39th IAHR WORLD CONGRESS
GRANADA, SPAIN 2022
From Snow to Sea
19-24 June 2022

Proceedings of the 39th IAHR World Congress
19–24 June 2022, Granada, Spain
doi://10.3850/IAHR-39WC2521716X2022926

A Reliable Monitoring Approach of Floods in Small and High Slope Watersheds

Gonçalo Jesus(1), Anabela Oliveira(2), João Rogeiro(3), Rui Rodrigues(4) and João Fernandes(5)

(1,2,3) Information Technology Research Group, National Laboratory for Civil Engineering, Lisbon, Portugal,
gjesus@lnec.pt, aoliveira@lnec.pt, jrogeiro@lnec.pt

(4,5) Water Resources and Hydraulic Structures Unit, National Laboratory for Civil Engineering, Lisbon, Portugal,
rjrodrigues@lnec.pt, jnfernandes@lnec.pt

Abstract

An implementation to instantiate a dependable data quality-oriented methodology in the Vinhas Creek monitoring network is presented herein. Redundancy was taken as a core aspect of network reliability. In this instantiation, we implement several machine learning mechanisms to process measurements from the multiple sensors while correlating them according to their geographical position, monitoring timing and the relevant physical processes involved. As an output, we are able to predict the sensor measurements and compare them with the actual sensing value obtained in the monitoring network station. Moreover, in case of any sensor failure, one or more replacement values can be issued. These are important for the correct simulation of the hydrologic and hydraulic processes of the dendritic watershed systems and to predict the inundation characteristics such as levels and flow velocities.

Keywords: Dependable monitoring; Machine learning; Flash floods; Data quality

1. INTRODUCTION

In the scope of the funded SUDOE Project INUNDATIO, we aim for the development of a management support system for flash floods in high slope watersheds based on the hydromorphology, which includes the component of hazard (by formulating the associated risk), the monitoring data analysis from precipitation and river flow (historical and real-time), the risk scenarios simulation and the vulnerability analysis regarding human impact and environmental and structural assets.

INUNDATIO project will support local administration along the case study site in improving risk management for flash flood events, according to the prediction and protection for this type of events and the resilience increase to reduce the human and resources impact. Flash floods are one of the most dangerous natural hazards. In the specific situation of the Portuguese case study of the INUNDATIO project, a small watershed with high slopes and small concentration time, there is a great focus on the development of a reliable early warning system that must be based on real-time monitoring of the forcing variables and hydrology and hydraulic modeling to evaluate the flooded areas and the inundation levels in the vulnerable area. In order to attain this objective, there are three main actions that are considered in the project.

First, the definition of innovative methodologies for hydrological analysis in high slope watersheds and data acquisition, consisting on i) a data acquisition methodology for historic data and a monitoring network, and ii) a geophysical, geomorphological, hydrological and river-hydraulics analysis.

The second main action is the development of a methodology to characterize the outputs of the previous action in terms of probability, vulnerability, frequency and impact by developing numerical and computational models of the hydrological processes in high slope watersheds and, therefore, construct a scenarios database for the risk management (Cheng et al., 2021).

Lastly, the third main action is the development of the emergency and evacuation plans in case of a catastrophe that include damage reports and repair actions and the dissemination of the project results, including workshops for the responsible entities.

For these three actions and, mainly, for the development of an effective emergency management support system for flash floods in small area watersheds with high slopes, we must consider a real-time monitoring network of both forcing and response variables in the basin to generate timely warnings. This network should be complemented with hydrological modeling and vulnerability analysis of flood prone areas within the watershed. In the case study of the Vinhas Creek basin, a low-cost monitoring network comprising multiple sensor stations was deployed along the area covering the pristine headwaters and the highly urbanized downstream floodplain, located at the Tagus estuary mouth. Despite its application in a specific case study, the reasoning for the monitoring system may be replicated for similar watersheds.

The effectiveness of the forecast procedures and emergency warning for these natural and hazardous events may however be hampered by inconsistent real-time observed data. Ensuring the quality of monitoring data is fundamental to avoid false alarms or to ignore relevant events (Jesus, 2017). In order to increase confidence in the sensory and sensor network technologies, considering that these are subject to potentially harmful environmental factors, a dependable data quality-oriented methodology (Jesus, 2021) was applied to the monitoring network datasets.

