Data Engineering and Applications

Research Paper

Wind Turbine Digital Twin Virtual Sensing Construction Based on a Koopman Autoencoder Model

Jassima Al-Buflasa[1,*], Essaida Al-Harmoodi[1] and Karima Al-Riyami[2]

[1] School of Information Technology, American University of Sharjah, Sharjah, PO Box 26666, United Arab Emirates
[2] College of Engineering, Information Technology Division, University of Sharjah, Sharjah, PO Box 27272, United Arab Emirates
*Corresponding author: jassima.buf@aus.edu

Abstract. Physical sensors are unable to meet the demand for dense measurement channels in wind turbine condition monitoring because of their high cost, challenging installation, harsh nacelle environment, and long-term drift. A Koopman Autoencoder approach is suggested for building virtual sensors in a digital twin of a wind turbine. The technique creates a latent space for linear Koopman evolution of nonlinear turbine behavior by mapping multi-source supervisory control, vibration, temperature, power, and pitch-yaw data. The digital twin provides residual feedback and operating-context constraints, while the decoder reconstructs the unmeasured or missing variables. Examine the framework in the following scenarios: normal operation, wind speed variations, sensor drift, missing sensors, and cross-turbine transmission. As a result, the virtual sensor's accuracy has improved and its reconstruction error for missing channels has decreased, making it appropriate for edge-assisted turbine monitoring.

Keywords: *Wind Turbine Digital Twin, Virtual Sensing, Koopman Autoencoder, Sensor Reconstruction*

Received on 25 December 2022, Accepted on 29 March 2023, Published on 05 April 2023

Introduction

In addition to having variable pitch control, drivetrain load fluctuations, and environmental disturbances, modern wind turbines operate in non-stationary wind fields. Many of the important states cannot be directly observed at high spatial or temporal resolution, despite the fact that supervisory control and data acquisition systems offer some measurements. A turbine model may incorporate its physical structure and operational data over time to facilitate online monitoring, according to research on digital twins [1]. By treating each channel as a separate numerical sequence, a twin can be used to transform sparse measurements into structured state estimates, making virtual sensing relatively simple to implement [2].

The dynamics of wind turbines often include nonlinear and regime-dependent difficulties. At start-up, rated-power, partial-load, curtailment, and shutdown, the relationships between rotor speed, generator torque, nacelle temperature, pitch angle, blade load proxy, vibration amplitude, and power output vary. Traditional regression methods can recover missing channels in the typical operational range, but they are not particularly stable under various input circumstances. The online model must be compact enough for near-equipment operation and store long-term state information, according to research on edge intelligence and industrial monitoring [3].

This research suggests a Koopman Autoencoder-based virtual sensing architecture for wind turbine digital twins in light of the issues. The decoder reconstructs physical variables as virtual measurements, the encoder learns a latent representation of aligned monitoring sequences, and the Koopman transition matrix models the evolution of latent states. It features drift correction, early abnormal condition detection, and a missing sensor compensation function. To enhance virtual sensing's precision, deployability, and interpretability in the context of turbine operation, it incorporates a latent linear dynamics model and twin-state restrictions [4].

Related Work

Wind Turbine Monitoring and Sensor Redundancy

Data-driven diagnosis and prediction have replaced threshold-based alerts in wind turbine monitoring. The actual turbine has been linked to a simulation model, past maintenance logs, and online monitoring data using digital twin systems [5]. In this instance, the unknown quantities—such as blade-root load, local temperature increase, drivetrain condition, and hidden damage indicators—are approximated using virtual sensors. Sensor redundancy must be used when extreme operating conditions result in missing data or communication delays, according to the predictive maintenance work mentioned above [6].

Multivariate time-series reconstruction is also connected to virtual sensing. If a channel is absent, use the model's correlated variables to infer it rather than just interpolating earlier values. The following traits are also present in wind turbines: Nacelle vibration is influenced by drivetrain conditions, generator temperature varies with load and ambient temperature, and power is correlated with wind speed and rotor speed. Operating-regime knowledge is linked to reconstruction quality, according to recent research on wind turbine monitoring [7] and SCADA-based diagnostic investigations [8].

Latent Dynamic Models for Virtual Sensing

In industrial reconstruction, autoencoders are frequently used to reduce the dimensionality of high-dimensional measurements to a small number before reconstructing the original variables. In anomaly identification and fault diagnosis, reconstruction error may be utilized to show that the typical operating pattern has not been accurately replicated [9]. A standard autoencoder is constrained since it does not explicitly describe state transitions. Although it can precisely reconstruct every window, it lacks the dynamic continuity needed for online virtual sensing and forecasting.

An excellent alternative that maps nonlinear dynamics to a latent space with almost linear development is Koopman learning. This feature separates temporal propagation from representation learning, making it appropriate for wind turbine virtual sensing. Latent evolution may preserve the characteristics of nonlinear systems and be more suited for prediction, as shown by deep Koopman models [10], linearly recurrent autoencoders [11], and universal linear embedding investigations [12]. Compact long-context models are necessary for monitoring windows with delayed thermal or mechanical responses, according to effective sequence studies [13].

Koopman Autoencoder-Based Virtual Sensor Construction

Monitoring Sequence Encoding and Twin-State Mapping

First, match the present state of the digital twin in the suggested framework with the multi-source sequences of wind turbine monitoring. Wind speed, rotor speed, generator torque, active power, pitch angle, yaw deviation, nacelle temperature, gearbox temperature, generator temperature, vibration amplitude, and available maintenance context are all included in each sample window. Component identification, operating circumstances, rated-power limitations, and expected-state windows are all provided by the digital twin. The field measurements, twin-state background, and historical data are arranged prior to being included into the virtual sensor model, as seen in Figure 1.

