Journal of Physics: Conference Series

PAPER • OPEN ACCESS

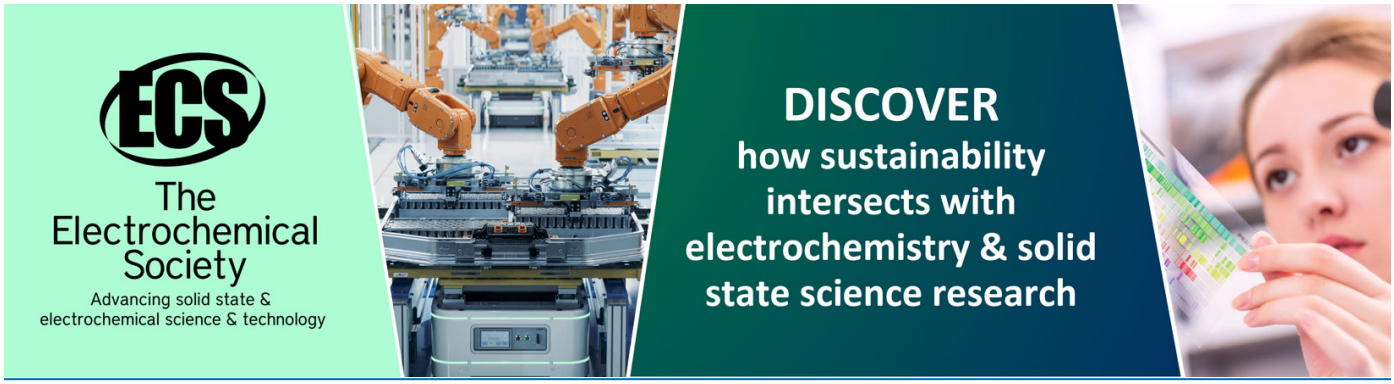
Temporal Causal Inference in Wind Turbine SCADA Data Using Deep Learning for Explainable AI

To cite this article: Joyjit Chatterjee and Nina Dethlefs 2020 *J. Phys.: Conf. Ser.* **1618** 022022

View the article online for updates and enhancements.

You may also like

- Investigation of deep transfer learning for cross-turbine diagnosis of wind turbine faults
Ping Xie, Xingmin Zhang, Guoqian Jiang et al.
- Wind turbine gearbox fault prognosis using high-frequency SCADA data
Ayush Verma, Donatella Zappalá, Shawn Sheng et al.
- Transfer condition monitoring across wind turbines using feature alignment and parameter fine-tuning
Shuai Chen, Zijian Qiao, Chongyang Xie et al.



This content was downloaded from IP address 3.135.183.187 on 24/04/2024 at 06:02

Temporal Causal Inference in Wind Turbine SCADA Data Using Deep Learning for Explainable AI

Joyjit Chatterjee* and Nina Dethlefs

University of Hull, Department of Computer Science & Technology, Cottingham Road, Hull, HU6 7RX, United Kingdom

E-mail: j.chatterjee-2018@hull.ac.uk

Abstract. Machine learning techniques have been widely used for condition-based monitoring of wind turbines using Supervisory Control & Acquisition (SCADA) data. However, many machine learning models, including neural networks, operate as black boxes: despite performing suitably well as predictive models, they are not able to identify causal associations within the data. For data-driven system to approach human-level intelligence in generating effective maintenance strategies, it is integral to discover hidden knowledge in the operational data. In this paper, we apply deep learning to discover causal relationships between multiple features (confounders) in SCADA data for faults in various sub-components from an operational turbine using convolutional neural networks (CNNs) with attention. Our technique overcomes the black box nature of conventional deep learners and identifies hidden confounders in the data through the use of temporal causal graphs. We demonstrate the effects of SCADA features on a wind turbine's operational status, and show that our technique contributes to explainable AI for wind energy applications by providing transparent and interpretable decision support.

1. Introduction

With the global increase in carbon emissions caused by conventional sources of energy, there has been a recent, yet fast paced move towards renewable energy sources [1]. Wind energy will play an integral role in the energy revolution [2], encouraging research and development in this domain. Wind turbines are highly complex systems [3] consisting of electrical and mechanical components that regularly suffer from operational inconsistencies, making Operations and Maintenance (O&M) challenging, which in turn can cause major costs and downtime [4].

Despite a growing interest in the wind industry in using machine learning for condition-based monitoring of turbines using Supervisory Control & Acquisition (SCADA) data[1] [5], the key problem with such data is the presence of multiple features which share a common cause (confounders) and can be hidden within complex and non-linear operational data. Although neural networks have recently demonstrated high accuracy in predictive maintenance [6], the black-box nature of such models has limited their ability to support transparent decision making. Actions, intervention and decisions have important consequences in data-driven decision making [7], and causal reasoning is an integral step in achieving human-level machine intelligence. Causal inference has played an important role in applications in medical treatments [8] and

[1] SCADA data consists of multiple simultaneous measurements of parameters like operational status of sub-components, meteorological variables, sensor measurements etc., see Section 3 for details.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

