

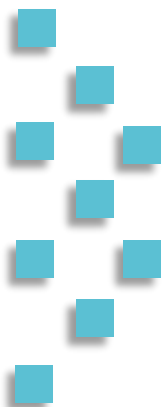


D3.2 FIWARE-enabled application for Water Distribution

Author: Lluís Echeverría (EUT), Marc Ribalta (EUT), Danillo Lange (EUT)

Co-Authors: Stéphane Deveughèle (3S), Abel Dembele (3S), Chouaib Mkireb (3S), Stéphane Aguiar (3S), Martin Wagner (TZW), Theresia Meltzer (TZW) Brett Snider (UNEXE) Didier Dumet (EGM)

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

© 2022, FIWARE4Water consortium

All rights reserved.

This document contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

This document is the property of the FIWARE4Water consortium members. No copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights. In addition to such written permission, the source must be clearly referenced.

Project Consortium



Executive Summary

Deliverable 3.2 technically describes the efforts and developments carried out in task T3.2 ‘Smart Applications for Water Distribution’, in WP3 ‘Smart Applications and Devices’ of the FIWARE4Water (F4W) project.

The objective of this task is to develop a set of smart applications that improve operational and tactical decision support and infrastructure management in the Urban Water Supply domain. In this scenario, four topics have been focused: i) forecasting water demand, ii) prediction of water resource availability, iii) the detection and location of water leakage events, and iv) the detection of water contamination events.

To demonstrate the advanced capabilities of the implemented applications, its development has been aligned with the integration of the FIWARE ecosystem in the so called FIWARE4Water architecture and deployed and validated into two different real Demo Cases (DCs). Mainly, the developments detailed in this document are demonstrated in the Cannes (France) DC, but the third topic (the detection and location of water leakage events) is also implemented in the Great Torrington (UK) DC, since both cases deal with the Water Supply complexities, either from cities, water utilities, water authorities, citizens, or consumers point of view. The innovative and advanced properties of the FIWARE4Water digital architecture are reported in the corresponding deliverables of WP2 ‘Architecture/Data/Ontology/API/Legacy links/Standards’. Furthermore, the deployment of the smart applications into the real systems, in the proposed DCs, is presented in deliverables D4.2 ‘FIWARE4 Leakage Management’, D4.3 ‘FIWARE4 Water Quality Monitoring and Pollution Response’ and D4.5 ‘FIWARE4 Smart Metering and Citizen Engagement’, all being outcomes of the work done in the WP4 ‘Demonstrating FIWARE4Water in the Real (Water) World’.

Essentially, the applications are based on predictive models, which provide them with a kind of intelligence, thus generating novel, effective and efficient solutions to the proposed water domain issues. These predictive models are boosted principally by data-driven algorithms even though, in some cases, they are also combined with numerical mechanisms (e.g. hydrological or hydraulic models) in order to obtain advanced capabilities.

The main objective of this document is to provide a detailed explanation (at technical level) of how these predictive models are designed, implemented, developed, and evaluated, thus enabling reproductivity and upscaling to other cases of similar characteristics and challenges. To do so, the description of these phases follows the same process and methodology used during the same project development. This process is an adaptation of the CRISP-DM (Cross-industry standard process for data mining), an open standard and data science methodology, which defines six sequential and iterative steps focused on the development of data mining projects: i) Business understanding, ii) Data understanding, iii) Data preparation, iv) Modelling, v) Evaluation, and vi) Deployment.

Five partners have contributed to this task to the development of the proposed smart models by demonstrating different innovative properties and characteristics, addressing different requirements, using different technologies, and demonstrating the FIWARE integration and interoperability capabilities (e.g. external sensors, legacy systems, different technologies, etc), as shown below:

- I. Water demand forecasting:
 - a. AI & ML: A semi-supervised approach for water demand forecast. Algorithms are implemented through MATLAB technology.
 - b. Big Data: Regressive algorithms for water demand forecast implemented through Python technology with horizontal scalability and high-performance properties.