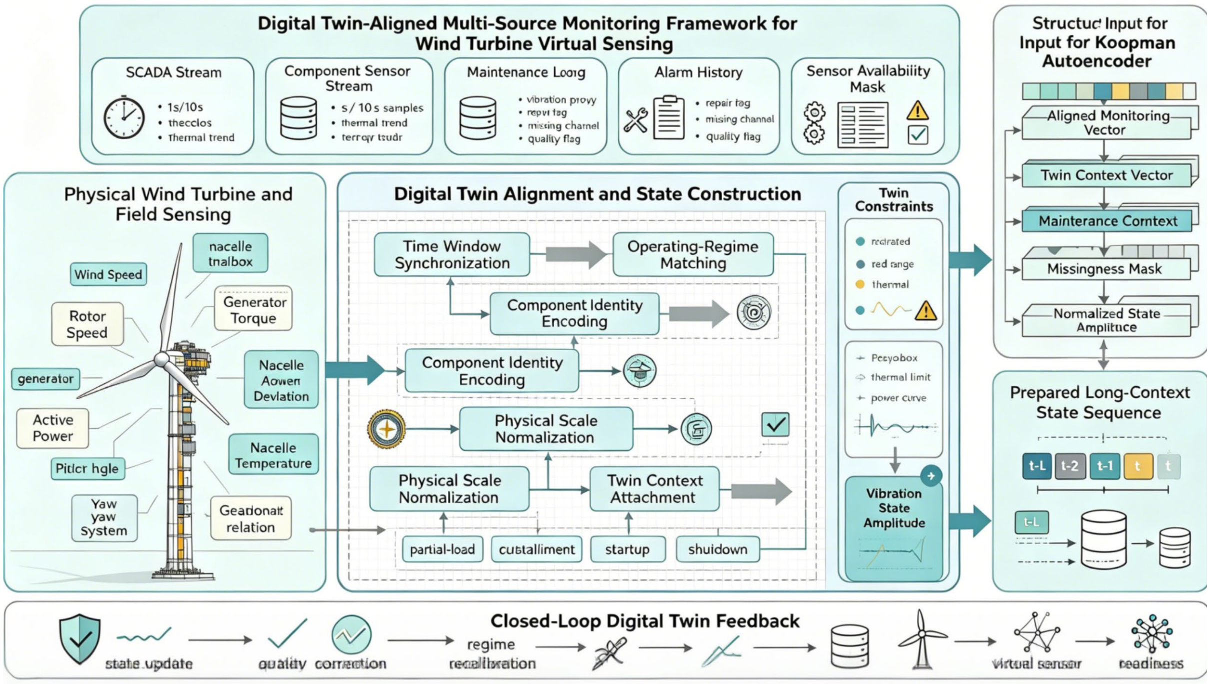

Figure 1. Multi-source turbine monitoring and twin alignment process.

At time t, the raw monitoring vector is expressed as follows:

$$x_t = [w_t, r_t, p_t, q_t, \theta_t, \psi_t, T_t, v_t] \qquad \text{Eq.(1)}$$

In this vector, w_t denotes wind-related input, r_t denotes rotor-state information, p_t and q_t describe power and torque response, while θ_t, ψ_t, T_t, and v_t represent pitch-yaw, thermal, and vibration-related measurements. This vector is not treated as a simple feature list. It is regarded as a turbine-state observation whose meaning depends on the current operating regime.

To connect measurement data with the digital twin, the aligned state is defined as:

$$s_t = [x_t, c_t, m_t] \qquad \text{Eq.(2)}$$

Here, c_t represents twin-context information such as operating mode, component identity, rated-power status, and environmental condition. The term m_t represents maintenance or data-quality information, including recent inspection labels, sensor availability, and channel reliability. By combining these elements, the model can distinguish a normal thermal increase during rated operation from an abnormal thermal rise under low-load operation.

Because turbine variables have different physical scales, the aligned state is normalized by operating regime:

$$\bar{s}_t = \frac{s_t - \mu_g(t)}{\sigma_g(t) + \epsilon} \qquad \text{Eq.(3)}$$

The mean $\mu_g(t)$ and standard deviation $\sigma_g(t)$ are estimated within the current operating group. This prevents signals from different regimes from being forced into one global distribution. For example, generator temperature during high-power production and generator temperature during shutdown cooling should not be interpreted under the same statistical baseline.

The observation mask is introduced as:

$$M_t \in \{0,1\}^d \qquad \text{Eq.(4)}$$

Each channel's availability, absence, delay, or purposeful concealment during training are all recorded by the mask. Convenient virtual sensing is essential, but the missing data should also point to a connectivity issue or sensor failure. Save the partial samples and train a model to deduce the missing characteristics from the relevant channels instead of discarding them.

Then, the input monitoring window looks like this:

$$X_{t-L+1:t} = [\bar{s}_{t-L+1}, \bar{s}_{t-L+2}, \dots, \bar{s}_t] \quad \text{Eq.(5)}$$

The length L determines how much historical context is retained. A longer window allows the model to observe delayed thermal accumulation, pitch-control response, and vibration propagation. This window-level representation is more suitable than single-frame input because many wind turbine states evolve gradually rather than instantaneously.

Koopman Latent Transition Learning

After aligning the sequences, a latent space is mapped to the normalized turbine states. Identify a small-scale dynamic coordinate system that makes it simpler to explain the wind turbine's nonlinear motion. Information about energy conversion, thermal response, drivetrain vibration, and control action must be stored in the latent state.

This is the encoder map:

$$z_t = E_\phi(\bar{s}_t) \quad \text{Eq.(6)}$$

The latent variable z_t is a compressed representation of the turbine state. Although its dimension is smaller than the original input, it should preserve the dominant dynamic evidence required for reconstruction. If the latent state loses load or thermal information, the virtual sensor will become accurate only under easy operating conditions and unstable under regime transition.

The Koopman transition is then introduced in the latent space:

$$\hat{z}_{t+1} = K z_t \quad \text{Eq.(7)}$$

The matrix K describes one-step latent evolution. The important point is that the original turbine dynamics may be nonlinear, but the learned latent dynamics are encouraged to evolve approximately linearly. This makes the model easier to propagate, easier to stabilize, and more suitable for edge-side deployment than a heavy recurrent model.

For multi-step latent prediction, the transition is extended as:

$$\hat{z}_{t+h} = K^h z_t \quad \text{Eq.(8)}$$

Through this expression, the model repeatedly applies the same transition operator to predict future latent states. It is appropriate for maintaining a virtual sensor's stability for a brief period of time. The reconstructed channel should not be regarded as a trustworthy turbinestate estimate if it is accurate at the moment but inconsistent in the subsequent steps.

This is how the dynamic consistency loss is expressed:

$$L_{dyn} = \sum_{h=1}^{H} \|z_{t+h} - K^h z_t\|_2^2 \quad \text{Eq.(9)}$$

As a result, the learnt latent representation will follow a regular time series due to the loss function. It does not stop the encoder from generating a latent vector that is capable of accurately reconstructing the present but is unable to characterize behavior in the future. It is necessary for the monitoring system to react to temperature, vibration, and power-response coupling changes for wind turbines somewhat quickly.

The reconstruction loss can be found using:

$$L_{rec} = \left\|\bar{s}_t - D_\psi(z_t)\right\|_1 \quad \text{Eq.(10)}$$

The decoder D_ψ reconstructs the normalized turbine state from the latent representation. This loss keeps the latent space connected to observable physical variables. Without this reconstruction constraint, the Koopman transition may become mathematically smooth but physically meaningless.

To avoid unstable latent growth, a relaxed spectral constraint is used:

$$\rho(K) \leq 1 + \delta \quad \text{Eq.(11)}$$

The term $\rho(K)$ denotes the spectral radius of the Koopman transition matrix. The small tolerance δ allows transient response but discourages uncontrolled expansion. This constraint improves stability when the model is applied to long monitoring windows or unfamiliar operating regimes.

Virtual Sensor Decoding and Residual Calibration

The target variables that are absent, erratic, or not immediately visible are reconstructed by the decoder. These goals might include gearbox vibration proxy, bearing temperature, blade-load proxy, or other digital twin measurement channels. The online virtual sensing architecture is composed of latent transition, decoding, confidence gating, and twin residual feedback, as seen in Figure 2.

