Data Fault Detection for Digital Twin Learning Action Decision of a Wind Turbine

William D. Chicaiza
Dept. of System Eng. & Automatic Control
Universidad de Sevilla
Seville, Spain
wchicaiza@us.es

Fabio Rodríguez
Dept. of Applied Mathematics II
Universidad de Sevilla
Seville, Spain
frodriguex@us.es

Adolfo J. Sánchez
Dept. of Mech., Biomedical & Manufacturing Eng.
Munster Technological University
Cork, Ireland
adolfo.sanchezdelpozofernandez@mtu.ie

Juan M. Escaño
Dept. of System Eng. & Automatic Control
Universidad de Sevilla
Seville, Spain
jescano@us.es

***Abstract*—This paper presents the design of a classifier of variable failures in a wind turbine system. The classifier is based on a structure formed by several TS fuzzy inference systems, with projections of the data onto components of a principal component analysis. The classifier is part of a discrepancy evaluator for triggering the learning mechanism of the digital twin of the wind turbine.**

***Index Terms*—Fault Detection, Digital Twin, Neuro-Fuzzy**

I. Introduction

The process of achieving higher levels of renewable energy penetration in industry is key to achieving substantial reductions in the sector's demand for fossil fuels and the sector's CO_2 emissions [4]. The optimal use of renewables in the industrial grid requires advanced control techniques. Model-based predictive control is a technique that has proven to be very efficient in the real-time optimisation of energy and industrial systems [1], [2]. Predictive control techniques require a dynamic model of the system to be controlled. The inclusion of these models in digital twins [6] allows the construction of virtual plants that help to uncover dynamic system behaviour problems, including process control problems. The digital twins need to communicate with the real plants, in order to update their parameters according to changes in the real plants. This update allows to have dynamic models perfectly adapted to the system, in order to control it efficiently. The question that arises then is the following: When is it considered that the change in real variables should trigger the update of the digital twins? The answer is not easy and needs analysis on the data collected in real time. For example, a failure or simply a drift in a sensor can trigger the learning mechanism of the twin in an undesired way. An important first step is to be able to detect such faults or false data in order to evaluate the decision to start the learning routine. System model-based fault detection is well known and widely used [3]. The problem of detection in them lies in the impossibility of detecting two simultaneous faults. In order to solve this problem, another type of technique is proposed in this work. A fuzzy classifier-based evaluator has been proposed and designed for the detection of drifts in energy values produced by a wind turbine, from which only some variables are obtained in real time. In a previous work [5], it has been proved that, in order to obtain a model of the active power, the following variables are necessary: wind direction, ambient temperature, wind speed, pitch angle and the operational state of the machine. In this case, in addition to the active power, only the following variables will be available every 10 minutes, by means of SCADA: wind speed, rotor speed and reactive power. The rest of the article is organised as follows. Section II deals with the preparation of data for fuzzy inference systems and principal component analysis. Section III will discuss the design of fuzzy inference systems. Section IV will show evaluation results and the article will end with a conclusion section.

This work has received funding from the European Union's H2020 research and innovation programme under Grant Agreement n. 958339 and the Spanish Ministry of Science, Innovation and Universities under grant PID2019-104149RB-I00

II. Preparation of operational data

As mentioned above, data from the following interrelated variables are used to design the proposed evaluator: Wind speed (W_S), rotor speed (R_S), reactive power (Q), and active power (P). Individual faults have been added to the active power measurement. An additive offset wich following the measured power profile under normal operating conditions. The faults induced in the active and reactive power (offset) are of 20% in the SCADA measurement of P (and Q). Additionally, real data from normal operation on different months are used, to which such faults in the active power variable have been added. In this work, we will initially detect specific added offset faults, classifying them in two different classes: normal operation and fault in the active and reactive power variables.

The proposed method combines principal component analysis (PCA) with adaptive neuro-fuzzy inference systems (ANFIS). PCA is a multivariate statistical technique, which optimally represents observations from an n-dimensional space in a low-dimensional space, allowing the identification of latent or unobserved variables in the data. In addition, it transforms variables that may be generally correlated into new uncorrelated variables, allowing ANFIS to better converge in learning. The variables obtained from the SCADA must undergo an initial data normalisation process to prevent the different nature and magnitude of the variables, noise and inconsistencies from affecting the learning process.

