Mathematical Modelling of a TENG-Powered Data Buoy

Iain McLeod
Dept. of Electronic Engineering, Maynooth University
Centre for Ocean Energy Research
Maynooth, Ireland
iain.mcleod.2022@mumail.ie

Daniel Clemente
Faculty of Engineering of the University of Porto (FEUP)
CIIMAR - Interdisciplinary Centre of Marine
and Environmental Research, Matosinhos
Porto, Portugal
up201009043@edu.fe.up.pt

Paulo Rosa Santos
Faculty of Engineering of the University of Porto (FEUP)
CIIMAR - Interdisciplinary Centre of Marine
and Environmental Research, Matosinhos
Porto, Portugal
pjrsantos@fe.up.pt

Francisco Taveira Pinto
Faculty of Engineering of the University of Porto (FEUP)
CIIMAR - Interdisciplinary Centre of Marine
and Environmental Research, Matosinhos
Porto, Portugal
fpinto@fe.up.pt

Marco Rosati
Dept. of Electronic Engineering, Maynooth University
Centre for Ocean Energy Research
Maynooth, Ireland
marco.rosati.2021@mumail.ie

John V. Ringwood
Dept. of Electronic Engineering, Maynooth University
Centre for Ocean Energy Research
Maynooth, Ireland
john.ringwood@mu.ie

Abstract—Triboelectric nanogenerators, or TENGs, are prominent among power take off (PTO) systems being researched for use in specifically small scale wave energy converters (WECs). The low power requirements, and modest size of data buoys, suggests TENGs as an ideal PTO for this small-scale wave energy application. However, a physics-based model of the relationship between the movement of the buoy and the performance of the onboard TENG is difficult, particularly for rolling-ball-type TENGs. This paper develops three multiple input single output system identification (SI), black-box models for a rolling-ball-type TENG. The SI models are identified, and validated, using input-output data gathered from real wave tank tests on a 1:8 Froude-scaled model of a navigation buoy. The models described in this paper, relate the TENG's current output to the displacement, in 6 degrees of freedom (DoFs), of the buoy to which it is mounted, across a range of wave conditions. The SI models used include linear autoregressive with exogenous input (ARX), and nonlinear Kolmogorov-Gabor polynomial (KGP).

Index Terms—navigation buoy, wave energy converter, triboelectric nanogenerator, TENG parameterization, data-based modelling, linear ARX model, nonlinear KGP model, system identification

I. Introduction

Large scale wave energy projects carry with them a large financial risk, and present a significant logistical challenge. This has resulted in a growing interest in smaller scale wave energy applications, such as wave powered data buoys [1]. Wave powered data buoys are exposed to the same environmental conditions, and must harness the same source of energy, as a large scale wave energy converter (WEC), but can be built and deployed for a fraction of the capital cost. However, harvesting wave energy at a small scale carries its own challenges, as effects from drag and friction increase. A trend observed in [2] was a split in the research landscape between those advocating for the scaling down of larger WEC designs, such as point absorbers and oscillating water columns (OWCs), and those arguing that small scale applications require power take off (PTO) systems designed specifically for the small scale. Among the most prominent of the PTO systems, identified in [2], specifically for small scale applications, is the triboelectric nanogenerator, commonly abbreviated to TENG. TENGs are designed to harvest energy from sources of irregular motion, with potential applications in wearable technology, environmental monitoring, and medical science [3]. The ability of TENGs to generate energy from irregular, chaotic motions has led to a large number of research papers focussing on their potential as a PTO for WECs (see reviews [4], [5], [6]), where the stochastic nature of waves creates a challenge for more conventional generators.

Although TENGs have many features which make them favourable for use as PTOs in small-scale wave energy applications, modelling their behaviour presents a significant practical challenge. For rolling-ball-type TENGs, such as the

ones shown in Fig. 1, this challenge is particularly pronounced, as the complex movement of the spheres within the device is highly chaotic. Usually, modelling the movements of these spheres requires visual recordings or measurements from experimental testing, or the use of a high fidelity solver such as COMSOL. Furthermore, the relationship between the movement of the buoy and the performance of the TENG is not always clear, as rolling-ball-type TENGs are multi-modal devices, utilising motions in multiple degrees of freedom (DoFs).

This obfuscated relationship makes it difficult to assess the feasibility of using the same TENG across a range of different WEC devices and device shapes, and presents challenges for developing energy maximising control systems. For complex systems whose behaviour is difficult to model from first principles system identification (SI) can be used [7]. SI is a data-based modelling approach, which deals with mathematically modelling dynamic systems based on data observed from those systems. SI models which rely only on input-output data are called "black box" models, while "grey box" SI models incorporate, up to a *variable* extent [8], known physics influencing the behaviour of the system. In either case, the identified SI model has to be validated against a separate set of validation input-output data, which are different from those used in the model identification phase. It should also be noted that the choice of suitable identification data is of critical importance, as the SI model validity is limited only to the range of equivalent frequencies, and amplitudes, of the input signals used in the identification phase.

A previous approach to parameterizing rolling-ball-type TENGs in relation to buoy motion is detailed in [9], where an optimisation approach was used to find parameter coefficients that correlate some of the buoy motions with the TENG voltage outputs for a number of wave tank tests. The model in [9] is based on the results of wave tank tests carried out on a 1:8 Froude-scaled model of a Sealite Atlantic-2600 navigation buoy in regular waves. However, the parameterized model in [9] relied on the ability of particle swarm algorithms to handle a large number of inputs. The model described in [9] required 18 inputs covering the displacement, velocity and acceleration in each of the 6 DoFs.

As with [9], the data for the parameterized TENG described in this paper comes from the wave tank tests in [10]. In [10], a number of different rolling-ball-type TENGs were tested, using regular waves with amplitudes and periods reflective of those observed off the coast of Póvoa de Varzim, Portugal. The TENG type selected for parameterization in this paper was the unidirectional lateral-based TENG (UL-TENG), described in Section II.

The SI approach described in this paper reduces the number of model inputs required, by using an input selection process to identify only the key DoFs that affect the power capture of the TENG. As SI models are dynamic, the model can also be simplified by considering only the displacements in each DoF as potential inputs.

