Towards CLEAN: Cycle Learning Energy-Emission Assessment Network

Yanming Guo[1], Qian Xiao[2], Kevin Credit[2], Jin Ma[1]
[1]University of Sydney, [2]Maynooth University
{yguo0337, j.ma}@sydney.edu.au, {qian.xiao, kevin.credit}@mu.ie

Abstract

Formulating effective climate policies is challenging due to the limitations of current decision-making models, such as Computable General Equilibrium models, which often rely on static, linear assumptions that oversimplify complex, real-world scenarios. To address these limitations, we introduce the Cycle Learning Energy-Emission Assessment Network, a scalable deep-learning framework trained on the proposed Carbon Monitor Energy-Emission time series dataset. CLEAN captures global patterns between energy structures and sectoral carbon emissions, enabling accurate energy-emission predictions at national, provincial, city, and sectoral levels. By framing the problem as a bidirectional supervised regression task, our model achieves a Mean Absolute Percentage Error of approximately 5%, significantly outperforming traditional models like XGBoost, especially with our novel data augmentation method, Densify. This demonstrates CLEAN's superior performance and generalization capabilities, making it a powerful tool for climate policy analysis. The CMEE dataset, CLEAN model, and Densify augmentation methods are open-sourced at `https://github.com/Yvnminc/CLEAN`.

1 Introduction

The United Nations' Sustainable Development Goal 7 calls for a global shift to affordable and clean energy, emphasizing its role in reducing dependence on fossil fuels and highlighting the broader environmental and economic benefits of sustainable energy [3, 23]. In alignment with this objective, the Paris Agreement was established with the ambitious target of keeping the global temperature rise well below 2 degrees Celsius above pre-industrial levels [17, 9]. This commitment has led countries worldwide to implement substantial energy and industrial transition policies [5, 12, 15].

Traditional policy-making in this context often relies on economic models like Computable General Equilibrium (CGE) models [24, 26]. While these models offer a structured approach to understanding economic outcomes, they are built on simplified assumptions and static frameworks [1, 19, 7]. As a result, they fall short in capturing the complex, dynamic interactions between energy systems and markets, and they lack the flexibility to adapt to varying future scenarios [14, 18]. This limitation poses significant challenges for climate policy development, which requires adaptive strategies to accommodate long-term goals, such as the commitment of many developed nations to achieve net-zero emissions by 2050. The uncertainties inherent in a 30-year planning horizon further complicate the creation of effective policies, as they must account for unpredictable technological, economic, and social changes [21, 6, 20].

2 Pathway Impact of Machine Learning for Climate Policy Making

This research focuses on a data-driven machine-learning approach. The proposed model, CLEAN for climate policy-making, demonstrates clear advantages and impactful pathways over CGE models

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2024.

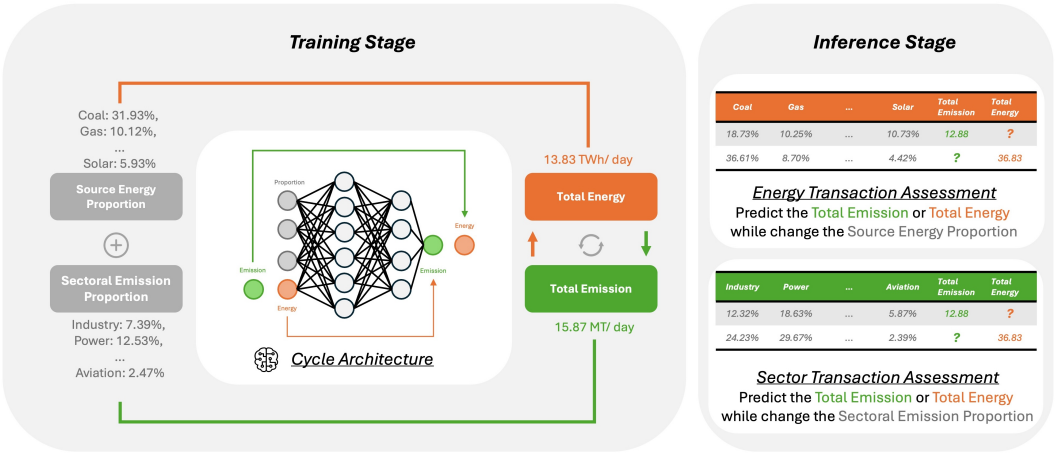


Figure 1: The Cycle Learning Energy-Emission Assessment Network (CLEAN) framework predicts national-level energy demand and emissions based on energy mix and sectoral emission structure. This model learns global patterns from the Carbon Monitor Energy-Emission dataset across scales, offering insights for policy formulation and feasibility assessments.

