

Building Suitable Datasets for Soft Computing and Machine Learning Techniques from Meteorological Data Integration: A Case Study for Predicting Significant Wave Height and Energy Flux

Energies

Paper reference: energies-1023087

Antonio Manuel Gómez-Orellana¹, Juan Carlos Fernández¹, Manuel Dorado-Moreno¹,
Pedro Antonio Gutiérrez¹ and César Hervás-Martínez¹.

¹Department of Computer Science and Numerical Analysis, University of Cordoba, 14071,
Cordoba, Spain.

✉ am.gomez@uco.es, jfcaballero@uco.es, manuel.dorado@uco.es, pagutierrez@uco.es,
chervas@uco.es

Contents

1	Responses to the Editor	1
2	Responses to reviewers' comments	2
3	Reviewer #1	2
4	Reviewer #2	4
5	Reviewer #3	13

1 Responses to the Editor

We would like to thank the Editor for giving us the opportunity to improve our paper with this review and resubmission. Following all indications, we have prepared a revised version of the paper, where we have addressed the points required by the reviewers to obtain an improved version of the manuscript.

Please find below the responses to the reviewers' comments.

2 Responses to reviewers' comments

We sincerely thank the anonymous reviewers for their careful reviews, constructive comments and hints to improve the impact of this work. We honestly believe that the quality of the paper has increased after addressing the comments and issues raised by the reviewers.

Most important changes in the paper have been highlighted in blue color in the revised version of the manuscript. Please, find below detailed responses to the comments.

Additionally, in order to improve the English quality and style we have carried out a proof-reading.

3 Reviewer #1

Comment 1

This work develops a framework to collect, integrate, and preprocess meteorological observation data from NDBC and NNRP. Moreover, this framework uses machine learning techniques to do predictions based on these pre-processed datasets. This work saves researchers from tedious and repetitive data collection and pre-processing work. Also, the use of machine learning in this framework is very useful given that NDBC and NNRP contain missing data in their observation datasets.

My only comment is that the wave prediction from the machine learning technique is not well validated.

Response: Thank you very much for your review and for your comment, which we try to clarify below.

This work does not focus on carrying out an extensive experimentation and validation of results obtained by different ML and SC techniques, which should be combined with statistical tests to check if significant differences exist. On the contrary, this work presents a very versatile software tool that facilitates the automated building of meteorological datasets from which researchers can perform prediction studies and apply them to different tasks related to coastal and ocean engineering and to marine energy prediction.

Nevertheless, although it is not the primary objective, this software tool also allows the use of ML and SC methods available in WEKA from SPAMDA itself or from an independent WEKA instance. In this way, researchers can carry out a preliminary study or even a complete study on the prediction problem which they are working on. Moreover, SPAMDA also gives the possibility to save the integrated datasets in .csv format, and the researchers could use these datasets with other ML and SC tools that they consider appropriate.

In any case, to further study the validation as requested by the reviewer, **we have extended Section "4. A case study applied to Gulf of Alaska" with a short-term forecasting problem**

of the energy flux generated by the waves in a specific location of that zone. Besides, **Sections "3.4. Matching configuration"** and **"3.5. Final datasets"** have been also rewritten for this purpose. In this way, it would be possible to perform prediction studies 6h in advance (note that the prediction horizon is customisable) of the amount of marine energy that could be extracted by the WECs. Given that this work does not focus on proposing new models or methodologies, a more extended validation or a comparison study of the results obtained in such example does not make sense. However, typical performance metrics as CCR (Accuracy), Kappa, RMSE (Root Mean Squared Error) or Correlation coefficient are shown.

We would like to thank you again for the suggestion and review carried out to our work.

4 Reviewer #2

Comment 1

Dear authors, Thank you for submitting your paper to the Energies. I think your paper is an informative paper that can be published after a major revision.

It is not clear what the research gap that the paper is addressing. What is the objective of this paper? Please clarify somewhere clearly all your contributions.