The paper is laid out as follows: Section II describes the experimental setup from which the test data used to train and validate the ARX and KGP models was gathered. In Section III, a second navigation buoy is introduced as a potential testbed to assess the performance of the black box model across a different floater geometry. Section IV, which describes the data-based modelling procedure, contains the majority of the content of this paper: it covers the signal preprocessing; a description of the input selection process used to define the key variables affecting TENG power output; and finally the identification and validation of a multiple input single output (MISO) linear ARX model, and a MISO nonlinear KGP model. Section V considers the best-fit simulation model to assess the TENG's performance on the alternative buoy geometry. This is followed by a discussion of the results in Section VI. Finally Section VII concludes the paper with a summary of the findings and suggestions for future work.

II. EXPERIMENTAL SETUP

A full description of the experimental setup used to generate the data used in this paper can be found in [10]. In summary, a number of TENG configurations were tested on a 1:8 Froude-scaled model of a Sealite Atlantic-2600 navigation buoy, in a wave basin measuring 28.0m long, 12.0m wide and 1.2m deep. The specifications for the buoy can be found at [11]. The wave conditions determined for use in the wave tank tests were based on scaled down regular waves with amplitudes and periods reflective of sea states observed at Póvoa de Varzim, Portugal. The 8 regular wave conditions tested in [10] are summarised in Table I, using the same numbering system for the wave conditions as applied in [9].

The three rolling-ball-type TENGs described in [10], all used a similar basic working principle, where the movement of spheres promotes the periodic contact separation of the two triboelectrically opposed layers which form the surface upon which the spheres move. In the case of the experiments in [10], the two triboelectrically opposed layers were constructed from Nylon 6,6 and PTFE. The electrodes were made of conductive silver thin films (Ag), with the supporting structure 3-D printed using PLA. Fig. 1 shows the three different rolling ball TENG configurations tested in [10].

Of the TENG designs tested in [10], the UL-TENG was found to be the best performing, with a maximum power of 230μW. The UL-TENG is a rolling-ball-type TENG, where the rolling spheres are guided in a straight line along a series of parallel channels, as demonstrated in Fig. 1a3). In addition the UL-TENG featured additional hinged TENGs at the ends of each channel, shown in Fig. 1e) which creates an additional source of power by utilising the impact force as the spheres within the TENG collide with the end stops. This collision causes the hinged Nylon 6,6/Ag flap to contact with, then separate from, the triboelectrically opposed PTFE/Ag wall.

The TENGs tested in [10], were each tested at two locations: the first mounted at the location of the centre of gravity (CoG) of the buoy, and a second above the CoG. As the TENG model in this paper aims to create a model of just the TENG's behaviour, the CoG position was chosen, as this was
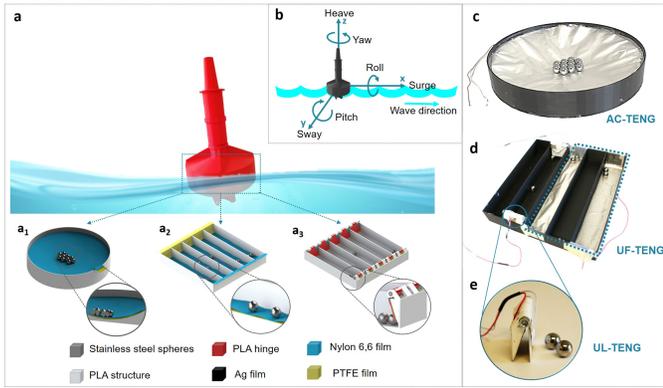
Fig. 1. a) Design of the navigation floating buoy-based on unidirectional and anisotropic triboelectric nanogenerators: a1) Anisotropic circular-based TENG (AC-TENG), a2) unidirectional flat-based TENG (UF-TENG) and a3) Unidirectional lateral-based TENG (UL-TENG). b) Six-degrees of freedom motions of the buoy. Photographs of the c) AC-, d) UF- and e) UL-TENGs. Reprinted with permission from [10], Copyright: Elsevier 2021.

the reference point for monitoring the motions of the buoy, and therefore there was a direct match between the motions of the buoy and the motions experienced by the TENG.

III. HYDRODYNAMIC MODEL OF ALTERNATIVE BUOY FOR TESTING BLACK BOX MODEL APPLICABILITY

As the aim of this paper was to develop a black box model of the 'isolated' UL-TENG, a different buoy was chosen to determine whether the TENGs performance can be assessed over a variety of floater types. The buoy chosen for testing an alternative geometry was the Balizamar B1600S, designed by Almarin [12]. The dimensions of the B1600S floater are shown in Fig. 2.

As the black box model would be using the displacements for each DoF as inputs, the motion of a 1:8 scale version of the B1600S needed to be determined. To do this Cummins' Equation was used, given in Equation (1):

$$
F_{moor} + F_{pto} + F_{ex} = (M + M_{a\infty})\ddot{x}(t) + \dots \\
\int_{-\infty}^{t} h_{rad}(t - \tau)\dot{x}(t)d\tau + kx(t)
$$

(1)

where, F_{moor}, F_{pto} and F_{ex} are the mooring-, PTO- and wave excitation forces respectively, M is the mass of the buoy, $M_{a\infty}$ is the infinite frequency added mass, h_{rad} is the radiation impedance impulse response function minus $M_{a\infty}$, k is the hydrodynamic stiffness, and x, $\dot{x}$ and $\ddot{x}$ are the displacement, velocity and acceleration of the buoy respectively. k is dependent on the cross sectional area of the buoy at the still water level. Experiments in [10] have shown that the PTO force exerted on a buoy by a rolling-ball TENG at this scale are negligible.