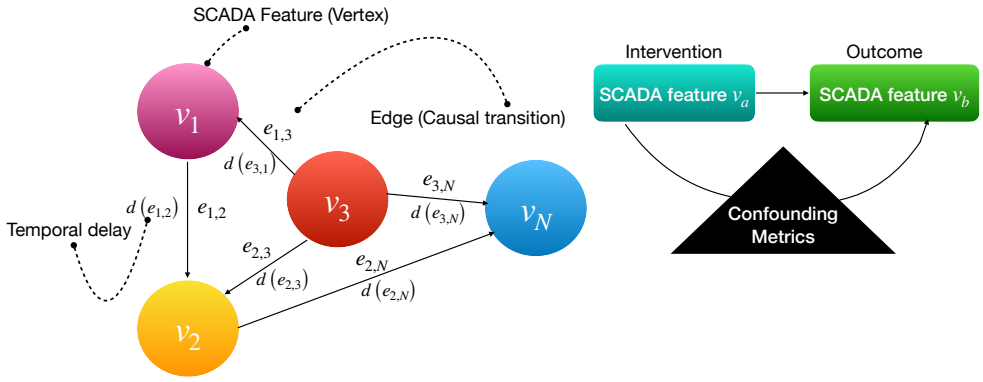


Figure 1. Depiction of the causality learning task.

stock markets [9], e.g. In this vein of research, Wang et al. [10] have previously proposed the *deconfounder*, an algorithm which uses probabilistic models for causal inference in real-world datasets. However, they only model the effects of univariate causes on an outcome and ignore situations where multiple causes can lead to multiple outcomes, as e.g. in SCADA features in wind turbines. The application of causal inference to wind energy has been limited. The closest is work by Felgueira et al. [11], who applied Normal Behaviour Models to study how causal inference improves the accuracy of fault-classification from SCADA, but they do not model the temporal causal graphs for this purpose. Also, Felgueira et al's approach is mostly hypothetical as they did not have access to labelled fault data.

Our primary contribution in this article is the use of temporal causal graphs that can accurately model the relationships between SCADA features and turbine faults as well as the time-delay between cause and effect. We harness the representational power of convolutional neural networks (CNN) with attention in discovering causal relationships [12] from observed time-series and historical error logs, and provide a comparison with recent state-of-art methods [8]. The proposed technique can identify hidden relationship in the data and thereby generate new insights on causal connections in wind turbine O&M. We show that not all these relations are obvious and some (that may potentially be discarded as insignificant noise) can contribute significantly to the model's learning process. Our study shows that causal inference can contribute towards enhancing the reliability of turbines, and support decision making in O&M by making neural machine learning models more transparent and interpretable.

2. Methodology

2.1. Learning Task

Consider a multivariate time series of various parameters (hence called features) in the SCADA data represented by $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N\} \in R^{N \times L}$, where N denotes the total number of SCADA features and L denotes the length (total number of observations) in each time-series. Every ith feature, where $i \in N$ is represented as $\mathrm{X}_1 \ldots \mathrm{X}_L$. Our learning task here is to identify the causal association between these N features (wherein each feature can have multiple cause-effect relationships), the time delay between the cause and effect and finally construct a temporal causal graph based on these relationships. As we are interested in identifying directionality between the SCADA features during different faults, our task requires the construction of a directed causal graph [13, 12] $\mathcal{G}_{\mathcal{SCADA}} = (V, E)$ with a collection of vertices and edges, with each vertex $v_i \in V$ denoting a SCADA feature, and each edge connection $e_{a,b} \in E$ describing causal transitions from vertex v_a to v_b. Further, in accordance with the notation in [12] for multivariate time-series observations, our goal is to identify the temporal delay $d(e_{i,j})$ in the

occurrence of cause and effect i.e. the number of time-steps (or observations), after which one SCADA feature causally affects another. Figure 1 illustrates the structure of a directed graph modelling the causal relationships in our scenario.

2.2. Learning Model

We propose attention-based convolutional neural networks (CNNs) as first proposed by Nauta et al.[12] for causal inference, which has been shown effective for financial and neuroscientific prediction making. We specifically explore the method's applicability to time-series SCADA features with continuous values and historical fault logs. We modify the original architecture to develop temporal causal graphs for identifying hidden confounders in an operational turbine. The description of different components of our learning model is provided below:-

(i) **Time-series prediction of SCADA features:** The first step of our causality learning model involves time-series prediction of SCADA features. As outlined before, given a sequence of N different SCADA features, ranging from $\mathrm{X}_1 \ldots \mathrm{X}_N$, with each feature consisting of L observations (total length of the time-series), the goal is to predict the future values $\hat{X}_i$ for each SCADA time-series, based on its present and past values. To accomplish this, we use a CNN architecture. CNNs are a special family of feed-forward neural networks which have achieved notable advances in computer vision [14], and have recently successfully handled sequential data such as time-series [15]. CNNs consist of various convolutional layers with a sliding kernel, giving them the ability to identify novel patterns in time-series data and forecast future values of a target time series. A kernel is a filter (or a matrix of weights in mathematical terms), computed as a convolution (dot product) of the input time series and the present filter weights W. Considering that future j values are to be predicted for each SCADA feature X_i i.e. $L - j$ observations in the feature are used as inputs (training data) to the CNN with a kernel size of S (usually determined experimentally during hyper-parameter optimisation), the dot product is computed as:

$$\mathcal{W} \odot \left[X_i^{j-S+1}, X_i^{j-S+2} \ldots, X_i^{j-1}, X_i^{j} \right]. \tag{1}$$

This process is specific to predicting a univariate time-series, such as a specific SCADA feature. However, as we aim to model a multi-variate time series, we utilise N independent CNNs denoted by $\mathcal{CNN}_1 \ldots \mathcal{CNN}_N$. Each CNN $\mathcal{N}_j$ predicts the corresponding time-series $\hat{\mathbf{X}}_j$ (i.e. $\mathcal{CNN}_1$ predicts the first SCADA feature, $\mathcal{CNN}_2$ predicts the second SCADA feature etc.). Alongside the predicted time-series, the kernel weights $\mathcal{W}_j$ for each SCADA feature are also output by the CNN. Please refer to Figure 2 for an illustration.

