

Research Article

Meta-Learning Enhanced Trade Forecasting: A Neural Framework Leveraging Efficient Multicommodity STL Decomposition

Bohan Ma ¹, Yushan Xue,¹ Jing Chen ¹ and Fangfang Sun²

¹School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, China

²Research Center for International Inspection and Quarantine Standards and Technical Regulations, Beijing 100013, China

Correspondence should be addressed to Jing Chen; chenjingma@cufe.edu.cn

Bohan Ma and Yushan Xue contributed equally to this work.

Received 25 October 2023; Revised 6 March 2024; Accepted 12 March 2024; Published 3 April 2024

Academic Editor: Yu-an Tan

Copyright © 2024 Bohan Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the dynamic global trade environment, accurately predicting trade values of diverse commodities is challenged by unpredictable economic and political changes. This study introduces the Meta-TFSTL framework, an innovative neural model that integrates Meta-Learning Enhanced Trade Forecasting with efficient multicommodity STL decomposition to adeptly navigate the complexities of forecasting. Our approach begins with STL decomposition to partition trade value sequences into seasonal, trend, and residual elements, identifying a potential 10-month economic cycle through the Ljung-Box test. The model employs a dual-channel spatiotemporal encoder for processing these components, ensuring a comprehensive grasp of temporal correlations. By constructing spatial and temporal graphs leveraging correlation matrices and graph embeddings and introducing fused attention and multitasking strategies at the decoding phase, Meta-TFSTL surpasses benchmark models in performance. Additionally, integrating meta-learning and fine-tuning techniques enhances shared knowledge across import and export trade predictions. Ultimately, our research significantly advances the precision and efficiency of trade forecasting in a volatile global economic scenario.

1. Introduction

Understanding the dynamics of global trade is paramount, as it interlinks with worldwide economic growth. Unforeseen economic or political shocks, such as the recent COVID-19 pandemic, can profoundly affect international trade. For instance, in the first half of 2020, China experienced a 6.3% decrease in total import and export value, compared to the previous year, due to pandemic-related disruptions (National Development and Reform Commission). Conflicts between countries, like Russia and Ukraine, also lead to economic disturbances, including inflation and reduced demand. These occurrences underline the significance of trade in economic growth and the necessity for accurate analysis and forecasting of foreign trade to mitigate potential adverse impacts on global economics.

Forecasting trade is inherently complex due to its susceptibility to unpredictable political and economic events

and the multifaceted nature of trading various commodities. Moreover, trade data are typically presented on a monthly basis, which means that the data volume is limited. Consequently, utilizing conventional deep learning techniques on these data often leads to overfitting. Nevertheless, accurate forecasting is crucial for both businesses and policymakers, as it aids in making informed decisions to mitigate adverse impacts on the global economy.

Historically, trade forecasting has employed traditional statistical techniques, such as ARIMA and fuzzy time series [1]. However, there has been a discernible shift towards machine learning and deep learning models [2], known for their enhanced accuracy in capturing complex data patterns.

The incorporation of deep learning has marked substantial advancements in time series forecasting [3–5]. Models like the Long- and Short-term Time series network (LSTNet) [6] have proven effective in multivariate time

series forecasting, ranging from solar plant energy outputs to traffic congestion predictions.

Recent explorations into spatiotemporal sequence forecasting offer promising avenues for trade forecasting. Treating import and export trades as nodes in spatiotemporal tasks and utilizing deep learning models like Convolutional Long Short-Term Memory (ConvLSTM), Spatiotemporal Graph Convolutional Networks (STGCNs), and Graph WaveNet [7–9] can possibly enhance trade prediction accuracy. Techniques such as Spatiotemporal Graph to Sequence (STG2Seq) [10] may further contribute to this evolving field, emphasizing the importance of continued research and development to adapt to the dynamic nature of global trade.

From these perspectives, we propose a neural model integrating Meta-Learning Enhanced Trade Forecasting with efficient multicommodity STL decomposition (Meta-TFSTL), which leverages the Transformer architecture and STL (Seasonal and Trend decomposition using Loess) decomposition, coupled with a dual-channel graph embedding with Meta-Learning. The structure of Meta-TFSTL is illustrated in Figure 1. The main contributions of this work are as follows:

- (1) *Novel Trade Forecasting Neural Network.* We develop a novel neural network that leverages the Transformer architecture and STL decomposition to capture intricate relationships and dependencies among various commodities and enhance the model's generalization ability.
- (2) *Construction of Temporal and Spatial Graphs.* We construct commodity spatial and temporal graphs based on Spearman correlation coefficients and temporal features, respectively, and employ graph embedding methods to capture the nodes' position and association and obtain high-dimensional representation vectors.
- (3) *Time Series Interpretability with Attention.* We use the self-attention mechanism of the Transformer architecture to capture drastic changes in the seasonal and trend components of the trade data, crucial for identifying sudden events and enhancing the interpretability of the time series neural network.
- (4) *Meta-Learning for Enhanced Generalization.* We integrate meta-learning techniques in response to the limited volume of monthly import and export data, aiming to enhance the model's proficiency in extracting insights from smaller datasets. Specifically, we adopt few-shot learning, a facet of meta-learning, to train our model such that it effectively generalizes to previously unseen datasets after minimal exposure to training examples.
- (5) *Meta Knowledge Adaptation in Import and Export Forecasting through Meta-Learning.* Building on the hypothesis that import and export value series can inform predictions for each other, we employ meta-learning to pretrain on one domain (either import or export) and subsequently fine-tune on

the other. The enhanced performance achieved through this meta knowledge adaptation approach, as compared to direct training on the target domain, reaffirms the existence of shared knowledge between imports and exports. This demonstrates the efficacy of meta-learning in harnessing this shared knowledge for improved forecasting accuracy.

- (6) *Advanced Performance relative to State of the Art.* Our model, Meta-TFSTL, outperforms advanced models on trade datasets (import and export), achieving lower Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). This results in more accurate trade forecasting, highlighting its practical superiority.

2. Related Work

2.1. Traditional Trade Forecasting. Trade forecasting plays a pivotal role in economic analysis and policymaking, offering insights into future trade trends and patterns. Over the years, numerous studies have assessed and compared various forecasting methods for predicting trade exports, imports, or both. These methods span from conventional statistical techniques such as autoregressive integrated moving average (ARIMA), fuzzy time series, and support vector regression (SVR), to more sophisticated approaches incorporating machine learning and deep learning technologies [11–13].

Classic time series models like the autoregressive integrated moving average (ARIMA) have been widely adopted for trade forecasting [14–16]. Research by Hasin et al. [17] and Farooqi [18] demonstrated ARIMA's efficacy in trade prediction. However, Fattah et al. [19] highlighted ARIMA's challenges, including its requirement for a large number of observations. Despite its capabilities, the exclusive use of ARIMA has declined due to its limitations in addressing modern data complexities.

Fuzzy time series models have also emerged as an alternative for trade forecasting. Wong et al. [20] assessed the performance of multivariate fuzzy time series models against traditional time series models for forecasting Taiwan's exports. Their findings suggested that fuzzy time series models could surpass ARIMA in short-term forecasting, although ARIMA was more adept for long-term predictions. Wang [21] conducted a similar comparison, concluding that fuzzy time series models offered superior accuracy for short-term forecasting but were limited in long-term projections. Wong et al. [22] demonstrated the effectiveness of fuzzy time series models in forecasting Taiwan's export volumes, showing more accurate predictions than ARIMA.

While fuzzy time series models excel in short-term forecasting, they struggle with long-term trends. In contrast, Support Vector Regression (SVR) outperforms both traditional and fuzzy models, especially with complex, nonlinear datasets. Guanghai [23] found SVR superior to

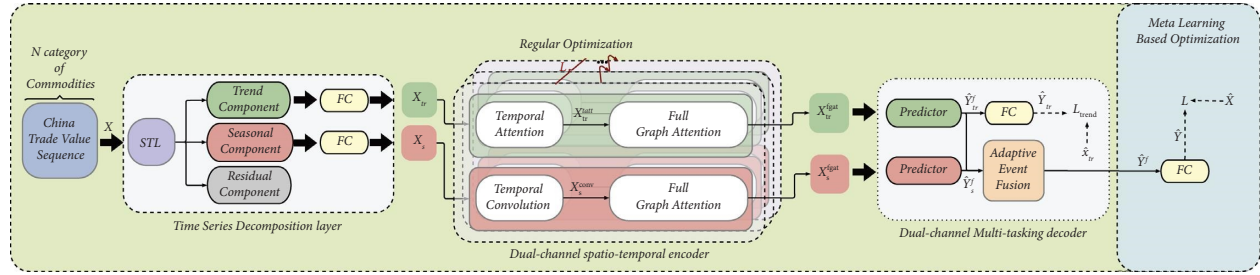


FIGURE 1: The architecture of the proposed Meta-TFSTL. FC: fully connected layer. In ANIL, the model’s feature-extracting backbone remains static (Regular Optimization), while the output head undergoes gradient descent updates (Meta-Learning-Based Optimization). This ensures stable foundational representations while allowing rapid task-specific adaptations.

ARIMA in demand forecasting. Studies by Lu and Wang [24] and Wu [25] emphasized SVR’s accuracy in product demand predictions. Kuo and Li [26] enhanced SVR’s performance by integrating it with various algorithms. However, SVR requires careful parameter tuning and can be computationally intensive.

Multivariate time series models, such as ARIMAX, are favored in trade forecasting for their ability to incorporate multiple variables [27–29]. Despite their prevalence, these models assume linear relationships and can be computationally demanding. Conversely, recent studies have shown machine learning models outperforming traditional time series methods in trade forecasting [30–32]. These models excel with complex, nonlinear data but necessitate extensive training data and parameter tuning. Their effectiveness largely depends on the quality of feature extraction.

2.2. Deep Learning in Trade Forecasting. Recent research has delved into the potential of deep learning for trade forecasting, attributed to its capacity for automatic feature extraction and high representation ability [33–35]. Lloret et al. [3] introduced models based on CNN and ED-RNN, both outperforming traditional methods in forecasting disaggregated freight flows. Similarly, Shen et al. [4] utilized an LSTM network for predicting trade volumes of 23 countries, demonstrating its superiority over conventional statistical models.

In the broader domain of time series forecasting, which includes trade forecasting, there has been a shift towards leveraging advanced deep learning models to improve predictive accuracy. For instance, Qin et al. [36] introduced the Dual-Stage Attention-Based Recurrent Neural Network (DA-RNN) in 2017, designed to enhance the effectiveness and interpretability of time series predictions. Subsequently, Lai et al. [6] proposed the Long- and Short-term Time series network (LSTNet) in 2018, a framework aimed at tackling the challenges associated with multivariate time series forecasting, crucial for predictions in various sectors such as solar plant energy output and traffic congestion. Following this, Oreshkin et al. [37] developed N-BEATS in 2019, a deep neural architecture specifically for univariate time series point forecasting. More recently, in 2021, Lim et al. [5] proposed the Temporal Fusion Transformer (TFT), an

innovative attention-based architecture for multihorizon forecasting, combining high performance with interpretability.

2.3. Spatiotemporal Sequence Forecasting. As the field of forecasting evolves and the complexity of data patterns increases, spatiotemporal sequence forecasting has become increasingly important. Li et al. [38] proposed a model that combines Graph Convolutional Networks (GCNs) with Recurrent Neural Networks (RNNs) to capture spatial and temporal dependencies, respectively. Yu et al. [8] introduced STGCN, utilizing graph convolutional filters and 1D convolutional neural networks. Wu et al. [9] proposed Graph WaveNet, incorporating a WaveNet-based architecture and GCN with attention mechanisms. Zhang et al. [39] developed deep spatiotemporal residual networks for citywide crowd flow prediction. Seo et al. [40] introduced structured sequence modeling with graph convolutional recurrent networks. Zheng et al. [41] presented a spatiotemporal sequence prediction model for large-scale traffic data using deep learning approaches. Nevavuori et al. [42] employed spatiotemporal deep learning architectures with UAV data for crop yield prediction, achieving promising results with a 3D-CNN architecture.

2.4. Optimization-Based Meta-Learning in Trade Forecasting. Meta-learning, or “learning to learn,” has emerged as a powerful paradigm for training models to quickly adapt to new tasks with minimal data. In trade forecasting, where data patterns can vary significantly, meta-learning offers a robust solution. The Model-Agnostic Meta-Learning (MAML) algorithm [43] is a pioneering method in this domain, designed to be model-agnostic and applicable to any model trained with gradient descent. Reptile [44] and Amortized Neural Inference for Learning (ANIL) [45] further extend this concept, offering scalable and efficient solutions.

