D3.1 FIWARE-enabled applications for Raw Water Supply

Author(s): Panagiotis Kossieris (NTUA), Vasilis Bellos (NTUA), Christos Makropoulos (NTUA), Elias Papakonstantis (NTUA)

Co-Author(s): Vasiliki Polychniatou (EYDAP), Panagiotis Dimas (NTUA), Sandra Baki (NTUA), Dionisios Nikolopoulos (NTUA), Ioannis Tsoukalas (NTUA), Kostas Eftaxias (NTUA)

April 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement No. 821036.

Disclaimer

This document reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Intellectual Property Rights

Project Consortium

Executive Summary

Fiware4Water (F4W) intends to link the water sector to the FIWARE smart solution platform by demonstrating its capabilities and the potential of its interoperable and standardised interfaces for both water sector end-users (cities, water utilities, water authorities, citizens and consumers), and solution providers (private utilities, SMEs, developers).

This deliverable reports the smart applications (algorithms and models) developed, in the framework of Task 3.1 (WP3), to support the optimal operation and management of the raw-water conveyance system that serves the metropolitan area of the city of Athens (Greece; Demo Case 1). The demo case part of the system is the “Giona – Dafnoula” aqueduct, with a total length of 131 km. The applications developed aim to support the operational staff of EYDAP (the Athens Water Supply and Sewerage Company that operates the system) in decision making with respect to flow and quality conditions of the conveyance system. In summary, three new applications have been developed to: a) provide advice on optimal sluice gate settings (openings) depending on the flow conditions and needs for water supply, b) provide early warnings for high turbidity events and one-hour ahead forecasts of the level of turbidity at the most downstream part of the system, which are close to the water treatment plants, and finally c) to provide one-day ahead forecasts of total water outflows from the water treatment plants. It is worth to highlight that these applications are innovative both with respect to the modelling approaches to address specific scientific problems, as well as to their operational character. Essentially, this is the first time that such decision support services become available to the operational staff of EYDAP, in an operational context.

Prior to F4W, the regulation of flow across the conveyance system was based on empirical rules. However, this management policy, which is strongly based on operators’ knowledge and experience, is neither sustainable nor safe from a resilience perspective. Furthermore, the system is subject to occasional failures, due to undesirable overflows resulting in non-negligible water losses. To support the optimal operation and scheduling of the regulation structures (Λ-type structures) which control the flow in the channel under study, in a systematic and automated way a model has been developed. The model follows a “grey-box” approach, combining physics-driven hydraulic equations, to simulate the flow through sluice gates and over spillways, with data-driven techniques for the estimation of the key parameters (e.g., coefficient of discharge) of the algorithm. The model estimates the new openings of sluice gates for a target flow, given the upstream and downstream water depths and flow, available from sensors on real-time basis. Furthermore, it estimates the time required for the flow to travel downstream, in other words, the response time between the time moment of a control action (either opening or closing the sluice gates) and the time moment when the downstream flowmeter captures this modification.

The project focused also on quality aspects of the raw-water system. Specifically, data-driven models have been developed to provide forecasts of the level of turbidity at the most downstream parts of the system under study, before raw-water reaching the 4 water treatment plants that serve the city of Athens. Specifically, we built two deep neural network models, using Long-short term memory kernels, to forecast one-hour ahead the level of turbidity at the most downstream quality metering stations, using as predictors the level of turbidity at the upstream distant metering stations. The model is accompanied by a threshold-based early warning system that notifies the operators for unusual high-turbidity events which may appear at the 6 water quality sensors existing in the system under study.

Furthermore, taking advantage of data provided by EYDAP, we developed a harmonic regression model to forecast one-day ahead the total daily water outflows from the 4 water treatment plants, and hence enabling operators to regulate the flow in the channel accordingly. The model gives special focus on

the reproduction of annual and weekly seasonality exhibiting in water outflows, as well as their significant variability, during periods of exceptional demand events (e.g., Easter holidays). It is worth mentioning that this model is not foreseen in the description of work of the Grant Agreement of the project, and it is developed as an initiative of NTUA, after the needs of operational staff of EYDAP.

The three above mentioned applications are currently in an operational context, consuming real-time data from the FIWARE Context Broker that has been deployed in the framework of F4W. The outcomes of the applications are available to operational staff of EYDAP via the new web platform that has been also developed in the project and is presented in Deliverable 4.1 ("D4.1: FIWARE4_Raw water supply system real-time operational management" [M35]).

As discussed in Section VI, the developed applications have great potential for further uptake both within EYDAP as well as in other large-scale raw-water conveyance systems in Europe.

Related Deliverables

D1.1: "Requirements from Demo Cases" and D1.2: "Requirements from end-users", which describe the requirements of the smart solutions developed in WP3.

D2.1: Specification of system architecture for water management, cybersecurity and quality monitoring, which provide guidelines to implement FIWARE-enabled architectures.

D4.1: FIWARE4_Raw water supply system real-time operational management, which describes the FIWARE-enabled deployed in Task 4.1 (WP4) for Athens Demo Case.

Document Information

Programme	H2020 – SC0511-2018
Project Acronym	**Fiware4Water**
Project full name	FIWARE for the Next Generation Internet Services for the WATER sector
Deliverable	**D3.1 FIWARE-enabled applications for Raw Water Supply**
Work Package	**WP3: Smart applications and devices**
Task	Task 3.1: Smart Applications for RawWater Supply
Lead Beneficiary	NTUA
Author(s)	Author(s): Panagiotis Kossieris (NTUA), Vasilis Bellos (NTUA), Christos Makropoulos (NTUA), Vasiliki Polychniatou (EYDAP)
Contributor(s)	Vasiliki Polychniatou (EYDAP), Panagiotis Dimas (NTUA), Sandra Baki (NTUA), Dionisios Nikolopoulos (NTUA), Ioannis Tsoukalas (NTUA), Kostas Eftaxias (NTUA)
Quality check	Lluís Echeverria (EURECAT)
Planned Delivery Date	30/04/2022
Actual Delivery Date	30/04/2022
Dissemination Level	Confidential, only for members of the consortium (including the Commission Services)

Revision history

Version	Date	Author(s)/Contributor(s)	Notes
Draft1	19/04/2022	Author(s): Panagiotis Kossieris (NTUA), Vasilis Bellos (NTUA), Christos Makropoulos (NTUA), Vasiliki Polychniatou (EYDAP)	
Draft2	25/04/2022	Lluís Echeverria (EURECAT)	Quality check and revision.
Final	30/04/2022	Author(s): Panagiotis Kossieris (NTUA), Vasilis Bellos (NTUA), Christos Makropoulos (NTUA), Vasiliki Polychniatou (EYDAP)	
Final V2	21/07/2022		Revision after reviwers’ comments

Table of content

Executive Summary ... 1
List of figures ... 5
List of tables ... 6
List of Acronyms/Glossary ... 6
Introduction ... 7
I. The raw-water conveyance system of Athens ... 8
I.1. Demo case description ... 8
I.2. Data sources and monitoring system ... 11
I.3. Applications to support the management of raw-water supply system of EYDAP ... 11
II. Optimal sluice gate openings and scheduling ... 12
II.1. Description of the case study ... 12
II.1. Available data sources ... 16
II.2. Theoretical background ... 16
II.3. Pre-processing and data cleaning of data ... 18
II.4. Global calibration: the sluice gate discharge coefficient ... 21
II.5. Global calibration: the Manning coefficient ... 22
II.6. Operational model for optimal sluice gate opening and scheduling ... 23
i. Fundamentals ... 23
ii. Input data ... 24
iii. Estimate sluice gate openings ... 25
iv. Estimate response time ... 28
III. Forecast level of raw-water turbidity and warning for high-turbidity events ... 29
III.1. Data sources and features extraction ... 29
III.2. Model to forecast the level of turbidity ... 33
IV. Analysis and forecast of water supply volumes ... 38
IV.1. Data sources and features extraction ... 38
IV.2. Forecast daily water outflows ... 40
V. Conclusion and Perspectives ... 43
VI. European added value (EAV) and upscaling ... 44
References ... 46

List of figures

Figure 1: Schematic representation of the external water supply system of EYDAP 8
Figure 2: Schematic representation of the external water supply system of EYDAP 10
Figure 3: Part of the external raw water supply system of EYDAP (from Giona to Dafnoula) where the F4W solutions will be demonstrated. The markings along the supply system represent the installed water level meters.. 10
Figure 4: General view of the Λ7 - Λ11 aqueduct. 13
Figure 5: Part of the Mornos channel, between Λ7 - Λ11 regulation structures. 13
Figure 6: Trapezoidal cross-section of the conveyance system 14
Figure 7: Λ-type regulation structures composed by sluice gates and lateral broad crested weirs 14
Figure 8: Flow over the lateral spillways 15
Figure 9: Schematic representation of water depths at the case study 17
Figure 10: Top view and longitudinal section of a Λ-type structure 18
Figure 11: First step of time series handling 19
Figure 12: Second step of time series handling 20
Figure 13: Third step of time series handling 20
Figure 14: Calibrated sluice gate discharge coefficient (global calibration phase) and comparison with theoretical values 22
Figure 15: Calibrated Manning coefficient against theoretical values 23
Figure 16: Flow chart of the proposed numerical model 24
Figure 17: Output of the numerical model: target discharge is bigger than the current. 26
Figure 18: Output of the numerical model: target discharge is smaller than the current. 27
Figure 19: Output of the numerical model: significant discrepancy between left and right opening. . 27
Figure 20: Water quality meters across the demo part of the external raw water supply system of EYDAP (from Giona to Dafnoula)... 29
Figure 21: Indicative example of the spiky behaviour of turbidity in the external conveyance system of EYDAP (red dots indicates sudden peaks). 30
Figure 22: Sudden peaks (local maxima) detection of turbidity time series at L8 metering station. ... 30
Figure 23: Turbidity time series after detecting sudden peaks and replacing them with neighboring values.. 31
Figure 24: Histogram of turbidity data at the 6 metering stations. Blue dotted line indicates the median, while red dotted lines indicate the 5% and 95% quantiles. 31
Figure 25: Box plots comparing turbidity data at the 6 metering stations. 32
Figure 26: High turbidity event travelling across the “Giona-Dafnoula” aqueduct. 32
Figure 27: Model accuracy for the prediction of the level of turbidity (4 classes) at L10 station, 1-hour ahead, using 320 observations from the 4 upstream stations.. 35
Figure 28: Model accuracy for the prediction of the level of turbidity (4 classes) at L10 station, 1-hour ahead, using 100 observations from the 4 upstream stations.. 35
Figure 29: Model accuracy for the prediction of the level of turbidity (2 classes) at L10 station, 1-hour ahead, using 320 observations from the 4 upstream stations.. 36
Figure 30: Model accuracy for the prediction of the level of turbidity (2 classes) at L10 station, 1-hour ahead, using 100 observations from the 4 upstream stations.. 36
Figure 31: Model accuracy for the prediction of the level of turbidity (2 classes) at Dafnoula station, 1-hour ahead, using 100 observations from the 5 upstream stations. 37
Figure 32: Time series of total daily water outflows from the 4 WTPs of Athens. 38
Figure 33: Seasonal variation of mean daily outflow per month, for the period 2015 – 2021............ 38

Figure 34: Variation of mean daily outflow per day of the week, for the period 2015 – 2021. 39
Figure 35: Box plot of daily water outflows during Easter Holidays, for the period 2015 – 2021. 39
Figure 36: Box plot of daily water outflows during Christmas Holidays, for the period 2015 – 2021. . 40
Figure 37: Observed (black line) vs. predicted (red line) outflows for the training set. 41
Figure 38: Observed (black line) vs. predicted (red line) outflows for the validation set. 42

List of tables

Table 1: Channel characteristics 15
Table 2: Λ-type geometrical characteristics 15
Table 3: Monitoring system characteristics for the case study 16
Table 4: Additional gauges installed in the case study during the F4W project. 16
Table 5: Typical examples of input data and output derived by the tool 27
Table 6: Estimated time required for a turbidity event to propagate among metering stations 33
Table 7: Evaluation measures for the training and validation dataset 42

List of Acronyms/Glossary

CB Context Broker

DW Data warehouse (of EYDAP) – Database that stores the data from the legacy system of EYDAP

EYDAP Athens Water Supply and Sewerage Company

EAV European Added Value

F4W Fiware4Water project

IoT Internet of Things

LSTM Long-Short term Memory

NN Neural network

NTUA National Technical University of Athens

WPL Work Packages Leaders

WTP Water Treatment Plant

Introduction

This deliverable presents the smart applications (algorithms and models) developed, in the framework of Task 3.1 (WP3), to support the optimal operation and management of the raw-water conveyance system that serves the metropolitan area of the city of Athens (Greece; Demo Case 1). The scope, context and functionalities of the applications have been identified in close collaboration between the National Technical University of Athens (NTUA) and the Athens Water Supply and Sewerage Company (EYDAP), which operates the conveyance system. The key functionalities and system requirements of the applications were identified via a series of three workshops between NTUA (scientific partner) and EYDAP (end-user), as part of the activities in WP1, and they are described in the form of user stories and use cases in Deliverable 1.1 ("D1.1: Requirements from Demo Cases" [M8]). The use cases concern both functionalities for processing and visualization of data from the existing sensors, to enable the real-time monitoring of the conveyance system, as well as smart applications, based on algorithms and models, which take advantage of real-time data to provide support and advice to the operators of the system towards its optimal management. The FIWARE-enabled applications that have been developed, concern both hydraulic (e.g., flow, water depths) and quality (e.g., turbidity, temperature, conductivity) aspects of raw-water in the conveyance system.

The functionalities related to the analysis, processing and visualization of real-time and historical data from the existing sensors are described in Deliverable 4.1 ("D4.1: FIWARE4_Raw water supply system real-time operational management" [M35]), as part of analytics of the web platform, which has been also developed in the framework of Fiware4Water (F4W) for EYDAP. In this report, we focus on the development of algorithms and models that transform real-time data into meaningful information for the operators of the system. Specifically, here we present the pre-processing and analysis steps of the involved data and time series, as well as the algorithmic part of the following use cases:

- Advice provision for optimal sluice-gate operation (Section II)
- Early warning and forecast of turbidity events (Section III)
- Analysis and forecast of the water volumes conveyed by the system on daily basis (Section IV)

A brief description of the raw-water conveyance system under study is given in Section I, while Sections V and VI conclude the work conducted and present the upscale opportunities and EU-added value.

The implementation of the algorithms in an operational context, as part of the new FIWARE-enabled web platform, is presented in Deliverable 4.1.

I. The raw-water conveyance system of Athens

I.1. Demo case description

Athens Water Supply and Sewerage Company (EYDAP S.A.) is the largest water company in Greece, serving approximately 4,400,000 customers (2,160,000 water meters) in the region of Attica (Athens), through water pipelines of total length greater than 14,000 km. The sewerage sector serves 3,500,000 residents with sewers spreading at almost 9,500 km. The Company's main objects is to provide water supply and sewerage services and to design, construct, install, operate, manage, maintain, expand and upgrade water supply and sewerage systems.

EYDAP uses raw water mainly from 4 surface water resources:

- Marathonas reservoir (maximum capacity: 41 million m^3, operation capacity: 34 million m^3),
- Yliki lake (maximum capacity: 590 million m^3, operation capacity: 580 million m^3),
- Mornos reservoir (maximum capacity: 764 million m^3, operation capacity: 630 million m^3),
- Evinos reservoir (maximum capacity: 138 million m^3, operation capacity: 113 million m^3)

Additionally, the Company's water sources also include underground water resources which are exploited by 100 boreholes.

The transfer of raw water sources (reservoirs and boreholes) to the 4 Water Treatment Plants (WTP) of Attica is accomplished via an extensive external aqueduct system with a total length of 495km. The schematic representation of the external water supply system of EYDAP, along with all the constructions, is given in Figure 1.

