Process-Informed Dynamic Graph WaveNet for Industrial Sintering Processes Forecasting

Zhili Zhang[1], Liya Wang[1], Jie Li[2]
[1]*College of Science, North China University of Science and Technology, Tangshan, China*
[2]*College of Metallurgy and Energy, North China University of Science and Technology, Tangshan, China*

Abstract: Accurate multi-step forecasting of key state parameters is a critical prerequisite for optimizing energy consumption and reducing emissions in the sintering process, which is characterized by significant nonlinearity and large time-varying lags. However, applying general Spatiotemporal Graph Neural Networks faces a critical mismatch: unbounded dot-product attention lacks robustness to industrial noise, and purely data-driven structures often violate physical consistency. To address this, we propose a Process-Informed Dynamic Graph WaveNet. The model incorporates a Global Feature Broadcasting mechanism to integrate system-wide control variables. Crucially, it employs a Decoupled Two-Stream Topology combining a static adaptive graph initialized with physical time-delay priors and a noise-robust dynamic graph constructor using a bounded Radial Basis Function kernel. This architecture strictly constrains node connectivity to suppress interference while respecting material flows. Across two real-world sintering datasets, the proposed model attains the lowest prediction errors on the more difficult Sinter-B dataset and matches the performance of top-performing baselines on Sinter-A, highlighting its adaptability and reliability under varying industrial conditions.

Keywords: Industrial Internet of Things; Sintering Process; Spatiotemporal Graph Neural Networks; Time Series Forecasting

1. Introduction

The steel industry is facing dual challenges of energy consumption and emissions, with the sintering process accounting for approximately 10% to 20% of the total energy consumption in steel production and being the major contributor to SO_2 and NO_x emissions [1]. Sintering process exhibits significant nonlinearity, strong time-varying characteristics, and large lag properties, posing immense difficulties for its fine-grained regulation and emission control [2]. Therefore, constructing a high-precision spatiotemporal prediction model to achieve advance estimation of key emission and state parameters is a crucial technical prerequisite for optimizing production operations, reducing process fluctuations, and achieving emission reduction at the source [3].

Sintering process modeling has evolved from being mechanism-driven to data-driven, and recently to spatiotemporal graph neural networks. Although early mechanism models provided a solid theoretical framework based on thermodynamic and fluid mechanics principles, they were limited by complex chemical reactions and high computational costs, making it difficult to meet the real-time requirements of online control. With the rise of the Industrial Internet of Things, deep learning methods represented by CNNs and LSTMs have significantly improved prediction accuracy [4]. However, these models essentially treat multidimensional sensor signals as isolated time series or flat features, ignoring the physical coupling mechanism of sensors in spatial distribution [5]. STGNNs address this by modeling sensors as graph nodes and spatial correlations as edges, jointly capturing temporal dynamics and topological features [6]. This has not only achieved success in the field of traffic flow prediction [7] but also provided a new technical pathway for solving the modeling challenges of strongly physically coupled industrial systems such as sintering process [8].

The fundamental basis for the successful adaptation of STGNN from the transportation domain to the sintering process lies in the deep mathematical isomorphism between the two in terms of the overall system topology and fluid propagation dynamics. In the spatial topological dimension, the sintering system exhibits a complex non-Euclidean structure that couples the "linear process backbone" with the

"networked flue gas circuit" [9]: there is both a linear flow of materials from batching to the machine tail and a physical "loop" formed by the confluence of large flues and circulating flues, which is highly consistent with the main road + interchange loop topology in transportation networks. In the dynamic mechanism dimension, both systems evolve according to a similar convection diffusion mechanism [10]. In sintering process, disturbances originating from upstream operations like mixing and feeding, much like upstream traffic in transportation networks, affect downstream thermodynamic states only after propagating through significant physical time lags. This consistency in mathematical description of "upstream disturbance $\rightarrow$ transmission lag $\rightarrow$ downstream response", combined with the spatial feedback mechanism brought about by flue gas circulation, provides a solid theoretical basis for transferring mature architectures such as DCRNN to sintering process.

