Article

Identifying Data Dependencies as First Step to Obtain a Proactive Historian: Test Scenario in the Water Industry 4.0

Andrei Nicolae[ID]**, Adrian Korodi ***[ID]** and Ioan Silea**

Department of Automation and Applied Informatics, University Politehnica Timisoara, Piața Victoriei 2, 300006 Timișoara, Romania; andy_nicolae@yahoo.com (A.N.); ioan.silea@upt.ro (I.S.)
* Correspondence: adrian.korodi@upt.ro

check for updates

Abstract: Current efforts towards achieving better connectivity and increasing intelligence in functioning of industrial processes are guided by the Industrial Internet of Things paradigm and implicitly stimulate occurrence of data accumulation. In recent years, several researchers and industrial products have presented Historian application solutions for data accumulation. The large amounts of data that are gathered by these Historians remains mostly unused or used only for reporting purposes. So far, Historians have been focused on connectivity, data manipulation possibilities, and sometimes on low-cost solutions in order to gain higher applicability or to integrate multiple SCADA servers (e.g. Siemens–WinCC, Schneider Electric – Vijeo Citect, IGSS, Wonderware, InduSoft Web Studio, Inductive Automation – Ignition, etc.), etc. Both literature and industry are currently unable to identify a Historian solution that functions in fog and efficiently applies and is built upon Industry 4.0 ideas. The future is to conceive a proactive Historian that is able to, besides gathering data, identify dependencies and patterns for particular processes and elaborate strategies to increase performance in order to provide feedback through corrective action on the functional system. Using available solutions, determining patterns by the Historian operator in the context of big data is a tremendous effort. The motivation of this research is provided by the currently unoptimized and partly inefficient systems in the water industry that can benefit from cost reduction and quality indicator improvements through IIoT concepts related to data processing and process adjustments. As the first part of more complex research to obtain a proactive Historian, the current paper wishes to propose a reference architecture and to address the issue of data dependency analyses as part of pattern identification structures. The conceptual approach targets a highly customizable solution considering the variety of industrial processes, but it also underlines basic software modules as generally applicable for the same reason. To prove the efficiency of the obtained solution in the context of real industrial processes, and their corresponding monitoring and control solutions, the paper presents a test scenario in the water industry.

Keywords: industrial internet of things; Historian; industrial automation; industry 4.0; water industry; fog computing; big data

1. Introduction

1.1. Industrial Internet of Things (IIoT)

The Industrial Internet of Things (IIoT) [1] defines the same principles as the Internet of Things (IoT) concept but applies them to industry instead of the consumer, general-purpose approach of IoT. IIoT's main goals are to generate interoperability of physical objects and to improve system functioning by elaborating optimizing strategies through interoperation between different industrial

entities, data analyses, and learning. Improved interfacing and communication brought by IIoT to the industry enables interoperation and may increase the intelligence of systems, thus allowing to significantly improve technological processes. Water industry-specific structures and functioning provide the perfect environment for improvements in efficiency, quality, and availability using Industry 4.0/IIoT principles. The water industry is represented by highly heterogeneous and geographically dispersed processes and technical solutions. These include legacy systems and new structures that are in stringent need of connecting the digital and the physical worlds in the context of highly functional process dependencies without interoperation. The current transition towards the IIoT paradigm is stimulated by the benefits that lie ahead such as cost reduction and increases in safety, productivity, and availability. This transition is also revealing a series of problems for the water industry. Specific to drinking water facilities, these issues include: water source quality changes, high energy and substance consumptions in the treatment process, and maintenance. Under the IIoT paradigm, the fog computing concept is emerging and becoming more significant in industry. This new term defines solutions that are placed closer to local automation and, therefore, are much more accessible and reliable.

IIoT is currently one of the most important research and development topics; it manages to draw significant attention from both academia and industry [2]. This new paradigm is steering the industry towards more intelligent communication between different industrial entities by connecting computers, controllers, actuators, and sensors to the Internet [3]. This better-connected industrial environment allows for superior information exchange between all the components involved [4]. Consequently, the IIoT paradigm is facilitating the development of more sophisticated technical solutions and autonomous software algorithms [5–7] that can improve the working characteristics (energy consumption, time efficiency) for industrial processes (see [8–11]).

Identifying data dependencies is a process of analyzing stored data and discovering relations and dependencies between the characteristics/tags that have their values stored. This process is vital for developing a proactive Historian application because it is necessary to understand the correct ways to react and adjust the system. In order to adjust technical system working parameters, an understanding of how the potential adjustment impacts the entire system (the rest of the working parameters) is required. For example, if working parameter A is adjusted by the proactive Historian, the application must know if parameter B is related or dependent on parameter A. Lack of information regarding data dependencies can provide the possibility that the proactive Historian applies adjustments that make the technical system unstable. So, data dependency identification is crucial for the rest of the processes that follow inside a proactive Historian application.

Because of the heterogeneity of devices that are starting to be connected under the IIoT paradigm [1], there are many communication protocols used in the industry; research from [12] provides insight on this problem. However, in recent years, Open Protocol Communication Unified Architecture (OPC UA) has started to become the standard IIoT protocol (see [13–16]). The popularity of OPC UA is also sustained by the large number of available software development kits, which makes OPC UA-based development easy (see [17–19]).

1.2. Interoperability and Historian

The water industry contains a very large variety of systems/solutions. These solutions are also highly dispersed chronologically and by location. The authors presented in [20,21] solutions for OPC UA wrapping with a high technology readiness level (TRL) applied at water distribution companies. The solutions led to interoperability.

The superior interoperability provided by the OPC UA in the IIoT context enables horizontal interoperation between systems placed at the same hierarchical level (see [22,23]). The benefits of interoperation are proved for the water industry in study [24] on a wastewater network, which started from a cascaded wastewater pumping station (WWPS) towards the wastewater treatment plant (WWTP). In a fog computing scenario, by using a noninvasive control algorithm and interoperation, the solution from [24] optimizes clogged pipes failures, WWPS blockages, and supplementary stormwater

at the inlet of the WWTP. By using interoperability, a noninvasive approach over existing functioning structures, and data accumulation, the study from [25] also presents a solution to reduce energy consumption in a WWTP.

Emergence of data accumulation leads to a different view from the IIoT perspective. Currently, data gathering in industry is usually implemented with Historian applications placed at the supervisory control and data acquisition (SCADA) level. The need of Historian applications in the water industry is emphasized by the research from [26], which also proposes standardized directions for different types of water industry-specific objectives.

Currently, most Historian applications available for industry are still offered by well-known automation/SCADA software producers and are very expensive. Therefore, they are placed only at the top supervisory levels. However, recent research has proposed different approaches for Historians.

In [27], the authors proposed an improved Historian structure that could handle large amounts of data. However, they used International Electrotechnical Commission 61850 protocol, which is an electricity domain-specific protocol, thus making the Historian not suitable for other industries.

A more general approach is presented in [28], where the authors proposed a low-cost and lightweight Historian based on OPC UA, which made it potentially available for a wide range of industries. The solution embedded Node-RED into a Java application and stored the collected data inside an SQLite database. At the same time, it was a platform-independent and complete hardware/software solution. The proposed Historian application was successfully applied in the water industry.

In [29], the authors proposed a distributed Historian framework, which allowed configuration of a Historian application by using an organizational model of a hierarchical system.

On a slightly different note, research from [30] presented an efficient data archiving method designed specifically for storing historical sensor data.

Because of progress recently made in data accumulation, new opportunities are arising regarding usage of the collected data [31,32]. This data can be used as input for software algorithms that can run autonomously and eventually optimize the technical systems from where the data was collected. This kind of software algorithm can bring great benefit to the industry by reducing costs and improving efficiency of various technical systems. There are numerous possible development directions in the stored data analysis area, but few researchers are currently integrating the IIoT context.

1.3. Towards a Proactive Historian Application in the Water Industry

Practical implementations resulting from this research paper were considered and deployed in the water industry. The water industry currently needs improvement in system functioning that cannot be obtained by most manual analyses of available data in the context of currently deployed data gathering solutions and structuring because of several issues, which are briefly detailed as follows: there is a large geographical spread of systems in this industry; Historian applications are currently available only at the top supervisory level because of the high costs that a classic Historian solution implies; collected data are filtered according to the hierarchical level of vertical integration and the implicit local process understanding level (a data processing operation closer to the technical systems (fog computing) would enable cost reduction and efficiency improvements); there is lack of significant pattern identification capabilities that are adaptable to the highly heterogenous water industry processes; there is lack of a process-aware Historian; and there is lack of proactive solutions that identify an applicable recipe and react over the local process in the form of corrective actions.

The proposed solution is conceived according to Industry 4.0 principles by offering superior connectivity and flexibility; therefore, it is applicable to the wide variety of local systems in the water industry. This is achieved by using the Node-RED platform [33], which offers various nodes that enable interfacing local industrial systems (e.g., OPC UA, Modbus (TCP and serial), S7, TCP/IP Ethernet, etc.) and offer the possibility to add other interfacing nodes in case of a nonpopular protocol. The proposed solution can connect (gather data and also to react) to local and regional SCADA systems, and also

to programmable logic controllers (PLCs)/human machine interfaces (HMIs)/gateways, offering a low-cost alternative to the current Historian applications available.

The water industry requires data accumulation and optimization techniques based on knowledge from gathered data to increase functioning efficiency. Studies need to be industry focused and applicable because various research findings have been purely theoretical without a chance to connect or to apply them on real water facilities. Drinking water treatment and distribution are critical processes with many parameters, and improvements in functioning are necessary. The treatment process is intensively studied, and in [34] the authors proved that the treatment process itself determined the impact that climate changes had over the drinking water systems. Water quality indicators are also of high importance, and various methods are studied to implement their increase. Efficiency of the entire process relies also on the cost aspects. The cost issue is intensively analyzed and related to certain parameters (e.g., consumption of substances, energy, and maintenance costs). In [35], the authors proved the efficiency of substance usage over water quality indicators. Some studies (see [36]) present a fuzzy solution to determine water quality, and it would be interesting to observe if the method will be adopted by water distribution companies. The cost issue is studied in [37] but without a specific concern over the water sources and the treatment process. Some studies focused on the cost issue considering automation techniques. For example, proof of reducing costs was provided in [38] when pumps were used with frequency converters. This also impacted water sources and proved that the optimal solution for water wells was for them to be equipped with a pump that had frequency converters, which, therefore, contained local flow and leveled referenced closed-loop control algorithms. A very complex study regarding cost perception was provided [39]. The cost was presented from various perspectives including the impact of parameters and equipment evaluation on the cost (e.g., energy consumption of equipment, equipment faults, maintenance costs, etc.), optimizing techniques in the water distribution network, etc. But, the study from [39] was focused on the distribution network, and it did not deal with the water treatment process. In the same context, authors in [40] presented a strategy to optimize reservoir functioning to minimize economic losses caused by pollution, and authors in [41] detailed an energy consumption reduction conceptual solution and its impact on swimming pool water distribution. Another important issue was presented in [42] regarding the influence of raw water over the treatment process. The study from [43] showed water quality degradation for drinking water sources in a complex and long-term study (441 water supply systems using 18 years of data). The outcome of [43] was relevant to the current paper's research perspective: it was necessary for water quality of each water source to be determined considering that, in practice, there are no quality sensors on a water well, and all information must be derived from complex monitoring, analyzing, and learning procedures. The importance of data gathering, analysis, and learning is presented in several studies. In [44], turbidity levels of the water sources were predicted using data mining techniques, and an early warning system was implemented. In [45], a burst detection in metering areas of water distribution was presented based on functional patterns of water demand with supervised learning. The same type of research was developed in [46], where based on SCADA, bursts were detected over a long-distance network. As data are highly important in quantitative research, in [47], the authors presented a solution to impute missing data for water distribution systems.