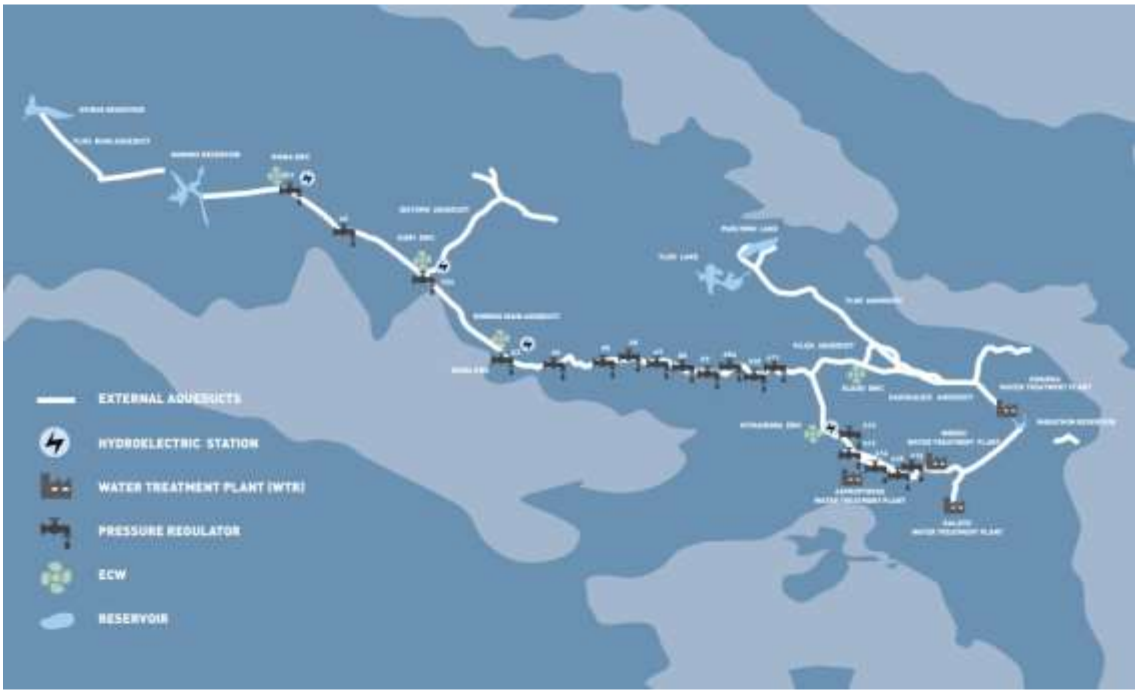

Figure 1: Schematic representation of the external water supply system of EYDAP

The Evinos reservoir communicates with the Mornos reservoir through the Evinos – Mornos Tunnel. Water from the Evinos Reservoir is transported and channelled to the Mornos reservoir to boost water reserves. The operation of the Tunnel is under pressure to provide 27 m^3/s of water. The Evinos-Mornos tunnel has a total length of 29.4 m and an inner diameter of 3.5 m. The construction of the tunnel started in 1992 and was completed in 2 years. It is one of the longest hydraulic tunnels in the world realised with the tunnel boring machine method.

The transfer of raw water sources (reservoirs and boreholes) to the Water Treatment Plants (WTP) of Attica is accomplished via an extensive external aqueduct system with a total length of 495km. The aqueducts are divided into:

- Main aqueducts (Mornos, Yliki) with a total length of 310 km
- Connecting aqueducts (Mornos-Yliki, Marathonas-Galatsi, Distomo) with a total length of 105 km
- Auxiliary aqueducts, with a total length of 80 km

The aqueducts of Mornos and Yliki communicate with each other via connecting aqueducts. The existence of connecting aqueducts allows control and maintenance of the two aqueducts, with the option of closure of one of the two. In addition, it provides the option of alternative modes of exploitation of water resources, depending on the hydrological conditions and consumption needs. The Yliki Aqueduct has a total length of 67 km and works with pumping. Water from the Mornos reservoir is carried to Athens via a gravity conveyance system with a total length of 188 km. The daily capacity of Mornos to Thiva aqueduct is 23 m^3/s and it splits on the 146th km in the Kethairon division. After the division, the main branch has a capacity of 11.5 m^3/s and a length of 43 km and it goes to Athens. The other branch meets with the Yliki aqueduct, which has a capacity of 4.2 m^3/s and a length of 17.85 km.

Despite the long distance of the main water sources from Attica, the largest amount of water is being carried via gravity without the financial and environmental burden that the energy-intensive pumps cause, which are activated only in case of an emergency.

Constant monitoring of the external aqueduct system operation is crucial for EYDAP, in order to ensure and certify the excellent quality of water and services provided to the citizens of Athens. For this reason, daily quality control tests of the raw water are conducted, with cutting edge analysis methods in the chemical and microbiological laboratories of EYDAP. In addition, systems for the on-line monitoring of critical qualitative parameters are installed at crucial positions at the external aqueducts, that send real-time data, with the use of telemetry.

To ensure high reliability of operations, EYDAP is upgrading its supervisory system and digital water strategy and is keen to look and test alternatives to (a) facilitate the integration of different sensors from different vendors into a common system; (b) facilitate the development of different applications (models, analytics) by different developers on the data available; and (c) interface seamlessly with and provide added value to legacy systems (sensors and online control systems).

Towards this direction, in the framework of F4W, FIWARE technology was employed to integrate data sources from different operational sensors (existing flow meters, water level and water quality sensors, as well as new sensors, purchased and installed in the context of the project) into a common operational information system (platform). Taking advantage of these integrated data sources, a series of FIWARE-compliant analytics and models were developed in F4W project to provide operational decision support to the operational staff of EYDAP, targeting both quantitative (flow) and qualitative aspects of raw-water of the conveyance system.

During the workshops conducted, between EYDAP and NTUA, in the framework of WP1 activities (see Deliverable D1.1 for further details), it was decided that the “Giona - Dafnoula” part of the conveyance system is the most suitable one for the development and deployment of FIWARE-enabled solutions (i.e., algorithms, models, analytics and web platform). It is worth to notice that in the Grant Agreement

(WP4.1, Task 4.1) it was initially stated that 'most probably the "Amfissa – Dafnoula" aqueduct for the water quantity applications and part of the Yliki aqueduct for the water quality applications' would be used for the Greek Demo Case. In the first place, the "Giona – Dafnoula" part of the aqueduct is actually the same part of the aqueduct as the "Amfissa – Dafnoula" part. It is just more accurately termed. Regarding the water quality application, it was decided that it would be better to use the "Giona – Dafnoula" part of the aqueduct as well. Thus, the demonstration of both F4W applications could be integrated and be more complete. Since the construction of the Mornos aqueduct, the Yliki aqueduct is used only in case of emergencies or during maintenance works of the main aqueduct, therefore monitoring of its water quality is useful only in these rare occasions. The part of the conveyance system studied in F4W is displayed in Figure 2, where the two red dots indicate the most upstream (Giona) and most downstream (Dafnoula) points of the demo case of Athens. A more detailed view of the demo case is given in Figure 3.

Figure 2: Schematic representation of the external water supply system of EYDAP

Figure 3: Part of the external raw water supply system of EYDAP (from Giona to Dafnoula) where the F4W solutions will be demonstrated. The markings along the supply system represent the installed water level meters.

I.2. Data sources and monitoring system

In the demo case part of the water supply system (Giona – Dafnoula aqueduct), the flow and quality conditions of raw-water are monitored via a network of sensors. The on-line monitoring system for this part includes in total:

- 5 open channel flowmeters
- 46 water level meters
- 6 water quality meters, measuring turbidity, conductivity and temperature

In the framework of F4W, 5 new water level meters were installed to support the development of the model that supports advice on optimal sluice gate settings. Analysis of data from existing water level meters showed that the transmitted series is characterized by a high percentage of missing values due to connection problems, while field visits to the conveyance system revealed that in some parts of the system the existing meters overestimate or underestimate the real water depth in the channel. Due to this, 4 out of 5 new meters were installed at positions upstream and downstream of two regulation structures, where water level meters already exist but their measurements were found of bad quality. The 5th level meter was installed in a position of the conveyance system which is not currently monitored (between two sluice gates) to allow further insights to derive on critical hydraulic parameters of the system (i.e., estimation of Manning coefficient). Further details on the metering system of the demo case part of the system, and the new devices installed, are given in Deliverable D4.1.

Further to real-time data from the sensors in the system, EYDAP also provided the available time series of daily inflows and outflows at the 4 WTPs. These time series were used for the development of analytics of the web platform for the analysis of water volumes (see Deliverable D4.1), as well as for the development of the model that estimates total water outflows of next day (Section IV).

Details of the time series used in the development of algorithms and models are given in the respective section of this document.

I.3. Applications to support the management of raw-water supply system of EYDAP

As discussed in the introductory section, the applications developed in the framework of F4W concern: a) a new web platform that integrates (using FIWARE technology) the data sources from different sensors of the system, hence allowing end-users (operational staff of EYDAP) to monitor and analyse data on real-time basis, and b) a series of models that use these data sources to produce outcomes that support end-users in decision making. A detailed description of the new web platform is given in Deliverable D4.1, as part of the deployment of the application, while here we present the models developed. Specifically, in this report, we present:

- A grey-box model that estimates new sluice gate openings to establish specific flow conditions in the conveyance system (see Section II).
- A data-driven model to forecast the level of turbidity at the downstream part of the conveyance system (see Section III).
- A model that estimates one-day ahead estimation of water supply volumes (see Section IV).

It is worth to mention that the last model is not foreseen in the description of work of the Grant Agreement of the project, and it was developed as an initiative of NTUA, taking into account the operational needs of EYDAP.

II. Optimal sluice gate openings and scheduling

As presented in Section I, the raw water conveyance system of Athens (Greece) is a complex infrastructure comprising around 500 km of aqueducts, conveying water from four reservoirs to four WTPs, while serving several other local users. Here, we focus on “Giona-Dafnoula” open-channel aqueduct of Mornos channel, which is one of the most important parts of the system.

This part of the system has a dual operation, namely water conveyance and flow regulation through temporary storage along the channel. This is achieved by a series of Λ-type structures (see Figure 6), each one comprising sluice gates for flow control and a lateral broad crested weir. The details of this type of structures is explained in the section below. Currently, the flow regulation across the channel is performed through empirical rules, according to the daily water volumes requested by the operators of the downstream WTPs.

However, this management policy, which is strongly based on expert’s knowledge, is neither sustainable nor safe from a resilience perspective. Furthermore, the system is subject to occasional failures, due to undesirable overflows resulting to non-negligible water losses.

In order to establish an optimal control policy, we developed an operational tool for the optimal real-time operation and scheduling of the sluice gates of Λ-type structures. The tool, along with other analytics and algorithms developed, has been seamlessly integrated with the existing legacy system (e.g., SCADA, databases) of the system’s operator using the FIWARE standardization protocol.

II.1. Description of the case study

Following the requirements of the operational staff of EYDAP, as case study, for the development and implementation of the tool for the real-time operation and scheduling of the sluice gates, we selected the aqueduct between the Λ7 and Λ11 regulation structure (Figure 4 and Figure 5). This consists of the downstream part of the “Giona – Dafnoula” aqueduct (see Section I.1) studied in the framework of F4W project. The case study has a length of 29.1 km and consists of a free-surface channel with a trapezoidal cross-section (Figure 6) and six Λ-type structures which regulate the flow through a twin sluice gate (Figure 7). Except for the sluice gates, in each Λ-type structure two lateral broad crested spillways exist. If the water level passes a threshold, water overflows through these spillways (Figure 8). The characteristics of the channel are presented in Table 1. The geometrical characteristics of each Λ-type structure are presented in Table 2.

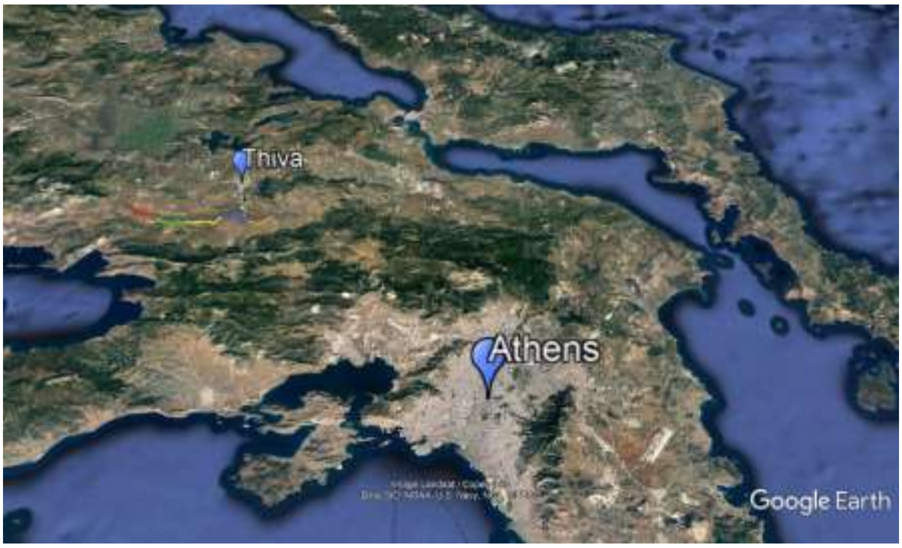

Figure 4: General view of the Λ7 - Λ11 aqueduct.

Figure 5: Part of the Mornos channel, between Λ7 - Λ11 regulation structures.

Figure 6: Trapezoidal cross-section of the conveyance system

Figure 7: Λ-type regulation structures composed by sluice gates and lateral broad crested weirs

Figure 8: Flow over the lateral spillways

Table 1: Channel characteristics

Section	Length (m)	Bottom slope	Bank slope (Horizontal : Vertical)	Bottom width (m)
Λ7 – Λ8	6300	0.000326	1.5	4
Λ8 – Λ9	5650	0.000326	1.5	4
Λ9 – Λ9A	6600	0.000326	1.5	4
Λ9α – Λ10	5700	0.0003	1.5	4
Λ10 – Λ11	4792	0.0003	1.5	4
SUM	**29,042**			

Table 2: Λ-type geometrical characteristics

Λ-type structure	Sluice gate height (m)	Sluice gate width (m)	Spillway length (m)
Λ7	2.52	2.2	60
Λ8	2.23	2.2	60
Λ9	2.17	2.2	60
Λ9A	2.17	2.2	60
Λ10	2.23	2.2	60
Λ11	2.29	2.2	60

II.1. Available data sources

The monitoring system of the case study consists of water elevation gauges (upstream of Λ7 and Λ9, upstream/downstream of Λ8, Λ9α, Λ10 και Λ11), one flow meter located between Λ10 and Λ11 and the openings of the sluice gates for each Λ-type structure (specifically Λ8, Λ9A, Λ10 και Λ11). This system is part of the EYDAP's SCADA, while its main characteristics are presented in Table 3. It is also noted that the time step for the records is 5 min.

Table 3: Monitoring system characteristics for the case study.

Monitor	Variable	Nº (EYDAP system)	Title (EYDAP system)	Location (m)
0EQL10	Discharge (m^3/s)	17963414	Flowmeter - Λ10	131265
0ANAB7	Water Elevation (m)	17681493	Upstream Λ7	106708
0BNAB9	Water Elevation (m)	17681505	Upstream Λ9	118658
0BNABC	Water Elevation (m)	17681497	Upstream Λ8	113008
0BNLBC	Water Elevation (m)	17681501	Downstream Λ8	113010
0CNABC	Water Elevation (m)	17681513	Upstream Λ9α	125258
0CNLBC	Water Elevation (m)	17681517	Downstream Λ9α	125260
0ENABC	Water Elevation (m)	17681521	Upstream Λ10	130958
0ENLBC	Water Elevation (m)	17681525	Downstream Λ10	130960
0FNABC	Water Elevation (m)	17681529	Upstream Λ11	135750
0FNLBC	Water Elevation (m)	17681533	Downstream Λ11	135752
0CPVRD	Opening (‰)	18125162	Λ9α - Right sluice gate	125258
0CPVRG	Opening (‰)	18125162	Λ9α - Left sluice gate	125258
0BPVRD	Opening (‰)	18125142	Λ8 - Right sluice gate	113008
0BPVRG	Opening (‰)	18125142	Λ8 - Left sluice gate	113008
0EPVRD	Opening (‰)	18125170	Λ10 - Right sluice gate	130958
0EPVRG	Opening (‰)	18125170	Λ10 - Left sluice gate	130958
0FPVRD	Opening (‰)	18125174	Λ11 - Right sluice gate	135750
0FPVRG	Opening (‰)	18125174	Λ11 - Left sluice gate	135750

In the framework of F4W, the latter monitoring system is updated with new water elevation gauges. These gauges are presented in Table 4, while a more description is given in Deliverable D4.1. The time step for the records is 5 min as well.

Table 4: Additional gauges installed in the case study during the F4W project.