To the best of our knowledge, meta-learning is yet to be extensively applied in the domain of import and export forecasting. This situation presents a dual-edged scenario: a challenge, due to the unique complexities of trade data influenced by myriad global factors, and an opportunity, given meta-learning’s adaptability to new tasks with sparse data. The unexplored application of meta-learning in this area holds significant promise, suggesting that its integration could revolutionize trade forecasting by tapping into the vast potential of emerging markets and new product categories.

Despite progress, trade prediction faces several challenges. Firstly, the heterogeneous nature of commodities and their susceptibility to unforeseen events complicate accurate forecasting. Secondly, most works in trade forecasting have not utilized spatiotemporal forecasting, thus missing out on capturing potentially crucial dependencies and relationships in the data. Thirdly, many trade forecasting scenarios suffer from limited data availability, especially for emerging markets or niche commodities. These challenges underscore the continuous need for innovation and methodological refinement to improve forecasting accuracy. In response to these challenges, we propose Meta-TFSTL, a novel spatiotemporal sequence forecasting neural network with Meta-Learning designed to capture intricate relationships and dependencies among various commodities, enhance model generalization, and provide more accurate and timely guidance for trade forecasting.

3. Preliminaries

3.1. Problem Definition. For the Meta-TFSTL model used to predict trade values, let $x_t^i \in \mathbb{R}^C$, $C = 1$ represent the trade value of the i th commodity (imported or exported) at time step t , and $X_t = [x_t^1, \dots, x_t^i, \dots, x_t^N]^T \in \mathbb{R}^{N \times C}$ represent the trade values of N commodities (imported or exported) at time step t , denoted simply as $X_t \in \mathbb{R}^N$, and in the following derivation, C will be omitted. Given the historical trade data $\mathcal{X} = \{X_1, \dots, X_{T_1}\} \in \mathbb{R}^{T_1 \times N}$, the purpose of trade prediction is to predict the trade values of all commodities in the next T_2 time steps, namely, $\hat{\mathcal{Y}} = \{\hat{Y}_{(T_1+1)}, \dots, \hat{Y}_{(T_1+T_2)}\} \in \mathbb{R}^{T_2 \times N}$, and its ground truth is denoted by $\mathcal{X}' = \{X_{(T_1+1)}, \dots, X_{(T_1+T_2)}\} \in \mathbb{R}^{T_2 \times N}$.

3.2. Self-Attention Mechanism. The self-attention mechanism is a widely employed attention technique that allows each token in a sequence to gather information from all other tokens. The inputs are represented by queries, keys, and values, each with dimension d . The mechanism computes the dot products between the query and all keys, scales down each product by \sqrt{d} , and applies a softmax function to obtain the weights for the values:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QW_QK^TW_K}{\sqrt{d}}\right)VW_V, \quad (1)$$

where W_Q, W_K , and W_V are the learnable projection parameters and $\text{Attention}(\cdot)$ denotes the self-attention operation.

3.3. STL Decomposition. STL decomposition, i.e., Seasonal and Trend decomposition using Loess [46], is a renowned technique for breaking down a time series into seasonal, trend, and residual components. The seasonal component identifies regular patterns recurring at fixed intervals, the trend component depicts the long-term direction, and the residual component captures random fluctuations that remain after extracting the seasonal and trend components.

The STL decomposition process includes two iterative phases: an internal cycle and an external cycle. The internal cycle, aimed at fitting the trend and determining the seasonal component, involves six steps:

- (1) *Remove the Trend.* Calculate a detrended series $y_v - T_v^{(k)}$. Initially, $T_v^{(0)} = 0$.
- (2) *Smooth Cycle-Subseries.* Apply Loess to each point in the seasonal cycle, denoted as $C_v^{(k+1)}$.
- (3) *Apply Low-Pass Filter.* Two moving average filters and a Loess smoother are used, resulting in $L_v^{(k+1)}$.
- (4) *Detrend Smoothed Cycle-Subseries.* Compute $S_v^{(k+1)} = C_v^{(k+1)} - L_v^{(k+1)}$.
- (5) *Remove Seasonality.* Process $y_v - S_v^{(k+1)}$.
- (6) *Smooth the Trend.* Apply Loess to the deseasonalized series to obtain $T_v^{(k+1)}$.

3.4. Optimization-Based Meta-Learning and ANIL. This section introduces Optimization-Based Meta-Learning (OBML) through the Almost-No-Inner-Loop (ANIL) algorithm [45], a streamlined variant of MAML. Adhering to the setup in Subsection 3.1, we consider N training tasks, each corresponding to N commodities, with n sample-label pairs denoted as $\{X_i, Y_i\}_{i=1}^N$. Typically, a query-support split is employed during training.

ANIL minimizes the loss over $\theta = \{\theta^{<L}, w\}$ as follows:

$$\begin{aligned} \min_{\theta} \mathcal{L}_{ANIL}(\theta) &:= \sum_{i \in [N]} \ell(\phi_{\theta^{<L}}^\top(X_i)w_i^*, Y_i) \\ \text{s.t. } w_i^* &= \text{InnerLoop}(w, \phi_{\theta^{<L}}^\top(X_i), Y_i, \tau, \lambda). \end{aligned} \quad (2)$$

This equation delineates the standard inner-loop optimization in OBML, involving τ steps of gradient descent on the loss $\ell(\phi_{\theta^{<L}}^\top(X_i)w, Y_i)$ with a learning rate λ .

Lin et al. [47] empirically showed that maintaining w static during training does not compromise ANIL's performance, suggesting that optimizing w in the outer loop is nonessential. Hence, ANIL's training objective, excluding the outer-loop optimization of w , becomes

$$\min_{\theta^{<L}} \mathcal{L}_{ANIL}(\theta) := \sum_{i \in [N]} \ell(\phi_{\theta^{<L}}^\top(X_i)w_i^*, Y_i). \quad (3)$$

4. Methodology

4.1. Overall Network Architecture. According to Figure 1, the historical trade data $\mathcal{X} \in \mathbb{R}^{T_1 \times N}$ are initially processed through a time series decomposition layer, employing STL decomposition to separate the data into trend ($\mathcal{X}_{tr} \in \mathbb{R}^{T_1 \times N}$), seasonal ($\mathcal{X}_s \in \mathbb{R}^{T_1 \times N}$), and residual ($\mathcal{X}_r \in \mathbb{R}^{T_1 \times N}$) components. This decomposition facilitates the nuanced capture of long-term patterns, periodic fluctuations, and sudden events within the trend and seasonal components, respectively. It also enables the verification of the completeness of information extraction by examining the residual component. To enable further processing, \mathcal{X}_{tr} and \mathcal{X}_s are linearly transformed into a higher dimensional

space $\mathbb{R}^{T_1 \times N \times D}$ through fully connected layers, ensuring a richer representation for subsequent layers.

The integration of STL decomposition significantly influences the model's architecture by offering a structured approach to manage different aspects of the time series data. The trend component $\mathcal{X}_{tr}^{tatt} \in \mathbb{R}^{T_1 \times N \times D}$, which captures long-term patterns, is fed into a time attention module to emphasize global dependencies over time. Meanwhile, the seasonal component $\mathcal{X}_s^{conv} \in \mathbb{R}^{T_1 \times N \times D}$, reflecting periodic fluctuations, is processed through a time convolution module to capture local and cyclic patterns. This dual-channel approach ensures a comprehensive representation of time series patterns.

Subsequently, both trend and seasonal components are integrated with graph information through the Full Graph Attention module. This step captures the intricate relationships and dependencies among various commodities over time, underscoring the STL decomposition's role in enriching the model's input and enhancing its predictive capabilities. The outputs from this module, $\mathcal{X}_{tr}^{fgat}, \mathcal{X}_s^{fgat} \in \mathbb{R}^{T_1 \times N \times D}$, incorporate both time series and graph-based features.

Finally, in the Dual-channel Multitasking decoder, the predictors yield $\hat{\mathcal{Y}}_{tr}^f, \hat{\mathcal{Y}}_s^f \in \mathbb{R}^{T_2 \times N \times D}$, which are aggregated through Adaptive Event Fusion interaction to capture the latent representation of dual-scale temporal patterns, resulting in $\hat{\mathcal{Y}}^f \in \mathbb{R}^{T_2 \times N \times D}$. The final prediction $\hat{\mathcal{Y}} \in \mathbb{R}^{T_2 \times N}$ is obtained through linear transformation.

4.2. Time Series Decomposition Layer. Given the dynamic nature of global trade, our model's architecture is precisely engineered to adeptly respond to sudden economic shifts or political events that may impact trade patterns. Through the application of STL decomposition, the model disentangles the input data \mathcal{X} into trend (\mathcal{X}_{tr}), seasonal (\mathcal{X}_s), and residual (\mathcal{X}_r) components, as shown in the following equation:

$$\mathcal{X} = \mathcal{X}_{tr} + \mathcal{X}_s + \mathcal{X}_r. \quad (4)$$

The trend component is tasked with capturing long-term changes, whereas the seasonal component addresses cyclical fluctuations. In instances where sudden events occur within a specific cycle, such anomalies trigger the temporal attention module, which focuses on the trend component, to detect deviations from established patterns. Concurrently, the variability in the seasonal component increases, reflecting the immediate effects of these events on trade dynamics. This nuanced response is further refined in the subsequent Full Graph Attention module, which captures anomalies by leveraging the interconnectedness and dependencies across various commodities and timeframes.

This dual-channel processing, combining trend stability with seasonal variability, renders the model particularly sensitive to abrupt economic or political occurrences. The focus of the temporal attention module on the trend component ensures that long-term shifts are accurately detected, while the increased variability in the seasonal

component signals short-term disruptions. The Full Graph Attention module enhances this mechanism by providing a broader context, allowing for an even more nuanced understanding and response to sudden changes.

Thus, the model supports rapid adaptation to such changes, ensuring its forecasting capabilities remain robust and responsive in the ever-changing landscape of global trade. This strategic design underscores our model's readiness to navigate the complexities of trade forecasting amidst economic variability and unforeseen events.

4.3. Dual-Channel Spatiotemporal Encoder. The dual-channel Spatiotemporal encoder, as shown in Figure 2, represents an innovative approach to trade forecasting, addressing the complexities of trade data through a comprehensive model. The "dual-channel" aspect of the encoder is designed to separately model the trend and seasonal components derived from STL decomposition, allowing for a nuanced understanding and representation of both long-term trends and cyclical patterns within trade data. This separation ensures that each component's unique characteristics are accurately captured and utilized for forecasting.

Moreover, the "Spatiotemporal" nature of the encoder incorporates both time slots with Struc2Vec graph embedding and a temporal attention mechanism, along with dilated causal convolution. This integration enables the model to capture not just the temporal correlations present within the trade data but also the spatial relationships between different commodities or markets. By embedding time slots and employing graph embedding techniques, the encoder enriches the model's input with both global and local patterns, significantly enhancing its forecasting capability.

Incorporating spatial and temporal graphs into the encoder addresses the limitations of previous single-method approaches, which may not fully capture the intricate dependencies and dynamics present in complex trade data. The dual-channel Spatiotemporal encoder's design philosophy is rooted in the need for a more robust and flexible modeling technique that can adeptly navigate the multifaceted nature of global trade, thereby offering a substantial improvement over traditional forecasting models.

4.3.1. Dual-Channel Temporal Pattern Recognition. Unlike previous works that employed single methods such as LSTM to model complex temporal patterns in entangled financial sequences, our approach utilizes both temporal convolutional layer and temporal attention to capture the temporal correlations of trend and seasonal components.

