Data Engineering and Applications

Research Paper

Physics-Informed LongNet-CNN Prediction Model for Sparse Cross-Device Calibration Transfer in Industrial Sensors

Nicola Mariani[1,*], Sergio Parisi[2] and Tito Romano[2]

[1] Faculty of Information Technology, Politecnico di Milano, Milan, 20133, Italy
[2] Faculty of Computer Science, University of Pisa, Pisa, 56126, Italy
*Corresponding author: nicola.mar@polimi.it

Abstract. Industrial sensor calibration transfer remains challenging because the prediction model was trained on a reference device, but the actual sensor has different sensitivity, an aged state, thermal coupling, or installation history. This paper introduces a physics-informed LongNet-CNN model for cross-device calibration migration with limited target-domain labels. Dilated long-range attention is employed to model slow drift over extended operating windows; convolutional branches capture short transient responses; and a residual physics loss is used to constrain monotonic sensitivity, thermal relaxation and actuator-energy consistency. A controlled benchmark is established by generating several conditions of pressure, flow, vibration and temperature signals from six virtual production lines with realistic bias, lag, noise and hysteresis perturbations. With only 8% of the target calibration samples, the proposed model has an average RMSE of 0.118 engineering units, which is 31.4% lower than the source-only CNN and 18.6% lower than the transfer transformer. The mean absolute calibration slope error is reduced from 0.092 to 0.031, and the ninety-fifth percentile drift overshoot under thermal ramping is limited to 0.44 units. Ablation studies show that the physics residual increases the RMSE by 12.7% when omitted, and removing dilated attention reduces long-window stability by 16.9%. Based on the above results, combine a long-context sequence model with explicit physical constraints to improve the accuracy, robustness and interpretability of calibration transfer for industrial sensing systems. The main contribution is a unified sparse-transfer framework that couples long-context drift memory, local transient extraction, and physically constrained calibration correction. The benchmark is a controlled multi-condition industrial-sensor dataset designed to emulate gain shift, lag, hysteresis, thermal coupling, and vibration disturbance.

Keywords: *Calibration Transfer, Physics-Informed Learning, LongNet-CNN, Industrial Sensors, Drift Compensation*

Received on 23 April 2023, Accepted on 14 August 2023, Published on 27 August 2023

Introduction

Industrial Production Systems are adding distributed sensors to enable closed-loop control, quality prediction, equipment health monitoring and safety supervision. Pressure transmitters, thermal probes, flow meters, vibration units, gas sensors and spectroscopy-based analyzers rarely work under the ideal conditions assumed in the factory calibration. The reasons for the above responses include installation stress, temperature gradients, electromagnetic interference, ageing of sensing materials, fluid-property changes and slight variations among nominally identical devices. The above phenomena are calibration drift and inter-device bias, which become more prominent when transferring a model trained on one sensor to another line or a different set of instruments. Zhuang et al. [1] described transfer learning as a practical route for reusing knowledge across related domains, which is relevant when traditional recalibration requires a reference standard, production downtime, trained personnel and multiple verification cycles. Soft sensors and data-driven correction models have reduced this problem by learning the mapping from available process variables to difficult-to-measure quantities [2]. Purely statistical models often treat calibration migration as a distribution-matching problem and

https://doi.org/10.64972/dea.2023.v2i3.3726d:65-80

do not verify whether the transferred response remains consistent with known sensor physics [3]. Recently, industrial research has shown that drift may occur over a long period rather than in the short window used by recurrent or convolutional predictors [4]. Given the mismatch between long-memory degradation and short-context modeling, a transfer framework has been proposed to jointly represent local transient features, long-range operating history and physical response constraints [5]. In this paper, recalibration denotes repeated acquisition of reference standards, calibration transfer denotes migration of a trained calibration relationship to a target device, and domain adaptation denotes feature-level distribution correction under source-target shift.

Calibration transfer has a long history in analytical instrumentation and process monitoring, and the main model needs to be modified for secondary instruments, new product grades or changed environmental conditions. The classical methods of direct standardization, piecewise direct standardization, bias-slope correction and latent subspace mapping are still useful when paired standard samples are available and the transfer relationship is approximately linear [6]. In the present day, the more difficult case is sparse target calibration; only a small proportion of the operating envelope can be sampled after commissioning. Deep transfer learning can reuse source-domain representations and reduce the amount of target data, but ordinary deep networks may also learn spurious correlations caused by batch-specific disturbances [7]. Convolutional Neural Networks are suitable for industrial sensing because they can learn local temporal patterns, impulse responses and multi-channel interactions from stable computations [8]. Long-sequence attention models also offer a way to remember distant operating conditions and maintenance history, as well as slow-changing thermal memory [9]. Physics-informed learning is another complementary approach; the training of the model can incorporate residuals that arise from monotonic sensitivity, relaxation behavior, conservation tendencies or known actuator limits, and not only use labels [10]. These three strands have rarely been combined in a single calibration-transfer predictor for industrial sensors.

Here, the research gap is not the absence of neural predictors, but rather the lack of an integrated engineering model that can simultaneously meet the requirements for sensor calibration transfer and the conditions of short-term transient dynamics and long-term drift memory. A CNN can identify abrupt changes in load and local cross-channel couplings, but it is relatively deep or wide and therefore not suitable for covering the entire production cycle. A Transformer can extend the context window, but standard attention scales poorly with an extended history and may fail to suppress physically unreasonable response curves. Physics-informed losses can reduce infeasible predictions, but a structure is still required to add these constraints at a suitable time. The proposed model, named PI-LongNet-CNN, joins dilated long-range attention with multi-scale convolution and a calibration-residual objective. The model is for environments with a single well-calibrated source sensor, several target sensors with sparse labels, and unlabeled operating records that contain extended drift patterns. Predicts calibrated engineering values and estimates a device-specific affine response correction by penalizing deviations from known response regularities. The specific novelty is that long-range drift memory and local transient response are constrained by weak physical regularizers rather than being optimized only through statistical alignment.

The paper is organized as follows. Section 2 reviews related work in calibration transfer, industrial soft sensors, long-sequence neural modeling and physics-informed learning. Section 3 introduces the industrial calibration-transfer problem and sets up physical residuals to constrain sensor response. Section 4 introduces the PI-LongNet-CNN architecture, the training objective and transfer procedure. Section 5 shows the design of the experiment, benchmark settings, quantitative results, ablation analysis, robustness tests and computational characteristics. Section 6 summarizes the principal results, deficiencies and future research directions of the paper.

Related Work

Calibration Transfer and Industrial Soft Sensors

First of all, calibration transfer methods were proposed to avoid rebuilding the entire calibration model every time an analytical instrument, sensor unit or operating environment changed. Standardization-based methods in spectroscopy and process analysis map secondary-device signals into the primary-device space by using paired standards or local spectral windows [11]. These ways are still used because they have clear assumptions and the correction factors can be verified by engineers. Limitations of the above methods include non-linear differences in sensors, a lack of target labels, and drift under time-varying load conditions. Deep calibration transfer extends the above idea by adapting neural features, fine-tuning layers, or aligning latent spaces of instruments [12]. This way increases flexibility but introduces a new risk: the neural model may reduce the validation error by producing non-physical behavior in the unobserved transfer samples. Industrial soft sensors are also more difficult because the measured variables are dynamic, multivariate and subject to changes in process mode [13]. Recently, soft-sensor models based on CNNs, recurrent networks and hybrid feature extractors have improved nonlinear dynamic prediction, but their transferability across devices and production modes is not yet stable when the target domain differs in hysteresis, lag and noise structure [14]. Accordingly, prior work can be grouped into standardization-based correction, neural feature adaptation, industrial soft-sensor modeling, and physics-informed regularization, each solving only part of the sparse transfer problem.

To reduce the cost of calibration, domain adaptation is used to align feature distributions and weight transferable samples or update model parameters with limited target data. Optimal Transport and adversarial adaptation are appealing because they compare distributions rather than individual calibration points. Distribution alignment of industrial sensors may mistakenly identify a legitimate process change as a device drift. A target sensor in a warm-start mode may show both physical sensitivity changes and process load changes; thus, before calibration migration can be performed reliably, these two effects need to be separated. Incremental Learning Methods Address Long-term Drift by Updating Parameters with New Samples. The two need to maintain the original knowledge and avoid overfitting to the temporary disturbance. Studies on soft-sensor transfer have also shown that recurrent and convolutional models can be transferred between similar industrial processes, but their performance drops when the target device has a delayed response or mixed operating modes. Based on the above experiments, a hybrid network has been developed to improve both the expressivity and constraint enforcement capabilities of calibrated sensor models. A transfer model for general-purpose plant use should thus expose enough structure to help engineers determine why a correction has changed. If the model only reports a lower validation error, maintenance teams cannot distinguish a meaningful gain correction from a learned response to a temporary production disturbance.