A. Principal Component Analysis

Principal component analysis (PCA) is performed offline on each of the classes, obtaining a matrix $\mathbf{R}_P^n$, which represents the new set of in-correlated variables, called principal components for each class. Each principal component is a linear combination of the original variables. The index n denotes each class $(Norm, Fault)$ and the subscript P denotes the variable associated with failure (in this case, active power).

The first principal component of each $\mathbf{R}_P^n$ matrix contains the largest variance of the data, specifically over 99% of the variability. Therefore, we use it to represent the data in one dimension. Additionally, the PCA calculation of all the classes that make up the training set is carried out, obtaining the matrix *total*: $\mathbf{R}_P^{Total}$.

B. Obtaining projected prototypes in the main components

Once the PCA calculation of each of the classes $\mathbf{R}_P^n$ has been carried out, as well as that of all the classes $\mathbf{R}_P^{Total}$, the projections of each class are determined for each of the components, that is, each class will be projected onto the first principal component of the different classes. We will call prototypes $\left(\mathbf{P}_{n2}^{n1}\right)$ to the different projections obtained from the variables that make up each class, obtaining three new variables per class on a single axis in space, which we will use to assign the degree of belonging to the corresponding class when a new piece of data is evaluated. The prototype is equal to the scalar product of each class with the first principal component of each matrix $\mathbf{R}_P^n$.

$$\begin{aligned}\mathbf{P}_{Norm}^{Norm} &= \prod\left(\mathbf{class}_{Norm}\cdot\mathbf{R}_P^{Norm}\right)\\ \mathbf{P}_{Fault}^{Norm} &= \prod\left(\mathbf{class}_{Norm}\cdot\mathbf{R}_P^{Fault}\right)\end{aligned} \tag{1}$$

$$\begin{aligned}\mathbf{P}_{Norm}^{Fault} &= \prod\left(\mathbf{class}_{Fault}\cdot\mathbf{R}_P^{Norm}\right)\\ \mathbf{P}_{Fault}^{Fault} &= \prod\left(\mathbf{class}_{Fault}\cdot\mathbf{R}_P^{Fault}\right)\end{aligned} \tag{2}$$

Additionally, a general prototype $\mathbf{P}_g^n$, of each class, is obtained from the product of each class with the matrix $\mathbf{R}_P^{Total}$.

$$\begin{aligned}\mathbf{P}_g^{Norm} &= \prod\left(\mathbf{class}_{Norm}\cdot\mathbf{R}_P^{Total}\right)\\ \mathbf{P}_g^{Fault} &= \prod\left(\mathbf{class}_{Fault}\cdot\mathbf{R}_P^{Total}\right)\end{aligned} \tag{3}$$

Finally, the training sets for ANFIS learning are made up of the prototypes of each of the classes and their general prototype.

$$\mathbf{T}_{Norm} = \left[\mathbf{P}_{Norm}^{Norm}\ \ \mathbf{P}_{Fault}^{Norm}\ \ \mathbf{P}_g^{Norm}\right] \tag{4}$$

$$\mathbf{T}_{Fault} = \left[\mathbf{P}_{Norm}^{Fault}\ \ \mathbf{P}_{Fault}^{Fault}\ \ \mathbf{P}_g^{Fault}\right] \tag{5}$$

The check set contains data from two days different from the training set and to obtain its prototypes, the procedure to follow is the same as described above. In addition, the $\mathbf{R}_P^n$ matrices of each of the classes already calculated in the training set are used.

The data provided are from different days, the flow variable is rising in the middle of the time and then falling, so the time series has been segmented and the rising part has been taken to obtain the projected prototypes in the first principal component of each class.

III. Drift classifier design in active power sensor

The prototype that defines the $\mathbf{class}_n$ is modelled by means of an ANFIS, using the prototypes of each class as inputs and the general prototype as output. In this way, several ANFIS are obtained, which, once trained, deliver the prototype estimation corresponding to each class. The training of each ANFIS is performed by first applying a clustering method (subtractive grouping), which initially estimates the number of clusters to determine the number of rules and membership functions (MF). Subsequently, the parameters of each ANFIS layer are updated using a hybrid learning method, which combines gradient descent to optimise the parameters of the antecedent and least squares to determine the linear parameters of the consequent. The parameters of the ANFIS architecture is the same in all cases due to the correlation of the data (similarity between each class, due to the addition of a positive and negative offset) and is shown in the table I.