Response: First of all, we would like to thank Reviewer #2 for his/her suggestions to improve the paper.

The reviewer is right about this comment, in the original version, we did not adequately indicated the purpose of this work. We think that this work can be useful as a support for applications in the fields of marine energy, engineering and environment studies, by creating well-constructed datasets ready to be used with ML and SC prediction techniques. Datasets can be used both in classification and regression problems based on the integration of NDBC observations and NNRP reanalysis data.

From our experience, manuscripts in marine energy and engineering that use ML and SC algorithms for prediction tasks do not provide software tools for incorporating and integrating the two data sources, considering the casuistry involved. These studies apply specific algorithms (extreme learning machine, metaheuristics, Bayesian networks, artificial neural networks, etc.), normally applying custom-made implementations or scripts in different programming languages. We have not found tools for building datasets ready to be used with any of those methodologies or algorithms by an automated and very versatile process. In this sense, it is convenient to indicate that the objective of these works is different than that of our manuscript. Although our software allows the use of ML and SC algorithms available in WEKA, datasets can also be exported to .csv format to researchers can use other tools apart from WEKA, and we neither focus on those methodologies nor on improving them.

Some of these clarifications have been included in **Section "1. Introduction" (lines 112-129 - "The main purpose ... tools")**, and others have been added to **"Abstract" (lines 10-19)** and **Section "5. Conclusions" (lines 759-766 - "The aim ... regression", lines 781-782 - "Given that ... carried out")**, so that the reader can better understand the scope of our work. **Comments 4 and 9** suggested by Reviewer #2 are also related to these clarifications and have been included in **"Abstract"** and **Section "5. Conclusions"**.

At the end of **Section "1. Introduction" (lines 133-168 - "The generation ... system")**, we have tried to summarize in a general way the contributions of SPAMDA, which are specified and detailed throughout the paper. Below, we comment on some contributions that have also been taken into account in this revised version of the manuscript (**highlighted in blue color**):

- SPAMDA allows preliminary studies of missing values (dates or measurements not

recorded) in buoys managed by NDBC, so that researchers have an idea of the quality of data observed by buoys and whether these would be suitable for the work to be carried out. Sometimes, buoys do not collect observations temporarily due to mechanical problems or weather conditions, so, depending on the problem to be solved, such raw data could not be appropriate for accurate modelling. Even so, SPAMDA allows data integration taking into account such missing values when needed by researchers. More details about how this task is performed can be found in **Section “2. Meteorological data sources”** (lines 208-215 - “**Note ...Appendix A.**”), **Section “3.2. Datasets”** (lines 283-286 - “**When an ... dates.**”), **Section “3.3. Pre-process”** (lines 315-328 - “**Recover ...values mean.**”), **Section “3.4. Matching configuration”** (lines 398-403 - “**Include missing ...as «?»**”), and **Section “Appendix A. Managing the casuistry of incomplete data”**.

- Although pre-processing is not the main objective of SPAMDA, the tool also provides some basic pre-processing filters on buoys measurements, such as normalisation and missing data recovery (see **Section “3.3. Pre-process”** (lines 323-328 - “**Replace missing...values mean**”) for more details).
- Building as many different datasets from the same meteorological data as needed, through a customisable matching process with respect to the input variables from NNRP or the output variable from NDBC, and ready to be used in regression and classification prediction tasks (more details in **Section “3.4. Matching configuration”** and **Section “3.5. Final datasets”**).
- The possibility of using several reanalysis nodes near to the NDBC buoy under study, which could provide a better description of the problem to achieve more accurate models (more details can be found in **Section “3.4. Matching configuration”** (lines 404-417 - “**Nearest reanalysis ...respectively.**”)).
- Estimation of the amount of energy flux that can be produced at different prediction horizons, for example, 6h, 12h, 18h, 24h, ... in advance (more details in **Section “3.5. Final datasets”** (lines 463-502 - “**Prediction horizon ... ($t + \Delta t$ in Eq. 5)**”). Although this work does not focus on model performance, it should be taken into account that models tend to generalise worse with greater prediction horizons.
- This software would be provided to the scientific community as a free and open source tool, so that any researcher could access its source code and extend its functionality (the github repository is shown in **Section “Additional material”**).