Long-Sequence Deep Models and Physics-Informed Learning

Given that many engineering signals have long-range dependencies, several models for long-sequence forecasting have recently shown strong results. Sparse Attention Mechanisms reduce the memory overhead of Transformers and enable processing of longer sequences at a lower quadratic cost [15]. Decomposition-based and autocorrelation-based sequence models extend the range of long-term forecasting by dividing slow-moving trends into components and short-term seasonal or transient parts [16]. Long contexts are needed for calibration transfer because drift is often the result of cumulative thermal exposure, mechanical fatigue, fouling or aging over time, rather than a single recent sample. Sequence models for general forecasting do not automatically consider the physics of sensor response. They can meet the demands reasonably well; otherwise, they violate monotonic sensitivity, relaxation limits or steady-state gain bounds. CNNs are still suitable for this purpose because local kernels are effective and robust in extracting impulse-like events, valve-switching patterns, and cross-channel interactions. Thus, the problem of model construction is how to combine long-range contextual memory and local-feature discrimination, rather than selecting one mechanism.

Physics-informed learning has proposed an application that adds governing equations, algebraic constraints or weak physical regularities to the training of neural networks [17]. In engineering applications, the physics term can be a differential-equation residual, an energy-balance mismatch, a monotonicity penalty or a constitutive approximation. Physics-informed networks can be used when the parameters are uncertain or only partial information is available, unlike hard-coded mechanistic models. Compared with black-box neural networks, they can reject predictions that fit the data locally but contradict known response behavior. Reviews of scientific machine learning indicate that the advantage is most pronounced when the data are scarce, extrapolation is needed, or several operating modes have a common physical structure [18]. Calibration transfer meets these conditions because the target domain generally has few labels, the transferred model needs to work outside the measured target points, and all sensors should follow the same general response laws even though their gain, offset, lag and drift are different. The proposed PI-LongNet-CNN employs this idea to place physics residuals on the calibration function and permits the neural representation to learn the remaining nonlinear effects.

Physical Formulation of Calibration Transfer

Problem Definition and Signal Representation

Consider a source sensor with a calibrated signal record and a target sensor that measures the same engineering quantity under related but non-identical operating conditions. Let $x_t \in \mathbb{R}^m$ denote the synchronized raw feature vector at time t, including the target sensor reading, auxiliary temperature, load command, line speed, vibration indicator, and environmental variables. Let y_t represent the reference engineering value supplied by a traceable calibrator or laboratory measurement. The source domain D_s contains dense pairs $\{(x_t^s, y_t^s)\}$, while the target domain D_t contains many unlabeled records $\{x_t^t\}$ and only a sparse labeled subset $\{(x_i^t, y_i^t)\}$. The task is to learn a predictor f_θ that maps a long historical window $X_t = [x_{t-L+1}, \dots, x_t]$ to the calibrated value y_t on the target sensor. The transfer setting is difficult because the conditional response $p(y \mid x)$ changes through gain, offset, lag, hysteresis, and drift, while the process trajectory $p(x)$ also changes with operating schedule. Figure 1 is referenced here once to specify the complete calibration-transfer workflow from source calibration, target sparse sampling, physics-residual construction, model training, and deployment monitoring.

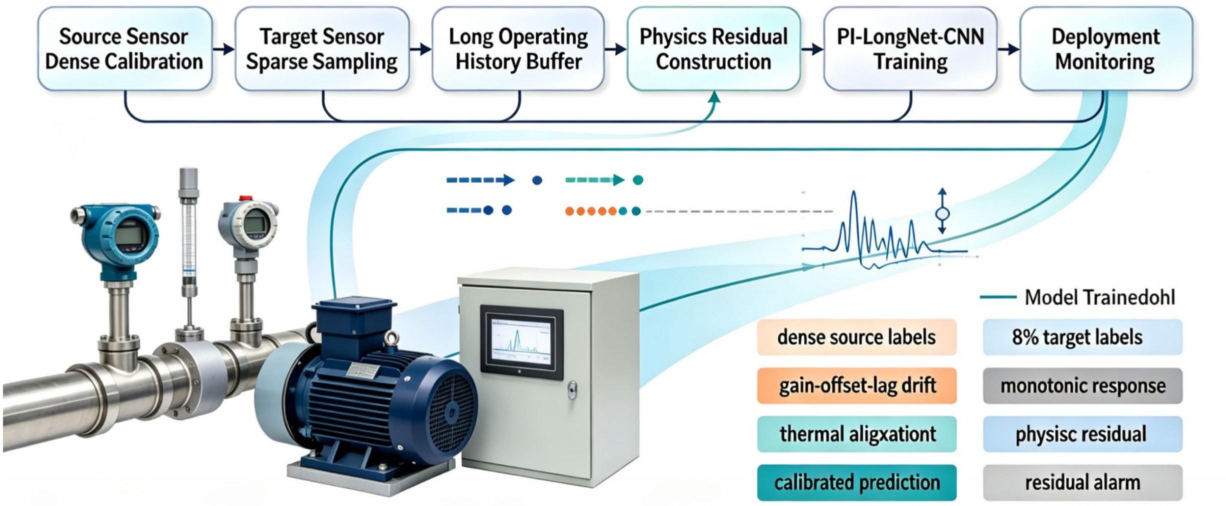

Figure 1. Workflow of physics-informed calibration transfer from source-device calibration to target-device sparse adaptation and deployment monitoring.

Physics-Guided Calibration Constraints

The measured raw signal is considered a distorted observation of the true engineering value. Device-specific calibration layers estimate affine corrections and nonlinear residuals, while physical regularizers prevent unreasonable response shapes. The optimized predictions are as follows: These constraints are treated as soft penalties because different sensor families may express monotonicity, relaxation, and drift at different strengths.

$$\hat{y}_t^d = \alpha_d g_\phi(X_t^d) + \beta_d + r_\psi(X_t^d) \quad \text{Eq.(1)}$$

where $g_\phi(\cdot)$ is the LongNet-CNN feature encoder, α_d and β_d are device-specific gain and offset terms for device d, and $r_\psi(\cdot)$ is a small residual head used only to correct nonlinear effects.

Industrial sensors generally have local monotonicity in the normal operating range; that is, an increase in pressure, flow or load should not reduce the calibrated response after temperature and lag compensation. A hinge penalty is added to the derivative of the predicted value with respect to the principal raw channel to impose this constraint.

$$L_{mono} = \frac{1}{N}\sum_t max\left(0, -\frac{\partial \hat{y}_t}{\partial s_t}\right)^2 \quad \text{Eq.(2)}$$

The same sensor also exhibits finite thermal relaxation. Let T_t be the measured body temperature and τ_d be a learnable relaxation constant constrained to a positive interval. The thermal drift component should evolve smoothly rather than jump abruptly between adjacent samples, giving the residual

$$L_{relax} = \frac{1}{N}\sum_t \left[\delta_t - exp\left(-\frac{\Delta t}{\tau_d}\right)\delta_{t-1} - \kappa_d(T_t - T_{t-1})\right]^2 \quad \text{Eq.(3)}$$

These restrictions are relatively loose. They do not provide a full mechanistic model of all the sensors because such a model is rarely available in production. Instead, they have regularities that are stable across devices: response monotonicity, bounded drift rate, and smooth relaxation with temperature change. Physical terms were evaluated in both labeled and unlabeled target windows, and even with limited reference measurements, the calibration reliability of unlabeled operational history can be improved.