[22, 27, 25]. Specifically, unlike CGE models, which are limited by their static assumptions and localized scope. **CLEAN identifies scalable global patterns across national, provincial, city, and sectoral**. By understanding that a nation's energy profile is an aggregate of its sectors, CLEAN effectively captures how different energy mixes influence total carbon emissions and how sector-specific emissions contribute to the overall energy demand. **CLEAN demonstrates high generalization ability for various downstream prediction tasks**. It can predict daily emissions for unseen countries where direct emissions data is unavailable by using accessible power generation data. **CLEAN supports bidirectional prediction for policy simulation**. The model is designed to help determine the optimal energy mix required to meet specific emission reduction targets, automatically identifying a feasible energy structure. This capability is crucial for developing sustainable energy policies and ensuring that emission goals are both realistic and achievable.

3 Carbon Monitor Energy-Emission Time Series Data

Table 1: Detail Information of Carbon Monitor Energy-Emission (CMEE) Time Series Dataset.

Characteristics	Number	Description
Spatial Coverage	12	Regions (e.g. China, France, United States)
Temporal Coverage	2008	Daily granularity from 01/01/2019 to 30/06/2024
Energy Type	8	Energy source (e.g. Coal, Gas, Solar)
Emission Type	6	Secotral emission (e.g. Transportation, Industry, Power)

We have developed a comprehensive, high-resolution Energy-Emission time series dataset CMEE based on the Carbon Monitor project to understand the relationship between power generation and CO_2 emissions, which is crucial for evaluating national energy and sectorial transition policies. The Carbon Monitor is an international initiative providing regularly updated, science-based estimates of daily carbon emissions and power generation data for 12 countries, covering the period from January 1, 2019, to June 30, 2024 [13, 10]. Our dataset includes data on eight types of energy sources and six categories of sectoral emissions, spanning a total of 24,096 days (2,008 days each for 12 countries). It integrates two independently collected but related datasets, ensuring consistency through structured formatting and noise reduction techniques such as Gaussian filtering. This unified dataset serves as a robust foundation for machine learning tasks and supports further research, including data mining, causal analysis, and time series forecasting. (See detailed data analysis in Appendix).

4 Methodology

Although the CMEE dataset is fundamentally time-series data, the conversion relationship between energy and emissions is independent of time. The CMEE dataset is formatted in a tabular structure and designed to have three main numerical components: proportion (p), a vector representing the proportions of various energy uses and sectoral emissions; total energy (e_n), indicating the daily power generation of a country; and total emissions (e_m), representing the daily CO_2 emissions.

Therefore, we formulate this problem as a bidirectional supervised regression task. Specifically, to predict total emissions (e_m), we aim to learn a mapping $f_\theta : [p_1, p_2, \ldots, p_n, e_n] \mapsto e_m$, where the input vector is the concatenation of the proportion vector (p), total energy (e_n), and θ denotes the model parameters. Similarly, to predict total energy generation (e_n), the mapping function is defined as $f_\theta : [p_1, p_2, \ldots, p_n, e_m] \mapsto e_n$. This bidirectional setup allows the model to utilize both energy mix and emissions data for accurate predictions, ensuring that it captures the mutual influence between energy use and emissions without being dependent on temporal order. This method also enhances the model's generalizability to different contexts, making it robust across various scales and regions, as illustrated in Figure 1.

To achieve this, we propose the Cycle Learning Energy-Emission Assessment Network (CLEAN), a deep neural network-based framework designed to learn global patterns between energy power generation and sectoral emissions. The model uses two encoders to capture the features for Energy-to-Emission and Emission-to-Energy mappings separately. The Cycle loss allows the model to learn more robust and generalized patterns. It is introduced as a regularization term to enforce consistency between these tasks inspired by CycleGAN [28]. Thereby Additionally, we proposed a data augmentation technique called Densify to address the sparsity inherent in tabular data, significantly enhancing model performance over traditional models like XGBoost (See detailed model selection, design and augmentation methods in Appendix).