However, when the STGNN paradigm in the general field is adapting to the sintering industrial scenario, it faces a mismatch between mathematical mechanism and data distribution. Current adaptive graph methods have two limitations: Firstly, dynamic graph learning models based on dot-product attention mechanisms lack robustness to industrial noise. Mainstream methods represented by ASTGCN and GMAN [11] widely adopt dot-product operations between query vectors and key vectors to capture node dependencies in real time. From a mathematical perspective, dot-product operations are unbounded, and their output is highly sensitive to the amplitude of input features. This is acceptable in smooth traffic data, but in sintering data accompanied by high-frequency electromagnetic interference and signal drift, this mechanism is prone to misjudging large-amplitude random noise as strong correlation, inducing the model to learn false topological connections at each level, resulting in significant divergence in prediction performance under non-stationary operating conditions. Secondly, pure data-driven dynamic graph construction lacks physical consistency constraints. Graph structures generated solely based on statistical correlation often ignore the inherent causal logic and time-delay characteristics of material transport in the sintering process. Under closed-loop control and strong coupling, pure data-driven models are prone to falling into the spurious correlation trap [12], generating topological relationships that are inverted causality and violate process mechanisms, severely limiting the interpretability and generalization ability of the model in industrial fields.

To address these challenges, this study proposes a physics-informed spatiotemporal prediction model named Dynamic Graph WaveNet. We introduce a Global Feature Broadcasting mechanism that broadcasts global scalar variables—such as machine speed—to all spatial locations, ensuring that local process dynamics are explicitly informed by the system-wide operational context. Crucially, we design a Decoupled Two-Stream Topology architecture to resolve the conflict between stability and adaptability. This approach includes a static adaptive graph initialized with physical time-delay priors to respect fundamental material flows, and a Dynamic Graph Constructor based on noise-robust RBF kernels that replaces unbounded attention scores with bounded similarity measures to strictly constrain node connectivity and suppress industrial sensor noise and process disturbances. Experimental results on two industrial datasets demonstrate that our framework significantly outperforms baselines, particularly in high-variance scenarios, by effectively decoupling stable physical structures from instantaneous thermodynamic fluctuations.

2. Dynamic Graph WaveNet

2.1 Problem Definition

The underlying topology of the sintering system is modeled as a graph $\mathrm{G}=(\mathrm{V},\mathrm{E})$, where V represents the set of N system nodes. These nodes encompass the full range of process variables, including state values collected by sensors and control setpoints originating from the Distributed Control System (DCS). We define the historical observation tensor as $\mathrm{X} \in \mathbb{R}^{T_P \times N \times C}$, where T_P, N, and C respectively denote the backtracking window length, the total number of nodes, and the feature dimension. Crucially, the feature dimension C integrates both original local sensor measurements and globally broadcasted driving factors to ensure each node explicitly perceives the overall process context. The main objective is to learn a parametric mapping function f_θ to forecast the

future system state Y based on the historical observations and the system topology G :