Transfer Objective

The training objectives are supervised source learning, sparse target learning, domain alignment and physics residuals. Source learning preserves the main calibration model and adapts the response for a new sensor in target learning. Domain alignment is performed on the latent features, not the raw signals; otherwise, valid process-mode differences in the raw signals would be lost. The entire reason is The weights of the alignment and physics terms are selected on a small target validation subset, and the same validation protocol is applied to every competing method.

$$L(\theta) = L_{sup}^s + \gamma L_{sup}^t + \lambda_a L_{align}(z_s, z_t) + \lambda_p(L_{mono} + L_{relax}) \quad \text{Eq.(4)}$$

where L_{sup} is the mean squared prediction loss on labeled samples, L_{align} is a maximum mean discrepancy penalty between source and target feature distributions within matched operating strata, and L_{phys} combines monotonic and relaxation residuals. The weights λ_a and λ_p are selected through a small validation set from the target device. The formulation is to prevent an excessively large transfer correction: if the target labels are too sparse, the physics term and source calibration will be prominent; when more target labels are available, the supervised target loss will gradually restrict the device-specific correction.

PI-LongNet-CNN Prediction Architecture

Multi-Scale Convolutional Feature Extraction

The first part of the proposed predictor is a multi-scale CNN branch that extracts local temporal features from each historical window. The three one-dimensional convolution streams have kernel sizes of 3, 7 and 15 samples, respectively, and cover rapid switching events, medium response lags and short thermal transients. Each stream applies depthwise convolution, pointwise channel mixing, layer normalization and gated activation. Concatenate the outputs and project them to a shared embedding dimension. The Design is relatively small to facilitate multiple training runs on different devices and to ensure deployment feasibility on supervisory computers or edge controllers. The CNN branch is used to learn features that are difficult to obtain through long-context attention, such as a short overshoot after a valve step, high-frequency vibration bursts, and local cross-correlation between the load command and raw sensor voltage. Figure 2 is referenced once here to show how the convolution streams, dilated LongNet blocks, calibration layer and physics-residual heads are connected in the complete PI-LongNet-CNN architecture. Kernel sizes 3, 7, and 15 are selected to cover rapid switching impulses, medium response lag, and short thermal transients without making the CNN branch excessively deep.

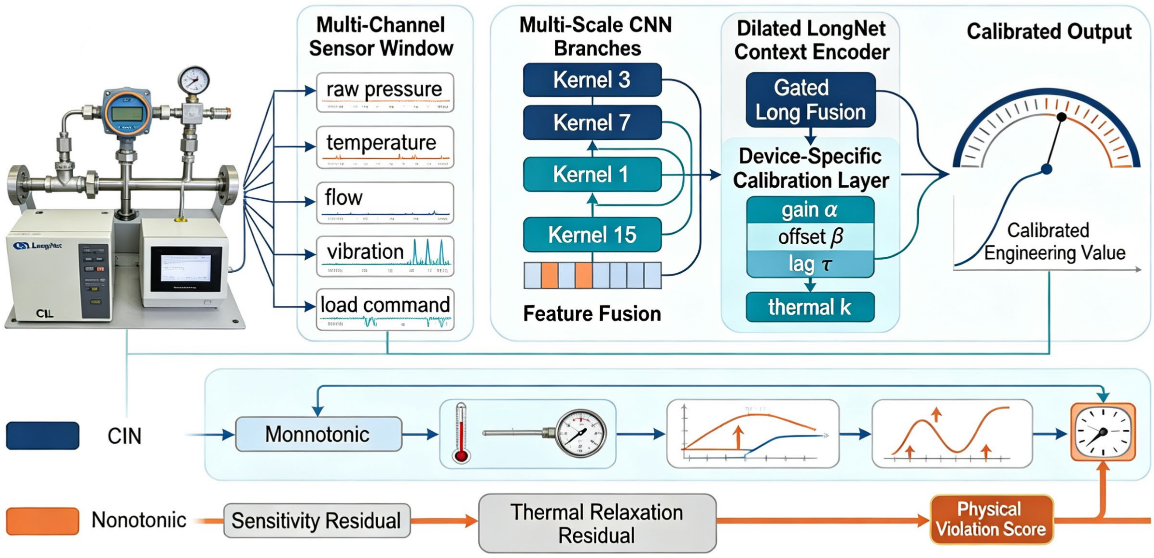

Figure 2. PI-LongNet-CNN architecture with multi-scale convolution, dilated long-context encoder, device-specific calibration layer, and physics-residual heads.

Dilated LongNet Context Encoder

The second part of the model has added dilated attention segments for long-term memory. Instead of considering each sample in an extended window separately, the encoder divides the sequence into several parts and connects near-field and far-field states with exponentially increasing dilation factors. This offers a reasonable-sized computationally feasible effective receptive field. For a token h_t at layer ℓ, the dilated attention output is represented as Dilated attention is preferred because it extends the effective context window while avoiding the full quadratic cost of dense attention over long industrial histories.

$$A_\ell(h_t) = \sum_{j\in S_\ell(t)} softmax\left(\frac{(W_Q h_t)^T (W_K h_j)}{\sqrt{d}}\right) W_V h_j \qquad \text{Eq.(5)}$$

where $S_\ell(t)$ is the set of keys selected by the segment length and dilation schedule, and W_Q, W_K, and W_V are learned projections. The early layers have a small dilation and keep local continuity; the later layers connect samples that are minutes or even hours apart. This mode is suitable for calibration migration because slow drift and thermal memory generally only appear after the model has observed a considerable amount of data. The LongNet branch receives the CNN embeddings, not the raw signals, and therefore attends to local features that already contain transient response and channel interactions.

A gated fusion module combines the final long-context representation and the most recent local CNN output in the architecture. The gate learns whether short transients are more informative than long drift history and when accumulated operating history should be used. The fused feature z_t is then input to a calibration head for the final engineered value and device-specific correction term. A small auxiliary head predicts the drift state and is not used as a label during deployment; it improves feature organisation during training.

Training and Deployment Procedure

The three Training periods are: First, the CNN and LongNet encoder are pre-trained on the source sensor with dense labels to learn a general mapping between raw multi-channel windows and calibrated values. Second, the encoder is trained using mixed source and target mini-batches, and on the target side, both sparse labelled windows and a larger set of unlabeled windows from the physics residual are included. Third, device-specific gain, offset and residual head are tuned, and some of the early convolution blocks are partially frozen. Thus, the model does not forget the calibrated source response but still has sufficient flexibility for target-domain migration. During adaptation, each mini-batch contains matched source windows, labeled target windows, and unlabeled target windows; early convolution blocks are partially frozen and early stopping is based on target validation RMSE and physical violation score.

The inference procedure is designed for ordinary industrial deployment. A rolling buffer stores the most recent L samples, updates normalized process variables, and calls the predictor at the sampling interval required by the control layer. A confidence flag is produced from the physics residual and from the distance between the current target feature and the source-target training strata. When either quantity exceeds its threshold, the system requests an additional reference calibration point rather than silently extrapolating. The prediction rule used during deployment is

$$C_t = f_\theta(X_t, \eta_d), \eta_d = [\alpha_d, \beta_d, \tau_d, \kappa_d]^T \quad \text{Eq.(6)}$$

where C_t is the calibrated estimate and η_d is the device-specific correction vector learned during transfer. This expression is implemented by the final calibration head, but writing it explicitly clarifies that the deployed model combines shared sequence representation with a compact target-sensor correction.

Experimental Data and Analysis

Experimental Design and Benchmark Protocol

A controlled industrial sensor benchmark was established to replicate the transfer conditions of pressure, flow, vibration and temperature monitoring for the evaluation. Six virtual production lines are generated from a hybrid simulator that has a base dynamics of first-order actuator response, thermal relaxation, friction-like hysteresis and stochastic process disturbances. The source line has 72h of dense calibration records sampled every 2s and generates 129,600 labelled windows after preprocessing. These five target lines have the same nominal sensing task, but differ in gain, offset, lag, noise color, thermal coupling, and drift slope. Only 8% of the target windows have reference labels; the rest are unlabeled and are only used for alignment and physics residual computation. The five or more operating conditions for the benchmark include: steady low-load, ramped-load, cyclic production, thermal warm-up and high-vibration disturbance. The first few are root mean square error, mean absolute error, calibration slope error, drift overshoot and inference latency. The benchmark is controlled rather than field-collected; the target lines are configured with different gain shifts, offsets, response lags, colored noise, hysteresis levels, thermal coupling coefficients, and drift slopes to approximate common commissioning differences.

