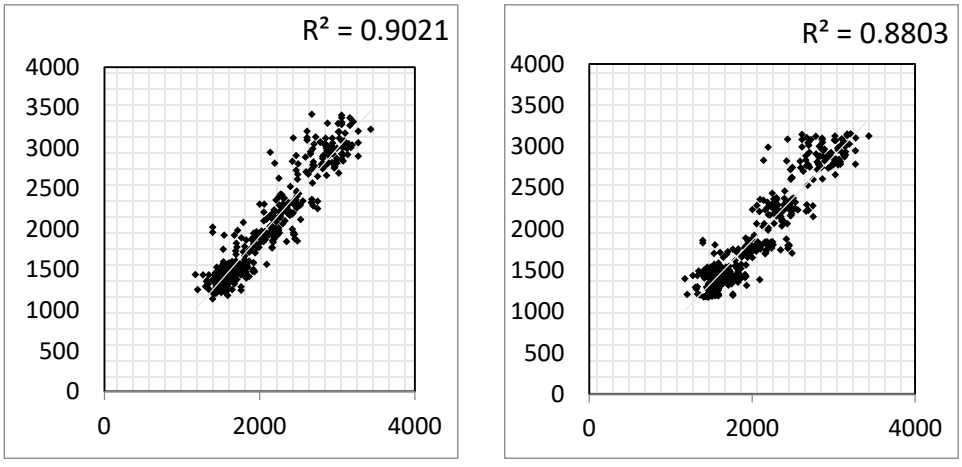


Figure 25. Scatterplots and R^2 for the two approaches ANFIS (left) and ANN (right) for the whole evaluation period

Afterword: In this subchapter, an approach for multivariate daily prediction of water demand in a highly touristic Mediterranean resort is introduced. The predictive adequacy of two forecasting methods, ANN and ANFIS is tested through various adequacy metrics. The drivers of water demand: mean and high temperature, precipitation, arrivals by any transportation means and NRW level, are proven to be capable of building up two quite satisfying models. The two methods seem to overcome any non-linearity or collinearity of the predictors issues that have been noticed in relative literature. The ANFIS method gives better results in all tested metrics. The methodology can be a useful forecasting tool that can be used either on prediction or water management by testing possible scenarios of expected shifts of the driving variables.

Parts of the content of chapter 3 is included in the following published works:

Kofinas, D., Mellios, N., Papageorgiou, E., & Laspidou, C. (2014). Urban water demand forecasting for the island of Skiathos. Procedia Engineering, 89, 1023-1030.

- The contribution of Mr. Kofinas, D. involves the conceptualization, the methodology, the validation, the formal analysis, the investigation, the writing, and the visualization.
- The contribution of Mr. Mellios, N. involves the provision and curation of data.
- The contribution of Associate Professor Papageorgiou, E. involves the review of the article.
- The contribution of Professor Laspidou, C. involves the scientific supervision.

Kofinas, D., Papageorgiou, E., Laspidou, C., Mellios, N., & Kokkinos, K. (2016, April). Daily multivariate forecasting of water demand in a touristic island with the use of artificial neural network and adaptive neuro-fuzzy inference system. In 2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater) (pp. 37-42). IEEE.

- The contribution of Mr. Kofinas, D. involves the conceptualization, the methodology,
- the validation, the formal analysis, the investigation, the writing, and the visualization.
- The contribution of Associate Professor Papageorgiou, E. involves the review of the article.
- The contribution of Professor Laspidou, C. involves the scientific supervision
- The contribution of Mr. Mellios, N. involves the provision and curation of data.
- The contribution of Dr. Kokkinos, K. involves the programming of involved code.

Mellios, N., Kofinas, D., Papageorgiou, E., & Laspidou, C. (2015, June). A multivariate analysis of the daily water demand of Skiathos Island, Greece, implementing the artificial neuro-fuzzy inference system (ANFIS). In *E-proceedings of the 36th IAHR World Congress* (Vol. 28, pp. 1-8).

- The contribution of Mr. Mellios, N. involves the provision and curation of data.
- The contribution of Mr. Kofinas, D. involves the conceptualization, the methodology,
- the validation, the formal analysis, the investigation, the writing, and the visualization.
- The contribution of Associate Professor Papageorgiou, E. involves the review of the article.
- The contribution of Professor Laspidou, C. involves the scientific supervision

This work was supported by the project ISS EWATUS—Integrated Support System for Efficient Water Usage and Resources Management—which is implemented in the framework of the EU 7th Framework Programme, Specific programme Cooperation Information and Communication Technologies; Grant Agreement Number 619228.

4. Building a spatio-temporal water distribution network simulation model

4.1 Spatio-temporal disaggregation of system input volume and non-revenue water of a water distribution network

Foreword: Pressure control management of a water distribution network is considered as an effective approach for the reduction of leakage in the network and for optimized savings in pumping energy. A successful pressure management scheme usually requires single- or multi-feed regulation through Pressure Reduction Valves, the network division in District Metered Areas and dense monitoring of pressure and flow throughout the network. A structural component of this approach would be a hydraulic model which would relate pressure and flow in the network, in accordance to the monitored values. In the ISS-EWATUS project ("ISS-EWATUS Integrated Support System for Efficient Water Usage and Resources Management", 2016) and specifically for the Skiathos, Greece case study, the water distribution network is simulated using EPANET software, so as to estimate the pressure-map of the island and later link pressure to leakage. The need to zoom-in spatially and temporarily makes the spatial and temporal disaggregation of aggregated System Input Volume (SIV), as provided by the water utility, a necessity. The daily aggregated water supply time series and consumer quarterly billing data for each water meter are used, in order to produce approximate, daily water consumption datasets of each household and daily Non-Revenue Water per household. The initial disaggregation is implemented roughly assuming that the leakage components as well as the metering inaccuracies and Unauthorized Consumption are proportionate to the billings (Billed Authorized Consumption, BAC). The more accurate formulas that link the leakage to pressure are not taken into account in this step. However, the produced data sets are used in the next chapter as initial values that are corrected after a nested double iterative process to meet the levels that can produce pressure values respective to the actual recorded ones. In this subchapter, the disaggregation methodology for the production of these data sets of initial values is presented.

4.1.1. Introduction: what does a spatio-temporal analysis of a Water Distribution Network offer?

The urban touristic phenomenon has long been investigated as a structural component of urban planning (Getz, 1993;Inskeep, 1987), while recently it has been linked to the idea of the smart city paradigm. La Rocca (2014) highlights the link of the dynamic and spatial variable touristic patterns to the planning of a smart city, through the interlinked touristic and infrastructure big data sets. Such infrastructure data sets would involve energy consumption, mobility and water consumption attributes, among others. In this context there is great potential in linking touristic and other land use dynamics to the management of a WDN.

The need to reduce leakage in WDNs on one hand and optimize all relevant objectives, such as energy, excessive pressure, etc., on the other hand, has turned the pressure control management method into the state-of-the-art solution for water companies. The idea is keeping pressure in the network to minimum heights demanded spatio-temporarily (Ulanicki et al., 2000). This facilitates leakage reduction, since the latter is directly related to pressure (Germanopoulos and Jowitt, 1989). Energy consumption reduction is also facilitated, since the higher pressure maintained the more energy is demanded. The specifications of the case study, in addition to the constraints in terms of budget, lead to a specialized pressure control scheme (Jowitt and Xu, 1990)

The Pressure Management (PM) is implemented with use of Pressure Reduction Valves (PRVs), which adjust their diameters to the water demand anytime. The number of PRVs, the location and the diameter adjustments are the three variables that need to be optimized, within the optimization process (Araujo et al., 2006). The PM scheme might be single-feed or multi-feed depending on the deviation of the network in DMAs. Large networks with intense spatial variation in altitude and water demand are generally expected to be divided into more DMAs. Each DMA is supposed to retain approximately common characteristics in terms of pressure demand profiles. Single-feed PRV schemes are preferable to more uniform and/or small networks, due to their ease of control and monitoring, while even the risk of a failure event is ` to such less flexible schemes (Abdelmeguid and Ulanicki, 2010). The aforementioned prove that a thorough spatio-temporal analysis of a WDN is a relevant, if not prerequisite, exercise.

The diameter of the PRV is usually an output of a multi-objective optimization process, which minimizes the leakage and energy consumption variables. The problem could be simplified to a single objective optimization, which minimizes the pressure in the network. Leakage—as well as energy—is expressed as a function of the pressure through variable empirical models (Giustolisi et al., 2008). A basic constraint of the optimization process would be the minimum pressure demand of a critical point of the network. This would be the lowest pressure point, either because it is the farthest away from the source, or because it has the highest latitude. It could even be a combination of the two conditions. The critical point, also, might change from time to time, depending on the dynamic consumption profiles variability. Nonetheless, this implies the need to map the pressure demand spatially and temporarily through the network at the highest possible space and time resolution. This can be achieved by mapping the water demand and then using a hydraulic model to calculate the pressure map.

Except for facilitating the application of an optimum PM scheme, spatio-temporal analysis of a WDN would be relevant to the overall process of the performance of a WDN. Active leakage control, Pipe and asset management, Speed and quality of repairs constitute, additional to Pressure Management, the four main tasks of the basic leakage management towards the elimination of the recoverable real losses

(Lambert, 2002). All four tasks are supported or even determined by a well-informed and updated spatio-temporal analysis. Other supplementary tasks that are also crucial for the operation of a network and can be facilitated by the knowledge of the spatial and temporal variability of the hydraulic parameters is the detection of theft across the network and the quantification of metering inaccuracies, both components burdening the Non-Revenue Water proportioning.

The purpose of the following task is to investigate the potential of using daily groundwater pumping data and quarterly consumer billing data in order to produce daily time series of System Input Volume that corresponds to each individual consumer. Additionally, a fraction of the network Non-Revenue Water is assigned to each water meter. This way, a disaggregation of water demand data sets spatially and temporally is performed in order to produce a rough estimation of the distribution of consumption and Non-Revenue Water across the network. The produced spatial and temporal values of the two main components may constitute the initial values of a more thorough analysis that takes into account the hydraulic properties of each component and its subcomponents, for example the fact that leakage is not proportionate to the demand flow, but follows a profile function of pressure. This analysis is implemented in the next subchapter 4.2. The final target would be to create a simulation model of the network at the level of the node of the hydraulic model, if not the individual consumer, and at the temporal resolution of an hour of less. Such a simulation model can constitute the basis for estimating and tracking a number of WDN performance indicators spatio-temporally and facilitate the detailed supervision of the network and its performance by the water utility or any other stakeholder.

4.1.2. Materials and Methods

Skiathos WDN is the study case for the suggested approach. The water distribution network of Skiathos is an aged network with significant levels of Non-Revenue Water, reaching up to 70% levels of the whole SIV during the winter months. The network is currently under reconstruction, works that are quite time-consuming, due to the importance of the maintenance of the traditional street infrastructure, the high touristic activity and other policy related matters. Skiathos has an intense temporal variability in SIV flowrates covering a range of hourly average maximums from 90 m^3/sec in winter months to 180 m^3/sec in summer months, approximately. The water uses throughout the town of Skiathos are not intensely variable, since there do not exist any exclusively residential, industrial or working areas and except for households they consist of small hotels and rooms to let, offices, shops and restaurants. However there is some variability of land uses that is linked to the touristic activity and all the relevant uses such as the touristic accommodation, food services, leisure, etc. The town is hilly and the water distribution system is significantly net-shaped and quite thick in terms of consumers (Laspidou et al., 2015).

The small extent of the network and the relative uniformity of uses have been the reasons why the water utility performs a single-DMA scheme. The whole town comprises approximately 3,500 water meters. Pressure control management is not applied yet, but is planned for a single-feed scheme. A PRV installed downstream the tank that supplies the town with water from a single groundwater drilling. In the future, a booster pump might be added to the pressure control management scheme, cutting off the hilly area of the town into a second DMA, for further localization of the pressure demand constraint.

The available data, for the specific exercise, consist of a data set of daily pumped groundwater from a single drilling filling a single tank; this data set is equal to the total daily SIV. Another data set includes quarterly billed water consumption for each household. The data sets are updated continuously and are currently approximately 5-years long.

For each water meter *k* and for day *t* of the trimester *tri*, the theoretical "local" Input Volume $d_{k,t,tri}$ is calculated, which contains the Billed Consumption (Revenue) and the Non-Revenue Water that theoretically corresponds to each household, taken the rough assumption of linearity:

$$d_{k,t,tri} = w_{k,tri} * D_{t,tri} \qquad \textit{(equation 18)}$$

where $D_{t,tri}$ is the input volume of the whole town for day *d* of the trimester, or the daily pumped water and $w_{k,tri}$ is the specific weight of each water meter for trimester *tri*. This weight is calculated by equation 19.

$$w_{k,tri} = \frac{d_{k,tri}}{\sum_t D_{t,tri}} \qquad \textit{(equation 19)}$$

where $d_{k,tri}$ is the local input volume of water meter *k* for the whole trimester *tri* including the Non-Revenue water that corresponds to that water meter. That is the trimester billing for water meter *k* with the theoretical Non-Revenue Water percentage added to it.

$$d_{k,tri} = d_{kbilled,tri} * (1 + a_{tri}) \qquad \textit{(equation 20)}$$

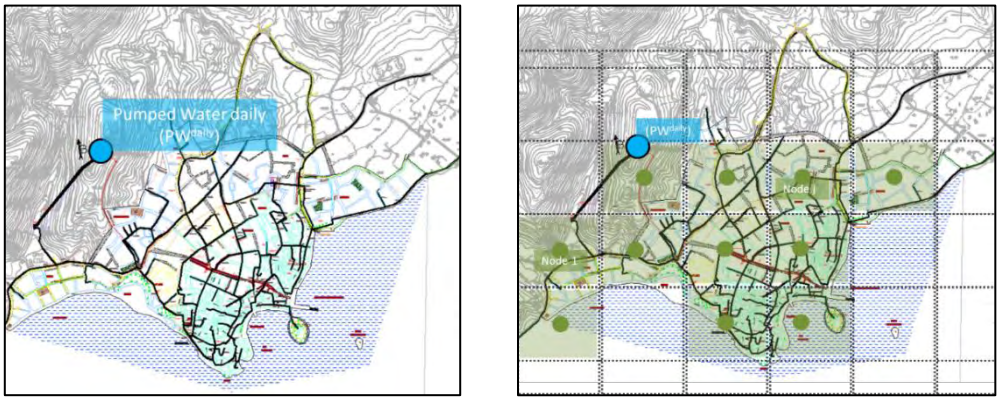
where $d_{kbilled,tri}$ is the trimester *tri* water meter billing, and

$$a_{tri} = \frac{\sum_t D_{t,tri} - \sum_k d_{kbilled,tri}}{\sum_k d_{kbilled,tri}} \qquad \textit{(equation 21)}$$

Daily Non-Revenue Water is calculated with equation 22.

$$Daily\ Non\ Revenue\ Water = D_{t,tri} \frac{\sum_t D_{t,tri} - \sum_k d_{kbilled,tri}}{\sum_t D_{t,tri}} \qquad \textit{(equation 22)}$$

The described methodology is also schematically depicted in Figure 26 and Figure 27.



(a) (b)

Figure 26. (a) Water distribution network of Skiathos Island supplied daily by a single drilling. (b) The network is divided into node areas

(a) (b)

Figure 27. (a) Each node will be a node of the hydraulic model that will be used to estimate the pressure map out of the water demand map. (b) Each node serves a billed consumption that comes out by adding the household billed consumptions that are supplied by the specific node of the network.

4.1.3. Results and Discussion

Implementing the above-described process, the whole SIV can be distributed to theoretical household demands including (or not) the Non-Revenue Water that "corresponds" to the specific households with use of trimester household consumption weights. Of course, it is not expected that this estimation is accurate, since throughout a trimester the consumption profile of a household might change dramatically from one day to the next. However, once the individual consumer daily demands are added at a node level (the required granularity for the creation of time series to be input in EPANET software for the hydraulic solution of the network), the error becomes less significant, since a node will include multiple consumers, even as many as a hundred. The expected divergences of the estimated and the actual node water demands will be revealed by comparing estimated data to monitoring data obtained by sensors; these are located, at a minimum, at critical points in the network, or at several points throughout the network. These divergences will be eliminated through an improved assessment of the SIV components taking into account their specific profiles and later on through the calibration process.

Using the approach described above the temporal resolution of the Billed Consumption have roughly zoomed in from a trimester level into a daily level. Once these Billed consumptions and respective Non-Revenue water are added at a node level, the initial water demand values needed as inputs to the hydraulic simulation will be available allowing for the pressure mapping of the network. Moreover the daily SIV is spatially disaggregated into node level. Respectfully, the Non-Revenue Water is also disaggregated the same rough and proportionate way. It should be noted at this point that the final spatio-temporal assessment of the Non-

Revenue Water, especially the Real Losses Component is expected to follow a rather divert profile than the estimations of the initial values, since the pressure dependent nature of leakage contradicts the proportionate formula used for the production of initial values.

Afterword: Limited data availability is a major obstacle in the management of distribution networks, since it demands a spatio-temporal knowledge of the network. In "low-tech" situations where smart meters are not available at the consumer level and customer billing is done manually with bills being issued once every three months or longer, data scarcity becomes a serious limitation. In this subchapter, a methodology is presented that can be useful in disaggregating bulk town water supply data to the level of individual consumer. The described process can offer a rough map of demands and their shifts throughout the seasons and facilitate some conclusions on the land use dynamics, such as the touristic activity. The methodology, although not accurate, provides a way to deal with the lack of data and produces initial value time series that can be refined later, as a more thorough analysis based on these initial values is implemented.

4.2 A Water Distribution Network Simulation Model with Key Performance Indicators for spatio-temporal analysis and operation of highly-stressed water infrastructure

Foreword: An annual and lumped water balance assessment of a water distribution network is recommended by the International Water Association as a first step and prerequisite for improving the performance of the network by minimizing real/physical water losses, bursts' incidents, water theft, non-revenue water, and energy consumption, among others. The current chapter suggests a modeling approach for developing the water balance of a network spatio-temporarily, in hour time-scale and neighborhood granularity. It exploits already established key performance indicators and introduces some new ones to highlight the potential in improving the management of a water distribution network when having a detailed spatio-temporal supervision, especially when the spatial and temporal conditions are variable. The methodology is applied in a seasonally touristic and hilly case study. Additionally, a pressure management scheme is applied to further exploit the potential of such a toolkit. For the investigated case study, Skiathos town, the annual real losses are estimated equal to 50.9-52.2 % of the System Input Volume, while Apparent losses are estimated to be about 5.6-6.6%. Losses depict intense seasonal variability. Real losses range from 38.8-39.6% in summer months to 63.3-64.7% in winter months, while Apparent losses range from 8.4-9.3% in summer to 1.3-2.5% in winter. Annual water theft is estimated to be at least 3.6% of System Input Volume. Spatial variability, which is linked to the elevation and the different urban land uses is proven to play a significant role in the neighborhoods' water balances and the various Key Performance Indicators that are suggested and applied for the pressure control scheme. The annual potential savings due to the applied scheme rise up to 51,300 m^3 for leakage and 53,730 m^3 for Pressure Driven Demand.Introduction

4.2.1. Introduction

A detailed component analysis of the annual System Input Volume (SIV) has increasingly been implemented for Water Supply Systems (WSSs) around the world with variable precision, depending on the available data, measurements, knowledge of the WDN and available expertise. The analysis typically focusses on completing the iconic IWA Water Balance (WB) table of SIV components, which are defined by a taxonomy that distinguishes between Revenue and Non-Revenue Water, Authorized and Unauthorized Consumption, Real and Apparent Losses, Metering Inaccuracies, Billed and Unbilled consumption, Metered and Unmetered Consumption, and Leakage in the transmission/distribution mains, service connections and tank overflows. The task has usually been implemented in an aggregated way for the whole WDN and for annual time step, although there have

also been few approaches that focus on the estimation of leakage, rather than on the construction of the whole, or part of the IWA table in a spatio-temporal manner. The difficulty in constructing a precise and trustworthy WB table lies in the difficulty of assessing the real, apparent losses, and unmetered water use.

Significant effort has been made towards suggesting a reliable methodological approach for estimating, and/or modelling leakage (physical losses in mains and service connections), metering inaccuracies and unauthorized consumption (water theft). Almadoz et al. (2005) proceeded to a real losses assessment by perceiving the apparent losses as a water-consumption-pattern-dependent component and used four different demand profiles: domestic, industrial, commercial and official. Cabrera & Pellejero (2003) and Ismail & Puad (2006) conducted bottom-up assessments of physical losses and highlighted that estimating leakage through night flow can minimize the error, while night consumption is minimum and easier to determine. Giustolisi et al. (2008) proposed an algorithm for the simultaneous simulation of Pressure-Driven Demand (PDD) and leakage by fully integrating the two demands in pipe level. Tabesh et al. introduced a methodology for distinguishing the pressure-dependent components from the independent ones.

Puust et al. (2010) distinguished the literature approaches of leakage assessment into top-down approaches where leakage is estimated through WB assessment, and the bottom-up approaches, where the leakage is estimated through the summing of its components. Kanakoudis and Tsitsifli, (2010) shared a thorough aggregate and annual component analysis for a series of case studies in WDNs in Greece following the established IWA methodology. They concluded that such an analysis is a useful tool for the local water utility, while a simple definition of SIV and billed water can be misleading for the performance of the network. Regarding the assessment of metering inaccuracies, they suggested laboratory investigation, while they distinguished under-registering and over-registering water-meters according to their age. Additionally, the practice of Pressure Management (PM) and division of the network in District Metered Areas (DMAs) occurred as a necessary approach for reducing real losses. Kanakoudis and Tsitsifli (2010b) implemented the IWA WB assessment in semiannual time scale, suggesting that a finer scale than the annual can be revealing for cases that depict seasonal demand peaks, while a few years later they proceeded to an even more detailed, bimonthly WB analysis (Kanakoudis and Tsitsifli, 2014). Cobacho et al. (2015) introduced a methodology for spatio-temporal simulation of leakage, using the EPANET software and the concept of simulating leakage through weighted leakage emitters. Sophocleous et al. (2019) introduced a leakage localization methodology based on head pressure and flow measurements tested in two case studies in the U.K. thus reducing the repair time for leakage; according to the article, this could lead to water savings as high as 70% of total loss.

More macroscopic investigation has also been implemented in an attempt to address the need of integrated water management linking the urban water supply to other urban water systems, such as wastewater and stormwater. Makropoulos et al. (2008) and Rozos and Makropoulos (2013) extended the boundaries of the typical investigation focus on the urban water cycle, involving wastewater, urban stormwater, potable water, grey water, and green water. They also incorporated indicators from all sustainability pillars: environmental, economic, social and technical, such as operational costs and energy. In their review on urban hydroinformatics, Makropoulos and Savíc, (2019) presented a scheme according to which the modeling is facilitated by the increasingly fluent flows of data and information in the context the developments in Information and Communication Technologies, cloud based information platforms, and remote monitoring. Such a technological landscape in combination to well-advanced methodological frameworks could lead to the next step in water distribution networks modeling, what has been referred as Water Distribution Network Digital Twins (Sun et al., 2020).

The importance of assessing the IWA standard WB table in combination with employing a PM scheme and the need to enhance leakage localization and detection of unauthorized consumption raise the need for a methodology that would facilitate the spatio-temporal supervision and would provide the WDN operator with insight on the various components of SIV. The temporal analysis is essential, especially in cases of intense seasonality in the water demand, such as in the highly touristic regions. Seasonality also becomes relevant in the spatial distribution of land uses within the urban landscape, which in turn causes an intense seasonal shift in the demand map.

In this chapter, a toolkit for the spatiotemporal analysis and simulation of SIV, SIV components and critical Key Performance Indicators (KPIs) throughout the WDN is presented that can serve as supervision support for the network and its properties, stresses and potential of improvement, or the basis for a Digital Twin. The well-established IWA WB table is assessed in neighborhood granularity and hourly time step. The WB components assessed include System Input Volume, Billed Consumption, Non-Revenue Water, Unauthorized Consumption, Real Losses, and Metering Inaccuracies. All the components are identified locally and instantly offering a useful tool to a water utility for creating dynamic hotspot maps that can contribute to the optimization of the WDN management through: i) the localization of leakage and water theft in neighborhood resolution and the identification of drivers of leakage; ii) the assessment of metering inaccuracies; iii) the quantification of tourism impact on water demand; and iv) the application of a PM scheme and the quantification of its beneficial effect. The assessment of the WB table is implemented by distinguishing the components in pressure-dependent and time-pattern-dependent and an application of a nested system of two loops, an inner and an outer, which through iterative runs, close when the simulated pressure best fits the actual network pressure. Spatio-temporal critical KPIs are introduced to facilitate a detailed supervision of the WDN. The KPIs include the energy

consumption in the water components (water-energy nexus), a pressure-driven demand indicator, and various expressions of water components such as "per-connection", "per-customer," and "per-network-length" indicators. The aforementioned KPIs are also used for the assessment of the PM scheme performance. The localization of the water components is used to conduct rough conclusions regarding the urban land uses, specifically the residential and the touristic ones..

4.2.2. Materials and Methods

- Case study specifications

Skiathos is a small island municipality in Thessaly, Greece. The island has approximately 6,000 registered inhabitants according to the 2011 census (Hellenic Statistical Authority, 2011), while the water utility counts approximately 3,500 consumers (Kofinas et al., 2016). The island population shows intense seasonal variability due to high touristic influx. The touristic season, from April to September, depicts a high peak in August, which often exceeds 90,000 tourists for the whole month with an annual average ascending trend (Kofinas et al., 2016). As expected, this is reflected in the water demand, which depicts seasonal oscillations in SIV, billings, non-revenue and non-revenue proportioning. Indicatively, the August peak summer water withdrawals often exceed a 130% increase compared to the winter minimum withdrawals, which shows a link between the touristic activity and the water demand profiles. The Water Distribution System (WDS) includes a single drilling a tank and the WDN, which is currently under renovation.

This research applies on the WDN of Skiathos before its renovation works had started, from 2011 to 2016. During these years the WDN is characterized by an extreme high rate of NRW gradually climbing up from 40%, during last trimester of 2011, to 70 % of SIV, during the last trimesters of 2015 and 2016 (Figure 28). The SIV values in Figure 28 from January 2011 to December 2016 are trimester sums of daily SIV readings. The SIV has an accuracy range of ±1% and the NRW has an accuracy range from ±1.5 to ±3.2% (95% confidence limits) depending on the billing period (Table 13). The WDN operates through a single-DMA scheme. The town is characterized by bold relief, with two hills. No PM scheme is applied at the WDN except for a switch of pressure every 6 months from touristic to non-touristic period.

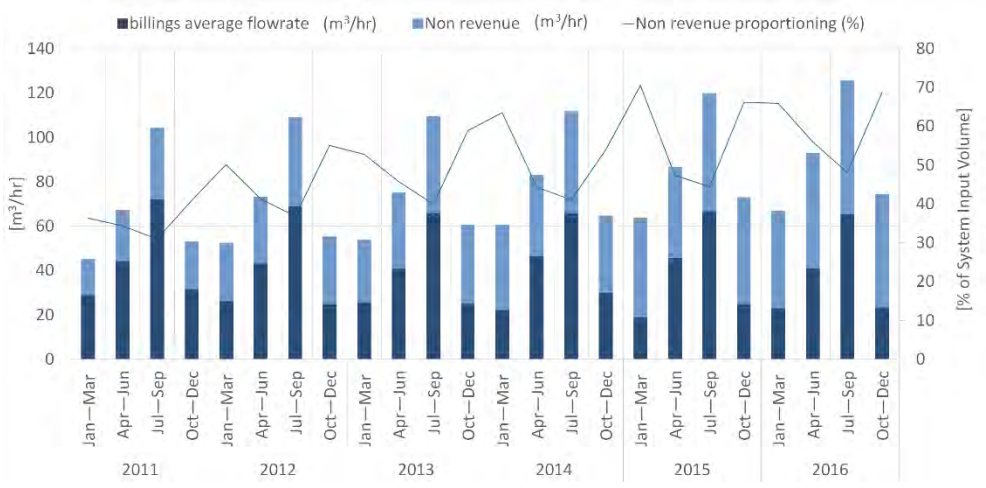


Figure 28. Skiathos WSS per 3-months SIV, revenue, NRW supply and fraction of SIV corresponding to NRW from January 2011 to December 2016

Table 13. Application of 95% Confidence limits to the WB and estimation of the accuracy ranges of NRW

trimester	SIV (m³/hr)	SIV standard deviation for ± 1% accuracy range	SIV variance	BAC variance	NRW = BAC variance + SIV variance	NRW standard deviation	NRW accuracy range
2011 1st	45.2	0.23	0.05	0	0.05	0.23	± 2.8
2011 2nd	67.1	0.34	0.12	0	0.12	0.34	± 2.9
2011 3rd	104.3	0.53	0.28	0	0.28	0.53	± 3.2
2011 4th	53.2	0.27	0.07	0	0.07	0.27	± 2.4
2012 1st	52.4	0.27	0.07	0	0.07	0.27	± 2.0
2012 2nd	73.2	0.37	0.14	0	0.14	0.37	± 2.4
2012 3rd	109.0	0.56	0.31	0	0.31	0.56	± 2.7
2012 4th	55.4	0.28	0.08	0	0.08	0.28	± 1.8
2013 1st	54.0	0.28	0.08	0	0.08	0.28	± 1.9
2013 2nd	75.3	0.38	0.15	0	0.15	0.38	± 2.2
2013 3rd	109.5	0.56	0.31	0	0.31	0.56	± 2.5
2013 4th	60.7	0.31	0.10	0	0.10	0.31	± 1.7
2014 1st	60.7	0.31	0.10	0	0.10	0.31	± 1.6
2014 2nd	83.0	0.42	0.18	0	0.18	0.42	± 2.3
2014 3rd	111.6	0.57	0.32	0	0.32	0.57	± 2.4
2014 4th	64.6	0.33	0.11	0	0.11	0.33	± 1.9
2015 1st	63.7	0.33	0.11	0	0.11	0.33	± 1.4
2015 2nd	86.5	0.44	0.19	0	0.19	0.44	± 2.1
2015 3rd	119.7	0.61	0.37	0	0.37	0.61	± 2.3
2015 4th	72.9	0.37	0.14	0	0.14	0.37	± 1.5
2016 1st	67.0	0.34	0.12	0	0.12	0.34	± 1.5
2016 2nd	92.9	0.47	0.22	0	0.22	0.47	± 1.8
2016 3rd	125.7	0.64	0.41	0	0.41	0.64	± 2.1
2016 4th	74.5	0.38	0.14	0	0.14	0.38	± 1.5

The deployment of a more efficient WDN in Skiathos has become an urgent need due to several reasons. The Island has been facing seasonal water scarcity, due to the increased water demand driven by tourism. At least two drying-out incidents occurred over the last five years, in August, due to the extreme drop of the water table of the aquifer level at the drilling location. Due to these incidents, the water supply was interrupted in the island, until the aquifer recovered. Naturally, such incidents have multiple social, financial, and environmental impacts, as well as sanitation, since water lies in the core of most everyday activities, especially in a touristic destination. On top of that, the overall aquifer level drop results in seawater intrusion and groundwater salinization, with detrimental consequences, such as the natural mineral mercury release into groundwater. Mercury concentration has repeatedly been recorded to exceed WHO limits, regarding toxicity, even by six times (Spyropoulou et al., 2018). The water utility of Skiathos (DEYASK) has declared the water in the island as non-potable. Overall, reducing water losses, optimizing water source withdrawals, investigating the potential of alternative water sources and applying innovative treatment methods are some of the tasks that are of utmost importance in Skiathos.

- *Infrastructure and data*

The following infrastructure installed in Skiathos is used for retrieving data needed for the simulation of the dynamic characteristics of the WDS: Flowrates are being monitored at the inflow point of the WDN, while pressure is being monitored at three points of the WDN. The daily SIV values from January 2011 to December 2016 are readings of a mechanical meter with accuracy of ±1% for each value. The values of SIV are also checked often by the water utility volumetrically by cross-checking with the tank's volume. The values of 2016 in particular are additionally compared to the readings of an electromagnetic flow meter with accuracy of ±0.25%. For year 2016, there is a Pressure Reducing Valve (PRV), a CSA XLC 410, DN150 PN16 in particular, that records pressure and flowrate (electromagnetic flowmeter) in the inflow of the WDN, after the tank, with 15 min time step—currently not regulating pressure automatically—and three cello sensors that record pressure at three points around the network with 15 min time step. The pressure data are used for the calibration of the model, as the fitting of the simulated pressure values to the recorded ones is used as the criterion for the reliability of the simulation. Billing data were provided by DEYASK. There are 3,500 water meters in the city used for billing purposes, with a billing period of three months, from January to March, April to June, July to September, and October to December, or 1st, 2nd, 3rd and 4th period, respectively. The procedure of manually recording the water meters lasts for approximately ten days, while this procedure starts before and finishes after the end of the billing period, which means that there is no systematic hysteresis between the records and the consumption. No fixed minimum consumption is imposed to the billings, so the billings represent what is metered and billed. All water-meters are geo-located with a hand-held GPS device of 4.9-m accuracy, during multiple field trips with the guidance of the utility's staff and matched to

corresponding billing data. In order to keep the privacy of water consumers, when these data are presented in public access, the scale is zoomed out to "neighborhood" level and the billed consumptions are summed.

- *Simulation of the WDN with EPANET*

The first step of the methodology includes the simulation of the static and dynamic characteristics of the WDN, through EPANET software (Rossman, 2000). For this purpose, historical maps of the WDN are used to incorporate network geometry and pipe diameter, length and material. The overview map in Figure 29 presents main elements of the WDS. The single water supply drilling is found in the north-western part of the city. Groundwater is pumped from the drilling and pumped up to the water tank located on the hill (Figure 29). From the tank water is distributed by gravity to the city area. The main pipeline of the WDN goes back down the hill to the SE direction and then to the E-NE around the city centre (Figure 29). The pipe diameters are presented in Figure 29 in three clusters and the pipe materials, amiant, metal (cast-iron) and PVC, are presented in Figure 30.