From an applicability point of view, SPAMDA could be used in coastal and ocean engineering applications and marine energy prediction, such as:

- Estimation of the energy supply sources in hybrid energy systems [3], based on the amount of energy that can be obtained by a marine energy system in the next 6h or 12h, for example.
- Regulation of WECs to avoid malfunction or breakage [27,28], depending on the significant wave height and/or energy flux expected, as well as the possibility of reconfiguring them in order to maximise the wave energy extraction.
- Decision-making in exploitation and environmental protection for the construction of

marine energy storage plants, future strategies on renewable energy and coastal planning [52].

- Support tool for the prediction of the energy that could be obtained from a certain maritime location, to know whether or not it is convenient to install WECs as power supply in marine structures, such as offshore oil and gas platforms or seawater desalination plants [24, 25].
- Design of offshore structures and ports [53].
- Decision-making and risk assessment about operational works in the sea [54], security systems for structures or naval security [55].

Some of these clarifications about the applicability of SPAMDA have been included in **Section "1. Introduction"** (lines 169-182 - "Therefore ... security").

Comment 2

The literature review is not goal-oriented. The process should be as follows:

i) Critical evaluation of the literature; ii) identifying the gap based on this critical evaluation of the literature; iii) proposing your hypothesis to address the identified gap; iv) posing the appropriate and relevant research question based on your proposed hypothesis, and finally explaining your proposed method to answer this research question. Therefore, you will have a systematic way of conducting your research. Right now, the literature review section has no clear objective.

Response: Thank you very much for questioning this fact and allowing us to clarify it.

Although we have cited several software tools related to energy prediction and environmental modelling, as well as works that use ML and SC methodologies in marine energy studies (now much more complete with the references provided by Reviewer #2 in **Comment 7**), we have not made a critical evaluation of the state-of-the-art because, to our knowledge, there are no open source software tools available to the scientific community, that perform the data integration provided by SPAMDA with the functionalities described in this work.

The reading order of the contents shown in **Section "1. Introduction"** has been carried out with the intention that is commented below, and the background has been reinforced with the interesting references suggested by Reviewer #2 (lines 37-39, 66, 74, 81, 86-94):

1. Firstly, readers are introduced to the importance of having quality and well-formed meteorological data and observations to perform studies of environmental comprehension, conservation, modelling and prediction, providing examples of this. It is also indicated that ML and SC methodologies, despite managing uncertainty, need adequate data to obtain accurate models (lines 23-43 - "A better ... quality models.").
2. Secondly, it is explained that researchers need special purpose software to be able to perform data adaptation tasks and to carry out studies related to energy and environmental modelling (lines 44-59 - "Continuing with ... solar tracker").