Herein, we detail in Section 3 the monitoring network that will be put in place in the Portuguese case study and discuss additional inputs that may be useful for the correct behavior of the reliable monitoring methodology. As mentioned, we will follow an approach based on the data quality-oriented methodology described in Jesus et al. (2021), overviewed in Section 4. Section 5 concludes the paper and discusses the project's future and expected results.

2. RELATED WORK

In order to develop a reliable environmental monitoring system, we promoted a detailed characterization of the involved physical phenomena and we must be able to cope with possible issues that affect sensing data. To do so, we defined related fault models according to the dependable data quality-oriented methodology and identified likely associated faults.

Following the comprehensive survey on fault models and strategies to test the reliance of sending data conducted in Jesus et. al (2017), we discussed in Section 3 the severe challenges on the operations of deployed sensors. Although understanding all possible impacts is not the focus of the project, their identification and mitigation is part of the reliable monitoring system development. Consequently, in this section, some possible solutions for particular sensor failure outputs are approached.

Both natural environmental-related and man-made related events are known to have a negative influence on sensors. This influence is often observed in meteorological events (Gomes et al., 2013), such as storms with strong winds and heavy rain affecting overall quality, either by causing total loss of the sensors or by relative sensor displacements, which commonly interferes with the intrinsic sensing processes and disturbs or invalidates the measurements. The presence of wildlife or human interference/tampering affecting both casing and sensor intake may also affect the case study sensors.

Moreover, as detailed in the above-mentioned survey, in the past decade many studies supporting the use of machine learning techniques and sensor fusion to identify or classify events, including failure situations, were presented. Likewise, we overview a few studies here considering these categories.

If we consider failures as events, a sensor-fusion solution to detect events in wireless sensor networks (WSN) is presented by Bahrepour et al. (2010), including machine learning techniques such as artificial neural networks (ANN), decision trees or Naive Bayes for event classification. These techniques have been proven to provide low computational complexity required for a real-time detection of the event, which was applied to a fire alarm network dataset with effective accuracy.

Given that an event can also be characterized by an abnormal reading or set of readings, we also considered fault detection related works. Moreover, these are mostly focused on a specific type of failure such as outliers/noise or systematic faults.

In outlier detection, studies can be further divided into the several machine learning approaches that were taken in consideration. These studies can be thoroughly overviewed in surveys such as the one by Ayad et al. (2017). The common machine learning categories include: statistical-based algorithms, using probability models to capture the distribution of data and be able to assess if a measurement is an outlier; nearest neighbor-based algorithms, calculating a similarity measure between two data measurements; clustering-based algorithms, used as a data mining approach to group similar sensor measurements into clusters with similar behavior, whereas outliers do not belong to any cluster; classification-based algorithms, training a classification model using a set of sensor measurements and classifying an unseen measurement into the learned class; and spectral decomposition-based algorithms, which aim to use principal components analysis (PCA) to find normal behaviors in sensors data.

Studies on systematic failures detection are rather limited when compared to outlier detection. Takruri et al. (2007) present a machine learning approach to detect offset and drifting faults in WSNs. Their method consists in using Hidden Markov Models to capture both the dynamics of the environment and the dynamics of the faults. Other works (Kumar et al., 2015; Rathore et al., 2018) focus specifically on drifting failures detection and correction in sensor networks, using several data fusion algorithms including statistical techniques, Kalman filters, Interacting Multiple Model algorithm, Recursive Bayesian algorithm, Spatial Kriging method, and ensembles of these techniques.

Although all these studies provide effective results in the field of detecting faults in sensor measurements, their focus is mostly for laboratory or indoor conditions. This contrasts with the work promoted by Jesus et al. (2021) that considers the high variability of environmental monitoring, which we adopted here as an implementation of the related methodology.

3. DATASETS

To demonstrate how we can instantiate the dependable data quality-oriented methodology, in this section we consider the available datasets, either collected from a real environmental monitoring system or extracted from a multiple scenarios database. Therefore, in Section 3.1 we briefly describe the considered watershed case study. After that, Section 3.2 describes the monitoring network comprising weather stations and surface water level sensors across the watershed. Lastly, we also discuss the additional datasets that will be available via simulations.