In order to fill gaps in the pipe diameter (and material) data, unknown pipe properties are imputed basing on properties of connected pipes. Specifically, if the pipe with unknown diameter (material) is between two pipes with known diameters (materials), the missing value is assumed to be the same as the bigger known diameter, or as the most recently-used material, while if it is connected to one pipe with known diameter (material) it is assumed that both diameters (materials) are the same, except for the situation it is downstream to the main pipe and all other pipes downstream have smaller diameter (other material) than the main pipe. This algorithm is empirical and it is set after discussion with the water utility director and employees that have relevant experience in the specific network. The head loss is estimated based on the empirical parameters described in and Table 16. Roughness coefficients assigned to the three materials of Skiathos WDN

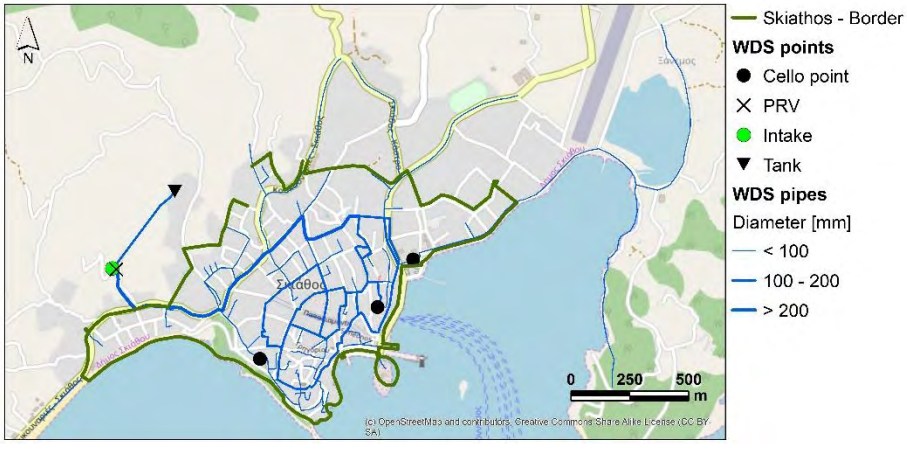


Figure 29. Skiathos WDN

Figure 30. WDN map of pipe materials

Table 14. Popular formulas for the estimation of head loss throughout a pipeline.

Formula	Resistance Coefficient	Flow Exponent
Hazen-Williams	$4.727\ C^{-1.852}\ d^{-4.871}\ L$	1.852
Darcy-Weisbach	$0.0252\ f(e,d,q)d^{-5}L$	2
Chezy-Manning	$4.66\ n^2\ d^{-5.33}\ L$	2

Where:

C = Hazen-Williams roughness coefficient

e = Darcy-Weisbach roughness coefficient (ft)

f = friction factor (dependent on e, d, and q)

n = Manning roughness coefficient

d = pipe diameter (ft)

L = pipe length (ft)

q = flow rate (cfs)

Table 15. Head loss formulas parameter values for different materials

Material	Hazen-Williams C	Darcy-Weisbach e		Manning's n
Unit	(-)	(feet x 10-3)	(mm)	(-)
Cast Iron	130 – 140	0.85	0.2591	0.012 - 0.015
Concrete or Concrete Lined	120 – 140	1.0 - 10	0.3048-3.048	0.012 - 0.017
Galvanized Iron	120	0.5	0.1524	0.015 - 0.017
Plastic	140 – 150	0.005	0.0015	0.011 - 0.015
Steel	140 – 150	0.15	0.0457	0.015 - 0.017
Vitrified Clay	110			0.013 - 0.015
PVC (Bishop, 1978)	150	0.007	0,0021	0.009

Table 16. Roughness coefficients assigned to the three materials of Skiathos WDN

	Hazen-Williams C	Manning's n
PVC	140 (optionally 150)	0.011 (optionally 0.009)
Metal	130-140 if new (60 corrugated)	0.012 (0.022 corrugated)
Amiant	140	0.011

(Asbestos cement)

- Water consumption, leakage and other SIV components spatio-temporal distribution simulation

The SIV components (Figure 31), as defined and established by IWA Blue Pages and Manuals of Best Practice (Hirner and Lambert, 2000; Alegre et al., 2000, respectively), are simulated in EPANET either as a time-pattern dependent demand or as a pressure-dependent demand, according to the distinction made by Cobacho et al. (2015). They decoupled the leakage from the base demand by introducing nodal emitters—open valves to the atmosphere—to better mimic the pressure-dependent behavior of leakage, as it is described by Germanopoulos (1985) and Germanopoulos & Jowitt (1989)

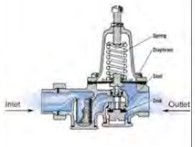
The components that are simulated as demand driven are the Billed Authorized Consumption (BAC) and the Apparent Losses, while Real Losses are simulated as pressure driven (Figure 31). This does not mean that the components which are simulated as demand driven are not also partially pressure dependent. For the Skiathos case study(Figure 32), BAC consists only of metered BACs. Unbilled Authorized Consumption (UAC) is considered as negligible. The main demand for UAC would be that of the fire department for fire-fighting; however during the simulation period, no significant fire occurred in the island. Apparent Losses, which consist of Unauthorized Consumption and Customer Metering Inaccuracies, along with reading and handling errors, cannot be perceived as negligible, since water theft in Skiathos has been repeatedly noticed and the recording of water-meters—that are older than 20 years—is implemented manually. These last two components are uniformly distributed to the WDN model following the time-pattern of the demand simulated as demand driven. Regarding Real Losses, "Leakage on Transmission and/or Distribution Mains" and "Leakage on Service Connections up to the point of Customer metering" are expected to constitute the largest part of NRW, since the WDN during the simulation period is characterized as a rather aging network in an overall bad condition—the construction of the WDN was initiated at the mid-60s and evolved in the following decades—. These two components are simulated as pressure driven flows. Leakage and overflows at utility's storage tanks and operational losses are considered negligible and occurs before the SIV metering in the specific case study. There is no adequate information on leakage after the point of customer metering, which are not considered to be negligible on the one hand and are pressure-driven on the other. Due to this lack of information, this hidden component in BAC is not decoupled by the demand driven components. However, the private property and infrastructure in Skiathos is constantly renovated to meet touristic demands; thus, leakage after the water meter is expected to be much less than the corresponding one before the water meter.

Another specification in Skiathos case study is the fact that a lot of houses and small enterprises owe small tanks that are set at the roofs. These tanks are not used in a regular basis, but only if there is a crisis i.e. when the utility is disrupting the supply. There are only few houses (less than twenty) at the top of the two hills that use the tanks in a more regular basis, when due to very high demand, during the summer peak, the pressure is very low. The aforementioned, regarding the tanks, are not considered to significantly influence the overall consumption pattern, neither the water metering in a three-months period of any customer.

<table>
<tr><td rowspan="7">System Input Volume (SIV)

(corrected for known errors)</td><td rowspan="3">Authorized Consumption</td><td rowspan="2">Billed Authorized Consumption (BAC)</td><td>Billed Metered Consumption</td><td rowspan="2">Revenue Water</td></tr>
<tr><td>Billed Unmetered</td></tr>
<tr><td rowspan="2">Unbilled Authorized Consumption (BAC)</td><td>Unbilled Metered Consumption</td><td rowspan="5">Non-Revenue Water</td></tr>
<tr><td>Unbilled Unmetered Consumption</td></tr>
<tr><td rowspan="4">Water Losses</td><td rowspan="2">Apparent Losses</td><td>Unauthorized Consumption (theft)</td></tr>
<tr><td>Customer Metering Inaccuracies</td></tr>
<tr><td rowspan="3">Real Losses</td><td>Leakage on Transmission and/or Distribution Mains</td></tr>
<tr><td>Leakage and Overflows at Utility's Storage Tanks</td></tr>
<tr><td>Leakage on Service Connections up to Customer Metering</td></tr>
</table>

Figure 31. The IWA standard SIV component analysis

<table>
<tr><td rowspan="5">System Input Volume (SIV) after Utility's Storage Tank (corrected for known errors)
  </td><td rowspan="2">Billed Authorized Consumption (BAC) Including leakage after Customer Metering

(every 3months)</td><td>Metered BAC</td><td rowspan="2">Revenue Water</td><td rowspan="2">Demand Driven Simulation</td></tr>
<tr><td>Unmetered BAC</td></tr>
<tr><td rowspan="2">Apparent Losses</td><td>Water Theft</td><td rowspan="3">Non-Revenue Water</td></tr>
<tr><td>Customer Metering Inaccuracies (manual recording)</td></tr>
<tr><td rowspan="2">Real Losses</td><td>Leakage on Transmission and Distribution Mains</td><td rowspan="2">Pressure Driven Simulation</td></tr>
<tr><td>Leakage on Service Connections up to Customer Metering</td></tr>
</table>

Figure 32. SIV components as specified for Skiathos Case Study and the respective simulation approaches.

Time-pattern spatio-temporal demands

The spatial distribution of BAC is implemented by linking the billing time-series of each water customer with the location of the respective water meter in the Skiathos WDN. Firstly, each customer code that appears in the billings is assigned to the corresponding water meter code, after the anonymization of the billings. Each water meter code is spatio-located with use of a GPS device. All locations are later on imprinted in AutoCad with the water meter and customer codes assigned to them as attributes. Billed water that is not assigned to a water meter (due to customers moving, or changing services) is introduced to the algorithm as "bulk" demand, which means demand that is not spatio-located but is distributed around the spatio-located water-meters in a weighted proportioning, according to the water-meters' assigned consumptions. According to this distribution of the "bulk" demand, the water meters that showed higher consumptions will be assigned bigger parts of the "bulk" demand and water meters with lower consumptions will be assigned smaller parts of the "bulk" demand in an analogous manner.

On the next step, the 3,500 water-meters are grouped into 121 neighborhoods, with an average of 26.6 water-meters/neighborhood. The rest of the billed water-meters are treated as "bulk" demand. The neighborhoods are designed, in a way that clusters service connections that are close to each other. These WDN neighborhoods are referred to as landzones in this article. Each landzone is assigned to a specific demand geonode, a node on the WDN that is considered to supply all the water-meters of the landzone. The geonodes are also simulated in EPANET as demand nodes. The output of these steps is the spatial distribution of three-month-step BAC time-series.

In order to temporarily disaggregate BAC maps from a three-month time-step to its daily values, the SIV daily time-series patterns are followed with the assumption that each landzone follows the same SIV pattern. In Figure 33, the intense seasonal variability can be seen as well as an ascending annually average SIV trend, also apparent in Figure 28. This procedure is described in detail in Kofinas et al. (2015), where the three-month BAC values were simply disaggregated to daily, following the SIV pattern and the aggregate SIV values were disaggregated to the landzones.

The daily BAC is not perceived to follow a uniform distribution within the day, but it is expected to follow hourly patterns that are formed by averages of PRV SIV recordings, after they are modified, by subtracting the leakage, which is estimated using the bottom up approach of minimum night flows combined to an iterative EPANET based process for the production of average WDN profiles. The description of leakage estimation is described in more detail in Section 2.5. PRV recordings are taken every 15 minutes, however it is decided that hourly time patterns are sufficiently detailed. The seasonality of the island touristic activity alters the daily SIV profiles, since rush hour times and the ratio of day- over night-consumption are

driven and co-shaped by touristic water demand. In Figure 34, , the twelve hourly SIV profiles are shown, as estimated by averaging PRV recordings, one for each month of 2016. It is apparent that for April to June and July to September billing periods, the relative night to day consumption is lower that the respective January to March and October to December. This does not imply that absolute summer night consumptions are lower than winter; on the contrary, they are much higher as shown in Figure 34.

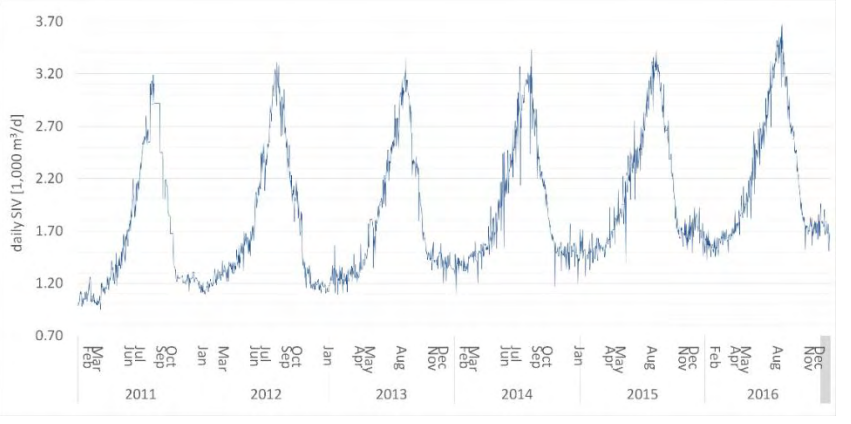


Figure 33. Daily SIV following an intense seasonal pattern with summer peaks and winter lows

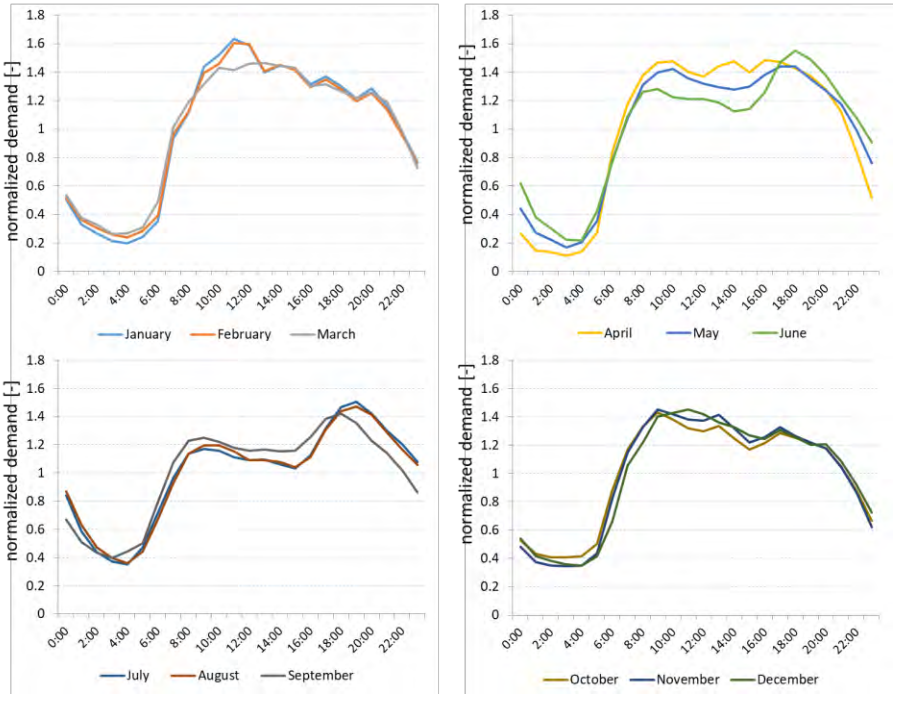


Figure 34. Normalized hourly demand profiles for every month after the leakage is estimated and subtracted by the SIV

Pressure-dependent spatio-temporal demands

Leakage is spatially distributed as pressure-dependent demand, through the introduction of emitters in every node. This means that every node is assigned with two demands, a time-pattern demand as described in the paragraph: *Time-pattern spatio-temporal demands* and a pressure driven demand (Figure 35). The procedure followed by Cobacho et al. (2015) is adopted. The emitters are described by a leakage coefficient or emitter coefficient K_j. The flowrate of the leakage demand is given by equation 23.

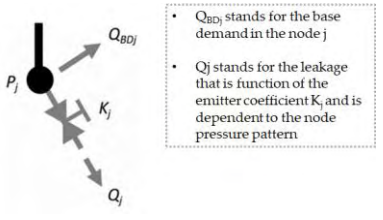


Figure 35. Scheme of the two types of node demands, where Q_{BDj} stands for the base demand in the node j and Qj stands for the leakage that is function of the emitter coefficient K_j and is dependent to the node pressure pattern (Cobacho et al., 2015)

$$Q_j = K_j * P_J^N \qquad \textit{(equation 23)}$$

where *Kj* is the emitter coefficient at node *j*, as defined by equation 24, *Pj* is the pressure at node *j*, and *N* is the pressure exponent, as defined in Table 17.

Table 17. Estimation of pressure exponent (N) based on partitioning of different materials' N according to their lengths

material	length [m]	length proportion	pressure exponent			
			from	to	average	N partitioning
PVC	12,266	0.64	0.40	1.85	1.13	0.72
metal	1,054	0.06	0.52	2.30	1.41	0.08
amiant	5,834	0.3	0.78	1.04	0.91	0.28
total	19,154					N = 1.08

$$K_j = K_{WDN} * \Gamma_j \qquad \text{(equation 24)}$$

where K_{WDN} is the WDN leakage coefficient, as defined by equation 25

and Γ_j is the normalized leak variable for node j, as defined by equation 26

$$K_{WDN} = \frac{Q_{WDN,\ real}}{\overline{P_{WDN}}^{N}} \qquad \textit{(equation 25)}$$

where $Q_{WDN,\ real}$ is the leakage of the whole WDN, which is estimated as described in the paragraph: *Pressure Control Management Scheme and key indicators for the performance of the WDN* and $\overline{P_{WDN}}$ is the average WDN pressure, which is defined as described in the same paragraph.

$$\Gamma_j = average\ (\Gamma_{j,\ service\ connections}, \Gamma_{j,\ pipes\ length}) \qquad \textit{(equation 26)}$$

where $\Gamma_{j,\ service\ connections}$ is the normalized leak variable of node j due to the variability of the number of service connections assigned to each node, through the respective landzone (equation 27) and $\Gamma_{j,\ pipes\ length}$ is the normalized leak variable of node j due to the variability of the total pipes' length assigned to its node (equation 28).

$$\Gamma_{j,service\ connections} = \frac{N_{j,\ service\ connections}}{N_{WDN,\ service\ connections}} \qquad \textit{(equation 27)}$$

where $N_{j,\ service\ connections}$ is the number of service connections assigned to node j and $N_{WDN,\ service\ connections}$ is the total number of service connections of the WDN

$$\Gamma_{j,pipes\ length} = \frac{L_j}{L_{WDN}} \qquad \textit{(equation 28)}$$

where L_j is the length of the pipes that are connected to node j, taken that if a pipe is connected to two nodes, its length is equally distributed to the two nodes and L_{WDN} is the total length of the WDN pipelines.

❖ Leakage estimation and SIV component analysis

Through the sequence of equations 23 through 28, the leakage is distributed to nodes according to the nodes' relative importance regarding leakage, as it is expressed by service connections, pipe length partitioning and node pressure. The $Q_{WDN,real}$ initial value is assumed to be equal to the NRW, which includes Real Losses and Apparent Losses. This is an initial value hypothesis for a first estimation of the WDN average pressure through EPANET. Once Real and Apparent Losses are better approximated, following the night flow methodology (described in more detail in the following section: *Estimation of leakage using the night-flow approach*), the Apparent Losses component is transferred to the time-pattern demands and added to the distributed BAC (Figure 36). This is achieved through two nested loops: an "inner" loop that iterates until convergence, at which time it triggers the "outer" loop that iterates and converges. Figure 37 presents the inner loop.

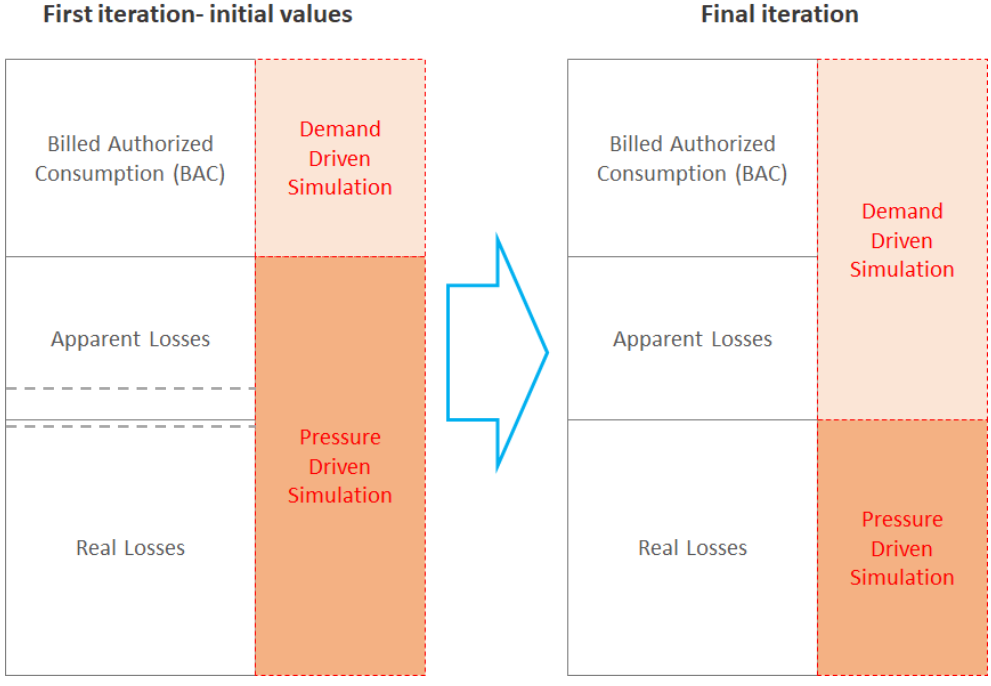


Figure 36. Schematic depiction of the purpose of the iterative process, the transfer of the apparent losses component from the pressure-driven simulated components to the demand driven simulated components. The dashed lines between apparent and real losses indicate that the two amounts are not defined in the first iteration separately, but only as a whole

EPANET initially produces the QWDN,real amount, which is initially equal to the whole nonrevenue component and, as it converges, eventually to the real losses component, following the inner loop shown in Figure 37. This loop ends, when simulated and input leakage converge with an acceptable error (ε) of 0.001. The input leakage is estimated through the bottom up approach of the minimum night

flow for each iteration. The average WDN pressure for the initial step can be a random value (preferably a value that is lower than the PRV inlet set pressure).

For decoupling real losses from apparent losses, the night flow leakage estimation approach is implemented, through the outer loop. This estimation assumes that night flow consists of leakage and night consumption. For this purpose, 2019 average SIV quarterly flowrate profiles are constructed for every month of the year, while the minimum value of flowrate recorded for every 15 minutes of the day throughout each month is also recorded. The minimum average flowrate of all months' night flowrates gives the leakage flowrate for the specific WDN pressure, after the estimated night consumption, 5.04 m^3/hr (see equation 30) is subtracted. Leakage profiles for every month are constructed based on WDN pressure profiles (an output of the inner loop). The new, decoupled from apparent losses, leakage is then input to the inner loop, while apparent losses are added to the time-pattern dependent demands, to give new WDN pressure profiles. The process iterates until the average actual recorded pressure values at the three points (variable in location and altitude) converge to the simulated ones. This closes both nested loops. Figure 38 presents the whole outer loop and its interactions with the inner loop. It is evident that increasing the number of pressure meters at variable points of the WDN improves the accuracy of the method.

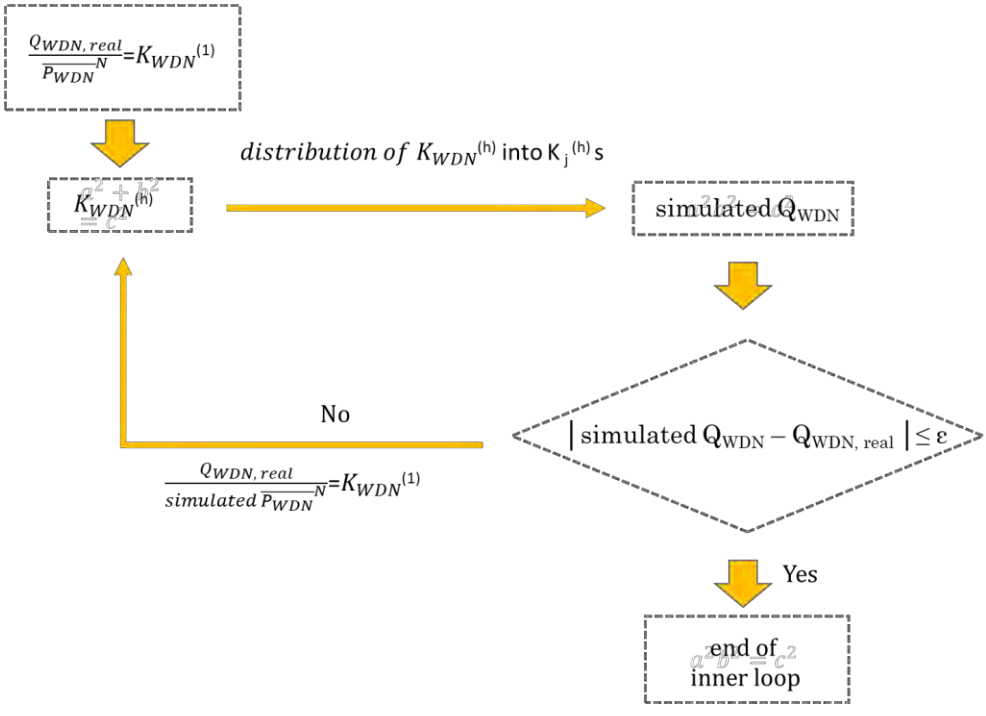


Figure 37. Scheme for the inner loop of the methodology as adapted from the corresponding Cobacho, et al. (2015) loop)

Daily pressure patterns

The leakage function is inextricably linked to the pressure dynamics. Estimating leakage through the night flow method implies that we know the pressure level at the time when the minimum night flow (MNF) occurs, while creating the daily leakage profile based on the MNF implies that we know the corresponding pressure pattern. With this in mind, twelve daily pressure profiles are created, one for each month. Skiathos intense touristic activity results in a rather intense shift of the residents' consumption patterns, since the composition of the consumer body changes seasonally from winter (permanent) residents to summer tourists. Except for the inflow of tourists, the change in weather causes changes in water consumption routines. The above cause in turn a seasonal change in the average daily water consumption profiles and the average daily pressure profiles. For the construction of the daily pressure profiles, the following steps are implemented (Figure 38):

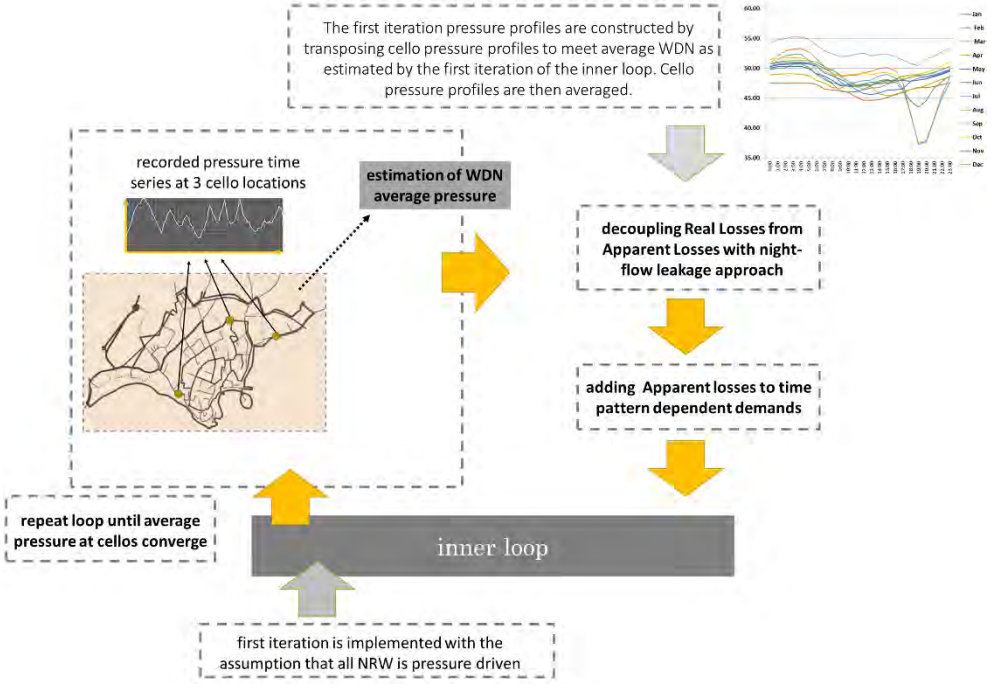


Figure 38. Outer loop for decoupling the Real Losses from Apparent Losses based on the night-flows leakage estimation methodology

a) The recorded pressure time series by the three cellos were processed; outliers are removed and replaced with averaged values of the adjacent time slot records, and the quarterly time step is transformed to hourly time step by averaging the four records of every hour. The recorded data use refers to a time period from February 2015 to March 2017.

b) From the above profiles, average daily pressure profiles are constructed for each one of the three locations, by averaging horizontally the records that correspond to the same hour of the day for the whole month. Days of the same

month but different year (e.g., all July days, independent of the year) are used for the construction of the respective month profile.

c) The average WDN pressure value of the monthly period is estimated through an EPANET initial run. This initial run holds the assumption of zero leakage, so the WDN pressure is expected to be overestimated. This is corrected through an iterative process, after the first estimation of leakage.

d) The cello pressure profiles are moved laterally in order to meet the average WDN pressure as estimated in the previous step.

e) The three produced WDN pressure profiles are averaged, so that a single WDN daily pressure profile is created for every month.

Estimation of leakage using the night-flow approach

For the estimation of leakage, the night-flow approach is implemented; this would imply the use of the MNF of SIV recorded to estimate night leakage through the subtraction of an estimate of the night consumption.

MNF = Real losses + Minimum Night Use (Liemberger and Farley, 2004),

where MNU= household use + commercial use + special use + after water meter burst (Fantozzi and Lambert, 2012)

For this estimation, the following steps are implemented:

a) The quarterly flowrate PRV time series are processed, outliers removed and missing values are imputed. The available time series refer to a period from January to December 2016.

b) Average flowrates for each 15 minutes are estimated separately for each month.

c) Minimum recorded flowrates for each quarter of the day are identified separately for each month.

d) At the average flowrates diagrams, the lowest estimated value is identified, as well as the timeslot it occurred. This value is defined as the MNF.

e) The pressure at that timeslot for the WDN is estimated with an EPANET run. The assumption of zero leakage is made for this first EPANET run, but it is corrected through an iterative process after the first estimated leakage values are input in the EPANET runs.

f) The night consumption is estimated according to the following empirical model (equations 29 and 30).

$$MNU = houselold\ use + commercial\ use + after\ water\ meter\ burst + special\ use$$

(equation 29)

$household\ night\ cons.$

$$= typical\ night\ use * number\ of\ houdeholds \qquad \textit{(equation 30)}$$

$$= \frac{0.0018 \frac{m^3}{hr}}{household} * 2800\ households = 5.04\ m^3/hr$$

Taken that during winter time, commercial night consumption is negligible in Skiathos and there is no particular special use as claimed by DEYASk, equations 29 and 30 transform to equation 31.

$$night\ consumption\ + after\ water\ meter\ bursts = 5.04\ m^3/hr \qquad \textit{(equation 31)}$$

g) The leakage is estimated as the subtraction of MNF-night consumption.

$$Real\ Losses = MNF - MNU = 49.6 \frac{m^3}{hr} - 5.04\ \frac{m^3}{hr} = 44.52\ \frac{m^3}{hr} \qquad \textit{(equation 32)}$$

where 49.6 m³/hr is the minimum of averaged recorded flowrates and it occurs in January 4.15 am time slot

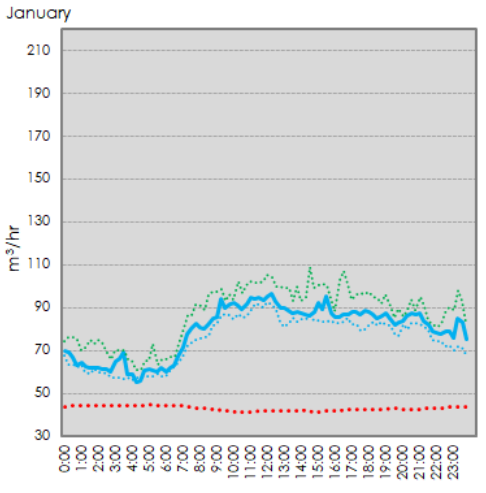
h) Equation 33 is used to construct the leakage daily profiles (as suggested in (Liemberger and Farley, 2004)) according to the pressure profiles produced as described in 2.5.2, the night leakage and the respective night leakage pressure value. N_1 is estimated equal to 1.08

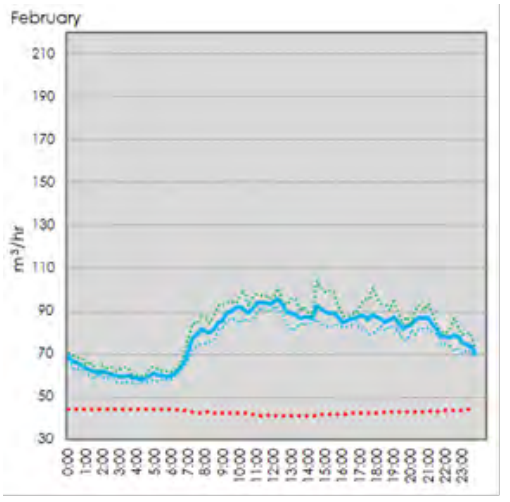
$$\frac{L_1}{L_2} = \left(\frac{P_1}{P_2}\right)^{N_1} \qquad \textit{(equation 33)}$$

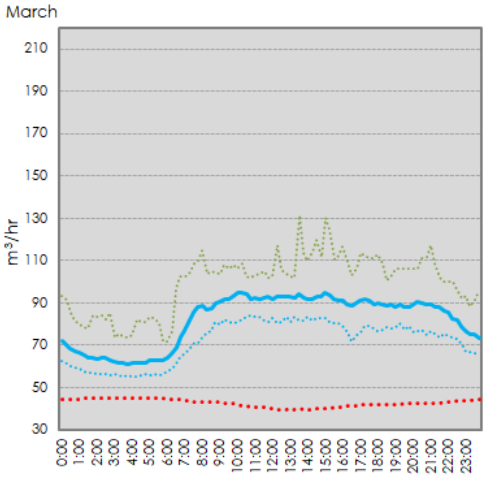
i) For the twelve produced leakage profiles (Figure 39), the leakage estimated is compared to the minimum recorded night flow. The estimated amount should be less or at most equal, assuming that there might be a quarter of night hours, especially in winter time that for such a small village the consumption is negligible, if not zero.

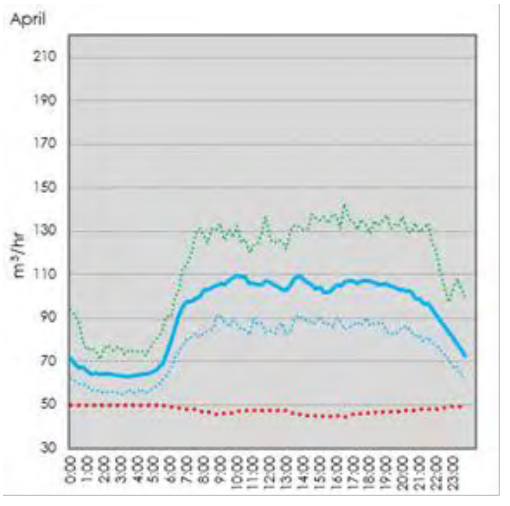
j) The leakage is integrated for every one of the 12 months and a percentage of leakage over SIV is estimated.