3. Thirdly, marine energy is introduced as one of the most important renewable and sustainable energy sources today, where meteorological data is used for prediction purposes. WECs are also introduced as the main mechanical devices for capturing and converting energy from waves. It is concluded that for both the correct design and use of WECs and for other applications related to marine energy, it is very useful to have accurate prediction models, for which it is necessary to have both data and techniques that use them (lines 60-77 - "**Marine ...wave energy period (T_e)**").
4. Fourthly, it is explained that ML and SC methodologies are supporting traditional statistical modelling procedures and that there are tools, such as Weka, that can be used even by researchers who are not specialized in the field of artificial intelligence or in the use of programming languages. Some studies, related to marine energy, using ML algorithms for wave height and energy flux prediction, are also mentioned. In this part, it is emphasized that datasets are essential both for these tools that offer an user-friendly interface for researchers not specialized in ML and for more specific works, and that, usually, they are not publicly available in an appropriate format. This causes them to present inconsistencies that prevent their treatment, and a great deal of effort and time is spent on solving them, which the software tool developed avoids (lines 78-111 - "**Currently ... tedious task**").
5. Next, it is indicated that the main purpose of this work is to solve in an automatic, versatile and customisable way, the integration of observations from NDBC and reanalysis data from NNRP to build datasets (address the gap), without focusing on the ML and SC methodologies that use them. To do that, a software tool has been developed that manages the casuistry involved in such data integration for marine energy prediction tasks (lines 112-129 - "**The main ... tools**").
6. Finally, the functionalities and the general contributions of the developed tool are described, as well as some examples of applicability (lines 130-168 - "**In order to ... system**", 169-182 - "**Therefore ... security [55]**"). The remaining sections of the manuscript explain in detail the contributions of SPAMDA and how they are carried out, ending with a case study in the Gulf of Alaska tackling significant wave height and energy flux prediction.

Comment 3

A thorough editorial check and English improvement are needed. Please kindly proofread the entire manuscript.

Response: Thanks for your suggestion, a revision of the English language has been carried out in this version of the manuscript.

Comment 4

The conclusion part is also needed to be revised; which questions are answered, what is the value/originality/contribution of the paper, how the proposed method answers the research questions that previous methods are not able to answer?

Response: Following the instructions of the reviewer, we have modified **Section “1. Introduction”** and **Section “5. Conclusions”**, so that readers will better understand the contribution of the developed tool.

Comment 5

It feels you need a king of aggregating results somewhere clearer.

Response: We would like to thank the reviewer for her/his suggestion. As this work is not a study that presents a prediction methodology related to marine energy, and whose results must be compared with other methodologies in the state-of-the-art, there is no a Section “Results”.

Nevertheless, a case study is carried out to predict significant wave height and energy flux in the Gulf of Alaska through data integration from NDBC and NNRP using SPAMDA. Besides, **we have extended the case study** of the initial version of this paper with experimentation about energy flux prediction that reinforces the usefulness of this software, without comparing the results achieved by the different algorithms applied, because this is not the purpose of this work. Only typical performance metrics as CCR (Accuracy), Kappa, RMSE (Root Mean Squared Error) or Correlation coefficient are shown.

Such experimental extension has been included in **Section “4. A case study applied to Gulf of Alaska”**. Besides, **Sections “3.4. Matching configuration”** and **“3.5. Final datasets”** **have been rewritten** for this purpose trying to add as few figures as possible and unifying existing ones, as the Reviewer #2 also suggests in **Comment 6**.

Comment 6

Please provide more explanation for Figure 21. Also, there are too many figures in the paper. Please reconsider combining some of them together.

Response: Many thanks for the suggestion. We have clarified the explanation of **Figure 21, now Figure 20 (lines 642-657 - “After configuring ... ARFF”)**. By extending the case study the number of figures has increased, but **we have combined Figures 8 and 9**.

Comment 7

An adequate literature review and a clear gap identification have been tried to be conducted. However, authors have ignored some research that has been done in the area. I strongly recommend the authors to provide a more comprehensive literature review in the introduction section. The following papers are recommended:

- Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renewable and Sustainable Energy Reviews*, 104, 281-295.
- A wavelet-Particle swarm optimization-Extreme learning machine hybrid modeling for significant wave height prediction. *Ocean Engineering*, 213, 107777.
- Managing computational complexity using surrogate models: a critical review. *Research in Engineering Design*, 31(3), 275-298.
- Near real-time significant wave height forecasting with hybridized multiple linear regression algorithms. *Renewable and Sustainable Energy Reviews*, 132, 110003.
- Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm–Extreme Learning Machine approach. *Renewable Energy*, 97, 380-389.
- Outlook on biofuels in future studies: A systematic literature review. *Renewable and Sustainable Energy Reviews*, 134, 110326.
- Statistical models for improving significant wave height predictions in offshore operations. *Ocean Engineering*, 206, 107249.
- Regional ocean wave height prediction using sequential learning neural networks. *Ocean Engineering*, 129, 605-612.
- Ensemble of surrogates and cross-validation for rapid and accurate predictions using small data sets. *AI EDAM*, 33(4), 484-501.
- Prediction of significant wave height; comparison between nested grid numerical model, and machine learning models of artificial neural networks, extreme learning and support vector machines. *Engineering Applications of Computational Fluid Mechanics*, 14(1), 805-817.

