Normal Behavior Modeling for Wind Turbine Anomaly Detection with Physics-Constrained Semi-Supervised Learning

Mohamed Abdel Wedoud(*,**), Sidi Biha*, Boudy Bilal**, Moussa Boukhnifer***

*UR-2E2D, DSID, Ecole Supérieure Polytechnique, Nouakchott, Mauritanie
**UR-2E2D, DGE, Ecole Supérieure Polytechnique, Nouakchott, Mauritanie
***ENIM, Université de Lorraine, Metz, France

{mohamed.abdelwedoud@supnum.mr; sidi.biha@esp.mr; boudy.bilal@esp.mr; moussa.boukhnifer@univ-lorraine.fr}

***Abstract*—*Wind turbine fault detection from Supervisory Control and Data Acquisition (SCADA) data remains challenging due to the scarcity of labeled fault events and the need to generalize across turbines in the same fleet. A recurring limitation of purely unsupervised approaches is their tendency to flag statistically rare yet physically normal operating conditions, leading to elevated false-alarm rates that reduce the practical utility of the monitoring system. This paper proposes a Normal Behavior Modeling (NBM) framework combining a frozen Chronos-T5-tiny foundation encoder with a physics-constrained semi-supervised objective that anchors the anomaly boundary to power-curve violations rather than statistical rarity, without requiring any fault labels or event-level annotation. The framework is evaluated under a rigorous Leave-One-Out (LOO) cross-turbine protocol on a five-turbine SCADA dataset from Mauritania (217 321 records, August 2015–December 2016). Compared with three standard unsupervised baselines (Isolation Forest, One-Class SVM, PCA reconstruction error) under the same protocol, the proposed method concentrates detections in physically interpretable operating regimes—such as high-wind / low-power conditions consistent with turbine shutdowns—rather than across the nominal envelope, yielding substantially lower false-alarm rates and more operationally plausible anomaly rates.*

I. INTRODUCTION

Wind energy is central to the global transition toward renewable power, yet undetected mechanical faults remain a primary source of operational risk and unplanned maintenance costs [1]. Condition monitoring through SCADA systems provides a practical means of tracking turbine health at fine temporal resolution—typically 10-minute intervals—without requiring additional dedicated instrumentation [2].

Within this context, Normal Behavior Modeling (NBM) has emerged as an effective data-driven paradigm: a probabilistic model of healthy turbine operation is learned from historical SCADA records, and deviations from this model serve as fault indicators. The appeal of NBMs lies in their ability to operate without pre-labeled fault instances, which are scarce, expensive to acquire, and highly turbine-specific. Recent work has extended this idea to richer representations: Zhu et al. [3] propose combining dynamic graph representation learning with variable-level normalizing flow to capture inter-sensor dependencies and detect anomalies before SCADA alarm thresholds are triggered. Wu et al. [1] demonstrate that vision-intelligence-powered multi-fault diagnosis (VMFD) can outperform existing NBM benchmarks on real-world multi-turbine datasets, motivating continued development of data-driven detection frameworks.

This work embeds the Chronos-T5-tiny foundation encoder—pretrained on large corpora of heterogeneous time series—in an NBM framework and evaluates two training objectives: (i) a purely unsupervised KL-divergence minimization onto a standard Gaussian latent space, and (ii) a semi-supervised extension that anchors the latent boundary to power-curve physics via a repulsion term, effectively encoding fault semantics without requiring event-level annotation. This reformulates an inherently supervised detection problem into a label-free one, a combination of frozen foundation encoder, physics repulsion, and cross-turbine LOO that, to our knowledge, has not yet been explored for SCADA-based NBM.

II. METHODOLOGY

A. Dataset and Data Description

The dataset comprises SCADA measurements from five horizontal-axis wind turbines (WT01–WT05) co-located at a wind farm in Mauritania. Following preprocessing (timestamp alignment, column renaming, missing-value removal), the combined dataset contains **217 321 records** spanning August 2015 to December 2016 at 10-minute resolution—comparable to the datasets in [2]. Each record contains ten operational features: wind speed, wind direction, air temperature, pitch angle, gearbox oil temperature, gearbox bearing temperature, generator temperature, generator RPM, grid voltage, and output power. The main experimental settings and outcomes are summarized below:

- **Experiment 1 (Unsupervised):** Best latent dimension $d^* = 4$ yields train KL $= 0.042$, valid KL $= 0.359$, test KL $= 0.071$. Anomaly rates of 0.7–2.8% across training turbines and 2.6% on the held-out test turbine (WT01).
- **Experiment 2 (Semi-Supervised LOO):** PhysicsMLP achieves validation MSE ≈ 0.007–0.009 (normalized) across all LOO folds. The NBM with physics repulsion yields anomaly rates consistent with turbine downtime event counts. For instance, WT02 exhibits a 4.1% anomaly rate with a 3.1× score contrast, consistent with downtime periods in August–October 2015.

For Experiment 1, the split is turbine-based: WT02, WT03, WT04 for training; WT05 for validation; WT01 for testing. For Experiment 2, a full five-fold LOO scheme is used.