The trend component (X_{tr}), embodying long-term patterns, is adeptly captured using temporal attention after being decomposed by the Time Series Decomposition layer via STL decomposition, as illustrated in Figure 2. This process considers global relationships across the entire series, thereby effectively capturing the overall trend. Conversely, the seasonal component (X_s), representing short-term patterns and specific events, is best modeled using temporal convolutional layers that focus on local patterns. These layers operate on the decomposed seasonal

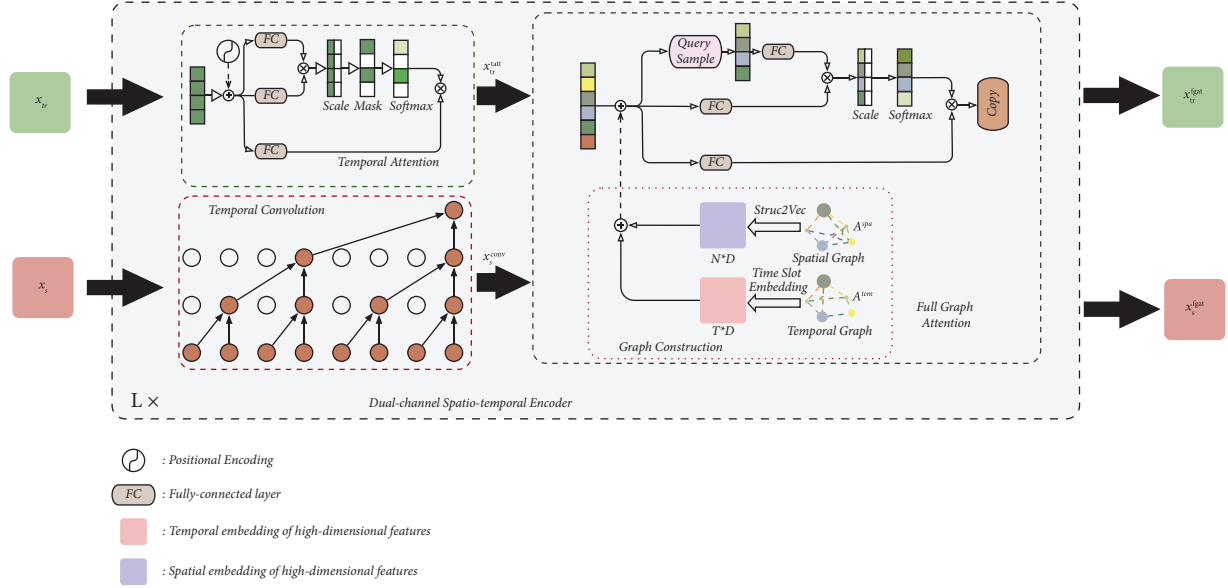


FIGURE 2: Architecture of the Dual-channel Spatiotemporal Encoder showcasing the integration of temporal attention mechanism, dilated causal convolution, and Struc2Vec graph embedding to capture both the temporal correlations and spatial relationships in trade data. The encoder effectively processes and integrates trend and seasonal components decomposed via STL, facilitating accurate trade forecasting by leveraging global and local patterns.

component, accurately capturing seasonal patterns and sudden events. This combination leverages the strengths of both methods, enabling comprehensive and accurate forecasting of entangled financial sequences. The distinct processing of X_{tr} and X_s components through the dual-channel Spatiotemporal encoder ensures a nuanced approach to modeling the intricate dynamics of trade data.

The temporal convolutional layer employed in this study is a one-dimensional convolution that slides over the input by skipping values at specific strides, as illustrated in Figure 2. Theoretically, given a one-dimensional sequence input $x \in \mathbb{R}^T$ and a filter $f \in \mathbb{R}^J$, the temporal convolution operation between x and f at time step t is defined as

$$x * f(t) = \sum_{j=0}^J f(j) \cdot x(t - c \times j), \quad (5)$$

where c is the dilation factor. The temporal convolution layer for the seasonal component is represented by

$$\mathcal{X}_s^{\text{conv}} = \text{ReLU}(\Theta * \mathcal{X}_s + b), \quad (6)$$

where Θ and b are learnable parameters and $\text{ReLU}(\cdot)$ denotes the rectified linear unit. Moreover, we apply masked self-attention to the temporal dimension of the trend component. This approach is motivated by the trend component's inherent stability, allowing it to clearly represent long-term trends:

$$\mathcal{X}_{tr}^{\text{tatt}} = \text{Concat}(ta_1, \dots, ta_n, \dots, ta_N), \quad (7)$$

where $ta_n = \text{Attention}(X_{tr}^n, X_{tr}^n, X_{tr}^n)$.

4.3.2. Global Spatial Feature Extraction. As depicted in the Graph Construction module of Figure 2, for spatial correlation of commodity series, we initially considered adopting the vanilla Graph Attention Network (GAT) to dynamically calculate weights between connected nodes. However, the spatial receptive field of the vanilla GAT is limited to immediate neighbors. Thus, we utilized the full GAT to dynamically capture global spatial dependence by performing self-attention on the spatial dimension of X_s^{conv} and X_{tr}^{tatt} . This approach enables the model to understand and leverage the complex spatial relationships within the trade data, enhancing its predictive capability by incorporating a broader context of intercommodity influences.

$$X^{\text{fgat}} = \text{Concat}(sa_1, \dots, sa_t, \dots, sa_{T_1}), \quad (8)$$

where $sa_t = \text{Attention}(X_t, X_t, X_t)$.

4.3.3. Temporal Graph Construction. Following the strategies outlined in the Graph Construction module of Figure 2, given the periodic fluctuations in commodity value sequences, with each monthly timestamp t acting as a unit of time in the dataset, we adopt the time-slot method inspired by Yuan et al. [48] to represent timestamps, simplifying the model to focus on quarterly cycles. With Δt set as one quarter, we construct a directed temporal graph of size $4 \times 3 = 12$ to represent the three months in each quarter. The time graph is embedded into the feature space using the Struc2Vec graph embedding technique, obtaining the high-dimensional temporal graph embedding $\rho^{\text{tem}} \in \mathbb{R}^{T \times d}$, which is then replicated N times to $\rho^{\text{tem}'} \in \mathbb{R}^{T \times N \times d}$.

4.3.4. Spatial Graph Construction. For computing the spatial correlation graph of import and export commodity value series, we calculate the Spearman correlation coefficient from the value series data of N commodities, forming an adjacency matrix that represents the correlation between commodities. Using the Struc2Vec algorithm, the vector representation of each node (commodity) is updated iteratively, capturing structural information and neighboring influence. The final high-dimensional commodity association graph embedding $\rho^{\text{spa}} \in \mathbb{R}^{N \times d}$ is replicated T times, resulting in $\rho^{\text{spa}'} \in \mathbb{R}^{T \times N \times d}$.

Finally, our full graph attention can be formulated as

$$\tilde{\mathcal{X}} = \mathcal{X} + \rho^{\text{spa}'} + \rho^{\text{tem}'}. \quad (9)$$

This equation combines raw trade data (\mathcal{X}) with spatial ($\rho^{\text{spa}'}$) and temporal ($\rho^{\text{tem}'}$) graph embeddings, enhancing the model's ability to capture spatial dependencies and temporal dynamics in trade data. By incorporating these embeddings, the model benefits from a richer representation that can take into account the interactions of commodities over time and space. This integrated approach enriches the model's inputs by taking advantage of the nuanced relationships and changing patterns inherent in global trade, thereby enhancing the model's generalization capabilities and its adaptability to complex trade data scenarios.

4.4. Dual-Channel Multitasking Decoder. To transform the representation encoded by the dual-channel encoder into future representations for multistep import and export commodity value series prediction, this study employs a predictor (i.e., a fully connected layer) on the time dimension of $\mathcal{X}_{\text{tr}}^{\text{gat}}$ and $\mathcal{X}_{\text{s}}^{\text{gat}} \in \mathbb{R}^{T_1 \times N \times d}$. Through the predictor, future representations of the trend and seasonal components, $\hat{\mathcal{Y}}_{\text{tr}}^f$ and $\hat{\mathcal{Y}}_{\text{s}}^f \in \mathbb{R}^{T_2 \times N \times d}$, are obtained. Subsequently,

fusion attention and multitasking supervision merge the information of trend and seasonal components, acquiring knowledge through the supervision of the trend component.

In the specific frequency decoder, we predict the future of the trend and seasonal components outputted by the dual-channel encoder using a predictor. The trend and seasonal information are integrated through fusion attention and multitasking supervision.

The trend component, generally more stable and less volatile than the seasonal component, provides a reliable basis for model training, helping stabilize the training process. Moreover, the trend, embodying the persistent and stable global patterns in the data, plays a crucial role in overall predictions. While the seasonal component is pivotal for capturing short-term fluctuations and sudden events, its inherent uncertainty and volatility make it less reliable for overall predictions. Therefore, by prioritizing the supervision of the trend component, the model can extract more valuable and reliable information for long-term predictions, crucial for decision making in many practical applications.

4.4.1. Decomposed Temporal Feature Fusion. The goal is not merely to predict the trend and seasonal components but to forecast the future trade value series based on these components and other factors. We propose a fusion attention mechanism, as illustrated in Figure 1, which integrates the representations of the trend and seasonal components, denoted by $\hat{\mathcal{Y}}_{\text{tr}}^f$ and $\hat{\mathcal{Y}}_{\text{s}}^f$, into the future trade value sequence $\hat{\mathcal{Y}}^f \in \mathbb{R}^{T_2 \times N \times d}$. This mechanism captures future internal dependencies by considering the trend component as the query, extracting useful long-term and short-term information from both the trend and seasonal components within two temporal attentions. The fusion attention mechanism is expressed as follows:

$$\hat{\mathcal{Y}}^f = \text{Concat}(fa_1, \dots, fa_n, \dots, fa_N),$$

$$\text{where } fa_n = \text{Attention}\left(\hat{\mathcal{Y}}_{\text{tr}}^{f^n}, \hat{\mathcal{Y}}_{\text{tr}}^{f^n}, \hat{\mathcal{Y}}_{\text{tr}}^{f^n}\right) + \text{Attention}\left(\hat{\mathcal{Y}}_{\text{tr}}^{f^n}, \hat{\mathcal{Y}}_{\text{s}}^{f^n}, \hat{\mathcal{Y}}_{\text{s}}^{f^n}\right). \quad (10)$$

4.4.2. Multitasking and Loss Function. During the training process, a fully connected layer is employed to convert the future representation of trade value sequences, denoted as $\hat{\mathcal{Y}}^f$, into the predicted values $\hat{\mathcal{Y}}$. The model utilizes L_1 loss for supervision. By leveraging knowledge from the more stable trend component, the model effectively enhances its capability to learn the long-term trends of the commodity value sequence, thus improving performance. Consequently, the optimization objective of Meta-TFSTL is to minimize the loss function shown in the following equation:

$$L = \sum_{t=T_1+1}^{T_1+T_2} \sum_{n=1}^N |y_t^n - \hat{y}_t^n| + |y_{tr_t}^n - \hat{y}_{tr_t}^n|. \quad (11)$$

This loss function computes the L_1 distance loss between the predicted and actual values for each timestep t and each node n , where y_t^n represents the actual value, and \hat{y}_t^n denotes the predicted value for each commodity in the subsequent months. Moreover, $y_{tr_t}^n$ signifies the real trend component, and $\hat{y}_{tr_t}^n$ its predicted counterpart. Minimizing this loss function enables the model to better fit the future trends in value, thereby enhancing model performance.

4.5. Meta-Learning Framework for Trade Forecasting. Faced with a new commodity or a shift in economic conditions, Meta-TFSTL applies its meta-learned knowledge for initial predictions, demonstrating the model's quick

adaptability by fine-tuning on a limited dataset specific to the new context. This adaptation mechanism is crucial for maintaining high forecasting accuracy in the dynamic and unpredictable domain of global trade, emphasizing our approach's effectiveness in navigating evolving market trends and economic shifts.

In trade forecasting, a task \mathcal{T}_i is defined as the prediction of trade values under specific economic conditions. Each task comprises a support set \mathcal{S}_i for training and a query set \mathcal{Q}_i for testing, illustrating the model's readiness for various forecasting scenarios. For task \mathcal{T}_i , the support set \mathcal{S}_i includes n pairs of historical trade data and corresponding trade values, represented as $\{\mathcal{X}_i^s, \mathcal{Y}_i^s\}$, where $\mathcal{X}_i^s \in \mathbb{R}^{T_1 \times N}$ and $\mathcal{Y}_i^s \in \mathbb{R}^{T_2 \times N}$.

The meta-learning framework categorizes model parameters into two groups: $\theta^{<L}$, the parameters of all layers except the last, termed Backbone parameters, and w , the last layer or task-specific head parameters, termed Output parameters. The main objective is to optimize $\theta^{<L}$ across multiple tasks, allowing w to be rapidly adjusted for each specific task \mathcal{T}_i .

The adaptation process for task \mathcal{T}_i is as follows:

$$w_i^* = w - \alpha \nabla_w \mathcal{L}_{\mathcal{T}_i}(\mathcal{S}_i; \theta^{<L}, w), \quad (12)$$

where α is the learning rate for the inner-loop optimization and $\mathcal{L}_{\mathcal{T}_i}(\mathcal{S}_i; \theta^{<L}, w)$ is the loss computed on the support set \mathcal{S}_i for task \mathcal{T}_i using the current model parameters.

