Data Drift Detection and Assessment for AI-hybrid Models Applied on Electrical Energy Consumption

Faouzi Adjed[1], **Milad Leyli-Abadi**[1], **Elies Gherbi**[1], **Martin Gonzalez**[1]
[1] IRT SystemX, 2 Boulevard Thomas Gobert 91120 Palaiseau, France
{faouzi.adjed, milad.leyli-abadi, elies.gherbi, martin.gonzalez}@irt-systemx.fr

Abstract

Data drift evaluation is crucial in the operational step in the industry. In the real world, several drift types are usually contributing to drift detection, which may come from input data and output data distributions. In addition, the application context and the interpretation of these drift types add complexity to drift analysis. In this work, we apply drift detection in the specific domain of electrical transmission network systems. Three drift types, covariate, label, and concept drift, are considered and implemented on systems based on Physics-Informed Neural Networks (PINNs). The experimental results show the impact of each drift type and the evolution of their contributions when drift occurs in the industrial system. A contextual interpretation of the obtained results is also developed in this specific application domain for the three drift types.

Introduction

The data quality is the key for estimating the relevance, the health, the reliability, and trustworthiness of the usage of a dataset in the building of an Artificial Intelligence (AI)-based system (Ehrlinger and Wöß 2022). It encompasses several technical metrics, such as representativeness, completeness, diversity (Chaouche et al. 2024), which intervene during the entire AI life cycle. Therefore, high-quality data is essential for building a robust and trustworthy AI system (Awadid et al. 2024; Mattioli et al. 2024).

Data drift is a specific and critical data quality issue that arises over time when the statistical properties of the data a model receives in production or in operation diverge from the data it was trained on (Ackerman et al. 2020). The monitoring and the maintaining of the required performance are often based on drift detection (Klaise et al. 2020). This data drift can be expressed in different ways synthesized into four types: sudden, gradual, incremental, and reoccurring drifts (Lu et al. 2018). This drift means that the patterns learned by the model are no longer relevant, leading to the degradation in data quality and a significant decline of the model performance and reliability. Consequently, a drift typically occurs gradually over time and is influenced by both the model and the data (Mirza et al. 2024). Therefore, distinguishing whether the drift originated from data changes or from model/system degradation (Bayram, Ahmed, and Kassler 2022) remains a challenge.

Furthermore, in the best scenario, this drift would affect only slightly the performance of a model, and if it is identified sufficiently soon after its occurrence, it could be managed properly by the model. In such a scenario, the model adaptation time and performance drop may give indications concerning the drift severity. A visual illustration of data drift is shown in Figure 1, where this drift phenomenon may occur during the life cycle of an AI model.

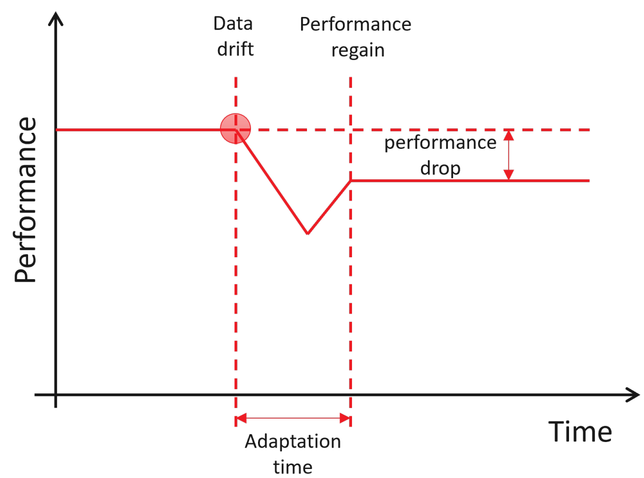


Figure 1: Data drift illustration and corresponding indicators quantifying the model adaptability / drift severity.

The data drift terminology may not be well established to unify concepts and definitions among researchers. Similar strategies have been developed independently under different names in different applications and contexts (Gama et al. 2014). Therefore, these works can be divided into two clusters, which are concept drift and covariate drift. Concept drift can be defined as the change of statistical properties of the target variable, conditional or joint distribution, which the model is predicting (Gama et al. 2014; Lu et al. 2018; Webb et al. 2018). Covariate drift describes the change in the distribution of the inputs between the learning and the operation phases (Nair et al. 2019). A third drift known also as

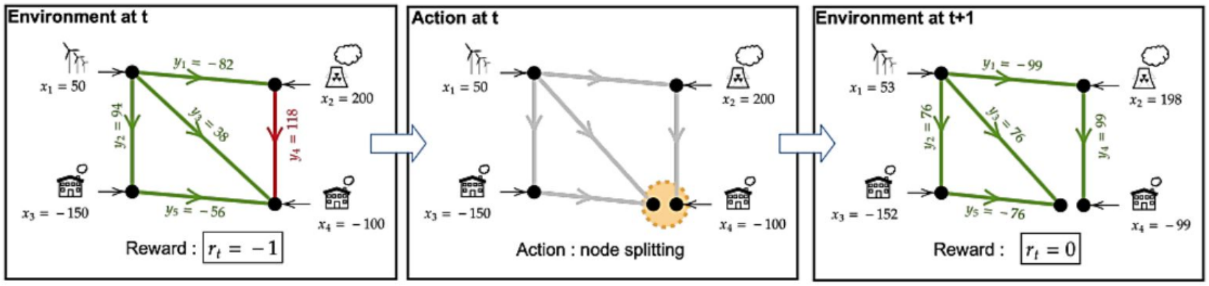


Figure 2: Toy example to demonstrate a power line overload and an remedial action (topology change).

label drift is also proposed in the literature, where the focus is more on the target variable (Wu et al. 2021). These three drifts are detailed and implemented in the current work. We draw attention that the word *drift* and the word *shift* are used interchangeably in the literature, where we noticed that usually *concept drift* and *label shift* are used for the output distributions change and *covariate shift* further for the input distribution change.