3.1 Case Study

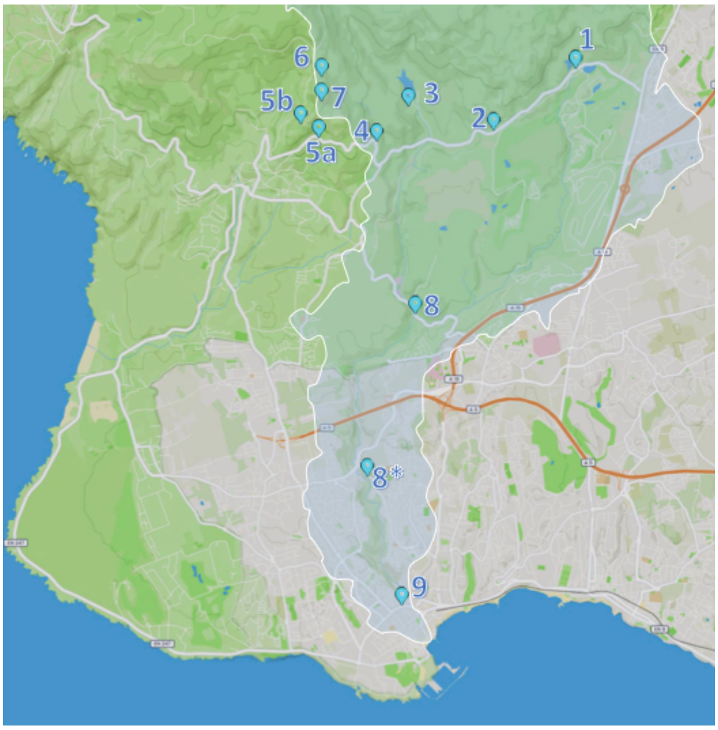

Figure 1. Vinhas creek basin monitoring network.

We apply our instantiation of the methodology using the online environmental monitoring network of the INUNDATIO portuguese case study of Vinhas Creek basin. This basin is located within the municipalities of Sintra and Cascais, around Lisbon, Portugal and is characterized by the steep slopes from the mountainous headwaters converging to a final flat reach located in the highly urbanized area of the city of Cascais, close to the Tagus estuary mouth. With this configuration, the area is rather prone to torrential floods that inundate the lower part of the city, which was witnessed several times in the last decades.

Along the Vinhas Creek basin, we are putting in place an online monitoring network with several monitoring points, measuring multiple physical parameters at different frequencies. In Figure 1 we provide an overview of the sensor network distribution in order to monitor important locations that might be relevant as forcing variables or get affected during a flash flood event.

3.2 Monitoring Network

All planned locations present in Figure 1 will be considered as relevant input for the reliable monitoring methodology, as correlated data is available for the temperature, windspeed, wind direction, rain and water surface level parameters. In fact, the use of sensors providing data that is correlated is fundamental in the adopted methodology. Sensor correlation and redundancy are part of the solution to avoid sensing data loss, which, in an open environment such as the case study one, is particularly sensible. Sensors can be affected by external disturbances, such as abnormal weather conditions like storms surges, or impact of animals and objects.

We divided the Vinhas Creek monitoring network into the two types of equipment in place. First, we installed weather stations in all planned locations, which are linked together by LORA private network protocol. Figure 2 overviews the sensors and the LORA equipment in each of the locations.

Figure 2. Weather stations network equipment. LORA gateway and included sensors.

The following step will be to install water surface level sensors in some of the planned locations within the Vinhas creek basin in order to measure significant water rise in the monitored streams. The Staal equipment was chosen to measure this parameter, and it works within its own network, connected via mobile communications, making the values available in a cloud environment. Figure 3 presents an example of the chosen equipment.

Both types of equipment will output the sensing measurements in a time series format, ideal for the adopted methodology.

3.3 Virtual/Simulation Dataset

In addition to the monitoring network dataset, we will make use of numerical models to first construct a scenario database and then feed the machine learning algorithms as virtual sensors in the same location as previously described. In the following section we detail the methodology and the necessity of constructing valid prediction models in order to obtain a reliable monitoring system. Moreover, considering the temporal dimension of the flash floods events and the need to provide a rich dataset for the models' construction, we will use two popular simulation models, HEC-HMS (Chu, 2009) and HEC-RAS (Brunner, 2021), to calculate the data for the monitoring locations based on the historic data. The obtained datasets will provide useful data that we can then feed as virtual sensing data to the adopted methodology implementation, as discussed ahead.

HEC-HMS stands for Hydrologic Engineering Center - Hydrologic Modeling System and is a software used to simulate the complete hydrologic processes of dendritic watershed systems. It comprises the simulation of the processes associated with infiltration, evapo-transpiration, snowmelt, soil moisture accounting and hydrologic routing.

