RENN: A Robust Physics-Informed Framework for Unsupervised Denoising of Constrained Data

C Lokesh Reddy* **Vishnu Swaroopji Masampally** **Venkataramana Runkana**

TCS Research, Tata Research Development & Design Centre
Tata Consultancy Services Limited
Pune 411057, India.

Abstract

Data reconciliation is critical for ensuring the integrity of industrial process data, yet existing methods face a debilitating trade-off: traditional iterative techniques require precise error statistics, while conventional data-driven models demand unavailable ground-truth labels and often disregard fundamental physical laws. This paper introduces the Robust Estimator-guided Neural Network (RENN), a unified framework that overcomes these limitations by integrating data-driven learning with process physics and robust statistics. RENN uniquely embeds a Quasi-Weighted Least Squares (QWLS) robust estimator directly into a physics-informed learning objective. This novel approach enables the network to train end-to-end solely on noisy, unlabelled measurements, learning to simultaneously filter noise and enforce physical consistency. Once trained, RENN delivers reconciled estimates in a single, rapid feedforward pass, making it ideal for real-time applications. We demonstrate that on complex, industrial-scale benchmarks, our framework consistently matches or outperforms state-of-the-art robust estimators and supervised neural network baselines, achieving near-perfect mass balance closure without ever seeing true process values during training. This work presents a paradigm shift for data reconciliation that offers a fast, accurate, and physically consistent solution without curated training data.

1 Introduction

High-quality data is the cornerstone of modern industrial operations, underpinning critical tasks from process control and optimization to safety and environmental compliance. However, the deluge of data from high-frequency sensors is invariably corrupted by measurement noise, sensor drift, and gross errors, which can lead to flawed analyses, compromising reliability and leading to operational inefficiencies or even unsafe operating conditions [1]. Data reconciliation aims to mitigate these effects by adjusting measured variables to satisfy known physical laws, such as the conservation of mass and energy.

State-of-the-art approaches to this problem fall into two main categories, each with significant limitations. Classical methods rely on robust statistical estimators (e.g., M-estimators) framed as optimization problems [2–5]. While effective, they are iterative and hand crafted whereas purely data-driven methods like standard neural networks often fail in this domain because they require large datasets with clean, ground-truth labels [6, 7], a rarity in industrial settings and they learn correlations from historical data [8, 9] without any mechanism to enforce fundamental physical laws, leading to physically implausible predictions. The integration of data-driven models with physics has also been explored, often by combining data driven models and classical estimators sequentially. For instance, reconciliation can be applied as a preprocessing step to improve input quality for

*Corresponding author: clokesh.reddy@tcs.com

2nd Conference on Applied AI and Scientific Machine Learning (CASML 2025).

hybrid soft sensors that combine mechanistic rules with kernel-based learners [10, 11]. Mousa et al. proposed a framework that integrates physics with data reconciliation, treating neural network predictions as "soft sensors," which are then reconciled externally using a quadratic optimization problem, primarily for linear constraints [12]. Recent advancements in Physics-Informed Neural Networks (PINNs) have enabled the integration of physical laws into deep learning, typically by penalizing the governing equations' residuals in the loss function [13]. However many studies still rely on modular frameworks and ground-truth labels, either directly or indirectly and do not offer a single, integrated training framework. This leaves a critical gap: no existing method effectively addresses the three industrial desiderata: i) accurate and fast feedforward inference suitable for real-time correction ii) independence from ground-truth labels, and iii) end-to-end satisfaction of physical constraints.

To overcome these limitations, this paper introduces the Robust Estimator-guided Neural Network (RENN), a unified framework that embeds a robust Quasi-Weighted Least Squares (QWLS) estimator directly into a physics-informed training objective. This approach uniquely allows the network to learn a robust, non-linear reconciliation function directly from noisy, unlabelled measurements. The resulting model provides reconciled estimates via a single feedforward pass, making it suitable for real-time deployment. Our experiments on challenging industrial benchmarks show that RENN significantly outperforms traditional estimators and supervised baselines, achieving mass balance closures of over 99.75 % while effectively reducing measurement error.

2 Methodology

Our proposed RENN framework adapts the core principles of physics-informed learning to the specific challenges of industrial data reconciliation. We achieve this by designing a neural network guided by a composite loss function that simultaneously enforces robustness to outliers and adherence to physical laws. We begin by briefly summarizing the foundational concepts before detailing our novel contributions.