The detection of data drift remains a challenging topic for both industrial and academic communities. This challenge is a significant obstacle across a wide array of application domains, tailoring solutions for different contexts, such as medical (Kore et al. 2024), finance (Masegosa et al. 2020), and industry (Soller, Hölzl, and Kranz 2020), etc. This research aims to evaluate the role and quality of data in the task of drift detection in the industrial domain, specifically applied in electricity transmission network systems, by using physics-informed neural networks models (PINNs) (Huang et al. 2022). The study will explore methods to assess and attribute the source of drift, with a focus on identifying data-driven drift. Furthermore, it will investigate techniques to interpret and mitigate such drift, as effective mitigation offers a strategic advantage in maintaining model performance over time.

The remainder of the paper is organized as follows. We begin by reviewing related work. Next, we describe the context, including the scenarios and datasets used in our study. We then present the methodological approach employed to model and quantify concept drift, along with the corresponding results. Finally, we conclude with a summary of findings and propose directions for future research.

Related Work

Data drift in electricity transmission networks is a critical challenge driven by the de-carbonization and decentralization of the power grid. This drift is a direct result of the grid's evolving physical and operational characteristics. The wide integration of distributed energy resources (Johansson, Vendel, and Nuur 2020) and the ecological transition to renewable energy resources may cause a significant drift in the consumption prediction (Milano et al. 2018).

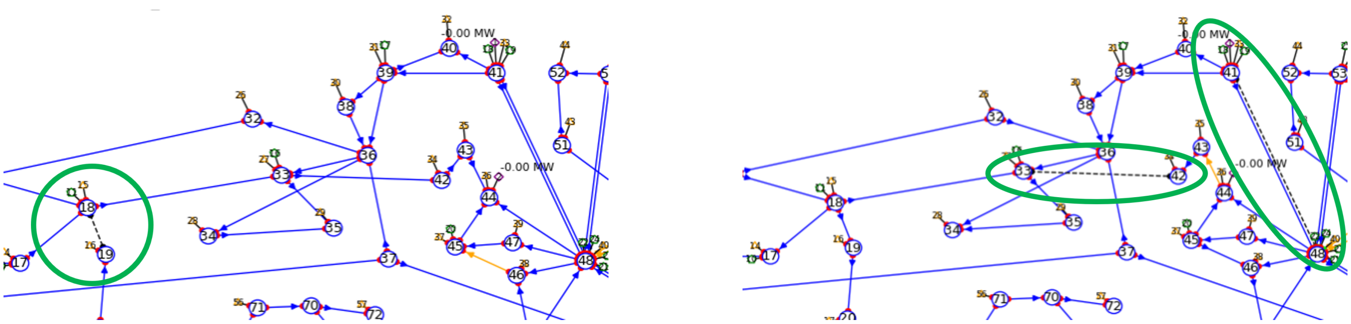
In the literature, the data drift for energy consumption, specifically for industrial machines, was explored by the scientific community. Indeed, Kahraman, Kantardzic, and Kotan proposed a framework to consider and integrate the repetitive regimes, which is based on dynamic modeling with memory and without memory to identify concept drifts applied on industrial machine energy consumption (Kahraman, Kantardzic, and Kotan 2022). Depending on the machine and the learning process, authors highlighted the link between the drift identification, with and without memory, and prediction performances of models. In the work of Zink, Ioshchikhes, and Weigold, five approaches of concept drift detection are evaluated for industrial machines' energy consumption. They explored these approaches on LSTM (Long-Short Time Memory) forecasting model, where the best result is obtained by Kolmogorov-Smirnov Windowing (KSWIN) method (Zink, Ioshchikhes, and Weigold 2024). In the same view, a mitigation drift approach, for LSTM forecasting model, was proposed for energy consumption forecasting in smart grids based on an AutoEncoder LSTM framework (Azeem et al. 2025).

Additionally, to industrial machines, the drift detection for electrical consumption forecasting in smart buildings (Mariano-Hernández et al. 2022), which is the largest consumer of electrical energy. Mariano-Hernández et al. integrated two drift detection methods, Adaptive Window (ADWIN) (Jagait et al. 2021) and KSWIN, based on the extension of (Togbe et al. 2021), in the forecasting model. Fenza, Gallo, and Loia proposed a methodology to adopt models to profile and forecast consumers by considering the concept drift (Fenza, Gallo, and Loia 2019).