In conventional CNNs, the number of time steps used by the sliding kernel (referred to as receptive field) is the same as the user specified kernel dimensions. However, for temporal causal inference, the receptive field needs to be larger than the delay between cause and effect identified by the model. To this end, a dilation mechanism is used that helps increase the dimensions of the receptive field, and identifies causal relationships even in situations with large delay between multiple causes. With the dilation rate denoted as K, $K = 1$ denotes a conventional CNN, while $K = 2$ skips 1 time step (observations), $K = 3$ skips 2 time steps etc., and $K = n$ skips $n - 1$ time steps. This is helpful whenever the time delay affecting causality between the SCADA features is large, and also reduces computational expense in updating the kernel weights over time [16]. For a given set of features from the exogenous SCADA time series, the goal is to predict the target time series, ensuring that the loss L between true values ($\mathbf{X}_j$) and predicted values ($\hat{\mathbf{X}}_j$) is minimised. Although CNNs can model time-series with high accuracy [15], like other deep learning models, they are inherently black-box in nature and do not provide transparent and interpretable rationales for their predictions. We therefore use an attention mechanism, as outlined below:

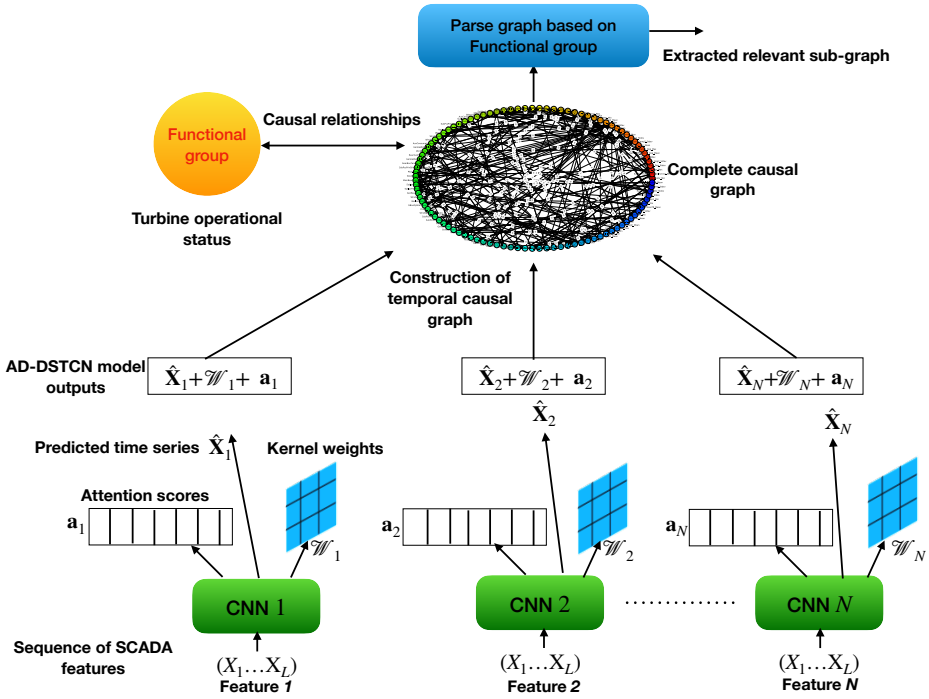


Figure 2. Illustration of a CNN with attention for temporal causal inference in SCADA data

(ii) **Attention mechanism for discovering potential causes:** An *attention mechanism* is used on top of the existing CNN architecture in order to identify the specific features in the SCADA dataset which are causally related to the target predictions. The attention mechanism helps identify the particular inputs which the model focuses on while making its predictions. We base our work on [17], and integrate it with the causality identification framework by [12] for our study. Consider the N SCADA features which are predicted in step (1) above by the N individual CNNs, amongst which the causal relationships are to be identified. The attention vector is basically a row vector with the same number of elements as our SCADA features (dimensions $1 \times N$), and at each time-step of the SCADA time-series with L observations (i.e. from time-step 1 to time-step L), the attention score is computed by multiplication of the input feature $\hat{\mathbf{X}}_j$ with the updated kernel weights in Equation 1. From this, we obtain an individual attention vector for each of our predicted features as in Equation 2, where, j denotes the individual time steps per feature.

$$\mathbf{a_j} = [a_{1,j}, a_{2,j}, \ldots, a_{N,j}] \tag{2}$$

Finally, all these row vectors are concatenated together to generate an attention score square matrix as per Equation 3. As a simple representation, given that we have N SCADA features, assume an $N \times N$ attention square matrix (similar to an adjacency matrix), wherein, every element $a_{i,j}$ denotes the attention score of feature X_i to be a probable cause of X_j. Here, the diagonal elements of the matrix (wherein $i = j$) denote self-causation [12].

$$\mathbf{A_{i,j}} = \begin{bmatrix} \mathbf{a}_{1,1} & \mathbf{a}_{1,2} \ldots & \mathbf{a}_{1,N} \\ \mathbf{a}_{2,1} & \mathbf{a}_{2,2} \ldots & \mathbf{a}_{2,N} \\ \ldots & & \\ \mathbf{a}_{N,1} & \mathbf{a}_{N,2} \ldots & \mathbf{a}_{N,N} \end{bmatrix} \tag{3}$$

Given these attention scores, when our model predicts a time-series of SCADA features,

the relative importance of all other features is obtained, with higher scores signifying that a feature contributes more to the cause. The user can manually specify the top k values to use for causal inference from the attention matrix. The scores, together with the kernel weights in step (1) are then used to construct the temporal causal graph in step (3) below.