TABLE I
ANFIS PARAMETERS FOR P AND Q

Description		**ANFIS**
MF type:	$Gaussian$	
Optimization method:	$hybrid$	
Output MF type:	$linear$	
$\mathbf{FIS}_v^n$	$\mathbf{P}^n$	$\mathbf{Q}^n$
Number MFs:	3	3
Number rules:	3	3
Influence range	0.6	0.70
Epoch number:	100	100

The ANFIS learning process uses a relatively low number of epochs and the errors obtained in the training and check sets are low, indicating that the learning was generalised. The error rates obtained in the training of each model are shown in the table II.

The aim after having obtained the $\mathbf{FIS}_P^n$ systems is to determine the class to which a new datum belongs, so we define a cost function that determines which system reaches the real prototype.

$$J_i = \left\|\mathbf{P}_{real}^* - \mathbf{P}_{\mathbf{FIS}_P^n}\right\|_2^2 \tag{6}$$

TABLE II
ANFIS VALIDATION FOR P AND Q

RMSE min of Training & Checking	ANFIS obtained	
	$\mathbf{FIS}_P^{Norm}$	$\mathbf{FIS}_P^{Fault}$
$RMSE_{Train}$	0.00319721	0.00324574
$RMSE_{Check}$	0.00368549	0.00514447
	$\mathbf{FIS}_Q^{Norm}$	$\mathbf{FIS}_Q^{Fault}$
$RMSE_{Train}$	0.00093166	0.00093149
$RMSE_{Check}$	0.00146925	0.00133776

TABLE III
CONFUSION MATRIX $\mathbf{P}$.

REAL \ PREDICTED	Norm	Fault
Normal	3418	107
Fault	352	3663

Presicion	90.66 %	97.16 %
Reacall	96.96%	
Accuracy	93.91 %	

TABLE IV
CONFUSION MATRIX $\mathbf{Q}$.

REAL \ PREDICTED	Norm	Fault
Normal	4017	3
Fault	203	4217

Presicion	95.19%	99.93%
Reacall	99.93 %	
Accuracy	97.56 %	

We use an Exhaustive Searching Algorithm (ESA) to determine the class to which a new piece of data to be classified belongs.

This algorithm evaluates the cost function with the output value provided by each fuzzy inference system $\mathbf{FIS}_P^n$ and determines the class by direct comparison, choosing the one that minimises the cost function. The model with the minimum J_i defines the class to which the new data belongs.

A. Neuro-fuzzy classifier

The ANFIS-based classifier uses TS-type FIS. Figure 1 shows the structure of the wind turbine active and reactive power offset fault detection classifier. The first component of the PCA, obtained from each class $\mathbf{R}_P^n$ and $\mathbf{R}_Q^n$, calculated previously, are kept as fixed parameters and serve to define the new prototypes in the dimension of each class. When the normal operation data set is multiplied by the first principal component $\mathbf{R}_P^{Fault}, \mathbf{R}_Q^{Fault}$, the projected prototype in the class $\mathbf{class}_{F_{neg}}$ will be obtained, multiplying the data by $\mathbf{R}_P^{Norm}$ (and $\mathbf{R}_Q^{Norm}$), we will obtain the projection in the class $\mathbf{class}_{Norm}$ and when it passes through $\mathbf{R}_P^{Fault}$ ($\mathbf{R}_P^{Fault}$), it is projected in the class $\mathbf{class}_{Fault}$. The input data passes through the first component $\mathbf{R}_P^n$ ($\mathbf{R}_Q^n$), where P and Q denote the active and reactive power, respectively, measured by the SCADA, and is projected into the class $\mathbf{class}_n$, thus obtaining the corresponding prototype. The calculated prototypes form the input vector with which each of the $\mathbf{FIS}_P^n$ ($\mathbf{FIS}_Q^n$) systems are evaluated. Finally, passing the data through the first principal component of $\mathbf{R}_P^{Total}$ ($\mathbf{R}_Q^{Total}$), we obtain the real prototype with which we have to evaluate the cost function in the **ESA** to determine the class to which it belongs. **ESA** performs a direct comparison by evaluating the cost function with the estimated value of each $\mathbf{FIS}_P^n$ ($\mathbf{FIS}_Q^n$) and assigns the label of the FIS that minimises the cost function. For example, when the evaluated data has Negative Faults, the $\mathbf{FIS}_P^n$ ($\mathbf{FIS}_Q^n$) system minimises the cost function, and that class is assigned to the data.