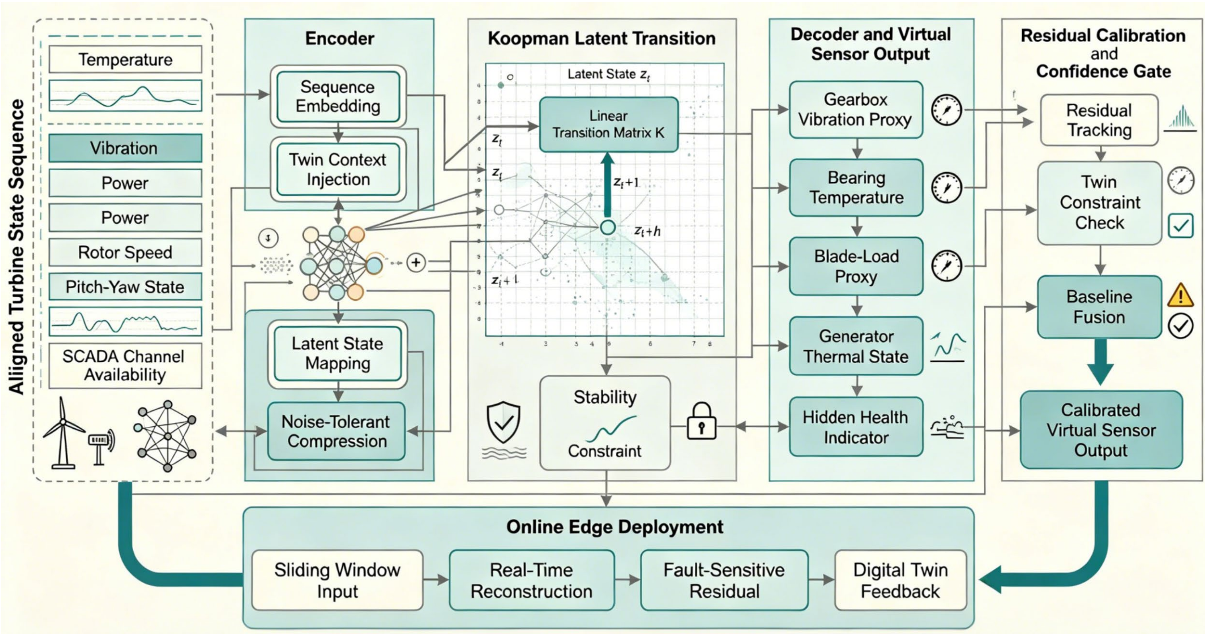

Figure 2. Online virtual sensor construction architecture.

The virtual sensor output is decoded as:

$$\hat{y}_t = D_\psi(z_t, c_t) \quad \text{Eq.(12)}$$

The predicted value $\hat{y}_t$ depends on both the latent state and the twin context. This means that the same latent pattern can lead to different interpretations under different operating regimes. For example, a vibration increase may be acceptable during transient wind fluctuation but suspicious during stable low-load operation.

When the true target channel is available during training or validation, the residual is computed as:

$$r_t = y_t - \hat{y}_t \quad \text{Eq.(13)}$$

The residual is the difference between the virtual sensor's actual value and its predicted value. It is used to guarantee stability while also optimizing. Sensor drift, unmodeled deterioration, or a mismatch between the present turbine and the learnt twin state might be the cause of persistent residual increase.

This is how the Confidence Gate is configured:

$$\alpha_t = \sigma(W_r|r_t| + W_c c_t + b) \quad \text{Eq.(14)}$$

The gate decides how much weight to apply to the virtual sensor's output. The latent reconstruction, residual behavior, and twin context are all in accord when the confidence level is high. A low confidence score suggests that the model should exercise caution since the present condition could be uncommon, unstable, or physically irrational.

The following are the outputs of the calibrated virtual sensor:

$$\tilde{y}_t = \alpha_t \hat{y}_t + (1 - \alpha_t) y_{base,t} \quad \text{Eq.(15)}$$

The baseline estimates $y_{\text{base},t}$ may come from a simple physical rule, historical average, or conservative interpolation. This weighted output prevents the model from producing overconfident predictions under unfamiliar conditions. It also makes the virtual sensor safer for practical monitoring because uncertain outputs are softened rather than blindly accepted.

To improve missing-channel robustness, the masking loss is written as:

$$L_{\text{mask}} = \|M_t \odot (s_t - \hat{s}_t)\|_1 \quad \text{Eq.(16)}$$

Based on the data that are available, use the trained model to recreate the channels that are absent or unavailable. To help the model understand cross-channel dependencies, certain channels are purposefully hidden during training. As a result, issues like sensor disconnection, packet loss, and delayed upload may be replicated in a real-world wind farm setting.

The following is the twin-guided feasibility constraint:

$$L_{twin} = \sum_j \max(0, \hat{y}_{t,j} - u_{j,c}) + \max(0, l_{j,c} - \hat{y}_{t,j}) \quad \text{Eq.(17)}$$

The bounds $u_{j,c}$ and $l_{j,c}$ are determined by operating context. They do not force all reconstructed values to be normal; instead, they penalize physically unreasonable outputs. This allows the model to represent abnormal behavior while still avoiding impossible virtual measurements.

The final training objective is:

$$L = L_{rec} + \lambda_1 L_{dyn} + \lambda_2 L_{mask} + \lambda_3 L_{twin} \quad \text{Eq.(18)}$$

Reconstruction error, latent dynamic consistency, missing-channel robustness, and digital twin feasibility are all included in the overall loss function. The four practical criteria for virtual sensing are as follows: the output must match the actual measurement, change steadily over time, be useful even in the absence of data, and continue to have significance in the context of turbine operation.

Experimental Results and Analysis

Virtual Sensor Reconstruction Performance

The virtual sensor reconstruction findings during turbine operation are shown in Figure 3. Create monitoring windows using simulation in the experimental dataset for wind turbines operating under normal, partial-load, rated-power, curtailment, and transient wind circumstances. Wind power curve modeling [15] further facilitates condition-specific performance evaluation, and previous deep manufacturing research [14] have suggested that equipment-state estimation should connect observed signals with model context. The RMSE for thermal virtual sensors is shown in Figure 3(a), while the MAE of vibration proxy reconstruction is shown in Figure 3(b). Since a digital twin of a wind turbine is the focus of this paper, each monitored signal is examined based on its component location, operating conditions, and maintenance needs rather than by isolated numerical correlations. If not, the virtual sensor will learn a handy but brittle mapping that only functions in the most common partial-load windows and fails when the relationship between channels shifts as a result of thermal lag, curtailment, rated-power transition, or yaw correction. Thus, Koopman Autoencoder is a kind of dynamic representation model that predicts the development of latent variables, reconstructs hidden or missing measurements, and confirms if the reconstructed values are compatible with the twin state. A virtual sensor is necessary for a design that can be used to wind farms in order to carry out daily monitoring, identify anomalies, fix problems with data quality, and facilitate maintenance analysis in the future.

The baselines of LSTM, GRU, and conventional autoencoder are 2.47, 2.39, and 2.91 °C, respectively, whereas the generator-temperature virtual sensing RMSE of the suggested model is 1.84 °C. Dynamic continuity gives it an edge over the Koopman latent state for rated-power transitions. Blade-load proxy accuracy is assessed in Figure 3(c), and normalized reconstruction error across operating regimes is shown in Figure 3(d). The two findings demonstrate that thermal, mechanical, and load-related virtual quantities all improve. Since a wind turbine digital twin is the focus of this study, all monitored signals will be examined based on component locations, operating circumstances, and maintenance needs rather than specific numerical correlations. If not, the virtual sensor will pick up a handy but brittle mapping that only works in the most common partial-load windows and fails when rated-power transition, curtailment, yaw correction, or thermal lag change the connection between channels. Therefore, Koopman Autoencoders are considered dynamic representation models that predict the evolution of latent variables, reconstruct missing or unobserved data, and assess whether the reconstructed values are consistent with the twin state. For a wind farm to be helpful in daily operations, anomaly detection, data-quality remediation, and future maintenance analysis, such a virtual sensor must be designed.