(iii) **Construction of temporal causal graph:** A combination of CNNs with dilation in step (1), and the attention mechanism in (2) leads to the Attention-based Dilated Depthwise Separable Temporal Convolutional Network (AD-DSTCN) [12]. Till this stage, we had N independent CNNs which predict the N different time-series SCADA features, and output the attention scores and kernel weights. These parameters are now used construct the temporal causal graph and discover delays, see Figure 2. As constructing the causal graph from all parameters will lead to too many relationships (many of which being non-sensible), we compute a *significance measure* $s(c, e)$ for a specific cause-effect pair by isolating the impact of cause c on effect e [18]. The significance measure (in a range $[0, 1]$) determines when the increase in model loss between the causing feature, where sudden events/changes in time occur, and the outcome (the feature which is affected causally) is sufficient for labelling potential causes as true (after validation) [12]. The higher the significance measure, the lesser the constraints on the model to identify hidden confounders during inference. Finally, once the temporal causal relationships are identified, the relevant sub-graph for a given anomaly can be extracted from the complete causal graph depending on the type of fault in the turbine sub-component (referred to as Functional Group in our study). Given a type of fault which occurs in the turbine, if a feature is important, any alteration in its values will cause a change in the confounding features, and will thus be included in the graph. Conversely, if the feature is not important, any alterations in its time-series values would not affect other confounding features, and would not be included in the graph.

3. Data Description and Preprocessing

We use SCADA data from an operational offshore turbine rated at 7 MW (Levenmouth Demonstration Turbine (LDT))[2]. For preprocessing, data from the substation, met mast and turbine were merged at identical timestamps to facilitate identification of the turbine operational status. Corrupted, missing values and outliers were also dealt with at this stage, see [6] for details. Features include electrical, pressure, and temperature readings etc. obtained from sensors, and meteorological information like wind speed and direction, outside temperature etc. We had historical logs of processed events for the turbine, where a fault can occur in any of 14 different functional groups (such as pitch system, gearbox, yaw brake etc.). For the purpose of our study, an anomaly is considered to have occurred whenever a fault is raised in the processed events data for alarms in between a specific time duration (*TimeOn*, when the alarm was started and *TimeOff*, when the alarm was cleared). All other circumstances were considered normal operation. To ensure that only anomalies are considered which have occurred as a consequence of an actual fault (and not due to requested shutdowns), the forced outages in the turbine are considered based on the unique alarms list to be an indicator of an anomaly. We obtained SCADA data with 21,392 measurements at generally 10-minute intervals, each containing 102 features. Any fault which occurs in a turbine sub-component belongs to one of the 14 functional groups mentioned above. Please see Github[3] and Chatterjee and Dethlefs [6] for more details on data preprocessing, power curve and fault labelling in SCADA data used in this study. Table 1 outlines an example of the structure of our data.

[2] Special Acknowledgment: Platform for Operational Data (POD) Disseminated by ORE Catapult: https://pod.ore.catapult.org.uk

[3] Details of preprocessing: https://github.com/joyjitchatterjee/TurbineSCADACausality/

Time Stamp	**Feature 1** (X_1)	**Feature 2** (X_2).....	**Feature N** (X_N)	**Functional group**
dd/mm/yyyy hh:mm:ss	2.104	0.890	8.124	Pitch System Interface Alarms
dd/mm/yyyy hh:mm:ss	1.245	3.753	9.509	Hydraulic System
dd/mm/yyyy hh:mm:ss	0.156	1.234	7.120	No fault

Table 1. Example of SCADA data structure from the LDT used for our study.

4. Experiments

We use the data for time-series prediction of different features corresponding to different functional groups (i.e. predicting values of the features under normal operation/anomaly). We then apply causal inference to gain a better understanding of underlying causes, as described in Section 2. We implement our learning model in PyTorch [19]. For training, we used a learning rate of 0.01, Adam optimisation, kernel size of 2x2, dilation coefficient of 2 and a 80-20% train-test split for all experiments. These parameters were obtained through hyper-parameter optimisation and empirical tuning. As described, we use the trained CNN features to extract attention matrices for unsupervised causal discovery of SCADA features on faults in turbine-sub components. We use the top 20 attention weights in constructing the causal graph. As our goal is to identify hidden confounders from the data (rather than obvious relationships), we base our significance measure on existing literature e.g. [10, 12]. Previous work has found that a significance measure of $s = 0.8$ gives good results that will reasonably bridge the gap between relevance and number of hidden confounders identified.

We evaluate our model based on the average Mean Absolute Scaled Error (MASE) and average standard deviation, and compare it against a state-of-the-art baseline, the deconfounder in [10]. The deconfounder is a statistical algorithm for causal inference based on a probability factor model which uses a controlled study to identify causal effects in different groups in the population. Finally, based on our attention scores and kernel weights, the complete temporal causal graph is constructed. We extract the relevant sub-graphs from the complete graph based on functional groups for interpretation of the confounders, see discussion in Section 5.

5. Results

Table 2 shows our results. Our AD-DSTCN model with 1 layer in Depthwise Convolution performed the best with a MASE of 1.066 and standard deviation of 2.948. In comparison, our deconfounder baseline [10] achieved a MASE of 3.901 – 72.67% worse than our proposed model.