HEC-RAS is the Hydrologic Engineering Center - River Analysis System (HEC-RAS), a software that performs one and two-dimensional unsteady flow calculations. Taking as input the geometry, the roughness and the boundary conditions (flow discharge and upstream and downstream levels), HEC-RAS predicts the inundation characteristics like levels and flow velocities.

Both models are calibrated to the case study, described above, and will run multiple input scenarios and thus generate a comprehensive pool of results related to river flow, inundation levels and river velocities.

Figure 3.Staal water surface level sensor.

4. METHODOLOGY

A methodology for processing measurements from multiple sensors is proposed in Jesus et al.(2021) that addresses essentially three main functions: failure detection, data quality assessment and data correction. The methodology was proposed to be applied during runtime, for the detection of faulty measurements and the respective mitigation in a continuous way, during a monitoring network operation. Moreover, the data correction function promises to provide an estimation of a replacement measurement with better quality and thus being a solution for dependable data quality with prediction mechanisms to model the sensors behaviors and estimate expected values for the measurements.

As observable in the related work, most of the failure detection and correction solutions are based in machine learning approaches. Although some steps or techniques in the adopted methodology do not necessarily need to be based on machine learning, this is mandatory for the detection and estimation steps.
In order to instantiate the methodology, we require several models to characterize the correct behavior of each sensor. These models are composed of one or more supervised learning techniques.

The first step of the instantiation is to construct the sensor models which requires datasets with clean and valid sensing data from all the sensors. The goal is that the machine learning algorithms are able to explore temporal correlations between consecutive past measurements of a determined target sensor, spatial and value correlations between past measurements of the determined target sensor and past measurements from a variable number of other sensors. Spatial correlations are used when the target sensor and other sensors are in different geographical locations. Value correlations exist for example between the target sensor and sensors placed at the same location.

Consequently, the methodology is designed as an ensemble of supervised learning methods, which require an offline initial training phase for each node's model construction.

Implementation of this methodology follows a two-step procedure. First, we need to promote model training, consisting of the neural networks (models) training with previous data from the same sensor. This will be performed in an off-line manner and, therefore, before the online application. We considered here, as in Jesus et al.(2021), the adoption of artificial neural networks as the machine learning process to construct the sensor models. The second step is the actual application of the trained models, using these in real-time by processing the incoming sensing data. These steps are depicted in Figure 4.

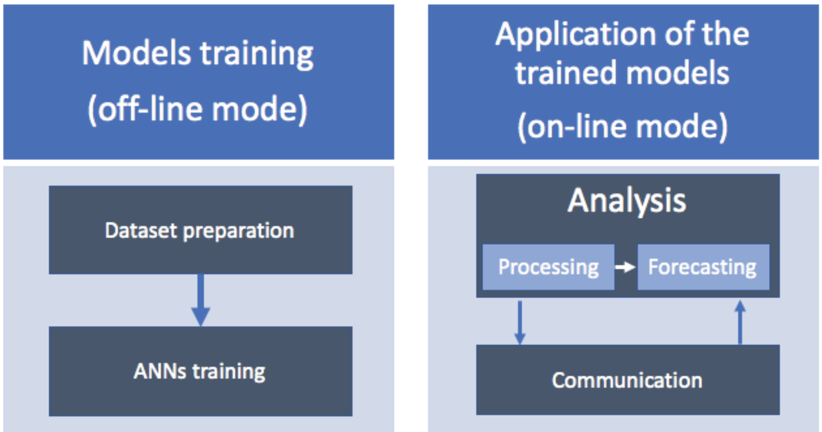

Figure 4. Generic two steps for the implementation of the adopted methodology.

The first step prepares the received sensing datasets for the model training. The outputs of the initial step are trained neural networks that introduce a redundancy method based on the sensors' correlation. The second step processing is based on these models.

In an online mode, we assume the monitoring network to be continuously producing information (sensing data) and in a timely manner we have a broker that receives each sensor data and organizes the data in order to run the trained models. This analysis part will be performed in a server due to the computing requirements and it comprises the data processing and the measurement estimation procedures. The methodology uses classification algorithms to detect anomalies in the measurements by comparing the received values with trained models. Predictions, as mentioned, are estimated via neural networks. In addition, another methodology output is a quality coefficient. For each received measurement, its quality is quantified. Whenever a measurement is faulty, we need to promote its correction, which can be done by using the models' predictions.