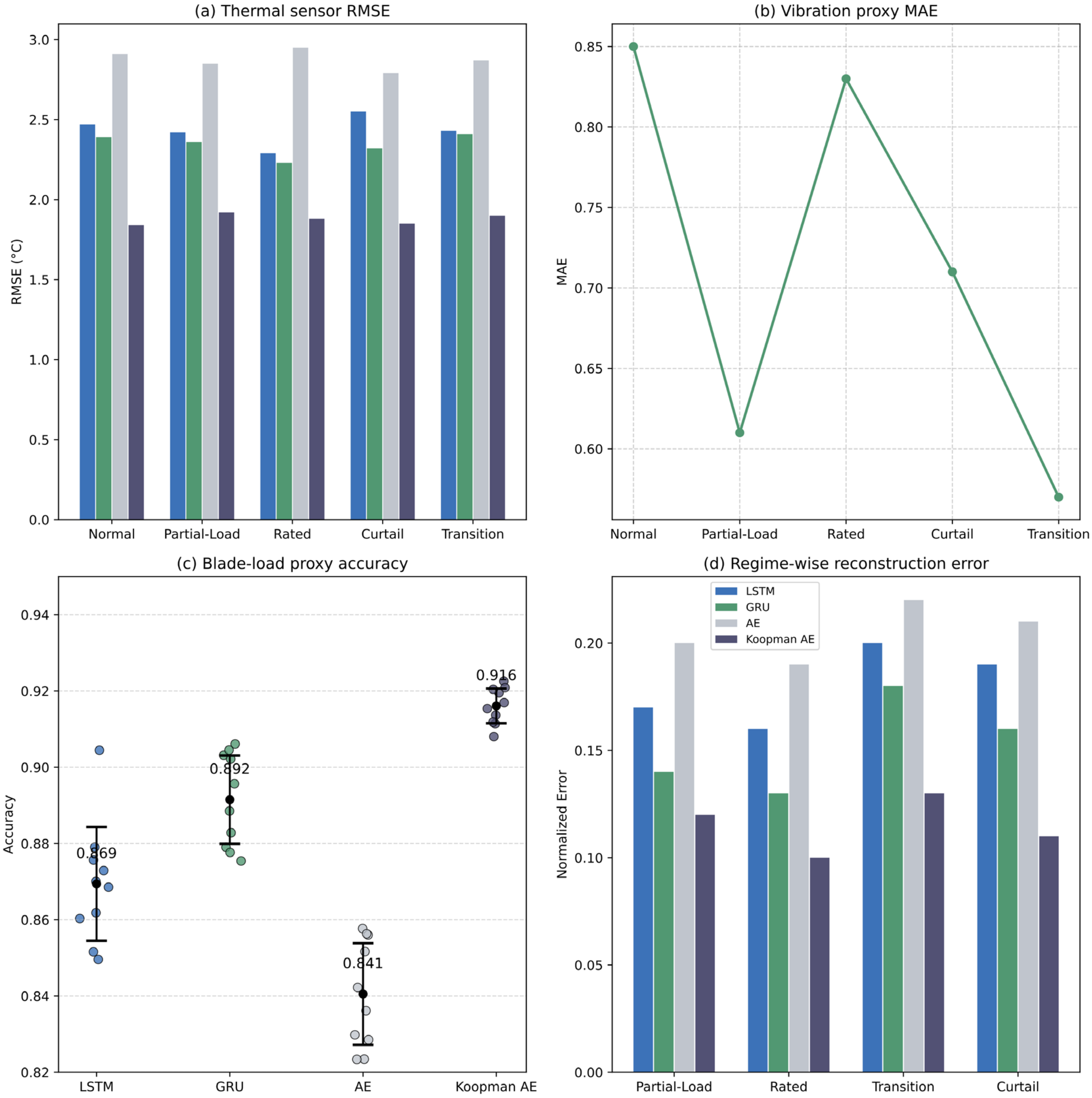

Figure 3. Virtual sensor reconstruction performance. (a) Thermal sensor RMSE. (b) Vibration proxy MAE. (c) Blade-load proxy accuracy. (d) Regime-wise reconstruction error.

The state-specific virtual sensing behavior under typical turbine circumstances is shown in Figure 4. The time-series reconstruction curve under steady partial-load operation is shown in Figure 4(a). While all models are similar, the residuals of the suggested approach are smoother. Rated-power transition performance is shown in Figure 4(b), where delayed thermal accumulation is underestimated by the standard autoencoder. The idea that the delayed emergence of deterioration should be modeled as a dynamic process rather of a static feature snapshot is supported by research on drivetrain reliability analysis [17] and predictive maintenance [16]. The

https://doi.org/10.64972/dea.2023.v2i2.3531d:1-14

topic of discussion is always a digital twin of a wind turbine, and rather than being understood as separate numerical correlations, all monitored signals are interpreted based on operational circumstances, component locations, and maintenance relevance. If not, the virtual sensor will pick up a handy but brittle mapping that only functions in the most common partial-load window and breaks when the channel relationship is changed by rated-power transition, curtailment, yaw correction, or thermal lag. In order to rebuild missing data, forecast the development of latent variables, and confirm if the reconstructed values match those of the twin state, Koopman Autoencoders are used as dynamic representation models. For virtual sensors to be helpful in routine monitoring, abnormal condition screening, data quality correction, and subsequent maintenance analysis, wind farms must have a solid design.

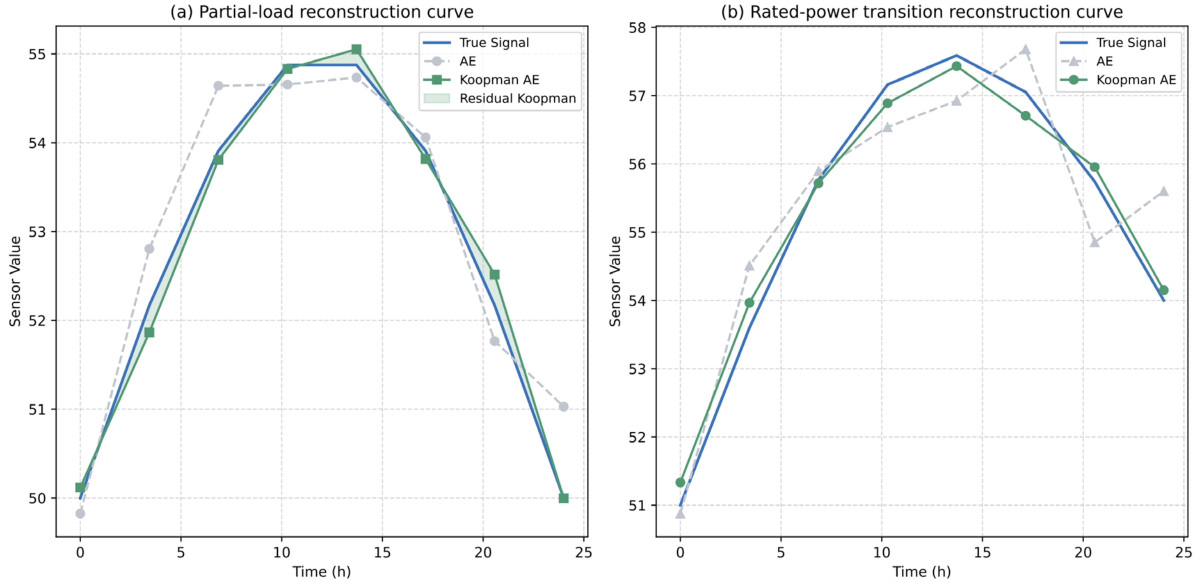

Figure 4. State-specific virtual sensing behavior. (a) Partial-load reconstruction curve. (b) Rated-power transition reconstruction curve.