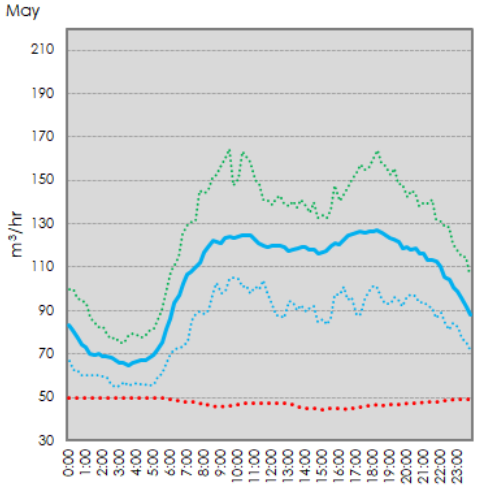
k) The difference of NRW percentage and leakage percentage is considered as the decoupled apparent losses percentage at this iteration and is added to the billed consumption of each EPANET node for the next EPANET run iteration.

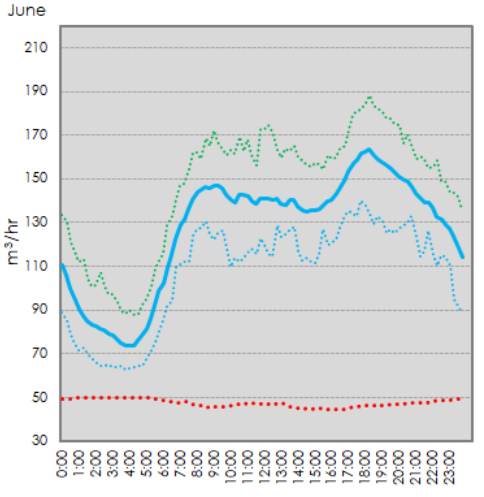












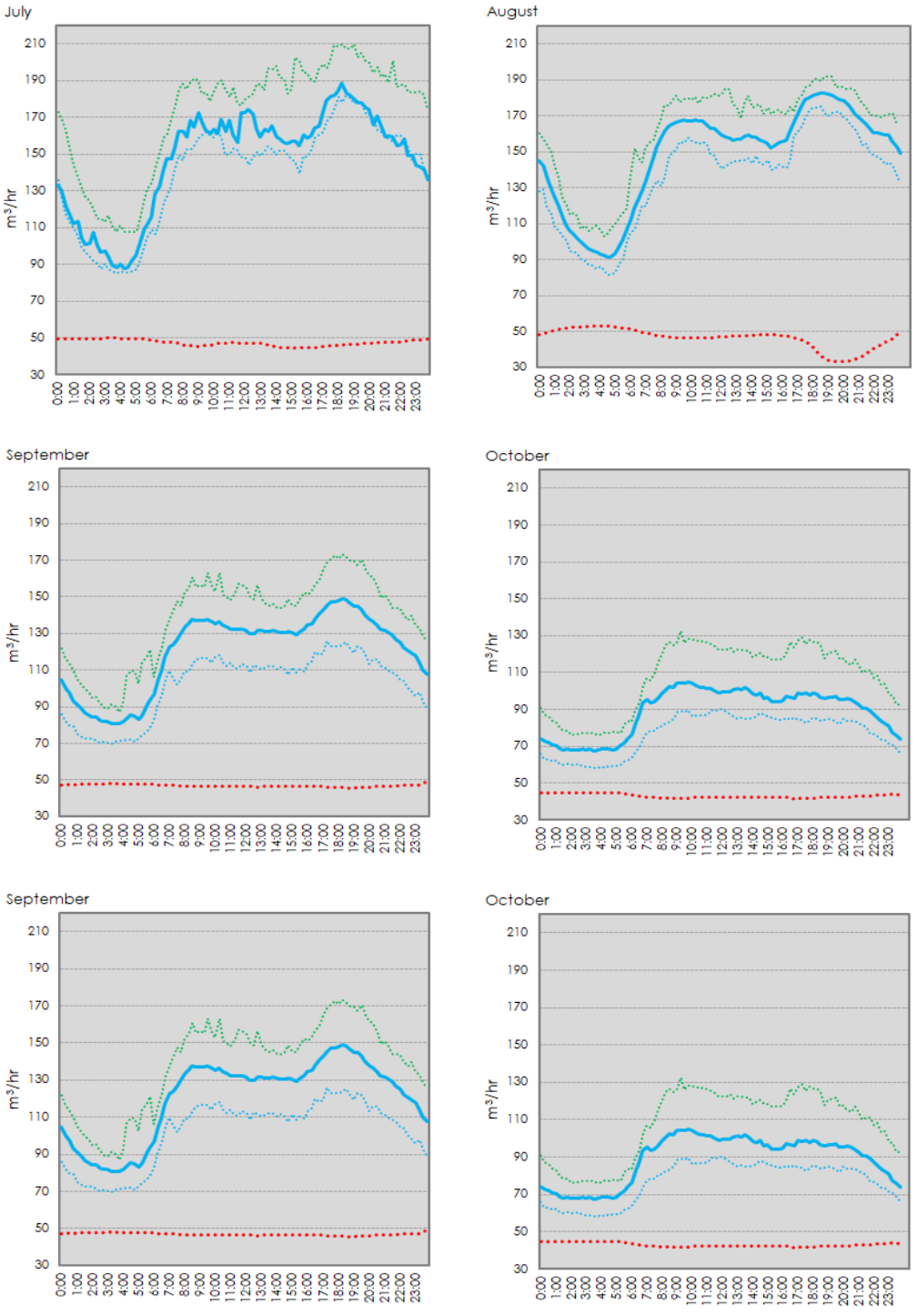


Figure 39. Flow daily patterns as estimated by average PRV records (maximum recorded in green dotted line, average in blue solid line, minimum recorded in blue dotted line) and leakage as estimated through the EPANET WDN average pressure patterns at the last iteration, after apparent losses are completely decoupled from leakage

The produced monthly diagrams of PRV recordings and leakage of Figure 39 (which are the final ones, after the end of the iterative process) show the intense seasonality of the case study. The winter months show relatively low SIV (≈90m³/sec) during the day and an extended minimum night SIV zone (from 60 m³/sec to 70 m3/sec) which lasts from 00:00 to 06:00. The level of the leakage during the winter months is depicting small variability (≈45 m³/sec). During the summer months the overall SIV is getting higher. The summer profile has two peaks, a morning one around 09:00 and an evening one around 19:00. The evening one is the highest (≈180 m³/sec). The summer months do not depict an extended night minimum SIV zone, but a nadir which occurs around 04:00. Summer leakage level is higher in absolute values (≈50 m³/sec), while it gets more variable during the day in August (from 30 m³/sec to 50 m³/sec). The ratio max SIV/min SIV is higher in summer (≈2) than the winter months (≈1.5), which is indicative of the more intense variability of water demand linked to the touristic activities.

Pressure Control Management Scheme and key indicators for the performance of the WDN

The utility and value of such a tool among many other aspects that will be analyzed in the Results and Discussion chapter, can also lie within the assessment of alternative PM schemes or even scenario on dividing the network into DMAs. To this purpose, a number of established KPIs such as the water losses per connection, the water losses per network length, the absolute leakage, and the electrical energy costs have been used (Kanakoudis et al., 2011a; 2011b; 2012 and 2013) and some original KPIs such as the absolute leakage difference between two scenarios, the PDD reduction (BAC or BAC + Apparent losses) and the sum of leakage and energy cost are introduced and estimated spatio-temporally or in an aggregate value depending on the nature of the KPI.

The absolute leakage and the absolute leakage difference between two scenarios in m³ would be the most common and useful KPI that the water utility manager would be interested to know. The amount of real losses saved due to an application of a different managerial scenario corresponds to water that is saved for the water resources. It is a variable that can easily be translated into groundwater level difference and help assess the issues of seawater intrusion, due to aquifer level drop.

Leakage per network mains length and the respective difference between two scenarios, a KPI suggested by the IWA best practice as indicative for assessing the performance of a management scheme (Farley and Trow, 2003) is also estimated.

An aspect that becomes increasingly interesting regarding the management of resources in general and specifically the management of water in an urban WSS is the resource nexus, which means the interlinkage of resources though a complex system of interrelations that may lead to synergies or trade-offs (Laspidou et al.,

2018). The water-energy nexus within the WSS can be found in the impact of leakage on head loss throughout the network (Colombo and Karney, 2002) and the coupling of energy consumption and water withdrawals, since every cubic meter of SIV has used energy to be pumped, transferred and/or treated. The second energy component is taken into account and estimated for this analysis, according to the user requirements. This way, the two commodities can be managed synergistically in such a system, since saving water almost proportionally would save energy. The KPI that is used to express the water-energy relation is defined as the energy saved due to leakage reduction comparing various managerial schemes. The actual value of energy that is spent for a cubic meter of water in kWhr/m^3 is estimated by the actual energy consumption data of DEYASK and is perceived as flat rate equal to 0.423 kWh/m^3. This assumption needs to be reconsidered if the variability in short term pumping depths were more significantly influencing the pumping energy.

An insightful KPI for the performance of a PM scheme is the reduction of PDD. This can be estimated for a geonode based on the actual pressure of the node, the PM scheme pressure of the node, the actual consumption and the pressure exponent (Morley and Tricarico, 2008). Two kinds of PDD volumes can be estimated: the reduction of the BAC and the reduction of the BAC plus the apparent losses, since they are expected to reduce as well even if this is not depicted to revenue. For the purpose of this work, the overall reduction of both components is estimated following equation 34.

PDD reduction = (BAC + Apparent Losses)*

(equation 34)

$$*[1-\left(\frac{P_{PCM\ scheme}}{P_{act}}\right)^{1.08}]$$

In a similar manner, a KPI regarding the economic impact of leakage reduction is introduced. This KPI is expressed in euros and represents the cost of the leakage difference between two scenarios as this is priced by the utility, also including the respective price of energy consumed. For Skiathos, this cost is estimated as 0.112 euros/m^3 as the sum of the price of water per m^3 and the cost of energy consumed for the pumping, transfer and treatment of water per m^3. The aforementioned prices are based on the Skiathos water utility relevant invoices for the reference year.

Other KPIs, that are used to assess the performance of a PM scheme, are relevant to the pressure fluctuations, since they are considered to relate to the networks stress and eventually bursts. Indicatively, Kanakoudis and Tolikas (2001)argue that velocity variability in small delivery pipes has a range of -50% to +50% of the

average value, while the relation of this variability to break rates is established. These KPIs are the decrease in pressure fluctuation, the decrease in pressure at nights and the decrease in the number of nodes where a pre-set maximum pressure value is exceeded.

All these metrics are applied for a tested scenario of PM (compared to the actual operation), in which a virtual CP, a different one for every time step, is set to never fall under the value of 5 m. This pressure level is not respectful to the legislative framework in Greece, which requires a pressure value at the ground of at least 4m*(number of building's floors + 1), but is quite realistic according to the utility. Nevertheless, the purpose of this task is not to suggest a valid PM, but to make a proof of concept regarding the usability of the suggested tool to assess scenario performance.

4.2.3. Results and Discussion

Figure 40 and Table 18 present the fitting of the simulated pressure at three points of the WDN to the actual recorded pressure as measured by pressure sensors for hourly values of the year 2016. The fitting is adequate, with R^2 values equal to 0.6729 (moderate fitting) for the central point, 0.7104 (strong fitting) for the eastern point, and 0.8852 (strong fitting) for the western point (Moore et al., 2013).

In Figure 41, SIV, BAC, Apparent losses, Real Losses, Revenue, and NRW are calculated for each trimester of 2016, using the developed model. One of the results that emerge is that in the warm trimesters (April to September), the SIV increases by approximately 50% compared to the winter and fall periods. The BAC and Revenue Water percentages are higher during these trimesters (44-52% of SIV), while the respective percentages for the winter and fall decrease to the levels of 32-34%. Real losses fluctuate from the summer level of 38.8-39.6 % of SIV to the high winter level of 63.3-64.7 % of SIV. Apparent losses fluctuate from 1.3-2.5% (first trimester) to 8.4-9.3% (third trimester) with the highest values noticed from July to December. NRW is inversely proportional to SIV, thus the former is at lowest levels during the summer trimester when SIV is the highest. Regarding the overall annual analysis (Figure 42 and 43), BAC is estimated equal to 42.1-42.3% of SIV, Real Losses equal to 50.9-52.2%, theft higher than 3.6% and metering accuracies lower than 2.4% (Apparent losses equal to 6%).

Regarding the uncertainties of the measured and estimated values of the WB components, the error of SIV is taken equal to ±0.25% according to the flowmeter specifications, the error of BAC and Revenue Water is taken equal to 0, since they represent the exact amount of water that is billed, and the Real Losses error is estimated following the rules of error propagation for the equation that produced their values, as follows:

$$L_1 = \left(\frac{P_1}{P_2}\right)^{N_1} * L_2 \,, \qquad \text{Error}(L_1)=2*\text{Error}(P) * N_1 + \text{Error}(L_2)$$

where

Error (L_1) is the error of Real Losses,

Error (P) = Error (P_1) = Error (P_2) is the error of the simulated pressure values as estimated for each trimester and for the whole year of simulation,

Error (L_2) = ±0.25% is the error of the minimum night flow measurement (the night use is a relatively small amount that does not change the error of the measurement even with a ±5% error),

and N1 = 1.08.

The highest error estimated for the real losses equals to 1.55% and corresponds to the 4th trimester. The error of NRW is estimated, with 95% confidence limits, from ±0.37% to ±0.52% for the trimester values and equal to ± 0.43% for the annual value. The error of the Apparent Losses is the most significant one with highest value equal to ±30.3% for the first trimester, while it is lower, equal to ±9.46% for the annual WB.

From the above it can be conducted that higher accuracy of the WB assessment can be achieved by increasing the accuracy of the model. The methodology followed for developing the simulation of the WDN uses average values in some specific steps, such as the monthly averaged consumption profiles and the monthly averaged pressure profiles. These average values are used as initial values at the beginning of its iteration, for disaggregation purposes and for the estimation of leakage variation during the day. The error in this specific application is low and leads to acceptable fitting, however if the N1 value was higher and not close to 1, it would introduce a more significant error to the simulation. For this reason, it is suggested that the methodology developed can perform better with real time data and a denser grid of monitoring devices. This way the tool can evolve to a real Digital Twin.

Assuming a linear relation between metering inaccuracies and the BAC (the water-meters in Skiathos are older than 10 years, thus expected to be under-registering flow with a rather linear relation (Kanakoudis and Tsitsifli, 2010a; Stakiadis and Papanikolaou, 2007; Arregui et al., 2015), theft can be estimated to take its higher values during the third and fourth trimester, higher than 6% and 5.2% of SIV respectively (Figure 43). This implies that theft cannot be linked wholly to the touristic activity but may be influenced by other social and/or economic factors such as the demand patterns of the local gardening and orcharding. The metering inaccuracies are smaller than 3% (or even smaller) in all trimesters, and smaller than 2.4% annually. For this analysis it should not be forgotten that Apparent losses as a whole have a relatively high level of uncertainty.

Except for the percentage of SIV that makes up Real Losses, the Infrastructure Leakage Index (ILI) is calculated and is found equal to 8.66, larger than 8 which is the limit between technical performance categories C and D. This would set the Skiathos WDN technical performance to the Category D, which implies inefficient use of resources, indicative of poor maintenance and system condition in general (Seago et al., 2005). Table 19 presents the parameters used for the estimation of ILI. A validation of the minimum night-flow analysis leakage assessment can be also made through the component analysis of Real Losses, as suggested by A. Lambert (A. Lambert, 2002) who claimed that leakage assessment should preferably be implemented with more than one methodologies to avoid errors. The component analysis of Real Losses is based on the assessment of the Unavoidable Annual Real Losses (UARL) due to leakage in mains and service connections. The component-analysis approach, based on the details of Table 19, gives a leakage of 45.57 m^3/hr (=0.5 m^3/d/connection *2,187 connections /24 hr), taken into account the suggested rough value of 0.5 m^3/d/connection for technical performance category D (ILI very close to the benchmark of 8 as estimated in Table 19), developed countries, and average WDN pressure of 50 m (Seago et al., 2005). This value is quite close to the minimum night-flow approach estimation that equals 44.99 m^3/hr.

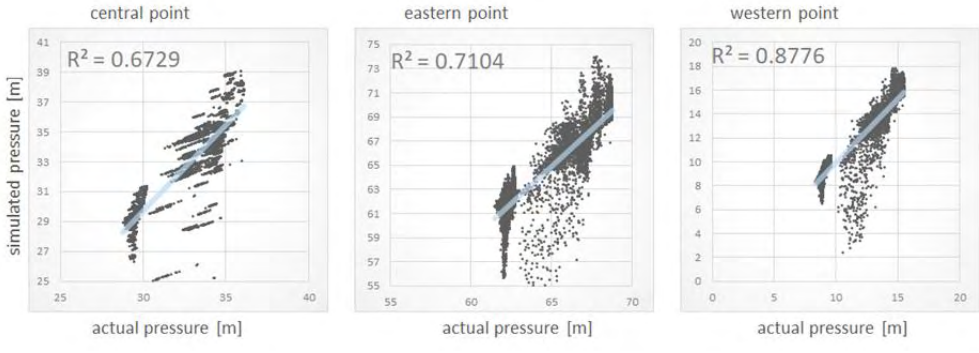


Figure 40. Scatterplots of actual and simulated hourly pressure values in m for three cello points: the central at the left and the eastern at the middle, and the western at the right

Table 18. Simulated and actual trimester-average pressure values in m for the three cello points

		simulated	actual	% error
central point	January-March	29.711	29.708	0.01

	April-June	34.810	34.865	-0.16
	July-September	33.371	33.236	0.41
	October-December	29.497	29.355	0,00
eastern point	January-March	62.384	61.913	0.76
	April-June	67.442	67.486	-0.07
	July-September	65.984	65.938	0.07
	October-December	62.173	61.752	0.68
western point	January-March	9.155	9.147	0.09
	April-June	14.349	14.417	-0.47
	July-September	13.056	12.991	0.50
	October-December	8.965	8.852	1.28

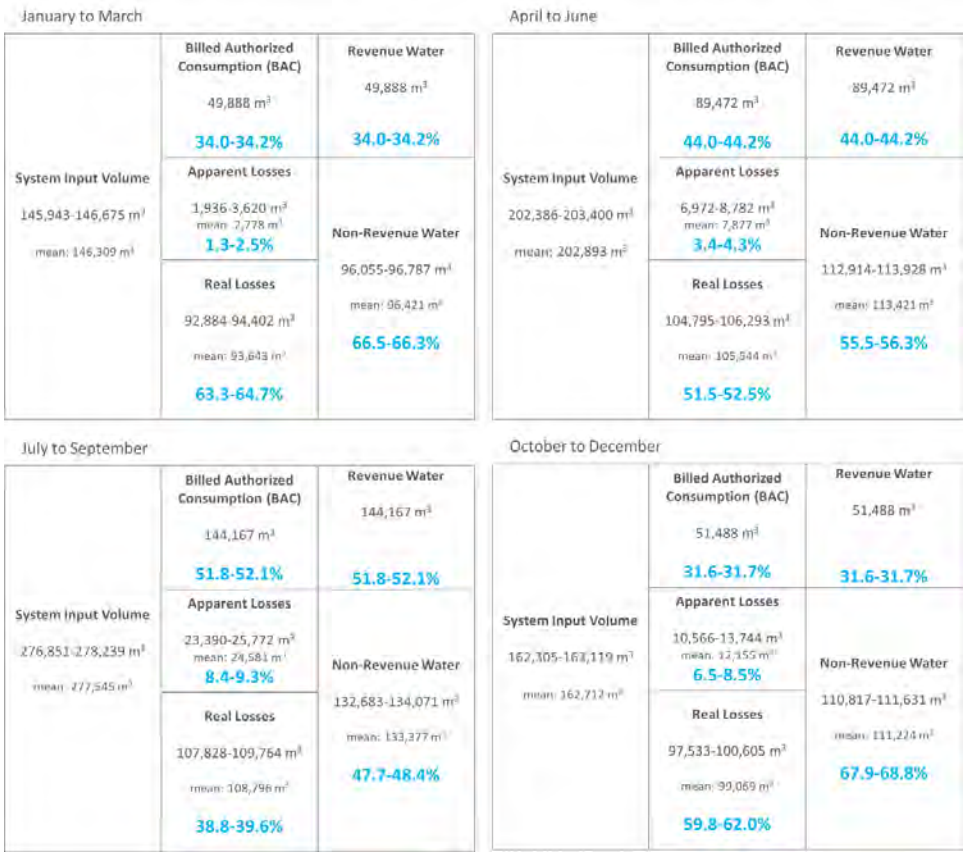


Figure 41. Seasonal high-level component analysis of SIV into BAC, Apparent Losses, Real Losses, Revenue Water and Non-Revenue Water (All components are also presented as percentages of SIV)

System Input Volume 787,485-791,433 m³ mean: 789,459 m³	**Billed Authorized Consumption (BAC)** 333,015 m³ **42.1-42.3%**	**Revenue Water** 333,015 m³ **42.1-42.3%**
	Apparent Losses 42,904-51,878 m³ mean: 47,391 m³ **5.6-6.6%**	**Non-Revenue Water** 452,469-456,417 m³ mean: 454,443 m³ **57.2-58.0%**
	Real Losses 403,022-411,082 m³ mean: 407,052m³ **50.9-52.2%**	

Figure 42. Annual component analysis of SIV into BAC, Apparent Losses, Real Losses, Revenue Water and Non-Revenue Water (All components are also presented as percentages of SIV)

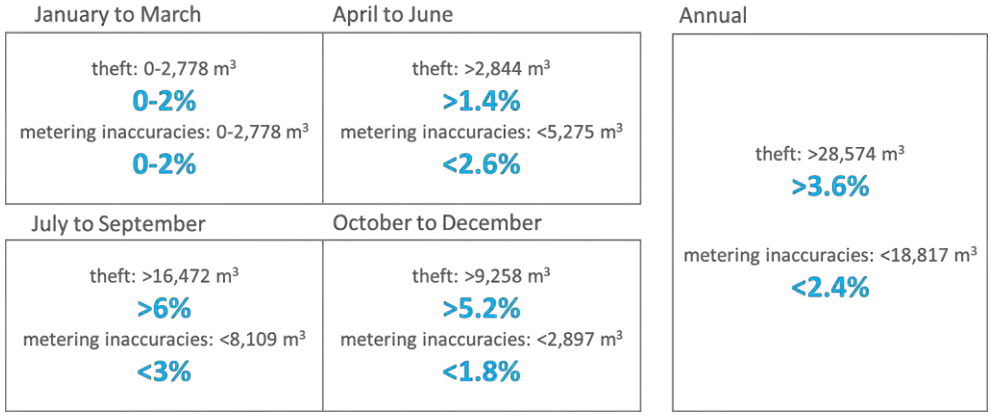


Figure 43. Split of apparent losses (mean values) with the assumption that metering inaccuracies are proportional to BAC, where the limits are true if theft during January to March is negligible

Table 19. Characteristics of Skiathos WDN used for the estimation of Background Losses and ILI

UARL (IWA) (m³/hr) =(18**total pipe length*+0.8**number of connections*+25**length of connections*)* **average WDN pressure*/24/1,000	5.19
total pipe length (km)	2,187 (<5,000,>3,000)
number of connections	42 (>25)
length of connections (km)	10.94
average WDN pressure (m)	53.5
length per connection (m/con)	5
CARL (m³/hr) *as estimated by the bottom up methodology*	44.99
ILI= CARL/UARL	8.66
reported leaks and bursts (m³/hr)	20.27
connection density (con./km)	129
population	6,100
average night pressure (m)	17
number of bursts in mains/year (min,max)	5-10
number of bursts in mains/year	7.5
number of bursts in distribution pipes/year	180
flow Rate for Reported main bursts (m³/hour/m pressure)	0.24

flow Rate for Reported distribution bursts (m^3/hour/m pressure)	0.032
repair duration (hr)	48
unreported leaks and bursts (m^3/hr) = *CARL-reported leaks and bursts*	24.72

A major function of the produced tool is the spatio-temporal supervision of the WDN regarding its SIV distribution throughout different landzones and different time intervals of selected lengths. The analysis scales down to hourly steps, while the longer the selected period is, the safer the results can be. The same also applies to the spatial scale, meaning that, even though the tool can zoom down to household level, the larger the area of focus is the more accurate the estimations are. This is because for greater amounts of water in WBs of wider areas or longer periods the different uncertainties, such as the uncertainty of the water meter locations and readings, play a less significant role. Additional features that can be supervised are the BAC, which for the case study of Skiathos equals to the sum of Revenue water, Real Losses, and Apparent Losses.

Discerning Apparent losses into Unauthorized Consumption and Metering Inaccuracies in smaller time intervals (daily) would require additional information regarding the performance of meters and transparent past reporting of unauthorized consumption incidents (Kanakoudis and Tsitsifli, 2010a). Such a spatio-temporal analysis can enhance the operator's understanding of the WSS and reveal the drivers of NRW and its components. The comparative diagrams of daily profiles, January 1st and August 1st, for two landzones of different elevation, 0.5 m and 39.5 m (Figure 18) can shed light on the intense seasonal and spatial variability of the investigated features: Comparing the daily Landzone Input Volumes (LIV) for the two landzones, one can notice that the land-zone of higher elevation requires less LIV for both dates. The reason for this might be three-fold. Firstly, the two landzones include different number of water-meters and different land-uses, secondly the elevation may have an impact both in leakage and pressure driven demand, and thirdly the infrastructure condition may also vary. Further inspection of the rest of the components can shed more light to the situation. Regarding the LIV increase for the two landzones, it can be estimated that the hilly one has an increase of 70%, while the coastal has an increase of 78%. This difference in the seasonal LIV increases of the two landzones can imply that the coastal landzone attracts much more touristic activity than the hilly one. This is actually true, since the coastal landzone in Skiathos is a part of the highly visited Skiathos promenade that hosts a large number of tourist-related small businesses, such as restaurants, cafes, pubs, and taverns. This is reinforced when revealing that the elevated landzone includes 25 water-meters, while the coastal includes eight. More evidence regarding the land uses of Skiathos town would require a spatial cluster analysis that would be implemented with the criterion of water-meter labeling into

"households", "hotels", "café-restaurants", etc. An urban land-use clustering analysis would also improve the spatio-temporal assessment of the WB, since it would facilitate the linking of different consumption patterns with different land-zones. Such a work is already conducted with the use of the Self-Organized Maps (SOM) methodology (Laspidou et al., 2015). The impacts of the intensity of touristic consumption and pressure driven consumption due to elevation difference are furtherly reinforced, when comparing the respective BAC seasonal increases, which are 111% for the hilly landzone and 203% for the coastal one. The comparison of Real Losses throughout time and space may prove a very useful tool for the prioritization of the network maintenance with quantifiable evidence. For example, the analysis for the two landzones of Figure 44 offers all the evidence needed to prioritize maintenance works for the coastal landzone. Specifically, the coastal landzone depicts real losses as high as 70 % of SIV in 1st of January that fall down to 44% in 1st of August. Respectively the hilly landzone depicts 56% in 1st of January that falls decreases to 40%.

Absolute quantities of water can also be converted in expressions relative to service connections, or main length (Table 20), which—depending on the purpose—are more useful than their expressions as total amounts. For instance, for evaluating the management of the WDN, IWA recommends the annual expression of the Real Losses in m³/connection(Winarni, 2009). However, such an expression would be more helpful for comparing the overall network performance to that of another

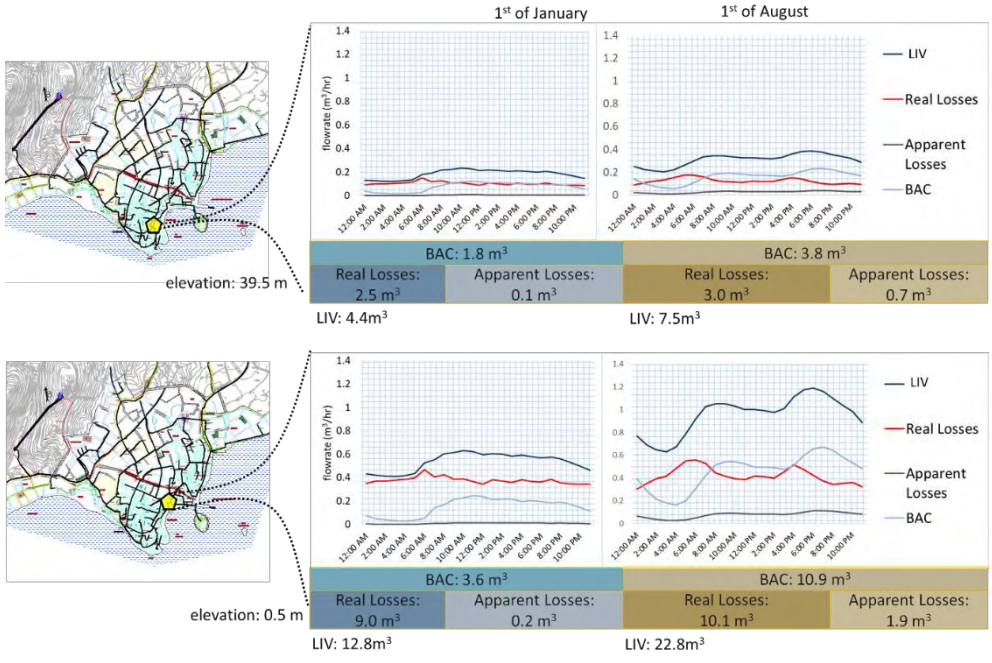


Figure 44. Hourly profiles of Landzone Input Volumes, Real and Apparent Losses, and BAC of 1st of January and 1st of August, for two indicative neighborhoods of the network with the elevation of 39.5 m and 0.5 m respectively

network, rather than the internal comparison of two landzones, especially in the absence of DMAs. Anyway, these metrics can be indicative of what causes leakage in different landzones (e.g. altitude vs. infrastructure condition). On top of that, they can be indicative of the cause of leakage, since on one hand, leaks and bursts are more frequent on the service connection links, but on the other hand bursts in mains result in much higher flowrates (Lambert, 2002). Taking this into account, comparing the two landzones, one can suspect that the coastal zone might be more stressed due to bursts in mains, while the hilly landzone due to leaks in the service connections, since the coastal zone has much higher values of Real Losses/ service connection than the hilly, while they have almost the same Real Losses/ network length. The much higher value in Real Losses/ service connection could be justified either by more incidents of bursts or by the elevation difference.

Additionally, the per-service-connection expression is more useful for conclusions regarding the level of touristic consumption and pressure-driven consumption, since the need to specify the number of households within a landzone is eliminated. In particular, comparing 0.072 m^3/service connection to 0.450 m^3/service connection for the hilly and coastal landzones respectively is more safe for conducting the conclusion that the coastal zone has more consumption in the non-touristic winter due to pressure driven demand, while comparing the corresponding amounts in the summer, 0.152 and 1.362 m^3/service, can shed light on the more intense touristic activity of the coastal landzone.

Table 20. The daily amounts of LIV, BAC, Real Losses, and Apparent Losses for two indicative landzones expressed in per-network-length and per-service-connection metrics.

landzone	hilly (39.5 m)		coastal (0.5 m)	
network length	60.34 m		201.44	
service connections	25		8	
	Jan, 1	Aug, 1	Jan,1	Aug,1
LIV/network length (m^3/d/m)	0.073	0.124	0.064	0.113
BAC/network length (m^3/d/m)	0.030	0.063	0.018	0.054
Real Losses/network length (m^3/d/m)	0.041	0.050	0.045	0.050
Apparent losses/network length (m^3/d/m)	0.002	0.012	0.001	0.009

LIV/ service connections (m^3)	0.176	0.300	1.600	2.85
BAC/ service connections (m^3)	0.072	0.152	0.450	1.362
Real Losses/ service connection (m^3)	0.100	0.12	1.125	1.263
Apparent Losses / per service connection (m^3)	0.004	0.028	0.025	0.238

Figure 45, figure 46 and Table 21 present various KPIs indicative for the performance of the simulated PM scheme. The PM scheme does not involve any DMA division, but only a temporal adjusment of the pressure to satisfy the pressure needs of the node that for each timestep is estimated to have the lowest pressure of the WDN. The pressure at the starting point of the network, at the PRV, is set to keep the pressure at this virtual critical point higher than 5 meters. This may not be respective to the relevant legislation, that requires pressure higher than 16 meters (=4m*(3 floors+1), but according to the water utility, it is rather realistic and applies to the current practical requirements of the WDN operation. At this point, it should be noticed that the PM scheme is not suggested, but rather applied to make a proof of concept.

Figure 45 shows the effect of the PM sheme in a high temporal scale of four trimesters. The annual potential leakage reduction is estimated to reach 51,300 m^3, which corresponds to 21,763 kWhs of energy for pumping, transfering, treating and other operational costs, and to 5,671 euros of revenue (Table 21). By the trimester analysis, it can be deduced that the highest potential for leakage reduction exists in April to June and July to September, 17.5 % and 16 %, respectively (Figure *45*, right). The absolute amounts in cubic meters are presented in the left of Figure 45, where the deep blue shaded area shows the quartely variation of the leakage reduction potential.

Table 21 presents the trimester values of energy savings and revenue for each trimester as well as two KPIs that often are interesting to water utilities, the night pressure decrease and the pressure variation decrease. The two KPIs that involve pressure are more relevant to the protection of the WDN against bursts.

Figure 46 proves that a spatial analysis of the critical designing parameters can be an important tool that quantifies needs and prioritizes improvements for the water utility to help in asset management. The constructed tool can quantify the benefits of a PM scheme in time and in space, at the scale of a landzone. By comparing the two indicative landzones (or any other neihboring landzone), the operator can realize that the touristic zone in the promenande of Skiathos has significantly more

room for improvement in terms of leakage reduction, than the elevated zone. At the same time, comparing the temporal differences it can be deduced that the amounts of leakage that can be saved through a PM scheme can double in August 1st compared to January 1st for both landzones.

