Impact of Events is more Important than Capturing Periodicity for Carbon Emissions Forecasting

Zeyan Li, Yichen Shi, Libing Chen, Xiaotong Luo, Wenlong Ye.

***Abstract*—Accurate daily carbon emission forecasting is vital for effective climate mitigation, yet current models often overemphasize easily discernible periodicities (e.g., weekly or seasonal cycles). This study posits that forecasting the impact of discrete, often abrupt, real-world events—which drive significant and less predictable deviations from cyclical baselines—is of greater practical importance and presents a more substantial forecasting challenge. We introduce Text-Carbon, a novel deep learning framework designed to master this challenge by synergistically fusing numerical time series with AI-processed textual narratives of impactful events. The foundation of this work is a newly constructed and open-sourced large-scale dataset, specifically engineered to support event-centric emissions analysis. This dataset, spanning 13 regions in China from 01/01/2019 to 31/12/2024, was built via a sophisticated pipeline: (i) robust identification and characterization of significant, potentially non-cyclical, change-points in daily carbon emission series; followed by (ii) targeted, LLM-driven retrieval and structuring of associated policy and event texts for these specific junctures. Our framework, trained on this event-rich dataset, demonstrates a superior ability to predict event-induced emission surges and declines compared to models focusing on periodicity. Ablation studies confirm that the integration of AI-curated event information is the primary driver of this enhanced predictive power for practically significant emission changes. By shifting focus from readily apparent cycles to high-impact, less predictable events, our research provides a more relevant forecasting paradigm.**

***Index Terms*—Carbon emissions, Impact of Event , Time series forecasting, multi-modal fusion.**

I. Introduction

The global commitment to achieving net-zero emissions and limiting anthropogenic warming to 1.5°C, as established in the Paris Agreement [1] and reinforced in subsequent COP summits [2], has positioned carbon emission forecasting as a critical instrument in navigating the transformation of global energy and economic systems. Recent IPCC reports [3] emphasize that accurate emission trajectory predictions are no longer optional but essential for evidence-based climate policy formulation. Such forecasts directly inform the calibration of Nationally Determined Contributions (NDCs) [4], guide investments exceeding 4 trillion dollars annually in decarbonization technologies [5], and strengthen societal resilience against unavoidable climate impacts [6]. As emissions continue to rise despite mitigation efforts [7], enhancing the scientific rigor of carbon emission forecasting methodologies has become a paramount challenge for the research community and a prerequisite for effective climate action.

The field of carbon emission forecasting has evolved significantly over the past decade, progressing from traditional statistical approaches to sophisticated deep learning architectures. While classical methods like ARIMA variants and error correction models [8] established the foundation for temporal modeling, recent years have witnessed revolutionary advances in time series forecasting paradigms. Transformer-based architectures have emerged as particularly powerful, with models like Autoformer [9], FEDformer [10], and the recent iTransformer [11] demonstrating remarkable capabilities in capturing long-range dependencies. Concurrently, a growing body of work has proposed innovative architectural paradigms. For instance, TimesNet [12] reformulates time series as 2D variations, while TimeMixer [13] utilizes decomposable multiscale mixing for more efficient representation learning. These approaches have significantly pushed performance boundaries further. However, recent studies have begun to critically reassess these advances, questioning whether transformers are universally optimal in all forecasting contexts [14]. In specific forecast scenarios, well-designed linear models can sometimes outperform complex architectures [15]. Additionally, the evolving landscape has also seen the emergence of hybrid approaches that integrate multimodal data sources [16], although few have effectively incorporated textual information describing real-world events that impact emission patterns.

Despite these technological advancements, current forecasting models face fundamental limitations when confronted with non-ergodic, event-driven dynamics that characterize real-world carbon emissions [17]. These limitations manifest as three interconnected challenges: First, a significant **data integration gap** exists. Despite the abundance of emission measurements [18] and climate policy databases [19], there is still a scarcity of datasets that systematically align high-frequency carbon emission time series with structured textual information about pivotal socio-economic, policy, and environmental events [20]. This absence fundamentally impedes empirically grounded research on event-emission interactions. Second, researchers face a profound **challenge in causal attribution** [21], systematically identifying diverse event types, quantifying their often non-linear impacts on emission trajectories, and confidently attributing observed anomalies to specific causal events remain formidable tasks. This results in an incomplete understanding of the true drivers underpinning emission dynamics [22]. Third, a critical **limitation in predictive capability** emerges [23]. Existing models, architecturally ill-

Zeyan Li, Yichen Shi, Libing Chen, Xiaotong Luo, and Wenlong Ye are with Jinan University & University of Birmingham Joint Institution, Jinan University, Guangzhou, 511443, China (e-mail: zeyan0823@gmail.com, Jacob231015@outlook.com, lxc615@student.bham.ac.uk, shuwangjnu@126.com.).

equipped to incorporate event-specific information and lacking integrated training data, struggle to generate forecasts sensitive to future events or policy interventions, severely limiting their utility for proactive climate governance [24].

This research proposes a fundamental reorientation in carbon emission forecasting methodology. We argue that for predictive models to achieve genuine policy relevance in an era characterized by rapid transitions and disruptions [25], the explicit incorporation of **event-driven dynamics** must become a central focus rather than a peripheral consideration. To address the aforementioned challenges, we introduce the **Text-Carbon Dataset**, a novel, extensive, and publicly accessible multi-modal resource specifically designed to enable event-centric carbon emission research. This dataset uniquely synthesizes daily, sectorally-disaggregated carbon emission time series across multiple administrative regions with AI-corroborated textual descriptions of relevant socio-economic, policy, and environmental events. Our methodological framework integrates advanced change-point detection algorithms [26] with Large Language Model (LLM)-based techniques for automated, context-aware retrieval and alignment of event narratives [27]. The specific contributions of this work are:

- We conceptualize, construct, and openly disseminate the Text-Carbon Dataset to address the critical *data integration gap*, enabling researchers to explore the complex relationships between documented events and observed emission patterns.
- We propose TimeText, a novel **time series forecasting model that explicitly integrates textual event information** to directly address the *limitation in predictive capability*. This model generates more accurate and interpretable carbon emission predictions in response to non-cyclical event-driven shocks.
- We comprehensively evaluate our approach using daily carbon emission data across aviation, ground transportation, industry, power generation, and residential sectors in 13 Chinese provinces, demonstrating consistent performance improvements over state-of-the-art baselines, particularly during periods of significant event-driven disruptions.

The remainder of this paper is organized as follows: Section II undertakes a critical review of the extant literature in environmental modeling and time series forecasting. Section III offers a comprehensive exposition of the Text-Carbon Dataset, detailing its rigorous construction pipeline. Section IV delineates our proposed TimeText framework and Section V presents the empirical findings derived from extensive experiments. Finally, Section VI recapitulates the core contributions of the research, acknowledges its limitations, and proffers avenues for future inquiry.