After adjusting w to w_i^* for task \mathcal{T}_i , the model's performance on the query set \mathcal{Q}_i informs the update of $\theta^{<L}$ based on overall task performance:

$$\theta^{<L*} = \theta^{<L} - \beta \nabla_{\theta^{<L}} \sum_i \mathcal{L}_{\mathcal{T}_i}(\mathcal{Q}_i; \theta^{<L}, w_i^*), \quad (13)$$

where β is the learning rate for outer-loop optimization and $\mathcal{L}_{\mathcal{T}_i}(\mathcal{Q}_i; \theta^{<L}, w_i^*)$ is the loss on the query set \mathcal{Q}_i for task \mathcal{T}_i , using the adapted parameters w_i^* .

This iterative two-phase optimization process—adjusting w for each task followed by updating $\theta^{<L}$ based on aggregated task performance—enables Meta-TFSTL to acquire generalized parameters $\theta^{<L*}$, facilitating rapid adaptation to new tasks. This capability significantly enhances the model's forecasting accuracy, especially for new commodities or changing economic conditions.

5. Experiments

5.1. Dataset. This study utilizes the monthly trade value series of imported and exported commodity from China between January 2005 and January 2023, sourced from *Global Trade Flow* (<https://gtf.sinoimex.com>). The trade values are tallied once a month and are denominated in US dollars. This dataset encompasses all of China's trade commodities in recent years. Based on the 2022 customs duty specifications, inspection and quarantine codes set by China Customs, and the globally accepted HS8 codes, this study categorizes the commodities into 13 major classes for both imports and exports. The commodities and their respective abbreviations are presented in Table 1.

In this study, we use the values of imported (or exported) commodities from the first 10 time steps to predict the values in the subsequent 2 time steps. These datasets are then chronologically split into training (70%), validation (20%), and test (10%) sets. Performance of all methods is evaluated using three standard metrics, namely, MAE, RMSE, and MAPE.

5.2. Experimental Settings

5.2.1. Baselines. In this paper, we benchmark the performance of our proposed Meta-TFSTL model against a comprehensive suite of established baseline models. These baselines span from traditional statistical methods to the latest neural network architectures in time series forecasting. Our selection includes a total of 11 models, providing a broad overview of the field's evolution and current state of the art. Here is a brief overview of each model, including their publication year to highlight recent advancements:

- (1) LastValuePredictor: a basic forecasting method using the most recent observation to predict future values. This approach serves as a simple baseline for comparison.
- (2) ARIMA [49] (1976): a well-established statistical method for time series forecasting, known for its effectiveness in capturing linear relationships and trends.
- (3) VAR [50] (1980): a model that captures linear interdependencies among multiple time series, widely used in econometrics and financial analysis.
- (4) Bagging [51] (1996): an ensemble technique that improves the stability and accuracy of machine learning algorithms by combining multiple models.
- (5) LSTM [52] (1997): a recurrent neural network architecture designed to learn long-term dependencies, marking a significant advancement in sequence modeling.
- (6) GRU [53] (2014): an efficient variant of LSTM with fewer parameters, making it faster and simpler while retaining the capability to capture temporal dependencies.
- (7) DeepAR [54] (2020): a probabilistic forecasting model that leverages deep learning for uncertainty estimation, representing a recent trend towards more adaptable and nuanced forecasting methods.
- (8) DeepVAR [54] (2020): an extension of the VAR model that incorporates deep learning techniques for enhanced multivariate time series forecasting. While not associated with a specific publication year, it is part of the recent push to integrate deep learning into traditional forecasting models.
- (9) N-Beats [37] (2019): a purely neural network-based model for time series forecasting that uses a stack of feed-forward networks, showcasing the increasing trend towards deep learning in the field.

TABLE 1: Abbreviations for imported and exported commodity.

Imported commodity	Abbreviation	Exported commodity	Abbreviation
Unwrought copper and copper materials	Cu	Other agricultural products	Agri
Natural gas	NG	Electrical equipment	EE
Metal ores and concentrates	Metal	Other	Other
Automatic data processing equipment and its parts	ADPE	Electrical appliances	EA
Integrated circuits	IC	Refined oil	RO
Grain	Grain	Plastic products	PP
Other agricultural products	Agri	Furniture and its parts	FP
Coal and lignite	Coal	Steel	Steel
Other	Other	Clothing, accessories, and textile	Textile
Pulp, paper, and paper products	Pulp	Automatic data processing equipment and its parts	ADPEP
Automobiles and parts	Auto	Grain	Grain
Primary shaped plastic	Plastic	Integrated circuit	IC
Crude oil	Crude	Automobiles and parts	Auto

- (10) N-Hits [55] (2021): a recent hierarchical neural forecasting model that emphasizes the model’s capacity for time series interpolation and extrapolation, reflecting the ongoing innovation in neural network architectures for forecasting.
- (11) TFT [5] (2021): a model that combines the temporal structure of time series data with the Transformer architecture, exemplifying the latest advancements in applying attention mechanisms to forecasting problems.

This diverse set of baselines, especially including models from the last three years (DeepAR, DeepVAR, N-Hits, and TFT), ensures that our comparison covers a wide spectrum of time series forecasting methodologies, from classical approaches to cutting-edge neural network models.

5.2.2. Experimental Settings. In this work, we implemented the Meta-TFSTL model using the PyTorch framework and trained it using the Adam optimizer for a total of 1000 iterations, each iteration including 5 adaptations. Within the Meta-TFSTL model, the number of heads e in the attention mechanism was set to 1, with an initial dimension d_e of 128. Additionally, the number of layers L in the spatiotemporal encoder was set to 2. To capture cyclical time dependencies, we employed dilated causal convolution layers with a kernel size of $J = 2$. The initial learning rate was set to 0.001, adjusted with a decay rate of 0.1. Dropout was also incorporated, with a dropout rate of 0.2, to mitigate the risk of overfitting in the model.

5.2.3. Training Environment. In this study, we utilized a computer equipped with two V100 GPUs and a Hygon C86 7380 32-core Processor CPU as our training environment. Each GPU boasts 32 GB of available memory, offering robust parallel computing capabilities to expedite the training of deep learning models.

5.3. Results. This study conducted experiments to investigate the Meta-TFSTL model, addressing the following six research questions:

RQ1. How should the periodicity and robustness of the STL time series decomposition be chosen and determined?

RQ2. How are the support set and query set selected and determined for the meta-learning algorithm ANIL?

RQ3. Does Meta-TFSTL outperform the baseline models, and what role does meta-learning play in enhancing the model’s performance?

RQ4. How do different components of Meta-TFSTL (e.g., sequence decomposition methods and graph embeddings) impact its performance?

RQ5. How do hyperparameters influence the performance of Meta-TFSTL?

RQ6. Is our proposed Meta-TFSTL more efficient than baseline models?

5.3.1. Determination of Periodicity and Robustness in STL Decomposition (RQ1)

(1) Periodically Determined. STL time series decomposition dissects a time series into seasonal components X_s , trend components X_{tr} , and residual components X_r . As mentioned earlier in Subsection 4.2, our focus in this inquiry is chiefly on the residual component, which captures the random fluctuations in the series that are not explained by its trend or seasonality.

To identify the dominant cycle in import (or export) value series, we examined periods ranging from 2 to 50. Using the STL decomposition, we tested the residuals for each period with a ten-order lag in the Ljung–Box test. A period was considered suitable if all series residuals showed white noise characteristics, indicating that the seasonal and trend components have effectively captured most of the series information.

After contrasting the test outcomes across varying periods, it was observed that the residuals for all product series distinctly passed the Ljung–Box test when the period was set to 10 months (i.e., period = 10). This period can be construed as the typical cyclicity for import (or export) value series. The results of the lagged ten-order Ljung–Box test under this

period for all the import/export product value series are delineated in Table 2.

From the perspective of national development and openness, trade series for commodities not only exhibit clear periodicity but also show an upward trend over time, reflecting the impact of globalization and economic growth on trade activities. In this context, STL becomes particularly important as it can precisely decompose economic time series into periodic and trend components, thus showing good adaptability to trade data [56].

Applying STL decomposition to all import commodities, as shown in Figure 3, and selecting a period of 10 months for analysis, we can not only clearly see the long-term growth patterns in the trend components of each commodity but also observe the regularity of periodic fluctuations. This identification of periodicity not only validates the accuracy of the chosen period length but also provides key prior knowledge for the construction of prediction models based on temporal attention. Particularly, the regular fluctuations observed in the seasonal components provide clear guidance for temporal convolutional layers in capturing periodic changes, ensuring that the model can effectively adapt to and recognize the periodic features in time series data.

Observing Figure 3, we can identify periods where the seasonal components exhibit noticeable shifts from previous trends, denoted as “Significant Regimes” (highlighted in purple in the figure). This observation aligns with our further analysis of the seasonal components in STL decomposition, emphasizing the model’s ability to discern substantial market fluctuations during specific periods. For instance, the rise in Natural Gas (NG) imports in 2021 reflects China’s energy demand and policies to reduce air pollution by shifting from coal to cleaner energy sources (US Energy Information Administration). Similarly, the increase in Metal Ores and Concentrates (Metal) imports from March 2020 onwards aligns with China’s economic recovery efforts and infrastructure projects post-COVID-19 (Reuters (2021) “China 2020 iron ore imports hit record on robust post-virus demand”). The growth in Grain imports from January 2020 is attributed to securing food supplies amid global uncertainties (World Grain (2020) “China imports record amount of grains in 2020”), while the spike in Coal and Lignite (Coal) imports by July 2020 corresponds to the demand for energy as the economy recovered (Reuters (2020) “China’s July coal imports surge on heatwaves, power use”). The volatility in Automobile and parts (Auto) imports between July 2017 and June 2021 could be due to domestic demand shifts, tariff adjustments, and global supply disruptions, particularly due to the pandemic and trade tensions (U.S. Department of Commerce “China—Automotive Industry”). The initial decline and subsequent rapid increase in Crude Oil (Crude) imports from January 2020 reflect global oil price fluctuations, strategic reserves replenishment, and support for domestic recovery (Reuters (2021) “China 2020 crude oil imports surge to record on buying binge”). These periods of significant changes in commodities imports underscore the STL decomposition’s effectiveness in capturing the dynamics of the market, providing valuable insights for the model’s attention

TABLE 2: Ljung–Box test results.

Lag order	lb_stat	lb_pvalue
1	7.61	$5.80e-03$
2	10.29	$5.83e-03$
3	10.82	$1.28e-02$
4	17.20	$1.77e-03$
5	26.50	$7.13e-05$
6	27.02	$1.43e-04$
7	27.21	$3.05e-04$
8	30.47	$1.75e-04$
9	32.73	$1.49e-04$
10	94.10	$8.20e-16$

mechanisms to focus on and learn from these key market changes.

Therefore, the application of STL decomposition in the analysis of multicommodity trade data showcases its superiority in revealing and utilizing the seasonal, trend, and random fluctuation components in time series data. This not only provides a solid foundation for subsequent model design and prediction but also offers a new perspective and method for understanding complex market behaviors.

(2) *Comparative Analysis of Robust and Nonrobust STL Decomposition.* STL time series decomposition primarily follows two distinct methods: Robust and Nonrobust decomposition. The Robust decomposition showcases enhanced robustness when dealing with data containing outliers or anomalies. Leveraging weighted algorithms, such as Local Weighted Regression (LOWESS), robust decomposition minimizes the impact of anomalies on the decomposition results. In contrast, the Nonrobust decomposition, relying on simple averaging, is more susceptible to outliers and can be adversely influenced by anomalous values. For a particular commodity series, the trend, seasonal, and residual components from both robust and nonrobust decompositions are illustrated in Figure 4 (with a period of 10, focusing on the imported Cu as an exemplary commodity).

From Figure 4, it can be discerned that the trend component from the robust decomposition is smoother, illustrating its insensitivity to anomalies. Conversely, the nonrobust decomposition’s trend component exhibits more pronounced local fluctuations, which contradict the trend component’s role in capturing overall tendencies. Examining the seasonal component, the robust decomposition’s seasonal fluctuations appear more pronounced. This can be primarily attributed to the reduced fluctuations captured by the robust decomposition’s trend component. As a result, sudden events or transient information might be incorporated into the seasonal features. Consequently, the seasonal component absorbs more volatility, reflecting sudden incidents in trade, aligning with the designed role of the seasonal component to detect periodic and abrupt events.