Monitor	Variable	Nº (EYDAP system)	Title (EYDAP system)
38421800WATL	Water Elevation (m)	199592	Λ9A E Λ10
38422168WATL	Water Elevation (m)	200452	Upstream Λ9A (NTUA)
38979144WATL	Water Elevation (m)	199570	Downstream Λ9A (NTUA)
38421808WATL	Water Elevation (m)	201179	Upstream Λ10 (NTUA)
38421817WATL	Water Elevation (m)	200451	Downstream Λ10 (NTUA)

II.2. Theoretical background

The model was based on the key hydraulic principles to represent the flow at the case study:

- Steady and non-uniform (either gradually varied or rapidly varied) free surface flow
- Flow through sluice gate

- Flow over a borad-crested spillway

If there is no action at the sluice gates, flow is considered steady and normal depth is achieved in between Λ-type structures. At the upstream part of a Λ-type structure, non-uniform gradually varied flow M1 profile is observed ($y>y_n>y_c$), where y_n is the normal depth and y_c is the critical depth. At the downstream part of a Λ-type structure, non-uniform rapidly varied flow is observed and specifically a submerged hydraulic jump (Figure 9).

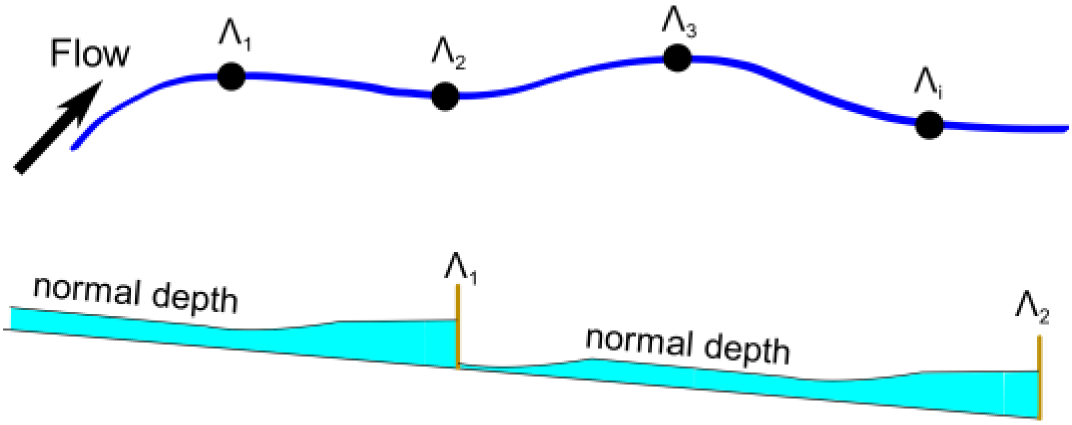

Figure 9: Schematic representation of water depths at the case study

Since there is a move at the sluice gates (either opening or closing), the flow becomes unsteady as long as the discharge is a function of time. An action consequence is a short wave surge which is propagated both upstream and downstream, with velocities *c-V* and *c+V* correspondingly, where *c* is the celerity and *V* is the flow velocity. Celerity can be calculated with the following relationship:

$$c = \sqrt{gD} \qquad \text{(II.1)}$$

where *D* is the equivalent hydraulic depth, which is equal to the ratio of the wetted cross-section area *A* to the free surface width *T*. Flow velocity can be calculated using the well-known Manning equation, for steady and uniform flows:

$$V = \frac{1}{n} R^{2/3} S_f^{1/2} \qquad \text{(II.2)}$$

where *R* is the hydraulic radius (equal to the ratio of the wetted cross-section area *A* to the wetted perimeter *P*) and S_f is the energy line slope, which can be substituted with the bottom slope S_0.

The Λ-type structure has two flow components, namely through the sluice gate Q_{sluice} and over the lateral spillway $Q_{spillway}$, which can be described by theoretical and semi-empirical hydraulic formulas, considering as unknown parameters the discharge coefficients of all sluice gates. The opening of the sluice gate is common for both left and right components. Figure 10 depicts a top view (left) and a longitudinal section of a Λ-type structure (right).

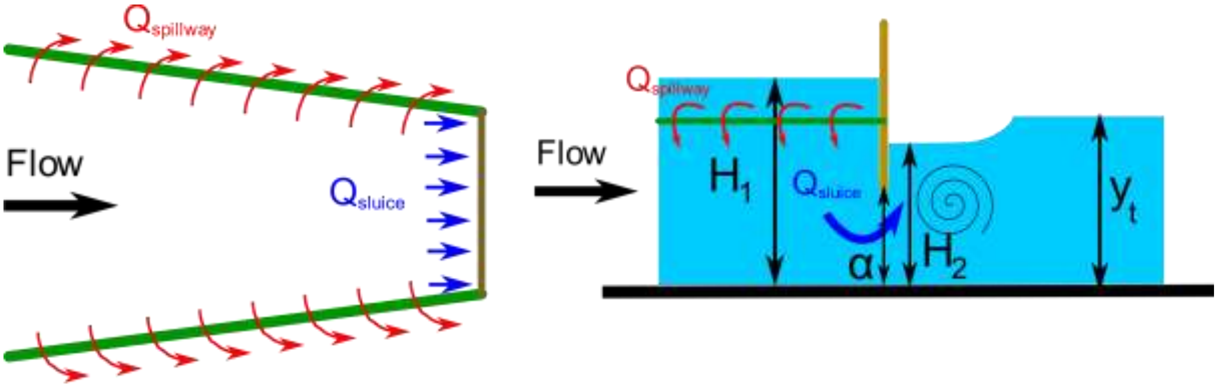

Figure 10: Top view and longitudinal section of a Λ-type structure

The theoretical background of the model is based on the formula proposed by Wu and Rajaratnam [1] for the flow through rectangular sluice gates and on the equation for flow over an ogee spillway (Hydraulics of Open Channel Flow, 2004). Specifically, when the hydraulic jump occurring downstream of the sluice gate is submerged (which is the case), the discharge can be calculated as follows:

$$Q_{sluice} = C_d ab\sqrt{2g(H_1 - H_2)} \qquad \text{(II.3)}$$

where Q_{sluice} is the flow through a sluice gate, C_d is the sluice gate discharge coefficient, a is the sluice gate opening, *b* is the sluice gate width, H_1 is the water depth just upstream of the sluice gate and H_2 is the water depth just downstream of the sluice gate. Since the variable H_2 cannot be easily measured, the latter equation is expressed using the variable y_t, which is the water depth when flow is settled and a measurement can be performed. Now the variable H_2 can be calculated as follows:

$$H_2 = C_d a\left(2C + \sqrt{4C^2 + A^2 - 4BC}\right) \qquad \text{(II.4)}$$

where:

$$A = \frac{y_t}{C_d a}; B = \frac{H_1}{C_d a}; C = 1 - \frac{C_d a}{y_t}\,{}_2 \qquad \text{(II.5)}$$

Although in the Mornos channel hydraulic jumps are always submerged, the equation which describes the flow in the case of a free hydraulic jump is also provided:

$$Q_{sluice} = C_d ab\sqrt{2g(H_1 - 0.61a)} \qquad \text{(II.6)}$$

It is noted that similar equations can also be found in [3].

Regarding the flow over an ogee spillway, the well-known spillway equation is used for estimating the discharge:

$$Q_{spillway} = CLH^{3/2} \qquad \text{(II.7)}$$

where *C* is the discharge coefficient, *L* is the length of the spillway crest and *H* is the hydraulic head over the spillway's crest.

II.3. Pre-processing and data cleaning of data

To support the development of the model, measurements from the sensors described in Section II.1 were collected and analysed. The available time series are the following:

- Flow recorded in the downstream part of Λ10
- Upstream water depth for each Λ-type structure (Λ7, Λ8, Λ9, Λ9A, Λ10 and Λ11)
- Downstream water depth for the Λ-type structures Λ8, Λ9A, Λ10 and Λ11
- Left sluice gate opening for each Λ-type structure
- Right sluice gate opening for each Λ-type structure
- Water depth in the middle between Λ9 and Λ9Aα

The analysis showed that the available time series are characterized by significant errors (namely flag values which indicate a problem related to either recording or transmission) and instabilities. The nature of the system (there are long time intervals in which no action is performed in the sluice gates) led the research team to perform a pre-processing handling of the time series, in order to filter the data and extract "characteristic" data instead of long "noisy" time series. The latter process can be summarized in the following steps:

- Filtering, smoothing and correction of the errors
- Searching for time intervals in which there is no sluice gate action for a specific Λ-type structure
- Mining “critical points” in the available time series. As critical point for a specific time interval, we define an equivalent value for the variable in this specific time interval, that is the median.

For the first step, sluice gate openings are classified with a step of 5‰, according to real world operation and a threshold is defined (usually 10‰). If the action is below than this threshold it is assumed that there is no action performed and this fluctuation is a monitoring system error. An example of this smoothening is depicted in Figure 11.

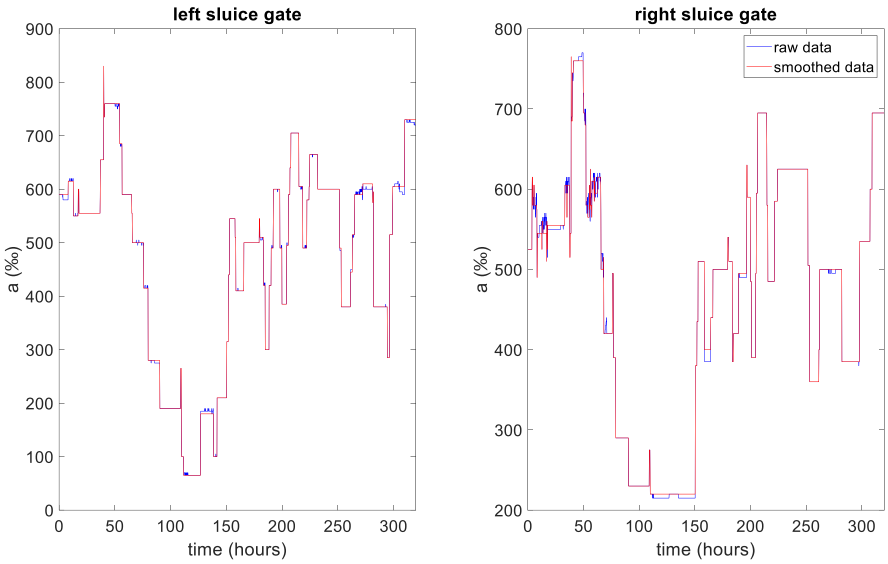

Figure 11: First step of time series handling

For the next step, the developed algorithm is searching for a time window in which both sluice gates for a Λ-type structure are not moved, for at least a time threshold, which is defined as 100 min. An example of this step is depicted in Figure 12. The dot indicates the median value (critical points).

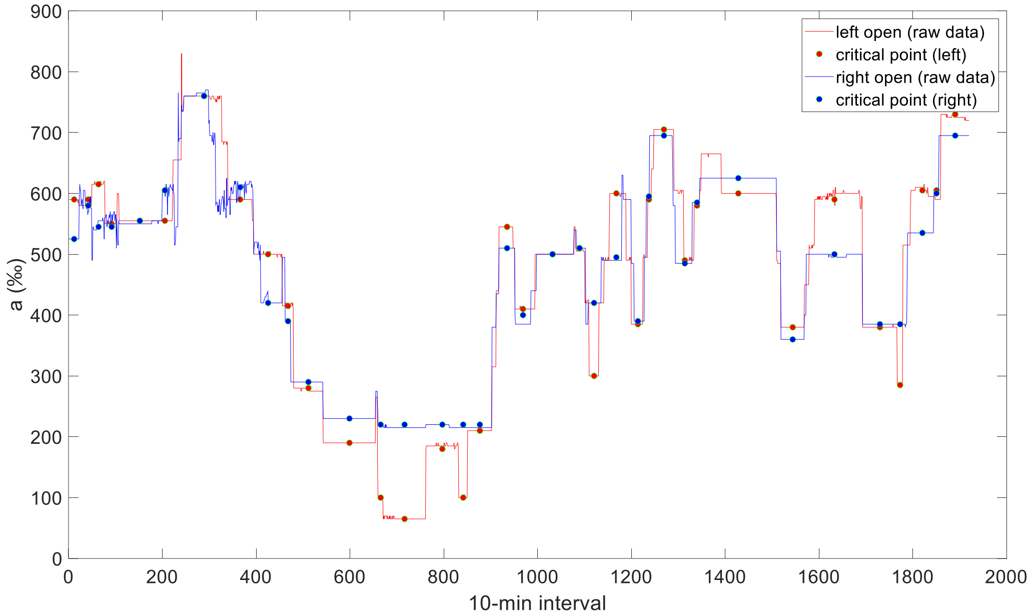

Figure 12: Second step of time series handling

Finally, in the third step, the critical points are found for the rest variables, namely the upstream water depth, the downstream water depth and the discharge (Figure 13). The data mining process is finalized by extracting from these long time series smaller equivalent data sets which are more efficient for the global calibration phase.

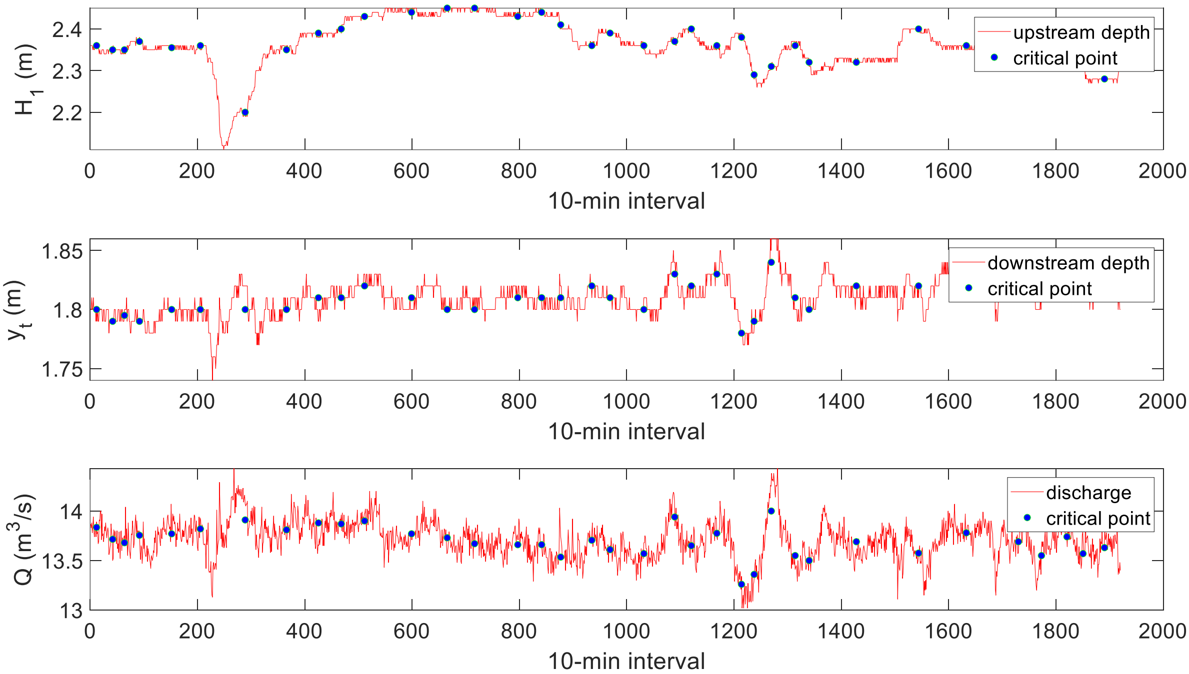

Figure 13: Third step of time series handling

Further to the above desk analysis, several field visits of data scientists along with the operators of the system took place to compare SCADA readings with in-situ measurements of water level. This comparative analysis showed that in some parts of the system the existing meters overestimate or underestimate the real water depth in the channel. As described in Deliverable D4.1, to improve the accuracy and validity of the model, EYDAP and NTUA decided to install 5 new water level meters.

Specifically, 4 of the new meters were installed at positions (upstream and downstream of two regulation structures) where water level meters already exist but their measurements were found of bad quality. The 5th water level meter was installed in a position of the conveyance system which is not currently monitored (between two sluice gates) to allow further insights to derive on critical hydraulic parameters of the system (i.e., estimation of Manning coefficient). The global calibration of the sluice gate discharge coefficient, described in next section, has been conducted on the basis of accurate data from these new water level meters.

II.4. Global calibration: the sluice gate discharge coefficient

The discharge coefficients for the sluice gate flow and the spillway overflow can be globally calibrated, based on the available data. According to the conveyance system contractor [4], the spillway discharge coefficient has been calibrated with good accuracy and is equal to C=1.85. Therefore, the latter coefficient is not included in the global calibration process.