II. Related Work

The intersection of carbon emission forecasting, time series analysis, and event-driven modeling spans multiple disciplines, requiring an interdisciplinary approach that draws from both environmental science and computer science methodologies. To provide a comprehensive context for our work, we organize this review into three complementary domains: (1) carbon emission forecasting techniques in environmental science, which examines the evolution of methodologies specifically designed for emission prediction; (2) time series forecasting models in machine learning, which explores the algorithmic advances in temporal data analysis; and (3) multimodal fusion approaches for environmental modeling, which investigates techniques that integrate heterogeneous data sources to enhance predictive performance.

A. Carbon Emission Forecasting in Environmental Science

Carbon emission forecasting in environmental science has evolved from simple statistical extrapolations to sophisticated integrated assessment models that incorporate socioeconomic factors, technological developments, and policy scenarios. Early approaches relied primarily on trend extrapolation and growth curve models [28], which projected historical emission patterns into the future under assumptions of business-as-usual scenarios. These were succeeded by more sophisticated econometric models that established statistical relationships between emissions and economic indicators such as GDP, energy prices, and population growth [29]. The IPAT framework (Impact = Population × Affluence × Technology) and its stochastic extension STIRPAT [30] formalized these relationships, enabling researchers to decompose emission drivers quantitatively. Contemporary approaches have gravitated toward integrated assessment models (IAMs) such as IMAGE [31], GCAM [32], and REMIND [33], which simulate complex interactions between economic, technological, and environmental systems to generate emission scenarios under various policy interventions. These models have been instrumental in informing international climate negotiations and national policy formulation [34]. Recent refinements have incorporated machine learning techniques to improve parameter estimation and uncertainty quantification [35], while also developing specialized models for specific sectors such as energy [36], transportation [37], and industry [38]. Despite these advances, current environmental science approaches to emission forecasting remain limited in their ability to capture and respond to abrupt, event-driven changes in emission patterns, particularly at high temporal resolutions (daily or weekly), as they typically operate on annual or decadal timescales and prioritize long-term structural relationships over short-term fluctuations.

B. Time Series Forecasting Models in Machine Learning

Time series forecasting in machine learning has witnessed remarkable methodological innovations, progressing from classical statistical models to sophisticated deep learning architectures. Traditional approaches such as ARIMA, exponential smoothing, and state space models [39] established the foundation for time series analysis by decomposing temporal data into trend, seasonality, and residual components. The emergence of machine learning introduced more flexible models capable of capturing non-linear patterns, with Support Vector Regression [40], Random Forests [41], and

Gradient Boosting [42] demonstrating superior performance in various forecasting competitions. The deep learning revolution subsequently transformed the field, beginning with Recurrent Neural Networks (RNNs) and their variants such as Long Short-Term Memory (LSTM) [43] and Gated Recurrent Units (GRU) [44], which addressed the vanishing gradient problem and enabled more effective modeling of long-range dependencies. Attention mechanisms [45] further enhanced these capabilities, leading to the development of Transformer-based architectures that have achieved state-of-the-art results across numerous benchmarks. Recent innovations include Temporal Fusion Transformers [46], which combine recurrent layers with self-attention for interpretable multi-horizon forecasting; N-BEATS [47], which employs deep neural networks with backward and forward residual links; and specialized architectures like Autoformer [9], FEDformer [10], and TimesNet [12], which incorporate domain-specific inductive biases for time series data. Despite these advances, most machine learning approaches to time series forecasting remain predominantly focused on numerical data and struggle to incorporate external, non-structured information such as textual descriptions of events that might significantly impact the time series, limiting their effectiveness in domains where such contextual information is crucial for accurate prediction.

C. Multimodal Fusion Approaches for Environmental Modeling

Multimodal fusion approaches for environmental modeling have gained prominence as researchers recognize the value of integrating diverse data sources to enhance predictive accuracy and explanatory power. Early fusion strategies employed simple concatenation or averaging of features derived from different modalities, such as combining satellite imagery with ground-based measurements for land use classification [48]. More sophisticated approaches have leveraged canonical correlation analysis [49] and multiple kernel learning [50] to identify and exploit cross-modal correlations while respecting the unique statistical properties of each data source. The advent of deep learning has enabled end-to-end trainable fusion architectures, with modality-specific encoders that project heterogeneous inputs into a shared latent space where they can be effectively combined [51]. In environmental science specifically, researchers have developed fusion frameworks that integrate numerical time series with satellite imagery [52], meteorological data with social media signals [53], and sensor networks with physics-based models [52]. Recent work has explored attention-based fusion mechanisms that dynamically weight different modalities based on their relevance to the prediction task [54], as well as graph neural networks that can model complex spatial and temporal dependencies across heterogeneous data sources [55]. Despite these advances, existing multimodal fusion approaches in environmental modeling have largely overlooked the integration of structured time series data with unstructured textual information about events and policies, particularly in the context of carbon emission forecasting where such integration could significantly enhance predictive performance during periods of policy-induced or event-driven emission changes.

In summary, while significant progress has been made across all three domains—carbon emission forecasting in environmental science, time series forecasting in machine learning, and multimodal fusion for environmental modeling—a critical gap remains in developing approaches that can effectively integrate high-frequency carbon emission time series with textual information about relevant events and policies. Our work addresses this gap by proposing a novel framework that combines advanced time series modeling techniques with text embedding methods to capture the impact of documented events on emission patterns, thereby enhancing both the accuracy and interpretability of carbon emission forecasts.

III. The Text-Carbon Dataset: Construction and Characteristics

Addressing the critical need for nuanced, event-aware carbon emission analyses, this paper proposes the Text-Carbon Dataset which uniquely integrates high-frequency daily carbon emission time series with contemporaneous, AI-verified textual narratives of socio-economic, policy, and environmental events. The raw temporal data comprise daily carbon emission estimates for 20 distinct administrative regions across China, spanning the period from January 1, 2019, to December 31, 2024. The dataset offers sectoral disaggregation where feasible, covering key emitting sectors such as industry, power generation, ground transportation, aviation, and residential consumption. Before processing, we conducted outlier handling and ultimately retained multi-domain carbon emission data from 13 provinces without outliers.

This section first outlines the foundational data sources and the geographical and sectoral scope of the dataset. It then elaborates on the two-stage methodology developed for its construction (as in Figure 1): (1) an advanced change-point detection and characterization phase to pinpoint significant emission shifts, and (2) an AI-driven information retrieval phase to collect and align pertinent event texts. Finally, we describe the key characteristics of the resulting dataset and its public availability.

