

Paper Technology

Rosario Othen*, Chen Song, Jan Schlake, Christian Möbitz and Thomas Gries

Hybrid soft sensor modeling for the wire section in tissue manufacturing: integrating first-principles and machine learning models

<https://doi.org/10.1515/npprj-2025-0045>

Received July 10, 2025; accepted October 17, 2025;

published online October 31, 2025

Abstract: Modeling and control of industrial processes are essential for product quality, energy efficiency, and operational stability. With growing sensor availability and computational power, machine learning (ML) has emerged as a powerful tool. However, purely data-driven models often lack interpretability and robustness, especially under varying conditions. A key example is the wire section of a tissue machine, where basis weight is determined but not directly measurable in real time. To address this challenge, we propose a hybrid modeling framework that integrates a first-principles model (FPM) with a one-dimensional convolutional neural network (Conv1D) in a parallel structure. A secondary gated recurrent unit (GRU) network dynamically adjusts the weighting between the two predictions. Evaluated on seven months of industrial data, the hybrid model achieved a root mean square error (RMSE) of 2.12 g/m^2 , outperforming standalone FPM (4.72 g/m^2) and ML (2.68 g/m^2) models. Designed for real-time deployment, the approach provides a scalable soft-sensor solution, enabling earlier intervention and more adaptive operation in tissue manufacturing.

Keywords: hybrid modeling; machine learning; first-principle model; wire section; paper machine; soft sensor

1 Introduction

Modeling and simulation are essential tools for improving process efficiency, product quality, and cost-effectiveness in

manufacturing industries. In the process industry, two dominant paradigms have emerged: first-principle modeling (FPM) and machine learning (ML). FPMs are grounded in physical laws such as mass and energy balances, offering high interpretability and consistency with known system behavior (Pantelides and Renfro 2013). However, building and calibrating such models requires detailed process knowledge and can become computationally intensive for complex, nonlinear systems (Bikmukhametov and Jäschke 2020).

Machine learning, by contrast, enables data-driven modeling with minimal reliance on process-specific assumptions. These models often achieve strong predictive accuracy, especially when sufficient historical data are available. Yet they typically function as black boxes, lacking interpretability and the ability to extrapolate outside trained operating regimes (Bikmukhametov and Jäschke 2020; Thai 2022). To address these limitations, hybrid modeling approaches have been proposed that integrate FPM and ML techniques – leveraging the strengths of both and mitigating their individual weaknesses (Hotvedt et al. 2021). In tissue manufacturing, the wire (forming) section governs initial web formation and early dewatering and thus strongly influences downstream quality, energy use, and runnability. Classic sources emphasize that the wire section removes more than 80 % of the web's water (Ramaswamy 2003), and contemporary operations guidance and recent modeling work indicate that this is typically on the order of 90 % to 95 % of the total water handled on the machine, underscoring the need for stability and controllability at the wet end (ABB Ltd 2025; Sjöstrand 2025; Sjöstrand and Bergström 2024; Valmet 2025). Process variables in this section – particularly basis weight, a key quality parameter – have a major influence on downstream product performance and energy use. However, basis weight is not measured directly in the wire section but typically downstream, near the reel, introducing delays that limit responsiveness during transients or machine start-ups (Holik 2006). At typical production speeds (e.g., 2,000 m/min) and a machine width of 5.34 m, such a scanner feedback delay of 20 s–30 s results in 667 m–1,000 m of paper being produced without real-time quality feedback. For a basis weight of 20 g/

*Corresponding author: Rosario Othen, Institut für Textiltechnik of RWTH Aachen University, Otto-Blumenthal-Straße 1, 52074 Aachen, Germany, E-mail: rosario.often@ita.rwth-aachen.de. <https://orcid.org/0000-0002-7844-7721>

Chen Song and Jan Schlake, ABB AG, Kallstadter Str. 1, 68309 Mannheim, Germany

Christian Möbitz and Thomas Gries, Institut für Textiltechnik of RWTH Aachen University, Otto-Blumenthal-Straße 1, 52074 Aachen, Germany

m², this corresponds to 71 kg–107 kg of potentially off-spec material per event. Besides the material loss, this also implies unnecessary energy usage and reduced start-up efficiency. Wet-end stabilization via retention and short-circulation control (e.g., headbox and white-water consistency, first-pass retention, ash) is therefore central to reducing MD variability upstream of the scanner (ABB Ltd 2025; Valmet 2025). Building on this, predictive modeling of basis weight in the wire section enables soft sensors that see quality before the reel scanner, improving start-ups and transient control. Recent tissue-relevant work spans machine-learning models of tissue-machine energy/efficiency and advances in soft-sensor methodology that improve temporal feature learning and drift handling (Stanišić et al. 2024; Viitala et al. 2025; Zhang et al. 2021, 2025). In line with this direction, predictive modeling of basis weight in the wire section can support early-stage control, reducing start-up losses and improving process stability (Sonsale et al. 2023; Zhang et al. 2021). Against this backdrop, our study develops a dynamically weighted hybrid soft sensor that fuses a simplified first-principles mass-balance model with a temporal Conv1D learner; a GRU-based component adapts the FPM-ML weighting online to remain robust during start-ups and regime shifts. In our experiments, this hybrid reduced basis-weight RMSE by more than 40 % versus an FPM-only baseline, enabling faster convergence to stable operation and better transient performance.

This paper presents a hybrid modeling framework that integrates a simplified first-principle model with a machine learning model in a dynamically weighted parallel architecture. The first-principle component is based on mass balance equations and implemented in MATLAB/Simulink, while the data-driven model consists of a one-dimensional convolutional neural network (Conv1D) trained on industrial process data. A third model – a gated recurrent unit (GRU) network – dynamically computes the weighting factor between the physical and data-driven predictions, enabling adaptive model fusion based on recent performance. The framework is evaluated using real production data from a tissue paper machine, demonstrating that the hybrid model outperforms both individual approaches in predicting basis weight.