The reduction of PDD is also presented in Table 22, in macro scale and in Figure 46 proves that a spatial analysis of the critical designing parameters can be an important tool that quantifies needs and prioritizes improvements for the water utility to help in asset management. The constructed tool can quantify the benefits of a PM scheme in time and in space, at the scale of a landzone. By comparing the two indicative landzones (or any other neihboring landzone), the operator can realize that the touristic zone in the promenande of Skiathos has significantly more room for improvement in terms of leakage reduction, than the elevated zone. At the same time, comparing the temporal differences it can be deduced that the amounts of leakage that can be saved through a PM scheme can double in August 1st compared to January 1st for both landzones. In Table 22, two alternative expressions of PDD reduction are intorduced. Expressing PDD in m^3 per m^3 of BAC + Apparent losses is more helpful when it comes to drawing conclusions on the impact of altitude in PDD reduction. In these two landzones, it seems that the specific scheme has more impact in higher altitudes than in lower. This is reasonable because the chosen PM scheme, which is lumped regarding DMA division, causes much greater percentage pressure reduction in high attitudes than in lower ones. This fact indicates that a multi-DMA scheme would be beneficial, since with a single DMA relative savings are lower where the potential for savings is higher. The second expression introduced for PDD reduction is in m^3 per service connection. This is a more helpful expression for assessing the impact of urban land uses in PDD reduction. In particular, we can see that the touristic zone, with all the restaurants, bars and cafes, has much higher PDD reduction both in winter and summer than that of the residential zone.

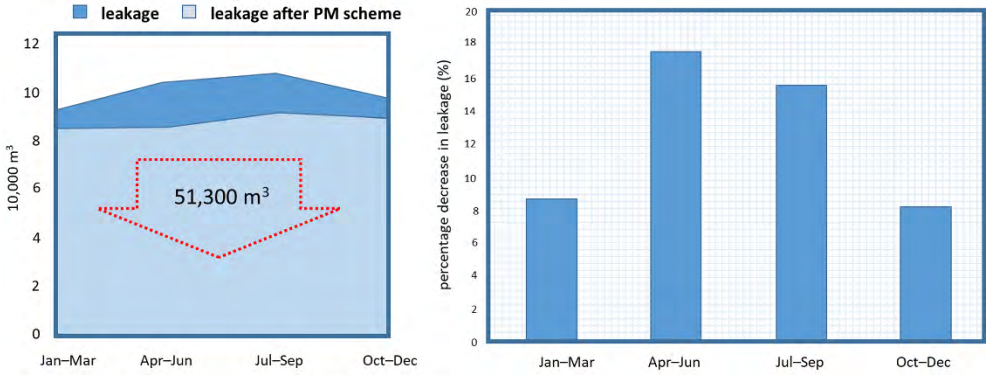


Figure 45. Comparative diagram of estimated leakage per billing period before and after the modelling application of the theoretical PM scheme for 2016 (left) and bar graph of the corresponding percentage decrease for each billing period (right)

Table 21. Indicative Trimester KPIs for the performance of a theoretical PM scheme for 2016.

trimester	night pressure decrease (%)	decrease in pressure fluctuation (%)	energy savings (kWh)	economic savings (euro)	PDD reduction (m^3)
Jan-Mar	9.5	43.2	3,389	897	4,786
Apr-Jun	19.2	41.2	7,809	2,067	17,462
Jul-Sep	17.6	38.7	7,131	1,888	25,968
Oct-Dec	9.28	47.8	3,435	909	5,514
annual	13.9	42.7	21,763	5,761	53,730

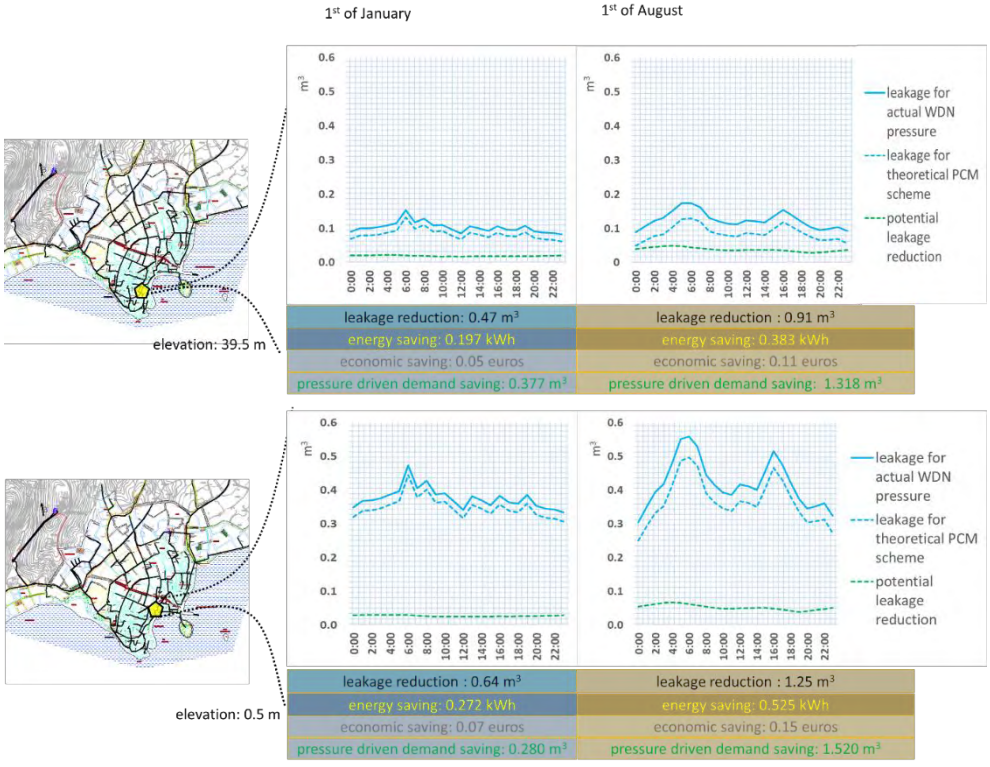


Figure 46. Beneficial effect of a theoretical PM scheme in two land-zones of different elevation presented through different KPIs: hourly comparative curves of leakage for actual pressure scenario and PM scenario, the curve of the respective leakage reduction, daily values of leakage reduction, and daily values of savings in energy and euros, for January 1st and August 1st

Table 22. PDD reduction for two indicative landzones expressed in m^3 per m^3 of BAC + Apparent Losses and in m^3 per service connection

landzone	hilly (39.5 m)		coastal (0.5 m)	
service connections	25		8	
	1st of Jan	1st of Aug	1st of Jan	1st of Aug
PDD reduction in m^3/m^3 of BAC + Apparent losses	0.199	0.293	0.074	0.119
PDD reduction in m^3/service connection	0.015	0.053	0.035	0.190

Through the results that have been presented, it becomes apparent that a spatio-temporal analysis of the key SIV components and a series of relevant KPIs can make a conclusively insightful tool for the improvement of the WDN operation. The quantification of real losses, apparent losses, authorized consumption, the respective quantities in energy consumption and revenue mapped spatially and temporally can give a full picture regarding the weaknesses of the WDN and help the operator prioritize asset replacement or renovation needs. Through such an analysis, the localization of leakage, metering inaccuracies and theft can be realized at least at the scale of a landzone. Additionally, through such a tool the benefits of a PM scheme can be quantified in spatial and temporal detail.

The potential of this tool can extend to the investigation for an optimum DMA division scheme where that can be achieved by a multi-objective optimization algorithm that would involve leakage minimization, energy consumption minimization, revenue maximization and other KPIs that are presented (Creaco and Haidar, 2019; Creaco et al., 2019; Di Nardo et al., 2017). Such an attempt would firstly require the decoupling of energy consumption and revenue from leakage. The two quantities shall alternatively be estimated also as a function of elevation, taking into consideration the dynamic energy needed to transfer water to higher levels, rather than calculate them with use of a per-cubic-meter flat rate throughout the whole WDN. This can suggest potential future work.

Afterword: A spatio-temporal simulation of a WDN is implemented, with the use of EPANET. In order to decouple the demand driven and pressure driven components of SIV, leakage “valves” are used in every EPANET node. The leakage is estimated through the IWA suggested night-flow approach that is based on the night-flow consumption. The component analysis approach is also used to validate the first estimation. A scheme of iterative processes has been constructed with two nested loops. The inner iterates until leakage converges to a given value. The outer iterates until the simulated pressure at specific points matches the actual known pressure values. The results of such an analysis

can offer conclusive insight to the operator of the WDN regarding a series of designing parameters and KPIs of the network. The produced toolkit can serve as supervision support for the WDN or the basis for construction of a Digital Twin. Leakage, consumption, and theft can be localized at a relatively small scale, while the intense seasonal impact of tourism on the WDN can be quantified, as well as a series of useful conclusions can be conducted. Some of these conclusions regard the impact of elevation, the change and impact of land uses, and the respective amounts of energy consumption and revenue. Annual and trimester The estimated WB tables depict the intense seasonal variability of the components and highlight the need to increase the time scale in the relevant assessments at case studies with seasonally variable demographics or weather, such as the Greek touristic islands. For Skiathos case study the analysis has shown that the annual Real Losses are from 50.9% to 52.2% and the NRW from 57.2% to 58.00 %. The first trimester shows the highest percentage of real losses, from 63.3% to 64.7%, while the third trimester shows the highest percentage of BAC, from 51.8% to 52.1%. The theft is estimated higher than 3.6%, while it increases significantly from July to December, both in absolute values and percentages. A theoretical PM scheme is applied to prove the value of such a tool, for quantifying the relevant benefits spatio-temporarily. Leakage reduction, PDD reduction, energy and economic savings are four KPIs examined for the PM. An annual potential of 51,300 m^3 in leakage reduction is estimated with the application of the PM scheme. The highest leakage reduction in absolute values can happen from April to September. Annual energy savings, economic savings, and PDD reduction potentials are estimated equal to 21,763 kWh, 5,761 euro and 53,730 m^3, respectively The form of the most appropriate expression for each KPI is also investigated, concluding that different expressions help reach conclusions in different aspects.

The content of chapter 4.1 is included in the following published work:

Kofinas, D.; Mellios, N.; Laspidou, C. Spatial and temporal disaggregation of water demand and leakage of the water distribution network in Skiathos, Greece. In Proceedings of the MDPI AG in 2nd International Electronic Conference on Sensors and Applications Session Sensing Technologies for Water Resource Management, 5–30 November 2015; p. S7001.

- The contribution of Mr. Kofinas, D. involves the conceptualization, the methodology, the review of the software, part of the data curation, the validation, the formal analysis, the investigation, the writing, and the visualization.
- The contribution of Mr. Mellios, N. involves the data provision.
- The contribution of Professor Laspidou, C. involves the scientific supervision.

This work was supported by the project ISS EWATUS—Integrated Support System for Efficient Water Usage and Resources Managementdwhich is implemented in the framework of the EU 7th Framework Programme, Specific programme Cooperation In- formation and Communication Technologies; Grant Agreement Number 619228.

The content of chapter 4.2 is included in the following published article:

Kofinas, D., Ulanczyk, R., & Laspidou, C. S. (2020). Simulation of a Water Distribution Network with Key Performance Indicators for Spatio-Temporal Analysis and Operation of Highly Stressed Water Infrastructure. *Water*, *12*(4), 1149.

- The contribution of Mr. Kofinas, D. involves the conceptualization, the methodology, the review of the software, part of the data curation, the validation, the formal analysis, the investigation, the writing, and the visualization.
- The contribution of Mr. Ulancyk, R. involves the software, the data curation, part of the investigation (relevant to the programming), and the review.
- The contribution of Professor Laspidou, C. involves the scientific supervision and review.

The work described in this paper has been conducted within the project Water4Cities. This project has received funding from the European Union's Horizon 2020 Research and Innovation Staff Exchange programme under grant agreement number 734409. This paper and the content included in it do not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

5. Modelling household consumption

5.1 A methodology for household water consumption modeling

Foreword: In the smart cities context, real-time knowledge of residential water consumption has become increasingly important, especially given the fast evolution of sensors, ICT and the production of big, high-resolution data coming from the urban environment. A variety of reasons often leads to the creation of continuity gaps in these data series, thus making the need for a methodology that produces reliable and realistic synthetic data urgent. In this article, we present a methodology that generates synthetic household water consumption data; we showcase it in two case studies, Skiathos, Greece and Sosnowiec, Poland, which exhibit significant differences in water consumption patterns. The methodology captures the stochasticity of daily residential water use. Algorithm validation is implemented through the comparison of various metrics for actual and generated data; this way, we show that the suggested approach is capable of adequately simulating water consumption in both micro- and macro- time scale.

5.1.1. Introduction

In fields such as DM, ML and Knowledge Discovery from Databases (KDD), a commonly emerging issue, which is the main focus of this chapter, is that of missing values or missing data. Numerous reasons can lead to such a problem: Equipment malfunctions, refusal of respondents to fill in questionnaires and gathering of erroneous data, etc. (Schafer and Graham, 2002; Batista and Monard, 2002). Demand management initiatives rely on good comprehension of water usage practices, as well as of factors influencing water demand (White et al., 2003). The emerging Data-Driven Demand Management has been supported by cloud-based data platforms and represents a new, critical element to improve decision-making in today's water industry. Utility managers can achieve the sustainability and affordability objectives they desire through the practical application of data analytics (Fielding et al., 2012). In this context, the implication of data gaps is really important, since the decision-making process relies on continuous data sets. Such continuous data sets improve the resilience of new decision-making schemes.

Based on the reason why a gap is created, missing data is categorized into three classes, depending on the level of randomness of the incident: Missing Completely At Random (MCAR), Missing At Random (MAR) and Not Missing At Random (NMAR) are commonly used classes that imply that the incident either does not depend on the missing value, or depends on a related to the value attribute, or directly depends on the value, respectively (Little and Rubin, 2002). An example for a MCAR would be the interruption of functioning of a sensor. That would create a gap no matter what the measurements would be. An example for MAR would be the absence of answer in a questionnaire about an attribute that is indirectly related to the gender of the respondent. An example for NMAR would be the case

of a sensor not recording a value, because it lies outside its measuring capacity range. Thus, a missing or erroneous value would imply that it is out of this range. The level of randomness is conclusive of the method that the missing data are treated. Depending on the class that the data gap belongs to, a different methodology for treating the missing data is selected.

Another criterion for choosing the method to treat an incident of missing value is the nature of the attribute. Specifically, if the attribute were a time-series, the treatment would involve analysis of components, such as trend and seasonality. Moreover, if the missing attribute value were correlated to another known attribute, then the method of treatment would be selected based on this correlation, which would imply the implementation of multivariate analysis, as opposed to univariate. Lastly, a criterion is the "length" of the missing part—this can vary from a single missing value to a larger gap of data. The aforementioned criteria are decisive of the treatment of an incident: variable methods are applied for this purpose. Some commonly applied tactics include ignoring and discarding the incident, case substitution mean or mode imputation, hot deck and cold deck method, applying a predictive model and others (Batista and Monard, 2003; Grzymala-Busse and Hu, 2001;Lakshminarayan et al., 1999). The imputation of a missing value is generally classified into deterministic or stochastic (Rao, 1996).

Other than filling missing data gaps, the production of data that mimic the properties of a data set (synthetic data) can be essential in situations in which available real data are limited and longer data sets are required for evaluation, validation and/or testing of models, platforms, algorithms, or Decision Support Systems (DSSs). Barse et al. (2003) define synthetic data as generated data by simulated users in a simulated system, performing simulated actions. A typical example of need for synthetic data is the case when privacy constraints block the direct use of original sets. In other words, water utilities may not agree to provide actual water consumption data, being concerned about violating the privacy of their customers, even if data is anonymized. Synthetically generated data overcome problems related to data privacy (Cominola et al., 2016). In such cases, the use of a tool that provides synthetic water consumption data will serve well the needs of water utilities, including decision-making platforms used in data-driven demand management schemes. However, privacy is not the main issue. That would be the unavailability of long enough series to do uncertainty based analysis. Another example is the training and adapting of a Fraud Detection System (FDS) on a synthetic data set, testing its properties by injecting synthetic frauds or comparing the performance of different FDSs (Barse et al., 2003).

Past research works have focused on investigating whether urban water consumption time series can be simulated in multiple temporal and spatial scales. Pulse models, developed for creating such artificial time series, generally consider two variables, the duration and intensity of a consumption event. These models can be divided into two categories: the ones that simulate pulses of specific end

use fixtures and the ones that are parameterized to simulate overall household water demand (Creaco et al., 2017). Buchberger and Wu (1995) used the Poisson Rectangular Pulse (PRP) methodology to simulate the behavior of the water consumer, considering actual monitored household consumption. Alvisi et al., (2003) introduced a cluster Neyman-Scott stochastic process to simulate residential water demand, respectfully to the cyclical behavior observed during a typical working day. Blokker et al. (2009) developed a methodology for simulating water consumption at residence level using 8 end uses (bathtub, dishwasher, etc.) patterns, based on survey data and technical characteristics of the appliances. Cominola et al. (2016) used data from 300 households in 9 U.S. cities and developed a stochastic simulation model for the generation of residential water end-uses based on the assumption that each end-use is characterized by a unique signature. Creaco et al. (2015) proved that taking into account dependence between duration and intensity variables can improve the pulse approach performance. Creaco et al. (2016) focused on parameterizing the values of the aforementioned model which were associated with the model variables respectfully to the water balance characteristics in multiple time scales. They concluded that high accuracy in smart metering relates to better performance of the model. Kossieris et al. (2016) applied the Bartlett-Lewis clustering mechanism for the simulation of residential water demand. The model variants were assessed to preserve the main properties of the actual time-series at a range of sub-hourly fine time scales, from 1-min to 15-min. The difference between Bartlett-Lewis and Neyman-Scott approaches lies in the way pulse incidents are distributed to clusters (Rodriguez-Iturbe, 1987). In the 2017 study of Di Palma et al, the "Overall Pulse model" was introduced to describe aggregated water consumption. This model does not generate single end use pulses but the water consumption of a whole household as recorded at a water meter. Kossieris and Makropoulos (2018) investigated the statistical properties of 15-minute and hourly water demand data of eleven Greek households. Among their findings they identified Gamma and Weibull distributions as superior to describe the non-zero demand values. Cominola et al. (2019) used smart water data from 327 households in Australia to identify water end-use signatures and concluded to three distinctive profile clusters, these of shower, washing machine and irrigation.

A methodology on filling a gap of water consumption data with meaningful values is presented in this chapter. The suggested methodology is in accordance to the established residential water demand pulse models, since it is based on simulating the user behavior, while considering characteristic variables such as intensity of flow and duration. The methodology differs from other established approaches in the fact that it firstly captures consumption patterns throughout the day and then introduces a novel algorithm that simulates the duration of the incident. The suggested innovative algorithm for simulating the duration attribute gives a valuable degree of freedom that allows dealing with the consumption pattern beforehand. Thus the methodology overall facilitates capturing precisely the

pattern. The water consumption data concern household water consumption and are collected for the purpose of investigating the effect of real time monitoring and informing consumers about their own water consumption on their water use behavior, through an ICT-supported consumer-awareness process. The methodology is developed, validated and applied to the created gap in order to retrieve required data sets to be used for the development of a DSS platform, developed in the context of the EC funded project ISS-EWATUS ("ISS-EWATUS Integrated Support System for Efficient Water Usage and Resources Management", 2016).

In this article, the two study areas and the data characteristics are firstly described. Next, the methodology is presented in a step-by-step format for the base case study of Skiathos households, as well as its modification for Sosnowiec households. The method validation process and results are presented and discussed. The article finishes with a series of conclusions and implications of the new data generation tool for the urban water sector.

5.1.2. Materials and Methods

- *Case study specifications*

In the context of the ISS-EWATUS project, sensors were installed in faucets, showers and appliances in multiple households in Skiathos Island, Greece and Sosnowiec, Poland. For the purpose of this article, data sets from 16 households were used, 10 and 6, with 10 and 9 sampling points, respectively (faucets at Skiathos and faucets and appliances at Sosnowiec). The monitoring period was initiated on February 1st, 2015, while data collection is still ongoing, as of November 2016. Technical issues during the installation in Skiathos Island delayed the initiation of data recording, which officially began on April 14, and created a data gap of 72 days (from February 1st to April 14, 2015) for the Greek case study. The initial motivation for the Synthetic Data-Generating (SDG) methodology presented in this article was to fill the data gap that was created in the Greek case study. Once this goal was reached, we extended the methodology to a generic SDG tool capable of producing synthetic data based on historical water consumptions. To ensure that the applicability of the developed SDG tool is not limited to the water consumption profiles of Skiathos for which it was developed, we further tested its robustness by generating data for the Polish case study that exhibited different water consumption profiles.

The two case studies are very different in terms of socio-economics, demographics, climate and geography, factors that are all expected to influence water consumption patterns. Skiathos is an island with high seasonal touristic activity and seasonal weather variability (Kofinas et al., 2014; 2016; Mellios et al., 2015). However, for the Skiathos case study, water sensors are installed at typical non-touristic households, where the impact of tourism and weather variables is not that

significant. It should be noted that in Skiathos, the sensors are installed at kitchen faucets, a factor that diminishes the seasonal weather impact on water consumption, in contrast to water consumption in the bathroom, which is related to showers and baths (influenced by weather). A third factor to conclusively define the pattern of water consumption is that water in Skiathos is announced to have high concentrations of mercury, thus not potable—water is used mostly for washing and cleaning. Furthermore, Skiathos is a small village of about 6,000 people with small family-owned businesses, without large companies and corporations; this corresponds to a traditional lifestyle with extended families living together and stay-at-home parents or grandparents. As a result, the water use pattern differs from one of a large city with urban lifestyle and exhibits no detectable weekday/weekend water consumption variability. Water use in the Skiathos case study shows no seasonality or trend for the aforementioned reasons. Finally, pilot households in Skiathos are located in the old town where houses are often over 100 years old and amenities are limited; usually, houses are equipped with a total of 3 or 4 faucets (mostly kitchen and bathroom), with dishwashers not being a typical appliance. The significance of this is that the total household water consumption is split among these few faucets and the water consumption in the kitchen, where the sensors were installed, is a significant part of the total household consumption.

Sosnowiec is a city in the Katowice urban area with 2.7 million people with strong urban dynamics; it is a typical industrial city with heavy industries, companies and other associated economic activity. As far as the water consumption case study is concerned, Sosnowiec reveals an urban lifestyle and the corresponding water use mode is characterized by variability in water consumption between weekdays and weekends. The sensors installed at Sosnowiec are in showers, bathtubs, kitchens, appliances, balconies, etc. thus the patterns of use are expected to be more variable, offering a challenging case study for testing the developed SDG methodology. Moreover, the fact that the Sosnowiec pilots have multiple water supply points in the household, especially when compared to those in the Skiathos pilots, the total average daily consumption in each Sosnowiec water supply point is overall lower than those in Skiathos.

- *Data availability*

The data used in this article are water consumption data collected by a total of 16 households for a period of 13 months—starting from February 1st, 2015—in two locations: Skiathos, Greece (10 sampling points—faucets—each one corresponding to a different household) and Sosnowiec, Poland (9 sampling points—faucets and appliances—in 6 households). The water consumption monitoring system was installed in a diverse group of households that were specifically chosen in order to provide needed data to help comprehend human behavior and water consumption patterns by different users in a household in various socio-economic settings. The criterion for the selection of the households was the availability and promptness of the housekeepers. Additionally, the

households were chosen so that they were diverse, regarding their location in the network and number of occupants. Wireless sensors were installed in various sampling points in the households, i.e. faucets, washers and showers. 30-second step records were transmitted to a remote central server in real time. Technical details on the water consumption monitoring system are provided in Chen et al. (2015). The data used are available online, at validation.issewatus.eu/data-re-use/.

- *Data description*

The data sets refer to water flowrate values that are recorded every 30 sec of water consumption. Once there is flow in the monitored faucet/appliance, the sensor generates a record with the corresponding timestamp; at the end of the 30-sec period, it records the water consumption (in liters/min) during the 30-sec period. If the faucet/appliance is still on, the next timestamp is recorded and then the corresponding consumption, and so on. Every 30 sec, the sensor checks for flow and when the faucet/appliance is off, no record is produced, thus finalizing the creation of a water consumption incident record. When the next water consumption incident starts, the procedure repeats itself. For the period in-between the two incidents, no record is produced. All incidents themselves have a 30-sec time step, but the starting time of an incident might be for example 45 seconds after the previous one. This means that sensor-produced data are not in the form of a single time series with a 30-sec time step, but are recorded in the form of numerous clusters each one representing a small time series of the equivalent incident. In reality, in order to distinguish between incidents, one detects when the time distance between two consecutive records is greater than 30 seconds, as shown in Figure 47a. The number of records per incident is used to calculate its duration.

- *Method description*

An algorithm that would generate flowrate records for a household water supply point should simulate all meaningful qualitative and quantitative characteristics of the actual records: the number of incidents per day, the duration of each incident, and the time of the day most likely for an incident to occur and the flowrate of the event, at a time scale which fits the objective of the model. An additional criterion to such a simulation is that when summing all simulated consumptions, the total water consumed should follow closely the actual total water consumption, both during the day and during the year as a whole. Components of the overall time-series, such as trend and seasonality, if any, should be traced as well.

The methodology developed for building the flowrate time series table follows a stochastic approach, agnostic to behavioral or exogenous factors that satisfies each one of the aforementioned characteristics. It is extensively described throughout the following steps of phase 1 and phase 2:

Steps 1 to 4 correspond to the calibration and distribution fitting (phase 1)

Step 1- preprocessing: The procedure starts with the transformation of raw data into a time series. For the time-series mode, a 30 sec time step is kept constantly. Time continuity is not interrupted when there is no record; on the contrary, time periods when the faucet is off are denoted with the phrase «no record». In order to keep the same time frame for all incidents the starting point of each incident is moved to the next 30-sec time step, thus transforming all time stamps to time steps in the produced time series (Figure 47).

Step 2- creating incident tables: Tables are created for each day of the recorded data, in which time steps with a recorded flow-rate value are denoted with «1» and time steps with no records are denoted with «0». This means that 1s stand for water taps being on and 0s for water taps being off (Figure 48).

Step 3- estimating binomial distribution: The binomial distribution of 1/0 is estimated for each time step across all 400 recorded days. This way, the probability of occurrence of an incident at each 30 sec time step is defined (Figure 49). At this point, it is noted that the probability of occurrence throughout the day is indicative of the consumption pattern for each household faucet, which means that rush-hours in terms of water use are expected to behold higher probabilities of a water incident, consistent to the combined day routines of the householders. In Figure 50a and Figure 50b, two consumption profiles in two households are presented. The specific consumption profiles are chosen to be presented, among all, because they indicate distinctive differences. The first profile shows 1 peak in the afternoon around 2 pm, while the second shows two peaks, one in the afternoon and one in the evening. Moreover, the second profile shows throughout the whole day higher level of probability of an incident (faucet use). This does not necessarily mean that the water *consumption* is higher in the second profile, since the flowrates for the given incidents might be significantly higher in the first case; it only means that the faucet in the second householder is used more often than the first householder.

Step 4-investigating the flowrate values' distribution: The distribution of the flowrate variable is investigated by using the steps described in the Kolmogorov-Smirnov (K-S) methodology (Hollander and Wolfe, 1973). Popular distributions, namely gauss, gamma, exponential, and beta are tested in order to conclude on the most suitable distribution that simulates the values of flowrate that occur in each water supply point. In order to define the distribution of recorded flowrate values, *all* recorded values available of every incident are set from lower to higher value and are divided into 20 classes (highest suggested number of classes), following to the typical procedure for building a Pearson histogram (Dean and Illowsky, 2009). The frequency of occurrence for each class is estimated (Figure 51). Equivalent frequencies of a hypothetical sample, which keeps the same mean and standard deviation with the real flowrate values are estimated for the 4 tested popular distributions (gauss, gamma, exponential, and beta). The maximum vertical

distances (D) between the cumulative frequencies at the center of each class of actual data distribution and the ones tested are estimated. The minimum of these distances is considered to give the distribution to better mimic the actual distribution (Figure 52).

	hrs/min/sec	flowrate [lt/min]
incident 1	13:18:32	1.78
	13:19:02	0.70
	13:19:32	0.20
	13:20:02	0.22
incident 2	13:22:16	0.34
	13:22:46	0.20
	13:23:16	0.96
incident 3	13:59:07	0.16
...	...	...

a) raw data

	hrs/min/sec	flowrate [lt/min]
incident 1	13:18:30	1.78
	13:19:00	0.70
	13:19:30	0.20
	13:20:00	0.22
closed water tap	13:20:30	no record
	13:21:00	no record
	13:21:30	no record
	13:22:00	no record
incident 2	13:22:30	0.34
	13:23:00	0.20
	13:23:30	0.96
closed water tap	13:24:00	no record
...	...	...

b) recorded data in a time-series mode

Figure 47. Transforming raw data into a time series

	day 1	day 2	day 3	...
13:18:30	1.78	no record	no record	
13:19:00	0.70	2.12	no record	
13:19:30	0.20	2.32	no record	
13:20:00	0.22	0.78	no record	
13:20:30	no record	0.1	no record	
13:21:00	no record	0.33	no record	
13:21:30	no record	0.32	0.17	
13:22:00	no record	0.35	1.43	
13:22:30	0.34	0.33	2.64	
13:23:00	0.20	0.56	0.56	
13:23:30	0.96	no record	0.52	
13:24:00	no record	no record	no record	

a) time series of flow-rates day after a day

	day 1	day 2	day 3	...
13:18:30	1	0	0	
13:19:00	1	1	0	
13:19:30	1	1	0	
13:20:00	1	1	0	
13:20:30	0	1	0	
13:21:00	0	1	0	
13:21:30	0	1	1	
13:22:00	0	1	1	
13:22:30	1	1	1	
13:23:00	1	1	1	
13:23:30	1	0	1	
13:24:00	0	0	0	

b) time series of 1s and 0s day after a day

Figure 48. Time series tables of 1s and 0s for denoting consumption and no consumption

The following steps concern the generation of synthetic data for a 24-hr period (phase 2):

Step 5-generating table of incidents: A table of 1s and 0s is produced through a random generator following the binomial distribution estimated for each time step (Step 3). This way, a number of 30-sec records is produced consistent to the probability of incident occurrence for a household water supply point. Naturally,

more 1s are expected to be generated during the "rush hour" and more 0s when faucet use is low.

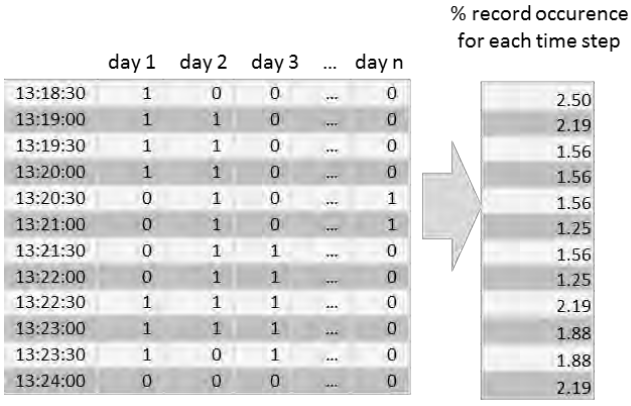


Figure 49. Estimating the binomial distribution of water use incident for each 30 seconds period

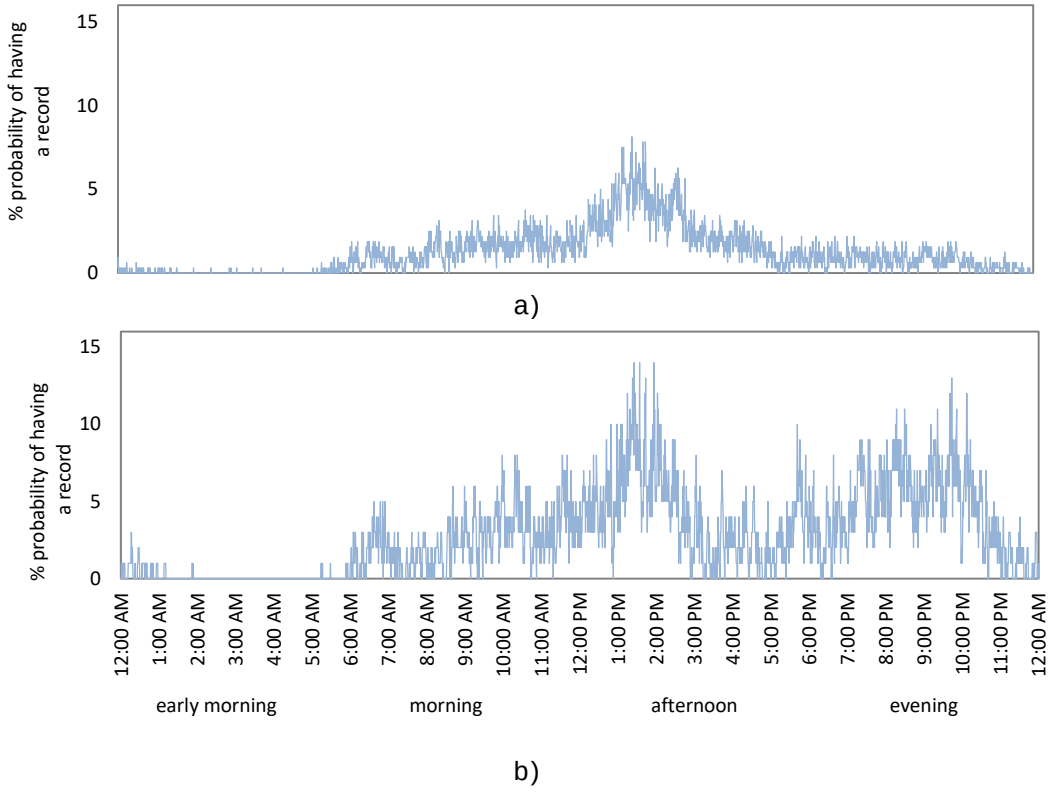


Figure 50. Distinguishing different water use patterns through the probability of faucet use diagrams in two examples: a) water use probability for Skiathos pilot 1 shows one peak around 02:00 pm and maximum probability of a water use incident to occur up to 8.15 %, b) water use probability for Skiathos pilot 2 shows two peaks around 02:00 pm and 10:00 pm and maximum probability of a water use incident to occur up to 14 %.