Baselines include source-only CNN, target-only CNN, fine-tuned CNN, LSTM transfer, generic transfer transformer, CNN with domain alignment, and PI-LongNet-CNN variants without the selected components. All models use identical train, validation, and test splits. Stratify the target labels by operating mode and provide the same sparse calibration data for all methods. Hyperparameters are chosen based on validation RMSE and a physical violation score that counts negative sensitivity events and excessive relaxation jumps. Figure 3(a) shows the RMSE of the five target lines, and the proposed model reduces the average RMSE from 0.172 for source-only CNN and 0.145 for the transfer transformer to 0.118. The increase is most pronounced at T4; here, the lag and hysteresis are relatively larger, and the proposed model reduces RMSE from 0.193 to 0.132 by addressing them. All baselines use identical source-target splits, the same sparse target-label budget, and repeated target-label selections to reduce accidental advantages from sampling variation.

https://doi.org/10.64972/dea.2023.v2i3.3726d:65-80

Figure 3(b) shows MAE and calibration slope error in the same coordinate system. PI-LongNet-CNN moves to the lower-left with MAE 0.074 and slope error 0.031, and the transfer transformer is still at MAE 0.087 and slope error 0.048. Therefore, the above two steps will improve the accuracy of the point-by-point prediction and also enhance the slope of the transferred calibration curve. Figure 3(c) shows the target-wise slope-error comparison and indicates that four out of the five target sensors are still below the 0.04 engineering limit after transfer. T4 is slightly above the upper bound of 0.040 and has stronger hysteresis and a slower response. Figure 3(d) shows the distribution of absolute errors at all operating conditions; the interquartile range decreases from 0.061-0.146 for the transfer transformer to 0.043-0.091 for PI-LongNet-CNN, and the largest outliers occur under high-vibration disturbance rather than in steady low-load operation.

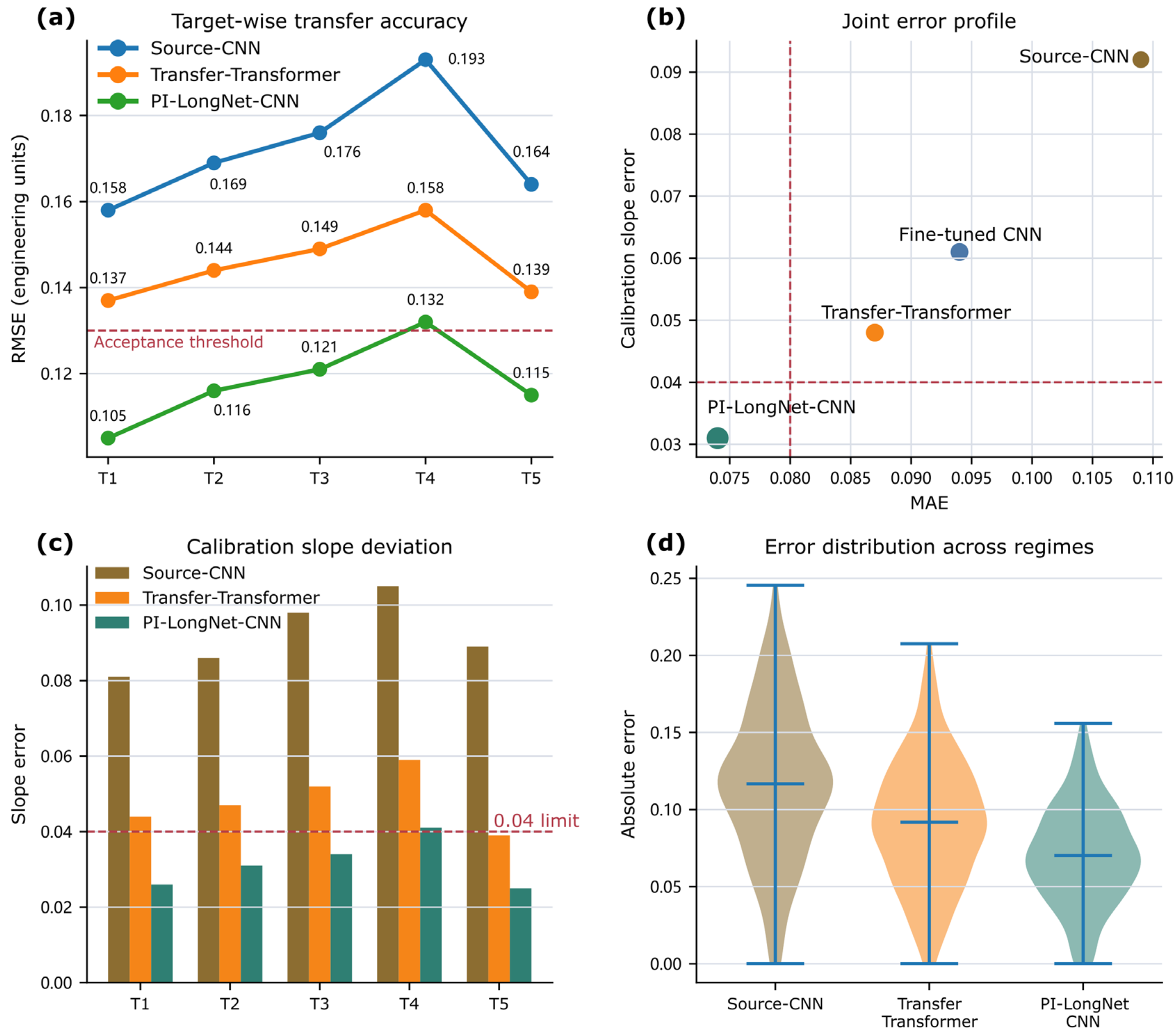

Figure 3. Overall prediction comparison: (a) RMSE across five target lines; (b) MAE and slope-error paired comparison; (c) target-wise calibration slope error; (d) absolute-error distribution under all operating regimes.

Prediction Accuracy and Calibration-Transfer Behavior

The specific prediction traces show why a long context and physical residuals improve transfer behavior. Under a 40-minute thermal ramp, the source-only CNN shows slight fluctuations but develops a positive bias after the sensor body temperature exceeds 54 °C. The transfer transformer corrects some of this bias but has a slow response when switching from ramped load operation to cyclic production mode. PI-LongNet-CNN has a peak absolute error of 0.37 units for the reference signal during the switch and is better than fine-tuned CNN (0.62) and transfer transformer (0.51). Figure 4(a) is the predicted-versus-reference scatter plot with a fitted slope of

https://doi.org/10.64972/dea.2023.v2i3.3726d:65-80

0.992 and an intercept of 0.011. The points are still close to the identity line throughout the entire range of 2.0-8.0, and thus the transferred model maintains the scale rather than just matching the central operating area.

Figure 4(b) shows the Bland-Altman residual band for the same dynamic test. The residual mean is -0.006, and the 95% limits of the interval are -0.232 and 0.219. The band is almost symmetrical, so the transfer error is not mainly due to a continuous positive or negative bias after adaptation. Figure 4(c) is the cumulative drift curve after thermal ramping. PI-LongNet-CNN limits the overshoot to 0.44 units after 12 hours, while the source-only CNN reaches 0.91 units and the transfer transformer remains above 0.55 units near the end of the ramp. Figure 4(d) is the regime-wise normalized error, and the largest remaining error is 0.134 in the high-vibration regime. The warm-up and cyclic regimes are still less than 0.12, so mechanical disturbance is harder to compensate for than smooth thermal memory in this benchmark.

