Ref. Ares(2022)1628774 - 04/03/2022

SYnergy of integrated Sensors and Technologies for urban sEcured environMent

D3.3
FINAL DESIGN OF THE DATA FUSION METHODOLOGY

04 March 2022 **V3.0**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 787128

Project title	SYnergy of integrated Sensors and Technologies for urban sEcured environment
Project acronym	SYSTEM
Project number	787128
Start date of the project	1st September, 2018
Duration	42 months
Topic	SEC-10-FCT-2017. Integration of detection capabilities and data fusion with utility providers' network

Deliverable number	D3.3
Deliverable title	FINAL DESIGN OF THE DATA FUSION METHODOLOGY
Leading partner	WUT
Partners contributing	WUT, BTEC
WP of reference	WP3
Title of the WP of reference	DATA ANALYSIS FROM UTILITY NETWORKS: MODELLING AND VALIDATION
Task of reference	T3.3 and T3.4
Title of the task of reference	T3.3 DESIGN OF THE DATA FUSION CONCEPT T3.4 EVALUATION OF THE DATA FUSION CONCEPT
Deliverable type	Report
Dissemination level	PUBLIC
Due date	M42 – February 2022

Keywords	Data fusion, sensor network planning
Abstract	This document details the achievements of WP3, viz. three computational methods for fusing sensor data, and provides an overview of the integration of the data fusion framework with RESI Monitoring Center. The purpose of this document is to provide an overview of the achievements of WP3 to the public. A specification of the data fusion algorithms was presented in detail in Deliverable 3.1 and its evaluation presented in Deliverable 3.2. Readers already familiar with the content of D3.1 and D3.2 may omit this deliverable.

Editor	Fernando Solano (WUT/BTEC)
Contributors	Robert Nowak (WUT), Krystian Chachuła (WUT), Tomasz Słojewski (WUT), Artur Tomaszewski (WUT), Fernando Solano (WUT/BTEC)
Reviewers	Susanna Fontana (FORMIT)
Submission date of the draft to reviewers	24/02/2022
Submission date of the draft to the SAB (if required)	Not required

Register of document versions

Partner acronym	Version number	Date	Suggested relevant changes	Notes
WUT/BTEC	V1.0	24/02/2022	//	Preliminary version
FORMIT	V2.1	26/02/2022	//	Minor remarks and adjustments
WUT/BTEC	V2.2	28/02/2022	//	Second draft
FORMIT	V3.0	4/03/2022	//	Final version

Every information is updated to the date of issue of this document

This document is composed by 11 pages

Table of Contents

List of figures

List of acronyms and abbreviations

API	Application programming interface
CA	Consortium Agreement
DAG	Directed Acyclic Graph
DoA	Description of Action
EB	Executive Board
EC	European Commission *or* Electrical Conductivity
ES	Exploitation Strategy
GA	Grant Agreement
LEA	Law Enforcing Agency
REST	REpresentational State Transfer
SME	Small and Medium Enterprise

EXECUTIVE SUMMARY

This document details the achievements of WP3 and provides an overview of the integration of the data fusion framework with RESI Monitoring Center. A specification of the data fusion algorithms was presented in detail in Deliverable 3.1 and its evaluation presented in Deliverable 3.2.

Data fusion algorithm processes measurements from sensors and calculates the most important and the most probable events (in the context of the project i.e. sources of pollution). The evaluation includes several simulated scenarios on network topologies typically used in this field as well as evaluation on real sewage network and measurements obtained from experiments.

1. MAIN ELEMENTS OF THIS DELIVERABLE

1.1 INPUT FROM OTHER PROJECTS

None.

1.2 INPUT FROM OTHER WPS AND RELATION WITH OTHER SYSTEM DELIVERABLES

This report is related to the following two deliverables:

- D3.1 SPECIFICATION OF THE METHODOLOGY USED FOR FUSING SIGNAL PATTERNS FROM THE THREE UTILITY NETWORKS which depicts the algorithm (in pseudocode).
- D3.2 EVALUATION OF THE DATA FUSION METHODOLOGY which depicts the results on simulated and real data.