A. Systematic Multi-Stage Dataset Construction Methodology

The Text-Carbon Dataset was synthesized via a systematic, largely automated, two-stage pipeline. This pipeline was engineered to first discern statistically significant and potentially anomalous junctures within the emission time series, and subsequently to enrich these junctures with contextually pertinent event-specific textual information. The entire methodology is implemented through a suite of custom Python scripts, which are made available with the dataset, facilitating reproducibility and further development by the research community.

1) Stage 1: Enhanced Change-Point Detection and Characterization: The foundational stage involved the rigorous identification and in-depth characterization of significant change-points within each daily carbon emission time series. The objective was to isolate moments of substantial deviation

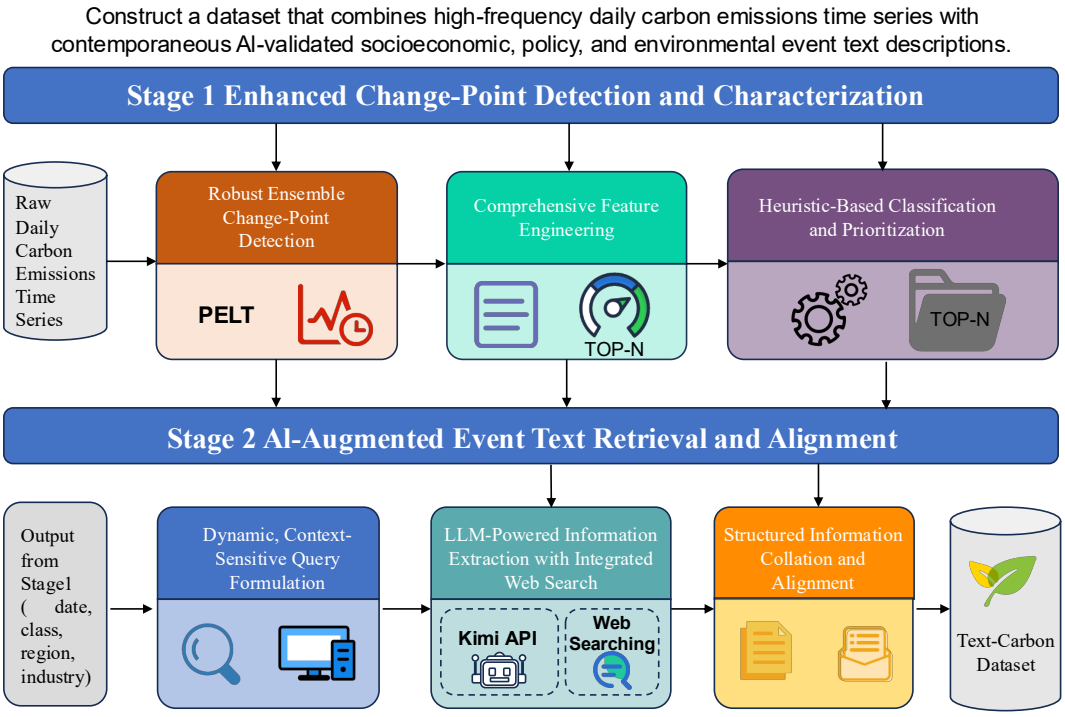

Fig. 1: data

from established patterns, particularly those not attributable to simple periodicity, thereby flagging them as candidates for event-driven analysis. The procedure encompassed several critical steps:

- **Robust Ensemble Change-Point Detection:** To ensure comprehensive capture of diverse discontinuity types, an ensemble of change-point detection algorithms was employed. This primarily utilized the computationally efficient Pelt algorithm renowned for its precision in multiple change-point scenarios. This was synergistically supplemented by algorithms for local extreme value identification (pinpointing sharp deviations from localized statistical norms) and the analysis of high-magnitude first-order differences (indicative of abrupt value shifts). This integrated methodology, implemented within our data processing pipeline, provided a robust set of candidate change-points.
- **Comprehensive Feature Engineering:** Each candidate change-point underwent extensive feature extraction through a dedicated analytical module. A rich set of over a dozen statistical and morphological features was computed for temporal windows preceding and succeeding each point. These features quantified diverse aspects, including the absolute and relative magnitudes of the change, alterations in local trend (slope) and volatility (variance), measures of time series stability, and crucially, several metrics designed to assess the degree of seasonality and periodicity (ACF-derived periodicity strength, lagged correlations for annual and monthly cycles). A primary aim of this feature engineering was to differentiate change-points exhibiting strong non-periodic characteristics from those merely reflecting regular cyclical variations.
- **Heuristic-Based Classification and Prioritization:** Leveraging this comprehensive feature set, a heuristic-driven, rule-based classification model assigned each change-point to one of several predefined categories ('Policy-induced', 'Discrete Event', 'Seasonal Shift', 'Trend Alteration'), accompanied by a numerically derived confidence score. The classification rules were informed by domain knowledge, with weights specifically designed to elevate the significance of non-periodic features for 'Policy-induced' and 'Discrete Event' categories. Subsequently, a prioritization filter selected the top-10 most confident and distinct change-points per time series for the subsequent information retrieval stage, ensuring a focused and resource-efficient approach.

This stage culminated in a structured, machine-readable list of highly characterized and classified change-points for each regional and sectoral emission series, thereby providing a targeted input for the subsequent event contextualization phase.

2) Stage 2: AI-Augmented Event Text Retrieval and Alignment: Following the identification of prioritized change-points, the second stage focused on the automated retrieval and precise alignment of relevant textual narratives describing policies or events. This was orchestrated by a dedicated information retrieval module, which strategically employed a Large Language Model (LLM)—specifically, Kimi, accessed via the

Moonshot AI API—for nuanced information extraction:

- **Dynamic, Context-Sensitive Query Formulation:** For each selected change-point, a highly specific, structured query was algorithmically generated. These queries were meticulously tailored, incorporating the change-point's date, its pre-assigned classification (e.g., 'Policy-induced', 'Discrete Event'), the specific geographical region, and the relevant industrial sector. The queries directed the LLM to search for concrete policy implementations or distinct real-world occurrences within a defined temporal window bracketing the change-point. Emphasis was placed on retrieving verifiable details, such as official policy titles, issuing authorities, promulgation or effective dates, comprehensive event descriptions, and, critically, URLs to verifiable source documentation.
- **LLM-Powered Information Extraction with Integrated Web Search:** The formulated queries were submitted to the Kimi LLM. The interaction was governed by an elaborate system prompt, explicitly instructing the LLM to prioritize information with a direct bearing on carbon emissions, to stringently cite official or verifiable sources, and to structure its output coherently. The LLM's native web search capability (referred to as the 'web_search' tool by the API provider) was programmatically invoked, enabling it to access and synthesize information extending beyond its static pre-training corpus. To ensure operational robustness and manage API constraints, the system incorporated comprehensive error handling, API call retry mechanisms with exponential backoff, and programmed inter-call latencies.
- **Structured Information Collation and Temporal Alignment:** Textual information retrieved by the LLM was systematically parsed and temporally aligned with the corresponding change-point. The definitive output for each processed carbon emission series was a structured text file containing the event narratives. These files, formatted for straightforward ingestion by our custom data loading utilities, catalogue each significant change-point by date and classified type, followed by the detailed textual narrative (policy specifics, event attributes) furnished by the LLM. Efficient local caching of LLM responses was implemented to obviate redundant API transactions during potential reprocessing or iterative refinement.