2 Background and related work

2.1 Papermaking and the role of the wire section

Papermaking is a continuous industrial process in which a dilute pulp suspension is progressively dewatered, formed, and dried into a continuous web of paper. A typical paper machine consists of several main sections: the headbox, wire

section, press section, dryer section, creping unit (in tissue production), and the reel (Gullichsen and Paulapuro 1998). Among these, the wire (forming) section governs the initial development of key sheet characteristics – most notably *basis weight* and formation – that directly drive downstream quality, runnability, and energy efficiency. Recent industrial guidance and modeling studies emphasize that early dewatering dynamics and wet-end stability are decisive levers for overall performance (ABB Ltd 2025; Sjöstrand 2025; Sjöstrand and Bergström 2024; Valmet 2025). In the wire section, the stock delivered by the headbox is uniformly distributed onto a continuously moving wire, where rapid water removal occurs via gravitational drainage, foils, vacuum boxes, and suction elements. Contemporary models capture these mechanisms at engineering scale (e.g., compact relations for vacuum dewatering and TAD solids development), enabling their integration with estimation and control frameworks (Sjöstrand 2025; Sjöstrand and Bergström 2024). The process simultaneously sets fiber orientation and initial retention, both sensitive to mechanics (jet-to-wire, turbulence, vacuum loading) and chemistry (retention aids, charge/ash balance), which are monitored and stabilized in practice through retention and ash consistency measurement and control (ABB Ltd 2025; Valmet 2025). Modeling this section remains challenging due to limited direct sensors in the forming zone, strong multi-phase interactions, and variability in furnish properties. Consequently, recent work increasingly combines physics-based relations with data-driven soft sensors to improve early quality observability and robustness under transients, leveraging temporal deep learning and attention mechanisms (Stanišić et al. 2024; Zhang et al. 2025). Tissue-focused machine learning studies further demonstrate how such hybrid and ML approaches can translate into tangible efficiency gains at machine level (Viitala et al. 2025; Zhang et al. 2021).

In this work, real industrial data from a tissue manufacturing plant was used to build and validate models for predicting basis weight in the wire section. The machine is equipped with a twin-wire former and operates without external vacuum during dewatering. The headbox delivers thick stock at a flow rate of 100,000 L/min, and the consistency of the suspension leaving the headbox ranges from 0.1 % to 1 %. After the wire section, the paper passes through a pressing section and a Yankee dryer, with the basis weight measured at the reel.

The available dataset spans seven months of production and includes five key process variables originally recorded at one-second intervals, but aggregated to 30 min intervals for analysis. Among them, the basis weight at the reel serves as the prediction target, while thick stock flow, fines content,

moisture content, and wire speed act as key input variables. Due to measurement constraints and noise, the raw data underwent extensive cleaning using thresholding and median filtering methods.

Figure 1 illustrates the layout of a typical paper machine, with emphasis on the wire section and key dewatering components.

2.2 First-principle modeling in papermaking

First-principle models describe process behavior based on fundamental conservation laws and physical principles. In the wire section, most FPMs use simplified mass balance equations to model water removal and estimate basis weight development over time (Cho 2005). These models are valued for their interpretability and alignment with physical reality, making them attractive for control and simulation tasks. However, their predictive accuracy is often limited by simplifying assumptions such as steady-state conditions, constant material properties, or ideal drainage behavior (Lennartsson et al. 2013). As a result, FPMs may struggle to capture nonlinear dynamics and variations caused by furnish composition or operational transients.

In the wire section of a paper machine, water removal and basis weight development are primarily governed by the dynamics of flow rate and consistency. The basis weight BW (in g/m²) can be approximated by:

$$BW = \frac{\dot{m}_s \cdot (1 - SR)}{v \cdot w \cdot \epsilon} \quad (1)$$

Here, $\dot{m}_s$ is the solid mass flow rate entering the press section, SR is the shrinkage rate of the wet web, v is the

machine speed, w is the web width, and ϵ is a lumped efficiency factor accounting for retention and sheet consolidation.

The underlying mass and consistency balances are derived from the dynamic model proposed by (Cho 2005), they developed a comprehensive retention and dewatering model based on pilot paper machine data. The key equations used in our adapted model are summarized in Table 1.

To apply this model to a real industrial setting, modifications were necessary. Our plant features a twin-wire roll former operating without external vacuum, no filler or retention aid dosage, and a significantly higher machine speed (2,000 m/min vs. 40 m/min). Therefore, delay elements were removed, empirical retention functions simplified, and machine-specific constants updated to reflect current operating conditions. These changes preserve the core mass balance structure while increasing applicability to real-time industrial data.

Table 1: Selected mass and consistency balance equations adapted from (Cho 2005).

Variable	Equation
Headbox consistency c_{hb}	$c_{hb} = \frac{\dot{m}_{hb}}{V_{hb}}$
Thick stock consistency c_{ts}	$c_{ts} = \frac{\dot{m}_{ts}}{V_{ts}}$
White water consistency c_{ww}	$c_{ww} = \frac{\dot{m}_{ww}}{V_{ww}}$
First-pass retention FPR	$FPR = \frac{\dot{m}_{hb} - \dot{m}_{ww}}{\dot{m}_{hb}}$
Basis weight BW	$BW = \frac{\dot{m}_{ret}}{v \cdot w}$
Retained mass flow $\dot{m}_{ret}$	$\dot{m}_{ret} = \dot{m}_{hb} \cdot FPR$