The monitoring and the drift detection of energy consumption of AI-based, specifically PINNs models, are not well explored in the literature. The current work addresses this challenge in terms of detection approaches and the resilience of PINNs models, in the context of power grids, under the data drift phenomenon.

Context and Scenario Design

This section presents the scenario and associated datasets that have been designed for the sake of this study. As mentioned earlier, we focus on the analysis of data drift in the context of power grids which, as a safety-critical infrastructure, could continuously be affected by multiple factors, e.g., environmental changes, renewable energy integration, and energy transition. All these factors may implicitly or explicitly change the power flow dynamics and should be managed in time by grid operators to maintain stability and avoid problems such as cascading failures (causing a power



(a) Target dataset topology. Only one disconnected power line. The same distribution as training dataset.

(b) Under-Drift-Distribution (UDD) dataset distribution. Two simultaneous power line disconnections.

Figure 3: Data distribution (grid topology) used for test and Under-Drift-Distribution (UDD) datasets. In UDD dataset, a non-observed topology induces a drift in power flow distributions which should be managed by the models.

outage). In this context, the human operators are equipped with simulators allowing them to analyze the impact of their actions before their implementation in real-world scenarios. Thousands of such simulations are required in near real-time in order to make a correct decision.

The power flow simulators are mainly based on the resolution of physics equations which are solved using iterative algorithms such as Newton-Raphson. Although being very precise, their computation time may be mitigated by their real-time application. Hence, the machine learning (ML) algorithms exploiting the physics knowledge, using PINNs, are investigated as surrogates with the advantage of having fast inference time. However, these algorithms should also be robust to some changes to the data distribution. It is because of the large observation and action space size, where all the possible scenarios and grid configurations could not be included in the training dataset. In this work, we analyze the capability of these ML-based models to scale-up over some unseen data under-drift-distribution (UDD) scenarios which can be similar to a set of out-of-distribution (OOD) scenarios.

To design the required scenarios, we make the steady-state hypothesis where the observations are independent in each dataset and are differently distributed over training and under-drift-distribution datasets. There are two different types of changes that are included in the datasets: (1) changing the connectivity of power lines in the grid (connected or disconnected), (2) changing the topology of the grid by changing the bus-bars at each substation (node) in the power grid. The first type of action could be the result of some overloads or maintenance actions, and the second type concerns the action allowing the grid operators to avoid the congestion (see Figure 2). The data generation process uses a power flow simulator and integrates one or both types of actions. Herein, we describe the different datasets that are used:

- Source dataset: Comprises $100,000$ observations by integrating 4 to 5 topological changes as reference and could include at most one power line disconnection;
- Target dataset: Comprises $10,000$ observations by integrating some random topological changes and includes one power line disconnection per observation (see Figure 3a). It is different from training dataset that could include or not one power line disconnection;
- Under-Drift-Distribution (UDD) dataset: Comprises $10,000$ observations by integrating some reference topological changes. To consider data drift, this dataset includes two simultaneous power line disconnections per observation (see Figure 3b). This changes the power flow dynamics, and the generalization capacity and adaptability of the developed model could be quantified using this dataset.

Table 1 summarizes the notation and variables used throughout the article. These variables are used in the operational step to describe inputs and outputs of models in terms of electrical power and grid topology.

Variable	Description	Type
P_{prod}	Production active power (continuous)	Input
P_{load}	Load active power (continuous)	Input
L_s	Vector indicating line connectivity (categorical)	Input (topology)
τ	Vector indicating bus connectivity (categorical)	Input (topology)
P_{or}	Active power flow at line origin (continuous)	Target
P_{ex}	Active power flow at line extremity (continuous)	Target

Table 1: Notations and variables.

Methodology

The proposed approach to quantify and characterize the resilience of PINNs under data drift, which uses the data drift decomposition framework, is based on the following three principal steps:

- Investigate how changes in input distributions during model deployment impact the predictive accuracy of PINNs, compared to purely data-driven models.
- Decompose overall performance degradation into constituent drift components (covariate shift, label shift, and concept drift) to pinpoint the dominant sources of error and enabling more granular interpretation of UDD impact and performance.
- Evaluate how the incorporation of physical equations within the learning process modulates sensitivity to different drift types, thereby elucidating the synergy between physical priors and data-driven shifts.

The following section details the metric used to evaluate each drift type.

Drift Detection Background

We implement a three-way drift decomposition framework that separates total drift into covariate shift, label shift, and concept drift components. Based on the equation (1), we define three distinct types of drift.

$$P(X,Y) = P(Y|X) \times P(X), \tag{1}$$

where changes in this joint distribution can be attributed to one or more of these three distinct types of drift:

- *Covariate drift:* It appears when the shift is introduced in the input data $P(X)$ and the label given the data $P(Y|X)$ remains unchanged, and described by the following equations:

$$\begin{cases} P_{\text{source}}(X) \neq P_{\text{target}}(X) \quad \text{and} \\ P_{\text{source}}(Y|X) = P_{\text{target}}(Y|X) \end{cases}$$