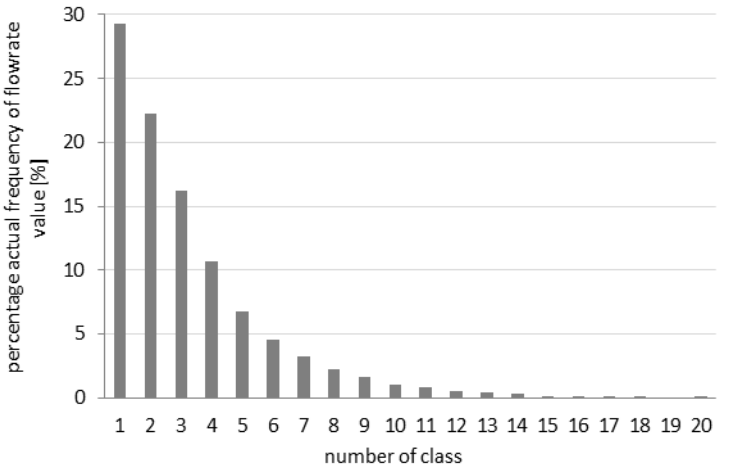


Figure 51. Distribution of frequency of water flow rate of Skiathos faucet 1

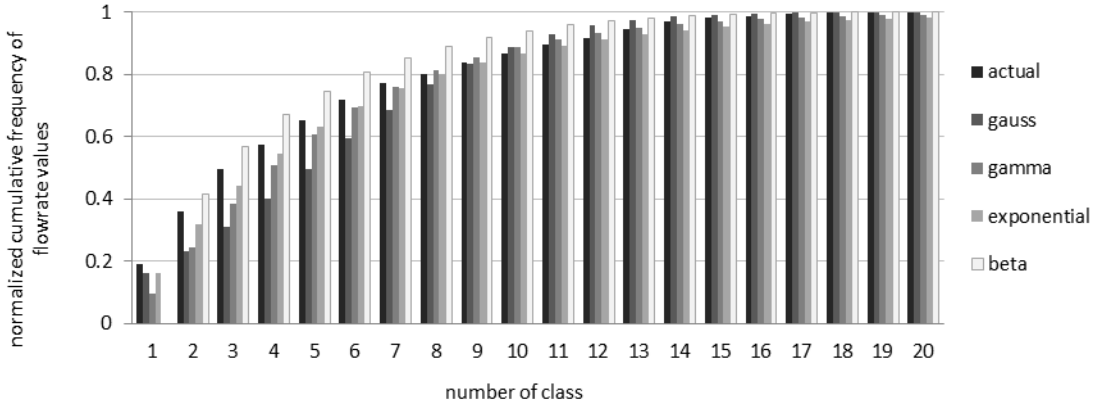


Figure 52. Identifying the most suitable distribution that would best simulate actual distribution of water flowrate. The following distributions are considered: gauss, gamma, exponential and beta for faucet 1.

Step 6-filling incidents with flowrate values: The time steps with 1s (Step 5) are now filled in with values (records) generated by a random generator following the flow-rate distribution as it was derived in Step 4. This way, main selected statistical characteristics, such as average level, standard deviation and distribution of simulated flowrates, are kept consistent with actual flowrates, since they are based on all flowrates in the sample. The same is true for the number of 30-sec records that the faucet was on. Step 5 and 6 could actually be merged into a single step by using a joint probability distribution of the two variables (incident occurrence and flowrate level). However, this is not applied in the present work. Keeping the steps separate helps having a more detailed overlook of the procedure and specify its weaknesses and strengths in parts. At this point, the simulated total average daily consumption is also kept consistent with actual data. The problem that remains is that generated 30-sec records are mostly isolated and widely

spread throughout the day and are not properly clustered to simulate the observed "water consumption incidents". This is addressed in the next step of the algorithm.

Step 7- clustering incident values (independent pulses): The way the algorithm is set up (Steps 1 through 6), it randomly produces 30-sec records throughout the day; these records are mostly isolated and not clustered in continuous events—water consumption incidents—that last a few minutes, for example. It is possible, although unlikely, that the algorithm will randomly produce two or more 30-sec records in a row, creating a simulated incident. Our intention is to reproduce not only the cumulative water consumed in a day, but also to reproduce the number of water consumption incidents (in a sense, the number of times the faucet is used during the day). To achieve this, we created a "clustering step" (Step 7) that clusters generated isolated records into realistic multiple time-step incidents. We use the law of inverse square distance (Newton, 1999) in order to produce these clusters. Specifically, each incident (either an isolated record, or multiple continuous records) generated is treated as a "particle" with attribute level equal to its flowrate. Particle will be considered an incident that is generated and is followed by a "no record", while its attribute will be the sum of all flowrates (if it comprises more than one record). The number of time steps between incidents is considered to be the distance between the particles. For every pair of consequent incidents, of m_1 = flowrate$_1$ and m_2 = flowrate$_2$, which have a distance of d= number of time steps, the "attractive forces" are calculated for *all* neighboring particles throughout a 24-hr period, according to the inverse distance square law (equation 35). The highest "force" drives the first "particle movement", stacking together the two neighboring particles on which the highest force is exerted and creates a new particle (cluster) keeping the number of original time steps the same. The newly created particle is placed in the timeline at a location that lies in between their initial positions. This location is defined by the fraction of their flowrates so that the resulting cluster is placed closer to the larger particle, according to equation 36. The clustering procedure, shown schematically in Figure 53, iterates for the generation of a day's data until the number of clusters reaches a desired value. That value comes out by a random generator consistent to the Gaussian distribution of cluster numbers per day for the whole sample of each household water supply point so that the mean value and standard deviation of the number of clusters of a day are preserved.

$$F = \frac{m_1 * m_2}{d^2} = \frac{flowrate_1 * flowrate_2}{number\ of\ timesteps^2} \qquad \textit{(equation 35)}$$

$$\frac{m_1}{m_2} = \left(\frac{r_2}{r_1}\right)^2 \qquad \textit{(equation 36)}$$

where r_1 and r_2 are the distances (number of time steps) from m_1 and m_2 equivalently of the point the two particles will meet.

A schematical flowchart of the developed methodology, its phases and steps, is presented in Figure 54.

- *Method modification for Sosnowiec data*

We initially suspected that water consumption in Sosnowiec exhibits weekly seasonality, since the area is strictly urban-residential and the working and lifestyle routines result into different water consumption patterns from weekdays to weekends—a statement also supported in the literature (Arampatzis et al., 2014). In Figure 55, we show a sample of water consumption patterns for a household in Sosnowiec, where we see that the two probability-of-having-a-record functions are different. Further investigation into this is presented in the Results and Discussion section, where we show that indeed water consumption shows weekly seasonality. Accordingly, the algorithm was modified to reproduce results consistent with this variable pattern: Initially, data were separated into weekday and weekend observations. Steps 1 and 2 were kept the same, while Steps 3 and 5 were modified to implement separately weekday and weekend binomial distributions. As a result, separate data series for weekdays and weekends are generated by different incident binomial distributions. Step 4 does not change, since we assume that the user will not modify faucet use (he/she will not turn the faucet on to a higher setting, for example) depending on the day of the week; therefore we used the same flowrate level distributions for all days of the week, just like the Greek case study. The approach with a separate dataset for each day of the week can be adopted, once longer time series of data will be available. The weekday/weekend-pattern approach is supported against the basic (no pattern) approach by a trial and error process. This means that data were generated with both approaches and the weekday/weekend approach gave better fitting according to the validation procedures as described next. Distinguishing seasonality by observing the week and weekend profiles of the households consumption is not always conclusive, since in most of the cases the mean and its high and low intervals are the same, but the diffence may lie within the distribution of the same consumption timely. The inspection of possible seasonality within the week, is prerequisite for the model to perform optimally and it enhances the generality of the model for its application in seasonal and non-seasonal consumption patterns.

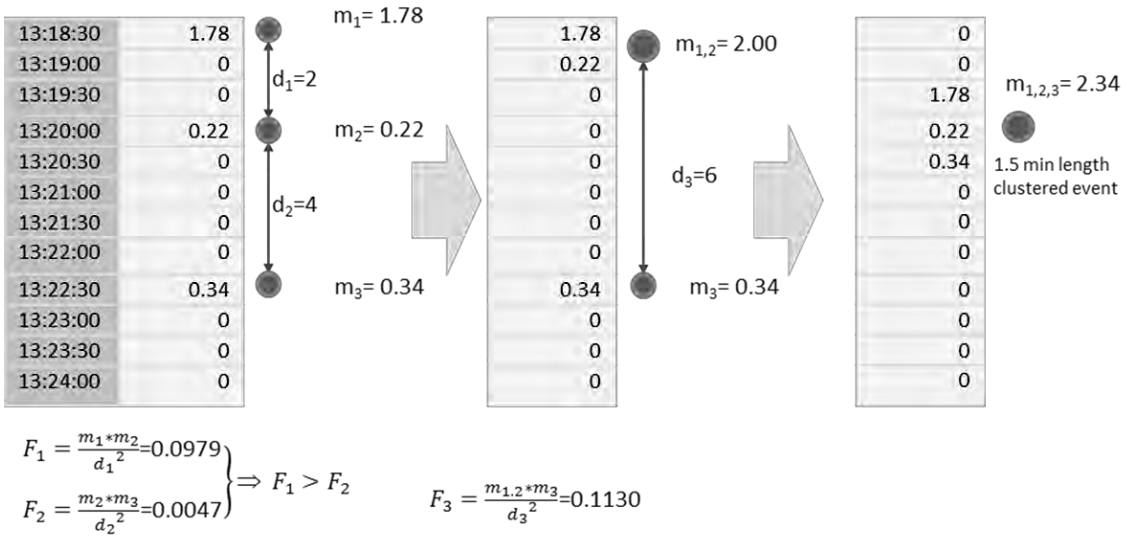


Figure 53. Routine for clustering the produced water use incidents according to the source data clustering patterns

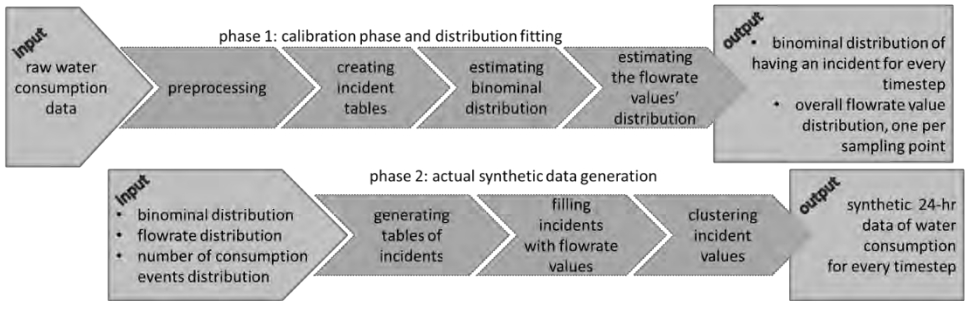


Figure 54. Flowchart of methodology's phases and steps.

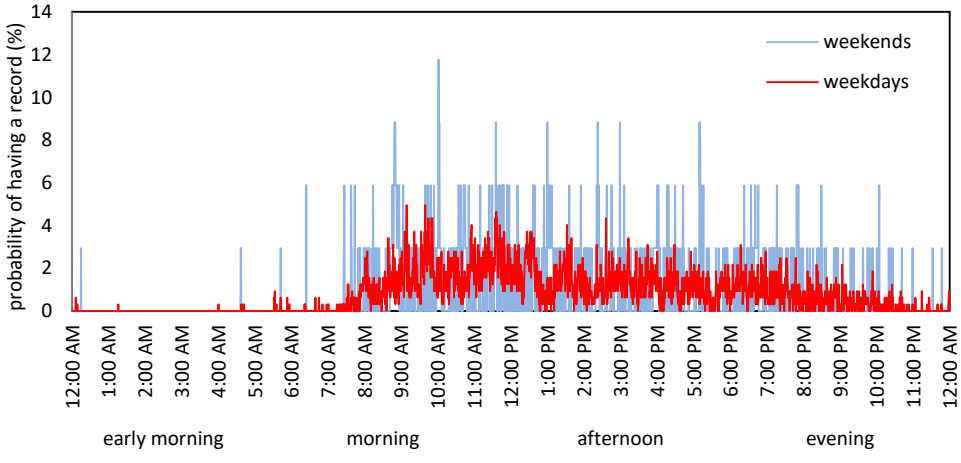


Figure 55. Indicative comparative diagram of weekend and weekday pattern of water use for Sosnowiec water supply point 1

❖ *Validation methodology*

For validating the methodology, data sets for one year of water consumption are generated for each water supply point, for both case studies. For these data sets, we produced curves of the probability of occurrence of an incident for both simulated and actual consumptions; this way we validated the "incident occurrence" variable. Next, the generated flowrate values are validated by checking the fitting in simulating actual average flowrates. An important criterion for validating the methodology is based on the difference between total water consumptions generated by the SDG algorithm and actual water consumptions recorded for each pilot; the generated data should neither overestimate nor underestimate significantly the total water consumption. This difference in cumulative consumptions is calculated for different time scales, so it is not only calculated for the whole simulating period; additionally, we divide the day into 4 quarters—early morning (24:00-06:00), morning (06:00-12:00), afternoon (12:00-18:00) and evening (18:00-24:00)—and calculated cumulative consumptions for these quarters. This way, we group consumptions and check the algorithm ability to capture consumption patterns throughout the day. The fraction of average quarterly consumption relative to the average daily consumption is also considered as a critical indicator for capturing the water consumption distribution throughout the pilots' 24 hour periods. Methodology validation steps and metrics used are presented in Table 23.

Table 23. Methodology validation steps and metrics

validation metric description	metric formulation
qualitative presentation of the ability of SDG to mimic the water consumption patterns (Figure 56 and Figure 57)	comparative diagrams of actual and generated 400 days average flowrate for Skiathos and Sosnowiec pilots
qualitative presentation of the ability of SDG to simulate the flowrate values of each 30-sec (Figure 58)	scatterplots of average generated and actual flowrate values of each 30-sec step for Skiathos and Sosnowiec pilots
ability of SDG to simulate the	R^2 values of generated and actual values of the probability of having a record of each 30-sec

probability of having a record of each 30-sec (Table 25)	
ability of SDG to simulate the flowrate values of each 30-sec of the pilots (Table 25)	R^2 values of generated and actual values of the flowrate values of each 30-sec
water balance preserved through percentage differences between 400 days generated and actual water consumption (Table 26)	$\frac{Total\ generated\ volume\ -Total\ actual\ volume}{Total\ actual\ volume} * 100\%$
difference between 6 hour percentage quota actual consumption and 6 hour percentage quota generated consumption (Table 27)	$\frac{1}{400} * \sum_{1}^{400} \left(\frac{actual\ midnight\ to\ 06{:}00\ cons.}{actual\ daily\ consumption} - \frac{gen.\ midnight\ to\ 06{:}00\ cons.}{generated\ daily\ consumption} \right) * 100\%$ $\frac{1}{400} * \sum_{1}^{400} \left(\frac{actual\ 06{:}00\ to\ noon\ cons.}{actual\ daily\ consumption} - \frac{gen.\ 06{:}00\ to\ noon\ cons.}{generated\ daily\ consumption} \right) * 100\%$ $\frac{1}{400} * \sum_{1}^{400} \left(\frac{actual\ noon\ to\ 18{:}00\ cons.}{actual\ daily\ consumption} - \frac{gen.\ noon\ to\ 18{:}00\ cons.}{generated\ daily\ consumption} \right) * 100\%$ $\frac{1}{400} * \sum_{1}^{400} \left(\frac{actual\ 18{:}00\ to\ midnight\ cons.}{actual\ daily\ consumption} - \frac{gen.\ 18{:}00\ to\ midnight\ cons.}{generated\ daily\ consumption} \right) * 100\%$

5.1.3. Results and Discussion

Before the actual generation of consumption data, and in order to perform Step 4 of the algorithm, the distribution of the flowrate variable is investigated. Gaussian, exponential, gamma and beta distributions are examined in order to identify the one that describes best the probability distribution of water consumption values. For each data set, one of the four distributions emerges as closest to the actual one. In Table 24, it is shown that almost for all pilots and for both cities, the exponential distribution is the closest one. Only or few water supply points, namely Skiathos 1, 6 and 10, gamma distribution was proved to be the most suitable; however, improvement over the exponential distribution was only marginal. Therefore, in order to keep the algorithm uniform, we selected the exponential distribution for generating actual flowrate values for both cities. However, we do not suggest generalizing that the probability distribution of any faucet flowrate would be exponential, since different kinds of fixtures, such as the ones operated by foot, or users with different habits, mentality or physiology, might produce values of another distribution. The flowrate value depends on the way the faucet is operated: some users turn it on all the way to the highest setting, some let it drip slowly, while different fixtures could affect this value as well: sink or bath sets with separate levers for hot and cold water will probably have different flowrate profiles than mixer faucets. This kind of distribution investigation is suggested as a pre-processing step for dealing with a new household faucet tap, before implementing the data generation algorithm.

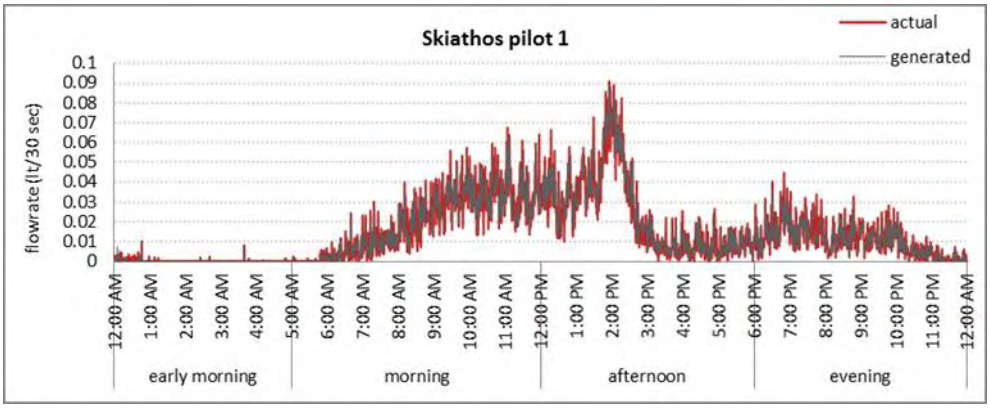
Table 24. Best and 2nd best fitting distributions for each pilot flow rate and the maximum vertical distances (D) between the cumulative frequencies at the center of each class of actual data distribution and the ones tested, respectively

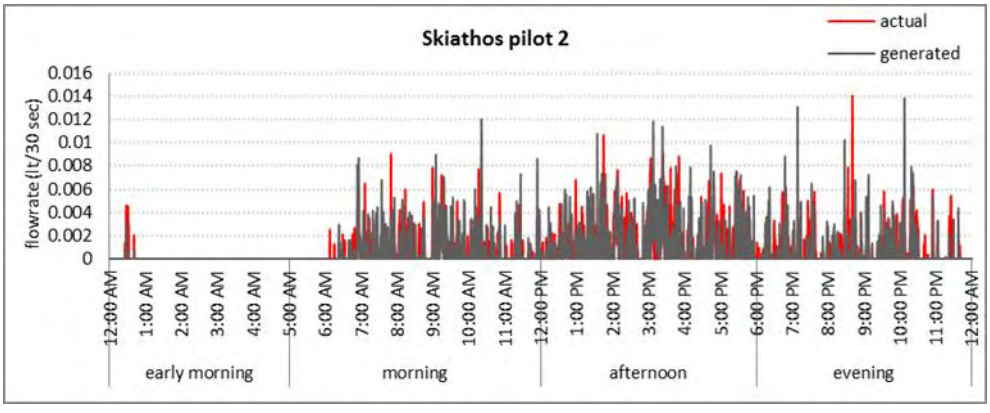
	pilot	best fitting distribution	minimum D	2nd best fitting distribution	2nd minimum D
Skiathos case -study	1	gamma	0.101	exponential	0.111
	2	exponential	0.041	gamma	0.099
	3	exponential	0.051	gamma	0.116
	4	exponential	0.095	gamma	0.013
	5	exponential	0.089	gauss	0.122
	6	gamma	0.099	exponential	0.102
	7	exponential	0.041	gamma	0.132

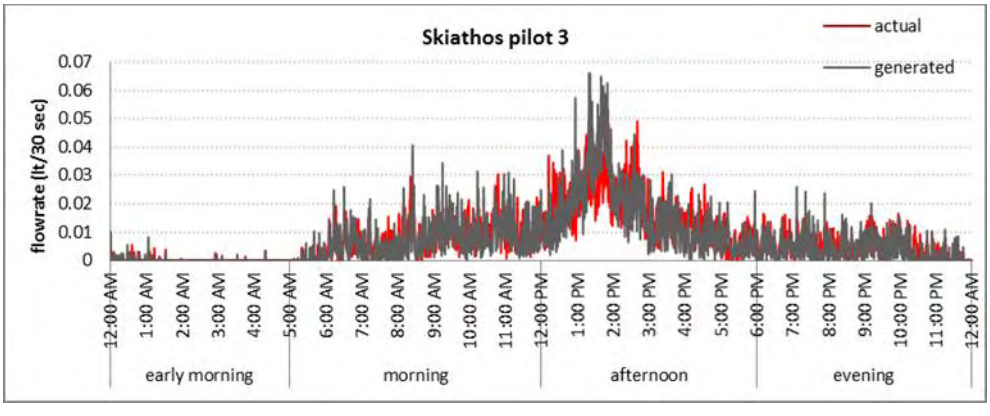
	8	exponential	0.078	gamma	0.087
	9	exponential	0.068	gamma	0.150
	10	gamma	0.101	exponential	0.113
Sosnowiec case -study	1	exponential	0.038	gamma	0.102
	2	exponential	0.084	gamma	0.116
	3	exponential	0.072	gamma	0.101
	4	exponential	0.047	gamma	0.098
	5	exponential	0.065	gamma	0.117
	6	exponential	0.040	gamma	0.095
	7	exponential	0.047	gamma	0.103
	8	exponential	0.045	gamma	0.099
	9	exponential	0.059	gamma	0.151

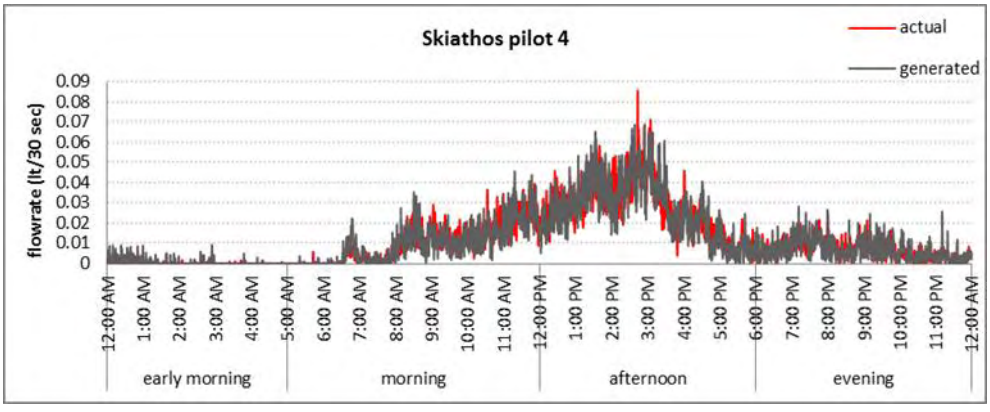
The SDG algorithm is used to generate water consumption data for about one year for each of the 10 Skiathos and 9 Sosnowiec pilots. In Figure 56 and Figure 57 comparative curves of average flowrate for actual (denoted with red) and generated (denoted with grey) data are presented. It is apparent that household consumption patterns are quite variable, with some of them exhibiting one or two distinct peaks throughout the day and some showing no distinct peak at all. It is also apparent that the suggested methodology can simulate quite successfully the water consumption pattern of each household. Not only does the method capture the peaks of the household pattern and the usual early morning very low or even zero consumption, but it also captures slight variations in microscale granularity.

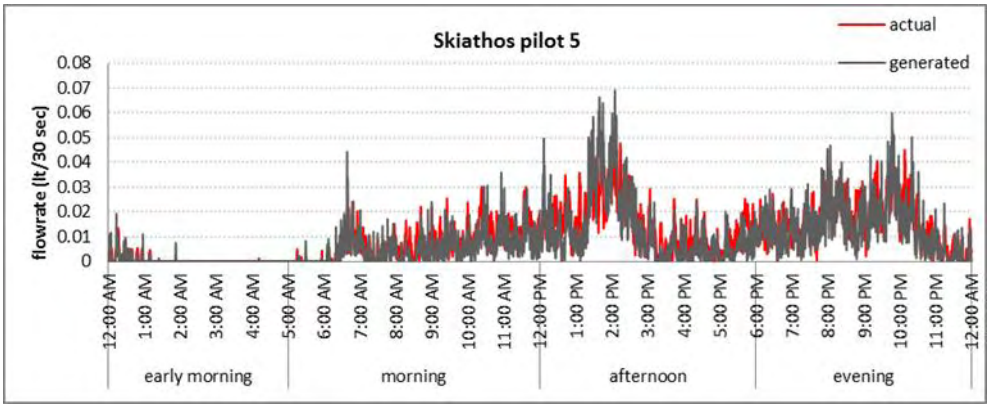
For a quantitative comparison of actual and generated data, calculations are made for both actual and simulated values for the whole data set (i) the probability of having a record at each 30-sec time step and (ii) the average flowrate for each 30-sec time step, including the zero values that correspond to a "no-record". Scatter plots of generated and simulated data are then produced for (i) and (ii) and the R^2 values are calculated. All scatter plots of flowrate values are shown in Figure 58 while a summary table, including the R^2 of the probability-of-having-a-record variable, is also provided (Table 25).

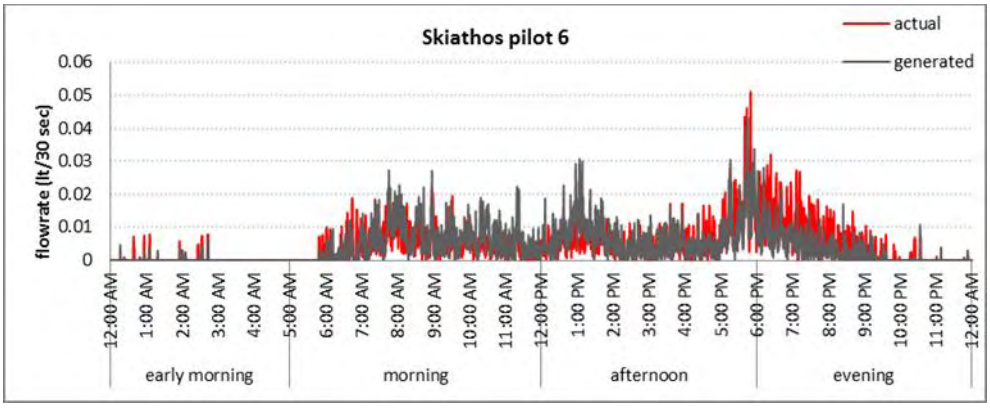


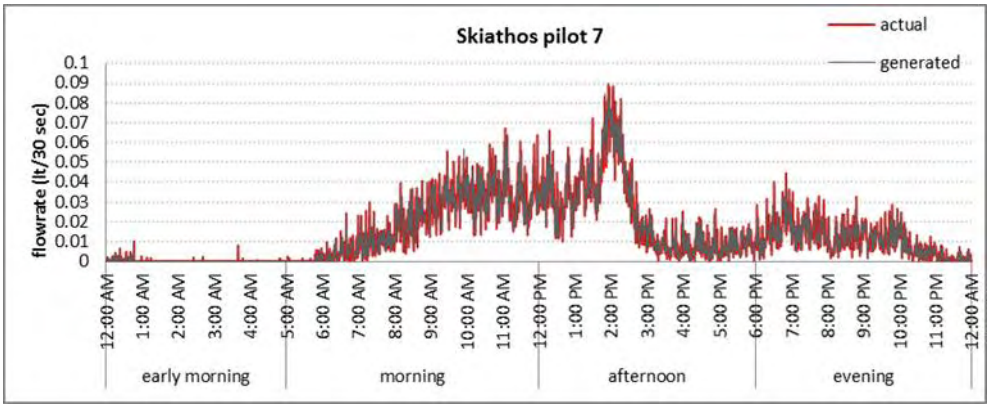


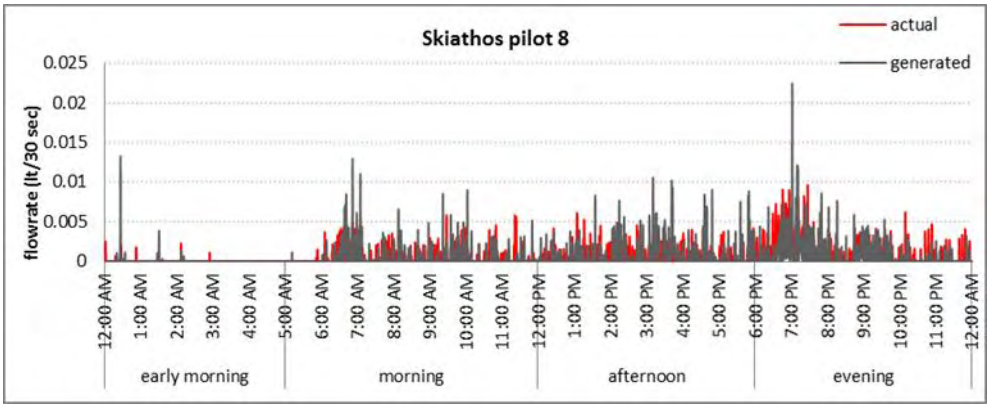


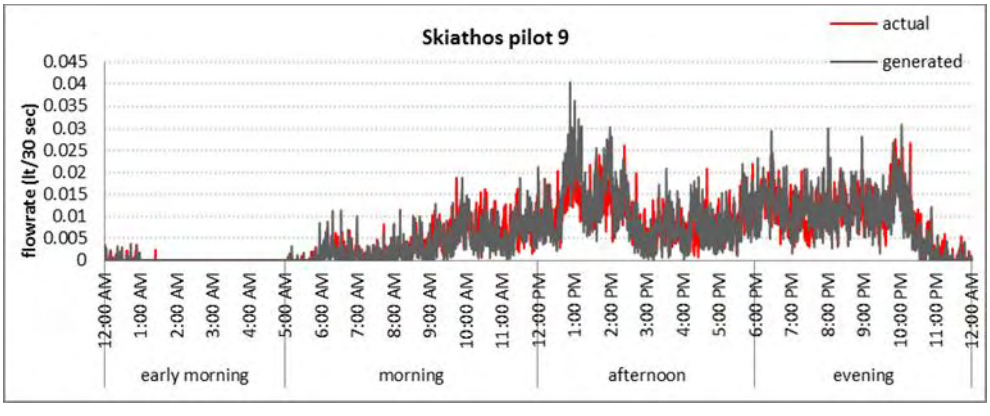












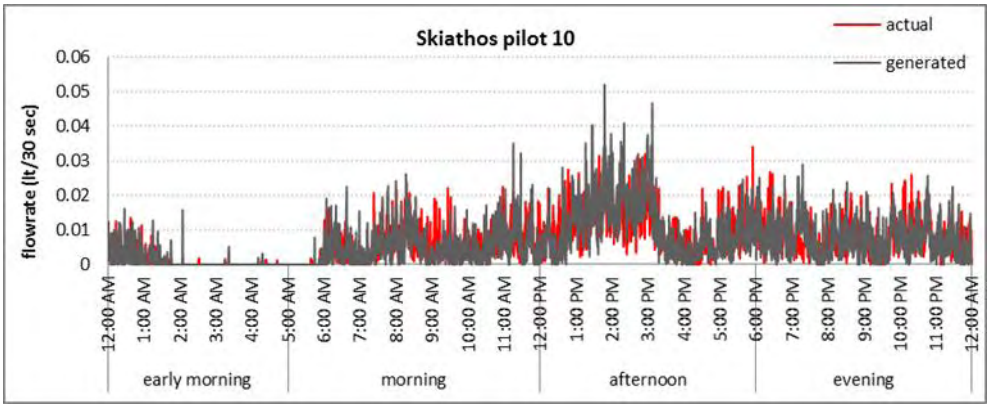
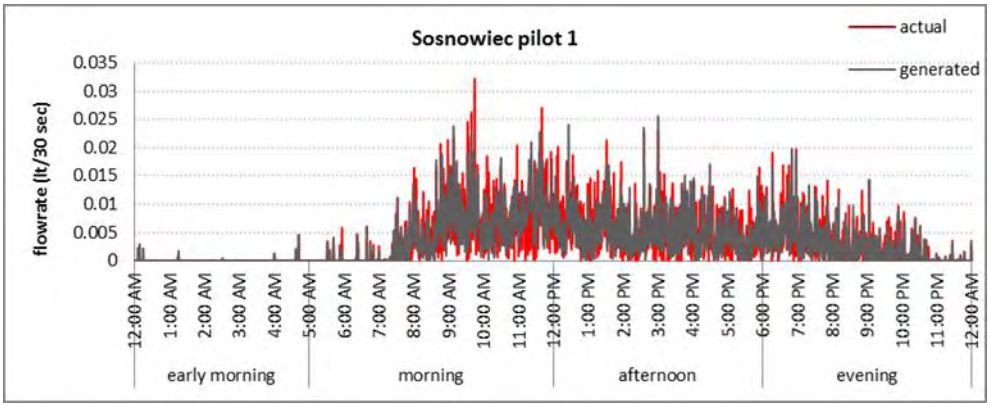
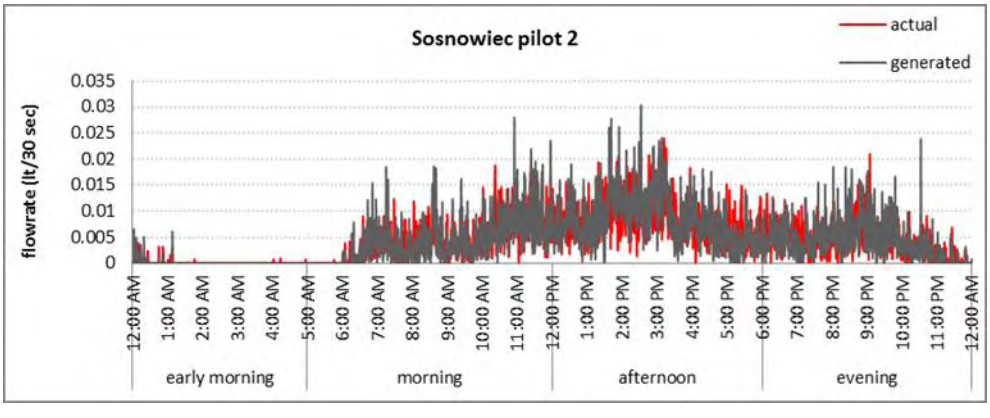
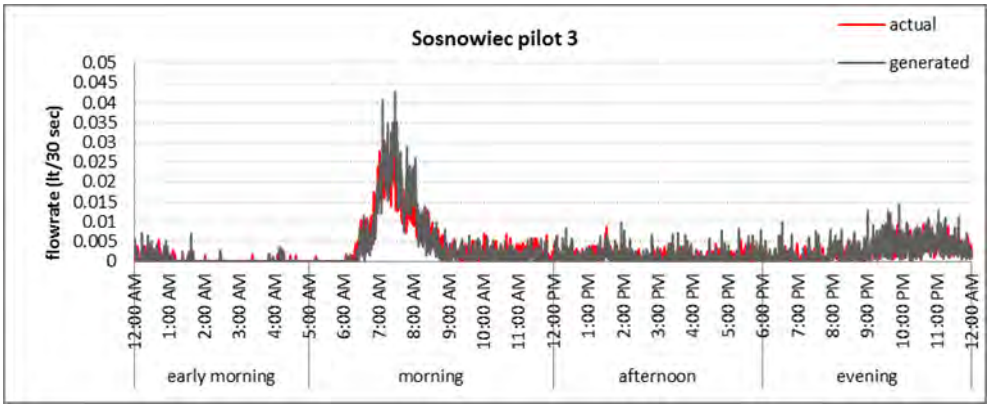
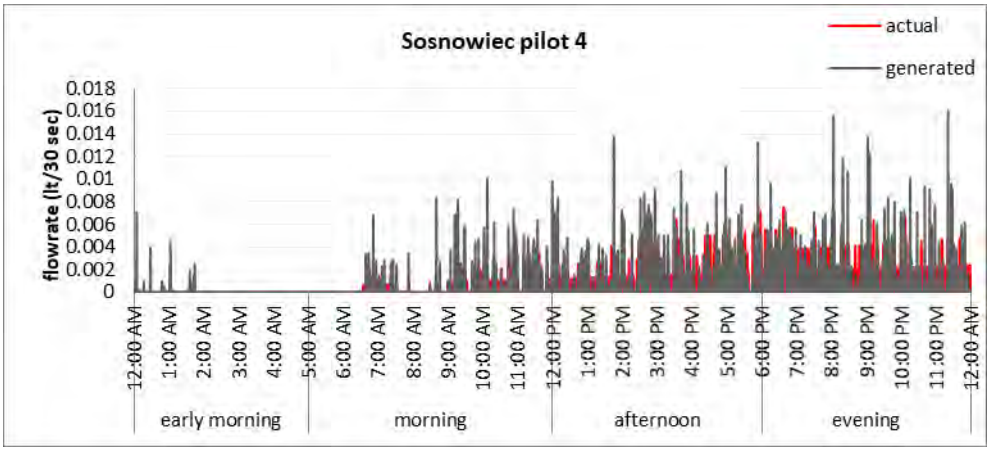


