

Failure Warning for Offshore Wind Turbines Based on Autoregressive Models

Hongxu Ye ^a, Wenjin Zhu ^{a *}, He Li ^b, Weidong Ji ^c, C. Guedes Soares ^d, Jin Wang ^b

^a School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an, Shaanxi, 710072, China

^b School of Engineering, Liverpool John Moores University, Liverpool, 3 Byrom Street, L3 3AF, UK

^c Goldwind Sci & Tech Co., Ltd., China

^d Centre for Marine Technology and Ocean Engineering (CENTEC), Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

Abstract: In this paper, a failure warning model for offshore wind turbines is constructed based on a hybrid model consisting of the Autoregressive Integrated Moving Average Model (ARIMA), Least Absolute Shrinkage and Selection Operator (LASSO), and multi-condition Exponentially Weighted Moving Average (EWMA) control charts, namely ARIMA-LASSO-EWMA. A correlation analysis is conducted to select variables that are strongly correlated with others. The ARIMA model is developed for selected variables and their corresponding residuals are computed. The model is then transferred to LASSO regression to extract the secondary residual as the indicator for the health state recognition. Anomalies in wind turbines are founded when secondary residuals deviate from the normal range. Consequently, an abnormal detection mechanism based on multi-condition EWMA is established, which sets different warning thresholds for various operating conditions of the wind turbines. Failure warning is issued when the statistic of the secondary residual exceeds the threshold. The results indicate that the proposed failure warning model can provide alarms more than 10 days in advance of actual faults. The proposed model contributes to the condition monitoring, state recognition, and failure warning for offshore wind turbines.

Keywords: Offshore Wind Turbine, Failure Warning, Operation and Maintenance, LASSO, Multi-Condition EWMA

* Corresponding Author Email: wenjin.zhu@nwpu.edu.cn (Wenjin Zhu)

Nomenclature

z_i	the observation of variable in SCADA systems at time i
$\hat{z}_i$	the predicted expected value for y_i by ARIMA prediction at time i
ϕ_p	the p -th auto-regressive coefficient
φ_q	the q -th moving-average coefficient
ε_i	the residual between z_i and $\hat{z}_i$ at time i
y_i	the observation of state variable at time i
x_{ij}	the observation of j -th selected variable at time i
ε_{y_i}	the residual of the state variable by ARIMA prediction at time i
$\varepsilon_{x_{ij}}$	the residual of the j -th selected variable by ARIMA prediction at time i
β_0, β_j	the coefficients of the LASSO model
λ	the positive regularization parameter of the LASSO model
m	the number of input residuals
$\hat{\varepsilon}_{y_i}$	the predicted expected value for ε_{y_i} by LASSO regression at time i
k	the number of features selected by LASSO regression
e_i	the secondary residual between ε_{y_i} and $\hat{\varepsilon}_{y_i}$ at time i
E_i	the EWMA statistic at time i
r	the smoothing weight parameter
UCL_k	the upper control limit for the k -th operating condition
LCL_k	the lower control limit for the k -th operating condition
μ_k	the centerline for the k -th operating condition
σ_k	the standard deviation for the k -th operating condition
L	the control limit coefficient

23 1 Introduction

24 Wind energy is taking over conventional energy's market share and stepping into the sea [1]. Offshore wind holds a
25 better wind profile, is land-free, and has a higher capability factor than onshore [2, 3]. Therefore, while there are
26 extensive offshore parks of bottom-fixed wind turbines, the trend is to increase floating offshore wind turbines
27 significantly [4, 5]. Offshore Wind Turbines (OWTs) fail more frequently than onshore ones as a consequence of
28 harsh marine environments and complicated structural configurations, which introduce additional operation and
29 maintenance costs for offshore wind projects [6, 7]. The reality is that a failure warning is required [8], which can
30 report the abnormal state of OWTs before a failure turns out to be catastrophic to reduce the failure probability and
31 unwanted downtimes [9, 10].

32 Physics-based methods, statistical-based methods, data-driven-based methods, and hybrid models are typical failure
33 warning methods. Specifically, physics-based techniques use state equations and parameter models to evaluate the
34 equipment condition by predetermined standards or thresholds. For instance, Zhao et al. [11] developed a model
35 based on the physical characteristics of wind turbines (WTs) to predict failures, demonstrating its accuracy in
36 predicting both the occurrence and location of failure in hydraulic pitch actuators. Bechhoefer et al. [12] utilized a
37 simplified Paris law and Kalman smoother to design a failure prediction model based on failure and torque
38 information, providing trends in bearing health. Montesano et al. [13] developed a physics-based multi-scale
39 progressive damage model to predict the evolution of damage in the critical regions of the WT blade structure.
40 Physics-based methods provide highly accurate and interpretable failure warnings, increasingly integrating with data-
41 driven techniques.

42 Statistical-based approaches acquire factors such as variance, kurtosis, skewness, mean value, and root mean square
43 values from time-series data to monitor component failures of the WTs by tracking deviations in these statistical
44 values online during operation. Herp et al. [14] developed a statistical approach to predict WT states online and
45 anticipated the failure state of bearings by performing online inference on temperature measurement residuals. A
46 failure detection and warning strategy was proposed based on fuzzy clustering and the Mahalanobis distance [15],
47 emphasizing its globalization by delivering results for multiple WTs. Bhardwaj et al. [16] proposed a method to
48 estimate the failure rate of an offshore wind turbine gearbox whose failure may cause long downtime and,
49 consequently, high operational and maintenance costs. Zou et al. [17] proposed a combined modified inverse-variance
50 weighting predicting method for oil monitoring and fault pre-warning of WT gearboxes, which accurately predicted
51 a gearbox failure and reduced the error by 10% compared to the single prediction method. Statistical-based methods

provide quantitative failure warnings by identifying patterns and trends from historical data and increasingly integrate with machine learning techniques.

Data-driven-based methods focus on simulating the hidden relationship between the inputs and outputs of the system. Kong et al. [18] proposed a novel condition monitoring method of WTs by convolutional neural networks and gated recurrent units, which accurately recognized the gearbox's condition through the entire residual. Zhu et al. [19] proposed a combined method for predicting operating states based on deep learning and deep domain adaptive networks, detecting potential fault information of WT gearboxes in advance of failures. Wu et al. [20] proposed a data-driven method based on Long Short-Term Memory (LSTM) with Kullback-Leibler divergence to detect failures, providing a 5-day early warning. Fu et al. [21] proposed a two-stage cascaded fault early warning method using machine learning and data visualization, providing early warnings 50 to 60 hours ahead of the Supervisory Control and Data Acquisition (SCADA) system reports such failures. Data-driven-based methods achieve efficient and adaptive failure warnings by utilizing large datasets independently of specific physical models.

Hybrid methods have been developed to enhance the accuracy and robustness of condition monitoring methods. Sun et al. [22] developed an early fault warning method for key transmission components based on nonlinear global local preservation projection and metric learning with Mahalanobis distance, demonstrating its significant utility in preventing mechanical accidents in WTs. Pan et al. [23] proposed a novel performance degradation prediction method based on maximum mean discrepancy and multi-sensor transfer learning, demonstrating its accuracy and adaptation of predicting degradation tendencies. An et al. [24] developed a fault prediction method using tree-based algorithms and unsupervised statistical learning methods based on alarm logs, achieving accuracy and recall of 0.784. Hybrid models integrate the advantages of multiple methods to provide more comprehensive, efficient, and accurate failure warnings.

The failure warning of WTs is based on extracting the sensitive indicator and setting the alarm threshold. The indicator extracted from the SCADA data can reflect the health state of the wind turbine based on the warning threshold calculated based on the historical data and the industrial practice corresponding to various working conditions of wind turbines. In real-time monitoring, the indicator will be extracted from the online SCADA data and compared with the warning threshold. The residual between the actual monitoring value and the predicted value is usually used as the indicator for warning [18, 25, 26], which may not be sensitive enough to detect subtle changes and anomalies in certain types of failures. The warning threshold without considering variations in actual operational conditions is commonly applied [19-21, 24], potentially resulting in failed failure warnings. To this end, a hybrid model consisting of the Autoregressive Integrated Moving Average Model (ARIMA), Least Absolute Shrinkage and Selection Operator

(LASSO), and multi-condition Exponentially Weighted Moving Average (EWMA) control charts, namely ARIMA-LASSO-EWMA, is proposed to consider the temporal nature of the data and excel in dealing with multiple features.

The main contributions of this research are as follows:

- (i) An ARIMA-LASSO model is proposed to select multiple features and extract secondary residuals as indicators.
- (ii) An anomaly detection mechanism based on a multi-condition EWMA control chart is proposed, which sets different thresholds according to the OWT's operational conditions.

The rest of this paper is organized as follows. Section 2 introduces the extraction of the indicator based on the ARIMA-LASSO model and the anomaly detection mechanism based on the multi-condition EWMA control chart. Section 3 demonstrates the results, discussions, and comparisons. Section 4 provides conclusions.

2 Methodology

2.1 The overall framework of the failure warning model

The overall framework and procedure for the failure warning model are outlined as follows and are illustrated in Figure 1:

Step 1: Establish ARIMA prediction models for preliminary variables and state variables selection and as a basis to obtain residuals between actual data and predicted values.