- *Label drift:* It appears when the shift is introduced in $P(Y)$ and the input data given the label $P(X|Y)$ is remained unchanged. This label is described by following equations:

$$\begin{cases} P_{\text{source}}(Y) \neq P_{\text{target}}(Y) \quad \text{and} \\ P_{\text{source}}(X|Y) = P_{\text{target}}(X|Y) \end{cases}$$

- *Concept drift:* It appears when the shift is introduced in the label given the input $P(Y|X)$ and the input $P(X)$ remains unchanged. The following equations describe this concept drift:

$$\begin{cases} P_{\text{source}}(Y|X) \neq P_{\text{target}}(Y|X) \quad \text{and} \\ P_{\text{source}}(X) = P_{\text{target}}(X) \end{cases}$$

The evaluation of these three drifts is performed by computing distribution distances between datasets source and target. For each drift, the details of computation are given below, and the implementation is based on the Python Alibi-Detect library (Van Looveren et al. 2019).

Covariate Shift Computation. For detecting changes in input distributions $P(X)$, we employ specialized statistical tests adapted to different feature types:

- *Maximum Mean Discrepancy (MMD)*: MMD measures the distance between distributions, for Continuous Features, by embedding them in a reproducing kernel Hilbert space:

$$\begin{aligned} \text{MMD}^2(P,Q) \quad &= \quad \mathbb{E}_{x,x'\sim P}[k(x,x')] + \mathbb{E}_{y,y'\sim Q}[k(y,y')] \\ &\quad -2\mathbb{E}_{x\sim P,y\sim Q}[k(x,y)], \end{aligned}$$

  where $k(\cdot,\cdot)$ is a characteristic kernel (we use RBF kernel).

- *Chi-Squared Test for Categorical Features:* For binary line status variables, we use the chi-squared test for independence:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

  where O_{ij} are observed frequencies and E_{ij} are expected frequencies under the null hypothesis of identical distributions. For multivariate distribution, the Nikulin-Rao-Robson test can be used.

- *Covariate Shift Score Calculation:*
  The computation covariate shift score combines the individual MMD and Chi-squared distances through a weighted averaging scheme as given by the equation (2):

$$\text{Covariate Score} = \frac{1}{3}\left(d_{\text{cont}} + \bar{d}_{\text{cate}} + d_{\text{topo}}\right), \tag{2}$$

  where d_{cont} is MMD distance for continuous features (P_{prod}, P_{load}), $\bar{d}_{\text{cate}} = \frac{1}{N}\sum_{i=1}^{N}\chi_i^2$ is Mean Chi-squared distance across N line status features L_s, and d_{topo} is MMD distance for topology vector features τ. This approach ensures that each feature type contributes equally to the final covariate shift assessment, regardless of the different scales of MMD and Chi-squared statistics.

Label Shift Computation. We employ the Kolmogorov-Smirnov test, given in equation (3), to detect changes between source and target distributions:

$$D = \sup_z |F_{\text{source}}(z) - F_{\text{target}}(z)|, \tag{3}$$

where $F(\cdot)$ represents the empirical cumulative distribution function.

Concept Shift Computation. Concept shift is detected through residual analysis using a linear regression baseline, given by equation (4), between target and source datasets:

$$\text{Residual Score} = \mathbb{E}[|Y_{\text{target}} - f_{\text{source}}(X_{\text{target}})|], \tag{4}$$

where f_{source} is a model fitted on training data.

Theoretical Framework Connection. The data drift decomposition $P_{new}(X,Y) = P_{new}(Y|X) \times P_{new}(X)$ reveals the following correspondences and interpretations:

- *Covariate Shift:*

$$P_{new}(X) \neq P_{source}(X) \quad \text{(topology changes);}$$

- *Label Stability:*

$$P_{new}(Y|X) \approx P_{source}(Y|X) \quad \text{(physics preserved)},$$

where new describes $target$ and UDD datasets.

By applying data drift decomposition to controlled shifts in input topology and physical parameters, we can systematically assess the impact of the addition of physics-based loss terms, and how the explicit encoding of grid topology and governing constraints, in the network architecture, enhance stability. Thus, pinpointing the exact distributional regimes to which PINNs remain robust and those in which they degrade.

Experiments and Results

Experiments

This section presents the experimental settings and corresponding results. For the sake of the experimentation, two different grid sizes are considered to evaluate the scale-up capability of models. The first one includes 14 nodes and 20 power lines (small grid) and the second one includes 36 nodes and 58 power lines with more complex power flow dynamics (large grid).

Two different categories of models have been employed. The first category concerns the pure ML-based models such as a simple multi-layer perceptron (MLP) neural network. The second category includes PINNs using two different strategies: using the physical constraint (local conservation law) as a regularization term (MLP-regularized), and using the physical constraint as the optimization problem during the learning of a graph neural network (GNN Init) model. The MLP-based models take as input a vector combining the injections (P_{prod}, P_{load}) and categorical topological variables (L_s and τ). In contrast, the GNN-based models use the injection as node features and construct the graph based on the topological information. For a more detailed description of these models, the reader may refer to (Leyli-abadi, Marot, and Picault 2025).

Obtained Results

We evaluated and compared performance using a comprehensive set of evaluation criteria based on the LIPS framework (Leyli-abadi et al. 2022) and the results are shown in Table 2. Besides the classic ML-related criteria to quantify the model's precision, we considered the adaptability and generalization capacity of the model toward data drift using the designed UDD dataset, compliance with physics constraints, and also the industrial readiness through two grid sizes.