Layers in Depthwise Convolution	**Optimiser**	**Epochs**	**MASE**	**Std. Deviation**
1	Adam	500	3.292	4.731
		1000	**1.066**	**2.948**
		2000	1.212	2.985
	RMSprop	1000	2.493	3.829
2	Adam	1000	2.609	2.356

Table 2. Evaluation of model performance for temporal causal inference.

We extracted the subsets of temporal causal graphs from the complete causal graph for various cases of faults in functional groups. This was used to discover hidden relationships in the SCADA corpus during faults in different sub-components of the turbine. Further, our model identifies the time delay in causation as a function of time-steps in the SCADA time-series (wherein 1 time-step is equivalent to 10 minutes of SCADA measurements). Note that temporal causal graphs, unlike any other form of data representation in coordinate systems (e.g. time-series/images/audio signals) do not exist in Euclidean space. Also, they do not

Functional Group	Hidden Conf. discovered	Avg. % Relevance
No Fault	✓	**86.95**%
Partial Performance-Degraded	✓	45.45%
Pitch System Interface Alarms	✓	**63.15**%
Gearbox	✓	33%
Pitch System EFC Monitoring	✓	**60%**
PCS	✓	54.54%
MVTR	×	N/A
Yaw Brake	✓	**60%**
Hydraulic System	✓	58%
Yaw	✓	44.44%
Wind Condition Alarms	×	N/A
Pitch	✓	**66.66%**
IPR	×	N/A
Test	✓	28.57%

Table 3. Qualitative evaluation of confounders. Cases $\geq 60\%$ relevance shown in bold face.

have a fixed structure, which makes both, the construction and quantitative analysis of the graph a difficult task. However, the relationships between visualised SCADA features can be qualitatively analysed by turbine engineers and technicians to facilitate explainability of decision-making in the learning models. Our deep learning approach for causality identification can provide completely new knowledge based on the identified relationships, and shed light on integral relationships that may otherwise be discarded as noise.

As it was not possible to verify the identified causal relationships quantitatively (due to the lack of ground truth on hidden confounders), we performed a qualitative evaluation of the temporal causal graphs generated for 14 different categories of faults (Functional Groups) for our study. Note that some relationships are obvious (e.g. time-series of active power mean value affecting active power minimum value, but may not make sense in the context of a given fault at all). For our qualitative evaluation, we only consider those relationships to be relevant where the identified confounders are in-line with the context of the anomaly. Additionally, in some cases, our causal learning model was not able to identify any hidden confounders in line with the fault context. Percentage relevance is evaluated on a random, but representative portion of our test set for our study and represents the human evaluation on how likely a relation is to occur based on domain understanding, though we understand that this might differ from how the machine interprets during the prediction process. The percentage relevance is therefore only a reflection of human understanding. Table 3 summarises the percentage relevance of identified causal relationships (if any) in different Functional Groups. Note that the temporal graphs used are based on 80% significance measure, as these were found to be the most reasonable in terms of sufficient number of hidden confounders identified as well as relevance to the fault context.

As can be seen from Table 3, our model identified the most relevant causal relationships during normal turbine operation (no fault), followed by pitch system, pitch interface alarms, pitch system EFC monitoring and yaw brake, all with higher than 60% relevance. The model was not able to identify hidden confounders for faults in the Wind Condition Alarms and Moisture Vapour Transmission Rate (MVTR), possibly due to a lack of multiple latent variables (other than the obvious wind speed in the former, and vapour transmission parameters in the latter). The model performed poorly for some cases like Test Rig and Gearbox (with relevance as low as 28.57%) during causality inference, likely due to the complex nature of the faults (and errors in data produced by Test Rigs). Below, we discuss some interesting cases of hidden confounders

observed by our model during causal inference. The interested reader is referred to Github[4] for temporal causal graphs for multiple other cases under different categories of anomalies.[5] [6]

Case 1: Hidden confounders identified during normal operation of turbine with significance measure 0.80 Figure 3 shows the causal relationships for turbine normal operation. Most of the identified relationships are as expected, such as rotor speed being causally related to pitch angle, nacelle angle to rotor speed and wind direction, active power to grid voltage etc. all with a delay of 0 time steps (instantaneous causation). These relations are reasonable as during power limitation and optimisation control of the turbine, such parameters essentially ensure optimal performance [20], and turbines use blade pitch to vary rotational speed during power generation. Many relationships from the same time series share causality (e.g. reactive power mean to standard deviation relation, wind direction min. to standard deviation etc.), mainly due to similar patterns (as 10 min. ave. features from the same context would share causal confounders).