- II. Prediction of water resource availability:
 - a. AI & ML: long-term forecasting of water resource availabilities. Algorithms are implemented through MATLAB technology.
 - b. Hybrid approach: a combination of data-driven & hydrological models. Real-time, continuous water availability forecast.
- III. Water leaks detection:
 - a. AI & ML: A supervised classification problem implemented through MATLAB technology.
 - b. Big Data: A supervised classification algorithm for water leaks detection implemented through Python technology with horizontal scalability and high-performance properties.
 - c. Hydraulic models: A model-driven methodology that relies on the EPANET model of the water distribution for leak detection.
 - d. Workforce tool: A planner responsible for finding the optimal scheduling of workforces based on resources optimization.
- IV. Detection of water contamination events:
 - a. AI & ML: drinking water quality events detection through an unsupervised approach. Algorithms are implemented through MATLAB technology.
 - b. AI & ML: A prediction-based approach combining Univariate Signal Evaluation and Forecasting-based Detector methodologies for water quality data.

As a result, four smart applications have been developed, one for each Water Supply topic, which combine the aforementioned capabilities in an ensembled solutions with the aim to provide the involved water utilities with an advanced Decision Support Systems for improved decision making, and resources and infrastructure management.

Finally, the EU added value and policies recommendations related to this document are detailed in the Conclusion sections.

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.

D4.2 – “FIWARE4_Leakage Management”, where the deployment and end-testing of the developed smart solutions for the French demo case are described.

D4.3 – “FIWARE4_Water Quality Monitoring and Pollution Response”, where some of the anomaly detection algorithms are tested under practical conditions.

D4.5 – “FIWARE4_Smart Metering and Citizen Engagement”, where the deployment and end-testing of the smart solutions for the United Kingdom demo case are described.

Document Information

Programme	H2020 – SC0511-2018
Project Acronym	FIWARE4Water
Project full name	FIWARE for the Next Generation Internet Services for the WATER sector
Deliverable	D3.2: FIWARE-enabled applications for Water Distribution
Work Package	WP3: Smart Applications and Devices
Task	Task 3.2: Smart Applications for Water Distribution
Lead Beneficiary	Eurecat
Author(s)	Lluis Echeverria (EUT), Marc Ribalta (EUT), Danillo Lange (EUT)
Contributor(s)	Stéphane Deveughèle (3S), Abel Dembele (3S), Chouaib Mkireb (3S), Stéphane Aguiar (3S), Martin Wagner (TZW), Theresia Meltzer (TZW) Brett Snider (UNEXE), Didier Dumet (EGM)
Quality check	Panagiotis Kossieris (NTUA)
Planned Delivery Date	30/04/22
Actual Delivery Date	30/04/22
Dissemination Level	Public (Information available in the Grant Agreement)

Revision history

Version	Date	Author(s)/Contributor(s)	Notes
1.0	21/04/2022	Lluis Echeverria (EUT), Marc Ribalta (EUT), Danillo Lange (EUT), Stéphane Deveughèle (3S), Abel Dembele (3S), Chouaib Mkireb (3S), Stéphane Aguiar (3S), Martin Wagner (TZW), Theresia Meltzer (TZW) Brett Snider (UNEXE), Didier Dumet (EGM)	A first version of the document, with the integration of all the work related to this task.
2.0	22/04/2022	Panagiotis Kossieris (NTUA)	Internal review of the document
3.0	29/04/2022	Lluis Echeverria (EUT), Marc Ribalta (EUT), Danillo Lange (EUT)	Final version of the document

Table of content

Executive Summary	3
List of figures	9
List of tables	12
List of Acronyms/Glossary.....	14
Introduction	16
I. Water demand forecasting.....	19
I.1. Business understanding.....	19
I.2. Data understanding.....	19
I.3. 3s modelling	22
i. Data preparation	22
ii. Modelling.....	23
iii. Evaluation	24
I.4. Big data modelling.....	28
i. Data preparation	28
ii. Modelling.....	30
iii. Evaluation	31
II. Water resource availability forecasting	33
II.1. Business understanding.....	33
II.2. 3s modelling	34
i. Data understanding.....	35
ii. Data preparation	36
iii. Modelling.....	37
iv. Evaluation	40
II.3. Combined LSTM-Hydro modelling.....	43
i. Data understanding.....	44
ii. Data preparation	47
iii. Modelling.....	48
iv. Evaluation	51
III. Water leak detection and location.....	56
III.1. French demo case.....	56
i. Business understanding.....	56
ii. Data understanding	56