F_{ex}, h_{rad}, and the added mass, M_a, are frequency dependent terms which require the use of a boundary element method (BEM) solver, in this case determined using ANSYS AQWA. As the Balizamar B1600S is considerably smaller than
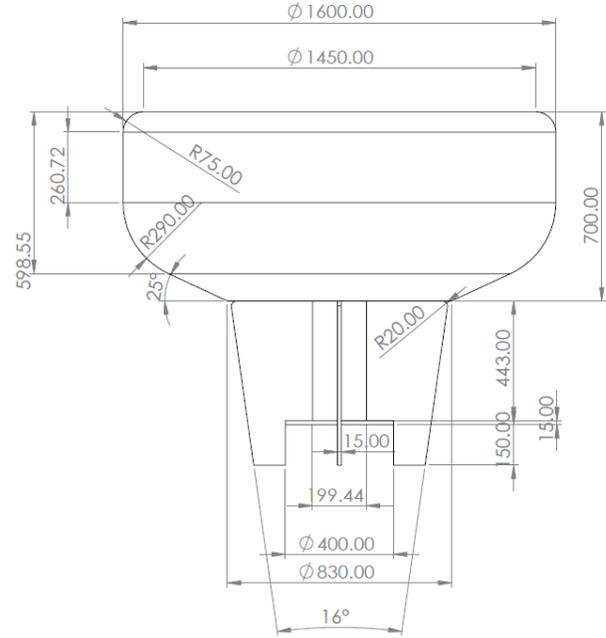
Fig. 2. Dimensions of B1600S buoy to be used as an alternative floater to the Sealite Atlantic-2600.

the Sealite Atlantic-2600, sea states with a lower significant wave height than those reported in [10] were chosen to keep the displacements of the B1600S within the same magnitude range as the Sealite Atlanticc-2600. The mooring force was also adjusted to tune the behaviour of the B1600S, by changing the line stiffness selected for the mooring.

IV. DATA-BASED MODELLING

In the wave energy field, a common approach for data-based WEC modelling is to use the free surface elevation, or the wave excitation force, as the SI model input. For instance, OWC WEC data-based hydrodynamic models which use the free surface elevation as the model inputs have been identfied in [13] and [14]. However, this approach limits the applicability of the SI model to only the device it was trained on. To keep the model described in this paper as generic as possible it was decided to identify a black box model only for the isolated TENG, using the displacement of the buoy in 6 DoFs as the model inputs, measured at its CoG, which were assumed to be equivalent to the displacement of the TENG mounted at the buoy CoG.

In the initial model, shown in Fig. 3, the displacement in 6 DoFs is selected as the input, and the current produced by the TENG is selected as the output. The time series values for the displacements in 6 DoFs used to validate the SI models in this paper came from experimental data. However, the identified SI models of the TENG can be used on any suitable buoy whose hydrodynamic model is given by Equation (1), as described in Section V.
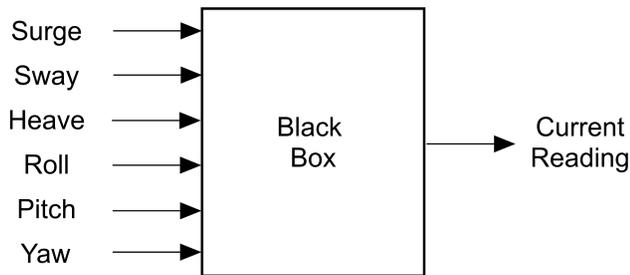
Fig. 3. Initial black box model with all 6 DoFs considered.

The previous approach to parameterizing a TENG, described in [9], focussed on voltage output, relating it with the frequencies of the displacements, velocities and accelerations in 6 DoFs. Unfortunately, accurate voltage data was unavailable for the construction of the model designed in this paper, so the alternating time series current response was used instead. The current readings provided, were, however, in almost raw form from the sensor used in the experiments, with only the resistance having been taken into account. An exact conversion formula to relate these current readings to the actual current values in µA, was not provided alongside the data. Therefore any graphs will refer to the current, purely as "Current Reading" and will not assign explicit units to the recorded values. Future work will aim to remedy this.

A. Signal Preprocessing

A number of steps needed to be taken to prepare the experimental signals for use in an SI model. Using data from the experiments in [10] presented some challenges, as the monitoring system used to sense the movement of the floater in the wave tank was switched on at a different time to the TENG sensor measuring current and voltage, creating a misalignment in the motion and current data. To temporally align the input and output traces the beginning of the transient behaviour was compared, as this marked the time at which the waves from the tank reached the floater and triggered both movement and power production. A cross correlation of the heave and current signals was then used to check the displacement and current signals were aligned for all 8 wave conditions.

A crucial step in preparing the data for an SI model is to remove noise, coming from sensors and/or external disturbances, from the signal. Indeed, a relatively poor signal-to-noise ratio in the input-output data can negatively affect the identification/validation performance of a model [13]. A Butterworth filter was used as a low pass filter to remove most of the high frequency noise from the data. A half-power frequency was selected for each of the wave conditions individually, as there is a significant variation in the wave frequencies between the different wave conditions (from 0.28 Hz to 1.34 Hz). The cut-off frequencies were tuned to remove the majority of noise spikes from the signal while preserving the frequency content of interest. A comparison of the raw signal and filtered signal for the current reading for wave condition 1 is shown in Fig. 4.
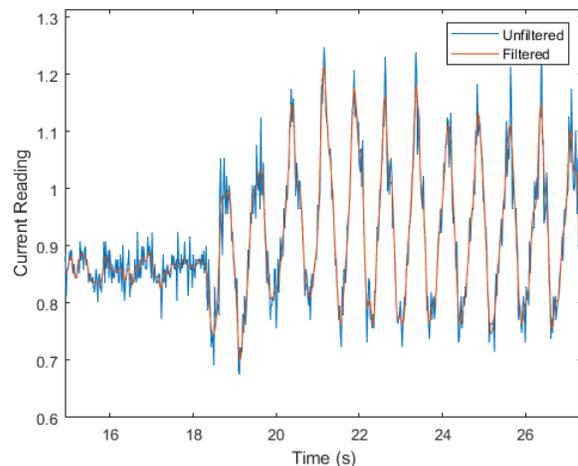
Fig. 4. Comparison of current signal with and without noise filtering applied.

Fig. 4 shows some noise variation of the signal is still present, particularly around the peaks and troughs, where the largest spikes in noise can be observed in the unfiltered signal. However, lowering the half-power frequency any further, begins to cause the signal to lose information around the peaks. The presence of noise in the current data was found to increase with a rise in significant wave height and period.

SI models are designed to capture the dynamic behaviour of a system. A prerequisite for training an SI model is that the output is zero when the system is at rest. Therefore any offset in the current readings and displacements was also removed at this stage.