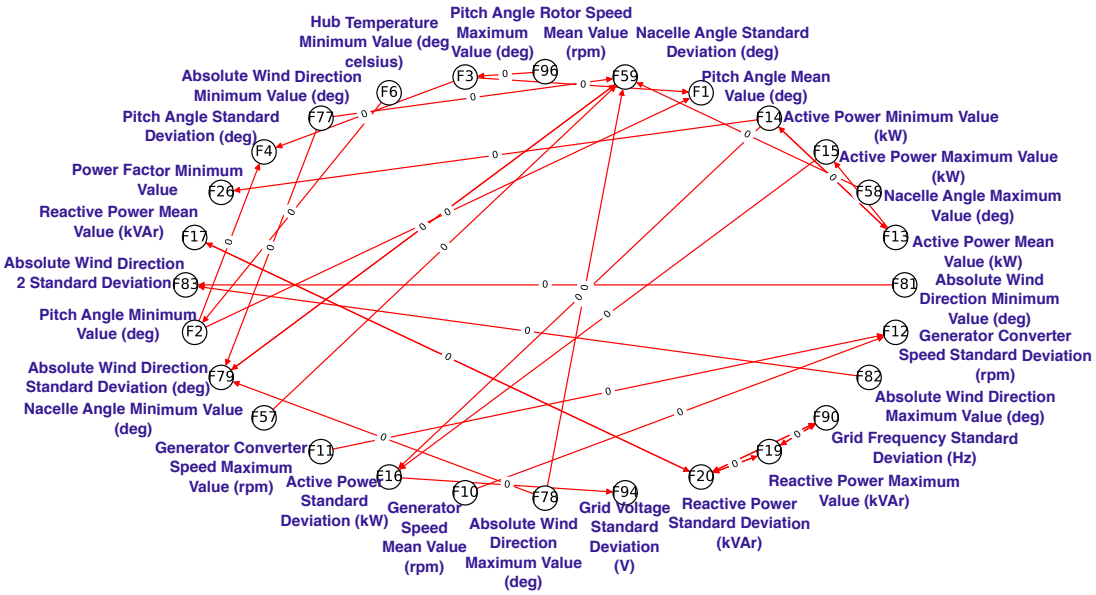


Figure 3. Hidden confounders for normal operation of turbine, wherein increase in loss between original time-series and intervened AD-DSTCN output is 80% significant.

Case 2: Hidden confounders identified during anomaly in yaw system with significance measure 0.80 Figure 4 shows hidden confounders for an anomaly in the yaw

[4] Various temporal causal graphs for our study: https://github.com/joyjitchatterjee/TurbineSCADACausality/

[5] We focus our discussion on the relevant hidden confounders which can be interpreted by a human, instead of all confounders which are used by the AI model, the qualitative evaluation of which is difficult to establish. The irrelevant parameters can be investigated further by engineers & technicians to see why the model uses the causally-related features in some contexts.

[6] Note: In our case-discussions for the temporal causal graphs, the relationships shown should be interpreted as follows: change in Y (outcome) is caused by any intervention (change) in X, where the identified causal relationship is $Y- > X$. We opted for this notation for the sake of simplicity and suitability to our SCADA features. Also, the identified delay $d(y, x)$ denotes that the outcome feature experiences the change d time steps after the intervening feature experienced it. E.g. a relation from rotor speed (Y) to pitch angle (X) $Y- > X$ at a delay of 1 time-steps would signify that the rotor speed is causally affected/experiences a change due to variation in pitch angle, and this effect occurs 1 time-step (10 minutes for our SCADA data) after variation of pitch angle.

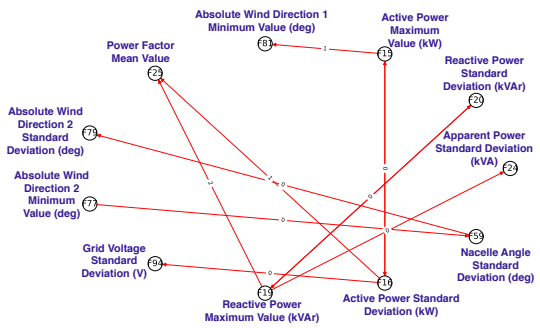


Figure 4. Hidden confounders for yaw system anomaly, wherein increase in loss between original time-series and intervened AD-DSTCN output is 80% significant.

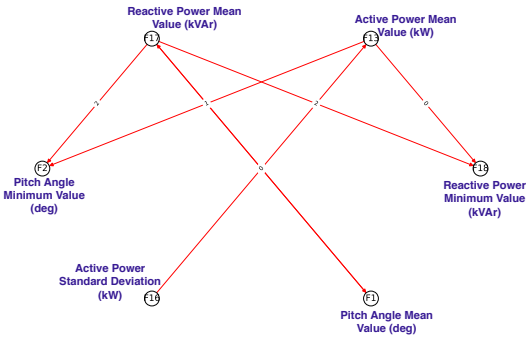


Figure 5. Hidden confounders for pitch system anomaly, where increase in loss between original time-series and AD-DSTCN output is 80% significant.

system. The absolute wind direction is causally related to the nacelle angle during this anomaly. This is reasonable as wind is a highly variable parameter, and depending on the present wind direction measured by the nacelle measuring instrument, the turbine's control system directs appropriate yawing. With any change in wind direction, frequent start/stop of the yaw system to keep the turbine aligned with the wind leads to rotor torque fluctuations and resistance torque variations, causing load fluctuations in the yaw system [21]. These can directly lead to speed fluctuations in the system, making it unstable, and cause vibrations in the nacelle. As the nacelle position signifies the angle between the turbine rotor axis and true north [22], this instantaneous causal relation *(0 time steps delay)* signifies possible high aerodynamic yaw loads, and slippage of the nacelle. The yaw anomaly further affects the turbine's power generation efficiency [23] causing production losses shortly after the anomaly occurs *(1-2 time steps delay)*.

Case 3: Hidden confounders identified during anomaly in pitch system with significance measure 0.80 Figure 5 shows the temporal causal graph for an anomaly in the pitch system. The turbine pitch angle value is causally related to the reactive power. This is reasonable given that any deviation in pitch angle from a predefined optimum value (under given wind speed) affects the power dynamics (such as active power and reactive power) of the turbine [24]. Additionally, we observe that these causal relationships arose during an actual alarm (Pitch System Interface Alarm), which can signify that the pitch angle's dynamic response faced a significant anomaly, resulting in potential failure of the control system.