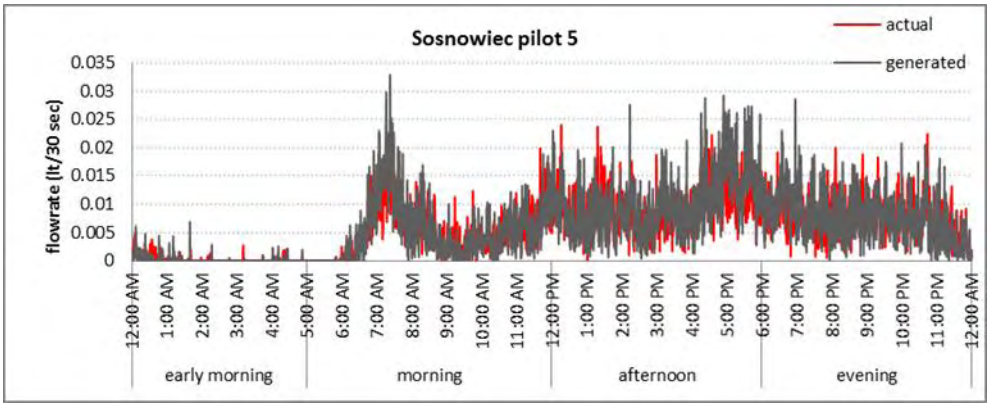
Figure 56. Comparative diagrams of actual and generated 400 days average flowrate for Skiathos pilots 1-10

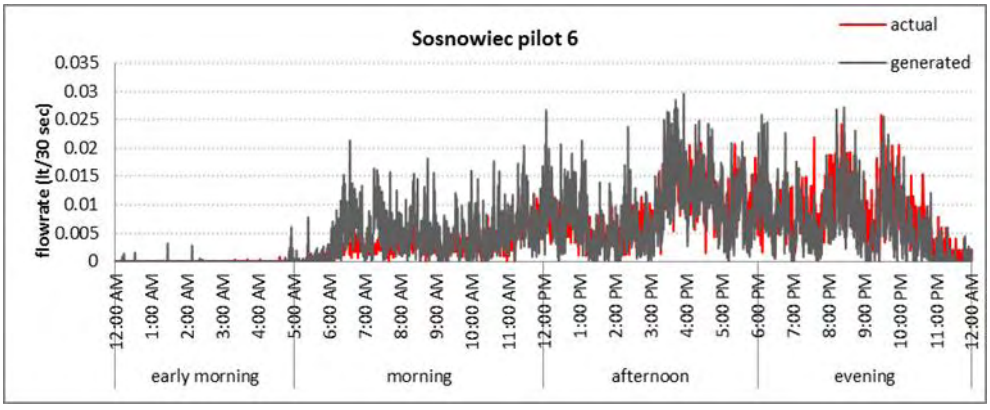


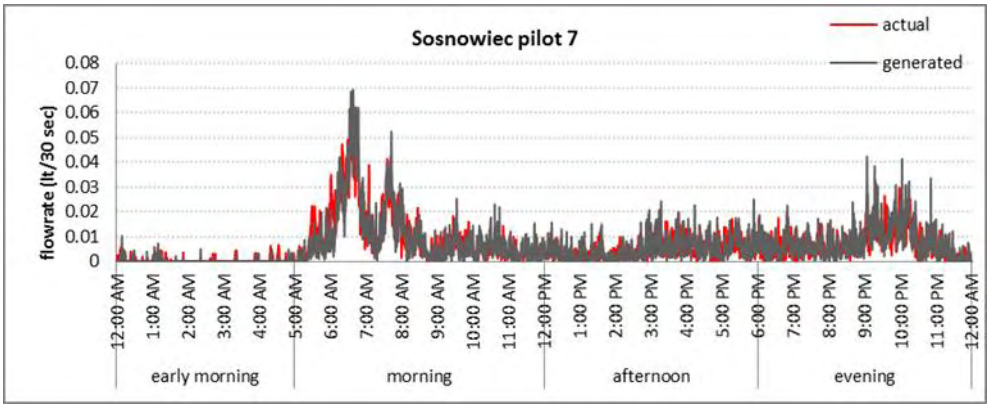


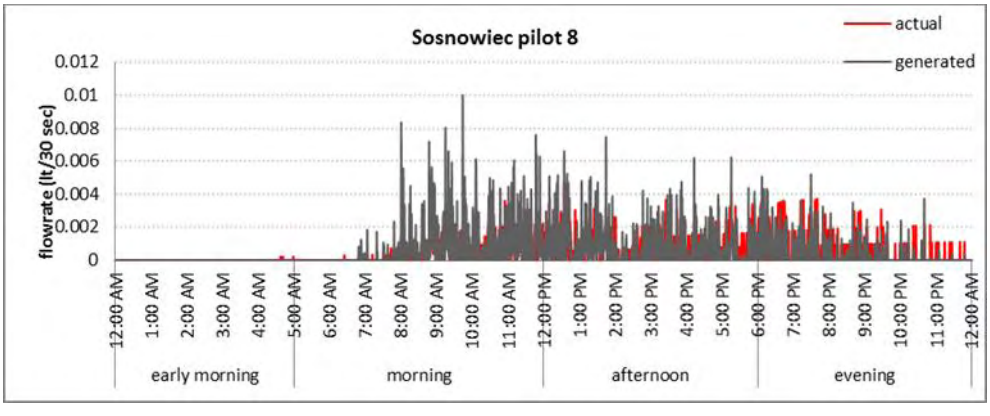












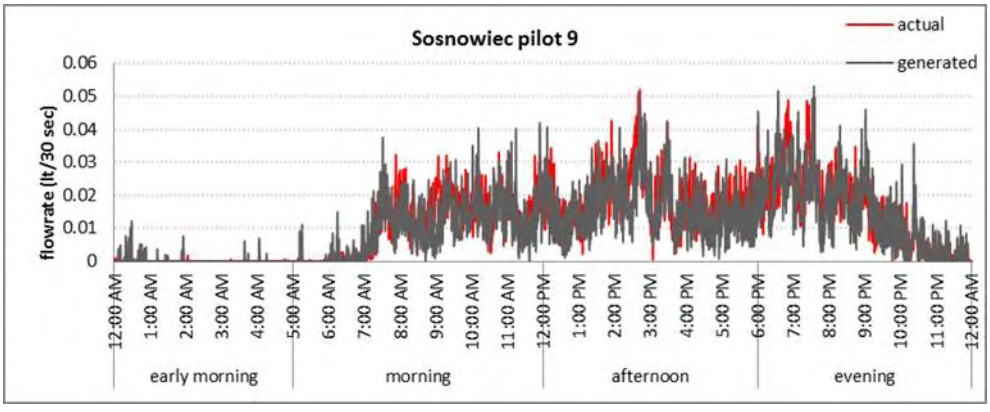
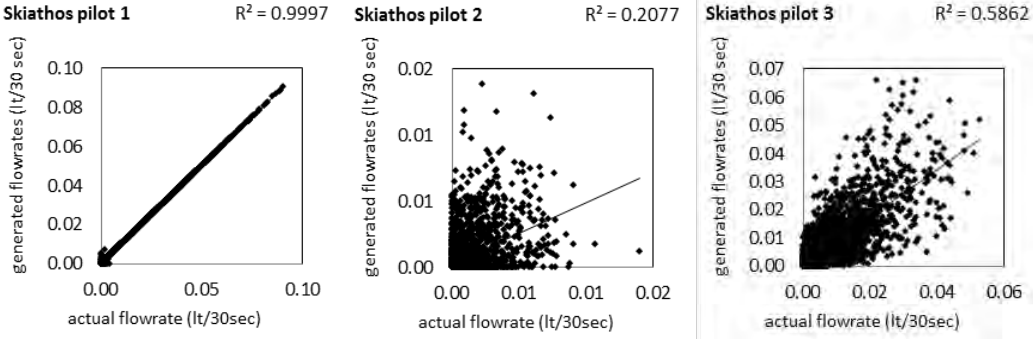
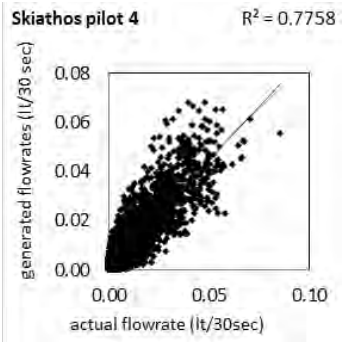
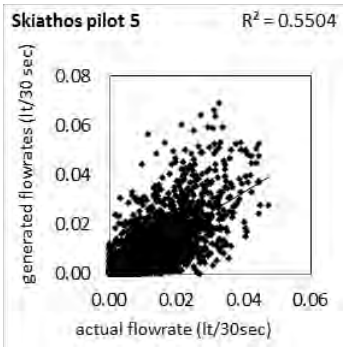


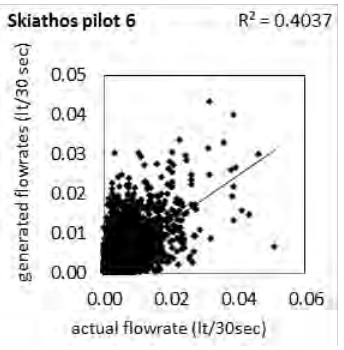
Figure 57. Comparative diagrams of actual and generated 400 days average flowrate for Sosnowiec pilots 1-9

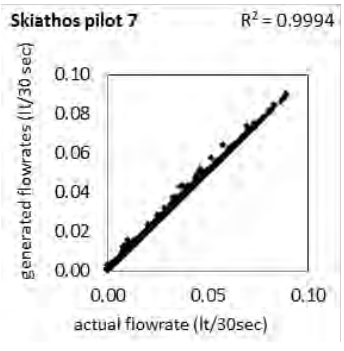
It should be noted that the R^2 values are indicative of a very detailed resemblance, a 30-sec granularity resemblance; thus, it is a quite strict metric for the validation of the method. Skiathos R^2 probability-of-having-a-record values are very high; most of them over 0.80 and two of them even reaching 1.00. Two values are relatively low, 0.37 for pilot 2 and 0.30 for pilot 8, even if the corresponding comparative diagrams do not show any particular inconsistency pattern between actual and simulated. This is due to the fact that consumption values for those houses are lower than those of the other households in the Greek case study. For Sosnowiec, the performance of the model in capturing the record probability—or faucet/appliance use during the day—is high with R^2 values higher than 0.70 except for pilot 7 which has a lower R^2 equal to 0.43.

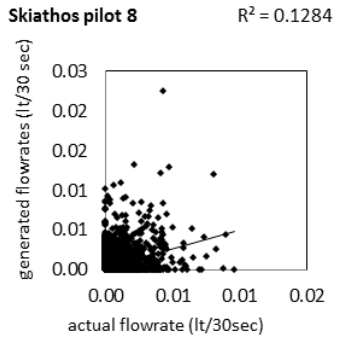


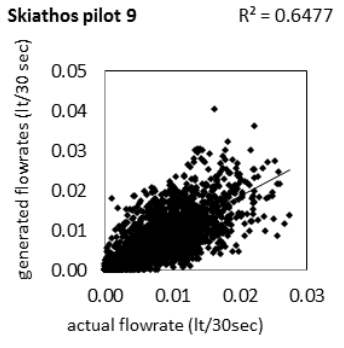


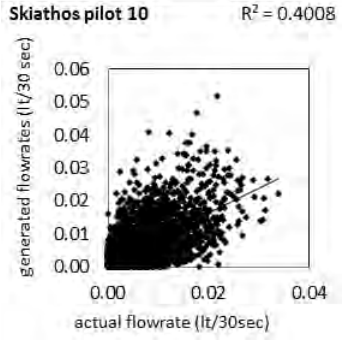


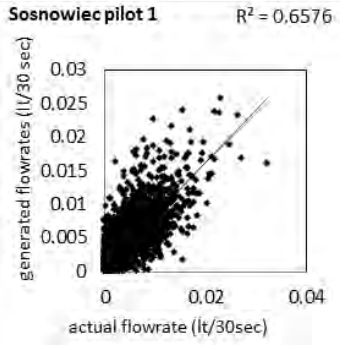


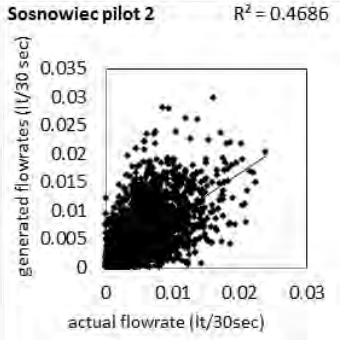


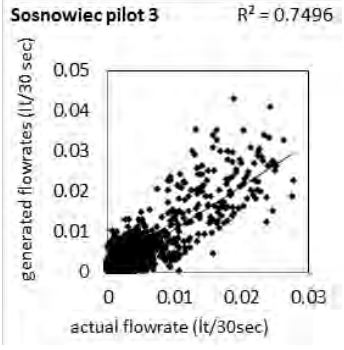


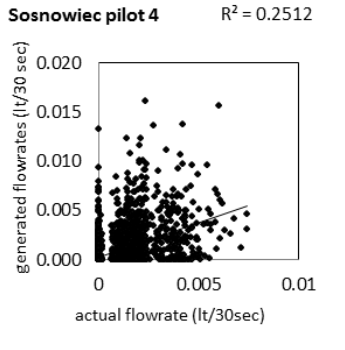


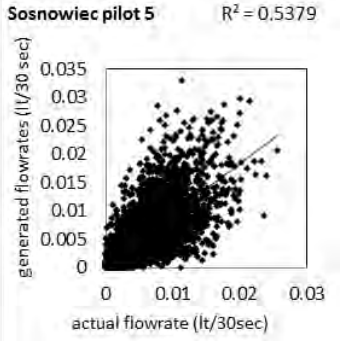












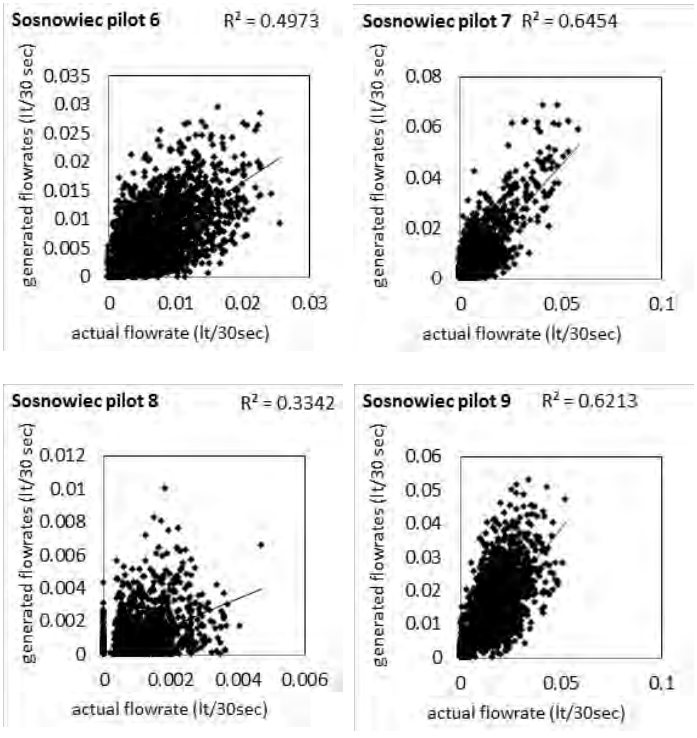


Figure 58. Scatterplots of average generated and actual flowrate values of each 30-sec step for Skiathos and Sosnowiec pilots

Generally, however, the method for Sosnowiec performs slightly worse than for Skiathos. This is because the method performance seems to be related to the frequency of incident occurrence. As described in the *case study* section, Sosnowiec households have more water taps and as a result, total household consumption is divided among more water supply points than in Skiathos. As a result, each water tap is used less frequently when compared to Skiathos, resulting in less data overall. In Table 25 this relevance can be denoted by comparing the incident frequency in incidents/month for Skiathos and Sosnowiec. It is also quite reasonable that in Skiathos pilots, the ones performing relatively poorly are the ones with very low use frequency, namely pilot 2 with 129 incidents per month and pilot 8 with 153 incidents per month.

Lower but still high performance results are derived from the flowrate R^2 values (Table 25). Most of the R^2 values (13 out of 19) are higher than 0.5 and go as high as 0.99 or 1.00 for the case of Skiathos pilots 1 and 7. Lower performance for this variable is reasonable, since its performance is contingent on the performance of the probability of having a record: in order to capture the exact average flowrate value, one needs to first capture the probability of having a record and then produce a realistic flowrate value. If the algorithm predicts a low probability of having a record at a time step, then we expect it to produce multiple 0 values (no

records), which will bring the average flowrate value down to lower values, thus reducing the performance of the flowrate variable. As expected, the pilots that performed more weakly for the probability-of-having-a-record variable are the ones that also performed relatively weak for the flowrate variable.

Table 25. Actual incident frequency and R^2 values for actual and generated incident probability and average flowrate for Skiathos pilots

	pilot	incidents/ month	probability of having a record	flow rate
			R^2	
Skiathos case study	1	1784	1.00	1.00
	2	129	0.37	0.21
	3	997	0.80	0.59
	4	2086	0.89	0.78
	5	998	0.72	0.55
	6	1015	0.73	0.44
	7	2381	1.00	1.00
	8	153	0.30	0.13
	9	1682	0.83	0.65
	10	4416	0.67	0.40
Sosnowiec case study	1	650	0.80	0.66
	2	845	0.70	0.47
	3	679	0.74	0.75
	4	414	0.77	0.25
	5	1237	0.89	0.54
	6	1249	0.57	0.50
	7	988	0.43	0.64
	8	187	0.78	0.33

9	1682	0.77	0.62

The SDG algorithm generally performs slightly worse for Sosnowiec than for Skiathos pilots, regarding the R^2 values. An important reason for this difference is the weekday/weekend routine, which, although necessary, it creates the issue of less data being available for the algorithm, since it splits the sample in two. Statistical measures that describe the incident binomial and flowrate exponential distributions come out of a lower number of samples overall. Additionally, and for the same reason, it is observed that rarely used water taps make nosier flowrate curves. We expect that higher use frequency of a faucet will lead to more data being collected for that pilot, thus leading to smoother average flowrate curves. Finally, Skiathos consumption profiles depict more characteristic signatures easier for the algorithm to capture.

In Table 26 the percentage differences of generated minus actual water volumes for all simulated days are presented; this comparison is done for validating the ability of the model to simulate cumulative water consumptions for each pilot. For most of the pilots, the difference is kept really low, lower than 3%. We can see that this percentage shows sometimes slight overestimation and sometimes slight underestimation of total water consumption, which means that the method in total is not showing some biased tendency to go in either direction. The higher differences are shown for Skiathos water pilot 3 (17.76%) and 6 (23.39%) and for Sosnowiec water pilot 9 (12.6%), while very low differences, less than 1 % are achieved for 5 out of 19 pilots. The overall better performance of Sosnowiec pilots, regarding percentage differences, is more deceptive than indicative of the actual comparative performance. This is because the flowrate average levels in Sosnowiec pilots are generally significantly lower than those of Skiathos pilots, thus resulting to higher apparent percentage differences that, however, correspond to lower actual volumes of water.

Table 26. Percentage differences between 400 days generated and actual water consumption (%)

pilot	1	2	3	4	5	6	7	8	9	10
Skiathos	0.01	6.43	17.76	-0.11	2.6	23.39	1.69	7.88	-7.12	1.40
Sosnowiec	0.00	-2.62	-1.14	0.01	-3.42	-0.86	-1.01	-1.41	12.06	

In Table 27 the quarter percentages of actual and generated data are compared. Here, we can see, which fraction of daily water consumption is realized in each quarter of the 24-hr period. So, for Skiathos pilot 1, 0.41% of the daily water

consumption happens from midnight to 6 am, while 39.52% of it happens from 6 am to noon. In the following row, the same percentages are calculated for the generated data, while in the row below, the differences between actual and generated fractions are calculated. As it is expected, the pilots that performed highly in the R^2 values of incident probability and flowrate variables perform even better in this less strict performance indicator. We also see that pilots that performed the poorest in terms of R^2 for both incident probability and flowrate (Skiathos pilots 2 and 8 and Sosnowiec pilot 7) actually perform quite well at the quarter scale. This might be a less detailed view on the method validation; however, it is more meaningful, since it is more indicative of the method's ability to capture the usage pattern rather than each single per 30-sec flowrate value.

Table 27. Mean actual and generated fractions of daily water consumption realized in each quarter of the 24-hr period for the two case studies

	pilot		midnight-06:00	06:00-noon	noon-18:00	18:00-midnight
Skiathos case study	1	actual	0.41	39.52	41.59	18.54
		generated	0.40	39.52	41.60	18.54
		difference	-0.01	0.00	0.01	0.00
	2	actual	0.74	28.01	56.21	15.07
		generated	0.89	27.90	58.43	12.85
		difference	0.15	-0.11	2.22	-2.22
	3	actual	1.23	21.62	37.13	40.03
		generated	0.97	18.78	38.30	41.92
		difference	-0.26	-2.84	1.17	1.89
	4	actual	0.38	39.49	41.60	18.58
		generated	0.38	39.50	41.63	18.55
		difference	0.00	0.01	0.03	-0.03
	5	actual	0.43	19.20	41.68	38.77
		generated	0.41	16.26	43.96	39.44

pilot			midnight-06:00	06:00-noon	noon-18:00	18:00-midnight
		difference	-0.02	-2.94	2.28	0.67
		actual	1.30	24.20	51.16	23.34
	6	generated	0.33	27.84	50.57	21.31
		difference	-0.97	3.64	-0.59	-2.03
		actual	0.69	24.87	60.66	13.79
	7	generated	1.28	24.99	59.06	14.70
		difference	0.59	0.12	-1.6	0.91
		actual	1.42	29.68	39.74	29.21
	8	generated	0.60	38.44	44.47	16.50
		difference	-0.82	8.76	4.73	-12.71
		actual	1.48	22.50	24.50	51.86
	9	generated	1.77	24.39	33.37	40.48
		difference	0.29	1.89	8.87	-11.38
		actual	3.55	23.73	41.55	31.20
	10	generated	3.33	22.28	44.44	29.96
		difference	-0.22	-1.45	2.89	-1.24
Sosnowiec case study		actual	0.49	38.63	41.65	19.28
	1	generated	0.42	40.00	41.04	18.63
		difference	-0.07	1.37	-0.61	-0.65
		actual	0.40	28.02	46.87	24.73
	2	generated	0.35	27.51	48.49	23.69
		difference	-0.05	-0.51	1.62	-1.04

pilot		midnight-06:00	06:00-noon	noon-18:00	18:00-midnight
3	actual	3.56	58.29	13.77	24.33
	generated	2.88	58.23	11.56	27.29
	difference	-0.68	-0.06	-2.21	2.96
4	actual	0.17	11.42	38.50	49.91
	generated	1.76	17.74	37.56	42.95
	difference	1.59	6.32	-0.94	-6.96
5	actual	0.88	24.80	40.87	33.47
	generated	0.70	25.06	42.57	31.70
	difference	-0.18	0.26	1.70	-1.77
6	actual	0.41	16.55	43.52	39.56
	generated	0.66	25.75	43.02	30.65
	difference	0.25	9.20	-0.50	-8.91
7	actual	7.05	46.59	18.60	27.78
	generated	3.77	44.90	20.17	31.17
	difference	-3.28	-1.69	1.57	3.39
8	actual	0.05	30.22	43.31	26.48
	generated	0.00	42.80	40.11	17.08
	difference	-0.05	12.58	-3.20	-9.40
9	actual	0.24	28.07	39.52	32.28
	generated	0.57	27.12	38.35	34.03
	difference	0.33	-0.95	-1.17	1.75

In an overall evaluation of the method, it is deduced that the proposed SDG algorithm has the ability to perform adequately, simulating with satisfying precision

the water use consumption patterns in the household. The stochastic structure of the model allows to agnostically simulate consumption regardless of behavioral or exogenous factors. Two factors, apparently linked to the performance of the methodology, are the density of recorded incidents (number of records) and the time length of the recorded data. The method can also accommodate previously identified, regular variations such as seasonal weekday/weekend features. Annual seasonality, if any, could also be treated either by producing days from equivalent days recorded data or by estimating the seasonality curve, producing the residual data from residual recorded data (recorded data minus estimated seasonality) and then adding the seasonality to the generated residuals. The former approach is expected to be more accurate but would need much more data to be implemented than the latter approach. A possible detection of a trend component could also be dealt with the latter approach, specifically by abstracting trend by the raw data, generating de-trended residuals and then adding trend to locate the generated consumption data to time wanted.

With further investigation of the applicability of the algorithm to the de-seasonalizing and de-trending techniques, the potential of using the proposed SDG method as a *forecasting tool* is enhanced. Especially when combined with advanced water consumer clustering techniques such as Self-Organizing Maps (SOM) (Laspidou et al., 2015), a user might be able to generate tap water-data by feeding the data generator with known or easy to find parameters, even nominal, such as the size or age of the household, or the working status and ages of the residents of the household, etc.

Another possible use of the algorithm could be that of implementing a Monte Carlo type simulation of daily water consumption of all individual households in an agglomeration, or District Metered Area (DMA). This way, we can identify extreme values (lowest and highest possible) for water consumption of all households. This might be a useful planning tool that can help the water provider identify maximum possible consumptions and corresponding times of the day that they can occur, thus quantifying the risk of failing to meet water demand. Such information might be useful in identifying "weak links" in a water supply network, which could be related to reservoir or groundwater levels, network capacity, meeting required pressure levels in the network, etc. Such a direction could give a useful tool to the water security domain. This, however, needs further investigation to make any safe conclusions.

One might wonder: Could this tool be used to model water consumption in a whole city? And in that case, how many households would we need to have data from, in order to be able to reliably draw conclusions about the whole city? The answer is that the parameters of population (N) and tolerated error (e) of simulated households in a city should be used to define the size of the sample that is

representative of the population (Särndal et al., 2003), i.e. the number of households with data needed. Obviously, such a tool allows the user to use only a limited number of households in order to draw conclusions for a much larger sample. Further to that, stratified sampling (Botev and Ridder, 1997) can be used; this technique suggests that a sample can be more representative when the population is divided into strata of common defined characteristics. Clustering algorithms can be considered pre-cursors of stratified sampling. Laspidou et al. (2015a) applied the SOM algorithm to cluster water meter data of a small town into different categories, such as households, hotels and small enterprises using 3-month water consumption data. The aforementioned imply that the scalability of the model can be enhanced by the use of clustering algorithms which would define the number of representative pilots needed respectively to the number of taxa suggested. There is clearly potential in the use of such an algorithm, especially when combined with clustering techniques; given enough representative households with water consumption data of fine granularity, the possibility of expanding its use to the whole city is to be considered.

Afterword: A method for generating meaningful water consumption data is introduced. The SDG algorithm differs from past algorithms within the pulse model group because it introduces early-on the probability of having a water consumption incident at each time step as a critical variable to the simulation. That variable is modeled with the use of the binomial distribution; this way, it ensures that the consumption pattern is captured. The flowrate values are modeled with the use of the estimated frequency distributions. A reverse-square-distance-based subroutine is introduced for the reproduction of the duration variable of each event, by clustering the reproduced incidents. The challenges of such a task are to mimic the characteristic patterns of each end-use, to simulate the water flowrate levels and durations of water consumption events and to preserve the water volume consumed at a large time scale, such as that of 400 days.

The methodology is implemented and tested for two case studies with different characteristics, namely Skiathos, Greece and Sosnowiec, Poland. Pilots in Skiathos (a small Greek island with rural life style) exhibit higher consumptions that show no seasonality, while pilots in Sosnowiec (an industrial urban center) are characterized by lower consumptions and a weekday / weekend seasonality pattern. A number of qualitative and quantitative tests are implemented to check whether the objectives of the study are satisfied. The suggested methodology can successfully reproduce data that mimic the source consumption patterns following the actual peaks and lags recorded and consumption characteristics, such as the number and length of water use incidents per day. The method successfully captures the cumulative yearly water balance for each pilot. The algorithm becomes more robust when more recorded data is available and when more frequent water use incidents are recorded. The method seems to adjust well when the weekday/

weekend pattern is evident. The tool developed can provide meaningful water consumption data for reproducing gaps of data and missing values. The suggested approach can constitute a potential basis for building up a tool, which will support tasks demanding long water consumption data series, such as implementing and testing DSSs, supporting the development of models of user behavior, risk assessment of critical water consumption design parameters, or developing water use fraud detection systems. Finally, the algorithm can be used to reproduce water consumption data when privacy constraints hinder the use of actual time series.

5.2 Assessment of pressure driven demand savings with a theoretical pressure control management scheme

Foreword: In this subchapter, an assessment of the potential of pressure driven demand reduction for the ten Skiathos pilot water taps due to the application of the theoretical pressure control management scheme of chapter 4 is implemented. The assessment is based on the rational of the relevant pressure driven demand reduction KPI of Chapter 4. The possible relation of the altitude of the corresponding households to the potential savings is discussed.

5.2.1. Introduction

Attempts to incorporate PDD simulation into WDN models has been around the last three decades (Giustolisi et al., 2008; Wu et al., 2007; Kalungi and Tanyimboh, 2003; Ackley et al., 2001; Gupta and Bhave, 1996; Chandapillai, 1991). The common perception is that even though PDD is following the time pattern profile of the BAC, however it is pressure dependent and works as a function of pressure similar to the nature of leakage as described by (Germanopoulos and Jowitt, 1989; Germanopoulos, 1985). The distinction of SIV components into pressure dependent and time pattern dependent is not that absolute as perceived for the needs of simulating a WDN and described in subchapter 4.1. In reality all components that constitute a flow out of the WDN, including consumptions, include a part that is caused by excessive pressure, the overhead of the established minimum pressure needed to satisfy customers (Figure 59). A PM scheme would facilitate the reduction of the PDD other than the leakage reduction. In this subchapter, the profiling of such PDD savings is implemented for the Skiathos WT pilots.

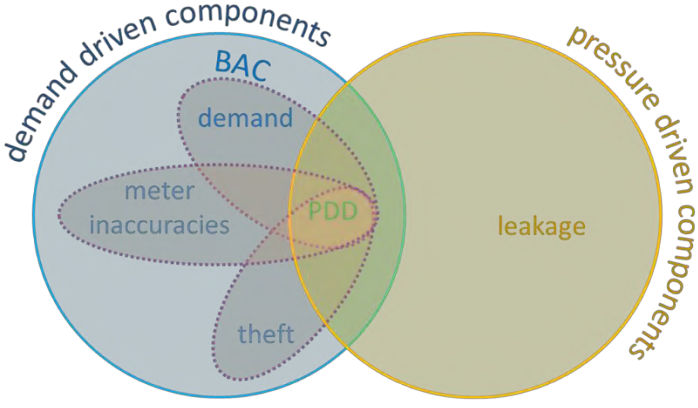


Figure 59. The Pressure Driven Demand shares a pressure driven and demand driven nature

5.2.2. Materials and Methods

For a more in depth understanding of the PDD reduction potential, an assessment of the actual theoretical reduction due to the simulated PM scheme of Chapter 4 is implemented for the actual Skiathos water tap pilot cases (Figure 60) described in sub-chapter 5.1. The pilot water taps are located diversely across the WDN and have different altitudes from 0.5 m to 22.35 m (Table 28). This variability in altitude and location imply variable local WDN pressure profiles and diverse urban land uses. The diverse conditions are also depicted in the average daily flowrate profiles of Figure 56, which are discussed in the previous sub-chapter.