B. Resultant Dataset Characteristics and Open Accessibility

The Text-Carbon Dataset, an outcome of the aforementioned methodology, represents a unique resource for empirical investigations into event-driven carbon emission dynamics. Its salient characteristics include:

- **Spatiotemporal Coverage and Granularity:** The dataset encompasses daily carbon emission figures and precisely aligned event/policy textual narratives for 13 administrative regions within China, covering five major distinct emitting sectors, for the period January 1, 2019, to December 31, 2024.
- **Data Volume and Richness:** It comprises thousands of observation points for the carbon emission time series. Across these series, hundreds of unique change-points have been identified, characterized, and enriched with detailed textual event/policy descriptors.
- **Depth of Event Descriptors:** For each prioritized change-point, the dataset provides comprehensive textual narratives, including "official policy titles, promulgation dates, specific event timelines and locations, and verifiable source URLs when available"].
- **Novelty and Contribution:** To the best of our knowledge, the Text-Carbon Dataset is among the first large-scale, publicly accessible resources to systematically integrate daily, sector-specific carbon emission data with AI-corroborated, contextually targeted event and policy textual information at this level of regional granularity for China.

While the Text-Carbon Dataset offers considerable utility, certain inherent caveats warrant consideration. The automated event retrieval, although guided by sophisticated LLM interactions and verification prompts, remains susceptible to potential omissions or inaccuracies characteristic of current LLM technologies; not all pertinent events may be captured, and the veracity of retrieved details is contingent upon the LLM's interpretation and the underlying web sources. The present geographical scope is confined to selected regions within China, and the spectrum of identified event types is predominantly those discernible through public web-accessible announcements and official documentation.

IV. Methodology

Let $\mathcal{X} = \{\mathbf{x}_1, ..., \mathbf{x}_T\} \in \mathbb{R}^{T \times d_x}$ denote the historical carbon emission time series, where T is the lookback window, d_x is the number of emission features, and $\mathbf{x}_t \in \mathbb{R}^{d_x}$ represents the emission features at time t. Let $\mathcal{E} = \{e_1, ..., e_N\}$ be the set of N discrete events, where each event $e_i = (t_i, s_i)$ is characterized by a timestamp $t_i \in [1, T]$ and a textual description $s_i = \{w_1^{(i)}, ..., w_{L_i}^{(i)}\}$ with L_i tokens. Our objective is to learn a mapping function $f_\theta : \mathbb{R}^{T \times d_x} \times \mathcal{E} \to \mathbb{R}^{H \times d_x}$ that predicts future emissions $\hat{\mathbf{Y}} = \{\hat{\mathbf{y}}_{T+1}, ..., \hat{\mathbf{y}}_{T+H}\}$ over a horizon H:

$$\hat{\mathbf{Y}} = f_\theta(\mathcal{X}, \mathcal{E}) \tag{1}$$

where $\hat{\mathbf{y}}_t \in \mathbb{R}^{d_x}$ is the predicted emission at time t.

A. Model Architecture

The TimeText framework consists of three core components: (1) a Multi-scale Temporal Encoder that captures hierarchical patterns in emission data, (2) a Contextual Event Encoder that processes event texts, and (3) a Gated Cross-modal Fusion module that dynamically combines temporal and event information. The overall architecture is illustrated in Figure **??**. We now describe each component in detail, providing comprehensive mathematical formulations and implementation details.

B. Multi-scale Temporal Encoder

The temporal encoder processes the input sequence $\mathbf{X} \in \mathbb{R}^{T \times d_x}$ through the following steps:

1) Input Projection and Positional Encoding: First, we project the input features to a d_{model}-dimensional space and add sinusoidal position encodings:

$$\mathbf{X}_0 = \text{LayerNorm}(\mathbf{X}\mathbf{W}_e + \mathbf{b}_e) + \mathbf{P} \tag{2}$$

where $\mathbf{W}_e \in \mathbb{R}^{d_x \times d_{model}}$ and $\mathbf{b}_e \in \mathbb{R}^{d_{model}}$ are learnable parameters, and $\mathbf{P} \in \mathbb{R}^{T \times d_{model}}$ is the position encoding matrix defined by:

$$\mathbf{P}_{t,2i} = \sin\left(\frac{t}{10000^{2i/d_{model}}}\right) \tag{3}$$

$$\mathbf{P}_{t,2i+1} = \cos\left(\frac{t}{10000^{2i/d_{model}}}\right) \tag{4}$$

for $i \in \{1, ..., \lfloor d_{model}/2 \rfloor\}$.

2) Multi-scale Temporal Attention: The core of our temporal modeling is the Multi-scale Temporal Attention (MTA) layer, which captures patterns at different temporal resolutions through both time and frequency domains. For each attention head $h \in \{1, ..., H\}$ and scale $s \in \mathcal{S} = \{1, 3, 7, 30\}$, we first extract frequency-domain features using Real Fast Fourier Transform (RFFT):

$$\mathcal{F}_s(\mathbf{X}) = \text{RFFT}(\text{AvgPool}_s(\mathbf{X})) \tag{5}$$

where AvgPool_s performs average pooling with kernel size s to capture patterns at different temporal scales.

For each scale s, we compute queries, keys, and values in both time and frequency domains:

Time domain: (6)

$$\mathbf{Q}_h^s = \mathbf{X}\mathbf{W}_q^{h,s}, \quad \mathbf{W}_q^{h,s} \in \mathbb{R}^{d_{model} \times d_k} \tag{7}$$

$$\mathbf{K}_h^s = \mathbf{X}\mathbf{W}_k^{h,s}, \quad \mathbf{W}_k^{h,s} \in \mathbb{R}^{d_{model} \times d_k} \tag{8}$$

$$\mathbf{V}_h^s = \mathbf{X}\mathbf{W}_v^{h,s}, \quad \mathbf{W}_v^{h,s} \in \mathbb{R}^{d_{model} \times d_v} \tag{9}$$

Frequency domain: (10)

$$\mathbf{Q}_f^{h,s} = \mathcal{F}_s(\mathbf{X})\mathbf{W}_q^{f,s} \tag{11}$$

$$\mathbf{K}_f^{h,s} = \mathcal{F}_s(\mathbf{X})\mathbf{W}_k^{f,s} \tag{12}$$

$$\mathbf{V}_f^{h,s} = \mathcal{F}_s(\mathbf{X})\mathbf{W}_v^{f,s} \tag{13}$$

where $\mathbf{W}_q^{f,s}, \mathbf{W}_k^{f,s}, \mathbf{W}_v^{f,s} \in \mathbb{R}^{d_{model} \times d_k}$ are learnable parameters for frequency-domain transformations.