IV. EVALUATION OF THE NEURO-FUZZY CLASSIFIER

As can be seen in Figures 2 and 3, the new data is multiplied by each principal component, the input vector is formed and each FIS model is evaluated. Each $\mathbf{FIS}_P^n$ (and $\mathbf{FIS}_Q^n$)system is evaluated with data from each class. If the input data contains active power faults, the FIS whose output will be closest to the actual prototype will be the $\mathbf{FIS}_P^{Fault}$ ($\mathbf{FIS}_Q^{Fault}$), as seen in Figures **??**, 3(b). Figures 2(a) and 3(a) show that, if the input data has no fault, the values estimated by the $\mathbf{FIS}_P^{Norm}$ ($\mathbf{FIS}_Q^{Norm}$) will follow the actual prototype characterising the Normal class.

Figure 4 shows the difference of the real prototype with each of the FIS models. If the evaluated data contains faults, the $\mathbf{FIS}_P^{Fault}$ will present the minimum error resulting from the difference as can be seen in Figure 4(b). Figure4(a) similarly shows that, if the evaluated data is not faulty, the $\mathbf{FIS}_P^{Norm}$ will present the minimum error. Moreover, if we evaluate the prototypes obtained from each $\mathbf{FIS}_Q^n$, with the real prototype of Q (see Figure 5), the minimum error resulting from the difference can be seen in Figure 5(a) with normal operation data and Figure 5(b) with failure data. As can be seen in the Figure 4 and 5, in both cases the minimum error (closest to zero) that defines the class to which the evaluated data belong is below the error of the class to which it does not belong.

The proposed algorithm for assigning the class to which the evaluated data belongs performs well as can be seen in Figure 6. $\mathbf{ESA^P}$ performs a direct comparison by evaluating the cost function with the estimated value of each $\mathbf{FIS}_P^n$ and assigns the label ($class_{Norm} \in 1, class_{Fault} \in 2$) of the FIS that minimises this function. Similarly, the assignment of classes through $\mathbf{ESA^Q}$ by evaluating the cost function with the estimated value of each $\mathbf{FIS}_Q^n$ performs well as can be seen in Figure 7.

Additionally, the confusion matrices (see table III and IV) have been performed to evaluate the proposed classifier, those matrices show the distribution of the classes assigned to the predicted data, compared to their true classes.

V. CONCLUSION

A classifier of offset faults in one of the variables of a wind turbine has been designed. The classifier is based on a structure formed by several TS fuzzy inference systems,

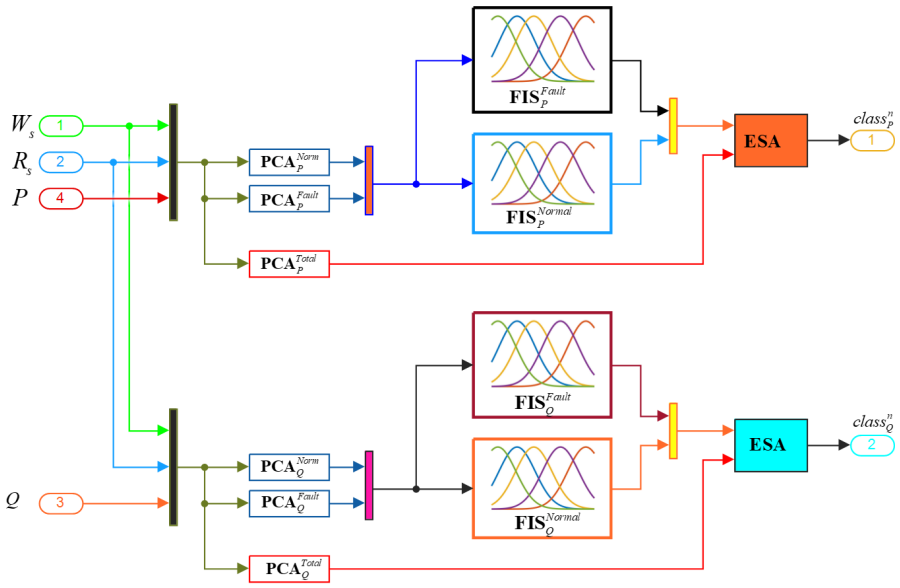


Fig. 1. Neuro-fuzzy classifier structure

with projections of the data on components of a principal component analysis. The classification results show very good rates of accuracy and sensitivity. The classifier is part of a discrepancy evaluator for triggering the learning mechanism of the digital twin of the wind turbine. As future work, it is proposed to extend this methodology to the rest of the critical variables in all assets susceptible to digital twinning.