The final step of preparing the current data was to resample it at the same timestep of 0.01s that the movement data was recorded at. The original current data timesteps were seemingly random and varied from 0s (where data points were repeated) up to 0.05s. The data was resampled using linear interpolation, and a quick check was carried out plotting the resampled signal and the original signal, to ensure that no significant changes had been imparted through the resampling process.

With the data prepared, the final step was to select which data would be used to train and validate the model. Both the training and the test data should be taken from the steady-state response of the system. Therefore the data was selected only where any transient behaviour has ended. The steady state behaviour for each wave condition was identified and split at the midpoint of the steady state response, into two equal sets of training and validation data. As the experiments in [10] were run for different lengths of time for each wave condition, the training and validation data were both limited to 10s, to ensure the identified model was not biased towards the tests run for longer time periods.

B. Input Selection

To keep a multiple input SI model as accurate and simple as possible, only those inputs which significantly affect the

output should be included. In order to identify those inputs which most affect the output, forward selection was used. Forward selection is a common input selection method, and has applications in both linear and nonlinear modelling [15]. Forward selection considers a set of inputs, X, that are related to an output, Y. To ascertain which inputs in X, Y is dependent on, a test criterion must be established. The process of forward selection is to start from a null model and sequentially select features from X that improve the fit to the selection criteria [16].

For the forward selection procedure, a generalized linear model (GLM) was used to relate the output current to a matrix of the 6 input DoFs for each wave condition. A chi-squared test was then determined as the fitting criterion. This criterion was used to cross-validate the deviance of the GLM with the 95th percentile of the chi-squared test for each degree of freedom. A 10-fold cross validation was sequentially carried out, comparing the deviance of the GLM with the 95th percentile of the chi-squared fit, iteratively selecting DoFs from the input matrix until adding more features did not further improve the fit.

A summary of the most important DoFs for each wave condition is given in Table I.

TABLE I
INPUTS AFFECTING OUTPUT CURRENT DEPENDING ON WAVE CONDITION

Wave Condition	Wave Period (s)	Significant Wave Height (m)	DoFs Affecting Current
1	0.743	0.0625	Surge, Heave, Roll and Pitch
2	0.920	0.1250	Heave and Pitch
3	1.097	0.1250	Surge and Heave
4	2.121	0.1250	Surge
5	2.828	0.1250	Heave
6	3.536	0.1250	Heave and Pitch
7	2.828	0.2500	Surge, Roll and Pitch
8	3.536	0.2500	Surge, Heave, Roll and Pitch

From Table I it can be observed that some combination of surge, heave, roll and pitch is required to predict the current output of the UL-TENG. Therefore, sway and yaw may be disregarded as inputs from the model shown in Fig. 3, resulting in a black box model with four inputs and a single output.

C. MISO Linear ARX Model Identification

A linear ARX model is a well known black box model structure with a linear input-output relationship [7]. It is described by Equation (2), where a_i and b_i are the unknown parameters:

$$\hat{y}(k) = -\sum_{i=1}^{n_a} a_i y(k-i) + \sum_{i=0}^{n_b} b_i u(k-n_k-i) \quad (2)$$

where, $\hat{y}(k)$ is the model output predicted at time k, and $y(k)$ and $u(k)$ are the kth output and input samples

respectively. n_a and n_b are the orders of the ARX model, again relating to the output and the input respectively. n_k represents the input delay and is used to adjust for any delay that may exist between the input and output signal. The number of unknowns in Equation (2) is equal to $n_a + n_b + 1$.

For a MISO linear ARX model, Equation (2) can be expanded, as shown in Equation (3):

$$\hat{y}(k) = -\sum_{i=1}^{n_a} a_i y(k-i) + \sum_{j=1}^{q} \sum_{i=0}^{n_{bj}} b_{ij} u_j(k-n_{kj}-i) \quad (3)$$

where, q is the number of model inputs.

To determine the orders of n_a and n_b for each of the model inputs, a sequence of model identification trials is carried out with incremental changes in n_a, and the n_b and n_k values for each input. To compare the model performance during the training and validation phases of the model identification trials the normalised root mean square error (NRMSE) is used, given in Equation (4):

$$NRMSE = \sqrt{\frac{\sum_k |y(k) - \hat{y}(k)|^2}{\sum_k |y(k)|^2}}, \quad (4)$$

Through an iterative process minimising the validation NRMSE [7], an n_a value can be found, and corresponding n_b values for the four inputs: n_{b1} (surge), n_{b2} (heave), n_{b3} (roll), and n_{b4} (pitch). The values of n_{k1}, n_{k2}, n_{k3} and n_{k4} were also determined at this step. In selecting the model orders, the principle of parsimony, defined in [17], was used, which considers the trade-off between model accuracy and model complexity, as, in general, the loss function will continue decreasing with the number of parameters. The MATLAB function $arxstruc$ allows for efficient calculation of a parsimonious ARX model, by iterating models across the user defined order ranges. In this paper the order ranges considered were $n_a = 1:10$ for the output, and $n_b = 0:9$ and $n_k = 0:5$ for each input. An initial ARX model using all 8 wave conditions determined a parsimonious model to have orders: $n_a = 5$, $n_{b1} = 1$, $n_{b2} = 5$, $n_{b3} = 4$, and $n_{b4} = 7$ with delays of $n_{k1} = 2$, $n_{k2} = 0$, $n_{k3} = 1$, $n_{k3} = 0$.