Step 2: Utilize LASSO to select important features from multiple residuals of selected variables and extract the secondary residual of state variables to indicate potential failures.

Step 3: Apply K-means clustering analysis considering the actual operation process to classify OWT operating conditions.

Step 4: Construct a multi-condition EWMA control chart with various thresholds for different operating conditions to detect anomalies and issue failure warnings.

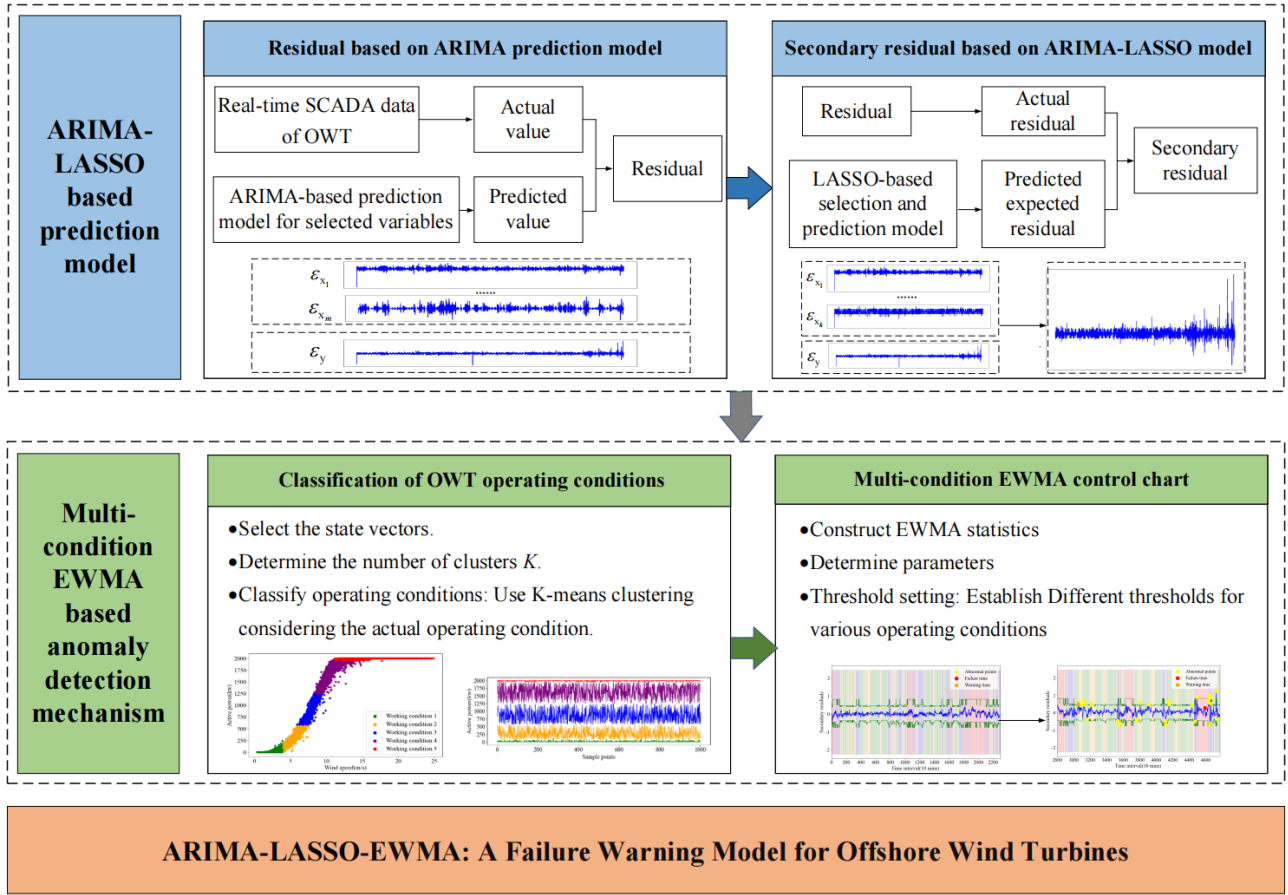


Fig.1. The framework of the proposed failure warning model

2.2 Indicator extraction based on ARIMA-LASSO

The failure warning of wind turbines relies on extracting sensitive indicators and establishing alarm thresholds. In this section, an ARIMA-LASSO model is developed to extract the secondary residual, which is utilized as an indicator for health state recognition. The framework of the ARIMA-LASSO model is illustrated in Fig. 2 and the following steps:

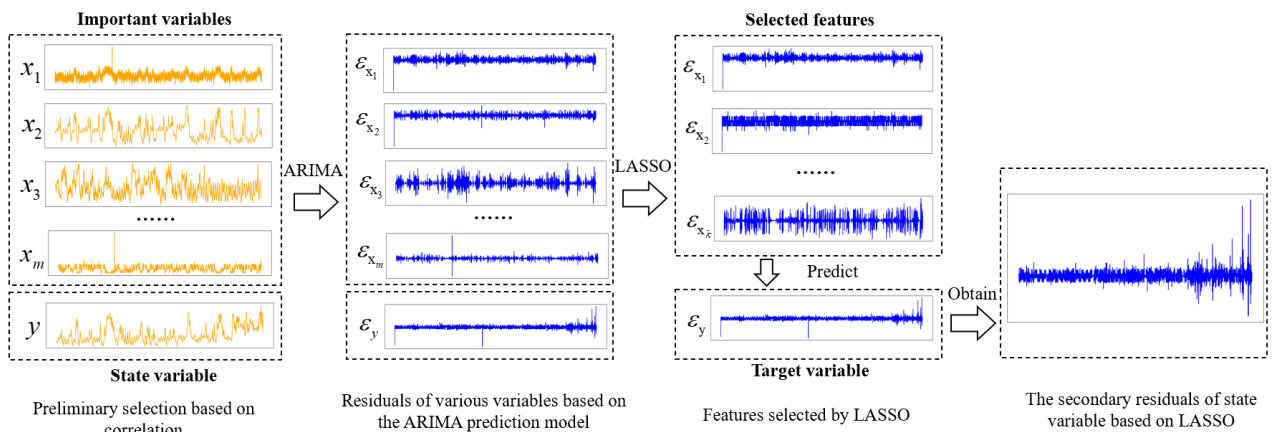


Fig.2. The extraction process of secondary residual

114 **Step 1:** Select important variables and a state variable based on the variation in correlation under different states.

115 **Step 2:** Establish an ARIMA prediction model for each selected variable to obtain the corresponding residual.

116 **Step 3:** Employ LASSO to select the important features from multiple features composed of the residuals from the

117 ARIMA model.

118 **Step 4:** Extract the secondary residual of the state variable as the indicator.

119 **Preliminary selection:** Correlation analysis is conducted to identify target variables using the Spearman correlation

120 coefficient to assess the degree of association among signals. The variables will be selected in case the Spearman

121 correlation coefficients with other signals exhibit significant changes in different states, indicating strong influences

122 of WT states on the target variables. The state variable is the one that shows the greatest change in coefficient. Signals

123 with slight changes in coefficients will be discarded.

124 **ARIMA:** The ARIMA model is a classical statistical model for time series analysis and forecasting, formed by

125 combining the Autoregression (AR) model, the Moving Average (MA) model, and differential Integration (I) [27].

126 ARIMA models can be denoted as $ARIMA(p, d, q)$, where AR: p = order of the auto-regression of the model and

127 $p \in [0, 1, 2, \dots]$; I: d = order of differencing to guarantee the model stationary and $d \in [0, 1, 2, \dots]$; MA: q = order

128 of the moving average aspect of the model and $q \in [0, 1, 2, \dots]$. the observation of variables in SCADA systems at

129 the time i , Z_i is given as follows:

$$Z_i = \sum_{p=1}^P \phi_p Z_{i-p} + \sum_{q=1}^Q \varphi_q \varepsilon_{i-q} + \varepsilon_i \quad (1)$$

131 where ϕ_p is the p -th auto-regressive coefficient, φ_q is the q -th moving-average coefficient, and ε_i is the error

132 term at the time i .