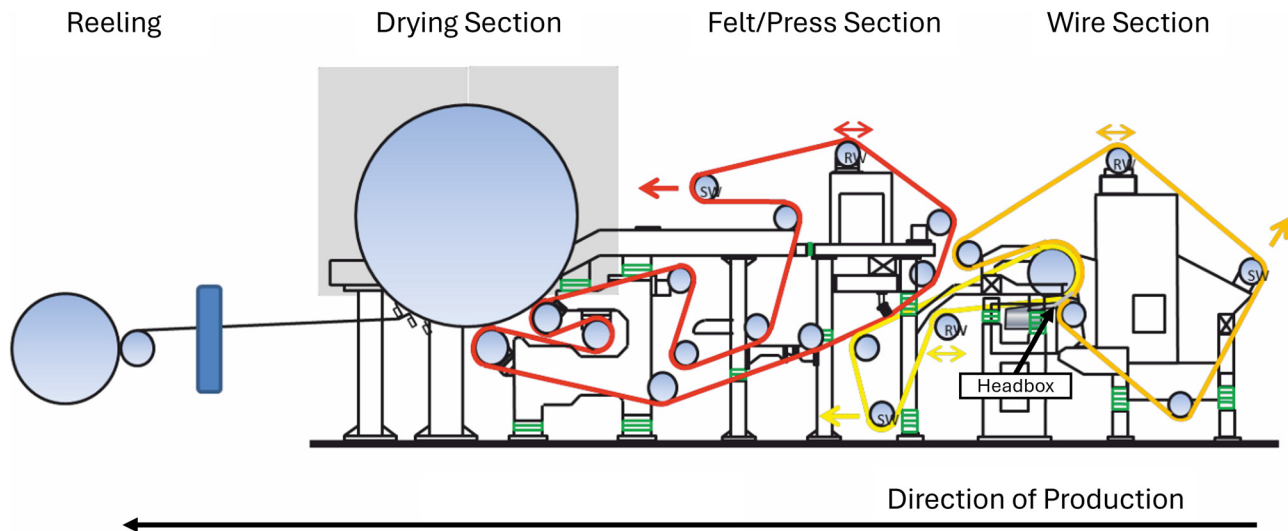


Figure 1: Schematic overview of a tissue machine highlighting the wire section where basis weight and formation are initially developed.

2.3 Data-driven modeling approaches

Data-driven modeling, particularly with machine learning, has gained momentum for complex process modeling in papermaking. With the availability of high-resolution sensor data, ML techniques such as decision trees and ensemble methods can infer relationships between process variables and product properties without requiring explicit physical models. Gradient boosting approaches like XGBoost are particularly effective in industrial settings due to their ability to handle nonlinearities, missing data, and variable importance evaluation (Chen and Guestrin 2016).

Nevertheless, these models often lack physical interpretability and require large amounts of representative training data. Their reliability deteriorates when extrapolating beyond trained conditions or under sensor noise and drift, limiting trust in safety-critical applications.

2.4 Hybrid modeling approaches

Hybrid modeling approaches aim to combine the interpretability of FPMs with the flexibility and accuracy of ML. In papermaking and other process industries, such models are used when partial physical knowledge is available but insufficient to fully describe system dynamics (Hotvedt et al. 2021). Hybrid models are particularly beneficial in scenarios with nonlinear behaviors, unmeasured disturbances, or incomplete instrumentation.

Common hybrid model structures include:

- **Parallel hybrid models:** The FPM and ML models run independently on the same inputs, and their outputs are fused via a fixed or dynamically adjusted weighting mechanism (Psychogios and Ungar 1992; Willard et al. 2022). This structure enables balancing physical consistency and data-driven correction under variable conditions.
- **Series hybrid models:** One model (typically the FPM) provides intermediate predictions which serve as additional input features for the ML model (Sharma and Liu 2021). This cascaded approach lets the ML model learn context-aware corrections based on physical predictions.
- **Residual hybrid models:** The ML model is trained on the residual error between the FPM output and the actual measurement. It acts as a corrective block to compensate for model mismatch or missing physics (Chaffart and Ricardez-Sandoval 2019).

Figures 2–4 illustrate these hybrid modeling architectures.

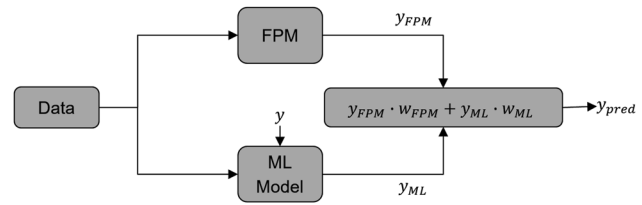


Figure 2: Parallel hybrid model: the outputs of FPM and ML are combined via a (possibly learned) weighting function.

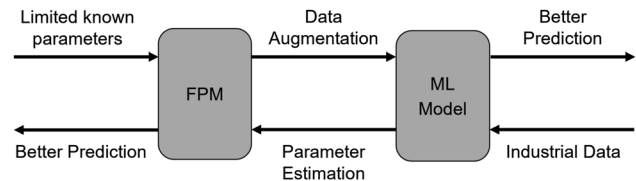


Figure 3: Series hybrid model: the output of the FPM is passed as input to the ML model.

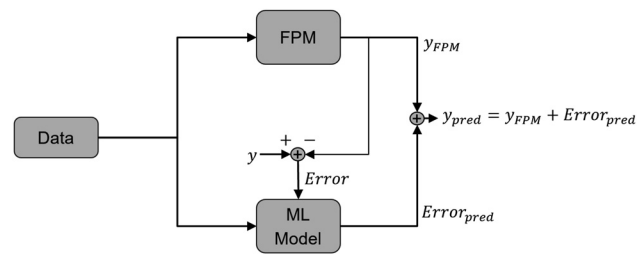


Figure 4: Residual hybrid model: the ML model learns the error between FPM predictions and real output to provide a corrective term.

In this work, a *parallel hybrid model* is developed to predict basis weight in the wire section. This approach was chosen for several reasons: First, the parallel hybrid architecture offers high flexibility by allowing both the first-principles model (FPM) and the data-driven model (ML) to independently contribute to the prediction. Unlike residual models, which require the FPM to closely approximate the target signal to avoid biasing the residual, or series models, which introduce dependency and error propagation between models, the parallel structure enables robust fusion without imposing strict constraints on either component. Second, the weighting mechanism – realized via an auxiliary ML model – enables dynamic adaptation to changing operating conditions and sensor drift. This is particularly important in papermaking processes where regime shifts (e.g., grade changes, startup phases) can significantly affect model reliability. By adjusting the contribution of FPM and ML based on recent performance, the hybrid model maintains both physical consistency and predictive accuracy across

varying process regimes. A simplified mass balance model is combined with a temporal convolutional neural network (Conv1D) trained on industrial data. An auxiliary recurrent neural network dynamically computes the weighting factor between the two models' outputs, allowing the hybrid model to adapt to changing process conditions while maintaining physical consistency (Rudolph et al. 2024).

