proceedings

Proceedings

On-line Digitalization Technologies for Monitoring Activities in the Marine Environment †

Ivan Felis Enguix [1,*], Pablo Ruiz [1] and Manuel de la Torre [1]

[1] Centro Tecnológico Naval y del Mar (CTN), 30320 Fuente Álamo, Murcia, Spain.
* Correspondence: ivanfelis@ctnaval.com; Tel.: +34-968-19-75-21

† Presented at the 6th International Electronic Conference on Sensors and Applications, 15–30 November 2019; Available online: https://ecsa-6.sciforum.net/

Published: 14 November 2019

Abstract: This proceeding shows the results of the investigation of techniques of integration, management and visualization of massive data from the digitalization of environmental and procedural parameters of facilities that operate in the marine environment. The work focuses on three main lines: (1) research on the development of a Cloud-based system for Big Data that allows the hosting of the data generated by the different devices to be monitored (GPS, sounds, vibrations, video, temperature, emissions, consumption, power, etc.); (2) implement a first layer of analysis and visualization of information; and (3) BigData analytics research for post-processing of information. The studies will be applied to underwater noise monitoring. With this, progress is made in another of the pillars of Web 4.0: use of context information, since the application is in charge of intelligently processing the data of the different variables together although they are not, in principle, directly related.

Keywords: Underwater noise; Acoustic monitoring; MSFD; Data compression

1. Introduction

The generation of new knowledge, related to the influence of human activity on the biodiversity of marine ecosystems and the exploitation of its resources, is essential to move towards the planning, management and sustainable exploitation of the seas and coasts. In addition, the lack of information on the different elements that characterize the marine environment and its interaction with activities of anthropogenic origin is presented as an opportunity to investigate and improve the state of knowledge of this environment.

However, the large volume of data generated in the field of planning, management and exploitation of natural resources located in the marine environment, requires technologies capable of capturing, storing, processing, analyzing, distributing and displaying such information quickly and efficient, contributing to the efficient management of the seas and oceans and the ecosystems dependent on them, avoiding the numerous and costly conflicts arising from poor data management in the face of complex decision-making processes.

Therefore, the integration of new programming techniques that allow the integration of advanced signal processing into a server will improve the knowledge of these ecosystems by society thus contributing to awareness, the first step to generate responsible behaviors that result in better conservation of these natural spaces.

In this context, the general objective of this work is the development of new programming techniques that allow the integration of advanced signal processing (machine learning among others) automatically into a web server. This study is focused in three lines:

1. Cloud-based system for Big Data environmental monitoring. There are discussed the key aspects of the backend web technologies that allow the management of large amounts of data
2. Layer of processing and on-line display of information. The processing of the necessary data for its on-line visualization will be explained, as well as the technologies that will allow an adaptable and agile representation.
3. Machine-Learning for Big Data. They are explained the Machine Learning algorithms for the treatment of massive data, which are integrated into a GIS platform. Tus, it is shown the results of the geospatial analysis with these algorithms.

In this document, it is shown the results of applying these different techniques and technologies to the case of underwater noise monitoring, one of the environmental indicators that has great influence in the maritime field. Indeed, unlike the other indicators of environmental impact (contaminants, temperature, etc.), underwater noise perseveres the critical aspect of generating large amounts of data in short time periods, since its acquisition is carried out at frequencies of the order of 10 to 100 thousand samples per second. Therefore, the results shown may apply to the case of on-line monitoring of other less demanding environmental parameters.