Response: We have taken into account and included all of them in **Section “1. Introduction” (lines 37-39, 66, 74, 81, 86-94)** of this new revised version of the manuscript. The following one was already in the original version:

[40] Kumar, N.K.; Savitha, R.; Al Mamun, A. "Regional ocean wave height prediction using sequential learning neural networks". *Ocean Engineering*, vol. 129, pp. 605–612, 2017. doi:10.1016/j.oceaneng.2016.10.033.

Comment 8

If you can, please make a small comparison between what did you do and what others did before, as a conclusion.

Response: Many thanks for the appreciation. To our knowledge, we can not make a comparison with other applications as there are no other software tools with the functionality and characteristics presented in this work. There may be console programs or scripts that use NDBC or NNRP data to perform some type of application on data, but we have not found in the literature any open source software available to the scientific community that performs the data integration process described in this paper. Moreover, we have not found any software that exhibits the functionalities and configurations available in SPAMDA, not only for the data integration but also to manage buoys and reanalysis data in a simple and versatile way, creating datasets in a proper format to be easily used with ML and SC algorithms.

Comment 9

The abstract is not deep enough and is not well prepared. Please try to re-write it better. The problem should be clearly stated and the gap in which you are going to address the need to be clarified. Simply explain your contributions and key findings.

Response: Thanks for this comment, we have rewritten the “**Abstract**” (lines 10-19), trying not to exceed the allowed size. We hope that, with this version of the abstract, readers can better understand what this work offers.

Comment 10

There are some errors in your reference list. Please check and fix the errors.

Response: The reviewer is completely right, we have revised the bibliography and corrected the errors detected. Please, excuse us if there is an omission in any type of character or compound name that could be due to the latex bibliographic format established by the journal.

These are the references that have been modified:

[3] Shivam, K.; Tzou, J.C.; Wu, S.C. Multi-Objective Sizing Optimization of a Grid-Connected Solar–Wind Hybrid System Using Climate Classification: A Case Study of Four Locations in Southern Taiwan. *Energies* 2020, 13, 2505. doi:10.3390/en13102505

[4] Dorado-Moreno, M.; Cornejo-Bueno, L.; Gutiérrez, P.A.; Prieto, L.; Hervás-Martínez, C.; Salcedo-Sanz, S. Robust estimation of wind power ramp events with reservoir computing. *Renewable Energy* 2017, 111, 428–437. doi:10.1016/j.renene.2017.04.016

[5] He, Q.; Zha, C.; Song, W.; Hao, Z.; Du, Y.; Liotta, A.; Perra, C. Improved Particle Swarm Optimization for Sea Surface Temperature Prediction. *Energies* 2020, 13, 1369.

doi:10.3390/en13061369

[7] da Silva, V.d.P.R.; Araújo e Silva, R.; Cavalcanti, E.P.; Braga Campos, C.; Vieira de Azevedo, P.; Singh, V.P.; Rodrigues Pereira, E.R. Trends in solar radiation in NCEP/NCAR database and measurements in northeastern Brazil. *Solar Energy* 2010, 84, 1852-1862.

doi:10.1016/j.solener.2010.07.011

[11] Dhanraj Bokde, N.; Mundher Yaseen, Z.; Bruun Andersen, G. ForecastTB-An R Package as a Test-Bench for Time Series Forecasting-Application of Wind Speed and Solar Radiation Modeling. *Energies* 2020, 13, 2578. doi:10.3390/en13102578