3 Materials and methods

This section describes the data preprocessing pipeline, the development of the first-principles model, the design of the machine learning architecture, and their integration into a hybrid modeling framework.

3.1 Industrial dataset and preprocessing

The dataset was collected from a tissue production facility operating a twin-wire former. Five key process variables (PVs) were recorded over a period of seven months at 30-min intervals: basis weight, thick stock flow, fines content, moisture content, and wire speed.

To ensure data reliability, the following preprocessing steps were applied:

- **Threshold filtering:** Removal of measurements outside physically plausible bounds.
- **Time alignment:** Compensation for seasonal time shifts and ensuring uniform sampling.

As the dataset represents a chronological time series, random shuffling was avoided to preserve temporal dependencies. Instead, the dataset was split sequentially in a hindcasting-like manner: the first 40 % were used for training, followed by 20 % for validation, 20 % for weight-model training, and the final 20 % for testing. This setup ensures that each model is evaluated only on future, unseen process behavior, reflecting realistic deployment scenarios.

3.2 First-principle modeling approach

The FPM was adapted from the wire section model originally proposed (Cho 2005). It is based on mass conservation principles and simulates short circulation dynamics, consistency evolution, and initial basis weight formation. The model was implemented in MATLAB/Simulink using a block-diagram structure.

Several simplifications were made to adapt the model to a modern hydraulic headbox configuration:

- No headbox volume (ideal hydraulic behavior)
- Absence of filler or retention aid dosing
- Updated operating parameters for flow rate and machine speed

These modifications improved model generality and enabled realistic simulation using actual plant input data.

3.3 Machine learning architecture

To model the nonlinear relationship between process variables and basis weight, temporal neural network architectures capable of capturing sequential dependencies and dynamic process behavior were investigated. The motivation for this choice lies in the temporal nature of the process: basis weight at the reel is influenced by upstream conditions with time delays and smoothing effects.

Three neural network architectures were considered:

- **Conv1D:** Captures local temporal patterns using convolutional filters.
- **Long Short-Term Memory (LSTM):** Recurrent network suitable for modeling long-term dependencies.
- **Gated Recurrent Unit (GRU):** A simplified recurrent model with fewer parameters than LSTM.

To justify the architecture selection, a structured benchmarking study was conducted. A hyperparameter search over model depth, number of units, window size, and learning rate was performed using a combination of grid search and automated tuning with Keras Tuner. Each model was trained and validated using the same training pipeline and input features, with early stopping based on validation RMSE.

The Conv1D model consistently outperformed LSTM and GRU in both training time and predictive accuracy on the validation set. This is attributed to the relatively short effective memory required for the prediction task (less than 4 h) and the stationary structure of the input data after preprocessing.

The final configuration for the selected architecture was:

- **Model:** 1D Convolutional Neural Network (Conv1D)
- **Window size:** 7 time steps (equivalent to 3.5 h)
- **Number of filters:** 64
- **Kernel size:** 3
- **Learning rate:** 0.001
- **Loss function:** Mean Squared Error (MSE)

Model performance was evaluated based on Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) on the

validation set. The Conv1D model demonstrated better generalization and robustness under regime shifts compared to the recurrent models.

3.4 Hybrid model integration

To fuse the predictions from the FPM and ML models, a weighted parallel model was implemented. The hybrid prediction is computed as:

$$\hat{y}_{\text{hybrid}} = w_{\text{fpm}} \cdot \hat{y}_{\text{fpm}} + (1 - w_{\text{fpm}}) \cdot \hat{y}_{\text{ml}} \quad (2)$$

Two strategies were used to define the weight w_{fpm} :

- **Static weighting:** A fixed weight computed to minimize RMSE over a tuning dataset.
- **Dynamic weighting:** A secondary ML model was trained to predict the optimal weight at each time step, using the recent history of FPM and ML outputs as input features.

Dynamic weighting was realized using a secondary GRU model. The GRU weighting model was implemented using a stacked architecture. It consisted of two GRU layers with 64 units each. Between the recurrent layers, a dropout rate of 0.2 was applied to reduce overfitting. After the recurrent blocks, the network included two fully connected layers (64 units, ReLU activation and 8 units, ReLU activation) followed by a final dense layer with linear activation for the output weight prediction. The model was trained with the Adam optimizer using learning rates of 0.001 and 0.0001, with mean squared error as the loss function. Early stopping (patience = 10) and ReduceLROnPlateau callbacks were employed to prevent overfitting and adapt learning rates during training. The input was based on 50-step time windows including recent feature and target histories, and the dataset was split chronologically into 40 % training, 10 % validation, and 50 % testing.

The GRU receives as input:

- recent prediction errors (residuals) of FPM and ML,
- model confidence indicators (e.g., moving average of residuals),
- current values of key process variables (PV2–PV5).

This enables the system to adjust the weight w_{fpm} at every time step based on the local reliability of each model. The weight model is trained to minimize the hybrid RMSE on a separate validation set. This improves adaptivity under regime shifts and transient startup behavior.

The dynamic weighting approach yielded the best performance, particularly in scenarios with transient process behavior or regime shifts.

4 Results and discussion

The performance of all models was evaluated using RMSE and MAE, computed on normalized basis weight predictions. All outputs were scaled using Min-Max normalization to enable direct comparison.

4.1 First-principle model performance

The FPM was evaluated using normalized industrial data for the target variable (basis weight). While the FPM captured key trends and general system dynamics, its accuracy was limited by structural simplifications and fixed parameter settings (see Figure 5).

4.2 Machine learning model performance

The Conv1D model outperformed the FPM in both RMSE and MAE, producing smoother and more responsive predictions. However, its performance deteriorated in high-variance regions and during rapid transitions, suggesting challenges in capturing short-term dynamics (see Figure 6).