As can be seen in the Table 2, the MLP-based model without physics consideration obtains fairly better performance on the target dataset and ML-related criteria than UDD dataset. However, its performance remains insufficient regarding the physics compliance criteria. Adding a regularization term based on physical constraint (local conservation law) allows for slightly improving its performance, but the physics law violation percentage still remains too high. It can also be easily observed that the model performance degrades when they are applied on the UDD dataset, which introduces the drift in the data distribution by considering new grid topologies.

Using the GNN-based approach which optimizes the physical constraint directly, we obtained more accurate performances in terms of ML-related and UDD generalization criteria. It also allowed us to improve significantly the physics compliance in comparison to MLP-based approaches. This performance gain could be explained by the fact that GNN is invariant to the power grid topology changes and the physics equations are solved at each message-passing layer. It also delivers more consistent performance when the grid is scaled to larger sizes. Finally, the GNN-based approach entails higher computational costs compared to MLP-based methods.

		Evaluation Criteria Categories			
		ML-related	**UDD Gen.**	**Physics Comp.**	
	Methods	MAE	MAE	Target	UDD
Small grid	MLP	1e-2±1e-3	4e-1±1e-2	40%±1%	47%±.5%
	MLP Reg	6e-2±2e-2	6e-1±4e-2	35%±1%	42%±1%
	GNN Init	**2e-3±2e-5**	**2e-2±1e-4**	**1%±.1%**	**13%±.7%**
Large grid	MLP	1e-1±2e-3	2e-1±2e-3	50%± 1%	50%±1%
	MLP Reg	7e-1±2e-1	9e-1±2e-2	45%±1%	47%±1%
	GNN Init	**4e-2±3e-5**	**6e-2±1e-3**	**20%±.5%**	**26%±.5%**

Table 2: Benchmark table. The results are presented using mean ± standard deviation (over 5 runs). The compared methods are Multi-layer perceptrons (MLP), Regularized MLP with physics loss (MLP Reg), Graph neural network with constraint as optimization (GNN Init). The considered criteria are Mean Absolute Error (MAE), Under-Drift-Distribution Generalization (UDD Gen.), and physics compliance category (Physics comp.) computing local conservation error on both target and UDD datasets.

Datasets	Drift type	Distance result
Source → Target	Covariate Shift	2.78
	Label Shift	0.07
	Concept Shift	0.73
Source → UDD	Covariate Shift	32.96
	Label Shift	0.21
	Concept Shift	1.27

Table 3: Drift analysis results. The different shifts are computed between two datasets. Source dataset is used for training of the models, Target dataset present the same distribution as source, UDD dataset includes a distributional drift from Source and Target.

Comparative Analysis. The UDD scenarios exhibit 9.6 times stronger drift than the test scenarios, as can be seen

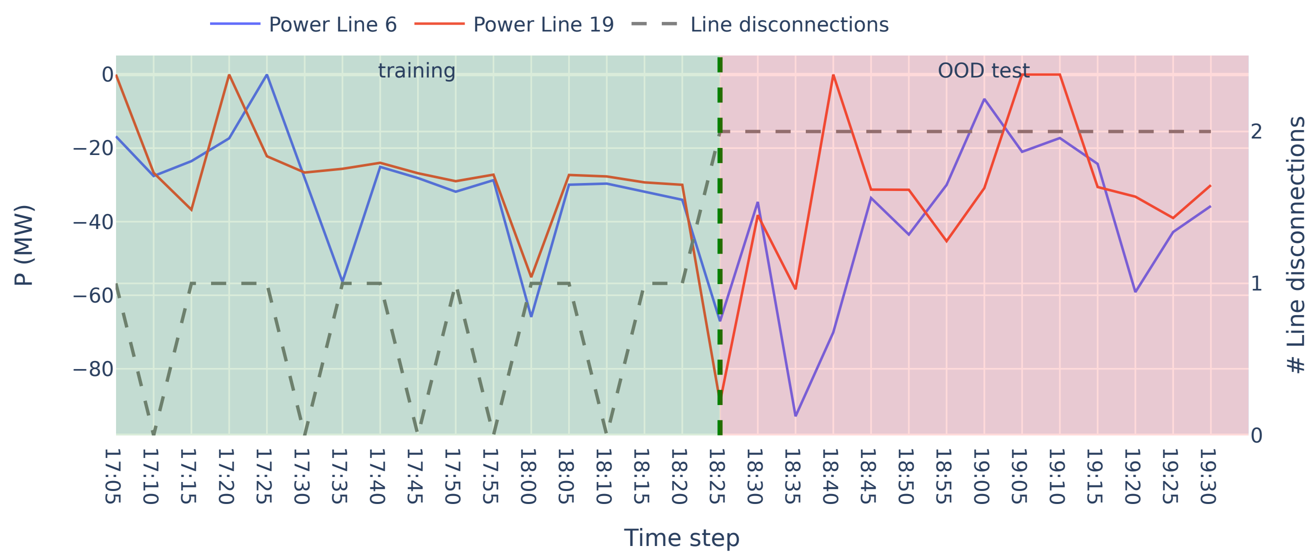


Figure 4: Data distribution before and after the data drift (vertical green dashed line). The left y-axis represents the power flow for two power lines 6 and 19 shown in blue and red, respectively. The right y-axis represents the number of the disconnections in the grid and corresponds to the dashed gray curve representing the line disconnections. It can be seen that before the data drift we have at most one disconnection, but after the drift, we have two simultaneous power line disconnections.