Upon careful observation and interpretation, the Robust STL decomposition emerges as the more suitable method. Its trend component aptly captures the overall tendencies without being hindered by transient information, while the

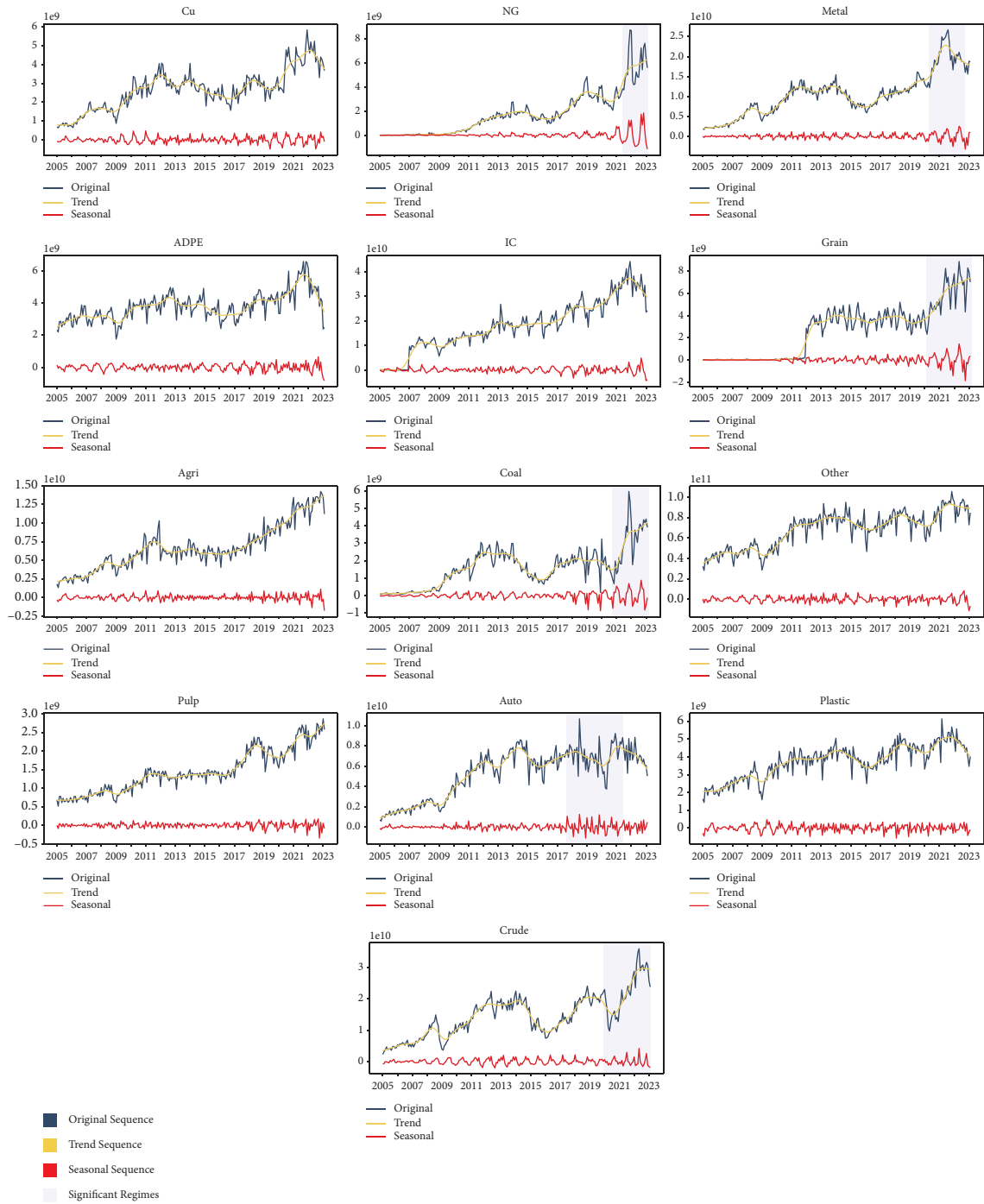


FIGURE 3: STL decomposition analysis across multiple commodities showcasing the identified 10-month cyclicity and the distinct upward trends, reinforcing STL’s adaptability in capturing economic series characteristics.

seasonal component proficiently identifies periodic fluctuations and unexpected occurrences.

5.3.2. *Determination of Support and Query Sets in the Meta-Learning Algorithm ANIL (RQ2).* From a meta-knowledge adaptation perspective, our aim is for the model to adapt to more complex scenarios. Therefore, we designate the more intricate situations as the query set [43]. The advantage of

enhancing generalization through meta-learning is evident in that training a model on a known data distribution (support set) can yield favorable results on an unknown data distribution (query set). Temporally speaking, the forecasting process often encompasses periods that are relatively straightforward to predict, as well as more challenging intervals. The overall performance can be adversely affected by these harder-to-predict time spans, leading to suboptimal model outcomes. To address this challenge, we strategically

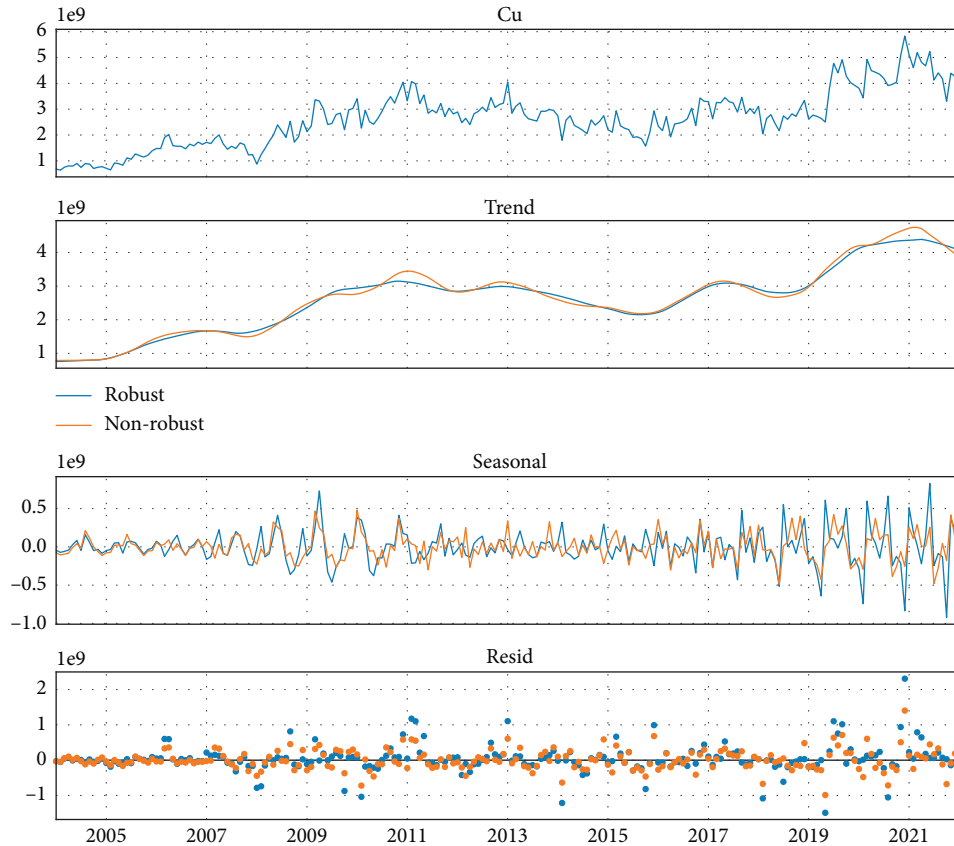


FIGURE 4: STL robust and unrobust decomposition of the value sequence of imported commodity Cu.

design our approach to leverage the strengths of meta-learning.

Adopting this approach for meta-learning modeling more effectively captures the intricate characteristics of the data. Initially, we employ the ARIMA algorithm to model all commodities in import and export, computing the monthly MAPE for each commodity. A subset of the results is illustrated in Figure 5.

We systematically computed the monthly MAPE between the predicted and actual values for all imported and exported commodities. Additionally, we derived the average MAPE across all commodities. By aggregating instances where the monthly MAPE exceeded the average MAPE for each commodity, a cumulative count was established, as illustrated in Figure 6. This metric serves as an indicator, highlighting specific months that are inherently more challenging to forecast compared to others. From this analysis, it is clear that the ARIMA forecasts for import and export commodity values are more accurate from April to September, with fewer instances where monthly MAPE exceeds average MAPE. Conversely, the months from January to March and October to December present greater forecasting challenges, likely influenced by global events such as New Year, Chinese Lunar New Year, and Christmas, which can disrupt trade patterns. Given these insights, we designate April to September as the support set and the remaining months as the query set for Meta-Learning.

5.3.3. Performance Comparison and Meta Knowledge Adaptation (RQ3)

(1) *Performance Comparison.* From Tables 3 and 4, both the TFSTL and Fine-Tuned Meta-TFSTL models excel in predicting import and export value series. Notably, the Fine-Tuned Meta-TFSTL surpasses TFSTL, demonstrating effective knowledge adaptation through fine-tuning.

For imported commodity value series, traditional models such as LastValuePredictor, ARIMA, and VAR tend to have higher error rates, with VAR underperforming significantly. This discrepancy may stem from the series' inherent nonlinearities and a lack of manually engineered features. Machine learning models show reliable results, with Bagging being noteworthy. Deep learning models generally perform comparably, but N-Beats edges ahead, potentially due to its sequence decomposition approach. This subtly reaffirms the robustness of our STL-based decomposition in TFSTL and Meta-TFSTL.

Predicting the exported commodity value series, traditional models show varied results, with VAR's performance being notably poor. Among deep learning models, while differences are minimal, N-Beats holds a slight edge, reaffirming its efficacy in such prediction tasks.

(2) *Meta Knowledge Adaptation.* Building on the premise that the Meta-TFSTL model leverages the predictability of certain months to establish foundational understanding of

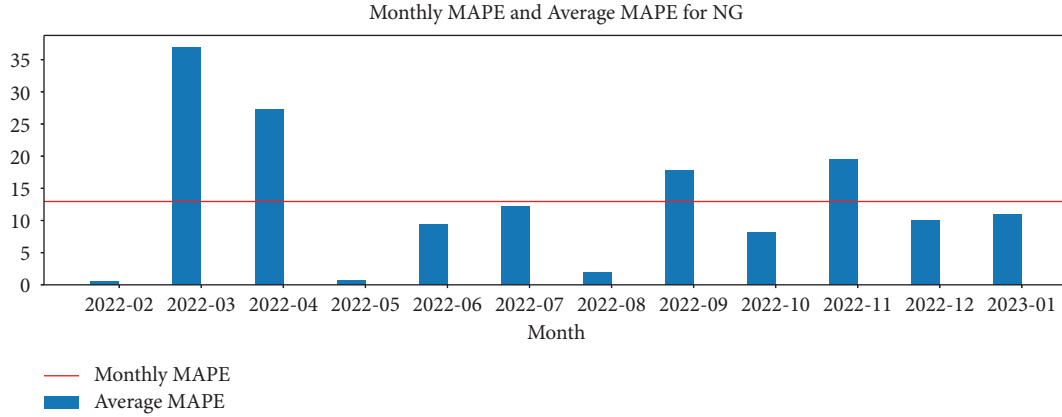


FIGURE 5: The red horizontal line represents the Mean Absolute Percentage Error (MAPE) of the ARIMA prediction results across the entire dataset (comprising the training, validation, and test sets). The blue bars depict the MAPE values for specific months, clearly highlighting several months where the prediction values significantly exceed the MAPE. These months are of particular interest and warrant further investigation.

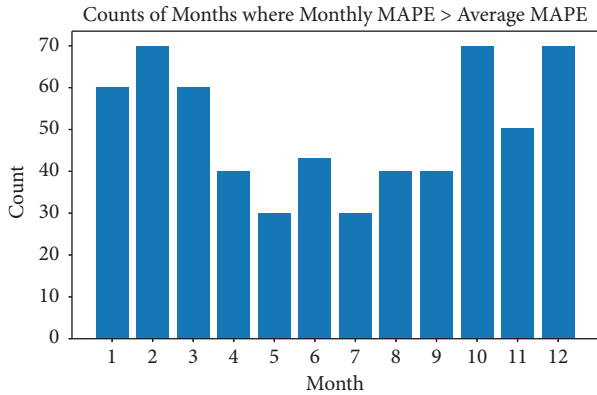


FIGURE 6: Cumulative count of months where commodity MAPE exceeds average, indicating forecast difficulty from January to March and October to December.

TABLE 3: Performance comparison of various models for imported commodity value series.