133 For non-stationary time series, d -order difference will be performed first to convert it into a stationary series.

134 Parameters p and q are estimated and determined according to the Auto Correlation Function (ACF) plots, the

135 Partial Auto Correlation Function (PACF) plots, and the Bayesian Information Criterion (BIC) [28]. The residual

136 between the actual monitoring value Z_i and the predicted expected value $\hat{Z}_i$ at time i can be denoted as:

$$\varepsilon_i = Z_i - \hat{Z}_i \quad (2)$$

138 **ARIMA-LASSO:** The classical ARIMA model deals with a single variable and is not capable of modeling

139 interactions among multiple variables. An ARIMA-LASSO model is established based on the classical ARIMA model

140 to consider multiple variables. LASSO is a regression analysis technique that excels in dealing with multiple features.
 141 By adjusting model parameters and shrinking some coefficients towards zero, LASSO can effectively select
 142 important features and reduce the complexity of the model when facing high-dimensional data [29].
 143 ARIMAs are established for the selected variables by correlation analysis. The state variable is selected by correlation
 144 analysis according to the Spearman correlation coefficient, and then residuals of the rest variables are employed as
 145 input using LASSO to retain the important features associated with potential failures. Then the residual of the state
 146 variable serves as the target variable, as:

$$147 \quad \{\hat{\beta}_j, \hat{\beta}_0\} = \arg \min_{\beta_j, \beta_0} \frac{1}{2n} \sum_{i=1}^n \left(\varepsilon_{y_i} - \left(\beta_0 + \sum_{j=1}^m \varepsilon_{x_{ij}} \beta_j \right) \right)^2 + \lambda \sum_{j=1}^m |\beta_j| \quad (3)$$

$$148 \quad \hat{\varepsilon}_{y_i} = \beta_0 + \sum_{j=1}^k \varepsilon_{x_{ij}} \beta_j \quad (4)$$

$$149 \quad e_i = \varepsilon_{y_i} - \beta_0 - \sum_{j=1}^k \varepsilon_{x_{ij}} \beta_j \quad (5)$$

150 where m is the number of input residuals, n represents the sample size, k denotes the number of features selected
 151 by LASSO and $k \leq m$. β_0 and β_j reflect coefficients of the LASSO model, λ is the positive regularization
 152 parameter, ε_{y_i} is the residual of the state variable from ARIMA prediction at the time i , $\hat{\varepsilon}_{y_i}$ is the predicted
 153 residual of the state variable by LASSO at the time i , $\varepsilon_{x_{ij}}$ displays the residual of the j -th selected variable from
 154 ARIMA prediction at the time i , e_i is the difference between the actual and predicted residuals at the time i .

155 **The secondary residual:** The residual between the predicted and actual value is used as the indicator, which is
 156 affected by the health state variations and random factors and is insensitive to minor changes in the operational state.
 157 To this end, the secondary residual, which represents the difference between the actual and predicted residuals is
 158 proposed in this paper. LASSO regression is employed to select k features from the m input residuals ($k \leq m$).
 159 Based on the selected features, the residual of the state variable ε_{y_i} is predicted to obtain the secondary residual e_i
 160 of the state variable. The model is trained by the data corresponding to a healthy state of the OWTs, and the parameters
 161 are estimated to adapt to the features of the healthy state. Conversely, when the data indicating potential failures are
 162 tested by the trained model, it will result in a large output e_i compared with the secondary residual using the data
 163 from the healthy state. Consequently, the secondary residual is extracted as the indicator for further warning analysis,

164 reflecting the variation of the health states of the wind turbine.

165 **2.3 Anomaly detection mechanism based on multi-condition EWMA chart**

166 This section introduces an anomaly detection mechanism based on the multi-condition EWMA control chart. The K-
167 means clustering algorithm is applied and integrated with the actual operation of the OWTs to classify each OWT's
168 operating conditions. The multi-condition EWMA control chart is constructed to detect the anomalies, establishing
169 diverse warning thresholds corresponding to different operating conditions to achieve failure warning.

170 **Operating condition classification of OWT:** OWTs continuously switch operating conditions during their operation
171 due to changes in the external environment and control strategy. After some data preprocessing, such as removing
172 abnormal values like power curtailment and outliers, the operating conditions of OWTs are classified by combining
173 the actual operating conditions of the OWTs with the K-means clustering.

174 The fundamental concept of K-means clustering is to classify data based on similarity and iterate until the differences
175 between clusters are maximized [30]. The operating conditions of the OWT can be viewed as state vectors composed
176 of a series of data from sensors, and the similarity between these state vectors is measured by their Euclidean distance.
177 The representative sensor parameters from the operation of OWTs, including wind speed and power, are selected for
178 clustering analysis [31]. The number of clusters, K , is calculated using the Calinski–Harabasz criterion and
179 considering the actual OWT operating procedures [30].

180 **Multi-condition EWMA control chart:** The secondary residual is proposed as an indicator that exhibits differences
181 between health data and fault data. The abnormal conditions of OWTs are indicated when the monitor indicator
182 deviates from the normal range. The setting of warning thresholds is the foundation of the stability of the warning
183 model, which is often set based on industrial experience. A multi-condition EWMA control chart is proposed in this
184 paper to avoid delayed or false warnings caused by experiential thresholds. The method utilizes Statistical Process
185 Control (SPC) and considers the changes in the operational states of wind turbines by setting various fault thresholds
186 according to different operating conditions.

187 The SPC chart is a statistical tool for monitoring process stability and quality control [32]. Exponentially Weighted
188 Moving Average (EWMA) is applied to monitor residuals as it is sensitive to small shifts in data and is not sensitive
189 to the normality assumption of variable data [33]. The secondary residuals e_i are smoothed using EWMA, where
190 prior residual information is fully incorporated by setting weight coefficients and computing the weighted average of
191 the current observation and historical statistics to reflect recent trends in residuals [34]. The EWMA statistic for e_i
192 is obtained as follows:

$$E_i = (1-r)E_{i-1} + re_i \quad (6)$$

where E_i is the EWMA statistic at the time i , i.e., the EWMA statistic of e_i , and r is the smoothing weight parameter and $r \in (0,1)$.

An important performance metric for assessing control charts is the Average Run Length (ARL), which denotes the average number of samples taken before an alarm is triggered from monitoring begins [35]. A failure warning is triggered when E_i exceeds the corresponding upper and lower control limits UCL_k and LCL_k .

$$UCL_k = \mu_k + L\sigma_k \sqrt{\frac{r}{(2-r)}[1-(1-r)^{2i}]} \quad (7)$$

$$LCL_k = \mu_k - L\sigma_k \sqrt{\frac{r}{(2-r)}[1-(1-r)^{2i}]} \quad (8)$$

where μ_k is the centerline for k -th operating condition and usually the average of k -th operation condition is chosen as the centerline. σ_k is the standard deviation for k -th operating condition, L is the control limit coefficient.

Based on the differences in e_i between healthy and fault states, and considering the ARL value, the smoothing parameter r and the control limit parameter L are set to avoid triggering out-of-control signals during the healthy state [36]. Furthermore, different control limits are defined based on varying means and standard deviations under different operating conditions.

3 Results, discussion, and comparisons

3.1 The dataset

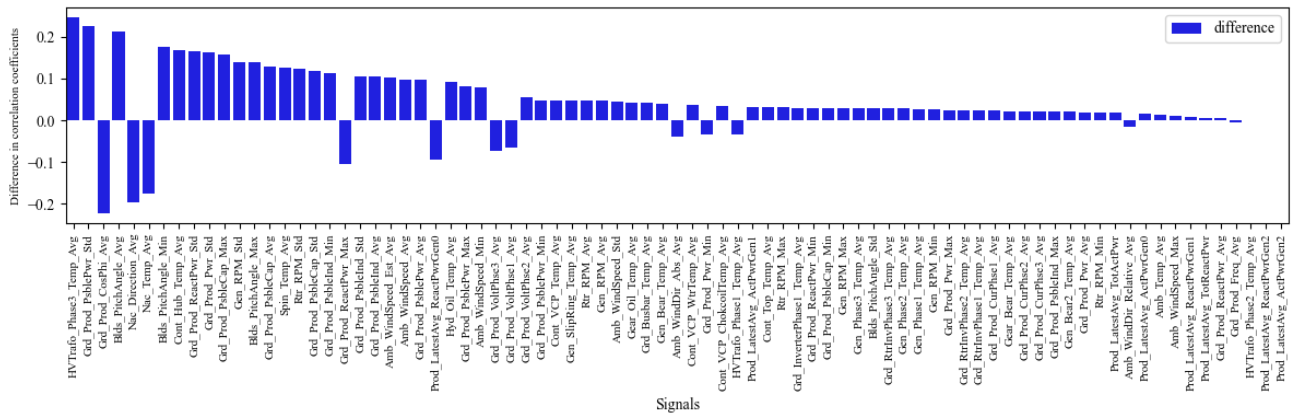
The data used in this study originates from an open dataset provided by EDP (Energias de Portugal). The offshore wind farm is located in the sea area near the Gulf of Guinea in Africa, and the OWTs operate with a cut-in wind speed of 4 m/s, cut-out wind speed of 25 m/s, rated power of 2 MW, and rated wind speed of 12 m/s. The dataset collects SCADA monitoring signals recorded at 10-minute intervals covering operational signals from 4 OWTs between January 1, 2016, and December 31, 2017. It comprises environmental parameters such as ambient temperature, wind speed, and wind direction, as well as operational parameters like gearbox oil temperature, generator bearing temperature, transformer temperature, and hydraulic unit oil temperature, among a total of 82 signals.

218

219 **3.2 Case illustration**

220 **(1) Extraction of indicator based on ARIMA-LASSO model**

221 After the SCADA data are cleaned to remove abnormal values such as power curtailment and outliers [37], differences
222 in Spearman correlation coefficients for 82 signals between the health and fault states of the transformer are calculated,
223 as shown in Fig. 3. The signal with the greatest difference is ‘HVTrafo_Phase3_Temp_Avg’, representing the average
224 temperature on the phase L3 of the high-voltage transformer. Hence ‘HVTrafo_Phase3_Temp_Avg’ is selected as the
225 state variable, effectively characterizing the transformer state. Some variables exhibit slight changes (less than 0.0274)
226 in coefficients, indicating that they are insensitive to the change in health state and can be discarded. Based on a
227 statistical significance test using p -values, 51 variables are selected, as shown in Table 1.