1.3 APPLICABILITY

This report details the achievements of WP3.

1.4 REFERENCE DOCUMENTS

The following hierarchy will be applied for conflict resolution between the Project Operational and Management Plan (D12.1) and other documents, such as the Description of Actions (DoA) or the Grant Agreement:

2. Grant Agreement (GA);
3. Consortium Agreement (CA);
4. The Project Operational and Management Plan (D12.1).

The hierarchy related to the documents above implies that the latter document needs to be consistent with the former. In case of issues, this hierarchy of documents is mandatory.

1.5 PURPOSE OF THE DOCUMENT

The purpose of this document is to provide an **overview** of the achievements of WP3 to the *public*. Specification of the algorithms and their evaluation can be found in D3.1 and D3.2, respectively. Readers already familiar with the content of D3.1 and D3.2 may omit this deliverable.

1.6 STRUCTURE OF THE DOCUMENT

The document is structured in the following way:

- Chapter 1 highlights the main elements of this deliverable and its rationale within the WP3 and within the SYSTEM project
- Chapter 2 summarises the main information on data fusion algorithm and evaluation metrics
- Chapter 3 describes the data fusion module integration with System
- Chapter 4 describes the iterative network planning and localisation, another methodology to localise the source of pollution
- Chapter 5 outlines the machine learning method for estimating of the localisation of illegal discharges
- A final chapter concludes and focuses on the next steps.

2. Discussion on data fusion algorithm and evaluation metrics

Data fusion algorithm, presented on Figure 1, depicted in Deliverable D3.1, was tested on 5 network topologies, example is presented on Figure 2 (for simulated data) and on the Bauerbach sewage network (for real data). Testing includes quality measurements as well as performance evaluation.

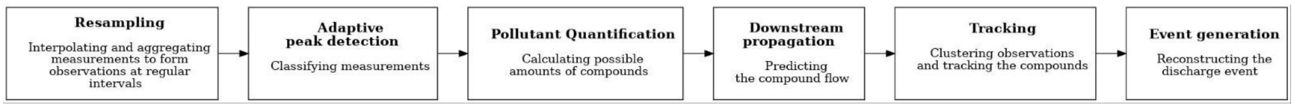

Figure 1 Data Fusion algorithm

The figure shows (1) the resampling (for properly handle asynchronous measurements), (2) the adaptive peak detection (to properly handle background level flow and noise), (3) the pollution quantification (to handle different substances), (4) the downstream propagation (to work with networks with sparse sensor coverage), (5) the tracking (using Kalman filter), and (6) event generation (to reduce unimportant alarms).