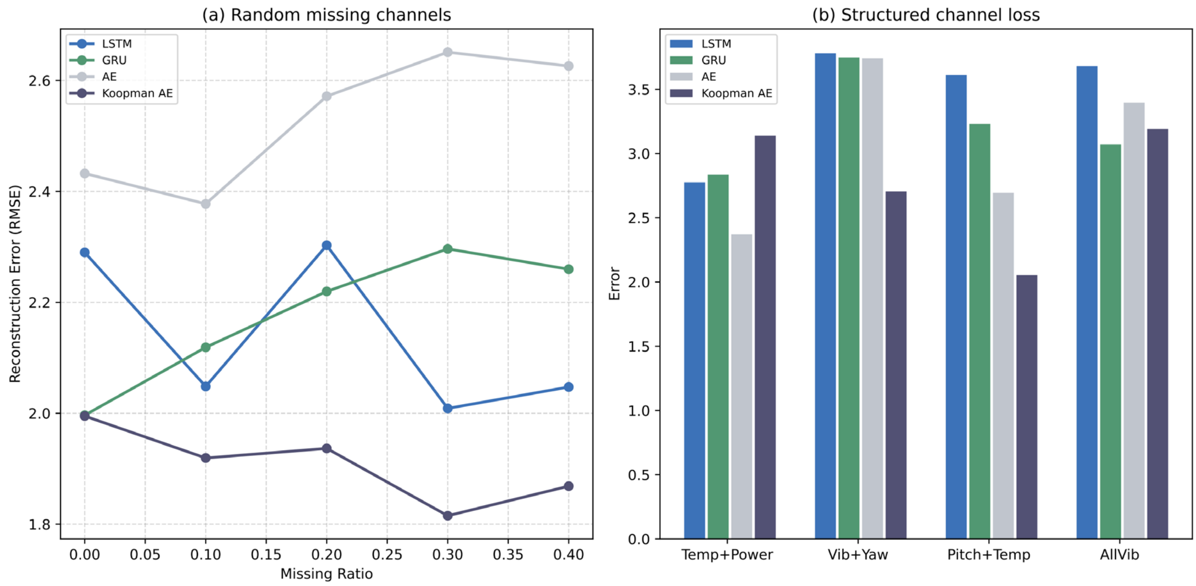

Figure 5. Robustness under missing sensor conditions. (a) Random missing channels. (b) Structured channel loss.

Cross-Condition Robustness and Fault Sensitivity

The assessment in the case of a missing sensor is shown in Figure 5. Missing data and drift in sensor data are common issues with wind farm operations. Robustness tests were performed for temperature, vibration, and power channels at different missing ratios based on reconstruction-oriented anomaly detection research [18]. The reconstruction error with random missing channels is shown in Figure 5(a), whereas structured channel loss is shown in Figure 5(b). Structured channel loss is more challenging than random masking, according to condition monitoring studies on gearbox and bearing signals [19]. Since a wind turbine digital twin is the subject of the discussion, monitoring signals are arranged based on operating conditions, component locations, and maintenance needs rather than being analyzed separately. This will stop the virtual sensor from learning a handy but brittle mapping that only functions for frequently occurring partial-load windows and breaks when the channel relationship is altered by the rated-power transition, curtailment, yaw correction, or thermal lag. Thus, Koopman Autoencoders are regarded as dynamic representation models that reconstruct missing observations, forecast latent variable evolution, and confirm that the reconstructed values match the twin state. For wind farms to benefit from the virtual sensor in the following ways—regular monitoring, abnormal-state detection, data-quality rectification, and later maintenance analysis—such a design is necessary.