Case 4: Hidden confounders identified during anomaly in yaw with significance measure of 0.95 Shown in Figure 6, while our model identifies the highest number of hidden confounders (with an increased signficance measure), this comes at the cost of explainability/credibility of certain features. Reasonable temporal features identified are: reactive and apparent power are causally related to power factor[7], active power is causally related to rotor speed etc.[8], and active power changes with generator converter speed. The graph shares other relations from Case 2. However, many relationships are either non-credible or difficult to explain, such as: generator stator temperature being causally related to nacelle angle and wind direction, outdoor temperature to active power, or pitch angle to reactive power. These seem

[7] Power factor is by definition the ratio of cosine angle from apparent to reactive power.
[8] Any change in active power is caused by variation of rotor speed, which is reasonable as the yaw error affects running characteristics such as rotor speed and active power [21].

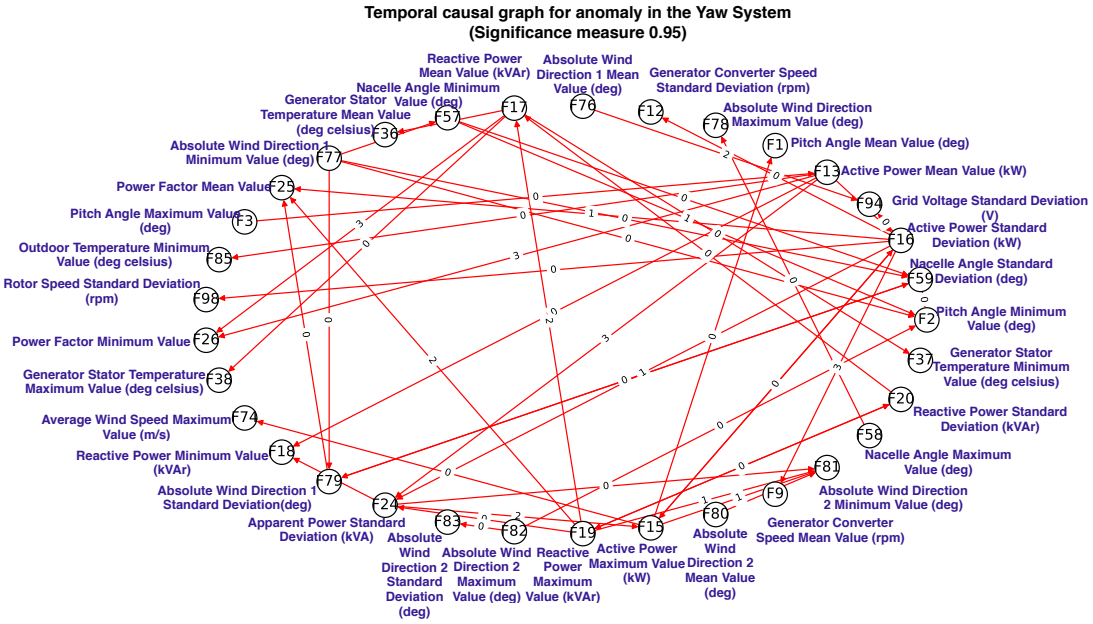


Figure 6. Hidden confounders for yaw system anomaly, wherein increase in loss between original time-series and intervened AD-DSTCN output is 95% significant.

to signify that a change in nacelle angle causes a change in generator stator temperature, which seems meaningless if seen stand-alone. However, we can see from the graph that nacelle angle is also causally related to the pitch angle (whereby generator stator temperature also indirectly shares cause-effect relationships to pitch angle), which can describe a situation where the yaw error affects the pitch control system, and causes the generator stator temperature to share this relationship.[9] We can see these relationships as situations where a co-located anomaly appeared in another turbine sub-component at the same time (either as an effect of the present anomaly, or independently), which led to variation in the statistical properties of the other features (either directly or as a consequence of causation between multiple confounders). We do not claim that all identified relationships are credible to a human expert and some can be hard to explain [26], but we believe that they can help probe a learning models' predictions and enhance transparency.

6. Discussion and Conclusion

We have proposed a novel application of deep learning for temporal causal inference using wind turbine SCADA data. In particular, we have demonstrated that using the operational status of a turbine as the target for causal inference, we can discover numerous hidden predictor variables that are not picked up by existing models in the literature, as far as we are aware. Earlier related work by Felgueira et al. [11] has applied Autoregressive Causal Normal Behaviour Models to temporal causal inference, however their model neglects crucial temporal cause-effect relationships which can be identified via temporal causal graphs. We have also found that

[9] According to [25], the pitch angle and generator stator temperature signals do indeed have correlation during dynamic monitoring of faults in pitch system and are related to a certain degree.

although a lower significance measure identifies fewer hidden confounders, more of these are relevant to the domain. Our temporal causal inference model is not perfect and we do not claim that all identified relationships are directly relevant for the wind turbine operators. Nonetheless, we show up a clear path towards explainable AI for the wind industry, as the internal parameters used by current black-box deep learning models are transparently revealed with our approach. By utilising larger datasets, and tuning the models with human-specified optimisations, we envisage that the explainablity of the AD-DSTCN model be can further improved in the future. By helping in explainable AI for wind turbines, and inspiring more wind farm operators to employ data-driven decision making for significant reductions in O&M costs, we hope to contribute to a fast-paced move to renewable energy sources. In future work, we plan to extend our model with a complete knowledge graph database for wind turbines, which can help in generating effective policies and maintenance strategies for O&M through natural language generation for further accessibility of turbine logs.