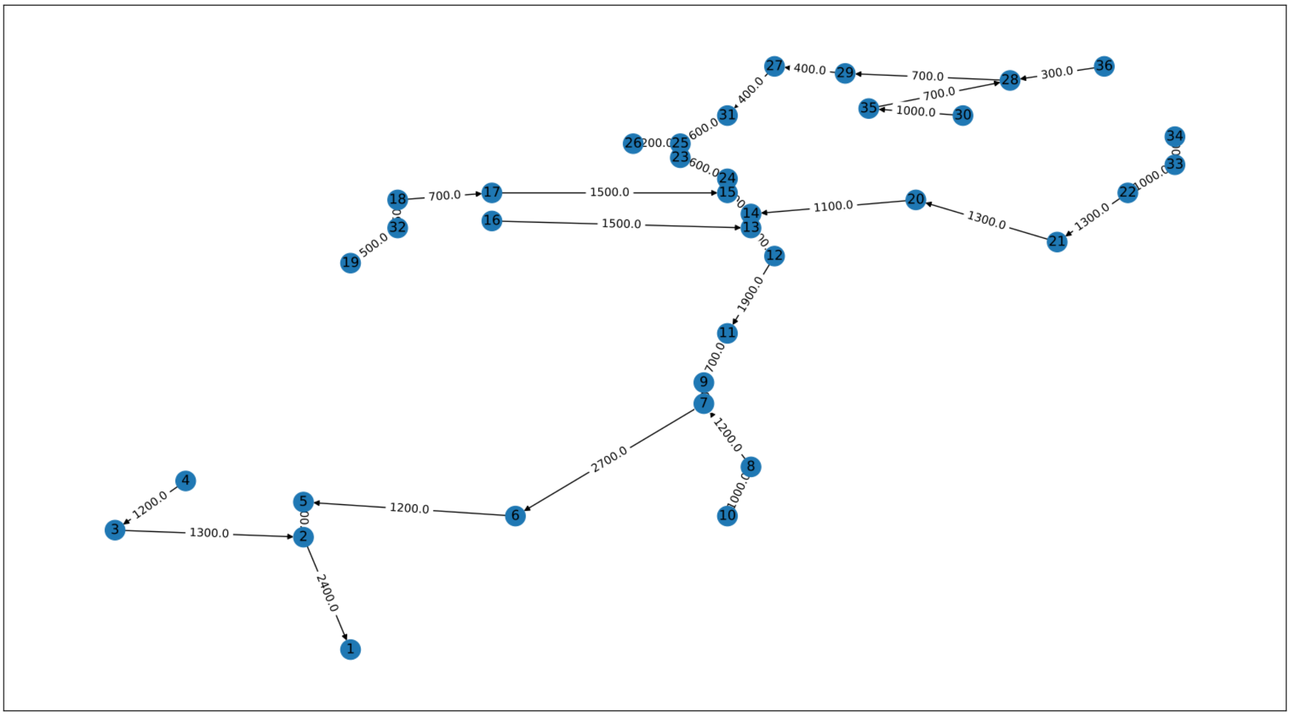

Figure 2 Simulated network topology used for testing (example)

The developed software allows to propose the sensors localisation strategy to achieve high sensitivity with a small number of sensors. Details are depicted in D3.2.

We developed our algorithm to detect pollution in the sewage network, and according to domain experts, the aim is to process the changes of measured values with high frequency. Our system calculates the sewage network state very fast. Therefore, the resampling error is negligible. We assume that the period will be small enough to introduce a minimal error in real scenarios. Moreover, we used simulated data, and we assumed that the generated measurements are periodic, and the period is equal to the sampling period.

All data included in the measurement domain are provided by sensors. We do not use uncertainty in the current version of the algorithm. However, such information can be used in future work to calculate the detection confidence coefficient that we already used in the event generation step of the algorithm.

In our approach, we do not model pollutant spreading. We require a short list of all possible pollutants. In the quantification process, each pollutant is considered for every measurement. Finally, the best pollutant is reported. In recent years, especially for air pollutants, hierarchical methods that model spatio-temporal processes and measurement noise were popular (Blangiardo, Li). We plan to apply such models in the future to make our results more accurate.

Further research should focus on evaluating the data fusion algorithm in real-world scenarios, which means running fusion using data from real sensors in live wastewater networks. This will require defining a procedure for choosing the best values of the parameters of the algorithm. Parallel processing techniques could be applied to the proposed solution in case current performance is not enough in large networks with many sensors. Investigation of the influence of sensor placement and discharge amount on the uncertainty of reconstructed events could improve the calculation of event confidence coefficients. Similarly, the uncertainty of measurements, as reported by sensors, could be considered when calculating these coefficients.

The proposed improvements to the data fusion algorithm significantly broadened its applicability. The new, adaptive peak detection algorithm can adapt to changing background values (D3.2). It is a milestone for testing this system in real-world networks. Implementing tracking in DAGs poses a significant improvement from previous achievements as real-world networks no longer have to be modified to perform fusion as shown in D3.2. Results in D3.2 suggest that accounting for flow rate by introducing the attenuation coefficient helps when detected pollutant amounts are consistently low because of a high flow rate. Unlike the previous iteration, these improvements make the proposed multisensor data fusion algorithm applicable in real-world scenarios.

