Improving sensor-fusion with environmental models

Gonçalo Jesus, Anabela Oliveira, Alberto Azevedo
Department of Hydraulics
LNEC
Lisbon, Portugal
Email: {gjesus, aoliveira, aazevedo}@lne.pt

Abstract— This paper presents an algorithm to improve sensor fusion results in outdoor WSNs using environmental models to redefine periodically which sensors to use and the amount of weight for each particular sensor in the fusion solution. Using daily forecast simulations of the monitored environment dynamics, clusters of sensor nodes sharing data correlation can be defined, for priority sensor selection in a fusion algorithm. The strategy was validated in an operational aquatic sensor network comprised of several sensors scattered over a multiple square kilometers area. It allowed the use of geographically separated nodes with confidence and sometimes even in alternative of closer nodes, unlike most fusion approaches that use quasi-redundant information provided by nearby sensor nodes. This approach represents a major step in the creation of a framework to assess the validity of the monitoring data in WSNs.

Keywords— Data Fusion, Outdoor WSNs, Faulty Sensors, Sensor Quality

I. INTRODUCTION AND MOTIVATION

Sensor fusion algorithms compare a set of observations from different sensors in the same monitoring area to assess what should be the accurate observation and to detect the faulty sensors. When monitoring in harsh conditions or large areas, where users generally prefer to scatter the sensors along the monitored body according to the specific environment dynamics, it is difficult to choose which sensor observations should be considered in the fusion algorithms, especially when sensor nodes are placed hundreds of meters apart [1].

This work is part of a broader study to develop automatic mechanisms to ensure dependable sensor networks data, through data fusion approaches using two types of redundancy:

- analytical redundancy - uses mathematical relationships between measurements to predict a value or to classify a sensor as faulty if a value changes abruptly [2];
- model-based redundancy - using simulation models of the monitored environment one can create virtual sensors and get values to validate real sensor data [3].

This broader study aims to design and develop a framework to assess and assert the quality of sensor outputs, assuming that a sensor may be induced to output an abnormal measurement due to external factors, intrinsic to the monitored environment.

II. RELIABLE MONITORING FRAMEWORK

A. Sensor faults classification and Failure modes

Faults on monitoring networks can be originated at sensor operational or at communication levels. Regardless of the importance of both levels, our main interest is the first one. The goal herein is to identify possible origins of faults in sensors and then characterize into modes related failures. The transducing process, which is the conversion of the various physical effects into electric signals, is one of the most basic

Our case studies are aquatic systems (estuaries). These types of environments offer a high variability of measured parameters such as salinity, temperature or water-quality-related variables, due to the exposure of the sensors to the natural phenomena (tides, waves, wind, river flow) and human-caused events (for instance, urban discharges), becoming difficult to co-relate every disturbance to the respective sensor behavior [4].

To achieve dependability on a aquatic monitoring network, we must use strategies that are either for stopping fault events, or despite the occurrence of one or more faults, to prevent their effect (failures) on the sensor output, simply by inferring or predicting the correct measurement. These strategies are based on automatic mechanisms using data fusion.

While designing and implementing these mechanisms to ensure dependable data, the possibility of using forecasting models to improve the application of analytical algorithms for outlier detection and sensor measurements inference arose. Forecast models are able to predict water levels and related variables with low errors, but we have to take in consideration that forecast models are not always available in every monitoring site or, when available, the models resolution may not be enough to provide low-error predictions. They also depend on a continuous confirmation with monitoring data to not deviate from the reality. So herein, we want to prove that circulation patterns or “heat maps” outputs can help evaluate dynamically the best sensors to use in multi-sensor fusion techniques, and, at the same time, to improve their outcomes.

In section 2 we describe summarily the dependability framework, its architecture, and innovative aspects. Section 3 provides a background on environmental models, more specifically on hydrodynamic forecast, identifying the key aspects to be addressed. Section 4 presents the general concepts of two sensor fusion techniques and presents the case study. The application and results of both techniques is given in Section 5, and Section 6 closes with some considerations.

This work was funded by the Portuguese National Foundation for Science and Technology through project MOLINES (PTDC/AAG-MAA/2811/2012) and through PhD Grant (SFRH/BD/82489/2011)
origins of faults in sensors. Certain sensors struggle to recognize a moving target in dusty conditions whereas others have difficulties obtaining a valid level measurement in liquids. From a dependability perspective it is thus important to distinguish sensors in terms of their operation and robustness to environment conditions. For instance, capacitive sensors are extensively used in various applications, due to their high sensitivity and low power consumption. However, its drawbacks can be expressed in faults: i) the change in capacitance of the sensor due to applied pressure is small compared to the offset capacitance; and ii) their response characteristics are highly nonlinear [5]. Several examples show that sensors vary a lot in their precision and accuracy, tolerance to hardware and external noise, etc., based on their type, robustness and application. Numerous other external and uncontrollable factors may affect the quality of the sensor values, which may lead to inaccurate data.