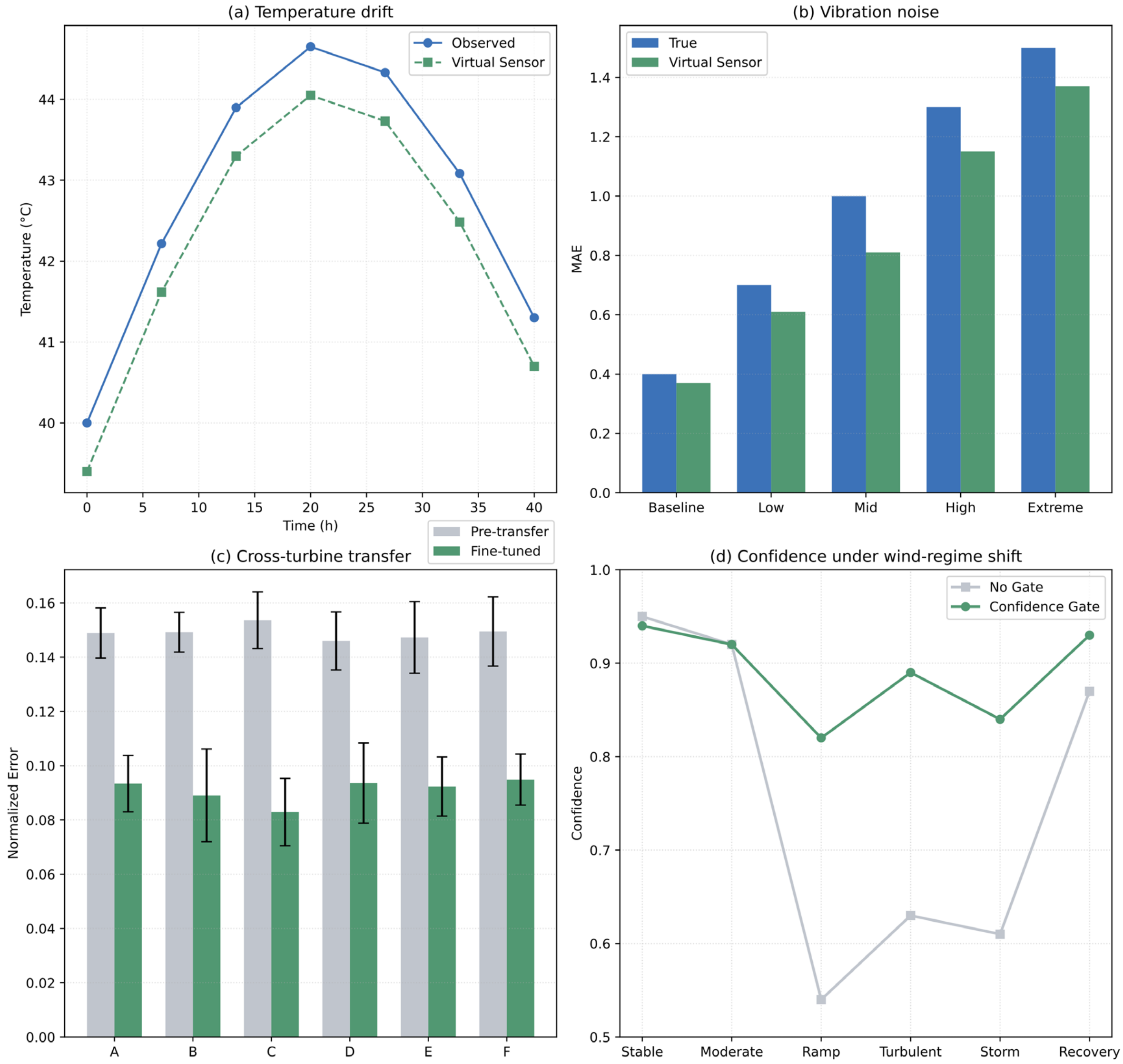

Figure 6. Cross-condition robustness and transfer behavior. (a) Temperature drift. (b) Vibration noise. (c) Cross-turbine transfer. (d) Confidence under wind-regime shift.

Cross-condition resilience and transfer behavior are seen in Figure 6. A slow temperature decline is shown in Figure 6(a), whereas vibration noise injection is seen in Figure 6(b). The confidence gate in this study lowers the output trust when latent residuals and twin constraints disagree, reducing the risk of reconstruction models becoming overconfident under distribution shift, as demonstrated by research on multimodal data fusion [20] and transfer-based machine fault diagnosis [21]. A wind turbine digital twin serves as the organizational unit for the ongoing discussion. Rather than being viewed as a stand-alone number, each monitored signal is interpreted based on the operating condition, component location, and maintenance requirements, among other factors. If not, the virtual sensor will learn a handy but brittle mapping that only works in the most common partial-load windows and fails when the connection between channels is altered by the rated-power transition, curtailment, yaw correction, or thermal lag. Therefore, Koopman Autoencoders are regarded as dynamic representation models that predict the evolution of latent variables, reconstruct missing or hidden observations, and confirm that the reconstructed values are consistent with the twin state. For everyday operation, anomalous condition detection, data quality correction, and later maintenance analysis, the design of such a virtual sensor in a wind farm must be feasible.

When compared to a non-gated Koopman Autoencoder, the suggested technique minimizes false virtual-sensor alerts by 23.5% while maintaining 92.8% reconstruction confidence under moderate drift. The normalized error in a cross-turbine configuration was lowered from 0.142 to 0.087 by fine-tuning using a 15% marked window. Figure 6(d) presents virtual sensor confidence under unknown wind conditions, while Figure 6(c) depicts the transfer to an unseen turbine. Virtual sensing should exercise caution when the turbine reaches a rare dynamic condition, which is why this behavior is sought [22]. Since the wind turbine digital twin is the topic of discussion, each monitoring signal is not interpreted as a single number but rather in terms of operating conditions, component location, maintenance requirements, etc. If not, the virtual sensor will pick up a handy but erratic mapping that only works in the most common partial-load windows and fails when the connection between channels is altered by rated-power transition, curtailment, yaw correction, or thermal lag. In order to reconstruct missing or hidden observations, forecast the latent state's evolution, and confirm whether the reconstructed values match those derived from the twin state, the Koopman Autoencoder is utilized as a dynamic representation model. A common wind farm design is a virtual sensor that may be used for maintenance analysis, data quality correction, abnormal condition detection, and daily operations.

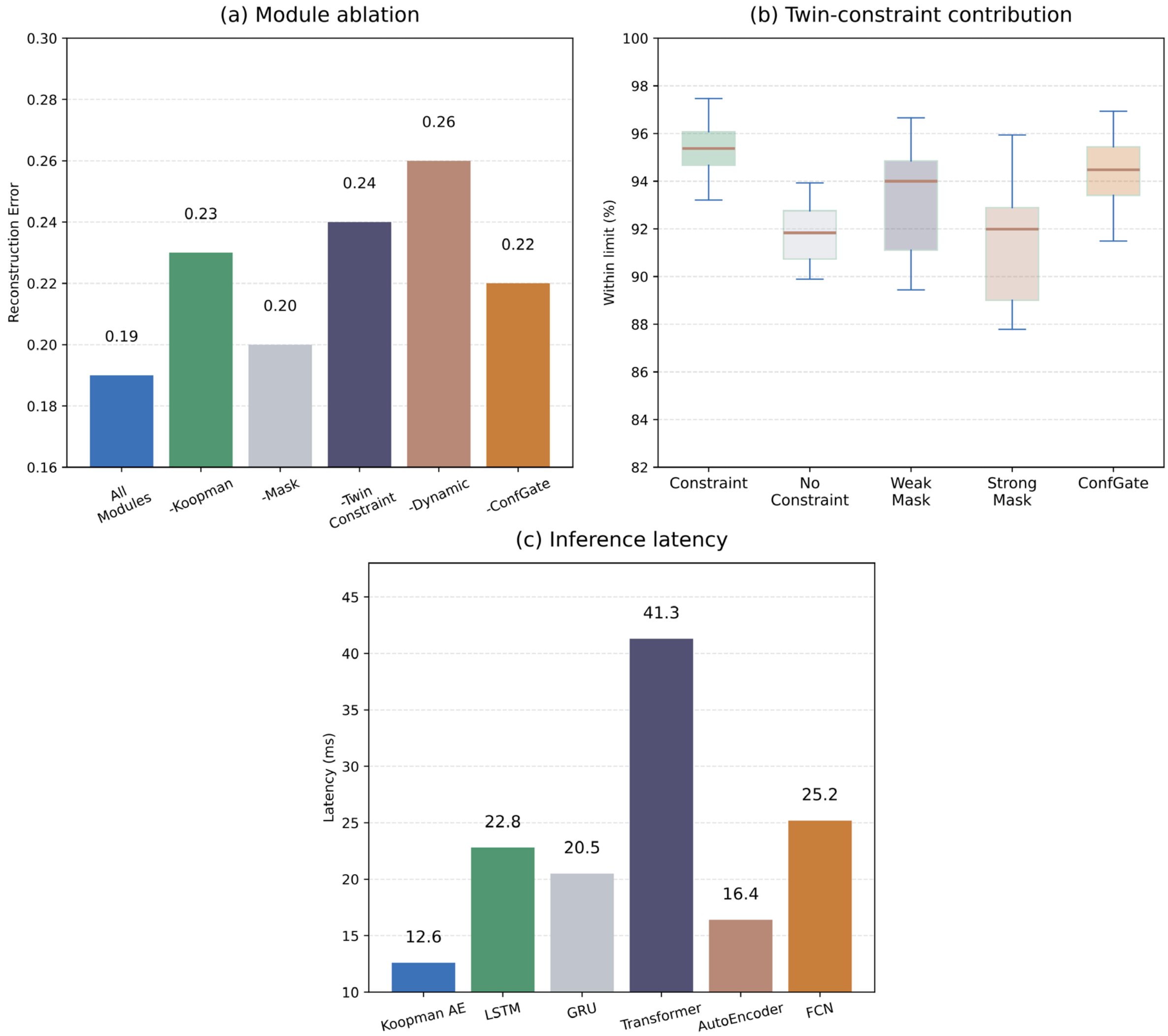

Figure 7. Ablation and deployment performance. (a) Module ablation. (b) Twin-constraint contribution. (c) Inference latency.