iii.	3S modelling	62
1.	Data preparation	62
2.	Modelling.....	64
3.	Evaluation	65
iv.	Big data modelling.....	67
1.	Data preparation	67
2.	Modelling.....	68
3.	Evaluation	70
III.2.	United kingdom demo case.....	73
i.	Business understanding.....	73
ii.	Water leak detection.....	74
1.	Data understanding	74
2.	Data preparation	75
3.	Modelling.....	77
4.	Evaluation	77
iii.	Leak Localization.....	78
1.	Data understanding	78
2.	Data preparation	78
3.	Modelling.....	79
4.	Evaluation	80
iv.	Mitigation actions.....	83
1.	Data Understanding	83
2.	Modelling.....	83
3.	Evaluation	86
v.	Leak Management Model Conclusion	86
IV.	Water contamination events.....	87
IV.1.	Detection of water quality anomalies	87
i.	Business understanding.....	87
ii.	3S modelling	89
1.	Data understanding	89
2.	Data preparation	91
3.	Modelling.....	92
4.	Evaluation	94
IV.2.	TZW modelling.....	96
1.	Data understanding	96
2.	Data preparation	100

3.	Modelling.....	100
4.	Evaluation.....	116
V.	Workforce tool to manage operator’s tasks.....	117
V.1.	Data understanding.....	118
V.2.	Planner.....	120
i.	Technologies.....	120
ii.	Cost function	120
V.3.	Results	121
	Conclusions and perspectives.....	123
	European added value (EAV) and upscaling	125
	Annex I: Supplementary Information – Explanation of CRISP-DM Methodology.....	128
	Annex II: Supplementary Information – Water demand forecasting.....	132
	Annex III: Supplementary Information – Water leakage.....	139

List of figures

Figure 1 Water demand time series (DMS 1)	21
Figure 2 Distribution (density) plots of water demand for sectors 1-16.....	22
Figure 3: Model architecture.....	23
Figure 4: Scheme of water demand estimation (Computation Step)	24
Figure 5 Prediction vs observation of DMS 1, 6, 18 and 25	25
Figure 6: Observations vs Predictions DMS 13.....	26
Figure 7:Example with and without weather data.....	26
Figure 8 ACF and PACF plots of DMS 1 time-series.....	29
Figure 9 Cross-correlation plots of DMS 1 against 27, 30 and 32	29
Figure 10 Cross-correlation heatmap considering all DMS.....	30
Figure 11. Spark structure	31
Figure 12 Demand prediction plot for sector 1 in 3 days (from 11/11 to 13/11).	32
Figure 13. Demand models performance graph during training.	33
Figure 14: Description of the water resources considered in this document.....	35
Figure 15: Evolution of water resources volumes from January 2018 to August 2020	37
Figure 16: Machine learning models tested for the water availability forecasting approach.	38
Figure 17: Result of the prevision model for the Bramafan water resource	40
Figure 18: Long-term forecasted weather data for Bramafan source	41
Figure 19: Perfect knowledge of weather data for Bramafan source.....	42
Figure 20: Perfect knowledge of weather data for Foux Saint-Cézaire source.....	42
Figure 21: Long-term forecasted weather data for Foux Saint-Cézaire source	43
Figure 22 Map of the analysed zone	45
Figure 23 Discharge, precipitation, and temperature timeseries in historical years	45
Figure 24 Random monthly aggregated years	46
Figure 25 Correlation matrix	46
Figure 26 Real time data	47
Figure 27 Météo France aggregated data	48
Figure 28 LSTM predictive model architecture	49
Figure 29 Hydrological model structure.....	50
Figure 30 Total monthly precipitation prediction	52
Figure 31 Simulated monthly temperature.....	53
Figure 32 Simulation of the hydrologic model	54
Figure 33 Number of variables per sector.....	57
Figure 34 Residual chlore pumping (Sector 901)	61
Figure 35 Rain and leakages values for sector 904.	62
Figure 36 Correlation heatmap of data fields of sector 904.	62
Figure 37: Transformation process	63
Figure 38: Example of training phase with bagged trees	64
Figure 39: Model process	65
Figure 40: Confusion Matrix DMS902	66
Figure 41: Confusion Matrix DMS904	66
Figure 42 Active leakages plot for sector 901.	67
Figure 43 Class count distribution for each sector.....	68
Figure 44 Spark structure.	69
Figure 45 Weighted metrics for every sector.....	72