Acknowledgment

The authors would like to thank the Spanish Ministry of Science, Innovation and Universities under grant PID2019-104149RB-I00 and the European Union's Horizon 2020 research and innovation programme under grant agreement no. 958339 for funding this work.

References

[1] C. Bordons, F. Garcia-Torres, and M. A. Ridao, *Model Predictive Control of Microgrids*. Springer International Publishing, 2020. [Online]. Available: http://link.springer.com/10.1007/978-3-030-24570-2

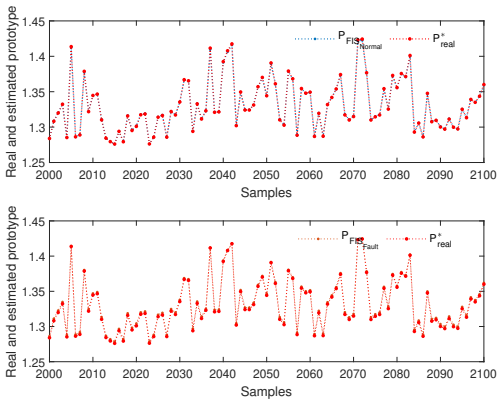
[2] E. F. Camacho and C. Bordons, *Model predictive controllers*. Springer International Publishing, 2007, vol. 2.

[3] S. X. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008.

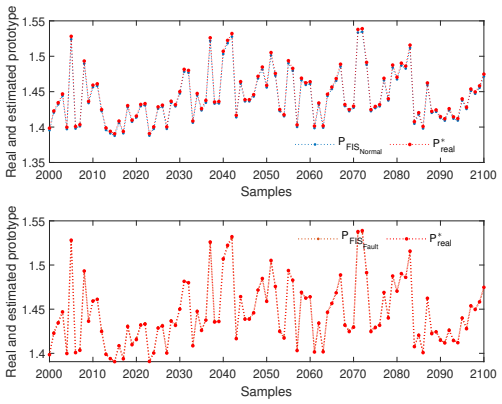
[4] R. Kempener and D. Saygin, "Renewable energy in manufacturing: A technology roadmap for remap 2030," *IRENA*, June 2014. [Online]. Available: https://www.irena.org/publications/2014/Jun/Renewable-Energy-in-Manufacturing

[5] M. G. Moreno, "Desarrollo y validación experimental de un gemelo digital para un aerogenerador," 2021. [Online]. Available: https://idus.us.es/handle/11441/127622

[6] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 2405–2415, 2019.

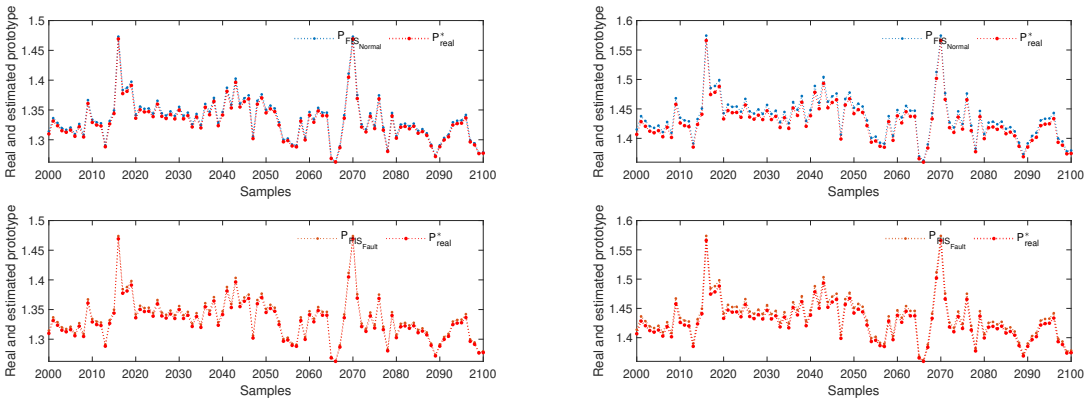


(a) Actual and estimated prototype for each FIS (of P) after evaluation with normal operating data