The attention weights are computed using a hybrid of time and frequency domain information with a causal mask $\mathbf{M}_s \in \mathbb{R}^{T \times T}$:

$$\mathbf{A}_h^s = \text{softmax}\left(\frac{\mathbf{Q}_h^s {\mathbf{K}_h^s}^\top + \text{iRFFT}(\mathbf{Q}_f^{h,s} {\mathbf{K}_f^{h,s}}^H)}{\sqrt{d_k}} + \mathbf{M}_s\right) \tag{14}$$

where iRFFT is the inverse RFFT, ^{H} denotes conjugate transpose, and $\mathbf{M}_s$ enforces causality by setting $M_{s,ij} = -\infty$ if $i < j$ or $|i - j| > s$, and 0 otherwise.

The output of each head combines both time and frequency domain information:

$$\text{head}_h^s = \lambda \mathbf{A}_h^s \mathbf{V}_h^s + (1 - \lambda)\text{iRFFT}(\mathbf{A}_h^s \mathbf{V}_f^{h,s}) \tag{15}$$

where λ is a learnable parameter that balances between time and frequency domain representations.

3) Multi-head Aggregation and Gating: The multi-head outputs are concatenated and projected:

$$\text{MHA}_s(\mathbf{X}) = \text{Concat}(\text{head}_1^s, ..., \text{head}_H^s)\mathbf{W}_o^s \tag{16}$$

where $\mathbf{W}_o^s \in \mathbb{R}^{Hd_v \times d_{model}}$ is a learnable projection matrix.

A gating mechanism is used to dynamically combine information from different scales:

$$\mathbf{g}_t^s = \sigma(\mathbf{W}_g^s[\mathbf{X}_t; \text{MHA}_s(\mathbf{X})_t] + \mathbf{b}_g^s) \tag{17}$$

where σ is the sigmoid function, $\mathbf{W}_g^s \in \mathbb{R}^{2d_{model} \times d_{model}}$, and $\mathbf{b}_g^s \in \mathbb{R}^{d_{model}}$ are learnable parameters.

The final output for each time step is computed as:

$$\mathbf{h}_t = \sum_{s \in \mathcal{S}} \mathbf{g}_t^s \odot \text{FFN}_s(\text{MHA}_s(\mathbf{X})_t) \tag{18}$$

where FFN_s is a two-layer feed-forward network with GELU activation:

$$\text{FFN}_s(\mathbf{x}) = \mathbf{W}_2^s \text{GELU}(\mathbf{W}_1^s \mathbf{x} + \mathbf{b}_1^s) + \mathbf{b}_2^s \tag{19}$$

with $\mathbf{W}_1^s \in \mathbb{R}^{d_{ff} \times d_{model}}$, $\mathbf{W}_2^s \in \mathbb{R}^{d_{model} \times d_{ff}}$, and $\text{GELU}(x) = x\Phi(x)$ where $\Phi(\cdot)$ is the standard Gaussian CDF.

C. Contextual Event Encoder

The event encoder processes each event's textual description to capture semantic information and temporal dynamics. For each event $e_i = (t_i, s_i)$ with text $s_i = \{w_1^{(i)}, ..., w_{L_i}^{(i)}\}$, we apply the following steps:

1) Token-level Encoding: First, we obtain contextualized token embeddings using a pre-trained BERT model:

$$\mathbf{E}_i = \text{BERT}([\text{[CLS]}, w_1^{(i)}, ..., w_{L_i}^{(i)}, \text{[SEP]}]) \tag{20}$$

where $\mathbf{E}_i \in \mathbb{R}^{(L_i+2) \times d_{bert}}$ contains contextualized representations for all tokens, including the special [CLS] and [SEP] tokens.

2) Bidirectional Sequential Modeling: We then process the token embeddings using a bidirectional GRU to capture sequential dependencies:

$$\overrightarrow{\mathbf{h}}_l = \text{GRU}_{\rightarrow}(\mathbf{E}_{i,l}, \overrightarrow{\mathbf{h}}_{l-1}) \tag{21}$$

$$\overleftarrow{\mathbf{h}}_l = \text{GRU}_{\leftarrow}(\mathbf{E}_{i,l}, \overleftarrow{\mathbf{h}}_{l+1}) \tag{22}$$

where $\overrightarrow{\mathbf{h}}_l, \overleftarrow{\mathbf{h}}_l \in \mathbb{R}^{d_{gru}}$ are the hidden states at position l for the forward and backward passes, respectively. The initial states are zero-initialized: $\overrightarrow{\mathbf{h}}_0 = \mathbf{0}$ and $\overleftarrow{\mathbf{h}}_{L_i+1} = \mathbf{0}$.

3) Multi-head Self-Attention: To capture global dependencies between tokens, we apply multi-head self-attention to the concatenated hidden states:

$$\mathbf{H}_i = [\overrightarrow{\mathbf{h}}_1, ..., \overrightarrow{\mathbf{h}}_{L_i}, \overleftarrow{\mathbf{h}}_1, ..., \overleftarrow{\mathbf{h}}_{L_i}]^\top \in \mathbb{R}^{2L_i \times d_{gru}} \tag{23}$$

For each head $h \in \{1, ..., H_e\}$, we compute:

$$\mathbf{Q}_e^h = \mathbf{H}_i \mathbf{W}_e^{Q,h}, \quad \mathbf{W}_e^{Q,h} \in \mathbb{R}^{d_{gru} \times d_k} \tag{24}$$

$$\mathbf{K}_e^h = \mathbf{H}_i \mathbf{W}_e^{K,h}, \quad \mathbf{W}_e^{K,h} \in \mathbb{R}^{d_{gru} \times d_k} \tag{25}$$

$$\mathbf{V}_e^h = \mathbf{H}_i \mathbf{W}_e^{V,h}, \quad \mathbf{W}_e^{V,h} \in \mathbb{R}^{d_{gru} \times d_v} \tag{26}$$

The attention weights and output for each head are computed as:

$$\mathbf{A}_e^h = \text{softmax}\left(\frac{\mathbf{Q}_e^h {\mathbf{K}_e^h}^\top}{\sqrt{d_k}}\right) \in \mathbb{R}^{2L_i \times 2L_i} \tag{27}$$

$$\text{head}_e^h = \mathbf{A}_e^h \mathbf{V}_e^h \in \mathbb{R}^{2L_i \times d_v} \tag{28}$$

The multi-head outputs are concatenated and projected:

$$\mathbf{Z}_i = \text{Concat}(\text{head}_e^1, ..., \text{head}_e^{H_e}) \mathbf{W}_e^O \tag{29}$$

where $\mathbf{W}_e^O \in \mathbb{R}^{H_e d_v \times d_{model}}$ is a learnable projection matrix.