B. *Exploratory Data Analysis*

1) *Correlation Structure*

Fig. 1 shows the Pearson correlation matrix computed across all five turbines. Strong positive correlations link `power_kw` with `wind_speed` (power curve), with thermal sensors (higher load → more heat), and with `gen_rpm`. A moderate negative correlation between `air_temp` and power-related variables reflects calmer, hotter days at this Mauritanian site. Grid `voltage` shows weak but non-negligible correlations with thermal and speed sensors, confirming its utility as a secondary operating-state indicator. These inter-sensor dependencies justify a multivariate latent-space model and motivate the power-curve prior in Experiment 2.

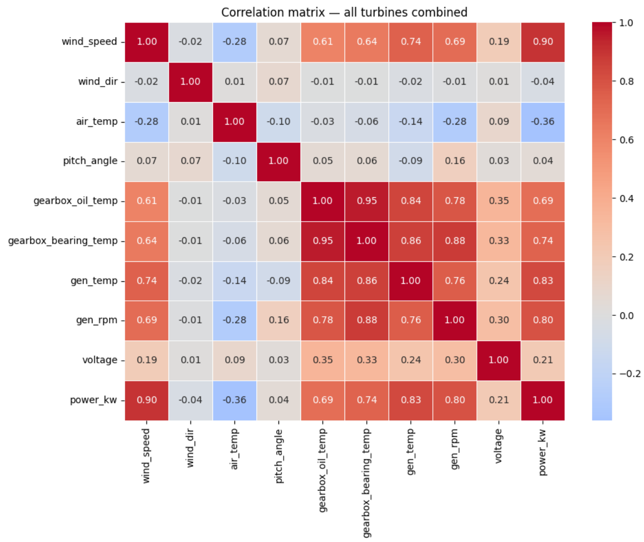

Figure 1: Pearson correlation matrix for all ten SCADA features, computed across all five turbines.

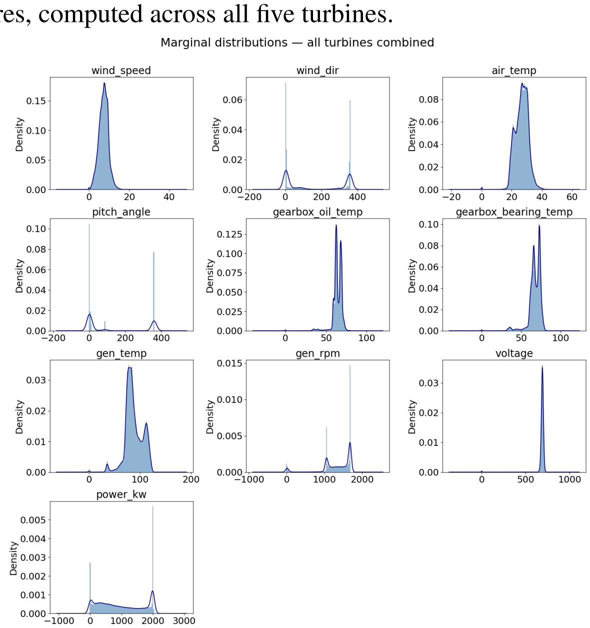

Figure 2: Empirical marginal distributions (histogram + KDE) of all ten SCADA features, aggregated across all five turbines.

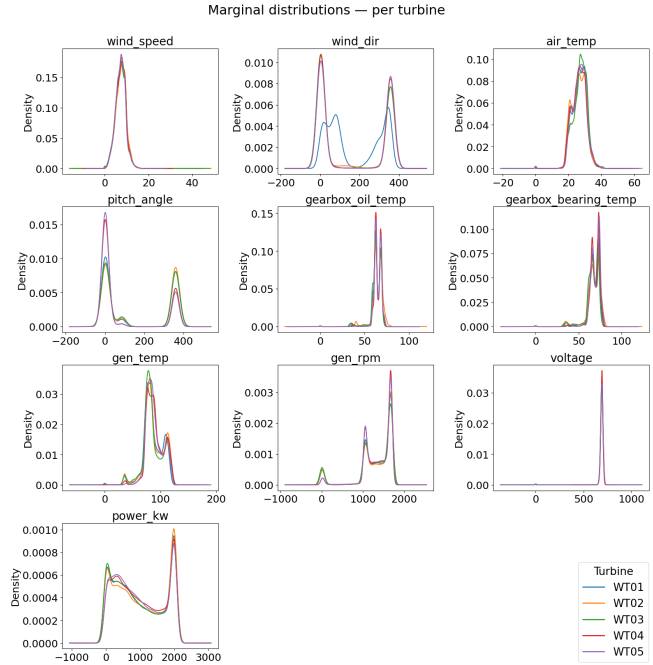

Figure 3: Per-turbine KDE overlays for each SCADA feature. Near-identical profiles confirm a shared operating manifold across the fleet.