Model	MAE ($\times 10^9$)	RMSE ($\times 10^9$)	MAPE
LastValuePredictor	3.01	5.25	0.2011
ARIMA	2.71	3.23	0.1921
VAR	10.18	11.79	0.6878
Bagging	2.97	2.39	0.1555
LSTM	2.31	3.70	0.1874
GRU	3.02	5.10	0.2272
DeepAR	2.62	4.38	0.2020
DeepVAR	2.96	5.26	0.2063
N-Beats	2.04	3.54	0.1705
N-Hits	1.94	3.43	0.1641
TFT	2.07	3.80	0.1628
TFSTL	<u>1.77</u>	<u>3.15</u>	<u>0.1263</u>
Meta-TFSTL (fine-tuning)	1.47	2.97	0.1074

Bold: best; underline: second best; italics: best in baseline.

import trends and subsequently fine-tunes this knowledge with the more challenging months, we further explored its adaptability.

TABLE 4: Performance comparison of various models for exported commodity value series.

Model	MAE ($\times 10^9$)	RMSE ($\times 10^9$)	MAPE
LastValuePredictor	3.50	7.36	0.1854
ARIMA	2.64	5.29	0.1811
VAR	7.15	8.13	0.3802
Bagging	3.63	5.14	0.1983
LSTM	3.70	7.18	0.2064
GRU	3.82	8.15	0.2344
DeepAR	3.00	6.60	0.2210
DeepVAR	2.65	5.96	0.2185
N-Beats	2.92	6.08	<i>0.1763</i>
N-Hits	3.04	6.77	0.1804
TFT	2.76	4.89	0.1781
TFSTL	<u>2.01</u>	<u>4.55</u>	<u>0.1512</u>
Meta-TFSTL (fine-tuning)	1.78	3.64	0.1023

Bold: best; underline: second best; italics: best in baseline.

Shen et al. [4] posited the potential of leveraging economic formulas to predict export data using import data and vice versa, achieving commendable results. This highlighted plausible knowledge adaptation between import and export data, suggesting that reusing such knowledge could enhance prediction accuracy. To empirically validate this hypothesis, we adopted a meta-learning approach in our study to harness this knowledge adaptation.

Building on this foundation, our experiments with the Meta-TFSTL model for both import and export predictions were designed to strategically utilize training and validation sets from one domain and fine-tune on the validation set of the other. Specifically, for import predictions, we trained on the export dataset and fine-tuned using the import validation set, achieving a performance boost with a nearly 2 percentage point reduction in MAPE over the TFSTL model. Conversely, for export predictions, the model was initially trained on the import dataset and fine-tuned with the export validation set, resulting in a significant improvement with a reduction of nearly 5 percentage points in MAPE compared to the TFSTL model. This approach not only validated

Shen et al.'s findings but also underscored the efficacy of knowledge adaptation between import and export data domains using the Meta-TFSTL framework.

In conclusion, Meta-TFSTL distinguishes itself as a superior model in forecasting the value of imports and exports of commodities when compared to other models, attributed to its ability to

- (1) Extract and model the trend and seasonal components separately using a dual-channel encoder.
- (2) Utilize an integration of attention mechanisms and multilevel supervision for effective information merging.
- (3) Incorporate positional encoding through time graphs and commodity association graphs for capturing global dependencies.
- (4) Enhance generalization capabilities for few-shot learning scenarios like monthly data through meta-learning.
- (5) Facilitate knowledge adaptation between import and export datasets, optimizing adaptability for accurate predictions.

The superior prediction performance for imported commodities over exported ones may reflect the relative stability of domestic demand influencing imports, the steady impact of long-term tariffs and trade agreements, and the more exhaustive data acquisition for imports due to mandatory customs checks.

5.3.4. Ablation Study (RQ4). To investigate the effectiveness of various components of Meta-TFSTL, we compared it with six distinct variants:

- (1) Meta-TFX11 (Trade Forecasting via X-11-Decomposition-based Networks): this variant uses the classical X-11 decomposition method [57] for analyzing and adjusting seasonal fluctuations in the trade value series.
- (2) Meta-TFVMD (Trade Forecasting via Variational Mode Decomposition-based Networks): this model employs Variational Mode Decomposition (VMD) [58] for decomposing the trade value series into a set of intrinsic mode functions.
- (3) Meta-TFWavelet (Trade Forecasting via Wavelet-Decomposition-based Networks): this variant employs the Discrete Wavelet Transform (DWT) [59] instead of STL for decomposing the trade value series.
- (4) w/o G: a version of Meta-TFSTL without both spatial and temporal graphs.
- (5) w/o D: a version of Meta-TFSTL without the time series decomposition layer.
- (6) w/o F: a version of Meta-TFSTL where fusion attention is replaced with additive operations.

The ablation study results, presented in Tables 5 and 6, are organized into two distinct sections to evaluate the

effectiveness of different components within the Meta-TFSTL framework. The upper section of each table, above the line, comprises variants that employ alternative decomposition methods, including classical X-11 decomposition, Variational Mode Decomposition (VMD), and Discrete Wavelet Transform (DWT). The lower section assesses models from which key components have been removed, such as spatial and temporal graphs, the time series decomposition layer, or the fusion attention mechanism. This structured comparison highlights the integral role of these components, with the complete Meta-TFSTL model outperforming all its variants on import and export forecasting tasks, thereby underscoring the composite model's robustness and efficiency.

The ablation study reveals that while the Meta-TFX11 and Meta-TFVMD variants offer innovative approaches by employing X-11 decomposition and Variational Mode Decomposition (VMD), respectively, they do not match the performance of the full Meta-TFSTL model. The Meta-TFX11 variant, despite utilizing the classical X-11 decomposition method for adjusting seasonal fluctuations, may not be as effective in capturing the nonlinear and complex patterns present in trade value series, leading to its lower performance. Similarly, the Meta-TFVMD variant, while adept at decomposing the trade value series into intrinsic mode functions, might oversimplify the intricate economic trends and seasonal dynamics, which are crucial for accurate forecasting. This simplification could be the reason for its suboptimal results compared to Meta-TFSTL. Furthermore, the Meta-TFWavelet variant significantly underperforms relative to Meta-TFSTL, likely due to the reduction in time step post-Discrete Wavelet Transform (DWT) and the potential loss of series information during inverse filtering for upsampling. Additionally, wavelet decomposition may not aptly capture economic trends and seasonal fluctuations as effectively as STL, contributing to TFWavelet's inferior performance.

The "-G," "-F," and "-DF" models do not perform as well as the Meta-TFSTL model, likely due to the absence of graph embedding information, replacement of fusion attention, and omission of the disentangling flow layer. These components are crucial for the model's capability in information integration, complex pattern modeling, and relationship extraction, underscoring their importance within the model.

5.3.5. Parameter Sensitivity Analysis (RQ5). Figure 7 presents the results of a parameter sensitivity analysis for merchandise import and export value sequences. The top row of three graphs relates to the import merchandise's model hyperparameter variations, while the bottom row pertains to the export merchandise. For the import model, the hidden layer size and batch size were varied within a search space of [32, 64, 128, 256], while for the export model, the search space was extended to [32, 64, 128, 256, 512]. The import model achieves minimum prediction error with both hidden layer size and batch size set at 64, suggesting that further increases may lead to overfitting and decreased predictive performance. Conversely, the optimal

TABLE 5: Ablation study results for imports.

Model	MAE ($\times 10^9$)	RMSE ($\times 10^9$)	MAPE
Meta-TFX11	2.36	3.95	0.1652
Meta-TFVMD	2.17	2.62	0.1561
Meta-TFWavelet	2.25	3.85	0.1417
Meta-TFSTL-w/o G	2.11	3.79	0.1266
Meta-TFSTL-w/o D	2.13	4.05	0.1469
Meta-TFSTL-w/o F	2.29	4.33	0.1348
Meta-TFSTL	1.47	2.97	0.1074

The bold value indicates best result.

TABLE 6: Ablation study results for exports.

Model	MAE ($\times 10^9$)	RMSE ($\times 10^9$)	MAPE
Meta-TFX11	2.50	4.42	0.1713
Meta-TFVMD	2.41	4.39	0.1681
Meta-TFWavelet	4.86	11.89	0.1781
Meta-TFSTL-w/o G	4.94	14.65	0.1512
Meta-TFSTL-w/o D	7.45	19.39	0.2290
Meta-TFSTL-w/o F	2.43	6.06	0.1836
Meta-TFSTL	1.78	3.64	0.1023

The bold value indicates best result.

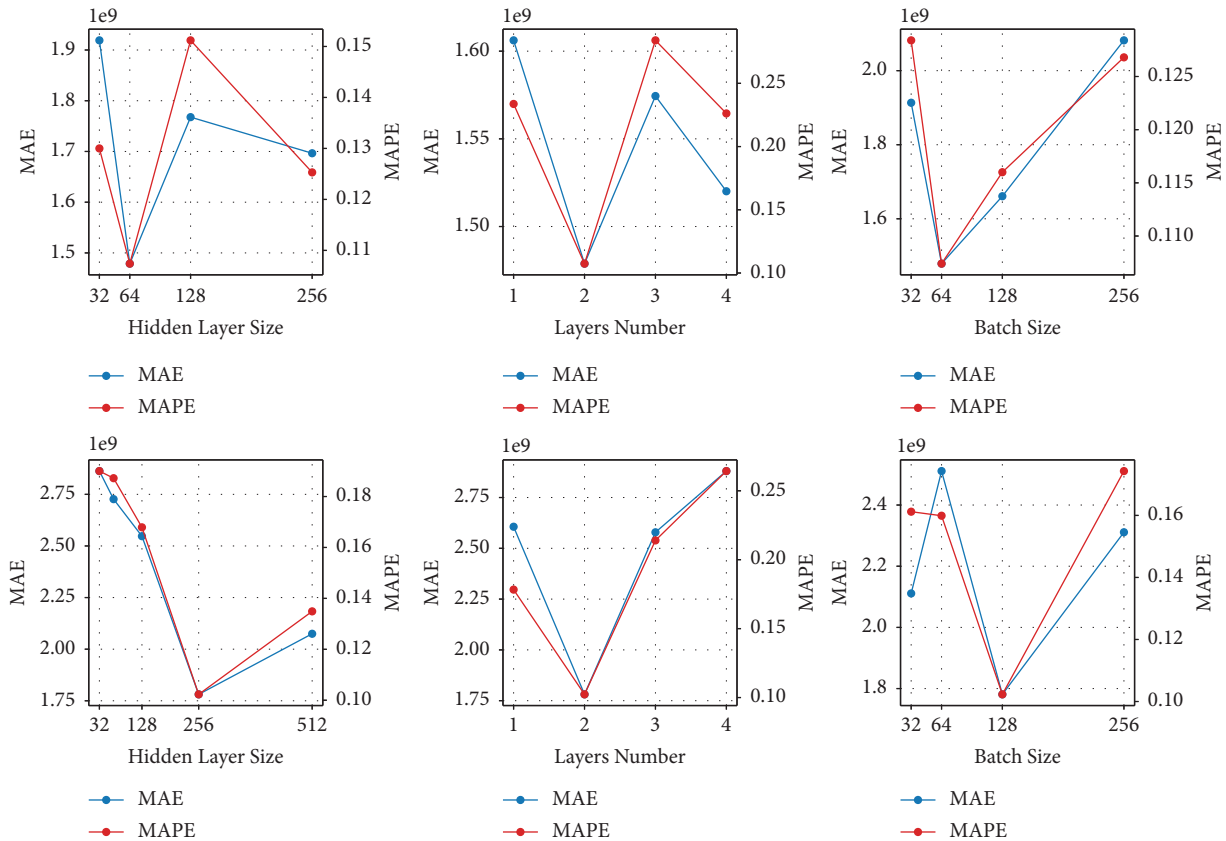


FIGURE 7: Hyperparameter study on import and export commodity value prediction: the blue and red lines represent the MAE and MAPE results on the test set, respectively.

outcome for the export model is achieved with a hidden layer size of 256 and a batch size of 128, indicating a higher predictive complexity for exports. Additionally, the performance of Meta-TFSTL improves with an increasing number of layers, stabilizing at a count of 2.