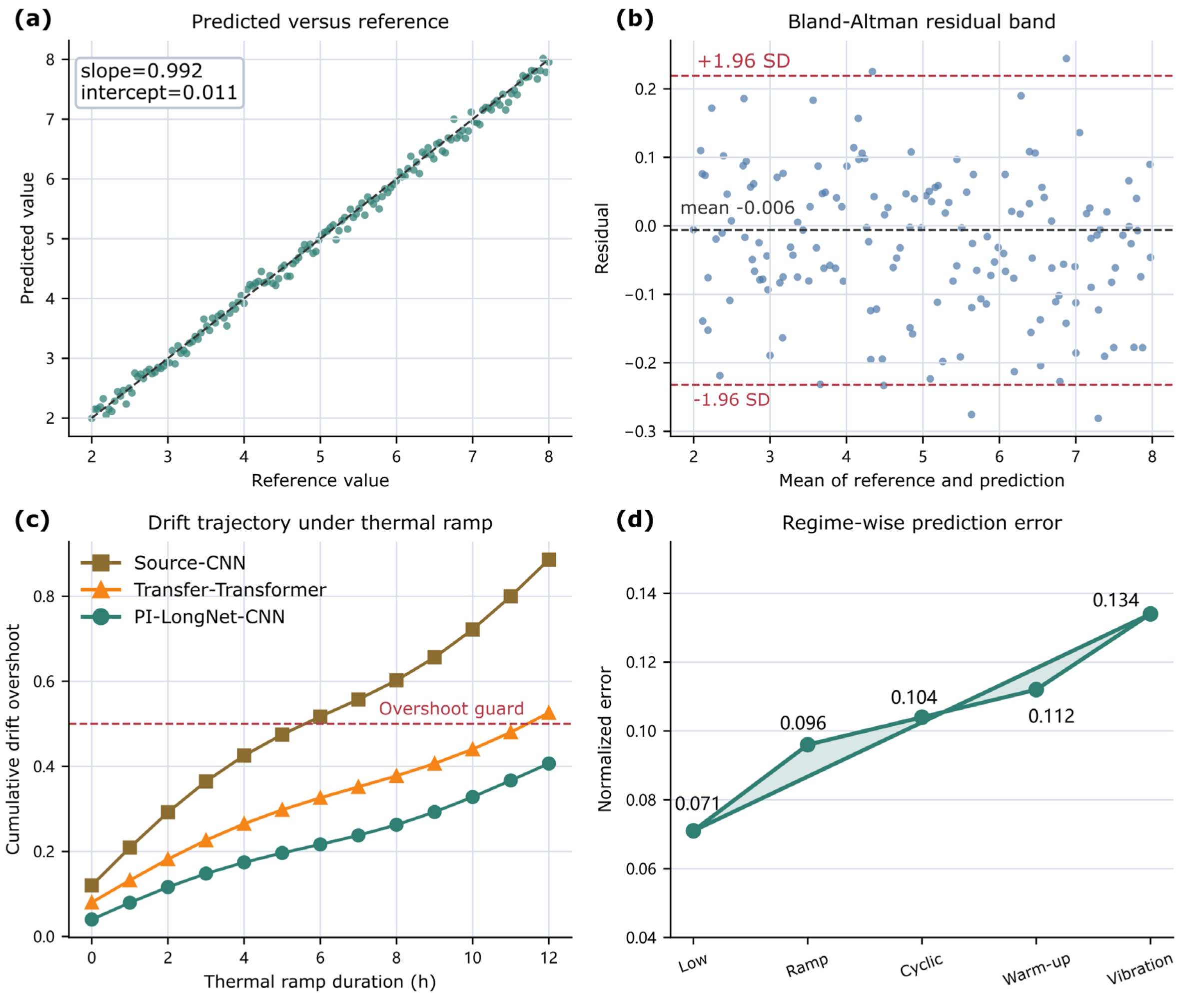

Figure 4. Calibration-transfer behavior under dynamic operating conditions: (a) predicted versus reference values; (b) Bland-Altman residual analysis; (c) drift trajectory under thermal ramping; (d) regime-wise normalized prediction error.

Ablation analysis can find out which part of the main architecture is responsible. As shown in Figure 5(a), the complete PI-LongNet-CNN has an RMSE of 0.118; without the addition of the physics residual, it is 0.133, and the rate of physical violation increases from 1.8% to 7.6%. This change is larger than the effect of removing domain alignment, and it has an RMSE of 0.127; thus, the physical residual also regularizes the latent space. It avoids fitting the device-specific correction to the sparse target labels via locally accurate but physically unstable response curves. The LongNet removal case has a relatively high RMSE of 0.138 and shows the largest drift-related error; that is to say, when calibration drift accumulates over thermal and production cycles, it is expressed more strongly than as a short local perturbation. The higher violation rate in the no-physics variant means that lower numerical error alone is insufficient for safe calibration migration when the transferred response curve becomes physically unstable.

Figure 5(b) compares the methods across normalized engineering metrics rather than a single error index. The scores of the proposed model are as follows: accuracy 0.92, drift stability 0.88, violation control 0.91, latency 0.76 and data efficiency 0.89. The transfer transformer has a competitive accuracy score of 0.78, but its latency is 0.48 and violation control is 0.62; it is a general long-sequence attention mechanism without response constraints. The fine-tuned CNN has the lowest latency score of 0.86, but its drift-stability score is only 0.56 because its receptive field is too small for slow-drift history. Therefore, the radar profile shows a trade-off that the RMSE table cannot, namely, that the proposed architecture sacrifices a small amount of computational simplicity to achieve a more balanced deployment profile.

Figure 5(c) examines target-label efficiency under sparse calibration. With only 2% target labels, PI-LongNet-CNN gives 0.162 RMSE, which is already lower than the transfer transformer at 0.191 and the no-physics variant at 0.184. At 6% target labels, the proposed model reaches 0.125 RMSE and approaches the engineering acceptance threshold, while the no-physics variant remains at 0.142. Increasing target labels from 8% to 24% improves the proposed model only from 0.118 to 0.109, whereas the no-physics model improves from 0.133 to 0.118. This difference indicates that physical residuals are most valuable in the low-label region, where the target device does not provide enough reference points to define the response shape by supervision alone.

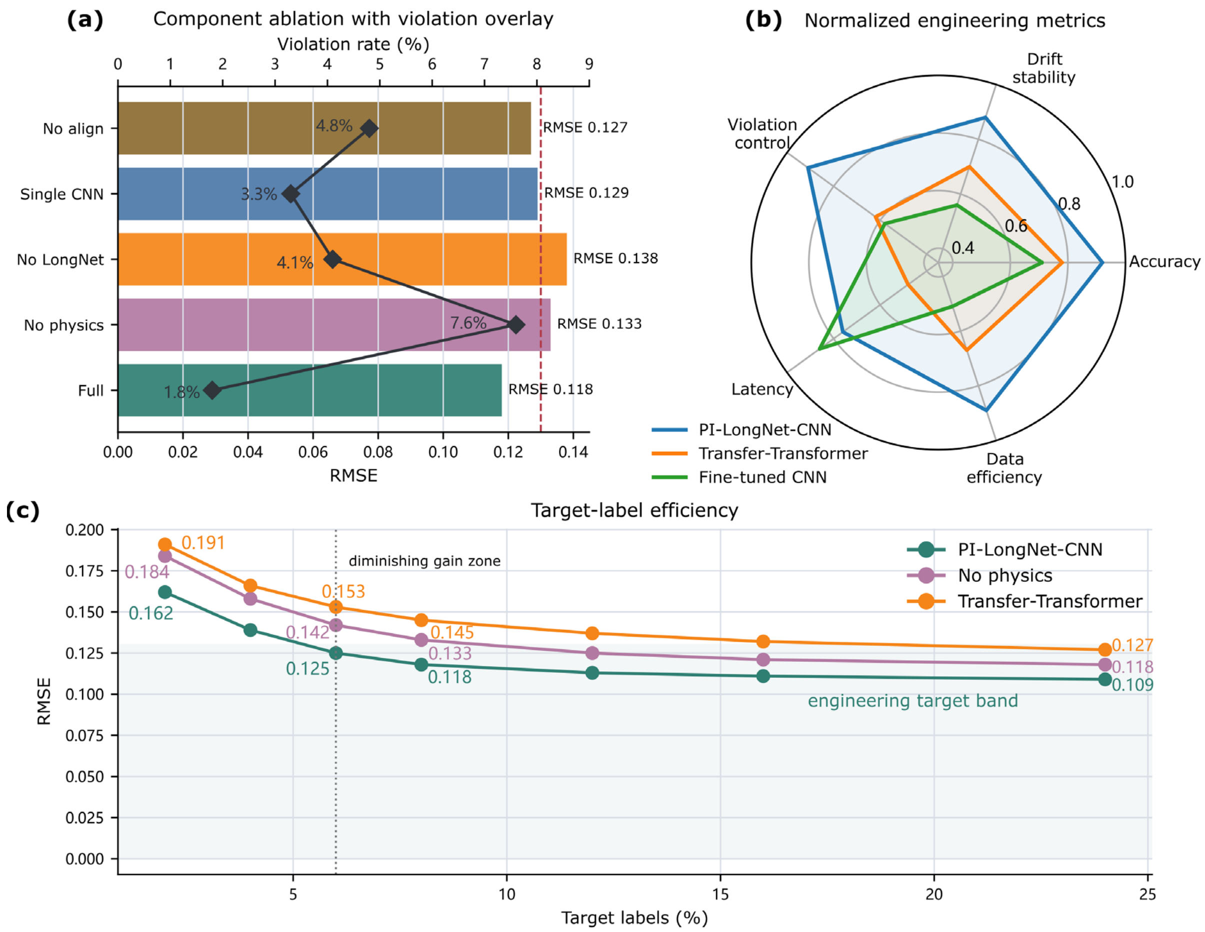

Figure 5. Ablation and component contribution: (a) RMSE after removing major components; (b) multi-metric radar comparison; (c) target-label efficiency curves.