For the global calibration, the sum of the sluice gate flow and the spillway flow shall be equal to the discharge recorded by the flowmeter, since the flow is assumed to be steady (there is no action at the sluice gates and therefore flow is steady indeed). Since the problem is not explicit, the following objective function is developed, which shall be minimized:

$$OF_1 = (Q_{mod} - Q_{obs}) \tag{II.8}$$

where Q_{obs} is the discharge recorded by the flowmeter and Q_{mod} is the discharge calculated by the proposed numerical model and is equal to the sum of sluice gate and spillway flow:

$$Q_{mod} = Q_{sluice} + Q_{spill} \tag{II.9}$$

Herein, it is noted that due to the fact that the sluice gate openings (left and right) are not always equal, the equivalent opening is assumed to be the average of the two openings.

The fundamentals for the flow estimation are given in the previous section and specifically in Eq. (II.3)-(II.7). Since the discharge coefficient of the spillway flow is not included for the global calibration phase (it is chosen C=1.85 as previously mentioned) just the sluice gate coefficient was calibrated. The minimization of the objective function was achieved using the optimization technique proposed by Lagarias et al [5].

Sluice gate discharge coefficients were calibrated for the mined critical points dataset. According to theory [1], these values depend on the ratio of the equivalent opening of the sluice gates to the upstream water depth. The results of the calibration phase are depicted in Figure 14. It seems that there is a discrepancy between the theoretical calibrated values (those obtained in laboratory experiments), which are not correlated with the ratio α/H and they are systematically lower. The output of the global calibration phase is a value of the discharge coefficient equal to C_d=0.40.

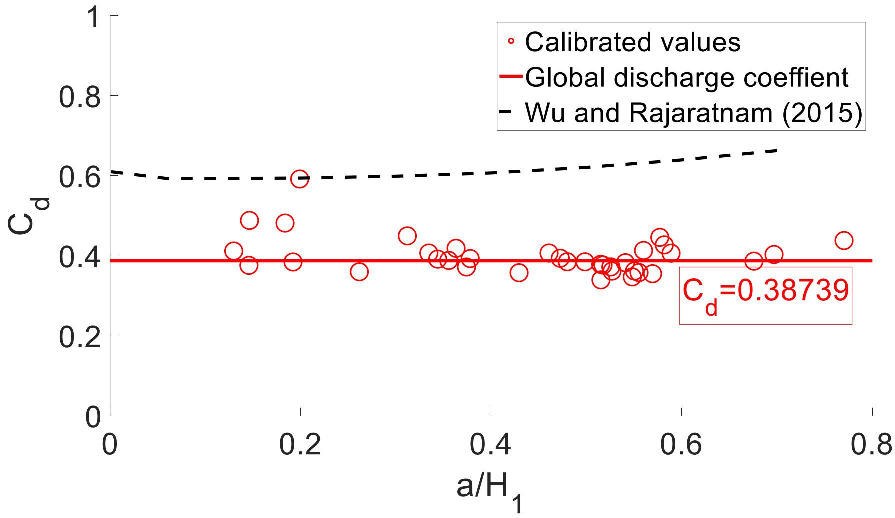

Figure 14: Calibrated sluice gate discharge coefficient (global calibration phase) and comparison with theoretical values

II.5. Global calibration: the Manning coefficient

Furthermore, to indicate the time response between the time instant when a sluice gate is moved (either opening or closing) and the time instant when the latter move is captured by the flow meter (installed downstream of Λ10), the flow velocity is used. The latter variable is estimated by the Manning equation, which is valid for steady and uniform flow and depends on the geometrical characteristics of the channel and the roughness coefficient *n*.

Although there are several values reported in classic hydraulic handbooks, such as Chanson [2] and Chaudhry [6], for several categories of a concrete channel, in this work Manning coefficient is globally calibrated in order to preserve accuracy. The data used for the calibration phase are the discharges measured by the flowmeter and the water depths recorded by the water elevation gauge installed in the framework of the F4W project in the middle of Λ9α and Λ10. Since there is no action in the sluice gates, the water depth at this point can be considered equal to the normal depth.

For the data mining, a similar procedure is followed, as previously described for the sluice gate coefficient calibration, namely the three steps: a) filtering, smoothing and correction of the errors; b) searching for time intervals in which there is no sluice gate action; c) mining critical points. The output of this process can be depicted in Figure 15. First, it seems that for a range of discharges from 13 to 15 m^3/s, Manning coefficient is steady and equal to 0.02. Second, calibrated values are systematically bigger than the corresponding values proposed by the bibliography.

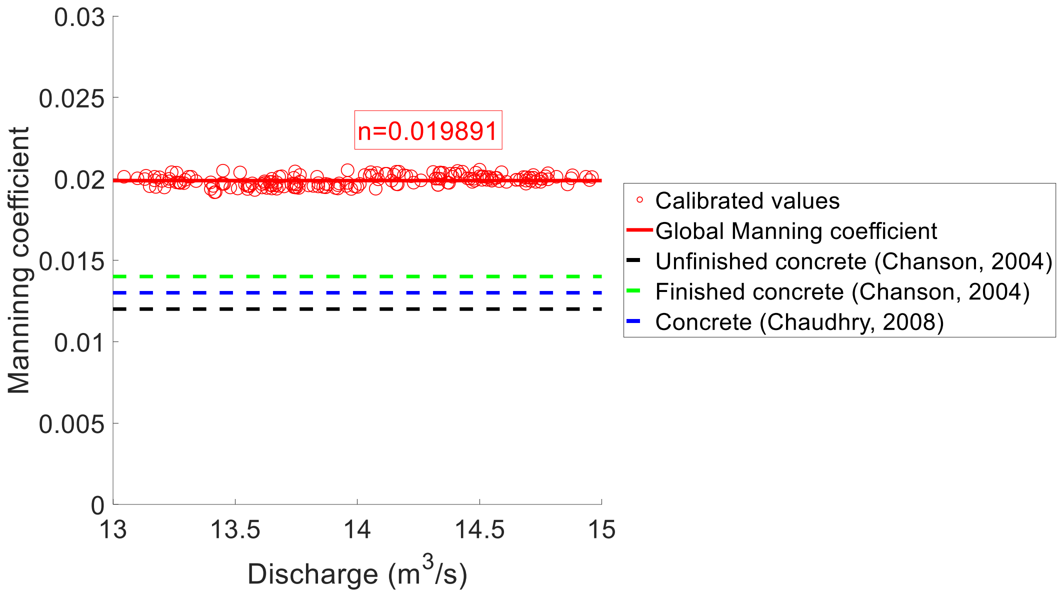

Figure 15: Calibrated Manning coefficient against theoretical values

II.6. Operational model for optimal sluice gate opening and scheduling

i. Fundamentals

The aim of this study is the development of an operational tool which can be a potential assistant to the operational staff of EYDAP, in order to estimate the new opening of the sluice gates when flow regulation is required. As previously described, the current practice depends on the subjective view and the experience of the technician.

The developed tool is based on a numerical model which first estimates the current situation (regarding the flow characteristics) with a small residual between reality (measurements) and model, and second gives an estimation for the new openings with respect to a new target discharge. The basic principles of the model are the following:

- The function for every Λ-type structure does not affect the flow characteristics observed either in the upstream Λ-type structure or in the downstream Λ-type structure.
- Flow is considered as steady (regarding the time) and uniform (regarding the space).
- Real-time flow is the discharge captured by the unique flowmeter of the case study, which is installed downstream of Λ10. Hence, there shall be no sluice gates moves for a considerable time before every action, in order to preserve steady and uniform conditions for the flow.
- The discharge which passes through each Λ-type structure has two components: flow through the sluice gates and flow over the lateral broad-crested weirs (spillways).
- The hydraulic head over the spillways is equal to the water elevation captured by a gauge in real-time, minus the crest of the spillway which is a geometrical characteristic of the Λ-type structure.
- For the sluice gate flow, Wu and Rajaratnam equation [1] is used. The discharge coefficient is calibrated every for every action (real-time calibration). In order to preserve the consistency,

there is a control process: if the real-time calibrated value is unrealistic, it is substituted by the globally calibrated value (off-line calibration).

- For the spillway flow, discharge coefficient was chosen equal to 1.85 as reported by the contractor of the channel.
- Each Λ-type structure is comprised from two sluice gates. In the case in which the openings (left and right) are not equal, the equivalent opening is considered to be the average of the two openings.

The above principles can be summarized in the following Figure 16, in which the flow chart of the developed numerical model is depicted.

Finally, in the design and control of hydraulic works, modelling approaches are bounded by two extreme cases: the pure "white-box" and the pure "black-box" approach. In the first case, the system is described via physically-based equations, whose parameters are obtained on the basis of hydraulic handbooks. In the second case, data-driven models (such as Machine Learning algorithms) are calibrated against field data, without any prerequisite to obey physical laws and provide physically meaningful parameters. In between these two approaches lies the "grey-box" approach [7] that combines the advantages of these two extreme cases to develop more robust and scientifically sound models.

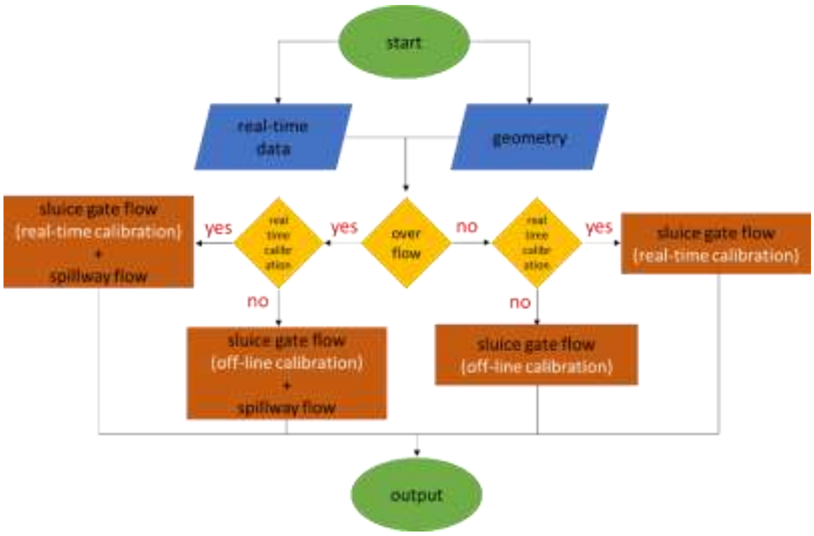

Figure 16: Flow chart of the proposed numerical model

ii. Input data

Input data consists of two categories:

- Real-time data (hydraulic characteristics of the system)
- Geometrical data for every Λ-type structure (do not depend on time)

The real-time data are the following:

- Channel discharge, which is captured by a flowmeter installed downstream of Λ10, given by the monitoring system.
- Upstream water depth for each Λ-type structure, given by the monitoring system.
- Downstream water depth for each Λ-type structure, given by the monitoring system.
- Left opening (percentage ‰ of the sluice gate height), given by the monitoring system.

- Right opening (percentage ‰ of the sluice gate height), given by the monitoring system.
- The new target discharge (desirable), given by the user.

The geometrical data for every Λ-type structure are the following:

- Sluice gate width
- Sluice gate height
- Spillway height
- Spillways length

iii. Estimate sluice gate openings

First, sluice gate discharge coefficients are real-time calibrated in every Λ-type structure. The objective function which shall be minimized is the absolute value of the residual between the observed (Q_{obs}) and the modelled flow (Q_{mod}):

$$OF_1 = |Q_{obs} - Q_{mod}| \tag{II.10}$$

where Q_{mod} is the simulated flow as derived by the numerical model and Q_{obs} is the observed flow as captured by the flowmeter. The simulated flow is the sum between the two components of the flow, namely sluice gate flow (Q_{sluice}) and spillway overflow (Q_{spill}):

$$Q_{mod} = Q_{sluice} + Q_{spill} \tag{II.11}$$

The theoretical background for each component has been extensively described in previous chapter. Since the spillway coefficient is constant and equal to C=1.85, just the sluice gate coefficient C_d shall be calibrated. The latter problem is implicit and therefore an optimization technique shall be used in order to minimize the objective function. In this study, the algorithm proposed by Lagarias et al. [5] is used.

In the case in which the real-time calibrated sluice gate coefficient has a significant divergence from the globally calibrated sluice gate coefficient (up to 50%), the latter coefficient takes the global value (off-line calibration).

An additional output, except the real-time calibrated sluice gate coefficient, is the model error, which is the residual between the observed and the simulated flow. If the latter error is above a threshold (in this work this threshold is defined to be 10%), the user is warned.

Next, the model calculates the new equivalent opening of the sluice gates, based on the new target discharge, which is given by the user. The problem is implicit as well and therefore a new objective function OF_2 is developed, which is the absolute residual between the new desirable discharge (Q_{des}) and the modelled flow (Q_{mod}):

$$OF_2 = |Q_{des} - Q_{mod}| \tag{II.12}$$

The challenge here is that the sampling space of the openings is not continuous but discrete, with a step of 5‰. Taking into account the small computational cost and the small sampling range, a grid-search calibration scheme is selected. In practice, the modeled flow is calculated with all the 200 possible openings (the sampling range is [0, 1000]‰ and therefore a step of 5‰ gives 200 values). The final new opening is the value which minimizes OF_2.

The following figures (Figure 17 - Figure 19) present some indicative examples of the model output. Figure 17 depicts a typical example when the target discharge is bigger than the current and hence sluice gates shall be open. The discontinuities observed for big opening values (more than 700‰) are due to the fact that for these openings the hydraulic jump becomes free, which of course is not an option for the current practice. However, the model is able to capture this transmission (from submerged to free jump) in order to preserve consistency. Figure 18 depicts a typical example when the target flow is smaller than the current and therefore sluice gates shall be closed. Figure 19 depicts a typical example when left and right opening positions are quite different and therefore the equivalent new opening is in between the left and right positions. Finally, some typical examples of input data and output derived by the tool are presented in Table 5.

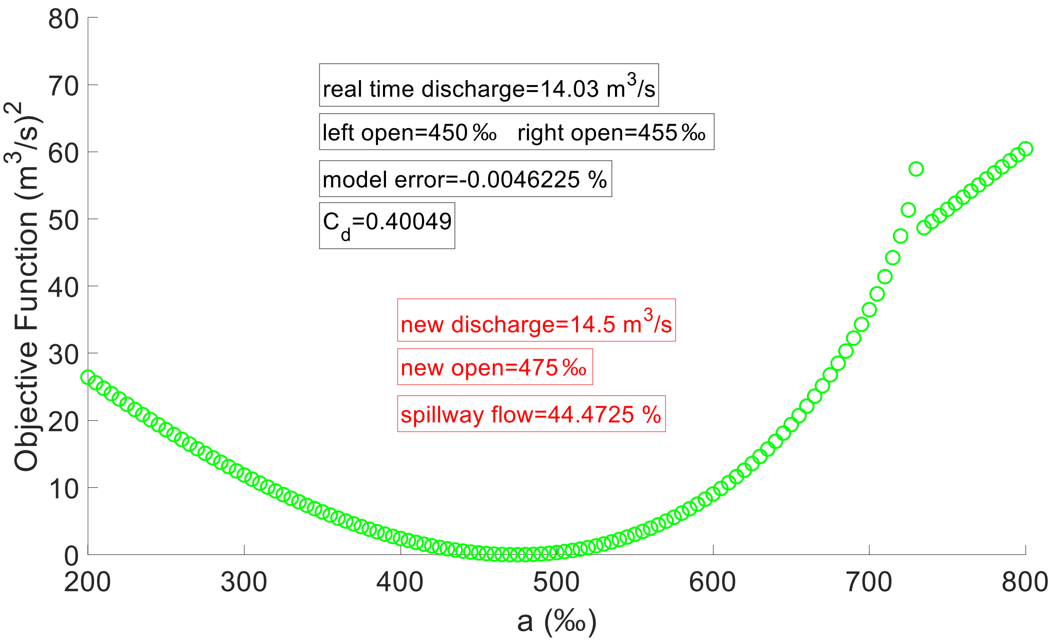

Figure 17: Output of the numerical model: target discharge is bigger than the current.