4.3 Hybrid model with static weighting

Two static weighting strategies were evaluated:

- **Per-row selection:** At each data point, the model (FPM or ML) with the lower squared error was selected.
- **RMSE minimization:** A fixed weight was tuned to minimize RMSE over the validation set.

Both strategies yielded modest improvements over the individual models, with RMSE-based weighting performing slightly better (see Figure 7).

4.4 Hybrid model with dynamic weighting – sliding window

A dynamic hybrid model using a sliding window to compute local performance-based weights outperformed static strategies. The optimal window size was selected based on a grid search using the designated validation set for the weighting model (20 % of the data). As the data represent a time series, the split was performed chronologically to preserve temporal dependencies and avoid data leakage. Randomized cross-validation was not applied, as it is unsuitable for time-dependent datasets (see Figure 8).

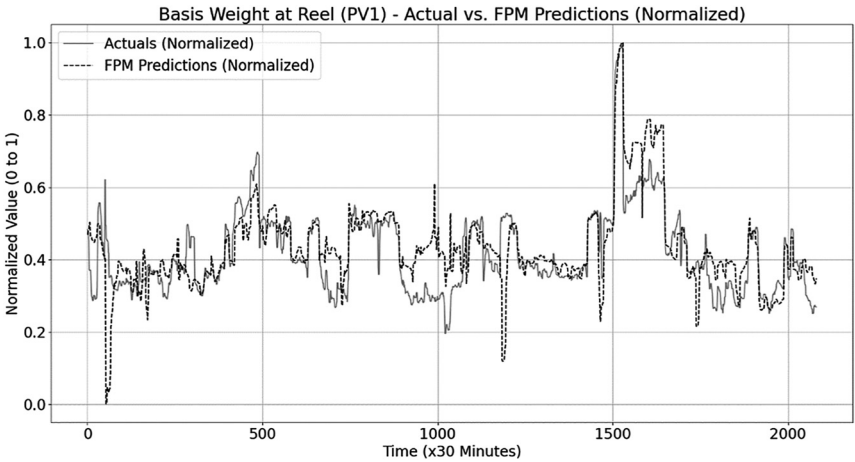


Figure 5: Normalized comparison of industrial data versus FPM prediction for basis weight.

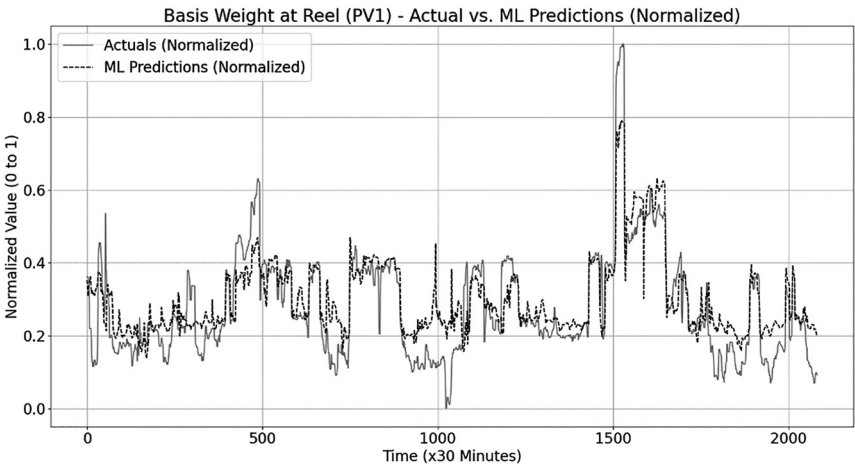


Figure 6: Normalized comparison of industrial data versus ML prediction (Conv1D) for basis weight.

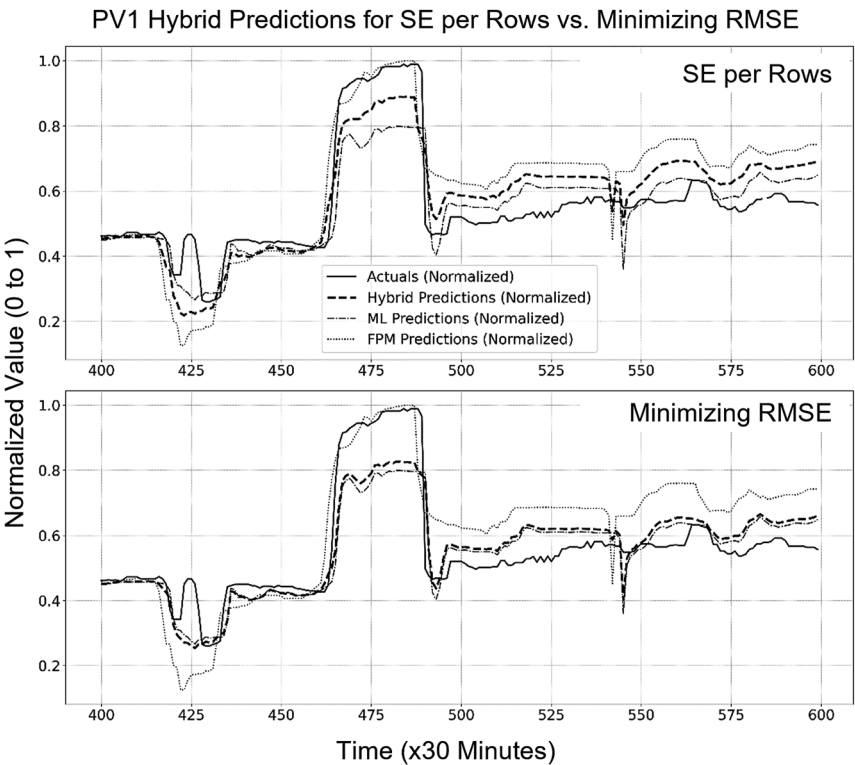


Figure 7: Normalized comparison of industrial data, static hybrid predictions (per-row and RMSE-based), ML predictions, and FPM predictions for basis weight.