$$Y = f_\theta(X, G) \quad (1)$$

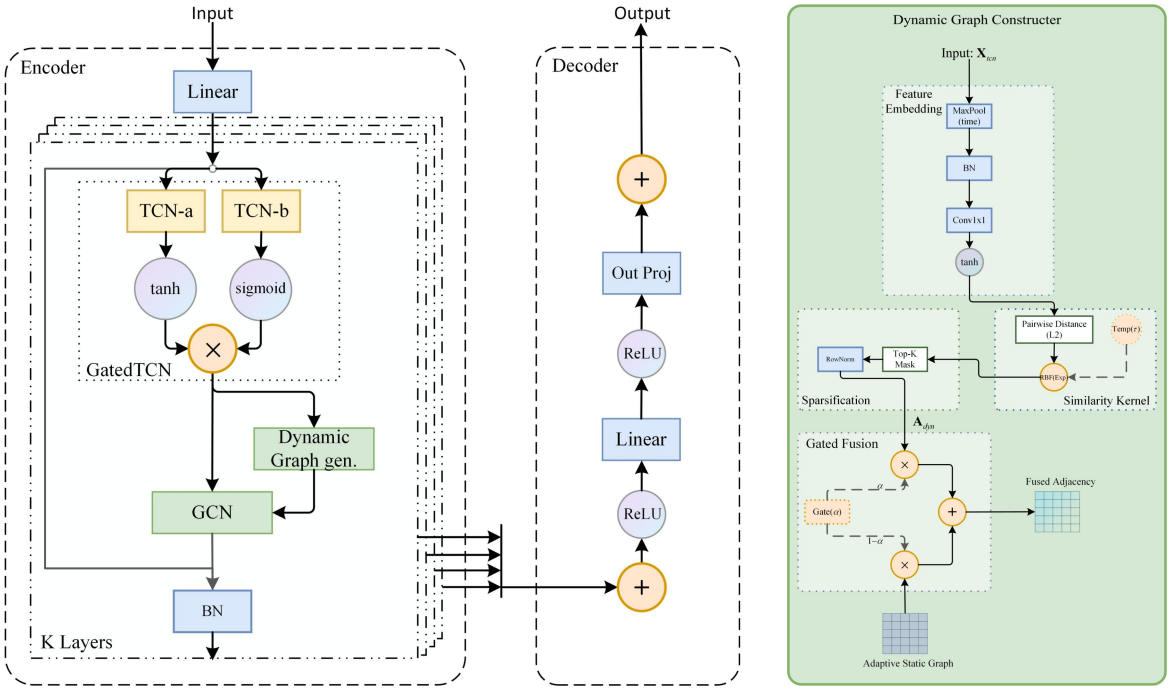
Here, $Y \in \mathbb{R}^{T_Q \times N}$ represents the predicted system states over the prediction horizon T_Q , and f_θ denotes the deep neural network parameterized by θ . The model aims to capture complex spatiotemporal dependencies and non-stationary evolution patterns within the historical data to achieve precise multi-step forecasting of sintering process.

2.2 Overall Framework

Building on the Graph WaveNet architecture, the proposed Dynamic Graph Network employs an encoder-decoder framework designed to jointly model multiscale temporal patterns and decoupled spatial dependencies. The model first processes the input history data through an Input Projection Layer, where a learnable linear transformation projects the low-dimensional raw input features into a high-dimensional hidden space to generate unified feature embeddings for the subsequent encoding stages.

The core encoder comprises L stacked Spatiotemporal Blocks. Within each block, the feature flow is processed sequentially: first, Gated Dilated Causal Convolutions capture multi-scale temporal patterns via exponentially expanding receptive fields. These temporal features are then spatially aggregated by the Decoupled Two-Stream Graph Module, where Diffusion Graph Convolutions operate on a noise-robust topology generated by the Dynamic Graph Constructor to model the system's underlying flow field dynamics.

In the output stage, the model utilizes a skip connection architecture to synthesize hierarchical representations from varying network depths. The intermediate features from the skip-connection layers of all stacked blocks are aggregated via element-wise summation to form a unified feature map. To generate the final multi-step predictions, the fused representation is passed through a nonlinear regression head consisting of convolutional layers separated by ReLU activations, effectively transforming high-dimensional latent features into the output sequence. The integrated structure of the proposed framework is illustrated in Figure 1, where Figure 1a presents the main pipeline and Figure 1b details the Soft Dynamic Graph Constructor.



(a) Overall Architecture **(b) Dynamic Graph Constructor**

Figure 1. Overall Architecture of the Model

2.3. Decoupled Two-Stream Topology

The model addresses spatial dependencies through two distinct components: the fixed physical topology and dynamic state variations. A Decoupled Two-Stream Topology Learning mechanism is employed to model these static and dynamic correlations separately.