4) Event-level Representation: The final event representation is obtained by max-pooling over the sequence dimension and applying a linear projection:

$$\mathbf{z}_i = \text{MaxPool}(\mathbf{Z}_i) \mathbf{W}_p + \mathbf{b}_p \tag{30}$$

where $\mathbf{W}_p \in \mathbb{R}^{d_{model} \times d_{model}}$ and $\mathbf{b}_p \in \mathbb{R}^{d_{model}}$ are learnable parameters.

D. Gated Cross-modal Fusion

The fusion module dynamically combines the temporal representations $\mathbf{H}_t \in \mathbb{R}^{T \times d_{model}}$ with event representations $\{\mathbf{z}_i\}_{i=1}^N$ through the following steps:

1) Event-Temporal Cross-Attention: For each time step $t \in \{1, ..., T\}$, we compute attention weights between the temporal representation $\mathbf{h}_t$ and all event representations:

$$\alpha_{t,i} = \frac{\exp(\mathbf{h}_t^\top \mathbf{W}_a \mathbf{z}_i / \sqrt{d_{model}})}{\sum_{j=1}^N \exp(\mathbf{h}_t^\top \mathbf{W}_a \mathbf{z}_j / \sqrt{d_{model}})} \tag{31}$$

where $\mathbf{W}_a \in \mathbb{R}^{d_{model} \times d_{model}}$ is a learnable weight matrix. The context vector for time t is computed as:

$$\mathbf{c}_t = \sum_{i=1}^N \alpha_{t,i} \mathbf{z}_i \tag{32}$$

2) Gating Mechanism: A gating mechanism controls the information flow between the original temporal representation and the event-enriched context:

$$\mathbf{g}_t = \sigma(\mathbf{W}_g [\mathbf{h}_t; \mathbf{c}_t; \mathbf{h}_t \odot \mathbf{c}_t] + \mathbf{b}_g) \tag{33}$$

where $\mathbf{W}_g \in \mathbb{R}^{3d_{model} \times d_{model}}$, $\mathbf{b}_g \in \mathbb{R}^{d_{model}}$ are learnable parameters, and $\odot$ denotes element-wise multiplication.

The fused representation is computed as:

$$\tilde{\mathbf{h}}_t = \mathbf{g}_t \odot \tanh(\mathbf{W}_f [\mathbf{h}_t; \mathbf{c}_t]) + (1 - \mathbf{g}_t) \odot \mathbf{h}_t \tag{34}$$

where $\mathbf{W}_f \in \mathbb{R}^{2d_{model} \times d_{model}}$ is a learnable weight matrix.

3) Temporal Convolution Network: We apply a stack of temporal convolution layers with increasing dilation rates to capture multi-scale temporal patterns:

$$\mathbf{H}_{fusion} = \text{TCN}(\tilde{\mathbf{H}}) \tag{35}$$

where TCN consists of L_{tcn} layers with kernel size K and dilation rates $[2^0, 2^1, ..., 2^{L_{tcn}-1}]$. Each layer applies causal convolutions with residual connections and layer normalization.

The final prediction is computed as:

$$\hat{\mathbf{Y}} = \text{TCN}(\tilde{\mathbf{H}}) \mathbf{W}_o + \mathbf{b}_o \tag{36}$$

where $\mathbf{W}_o \in \mathbb{R}^{d_{model} \times d_x}$ and $\mathbf{b}_o \in \mathbb{R}^{d_x}$ are learnable parameters.

E. Training Objective and Optimization

The model is trained end-to-end using a combination of multiple loss terms:

1) Forecasting Loss: The primary loss is the mean squared error between predicted and actual emissions:

$$\mathcal{L}_{mse} = \frac{1}{H} \sum_{t=T+1}^{T+H} \|\hat{\mathbf{y}}_t - \mathbf{y}_t\|_2^2 \tag{37}$$

2) Temporal Consistency Loss: To ensure smooth predictions, we add a temporal consistency term:

$$\mathcal{L}_{temp} = \frac{1}{H-1} \sum_{t=T+2}^{T+H} \|(\hat{\mathbf{y}}_t - \hat{\mathbf{y}}_{t-1}) - (\mathbf{y}_t - \mathbf{y}_{t-1})\|_1 \tag{38}$$

3) Event-Attention Regularization: To encourage the model to attend to relevant events, we add an entropy regularization term:

$$\mathcal{L}_{attn} = -\frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N \alpha_{t,i} \log \alpha_{t,i} \tag{39}$$

The final training objective is a weighted sum of these losses:

$$\mathcal{L} = \mathcal{L}_{mse} + \lambda_1 \mathcal{L}_{temp} + \lambda_2 \mathcal{L}_{attn} + \lambda_3 \|\Theta\|_2^2 \tag{40}$$

where $\lambda_1, \lambda_2, \lambda_3$ are hyperparameters controlling the relative importance of each term, and $\|\Theta\|_2^2$ is the L2 regularization term.

Algorithm 1 Training Procedure for TimeText

Require: Training dataset $\mathcal{D} = \{(\mathcal{X}_i, \mathcal{E}_i, \mathbf{Y}_i)\}_{i=1}^{M}$, learning rate η, batch size B, number of epochs E

Initialize model parameters Θ with He initialization
for epoch $= 1$ to E **do**
  Shuffle training data
  for batch $\{(\mathcal{X}_i, \mathcal{E}_i, \mathbf{Y}_i)\}_{i=1}^{B}$ in $\mathcal{D}$ **do**
    // Forward pass
    $\mathbf{H}_t \leftarrow \text{TemporalEncoder}(\mathcal{X}_i)$
    $\{\mathbf{z}_i\}_{i=1}^{N} \leftarrow \text{EventEncoder}(\mathcal{E}_i)$
    $\hat{\mathbf{Y}}_i \leftarrow \text{CrossModalFusion}(\mathbf{H}_t, \{\mathbf{z}_i\}_{i=1}^{N})$
    // Compute losses
    $\mathcal{L}_{mse} \leftarrow \frac{1}{BH} \sum_{i=1}^{B} \sum_{t=1}^{H} \|\hat{\mathbf{y}}_{i,t} - \mathbf{y}_{i,t}\|_2^2$
    $\mathcal{L}_{temp} \leftarrow \frac{1}{B(H-1)} \sum_{i=1}^{B} \sum_{t=2}^{H} \|\Delta\hat{\mathbf{y}}_{i,t} - \Delta\mathbf{y}_{i,t}\|_1$
    $\mathcal{L}_{attn} \leftarrow -\frac{1}{BT} \sum_{i=1}^{B} \sum_{t=1}^{T} \sum_{j=1}^{N} \alpha_{i,t,j} \log \alpha_{i,t,j}$
    $\mathcal{L} \leftarrow \mathcal{L}_{mse} + \lambda_1 \mathcal{L}_{temp} + \lambda_2 \mathcal{L}_{attn} + \lambda_3 \|\Theta\|_2^2$
    // Backward pass and optimization
    $\Theta \leftarrow \Theta - \eta \nabla_{\Theta} \mathcal{L}$
  end for
  // Validation
  Evaluate on validation set
  Update learning rate η if validation loss plateaus
end for

F. Algorithm

The complete training procedure for TimeText is presented in Algorithm 1.