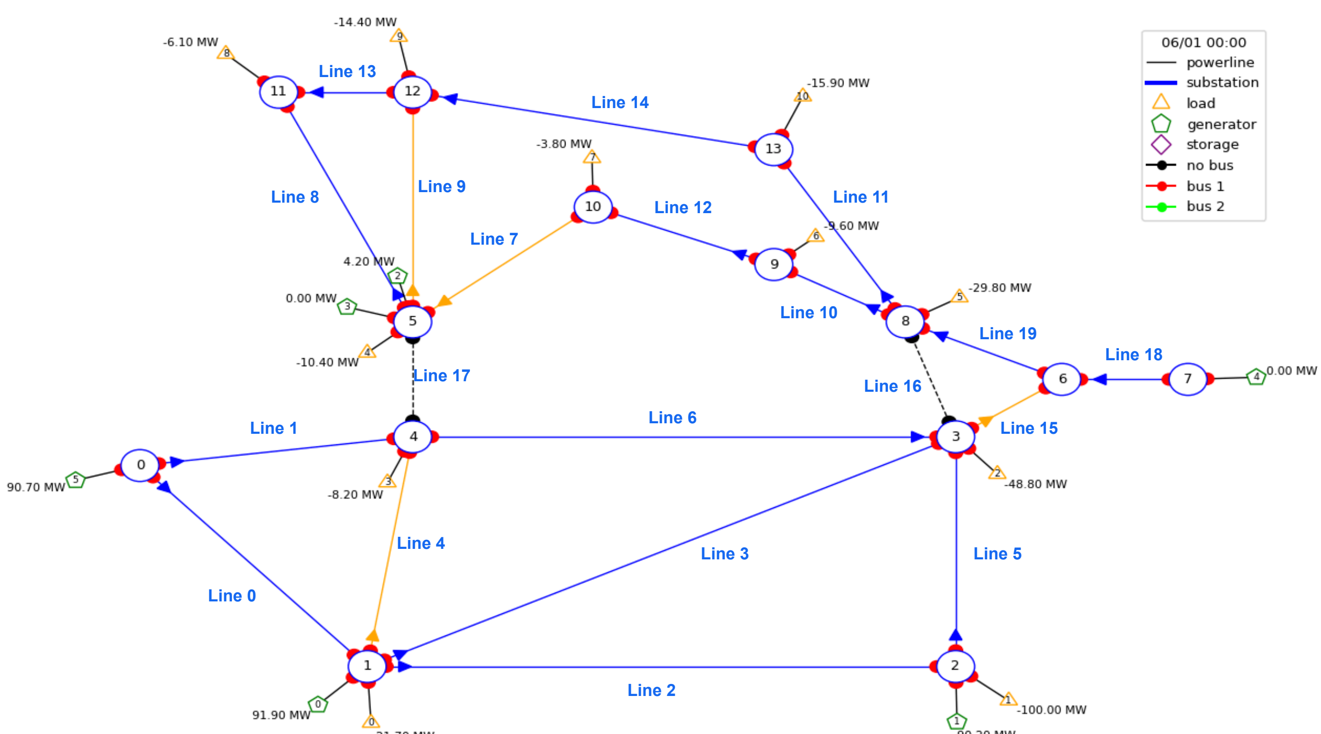


Figure 5: The Power grid topology at the data drift occurrence corresponding to timestamp 18:25 which corresponds to the vertical green dashed line shown in Figure 4. Two simultaneous power line disconnections (lines 16 and 17) impacts the flow in adjacent lines.

from the increase of the total drifts (sum of all drifts) from source-to-target to source-to-UDD detailed in Table 3. By comparing each drift type, the covariate shift showed the highest drift with a distance increase of 2.78 to 32.96, which represents 11.86 times of increase. While label and concept drifts showed a distance increase of 0.07 to 0.21 and 0.73 to 1.27, which represent only 3 times and 1.74 times of increase, respectively. This significant change and important increase of covariate shift are consistent with the change of the input data only.

Another reading of the obtained result, in the table, is the contribution of each drift type to the overall drift. It can be seen that for the source-to-target drift, the contributions of drift types are 77.6%, 2.0% and 20.4% for covariate, label, and concept drifts, respectively. Whereas, for source-to-UDD drift, the contributions of drift types are more concentrated on the covariate shift with a contribution of 95.7%, and only 0.6% and 3.7% for label and concept drifts, respectively. These results show the impact of UDD on the covariate drift, which absorbs a high percentage and corresponds to the theoretical formulation given in the previous section, due to significant changes in input data.