Acknowledgments
We would like to acknowledge the Offshore Renewable Energy (ORE) Catapult for providing us access to operational data from the turbine through Platform for Operational Data (POD). We are also grateful to Aura Innovation Centre and the University of Hull for their support.

References

[1] Christopher J Crabtree, Donatella Zappala, and Simon I Hogg. Wind energy: UK experiences and offshore operational challenges. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, 229:727–746, August 2015.

[2] Rafael Orozco, Shuangwen Sheng, and Caleb Phillips. Diagnostic models for wind turbine gearbox components using scada time series data. In *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pages 1–9, Seattle, WA, USA, June 2018.

[3] Nicolas Tchertchian and Dominique Millet. Eco-maintenance for complex systems: Application on system of renewable energy production. In *3rd International Symposium on Environmental Friendly Energies and Applications (EFEA)*, pages 1–6, St. Ouen, France, November 2014.

[4] Joyjit Chatterjee and Nina Dethlefs. A dual transformer model for intelligent decision support for maintenance of wind turbines (to appear). In *International Joint Conference on Neural Networks (IJCNN)*, Glasgow (UK), July 2020.

[5] Yingying Zhao, Dongsheng Li, Ao Dong, Dahai Kang, Qin Lv, and Li Shang. Fault prediction and diagnosis of wind turbine generators using scada data. *Energies*, 10(8):1210, 2017.

[6] Joyjit Chatterjee and Nina Dethlefs. Deep learning with knowledge transfer for explainable anomaly prediction in wind turbines. *Wind Energy*, 23(8):1693–1710, 2020.

[7] Judea Pearl. The seven tools of causal inference, with reflections on machine learning. *Commun. ACM*, 62(3):54–60, February 2019.

[8] Chad Hazlett. Estimating causal effects of new treatments despite self-selection: The case of experimental medical treatments. *Journal of Causal Inference*, 7(1):1–11, Dec 2018.

[9] Xiangzhou Zhang, Yong Hu, Kang Xie, Shouyang Wang, E.W.T. Ngai, and Mei Liu. A causal feature selection algorithm for stock prediction modeling. *Neurocomputing*, 142:48–59, 10 2014.

[10] Yixin Wang and David M. Blei. The Blessings of Multiple Causes. *Journal of the American Statistical Association*, Volume 0:1–71, 2019.

[11] Telmo Felgueira, Silvio Rodrigues, Christian S. Perone, and Rui Castro. The Impact of Feature Causality on Normal Behaviour Models for SCADA-based Wind Turbine Fault Detection. *Climate Change: How Can AI Help? ICML, California, USA*, June 2019.

[12] Meike Nauta, Doina Bucur, and Christin Seifert. Causal Discovery with Attention-Based Convolutional Neural Networks. *Machine Learning and Knowledge Extraction*, 1(1):312–340, Jul 2019.

[13] Michael Lewis and Alexis Kuerbis. An overview of causal directed acyclic graphs for substance abuse researchers. *Journal of Drug and Alcohol Research*, 5:1–8, 01 2016.

[14] Alom, Md. Zahangir et al. A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8:292, 03 2019.

[15] Qianjin Du, Weixi Gu, Lin Zhang, and Shao-Lun Huang. Attention-based lstm-cnns for time-series classification. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, SenSys '18, page 410–411, New York, NY, USA, 2018. Association for Computing Machinery.

[16] Kaiming He and Jian Sun. Convolutional neural networks at constrained time cost. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5353–5360, Boston, MA, USA, 06 2015.

[17] Y. Tang, J. Xu, K. Matsumoto, and C. Ono. Sequence-to-sequence model with attention for time series classification. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 503–510, Los Alamitos, CA, USA, dec 2016. IEEE Computer Society.

[18] Yuxiao Huang and Samantha Kleinberg. Fast and accurate causal inference from time series data. In *Florida Artificial Intelligence Research Society Conference*, pages 49–54, Florida, USA, 2015.

[19] Adam Paszkeet et al. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alche Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[20] Ramakrishna Balijepalli, V.P. Chandramohan, and K. Kirankumar. Optimized design and performance parameters for wind turbine blades of a solar updraft tower (sut) plant using theories of schmitz and aerodynamics forces. *Sustainable Energy Technologies and Assessments*, 30:192 – 200, 2018.

[21] Shuting Wan, Lifeng Cheng, and Xiaoling Sheng. Effects of yaw error on wind turbine running characteristics based on the equivalent wind speed model. *Energies*, 8(7):6286–6301, 2015.

[22] N. Mittelmeier and M. Kühn. Determination of optimal wind turbine alignment into the wind and detection of alignment changes with scada data. *Wind Energy Science*, 3(1):395–408, 2018.

[23] S. Reddy, Kishore Prathipati, and Young Lho. Transient stability improvement of a system connected with wind energy generators. *International Journal of Emerging Electric Power Systems*, 18, 01 2017.

[24] Jamie Godwin and Peter Matthews. Classification and detection of wind turbine pitch faults through scada data analysis. *International Journal of Prognostics and Health Management*, 4, 01 2013.

[25] Meng Li and Shuangxin Wang. Dynamic fault monitoring of pitch system in wind turbines using selective ensemble small-world neural networks. *Energies*, 12:3256, 08 2019.

[26] Nadya Vasilyeva, Thomas Blanchard, and Tania Lombrozo. Stable causal relationships are better causal relationships. *Cognitive Science*, 42(4):1265–1296, 2018.