2. Methods

2.1. Cloud-Based System for Big Data Environmental Monitoring

The architecture chosen for the cloud-based system for environmental monitoring of Big Data follows the principles of the Lambda architecture [1]. This is adequate because it combines real-time information (transmission type processing) with heavy processing information (batch type processing). The objective of the chosen architecture is to have a robust fault-tolerant system, both human and hardware, that is linearly scalable and that allows writing and reading with low latency. A typical Lambda architecture scheme is as follows:

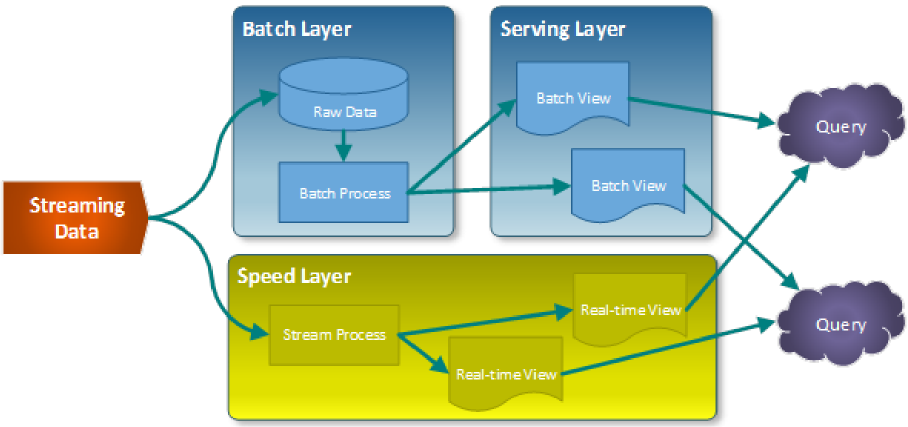

Figure 1. Typical Lambda architecture.

As seen in the scheme, it has three distinct layers:

1. The Batch Layer where the raw information is managed.
2. The Serving Layer where the information to be presented in the queries is prepared and has as its main feature the low latency.
3. The Speed Layer where only the latest information is used to provide real-time information.

From the general scheme, for the case at hand, an adaptation has been made using web technologies such as Django, MongoDB and VUE, remaining as shown in the following scheme:

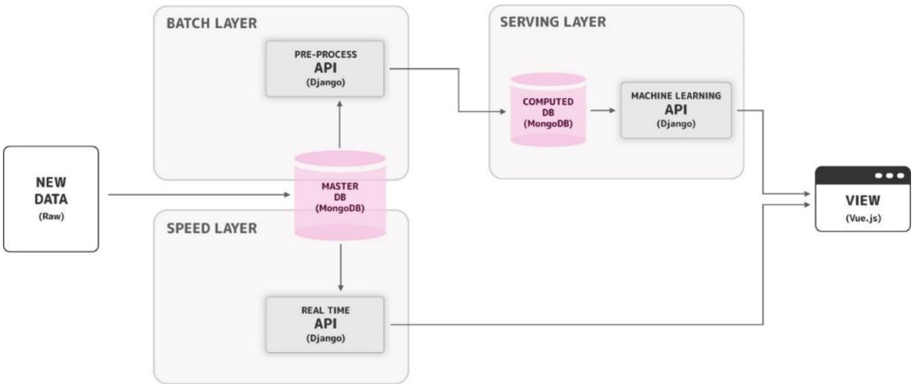

Figure 2. Adapted Lambda Architecture for monitoring activities in the marine environment.

In our case, the starting point is the master database where the raw data is stored. Here, an immutable data set is built, only for the annexation of raw data.

As Batch Layer, we have a Pre-Processed API that makes a first basic data treatment to select the most relevant data according to our objective. From there, the Serving Layer is responsible for executing the machine learning algorithms required by the View, always leaving the last calculated available, thus ensuring low latency. On the other hand, the architecture has a Speed Layer that, when required by the view, reads only the last data received and makes a fast processing to show them in real time. In this way, View combines Real-Time data with the heaviest data from Machine Learning.

2.2. Layer of On-Line Processing And Display of Information

When displaying the information, we have chosen to use a single-page application (SPA) in order to offer a fluid experience to users, just as a desktop application does. In this sense, all HTML, JavaScript and CSS codes are dynamically loaded according to the User's interactions on the page.

To achieve this implementation, the JavaScript-based FronEnd framework VUE.js [2] was used. Its main advantages over others of the same style are: its progressive adaptation as the complexity of the application increases; intuitive functionality, modern and easy to use; varied ecosystem that covers everything you need; a very active community; and very well componentized.