Figure 46 Leakage models performance graph during training	73
Figure 47 Greater Torrington EPANET Schematic	74
Figure 48 Search Radius versus Leaks Located within Search Areas.....	81
Figure 49 Comparing Search Areas for Various Localization Methods	83
Figure 50: Probes position on the network.....	89
Figure 51: Raw probe data	90
Figure 52 Conductivity - pH - UV observations, from left to right.	90
Figure 53 Pressure - Chlore - Temperature observations	90
Figure 54 Turbidity and TOC observations	91
Figure 55 From left to right, UV data before process, UV after process and the Boxplot of the UV observations before processing	91
Figure 56: Model description	92
Figure 57: Illustration of threshold module	92
Figure 58: Illustration of drift module	93
Figure 59: Chlore event	94
Figure 60: Turbidity event	95
Figure 61: TOC/Turbidity event	95
Figure 62 Time series of the BDD_Zone_804 zone of the Cannes Demo Case for the year 2011	98
Figure 63: Time series of free chlorine in all four zones of the Cannes Demo Case over a two-week period.	98
Figure 64: Correlation matrix (Pearson's correlation coefficient) of water quality parameters of zone 802.....	99
Figure 65: different types of anomalies found after visual inspection of the data during exploratory data analysis (free chlorine, zone 802).	99
Figure 66: Outliers, two different event types (peaks, step), states.....	101
Figure 67: The learning process has a large impact on how much and which events can be captured. Models that are incrementally updated can learn new states (top). Models that are trained only once capture new states completely as events (bottom).....	102
Figure 68: Basic structure of the model, which can be used to recognize four different types of events.	103
Figure 69: Example of a time series with limit values (legal requirements) and value ranges (measuring range of the sensor).	104
Figure 70: Difference between single (outlier) and permanent limit violation (event).....	104
Figure 71: Removal of outliers from a time series using Median Absolute Deviation (MAD) algorithm.	105
Figure 72: Detection of changes in the quality (noise) of the signal.....	106
Figure 73: Variance of the residuals between normal period and period of the event.....	106
Figure 74: Workflow of UnivariateDetector.....	108
Figure 75: Workflow of ForecastDetector (Training). Please note that the process for training is performed twice. After the first training, outliers are removed from the training data set and then the training is repeated.	109
Figure 76: Transformation of the input data (X, left). A shift adds new columns to X (X0, center). This dataset forms the input for the subsequent principal component analysis. The target variables (next time step) is also created by a shift (Y, right). The letters f1, f2, etc. stand for feature 1, feature 2, etc. and t1, t2, etc. for time step 1, time step 2, etc.....	110
Figure 77: Iterative calculation of principal components and self-filling missing values.....	112

Figure 78: Time series of the turbidity (Cannes Demo Case). The original signal is very stable and shows small but negligible fluctuations. Adding a small amount of noise in the training phase of the model results in small fluctuations not being classified as events..... 114

Figure 79: Example of applying the ForecastDetector to data from the Cannes Demo Case (zone 802). 116

Figure 80: Example of manual classification of events based on the time series of free chlorine in zone 801 (Cannes Demo Case)..... 116

Figure 81: For the determination of recall, precision and F1 score, it is checked whether the detected events overlap with the manually classified events. A minute-by-minute comparison of individual time steps is not performed. FN : false negative, TP : true positive, FP : false positive..... 117