The unknown parameters a_i, and the b_{ij} values for the four inputs are found by solving a least squares (LS) optimisation problem. In order to simplify Equation (3), it can be split into vector form, with the regression vector, ϕ, given by Equation (5):

$$\phi(k) = [-y(k-1)... - y(k-n_a) \quad u_1(k-1+n_{k1})...$$
$$u_1(k-n_{b1}+n_{k1}) \quad ... \quad u_q(k-1+n_{kq})...$$
$$u_q(k-n_{bq}+n_{kq})]^T, \quad (5)$$

and the parameter vector, θ, given by Equation (6):

$$\theta(k) = [a_1 \, a_2 \, ... \, a_{na} \, b_{10} \, b_{11} \, ... \, b_{nb1} \, ... \, b_{q0} \, b_{q1} \, ... \, b_{nbq}]^T, \quad (6)$$

Having generated parameter and regression vectors to handle multiple inputs, Equations (3), (5) and (6), can be combined as (7):

$$\hat{\boldsymbol{y}}(k) = \boldsymbol{\theta}^T \boldsymbol{\phi}(k) = \boldsymbol{\phi}(k)^T \boldsymbol{\theta}, \tag{7}$$

Equation (7) is a simple linear regression, the solution of which is well documented [7]. Expanding for all time instants ($k = 1,2,3, \ldots N$) allows Equation (7) to be rewritten as Equation (8):

$$\hat{\boldsymbol{y}}(k) = \boldsymbol{\Lambda}\boldsymbol{\theta} \tag{8}$$

where, $\hat{\boldsymbol{y}}$ is the model output vector containing $[\hat{y}(1), \hat{y}(2), \hat{y}(3), \ldots \hat{y}(N)]^T$, and $\boldsymbol{\Lambda}$ is the regression matrix. The prediction error vector, $\boldsymbol{\epsilon}$, can then be defined as Equation (9):

$$\boldsymbol{\epsilon} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{\Lambda}\boldsymbol{\theta} \tag{9}$$

This allows the least squares optimisation, where $\boldsymbol{\theta}_{opt}$ can be found such that the LS error is minimised, to be written as Equation (10):

$$\boldsymbol{\theta}_{\mathrm{opt}} = \underset{\boldsymbol{\theta}}{argmin}(\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}), \tag{10}$$

The least squares optimisation can then be solved using the solution in Equation (11):

$$\hat{\boldsymbol{\theta}}_{opt} = (\boldsymbol{\Lambda}^T\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}^T\boldsymbol{y} \tag{11}$$

D. MISO Linear ARX Model Validation

To validate the MISO linear ARX model, three different criteria were assessed which link directly to the potential applications of the model. WEC models are typically used in model-based energy maximising control. To assess the model validity for use in control systems, the 1-step and 10-step ahead predicted outputs are computed, using the validation data. In a generic k-step ahead prediction, the ith output value is predicted using n_a measured output, up to time instant $i-k$, and n_b inputs, up to the ith time instant.

The final test of applicability is an uncorrected output simulation, where the model response is calculated only from the relevant input data and a set of initial conditions, and no measurement of the actual input is used to correct the model output. This type of output simulation is considered to assess the black box model's ability to be used across a range of devices, where the current output is not known.

As shown by Table I, each wave condition varied between being most dependent on surge, heave, roll, pitch or a combination of some or all four DoFs. The ARX model was constructed to relate the test data from across all 8 tested wave conditions, reading in surge, heave, roll and pitch as the four inputs. For validation of the model the NRMSE in Equation (4) was once again used. In Table II, model validation performance is reported in terms of model fidelity, defined as Equation (12):

$$\%_{fidelity} = 100(1 - NRMSE) \tag{12}$$

Note that negative values of $\%_{fidelity}$ can result when NRMSE exceeds 1. Using Equation (12), a comparison of the ARX models performance for each wave condition could be assessed. Table II summarises the fidelity values obtained for the single-step and 10-step ahead predictive models, and the fidelity of the uncorrected simulation model, for the linear ARX model trained on all 8 wave conditions. All NRMSE values for the models were calculated after 2s to ensure only the steady state behaviour of the model is included in the fidelity calculation.

TABLE II
FIDELITY FOR LINEAR ARX MODEL TRAINED ON ALL 8 WAVE
CONDITIONS (NEGATIVE VALUES INDICATE AN NRMSE > 1)

Wave Condition	Fidelity (%) 1-Step Ahead Prediction	Fidelity (%) 10-Step Ahead Prediction	Fidelity (%) Uncorrected Simulation
1	97.94	71.90	46.32
2	97.63	77.20	57.62
3	97.61	70.21	61.84
4	97.49	6.22	-126.02
5	98.31	51.78	14.16
6	97.81	54.34	16.41
7	98.10	38.37	-5.26
8	98.57	69.11	57.62

Table II shows that the overall fidelity for the model trained to fit all 8 wave conditions, is above 97% for all of the single-step predictions. The 10-step ahead prediction performs best at wave condition 8 and at the higher frequency wave conditions (1-3), with fidelities close to, or above, 70%. The uncorrected simulation output follows a similar pattern. As shown in Fig. 5, for wave conditions 4-7, the simulated outputs with the lowest fidelity align with the spikier validation waveforms, where the simulated linear ARX model is unable to capture the more complex behaviour of the current signals at these wave conditions.

Wave condition 4 is a particularly poor fit, both for 10-step ahead prediction and the uncorrected simulation. The model output for wave condition 4, shown in Fig. 5d), is a clear example of the linear ARX model struggling to fit the current output to all 8 wave conditions in a single model. As the model identified for all 8 wave conditions only had an associated surge input order of $n_{b1} = 1$, it is unsurprising that wave condition 8, which was found to be heavily dependent on surge in Table I, had the lowest fidelity for both 10-step ahead prediction and uncorrected simulation.