For the representation of the graphics of the application, the use of the Bokeh [3] library was chosen. This graphics display library provides us with information presentation in a modern, elegant and concise way. It is also very versatile, but above all, the critical aspect for what has been chosen is that it offers interactivity with the user with high performance over very large or streaming datasets.

Since underwater noise is usually recorded with a high sampling rate (of the order of 10^4 to 10^5 samples per second) compared to other maritime data (waves, temperature, pH, electromagnetic fields, etc.), the resulting log files have a large amount of data. In the first instance, for a better understanding of acoustic signals, processing techniques are used. These techniques can be applied to the entire signal as well as to temporary windows, so that they reduce the data to be displayed at the user's request, allowing it to adapt to the processing capabilities of the cloud server. These processing techniques have been classified in two groups:

- On the one hand, the application implements calculations of statistical parameters of end and centralization in the time domain, such as the peak-to-peak value or the rms value; In addition, it allows obtaining the spectrum of the signal in the frequency domain, as well as the estimation of the SPL [dB re 1 μPa] in different frequency bands and their corresponding percentiles. As will be seen, the application of this processing to different temporary windows of the signal allows to evaluate the stationarity of the measured parameters.
- On the other hand, the application allows the calculation of the D11C2 of the MSFD[1], associated with the quantification of the continuous noise in the third octave bands centered at 63 and 125

[1] Descriptor 11 Criteria 2 is the indicator for continuous underwater noise pollution in Marine Strategy Framework Directive.

Hz. It considers the different calculation parameters that can be extracted from different methodological guides and expert groups [4,5].

2.3. Machine-Learning for Big Data

Machine Learning (ML) techniques constitute a breakthrough in the field of analysis and classification of large data sets [6]. The main objective of this work was to incorporate the position variable (geographic coordinates) as an additional feature to the characteristics of the noise signal recorded by the application. This means that ML algorithms consider the coordinates in which noise measurements were made as one more variable in the classification process. The main idea is to improve the results in the assignment of groups by having a spatial variable since the place where a noise occurs can be defining depending on external agents, such as, for example, a channel where large numbers of ships pass, the docking area of the port, open sea area, etc. To do this, GIS tools have been used for grouping by position with supervised classification algorithms.

In this application, cluster analysis has been used using the ML algorithm called k-nearest neighbors algorithm (k-NN). As it is a supervised classifier, the estimation of the classes of the samples is carried out through training samples, that is, a grouping of data that has similar characteristics to each other, and that have an associated class known to the operator. The determination of the training samples modifies the result of the classification, so it is important that these samples are correctly defined [7].

Basically the algorithm works as follows: when it is necessary to establish a classification on a set of data, the algorithm k closest neighbors searches among the training samples for the instances closest to the sample that is intended to establish class value, up to a User defined value of number of neighbors. The most repeated class of the closest instances obtained will be the class awarded by the sample to be classified. Therefore, entering a different number of neighbors value in each classification will modify the results. Usually, the higher the number of neighbors value, the better the classification accuracy.

3. Results

3.1. Prototype Application

The built application has three main sections:

1. Home, which the objective and context of the Project is shown.
2. Processing. It allows to generate different information processes both in time and frequency by selecting the desired time and frequency intervals.
3. GIS. It allows you to perform advanced Machine Learning calculations with ESRI's own techniques on a map. An example of this section where a map focused on the Port of Cartagena with different measurement locations shown submarine noise shown.

Some of the images of the application are shown:

Figure 3. Home section of prototype application.

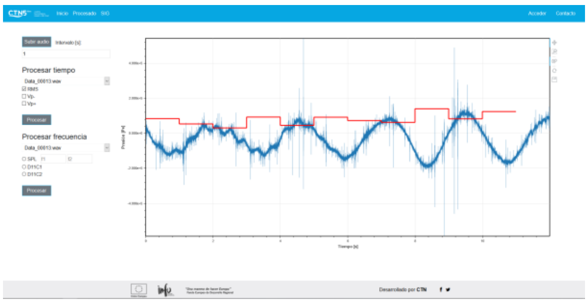

Figure 4. Example of processing section of prototype application. Time Analysis.