Physical Interpretation. Figure 4 exhibits a surprisingly intuitive visual representation of the qualitative distinction between covariate and concept shifts. In this figure, we have represented the power flows for two power lines (6 and 19) before (green region) and after (red region) the occurrence of data drift (vertical dashed green line). The number of line disconnections is also demonstrated using a gray dashed line with respect to the right y-axis. It can be seen that under the UDD regime, we have two simultaneous power line disconnections which have never been observed before. To better contextualize the problem, we have also illustrated, in Figure 5, a snapshot of the power grid for the timestamp 18:25 corresponding to the drift occurrence. We can observe, from the figure, that the power lines 16 and 17 are disconnected which may have some non-linear impact on the dynamics of adjacent power lines.

Let us put it in context: usually, formally defining such kind of shifts in terms of probability measure properties such as those presented in the methodology section of this article seems straightforward, but the question of concretely illustrating them over a single modality remains an illusive subject. Covariate shifts are predominant in computer vision - in the form of common corruptions (Hendrycks and Dietterich 2019) - but it remains unclear on how to illustrate concept shifts on them. On the other hand, concept shifts are typically ubiquitous in tabular data (Liu et al. 2023), as simply adding or deleting feature columns is ipso-facto manually producing a concept drift. In Figure 4, one can interpret the values of the power flow of lines 6 and 19, before line disconnection, as an instantiation of a covariate shift: the values are distinct while remaining somewhat correlated. But after the line has been disconnected, the ratio in the distribution shift corresponding to concept drift passes from small to being predominant. Qualitatively, what happens is that whatever correlation between power flow of lines 6 and 19 disappears. In other words, our use-case not only allows a clear visual representation of these kinds of shifts on the same data modality but even allows for qualitative distinctions to be analyzed.

Conclusions and Perspectives

The data drift evaluation is one of the trustworthiness keys of AI systems and an important operational need in the industry. Thus, in this paper, the drift detection is addressed in the context of energy consumption of electrical power, applied on data-driven and physics-informed neural network (PINNs) models. Two results are highlighted in this work, which are the resilience of PINNs to data drift compared to data-driven models, and the impact of data drift types, covariate, label, and concept drift, under changes.

Indeed, a contextual and an interpretable data drift distribution is simulated based on grid topology disconnection applied to a set of data-driven and hybrid (PINNs) models. The obtained result showed that the data drift is detected for each model, and an important resilience of PINNs is highlighted compared to data-driven models, despite the violation of physical laws and physical constraints. The obtained results also showed the contribution of each drift type and its evolution from target to under-drift-distribution datasets, which is beneficial for the drift understanding, physical interpretation, and the trustworthiness of AI systems during the industrialization phase. By monitoring data drift through these metrics, operators can anticipate when predictive models may no longer reflect real grid behavior, enabling timely adjustments before reliability is compromised.

The presented methodology to detect the data drift and corresponding metrics could be easily adapted to other domains processing structured data. For the future work, more rich approaches, such as PCA (principal component analysis) and VAE (variational auto-encoder), could be used to capture nonlinear and high dimensional drifts.

Acknowledgments

This work has been supported by the French Government under the "France 2030" program, as part of the SystemX Technological Research Institute.

References

Ackerman, S.; Farchi, E.; Raz, O.; Zalmanovici, M.; and Dube, P. 2020. Detection of data drift and outliers affecting machine learning model performance over time. *arXiv preprint arXiv:2012.09258*.

Awadid, A.; Amokrane-Ferka, K.; Sohier, H.; Mattioli, J.; Adjed, F.; Gonzalez, M.; and Khalfaoui, S. 2024. AI systems trustworthiness assessment: State of the art. In *Workshop on Model-based System Engineering and AI, 12th International Conference on Model-Based Software and Systems Engineering (Modelsward)*.

Azeem, A.; Ismail, I.; Mohani, S. S.; Danyaro, K. U.; Hussain, U.; Shabbir, S.; and Jusoh, R. Z. B. 2025. Mitigating concept drift challenges in evolving smart grids: An adaptive ensemble LSTM for enhanced load forecasting. *Energy Reports*, 13: 1369–1383.

Bayram, F.; Ahmed, B. S.; and Kassler, A. 2022. From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowledge-Based Systems*, 245: 108632.

Chaouche, S.; Randon, Y.; Adjed, F.; Boudjani, N.; and Khedher, M. I. 2024. DQM: Data Quality Metrics for AI components in the industry. In *Proceedings of the AAAI Symposium Series*, volume 4, 24–31.

Ehrlinger, L.; and Wöß, W. 2022. A survey of data quality measurement and monitoring tools. *Frontiers in big data*, 5: 850611.

Fenza, G.; Gallo, M.; and Loia, V. 2019. Drift-aware methodology for anomaly detection in smart grid. *IEEE Access*, 7: 9645–9657.

Gama, J.; Žliobaitė, I.; Bifet, A.; Pechenizkiy, M.; and Bouchachia, A. 2014. A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46(4): 1–37.

Hendrycks, D.; and Dietterich, T. 2019. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. *Proceedings of the International Conference on Learning Representations*.

Huang, G.; Zhou, Z.; Wu, F.; and Hua, W. 2022. Physics-informed time-aware neural networks for industrial nonintrusive load monitoring. *IEEE Transactions on Industrial Informatics*, 19(6): 7312–7322.