Robustness, Data Efficiency, and Computational Cost

Noise level, thermal range shift, target-label percentage reduction, and lag between raw signal and reference measurement are all used to assess robustness. Research on industrial calibration has pointed out that the practical transfer method must remain stable when collecting target data under imperfect commissioning conditions [19]. At the present time, the Gaussian measurement noise has been increased to 0.12 units, and vibration bursts are randomly introduced. Figure 6(a) is the noise-sensitivity curve under this perturbation. PI-LongNet-CNN has an RMSE less than 0.16 until the injected noise standard deviation reaches 0.10 units, while the transfer transformer exceeds 0.19 at 0.08 units and reaches 0.259 at 0.12 units. Robustness values are averaged over repeated disturbance seeds and target-label selections, and confidence bands are used where the label sampling effect is visible.

Figure 6(b) examines sparse-label robustness with a confidence band over repeated target-label selections. The proposed model reaches RMSE 0.125 with only 6% target labels, and the upper confidence boundary remains close to 0.131 at that point. Figure 6(c) gives the source-target transfer matrix and identifies T4 as the hardest receiver, with a mean RMSE of 0.139 across donor lines. The matrix also shows that using S4 as a donor produces generally higher errors, suggesting that source-device quality affects migration even when the target adaptation procedure is unchanged.

Figure 6(d) presents the residual density under mixed noise and vibration disturbance. The distribution remains centered near zero with a mild right tail, and most residual mass stays inside the ±0.23 band used for monitoring. Figure 6(e) reports physical violation counts under increasing thermal shift. Without the physics residual, violations rise from 7 at zero shift to 54 under a +20 °C shift, whereas PI-LongNet-CNN limits the count to 13 under the same condition. Huang et al. [20] showed that domain-adaptive regression can reduce cross-domain mismatch, and the present results further indicate that adaptation becomes more reliable when domain-relevant structure restricts feature alignment.

Computational cost is important because calibration models may be retrained for many sensors after maintenance or replacement. The proposed model has a 256-sample input window, three convolution streams, four dilated attention blocks and a 128-dimensional hidden representation. Training on the benchmark takes 23.4 minutes with a single mid-range GPU, and target adaptation takes 4.8 minutes per sensor when reusing the source encoder. Inference latency is about 3.7ms per window on the GPU and 18.6ms on the industrial CPU-class processor. Figure 7(a) is the latency scaling when the input window length increases from 128 to 512 samples. The latency increase of the ordinary transformer is from 11.2ms to 42.7ms; that of PI-LongNet-CNN is 2.6ms to 6.9ms, and both are well within the 20ms control limit. The latency values are measured for a 256-sample window on a single mid-range GPU and an industrial CPU-class processor, which makes the deployment comparison interpretable.

Figure 7(b) presents the calibration-cost waterfall for one target sensor. Full recalibration requires 1,200 reference measurements, whereas sparse transfer reduces the measurement requirement by 1,104 points before adding a small residual-review budget. The final protocol uses 96 reference measurements while preserving RMSE below 0.13. Figure 7(c) compares energy-normalized inference cost and shows 0.84 mJ per prediction for the proposed model, lower than the 1.47 mJ transformer baseline but higher than the 0.51 mJ source-only CNN. The extra cost is justified in lines where drift or sparse labels would otherwise cause frequent recalibration, although ultra-low-power sensors may require pruning or integer quantization [21].

https://doi.org/10.64972/dea.2023.v2i3.3726d:65-80

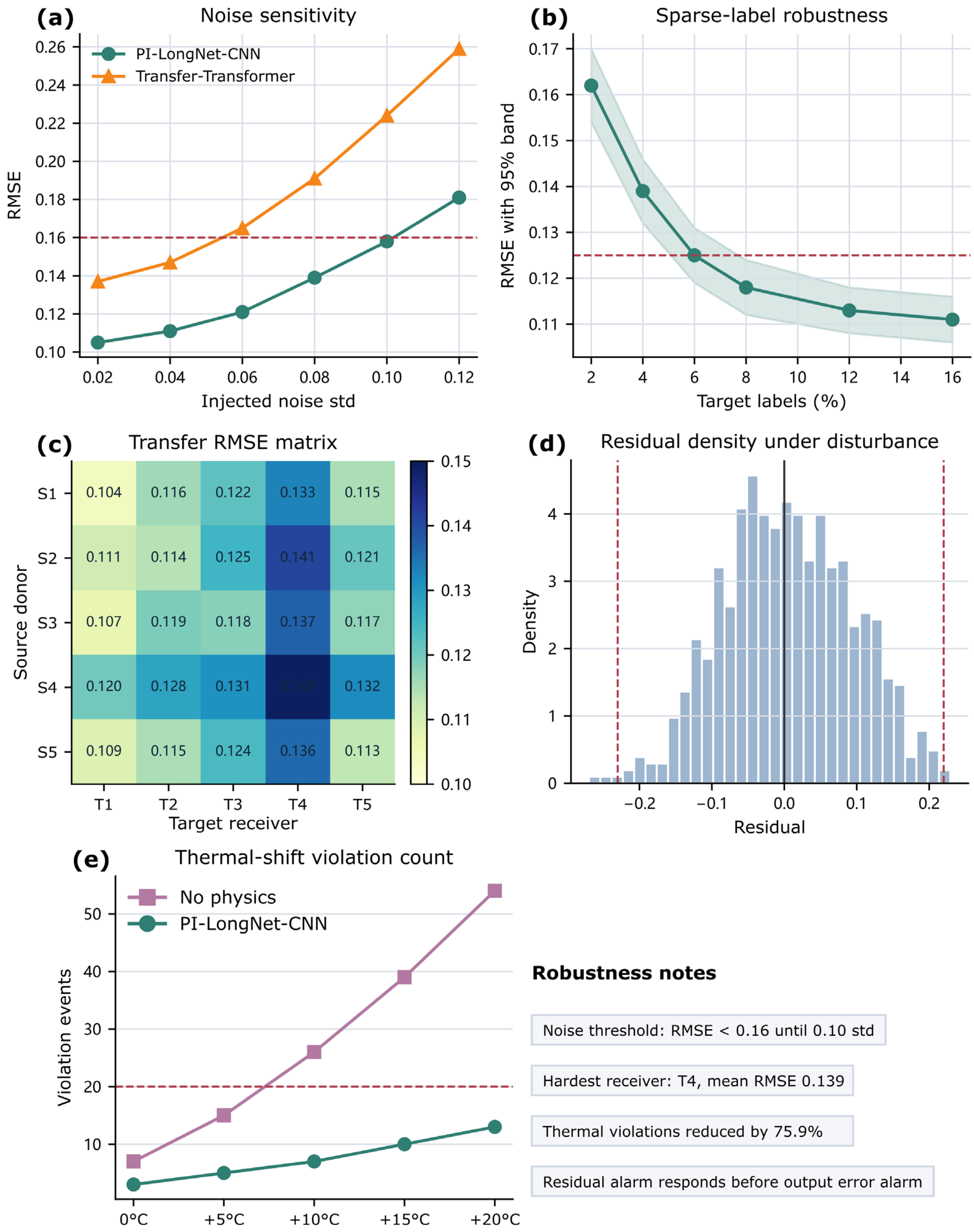

Figure 6. Robustness analysis: (a) noise sensitivity; (b) sparse-label efficiency; (c) source-target transfer matrix; (d) residual density under disturbances; (e) physical violation count under thermal shift.