In the context of water resource management, various issues are considered and analyzed. Using hydraulic modelling, research from [48] studied hydraulic regimes using different datasets associated to flood and low-flow events. The study from [49] presented a historical review of the evolution and problems of water sources and distribution. In [50], authors presented the importance of meteorological data integration in the water domain. In [51], the seepage effect on river dikes was investigated, which could affect the quality and availability of surface of water sources. Through IIoT, the impacts of the encountered problems may be reduced, or even eliminated, while the functioning of respective water treatment and distribution processes may be improved.

Following the above-mentioned status regarding IIoT, Historians, and the water industry, the next question arises: how do we efficiently use the accumulated amount of information to obtain maximum benefits for the water industry?

Data accumulation refers to Historians in automation and SCADA domains. An industry-oriented analysis, which considered both literature study and the authors' many years of experience in the water industry, concluded that classic Historian solutions from automation-/SCADA-producing companies were: very expensive and therefore placed only at the top supervisory level (less than 5% of the encountered SCADA control rooms had separate Historians); used only for data accumulation and reporting purposes; provided only manual (Excel-type) data manipulation possibilities; were mostly unusable by operators because of process and application understanding issues; and were, many times, platform-dependent. No Historian was encountered in the water industry that used accumulated data to automatically identify data dependencies, that elaborated on optimizing strategy-related conclusions, nor that reacted over local process controls to increase its efficiency in any way. The currently implemented Historians in the water industry were not aware of interfaced process characteristics; therefore, they cannot have any process-related objective [52,53].

Industry requirements from Industry 4.0, and implicitly from IIoT and industrial automation, include advantages such as cost reduction, improved safety, wider availability, and an increase in productivity. Interoperability is essential and it is dependent on equipment and automation/SCADA solutions. Further research towards interoperation, data analysis and pattern identification, objective function definition, and model-based analysis is process-dependent. Water distribution companies are concerned about cost reduction (energy efficiency, substance consumption, and good maintenance strategy) but also about water quality. The water industry includes various geographically and chronologically dispersed local processes and implementations. Consequently, the best solutions should be fog-based so they can be close to local automation, noninvasive over existing control solutions, and adaptable to the individual processes and control procedures in order to identify dependencies by analyzing and understanding gathered data and to react through algorithms that augment existing structures. No such solution was encountered in practice nor in the literature.

Therefore, the general objective of the initiated research is to answer the above-mentioned question by providing an approach that will increase industrial efficiency through data accumulation and analysis. The current paper proposes a reference architecture for a proactive Historian that consists of a multilevel algorithm hierarchy. The proposed architecture facilitates creation of an autonomous, proactive Historian able to optimize and influence a functional system without human assistance. Extending the research from [28], where a basic low-cost Historian was developed, the current research presents a stored data analysis algorithm that identifies data dependencies between measured characteristic and reference characteristic, establishes degrees of dependency, and exposes functional patterns. The proposed solution is aware of the process, so that computing or process-related degrees of freedom are considered (e.g., parameter and functional limitations, output possibilities). The associated process is from the water industry, a domain where physical and digital entities are currently trying to find common ground in the context of Industry 4.0 (e.g., SC5-11-2018 Horizon 2020 European Commission research project call from 2018: "Digital solutions for water: linking the physical and digital world for water solutions"). Considering drinking water treatment and distribution processes, the specific objectives of this research are to provide the reference architecture for the Historian, the data dependency identification algorithm, degrees of dependency between characteristics, process awareness, interoperability and interoperation possibilities, an integrated solution that is applied and tested on a real system, and a step-ahead view in increasing the energy efficiency by improved water source manipulation.

The conceived research is applicable to any industry as long as the reference tags, the limitations, and the objectives are defined at the beginning so that the Historian will understand the purpose and the degrees of freedom of the analysis.

The following section presents processes that take place inside a typical drinking water treatment plant (DWTP), presents the proposed proactive Historian reference architecture, and provides an algorithm description. Section 3 details integration of the algorithm into the Historian application developed in [28], illustrates test scenarios from the water industry, and provides insight into improving a DWTP. Section 4 discusses the results and findings, while Section 5 concludes this paper.

2. Materials and Methods

2.1. Drinking Water Treatment Plant (DWTP) Typical Processes

Figure 1 presents processes that take place inside a typical drinking water treatment plant (DWTP). The inlet of a DWTP comes from water sources. Water sources are usually represented by underground water wells or surface water sources. In case of very dry climate, the water may come even after a complete recirculation of the water from wastewater facilities. The DWTP usually takes over the water from the sources and initiates the treatment. The phases of water treatment are aeration, filtering with sand and charcoal, disinfection (chlorine), and sludge treatment. Aeration is realized with blowers that maintain an oxygen level inside the tank. The first pumping station moves the water from the aeration tank to the sand filters, to reduce turbidity, and then to the charcoal filters. Aeration and charcoal filtering are essential to obtain the desired pH and conductivity levels of the treated water. The filters are cleaned frequently using air (additional blowers) or water. Cleaning with water is usually realized using other pumps, but in some situations the first pumping station may function in two functioning regimes (filter regime and washing regime). Chlorination is realized through a chlorine station. Chlorine levels are measured at different points in the DWTP, and corresponding dosing rates are associated inside closed-loop control procedures. Usually, an elementary water flow-based chlorine control strategy is added with a higher-level, closed-loop chlorine control loop that has residual chlorine on feedback. Continuous water flow is necessary at the DWTP inlet because large periods are necessary to obtain an efficient chlorine control (e.g., 30 min). Output of the DWTP is sent to the water distribution network using a second pumping station and eventually a water tower. Also, the sludge treatment phase consists of various energy consumer objects.
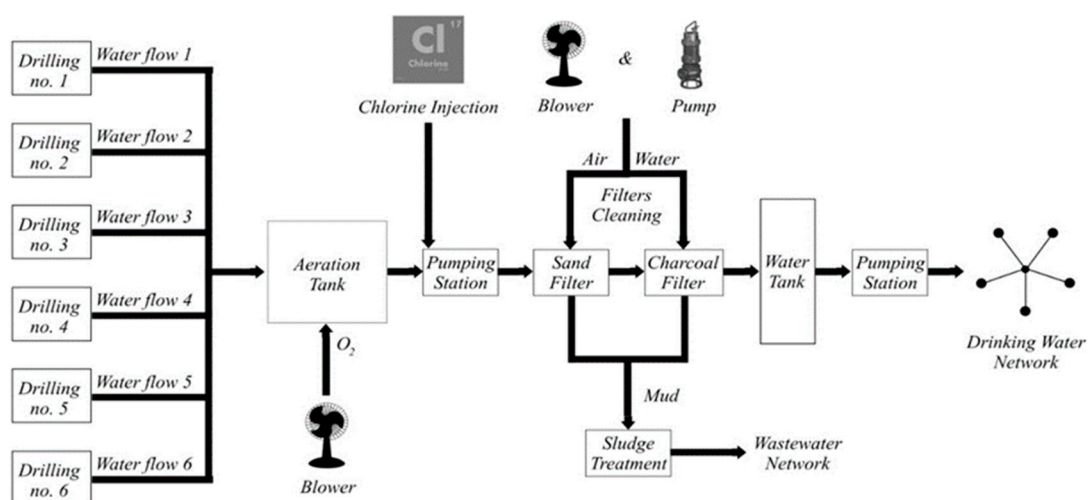

Figure 1. The processes inside a drinking water treatment plant.

Considering the mentioned steps in water treatment and practical DWTP monitoring using the developed Historian, the following dependencies and issues are identified:

- Growth of water acidity/alkalinity (pH) and lowering conductivity levels are implicitly assured through energy and chlorine consumption (e.g., aeration blowers, chlorine station, and maintenance of charcoal filters).

- A high turbidity level implies possible clogging, respectively high energy consumption and water losses, which result from cleaning the filters with air and water.
- A level control loop in the water distribution tank complemented with a flow-based control loop, which calls water sources according to the actual flow demand from the water distribution network, assures a very useful, anticipative character of the entire water distribution when high water consumption is identified in critical periods of a day. At night, however, water sources are usually stopped when the upper hysteresis limit is achieved in the water tank. If water losses are present in the water distribution network, then the flow control loop will activate water sources. If the flow setpoints in the flow-based closed-loop control algorithms at the water sources are too high, or are set at fixed values, the water source pumps may start and stop multiple times during the night, causing pump and water source wear out.
- Besides the previously mentioned problem, successive starting and stopping of the water sources causes activation/deactivation of the entire DWTP for short periods. This leads to perturbed filtering and chlorination processes.
- Water sources have different characteristics; therefore, some of them may provide higher flow values and some better water quality. Monitoring residual chlorine, blower functioning times, and filter washing cycles over longer periods of time, together with chosen water sources that are currently functioning (provided water flow values), water source quality indicators can be identified. Using suitable water sources, specific consumptions can be reduced (flow distribution).
- Water source quality indicators change over time.
- The water level in the water distribution tank cannot be kept inside two hysteresis limits because water consumption variation in the distribution network perturbs the level control algorithm. Consequently, inconsistencies of water reserves in the tank may be identified including higher energy consumption and possible water treatment process disturbances.
- Proper equipment functioning hours and number of starts is essential to consider because maintenance/replacement is expensive.

Many of the previously mentioned ideas are usually not solved nor identified in classical developments. The most proper and efficient recipe cannot be identified with classical structures. Implementation of automation/SCADA solutions for the physical dependencies in Figure 1 is usually made in phases. Entrepreneurs commission systems using their best knowledge at the time, following pending documentation without long-term testing and without any optimizing solutions.

Considering the above-mentioned ideas, monitoring and storing process values is essential, but more must be developed in order to increase efficiency. The best recipe has to consider the available variables (usually thousands of tags) and their dependencies following a cost objective. The cost objective is influenced by many factors (e.g., reducing functioning hours of one equipment decreases maintenance costs, but it may negatively influence quality indicators). Most times the recipe cannot be identified by operators or engineers, which is an essential scope of IIoT and Industry 4.0.