The static branch follows the self-adaptive adjacency matrix formulation of Graph WaveNet. It generates a global graph structure $\mathbf{A}_{static}$ using two learnable embedding matrices, $\mathbf{E}_1$ and $\mathbf{E}_2$, computed as:

$$\mathbf{A}_{static} = \text{Softmax}(\text{ReLU}(\mathbf{E}_1\mathbf{E}_2^T)) \quad (2)$$

To integrate physical priors, the node embeddings are initialized from a predefined matrix that encodes the mechanical connectivity of the sintering system. This design anchors the learned graph structure in real-world system constraints, establishing a physically plausible foundation for message passing.

In parallel, an RBF-Based Dynamic Graph Constructor is designed to capture time-dependent correlations. To mitigate the sensitivity of unbounded dot-product metrics to noise, this module utilizes a bounded Radial Basis Function kernel:

$$\mathbf{S}_{ij} = \exp\left(-\frac{\| \mathbf{v}_i - \mathbf{v}_j \|^2}{\tau}\right) \quad (3)$$

Here, τ is a learnable parameter controlling the kernel bandwidth. This formulation restricts similarity scores to the $[0,1]$ interval. Top-K sparsification is subsequently applied to remove low-weight connections associated with noise. Finally, the static and dynamic graphs are integrated via a learnable gating parameter to form the input for the spatial convolution layer.

$$\mathbf{A}_{final} = \alpha\mathbf{A}_{dyn} + (1-\alpha)\mathbf{A}_{static} \quad (4)$$

3. Experimental Results and Analysis

3.1 Datasets and preprocessing

We utilize two real-world datasets, designated as Sinter-A and Sinter-B, acquired from the DCS of two distinct sintering machines at a large-scale steel enterprise. Both datasets cover the complete production workflow, spanning from batching and sintering to flue gas circulation, with a 5-minute sampling interval. They encompass critical state variables such as feed rates, flow rates, temperatures, pressures, and exhaust compositions, alongside control setpoints. Sinter-A (Standard Scenario): Collected from Machine #1 over a period of 54 days, this dataset models the system with 183 nodes and 1,182 edges. It represents a standard industrial environment with a basic sensor layout and concentrated numerical distribution, serving as a stable benchmark for prediction accuracy. Sinter-B (High-Variance Scenario) was collected from the larger-scale Machine #2 over 52 days, comprising 221 nodes and 1,047 edges. Distinguished by dense flow monitoring in the return air branch, Sinter-B exhibits significant fluctuations and high-amplitude noise due to complex airflow dynamics. This scenario specifically evaluates the model's robustness under non-stationary and high-variance conditions.

To construct the model inputs, we first augmented the raw sensor data by broadcasting key global control variables—such as total feed rate and sintering machine speed—and concatenating them with the local features of each node, thereby ensuring the integration of system-wide context into local dynamics. Subsequently, a sliding window approach was applied to generate training samples, where a window slides step-by-step to extract historical sequences of length P as inputs to predict the ground truth values over the future horizon Q.

To explicitly model material flow and transmission relationships, a static adjacency matrix was pre-calculated based on the physical topology. We quantified the influence between nodes by calculating the generalized process lag d_{ij} which combine physical distance and transmission time and constructed the weighted graph using a Gaussian kernel function:

$$A_{ij} = \begin{cases} \exp(-\frac{d_{ij}^2}{\sigma^2}), & \text{if } A_{ij} \geq ò \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where $\sigma = 250$ is the attenuation coefficient and $ò = 0.1$ ϵ is the sparsity threshold. Self-loops were enforced to preserve each node's historical feature evolution.

3.2 Performance Comparison

Table 2 illustrates distinct performance patterns across the two datasets. On the stable Sinter-A dataset, models utilizing static adaptive graphs (specifically GWNet and AGCRN) yield the best performance. In contrast, DGCRN, which relies on a recursive mechanism to generate dynamic graphs at each time step, exhibits suboptimal results. This comparison between AGCRN and DGCRN—both recursive models—isolates the impact of topology: under steady-state conditions, the forced step-wise generation in DGCRN introduces unnecessary stochastic variance, destabilizing the dominant physical

structure that static graphs successfully capture.