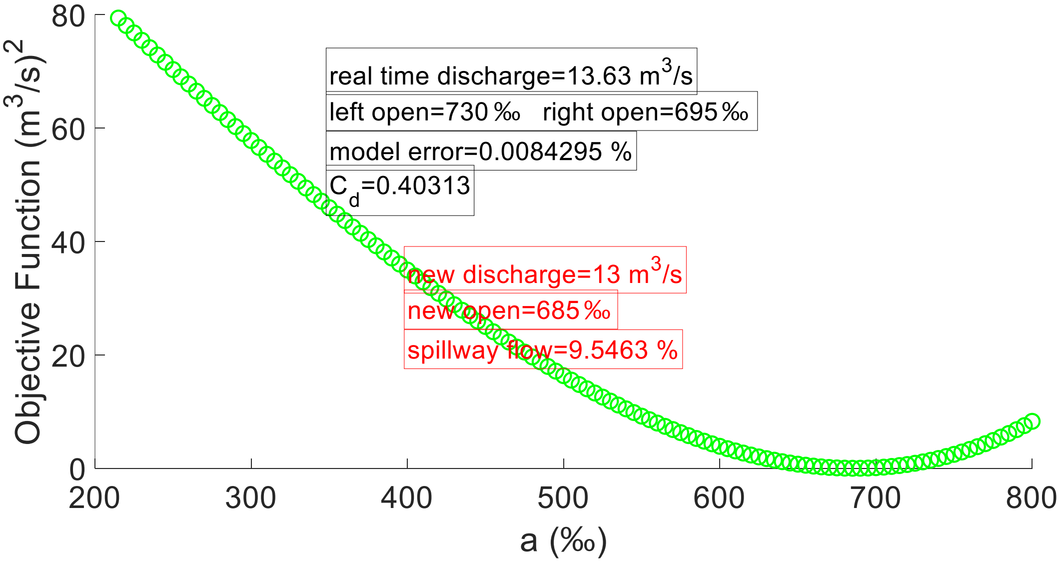

Figure 18: Output of the numerical model: target discharge is smaller than the current.

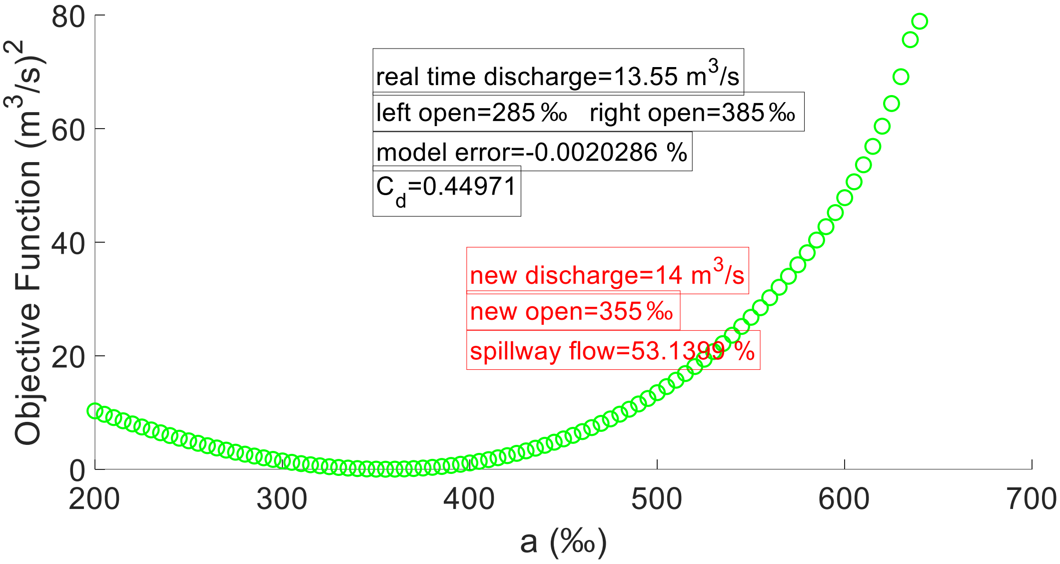

Figure 19: Output of the numerical model: significant discrepancy between left and right opening.

Table 5: Typical examples of input data and output derived by the tool

Scenario	Current discharge (m^3/s)	Left opening (‰)	Right opening (‰)	Upstream water depth (m)	Downstream water depth (m)	Target discharge (+)	Target discharge (-)	Proposed equivalent opening - (‰)	Proposed equivalent opening + (‰)
1	13.63	730	695	2.28	1.82	14.00	13.00	725	685
2	13.55	285	385	2.40	1.81	14.00	13.00	355	310
3	13.72	590	580	2.35	1.79	16.00	11.00	685	445
4	14.00	705	695	2.31	1.84	15.00	13.00	745	655
5	13.26	385	390	2.38	1.78	14.00	12.00	420	330

iv. Estimate response time

The developed tool estimates the response time between the time moment of an action (either opening or closing the sluice gates) and the time moment when the unique flowmeter (installed downstream of Λ10) captures this modification. If there is a divergence between the new record observed at the flowmeter after the response time, and the new target discharge, the user shall repeat the previously described process to fine-tune the openings.

The time response is calculated assuming uniform and steady flow condition, through the well-known Manning equation. The Manning coefficient is selected as the globally calibrated value, equal to 0.02. With given discharge and geometrical characteristics of the channel, the normal depth y_n can be calculated for all the sections between the Λ-type structures. Then, flow velocity *V* for each section can be easily calculated with the formula $V=Q/A$, where *Q* is the real-time discharge and *A* is the wetted cross-section area, which is trapezoidal and function of the normal depth. Needless to say that the latter method is characterized by several uncertainties. Hence, the estimated response time is rounded in the order of magnitude of 5 min.

The model presented above, for the estimation of optimal sluice gate opening and scheduling, have been implemented in an operational, FIWARE-enabled, context. The model is fed with real-time measurements from the sensors in order to provide the openings and time responses, which are presented to the operational staff of EYDAP in the dashboard of the web platform. This operational implementation of the model is presented in Deliverable D4.1.

III. Forecast level of raw-water turbidity and warning for high-turbidity events

This section focuses on the provision of early warnings for high turbidity events and forecast of the level of turbidity of raw-water at the most downstream parts of “Giona - Dafnoula”, before reaching the 4 water treatment plants that serve the city of Athens.

III.1. Data sources and features extraction

The quality parameters of raw water in the demo case part of the conveyance system (“Giona - Dafnoula”) are monitored via 6 metering stations, covering the entire aqueduct. The metering stations measure and log the instantaneous temperature, conductivity and turbidity of raw water, at a frequency of 5 min. The position of the 6 raw-water quality sensors is shown in Figure 20. As shown, the most upstream sensor is at Giona, while the two most downstream metering stations are in L10 and Dafnoula, respectively.

Figure 20: Water quality meters across the demo part of the external raw water supply system of EYDAP (from Giona to Dafnoula).

Time series of turbidity, of 5 min time step, were collected from the 6 metering stations. The available series spans from 01/07/2017 up to 20/05/2021 (a sample of approximately 4 years length). The pre-processing analysis of turbidity measurements revealed that the series exhibit sudden spikes (i.e., individual peaks) that last for a few time intervals, where measurements are extremely greater than the values preceded and followed. Indicatively, this behaviour is displayed in the plot of Figure 21. Evidently, such peaks do not consist high turbidity events with a permanent behaviour, but erroneous measurements which are attributed to sensor malfunction or bad sensor readings due to external factors (for instance, if a leaf stands upon the sensor or the sensor is very close to the surface of water in the channel).

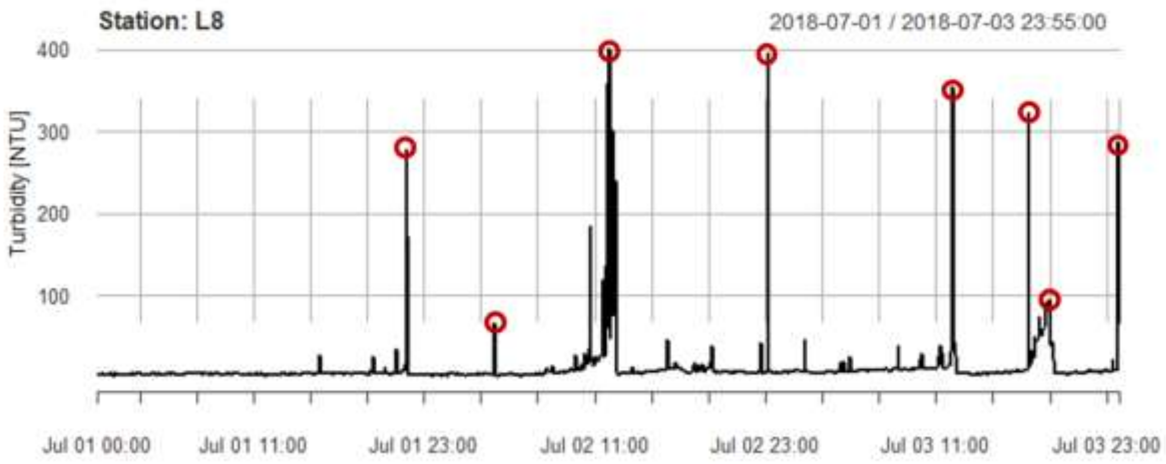

Figure 21: Indicative example of the spiky behaviour of turbidity in the external conveyance system of EYDAP (red dots indicates sudden peaks).

To determine the position and intensity of these spikes in the 6 available time series, we employed the "*.find_peaks()*" function, which belongs to the package "*.signal*" of the "Scipy" library [8] of Python. The function finds local maxima by comparison of neighboring values in the series. To ensure that only sudden peaks, and not persistent turbidity events of higher rates, will be isolated via this procedure, we set the window length that limits the evaluation area for each peak equal to 2. Such turbidity peaks, as appear in the series of L8 metering station, is presented in Figure 22.

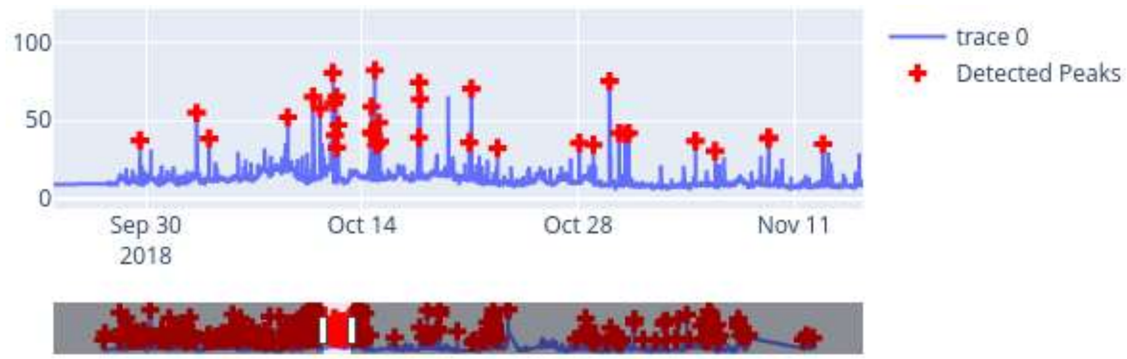

Figure 22: Sudden peaks (local maxima) detection of turbidity time series at L8 metering station.

For all 6 series, the detected peaks were removed, and the missing values were filled with the neighboring values. Figure 23 presents the same time series with Figure 21 after detecting and smoothening the sudden individual peaks. As we see, the pre-processing step affects only sudden spikes, without distorting the events of higher turbidity exhibiting a persistent behaviour.

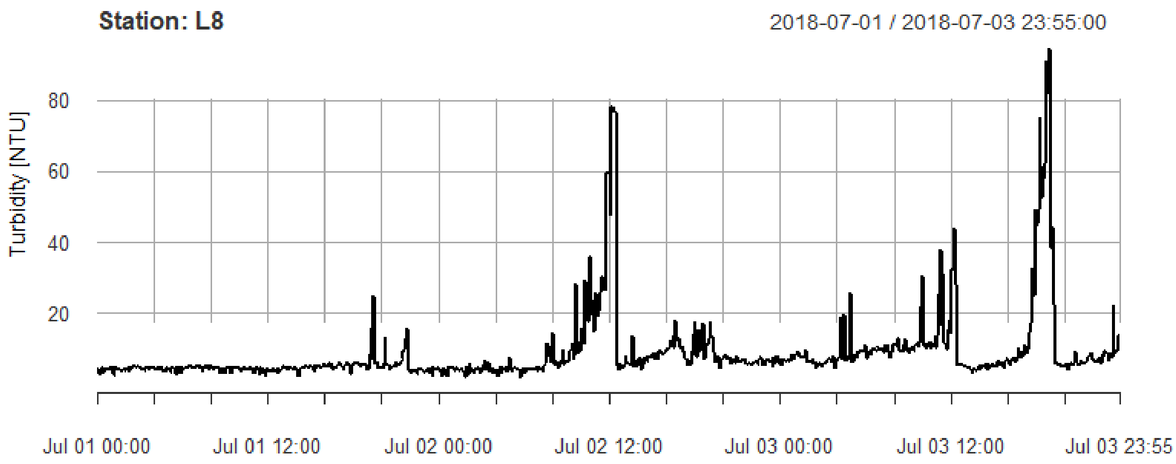

Figure 23: Turbidity time series after detecting sudden peaks and replacing them with neighboring values.

The processed series were further analysed to obtain insights on their statistical peculiarities. The histogram of turbidity series at the 6 metering stations are presented in Figure 24, while a direct comparison of the statistical behaviour is given in the box plots of Figure 25. In the histograms the blue dotted line indicates the median of the records, while the two red dotted lines indicate the 5% and 95% quantile points. As it is evident, the vast majority of turbidity values in all stations vary in the range [3, 15] NTU, while the range of variation is greater in the case of L8 and Dafnoula metering stations. As depicted via blue dots in the box plots of Figure 25, the average turbidity is close to 5 NTU at Giona and Saradi stations, while for the downstream stations (L8, L10, Dafnoula) average turbidity found a little higher than that of upstream and approximately equal to 7.5 NTU.

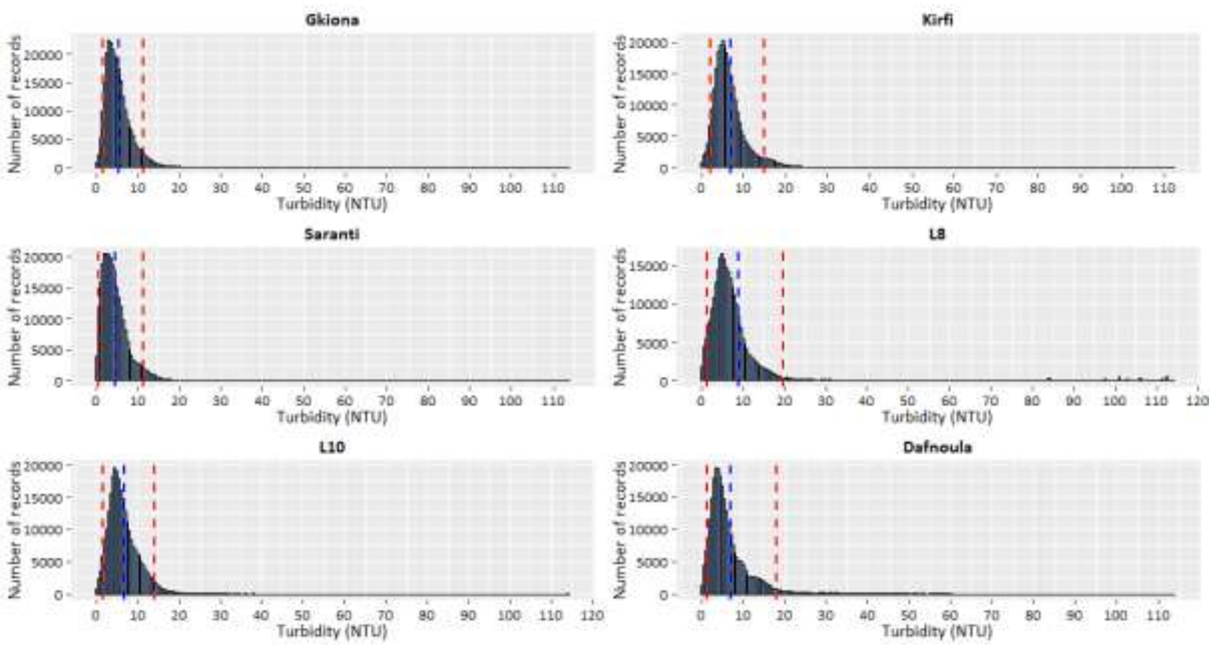

Figure 24: Histogram of turbidity data at the 6 metering stations. Blue dotted line indicates the median, while red dotted lines indicate the 5% and 95% quantiles.