(b) Actual and estimated prototype for each FIS (of P) after evaluation with faulty data

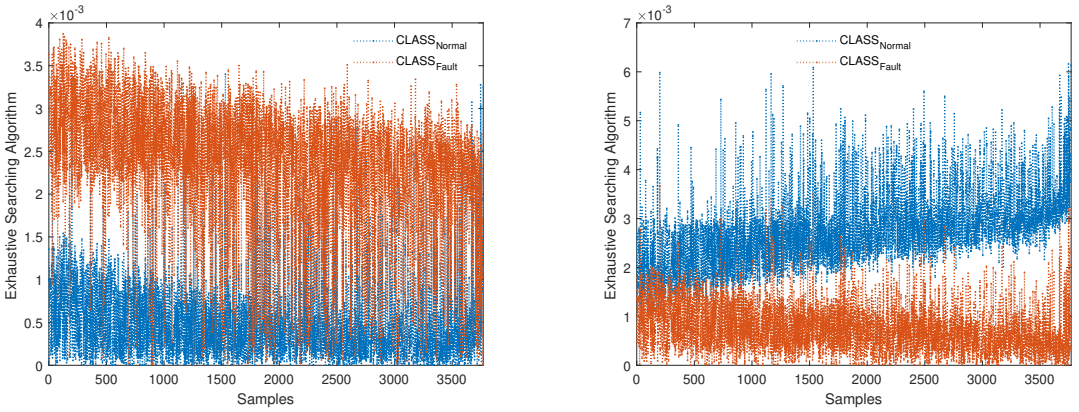
Fig. 2. Prototypes obtained with failure and normal operation of P.



(a) Actual and estimated prototype for each FIS (of Q) after evaluation with normal operating data

(b) Actual and estimated prototype for each FIS (of Q) after evaluation with faulty data

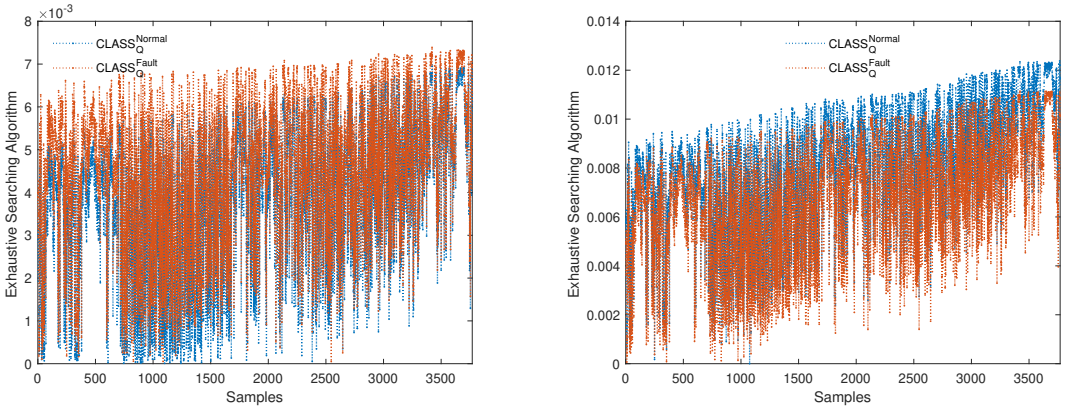
Fig. 3. Prototypes obtained with failure and normal operation of Q.



(a) Error obtained from direct comparison in the ESA between the prototype of each FIS and the real prototype with normal operating data.

(b) Error obtained from direct comparison in the ESA between the prototype of each FIS and the real prototype with faulty data.

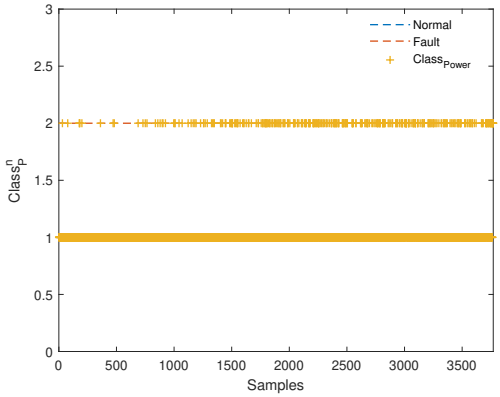
Fig. 4. Errors delivered by the ESA after minimising the cost function for $\mathbf{FIS}_P^n$.