Data reconciliation is formally an optimization problem that adjusts noisy measurements, y, to find the estimated values, x^r, that minimizes an objective function of the measurement errors, $\rho(\varepsilon_i)$, subject to a set of physical constraints, $G(x^r) = 0$. We use the Quasi-Weighted Least Squares (QWLS) estimator, a robust objective whose influence on the reconciliation result remains bounded even for large errors, making it highly effective at mitigating the impact of outliers [3]. Standard PINNs integrate physics by adding a penalty term to the loss function that quantifies violations of a governing PDE, guiding the network towards physically plausible solutions [13]. However, standard PINNs are ill-suited for process network reconciliation, which involves constraints on discrete nodes that lack any spatial or temporal domain rather than continuous PDE domains. Furthermore, their conventional mean-squared-error data loss term is not robust to the outliers and gross errors prevalent in industrial data. RENN modifies this paradigm for discrete process networks. It is built on two primary modifications. First, we replace the conventional supervised data-fidelity loss with an unsupervised, reconciliation-specific objective based on the robust QWLS estimator. Second, we enforce the system's physical laws by directly penalizing the residuals of the mass balance equations in the loss function. This synergistic design enables the network to learn a direct mapping from noisy measurements to reconciled, physically-consistent values without ever seeing ground-truth data.

The RENN architecture is a fully connected neural network, f_θ, that learns the direct mapping from a vector of noisy measurements, $\vec{y}$, to a vector of reconciled estimates, $\vec{x_\theta^r} = f_\theta(\vec{y})$. The network architecture is designed to be deep enough to capture complex, non-linear relationships between process variables but compact enough for efficient training and inference The architecture of our proposed framework is illustrated in Figure 1.

To train the network without ground-truth labels, we designed a composite loss function which is a weighted sum of a reconciliation loss and multiple constraint satisfaction losses.

$$L_{\text{total}}(\theta) = \lambda_{\text{recon}} L_{\text{recon}}(\theta) + \sum_{k=1}^{K} \lambda_{\text{cons},k} L_{\text{cons},k}(\theta) \tag{1}$$

where K is the number of process units (and thus mass balance constraints) and the weights, $(\lambda_{\text{recon}}, \lambda_{\text{cons},k})$ are hyperparameters that balance the trade-off between robust data fidelity and strict

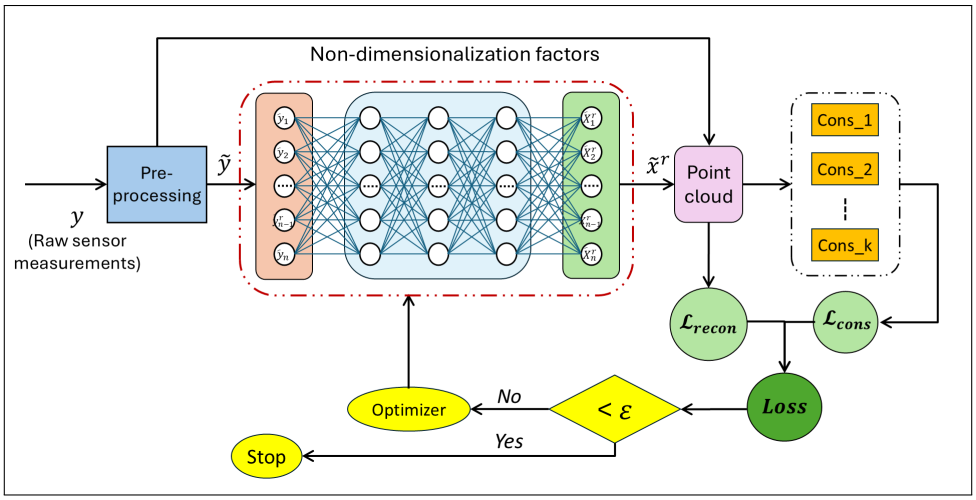

Figure 1: Architecture of the Robust Estimator-guided Neural Network (RENN) framework. Raw measurements are preprocessed and normalized before being fed into the network. The normalized reconciled outputs and measurements are used to reconciliation and constraint loss. Total loss guides the optimizer for training without true-labels

physical adherence. The first loss component is the reconciliation loss. The standard PINNs' unweighted L_2 data loss ignores sensor precision heterogeneity, biases reconciliation towards low precision sensors and is sensitive to outliers. To address this, we introduce a reconciliation-specific loss function using the QWLS estimator, which weights errors by measurement precision and incorporates robust regularization to suppress outliers. Its robustness under large errors is captured by the Influence Function (IF) [14]. The estimator is robust if the effect of large errors on IF are bounded and leads to a constant value when the errors increase.