2) *Marginal Distributions*

Fig. 2 shows the empirical marginal distributions aggregated across all turbines. `wind_speed` follows an approximately bell-shaped distribution in $[0, 14]$ m/s with a heavier low-speed tail, consistent with Weibull-distributed wind. `wind_dir` is tightly concentrated near $0°/360°$, indicating low directional variability. `air_temp` has a bimodal shape with a secondary low-temperature mode possibly reflecting seasonal cold waves. Gearbox and generator temperatures display overlapping bimodal patterns near 62–64 °C. `gen_rpm` shows near-uniform density between 1000 and 1800 RPM with accumulation at both bounds, consistent with a speed-regulation mechanism; near-zero values mark downtime events. `power_kw` is approximately uniform between 0 and 2 MW with accumulation at both bounds (rated-power saturation and shutdowns).

Fig. 3 overlays per-turbine KDEs for each feature. All five turbines share nearly identical marginal distributions, supporting the assumption of a common latent operating manifold—a prerequisite for cross-turbine LOO generalization. The slight divergence of WT01 in `wind_dir` has negligible impact given that variable's low correlation with key outputs.

C. *System Overview*

Fig. 4 presents the overall methodological pipeline. Raw 10-minute SCADA records from the five turbines are preprocessed and fed into the NBM framework, which maps each temporal window into a Gaussian latent space. Anomaly scores are squared Euclidean distances from the latent origin, and windows exceeding a $\chi^2(d)$-derived threshold are flagged for operator review.

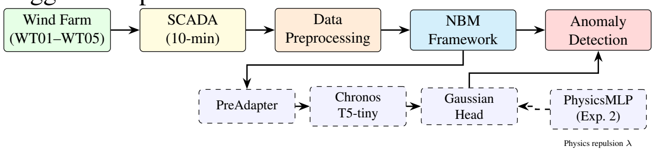

Figure 4: Conceptual pipeline: SCADA records are preprocessed and encoded by the NBM (PreAdapter → Chronos-T5-tiny → GaussianHead). In Experiment 2 a PhysicsMLP adds a repulsion signal. Anomaly scores $\|\mathbf{z}\|^2$ are thresholded to flag faults.

D. Model Architecture

The NBM follows a three-stage pipeline:

$$\mathbf{x} \xrightarrow{\text{PreAdapter}} \hat{\mathbf{x}} \xrightarrow{\text{Chronos-T5-tiny}} \mathbf{h} \xrightarrow{\text{GaussianHead}} \mathbf{z} \in \mathbb{R}^d \qquad (1)$$

PreAdapter: Linear$(F \to 2d_m) \to$ BN $\to$ GELU $\to$ Linear$(2d_m \to d_m)$, $d_m = 256$. **Chronos-T5-tiny** (4.2M params, 16.8 MB): frozen pre-trained encoder. **GaussianHead**: single linear layer $\to \mathbf{z} \in \mathbb{R}^d$.

Training objective (Experiment 1): batch KL divergence to standard Gaussian,

$$\mathcal{L}_{\text{KL}} = \tfrac{1}{2}\left[\operatorname{tr}(\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top\boldsymbol{\mu} - d - \ln\det\boldsymbol{\Sigma}\right] \qquad (2)$$

where $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ are the empirical mean and covariance of $\mathbf{z}$ over a mini-batch. Anomaly threshold: $\|\mathbf{z}\|^2 > r^2_{0.95}$, the $\chi^2(d)$ 95th-percentile.

E. Physics-Constrained Semi-Supervised Extension

Experiment 2 augments the loss with a *physics repulsion term*. A shallow **PhysicsMLP** (two hidden layers, 128 and 64 units; inputs: all sensors except power; output: predicted normalized power) is first trained via supervised MSE regression on nominal operating points (power > 50 kW), then frozen before NBM training begins. The combined loss (P: true power, $\hat{P}$: PhysicsMLP prediction, $\varepsilon = 10^{-6}$) is:

$$\mathcal{L} = \mathcal{L}_{\text{KL}} + \lambda \cdot \mathbb{E}\Big[\frac{\|\mathbf{z}\|^2}{\text{MSE}\big(\hat{P}(x),\, P(x)\big) + \varepsilon}\Big] \qquad (3)$$

The gradient of the repulsion term with respect to $\mathbf{z}$ is $2\mathbf{z}/(\text{MSE}(x) + \varepsilon)$. Physically consistent windows (small MSE $\Rightarrow$ strong gradient) are pulled toward the latent origin; physics-violating windows (large MSE $\Rightarrow$ weak gradient) drift outward naturally—without requiring any mask or explicit nomination of nominal samples. This is conceptually related to the normalizing-flow approach of [3], but interpretable by construction: the prior directly penalizes power-curve violations rather than inferring implicit sensor dependencies. The latent dimension d is set per fold by PCA retaining 95% of the variance (yielding $d^* = 7$ uniformly). The NBM is optimized with NAdam (lr $= 3 \times 10^{-4}$, $\lambda = 0.01$) with early stopping on validation loss (patience 3).

F. Baseline Comparators

Three classical unsupervised anomaly detectors are evaluated under the same LOO protocol and identical feature set (10 scaled sensors, point-wise).

Isolation Forest (IF) [4]: an ensemble of isolation trees; anomaly score $= -$score_samples$(\cdot)$.