When a sensor fault event occurs, the faulty measurement is the observable failure. The design of a framework that will ensure dependable measurements starts by setting the fault model of the sensors supported by the experts on the monitored system. The following failure modes were identified [1]: a) constant or offset; b) continuous varying or drifting; c) non-existing or jammed; d) trimming; and e) outliers and noise [7].

B. Framework overview

There are two basic operations to perform a correct measuring process in in-situ measurements: calibration and measurand reconstruction. Regarding calibration, the main problem is that sensors are exposed to the impact of various factors which are absent in laboratory measurements. Thus, the physical measurement elements must be adjusted or even dedicated to the monitored device or process.

The framework under development will perform the reliability work at the post-“sensor measurement” state. In this layer we will track the measurements and provide as outputs corrected values, when necessary, and the observation validity, through a set of assessments and corrections (Fig. 1).

First, the different kinds of faults that may affect the operation of sensors and their effect on the sensor data are identified. Then, this information will be used to enhance the solutions for processing sensor data, possibly allowing the detection of specific faults affecting the data and enabling the application of automatic adjustments to correct the disturbances and increase the data quality and validity. In addition to the detection and correction strategy, fault-tolerance strategies will also be applied through sensor data fusion procedures, exploiting the availability of redundant measurements. Several data processing techniques can be applied to define failures modes and apply adequate correction techniques. For instance, outliers and noise are corrected with signal filtering techniques to improve measurement precision and reduce the noise of the analogic-to-digital components. Correcting the observations without context knowledge on the variables involved is often the most difficult task. Sensor fusion is one of the options to combine context knowledge with other methods.

Introducing more complex data processing techniques and model-based redundancy schemes complements the “abstract sensor” notion [6], needed to assess the confidence in sensor information. This confidence will be delivered by the framework as the validity of the sensor data, which must process continuously the data. Thus data validity can be used to estimate the: usefulness of data; context information encapsulating the individual sensor properties; and the performance of the pre-processing algorithms.

We can also assess data quality from a monitoring network through multi-sensor fusion techniques, described earlier herein. Through data fusion techniques, this framework has the ability to provide virtual redundancy via simulation models and sensor abstraction strategies. This approach will provide an added value input to the data validity construction, as well as an alternative method to single sensor situations.

An innovative aspect is adding aquatic processes knowledge in the validity procedure. This information will feed the model-based redundancy strategies by either i) a validated numerical model that simulates the (hydro)dynamics of the water body and provides forecasts for the monitoring points, or ii) a simplistic pre-determined behavioral model that, instead of forecasts, provides insights on the expected data behavior progression in time, without predicting actual values [1].

Fig. 1 General overview of dependability framework for aquatic sensors
III. ENVIRONMENTAL FORECAST MODELS

Forecasting modeling techniques include simulation, estimation and syntactic methods. Simulation is used when the physical characteristics can be accurately predicted. Most studies are based on terrestrial (indoor) applications whereas the thematic of the work herein concentrates on the complexity of the aquatic environment which has not reached the same level of modelled accuracy. As such, current aquatic systems do not support real-time model-based data fusion. Nevertheless, forecasts represents a reference to validate the sensing data and can also be used for optimization.

In aquatic monitoring networks, forecasting tools are composed mainly by physics-based 3D modelling of the fluid dynamics, computing free-surface elevation and the three-dimensional fields of velocity, salinity and temperature. Current forecasting systems integrate well-established numerical models of riverine, estuarine, and ocean circulation, such as SELFE [7], which is used herein. Currently, we are only using visual model data to extract information to select which are the best sensors, but we can easily develop an algorithm to dynamically choose which sensors to use in the sensor fusion. The proof-of-concept will be provided below.

IV. MULTI-SENSOR FUSION

Multi-sensor fusion uses more than one sensor in a multi-sensor system to enable more accurate or additional data in the observation space of the sensors. With more than one sensor to monitor the whole observation space at all times, it is possible that: i) some sensors measure correctly but others do not; ii) the data of a sensor may be masked with respect to one sensor but not another; iii) one sensor may be blocked or not able to measure, but another sensor located elsewhere in space may be used to infer the correct data. In this case, the data from the correct sensor may be combined with past information from the other sensor to update the overall measurements.

Sensor fusion techniques include comparison, combination and/or smart voting schemes between sensors [3]. We chose for demonstration purposes one analytical algorithm based on statistics and a machine learning technique with artificial neural networks. Both methods are used to: i) detect outliers or anomalous measurements ii) infer on future measurements.