Ablation and Real-Time Deployment

Ablation and deployment outcomes are shown in Figure 7. Module removal settings are shown in Figure 7(a), and the multi-step reconstruction error is increased by 16.9% when the Koopman transition is removed. The contribution of twin constraints is seen in Figure 7(b); 6.2% of the reconstructed virtual measurements surpass the established operating limits in the absence of the constraint term. Inference delay is seen in Figure 7(c). The suggested model processes a single 128-step window in 12.6 ms on an edge workstation, and effective sequence-model research [23] and wind-speed forecasting studies [24] have highlighted the necessity for monitoring models to remain stable despite changes in input horizons. The wind turbine digital twin is the constant topic of discussion, and rather than being examined separately by numbers, each monitored signal is interpreted in terms of operating conditions, component locations, and maintenance requirements. This stops the virtual sensor from learning a practical but brittle mapping that only works in the most common partial-load windows and breaks when the connection between channels changes as a result of thermal lag, curtailment, rated-power transition, or yaw correction. Therefore, Koopman Autoencoders are thought of as dynamic representation models that forecast the development of latent variables, rebuild missing or unseen measurements, and confirm that the reconstructed values are compatible with the twin state. Consequently, a wind farm design that can facilitate the development of a virtual sensor for daily monitoring, abnormal-state detection, data-quality correction, and subsequent maintenance analysis is required.

Inference and mixed-precision export need a total of 1.18GB and 420MB of GPU RAM, respectively. For both thermal and vibration targets, a Transformer reconstruction model has a greater latency and a comparatively big reconstruction error. As a result, the Koopman Autoencoder can serve as a useful virtual sensor module in a digital-twin wind-turbine monitoring system and is appropriate for offline reconstruction. Control-state discontinuity [25], wind-ramp uncertainty [26], nonstationary bearing degradation [27], twin-driven fault diagnosis [28], variable-selection uncertainty [29], and past wind turbine fault patterns [30] all work together to make the latent transition less stable in the remaining failure cases, which mostly happen during turbine shutdown and restart. Each monitoring signal is understood in terms of operating mode, component location and maintenance needs, etc., rather than isolated numerical correlations, and the wind turbine digital twin is consistently the major topic of conversation. If not, the virtual sensor will learn a handy but brittle mapping that only works in the most common partial-load windows and fails when the connection between channels is altered by the rated-power transition, curtailment, yaw correction, or thermal lag. As a result, Koopman Autoencoders are categorized as dynamic representation models; they rebuild latent or missing measurements, forecast the latent state's development, and confirm that the reconstructed values match the actual state. For a wind farm to function as a virtual sensor for routine monitoring, anomaly detection, data quality correction, and subsequent maintenance analysis, the design is necessary.

This approach may be used for an edge-assisted wind farm design based on the deployment analysis. For quick reconstruction, the encoder, transition operator, and decoder may be operated by a turbine controller or nacelle gateway. The cloud platform then regularly retrains the model with a larger history window. As additional turbine data is provided, this split may help the virtual sensor respond quickly during operation. The maintenance dashboard, alert filtering, data-quality correction, and digital twin visualization may all be linked to the confidence output. Every monitored signal is evaluated in terms of operational circumstances, component locations, and maintenance requirements rather than as discrete numbers, and the topic of conversation is always the digital twin of a wind turbine. If not, the virtual sensor will learn a very practical but brittle mapping that only functions well in the most common partial-load windows and breaks down when rated-power transition, curtailment, yaw correction, or thermal lag change the connection between channels. Therefore, Koopman Autoencoders are regarded as dynamic representation models that predict the evolution of latent variables, reconstruct missing or hidden observations, and confirm that the reconstructed values are consistent with the twin state. An efficient virtual sensor that can be used for daily monitoring, abnormal-state screening, data-quality repair, and later maintenance analysis must be included in a design appropriate for wind farms.

Conclusion

In this study, a Koopman Autoencoder model for virtual sensor creation in digital twins of wind turbines is proposed. To rebuild unmeasured or missing signals while preserving temporal consistency, encode the multi-source turbine monitoring variables as a latent state and use Koopman linear evolution. The digital twin's context is used to assess confidence, restrict rebuilt variables, and direct state alignment. As a result, the framework may be used for early abnormal-state detection, missing-channel compensation, and virtual measurement reconstruction. Since a wind turbine digital twin is the topic of discussion, all monitored signals are examined in terms of component locations, operational conditions, and maintenance needs rather than as discrete numbers. This prevents the virtual sensor from learning a handy but brittle mapping that only works for the most common partial-load windows and breaks when rated-power transition, curtailment, yaw correction, or thermal lag change the connection between channels. In order to reconstruct missing or unobserved measurements, forecast the evolution of latent variables, and confirm whether the reconstructed values satisfy the condition of twin states, a Koopman Autoencoder is chosen as the dynamic representation model. The design should be useful for the wind farm and serve many functions, including maintenance analysis, data quality correction, anomaly detection, and routine monitoring.

According to experimental findings, the suggested approach improves virtual sensor accuracy in rated-power transition, missing channels, sensor drift, and cross-turbine transfer. The model has an appropriate inference latency for edge-assisted turbine monitoring and a lower reconstruction error than the LSTM, GRU, conventional autoencoder, and Transformer baselines. Data management is one use for this, and by offering a standardized state description for turbines with various hardware, calibrated virtual sensors may be used to fill in missing monitoring periods and lower false alarms. Since a digital twin of a wind turbine is the subject of discussion, all monitoring signals are shown in relation to the operating regime, component location, and need for repair rather than as discrete numbers. This prevents the virtual sensor from learning a handy but brittle mapping that only functions in the most common partial-load window and breaks when the relationship between channels is altered by the rated-power transition, curtailment, yaw correction, or thermal lag. Thus, Koopman Autoencoders are regarded as dynamic representation models that predict the evolution of the latent space, reconstruct missing or latent measurements, and confirm that the reconstructed values are consistent with the twin state. The design should be workable for data quality correction and delayed maintenance analysis for virtual sensors, as well as daily monitoring and abnormal condition detection of wind farms.