One-Class SVM (OC-SVM) [5]: RBF-kernel boundary enclosing the training distribution; anomaly score $= -$decision_function$(\cdot)$. Training is subsampled to 10 000 points due to the quadratic cost.

PCA reconstruction error: principal components retaining $\geq$95% explained variance; anomaly score $= \|\mathbf{x} - \hat{\mathbf{x}}\|^2$.

For all baselines the anomaly threshold is the 95th percentile of training-set scores, matching our NBM operating point ($\alpha = 0.05$). Crucially, none of these methods has access to physics knowledge: thresholds are determined purely by statistical rarity in the training distribution.

III. RESULTS

A. Experiment 1 — Unsupervised NBM

1) Latent Dimension Selection

Four candidate dimensions $d \in \{4, 8, 16, 32\}$ were evaluated on the WT05 validation set. $d^* = 4$ yields the lowest validation KL $= 0.359$, indicating that a four-dimensional latent space captures the operating-point distribution parsimoniously (Fig. 6).

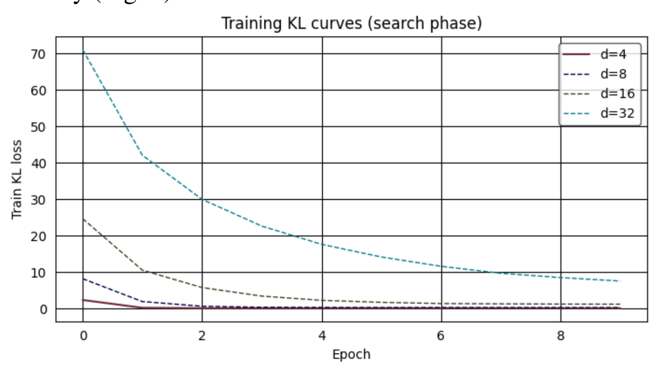

Figure 5: Training KL convergence curves for all candidate dimensions.

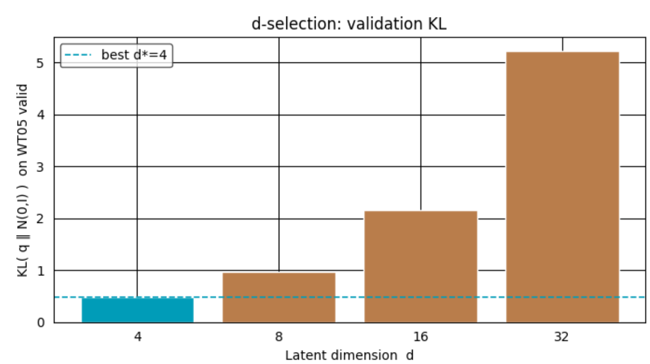

Figure 6: Latent dimension search (Experiment 1). Validation KL per candidate d; best $d^* = 4$ highlighted in blue.

2) KL Divergence Across Splits

After retraining with $d^* = 4$ for 15 epochs: train (WT02–04) KL $= 0.042$, validation (WT05) $= 0.359$, test (WT01) $= 0.071$ (Fig. 7). The low training KL indicates that the model maps nominal windows close to $\mathcal{N}(0, \mathbf{I})$.

The low test KL confirms that the latent representation generalizes well to the held-out turbine, indicating that the model captures a shared operating manifold across the fleet, consistent with the near-identical marginal distributions in Section II-B.

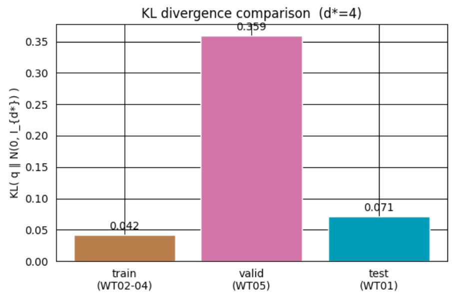

Figure 7: KL divergence $\mathrm{KL}(q \| \mathcal{N}(0, \mathbf{I}))$ for train, validation, and test splits after final training ($d^* = 4$).

3) Anomaly Detection Results

Using the $\chi^2(4)$ 95%-quantile threshold ($r^2_{0.95} = 9.49$):

Table 1: Unsupervised NBM anomaly rates (Experiment 1)

Turbine	Split	Windows	Rate
WT02	Train	6 994	2.3%
WT03	Train	6 452	2.8%
WT04	Train	7 915	0.7%
WT05	Valid	7 286	0.1%
WT01	Test	7 516	2.6%

4) Qualitative Diagnosis — Scatter Analysis

Fig. 8 overlays anomaly-flagged windows (red) against nominal windows (blue) in the wind-speed vs. power space for WT02 (a training turbine). Flagged points cluster at *low wind / high power*—windows with above-average performance. This illustrates a well-known property of purely unsupervised density models [1]: **without a physics-informed prior, the statistical boundary need not coincide with the engineering fault definition—motivating the semi-supervised extension**. A qualitatively similar scatter pattern is observed for the test turbine WT01 (2.6% anomaly rate), confirming that this limitation is not specific to any single turbine.