In the high-variance Sinter-B scenario, the performance ranking provides deeper insights. Our model achieves state-of-the-art results, followed by GWNet, while both AGCRN and DGCRN suffer significant degradation. Notably, GWNet outperforms DGCRN, confirming that a stable static anchor offers greater robustness against industrial noise than a purely dynamic, step-wise generation mechanism, which is prone to instability. Our model synthesizes these advantages by explicitly decoupling topological learning: it preserves the robustness of the static anchor (like GWNet) while employing an RBF-driven branch to capture instantaneous drifts. This design effectively avoids the instability of DGCRN's step-wise generation while compensating for the rigidity of static graphs, thereby achieving the optimal trade-off between robustness and flexibility.

Table 1. Experimental Results of the Proposed Model And Baseline Model.

Dataset	Sinter-A				Sinter-B			
Model	MAE	RMSE	MAPE	WAPE	MAE	RMSE	MAPE	WAPE
LSTM	15.994	61.581	4.750%	3.658%	138.293	1023.685	9.814%	**2.622%**
Transformer [13]	16.115	62.323	4.752%	3.687%	143.184	1065.538	10.183%	2.656%
DLinear [14]	16.226	62.723	4.776%	3.715%	142.844	1007.854	9.864%	2.787%
AGCRN [15]	15.755	**60.547**	**4.381%**	3.605%	144.254	1019.581	9.980%	2.869%
DGCRN [16]	16.745	65.019	4.889%	3.828%	141.603	1021.329	10.060%	2.808%
GWNet [17]	**15.737**	60.714	4.663%	**3.601%**	135.505	995.881	9.792%	2.696%
Ours	15.844	61.314	4.627%	3.626%	**135.223**	**971.472**	**9.682%**	2.679%

3.3 Ablation Study

To evaluate the specific contribution of each topological component, we conducted an ablation study on the Sinter-A and Sinter-B datasets. As summarized in Table 3, the complete Decoupled Two-Stream model consistently achieves the optimal trade-off between stability and flexibility.1. In the standard Sinter-A environment, the performance gains from learnable topologies are marginal. The "Physical Topology" baseline already achieves high precision (MAPE: 4.651%), and the introduction of the Static Adaptive Graph yields negligible improvement in MAE ($15.888 \rightarrow 15.878$). This indicates that under steady-state conditions, the predefined physical connections sufficiently capture the primary spatial dependencies. Consequently, the learnable static module saturates quickly, offering limited optimization space over the robust physical priors. In contrast, the limitations of fixed topologies become evident on the volatile Sinter-B dataset. The Static Adaptive Graph reduces the MAE from 142.451 to 140.956, confirming that learnable embeddings successfully capture latent long-term correlations missed by the rigid physical layout. Crucially, the integration of the Dynamic Graph delivers the most substantial performance boost, reducing the RMSE from 990.425 to 971.472. This significant drop in RMSE—a metric sensitive to large errors—demonstrates that the dynamic branch effectively captures instantaneous "operating condition drifts" and suppresses high-amplitude noise, validating the necessity of the dual-stream design for handling extreme regime switches.

Table 2. Performance Comparison of Different Model Variants on the Sinter-A and Sinter-B Datasets.

Dataset	Model	MAE	RMSE	MAPE (%)	WAPE (%)
Sinter-A	Fixed Graph Only	15.888	61.735	4.651	3.636
	Fixed + Static Adaptive	**15.878**	**61.580**	4.676	3.634
	Static + Dynamic Fusion (Ours)	15.844	61.314	**4.627**	**3.626**
Sinter-B	Fixed Graph Only	142.451	993.696	9.925	2.761
	Fixed + Static Adaptive	140.956	990.425	9.852	2.750
	Static + Dynamic Fusion (Ours)	**135.223**	**971.472**	**9.682**	**2.679**