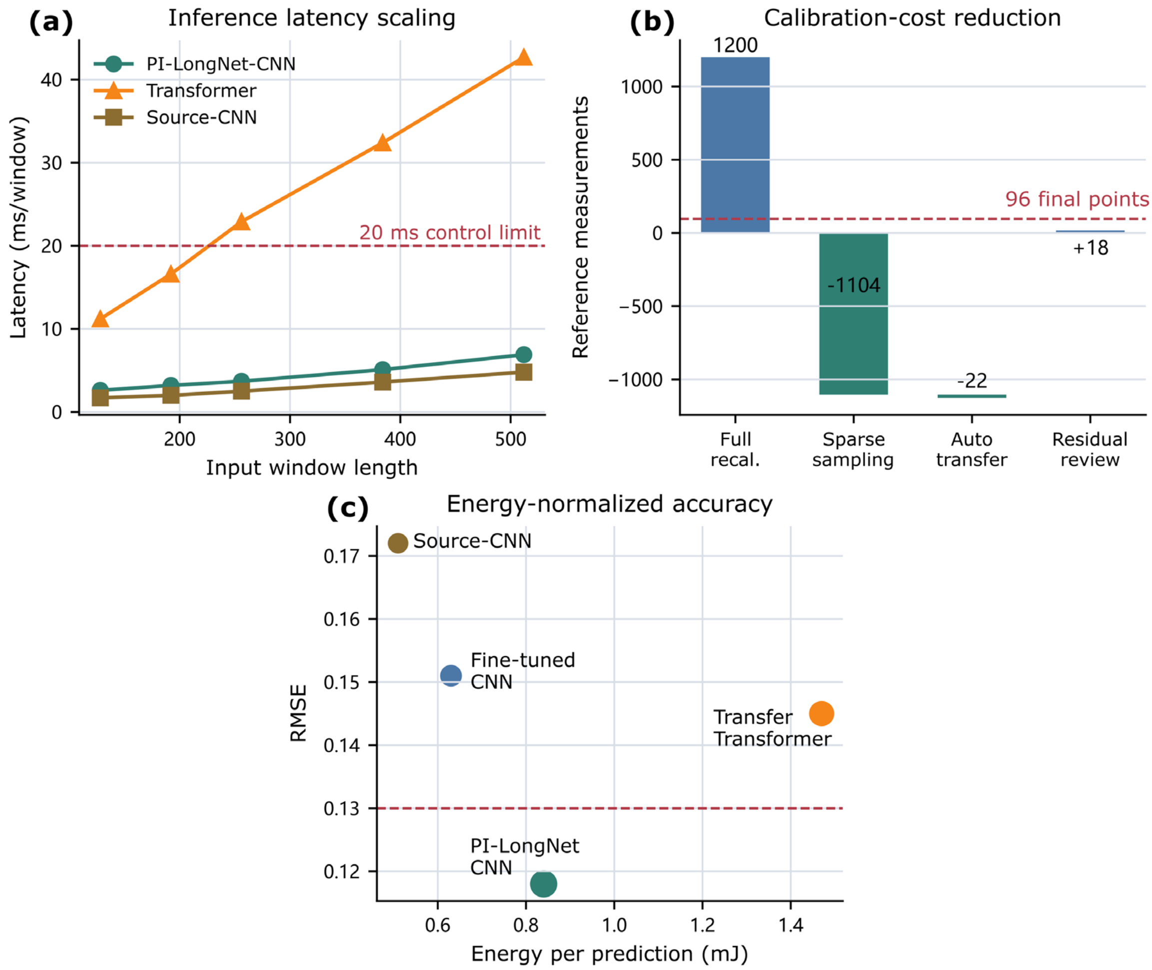

Figure 7. Deployment-oriented evaluation: (a) inference latency versus input-window length; (b) calibration-cost reduction from full recalibration to sparse transfer; (c) energy-normalized inference cost.

The fifth group of experiments investigates whether the model's behavior is still understandable to engineers. Attention maps show that long-range tokens are highlighted in the thermal warm-up and after maintenance-like resets, and CNN activation peaks appear around rapid command changes. The physics residual can also be used as a process-independent warning signal. When a target sensor is intentionally given an unmodeled offset jump of 0.6 units, the residual score exceeds the threshold in 18 samples, but the prediction error alone does not cross the alarm threshold until after 42 samples. Recently, similar residual monitoring has been proposed in physics-informed engineering models to detect deviations in the learned response structure before a significant accumulation of output error [22]. Calibration-transfer literature has also shown that small target sets can be misleading if they cover only a narrow range of the operating range [23].

The present results show that, when all target labels are taken from the steady low-load operation, the RMSE in the ramped-load test increases from 0.118 to 0.151, although the validation error remains low. Add only twelve ramped-load reference points, and the RMSE will be 0.124; therefore, sparse target labels should be chosen based on regime coverage rather than random convenience. Repeat the same experiment using only labels collected during high-vibration operation. In this case, cyclic-production error decreased slightly due to stronger local damping in the model, but warm-up error increased from 0.112 to 0.143 because the target correction underestimated temperature sensitivity. A balanced label set with 20% low-load points, 25% ramped-load points, 25% cyclic points, 20% warm-up points, and 10% high-vibration points produced the most stable validation-to-test behavior. Therefore, the calibration plan can be based on the actual sampling schedule and model accuracy instead of a general label-ratio approach.

Add comparisons with traditional standardization to clarify the function of the proposed neural network. Piecewise linear correction performs well when the target device differs only in gain and offset, achieving an RMSE of 0.141 for T1; however, due to a larger lag and hysteresis in T4, the error reaches 0.203. A domain-aligned CNN has improved the average to 0.136, but still shows 6.9% monotonicity violations under thermal shift. Deep calibration-transfer studies in spectroscopy have also found that flexible neural adaptation can outperform standardization when instrument differences are non-linear [24]. Yet the present benchmark shows that flexibility alone is insufficient: the transfer transformer without physics reaches RMSE 0.145 and produces the second-worst violation score. Research on sensor drift has found that cross-domain transfer samples can stabilize adaptation under mixed drift and shift [25].

In this paper, the unlabeled target windows also serve a similar function through latent alignment and physics residual evaluation, and the model can utilize the rich production data without needing a reference value for every target signal. The Unlabeled-to-labelled Ratio was also set at 1:1 and 20:1. Accuracy increased rapidly to a ratio of 8:1, and then the gain was small, and alignment loss occasionally over-smoothed rare high-load states. Therefore, the final protocol sets the alignment mini-batch size to ten unlabeled windows per labeled target window and retains regime labels in the sampler. The cap will prevent excessive steady-state records from dominating the adaptation process, and the experiments have also shown deployment constraints. PI-LongNet-CNN performs best when the target sensor has a relatively wide response characteristic compared with the source device. If the target hardware uses a different sensing principle or has serious saturation not seen in the source data, the physics residual flags an anomaly but cannot repair the missing calibration relationship. Multiple-instrument calibration-transfer studies have reported similar dependencies on the shared response structure [26]. The model assumes that the auxiliary variables of temperature and load command are aligned with the raw signal. When synchronization error exceeds four samples, RMSE increases from 0.118 to 0.146, and the relaxation residual becomes less reliable. Recent work on process soft sensors has shown that dynamic feature extraction is sensitive to lagged variables and mode boundaries [27]. Therefore, the deployment workflow will be aligned with the time of model adaptation. Another real-life problem is target-label quality. If the reference measurements have heteroscedastic noise exceeding 0.08 units, the learned device-specific correction will be unstable unless the target loss is down-weighted. Robust loss functions address this issue but reduce the fine-grained accuracy in steady-state operation [28]. Based on the above results, the proposed model is most suitable for sensor families with similar physical characteristics, a relatively small number of well-distributed target labels, and a long-enough operating history to exhibit drift memory [29].

Conclusion

This paper introduces a physics-informed LongNet-CNN prediction model for calibration transfer in industrial sensors. This way is to use multi-scale convolution for local transient extraction, dilated long-context attention for drift memory, and physical residuals for monotonic response and thermal relaxation consistency. The resulting framework addresses sparse target calibration by employing dense source labels, limited target labels, unlabeled target operating records, and stable physical regularities in a unified training objective. The manuscript developed the physical formulation, architecture, transfer procedure and experimental verification for the engineering problem of sensor calibration migration without full recalibration. The main implication is not that full recalibration can always be eliminated, but that reference measurements can be reduced when the target sensor shares the same broad response physics as the source device.