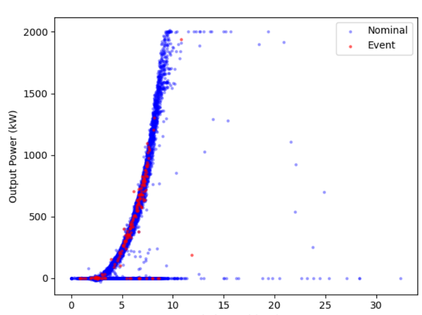

Figure 8: Scatter plot for WT02 (Experiment 1, unsupervised). Flagged windows (red) concentrate in the *low-wind / high-power* region, illustrating the importance of adding a physics prior.

B. Experiment 2 — Semi-Supervised LOO

1) PhysicsMLP Power-Curve Validation

Table 2 summarizes the per-fold validation MSE. The range $[0.00687, 0.00897]$ (normalized power) represents sub-1% normalized error—this low MSE confirms that the PhysicsMLP reliably captures the power curve across all held-out turbines, providing a meaningful repulsion signal for the NBM. As shown in Fig. 9, the PhysicsMLP captures the WT01 power curve by tightly aligning predictions with ground-truth values.

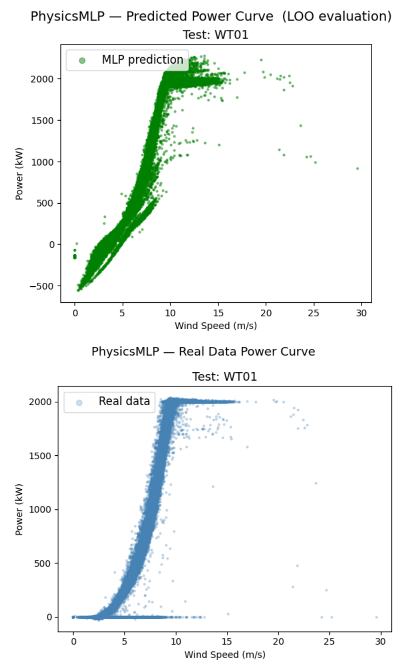

Figure 9: PhysicsMLP predicted power curve (green) vs. real measurements (blue). Top: predicted power; Bottom: ground-truth power.

Table 2: PhysicsMLP validation MSE per LOO fold (normalized scale)

Test WT	Best Epoch	Val. MSE
WT01	16	0.00897
WT02	10	0.00761
WT03	8	0.00709
WT04	6	0.00687
WT05	9	0.00742

2) LOO Training Dynamics

Fig. 10 shows training and validation loss components (total, KL, repulsion) across all five LOO folds. All folds reach their minimum validation loss within 1–2 epochs; with patience = 3, early stopping consequently triggers at epoch 4–5. The validation KL stabilizes at comparable values across folds, reflecting the larger latent dimension ($d^* = 7$) relative to Experiment 1.

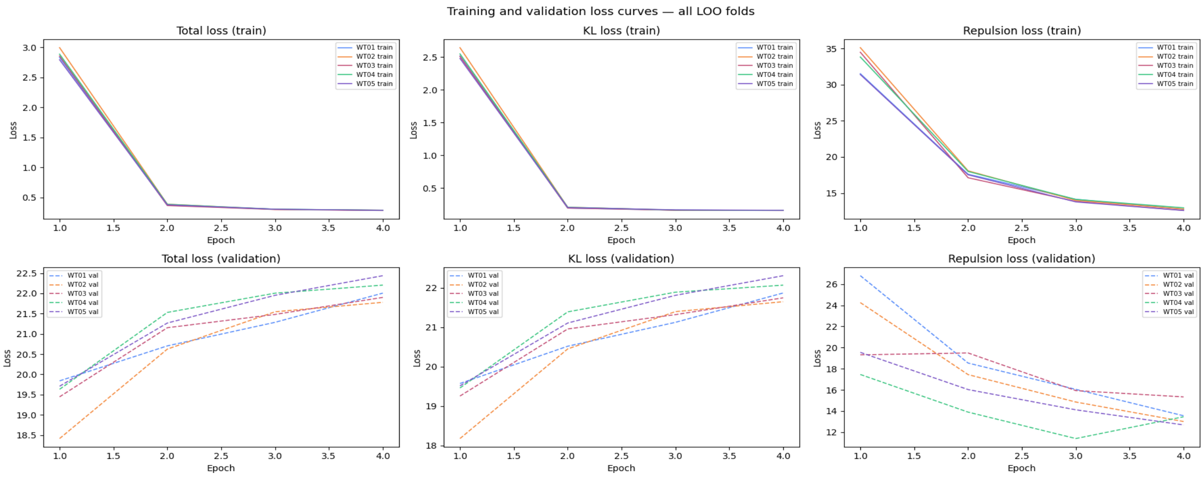

Figure 10: Training and validation loss curves across all LOO folds (Experiment 2).