The other important issue relates to interfacing, communication, and the ability to implement noninvasive changes over process automation. Practically, the ability of a system to react after connecting, collecting, and processing procedures makes it a proactive system that is able to apply the identified best recipe.

2.2. The Reference Architecture

In order to develop a proactive Historian, which can collect data from a technical system and then autonomously analyze the stored data and influence the technical system in order to meet clearly defined objectives with no human assistance, a reference architecture regarding the software algorithms is imperatively needed. A reference architecture defines the software modules involved, and it delimits the used algorithms into distinct categories based on their main functionalities. The finalized proactive Historian will be a complex software application consisting of numerous independent modules with

multiple interactions. Also, in order to compute the exact adjustment that needs to be applied to the technical system in order to meet the defined objective, multiple software algorithms must be involved. Each software algorithm will have different functionalities, characteristics, inputs, and outputs, and only the collective effort of all those algorithms can produce the desired end result (correct adjustment to the technical system). In this context, lack of a reference architecture leads to chaotic development with a lack of overview, which makes both software development and software maintenance much more difficult. Although [54] proposed a system architecture suitable for Industry 4.0, the approach was too general, and a more specific architecture must be defined for a proactive Historian application.

A simple Historian application can only be used to store different measured characteristic values (tags) of a technical system in a database, but it does not analyze data nor does it establish parameter behavioral patterns. As such, it does not optimize solutions nor does it influence in any way the technical system it monitors.

The architecture of a simple Historian application that only stores data from a technical system is already available [28], so this paper emphasizes the architecture of algorithms that optimizes the technical system. The proactive reference architecture proposed by this paper is detailed in Figure 2. The architecture is generic, so it does not limit itself to a specific industry or domain. The proposed architecture can be successfully applied in any technical domain.

The proposed architecture consists of a multilevel software algorithm structure that logically separates the involved algorithms into three distinct levels. Each of the three levels contain more than just one algorithm.

The main purposes of Level 1 algorithms are to identify relations and dependencies between technical system characteristics. This is important both for estimates of future evolution and for decisions regarding adjustments of the technical system. In order to predict future evolutions, Level 2 algorithms must understand what relations exist between different characteristics and how the characteristics influence each other. Level 1 algorithm outputs are a set of relations and dependencies between the measured characteristics. This output must also provide quantitative results, showing to what extent characteristics are related or influenced by other characteristics or context data.

Level 2 algorithms must predict the future evolution of measured characteristics. In order to make any adjustments to a technical system, it is required to know how the measured characteristic values evolve in the future. Level 2 algorithms must have two distinct inputs. First, relations and dependencies between characteristics identified by Level 1 algorithms are required because the technical systems are usually very complex with interconnected characteristics, so an isolation of the prediction to only one characteristic without analyzing the implications to the entire system (and on related characteristics) would lead to erroneous results. Secondly, in order to greatly improve the accuracy of the future estimated evolution, Level 2 algorithms must also receive context data as an input future. These data would also represent a prediction, so the accuracies of Level 2 algorithm outputs are influenced by the accuracy of this future context data. The future context data needed are closely related to the particularities of the industry in which the technical system runs. For example, water industry or agriculture are influenced by meteorological data. If the technical system is related to the water industry and it has a characteristic which represents the inlet of a wastewater treatment plant, then the meteorological rainfall forecast must be fed at the input of Level 2 algorithms. Level 2 algorithms can produce an erroneous wastewater inlet prediction if the rainfall forecast is not taken into account. Accuracy of the meteorological forecast influences the accuracy of Level 2 algorithm outputs. The outputs of Level 2 algorithms are the predicted future evolution of the measured characteristics of the technical system. For better accuracy, as many technical system characteristics as possible should be taken into consideration when starting the analysis to identify all possible implications/dependencies.
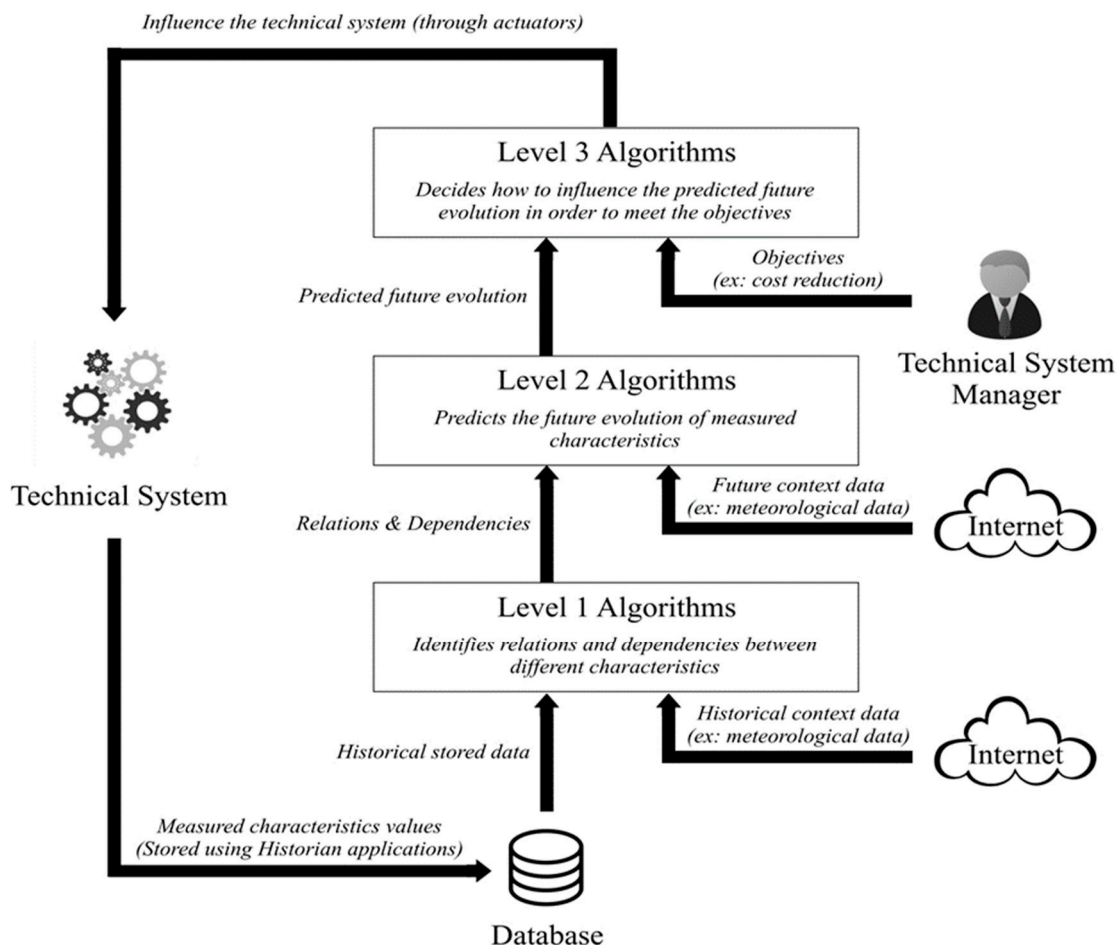
Figure 2. Reference architecture for a proactive Historian.

Level 3 algorithms are responsible for deciding how to influence the technical system in order to meet a defined objective. These algorithms must have two distinct inputs. The first input is the predicted future evolution of the technical system, computed by Level 2 algorithms. The second input is an objective (or even a set of objectives) that must be provided by a technical system manager. For example, the technical system manager can choose to reduce costs by reducing the value of a certain characteristic (e.g., in the water industry reduce the overall power consumption of water pumps) or to improve general cost efficiency. Based on those two inputs, Level 3 algorithms must compute how to adjust the system in order to achieve the desired objective. The outputs of Level 3 algorithms would be the exact influence that needs to be applied and can be applied to the technical system. The output of Level 3 must be directly applicable to the technical system; therefore, Level 3 algorithms must have a comprehensive understanding of the technical system and the tags that can be modified to augment the process setpoints without invasive/destructive interference.

The proposed proactive Historian architecture consists of a repetitive loop, whereby the evolution of the technical system is recorded, analyzed, predicted, and then technical system evolution is altered from the prediction in order to achieve predefined goals. The objectives can be changed along the way. From a different point of view, the algorithms that transform the simple Historian application into a proactive Historian application work in a pipeline architecture. Each algorithm level uses outputs of lower level algorithms as inputs.

2.3. The Implemented Solution—Algorithm Description

This section presents a Level 1 algorithm from the reference architecture detailed in the previous section. Being a Level 1 algorithm, its main goal is to identify data dependencies between different

characteristics of a technical system. In the implemented version, the algorithm only used historical, stored data as input without taking into account external context data, but this algorithm represented only a part of Level 1 from the reference architecture.

The proposed algorithm used a reference characteristic and, starting from the evolution of the reference measured values, it determined if the other characteristics were connected somehow to the reference. In case the algorithm determined that the measured values of a characteristic were related (regarding the evolution in time) to the measured values of the reference, it also computed the degree of impact regarding the dependency (the two characteristics could be very tightly related or could have a very low influence one over another).

The input for this algorithm was historically stored data that represented measured values of different characteristics (e.g., water pressures, water debits, water tank levels, energies, functioning hours, substance consumptions, etc.) sampled at different intervals (the sampling step should not necessarily be the same). In the analysis, the sampling step used was one day, so the input data were prepared before analysis by computing daily averages of the measured values. This allowed variation of the sampling period. Also, as input, the user indicated which characteristic would be used as reference.

Regarding the output, the algorithm returned two main pieces of information for each of the analyzed characteristics in relation to the reference: *Proportionality* and *Quantity.*

Proportionality information had three different possible values: directly proportional, inversely proportional, and not proportional. This information showed if the analyzed characteristic was proportional to the evolution of the reference.

Quantity information was provided by the algorithm in the form of a percent, which indicated to what extent the evolution of the measured characteristic values was affected by the evolution of the reference values. This quantitative information was relevant only when the first information (*Proportionality*) values were either directly proportional or inversely proportional. To substantiate *Quantity* information, some examples follow:

- *Quantity* information is 100%—indicates a 1:1 ratio between the analyzed characteristic and the reference, meaning that if the reference value changes by 20%, then the analyzed characteristic value also changes by 20%.
- *Quantity* information is 50%–150%—if the reference value changes by 20%, then the analyzed characteristic value changes by 10%–30%.

Some steps of the implemented algorithm are detailed in Figure 3.

To elaborate a conclusion, each step contained data structured in a one-day sampling period, and each analyzing step stored a one-day conclusion after data processing.

In the meantime, the Historian was connected with process variables to gather data. The current case scenario was built on the Industry 4.0 and IIoT main protocol: OPC UA. Therefore, the connection to each variable was realized in a publish–subscribe manner with a permanent link between the Historian tags and process tags. The Historian followed implicitly the PUSH principle for the variables. The PUSH principle means that, guided by the local automation sampling period, values were transmitted to the Historian only when they changed (e.g., when a blower started, the value was transmitted to the Historian). Thus, large numbers of data transfers and processing times were saved. For other interfacing options, the sampling times were adjustable.