228

229 Fig.3. Difference in correlation coefficients of 82 signals between health and fault states

230 Table 1 The selected variables by correlation analysis

	Signal	Description	Difference
1	HVTrafo_Phase3_Temp_Avg	Average Temperature of High-Voltage Transformer Phase L3 (°C)	0.2461
2	Grd_Prod_PsblePwr_Std	Standard Possible Active Power of Grid (kw)	0.2249
3	Grd_Prod_CosPhi_Avg	Average Actual Phase Shift of Grid	-0.2231
4	Blds_PitchAngle_Avg	Average Angle of Blades (°)	0.2117
5	Nac_Direction_Avg	Average Nacelle Direction (°)	-0.1974
6	Nac_Temp_Avg	Average Temperature Inside Nacelle (°C)	-0.1764
7	Blds_PitchAngle_Min	Minimum Angle of Blades in Recent Average Period (°)	0.1744
8	Cont_Hub_Temp_Avg	Average Temperature of Hub Controller (°C)	0.1664
9	Grd_Prod_ReactPwr_Std	Standard Reactive Power of Grid (kVar)	0.1661

10	Grd_Prod_Pwr_Std	Standard Power of Average Period (kw)	0.1617
11	Grd_Prod_PsbleCap_Max	Maximum Possible Capacitive Reactive Power (kVAr)	0.1573
12	Gen_RPM_Std	Standard RPM of Generator in Average Period (rpm)	0.1396
13	Blds_PitchAngle_Max	Maximum Angle of Blades in Recent Average Period (°)	0.1385
14	Grd_Prod_PsbleCap_Avg	Average Possible Capacitive Reactive Power (kVAr)	0.1288
15	Spin_Temp_Avg	Average Temperature Inside Nacelle Cone (°C)	0.1258
16	Rtr_RPM_Std	Standard Rotor RPM in Recent Period (rpm)	0.1237
17	Grd_Prod_PsbleCap_Std	Standard Possible Capacitive Reactive Power (kVAr)	0.1187
18	Grd_Prod_PsbleInd_Min	Minimum Possible Inductive Reactive Power (kVAr)	0.1115
19	Grd_Prod_ReactPwr_Max	Maximum Reactive Power of Grid (kVAr)	-0.1050
20	Grd_Prod_PsbleInd_Std	Standard Possible Inductive Reactive Power (kVAr)	0.1044
21	Grd_Prod_PsbleInd_Avg	Average Possible Inductive Reactive Power (kVAr)	0.1039
22	Amb_WindSpeed_Est_Avg	Average Wind Speed in Average Time Base (m/s)	0.1008
23	Amb_WindSpeed_Avg	Average Wind Speed over the Average Time Base (m/s)	0.0963
24	Grd_Prod_PsblePwr_Avg	Average Possible Active Power of Grid (kw)	0.0961
25	Prod_LatestAvg_ReactPwrGen0	Reactive Power at Generator Terminal (Varh)	-0.0939
26	Hyd_Oil_Temp_Avg	Average Temperature of Hydraulic Oil in the System (°C)	0.0925
27	Grd_Prod_PsblePwr_Max	Maximum Possible Active Power of Grid (kw)	0.0823
28	Amb_WindSpeed_Min	Minimum Wind Speed over the Average Time Base (m/s)	0.0797
29	Grd_Prod_VoltPhse3_Avg	Average Voltage of Phase 3 (V)	-0.0741
30	Grd_Prod_VoltPhse1_Avg	Average Voltage of Phase 1 (V)	-0.0665
31	Grd_Prod_VoltPhse2_Avg	Average Voltage of Phase 2 (V)	0.0542
32	Grd_Prod_PsblePwr_Min	Minimum Possible Active Power of Grid (kw)	0.0478
33	Cont_VCP_Temp_Avg	Average Temperature of Controller on VCP Board (°C)	0.0477
34	Gen_SlipRing_Temp_Avg	Average Temperature of Generator in Open Loop Chamber (°C)	0.0472
35	Rtr_RPM_Avg	Average Rotor RPM (rpm)	0.0466
36	Gen_RPM_Avg	Average Generator RPM (rpm)	0.0460
37	Amb_WindSpeed_Std	Standard Wind Speed over the Average Time Base (m/s)	0.0440
38	Gear_Oil_Temp_Avg	Average Temperature of Gearbox Oil (°C)	0.0427
39	Grd_Busbar_Temp_Avg	Average Temperature of Bus Segment (°C)	0.0414

40	Gen_Bear_Temp_Avg	Average Temperature of Generator Bearing 1 (Non-Drive End) (°C)	0.0401
41	Cont_VCP_WtrTemp_Avg	Average Temperature of VCS Controller Cooling Water (°C)	0.0375
42	Cont_VCP_ChokcoilTemp_Avg	Average Temperature of Controller on VCS Choke Coil (°C)	0.0343
43	Prod_LatestAvg_ActPwrGen1	Active Power Generator Connected in Delta (Wh)	0.0324
44	Cont_Top_Temp_Avg	Average Temperature of Top Nacelle Controller (°C)	0.0315
45	Rtr_RPM_Max	Maximum Rotor RPM in Recent Average Period (rpm)	0.0310
46	Grd_InverterPhase1_Temp_Avg	Average Temperature Measured for IGBT Driver on Grid-Side Inverter (°C)	0.0300
47	Gen_RPM_Max	Maximum Generator RPM in Recent Average Period (rpm)	0.0286
48	Gen_Phase3_Temp_Avg	Average Temperature of Phase 3 Stator Windings in Generator (°C)	0.0285
49	Grd_RtrInvPhase3_Temp_Avg	Average Temperature Measured for Phase 3 IGBT Driver on Rotor-Side Inverter (°C)	0.0278
50	Gen_Phase2_Temp_Avg	Average Temperature of Phase 2 Stator Windings in Generator (°C)	0.0277
51	Gen_Phase1_Temp_Avg	Average Temperature of Phase 1 Stator Windings in Generator (°C)	0.0274

231

232 Individual ARIMA prediction models are established for the selected variables listed in Table 1, with the parameter
233 d determined by a data stationarity test and p and d determined by ACF and PACF plots [28]. The respective residuals
234 between actual monitoring values and ARIMA predictions are then computed. Based on the residual of the state
235 variable ‘HVTrafa_Phase3_Temp_Avg’, the residuals of selected variables are screened to identify important features,
236 thereby reducing model complexity and preventing overfitting. The positive regularization parameter λ is
237 determined to be 0.0033 through cross-validation. Subsequently, the residual of the state variable is predicted based
238 on the 22 features selected by LASSO, and the secondary residual is extracted as the indicator for failure warning.
239 The overall procedure is summarized in Figure 4.

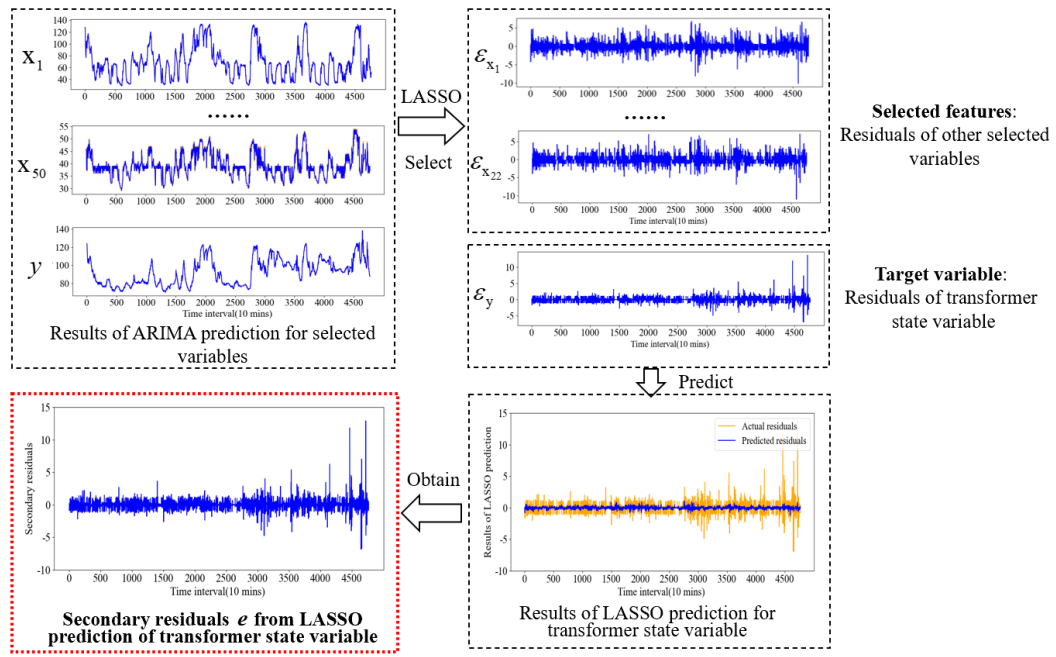


Fig.4. The extraction process of secondary residuals of transformer state variable

(2) Operating condition classification of OWT

The operational states of the OWT are classified using the K-means clustering method considering actual operational conditions. Following the Calinski–Harabasz criterion and engineering expertise [30], the number of clusters is set to be 5, and thus five operation conditions are defined as follows:

- Operating condition 1: Wind speed is below the cut-in speed, indicating the turbine is in stoppage mode.
- Operating condition 2: Wind speed meets the startup threshold with low power output.
- Operating condition 3: Wind speed gradually increases, and blade angles are adjusted to maximize wind capture, effectively converting wind energy into electricity.
- Operating condition 4: Wind speed approaches rated speed, and power output steadily approaches rated capacity.
- Operating condition 5: Stable wind conditions achieve consistently rated power output, with the turbine maintaining stable operation by adjusting rotor speed and other parameters to optimize wind capture.