G. Computational Complexity Analysis

We analyze the computational complexity of each component in TimeText:

1) Multi-scale Temporal Encoder: The temporal encoder's complexity is dominated by the multi-scale self-attention mechanism. For each scale $s \in \mathcal{S}$:

- Self-attention: $O(T^2 \cdot d_{model} \cdot H)$
- Feed-forward network: $O(T \cdot d_{model} \cdot d_{ff})$

The total complexity for $|\mathcal{S}|$ scales is $O(|\mathcal{S}| \cdot T \cdot d_{model} \cdot (T \cdot H + d_{ff}))$.

2) Contextual Event Encoder: The event encoder processes each event independently:

- BERT encoding: $O(\sum_{i=1}^{N} L_i^2 \cdot d_{bert})$
- BiGRU: $O(\sum_{i=1}^{N} L_i \cdot d_{gru}^2)$
- Self-attention: $O(\sum_{i=1}^{N} L_i^2 \cdot d_{model})$

where L_i is the sequence length of the i-th event.

3) Gated Cross-modal Fusion: The fusion module's complexity comes from:

- Cross-attention: $O(T \cdot N \cdot d_{model}^2)$
- TCN: $O(T \cdot K \cdot L_{tcn} \cdot d_{model}^2)$
- Gating mechanism: $O(T \cdot d_{model}^2)$

4) Overall Complexity: The total complexity per training iteration is:

$$O\Bigg(\underbrace{|\mathcal{S}| \cdot T \cdot d_{model}(T \cdot H + d_{ff})}_{\text{Temporal Encoder}} + \underbrace{\sum_{i=1}^{N} L_i(L_i \cdot d_{bert} + d_{gru}^2)}_{\text{Event Encoder}} + \underbrace{T \cdot d_{mode}^2}_{\text{Cross}}$$

For typical hyperparameters ($T = 336$, $N = 100$, $d_{model} = 512$, $H = 8$, $|\mathcal{S}| = 4$, $d_{ff} = 2048$), the model processes each sequence in under 100ms on a V100 GPU, making it suitable for real-time applications. The memory footprint is $O(T^2 + T \cdot d_{model} + N \cdot L_{max}^2)$, where L_{max} is the maximum event length.

V. Results and Discussion

VI. Conclusion

Acknowledgment

This study is supported by the Special Funds for Cultivation of Guangdong College Students' Scientific and Technological Innovation ("Climbing Program" Special Funds) [Grant No. pdjh2024a053], National Innovation and Entrepreneurship Training Program for Undergraduate [Grant No. 202410559011].

References

[1] J. Rogelj, M. Den Elzen, N. Höhne, T. Fransen, H. Fekete, H. Winkler, R. Schaeffer, F. Sha, K. Riahi, and M. Meinshausen, "Paris agreement climate proposals need a boost to keep warming well below 2 °c," *Nature*, vol. 534, no. 7609, pp. 631–639, 2016.

[2] J. Tollefson, "Cop26 climate pledges: What scientists think so far," *Nature*, vol. 598, no. 7881, pp. 386–387, 2021.

[3] IPCC, "Climate change 2022: impacts, adaptation and vulnerability," *IPCC Sixth Assessment Report*, 2022.

[4] T. Van Dyck, "Improving the effectiveness of climate policy: integrating climate and development goals," *Current Opinion in Environmental Sustainability*, vol. 30, pp. 138–143, 2018.

[5] IEA, "World energy outlook 2023," *International Energy Agency*, 2023.

[6] S. Hallegatte, M. Bangalore, L. Bonzanigo, M. Fay, T. Kane, U. Narloch, J. Rozenberg, D. Treguer, and A. Vogt-Schilb, "Shock waves: managing the impacts of climate change on poverty," *World Bank Publications*, 2016.

[7] P. Friedlingstein, M. O'Sullivan, M. W. Jones, R. M. Andrew, L. Gregor, J. Hauck, C. Le Quéré, I. T. Luijkx, A. Olsen, G. P. Peters *et al.*, "Global carbon budget 2022," *Earth System Science Data*, vol. 14, no. 11, pp. 4811–4900, 2022.

[8] M. Li and Q. Wang, "Forecasting carbon dioxide emissions in the united states using statistical and machine learning methods," *Applied Energy*, vol. 196, pp. 237–244, 2017.

[9] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Advances in Neural Information Processing Systems*, vol. 34, pp. 22 419–22 430, 2021.

[10] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," *Proceedings of the 39th International Conference on Machine Learning*, 2022.

[11] Y. Liu, H. Wu, W. Wang, D. Zha, Z. Tian, C. Chen, J. Hao, J. Chen, J. Jiang, and Z. Xu, "itransformer: Inverted transformers are effective for time series forecasting," *arXiv preprint arXiv:2310.06625*, 2023.

[12] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Timesnet: Temporal 2d-variation modeling for general time series analysis," *arXiv preprint arXiv:2210.02186*, 2022.

[13] D. Wang, X. Zhang, R. Yang, Y. Zhang, Z. Hu, Y. Xu, Z. Xiong, and Y. Xu, "Timemixer: Decomposable multiscale mixing for time series forecasting," *arXiv preprint arXiv:2308.15230*, 2023.

[14] Q. Zeng, J. Chen, L. Zhou, D. Zhang, C. Wu, Q. Wen, Y. Xie, Y. Bian, Y. Rao, J. Zhou *et al.*, "Transformers in time series: A survey," *arXiv preprint arXiv:2202.07125*, 2023.

[15] Z. Li, Y. Shi, L. Chen, X. Luo, W. Ye, and S. Wang, "Revisiting linear methods for time series forecasting: Simple but effective," *arXiv preprint arXiv:2312.06020*, 2023.

[16] Y. Zhang, J. Yan, Q. Wen, J. Chen, Y. Jiang, Y. Gao, Z. Wu, and Y. Fu, "Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting," *arXiv preprint arXiv:2303.04713*, 2023.

[17] G. P. Peters and R. M. Andrew, "Climate change and carbon dioxide emissions: An analysis of the performance of forecasting models and methodologies," *Environmental Research Letters*, vol. 17, no. 3, p. 034028, 2022.