Data were inserted for storage in a reduced manner on an hourly basis. The reduction strategy may be an average, but it also exceeded the related limit.

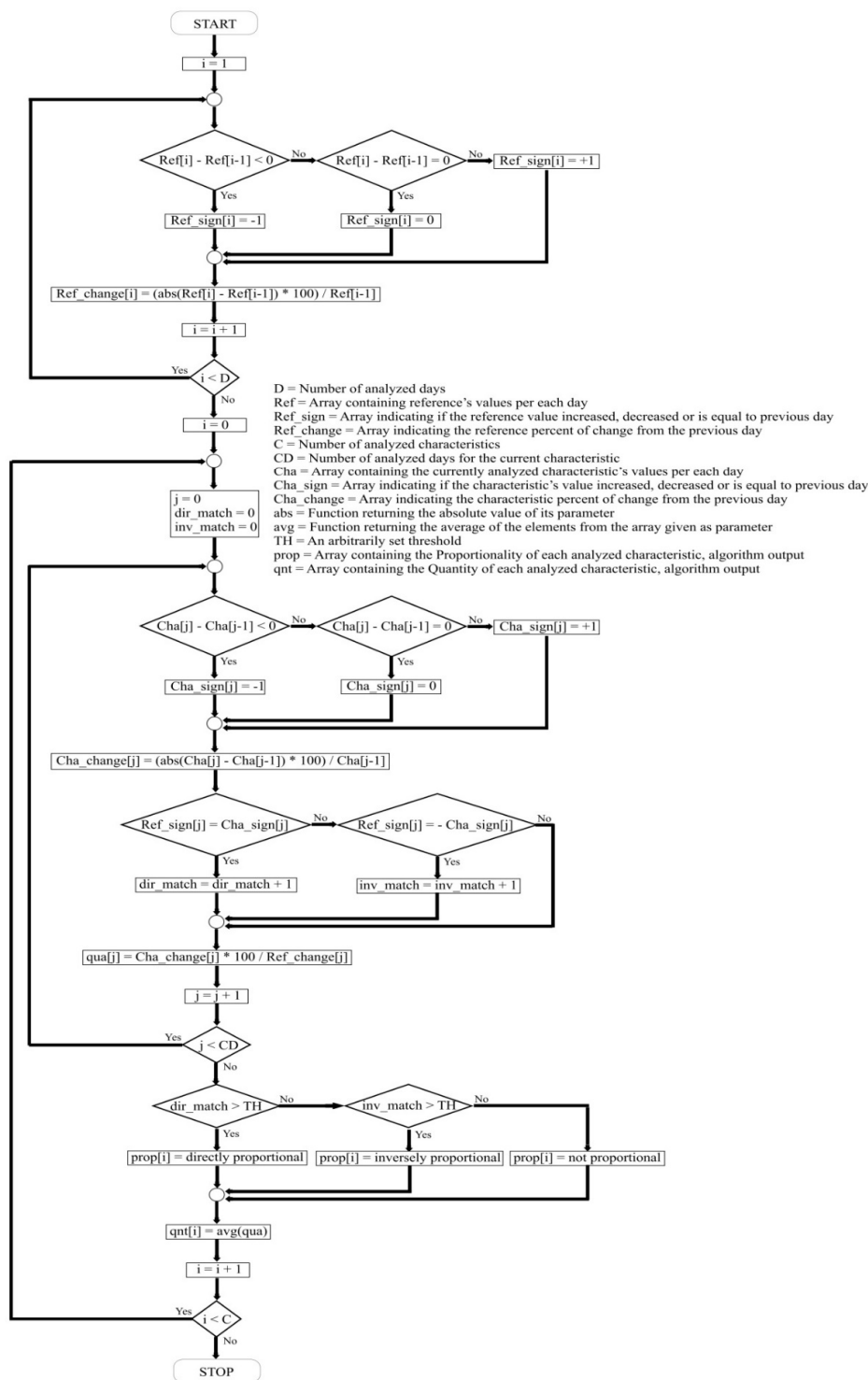The identified dependencies were stored using a one-day sampling period.

Figure 3. Steps of the data dependency identification algorithm.

3. Results

3.1. Water Industry Application

The developed algorithm was applied in the water industry integrated in the previously developed Historian application. Having data gathering capabilities, the data analysis module was implemented towards developing a completely autonomous software solution capable of optimizing different industrial systems.

　　　The augmented Historian was associated with a drinking water facility consisting of water wells as well as water treatment and distribution facilities. The existing process control structures in relation with the tested Historian (see Figure 4) were as follows:

- Water wells S7-314 type PLCs were all integrated in the S7-315 PLC, which was responsible for water distribution using the S7 protocol. Level- and flow-based control loops that were responsible for automatic water requests from the wells were implemented in the S7-315. Local flow-based control loops were implemented at each S7-314 PLC.
- The entire water treatment process was guided by two redundant S7-412-5H type PLCs.
- The S7-315 PLC responsible for water distribution and the redundant PLCs responsible for the water treatment process were integrated in the WinCC 7.2-based SCADA system, which consisted of two redundant servers and two clients. Since connectivity packs were configured at each server, OPC UA servers were available and assured interoperability.
- The solution was implemented on a Raspberry Pi 3 B (because of its reduced physical dimensions and the availability of industrial cases, which makes it suitable for industrial environments) using the Node-RED environment and an embedded Java application. The OPC UA client (several nodes were used: OpcUa-Browser, OpcUa-Client, TCP, etc.) was used to interface the entire system for data gathering and for noninvasive interventions over the local process. Having a complete local redundancy for the water treatment control structures, the Historian connection to the SCADA system was enough, but the S7 node was prepared for backup interfacing in case of a total SCADA failure. S7 protocol allowed for the current study to interface all PLCs. Details about the already developed Historian application can be found in [22]. This application was used as a starting point, to which the data dependencies identification algorithm was added.
- An OPC UA server (Node-RED flow was created to define the OPC UA folder/tag structuring inside the secured endpoint in order to constantly populate the address space with values and to propagate an eventual tag change) was implemented to assure higher-level interoperability of the Historian.
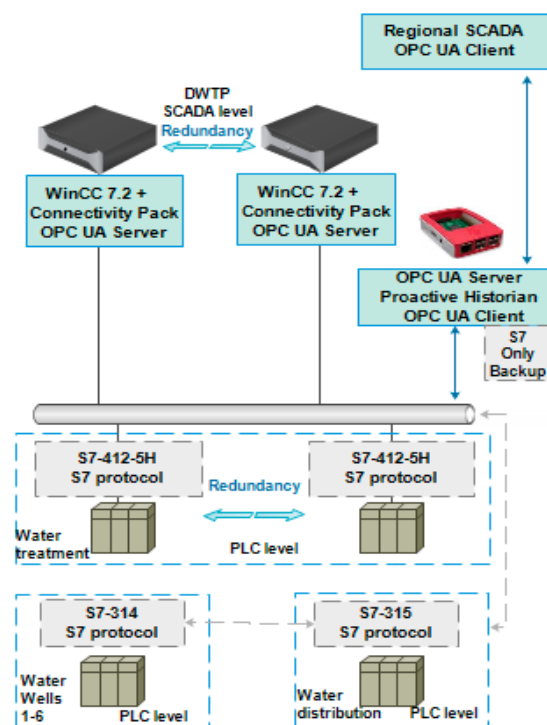

Figure 4. The Historian in relation with the local structures.

Regarding integration of the proposed algorithm into the Historian graphical user interface (GUI), a new tab ('Analyzer') was added to the interface. Figure 5 presents the Historian GUI with the newly added tab. The new tab's content was divided into two areas: 'Input' and 'Result'. In manual regimes, the 'Input' area allowed the user to choose the input that will be sent to the algorithm. From the 'Source' drop-down control, the user could choose the desired table from the SQLite database to be analyzed by the algorithm. This control was necessary because of the Historian internal data management system (the division of stored data in different tables, having one table per each stored variable list). More details about how the Historian handles data and stores it in different tables can be found in [28]. Also, in the 'Input' area, the 'Interval' control allowed the user to choose a subinterval of time from the entire interval available in the table chosen at 'Source'. This allowed the algorithm to run over custom periods, thus offering greater flexibility for the user. Finally, the 'Reference' control allowed the user to choose one of the variables that had values stored in the table selected at 'Source', a variable that was a reference for the algorithm. Input data that were chosen by the user were processed before being fed to the algorithm. Preprocessing of input data consisted of different checks and verifications, and it computed the daily averages of all the analyzed measured characteristics by using a complex, dynamically generated SQL query. After being processed, input data were inserted into specific data structures (classes and lists implemented in Java), which were then used by the algorithm.
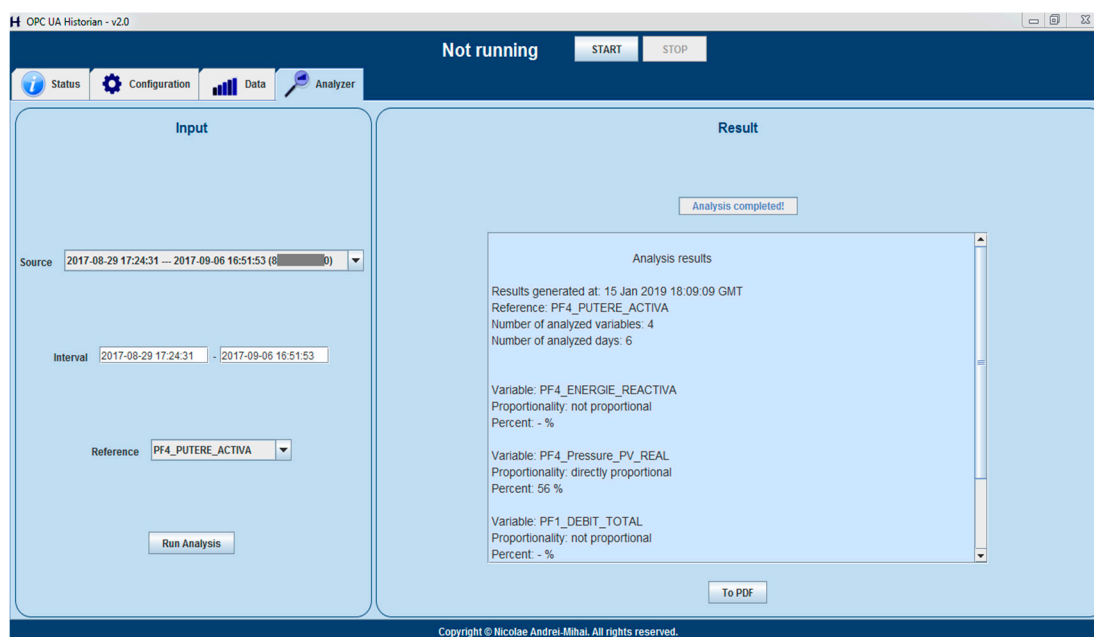

Figure 5. Integration of the proposed algorithm into the Historian application GUI.