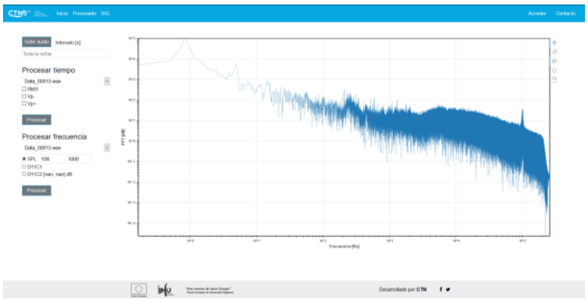

Figure 5. Example of processing section of prototype application. Frequency analysis.

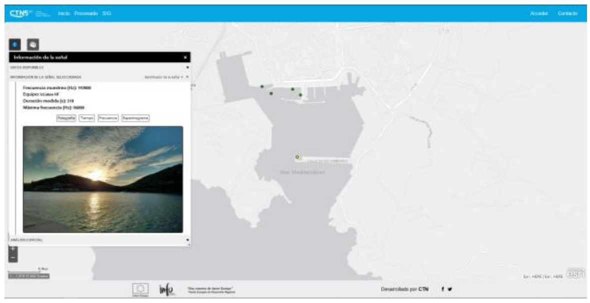

Figure 6. Example of GIS section of prototype application.

3.2. Processing Test

To measure the performance of the application developed in Django with Python, measurements have been made of the time required for the execution of each of the processes developed for a series of signals with different time duration, both stationary, for the entire signal, and non-stationary by setting time intervals.

The following figure shows the processing times of different parameters depending on the duration of the processed signal. In absolute terms (left), the calculation time of the FFT is much longer than the other processing, while the calculation of D11C2 and SPL are around an order of magnitude below, and the RMS two orders below. In relative terms (right), we can see that the RMS, SPL and D11C1 processes have approximately the independent duration of the original signal, even decreasing with it for the RMS.

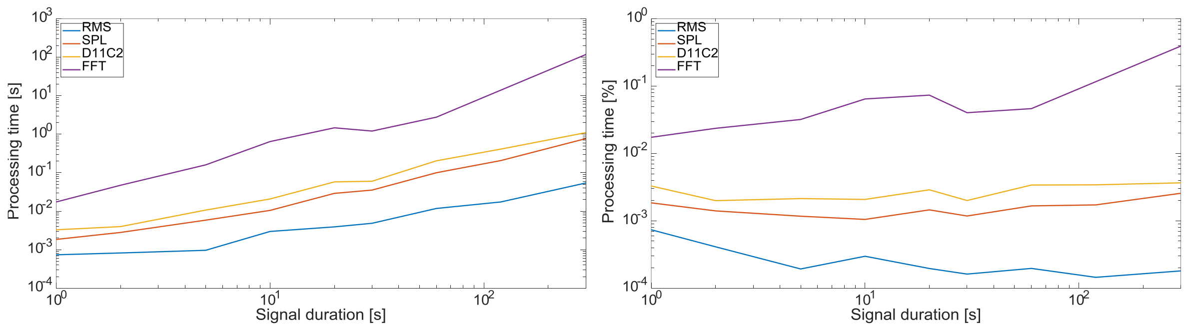

Figure 7. Computational cost of the processing of different parameters depending on the duration of the signal. On the left, calculation time. On the right, percentage of said time with respect to the duration of the signal.

To delve deeper into the calculation of the parameters under study, the processing of the same signals as above but cut at different time intervals (0.05, 0.1, 0.5 and 1 second) was tested. The results are shown in the following figure.