Fig. 5 depicts the classification of OWT operating conditions. Distinct ranges of wind speed and power output are clearly delineated across various operating conditions. One thousand sample points are taken from each condition for comparison.

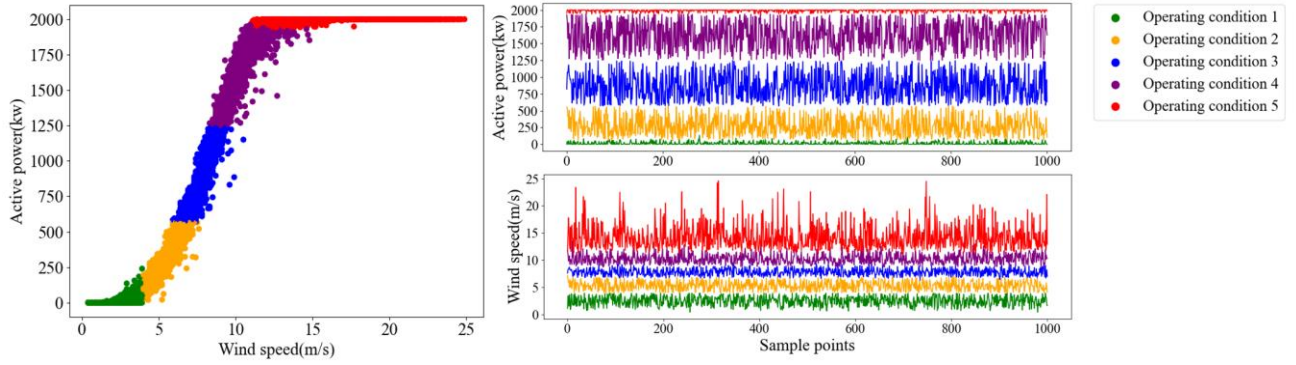


Fig.5. The classification results and comparisons

(3) Construction of multi-condition EWMA control chart

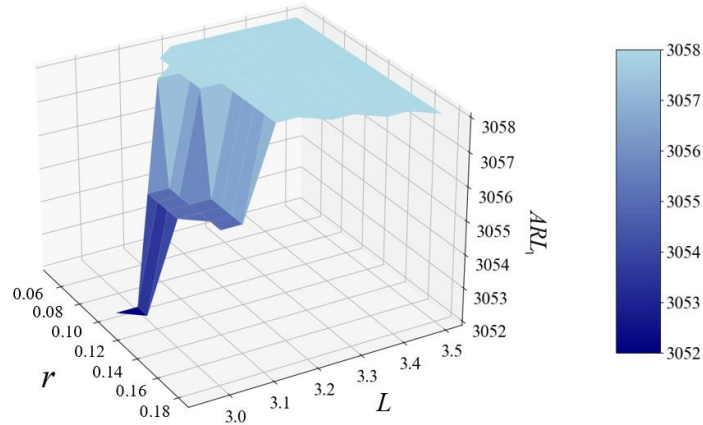
Statistical tests indicate that the secondary residual e_i follows a normal distribution. Different warning thresholds are established for various operating conditions to construct a multi-condition EWMA control chart. The parameters L and r are determined based on Average Run Length (ARL) [35], which denotes the average number of samples taken before an alarm is triggered since monitoring began. The EWMA statistic E_i and multiple control limits are subsequently calculated, and an alarm is triggered when $E_i < LCL$ or $E_i > UCL$. Typically, the ARL under the controlled state is denoted as ARL_0 , while the ARL under an out-of-control state is denoted as ARL_1 . A larger ARL_0 value in the healthy state is preferable, as it indicates a lower probability of false alarms. In general, the values of r in the interval $0.05 \leq r \leq 0.25$ work well in practice, and $L = 3$ (the usual three-sigma limits) work reasonably well, particularly with the larger value of r . When r is small there is an advantage in reducing the width of the limits [38]. Consequently, the value range of $0.05 \leq r \leq 0.25$ and $2 \leq L \leq 3.5$ are considered in this research. Since the monitor indicator of health data should be maintained in control, the detailed parameter combinations are obtained, as indicated in Table 2, ensuring that no alarms are triggered under a healthy state in these control charts.

Table 2 Parameter combinations for maintaining health data in control

L	r	L	r
2.95	0.1	3.25	0.05~0.15
3	0.1~0.11	3.3	0.05~0.16
3.05	0.09~0.12	3.35	0.05~0.16
3.1	0.08~0.14	3.4	0.05~0.17
3.15	0.06~0.14	3.45	0.05~0.17
3.2	0.05~0.15	3.5	0.05~0.18

276

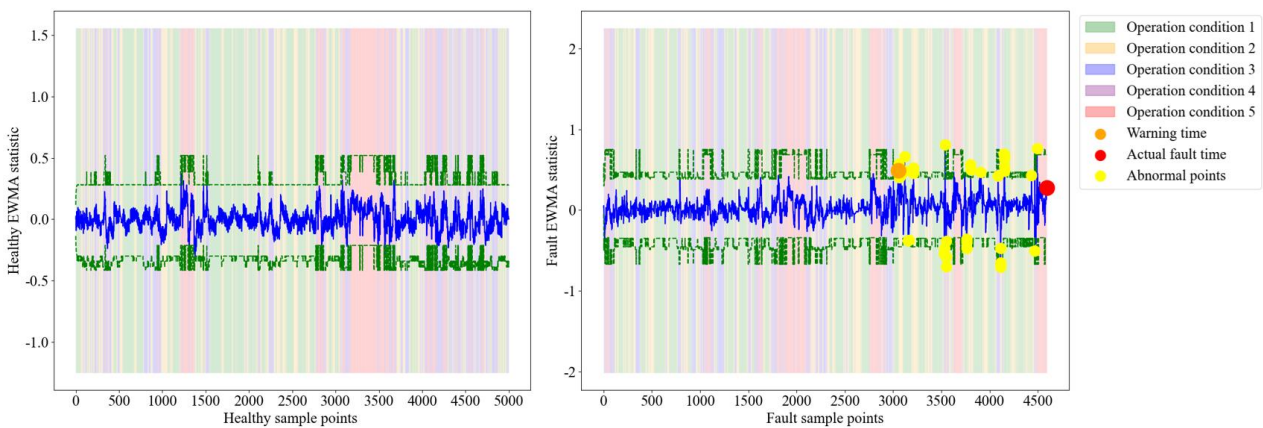
277 For failure data, a smaller ARL_1 value is preferable to ensure that anomalies can be detected earlier, leading to more
 278 efficient failure warnings. The respective selected parameter combinations under fault period data are presented in
 279 Fig. 6, with the actual failure occurring at the 4,592 sample point. ARL_1 decreases with decreased L and r ,
 280 leading to an earlier alarm. The earliest alarm in all control charts with the selected parameter combination is at the
 281 3,052th sample point. Considering a larger L can reduce the probability of false alarms, while a smaller r can
 282 enhance the sensitivity in detecting minor deviations, the optimal parameter combination is determined to be $L = 3$
 283 and $r = 0.1$. The multi-condition EWMA control chart is obtained based on the optimal parameter combination. The
 284 control chart for the healthy data and fault data is presented in Fig. 7.