The 'Result' area of the tab was used for displaying the output of the algorithm analysis. This area also implemented an export feature, which built a PDF document containing the results displayed in the text area on the GUI. This feature facilitated the extraction of the algorithm output outside of the Historian application. Figure 5 presents an example of the results displayed on the Historian GUI after a successful analysis made on real data, which was collected from the water industry by the Historian application.

3.2. Test Scenario

As it was integrated into the Historian application, the algorithm was successfully tested on different sets of real-world data with promising results, thus providing added value applications in the water industry. Figures 6–9 were added with the purpose of offering a graphical representation of some of the analyzed characteristic value evolutions in time, alongside output of the algorithm

after finishing the analysis. Graphical representations were made using real world data collected by the Historian.

The data dependencies identifying algorithm brought enhancements to the processes of a DWTP. For example, the following energy consumption-oriented scenario was considered. For a specific period, only water well no. 1 (WW 1) was used. During this time, the Historian recorded and analyzed the overall energy consumption and the water flow for the WW 1. After this first test period, only WW 2 was used over the same amount of time. The testing process included individual or concomitant functioning of multiple wells, and water flow differences were taken into consideration. Thus, dependency between the water flow of each well and the overall energy consumption was identified. By comparing those dependencies, it was possible to identify which water well generated minimum energy costs. In more detail, if WW 1 and WW 2 were both directly proportional to overall energy consumption, but WW 2 influenced the energy to a lower degree, then it should be prioritized in front of WW 1 in order to increase the efficiency of the DWTP. This result was difficult to identify in the absence of data dependency identifying algorithms because of both the large amounts of data that were recorded and because the water flows of each drilling varied depending on current drinking water network demands. Also, many other system tags influenced the general cost, and data analysis was complex. Figure 6 presents a water well output flow and a water facility overall energy consumption, while Figure 7 presents data representing the water flow from a different water well and the energy consumption. Although the result of data dependency analysis identified that both water flow values were directly proportional to the overall energy consumption, the "Quantity" information provided important details: the WW from Figure 6 required less energy to be treated than the one from Figure 7, so the first WW should be prioritized.

A different test scenario was presented in Figure 8, where WW 1 functioned throughout the entire test case time period, and WW 2 started occasionally during this period. Figure 8 presents WW 2 water flow as reference and the overall water turbidity (the turbidity of the mixed water coming from both sources) as an analyzed characteristic. Results of the algorithm (inversely proportional) signified that WW 2 had an inverse effect on turbidity when compared to the water from WW 1.
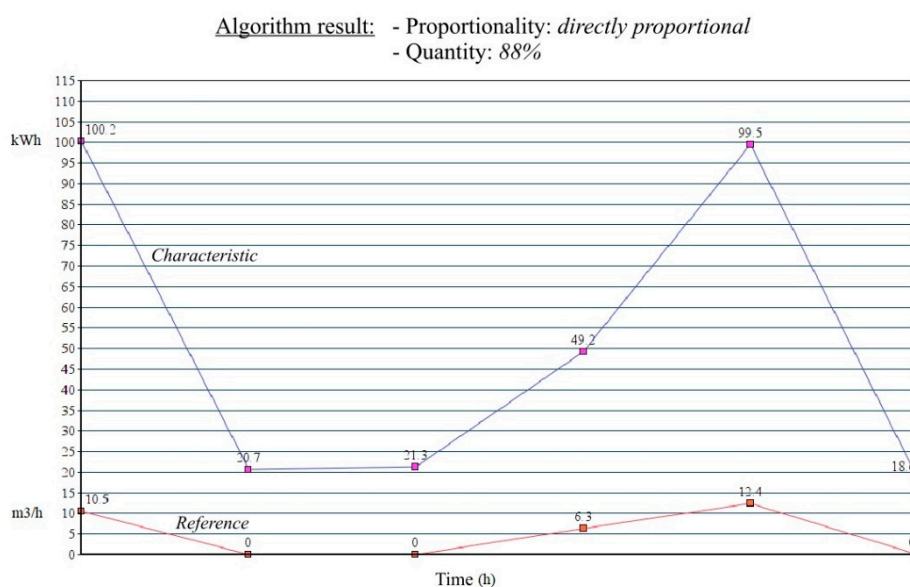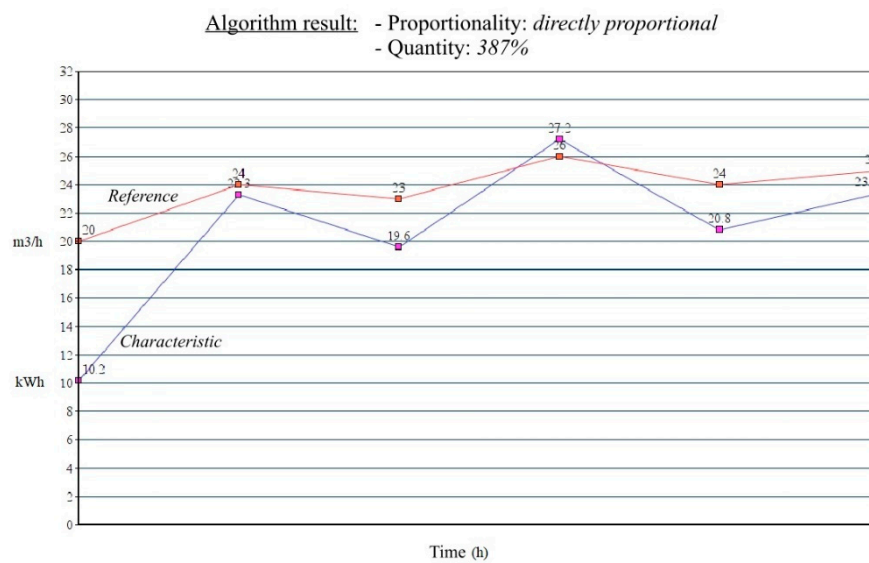

Figure 6. Test case graphical representation.

Algorithm result: - Proportionality: *directly proportional*
- Quantity: *387%*

Figure 7. Test case graphical representation.

Algorithm result: - Proportionality: *inversely proportional*
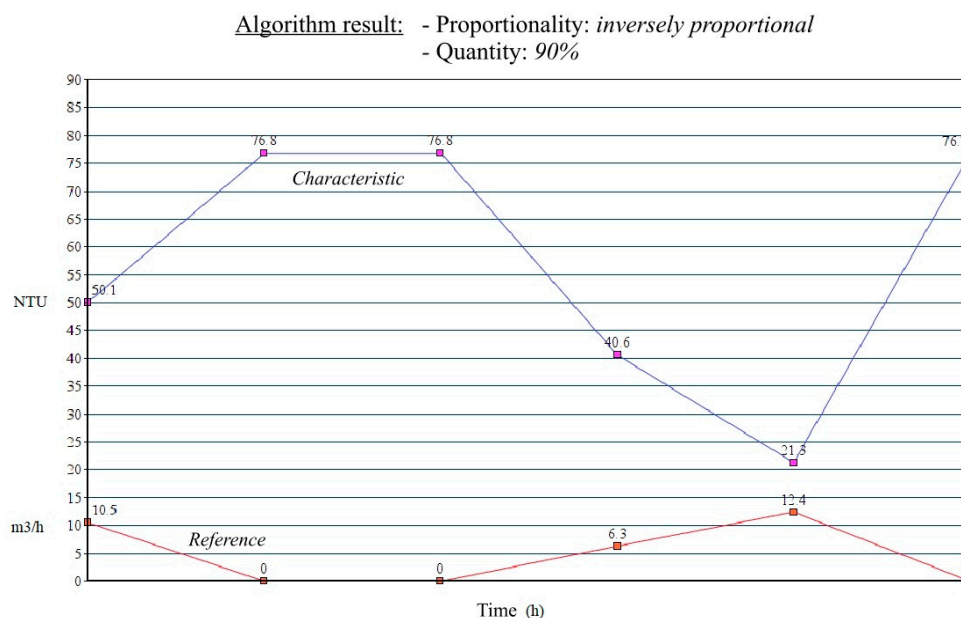- Quantity: *90%*

Figure 8. Test case graphical representation.

The test case scenario from Figure 9 presents a water well output flow as reference and the overall working time of a different equipment in the treatment plant as an analyzed characteristic. Because the algorithm found that those two values were not related, it meant that the usage of that specific drilling did not influence the maintenance of the equipment (did not wear down the technical system faster than usual).
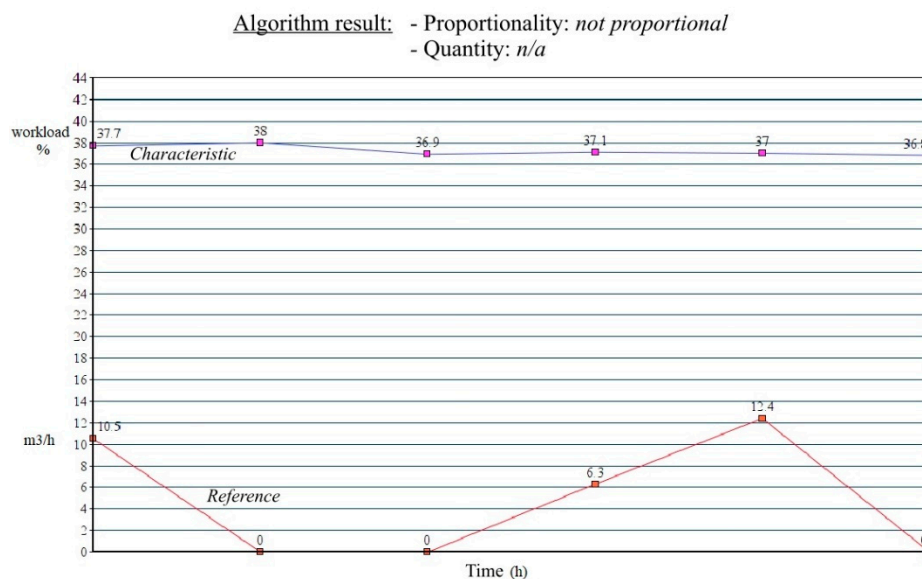
Figure 9. Test case graphical representation.

3.3. One Step Further to Improve a Drinking Water Facility

Taking the previously presented results of the current research one step further, the authors used the obtained dependencies and chose to experiment with one important cost factor reduction: the consumed energy. To test the impact on the consumed energy, the following steps were necessary:

- After analyzing the accumulated data using the presented solution, patterns were identified, and quality indicators for the water sources were conceived and set.
- Other research activity results (not published yet) were used to convert the water source quality indicators and the functioning hours into priorities that influenced the requested amount of water from each source in order to optimize water treatment. According to the determined priorities for the water sources, flow references were calculated and considered for each well.