Moreover, the development mentioned above did not negatively influence the correctness of the results. This was verified by measuring success rate the same way as in our previous study, using the same network, this time unmodified, in D3.2. In order to evaluate the algorithm in more depth, additional metrics were calculated.

The applicability was not the only aspect of the algorithm that was improved. Simplifying the implementation caused it to be able to process measurements 50 times faster than the previous one as presented in D3.2. The measured processing speed and memory requirements show that such a system could run on a regular computer and monitor the network in real-time.

3. Data fusion module integration with SYSTEM

Data fusion module is part of SYSTEM. Data fusion module provides API to (a) start data fusion calculations for given scenario (given network topology, given set of sensors, etc.), (b) stop data fusion calculations for given scenario, and (c) to return the current state for given scenario and given time. The protocol is based on HTTP and JSON.
Data fusion can communicate with the SYSTEM database to get the data from sensors. It uses a REST API. The deployment diagram is presented in Figure 3.

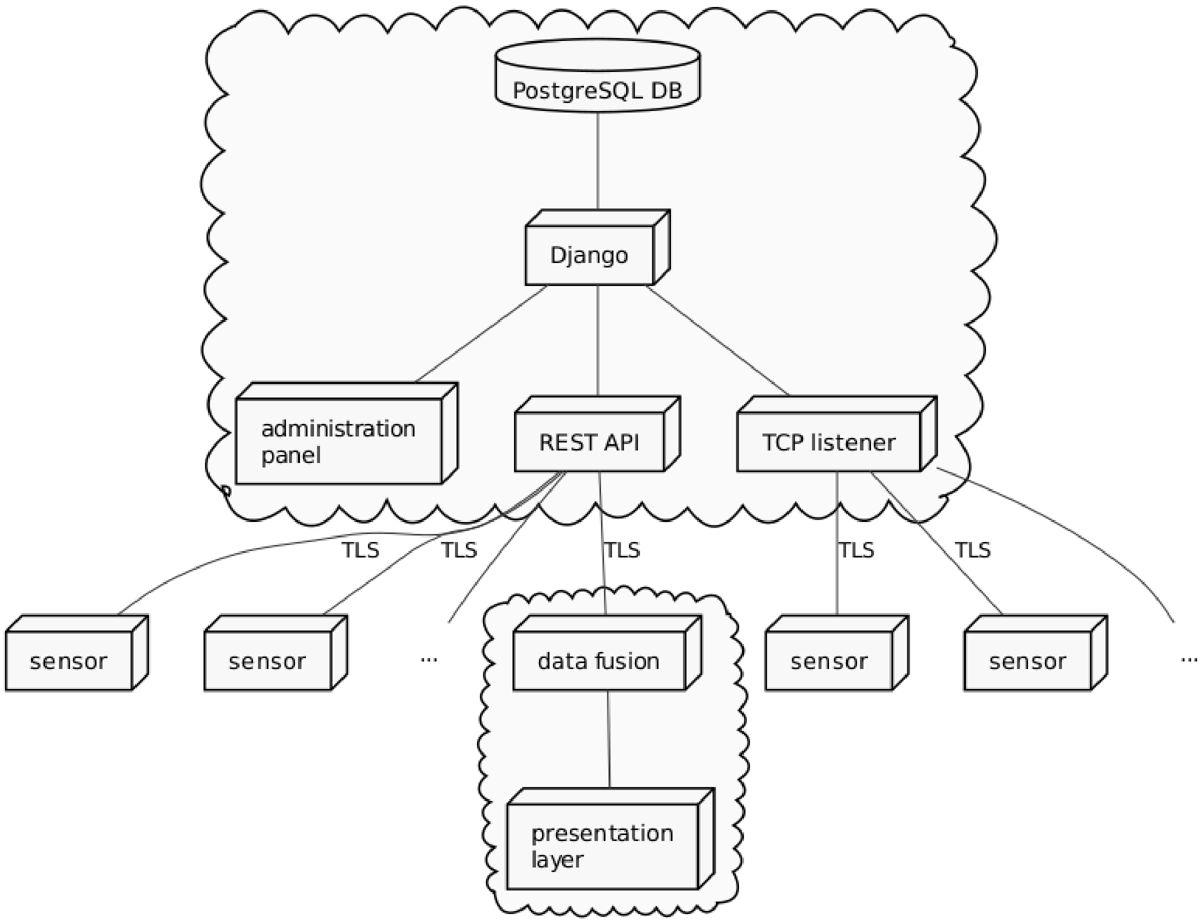

Figure 3 Data fusion module as part of SYSTEM

4. Iterative Network Planning and Localization

In WP3, a second computational method was developed for estimating the localisation of a polluting source. This method works on the assumption that a single polluting source continuously discharges illegal waste into the sewage network.

The method focuses on the planning of the localisation of sensor devices in the sewage network for the purpose of localising the source of pollution. The solution scheme we study considers that a limited number of sensor devices is available, and that the sensor devices can be ***re-arranged*** if a sensor device detects an illegal discharge. Therefore, the solution scheme we study consists of a decision tree where at each step a layout of a limited number of sensors is provided, considering the outcome of previously taken decisions.

The proposed methodology considers that the maximum number of sensor devices at any step is limited, and we aim at minimising the number of steps needed for any potential source of pollution that may occur in the network.

WUT worked on two implementations of this method. The first one is based on a mixed-integer programming model which guarantees finding the optimal solution to the optimization problem. The second is a branch-and-cut heuristic for larger networks.

These methods provide a plan for sensor deployment. Unfortunately, the RESI Monitoring Centre functionality does not include a network planning functionality. Therefore, these methods were not integrated with the RESI Monitoring Centre.