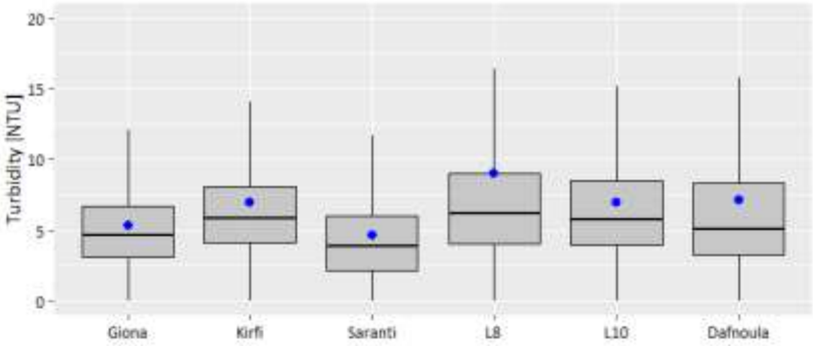

Figure 25: Box plots comparing turbidity data at the 6 metering stations.

The series were further analysed to identify high turbidity events with persistent behaviour, i.e., clusters of turbidity measurements with high values, which are detected at an upstream point and appear at a downstream point after some time intervals. It is worth to mention that, according to the operational rules of EYDAP, the threshold to define a turbidity event as high (severe) is the 100 NTU. In the case of such an event, the operatorial staff cut-off the inlet of raw-water in the water treatment plants to avoid damage. The available dataset does not contain severe events with a persistent behaviour that propagates to the downstream part of the aqueduct. Events of turbidity greater than 100 NTU appear to have local character (detected at a single metering station) and lasts for a few time intervals.

The dataset contains a small number of high turbidity events, thought not severe, which travel across the aqueduct. Such an event is presented in Figure 26. Evidently, there is a spike in turbidity starting from Giona station (the most upstream one in the demo case part of the system) and propagates to the next station on a downstream order. As we see in Figure 26, turbidity at Dafnoula (most downstream metering station) deviates from the behaviour of other metering stations, and especially the behaviour of the just upstream metering station at L10. This is attributed to the fact that the metering station at Dafnoula is close to the point where "Giona - Dafnoula" aqueduct is spitted to two sub-branches, the main branch (that goes to Athens) and the secondary branch (that meets Yliki aqueduct), respectively (see Figure 1).

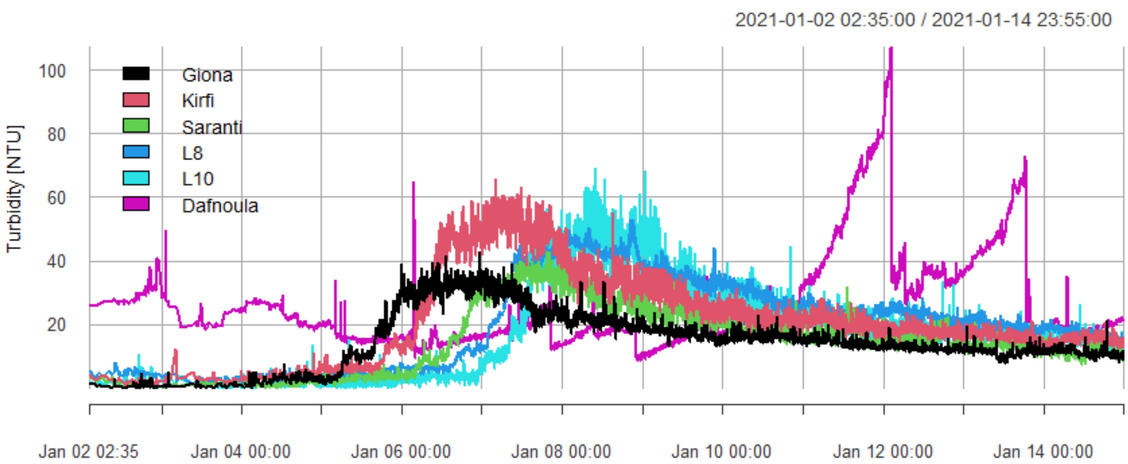

Figure 26: High turbidity event travelling across the "Giona-Dafnoula" aqueduct.

Furthermore, we took advantage of the persistent high-turbidity events to obtain evidence on the mean time required a turbidity event to propagate between metering stations, studying turbidity events from Giona up to L10 point. Specifically, we estimated the cross-correlation coefficients between the series of each metering station, searching for the time lags with the higher cross-correlation values. The results of this analysis, i.e., the estimated time required for an event to propagate between metering stations, are shown in Table 6.

Table 6: Estimated time required for a turbidity event to propagate among metering stations

	Giona	Kirfi	Saradi	L8	L10	Dafnoula
Giona	0	9.08	17.67	21	26.75	31.43
Kirfi		0	4.5	8.67	15.25	20.12
Saradi			0	4.83	11.42	15.90
L8				0	3.42	7.80
L10					0	4.10

It is highlighted that the above analysis has as a target to estimate how many samples of the past measurements we need to include as predictors to forecast turbidity at a downstream metering station.

III.2. Model to forecast the level of turbidity

A threshold-based early warning approach for extremely high turbidity events has been developed and implemented in an operational context. Following the operational rules of EYDAP, two levels of severity were defined. A medium severity alert (“orange alert”) when turbidity measurements exceed 80 NTU for 3 successive time intervals at a metering station, and a high severity alert (“red alert”) when turbidity measurements exceed 100 NTU for 3 successive time intervals at a metering station. More details on the implementation of the early-warning system in an operational context are given in Deliverable D4.1. In brief, when a medium or high severity turbidity event is detected in a metering station, alters (via e-mails) are sent to the operators of the system, while further information on the event is provided via the web platform. The platform allows the operational staff to define alternative levels of severity, by setting custom thresholds, depending on the needs.

Going on the step further, a data-driven model was developed and implemented to forecast the level of turbidity at the downstream part of the conveyance system, given the level of turbidity upstream. Specifically, the forecast problem reads as: forecast turbidity one hour ahead of the latest measurement available at the two most downstream metering stations (i.e., at L10 and Dafnoula stations), given the measurements of turbidity at the upstream stations (i.e., at Giona, Kirfi, Saradi and L8).

To address the above problem, we adopt a data-driven approach, and specifically, we employ Deep Learning (DL) Neural Network (NN) architectures, which have become particularly popular, holding today a prominent position in complex and non-linear prediction problems [9], [10]. At their core NNs build on an algorithmic analogy of the “human brain”, which in turn justifies the rationale behind the use of notions such as those of, neurons, connections and layers. To capture the temporal dependencies exhibited in the process under study, we build a long short-term memory (LSTM)-based model. This type of model belongs to the general category of recurrent neural networks, which, through feedback connections, allow the capturing of long- and sort-term dependences in the input sequences. This ability makes LSTM-based models more suitable compared to the conventional NN (e.g., feedforward neural networks) for time series predictions. LSTM models have been used in a wide range of applications and domains, where the sequence of appearance of information is of interest.

Indicative examples of LSTMs' implementations are for speech recognition, image recognition and language processing (e.g., see [11]). In the general water domain, LSTMs find wide applicability since the capturing of temporal and spatial dependences of physical (e.g., rainfall) and non-physical (e.g., water demand) processes is of paramount importance.

To forecast turbidity, we build a stacked LSTM network using Tensorflow[1] and Keras[2] [12] libraries in Python[3]. Due to the high local variability that turbidity exhibits at the most downstream metering station of the system under study (i.e., "Dafnoula"), we first focus on the development of a model that predicts the turbidity, 1 hour ahead, at the second most downstream station ("L10"), given turbidity at the 4 upstream stations (i.e., "Giona", "Kirfi", "Saradi" and "L8"). Initially, the problem was treated as a regression one, trying to forecast the exact value of turbidity. However, as discussed in the previous section, on the one hand, turbidity exhibits a highly noisy behaviour at the scale of the metering station (e.g., local scale), and on the other hand, this behaviour does not propagate to the downstream points, where local turbidity conditions dominate, in the vast majority of cases. This behaviour, along with the limited number of available turbidity events with persistent behaviour (such as that presented in Figure 26), hampered the development of a regression-type model of high accuracy. To provide a remedy to this issue, we addressed the problem as a classification one, trying to predict the range of values (class) in which turbidity will be at the downstream points, given the class of turbidity at the upstream points. This approach is consistent with the rules of operation of the conveyance system and of the water treatment plants, which are differentiated with respect to the level of turbidity in raw-water, and not the exact value of turbidity per se.

The data set was divided into a training set (70% of the available time series) and a validation set (30% of the available time series) to train and validate the model, respectively. In this context, the first set is used to train the model (obtain model parameters), while the validation set is used to early-stop the training algorithm, preventing overfitting or underfitting. To determine the optimal LSTM-based model, we analyse alternative model architectures, with different number of layers and number of nodes/units per layer. The batch size[4] was set equal to 256 and the number of epochs[5] equal to 1000, while the sigmoid function was used as an activation function[6]. As early-stopping measure[7], the validation accuracy was used, while as loss function[8], the categorical crossentropy loss[9] (and its binary version).

The alternative model architectures were first evaluated, by classifying the values of turbidity into 4 classes, i.e., [0, 4) NTU, [4, 8) NTU, [8, 12) NTU and [12,) NTU, labeled as "0", "1", "2" and "3", respectively. The final model is composed by two hidden LSTM layers, with 512 and 128 nodes,

---

1 TensorFlow is a free and open-source software library for machine learning and artificial intelligence (https://www.tensorflow.org/).

2 Keras is an open-source software library that provides a Python interface for artificial neural networks (https://keras.io/).

3 Python is a high-level, general-purpose programming language (https://www.python.org/).

4 Bach size: the number of training samples in one forward/backward pass (before updating the internal model parameters).

5 Epochs: the number times that the learning algorithm will work through the entire training dataset.

6 Activation function: also known as transfer function, it defines how the weighted sum of the input is transformed into an output from a node or nodes in a layer of the network.

7 Early stopping is a method that allows you to specify an arbitrarily large number of training epochs and stop training once the model performance stops improving on the validation dataset.

8 Loss function: a function quantifying the difference between the expected outcome and the outcome produced by the model.

9 Categorical crossentropy loss: A typical measure that quantifies the difference between the predicted and given categorical data (https://keras.io/api/losses/probabilistic_losses/#categoricalcrossentropy-class).

respectively, along with a hidden regular densely-connected layer with 64 nodes. To forecast the class of turbidity at L10, 1-hour ahead, we use as predictors (inputs) 320 5-min past observations from the 4 upstream stations. The number of past observations is consistent with the analysis presented in Section III.1, indicating that the time required for an event to travel from the most upstream station ("Giona") to "L10" station is approximately 27 hours. The performance of the model is presented in Figure 27. As we see, the achieved accuracy is approximately 60% for the training set. This accuracy is mainly attributed to the large number of turbidity values that belong to class "0" (i.e., [0, 4) NTU) and have been classified to class "1" (i.e., [4, 8) NTU), as well as those which have been incorrectly classified in class 3 (i.e., [12,) NTU) instead of 2 (i.e., [8, 12) NTU).

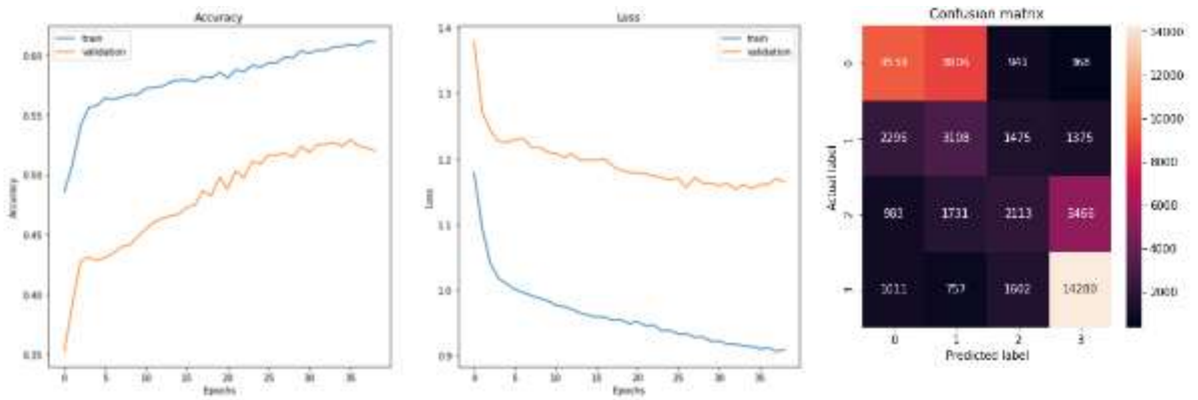

Figure 27: Model accuracy for the prediction of the level of turbidity (4 classes) at L10 station, 1-hour ahead, using 320 observations from the 4 upstream stations.

To improve the prediction accuracy, we tested other setups. We experimented with shortening the time window for the feature vector, thus making the training process simpler and more robust. The results of the model, using as predictors (inputs) 100 past observations, are shown in Figure 28. The overall accuracy found approximately equal to 55% for the training set. As we see, in the confusion matrix of Figure 28, this setup enables a better classification of values in classes "0" and "1", but it results in a larger number of values that have been incorrectly labelled as "3", instead of "2".

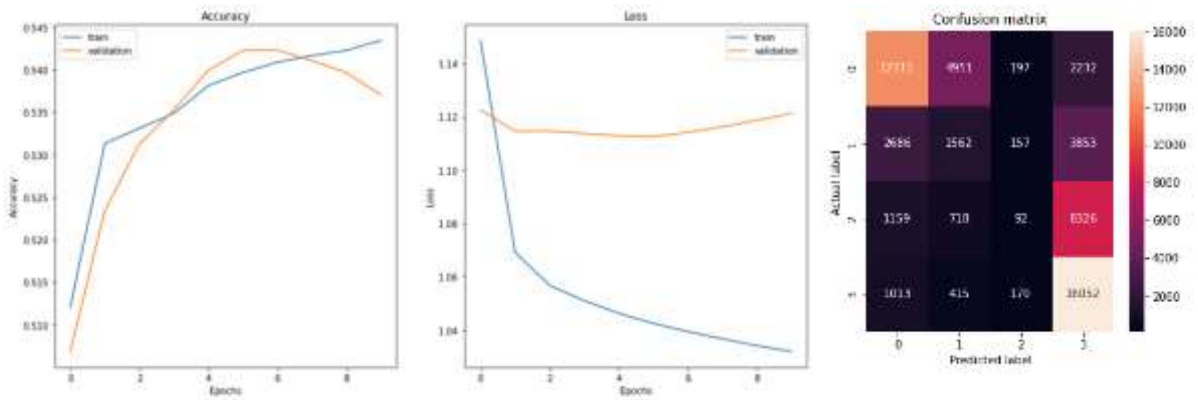

Figure 28: Model accuracy for the prediction of the level of turbidity (4 classes) at L10 station, 1-hour ahead, using 100 observations from the 4 upstream stations.

The results of the previous analysis indicate that the available time series can not support the prediction of the level of turbidity, on the basis of 4 classes. This is reasonable since, as shown in the histograms of Figure 24, the greatest percentage of observed turbidity values are less than 12 NTU (more than 95% of values in most metering stations), with the higher turbidity values being significantly less and sporadic in the sample.

A much higher accuracy was obtained for the classification of the level of turbidity in two classes, i.e., [0, 6) NTU and [6,) NTU, labeled as “0”, “1”, respectively. The deep neural network, composed by two hidden LSTM layers, with 512 and 128 nodes, respectively, and a hidden regular densely-connected layer with 64 nodes, was employed. As shown in Figure 29, we achieved 81% validation accuracy, however several false positives and negatives are obtained.

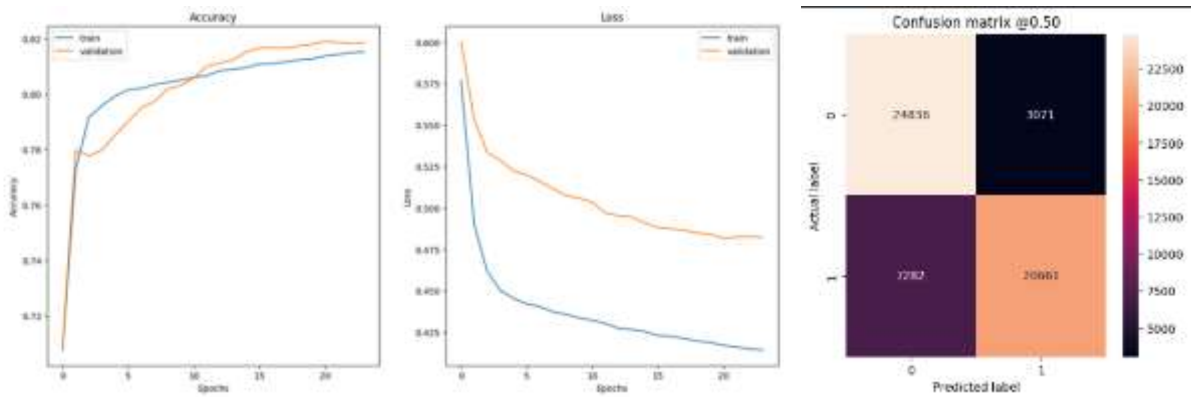

Figure 29: Model accuracy for the prediction of the level of turbidity (2 classes) at L10 station, 1-hour ahead, using 320 observations from the 4 upstream stations.