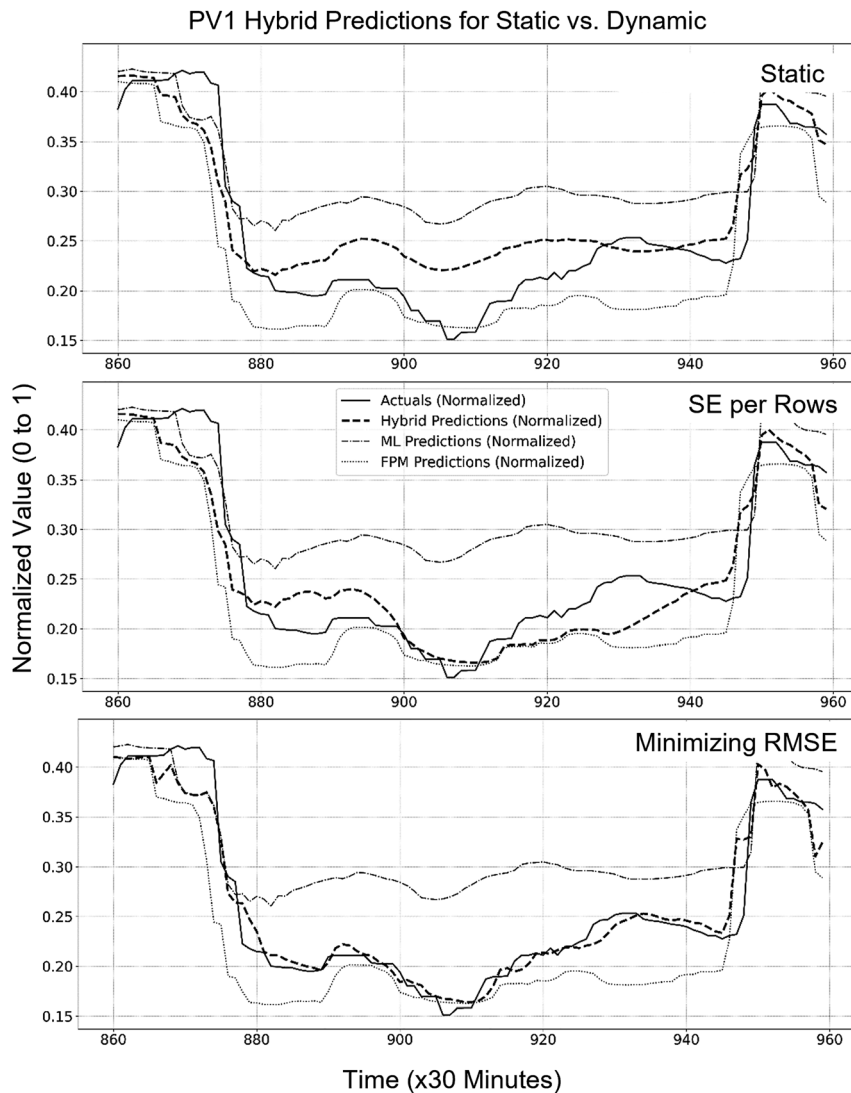


Figure 8: Comparison of static and dynamic sliding window hybrid predictions for basis weight.

4.5 Hybrid model with dynamic weighting – additional ML model

The highest-performing model integrated a secondary ML model (GRU) to dynamically predict the weighting between FPM and ML predictions based on recent performance history. This model provided robust and adaptive behavior across various operating conditions (see Figure 9).

4.6 Evaluation summary

The evaluation of different modeling strategies highlights the trade-offs between physical interpretability and predictive accuracy. The first-principles model (FPM) provides transparent results but suffers from structural simplifications that limit its precision. The data-driven Conv1D model, by contrast, delivers more accurate predictions but lacks

physical consistency and shows reduced robustness under regime shifts. Hybrid approaches combine the complementary strengths of both paradigms. Static weighting improves prediction accuracy moderately, while dynamic weighting enables adaptation to transient and non-stationary conditions. The highest-performing configuration is the hybrid model with ML-based dynamic weighting, which integrates a GRU-based mechanism to dynamically adjust the contributions of FPM and ML outputs.

Table 2 summarizes the performance of all tested models in terms of both normalized error metrics and absolute values expressed in g/m^2 . While normalized values facilitate direct comparison across modeling strategies, absolute metrics are more relevant for practical industrial assessment. The ML-based hybrid model achieves the lowest error with an RMSE of 2.12 g/m^2 and MAE of 1.53 g/m^2 , significantly outperforming both the standalone FPM (RMSE = 3.51 g/m^2) and Conv1D ML model (RMSE = 2.58 g/m^2). These

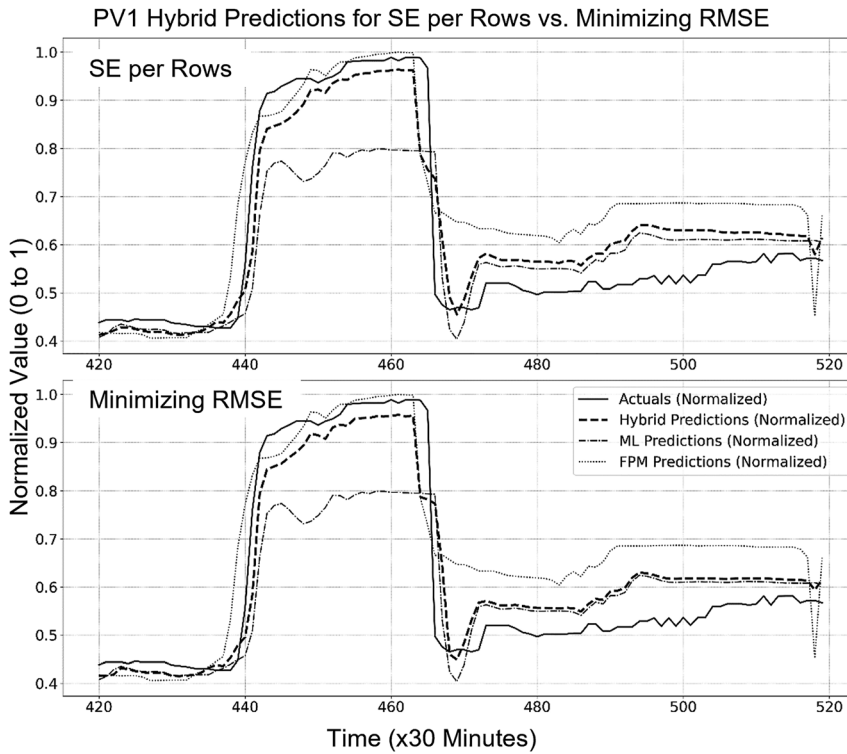


Figure 9: Comparison of hybrid predictions using ML-based dynamic weighting.