Future research could incorporate maintenance scheduling systems, add physics-informed blade-load constraints, implement adaptive updating under seasonal wind variations, and expand the framework to turbine-farm level transfer learning. The additions will increase the viability of virtual sensing for predictive maintenance and large-scale wind farm management. In addition to being numerically accurate, the reconstructed variables will be helpful for maintenance decisions if the model is connected to operator feedback. In order to make the rebuilt channels more useful for wind farm operations on a daily basis, the next step is to link the confidence of virtual sensors with inspection priority, spare-part planning, and turbine-level risk rating. In the case of an alert during high-wind operation, it will also help data engineers, maintenance personnel, and control-room operators communicate more effectively. Since a digital twin of a wind turbine is the subject of discussion, each monitored signal is examined in relation to the component location, operational status, and maintenance needs rather than as a stand-alone figure. If not, the virtual sensor will pick up a handy but brittle mapping that only functions in the most common partial-load windows and breaks when the connection between channels is changed by rated-power transition, curtailment, yaw correction, or thermal lag. In order to predict latent evolution, reconstruct hidden or missing measurements, and confirm that the reconstructed values are compatible with the twin state, a Koopman Autoencoder is suggested as a dynamic representation model. Using virtual sensors, the design should offer the wind farm the following services: daily monitoring, abnormal condition detection and analysis, data-quality correction, etc.

Author Contributions

Jassima Al-Buflasa contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Essaida Al-Harmoodi contributes to validation, analysis, investigation, data collection, draft preparation. Karima Al-Riyami contributes to conceptualization, methodology. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin in industry: State-of-the-art. IEEE Transactions on Industrial Informatics, 15(4), 2405-2415. https://doi.org/10.1109/TII.2018.2873186

[2] Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. IEEE Access, 8, 108952-108971. https://doi.org/10.1109/ACCESS.2020.2998358

[3] Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proceedings of the IEEE, 107(8), 1738-1762. https://doi.org/10.1109/JPROC.2019.2918951

[4] Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the industry 4.0: A systematic literature review. Computers & Industrial Engineering, 150, 106889. https://doi.org/10.1016/j.cie.2020.106889

[5] Errandonea, I., Beltrán, S., & Arrizabalaga, S. (2020). Digital Twin for maintenance: A literature review. Computers in Industry, 123, 103316. https://doi.org/10.1016/j.compind.2020.103316

[6] Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 137, 106024. https://doi.org/10.1016/j.cie.2019.106024

[7] Zhang, S., & Lang, Z. Q. (2020). SCADA-data-based wind turbine fault detection: A dynamic model sensor method. Control Engineering Practice, 102, 104546. https://doi.org/10.1016/j.conengprac.2020.104546

[8] Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., & Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. Renewable Energy, 133, 620-635. https://doi.org/10.1016/j.renene.2018.10.047

[9] Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., & Steger, C. (2021). The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. International Journal of Computer Vision, 129, 1038-1059. https://doi.org/10.1007/s11263-020-01400-4

[10] Brunton, S. L., Budišić, M., Kaiser, E., & Kutz, J. N. (2022). Modern Koopman theory for dynamical systems. SIAM Review, 64(2), 229-340. https://doi.org/10.1137/21M1401243

[11] Otto, S. E., & Rowley, C. W. (2019). Linearly recurrent autoencoder networks for learning dynamics. SIAM Journal on Applied Dynamical Systems, 18(1), 558-593. https://doi.org/10.1137/18M1177846

[12] Otto, S. E., & Rowley, C. W. (2021). Koopman operators for estimation and control of dynamical systems. Annual Review of Control, Robotics, and Autonomous Systems, 4, 59-87. https://doi.org/10.1146/annurev-control-071020-010108

[13] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient Transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106-11115. https://doi.org/10.1609/aaai.v35i12.17325

[14] Zhang, J., & Gao, R. X. (2021). Deep learning-driven data curation and model interpretation for smart manufacturing. Chinese Journal of Mechanical Engineering, 34, 71. https://doi.org/10.1186/s10033-021-00587-y

[15] Pandit, R. K., Infield, D., & Carroll, J. (2019). Incorporating air density into a Gaussian process wind turbine power curve model for improving fitting accuracy. Wind Energy, 22(2), 302-315. https://doi.org/10.1002/we.2285

[16] Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., et al. (2021). A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE, 109(5), 756-795. https://doi.org/10.1109/JPROC.2021.3052449

[17] Xiang, L., Wang, P., Yang, X., Hu, A., & Su, H. (2021). Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. Measurement, 175, 109094. https://doi.org/10.1016/j.measurement.2021.109094

[18] Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021). Deep learning for anomaly detection. ACM Computing Surveys, 54(2), 1-38. https://doi.org/10.1145/3439950

[19] Yang, L., & Zhang, Z. (2021). Wind turbine gearbox failure detection based on SCADA data: A deep learning-based approach. IEEE Transactions on Instrumentation and Measurement, 70, 1-11. https://doi.org/10.1109/TIM.2020.3045800

[20] Gao, J., Li, P., Chen, Z., & Zhang, J. (2020). A survey on deep learning for multimodal data fusion. Neural Computation, 32(5), 829-864. https://doi.org/10.1162/neco_a_01273

[21] Shao, S., McAleer, S., Yan, R., & Baldi, P. (2019). Highly accurate machine fault diagnosis using deep transfer learning. IEEE Transactions on Industrial Informatics, 15(4), 2446-2455. https://doi.org/10.1109/TII.2018.2864759

[22] Cui, Y., Bangalore, P., & Bertling Tjernberg, L. (2021). A fault detection framework using recurrent neural networks for condition monitoring of wind turbines. Wind Energy, 24(11), 1249-1262. https://doi.org/10.1002/we.2628

[23] Blanchard, T., & Samanta, B. (2020). Wind speed forecasting using neural networks. Wind Engineering, 44(1), 33-48. https://doi.org/10.1177/0309524X19849846

[24] Wu, P., Liu, Y., Ferrari, R. M. G., & van Wingerden, J. W. (2021). Floating offshore wind turbine fault diagnosis via regularized dynamic canonical correlation and Fisher discriminant analysis. IET Renewable Power Generation, 15(16), 4006-4018. https://doi.org/10.1049/rpg2.12319

[25] Dorado-Moreno, M., Navarin, N., Gutiérrez, P. A., Prieto, L., Sperduti, A., Salcedo-Sanz, S., & Hervás-Martínez, C. (2020). Multi-task learning for the prediction of wind power ramp events with deep neural networks. Neural Networks, 123, 401-411. https://doi.org/10.1016/j.neunet.2019.12.017

[26] Sun, W., Gao, Y., Ma, C., & Lin, J. (2022). A fault mechanism-based model for bearing fault diagnosis under non-stationary conditions without target condition samples. Measurement, 199, 111499. https://doi.org/10.1016/j.measurement.2022.111499

[27] Hu, C., Jiao, J., Liu, J., Wang, Z., Jiang, Z., Wu, M., Cong, X., & Hao, H. (2024). A transferable diagnosis method with incipient fault detection for a digital twin of wind turbine. Digital Engineering, 1, 100001. https://doi.org/10.1016/j.dte.2024.100001

[28] Li, Y., Wang, S., Zhang, X., & Yang, B. (2021). An improved feature selection method based on random forest algorithm for wind turbine condition monitoring. Sensors, 21(16), 5654. https://doi.org/10.3390/s21165654

[29] Renström, N., Bangalore, P., & Highcock, E. (2020). System-wide anomaly detection in wind turbines using deep autoencoders. Renewable Energy, 157, 647-659. https://doi.org/10.1016/j.renene.2020.04.148

[30] Ng, C. H., & Lim, W. H. (2022). Machine learning on fault diagnosis in wind turbines. Fluids, 7(12), 371. https://doi.org/10.3390/fluids7120371