5.3.6. Model Scalability and Computation Cost (RQ6)

(1) *Model Scalability.* The scalability of neural network models is a crucial factor in their applicability to time series forecasting, particularly as the volume of data available for

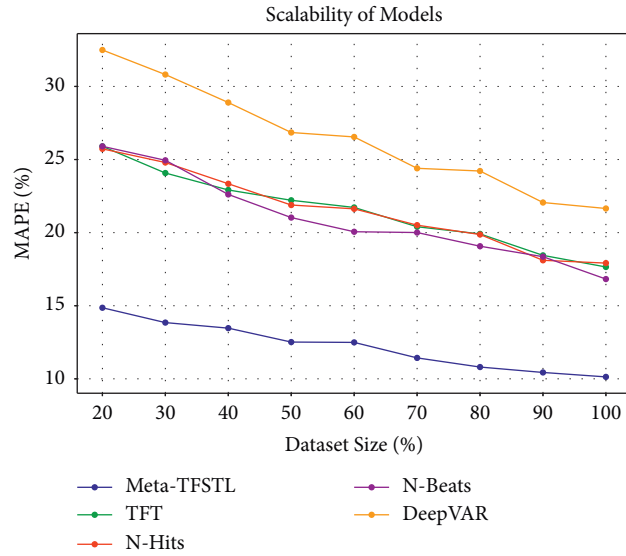


FIGURE 8: Comparative scalability analysis of neural network models for time series forecasting, highlighting Meta-TFSTL’s superior performance.

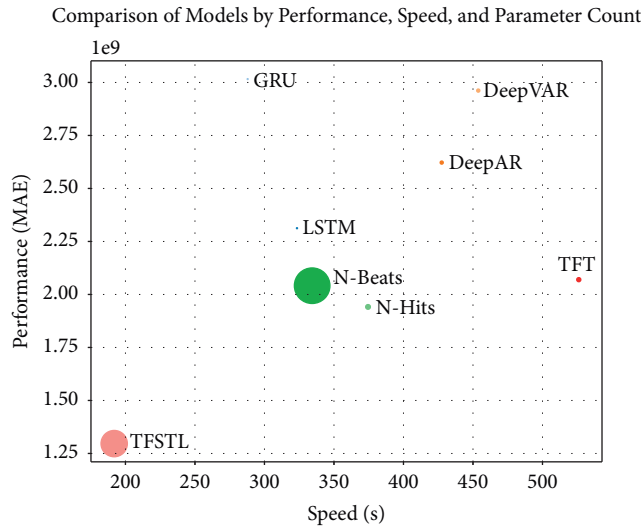


FIGURE 9: Performance (y axis), speed (x axis), and parameter count (size of the circles) of methods.

training increases. This study presents an empirical evaluation of the scalability of several advanced neural network models, including Meta-TFSTL, TFT, N-Hits, N-Beats, and DeepVAR, across varying dataset sizes from 20% to 100% in 10% increments. Our analysis, leveraging Mean Absolute Percentage Error (MAPE) as the performance metric, reveals Meta-TFSTL’s consistent superiority in scalability and predictive accuracy across all evaluated dataset sizes (Figure 8).

Our analysis, leveraging MAPE as the performance metric, reveals Meta-TFSTL’s consistent superiority in scalability and predictive accuracy across all evaluated dataset sizes. Starting with a MAPE of 14.86% at 20% dataset size, Meta-TFSTL exhibits a notable performance improvement, achieving a MAPE of 10.13% at full dataset utilization. This contrasts with other models, which, despite

showing improvements, do not match the efficiency and accuracy of Meta-TFSTL, highlighting its robustness and effectiveness in leveraging larger data volumes for enhanced forecasting accuracy.

(2) *Computation Cost.* The results from Figure 9 highlight a significant disparity in computational costs, manifesting through both speed and parameter count. Conventional RNN architectures, like LSTM and GRU, display moderate speeds with subpar performance. Their relatively smaller parameter count makes them computationally lightweight and simpler in design. On the other hand, models like DeepAR and DeepVAR seem to prioritize model intricacy with a more compact parameter footprint. However, their elevated MAE suggests potential compromises in their performance.



FIGURE 10: Imported commodity forecast result display (February 2022–January 2023). Each subplot illustrates the time series forecast for an imported commodity, including its actual values, predicted values, and MAPE of the model’s predictions.

The N-Beats model showcases a notably high parameter count, hinting at a complex model architecture. This complexity, however, does not necessarily translate to superior performance as its MAE is middling. The N-Hits and TFT models strike a balance between speed, performance, and parameter count.

Interestingly, the Meta-TFSTL model emerges as a frontrunner in terms of performance, boasting the lowest MAE. With its highly parallelized design and transformer-based architecture, it achieves the fastest speed among the models, despite its substantial parameter tally. Such a design choice is justifiable in applications where precision is

paramount, even if it means increased computational overhead within a given timeframe.

5.4. Enhancing Trade Forecasting through Meta Knowledge Adaptation: A Meta-TFSTL Case Study. Tables 3 and 4 have already highlighted the superior performance of Meta-TFSTL in comparison to baseline models. Delving into the predictive accuracy on the test set, incorporating Meta Knowledge Adaptation, the Meta-TFSTL model further cements its position by elevating forecasting precision for a wide array of commodities. This advancement is

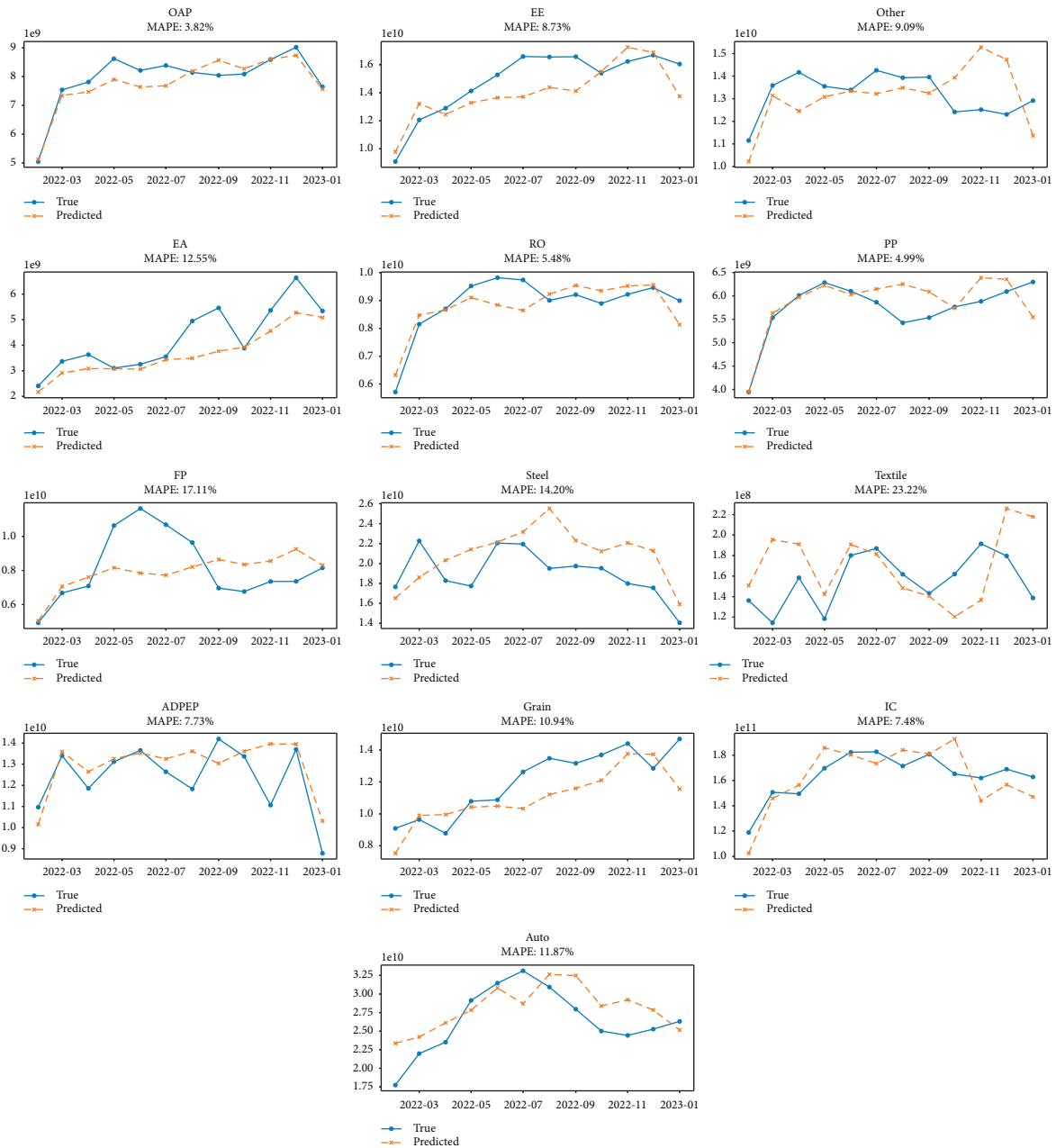


FIGURE 11: Exported commodity forecast result display (February 2022–January 2023). Each subplot illustrates the time series forecast for an exported commodity, including its actual values, predicted values, and MAPE of the model’s predictions.

significantly noticeable in both import and export sectors, as illustrated in Figures 10 and 11, with the model achieving exceptionally low MAPEs for commodities such as Cu (7.07%) and Agri (6.07%) in imports, and OAP (3.82%) and PP (4.99%) in exports, highlighting its predictive accuracy.

The model’s robustness is particularly noteworthy in its adept handling of commodities known for their market volatility, such as Coal and Textile in exports, with MAPEs of 13.21% and 23.22%, respectively. This showcases Meta-TFSTL’s capability to navigate and forecast within the unpredictable commodity markets effectively, where its adaptability and analytical prowess are paramount.

The essence of the Meta-TFSTL model’s success lies in its innovative adaptation of knowledge between import and export data, leveraging inherent patterns to enhance predictions. This adaptability is key, demonstrating the model’s superior analytical capabilities and consistent performance over baseline models in volatile market conditions.

This nuanced approach not only confirms the model’s supremacy but also underscores the critical role of knowledge adaptation in forecasting market trends accurately. Through meta-learning, Meta-TFSTL delivers dependable forecasts, crucial for strategic decision making, thus underscoring its indispensable value in commodities trading.

6. Discussion

While our Meta-TFSTL model demonstrates promising results in forecasting trade values, its practical applicability in real-world scenarios entails navigating a complex landscape of data availability, model interpretability, and adaptability to sudden market changes. Below, we detail the model's real-world applicability and delineate pivotal challenges alongside prospective enhancements.

6.1. Real-World Applicability of the Model

- (1) *Data Availability and Quality.* The performance of Meta-TFSTL heavily relies on access to accurate, detailed, and current trade data. Challenges such as delays in data collection, inconsistencies across international trade databases, and restrictive data policies can hinder model effectiveness. Enhancing collaborations with global trade organizations and exploring alternative data sources, like satellite imagery, could improve data quality and enrich model inputs.
- (2) *Model Interpretability.* The ability to interpret model predictions is crucial for trade policy and economic decision making. Despite its accuracy, the complex architecture of Meta-TFSTL may not be easily understandable, emphasizing the need to incorporate Explainable Artificial Intelligence (XAI) techniques to clarify the model's predictive processes and build trust among stakeholders.
- (3) *Adaptability to Market Fluctuations.* The dynamic nature of global trade, influenced by geopolitical, economic, and policy changes, requires a model that can quickly adapt. Integrating live economic indicators and sentiment analysis could enhance Meta-TFSTL's responsiveness, allowing for timely model updates in response to changing global trends.

6.2. Challenges and Prospective Developments

- (1) *Bilateral Trade Dynamics.* The model might not fully capture the complexities of bilateral trade agreements and policies. Developing a more nuanced approach that considers tariff negotiations, trade barriers, and bilateral agreements could offer a deeper understanding of global trade flows.
- (2) *Market Scalability.* While Meta-TFSTL shows promising results for China's trade data, extending its applicability to diverse economic systems and trade regulations worldwide is challenging. Future research should aim to test and adapt the model across different global markets to achieve broad applicability and scalability.

Conclusively, Meta-TFSTL represents a significant advance in trade forecasting. However, to fully realize its practical utility, it is essential to address these challenges through focused improvements, leveraging interdisciplinary

collaboration and innovation to enhance the model's real-world applicability and inform strategic trade policy and economic planning.