The scenario consisted of a DWTP (Figure 1) with an inlet provided by six water wells (WWs). Figures 10–12 present an example of real parameter evolutions. Over a long period, the authors identified that only four WWs functioned with a reference flow provided manually by the operator in automatic mode (WW2, WW3, WW4, and WW7), and the other two WWs were stopped manually (see Figure 10). Figures 11 and 12 show water treatment/distribution process input and output parameters.

The current solution provided results for the four functioning WWs and, therefore, the corresponding water quality indicators were determined. The quality indicators may have values in the 0–10 interval. Initially, all quality indicators were set to 10 since no related previous knowledge was available. The following main parameters guided the data dependency analysis:

- DWTP output water quality indicators that were kept inside limits: pH, conductivity, and turbidity.
- Overall energy consumption was the cost reduction objective.
- Chlorine consumption was kept under a limit.
- Filters were washed no more than 1 cycle/filter/day.

The algorithm required longer data gathering because the water wells functioned according to water requirements from the distribution network and followed a rotation algorithm based only on functioning hours. Data dependencies were identified by the algorithm, and water quality indicators for each well were determined according to the degree of dependence. The best encountered water well was always the first reference when adjusting the quality values.

Since the Historian could permanently gather and analyze data, the Historian was able to adjust water quality parameters as time passed. This was an important characteristic, considering that water quality parameters of water sources change over time.
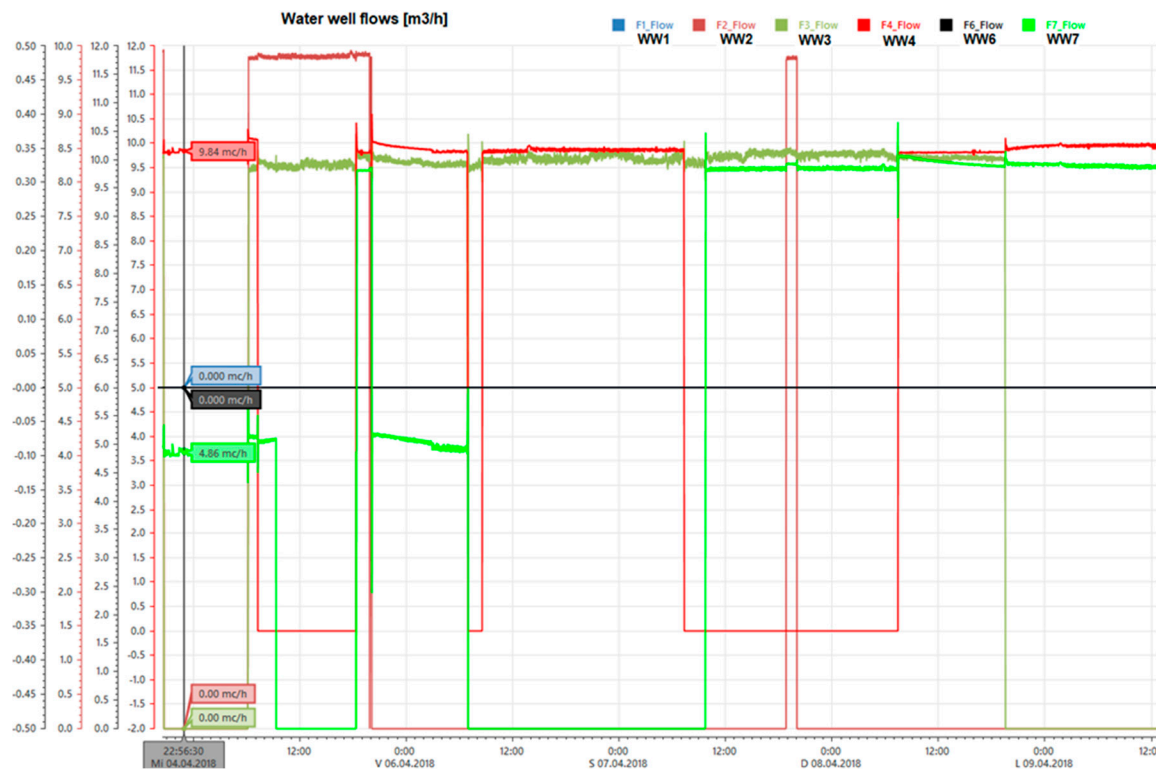

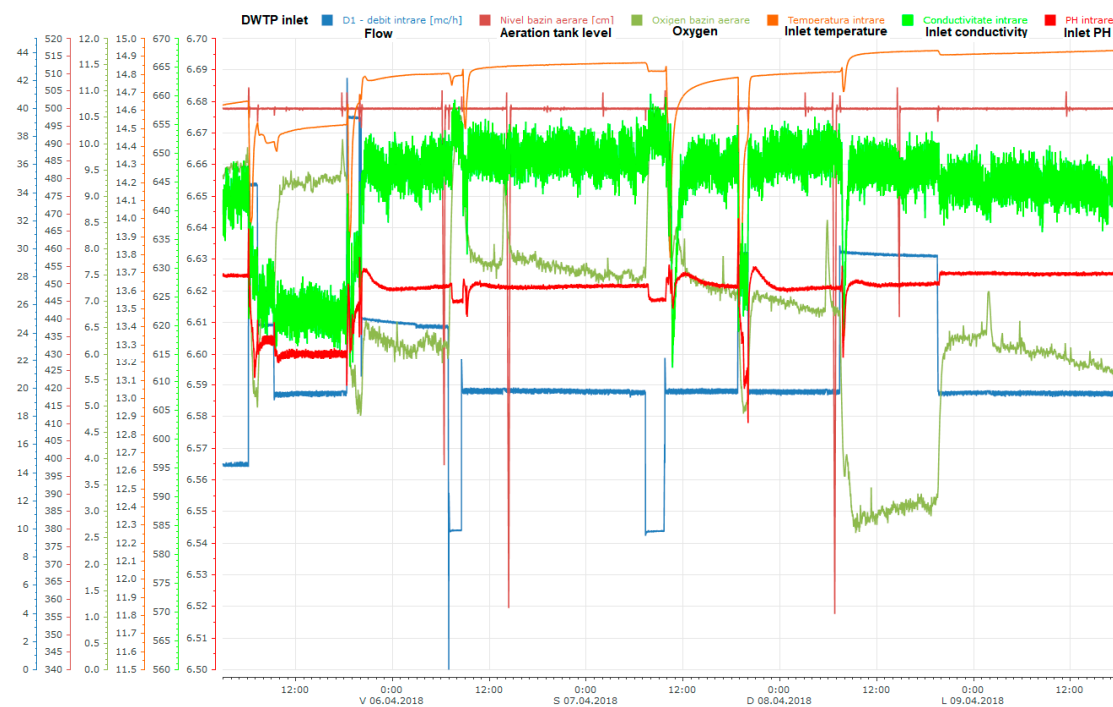**Figure 10.** Water well (WW) output flows.
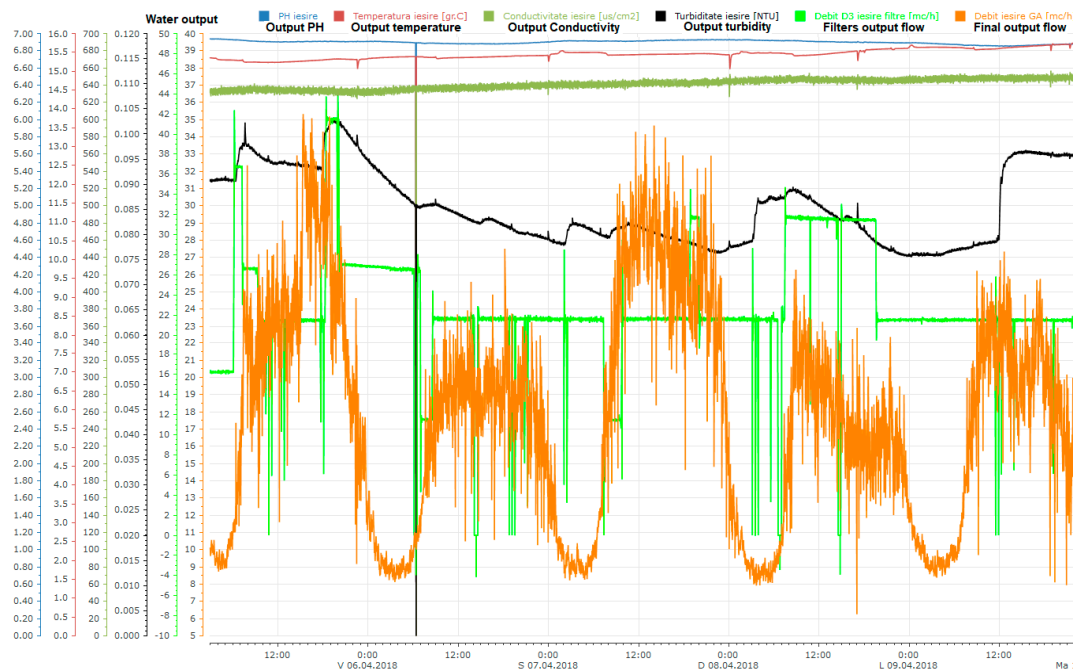

Figure 11. DWTP input parameters.

Figure 12. Treatment/distribution output parameters.

Water quality indicators were applied as input parameters for other research activities in order to set a WW priority based on quality indicators. Based on previously determined water quality indicators and the functioning hours of the water wells, priority indicators were defined by a normalizing procedure in the 0–10 interval. A total WW priority indicator was defined as the weighted sum of the quality priority indicator and the functioning hours-based indicator (e.g., $P_f = \alpha \times PH_f + \beta \times PQ_f$; where P_f, PH_f, and PQ_f are the total priority indicator, the functioning hours and the well quality based priority indicators, respectively; and α and β are weighting factors). The total priority indicator guided the decision regarding the activation of water sources. The setpoint regarding the requested flow for each WW was then determined using the WW quality priority indicator and the local flow limitations (e.g., $F_{w\_f} = F_{f\_min} + \gamma \times \left(F_{f\_max} - F_{f\_min}\right) \times \frac{PQ_f}{10}$; where F_{w_f}, F_{f_min}, and F_{f_max} are the flow setpoint, the minimum flow limit, and the maximal flow limit, respectively; and γ is a weighing factor that influences the maximum possible flow value for the WW). Each WW contained a local flow-based control loop that functioned according to the determined flow setpoint. The algorithm of the existing system applied the WW rotation rule based on functioning hours. This algorithm was augmented according to the energy consumption reduction strategy. Figure 13 presents an example of how the total priority indicator was influenced by the quality and functioning hours priority indicators.
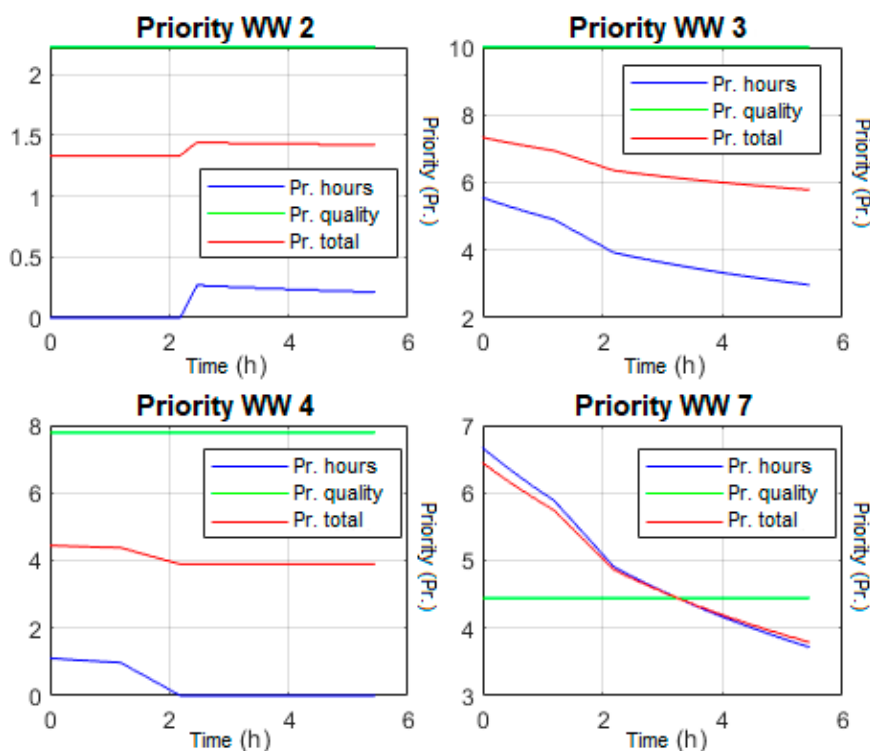
Figure 13. Water well priority indicators.