285

286 Fig.6. The ARL_1 value for selected parameter combination under fault period data

287



288

289 Fig.7. Multi-condition EWMA control chart based on the optimal parameter combination.

290

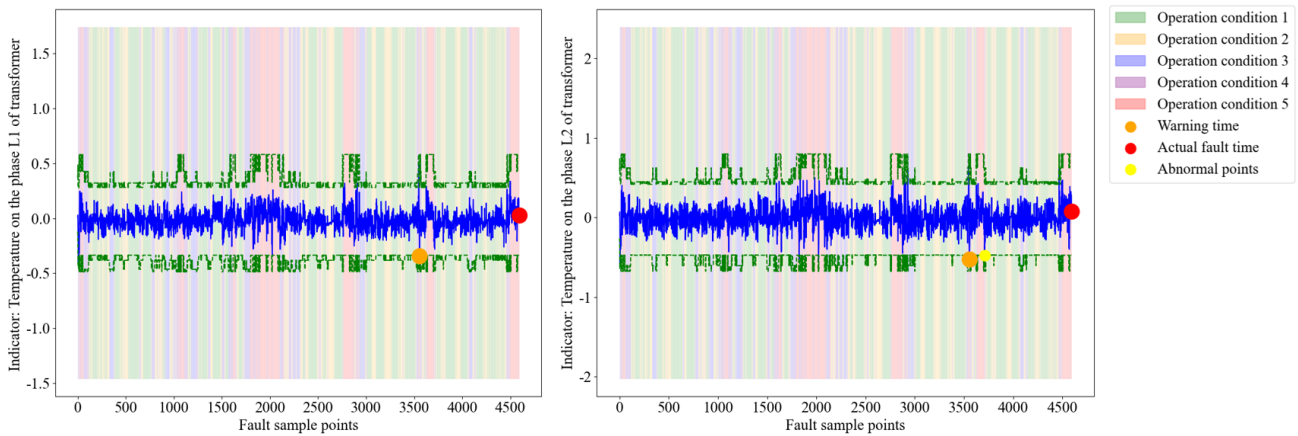
291 In the multi-condition EWMA control charts, different background colors represent various operating conditions of
 292 the OWT, with distinct thresholds set for each condition. As shown in Fig. 7, the EWMA statistic of the secondary

residuals remains within the control threshold under healthy conditions, indicating a controlled state. The statistic exceeds the upper control limit at the 3,052 sample point, and multiple anomalies are detected within a short period subsequently, while the actual failure occurs at the 4,592 sample point. Consequently, the alarm can be issued 256 hours and 40 minutes in advance, approximately 10.6 days prior to the actual failure, thereby achieving an effective failure warning.

(4) Analysis of state variable

The signal “HVTrafo_Phase3_Temp_Avg” is considered a state variable as it exhibits the greatest variation in correlation with other signals, which can characterize the transformer state effectively. In the dataset, other signals related to the transformer include “HVTrafo_Phase1_Temp_Avg” and “HVTrafo_Phase2_Temp_Avg”, which represent the average temperature on the phase L1 and L2 of the high-voltage transformer. To evaluate the impact of different state variables on the warning effectiveness, ARIMA-LASSO models are used to extract secondary residuals from these two signals, and multi-condition EWMA control charts based on the respective optimal parameter combination are constructed for failure warning. The results are shown in Fig. 8.

306



307

308

309

Fig.8. The warning results based on related signals

Fig. 8 shows that the secondary residual of “HVTrafo_Phase1_Temp_Avg” serves as the monitor indicator. The earliest alarm is issued at the 3,551 sample point, with no further anomalies detected afterward. The same situation happens when using the secondary residual of “HVTrafo_Phase2_Temp_Avg”. Therefore, the failure warning is more effective and sensitive when the signal “HVTrafo_Phase3_Temp_Avg” is a state variable that can provide the earliest alarm. Accordingly, the effectiveness of the method in identifying state variables based on the correlation variation under different states is also demonstrated.

315

3.3 Comparisons

The proposed model is compared with:

- (i) A classical univariate ARIMA-based warning model. An ARIMA model is used to predict the state variable. The mean of residuals with a sliding window of size ten between monitored and predicted values is used for assessment, with warning thresholds determined based on the residual range observed during the healthy state. This method aims to evaluate the superiority of the multi-feature approach compared to the single-variable method.
- (ii) Initial ARIMA-LASSO-based warning model. A model utilizes the secondary residual of the state variable as the indicator, with warning thresholds determined based on the residual range during the healthy state. This comparison assesses the validity of employing multi-condition EWMA control charts for threshold determination.
- (iii) LSTM-based warning model. A model utilizes an LSTM network to predict state variables considering multiple features. The mean of residuals with a sliding window of size ten between monitored and predicted values is used for assessment, with warning thresholds determined based on the residual range observed during the healthy state. This method assesses the efficacy of secondary residual in the proposed approach.

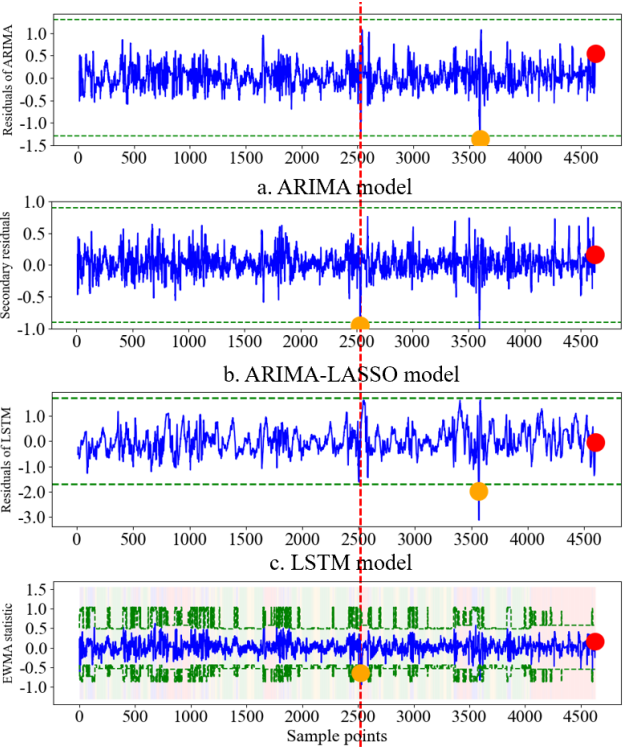
(1) Warning time

Failure warning tests are conducted for various types of failures of the four OWTs in the dataset: T01, T06, T07, and T11, encompassing gearbox faults, generator faults, and hydraulic unit faults. The above three approaches utilize the range of residuals observed during healthy conditions as warning thresholds, while, the proposed methods are formulated based on Equations (7) and (8). Specific warning threshold values and parameters are presented in Table 3. Comparisons of warning effectiveness are depicted in Figs. 9 and 10, with warning results detailed in Table 4.

Table 3 Warning thresholds and parameters of four methods for various faults

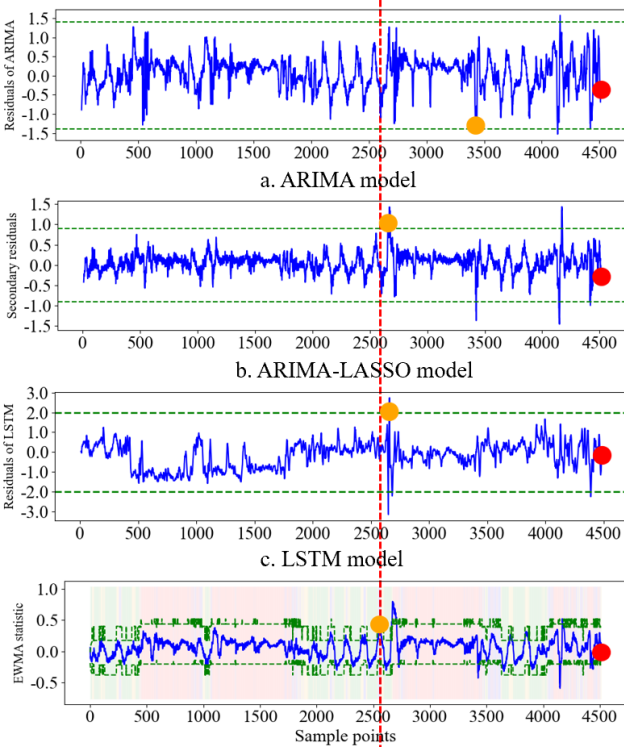
	Classical univariate ARIMA	Initial ARIMA- LASSO	LSTM	ARIMA-LASSO and Multi- condition EWMA	
				L	λ
T01_Gearbox	(-1.3,1.3)	(-0.9,0.9)	(-1.7,1.7)	3	0.1
T06_Generator	(-1.3,1.3)	(-0.9,0.9)	(-2.0,2.0)	3	0.05
T11_Hydraulic unit	(-1.1,1.1)	(-1.0,1.0)	(-1.2,1.2)	3	0.1
T07_Transformer_1	(-1.3,1.3)	(-0.9,0.9)	(-2.5,2.5)	3	0.1

T07_Transformer_2	$(-1.3,1.3)$	$(-0.9,0.9)$	$(-2.5,2.5)$	3	0.1
T07_Generator_1	$(-1.3,1.3)$	$(-0.9,0.9)$	$(-2.0,2.0)$	3	0.05
T07_Generator_2	$(-1.3,1.3)$	$(-0.9,0.9)$	$(-2.0,2.0)$	3	0.05