5. Machine Learning for Estimating of the Localisation of Illegal Discharges

In WP3, a third method was developed for estimating the localisation of a polluting source. The third method makes use of the capability of simulating wastewater flow and water quality parameters offline, as performed by UniBWM and IH-SAS. Our method makes use of such simulations for training a machine learning algorithm.

A full description of this method can be found in the published article: **Identifying and Estimating the Location of Sources of Industrial Pollution in the Sewage Network**, authored by Buras Magdalena Paulina, Solano Donado Fernando, in *Sensors*, 2021, vol. 21, no. 3426, pp.1-19, DOI:10.3390/s21103426.

The third method works with online data from static sensors, similarly to the algorithm described in Section 2. Nevertheless, it does require a large number of simulations for high precision. As a result, the consortium focused on the integration of the algorithm presented in Section 2.

6. Conclusions

WP3 presented three different data fusion methods, each of which was designed with different assumptions in mind. All methods focused on fusing data from different sensors located in the sewage network. All methods were evaluated using simulated data resulting from hydraulic models from WP4.

After internal consortium discussions, one of the methods was chosen as to be integrated with RESI Monitoring Centre, which is the one described in section 2 and 3. The integration with RESI Monitoring Centre was successful and tested in laboratory conditions. Unfortunately, it was not possible due to time limitations to carry out tests using online real sensors.

7. Bibliography

- Nowak, R., Misiurewicz, J. and Biedrzycki, R., 2011, July. Automatic adaptation in classification algorithms fusing data from heterogeneous sensors. In 14th International Conference on Information Fusion, pp. 1-7. IEEE.
- Blangiardo, M., Pirani, M., Kanapka, L., Hansell, A. and Fuller, G., 2019. A hierarchical modelling approach to assess multi pollutant effects in time-series studies. PloS one, 14(3), p.e0212565.
- Li, L., Wu, J., Ghosh, J.K. and Ritz, B., 2013. Estimating spatiotemporal variability of ambient air pollutant concentrations with a hierarchical model. *Atmospheric Environment*, *71*, pp.54-63.
- Krystian Chachuła, Tomasz Słojewski, Robert Nowak. 2022. Multisensor data fusion for localization of pollution sources in wastewater network. Sensors, 22(387), doi:10.3390/s22010387.
- Krystian Chachuła, Robert Nowak, Fernando Solano Donado. 2021. Pollution source localization in wastewater networks. Sensors, 21(826), doi:10.3390/s21030826.