Treated water was distributed to the population and took into consideration the water pressure in the distribution network. However, the actual flow of water requested from the sources and treated in the DWTP was established following the level in the water distribution tank and the output flow in the distribution network, respectively. Other limitations from the DWTP ensured the correct treatment process. Considering all the mentioned requirements, the flow setpoints for each WW over six hours of evolution of the priority indicators from Figure 13 resulted in the evolution in Figure 14. As experimented by the authors, by applying and adjusting flow setpoints for the local WW control loops following priority indicators, at least a 9% reduction in energy consumption was reached.

4. Discussion

Results of the current study were analyzed and appreciated by practitioners from the industry. Since drinking water is a critical infrastructure, and we did not know the impact of research on the existing functional systems, only data gathering from various drinking water facilities was first allowed by the company. After showing data dependency results that demonstrated pattern identifications and how cost can be reduced, two aspects were discussed: (1) how the solution could be integrated and usable in the fog of the local system in a noninvasive manner regarding existing developments and (2) how the structure will react over the local system without perturbing the treatment and distribution process. Using OPC UA, interoperability is assured following correct definition of the constraints (e.g., functional acceptable limits of parameters). Any conclusions from the data dependency analysis algorithm will not allow any damage over some process characteristic. Also, if correct output tags are defined for Historian reactions to local process functional adjustments (e.g., setpoints for the control loops), then no local algorithm structural disturbance is possible.
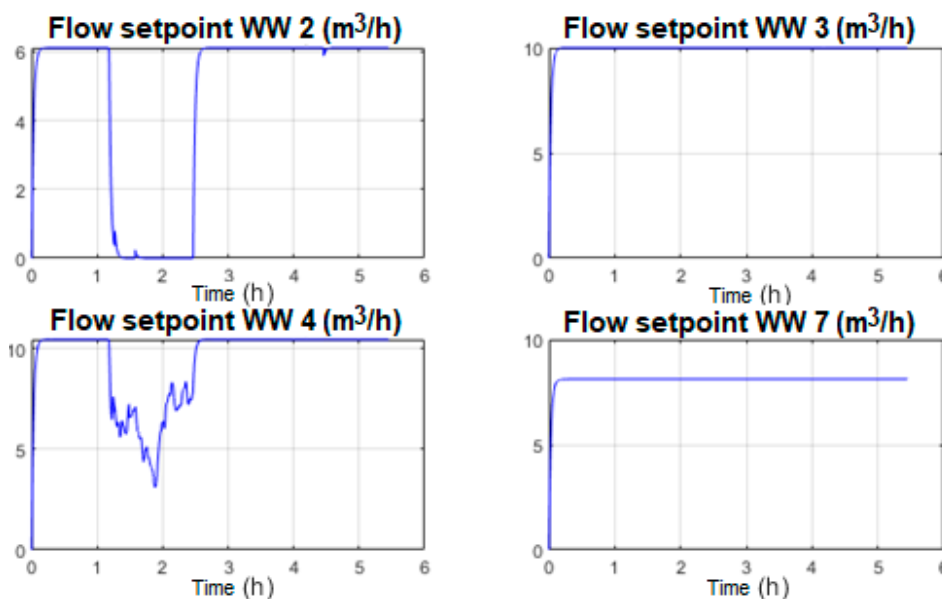
Figure 14. Water well flow references.

Hardware and software platforms were also important features in the solution validation process. Water companies and integrators require an industry-oriented solution, with a low-cost and a high TRL, to offer an easy transition towards automation/SCADA integrators. Raspberry Pi 3 B hardware (Quad Core ARM Cortex-A53 64 bit 1.2 GHz processor, 1 GB RAM, WiFi 802.11n, Bluetooth 4.1, Bluetooth LE, USB ports, Ethernet port, micro SD card, etc.) is a high-performance and very popular product, and with currently available enclosures it can become industry-oriented. Also, with available documentation and product maturity, it assures a high TRL for future integrators. From the software point of view, the operator's reaction was first studied, and the Historian GUI was proven to be very friendly and easy. The Node-RED environment, with its flow- and node-based structuring and browser-based development, can be assimilated easily by the integrators.

The obtained results have important practical impacts. Considering the case study on the influence of the water well's quality on the treatment process, the findings are important. The ability to associate quality indicators for each water well, following a dependency analysis towards the energy consumption as a cost objective, allowed the current well rotation algorithm to be augmented based on functioning hours that considered only maintenance cost reduction. It is also important that the data dependency conclusion may depend on the actual cost function of additional constraints, which assures that the process cannot be damaged (e.g., DWTP output water quality is in legal limits, the chlorine consumption is under a limit, etc.). Although efficiency is considerably increased by applying the presented study, the authors consider that energy can be further reduced in the following areas: water demand from the distribution network was very high in the current study, and only four activated wells were functioning for long periods; local automation in many other situations is poorly implemented; the applicability of the Historian to access and react on the real system was still reduced, which reduced the possibility to properly compare results.

Some interesting ideas may add to the complete view regarding proper prioritizing of water sources and cost efficiency following a long-term analysis. By having water wells in manual mode or not active for a long time (e.g., wells 1 and 6 from the tested process or a several days-based rotation strategy as encountered in some cases) it would be interesting to establish a minimal amount of data needed for a stable analysis. Considering filter washing cycles, so far the energy aspect is related to equipment energy consumption in accordance with the analyzed processes. However, in some situations the relation between flow and filter cleaning procedures is so critical that when filters are stopped, flow has to be significantly reduced. Also, after long-term usage, the solution will provide a

more detailed approach towards constituting the total priority indicator for the water wells in relation to the cost objective, which is a more exact determination of the weighting factors (e.g., α, β) associated with the quality and functioning hours priority indicators.

5. Conclusions

The current paper presented an essential step towards developing an autonomous, proactive Historian application that can record and analyze process parameters and then analyze and predict the future evolution of the system following the best encountered recipe. The approach assumes that the solution will react over the local system and adjust process functions in order to achieve predefined objectives.

A contribution of the paper is the reference architecture, which guides future development of an entire set of algorithms that will collectively offer the functionality of a proactive Historian. The proposed reference architecture is generic and can be applied for any industry.

According to the gather data, the algorithm identifies data dependencies according to reference characteristics and is able to associate degrees of dependency. Also, the algorithm allows constraint settings that consider functional parameter value limitations or output tag definitions. Constraint settings, together with the interfacing capabilities of the Historian, provide the noninvasive character of the solution towards the local processes and control structures. Also, the implemented interface assures vertical and horizontal interoperability. Eventually, contextual data integration from external sources may influence pattern identification and decision making.

The determined dependencies are evaluated, prioritized, and structured. With the initial settings and data dependencies, the Historian is becoming process-aware. By associating a cost objective, data dependencies help in determining the best recipe for improvement and the possibility to react over the local system in order to increase efficiency.

The developed solution is applied and tested with good results in the water industry. The main test scenario consists of a water treatment facility that receives water from several water wells and sends the treated water to the distribution network using a pumping station and reservoirs. Large amounts of data are gathered from the entire process, and several dependencies were identified with the purpose of reducing operational costs. The first results prove the efficiency of identifying data dependencies (using the available data) by considering energy, turbidity, or equipment functioning time as example reference characteristics. The next results are associated with a total energy reduction objective function. Following the data dependency identification algorithm, quality priority indicators are associated to each water well, and total priority indicators and flow setpoints for local control loops are determined. To go one step further, the evolution of the local flow setpoints for each water well are presented, which noninvasively changes local system behavior and proves at least 9% energy efficiency improvements.

Conflicts of Interest: The authors declare no conflict of interest.

References

1.   Gilchrist, A. *Industry 4.0: The Industrial Internet of Things*; Springer: Berkeley, CA, USA, 2016.
2.   Jeschke, S.; Brecher, C.; Meisen, T.; Ozdemir, D.; Eschert, T. Industrial Internet of Things and Cyber Manufacturing Systems. In *Industrial Internet of Things*; Jensche, S., Brecher, C., Song, H., Rawat, D.B., Eds.; Springer: Boston, MA, USA, 2016; pp. 3–19.

3.　Vandikas, K.; Tsiatsis, V. Performance evaluation of an IoT platform. In Proceedings of the IEEE Eighth International Conference on Next Generation Mobile Apps, Services and Technologies (NGMAST), Oxford, UK, 10–12 September 2014.

4.　Khan, W.A.; Wisniewski, L.; Lang, D.; Jasperneite, J. Analysis of the requirements for offering industry 4.0 applications as a cloud service. In Proceedings of the 26th IEEE International Symposium on Industrial Electronics (ISIE), Edinburgh, UK, 19–21 June 2017.

5.　Huh, J.H. PLC-based design of monitoring system for ICT-integrated vertical fish farm. *Human-centric Comput. Inf. Sci.* **2017**, *7*, 20. [CrossRef]

6.　Lee, H.G.; Huh, J.H. A Cost-Effective Redundant Digital Excitation Control System and Test Bed Experiment for Safe Power Supply for Process Industry 4.0. *Processes* **2018**, *6*, 85. [CrossRef]

7.　Wang, S.; Wan, J.; Zhang, D.; Li, D.; Zhang, C. Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. *Comput. Networks J.* **2016**, *101*, 158–168. [CrossRef]

8.　Ożadowicz, A.; Grela, J.; Wisniewski, L.; Smok, K. Application of the Internet of Things (IoT) technology in consumer electronics–case study. In Proceedings of the IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), Torino, Italy, 4–7 September 2018; pp. 1037–1042.