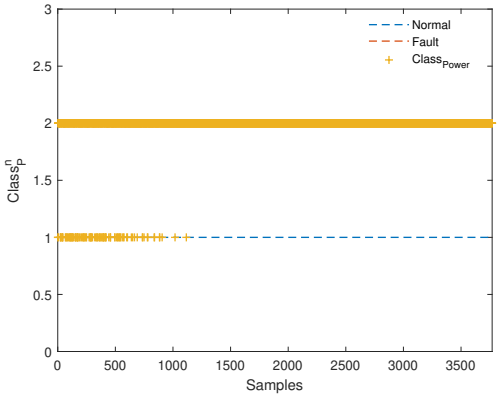
(a) Error obtained from direct comparison in the ESA between the prototype of each FIS and the real prototype with normal operating data.

(b) Error obtained from direct comparison in the ESA between the prototype of each FIS and the real prototype with faulty data.

Fig. 5. Errors delivered by the ESA after minimising the cost function for $\mathbf{FIS}_Q^n$.

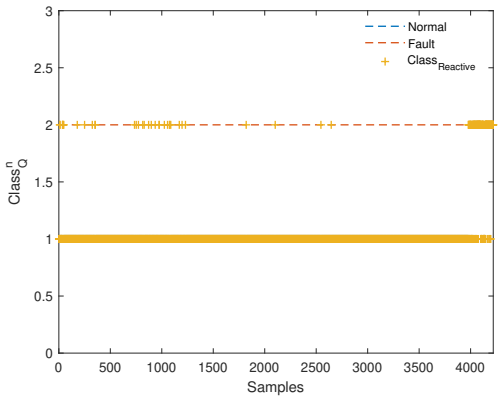


(a) Classification with normal operating data.

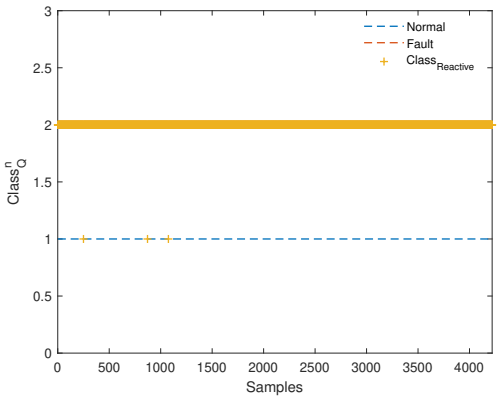


(b) Classification with faulty data.

Fig. 6. Class assignment by means of ESA with different $\mathbf{P}$ data.



(a) Classification with normal operating data.



(b) Classification with faulty data.

Fig. 7. Class assignment by means of ESA with different $\mathbf{Q}$ data.

Software de Desarrollo para el Análisis de los Desplazamientos de una Plataforma Flotante Sobre Ensayos en Modelo Físico

Vicente Negro Valdecantos
Enviroment Coast and Ocean Research Laboratory (ECOREL-UPM).
Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos.
Universidad Politécnica de Madrid.
Madrid, España
vicente.negro@upm.es

César Sastre Gil
Enviroment Coast and Ocean Research Laboratory (ECOREL-UPM).
Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos.
Universidad Politécnica de Madrid.
Madrid, España
cesar.sastre@ucm.es

Jose Luis Orts Egío
Enviroment Coast and Ocean Research Laboratory (ECOREL-UPM).
Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos.
Universidad Politécnica de Madrid.
Madrid, España
joseluis.orts@upm.es

***Abstract*— The 2030 Agenda and Sustainable Development Goals (SDG) warn about the climate emergency on the planet. The greenhouse gases with the increase in temperature and the change of pH in the oceans is causing a degradation that can be irreversible. If to this is added that the countries only offer minimal consensus positions at their COP 25 (Madrid, 2019), COP 26 (Glasgow, 2021) meetings, the risks are obvious. In this sense, the commitment to wind energy in the sea in great depth is a reality. For three years, different platforms have been tested to analyze their operation against the maritime climate and wind and to study their movements and turns. This research aims to describe its analysis and first results.**

Keywords*— *2030 Agenda, ODS 13, ODS 11, ODS 14, energy, marine energy, offshore platform, float facility, floating, wind waves, sensor, physical model, optitrack.