$$\rho(\varepsilon_i) = \frac{\varepsilon_i^2}{2+\beta|\varepsilon_i|}, \quad IF = \frac{d\rho(\varepsilon)}{d\varepsilon} = \begin{cases} \frac{4\varepsilon-\beta\varepsilon^2}{(2-\beta\varepsilon)^2}, & \varepsilon < 0 \\ \frac{4\varepsilon+\beta\varepsilon^2}{(2+\beta\varepsilon)^2}, & \varepsilon > 0 \end{cases} \tag{2}$$

$$\text{if } \varepsilon \to \infty, \quad IF = \begin{cases} -\frac{1}{\beta}, & \varepsilon < 0 \\ \frac{1}{\beta}, & \varepsilon > 0 \end{cases} \tag{3}$$

A key challenge is the wide disparity in the scales of process variables, which can cause training to be dominated by larger flows. To mitigate this, we non-dimensionalize all equations and variables involved in the loss computation. Therefore, hereafter all equations and variables are non-dimensionalized. The L_{recon} penalizes deviations between the network's scaled outputs ($\tilde{x}^r_{\theta,i}$) and the scaled noisy measurements ($\tilde{y}_i$) weighted by measurement precision ($\tilde{\sigma}$). Its purpose is to guide the network to produce estimates that remain faithful to the original measurements in a robust manner.

$$L_{\text{recon}}(\theta) = \frac{1}{N_{\text{recon}}} \sum_{i=1}^{N_{\text{recon}}} \sum_{j=1}^{S} \frac{\tilde{\varepsilon}_j}{2+\beta|\tilde{\varepsilon}_j|} \tag{4}$$

The second loss component is the mass balance constraint loss. For a constraint k, we define a residual from the non-dimensionalized mass balance equation and the loss term is shown in the equation below:

$$L_{\text{cons},k}(\theta) = \frac{1}{N_{\text{cons}}} \sum_{i=1}^{N_{\text{cons}}} \sum_{j=1}^{S} a_{kj}(\tilde{x}^r_j \delta_j + y_{min,j}) \tag{5}$$

Here, a_{kj} is the incidence factor, while δ_j and $y_{min,j}$ are constant scaling factors for stream j. This loss term penalizes any violation of the conservation of mass, forcing the network's scaled outputs to satisfy the system's physical laws.

We employ a deterministic, constraint-based pointcloud which simplifies sampling of collocation points in a discrete domain-less problem avoiding cross-sampling between different sets of measurements and managing memory usage for a high-dimensional data. The network is trained using full-batch evaluation with the Adam optimizer (learning rate of 0.001) until the change in reconciliation loss falls below a tolerance of 1e-04. Model performance is assessed using four key metrics: Coefficient of Determination (R^2) for accuracy, Root Mean Squared Error (RMSE) for noise reduction, Mass Balance Closure (%) for constraint satisfaction, and Relative Error Reduction (RER) for comparative analysis against other methods.

3 Results

The RENN framework was validated on several diverse industrial benchmarks ranging from (4-26) units and (6-50) streams. However to demonstrate the model's robustness and generalizability, this section focuses on its application to one of the complex, industrial-scale steam metering system [15]. This benchmark, derived from an actual methanol synthesis plant, is particularly challenging due to its large scale (11 units, 28 streams), highly interconnected topology, and wide dynamic range of flow rates captured from real plant data. We compare RENN's performance against the best-performing traditional methods from our study as well as a supervised Artificial Neural Network (ANN) baseline trained on ground-truth labels.