Several Deficiencies Remain. The benchmark is controlled and intentionally organized; thus, the response laws are more distinct than those in some field installations. The physics residuals are relatively weak approximations and not full mechanistic sensor models; therefore, they may only be suitable for the device family of the constraints chosen. The method also assumes synchronous auxiliary variables and representative target labels in the operating regime. If the target labels are concentrated in a narrow range, the model may perform well on the validation set but lack robustness to unobserved modes. Additional limitations include the absence of a full plant-scale field trial, simplified physics residuals, and the need to validate synchronization quality before deployment.

Future work will extend the framework to real multi-plant data, heterogeneous sensor systems and online adaptation after maintenance. More expressive physical constraints can be learned from digital twins or first-principles simulators, and uncertainty-aware heads can indicate when the transferred calibration should be trusted. Pruning, Quantization and residual-aware monitoring can also be applied at the edge to make calibration migration a regular feature of industrial sensing infrastructure rather than an exceptional offline model.

Author Contributions

Nicola Mariani contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Sergio Parisi contributes to analysis, investigation, data collection, draft preparation. Tito Romano contributes to software, validation, analysis. 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] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A comprehensive survey on transfer learning. Proceedings of the IEEE, 109(1), 43-76. https://doi.org/10.1109/JPROC.2020.3004555

[2] Sun, Q., & Ge, Z. (2021). A survey on deep learning for data-driven soft sensors. IEEE Transactions on Industrial Informatics, 17(9), 5853-5866. https://doi.org/10.1109/TII.2021.3053128

[3] Jiang, Q., Yan, S., Yan, X., Yi, H., & Gao, F. (2020). Data-driven two-dimensional deep correlated representation learning for nonlinear batch process monitoring. IEEE Transactions on Industrial Informatics, 16(4), 2839-2848. https://doi.org/10.1109/TII.2019.2952931

[4] Yan, W., Xu, R., Wang, K., Di, T., & Jiang, Z. (2020). Soft sensor modeling method based on semisupervised deep learning and its application to wastewater treatment plant. Industrial & Engineering Chemistry Research, 59(10), 4589-4601. https://doi.org/10.1021/acs.iecr.9b05087

[5] Yuan, X., Qi, S., Shardt, Y. A. W., Wang, Y., Yang, C., & Gui, W. (2020). Soft sensor model for dynamic processes based on multichannel convolutional neural network. Chemometrics and Intelligent Laboratory Systems, 203, 104050. https://doi.org/10.1016/j.chemolab.2020.104050

[6] Yuan, X., Qi, S., Wang, Y., & Xia, H. (2020). A dynamic CNN for nonlinear dynamic feature learning in soft sensor modeling of industrial process data. Control Engineering Practice, 104, 104614. https://doi.org/10.1016/j.conengprac.2020.104614

[7] Yuan, X., Li, L., Shardt, Y. A. W., Wang, Y., & Yang, C. (2021). Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. IEEE Transactions on Industrial Electronics, 68(5), 4404-4414. https://doi.org/10.1109/TIE.2020.2984443

[8] Wang, Y., Liu, D., Liu, C., Yuan, X., Wang, K., & Yang, C. (2022). Dynamic historical information incorporated attention deep learning model for industrial soft sensor modeling. Advanced Engineering Informatics, 52, 101590. https://doi.org/10.1016/j.aei.2022.101590

[9] Geng, Z., Chen, Z., Meng, Q., & Han, Y. (2022). Novel transformer based on gated convolutional neural network for dynamic soft sensor modeling of industrial processes. IEEE Transactions on Industrial Informatics, 18(3), 1521-1529. https://doi.org/10.1109/TII.2021.3086798

[10] Shen, B., & Ge, Z. (2021). Weighted nonlinear dynamic system for deep extraction of nonlinear dynamic latent variables and industrial application. IEEE Transactions on Industrial Informatics, 17(5), 3090-3098. https://doi.org/10.1109/TII.2020.3027746

[11] Mishra, P., & Passos, D. (2021). Deep calibration transfer: Transferring deep learning models between infrared spectroscopy instruments. Infrared Physics & Technology, 117, 103863. https://doi.org/10.1016/j.infrared.2021.103863

[12] Passos, D., & Mishra, P. (2021). Realizing transfer learning for updating deep learning models of spectral data to be used in new scenarios. Chemometrics and Intelligent Laboratory Systems, 212, 104283. https://doi.org/10.1016/j.chemolab.2021.104283

[13] Mishra, P. (2021). CT-GUI: A graphical user interface to perform calibration transfer for multivariate calibrations. Chemometrics and Intelligent Laboratory Systems, 214, 104338. https://doi.org/10.1016/j.chemolab.2021.104338

[14] Yap, X. Y., Chia, K. S., & Suarin, N. A. S. (2022). Adaptive artificial neural network in near infrared spectroscopy for standard-free calibration transfer. Chemometrics and Intelligent Laboratory Systems, 230, 104674. https://doi.org/10.1016/j.chemolab.2022.104674

[15] Shan, P., Zhao, Y., Wang, Q., Wang, S., Ying, Y., & Peng, S. (2021). A nonlinear calibration transfer method based on joint kernel subspace. Chemometrics and Intelligent Laboratory Systems, 210, 104247. https://doi.org/10.1016/j.chemolab.2021.104247

[16] Zhou, C., Yu, W., Huang, K., Zhu, H., Li, Y., Yang, C., & Sun, B. (2021). A new model transfer strategy among spectrometers based on SVR parameter calibrating. IEEE Transactions on Instrumentation and Measurement, 70, 1-13. https://doi.org/10.1109/TIM.2021.3119129

[17] Nikzad-Langerodi, R., & Sobieczky, F. (2021). Graph-based calibration transfer. Journal of Chemometrics, 35(4), e3319. https://doi.org/10.1002/cem.3319

[18] Nikzad-Langerodi, R., & Andries, E. (2021). A chemometrician's guide to transfer learning. Journal of Chemometrics, 35(11), e3379. https://doi.org/10.1002/cem.3379

[19] Nikzad-Langerodi, R., Zellinger, W., Saminger-Platz, S., & Moser, B. A. (2020). Domain adaptation for regression under Beer-Lambert's law. Knowledge-Based Systems, 210, 106447. https://doi.org/10.1016/j.knosys.2020.106447

[20] Huang, G., Chen, X., Li, L., Chen, X., Yuan, L., & Shi, W. (2020). Domain adaptive partial least squares regression. Chemometrics and Intelligent Laboratory Systems, 201, 103986. https://doi.org/10.1016/j.chemolab.2020.103986

[21] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422-440. https://doi.org/10.1038/s42254-021-00314-5

[22] Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: Where we are and what's next. Journal of Scientific Computing, 92(3), 88. https://doi.org/10.1007/s10915-022-01939-z

[23] Meng, X., Li, Z., Zhang, D., & Karniadakis, G. E. (2020). PPINN: Parareal physics-informed neural network for time-dependent PDEs. Computer Methods in Applied Mechanics and Engineering, 370, 113250. https://doi.org/10.1016/j.cma.2020.113250

[24] Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nature Machine Intelligence, 3, 218-229. https://doi.org/10.1038/s42256-021-00302-5

[25] Zhang, D., Guo, L., & Karniadakis, G. E. (2020). Learning in modal space: Solving time-dependent stochastic PDEs using physics-informed neural networks. SIAM Journal on Scientific Computing, 42(2), A639-A665. https://doi.org/10.1137/19M1260141

[26] 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

[27] Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

[28] Jiang, Y., Yin, S., Dong, J., & Kaynak, O. (2021). A review on soft sensors for monitoring, control, and optimization of industrial processes. IEEE Sensors Journal, 21(11), 12868-12881. https://doi.org/10.1109/JSEN.2020.3033153

[29] Schweidtmann, A. M., Esche, E., Fischer, A., Kloft, M., Repke, J. U., Sager, S., & Mitsos, A. (2021). Machine learning in chemical engineering: A perspective. Chemie Ingenieur Technik, 93(12), 2029-2039. https://doi.org/10.1002/cite.202100083