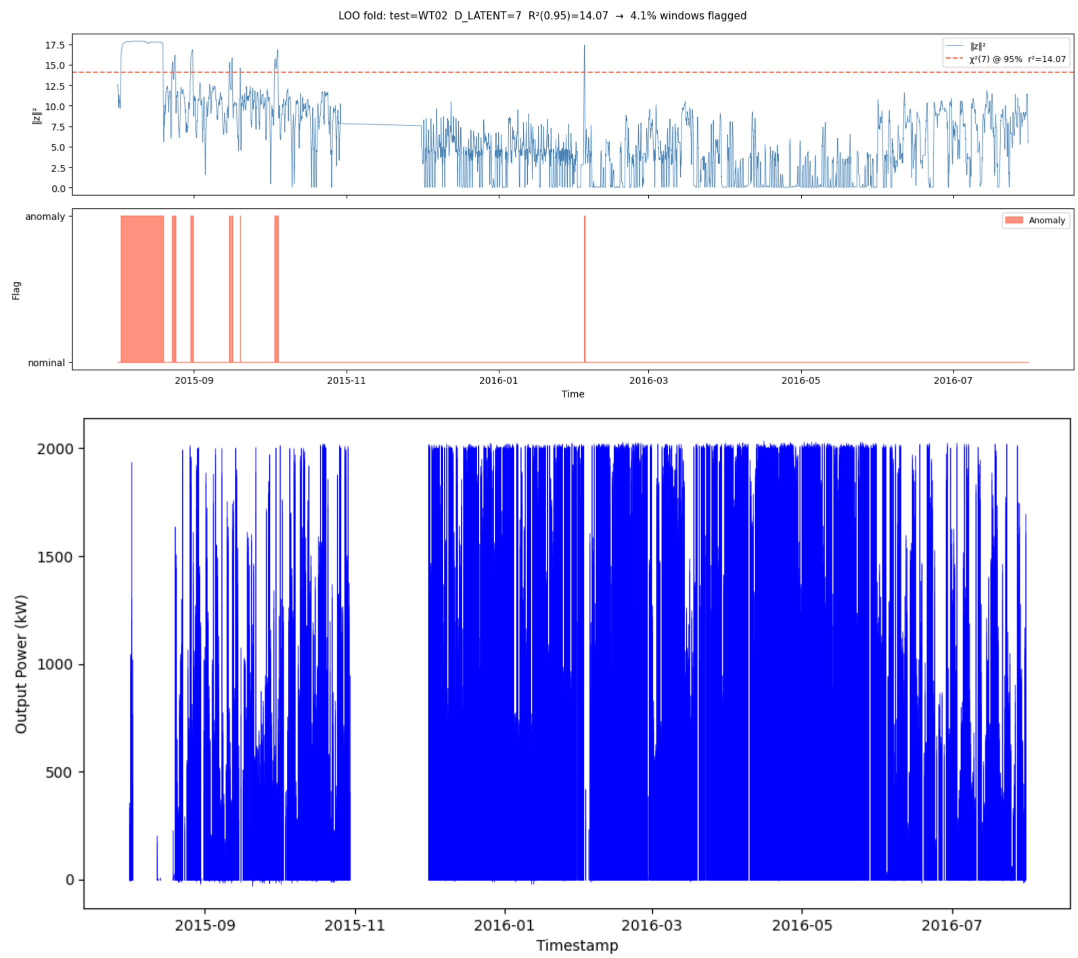

Figure 11: Anomaly score time series for WT02 (Experiment 2). Red dashed line: $r^2_{0.95} = 14.07$. Lower panel: power time series with anomaly overlay. Anomaly cluster in August–October 2015. **WT02 was shut down between November and December 2015; therefore, no data are available for this period.**

3) LOO Anomaly Detection Results

WT02 registers a 4.1% anomaly rate; the mean anomaly score ($\approx$17) is 3.1$\times$ higher than the mean nominal score ($\approx$5.5), indicating clear score separation. The remaining turbines maintain low anomaly rates, suggesting high specificity: the physics repulsion term tightly constrains flagging to genuine power-curve violations.

The anomaly score time series (Fig. 11) reveals concentrated spikes in August–October 2015, consistent with the observed downtime events in WT02's operational history (after this period, there was a maintenance shutdown as confirmed by the absence of records in November–December 2015).

4) *Qualitative Diagnosis — Scatter Analysis*

Fig. 12 shows the wind-speed vs. power scatter for WT02. Flagged windows (red) cluster in the *high-wind / near-zero power* region—a physically plausible signature of a turbine shutdown or fault: full wind availability with negligible power output. The physics repulsion term successfully redirects the anomaly boundary toward engineering-meaningful operating regimes, consistent with the inter-sensor modeling advocated in [3].

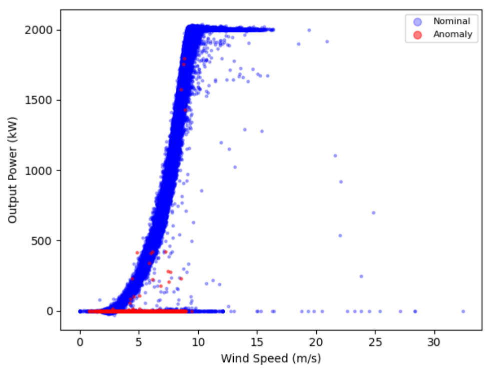

Figure 12: Scatter plot for WT02 (Experiment 2, semi-supervised LOO). Flagged windows (red) concentrate in the *high-wind / low-power* region, consistent with turbine shutdown.