4. Conclusion

This paper presents Process-Informed Dynamic Graph WaveNet, a novel spatiotemporal forecasting framework tailored for multi-step prediction in industrial sintering processes. By integrating global control variables through feature broadcasting and decoupling graph learning into a static anchor which initialized with process time-delay priors and a noise-

robust RBF-based dynamic module, the model effectively balances physical consistency with adaptability to transient fluctuations. Experiments on two real-world sintering datasets demonstrate that our approach achieves state-of-the-art performance, particularly excelling in high-variance conditions where it significantly outperforms strong baselines. The results validate that embedding process knowledge into graph topology enhances robustness against industrial noise and operating drifts. Future work will explore online adaptation to long-term process degradation and extend the framework toward multi-objective optimization of energy efficiency and emission reduction.

References

[1] Wen W, Deng Z, Ma X, et al. Analysis of the synergistic benefits of typical technologies for pollution reduction and carbon reduction in the iron and steel industry in the Beijing–Tianjin–Hebei region[J]. Scientific Reports, 2024, 14(1): 12413.

[2] Yan F, Zhang X, Yang C, et al. Data-driven modelling methods in sintering process: Current research status and perspectives[J]. The Canadian Journal of Chemical Engineering, 2023, 101(8): 4506-4522.

[3] Du S, Ma X, Fan H, et al. Intelligent prediction and soft-sensing of comprehensive production indicators for iron ore sintering: A review[J]. Computers in Industry, 2025, 165: 104215.

[4] Xiong D L, Ning H Y, Xie M, et al. A Predictive Model for Sintering Ignition Temperature Based on a CNN-LSTM Neural Network with an Attention Mechanism[J]. Processes, 2024, 12(10): 2185.

[5] Li Y, Yang C, Sun Y. Dynamic time features expanding and extracting method for prediction model of sintering process quality index[J]. IEEE Transactions on Industrial Informatics, 2021, 18(3): 1737-1745.

[6] Yu Q, Yang G, Wang X, et al. A review of time series forecasting and spatio-temporal series forecasting in deep learning: Q. Yu et al[J]. The Journal of Supercomputing, 2025, 81(10): 1160.

[7] Zhang Z, Wang X, Tan X, et al. DSSA-TCN: Exploiting adaptive sparse attention and diffusion graph convolutions in temporal convolutional networks for traffic flow forecasting[J]. PloS one, 2025, 20(11): e0336787.

[8] Zhu K, Zhao C. Dynamic graph-based adaptive learning for online industrial soft sensor with mutable spatial coupling relations[J]. IEEE Transactions on Industrial Electronics, 2022, 70(9): 9614-9622.

[9] Zhou H, Zhao J P, Loo C E, et al. Numerical modeling of the iron ore sintering process[J]. ISIJ international, 2012, 52(9): 1550-1558.

[10] Li Y, Yu R, Shahabi C, et al. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting[C]//International Conference on Learning Representations. 2018.

[11] Zheng C, Fan X, Wang C, et al. Gman: A graph multi-attention network for traffic prediction[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(01): 1234-1241.

[12] Xia Y, Liang Y, Wen H, et al. Deciphering spatio-temporal graph forecasting: A causal lens and treatment[J]. Advances in Neural Information Processing Systems, 2023, 36: 37068-37088.

[13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[14] Zeng A, Chen M, Zhang L, et al. Are transformers effective for time series forecasting?[C]//Proceedings of the AAAI conference on artificial intelligence. 2023, 37(9): 11121-11128.

[15] Bai L, Yao L, Li C, et al. Adaptive graph convolutional recurrent network for traffic forecasting[J]. Advances in neural information processing systems, 2020, 33: 17804-17815.

[16] Li F, Feng J, Yan H, et al. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution[J]. ACM Transactions on Knowledge Discovery from Data, 2023, 17(1): 1-21.

[17] Wu Z, Pan S, Long G, et al. Graph wavenet for deep spatial-temporal graph modeling[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. 2019: 1907-1913.