Table 2: Performance of different modeling approaches for basis weight prediction. Both normalized errors and absolute values (g/m^2) are reported.

Model	RMSE (normalized)	MAE (normalized)	RMSE (g/m^2)	MAE (g/m^2)
First-principles (FPM)	0.0911	0.0631	3.51	2.43
Machine learning (Conv1D)	0.0669	0.0507	2.58	1.95
Hybrid static (per- row)	0.0701	0.0528	2.70	2.04
Hybrid static (RMSE-based)	0.0662	0.0509	2.55	1.96
Hybrid dynamic (sliding window)	0.0641	0.0473	2.47	1.82
Hybrid dynamic (ML-based)	0.0550	0.0398	2.12	1.53

results confirm that the hybrid framework combines physical consistency with adaptability, providing a scalable foundation for real-time quality monitoring and control in papermaking (see Table 2).

5 Conclusion and outlook

This study demonstrates the potential of hybrid modeling for accurate and interpretable prediction of basis weight in the

wire section of a paper machine. By integrating a first-principles model (FPM) with a data-driven machine learning (ML) model, the proposed approach leverages the complementary strengths of both modeling paradigms.

The FPM, implemented in MATLAB/Simulink, provided transparent predictions based on mass and retention balances but was limited by structural simplifications. In contrast, the Conv1D-based ML model exhibited superior predictive accuracy but lacked physical interpretability. The hybrid model, realized through a weighted parallel model, effectively combined the benefits of both.

Both static and dynamic weighting schemes were evaluated. The highest accuracy (RMSE = 0.0550) was achieved with dynamic weighting based on a secondary ML model, confirming the hybrid framework's adaptability under varying process conditions.

While the hybrid model improves predictive accuracy, interpretability remains an important consideration for industrial deployment. The first-principles component ensures some transparency, but the machine learning models – particularly the GRU-based weighting mechanism – still operate as non-interpretable modules. Future work will explore the use of explainability tools such as SHAP or other feature attribution techniques to better understand how process variables influence model behavior. Enhancing interpretability will be essential for operator trust, model validation, and robust decision-making in production environments. Future work may also explore the

transition from hybrid soft sensors to fully autonomous process control systems. As outlined by (Kins et al. 2025), hybrid modeling with embedded uncertainty quantification, adaptation, and explainability forms the basis for trustworthy autonomous operation in technical systems. In addition, future research should address robustness under sensor faults and measurement noise, for example through fault-tolerant training strategies or redundancy in sensor fusion.

Future work will explore:

- Expanding prediction scope to additional quality metrics (e.g., ash content, energy use).
- Generalization across multiple paper grades and machine configurations.
- Integration into real-time control systems, e.g., via OPC-UA or Asset Administration Shell (AAS).
- Incorporation of uncertainty-aware modeling approaches, such as Bayesian Neural Networks or Gaussian Process Regression (GPR) with first-principles models as priors.
- Exploration of physics-informed neural networks (PINNs) to embed domain knowledge directly into the learning process.
- Robustness enhancement under concept drift and dynamic process changes.
- Evaluation of explainability tools (e.g., SHAP, LIME) to increase model transparency and operator trust.

In summary, the results underline the value of hybrid modeling as a scalable, interpretable, and accurate tool for advancing digitalization and smart monitoring in the pulp and paper industry. This perspective aligns closely with recent initiatives such as the FOREST project (Modellfabrik Papier gGmbH 2025), which develops a modular digital twin framework for energy and CO₂ optimisation in fibre-based industries. Recent contributions highlight how simulation, interoperability standards (AAS, FMI, OPC UA) and data ecosystems can enable sustainable papermaking and cross-vendor digital twin architectures (Juhlin et al. 2025; Othen et al. 2025). The integration of hybrid modeling approaches into such reference architectures represents a promising step towards real-time sustainability management and carbon-neutral paper production.

Acknowledgments: This work was partially supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) within the scope of the FOREST project (03EN2095 A–G). The authors would especially like to thank WEPA Hygieneprodukte GmbH, Arnsberg, Germany, for providing the data used in this study. The support of Modellfabrik Papier gGmbH is also gratefully acknowledged,

particularly for their strategic guidance and infrastructural contributions to the FOREST project.

Research ethics: Not applicable.

Informed consent: Not applicable.

Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Use of Large Language Models, AI and Machine Learning Tools: ChatGPT to improve language.

Conflict of interest: The authors state no conflict of interest.

Research funding: German Federal Ministry for Economic Affairs and Climate Action (BMWK) within the scope of the FOREST project (03EN2095 A–G).

Data availability: The raw data can be obtained on request from the corresponding author.