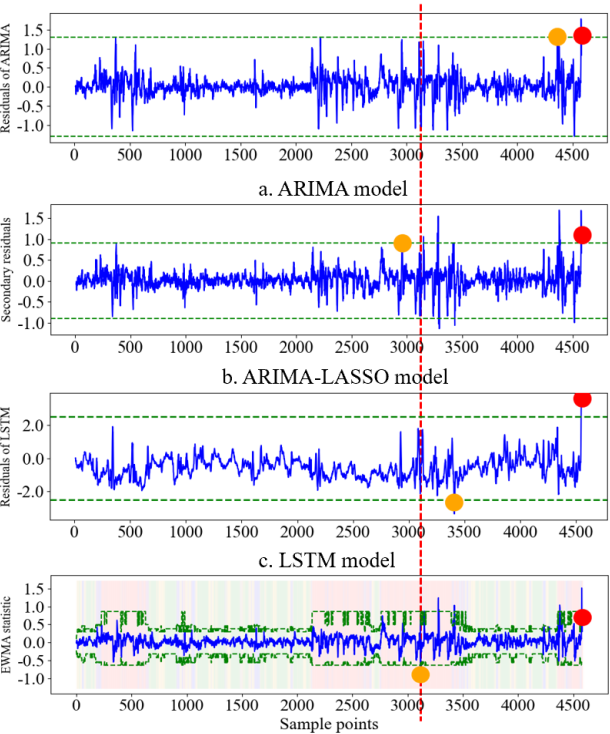
d. ARIMA-LASSO and multi-condition EWMA

(A)

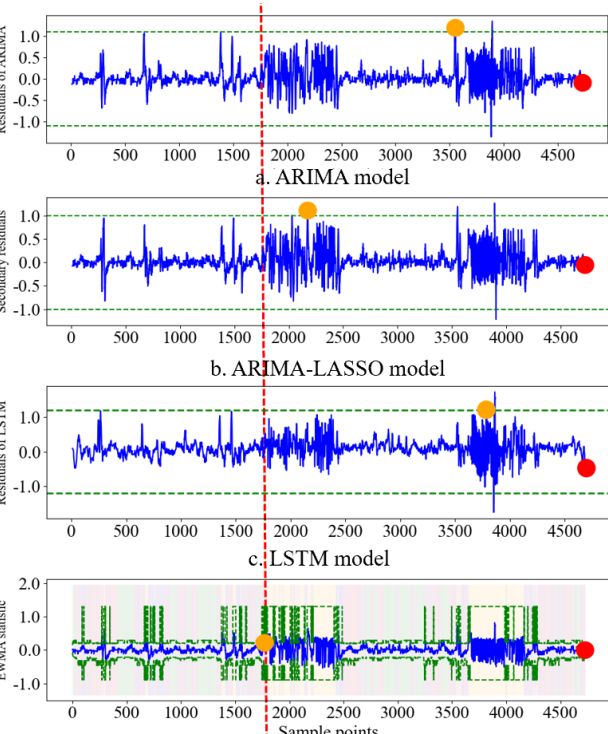


d. ARIMA-LASSO and multi-condition EWMA

(B)



d. ARIMA-LASSO and multi-condition EWMA



d. ARIMA-LASSO and multi-condition EWMA

342
343
344
345

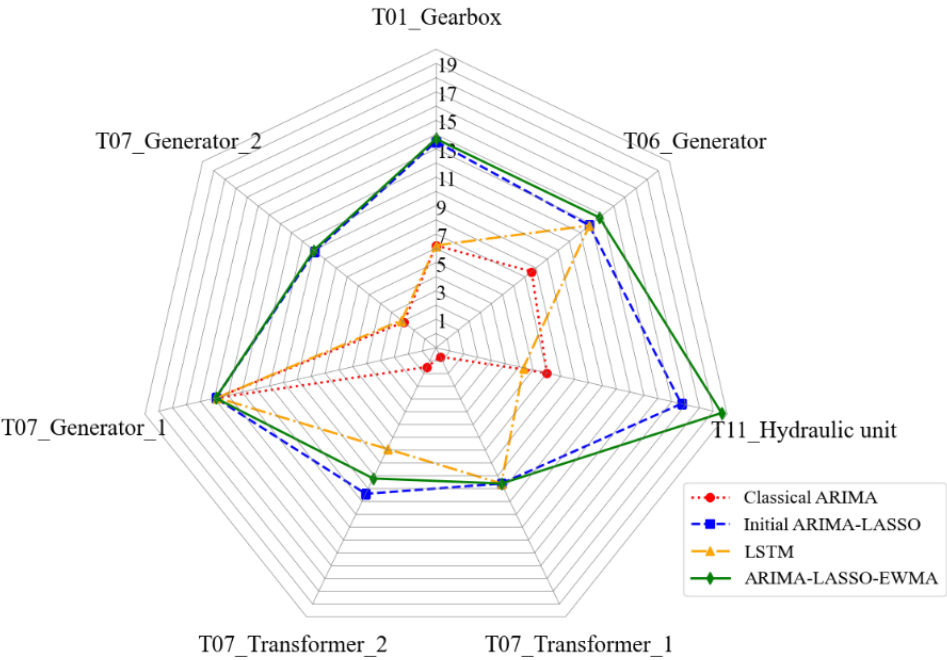
(C) (D)

Fig.9. Comparison of failure warning effectiveness on various OWTs. (A) Gearbox fault in T01. (B) Generator fault in T06. (C) Transformer fault in T07. (D) Hydraulic unit fault in T11.

Table 4 Comparison of time before actual failure on four OWTs / Days

	Univariate ARIMA	Initial ARIMA-LASSO	LSTM	ARIMA-LASSO-EWMA
T01_Gearbox	7.2	14.5	7.2	14.7
T06_Generator	8.6	13.8	13.8	14.7
T11_Hydraulic unit	8	17.7	6.3	20.6
T07_Transformer_1	0.7	10.6	10.6	10.6
T07_Transformer_2	1.5	11.4	7.9	10.2
T07_Generator_1	15.9	15.9	15.8	15.9
T07_Generator_2	2.9	10.9	3.1	11

346



347
348
349

Fig.10. Comparison of failure warning effectiveness

350 In the comparison of the aforementioned warning methods, the proposed ARIMA-LASSO-EWMA method can issue
351 the earliest warnings for various component faults in different OWTs. Compared to the LSTM-based warning model,
352 the effectiveness of using secondary residuals as the monitor indicator is clearly illustrated. The multi-condition

EWMA control chart incorporating the actual operational conditions of OWTs, can reduce the frequency of “missed warnings” and “false warnings”, while maintaining high warning efficacy. Utilizing secondary residuals extracted by the ARIMA-LASSO model as an indicator facilitates effective state monitoring of OWT components. By integrating this indicator into the anomaly detection mechanism based on a multi-condition EWMA control chart, failure warnings can be issued more than ten days in advance, thereby reducing failure probability and unwanted downtime and ultimately enhancing the energy production efficiency of OWTs.

(2) Precision

The precisions of the above prediction models for healthy data preceding various faults are presented in Fig. 11. The mean absolute error (MAE) and mean squared error (MSE) values of the proposed model are smaller and more concentrated, indicating that the proposed model's precision is superior to other methods in the healthy data. Consequently, the proposed prediction model enhances the prediction of component health trends, making the indicator more sensitive to fault data. This aids in more accurately recognizing equipment states, thereby reducing false alarms, avoiding unnecessary downtime, and ultimately improving the energy production efficiency of OWTs.

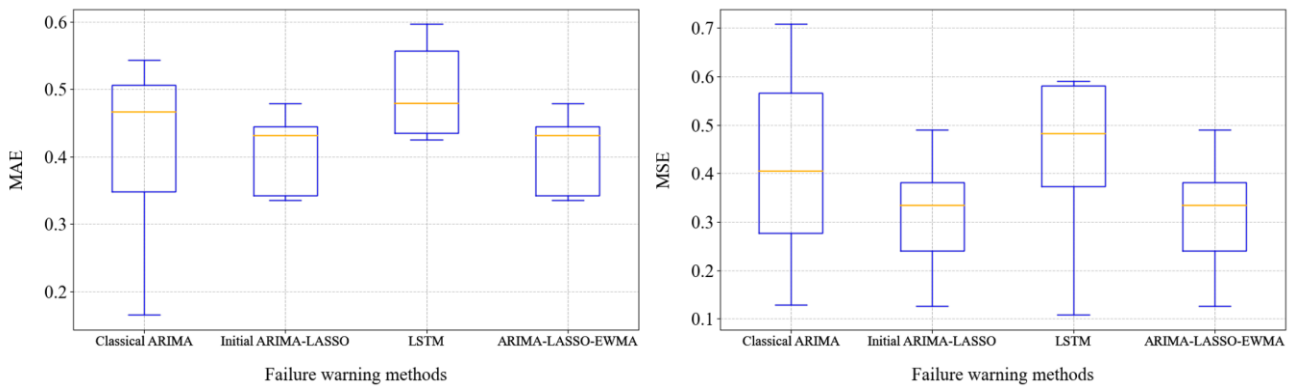


Fig.11. Precision comparisons of the prediction models

4 Conclusions

This research proposes a failure warning model for offshore wind turbines based on the Autoregressive Integrated Moving Average Model (ARIMA), Least Absolute Shrinkage and Selection Operator (LASSO), and multi-condition Exponentially Weighted Moving Average (EWMA) control charts. The ARIMA-LASSO model is employed to select from multiple features and extract secondary residuals as an indicator. The multi-condition EWMA control chart is utilized to detect anomalies and issue alarms, with different thresholds set according to the various operational conditions of the OWTs. The proposed method is applied to failure warning across various faults in different OWTs

and is compared with other methods. The results indicate that the ARIMA-LASSO-EWMA failure warning model provides the earliest warnings, more than ten days in advance. It is important to evaluate the severity of the potential failure regarding the random time interval between the first alarm and the actual failure in the future.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NO. 72471188 and NO. 72231008) and the Science and Technology Innovation Team of Shaanxi Provincial [NO. 2024RS-CXTD-28]. This paper is funded by the Horizon Europe Marie Skłodowska-Curie Postdoctoral Fellowship [DROMS-FOWT – 101146961] and UKRI under grant no, EPSRC EP/Z001501/1. This work contributes to the Strategic Research Plan of the Centre for Marine Technology and Ocean Engineering (CENTEC), which is financed by the Portuguese Foundation for Science and Technology (Fundação para a Ciência e Tecnologia - FCT) under contract UIDB/UIDP/00134/2020.