9.　Borgia, E. The Internet of Things vision: Key features, applications and open issues. *Comput. Commun.* **2014**, *54*, 1–31. [CrossRef]

10.　Kim, H.K.; So, W.Y.; Je, S.M. A big data framework for network security of small and medium enterprises for future computing. *J. Supercomputing* **2019**, *75*, 3334–3367. [CrossRef]

11.　Huh, J.H.; Seo, K. An indoor location-based control system using bluetooth beacons for IoT systems. *Sens. J.* **2017**, *17*, 2917. [CrossRef]

12.　Fysarakis, K.; Askoxylakis, I.; Soultatos, O.; Papaefstathiou, I.; Manifavas, C.; Katos, V. Which IoT Protocol? Comparing Standardized Approaches over a Common M2M Application. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA, 4–8 December 2016.

13.　Haskamp, H.; Meyer, M.; Mollmann, R.; Orth, F.; Colombo, A. Benchmarking of existing OPC UA implementations for Industry 4.0-compliant digitalization solutions. In Proceedings of the 15th IEEE International Conference on Industrial Informatics (INDIN), Emden, Germany, 24–26 July 2017; pp. 589–594.

14.　Moraes, E.; Lepikson, H.; Konstantinov, S. Improving connectivity for runtime simulation of automation systems via OPC UA. In Proceedings of the 13th IEEE International Conference on Industrial Informatics (INDIN), Cambridge, UK, 22–24 July 2015.

15.　Dai, W.; Vyatkin, V.; Christensen, J.H.; Dubinin, V.N. Bridging Service-Oriented Architecture and IEC 61499 for Flexibility and Interoperability. *IEEE Trans. Ind. Inf.* **2015**, *11*, 771–781. [CrossRef]

16.　Cavalieri, S.; Regalbuto, A. Integration of IEC 61850 SCL and OPC UA to improve interoperability in Smart Grid environment. *Comput. Stand. Interfaces* **2016**, *47*, 77–99. [CrossRef]

17.　Derhamy, H.; Ronnholm, J.; Delsing, J.; Eliasson, J.; van Deventer, J. Protocol interoperability of OPC UA in Service Oriented Architectures. In Proceedings of the 15th IEEE International Conference on Industrial Informatics (INDIN), Emden, Germany, 24–26 July 2017.

18.　Muller, M.; Wings, E.; Bergmann, L. Developing Open Source Cyber-Physical Systems for Service-Oriented Architectures using OPC UA. In Proceedings of the 15th IEEE International Conference on Industrial Informatics (INDIN), Emden, Germany, 24–26 July 2017.

19.　Leszczynski, I. OPC UA App Development for Android. Master' Thesis, Aalto University, Helsinki, Finland, 2015.

20.　Korodi, A.; Silea, I. Achieving Interoperability Using Low-Cost Middleware OPC UA Wrapping Structure. Case Study in the Water Industry. In Proceedings of the 15th IEEE International Conference on Industrial Informatics (INDIN), Emden, Germany, 24–26 July 2017; pp. 1223–1228.

21.　Toc, S.I.; Korodi, A. Modbus-OPC UA Wrapper using Node-RED and IoT-2040 with application in the water industry. In Proceedings of the 16th IEEE International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 13–15 September 2018.

22.　Veichtlbauer, A.; Ortmayer, M.; Heistracher, T. OPC UA Integration for Field Devices. In Proceedings of the 15th IEEE International Conference on Industrial Informatics (INDIN), Emden, Germany, 24–26 July 2017; pp. 419–424.

23. Seilonen, I.; Tuovinen, T. Aggregating OPC UA servers for monitoring manufacturing systems and mobile work machines. In Proceedings of the IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), Berlin, Germany, 6–9 September 2016.

24. Korodi, A.; Radu, M.A.; Crisan, R. Non-Invasive Control Solution inside Higher-Level OPC UA based Wrapper for Optimizing Groups of Wastewater Systems. In Proceedings of the IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), Torino, Italy, 4–7 September 2018; pp. 597–604.

25. Crisan, R.; Korodi, A. Noninvasive control solution for energy efficiency in wastewater treatment plants. In Proceedings of the IEEE International Conference on Industrial Technology (ICIT), Lyon, France, 20–22 February 2018; pp. 1604–1609.

26. Gray, A.D.L.; Pisica, I.; Taylor, G.A.; Whitehurst, L. A Standardised Modular Approach for Site SCADA, Applications Within a Water Utility. *IEEE Access* **2017**, *5*, 17177–17187. [CrossRef]

27. Mazur, D.C.; Entzminger, R.A.; Kay, J.A. Enhancing Traditional Process SCADA and Historians for Industrial and Commercial Power Systems with Energy (Via IEC 61850). *IEEE Trans. Ind. Appl.* **2016**, *52*, 76–82. [CrossRef]

28. Nicolae, A.; Korodi, A. Node-Red and OPC UA Based Lightweight and Low-Cost Historian with Application in the Water Industry. In Proceedings of the IEEE 16th International Conference on Industrial Informatics (INDIN), Porto, Portugal, 18–20 July 2018.

29. McGreevy, R.J.; Sadowski, D.A.; Herbst, R.J.; Hartman, J.R. Historian Module for Use in an Industrial Automation Controller. U.S. Patent 762,738,5B2, 1 December 2009.

30. Yan, Q.; Wang, Y. A Kind of Efficient Data Archiving Method for Historical Sensor Data. In Proceedings of the 3rd International Conference on Information Science and Control Engineering (ICISCE), Beijing, China, 8–10 July 2016.

31. Dwivedi, A.; Pant, R.P.; Pandey, S.; Kumar, K. Internet of Things' (IoT's) Impact on Decision Oriented Applications of Big Data Sentiment Analysis. In Proceedings of the 3rd International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Bhimtal, India, 24 February 2018.

32. Kaur, H.; Kushwaha, A.S. A Review on Integration of Big Data and IoT. In Proceedings of the 4th International Conference on Computational Sciences (ICCS), Wuxi, China, 11–13 June 2018.

33. Node-RED. Available online: https://nodered.org/ (accessed on 21 May 2019).

34. Stang, S.; Wang, H.Y.; Gardner, K.H.; Mo, W.W. Influences of water quality and climate on the water-energy nexus: A spatial comparison of two water systems. *J. Environ. Manag.* **2018**, *218*, 613–621. [CrossRef] [PubMed]

35. Rosinska, A.; Dabrowska, L. Selection of Coagulants for the Removal of Chosen PAH from Drinking Water. *Water* **2018**, *10*, 7. [CrossRef]

36. Shan, W.; Cai, S.; Liu, C. A New Comprehensive Evaluation Method for Water Quality: Improved Fuzzy Support Vector Machine. *Water* **2018**, *10*, 1303. [CrossRef]

37. Tsagarakis, K.P. Operating Cost Coverage vs. Water Utility Complaints. *Water* **2018**, *10*, 27. [CrossRef]

38. Wu, W.; Simpson, A.R.; Maier, H.R.; Marchi, A. Incorporation of variable-speed pumping in multiobjective genetic algorithm optimization of the design of water transmission systems. *J. Water Resour. Plan. Manag. ASCE* **2012**, *138*, 543–552. [CrossRef]

39. Mala-Jetmarova, H.; Sultanova, N.; Savic, D. Lost in Optimisation of Water Distribution Systems? A Literature Review of System Design. *Water* **2018**, *10*, 307. [CrossRef]

40. Ciolofan, S.; Militaru, G.; Draghia, A.; Drobot, R.; Dragoicea, M. Optimization of Water Reservoir Operation to Minimize the Economic Losses Caused by Pollution. *IEEE Access* **2018**, *6*, 67562–67580. [CrossRef]

41. Diaz-Perez, F.J.; Pino-Otin, M.R.; Mouhaffel, A.G.; Martín, R.D.; Chinarro, D. Energy and Water Consumption and Carbon Footprint in Tourist Pools Supplied by Desalination Plants: Case Study, the Canary Islands. *IEEE Access* **2018**, *6*, 1727–11737. [CrossRef]

42. Wang, D.S. Raw water quality assessment for the treatment of drinking water. *Environ. Earth Sci.* **2016**, *75*, 19. [CrossRef]

43. Chowdhury, S. Water quality degradation in the sources of drinking water: an assessment based on 18 years of data from 441 water supply systems. *Environ. Monit. Assess.* **2018**, *190*, 7. [CrossRef]

44. Delpla, I.; Florea, M.; Rodriguez, M.J. Drinking Water Source Monitoring Using Early Warning Systems Based on Data Mining Techniques. *Water Resour. Manag.* **2019**, *33*, 129–140. [CrossRef]

45. Huang, P.; Zhu, N.; Hou, D.; Chen, J.; Xiao, Y.; Yu, J.; Zhang, H. Real-Time Burst Detection in District Metering Areas in Water Distribution System Based on Patterns of Water Demand with Supervised Learning. *Water* **2018**, *10*, 1765. [CrossRef]

46. Cheng, W.; Fang, H.; Xu, G.; Chen, M. Using SCADA to Detect and Locate Bursts in a Long-Distance Water Pipeline. *Water* **2018**, *10*, 1727. [CrossRef]

47. Osman, M.; Ab-Mahfouz, A.; Page, P. A Survey on Data Imputation Techniques: Water Distribution System as a Use Case. *IEEE Access* **2018**, *6*, 63279–63291. [CrossRef]

48. Kuriqi, A.; Ardiclioglu, M. Investigation of hydraulic regime at middle part of the Loire River in context of floods and low flow events. *Pollack Periodica* **2018**, *13*, 145–156. [CrossRef]

49. Angelakis, A.N.; Zheng, X.Y. Evolution of Water Supply, Sanitation, Wastewater, and Stormwater Technologies Globally. *Water* **2015**, *7*, 455–463. [CrossRef]

50. Kuriqi, A. Assessment and quantification of meteorological data for implementation of weather radar in mountainous regions. *MAUSAM* **2016**, *67*, 789–802.

51. Kuriqi, A.; Ardiclioglu, M.; Muceku, Y. Investigation of seepage effect on river dike's stability under steady state and transient conditions. *Pollack Periodica* **2016**, *11*, 87–104. [CrossRef]

52. Sooley, B.; Little, C. Reinventing the Role of the SCADA Historian. In Proceedings of the ISA Water/Wasterwater and Automatic Controls Symposium (WWAC), Orlando, FL, USA, 6–8 August 2013.

53. Chardin, B.; Lacombe, J.M.; Petit, J. M Data Historians in the Data Management Landscape. In Proceedings of the 4th TPC Technology Conference (TPCTC), Istanbul, Turkey, 27 August 2012.

54. Lee, J.; Bagheri, B.; Kao, H.A. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manuf. Lett.* **2015**, *3*, 18–23. [CrossRef]