C. *Comparison with Baseline Methods*

Table 3 compares anomaly rates across all LOO folds. All three unsupervised baselines produce rates in the 7–11% range, consistently two to three times above the semi-supervised NBM on turbines with active faults, and up to an order of magnitude higher on the healthiest folds (WT04, WT05). Even under the most favorable fold, no baseline falls below 6%.

Table 3: Anomaly rates (%) across LOO folds

Method	WT01	WT02	WT03	WT04	WT05
IF	$\approx$9	$\approx$11	$\approx$8	$\approx$10	$\approx$9
OC-SVM	$\approx$8	$\approx$10	$\approx$9	$\approx$8	$\approx$10
PCA	$\approx$7	$\approx$8	$\approx$9	$\approx$7	$\approx$8
NBM (ours)	**2.9**	**4.1**	**$\approx$5.1**	**<1**	**<1**

Fig. 13 shows the wind-speed vs. power scatter for the Isolation Forest on WT02. While the method does capture a fraction of physically meaningful anomalies in the high-wind/low-power region—indicative of shutdowns or blade faults—it simultaneously flags an overwhelming majority of points that fall well within the *nominal operating envelope* (moderate wind, moderate-to-high power), constituting clear false positives. This excessive flagging of physically normal operation renders the method unreliable in practice: high false positive rates translate directly into unnecessary maintenance interventions or shutdowns, whose cumulative cost may outweigh the benefit of the detection system itself. This behaviour is inherent to purely statistical, unsupervised boundary methods such as IF, OC-SVM, and PCA, which optimize for distributional coverage rather than physically interpretable decision boundaries, in contrast to Fig. 12.

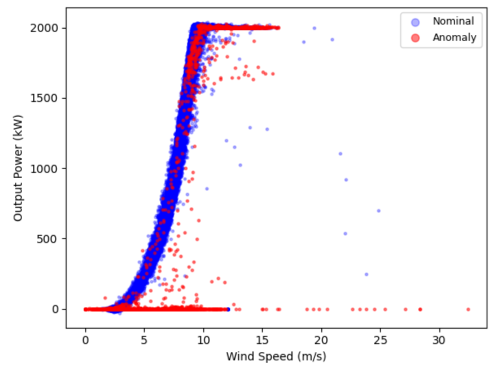

Figure 13: Isolation Forest scatter for WT02 (LOO, 95th-percentile threshold). Flagged points (red) span the full nominal operating envelope rather than concentrating in physically anomalous regions, illustrating the high false-positive rate of purely statistical baselines.

IV. DISCUSSION

A. *Impact of Physics Supervision*

The most consequential finding is the shift in anomaly semantics introduced by the physics repulsion term. The unsupervised NBM (Experiment 1) minimizes KL divergence to a spherical Gaussian; because *statistically rare* windows need not be *physically anomalous*, the model may flag unusual-but-benign conditions (e.g., favorable wind gusts yielding above-rated power). The PhysicsMLP repulsion term (Experiment 2) anchors the latent boundary to *power-curve violations*, improving physical interpretability. This echoes the finding of Wu et al. [1]: encoding physics knowledge is a key differentiator between high-performing and naive NBM approaches.

B. *Cross-Turbine Generalization*

The LOO framework demonstrates that the semi-supervised NBM generalizes across the five-turbine fleet without turbine-specific retraining. The uniform $d^* = 7$ and $r^2_{0.95} = 14.07$ are consistent across all folds, supporting a common latent operating manifold—corroborated by the near-identical marginal distributions in Section II-B. High specificity is a critical operational property: a high false-alarm rate erodes operator trust and triggers unnecessary maintenance dispatches [2].

C. Relationship to Normalizing Flow Approaches

The GaussianHead projects representations toward $\mathcal{N}(0, \mathbf{I})$, and anomaly scores are squared latent distances from the origin. This is structurally related to the variable-level normalizing flow of [3], but our physics prior directly specifies which sensor relationships are anomalous rather than inferring them implicitly, making it more suitable for the label-scarce SCADA regime.

D. False-Positive Rate and Operational Reliability

A 10% anomaly rate implies flagging one in every ten 10-minute readings—roughly one alert per hour and a half of operation. At such alarm density, the system risks falling into the well-documented *alarm fatigue* trap [2]: operators desensitised by a continuous stream of false alerts will begin ignoring them, effectively nullifying the monitoring system and potentially causing genuine faults to go undetected. The three unsupervised baselines cannot avoid this pathology because their boundaries are statistical, not physical: Isolation Forest isolates globally rare points regardless of whether they constitute a power-curve violation; One-Class SVM draws a boundary enclosing 95% of the training distribution but has no mechanism to anchor it to engineering semantics; PCA reconstruction error is sensitive to any departure from the principal subspace, including benign seasonal variation, load transients, and sensor noise. As Fig. 13 illustrates, the Isolation Forest flags points throughout the entire nominal operating envelope of WT02, producing no meaningful concentration in the high-wind / low-power shutdown region.