Given the strong prediction fidelity, but poor uncorrected simulation fidelity, for wave conditions 1-3, a linear ARX model was created using only these wave conditions for training and validation. Identifying this model resulted in orders of $n_a = 5$, $n_{b1} = 4$, $n_{b2} = 8$, $n_{b3} = 4$, and $n_{b4} = 7$ with delays of $n_{k1} = 2$, $n_{k2} = 2$, $n_{k3} = 2$, $n_{k3} = 0$. Using just the three high frequency wave conditions, resulted
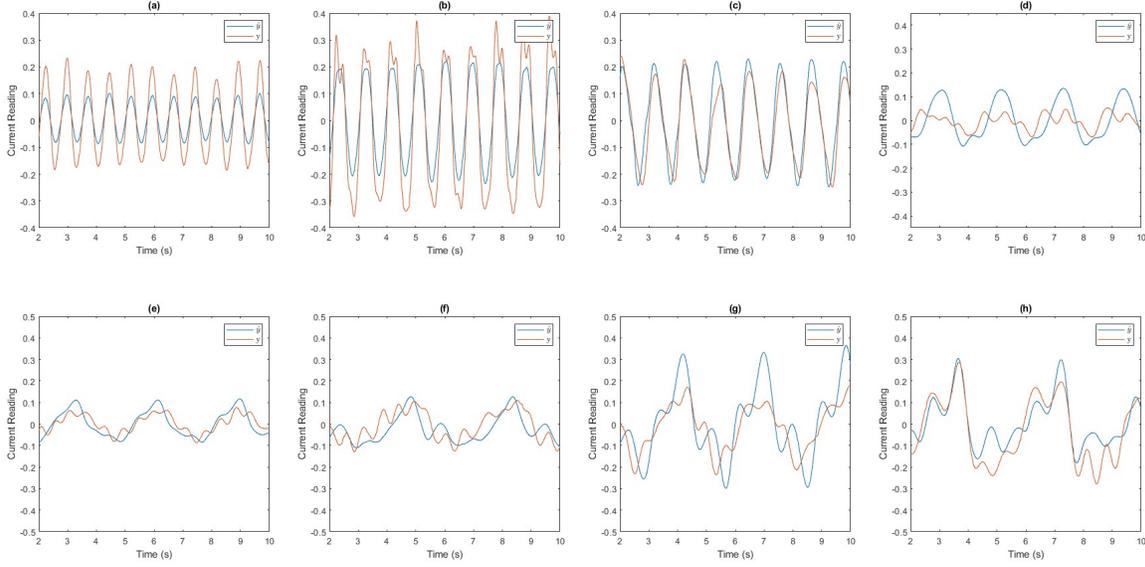
Fig. 5. Uncorrected simulation model output current ($\hat{y}$), compared with validation data (y) for the linear ARX model trained all 8 wave conditions. a) Wave condition 1 b) Wave condition 2 c) Wave condition 3 d) Wave condition 4 e) Wave condition 5 f) Wave condition 6 g) Wave condition 7 h) Wave condition 8.

in improvement for the uncorrected simulation fidelity across all 3 wave conditions, as summarised in Table III.

Plotting the uncorrected simulation current output of the linear ARX model trained on only wave conditions 1-3, shown in Fig. 6, shows a close fit across the three high frequency wave conditions.

TABLE III
FIDELITY FOR LINEAR ARX MODEL TRAINED ONLY ON WAVE
CONDITIONS 1-3

Wave Condition	Fidelity (%) Uncorrected Simulation
1	73.73
2	77.71
3	64.00

The fact that the linear ARX model performs considerably differently for each wave condition in multi-step prediction, suggests that there may be a dynamic nonlinearity within the data sets. This is also suggested by the results of Table I, where different DoFs become more or less dominant, depending on the wave conditions. Therefore a nonlinear MISO model was also identified, described in Section IV-E.

E. MISO Nonlinear KGP Model Identification

The model structure chosen to identify the dynamic nonlinearities in the system was a Kolmogorov–Gabor polynomial, which adds a polynomial nonlinearity to an ARX model. In [18], it was identified that the cross-product terms of the conventional KGP equation can lead to instabilities in the

model. Therefore, in this paper, the cross product terms of the KGP model were removed. The MISO KGP model used can then be described by Equation (13):

$$
\hat{y}(k) = \sum_{m=1}^{p} \left[-\sum_{i=1}^{n_a} a_{im} y^m(k-i) + \ldots \right.
$$
$$
\left. \sum_{j=1}^{q} \sum_{i=0}^{n_{bj}} b_{ijm} u_j^m(k - n_{kj} - i) \right], \tag{13}
$$

where, p is the maximum polynomial order. The number of unknowns in Equation (13) becomes $p(n_a + n_b + 1)$, where n_b is the sum of the n_b values for each input. Although the KGP model has a nonlinear input-output relationship, the model is still linear in the parameters a_{im} and b_{ijm}, meaning that the KGP unknown parameters are also identified by solving a linear regression problem. Given the close relationship between KGP and linear ARX models, and the complexities in identifying the orders for a non-linear model, it is common practice to use n_a, n_b, and n_k values for a KGP model equal to those for a linear ARX model trained on the same data set [19]. Therefore, the orders used were as identified in Section IV-C. A comparison of model outputs with polynomial orders 2 and 3, identified a best fit at a polynomial order, $n_p = 2$.

F. MISO Nonlinear KGP Model Validation

The model was again validated considering the model fidelity obtained in the 1-step ahead prediction, 10-step ahead prediction, and output simulation, using the validation data. The 10-step ahead prediction shows an improvement for every sea state other than 2, which remains high at 77.20%. The
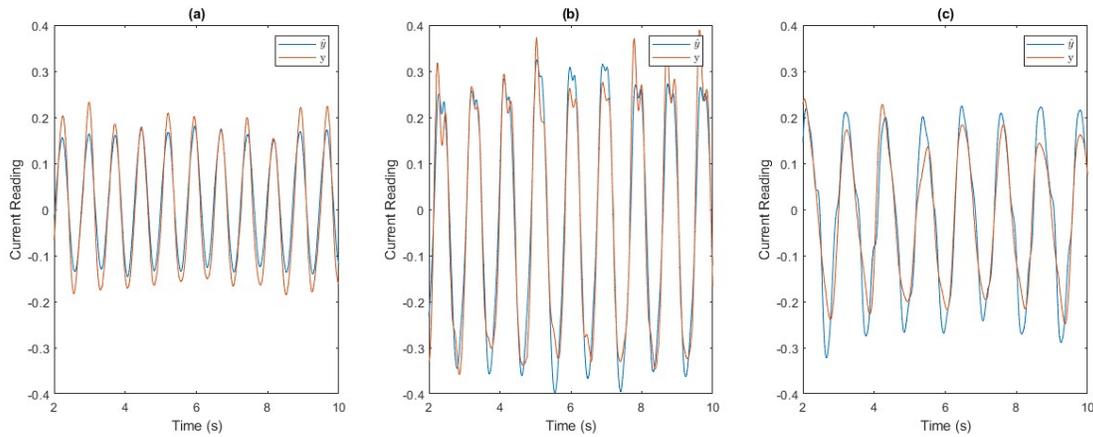
Fig. 6. Uncorrected simulation model output current ($\hat{y}$), compared with validation data (y) for the linear ARX model trained only on wave conditions 1-3. a) Wave condition 1 b) Wave condition 2 c) Wave condition 3.

uncorrected simulation outputs from the KGP model do not show the same improvement across all sea conditions, with some fits improving, and others becoming significantly worse. This suggests a KGP model is a good multi-step predictor of the behaviour of a UL-TENG, but that for applications simulating a UL-TENG with no knowledge of the output data, and using only the relevant inputs, a different approach is required.