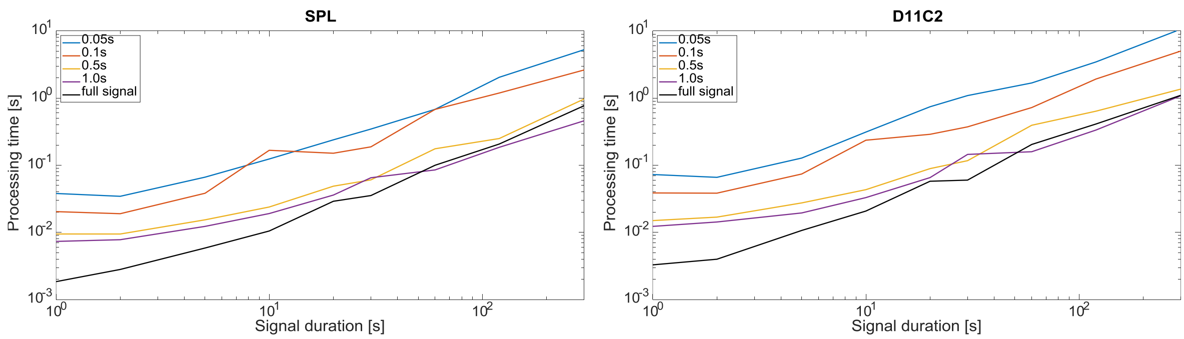

Figure 8. processing cost for several cuts of the original signal. On the left, SPL calculation. On the right, calculation of D11C2.

It is observed that the processing time does not improve, except for long durations of the original signal, where it is noted that for certain intervals the computational cost can improve. Although the time periods studied are quite short, their influence is important for the qualification of D11C1 and its influence on the resulting percentiles and values [8].

3.3. Machine Learning Test

The result of grouping without adding position variables can easily be seen in the following images. The results obtained with the use of the k-NN algorithm taking into account the position variable shows a more defined spatial differentiation (right image) than in the algorithm that only takes into account the characteristics of the signal (left image). The context of the example is the proximity of the Port of Cartagena (Spain).

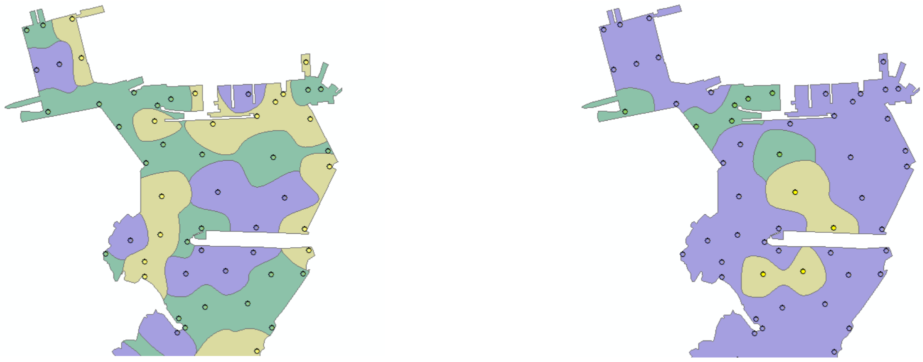

Figure 9. Example of application of the K-NN algorithm without considering adding position variables (left) and considering the position of the measuring points (right).

4. Conclusions

The implementation of a cloud-based architecture for the recording, processing and spatial analysis of underwater noise signals has been tested. Although the implemented application is a preliminary version with a view to R&D, they have been able to run with real signals and in real contexts.

Some results show processing techniques based on temporary windows that could reduce calculation times and adapt to the most current D11C2 definitions. In addition, although the results of applying ML techniques taking into account the position of the signals are preliminary, they show fairly consistent results that can be exploited in future studies.

Author Contributions: conceptualization, I.Felis. and P.Ruiz.; methodology, I.Felis. and P.Ruiz.; software, P.Ruiz and M.de la Torre.; formal analysis and data curation, I.Felis.; writing—review and editing, I.Felis, P.Ruiz and M. de la Torre.