[14] Di Bari, R.; Horn, R.; Nienborg, B.; Klinker, F.; Kieseritzky, E.; Pawelz, F. The Environmental Potential of Phase Change Materials in Building Applications. A Multiple Case Investigation Based on Life Cycle Assessment and Building Simulation. *Energies* 2020, 13, 3045. doi:10.3390/en13123045

[15] Astiaso García, D.; Bruschi, D. A risk assessment tool for improving safety standards and emergency management in Italian onshore wind farms. *Sustainable Energy Technologies and Assessments* 2016, 18, 48-58. doi:10.1016/j.seta.2016.09.009

[16] Raabe, A.L.A.; Klein, A.H.d.F.; González, M.; Medina, R. MEPBAY and SMC: Software tools to support different operational levels of headland-bay beach in coastal engineering projects. *Coastal Engineering* 2010, 57, 213-226. doi:10.1016/j.coastaleng.2009.10.008

[17] Motahhir, S.; EL Hammoumi, A.; EL Ghzizal, A.; Derouich, A. Open hardware/software test bench for solar tracker with virtual instrumentation. *Sustainable Energy Technologies and Assessments* 2019, 31, 9-16. doi:10.1016/j.seta.2018.11.003

[18] Cascajo, R.; García, E.; Quiles, E.; Correcher, A.; Morant, F. Integration of Marine Wave Energy Converters into Seaports: A Case Study in the Port of Valencia. *Energies* 2019, 12, 787. doi:10.3390/en12050787

[20] De Jong, M.; Hoppe, T.; Noori, N. City Branding, Sustainable Urban Development and the Rentier State. How do Qatar, Abu Dhabi and Dubai present Themselves in the Age of Post Oil and Global Warming? *Energies* 2019, 12, 1657. doi:10.3390/en12091657

[21] Brede, M.; de Vries, B.J.M. The energy transition in a climate-constrained world: Regional vs. global optimization. *Environmental Modelling & Software* 2013, 44, 44-61. doi:10.1016/j.envsoft.2012.07.011

[25] Fernández Prieto, L.; Rodríguez Rodríguez, G.; Schallenberg Rodríguez, J. Wave energy to power a desalination plant in the north of Gran Canaria Island: Wave resource, socio-economic and environmental assessment. *Journal of Environmental Management* 2019, 231, 546-551. doi:10.1016/j.jenvman.2018.10.071

[38] Frank, E.; Hall, M.A.; Witten, I.H. The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques, 2016

[49] Kalnay, E.; Kanamitsu, M.; Kistler, R.; Collins, W.; Deaven, D.; Gandin, L.; Iredell, M.; Saha, S.; White, G.; Woollen, J.; Zhu, Y.; Leetmaa, A.; Reynolds, R.; Chelliah, M.; Ebisuzaki, W.; Higgins, W.; Janowiak, J.; Mo, K.C.; Ropelewski, C.; Wang, J.; Jenne, R.; Joseph, D. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society* 1996, 77, 437-471. doi:10.1175/1520-0477(1996)077<0437:TNYP>2.0.CO;2

- [62] Quinlan, J.R. C4. 5: Programs for machine learning; Morgan Kaufmann, 1992.
- [63] Breiman, L. Random forests. *Machine learning* 2001, 45, 5-32. doi:10.1023/A:1010933404324
- [64] Cortes, C.; Vapnik, V. Support-vector networks. *Machine learning* 1995, 20, 273-297. doi:10.1007/BF00994018
- [68] Alcalá-Fdez, J.; Sánchez, L.; García, S.; del Jesús, M.J.; Ventura, S.; Garrell, J.M.; Otero, J.; Romero, C.; Bacardit, J.; Rivas, V.M.; Fernández, J.C.; Herrera, F. KEEL: a software tool to assess evolutionary algorithms for data mining problems. *Soft Comput.* 2009, 13, 307-318. doi:10.1007/s00500-008-0323-y

We would like to thank you again for the excellent review carried out to our work.