Jagait, R. K.; Fekri, M. N.; Grolinger, K.; and Mir, S. 2021. Load forecasting under concept drift: Online ensemble learning with recurrent neural network and ARIMA. *IEEE Access*, 9: 98992–99008.

Johansson, P.; Vendel, M.; and Nuur, C. 2020. Integrating distributed energy resources in electricity distribution systems: An explorative study of challenges facing DSOs in Sweden. *Utilities Policy*, 67: 101117.

Kahraman, A.; Kantardzic, M.; and Kotan, M. 2022. Dynamic modeling with integrated concept drift detection for predicting real-time energy consumption of industrial machines. *IEEE Access*, 10: 104622–104635.

Klaise, J.; Van Looveren, A.; Cox, C.; Vacanti, G.; and Coca, A. 2020. Monitoring and explainability of models in production. In *ICML*, Workshop on Challenges in Deploying and Monitoring Machine Learning Systems.

Kore, A.; Abbasi Bavil, E.; Subasri, V.; Abdalla, M.; Fine, B.; Dolatabadi, E.; and Abdalla, M. 2024. Empirical data drift detection experiments on real-world medical imaging data. *Nature communications*, 15(1): 1887.

Leyli-abadi, M.; Marot, A.; and Picault, J. 2025. Study Design and Demystification of Physics Informed Neural Networks for Power Flow Simulation. *ECML PKDD Machine Learning for Sustainable Power Systems*.

Leyli-abadi, M.; Marot, A.; Picault, J.; Danan, D.; Yagoubi, M.; Donnot, B.; Attoui, S.; Dimitrov, P.; Farjallah, A.; and Etienam, C. 2022. LIPS-learning industrial physical simulation benchmark suite. *Advances in Neural Information Processing Systems*, 35: 28095–28109.

Liu, J.; Wang, T.; Cui, P.; and Namkoong, H. 2023. On the Need for a Language Describing Distribution Shifts: Illustrations on Tabular Datasets. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Lu, J.; Liu, A.; Dong, F.; Gu, F.; Gama, J.; and Zhang, G. 2018. Learning under concept drift: A review. *IEEE transactions on knowledge and data engineering*, 31(12): 2346–2363.

Mariano-Hernández, D.; Hernández-Callejo, L.; Solís, M.; Zorita-Lamadrid, A.; Duque-Pérez, O.; Gonzalez-Morales, L.; García, F. S.; Jaramillo-Duque, A.; Ospino-Castro, A.; Alonso-Gómez, V.; et al. 2022. Analysis of the integration of drift detection methods in learning algorithms for electrical consumption forecasting in smart buildings. *Sustainability*, 14(10): 5857.

Masegosa, A. R.; Martínez, A. M.; Ramos-López, D.; Langseth, H.; Nielsen, T. D.; and Salmerón, A. 2020. Analyzing concept drift: A case study in the financial sector. *Intelligent Data Analysis*, 24(3): 665–688.

Mattioli, J.; Gonzalez, M.; Mattioli, L.; Quintero, K.; and Sohier, H. 2024. Leveraging tropical algebra to assess trustworthy ai. In *Proceedings of the AAAI Symposium Series*, volume 4, 81–88.

Milano, F.; Dörfler, F.; Hug, G.; Hill, D. J.; and Verbič, G. 2018. Foundations and challenges of low-inertia systems. In *2018 power systems computation conference (PSCC)*, 1–25. IEEE.

Mirza, S.; Nguyen, V. D.; Mantini, P.; and Shah, S. K. 2024. Data quality aware approaches for addressing model drift of semantic segmentation models. *arXiv preprint arXiv:2402.07258*.

Nair, N. G.; Satpathy, P.; Christopher, J.; et al. 2019. Covariate shift: A review and analysis on classifiers. In *2019 global conference for advancement in technology (GCAT)*, 1–6. IEEE.

Soller, S.; Hölzl, G.; and Kranz, M. 2020. Predicting machine errors based on adaptive sensor data drifts in a real world industrial setup. In *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 1–9. IEEE.

Togbe, M. U.; Chabchoub, Y.; Boly, A.; Barry, M.; Chiky, R.; and Bahri, M. 2021. Anomalies detection using isolation in concept-drifting data streams. *Computers*, 10(1): 13.

Van Looveren, A.; Klaise, J.; Vacanti, G.; Cobb, O.; Scillitoe, A.; Samoilescu, R.; and Athorne, A. 2019. Alibi Detect: Algorithms for outlier, adversarial and drift detection.

Webb, G. I.; Lee, L. K.; Goethals, B.; and Petitjean, F. 2018. Analyzing concept drift and shift from sample data. *Data Mining and Knowledge Discovery*, 32(5): 1179–1199.

Wu, R.; Guo, C.; Su, Y.; and Weinberger, K. Q. 2021. Online adaptation to label distribution shift. *Advances in Neural Information Processing Systems*, 34: 11340–11351.

Zink, R.; Ioshchikhes, B.; and Weigold, M. 2024. Concept drift monitoring for industrial load forecasting with artificial neural networks. *Procedia CIRP*, 130: 120–125.