Funding: This research was funded by the Instituto de Fomento de la Región de Murcia (INFO) under the Program of grants aimed at Technological Centers of the Region of Murcia for the realization of non-economic R&D activities. Modality 1: Independent R&D Projects, with File No.: 2017.08.CT01.0043 and 2019.08.CT01.0037.

Acknowledgments: Thanks to the Port Authority of Cartagena (APC), for its support and collaboration for the realization of measurements that will allow the implementation of new techniques and technologies for the sustainable ports of the future.

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

References

1. Hausenblas, M.; Bijnens, Y.N. Lambda architecture. Available online: http://lambda-architecture.net/(accessed on 11 November 2019).
2. MIT, Vue JS. Available online: https://vuejs.org/.(accessed on 11 November 2019).

3. Anaconda and Bokeh Contributors, Bokeh. Available online: https://docs.bokeh.org/en/latest/.(accessed on 11 November 2019).
4. Dekeling, R.P.A., Tasker, M.L., Van der Graaf, A.J., Ainslie, M.A, Andersson, M.H., André, M., Borsani, J.F., Brensing, K., Castellote, M., Cronin, D., Dalen, J., Folegot, T., Leaper, R., Pajala, J., Redman, P., Robinson, S.P., Sigray, P., Sutton, G., Thomsen, F., Werner, S., Wittekind, D., Young, J.V., *Monitoring Guidance for Underwater Noise in European Seas, Part II: Monitoring Guidance Specifications*, JRC Scientific and Policy Report EUR 26555.
5. Dekeling, R.P.A., Tasker, M.L., Van der Graaf, A.J., Ainslie, M.A, Andersson, M.H., André, M., Borsani, J.F., Brensing, K., Castellote, M., Cronin, D., Dalen, J., Folegot, T., Leaper, R., Pajala, J., Redman, P., Robinson, S.P., Sigray, P., Sutton, G., Thomsen, F., Werner, S., Wittekind, D., Young, J.V., *Monitoring Guidance for Underwater Noise in European Seas, Part III: Background Information and Annexes*, JRC Scientific and Policy Report EUR 26556 EN, Publications Office of the European Union, Luxembourg, 2014, doi: 10.2788/2808.
6. Arcila-Calderón, C.; Barbosa-Caro, E. and Cabezuelo-Lorenzo, F. *Técnicas big data: análisis de textos a gran escala para la investigación científica y periodística*, *El Prof. la Inf.*, **2016**, *25*, 4, 623,
7. A.Burkov. *The hundred-page machine learning book*. 2018.
8. Merchant, N.D., Barton, T.R., Thompson, P.M., Pirotta, E., Dakin, D.T. & Dorocicz, J. (2013). *Spectral probability density as a tool for ambient noise analysis*. *J. Acoust. Soc. Am.*,**2013**, *133*, EL262-7.
9. I.Felis, et al. *Acoustic characterization of impulsive underwater noise present in port facilities. Practical case of the Port of Cart*agena. Proceedings. 2018.
10. R.J.Urick. *Principles of underwater sound*. 3th edition. McGraw-Hill, Inc. Los Altos Hills, 1983.
11. P.Ruiz, et al. *Sound impacto of Mutriku oscillating water column plant: Unerstanding MSFD Descriptor 11*. XIV Jornadas españolas de ingeniería de costas y puertos. Alicante, 24 y 25 de Mayo de 2017.
12. MSFD calibration guidelines. Available online: http://www.quietmed-project.eu/wp-content/uploads/2019/01/QUIETMED_D3.1.-Best_practices_guidelines_on_sensor_calibration_final.pdf (accessed on 11 November 2019).
13. MSFD signal processing guidelines. Available online: http://www.quietmed-project.eu/wp-content/uploads/2019/01/QUIETMED_D3.2_Best_practices-guidelines_on_signal_processing_final.pdf (accessed on 11 November 2019).
14. MSFD acoustic modelling processing guidelines. Available online: http://www.quietmed-project.eu/wp-content/uploads/2019/01/QUIETMED_D3.3_Best-practices-guidelines-on-acoustic-modelling-and-mapping_final.pdf (accessed on 11 November 2019).