In the INUNDATIO implementation, we will also use another type of datasets in the training process. As mentioned in the section above, we have a scenario database comprising environment models' outputs for those timelines. Therefore, we have time series for the multiple parameters of interest, at any geographical point of the scoped area of the numerical model, representing an additional input as virtual sensor streams for the training process. Or, if necessary by any constraint or requirement, such as absence of information from a determined sensor, or simply to exploit the availability of an additional source of correlated data to increase the sensor fusion accuracy, these types of model outputs can be used as virtual sensors in the fusion process to achieve a more dependable monitoring system.

The virtual sensors provide a new source of redundancy in the sensor fusion process, allowing to mitigate errors in the real sensors' measurements. On the downside, these simulation models may not consider all existing processes with influence on the environment, and provide data that are limited by the accuracy of the underlying numerical methods which can cause significant errors depending on the accuracy of the numerical scheme and the model application setup. Nevertheless, a virtual sensor is better than not using it. Thus, a virtual sensor still provides very useful information that can be used to replace a real sensor in case of failure of the latter. In summary, virtual sensors are a useful source of redundancy, proving their usefulness in the sensor fusion process and as a guarantee of the quality of a monitoring network.

5. CONCLUSIONS AND FUTURE WORK

A reliable monitoring approach for flash floods management in small and high slopes watershed systems is presented, integrating a sensor monitoring network, a scenario database and a dependable quality-oriented methodology based in machine learning procedures. Moreover, we applied this approach in the Vinhas Creek basin and defined the needed steps for its successful implementation.

INUNDATIO project is still undergoing and in the recent future we will finish the monitoring network deployment, including the water surface level and velocity sensors. Also, we will add multiple simulations retrieved from the overviewed numerical models. Lastly, we will proceed with the machine learning models training and characterization of the flash floods impactful forcing variables.

6. ACKNOWLEDGEMENTS

This work was funded by SUDOE through project INUNDATIO (SOE3/P4/E0929).

7. REFERENCES

Bahrepour, M., Meratnia, N., Poel, M., Taghikhaki, Z., & Havinga, P. J. (2010, November). Distributed event detection in wireless sensor networks for disaster management. In 2010 international conference on intelligent networking and collaborative systems (pp. 507-512).

Brunner, P.E. (2021) HEC-RAS 6.0, Hydraulic Reference Manual, May, Hydrologic Engineering Center, US Army Corps of Engineers.

Cheng, X., Ma, X., Wang, W., Xiao, Y., Wang, Q., & Liu, X. (2021). Application of HEC-HMS Parameter Regionalization in Small Watershed of Hilly Area. Water Resources Management, 35(6), 1961-1976.

Chu, X., & Steinman, A. (2009). Event and continuous hydrologic modeling with HEC-HMS. Journal of Irrigation and Drainage Engineering, 135(1), 119-124.

Gomes, J., Rodrigues, M., Azevedo, A., Jesus, G., Rogeiro, J., & Oliveira, A. (2013). Managing a coastal sensors network in a nowcast-forecast information system. In 2013 Eighth International Conference on Broadband and Wireless Computing, Communication and Applications (pp. 518-523).

Jesus, G., Casimiro, A., & Oliveira, A. (2017). A survey on data quality for dependable monitoring in wireless sensor networks. Sensors, 17(9), 2010.

Jesus, G., Casimiro, A., & Oliveira, A. (2021). Using Machine Learning for Dependable Outlier Detection in Environmental Monitoring Systems. ACM Transactions on Cyber-Physical Systems, 5(3), 1-30.

Jesus, G., Casimiro, A., & Oliveira, A. (2021). Systematic Failure Detection and Correction in Environmental Monitoring Systems. Sensors & Transducers, 252(5), 28-34

Kumar, D., Rajasegarar, S., & Palaniswami, M. (2015). Geospatial estimation-based auto drift correction in wireless sensor networks. ACM Transactions on Sensor Networks (TOSN), 11(3), 1-39.

Rathore, P., Kumar, D., Rajasegarar, S., & Palaniswami, M. (2018, February). Bayesian maximum entropy and interacting multiple model based automatic sensor drift detection and correction in an IoT environment. In 2018 IEEE 4th World Forum on Internet of Things (WF-IoT) (pp. 598-603).

Takruri, M., & Challa, S. (2007, July). Drift aware wireless sensor networks. In 2007 10th International Conference on Information Fusion (pp. 1-7).