TABLE IV
FIDELITY FOR KGP MODEL TRAINED ON ALL 8 WAVE CONDITIONS
(NEGATIVE VALUES INDICATE AN NRMSE > 1)

Wave Condition	Fidelity (%) 1-Step Ahead Prediction	Fidelity (%) 10-Step Ahead Prediction	Fidelity (%) Uncorrected Simulation
1	97.86	75.45	36.67
2	97.56	77.20	62.30
3	98.23	70.21	43.98
4	97.09	45.91	-64.95
5	98.34	75.36	30.69
6	97.89	72.53	25.92
7	98.12	65.99	-0.48
8	98.54	80.32	48.82

The most significant improvement in the KGP model is with the 10-step ahead prediction for wave condition 4. Although far from a perfect fit at only 45.91%, the KGP model 10-step ahead prediction is much closer to the validation signal than that of the linear ARX model, as shown in Fig. 7.

V. SIMULATING BLACK BOX ON B1600S

As the uncorrected simulation fidelities for both the ARX model and the KGP model trained on all 8 sea states were very variable, it was decided to simulate the behaviour of the UL-TENG using the linear ARX model trained on only the three high frequency wave conditions (1, 2, and 3). This black box model was then used to estimate the UL-TENG's output when mounted on a 1:8 Froude-scaled model of the B1600S.
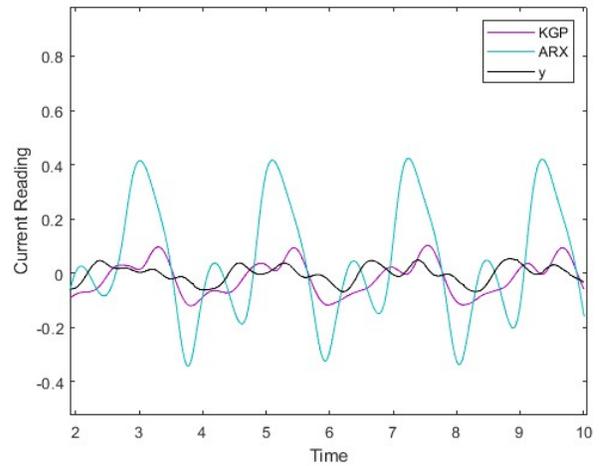
Fig. 7. 10-step ahead prediction model output currents for the KGP and ARX model compared with validation data (y) for wave condition 4.

To attempt to match the frequency and magnitude of the displacement responses of the 1:8 Froude-scaled Sealite Atlantic-2600, a variety of sea states were investigated. However, as the natural frequency of the Balizamar buoy itself is approximately double that of the Sealite one, it was unfortunately not possible to identify a sea state which produced a similar range of displacement amplitudes and frequencies to the training data, and the output current was overpredicted by an order of magnitude, as shown in Fig. 8.

The poor performance of the uncorrected simulation for the Balizamar buoy is perhaps unsurprising, particularly given the narrow range of training data that was required to get a the good uncorrected simulation fidelity in Table III.

VI. DISCUSSION

Tables II and IV, show that for applications where a single-step prediction model is sufficient, both the linear ARX and
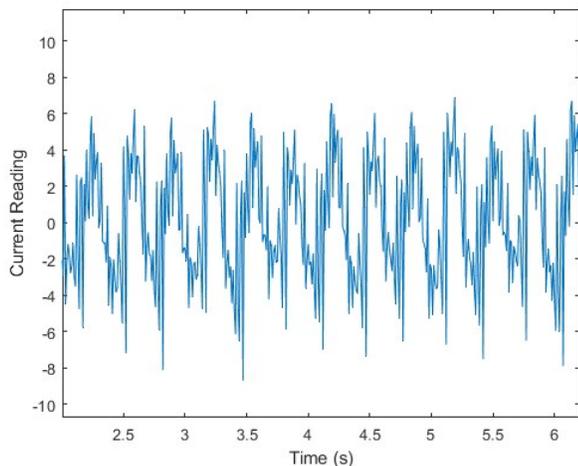
Fig. 8. Uncorrected simulation output using the displacements in 4 DoFs of a 1:8 Froude-scaled Balizamar B1600S buoy. Due to the much higher resonant frequency of the B1600S, the buoy could not be brought within the applicability range of the black box model.

the nonlinear KGP model give a high fidelity output (above 97% for all 8 wave conditions). As both KGP and ARX model fidelities are comparable for single step prediction, the linear ARX model should be used due to its simpler structure. However, apllications for single-step models are limited, as a single step ahead in this case is equivalent to 0.01 s.

The linear ARX model structure is also the most appropriate for cases where the model output can not be uncorrected by measured data, but only when trained on a limited range of wave conditions (1, 2 and 3), which in turn limits the model applicability. Unfortunately, the Balizamar B1600S characteristics differed too greatly from those of the Sealite Atlantic-2600 for the uncorrected black box model to be applicable for the limited range the three wave condition model was trained on. The low uncorrected simulation fidelities of the two models trained on all 8 wave conditions, suggest that a highly applicable, uncorrected black box model of a TENG for simulation purposes across a range of buoys would be difficult to achieve.

For a 10-step ahead prediction, the nonlinear KGP model has some level of validity for all but wave condition 4, where a fidelity of only 45.91% was achieved. The fact that the KGP model improved the predictive ability of the model across the 8 wave conditions, suggests it as the most appropriate choice for use with a predictive controller.

VII. Conclusions

In this paper, three black box models have been identified that predict the dynamic behaviour of a UL-TENG: a linear ARX, and a nonlinear KGP model, each combining the data from 8 different wave conditions; and a linear ARX model focussing only on the data from the 3 high frequency wave conditions. All three models identified were moderately successful in predicting the current output of the UL-TENG,

but only the model trained on just three high frequency wave conditions was able to provide a satisfactory estimation model fidelity for the uncorrected simulation case, where the output is solely dependent on the relevant inputs.