Resumen*— *La Agenda 2030 y los Objetivos de Desarrollo Sostenible de las Naciones Unidas alertan sobre la emergencia climática en el planeta. Los gases de efecto invernadero con el aumento de la temperatura y el cambio de pH en los océanos está provocando una degradación que puede ser irreversible. Si a eso se añade que los países solo ofrecen posiciones de consenso mínimas en sus reuniones COP 25 (Madrid, 2019), COP 26 (Glasgow, 2021), los riesgos son evidentes. En ese sentido la apuesta por la energía eólica en el mar en gran profundidad es una realidad. Durante tres años se han ensayado distintas plataformas para el análisis de su funcionamiento frente al clima marítimo y viento y analizar sus desplazamientos y giros. Esta investigación pretende describir su análisis y primeros resultados.

Palabras clave*— *Agenda 2030, ODS 13, ODS 11, ODS 14, energía, energía eólica marina, plataforma offshore, flotador, flotante, oleaje, sensor, modelo físico, optitrack.

I. INTRODUCCIÓN

Es un hecho que el planeta está en "emergencia climática" y que tanto la Agenda 2030 y los Objetivos de Desarrollo Sostenible muestran la preocupación por la situación, preferentemente en los conceptos de la "acción por el clima" ODS 13 y los efectos que pueden tener sobre las ciudades y comunidades (ODS 11) más cuando se encuentran en el litoral y sobre la vida en los ecosistemas, tanto marinos (ODS 13), como terrestres (ODS 14). Esta situación ha sido tratada ampliamente en la Conferencia de París (COP 21, 2015), en los acuerdos sobre Océano y Criosfera de Mónaco (2019) y es un efecto muy claro en el mundo [4] [12].

Los eventos extremos son cada vez más frecuentes, las inundaciones, sequías, temporales, maremotos y terremotos, nevadas empiezan a convivir de forma cotidiana con el ser humano.

Con un análisis de las previsiones de las referencias científicas y sobre la base de los datos de oceanografía tomados del análisis de clima en España y el Reino Unido, como de corrientes y mareas, así como de modelos numéricos de propagación y de corrientes se plantearon un análisis de dispositivos "limpios" basados en las siguientes hipótesis en función de la definición de cuatro nuevos escenarios de emisión, las denominadas Trayectorias de Concentración Representativas (RCP, por sus siglas en inglés). Éstas se caracterizan por su Forzamiento Radiativo (FR) total para el año 2100 que oscila entre 2.6 y 8.5W/m^2.

Las cuatro trayectorias RCP comprenden un escenario en el que los esfuerzos en mitigación conducen a un nivel de forzamiento muy bajo (RCP 2.6); dos escenarios de estabilización (RCP 4.5 y RCP 6.0) y un escenario con un nivel muy alto de emisiones de GEI (RCP 8.5). Los nuevos RCP pueden contemplar los efectos de las políticas orientadas a limitar el cambio climático del siglo XX y analizar qué sucedería al planeta en situaciones donde no se actúe; donde se limiten las emisiones para que la temperatura solo se supere en dos grados, en 1.50 grados entre otros aspectos.

En este sentido la energía eólica en el mar ha sido una evolución desde que el primer campo funcionase en aguas someras en 1991 en Dinamarca. Su desarrollo ha sido tan vertiginoso que parece el proceso evolutivo de las especies en muy poco tiempo (Fig. 1). [3] [5].

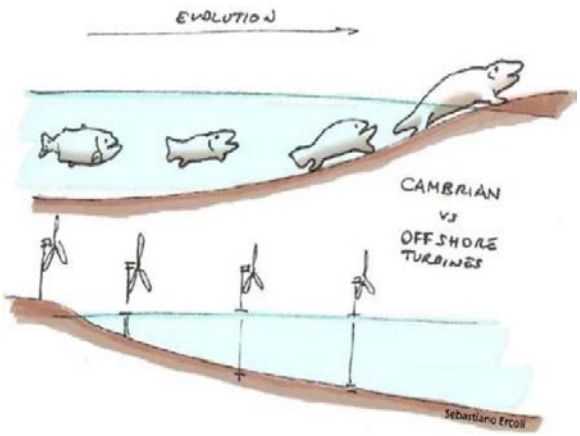


Fig. 1. Comparación de la evolución de la energía eólica con los períodos históricos