To improve further the performance of the model, we use a narrower time window of past observation, equal to 100. In this case, we noticed a performance gain with the accuracy reaching 83%. Moreover, the confusion matrix below is improved with less false classifications.

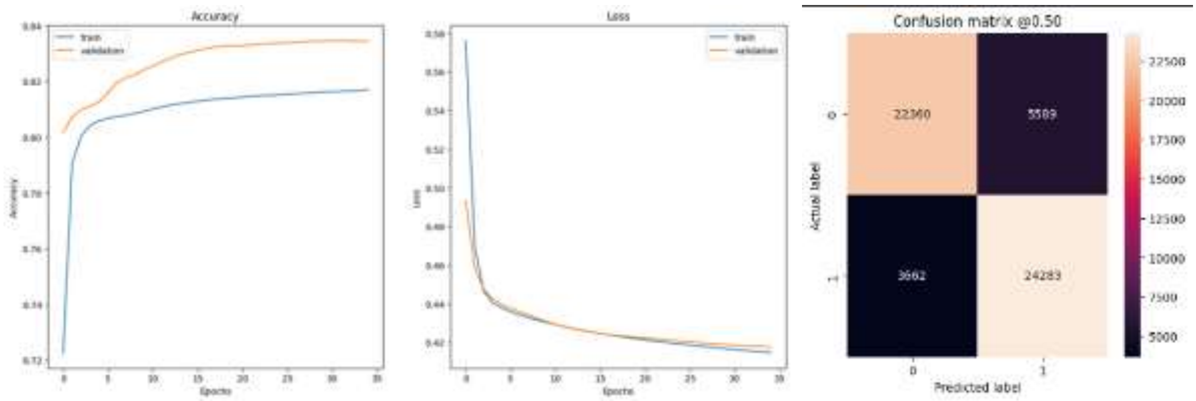

Figure 30: Model accuracy for the prediction of the level of turbidity (2 classes) at L10 station, 1-hour ahead, using 100 observations from the 4 upstream stations.

To support the prediction of the level of turbidity also at the most downstream metering station of the system (i.e., “Dafnoula”), a similar LSTM-based model was trained. In this case, as predictors we used past observations also from L10 station, along with turbidity at “Giona”, “Kirfi”, “Saradi” and “L8”. As indicated in Figure 31, again a high validation accuracy approximately equal to 80% was obtained, with a small number of false positives and negatives.

The above models consist a first endeavor to forecast the level of turbidity in the raw-water channel of EYDAP. There is a room for improvement (at least in theory) for the data-driven algorithms by training the models on the basis of larger datasets (comprising a higher number of high-turbidity events) and by incorporating other deep learning layers (e.g., convolutional networks).

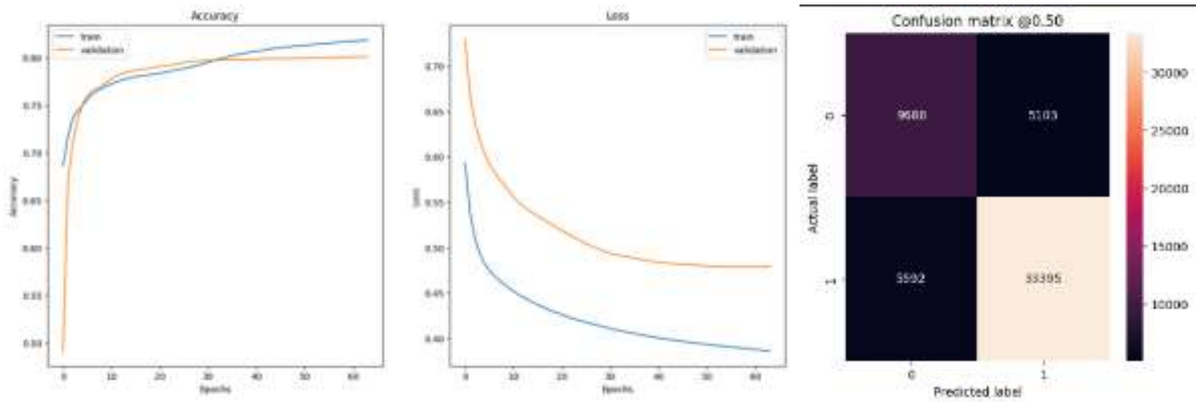

Figure 31: Model accuracy for the prediction of the level of turbidity (2 classes) at Dafnoula station, 1-hour ahead, using 100 observations from the 5 upstream stations.

The two deep LSTM-based models that forecast the level of turbidity at L10 and Dafnoula stations, 1-hour ahead of the latest measurement, have been implemented in an operational, FIWARE-enabled, context. The models are fed with real-time measurements of turbidity from the 6 metering stations and provide forecasts, which are presented to the operational staff of EYDAP in the dashboard of the web application. This operational implementation of the forecast models is presented in Deliverable D4.1.

IV. Analysis and forecast of water supply volumes

Further to the applications described in the previous sections, a tool for the analysis and forecast of total daily water supply volumes has been developed in the framework of the project. It is worth to mention that this application is not foreseen in the description of work of the Grant Agreement of the project, and it was developed as an initiative of NTUA, taking into account the operational needs of EYDAP.

IV.1. Data sources and features extraction

As discussed in Section I.2, further to real-time flow and quality data from the sensors in the conveyance system under study, EYDAP provided also access to the time series of daily inflows and outflows of the 4 WTPs that serve the city of Athens. The present analysis focuses on the total outflow from the 4 WTPs, as presented in Figure 32, for the period 01/01/2015 – 01/01/2022. The mean daily water outflow for Athens is approximately 1.075×10^6 m^3 per day.

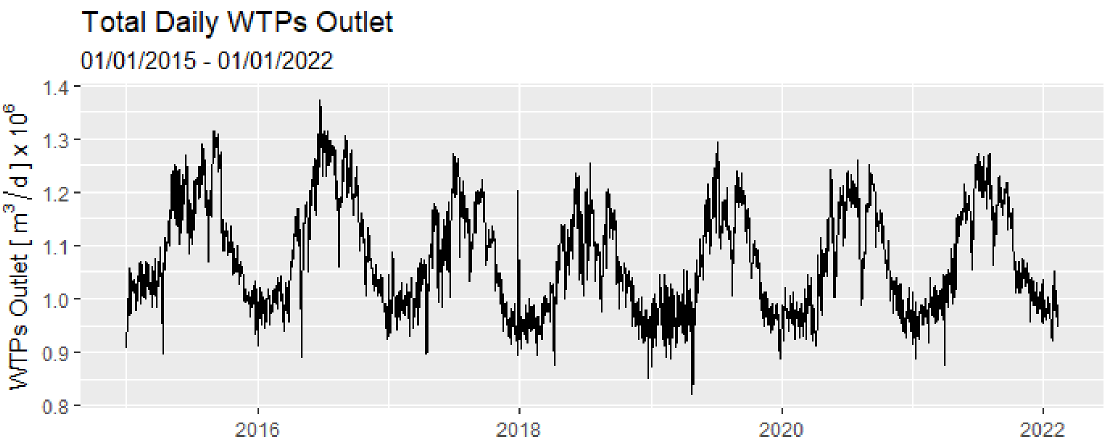

Figure 32: Time series of total daily water outflows from the 4 WTPs of Athens.

As it is evident in the above figure, the series is characterised by significant seasonal periodicity, with higher outflows during the summer months. This seasonal behaviour is further depicted in Figure 33 which presents the mean daily water outflow per month, for each year of the period 2015 – 2021.

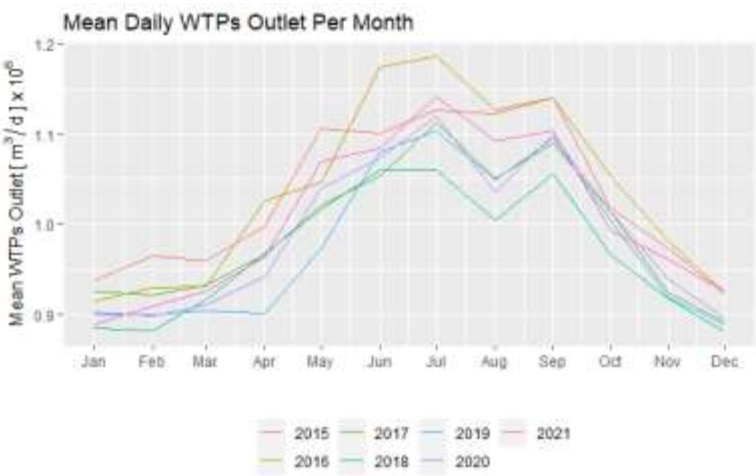

Figure 33: Seasonal variation of mean daily outflow per month, for the period 2015 – 2021.

Along with monthly periodicity, as presented in Figure 34, the series exhibits a weekly periodicity with higher mean water outflows on Saturday and lower mean water outflows on Sunday, compared to the working days of the week.

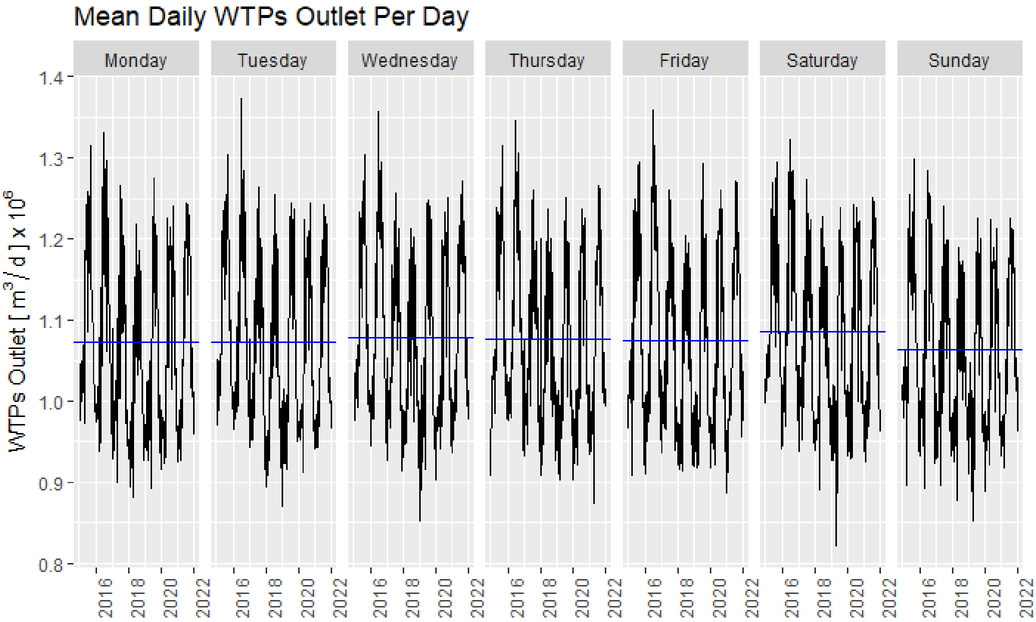

Figure 34: Variation of mean daily outflow per day of the week, for the period 2015 – 2021.

Furthermore, we gave special focus on the effect of special events in the variation of water outflows. Special events are mainly periods of public holidays (e.g., Easter holidays), during which water outflows vary significantly (both increases and decreases), compared to the days preceded and succeeded the event. Indicative examples of such events are given in Figure 35 and Figure 36, for the period of Easter and Christmas holidays, respectively. As shown in Figure 35, there is a significant decrease in water outflows from H. Friday to Monday, after the Easter day. This is mainly attributed to the fact that during this period there is a massive relocation of citizens outside the metropolitan area of Athens, while the days after Easter the outflow increases rapidly to the previous level.

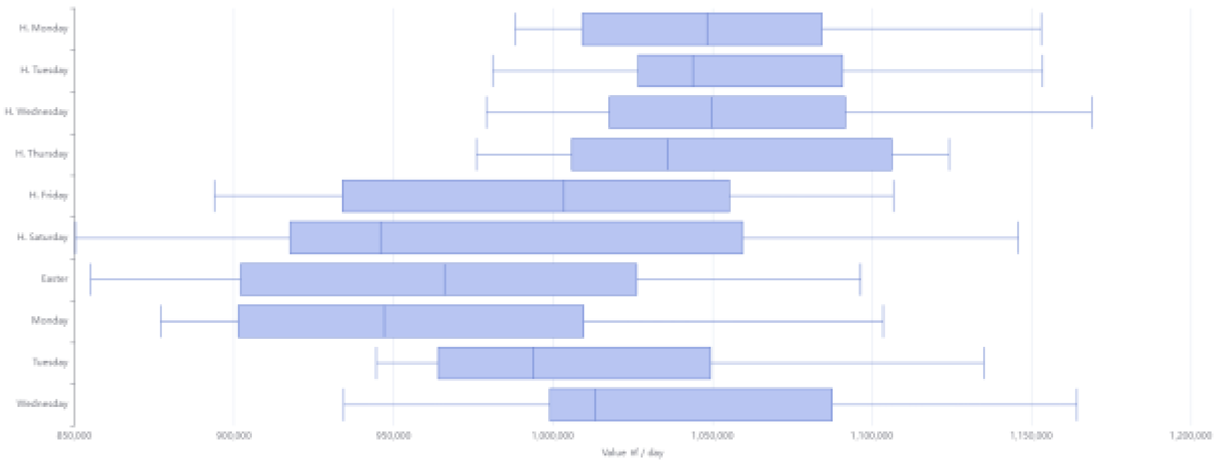

Figure 35: Box plot of daily water outflows during Easter Holidays, for the period 2015 – 2021.

An even higher variability is exhibited in the water outflows during Christmas holidays. As it is evident in Figure 36, a significant decrease in outflows appears on the Christmas day (25th of December) and the day after, while there is a significant increase on the 1st day of each year. Significantly higher is also the outflow on the 6th of January, which is related to the public holiday of Epiphany in Greece. It is also interesting to note that during the days of special events, the variability of outflow is high, as indicated by the larger length of boxes and whiskers.

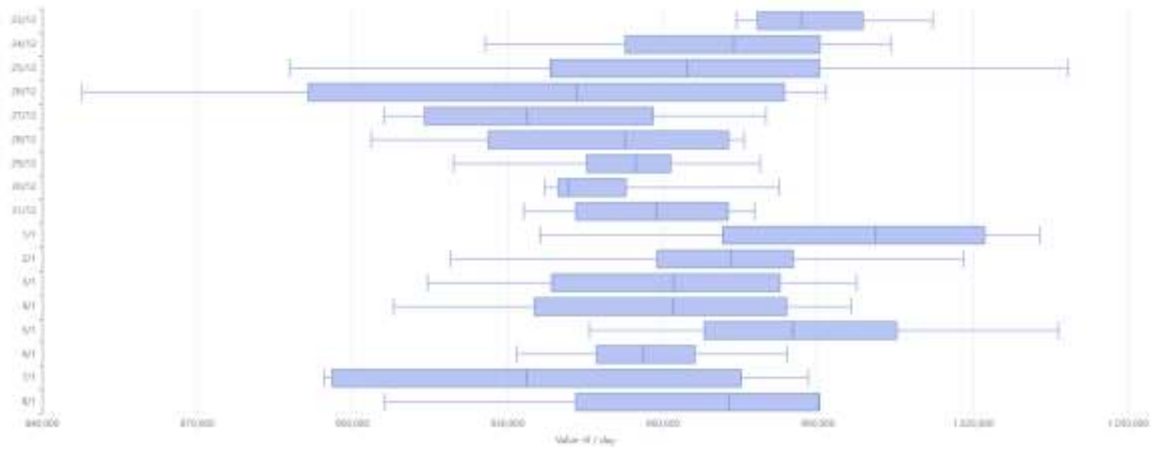

Figure 36: Box plot of daily water outflows during Christmas Holidays, for the period 2015 – 2021.