Figure 82 Models development strategy 128

List of tables

Table 1 Data source details	19
Table 2 General details of source data.....	19
Table 3 Fields general details	20
Table 4 Month with the highest consumption, per year (DMS 1).....	21
Table 5: Historical data.....	24
Table 6: Forecasting data	24
Table 7: Input descriptions.....	25
Table 8: Summary of results.....	26
Table 9 Models performance table.....	32
Table 10: Extraction of some values for weather data provided by our supplier.....	35
Table 11: Extraction of some values for water resource availability for one particular source	35
Table 12: Evaluation of the performance of the linear model for a particular water resource	39
Table 13: Train results for the Bramafan resource	39
Table 14: Forecasting results for summer 2018.....	40
Table 15: Forecasting results for summer 2019.....	41
Table 16: Forecasting results for summer 2020.....	41
Table 17 Meteorological historic data.....	44
Table 18 Joined meteo data sets.....	44
Table 19 Model calibration result	51
Table 20 Scoring results of the prediction performed by the LSTM	52
Table 21 Predicted value and scoring metrics for each variable.....	53
Table 22 Results of the evaluation for the historical period.....	54
Table 23 Evaluation of the real time data	55
Table 24 Data source details.	56
Table 25 Data availability across sectors.....	57
Table 26 General details of source data.....	57
Table 27 Statistical metrics of sector 901.	58
Table 28 Statistical metrics for sector 902.	59
Table 29 Statistical metrics for sector 903.	59
Table 30 Statistical metrics for sector 904.	60
Table 31 Statistical metrics for sector 905.	60
Table 32 Statistical metrics for sector 906.	61
Table 33: Table of parameters	65
Table 34 Training dataset features.....	69
Table 35 Multiclass classification algorithms accuracy (the average precision is weighted 1).....	70
Table 36 Metrics separated by class.	70
Table 37 Classification results of class 3 predictions.....	71
Table 38 Performance metrics results	72
Table 39 Greater Torrington Hydraulic Model Attributes.....	74
Table 40 Greater Torrington Simulated Pressure and Flow Network Sensors.....	75
Table 41 Impact of grid spacing on predicted leaks located by MCLP	81
Table 42 Impact of Search Areas on Anomaly Detection	82
Table 43 Comparing Mitigation tool’s proposed solution versus default repair method.....	86

Table 44: Input description	94
Table 45 Overview about the used sensors	96
Table 46: Limits of French drinking water guideline	97
Table 47: Hyperparameters of module UnivariateDetector	107
Table 48: Hyperparameters of forecasting-based event detection (ForecastDetector).....	114
Table 49 Work Order's data model	118
Table 50 Technician's data model	119
Table 51 Work Order Request data model	119
Table 52 Planification Result data model.....	121
Table 53 Scenario's data model.....	122
Table 54 Demand field statistical metrics - Debit distribution.....	132
Table 55. Water demand forecasting - Outliers table.....	133
Table 56. Best neighbours table	135
Table 57 Features of the training dataset for forecasting the demand	136
Table 58 Demand models performance metrics	137
Table 59 Fields details.	139

List of Acronyms/Glossary

AI	Artificial Intelligence
ACF	Autocorrelation Function
CRISP-DM	Cross-industry standard process for data mining
DM	Demo Case
DMA	District Metered Areas
DMS	District Metered Sector
DOC	Dissolved Organic Matter
EWMA	Exponential Weighted Moving Average
F4W	FIWARE4Water project
FPR	False Positive Rate
GBT	Gradient Boost Tree
k-NN	k Nearest Neighbours
LSTM	Long short-term memory
MAE	Mean Absolute Error
MAD	Median Absolute Deviation
MCLP	Maximum Coverage Location Problem
ML	Machine Learning
MSE	Mean Squared Error
NGI	Next Generation Internet <i>The Next Generation Internet (NGI) initiative, launched by the European Commission in the autumn of 2016, aims to shape the future internet as an interoperable platform ecosystem that embodies the values that Europe holds dear: openness, inclusivity, transparency, privacy, cooperation, and protection of data.</i>
NRMSE	Normalized Root Mean Squared Error
PACF	Partial Autocorrelation Function
PI	Performance Indicator
RMSPE	Root Mean Squared Percentage Error Loss
SLA	Service Level Agreement
SVM	Support Vector Machine
TOC	Total Organic Carbon

- UV** Ultraviolet
- WDS** Water Distribution System
- WO** Work Order
- WP** Work Package
- WPL** Work Packages Leaders