References

- ABB Ltd (2025). *Paper machine wet end stability: the importance of accurate retention measurement*, Online, Available at: <https://new.abb.com/pulp-paper/abb-in-pulp-and-paper/articles/paper-machine-wet-end-stability-the-importance-of-accurate-retention-measurement>.
- Bikmukhametov, T. and Jäschke, J. (2020). Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. *Comput. Chem. Eng.* 138: 106834, <https://doi.org/10.1016/j.compchemeng.2020.106834>.
- Chaffart, D. and Ricardez-Sandoval, L.A. (2019). A hybrid modelling framework for improving soft sensor design and control performance in process systems. *J. Process Control* 84: 54–68, <https://doi.org/10.1016/j.jprocont.2019.11.012>.
- Chen, T. and Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 785–794.
- Cho, B.U. (2005). *Dynamics and control of retention and formation on a paper machine using a microparticulate retention aid system*, Ph.D. thesis. University of British Columbia.
- Gullichsen, J. and Paulapuro, H. (1998). *Papermaking part 1: stock preparation and wet end. Papermaking science and technology series*. Fapet Oy, Helsinki, Finland.
- Holik, H. (2006). *Handbook of paper and board*, 1st ed. Wiley VCH, Weinheim, Germany.
- Hotvedt, M., Grimstad, B., and Imsland, L. (2021). Identifiability and physical interpretability of hybrid, gray-box models – a case study. *IFAC-PapersOnLine* 54: 389–394, <https://doi.org/10.1016/j.ifacol.2021.08.273>.
- Juhlin, P., Schlake, J., Song, C., Zehnpfund, A., Grüner, S., Schmeiser, A., Kratzer, K., Othen, R., Sejdija, J., and Kayser, P. (2025). CRADLE: cross-vendor asset digital twin architecture using industrial interoperability standards with application to paper manufacturing. In: *30th international conference on emerging technologies and factory automation (ETFA)*. IEEE.
- Kins, R., Möbitz, C., and Gries, T. (2025). Towards autonomous learning and optimisation in textile production: data-driven simulation approach for optimiser validation. *J. Intell. Manuf.* 36: 3483–3508, <https://doi.org/10.1007/s10845-024-02405-3>.

- Lennartsson, M., Åström, K.J., and Wittenmark, B. (2013). Modeling and control of the wet end in paper machines: a survey. *Control Eng. Pract.* 21: 1653–1672, <https://doi.org/10.1016/j.conengprac.2013.07.009>.
- Modellfabrik Papier gGmbH (2025). *Our research project FOREST*, Available at: <https://modellfabrikpapier.de/en/forest-en/>.
- Othen, R., Pohlmeier, F., Song, C., Schlake, J., Sejdija, J., Schmeiser, A., Möbitz, C., and Gries, T. (2025). Sustainability management in fibre-based web production. In: *Proceedings of the conference on production systems and logistics: CPSL 2025*. Publish-Ing. Hannover: Publish-Ing, pp. 221–232.
- Pantelides, C.C. and Renfro, J.G. (2013). The online use of first-principles models in process operations: review, current status and future needs. *Comput. Chem. Eng.* 51: 136–148, <https://doi.org/10.1016/j.compchemeng.2012.07.008>.
- Psichogios, D.C. and Ungar, L.H. (1992). Hybrid neural network models for modeling and control of nonlinear processes. *AIChE J.* 38: 1499–1511, <https://doi.org/10.1002/aic.690381003>.
- Ramaswamy, S. (2003). Vacuum dewatering during paper manufacturing. *Dry. Technol.* 21: 685–717, <https://doi.org/10.1081/drt-120019058>.
- Rudolph, M., Kurz, S., and Rakitsch, B. (2024). Hybrid modeling design patterns. *J. Math. Ind.* 14: 1–25, <https://doi.org/10.1186/s13362-024-00141-0>.
- Sharma, N. and Liu, Y.A. (2021). A hybrid science-guided machine learning approach for modeling and optimizing chemical processes. *arXiv preprint arXiv:2112.01475*, <https://arxiv.org/abs/2112.01475>.
- Sjöstrand, B. (2025). Easy-to-use numerical models of water removal in vacuum dewatering and molding during through air drying of tissue paper. *Dry. Technol.* 43: 668–678, <https://doi.org/10.1080/07373937.2024.2449177>.
- Sjöstrand, B. and Bergström, V. (2024). Monitoring solids content development in pilot-scale through air drying of tissue paper. *Nord. Pulp Pap. Res. J.* 39: 127–137, <https://doi.org/10.1515/nprrj-2023-0092>.
- Sonsale, A.N., Yashpal, Pohekar, S., and Purohit, J. (2023). Drivers to energy efficiency measures in recycled paper and pulp industry in India: an interpretive structural modelling-based framework. *Sustain. Energy Technol. Assess.* 55: 102961, <https://doi.org/10.1016/j.seta.2022.102961>.
- Stanišić, D., Mejić, L., Jorgovanović, B., Ilić, V., and Jorgovanović, N. (2024). An algorithm for soft sensor development for a class of processes with distinct operating conditions. *Sensors (Basel, Switz.)* 24: 1948, <https://doi.org/10.3390/s24061948>.
- Thai, H.T. (2022). Machine learning for structural engineering: a state-of-the-art review. *Structures* 38: 448–491, <https://doi.org/10.1016/j.istruc.2022.02.007>.
- Valmet (2025). *Total and true ash consistency for a stable wet end*, Online, Available at: <https://www.valmet.com/automation/analyzers-measurements/analyzers/retention-measurement/>.
- Viitala, R., Miettinen, M., Marquez, R., Hämäläinen, A., Karhinen, A., Barrios, N., Gonzalez, R., Pal, L., Jameel, H., and Holmberg, K. (2025). Integration of artificial intelligence and sustainable energy management in the pulp and paper industry: a path to decarbonization. *Renew. Sustain. Energy Rev.* 218: 115809, <https://doi.org/10.1016/j.rser.2025.115809>.
- Willard, J., Jia, X., Xu, S., Steinbach, M., and Kumar, V. (2022). Integrating physics-based modeling and machine learning: a survey of methods and applications. *ACM Comput. Surv. (CSUR)* 55: 1–34.
- Zhang, H., Li, J., and Hong, M. (2021). Machine learning-based energy system model for tissue paper machines. *Processes* 9: 655, <https://doi.org/10.3390/pr9040655>.
- Zhang, L., Ren, G., Li, S., Du, J., Xu, D., and Li, Y. (2025). A novel soft sensor approach for industrial quality prediction based TCN with spatial and temporal attention. *Chemom. Intell. Lab. Syst.* 257: 105272, <https://doi.org/10.1016/j.chemolab.2024.105272>.