[18] M. Crippa, D. Guizzardi, M. Muntean, E. Schaaf, E. Solazzo, F. Monforti-Ferrario, J. G. Olivier, and E. Vignati, "Edgar v6.0 greenhouse gas emissions," *European Commission, Joint Research Centre (JRC)*, 2021.

[19] L. Nascimento, T. Kuramochi, G. Iacobuta, M. den Elzen, H. Fekete, M. Weishaupt, H. L. van Soest, M. Roelfsema, G. de Vivero-Serrano, S. Lui *et al.*, "Global climate policy database: Mapping policies for energy efficiency and low-carbon technology," *Energy Research & Social Science*, vol. 89, p. 102646, 2022.

[20] W. F. Lamb, M. Grubb, F. Diluiso, and J. C. Minx, "Climate policy event database: A new tool to track climate policy progress," *Environmental Research Letters*, vol. 16, no. 6, p. 064005, 2021.

[21] U. Gasser, M. Ienca, J. Scheibner, J. Sleigh, and E. Vayena, "Understanding the temporal evolution of covid-19 research through machine learning and natural language processing," *Scientometrics*, vol. 125, pp. 3609–3631, 2020.

[22] Z. Liu, D. Guan, W. Wei, S. J. Davis, P. Ciais, J. Bai, S. Peng, Q. Zhang, K. Hubacek, G. Marland *et al.*, "Drivers of carbon emission intensity change in china," *Environmental Research Letters*, vol. 17, no. 2, p. 024043, 2022.

[23] Z. Hausfather, H. F. Drake, T. Abbott, and G. A. Schmidt, "Evaluating the performance of past climate model projections," *Geophysical Research Letters*, vol. 47, no. 1, p. e2019GL085378, 2020.

[24] N. Grant, A. Hawkes, T. Mittal, and A. Gambhir, "Cost-effective mitigation of climate change: The roles of economic growth, mitigation technologies, and uncertainty," *Environmental Research Letters*, vol. 15, no. 2, p. 024008, 2020.

[25] F. W. Geels, B. K. Sovacool, T. Schwanen, and S. Sorrell, "Sociotechnical transitions for deep decarbonization," *Science*, vol. 357, no. 6357, pp. 1242–1244, 2017.

[26] C. Truong, L. Oudre, and N. Vayatis, "Selective review of offline change point detection methods," *Signal Processing*, vol. 167, p. 107299, 2020.

[27] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong *et al.*, "A survey of large language models," *arXiv preprint arXiv:2303.18223*, 2023.

[28] R. Schmalensee, T. M. Stoker, and R. A. Judson, "World carbon dioxide emissions: 1950–2050," *Review of Economics and Statistics*, vol. 80, no. 1, pp. 15–27, 1998.

[29] D. Holtz-Eakin and T. M. Selden, "Carbon dioxide emission-intensity in climate projections: Comparing the observational record to socioeconomic scenarios," *Energy Policy*, vol. 122, pp. 723–734, 2018.

[30] R. York, E. A. Rosa, and T. Dietz, "Stirpat, ipat and impact: analytic tools for unpacking the driving forces of environmental impacts," *Ecological Economics*, vol. 46, no. 3, pp. 351–365, 2003.

[31] E. Stehfest, D. Van Vuuren, T. Kram, L. Bouwman, R. Alkemade, M. Bakkenes, H. Biemans, L. Bouwman, M. Den Elzen, J. Janse *et al.*, "Integrated assessment of global environmental change with image 3.0: Model description and policy applications," *Netherlands Environmental Assessment Agency (PBL)*, 2014.

[32] K. Calvin, P. Patel, L. Clarke, G. Asrar, B. Bond-Lamberty, R. Y. Cui, A. Di Vittorio, K. Dorheim, J. Edmonds, C. Hartin *et al.*, "Gcam v5.1: representing the linkages between energy, water, land, climate, and economic systems," *Geoscientific Model Development*, vol. 12, no. 2, pp. 677–698, 2019.

[33] G. Luderer, M. Leimbach, N. Bauer, E. Kriegler, L. Baumstark, C. Bertram, A. Giannousakis, J. Hilaire, D. Klein, A. Levesque *et al.*, "Description of the remind model (version 1.6)," *SSRN Electronic Journal*, 2015.

[34] K. Riahi, D. P. Van Vuuren, E. Kriegler, J. Edmonds, B. C. O'neill, S. Fujimori, N. Bauer, K. Calvin, R. Dellink, O. Fricko *et al.*, "The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview," *Global Environmental Change*, vol. 42, pp. 153–168, 2017.

[35] T. Burandt, B. Xiong, K. Löffler, and P.-Y. Oei, "Big data driven carbon emission reduction in smart cities: A framework for analyzing and implementing regional mitigation strategies," *Journal of Cleaner Production*, vol. 197, pp. 1259–1274, 2018.

[36] F. Creutzig, L. Niamir, X. Bai, M. Callaghan, J. Cullen, J. Díaz-José, M. Figueroa, A. Grubler, W. F. Lamb, A. Leip *et al.*, "Demand-side solutions to climate change mitigation consistent with high levels of well-being," *Nature Climate Change*, vol. 12, no. 1, pp. 36–46, 2022.

[37] H. Yin, J. Xie, and J. Li, "China's carbon emission trading scheme: context, status quo, and future prospects," *Climate Policy*, vol. 15, no. 6, pp. 817–838, 2015.

[38] N. Karali, N. Khanna, N. Xu, S. Repetto, S. Vaid, and J. E. McMahon, "Carbon implications of marginal oils from market-driven crude oil refining baseline," *Applied Energy*, vol. 264, p. 114683, 2020.

[39] R. J. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice," *OTexts*, 2018.

[40] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[41] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[42] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, pp. 1189–1232, 2001.

[43] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[44] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[45] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[46] 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.

[47] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-beats: Neural basis expansion analysis for interpretable time series forecasting," in *International Conference on Learning Representations*, 2019.

[48] A. Joshi, S. Saboo, S. Nair, and L. Vachhani, "A review on multi-sensor data fusion for object detection in autonomous vehicle applications," *IEEE Sensors Journal*, vol. 16, no. 18, pp. 6824–6836, 2016.

[49] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.

[50] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2211–2268, 2011.

[51] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

[52] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[53] D. Wang, B. K. Szymanski, T. Abdelzaher, H. Ji, and L. Kaplan, "Social sensing: A systematic quantitative review," *ACM Transactions on Internet Technology (TOIT)*, vol. 19, no. 3, pp. 1–28, 2019.

[54] N. Xu, W. Mao, and G. Chen, "Multimodal deep learning for natural language processing," *ACM Computing Surveys (CSUR)*, vol. 54, no. 9, pp. 1–36, 2018.

[55] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, "Connecting the dots: Multivariate time series forecasting with graph neural networks," *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 753–763, 2020.