A primary measure of model performance is its ability to recover the true process values from noisy measurements. We quantified this using the coefficient of determination (R^2) between the reconciled estimates and the ground-truth values. As presented in Table 1 the RENN model consistently achieves high R^2 scores across all streams, demonstrating robust performance even under large variations in flowrate spanning three orders of magnitudes. This indicates that the training process is not dominated by higher magnitude flows and that the model generalizes well across different scales. Overall, the consistently high R^2 values confirm that the RENN framework effectively rectifies measurement errors and yields estimates in excellent agreement with the ground truth.

Table 1: R^2 coefficients of RENN predictions relative to true values

Stream	R^2	Stream	R^2	Stream	R^2	Stream	R^2
1	0.9111	8	0.9910	15	0.9680	22	0.9695
2	0.9122	9	0.9181	16	0.9526	23	0.9188
3	0.9924	10	0.9724	17	0.9247	24	0.9166
4	0.9924	11	0.9229	18	0.9183	25	0.9484
5	0.9727	12	0.9725	19	0.9153	26	0.97
6	0.9920	13	0.9919	20	0.9203	27	0.9734
7	0.9164	14	0.9111	21	0.9190	28	0.9735

RMSE was used to evaluate the model's noise-filtering ability. RMSE for both raw measurements and reconciled estimates are calculated relative to the true values and the reduction in error is shown as a vertical line in Figure 2(a). A substantial reduction is evident and the uniform suppression across all streams confirms the model's effectiveness in error rectification. We also compare RENN's performance against the best-performing traditional methods from our study-QWLS, Contaminated Normal, Cauchy estimator, Absolute estimator as well as a supervised Artificial Neural Network (ANN) baseline trained on ground-truth labels. The comparative performance, summarized in Figure 2(b) highlights the superiority of our framework. RENN achieves the highest average Relative Error Reduction (RER), a metric that simultaneously quantifies the percentage of measurement error removed by reconciliation and its closeness to true value. This indicates that our unsupervised method is more effective at noise filtering than the iterative robust estimators. The baseline ANN, despite being trained with ground-truth data, performs similarly, underscoring the critical importance of embedding robust reconciliation objective directly into the learning process—a core feature of the RENN architecture. Critically, our method enforces physical laws with high fidelity, improving the mass balance closure for all 11 units to over 99.75% reported in Table 2.

Table 2: Achieved mass balance closure by RENN predictions compared to raw measurements

Unit	1	2	3	4	5	6	7	8	9	10	11
Measurements	96.13	97.49	95.14	97.4	96.2	94.16	97.39	97.57	96.55	97.26	96.58
RENN estimates	99.86	99.92	99.91	99.92	99.94	99.88	99.94	99.75	99.92	99.88	99.95

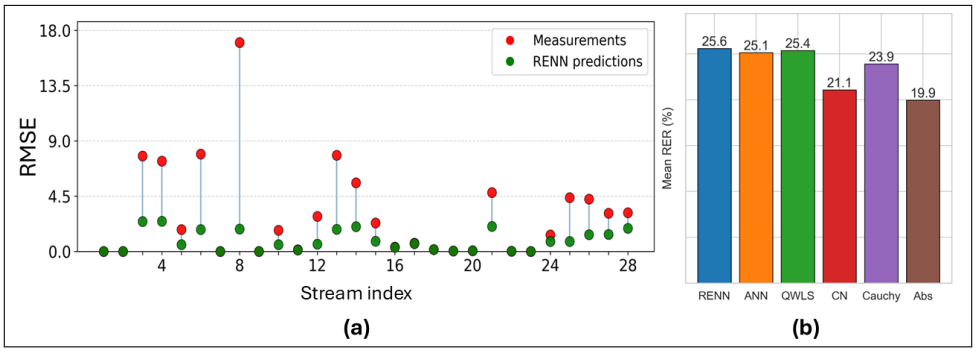

Figure 2: (a) Reduction in RMSE in RENN estimates compared to raw measurements, (b) Comparison of average Relative Error Reduction (in %) achieved by RENN against existing methods

An ablation study was conducted to isolate the contributions of RENN's components, with results summarized in Figure 3. The full RENN model with tuned loss weights achieves the highest RER. The performance degradation from removing a single component demonstrates the necessity of each loss component. The drop in performance when using equal weights indicates the importance of appropriately scaling the loss components relative to each other. Furthermore, supervised ANN performs similarly even after trained on true values. This underscores the synergistic effect of holistically enforcing all physical laws and appropriately balancing the loss components to achieve optimal performance.