7. Conclusion

In this study, we introduced Meta-TFSTL, a novel neural model that integrates Meta-Learning Enhanced Trade Forecasting with efficient multicommodity STL decomposition. Empirical evaluations demonstrated Meta-TFSTL's superiority over baseline models, offering significant improvements in forecasting accuracy with the added benefits of swift computational efficiency. Through strategic application of STL decomposition, dual-channel spatio-temporal encoding, and innovative use of Struc2Vec graph embedding for spatial graphs and temporal graphs construction, Meta-TFSTL successfully merges insights from trend and seasonal components. This is further augmented by the adoption of fused attention and multisupervision strategies during the decoding phase. Incorporating meta-learning and fine-tuning methodologies, we have established a framework for effective knowledge adaptation between import and export trade predictions, leveraging the shared insights across these domains. Looking ahead, we plan to introduce more complex methodologies to enhance the model's capabilities, ensuring that Meta-TFSTL continues to set benchmarks in trade forecasting accuracy and computational efficiency.

Data Availability

The detailed data used in this study were sourced from the website <https://gtf.sinoimex.com>. Currently, we are providing a subset of the data used in this research, which can be accessed from the following link: <https://pan.baidu.com/s/1hn6arX8oO6J9y4ZJJVCJ1A?pwd=v8b2>. For more detailed data, please contact the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Bohan Ma and Yushan Xue contributed equally to this work.

Acknowledgments

We would like to express our heartfelt gratitude to all those who have supported and encouraged us throughout the journey of this work. This research is a collaborative project, supported by the National Natural Science Foundation of China under grant no. 12001556, the National Key Research and Development Program of China under the "National Quality Infrastructure System" Key Special Project (no. 2023YFF0614700), the Program for Innovation Research at the Central University of Finance and Economics, the Beijing Social Science Fund Project (no. 15JGC184), the Disciplinary Funds at the Central University of Finance and

Economics, and the Emerging Interdisciplinary Project of CUFÉ.

References

- [1] C. Feng and M. Gao, "An improved arima method based on functional principal component analysis and bidirectional bootstrap and its application to stock price forecasting," *Academic Journal of Computing and Information Science*, vol. 5, no. 10, 2022.
- [2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [3] I. Lloret, J. A. Troyano, F. Enriquez, and J.-J. González-de-la Rosa, "Two deep learning approaches to forecasting disaggregated freight flows: convolutional and encoder–decoder recurrent," *Soft Computing*, vol. 25, no. 12, pp. 7769–7784, 2021.
- [4] M. L. Shen, C. F. Lee, H. H. Liu, P. Y. Chang, and C. H. Yang, "Effective multinational trade forecasting using lstm recurrent neural network," *Expert Systems with Applications*, vol. 182, Article ID 115199, 2021.
- [5] B. Lim, S. Ö Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.
- [6] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long-and short-term temporal patterns with deep neural networks," in *Proceedings of the The 41st international ACM SIGIR conference on research and development in information retrieval*, pp. 95–104, Ann Arbor, MI, USA, July 2018.
- [7] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W. Wang-chun, "Convolutional lstm network: a machine learning approach for precipitation nowcasting," in *Proceedings of the 28th International Conference on Neural Information Processing Systems- Volume 1, NIPS'15*, pp. 802–810, Montreal Canada, December 2015.
- [8] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI'18*, pp. 3634–3640, AAAI Press, Stockholm, Sweden, July 2018.
- [9] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI'19*, pp. 1907–1913, AAAI Press, Macao China, August 2019.
- [10] L. Bai, L. Yao, S. S. Kanhere, X. Wang, and Q. Z. Sheng, "Stg2seq: spatial-temporal graph to sequence model for multi-step passenger demand forecasting," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI'19*, pp. 1981–1987, AAAI Press, Macao China, August 2019.
- [11] I. Mbarek, J. Youness, B. Mohamed, and M. Driss, "A comparative simulation study of classical and machine learning techniques for forecasting time series data," *International Journal of Online and Biomedical Engineering*, vol. 19, no. 8, p. 57, 2023.
- [12] C.-H. Cheng, M.-C. Tsai, and C. Chang, "A time series model based on deep learning and integrated indicator selection method for forecasting stock prices and evaluating trading profits," *Systems*, vol. 10, no. 6, p. 243, 2022.
- [13] S. O. Olukanmi, F. V. Nelwamondo, and N. I. Nwulu, "Utilizing google search data with deep learning, machine learning and time series modeling to forecast influenza-like illnesses in South Africa," *IEEE Access*, vol. 9, Article ID 126822, 2021.
- [14] A. Manowska, A. Rybak, A. Dylong, and J. Pielot, "Forecasting of natural gas consumption in Poland based on arima-lstm hybrid model," *Energies*, vol. 14, no. 24, p. 8597, 2021.
- [15] O. Melnychenko, V. Matskul, and T. Osadcha, "The dynamics of trade relations between Ukraine and Romania: modelling and forecasting," *Virtual Economics*, vol. 5, no. 2, pp. 7–23, 2022.
- [16] N. Ersen and I. Akyüz, "Forecasting foreign trade of Bosnia and Herzegovina for wood and articles of wood, wood charcoal by seasonal arima model," *Periodicals of Engineering and Natural Sciences*, vol. 5, no. 1, 2017.
- [17] M. A. A. Hasin, S. Ghosh, and M. A. Shareef, "An ann approach to demand forecasting in retail trade in Bangladesh," *International Journal of Trade, Economics and Finance*, vol. 2, no. 2, pp. 154–160, 2011.
- [18] A. Farooqi, "Arima model building and forecasting on imports and exports of Pakistan," *Pakistan Journal of Statistics and Operation Research*, vol. 10, no. 2, pp. 157–168, 2014.
- [19] J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, and A. Lachhab, "Forecasting of demand using arima model," *International Journal of Engineering Business Management*, vol. 10, Article ID 184797901880867, 2018.
- [20] H.-L. Wong, Y.-H. Tu, and C.-C. Wang, "An evaluation of comparison between multivariate fuzzy time series with traditional time series model for forecasting taiwan export," in *Proceedings of the 2009 WRI world congress on computer science and information engineering*, vol. 7, pp. 462–467, IEEE, Los Angeles, CA, USA, April 2009.
- [21] C.-C. Wang, "A comparison study between fuzzy time series model and arima model for forecasting taiwan export," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9296–9304, 2011.
- [22] H.-L. Wong, Y.-H. Tu, and C.-C. Wang, "Application of fuzzy time series models for forecasting the amount of taiwan export," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1465–1470, 2010.
- [23] W. Guanghui, "Demand forecasting of supply chain based on support vector regression method," *Procedia Engineering*, vol. 29, pp. 280–284, 2012.
- [24] C.-J. Lu and Y.-W. Wang, "Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting," *International Journal of Production Economics*, vol. 128, no. 2, pp. 603–613, 2010.
- [25] Q. Wu, "Product demand forecasts using wavelet kernel support vector machine and particle swarm optimization in manufacture system," *Journal of Computational and Applied Mathematics*, vol. 233, no. 10, pp. 2481–2491, 2010.
- [26] R. J. Kuo and P. S. Li, "Taiwanese export trade forecasting using firefly algorithm based k-means algorithm and svr with wavelet transform," *Computers and Industrial Engineering*, vol. 99, pp. 153–161, 2016.
- [27] L. Narsimhaiah, P. K. Sahu, K. Sinha, S. Herojit Singh, S. Dey, and P. Pandit, "Forecasting of coconut production in India: an approach with arima, arimax and combined forecast techniques," *International Journal of Current Microbiology and Applied Sciences*, vol. 8, no. 11, pp. 1710–1719, 2019.
- [28] Y. Rashed, H. Meersman, E. Van de Voorde, and T. Vanelslander, "Short-term forecast of container throughput: an arima-intervention model for the port of antwerp," *Maritime Economics and Logistics*, vol. 19, no. 4, pp. 749–764, 2017.

- [29] M. Gopinath, F. A. Batarseh, J. Beckman, A. Kulkarni, and S. Jeong, "International agricultural trade forecasting using machine learning," *Data and Policy*, vol. 3, Article ID e1, 2021.
- [30] T. M. Ghazal, S. Noreen, R. A. Said et al., "Energy demand forecasting using fused machine learning approaches," *Intelligent Automation and Soft Computing*, vol. 31, no. 1, pp. 539–553, 2022.
- [31] M. Yahşi, E. Çanakoğlu, and S. Ağralı, "Carbon price forecasting models based on big data analytics," *Carbon Management*, vol. 10, no. 2, pp. 175–187, 2019.
- [32] H. Lu, X. Ma, K. Huang, and M. Azimi, "Carbon trading volume and price forecasting in China using multiple machine learning models," *Journal of Cleaner Production*, vol. 249, Article ID 119386, 2020.
- [33] S. Bouktif, A. Fiaz, A. Ouni, and M. Serhani, "Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches," *Energies*, vol. 11, no. 7, p. 1636, 2018.
- [34] D. M. Dimiduk, E. A. Holm, and S. R. Niezgodá, "Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering," *Integrating Materials and Manufacturing Innovation*, vol. 7, no. 3, pp. 157–172, 2018.
- [35] M. Chierigato, F. Frangiamore, M. Morassi et al., "A hybrid machine learning/deep learning covid-19 severity predictive model from ct images and clinical data," *Scientific Reports*, vol. 12, no. 1, p. 4329, 2022.
- [36] Y. Qin, D. Song, H. Cheng, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI'17*, pp. 2627–2633, AAAI Press, Melbourne, Australia, August 2017.
- [37] B. N. Oreshkin, D. Carпов, N. Chapados, and Y. Bengio, "N-beats: neural basis expansion analysis for interpretable time series forecasting," in *Proceedings of the International Conference on Learning Representations*, Addis Ababa, Ethiopia, April 2020.
- [38] Y. Li, Y. Rose, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: data-driven traffic forecasting," in *Proceedings of the International Conference on Learning Representations*, Vancouver, Canada, April 2018.
- [39] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, San Francisco, CA, USA, February 2017.
- [40] Y. Seo, M. Defferrard, P. Vandergheynst, and X. Bresson, "Structured sequence modeling with graph convolutional recurrent networks," in *Proceedings of the Neural Information Processing: 25th International Conference, ICONIP 2018*, pp. 362–373, Springer, Siem Reap, Cambodia, December 2018.
- [41] G. Zheng, W. K. Chai, and V. Katos, "A dynamic spatial-temporal deep learning framework for traffic speed prediction on large-scale road networks," *Expert Systems with Applications*, vol. 195, Article ID 116585, 2022.
- [42] P. Nevavuori, N. Narra, P. Linna, and T. Lipping, "Crop yield prediction using multitemporal uav data and spatio-temporal deep learning models," *Remote Sensing*, vol. 12, no. 23, p. 4000, 2020.
- [43] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of the International conference on machine learning*, pp. 1126–1135, PMLR, Sydney, Australia, August 2017.
- [44] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, Stockholm, Sweden, July 2018.
- [45] A. Raghu, M. Raghu, S. Bengio, and O. Vinyals, "Rapid learning or feature reuse? towards understanding the effectiveness of maml," in *Proceedings of the International Conference on Learning Representations*, Addis Ababa, Ethiopia, April 2020.
- [46] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "Stl: a seasonal-trend decomposition," *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.
- [47] Z. Lin, Z. Zhao, Z. Zhang, H. Baoxing, and J. Yuan, "To learn effective features: understanding the task-specific adaptation of maml," 2021, <https://openreview.net/forum?id=FPpZrRfz6Ss>.
- [48] H. Yuan, G. Li, Z. Bao, and L. Feng, "Effective travel time estimation: when historical trajectories over road networks matter," in *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, SIGMOD '20*, pp. 2135–2149, Portland, OR, USA, June 2020.
- [49] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis, Forecasting and Control*, Holden Day, Port Melbourne, Australia, 1976.
- [50] C. A. Sims, "Macroeconomics and reality," *Econometrica*, vol. 48, pp. 1–48, 1980.
- [51] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [52] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [53] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proceedings of the NIPS 2014 Workshop on Deep Learning*, Montreal, Canada, December 2014.
- [54] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "Deepar: probabilistic forecasting with autoregressive recurrent networks," *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020.
- [55] C. Challu, K. G. Olivares, B. N. Oreshkin, F. Garza, M. Mergenthaler-Canseco, and A. Dubrawski, "Nhits: neural hierarchical interpolation for time series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 6989–6997, AAAI, Washington, DC, USA, February 2023.
- [56] Z. Ouyang, P. Ravier, and M. Jabloun, "Stl decomposition of time series can benefit forecasting done by statistical methods but not by machine learning ones," *Engineering Proceedings*, vol. 5, no. 1, p. 42, 2021.
- [57] J. Shiskin, A. H. Young, and J. C. Musgrave, *The X-11 Variant of the Census Method Ii Seasonal Adjustment Program*, Bureau of the Census, Suitland, MD, USA, 1967.
- [58] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [59] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.