The production of PDD reduction amounts is based on the SDG algorithm introduced in chapter 5.1, modified to reduce the water flow following the reduction of the pressure as the latter is simulated in chapter 4. The rational is the same as explained in chapter 4 regarding the PDD KPI. The following equation describes the modification of the SDG algorithm.

$$Savings_{PM,WT\ in\ L} = Flowrate_{SDG,\ WTi\ in\ L} * \left[1 - \left(\frac{Pressure_{\ PM,L}}{Pressure_{L}}\right)^{1.08}\right] \qquad \textit{(equation 37)}$$

where $\mathrm{Savings_{PM,WT\ in\ L}}$ is the theoretical amount of PDD that can be saved due to the theoretical PM scheme at Water Tap WT, which is located in Landzone L, $\mathrm{Flowrate_{SDG,\ WTi\ in\ L}}$ is the simulated flowrate of Water Tap WT which is located in Landzone L, with use of the SDG algorithm. $\mathrm{Pressure_{\ PM,L}}$ is the pressure at Landzone L, as simulated for the PM scheme described in subchapter 4.2 by the Skiathos simulation model, and $Pressure_L$ is the pressure at Landzone L as simulated by the Skiathos model.

Figure 60. The locations of the Skiathos water tap pilots across the water distribution system

The assessment does not take into account the potential extra reduction of the pressure due to the reduction of demand—specifically the reduction of part of the PDD—, because it is assumed that the pressure is reregulated to its minimum level of customer satisfaction. The same assumption is also made for the leakage estimation in subchapter 4.2.

Table 28. The elevations of the Skiathos pilot water taps

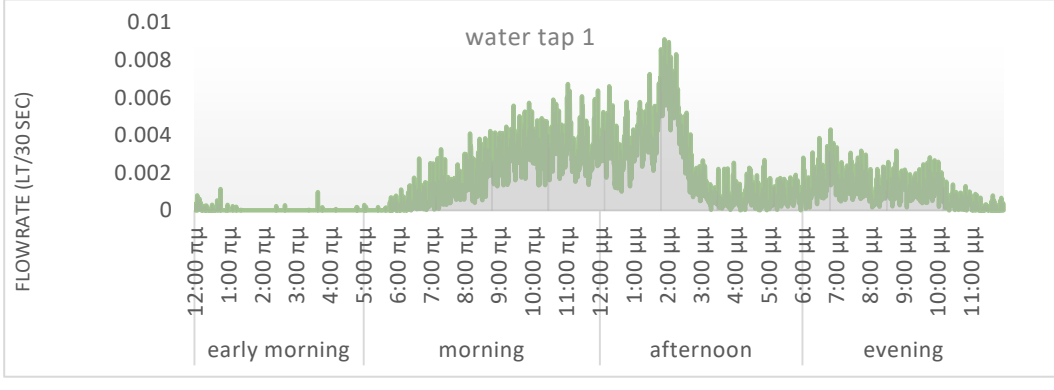
water tap	elevation
1	1.50 m
2	19.50 m
3	6.80 m
4	1.95 m
5	5.40 m
6	4.30 m
7	0.50 m
8	22.35 m
9	4.30 m
10	5.80 m

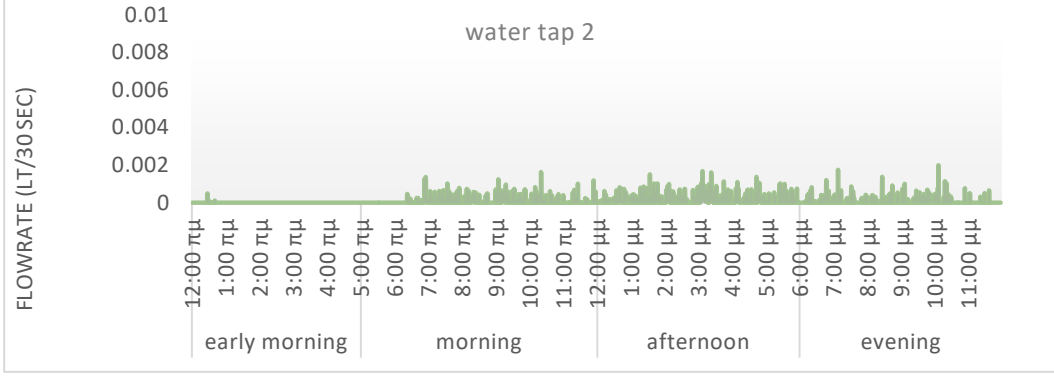
5.2.3. Results and Discussion

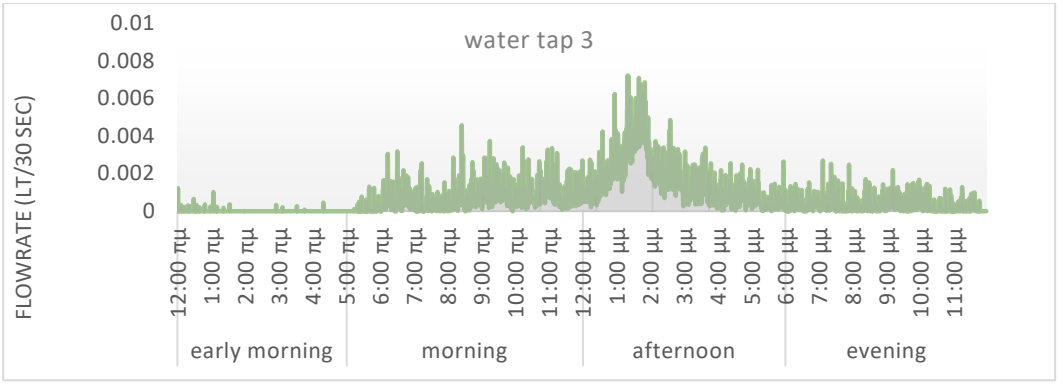
The PDD reduction, or else the savings due to the theoretical PM scheme are presented in Figure 61 and tables Table 29 and Table 30. In Figure 61, the annual average daily profiles of the PDD savings for each WT are presented. Compared to the respective diagrams of the initial flowrates, where no PM scheme is applied, (Figure 56, Subchapter 5.1), it can be concluded that the savings follow the profiles of the demands. Additionally, regarding absolute amount, higher savings are achieved where there are higher demand levels. This does not also apply for relative values of PDD savings over initial demands, which can be explained, because higher demands occur to low levels, where percentage pressure reduction is lower, since the theoretical PM scheme is lumped. In fact, the relation seems to be reverse (Figure 62). This agrees the findings of chapter 4.2, where in the hilly landzones the relative PDD savings is lower than the savings in coastal low levels.

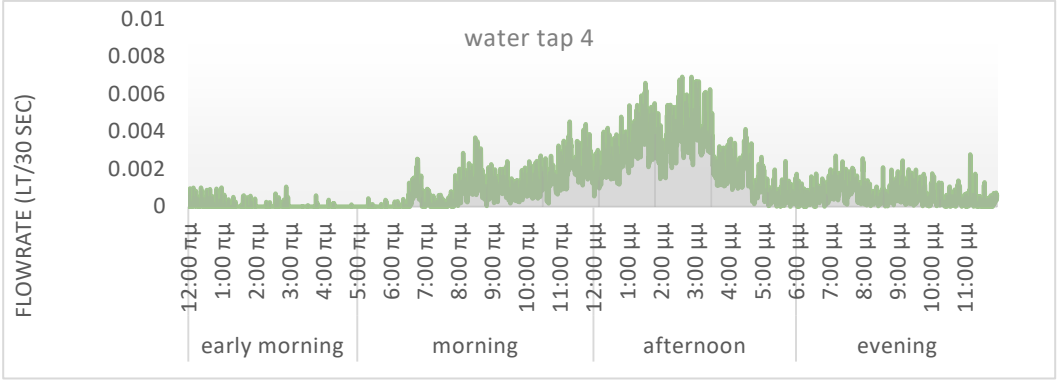
Table 30 presents the comparison of the relative PDD savings for the Skiathos WT pilots and the respective relative savings for the whole landzones where the WTs are located. Both savings amounts are expressed as a ratio over their initial demands. The WT savings are divided by the WT flowrates, while Landzone savings are divided by the landzones' *BAC + Apparent Losses* amounts. The comparison of the two agrees sufficiently indicating that the PDD savings KPI of chapter 4.2 is reliable. A more careful observation reveals that the WTs with higher demands agree better to the respective landzones than the WTs with lower demands whose savings seem to be slightly overestimated by the landzone KPIs. This can possibly be explained by the fact that low demand WTs are used less on one hand and are located in high elevations which force low flowrates.

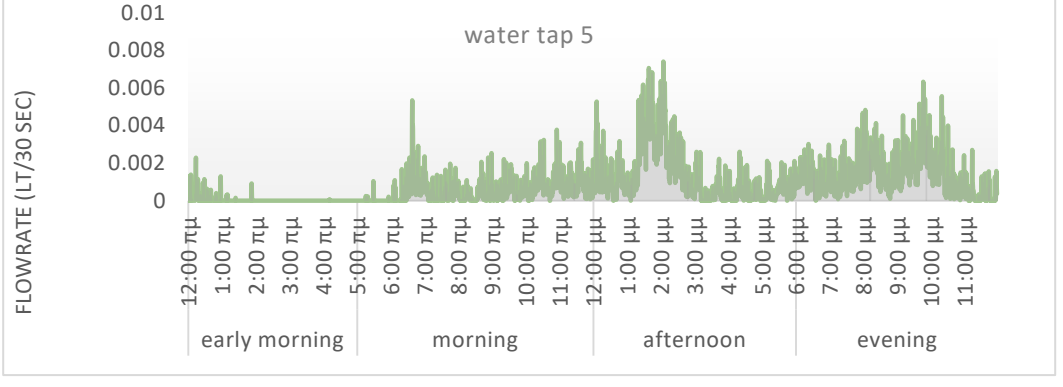
Table 29 presents the absolute potential savings amounts for the ten WTs for a whole year and the split of these amounts into four time-zones: early morning (24:00 to 06:00), morning (06:00 to 12:00), afternoon (12:00 to 18:00), and evening (18:00 to 24:00). The total annual amounts depict a very wide range from 82 lt/year to 1,500 lt/year. The higher savings correspond to the higher demands. The time zones that present the massive amounts of savings are the morning zone and afternoon zone. The evening zone depicts a relevantly significant level of savings, while the early morning zone depicts a rather insignificant amount of savings.

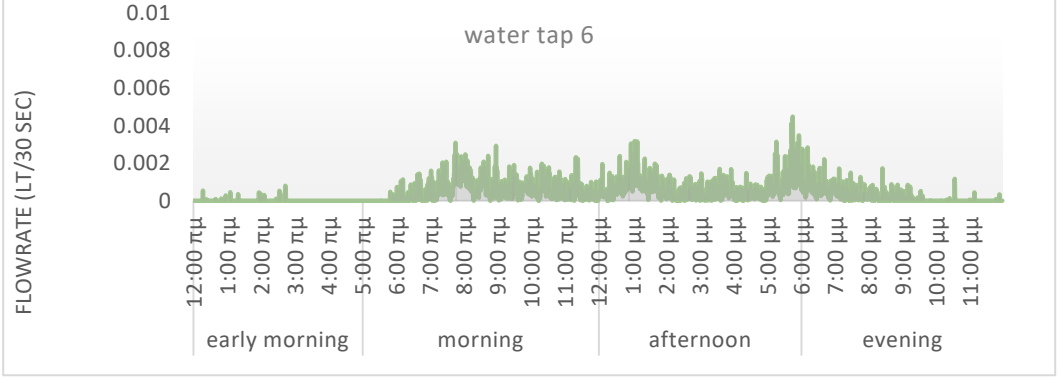












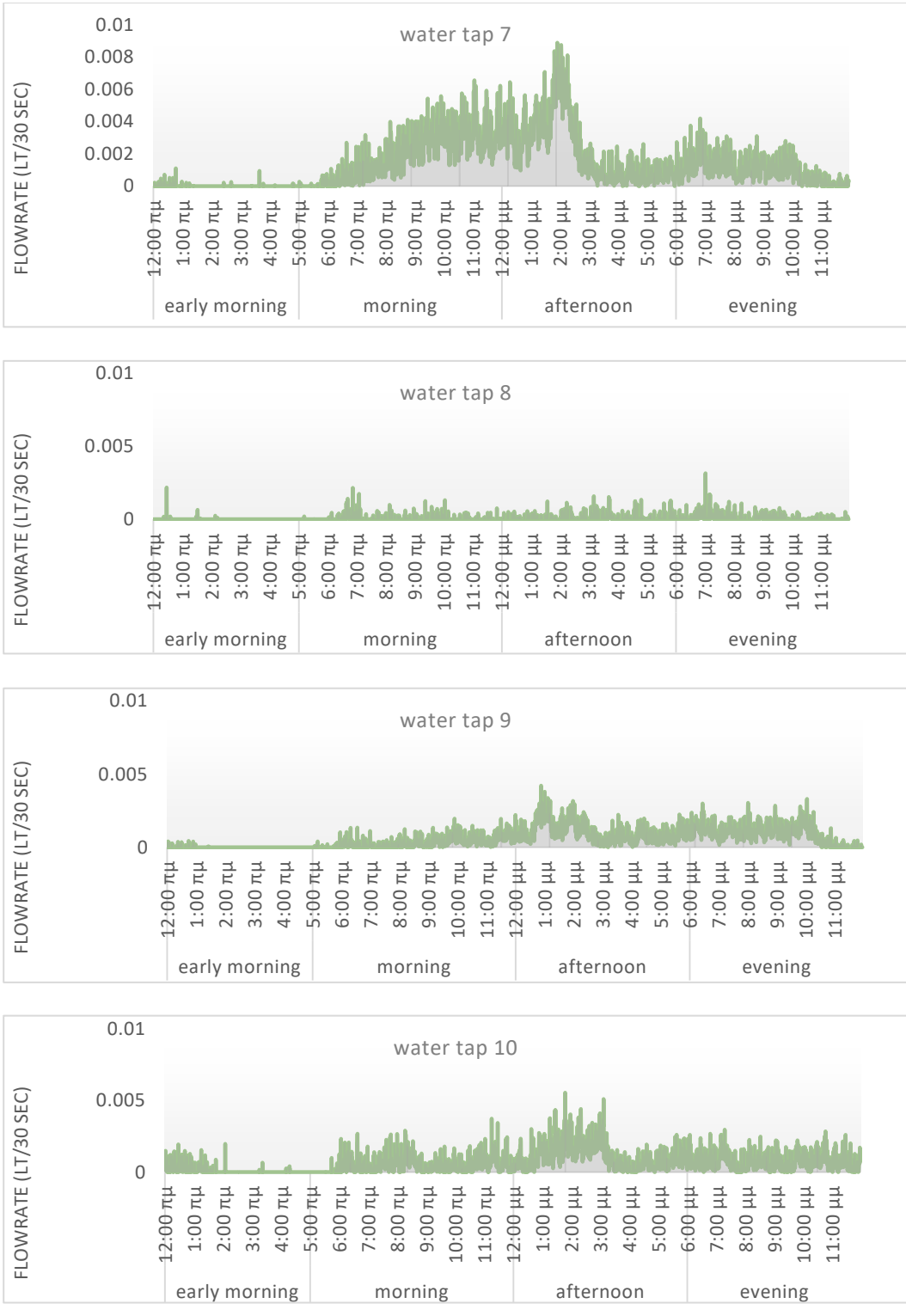


Figure 61. Potential annual average daily profile of potential saving of pressure driven demand due to the application of a pressure control scheme

Table 29. Annual potential PDD savings (lt) due to the theoretical PM scheme simulated, total and divided into day quarter time zones

water tap	annual totalo	annual early morning	annual morning	annual afternoon	annual evening
1	1,514	5	605	629	275
2	92	0	26	46	19
3	810	4	234	471	102
4	1,187	17	299	700	171
5	936	9	181	358	387
6	447	3	176	197	71
7	1,480	4	591	616	269
8	82	2	21	27	32
9	650	2	109	287	253
10	724	26	166	318	214

Table 30. Comparative table of the PDD reduction of the pilot water taps and the respective whole landzones, both expressed in m^3/m^3 of their respective initial demands.

water tap	PDD savings in m^3/m^3 of initial water tap demand	PDD savings in m^3/m^3 of initial BAC + Apparent Losses at the respective landzone
1	0.100	0.104
2	0.141	0.174
3	0.110	0.111
4	0.101	0.107
5	0.107	0.110
6	0.106	0.172
7	0.099	0.101

8	0.148	0.163
9	0.104	0.132
10	0.109	0.117

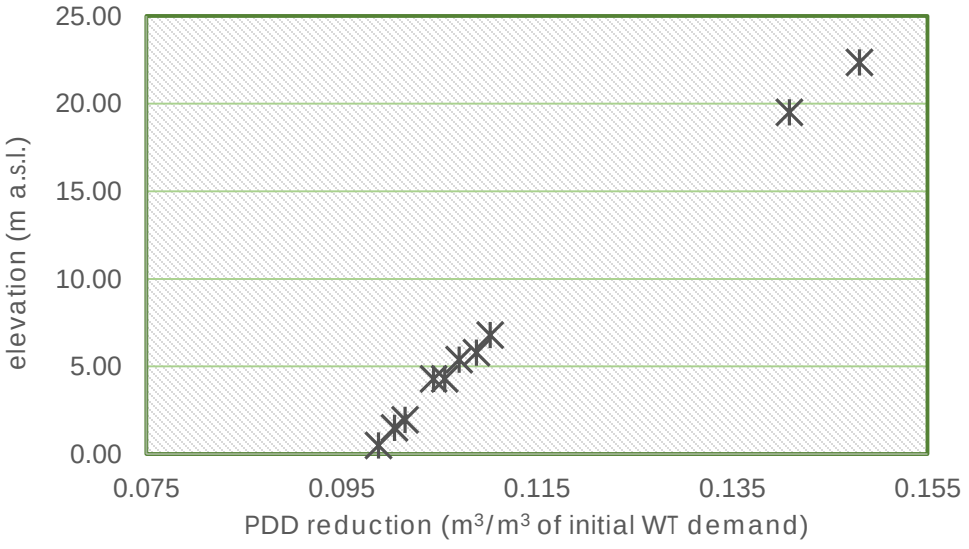


Figure 62. Scatter plot of the relative PDD savings of the 10 Skiathos WT pilots vs their respective elevations

Afterword: This subchapter is about the assessment of theoretical PDD reduction due to the PM scheme introduced in chapter 4. For the assessment the SDG algorithm, presented in chapter 5.1 is used as well as the rational of the PDD KPI also introduced in chapter 4. The pressure values used for the assessment are taken from the simulation model and are the values that have been estimated for the landzones each one for the respective WT that includes it. The analysis agrees with the findings of chapter 4 and proves the introduced PDD KPI to be reliable. The PDD savings potential for a WT showcases a wide range, approximately, from 80 lt/year to 1,500 lt/year depending on the demand level and the elevation. Higher elevations correspond to higher relative PDD savings, however this is reasonable because the PM scheme is lumped. The day time zones showcase higher amounts of PDD savings that the night zones in absolute amounts, since during the day there is much higher consumption than during the night.

The content of chapter 5.1 is included in the following published article:

Kofinas, D. T., Spyropoulou, A., Laspidou, C. S. (2018). A methodology for synthetic household water consumption data generation. *Environmental modelling & software*, *100*, 48-66.

- The contribution of Mr. Kofinas, D. involves the conceptualization, the methodology, the validation, the formal analysis, the investigation, the writing, and the visualization.
- The contribution of Dr Spyropoulou, A. involves the programming of the involved code.
- The contribution of Professor Laspidou, C. involves the scientific supervision

This work was supported by the project ISS EWATUS—Integrated Support System for Efficient Water Usage and Resources Managementdwhich is implemented in the framework of the EU 7th Framework Programme, Specific programme Cooperation In- formation and Communication Technologies; Grant Agreement Number 619228.

6. Conclusions and future work

The current thesis constitutes an investigation on the modeling potential of variable components of the WDNs from the drilling to the tap. Among the multiple objectives, on high level, the overall objective is to develop an innovative methodological framework on improving urban water management based on simulation tools that provide detailed overview of the system on multiple scales (Figure 63).

The investigation is implemented on an actual case study, the WDS of Skiathos Island. This offers some specifications that led to interesting results. Skiathos is a touristic resort with intense touristic influx and intense weather seasonality according to the Mediterranean profile. The infrastructure of the WSS is aged, while the town of Skiathos has a bold relief. The pumped groundwater is loaded with mercury, which is released in the aquifer by the bedrock, due to the salinity increase and sea intrusion which in turn is happening due to the over pumping.

The analysis begins with the development of forecasting algorithms of water demand from the perspective of water sources, meaning that water demand includes all the water that is pumped out regardless if it is consumed or lost across the network. The forecasting is implemented in three periodicities: trimester, monthly and daily. Various forecasting approaches are tested such as statistical and ML, univariate and multivariate. Namely, MR, ARIMA, SARIMA, ARIMAX, ANFISS, and ANN, Hybrid approaches. The predictors used in the multivariate analysis are weather variables—which have generally been used quite often in past multivariate analysis attempts—, a touristic activity indicator and an indicator of the WDN performance. The sum of arrivals into the island by all means is used as a touristic activity indicator. The Non-Revenue Water Percentage is used as an indicator of the performance of the WDN. The aforementioned predictors proved to be very reliable for the forecasting. The methodology is tested on a blind data set, different from the one used for the calibration. All methodologies provided adequate results, however a multivariate analysis based on ML algorithms has the benefit of providing a managerial tool that can be used for scenario assessment.

The analysis continues with an attempt to create a simulation model of the Skiathos WDN. Firstly initial values of daily water demands in numerous nodes are created by disaggregating the trimester customer billings through the daily SIV profile, assuming a linear distribution and that the NRW all equals the leakage. With the same rational, reversely, the daily SIV values are spatially disaggregated through the billings distribution. The initial values are used to feed an EPANET model. A system of two nested loops is iterating to produce the final values of BAC, real losses, and apparent losses based on the principles that real losses are pressure dependent and the apparent losses are rather time pattern dependent following the BAC profile. The pressure dependent components are simulated with use of open valves in each demand node, based on the minimum night flow approach

for leakage estimation. The loop system closes when the apparent losses and real losses are decoupled, summing up to the NRW amount, as the simulated pressure fits the actual pressure recorded in three points of the WDN. The simulation is used to conduct a series of conclusions regarding the seasonality of the WSS, the urban land uses, the spatio-temporal profile of the IWA WB components, the high level localization of theft and the estimation of the water meter inaccuracy level. On top of the assessment of the WB, a series of KPIs are introduced to facilitate the informed management of the WDN, as well as the performance evaluation of a theoretical PM scheme, which is also simulated. Such KPIs are the energy nexus, the PDD reduction, the leakage expressed per service connection and per network length, among others. The spatiotemporal analysis proved that such a simulation tool can be crucial for the improvement of WDN management especially in an intensely variable in time and/or in space case where the adjustments need to follow all the dynamic parameters.

The last part of the analysis focuses on tap water consumption. A Synthetic Data Generator algorithm is developed to simulate water tap data consumption. The SDG constitutes a modified pulse model that firstly focuses on mimicking the consumption time pattern. The level of flowrates and the instants when they occur are assessed based on the respective distributions. A Newton law of gravity based formula is later on used to attract instants of flowrates together to shape up incidents of realistic length. The algorithm is tested both on ten Skiathos WT pilots and on 9 Sosnowiec WT pilots. The two case studies depict significant variabilities since the first is rural and touristic, while the other is industrial. For the better fitting of the second case study a week/weekend seasonality routine is incorporated. The SDG is finally used to simulate the PDD reduction at the ten Skiathos WTs due to the PM scheme described earlier. The assessment is based on the same rational of the development of the PDD reduction KPI. The two approaches are compared and present adequate consistency.

Potential future work that emerges from the current thesis would involve

- the investigation regarding the variability of demand patterns in relation to the different urban land uses, using SOM clustering techniques;
- the investigation of the possibility of simulated based ML techniques application, which would imply the training of AI algorithms on the simulated data sets of the IWA WB components instead of training them on recorded actual data. This may reveale a potential in facilitating a WDN's simulation with the prerequisite of having available libraries of numerous hydraulic simulations of variable WDNs.
- the application with use of the simulation model and its KPIs of a multi-DMA PM scheme;
- a more detailed water-energy analysis which would explore the potential energy savings due to the division of the WDN into DMAs;

- the connection of the WDN simulation model and KPIs to sensors to continuously perform calibration and forecasts in real time, thus operating as a Digital Twin;
- the comparison of the performance of the SDG model to other pulse model approaches;
- the potential improvement of the SDG model by incorporating multivariate analysis;
- the identification and the quantification of the benefits of smoothening water demand peaks in WDN, landzone and water tap level.

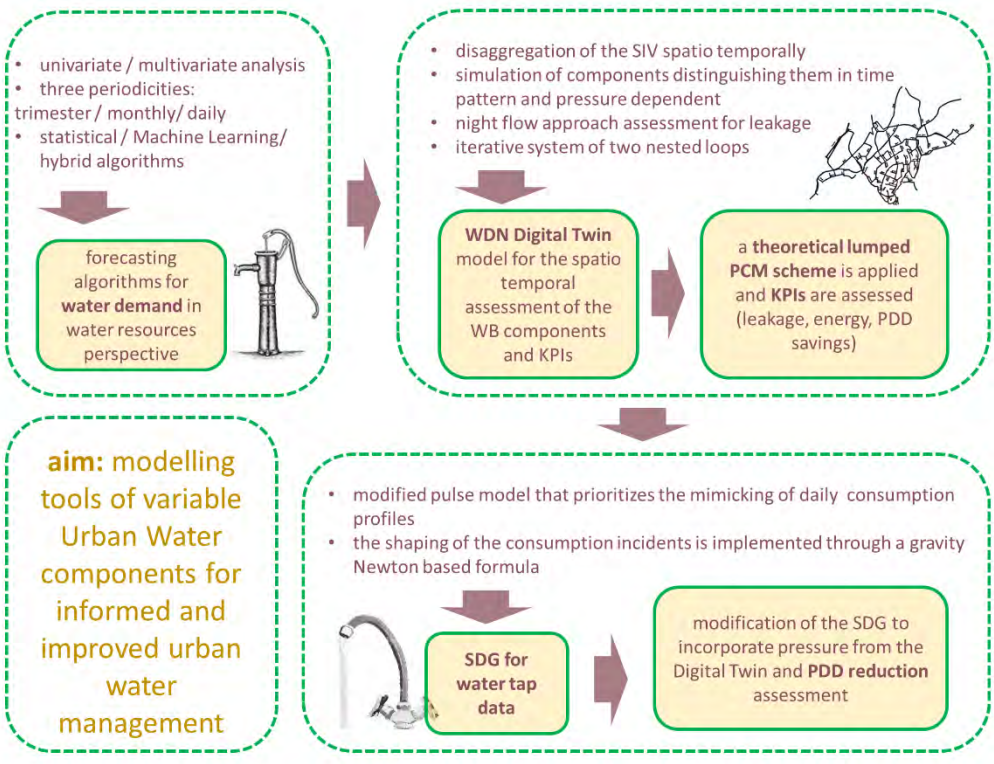


Figure 63. Schematic overview of the current thesis

Bibliography

Abdelmeguid, H., Ulanicki, B., 2010. Pressure and Leakage Management in Water. Simulation 1124–1139.

Ackley, J.R.L., Tanyimboh, T.T., Tahar, B., Templeman, A.B., 2001. Head-Driven Analysis of Water Distribution Systems. Int. Conf. Comput. Control Water Ind. CCWI 2001 1, 183–192.

Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol. 407, 28–40. https://doi.org/10.1016/j.jhydrol.2011.06.013

Adamowski, J., Karapataki, C., 2010. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. J. Hydrol. Eng. 15, 729–743. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000245

Adamowski, J.F., 2008. Peak daily water demand forecast modeling using artificial neural networks. J. Water Resour. Plan. Manag. 134, 119–128. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:2(119)

Adriaenssens, V., De Baets, B., Goethals, P.L.M., De Pauw, N., 2004. Fuzzy rule-based models for decision support in ecosystem management. Sci. Total Environ. 319, 1–12. https://doi.org/10.1016/S0048-9697(03)00433-9

Agthe, D., Billings, B., Dworkin, J., 1988. Effects of rate structure knowledge on household water use 24, 627–630.

Akaike, H., 1974. A New Look at the StatisticalModel Identification. IEEE Trans. Automat. Contr. AC-19, 716–723.

Al-Hoqani, N., Yang, S.H., 2015. Adaptive sampling for wireless household water consumption monitoring. Procedia Eng. 119, 1356–1365. https://doi.org/10.1016/j.proeng.2015.08.980

Alegre, H., 1999. Performance Indicators for Water Supply Systems, 3rd ed, In: Cabrera E., García-Serra J. (eds) Drought Management Planning in Water Supply Systems. Water Science and Technology Library. Dordrecht.

Allende, H., Moraga, C., Salas, R., 2002. Artificial neural networks in time series forecasting: A comparative analysis. Kybernetika 38, 685–707.

Almandoz, J., Cabrera, E., Arregui, F., Cabrera, E., Cobacho, R., 2005. Leakage Assessment through Water Distribution Network Simulation. J. Water Resour. Plan. Manag. 131, 458–466. https://doi.org/10.1061/(asce)0733-9496(2005)131:6(458)

Alvisi, S., Franchini, M., Marinelli, A., 2007. A short-term, pattern-based model for water-demand forecasting. J. Hydroinformatics 9, 39–50.

https://doi.org/10.2166/hydro.2006.016

Alvisi, S., Franchini, M., Marinelli, A., 2003. A stochastic model for representing drinking water demand at residential level. Water Resour. Manag. 17, 197–222. https://doi.org/10.1023/A:1024100518186

Arampatzis, G., Perdikeas, N., Kampragou, E., Scaloubakas, P., Assimacopoulos, D., 2014. A water demand forecasting methodology for supporting day-to-day management of water distribution systems, in: 12th International Conference "Protection & Restoration of the Environment. Skiathos, Thessaloniki, Greece, pp. 1–8.

Araujo, L.S., Ramos, H., Coelho, S.T., 2006. Pressure control for leakage minimisation in water distribution systems management. Water Resour. Manag. 20, 133–149. https://doi.org/10.1007/s11269-006-4635-3

Arbués, F., García-Valiñas, M.Á., Martínez-Espiñeira, R., 2003. Estimation of residential water demand: A state-of-the-art review. J. Socio. Econ. 32, 81–102. https://doi.org/10.1016/S1053-5357(03)00005-2

Arregui, F., Cabrera, E., Cobacho, R., García-Serra, J., 2015. Reducing Apparent Losses Caused By Meters Inaccuracies. Water Pract. Technol. 1. https://doi.org/10.2166/wpt.2006.093

Authority H. S., 2011. Population housing census.

Azadeh, A., Saberi, M., Anvari, M., Azaron, A., Mohammadi, M., 2011. An adaptive network based fuzzy inference system-genetic algorithm clustering ensemble algorithm for performance assessment and improvement of conventional power plants. Expert Syst. Appl. 38, 2224–2234. https://doi.org/10.1016/j.eswa.2010.08.010

Barse, E.L., Kvarnström, H., Jonsson, E., 2003. Synthesizing test data for fraud detection systems. Proc. - Annu. Comput. Secur. Appl. Conf. ACSAC 2003-Janua, 384–394. https://doi.org/10.1109/CSAC.2003.1254343

Batista, G.E.A.P.A., Monard, M.C., 2003. An analysis of four missing data treatment methods for supervised learning. Appl. Artif. Intell. 17, 519–533. https://doi.org/10.1080/713827181

Batista, G.E.A.P.A., Monard, M.C., 2002. A study of k-nearest neighbour as an imputation method. Front. Artif. Intell. Appl. 87, 251–260.

Baxter, C.W., Zhang, Q., Stanley, S.J., Shariff, R., Tupas, R.R.T., Stark, H.L., 2001. Drinking water quality and treatment: The use of artificial neural networks. Can. J. Civ. Eng. 28, 26–35. https://doi.org/10.1139/cjce-28-1-26

Bennett, C., Stewart, R.A., Beal, C.D., 2013. ANN-based residential water end-use demand forecasting model. Expert Syst. Appl. 40, 1014–1023. https://doi.org/10.1016/j.eswa.2012.08.012

Billings, R.B., Jones, C.V., 2008. Forecasting Urban Water Demand 367.

Bishop, M., 1995. Neural Networks for Pattern Recognition. Oxford University Press Inc., New York.

Bishop, R.R., 1978. Hydraulic Characteristics of PVC Pipe in Sanitary Sewers (A Report of Field Measurements) Part of the Civil and Environmental Engineering Commons, and the Water Resource Management Commons HYDRAULIC CHARACTERISTICS OF PVC PIPE IN SANITARY SEWERS (A Report of Field Measurements).

Blokker, E., Vreeburg, J., van Dijk, J., 2009. Simulating residential water demand with a stochastic end-use model. J. Water Resour. Plan. Manag. 136, 19–26.

Bonissone, P.P., Badami, V., Chiang, K.H., Khedkar, P.S., Marcelle, K.W., Schutten, M.J., 1995. Industrial Applications of Fuzzy Logic at General Electric. Proc. IEEE 83, 450–465. https://doi.org/10.1109/5.364490

Botev, Z., Ridder, A., 1997. Variance Reduction I 1, 118–141. https://doi.org/10.1002/9781118445112.stat07975

Bougadis, J., Adamowski, K., Diduch, R., 2005. Short-term municipal water demand forecasting. Hydrol. Process. 19, 137–148. https://doi.org/10.1002/hyp.5763

Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. Time series analysis: forecasting and control. John Wiley & Sons.

Brekke, L., Larsen, M.D., Ausburn, M., Takaichi, L., 2002. Suburban water demand modeling using stepwise regression. J. / Am. Water Work. Assoc. 94, 65–75. https://doi.org/10.1002/j.1551-8833.2002.tb09558.x

Buchberger, S.G., Wu, L., 1995. Model for Instantaneous Residential Water Demands. J. Hydraul. Eng. 121, 232–246.

Burney, N., Mukhopadhyay, A., Al-Mussallam, N., Akber, A., Al-Awadi, E., 2001. Forecasting of freshwater demand in Kuwait. Arab. J. Sci. Eng. 26, 99–113.

Cabrera, E., Pellejero, I., 2003. Evaluation of leakage by means of night flow measurements and analytical discrimination. A comparative study., in: Pumps, Electromechanical Devices and Systems Applied to Urban Water Management 1. p. 327.