Figure 3: Effect of constraint loss components and their weighting on RER performance

4 Conclusion

In this work, we introduced the Robust Estimator-guided Neural Network (RENN), a novel framework for data reconciliation that successfully operates without ground-truth labels. By integrating a robust statistical estimator and physical conservation laws directly into the training objective, RENN learns to produce highly accurate and physically consistent reconciled data from noisy measurements alone. Its single-pass, feedforward architecture makes it computationally efficient and suitable for real-time industrial applications. Our results on complex benchmarks demonstrate that RENN matches or exceeds the performance of state-of-the-art iterative methods and supervised deep learning models. The framework's primary limitations are its current applicability to fixed process topologies and its assumption that all variables are measured. Future work will focus on extending the framework

to handle dynamic plant configurations and unmeasured variables, further enhancing its industrial relevance. Ultimately, RENN offers a powerful and practical solution to a long-standing challenge in the process industries, paving the way for more reliable process monitoring, control, and optimization.

Acknowledgments and Disclosure of Funding

The authors would like to thank Dr. Harrick Vin and Dr. Sachin Lodha for their continuous encouragement and guidance. The authors declare no external funding for this research.

References

[1] Shankar Narasimhan and Cornelius Jordache. *Data reconciliation and gross error detection: An intelligent use of process data*. Elsevier, 1999.

[2] Nikhil Arora and Lorenz T Biegler. Redescending estimators for data reconciliation and parameter estimation. *Computers & Chemical Engineering*, 25(11-12):1585–1599, 2001.

[3] Zhengjiang Zhang, Zhijiang Shao, Xi Chen, Kexin Wang, and Jixin Qian. Quasi-weighted least squares estimator for data reconciliation. *Computers & chemical engineering*, 34(2):154–162, 2010.

[4] Claudia E Llanos, Mabel C Sanchez, and Ricardo A Maronna. Robust estimators for data reconciliation. *Industrial & Engineering Chemistry Research*, 54(18):5096–5105, 2015.

[5] Sen Xie, Chunhua Yang, Xiaofeng Yuan, Xiaoli Wang, and Yongfang Xie. A novel robust data reconciliation method for industrial processes. *Control Engineering Practice*, 83:203–212, 2019.

[6] Patricia A Terry and David M Himmelblau. Data rectification and gross error detection in a steady state process via artificial neural networks. *Industrial & engineering chemistry research*, 32(12):3020–3028, 1993.

[7] Mark A Kramer. Autoassociative neural networks. *Computers & chemical engineering*, 16(4):313–328, 1992.

[8] Guiting Hu, Luping Xu, and Zhengjiang Zhang. Correntropy based elman neural network for dynamic data reconciliation with gross errors. *Journal of the Taiwan Institute of Chemical Engineers*, 140:104568, 2022.

[9] Zhengjiang Zhang, Junghui Chen, Xiaofei Wu, Lei Xie, and Chun-I Chen. A novel strategy of correntropy-based iterative neural networks for data reconciliation and gross error estimation in semiconductor industry. *Journal of Process Control*, 131:103096, 2023.

[10] Sen Xie, Huaizhi Wang, and Jianchun Peng. A hybrid prediction model of recycled sodium aluminate solution concentration in evaporation process. *IEEE Transactions on Instrumentation and Measurement*, 70:1–13, 2021.

[11] Wangwang Zhu, Mingwei Jia, Zhengjiang Zhang, and Yi Liu. Dynamic data reconciliation for enhancing the performance of kernel learning soft sensor models considering measurement noise. *Chemometrics and Intelligent Laboratory Systems*, 246:105083, 2024.

[12] Jana Mousa, Stéphane Negny, Rachid Ouaret, Alessandro Di Pretoro, and Ludovic Montastruc. Incorporating physical constraints inside neural networks to improve their accuracy and physical reliability for chemical engineering unit operations modeling. *Computers & Chemical Engineering*, page 109156, 2025.

[13] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[14] Derya B Özyurt and Ralph W Pike. Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. *Computers & chemical engineering*, 28(3):381–402, 2004.

[15] RW Serth and WA Heenan. Gross error detection and data reconciliation in steam-metering systems. *AIChE Journal*, 32(5):733–742, 1986.