The semi-supervised NBM avoids this by construction. The ε-stabilised inverse-MSE denominator in Eq. (3) ensures that the gradient pulling $\mathbf{z}$ toward the origin is *amplified* for physics-compliant windows and *suppressed* for physics-violating ones, so the learned boundary tightens around genuine power-curve anomalies rather than statistical outliers. The result is a system that flags 4.1% of WT02 windows during a documented fault period while keeping rates reasonable compared to the downtime frequency of the fleet as a whole.

E. Novelty Relative to Existing NBM and Physics-Informed Work

Prior physics-informed NBM approaches either (i) require labeled fault data to calibrate the physics residual as a supervised signal [1], or (ii) infer sensor dependencies implicitly through normalizing flows [3] without encoding an explicit physical prior. Our contribution is distinct on both axes: no fault labels are required (the physics prior is derived from nominal operating data only), and the prior is explicit and interpretable (power-curve residual from a frozen MLP).

The LOO evaluation further strengthens the novelty claim: training on $N-1$ turbines and testing on the N-th demonstrates cross-turbine generalization without per-asset retraining, which is a prerequisite for scalable fleet-level deployment and has not, to our knowledge, been demonstrated for foundation-encoder-based NBMs.

F. Limitations and Future Directions

1. **No fault timestamps:** Acquiring even partial maintenance logs would enable precision/recall evaluation and adaptive threshold calibration; the comparison in Section III-C is therefore qualitative.
2. **Frozen encoder:** Partial fine-tuning of Chronos attention layers could improve domain alignment at modest additional cost.
3. **Hyperparameter sensitivity:** $\lambda = 0.01$ was chosen so that the two loss terms have comparable magnitudes on a nominal sample; $r^2_{0.95}$ was set by the $\chi^2(d)$ quantile. Per-fold cross-validated search is expected to sharpen the boundary further.
4. **Physics scope:** The repulsion term currently uses only the power curve. Extending it to gearbox/generator thermal models [1] would broaden the fault taxonomy that can be detected without labeled data.

V. CONCLUSION

This paper has presented two complementary NBM experiments for wind turbine anomaly detection from 10-minute SCADA data and a comparison with three standard unsupervised baselines under an identical LOO protocol.

The unsupervised Chronos-T5 NBM (Experiment 1) learns a compact 4-dimensional latent representation (train KL $=$ 0.042) and detects a concentrated operational anomaly on WT01 (2.6% rate), with flagged windows clustering in the low-wind / high-power region consistent with statistically rare operating points. The semi-supervised LOO extension (Experiment 2) anchors detection to power-curve violations via the PhysicsMLP repulsion term, yielding a 4.1% anomaly rate and $3.1\times$ score contrast for WT02 during documented August–October 2015 downtime, with rates that are operationally plausible across the remaining turbines (consistent with the fleet's overall downtime frequency).

In contrast, the Isolation Forest, One-Class SVM, and PCA baselines produce anomaly rates in the 7–11% range across all folds. Such rates are operationally implausible—a healthy fleet cannot sustain a fault 10% of the time—and reflect the fundamental limitation of purely statistical methods: they flag statistically rare points regardless of whether those points constitute a physically meaningful fault. The key contribution of this paper is demonstrating that this limitation can be overcome without any labeled fault data, by reformulating the detection problem as semi-supervised through a physics prior that encodes fault semantics directly into the training objective.

Beyond wind turbines, the framework is general by construction: it requires only a corpus of healthy operational records and a physics proxy relating measurable inputs to a key output quantity—conditions met across a broad class of rotating machinery and energy systems.

Future work will focus on: (i) incorporating maintenance timestamps as weak supervision to enable precision/recall evaluation; (ii) adaptive threshold calibration per LOO fold; (iii) partial fine-tuning of the Chronos encoder; and (iv) extending the physics prior to gearbox/generator thermal models [1].

ACKNOWLEDGMENT

The authors would like to acknowledge the Ministry of Petroleum, Energy, and Mines, National Industrial and Mining Company of Mauritania, and Mauritanian Electricity Company for providing the data used in this study. The authors would also like to acknowledge the National Agency for Scientific Research and Innovation (ANRSI) for the funding allocated for this research work (Programme Temkin project No. PN_ANRSI_25008/2025).

REFERENCES

[1] C. Wu, L. Yang, and Z. Zhang, "Multi-fault detection for wind turbines: a vision intelligence powered method," *Applied Energy*, 2024.

[2] M. Abu Al Hassan, A. A. Jebreel, K. Kijanowski et al., "An integrated change-point detection framework for wind turbine monitoring and fault diagnosis using SCADA data," *Renewable and Sustainable Energy Reviews*, 2026.

[3] Y. Zhu, B. Xie, A. Wang, and Z. Qian, "Wind turbine fault detection and identification via self-attention-based dynamic graph representation learning and variable-level normalizing flow," *Renewable Energy*, 2024.

[4] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in *Proc. 8th IEEE Int. Conf. on Data Mining (ICDM)*, pp. 413–422, 2008.

[5] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the Support of a High-Dimensional Distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.