Herein, we used CMOP’s Science and Technology University Research Network (SATURN) monitoring network (http://www.stccmop.org/damart/observation_network), to do the outlier detection and future measurements estimation. SATURN is an interdisciplinary, river-to-shelf observational network with freshwater stations and estuarine and plume stations, measuring everything from tides, salinity and temperature to biogeochemistry on a 24/7 basis. The data collected from SATURN feeds the Virtual Columbia River, a skill-assessed forecast modeling system across river-to-ocean scales. A human-based quality control procedure is available for all datasets, providing a ground-truth for comparison with the automatic quality control proposed herein. Stations JettyA, Lower Sand Island light, Saturn01 and Saturn07 (Fig. 2a) were selected as they have in common several variables (temperature, elevation and salinity) and were measuring at the same period of time and approximate depths. For this first application of the method the variable salinity was chosen, due to its high sensitivity to environmental changes.

A. Statistical approach

We applied a method based on the statistics of differences between sensor measurements to perform automatic event detection and data quality assurance [8]. A normal distribution function to fit the probability density of the differences in each studied sensor was used. The probability distribution is learned during a pre-determined period.

We performed a batch of runs with different periods of the circulation of Columbia, M2 and K1, and larger (Mf) periods. The minimum mean square error was obtained with a period of M2 (12 hours and 25.2 minutes). For predicting a sensor’ next measurement, the most natural estimator is through the mean differences. For outlier and anomalous detection, a simple p-value test was used. If the new difference fails the significance test it is flagged as an outlier or anomaly.

B. Artificial Neural Networks

Artificial neural networks (ANNs) have been applied to wireless sensor networks fault detection [8]. For nonlinear dynamic systems recurrent neural networks (RNNs) are considered to be the best: they have connections between nodes forming a direct cycle to create an internal state of the network that allows it to exhibit dynamic temporal behavior. Both types of neural networks were applied, with almost similar results.

![Image](https://example.com/image.png)

**Fig. 2a)** Columbia River WSN: Jetty A sensor in red, Saturn07 in green, Sand Island sensor in pink and Saturn01 in black. **b)** Sample circulation outputs at Columbia River, in the area of the considered sensors.
For the present test we only considered the feed-forward type, to predict a determined sensor output based on the previous \( n \) measurements of neighboring sensors. In the ANNs technique several batches were done to understand the effect of \( n \) in the outcomes. Results were not conclusive but showed that higher values lead to more complex learning rates and lower values may not capture the system dynamics. An important aspect is the use of M2 tidal constituent as the period to pick the evenly distributed \( n \) previous measurements of the neighbor sensor.

V. RESULTS

We present two examples of the proposed strategy. Firstly, the model outputs from the simulations characterize the area of interest, based on the water circulation of the Columbia river extracted from the forecast model (Fig. 2b). We can then infer which sensors should be used in the fusion processes. The circulation pattern shows that Saturn07 shouldn’t be used in the sensor fusion to infer both neighbor sensors (Jetty A and Sand Island sensors) as circulation cells are distinct.

Then, we tried to predict the measurements of the salinity sensor in Sand Island using the two sensor fusion techniques. In the statistical approach we did this through fusing Sand Island sensor own measurements with ones of the neighbor sensors separately and compared it. Fig. 3 shows the statistical fusion algorithm of both sensors with different mean square error (4.58 vs 19.03), showing that sensors’ distance may not be adequate to choose the most relevant sensors.

We trained the ANN neural networks with 8 months of data, with and without Sand Island measurements as inputs, combined with the other neighbor sensors. Regardless of our strategy assumptions there is one major conclusion: ANNs trained with previous measurements of the same sensor are best because the temporal co-relations overpower spatial co-relations, leading to lower error rates in these runs. We compare outputs of the ANNs trained with Saturn07 and Saturn01 (the more distant of all, Fig. 4). We can observe that the mean square error is quite similar.

VI. CONCLUSIONS

We showed that using environmental forecast models to select the neighbor sensors in sensor fusion techniques produce good results, as part of an on-going development of a framework to support the dependability on aquatic real-time monitoring and forecasting systems. Much remains to be done in evaluating individually the techniques and algorithms to fully understand its effectiveness in already deployed aquatic sensor networks. Future work will also contemplate automation of all methods dynamically in runtime.

ACKNOWLEDGMENT

The authors thank António Baptista and the CMOP SATURN team for their support in the Columbia river analysis.

REFERENCES


Fig. 3 Statistical fusion outputs including Jetty A sensor (left) versus fusion outputs including Saturn07 sensor (right)

Fig. 4 ANN fusion outputs including Saturn07 sensor (left) versus fusion outputs including Saturn01 sensor (right)