Chandapillai, J., 1991. REALISTIC SIMULATION O F W A T E R DISTRIBUTION SYSTEM By Jacob Chandapillai 1 (Reviewed by the Pipeline Division). J. Transp. Eng. 117, 258–263.

Chang, M., Liu, J., 2009. Water demand prediction model based on radial basis function neural network. 2009 1st Int. Conf. Inf. Sci. Eng. ICISE 2009 2, 5295–5298. https://doi.org/10.1109/ICISE.2009.1343

Chau, K.W., 2006. Particle swarm optimization training algorithm for ANNs in stage

prediction of Shing Mun River. J. Hydrol. 329, 363–367. https://doi.org/10.1016/j.jhydrol.2006.02.025

Chen, Xiaomin, Yang, S.H., Yang, L., Chen, Xi, 2015. A benchmarking model for household water consumption based on adaptive logic networks. Procedia Eng. 119, 1391–1398. https://doi.org/10.1016/j.proeng.2015.08.998

Cheng, C.H., Wei, L.Y., 2010. One step-ahead ANFIS time series model for forecasting electricity loads. Optim. Eng. 11, 303–317. https://doi.org/10.1007/s11081-009-9091-5

Cobacho, R., Arregui, F., Soriano, J., Cabrera, E., 2015. Including leakage in network models: An application to calibrate leak valves in EPANET. J. Water Supply Res. Technol. - AQUA 64, 130–138. https://doi.org/10.2166/aqua.2014.197

Colglazier, W., 2015. Sustainable development agenda: 2030. Science (80-.). https://doi.org/10.1126/science.aad2333

Colombo, A.F., Karney, B.W., 2002. Energy and costs of leaky pipes toward comprehensive picture. J. Water Resour. Plan. Manag. 128, 441–450. https://doi.org/10.1061/(ASCE)0733-9496(2002)128:6(441)

Cominola, A., Giuliani, M., Castelletti, A., Abdallah, A.M., Rosenberg, D.E., Cominola, A.;, Giuliani, M.;, Castelletti, A.;, Abdallah, A.M.;, 2016. Developing a stochastic simulation model for the generation of residential water end-use demand time series. 8th Int. Congr. Environ. Model. Softw.

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., Rizzoli, A.E., 2015. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. Environ. Model. Softw. 72, 198–214. https://doi.org/10.1016/j.envsoft.2015.07.012

Cominola, Andrea, Cominola, A, Nguyen, K., Giuliani, M., Stewart, R.A., Maier, H.R., Castelletti, A., 2019. Data mining to uncover heterogeneous water use behaviors from smart meter data Data mining to uncover heterogeneous water use behaviors from smart meter data Key Points : https://doi.org/10.1029/2019WR024897

Contreras, J., Espínola, R., Nogales, F.J., Conejo, A.J., 2003. ARIMA models to predict next-day electricity prices. IEEE Trans. Power Syst. 18, 1014–1020. https://doi.org/10.1109/TPWRS.2002.804943

Creaco, E., Blokker, M., Buchberger, S., 2017. Models for generating household water demand pulses: Literature review and comparison. J. Water Resour. Plan. Manag. 143. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000763

Creaco, E., Cunha, M., Franchini, M., 2019. Using Heuristic Techniques to Account for Engineering Aspects in Modularity-Based Water Distribution Network Partitioning Algorithm. J. Water Resour. Plan. Manag. 145, 1–11. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001129

Creaco, E., Farmani, R., Kapelan, Z., Vamvakeridou-Lyroudia, L., Savic, D., 2015. Considering the mutual dependence of pulse duration and intensity in models for generating residential water demand. J. Water Resour. Plan. Manag. 141, 1–9. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000557

Creaco, E., Haidar, H., 2019. Multiobjective Optimization of Control Valve Installation and DMA Creation for Reducing Leakage in Water Distribution Networks. J. Water Resour. Plan. Manag. 145, 1–10. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001114

Creaco, E., Kossieris, P., Vamvakeridou-Lyroudia, L., Makropoulos, C., Kapelan, Z., Savic, D., 2016. Parameterizing residential water demand pulse models through smart meter readings. Environ. Model. Softw. 80, 33–40. https://doi.org/10.1016/j.envsoft.2016.02.019

Darby, S., 2006. the Effectiveness of Feedback on Energy Consumption a Review for Defra of the Literature on Metering , Billing and. Environ. Chang. Inst. Univ. Oxford 22, 1–21. https://doi.org/10.4236/ojee.2013.21002

Dean, S., Illowsky, B., 2009. Descriptive Statistics: Histogram [WWW Document]. Connexions Proj.

DeLurgio, S., 1998. Forecasting principles and applications. McGraw-Hill Irwin Publishing, Burr Ridge, IL.

Deyfus, G., 2005. Neural Networks: Methodology and Applications, 2nd ed, International Journal of Engineering, Science and Technology. Springer, Paris. https://doi.org/10.4314/ijest.v2i10.64007

Di Nardo, A., Di Natale, M., Giudicianni, C., Laspidou, C., Morlando, F., Santonastaso, G.F., Kofinas, D., 2017. Spectral analysis and topological and energy metrics for water network partitioning of Skiathos island. Eur. Water 423–428.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. J. Am. Stat. Assoc. 74, 427. https://doi.org/10.2307/2286348

Donkor, E.A., Mazzuchi, T.A., Soyer, R., Alan Roberson, J., 2014. Urban water demand forecasting: Review of methods and models. J. Water Resour. Plan. Manag. 140, 146–159. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000314

Edelenbos, J., van Buuren, A., van Schie, N., 2011. Co-producing knowledge: Joint knowledge production between experts, bureaucrats and stakeholders in Dutch water management projects. Environ. Sci. Policy 14, 675–684. https://doi.org/10.1016/j.envsci.2011.04.004

EEA, 2018. EEA SIGNALS 2018, Water is life. European Environmental Agency. https://doi.org/10.2800/52469

Ehrhardt-martinez, A.K., Donnelly, K. a, 2010. Advanced Metering Initiatives and

Residential Feedback Programs : A Meta-Review for Household Electricity-Saving Opportunities. Energy 123, 128.

Fantozzi, M., Lambert, A., 2012. Residential Night Consumption-Assessment, Choice of Scaling Units and Calculation of Variability. Manila, Philippines.

Farley, M., Trow, S., 2003. Losses in water distribution networks. IWA publishing.

Fielding, K.S., Russell, S., Spinks, A., Mankad, A., 2012. Determinants of household water conservation: The role of demographic, infrastructure, behavior, and psychosocial variables. Water Resour. Res. 48. https://doi.org/10.1029/2012WR012398

Fischer, C., 2008. Feedback on household electricity consumption: A tool for saving energy? Energy Effic. 1, 79–104. https://doi.org/10.1007/s12053-008-9009-7

Flood, I., Kartman, N., Associate_Members, ASCE, 1994. Neural Networks in Civil Engineering. II: Systems nd Applications. University of Maryland Minta Martin awards and Maryland Industrial Partnerships.

Foster, H.S., Beattie, B.R., 1981. On the Specification of Price in Studies of Consumer Demand under Block Price Scheduling. Land Econ. 57, 624–629.

Foster, S., Beattie, B.R., 1979. Urban Residential Demand Demand for Water in the United States. Land Econ. 55, 43–58.

Gardiner, V., Herrington, P., 1990. Water Demand Forecasting, 1st ed. Spon Press.

Germanopoulos, G., 1985. A technical note on the inclusion of pressure dependent demand and leakage terms in water supply network models. Civ. Eng. Syst. 2, 171–179. https://doi.org/10.1080/02630258508970401

Germanopoulos, G., Jowitt, P.W., 1989. Leakage reduction by excess pressure minimization in a water supply network. Proc. - Inst. Civ. Eng. Part 2. Res. theory 87, 195–214.

Getz, D., 1993. Planning for tourism business districts. Ann. Tour. Res. 20, 583–600. https://doi.org/10.1016/0160-7383(93)90011-Q

Ghiassi, M., Zimbra, D., Saidane, H., 2008. Urban Water Demand Forecasting with a Dynamic Artificial Neural Network Model. J. Water Resour. Plan. Manag. 134(2), 138–146. https://doi.org/10.1061/(ASCE)0733-9496(2008)134

Giffinger, R., Fertner, C., Kramar, H., Meijers, E., 2007. Abigail_Final_Research_Papper 1–12.

Giustolisi, O., Savic, D., Kapelan, Z., 2008. Pressure-driven demand and leakage simulation for water distribution networks. J. Hydraul. Eng. 134, 626–635. https://doi.org/10.1061/(ASCE)0733-9429(2008)134:5(626)

Govindaraju, R.S., Rao, A.R., 2000. Artificial Neural Networks in Hydrology, Water

Science and Technology Library Vol. 36, Springer-Science+Business Media, B.V. https://doi.org/10.1017/CBO9781107415324.004

Grzymala-Busse, J.W., Hu, M., 2001. A Comparison of Several Approaches to Missing Attribute Values in Data Mining BT - Rough sets and current trends in computing. Rough sets Curr. trends Comput. 2005, 378–385. https://doi.org/10.1007/3-540-45554-X_46

Gupta, R., Bhave, P., 1996. Comparison of Methods for Predicting Deficient - Network Performance. J. Water Resour. Plan. Manag. 122, 214–217.

Hajeeh, M., 2010. Technical note: Water conservation in Kuwait: A fuzzy analysis approach. J. Ind. Eng. Int. 6, 90–105.

Hassoun, M.H., 1995. Fundamentals of Artificial Neural Networks. Cambridge, MA: MIT Press. https://doi.org/10.1177/1757913910379198

Hatzikos, E. V, Anastasakis, L., Bassiliades, N., Vlahavas, I., 2005. Applying neural networks with active neurons to sea-water quality measurements The above table shows the correlation between the variables at time t and hence it, in: 2nd Int. Scientific Conf. on Computer Science, Plamenka Borovska, Sofoklis Christofordis (Ed.), IEEE Computer Society, Bulgarian Section, 30th Sep-2nd Oct 2005. Halkidiki, Greece.

Haykin, S., 2005. Neural Networks - A Comprehensive Foundation - Simon Haykin.pdf.

Hazen and Sawyer and PMCL, 2004. The Tampa Bay Water Long-Term Demand Forecasting Model Water Management District.

Hirner, W., Lambert, A., 2000. Losses from Water Supply Systems: Standard Terminology and Recommended Performance Measures.

Hollander, M., Wolfe, D.A., 1973. Nonparametric Statistical Methods, 2nd ed. John Wiley & Sons, New York, Sydney, Tokyo, Mexico City.

Hollands, R.G., 2008. Will the real smart city please stand up? Intelligent, progressive or entrepreneurial? City 12, 303–320. https://doi.org/10.1080/13604810802479126

Inskeep, E., 1987. Environmental planning for tourism. Ann. Tour. Res. 14, 118–135. https://doi.org/10.1016/0160-7383(87)90051-X

Ismail, Z., Puad, W.F.W.A., 2006. Non-Revenue Water Losses: A Case Study, Asian Journal of Water, Environment and Pollution.

ISS-EWATUS Integrated Support System for Efficient Water Usage and Resources Management [WWW Document], 2016. URL http://issewatus.eu/ (accessed 11.10.16).

Jang, J.S.R., 1993. ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE

Trans. Syst. Man Cybern. 23, 665–685. https://doi.org/10.1109/21.256541

Jang, J.S.R., Sun, C.T., Mizutani, E., 2005. Neuro-Fuzzy and Soft Computing-A Computational Approach to Learning and Machine Intelligence [Book Review]. IEEE Trans. Automat. Contr. 42, 1482–1484. https://doi.org/10.1109/tac.1997.633847

Jentgen, L., Kidder, H., Hill, R., Conrad, S., 2007. Energy management strategies use short-term water consumption forecasting to minimize cost of pumping operations. J. / Am. Water Work. Assoc. 99, 86–94. https://doi.org/10.1002/j.1551-8833.2007.tb07957.x

Jowitt, P.W., Xu, C., 1990. Optimal valve control in water distribution networks. J. Water Resour. Plan. Manag. 116, 455–472. https://doi.org/10.1061/(ASCE)0733-9496(1990)116:4(455)

Kalungi, P., Tanyimboh, T.T., 2003. Redundancy model for water distribution systems. Reliab. Eng. Syst. Saf. 82, 275–286. https://doi.org/10.1016/S0951-8320(03)00168-6

Kanakoudis, V., Tsitsifli, S., 2019. Water networks management: New perspectives. Water (Switzerland) 11, 2–5. https://doi.org/10.3390/w11020239

Kanakoudis, V., Tsitsifli, S., 2014. Using the bimonthly water balance of a non-fully monitored water distribution network with seasonal water demand peaks to define its actual NRW level: The case of Kos town, Greece. Urban Water J. 11, 348–360. https://doi.org/10.1080/1573062X.2013.806563

Kanakoudis, V., Tsitsifli, S., 2010a. Results of an urban water distribution network performance evaluation attempt in Greece. Urban Water J. 7, 267–285. https://doi.org/10.1080/1573062X.2010.509436

Kanakoudis, V., Tsitsifli, S., 2010b. Water volume vs. revenues oriented water balance calculation for urban water networks: the Minimum Charge Difference component makes a difference! Proc. IWA Int. Spec. Conf. Water Loss 2010 v, 27–30.

Kanakoudis, V., Tsitsifli, S., Papadopoulou, A., Cencur Curk, B., Karleusa, B., 2017. Water resources vulnerability assessment in the Adriatic Sea region: the case of Corfu Island. Environ. Sci. Pollut. Res. 24, 20173–20186. https://doi.org/10.1007/s11356-017-9732-8

Kanakoudis, V., Tsitsifli, S., Papadopoulou, A., Curk, B.C., Karleusa, B., 2016. Estimating the Water Resources Vulnerability Index in the Adriatic Sea Region. Procedia Eng. 162, 476–485. https://doi.org/10.1016/j.proeng.2016.11.091

Kanakoudis, V., Tsitsifli, S., Samaras, P., Zouboulis, Demetriou, G., 2011. Developing appropriate performance indicators for urban distribution systems evaluation at Mediterranean countries. Water Util. J. 31–40.

Kanakoudis, V., Tsitsifli, S., Samaras, P., Zouboulis, A., Banovec, P., 2013. A new set

of water losses-related performance indicators focused on areas facing water scarcivty conditions. Desalin. Water Treat. 51, 2994–3010. https://doi.org/10.1080/19443994.2012.748448

Kanakoudis,V.; Tsitsifli, S.; Samaras,P.; Zoumpoulis, A.; Argyriadou, I.Water distribution system performance level integrated evaluation using new Performance Indicators based on a modified IWA Water Balance. In Proceedings of the 10th International Conference Hydroinformatics-HIC2012, Understanding Changing Climate and Environment and Finding Solutions, Hamburg, Germany, 14–18 July 2012.

Kanakoudis, V.K., Tolikas, D.K., 2001. The role of leaks and breaks in water networks: Technical and economical solutions. J. Water Supply Res. Technol. - AQUA 50, 301–311. https://doi.org/10.2166/aqua.2001.0025

Kang, H.S., Kim, H., Lee, J., Lee, I., Kwak, B.Y., Im, H., 2015. Optimization of pumping schedule based on water demand forecasting using a combined model of autoregressive integrated moving average and exponential smoothing. Water Sci. Technol. Water Supply 15, 188–195. https://doi.org/10.2166/ws.2014.104

Khan, A.S., Swerdlow, D.L., Juranek, D.D., 2001. Precautions against biological and chemical terrorism directed at food and water supplies. Public Health Rep. 116, 3–14. https://doi.org/10.1016/S0033-3549(04)50017-1

Khoshnevisan, B., Rafiee, S., Mousazadeh, H., 2014. Application of multi-layer adaptive neuro-fuzzy inference system for estimation of greenhouse strawberry yield. Meas. J. Int. Meas. Confed. 47, 903–910. https://doi.org/10.1016/j.measurement.2013.10.018

Kleczkowski, A., Maharaj, S., 2010. Stay at home, wash your hands: Epidemic dynamics with awareness of infection. Summer Comput. Simul. Conf. SCSC 2010 - Proc. 2010 Summer Simul. Multiconference, SummerSim 2010 141–146.

Kofinas, D., Mellios, N., Laspidou, C., 2015. Spatial and temporal disaggregation of water demand and leakage of the water distribution network in Skiathos, Greece S7001. https://doi.org/10.3390/ecsa-2-s7001

Kofinas, D., Mellios, N., Papageorgiou, E., Laspidou, C., 2014. Urban water demand forecasting for the island of skiathos, in: Procedia Engineering. Elsevier Ltd, pp. 1023–1030. https://doi.org/10.1016/j.proeng.2014.11.220

Kofinas, D., Papageorgiou, E., Laspidou, C., Mellios, N., Kokkinos, K., 2016. Daily multivariate forecasting of water demand in a touristic island with the use of artificial neural network and adaptive neuro-fuzzy inference system, in: 2016 International Workshop on Cyber-Physical Systems for Smart Water Networks, CySWater 2016. Institute of Electrical and Electronics Engineers Inc., pp. 37–42. https://doi.org/10.1109/CySWater.2016.7469061

Koiv, T.A., Toode, A., 2006. Trends in domestic hot water consumption in Estonian apartment buildings, in: Proc. Estonian Acad. Sci. Eng. pp. 12(1), 72–80.

Kossieris, P., Makropoulos, C., 2018. Exploring the statistical and distributional properties of residential water demand at fine time scales. Water (Switzerland) 10. https://doi.org/10.3390/w10101481

Kossieris, P., Makropoulos, C., Creaco, E., Vamvakeridou-Lyroudia, L., Savic, D.A., 2016. Assessing the Applicability of the Bartlett-Lewis Model in Simulating Residential Water Demands. Procedia Eng. 154, 123–131. https://doi.org/10.1016/j.proeng.2016.07.429

Kossieris, P., Tsoukalas, I., Makropoulos, C., Savic, D., 2019. Simulating marginal and dependence behaviour of water demand processes at any fine time scale. Water (Switzerland) 11. https://doi.org/10.3390/w11050885

Koutiva, I., Makropoulos, C., 2019. Exploring the effects of alternative water demand management strategies using an agent-based model. Water (Switzerland) 11. https://doi.org/10.3390/w11112216

La Rocca, R.A., 2014. The Role of Tourism in Planning the Smart City. TeMA, J. L. Use, Mobil. Environ. 7. https://doi.org/10.6092/1970-9870/2814

Lakshminarayan, K., Harp, S.A., Samad, T., 1999. Imputation of missing data in industrial databases. Appl. Intell. 11, 259–275. https://doi.org/10.1023/A:1008334909089

Lambert, A., 2002. Water Losses Management and Techniques International Report.

Lambert, A.O., 2002. International report: Water losses management and techniques, in: Water Science and Technology: Water Supply. IWA Publishing, pp. 1–20.

Lanzarone, G.A., Zanzi, A., 2010. Monitoring gas and water consumption through icts for improved user awareness. Inf. Commun. Soc. 13, 121–135. https://doi.org/10.1080/13691180902992962

Laspidou, C., Papageorgiou, E., Kokkinos, K., Sahu, S., Gupta, A., Tassiulas, L., 2015. Exploring patterns in water consumption by clustering. Procedia Eng. 119, 1439–1446. https://doi.org/10.1016/j.proeng.2015.08.1004

Laspidou, C.S., 2014. ICT and stakeholder participation for improved urban water management in the cities of the future. Water Util. J. 8, 79–85.

Laspidou, C.S., Kofinas, D.T., Mellios, N.K., Witmer, M., 2018. Modelling the Water-Energy-Food-Land Use-Climate Nexus: The Nexus Tree Approach. Proceedings 2, 617. https://doi.org/10.3390/proceedings2110617

Li, W., Huicheng, Z., 2010. Urban water demand forecasting based on HP filter and fuzzy neural network. J. Hydroinformatics 12, 172–184. https://doi.org/10.2166/hydro.2009.082

Liemberger, R., Farley, M., 2004. Developing a Non-Revenue Water Reduction

Strategy Part 1 : Investigating and Assessing Water Losses. Proc IWA 4th World Water Congr. Exhib. 1924 Sept. 2004 Marrakech Morocco 1–10.

Liemberger, R., Wyatt, A., 2019. Quantifying the global non-revenue water problem. Water Sci. Technol. Water Supply 19, 831–837. https://doi.org/10.2166/ws.2018.129

Lim, S.L., Quercia, D., Finkelstein, A., 2010. StakeSource: Harnessing the power of crowdsourcing and social networks in stakeholder analysis. Proc. - Int. Conf. Softw. Eng. 2, 239–242. https://doi.org/10.1145/1810295.1810340

Lima, A.R., Hsieh, W.W., Cannon, A.J., 2017. Variable complexity online sequential extreme learning machine, with applications to streamflow prediction. J. Hydrol. 555, 983–994. https://doi.org/10.1016/j.jhydrol.2017.10.037

Little, R., Rubin, D.B., 2002. Statistical Analysis with Missing Data Second Edition. John Wiley & Sons.

Liu, G., Zhang, L., He, B., Jin, X., Zhang, Q., Razafindrabe, B., You, H., 2015. Temporal changes in extreme high temperature, heat waves and relevant disasters in Nanjing metropolitan region, China. Nat. Hazards 76, 1415–1430. https://doi.org/10.1007/s11069-014-1556-y

M. B. Singleton, Liu, Q., 2017. Machine Learning and Intelligent Communications, Machine Learning and Intelligent Communications. https://doi.org/10.1007/978-3-319-52730-7

Machiwal, D., Jha, M.K., 2012. Hydrologic time series analysis: Theory and practice, Hydrologic Time Series Analysis: Theory and Practice. https://doi.org/10.1007/978-94-007-1861-6

Maidment, D.R., Miaou, S. -P, 1986. Daily Water Use in Nine Cities. Water Resour. Res. 22, 845–851. https://doi.org/10.1029/WR022i006p00845

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. Environ. Model. Softw. https://doi.org/10.1016/S1364-8152(99)00007-9

Makropoulos, C., Savíc, D.A., 2019. Urban hydroinformatics: Past, present and future. Water (Switzerland) 11, 1–17. https://doi.org/10.3390/w11101959

Makropoulos, C.K., Natsis, K., Liu, S., Mittas, K., Butler, D., 2008. Decision support for sustainable option selection in integrated urban water management. Environ. Model. Softw. 23, 1448–1460. https://doi.org/10.1016/j.envsoft.2008.04.010

Mamo, T., Juran, I., Shahrour, I., 2013. Urban Water Demand Forecasting Using the Stochastic Nature of Short Term Historical Water Demand and supply Pattern. J. Water Resour. … 2, 92–103.

Martinez-Espiñeira, R., 2002. Residential water demand in the Northwest of Spain. Environ. Resour. Econ. 21, 161–187. https://doi.org/10.1023/A:1014547616408

McKenzie, E., 1984. General exponential smoothing and the equivalent arma process. J. Forecast. 3, 333–344. https://doi.org/10.1002/for.3980030312

Mellios, N., Kofinas, D., Papageorgiou, E., Laspidou, C., 2015. a Multivariate Analysis of the Daily Water Demand of Skiathos Island, Greece, Implementing the Artificial Neuro-Fuzzy Inference System (Anfis).

Miaou, S. -P, 1990. A class of time series urban water demand models with nonlinear climatic effects. Water Resour. Res. 26, 169–178. https://doi.org/10.1029/WR026i002p00169

Milano, M., Ruelland, D., Fernandez, S., Dezetter, A., Fabre, J., Servat, E., Fritsch, J.M., Ardoin-Bardin, S., Thivet, G., 2013. Current state of Mediterranean water resources and future trends under climatic and anthropogenic changes. Hydrol. Sci. J. 58, 498–518. https://doi.org/10.1080/02626667.2013.774458

Minister of Maritime Affairs Islands and Fisheries- Greek Republic, 2018. hcg.gr [WWW Document]. Renov. Water Distrib. Networks 9 islandsfile///C/Users/user/AppData/Local/Temp/YF_15_02_18_12-1.doc. URL http://www.hcg.gr/node/17169 (accessed 11.19.19).

Mohammed, J.R., Ibrahim, H.M., 2012. Hybrid Wavelet Artificial Neural Network Model for 7, 1047–1065.

Mol, A.P.J., 2010. Sustainability as global attractor: The greening of the 2008 Beijing Olympics. Glob. Networks 10, 510–528. https://doi.org/10.1111/j.1471-0374.2010.00289.x

Moore, D.S., Notz, W.I., Flinger, M.A., 2013. The basic practice of statistics, 6th ed. W. H. Freeman and Company, New York.

Morley, M.S., Tricarico, C., 2008. Pressure Driven Demand Extension for EPANET (EPANETpdd).

Nam, T., Pardo, T.A., 2011. Conceptualizing smart city with dimensions of technology, people, and institutions. ACM Int. Conf. Proceeding Ser. 282–291. https://doi.org/10.1145/2037556.2037602

Newton, I., 1999. The Principia: mathematical principles of natural philosophy. University of California Press.

Novak, J., Melenhorst, M., Micheel, I., Pasini, C., Fraternali, P., Rizzoli, A.E., 2016. Behaviour change and incentive modelling for water saving : first results from the SmartH2O project Behaviour change and incentive modelling for water. Proc. 8th Int. Congr. Environ. Model. Softw.

Ömer Faruk, D., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Eng. Appl. Artif. Intell. 23, 586–594. https://doi.org/10.1016/j.engappai.2009.09.015

Palani, S., Liong, S.Y., Tkalich, P., 2008. An ANN application for water quality

forecasting. Mar. Pollut. Bull. 56, 1586–1597. https://doi.org/10.1016/j.marpolbul.2008.05.021

Perren, K., Yang, L., 2015. Psychosocial and behavioural factors associated with intention to save water around the home: A Greek case study. Procedia Eng. 119, 1447–1454. https://doi.org/10.1016/j.proeng.2015.08.1005

Puust, R., Kapelan, Z., Savic, D.A., Koppel, T., 2010. A review of methods for leakage management in pipe networks. Urban Water J. 7, 25–45. https://doi.org/10.1080/15730621003610878

Rao, J.N.K., 1996. On Variance Estimation with Imputed Survey Data. J. Am. Stat. Assoc. 91, 499–506. https://doi.org/10.1080/01621459.1996.10476910

Raskin, P., Gleick, P.H., Stockholm Environment Institute., 1997. Water futures : assessment of long-range patterns and problems. Stockholm Environment Institute.

Rockaway, T.D., Coomes, P.A., Rivard, J., Cornstein, B., 2011. Residential water use trends in North America. Am. Water Work. Assoc. 1992, 20–22.

Rodriguez-Iturbe, I., Cox, D. R., Isham, V., 1987a. Some Models for Rainfall Based on Stochastic Point Processes. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 410, 269–288. https://doi.org/10.1098/rspa.1987.0039.

Rojas, R., 1996. Neural Networks: A Systematic Introduction. Springer-Verlag, New York.

Rossman, L.A., 2000. EPANET 2 Users Manual. U.S. Environmental Protection Agency, Washington, D.C.

Rozos, E., Makropoulos, C., 2013. Source to tap urban water cycle modelling. Environ. Model. Softw. 41, 139–150. https://doi.org/10.1016/j.envsoft.2012.11.015

Rumelhart, D.E., Hinton, G.E., Mcclelland, J.L., 1986. A General framework for Parallel Distributed Processing. Parallel Distrib. Process. Explor. Microstruct. Cogn. 45–76.

Russell, S., Fielding, K., 2010. Water demand management research: A psychological perspective. Water Resour. Res. 46. https://doi.org/10.1029/2009WR008408

Sadat, N., 2012. Case study of Iranian immigrants in Sweden.

Särndal, C.E., Swensson, B., Wretman, J., 2003. Model assisted survey sampling. Springer Science & Business Media.

Schafer, J.L., Graham, J.W., 2002. Missing data: Our view of the state of the art. Psychol. Methods 7, 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Seago, C.J., Mckenzie, R.S., Liemberger, R., 2005. International Benchmarking of Leakage from Water Reticulation Systems IWA Methodology and Performance Indicators 1–14.

Sensus, SWAN, 2012. Water 20/20 Bringing Smart Water Networks into Focus [WWW Document]. URL https://www.swan-forum.com/wp-content/uploads/sites/218/2016/05/sensus_water2020-usweb.pdf (accessed 11.14.19).

Shan, Y., Yang, L., Perren, K., Zhang, Y., 2015. Household water consumption: Insight from a survey in Greece and Poland. Procedia Eng. 119, 1409–1418. https://doi.org/10.1016/j.proeng.2015.08.1001

Singh, R., Kainthola, A., Singh, T.N., 2012. Estimation of elastic constant of rocks using an ANFIS approach. Appl. Soft Comput. J. 12, 40–45. https://doi.org/10.1016/j.asoc.2011.09.010

Sitarz, D., 1993. Agenda 21: The Earth summit strategy to save our planet. Boulder, United States.

Skworcow, P., Ulanicki, B., 2011. Burst detection in water distribution systems via active identification procedure. Conf. Comput. Control Water Ind.

Sophocleous, S., Savić, D., Kapelan, Z., 2019. Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction. J. Water Resour. Plan. Manag. 145, 04019024. https://doi.org/10.1061/(asce)wr.1943-5452.0001079

Spilanis, I., Kizos, T., Koulouri, M., Kondyli, J., Vakoufaris, H., Gatsis, I., 2009. Monitoring sustainability in insular areas. Ecol. Indic. 9, 179–187. https://doi.org/10.1016/j.ecolind.2008.03.003

Spilianis, G., Kizos, T., 2015. Atlas of Islands. University of the Aegean, Mitilini.

Spilianis, G., Vayanni, E., 2004. 13 Sustainable Tourism: Utopia or Necessity? The Role of New Forms of Tourism in the Aegean Islands, in: Bramwell, B. (Ed.), Coastal Mass Tourism: Diversification and Sustainable Development in Southern Europe. Channelview publications.

Spyropoulou, A., Lazarou, Y.G., Laspidou, C., 2018. Mercury Speciation in the Water Distribution System of Skiathos Island, Greece. Proceedings 2, 668. https://doi.org/10.3390/proceedings2110668

Staats, H., Harland, P., Wilke, H.A.M., 2004. Effecting durable change: A team approach to improve environmental behavior in the household. Environ. Behav. 36, 341–367. https://doi.org/10.1177/0013916503260163

Stakiadis, G., Papanikolaou, P., 2007. Reliability laboratory exploration of household water meters, in: 6th National Conference of EEDYP, 14-16 June 2007. Chania, Crete, Greece.

Stevens, T.H., Miller, J., Willis, C., 1993. Effect of Price Structure on. Water Resour. Bull. 28, 681–685.

Su, C.T., Tong, L.I., Leou, C.M., 1997. Combination of time series and neural network for reliability forecasting modeling. J. Chinese Ind. Eng. 14, 419–429.

Sun, C., Puig, V., Cembrano, G., 2020. Real-Time Control of Urban Water Cycle under Cyber-Physical Systems Framework. Water 12, 406. https://doi.org/10.3390/w12020406

Takagi, T., Sugeno, M., 1985. Fuzzy Identification of Systems and Its Applications to Modeling and Control. IEEE Trans. Syst. man, Cybern. SMC-15, 116–132. https://doi.org/10.1016/s1474-6670(17)42481-5

Tillman, T., Larsen, T.A., Pahl-Wostl, C., Gujer, W., 2001. Interaction analysis of stakeholders in water supply systems. Water Sci. Technol. 43, 319–326. https://doi.org/10.2166/wst.2001.0316

Tsoukalas, I., Efstratiadis, A., Makropoulos, C., 2019. Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure. J. Hydrol. 575, 354–380. https://doi.org/10.1016/j.jhydrol.2019.05.017

Ulanicki, B., Bounds, P.L.M., Rance, J.P., Reynolds, L., 2000. Open and closed loop pressure control for leakage reduction. Urban Water 2, 105–114. https://doi.org/10.1016/s1462-0758(00)00048-0

Wang, Z., Capiluppi, A., 2015. A Social-Centred Gamification Approach to Improve Household Water Use Efficiency. VS-Games 2015 - 7th Int. Conf. Games Virtual Worlds Serious Appl. https://doi.org/10.1109/VS-GAMES.2015.7295761

White, S., Jha, M., Milne, G., 2003. Urban water demand forecasting and demand management : research needs review and recommendations. Water Services Association of Australia, Occasional Paper No. 9.

Wilhite, H., Ling, R., 1995. Measured energy savings from a more informative energy bill. Energy Build. 22, 145–155. https://doi.org/10.1016/0378-7788(94)00912-4

Winarni W., 2009. Infrastructure Leakage Index (ILI) as Water Losses Indicator. Civ. Eng. Dimens. 11, 126–134.

Winters, P.R., 1960. Forecasting Sales by Exponentially Weighted Moving Averages. Manage. Sci. 6, 324–342. https://doi.org/10.1287/mnsc.6.3.324

Wu, Z.Y., Wang, R.H., Walski, T.M., Yang, S.Y., Bowdler, D., Baggett, C.C., 2007. Efficient pressure dependent demand model for large water distribution system analysis. 8th Annu. Water Distrib. Syst. Anal. Symp. 2006 40941, 39. https://doi.org/10.1061/40941(247)39

Yang, L., Yang, S.H., Magiera, E., Froelich, W., Jach, T., Laspidou, C., 2017. Domestic

water consumption monitoring and behaviour intervention by employing the internet of things technologies. Procedia Comput. Sci. 111, 367–375. https://doi.org/10.1016/j.procs.2017.06.036

Zhang, G., Eddy Patuwo, B., Y. Hu, M., 1998. Forecasting with artificial neural networks: The state of the art. Int. J. Forecast. 14, 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

Zhang, P.G., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhou, Q., Jiang, H., Wang, J., Zhou, J., 2014. A hybrid model for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network. Sci. Total Environ. 496, 264–274. https://doi.org/10.1016/j.scitotenv.2014.07.051

Eurostat,2019: https://ec.europa.eu/eurostat/databrowser/view/t2020_rd220/default/table?lang=en

https://www.unwater.org/high-level-political-forum-sustainable-development-2019/?utm_source=IWA-NETWORK&utm_campaign=84336ba8a8-IWA+Newsletter+July+2019+members&utm_medium=email&utm_term=0_c457ab9803-84336ba8a8-%5BLIST_EMAIL_ID%5D&ct=t%28EMAIL_IWA+Newsletter+July+2019_members%29

https://ec.europa.eu/environment/pubs/pdf/factsheets/water_scarcity.pdf

https://ec.europa.eu/eurostat/web/products-datasets/-/t2020_rd220

http://www.fondazioneamga.org/public/janssensgemaggio06.pdf