An interesting observation made in this paper is the variability in which DoFs affect the current output most, given different wave conditions. This is a testament to the complexity of a rolling-ball-type TENG as a system. The variability between what DoFs most affect the output current proved a challenge, and even the nonlinear KGP model, which had the best 10-step ahead prediction fidelity for all 8 wave conditions, could not achieve a fidelity above 50% for wave condition 4. The fact that the oscillations of a buoy directly depend upon its resonance frequency, presents a significant challenge for identifying black box models which are applicable for a range of devices. This is coupled with the fact that as different frequencies of motion are experienced different DoFs became dominant for the Sealite Atlantic-2600, which suggests that the motions from other buoys will likely have similarly complex relationships with the TENG's output current.

This paper highlights the difficulty of modelling TENGs from experimental data, particularly the issues from filtering the noise. The lower frequency current wave-forms had significantly more spikes around each peak, where noise could not be further removed without losing signal information. Furthermore, the design of the UL-TENG includes end-stop TENGs which cause additional current spikes when the spheres collide with them. It is possible, that for high frequency motions, these end-stop spikes are not as clear, as a single peak current tends to dominate these signals, or it could be that the subtleties of the end-stop TENGs are removed from the higher frequency signals by noise filtering. High levels of noise are almost inevitable in experimental data from TENGs, as dataloggers need to be highly sensitive due to the low power output. Potential future work could be to develop a simulated test programme in COMSOL, covering a greater range of displacements and their concomitant frequencies, to increase the applicability of the uncorrected model across different WECs. Modelling in COMSOL could also be used to help identify whether the spiking behaviour observed in the lower frequency current data is purely from noise, or whether the additional spikes in current are also influenced by the end-stops in the UL-TENG.

VIII. Acknowledgments

References

[1] D. Clemente, P. Rosa-Santos, and F. Taveira-Pinto, "On the potential synergies and applications of wave energy converters: A review," *Renewable and Sustainable Energy Reviews*, vol. 135, p. 110162, 2021.

[2] I. McLeod and J. V. Ringwood, "Powering data buoys using wave energy: a review of possibilities," *Journal of Ocean Engineering and Marine Energy*, vol. 8, pp. 1–16, 08 2022.

[3] F. Fan, Z. Tian, and Z. L. Wang, "Flexible triboelectric generator," *Nano Energy*, vol. 1, pp. 328–334, 2012.

[4] Z. L. Wang, T. Jianga, and L. Xu, "Toward the blue energy dream by triboelectric nanogenerator networks," *Nano Energy*, vol. 39, pp. 9–23, 2017.

[5] C. Rodrigues, D. Nunes, D. Clemente, N. Mathias, J. M. Correia, P. Rosa-Santos, F. Taveira-Pinto, A. P. T. Morais, and J. Ventura, "Emerging triboelectric nanogenerators for ocean wave energy harvesting: state of the art and future perspectives," *Energy and Environmental Science*, vol. 13, pp. 2657–2683, 2020.

[6] T. Zhao, M. Xu, X. Xiao, Y. Ma, Z. Li, and Z. L. Wang, "Recent progress in blue energy harvesting for powering distributed sensors in ocean," *Nano Energy*, vol. 88, p. 106199, 2021.

[7] L. Ljung, *System Identification: Theory for the User (2nd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[8] ——, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, pp. 1–12, 2010.

[9] H. Cestaro, N. Mathias, N. D. Gonçalves, and T. Morais, "TENG estimation model of voltage production for buoys using particle swarm optimization," *Applied Ocean Research*, vol. 125, p. 103231, 2022.

[10] C. Rodrigues, M. Ramos, R. Esteves, J. Correia, D. Clemente, F. Gonçalves, N. Mathias, M. Gomes, J. Silva, C. Duarte, T. Morais, P. Rosa-Santos, F. Taveira-Pinto, A. Pereira, and J. Ventura, "Integrated study of triboelectric nanogenerator for ocean wave energy harvesting: Performance assessment in realistic sea conditions," *Nano Energy*, vol. 84, p. 105890, 2021.

[11] Sealite, "Sealite Atlantic-2600 navigation buoy," 2021, Accessed on: Jan. 08, 2023. [Online]. Available: https://www.sealite.com/large-ocean-buoy-2600mm-diameter/

[12] Almarin, "B1600S Balizamar buoy," 2021, Accessed on: Dec. 14, 2022. [Online]. Available: https://www.almarin.es/en/marine-aids-to-navigation/navigation-buoys/b1600s-buoys

[13] M. Rosati, T. Kelly, D. G. Violini, and J. V. Ringwood, "Data-based hydrodynamic modelling of a fixed owc wave energy converter," in *Proc. EWTEC 2021*, Plymouth, UK, Sept 2021, pp. 281–1–281–10.

[14] M. Rosati, H. Bingham, B. Joensen, K. Nielsen, and J. V. Ringwood, "A data-based modelling approach for a vented oscillating water column wave energy converter," in *Proc. RENEW 2022*, Lisbon, Portugal, Nov 2022, pp. 339–347.

[15] S. Chen, "Orthogonal-least-squares forward selection for parsimonious modelling from data," *Engineering*, vol. 01, pp. 55–74, 2009.

[16] N. N. Narisetty, "Chapter 4 - Bayesian model selection for high-dimensional data," in *Principles and Methods for Data Science*, ser. Handbook of Statistics, A. S. Srinivasa Rao and C. Rao, Eds. Elsevier, 2020, vol. 43, pp. 207–248.

[17] T. Söderström and P. Stoica, *System Identification*. Upper Saddle River, NJ, USA: Prentice Hall, 1989.

[18] S. Giorgi, J. Davidson, and J. V. Ringwood, "Identification of wave tank models from numerical wave tank data – part 2: Data-based model determination," *IEEE Transactions on Sustainable Energy*, vol. 7, pp. 1020–1027, 2016.

[19] S. Giorgi, J. Davidson, M. Jakobsen, M. Kramer, and J. V. Ringwood, "Identification of dynamic models for a wave energy converter from experimental data," *Ocean Engineering*, vol. 183, pp. 426–436, 2019.