Disclaimer

This paper is the authors' opinion and does not necessarily represent the beliefs and policies of their employers.

References

- [1]. B. Johnston, A. Foley, J. Doran, T. Littler, Levelised cost of energy, A challenge for offshore wind. *Renewable Energy* 160 (2020) 876-885.
- [2]. H. Díaz, C. Guedes Soares. Review of the current status, technology and future trends of offshore wind farms. *Ocean Engineering* 209 (2020): 107381.
- [3]. H. Li, C. Guedes Soares. Assessment of failure rates and reliability of floating offshore wind turbines. *Reliability Engineering and System Safety* 228 (2022): 108777.
- [4]. J. Carroll, A. McDonald, D. McMillan, Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines. *Wind Energy* 19 (6) (2016) 1107-1119.
- [5]. J. Kang, C. Guedes Soares. An opportunistic maintenance policy for offshore wind farms. *Ocean Engineering* 216 (2020): 108075.
- [6]. J. Kang, J. Sobral, C. Guedes Soares. Review of condition-based maintenance strategies for offshore wind energy. *Journal of Marine Science and Application* 18 (2019): 1-16.
- [7]. S. Koukoura, M. N. Scheu, A. Kolios, Influence of extended potential-to-functional failure intervals through condition monitoring systems on offshore wind turbine availability. *Reliability Engineering and System Safety* 208 (2021) 107404.
- [8]. H. Li, C. Huang, C. Guedes Soares. A real-time inspection and opportunistic maintenance strategies for floating offshore wind turbines. *Ocean Engineering* 256 (2022): 111433.
- [9]. X. Zhao, B. Guo, Y. Chen, A condition-based inspection-maintenance policy for critical systems with an

unreliable monitor system. *Reliability Engineering and System Safety* 242 (2024) 109710.

[10]. C. Wu, R. Pan, X. Zhao, X. Wang, Designing preventive maintenance for multi-state systems with performance sharing. *Reliability Engineering and System Safety* 241 (2024) 109661.

[11]. H. Zhao, S. Lian, L. Shao, Fault prediction of pitch actuator for wind turbines. *Applied Mechanics and Materials* 721 (2015) 397-401.

[12]. E. Bechhoefer, R. Schlanbusch, Generalized Prognostics Algorithm Using Kalman Smoother IFAC-Papers OnLine. 48(21) (2015) 97-104.

[13]. J. Montesano, H. Chu, C. V. Singh, Development of a physics-based multi-scale progressive damage model for assessing the durability of wind turbine blades. *Composite Structures*. 141 (2016) 50-62.

[14]. J. Herp, M. H. Ramezani, M. Bach-Andersen, N. L. Pedersen, E. S. Nadimi, Bayesian state prediction of wind turbine bearing failure. *Renewable Energy*. 116 (2018) 164-172.

[15]. R. R. de la Hermosa González, Wind farm monitoring using Mahalanobis distance and fuzzy clustering. *Renewable Energy*. 123 (2018) 526-540.

[16]. U. Bhardwaj, A. P. Teixeira, and C. Guedes Soares. Reliability prediction of an offshore wind turbine gearbox. *Renewable Energy* 141 (2019): 693-706.

[17]. X. Zou, J. Zhang, J. Chen, O. Orozovic, X. Xie, J. Li, Oil Monitoring and Fault Pre-Warning of Wind Turbine Gearbox Based on Combined Predicting Method. *Sustainability*. 15(4) (2023) 3802.

[18]. Z. Kong, B. Tang, L. Deng, W. Liu, Y. Han, Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units. *Renewable Energy*. 146 (2020) 760-768.

[19]. Y. Zhu, C. Zhu, J. Tan, C. Song, D. Chen, J. Zheng, Fault detection of offshore wind turbine gearboxes based on deep adaptive networks via considering Spatio-temporal fusion. *Renewable Energy*. 200 (2022) 1023-1036.

[20]. Y. Wu, X. Ma, A hybrid LSTM-KLD approach to condition monitoring of operational wind turbines. *Renewable Energy*. 181 (2022) 554-566.

[21]. Y. Fu, S. Wang, F. Jia, Q. Zhou, X. Ge, Two-Stage Cascaded High-Precision Early Warning of Wind Turbine Faults Based on Machine Learning and Data Graphization. *Journal of Electrical Engineering and Technology*. 19(3) (2024) 1919-1931.

[22]. G. Sun, Y. Wang, J. Li, Research on early fault warning methods for key transmission components of wind turbines. *Measurement Science and Technology*. 32(11) (2021) 115002.

[23]. Y. Pan, R. Hong, J. Chen, J. Feng, W. Wu, Performance degradation assessment of wind turbine gearbox based on maximum mean discrepancy and multi-sensor transfer learning. *Structural Health Monitoring*. 20(1) (2021) 118-138.

[24]. J. An, X. Hu, L. Gong, Z. Zou, L. R. Zheng, Fuzzy reliability evaluation and machine learning-based fault prediction of wind turbines. *Journal of Industrial Information Integration*. 40 (2024) 100606.

- [25]. G. A. Skrimpas, T. Ursin, C. Sweeney, K. Marhadi, N. Mijatovic, J. Holboell 2017. Residual signal feature extraction for gearbox planetary stage fault detection. *Wind Energy*. 20(8) (2017) 1389-1404.
- [26]. X. Wu, H. Wang, G. Jiang, P. Xie, X. Li, Monitoring wind turbine gearbox with echo state network modeling and dynamic threshold using SCADA vibration data. *Energies*. 12(6) (2019) 982.
- [27]. C. Velasco-Gallego, I. Lazakis, Real-time data-driven missing data imputation for short-term sensor data of marine systems. A comparative study. *Ocean Engineering*, 218 (2020) 108261.
- [28]. D. Song, J. Yang, Y. Liu, M. Su, A. Liu, Y. H. Joo, Wind direction prediction for yaw control of wind turbines. *International Journal of Control, Automation and Systems*. 15(4) (2017) 1720-1728.
- [29]. Z. Cai, L. Li, L. Yu, C. Li, M. Sun, Diversity, quality, and quantity of real ship data on the black-box and gray-box prediction models of ship fuel consumption. *Ocean Engineering*. 291 (2024) 116434.
- [30]. Y. Huang, S. Liu, H. Chen, J. Dai, X. Wang, H. Tao, Dynamic behavior analysis method for the normal and faulty drivetrain of wind turbine under multi-working conditions. *Energy Science and Engineering*. 11(10) (2023) 3619-3640.
- [31]. J. M. Ha, H. Oh, J. Park, B. D. Youn, Classification of operating conditions of wind turbines for a class-wise condition monitoring strategy. *Renewable Energy*. 103 (2017) 594-605.
- [32]. S. Zhang, L. Xue, Z. He, Y. Liu, Z. Xin, A sensitized variable selection control chart based on a classification algorithm for monitoring high-dimensional processes. *Quality and Reliability Engineering International*. 39(7) (2023) 2837-2850.
- [33]. Y. Mahmood, S. Ishtiaq, M. B. Khoo, S. Y. Teh, H. Khan, Monitoring of three-phase variations in the mortality of COVID-19 pandemic using control charts: where does Pakistan stand? *International Journal for Quality in Health Care*. 33(2) (2021) 1-8.
- [34]. L. Song, S. He, Z. Wang, Z. He, Dynamic monitoring of polynomial profiles with attribute responses and between-profile correlation. *IIE Transactions*. 56(8) (2023) 870-885.
- [35]. L. Xue, Q. Wang, Z. He, P. Qiu, A nonparametric EWMA control chart for monitoring mixed continuous and count data. *Quality Technology and Quantitative Management* 21(5) (2024) 749-765.
- [36]. L. Xue, Q. Wang, L. An, Z. He, S. Feng, J. Zhu, A nonparametric adaptive EWMA control chart for monitoring mixed continuous and categorical data using self-starting strategy. *Computers and Industrial Engineering*. 188 (2024) 109930.
- [37]. R. Morrison, X. Liu, Z. Lin, Anomaly detection in wind turbine SCADA data for power curve cleaning. *Renewable Energy*. 184 (2022) 473-486.
- [38]. D. C. Montgomery, *Introduction to Statistical Quality Control* (7th ed.). John Wiley and Sons, Inc. 2012.