5 Reviewer #3

Comment 1

The Authors present a new open source tool for the creation of datasets integrated by meteorological variables from two sources of information. Basically, a user-friendly software has been developed and described on the basis of pivotal parameters, carefully selected for the specific study of the wave height in the Gulf of Alaska. The statistical model can be in principle extended to other meteorological measurements and monitoring.

This study is very technical and of interest for a specific audience. By the way, I don't see any link between the topic of the manuscript and the journal *Energies*.

The software developed it is clearly presented as useful tool for meteorological application and no application to energy saving, production, conversion or similar it is presented.

Thus, I recommend to submit the manuscript to a more specialized journal.

Response: First of all, we would like to thank Reviewer #3 for his/her suggestions.

In the initial version we did not adequately explain what the purpose of this work is and its relation to energy, we hope that now it is better clarified and readers can understand the scope of our work. In this revised version of the manuscript we have tried to clarify the contributions of this work, emphasizing them mainly in **"Abstract"** (lines 10-19) and **Sections "1. Introduction"** (lines 130-168 - **"In order to ... system"**, 169-182 - **"Therefore ... security [55]"**) and **"5. Conclusions"** (lines 759-766 - **"The aim ... regression"**, lines 779-782 - **"The case ... carried out"**). Besides, in **Section "1. Introduction"** (lines 60-77) the relation of using this tool (to create datasets to model waves or flux of energy) with WECs, is discussed.

Prediction studies using ML and SC algorithms do not provide tools to incorporate and integrate, considering the casuistry that it entails, the two data sources used by SPAMDA. Such studies apply specific algorithms (extreme learning machine, metaheuristics, Bayesian networks, artificial neural networks, etc) on data, using custom-made implementations or scripts developed in some programming language, but they do not allow to build datasets, ready to be used with any of those methodologies or algorithms, through an automated and versatile process.

SPAMDA allows the creation of datasets to estimate the amount of energy flux that can be generated by waves at different prediction horizons. These datasets are well-formed and ready to be used with ML and SC prediction techniques. Therefore, we think that this work can be useful as a support for applications related to marine energy saving, production and conversion processes.

In that sense, and to better clarify the applicability and the relation to energy of this work,

we have extended Section "4. A case study applied to Gulf of Alaska" with a short-term forecasting problem of the energy flux generated by the waves in a specific location of that zone. Besides, Sections "3.4. Matching configuration" and "3.5. Final datasets" have been also rewritten for the purpose of flux of energy. In this way, it would be possible to perform prediction studies 6h in advance (the prediction horizon is customisable) of the amount of marine energy that could be extracted by WECs. Given that this work does not focus on models performance, a validation or comparison study of the results obtained in such example has not been carried out.

Having said that, we think that this software tool could be used in coastal and ocean engineering applications and marine energy prediction, such as (included in Section "1. Introduction", lines 169-182 - "Therefore ... security [55]"):

- EStimation of the energy supply sources in hybrid energy systems [3], based on the amount of energy that can be obtained by a marine energy system in the next 6h or 12h, for example.
- Regulation of WECs to avoid malfunction or breakage [27,28], depending on the significant wave height and/or energy flux expected, as well as the possibility of reconfiguring them in order to maximise the wave energy extraction.
- Decision-making in exploitation and environmental protection for the construction of marine energy storage plants, future strategies on renewable energy and coastal planning [52].
- Support tool for the prediction of the energy that could be obtained from a certain maritime location, to know whether or not it is convenient to install WECs as power supply in marine structures, such as offshore oil and gas platforms or seawater desalination plants [24, 25].
- Design of offshore structures and ports [53].
- Decision-making and risk assessment about operational works in the sea [54], security systems for structures or naval security [55].

We would like to thank you again for the suggestions and review carried out to our work.