Further to the above special events, we also analysed the variability of water outflow during other public holidays in Greece. Specifically, the day of Epiphany, the 25th of March, the 28th of October, the 1st of May, the Monday of Holy Spirit, the Clean Monday and the 15th of August. In these days, water outflows exhibit a significant decrease compared to the mean outflow of the relevant period. This analysis informed the requirements and functionalities of the analytic for the analysis of the daily time series of water inflows and outflows of the 4 WTPs. The analytic is part of the web platform developed in the framework of F4W for EYDAP and is presented in Deliverable 4.1. Furthermore, the above analysis provides insights on the features that should be used as predictors in a model to forecast the total daily water outflows.

IV.2. Forecast daily water outflows

This section presents the tool that has been developed to allow the operational staff of EYDAP to obtain estimations of the next-day total water outflows, and hence to regulate the flow in the conveyance system accordingly. We addressed the problem as a time series forecasting problem using as predictor variables both endogenous characteristics of the series (e.g., past observations), as well as exogenous parameters (e.g., whether the day to be forecasted is a working or a weekend day). To forecast water demand volumes, various approaches can be found in the literature, ranging the classical decomposition and regression models to the ARIMA-type models, and to the more advanced neural-network-based approaches, depending to the time scale and the time horizon of interest (e.g., see [13]–[15]).

For the purposes of the present work, we built a dynamic regression model, which can be easily re-trained and maintained by the operational staff of EYDAP. Alternative model structures were assessed by using different variables as predictors. Specifically, we examined the forecast accuracy gained by using as predictors the outflow of previous day, the outflow two days before, the month of the day of interest, whether the day of interest is a working or week-end day, and whether the day is characterised by a special event (e.g., Easter holiday). The selection of the most suitable model was conducted on the basis of Akaike's Information Criterion (see [16]), which is a typical measure of assessment of the forecast accuracy obtained, along with the complexity of the model and number of observations available. The model selected, to forecast the one-day ahead water outflow, Q_t, reads as:

$$Q_t = a_1 Q_{t-1} + a_2 W_t + a_3 D_t + a_4 H_t + \varepsilon_t \quad \text{(IV.1)}$$

where, Q_{t-1} is the outflow of previous day, W_t represents the Fourier terms to account for annual seasonality, D_t represents a dummy variable that takes value 1 for working days and 0 for week-end days, H_t a dummy variable to account for special events, and ε_t the error term. It is highlighted that H_t takes value 0 for a regular day, 1 for a day with a special event that significant increase is expected in water outflows (e.g., the 1st day of each year) and -1 for a day that a significant decrease is expected (e.g., Easter day).

Having said the above, it is worth to highlight that for the development of this application, the outliers, which, in this case, are the exceptionally high or low values, that appear on days of special events (e.g., Easter or Christmas holidays), did not smoothed or removed. On the contrary, the harmonic regression model (Eq. IV.1) was designed and calibrated so as to capture the daily sudden daily fluctuations due to special events, by using as predictor a predictor variable (i.e., H_t) that indicates whether a day is characterised by such an event or not.

To train the model we use the time series from 15/12/2015 to 15/12/2019, keeping the series from 16/12/2019 to 10/02/2022 for validation purposes. The model was fitted using the "forecast" [17] library in R, by minimizing the sum of the squared errors between observed and forecasted values. A first, visual, comparison between the observed the forecasted water outflows, is given in Figure 37 and Figure 38, for the training and validation period, respectively.

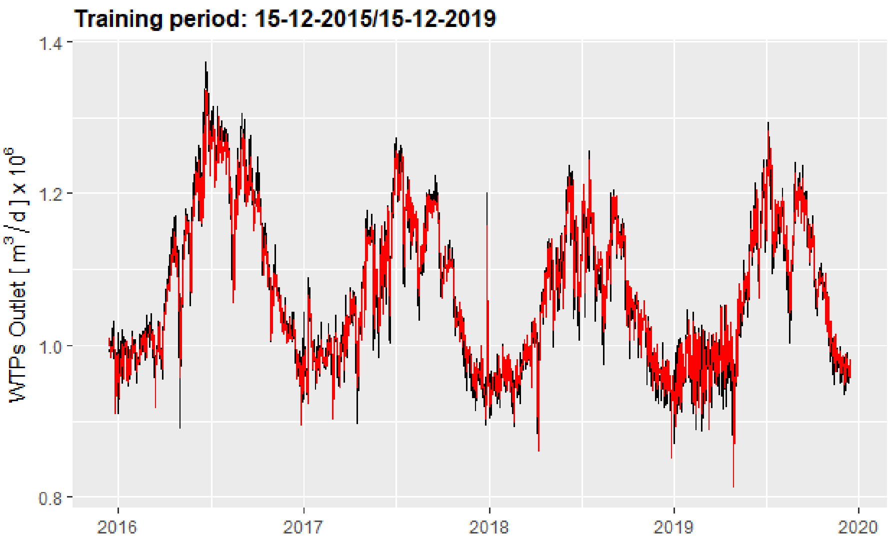

Figure 37: Observed (black line) vs. predicted (red line) outflows for the training set.

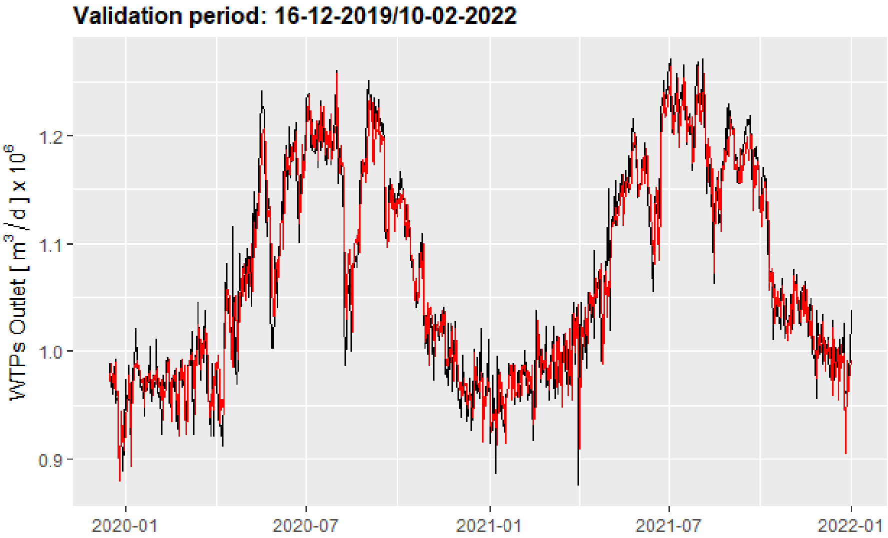

Figure 38: Observed (black line) vs. predicted (red line) outflows for the validation set.

To characterise the performance of the model, we use three measures that are commonly used by relevant studies. Specifically, we employed the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the coefficient of determination (R^2), as given via the following expressions:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \hat{Y}_i\right| \qquad \text{(IV.2)}$$

$$\mathrm{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{Y_i-\hat{Y}_i}{Y_i}\right| \qquad \text{(IV.2)}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Y_i-\hat{Y}_i)^2}{\sum_{i=1}^{N}(Y_i-\bar{Y}_i)^2} \qquad \text{(IV.3)}$$

where, Y_i is the observed value, $\hat{Y}_i$ the forecasted value and $\bar{Y}_i$ the mean of the observed values. As we see in Table 7, which summarises the values of evaluation measures for the training and validation dataset, the model forecasts next-day outflow with high accuracy.

Table 7: Evaluation measures for the training and validation dataset

	$\mathbf{MAE}\ (\mathbf{m^3/d})$	$\mathbf{MAPE}\ (\%)$	$\boldsymbol{R^2}(-)$
Training set	23,662.06	2.24	0.8965
Validation set	21,540.49	2.10	0.916

V. Conclusion and Perspectives

This deliverable presents the smart applications (algorithms and models) developed in the framework of F4W to provide decision operational support to EYDAP, with respect to both flow and quality aspects of the raw-water external conveyance system, that serves the metropolitan area of the city of Athens. The applications developed concern: a) the provision of advice on optimal sluice gate openings depending on the flow conditions and needs, b) the provision of early warnings for high turbidity events and one-hour ahead forecasts of the level of turbidity at the downstream part of the system, and finally c) the provision one-day ahead forecasts of water outflows from the 4 water treatment plants. The applications adopt different modelling approaches depending on the peculiarities and requirements of the problem. In this vein, a grey-box approach has been adopted to simulate the flow through and over the regulation structures of the channel, combining the well-known physics-driven equations with data-driven techniques for parameter estimation and data reconciliation. Moreover, to provide forecasts for key aspects of the system, data-driven approaches have been adopted, ranging from the LSTM-based neural network models, for the highly complex and non-linear problem of turbidity forecast, to the more traditional regression-type models to forecast the total daily water outflows. It is worth mentioning that this is the first time that such type of models is deployed in EYDAP, in an operation and interoperable context, aiming to go beyond the management and operation of the complex conveyance system on the basis of empirical rules and practices.

The activities towards the development of such applications unfold new perspectives and opportunities. The desk studies as well as field visits at key points of the conveyance system allowed to gather valuable insights and knowledge for the operation of such a complex real-world conveyance system. The information and data gathered within F4W are undoubtedly a valuable prior knowledge that can directly inform relevant developments in the future. Furthermore, the analysis of parameters that have never been examined systematically in the past, such as turbidity and hydraulic parameters of the system, provided unrevealed insights for both EYDAP and NTUA. The analysis of the available dataset, in combination with the accuracy of models developed, indicates the necessity for validation and reconciliation of data from the existing sensors, before using it in the training of models and their operational implementation. In this context, the development of a real-time and automated data validation application, which will allow detection and correction of faulty measurements, imputation of missing values and smoothing of series, could be a priority as future development. Taking advantage of the dataset collected, alternative modelling procedures can be easily implemented and evaluated, allowing for even more accurate outcomes. To conclude, the fact that the applications developed are FIWARE-compliant increases substantially their transferability, allowing their testing and implementation for other cases in a straightforward way.

VI. European added value (EAV) and upscaling

Raw-water conveyance system is a key element of water supply chain, conveying water from sources to water treatment plants. Compared to drinking water networks, raw-water systems are characterized by much higher diversity and are typically unique, tailored to the characteristics of a region and the purposes that aim to serve. Such systems are typically complex and of large-scale nature, composed by different interconnected infrastructures and hydraulic works (e.g., reservoirs, aqueducts, water regulation structures, diversion structures, energy production and dissipation units) that serve different, usually conflicting and variable, targets (e.g., reliable water supply, energy production, water storage, environmental target, flood protection).

In the framework of Fiware4Water, a series of smart applications and models have been developed for the raw-water conveyance system of Athens, that is one of the largest and more complex in Europe and worldwide. The conveyance system transfers raw-water from the sources (Mornos and Evinos reservoirs) located at the western regions of Greece to the eastern regions, and specifically to Attica (Athens), via a complex system of pressurized and free-flow aqueducts. The entire system has a total length of 495 km. In F4W we gave special focus on three challenges, and we developed smart applications, which are of high practical interest and may find applicability to any typical raw-water system. These are: a) the optimal regulation of flow in a system to convey timely specific water volumes, b) the forecast of water volumes that should be delivered on daily basis to serve the needs of customers, and c) the provision of early-warning for unusual (high-turbidity) quality events and forecasts for the level of turbidity at the downstream parts of the system given the quality conditions upstream.

Although, modifications are required for the applications to be tailored to the peculiarities of other raw-water systems in Europe, the potential for their further uptake is large. The operational "grey-box" methodology developed for the optimal regulation of flow can be deployed in other raw-water systems, as well as in irrigation channels, where flow control is performed via sluice gates. The developed tool can be easily extended and upscaled for other parts and regulation structures of the conveyance system of EYDAP, building a decision support system that will provide advice about optimal regulation settings across the entire conveyance system, from Mornos up to the water treatment plants. Furthermore, the approaches to build, train and evaluate the data-driven predictive models for critical processes of such systems (i.e., turbidity and water supply volumes) can be employed to other raw-water systems and for other processes with similar characteristics (e.g., water demand, conductivity).

The development of smart applications and models was achieved by close and extensive collaboration between NTUA (the scientific partner) and EYDAP (the operator of the system). This allowed the exchange of complementary knowledge both on the way that such a complex system operates under different conditions, as well as on the modelling approaches that can be adopted, depending on the requirements posed by the operators and the available data sources. During the development of smart applications, large datasets of observations for different key qualitative and quantitative aspects of raw-water were collected, analysed and pre-processed. Much of the experience and insights gained during this process consist valuable prior knowledge that can directly inform the development of similar services across Europe. It is worth to highlight that this knowledge does not concern only the development of accurate models, but also on their implementation to a live operational environment, aiming to bridge science to practice.

In the light of the above, the European added-value (EAV) has multiple aspects: (i) showcase the performance of different modelling approaches, from physically-based to pure data-driven models, to address key challenges and operational requirements for the raw-water sector, (ii) gathering of knowledge on practical issues and challenges, which can be found in almost all typical raw-water systems (i.e., dataset with high percentage of missing values, erroneous measurements, highly noisy behaviour), (iii) advancing decision support through smart applications towards the optimal operation of raw-water systems. Overall, the developed scientific models, along with the web platform, developed within F4W, serve as a paradigm of a complete digital, and interoperable, solution on how utilities can take advantage of the integration of different data sources, along with analytics, to improve the operational management of large and complex raw-water systems.

References

[1] S. Wu and N. Rajaratnam, “Solutions to Rectangular Sluice Gate Flow Problems,” *Journal of Irrigation and Drainage Engineering*, vol. 141, no. 12, p. 06015003, Dec. 2015, doi: 10.1061/(ASCE)IR.1943-4774.0000922.

[2] *Hydraulics of Open Channel Flow*. Elsevier, 2004. doi: 10.1016/B978-0-7506-5978-9.X5000-4.

[3] P. K. Swamee, “Sluice-Gate Discharge Equations,” *Journal of Irrigation and Drainage Engineering*, vol. 118, no. 1, pp. 56–60, Jan. 1992, doi: 10.1061/(ASCE)0733-9437(1992)118:1(56).

[4] Gersar and Tetraktys partnership, *Technical study for the first phase of the water conveyance system Mornos-Athens. Department D5, Hydraulic Works service, Greek Ministry of Public Works*. 1974.

[5] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, “Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions,” *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, Jan. 1998, doi: 10.1137/S1052623496303470.

[6] M. H. Chaudhry, *Open-Channel Flow*. Boston, MA: Springer US, 2008. doi: 10.1007/978-0-387-68648-6.

[7] V. Bellos, I. Nalbantis, and G. Tsakiris, “Friction Modeling of Flood Flow Simulations,” *Journal of Hydraulic Engineering*, vol. 144, no. 12, p. 04018073, Dec. 2018, doi: 10.1061/(ASCE)HY.1943-7900.0001540.

[8] P. Virtanen *et al.*, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.

[9] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[10] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[11] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.

[12] F. Chollet, “Keras,” *GitHub*, 2015. https://github.com/fchollet/keras (accessed Apr. 19, 2022).

[13] E. Pacchin, S. Alvisi, and M. Franchini, “A Short-Term Water Demand Forecasting Model Using a Moving Window on Previously Observed Data,” *Water (Basel)*, vol. 9, no. 3, p. 172, Feb. 2017, doi: 10.3390/w9030172.

[14] L. Mu, F. Zheng, R. Tao, Q. Zhang, and Z. Kapelan, “Hourly and Daily Urban Water Demand Predictions Using a Long Short-Term Memory Based Model,” *Journal of Water Resources Planning and Management*, vol. 146, no. 9, p. 05020017, Sep. 2020, doi: 10.1061/(ASCE)WR.1943-5452.0001276.

[15] G. Guo, S. Liu, Y. Wu, J. Li, R. Zhou, and X. Zhu, "Short-Term Water Demand Forecast Based on Deep Learning Method," *Journal of Water Resources Planning and Management*, vol. 144, no. 12, p. 04018076, Dec. 2018, doi: 10.1061/(ASCE)WR.1943-5452.0000992.

[16] R. J. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice," *3rd edition, OTexts: Melbourne, Australia.*, 2021. OTexts.com/fpp3 (accessed Apr. 19, 2022).

[17] R. J. Hyndman and Y. Khandakar, "Automatic Time Series Forecasting: The forecast Package for R," *Journal of Statistical Software*, vol. 27, no. 3, 2008, doi: 10.18637/jss.v027.i03.