5 Experiments & Results

To evaluate the effectiveness of the CLEAN framework, we conducted a series of experiments using a country-based holdout approach rather than a traditional random split. The data was divided into 60% training countries, 20% validation countries, and 20% test countries. This approach provides a more rigorous test of the model's generalization abilities across different geographical contexts. The results, summarized in Table 2, demonstrate that deep learning models with MLP and CNN encoders generally outperform traditional machine learning models.

Table 2: Mean Absolute Percentage Error (MAPE) for Energy to Emission and Emission to Energy models with densification levels n. Here, n represents the number of synthesized countries; the more synthesized countries, the denser the training data space, and vice versa. The top results are in **bold**. It can be observed that deep learning models generally outperform traditional learning models, and those trained with the CLEAN framework and Cycle loss achieve the best performance.

	CMEE$_0$		**CMEE**$_{10}$		**CMEE**$_{100}$	
Model	Emission	Energy	Emission	Energy	Emission	Energy
LR [8]	0.423	0.347	0.243	0.128	0.104	0.101
KNN [4]	0.309	0.461	0.349	0.316	0.070	0.102
DT [16]	0.768	1.256	0.402	0.450	0.094	0.128
XGB [2]	0.359	1.166	0.374	0.370	0.070	0.112
MLP [11]	0.136	0.157	0.183	0.112	0.056	**0.049**
CNN [11]	0.206	0.273	0.259	0.201	0.062	0.050
CLEAN$_{MLP}$	**0.118**	**0.111**	**0.125**	**0.110**	0.051	0.057
CLEAN$_{CNN}$	0.133	0.196	0.178	0.191	**0.048**	0.054

We evaluated various models across different levels of data augmentation, and the results show that as the amount of synthetic data (n) increases, all models experience a significant decrease in MAPE, indicating that data augmentation substantially improves generalization performance. Interestingly, even a simple Linear Regression (LR) model outperformed more advanced models like XGBoost (XGB) in our regression tasks. This is because LR effectively captures continuous relationships, whereas XGB, which is optimized for discrete data, has difficulty generalizing to continuous patterns.

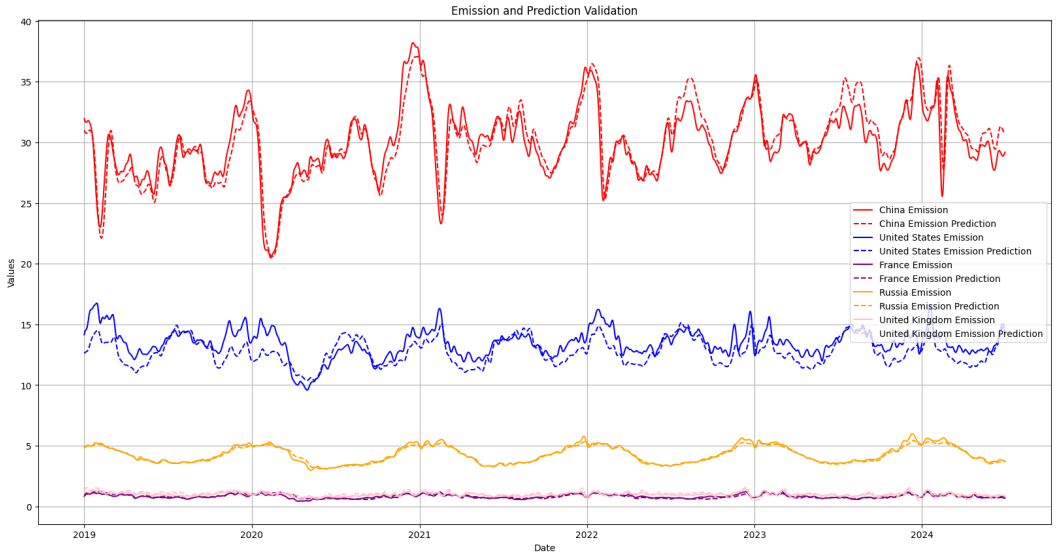


Figure 2: Daily emission predictions for Permanent Five using the CLEAN$_{CNN}$ model.

The CLEAN framework, incorporating Cycle loss, performed exceptionally well on sparser datasets (Densify$_0$ and Densify$_{10}$), suggesting that Cycle loss adds valuable inductive bias and regularization. However, on denser datasets such as Densify$_{100}$, the difference in performance decreases, and MLP achieves the best results, indicating that Cycle loss becomes less important with more densely populated data. Overall, the CLEAN framework consistently maintains a very low MAPE, even when predicting outcomes for unseen countries, demonstrating its robust generalization capabilities.

6 Conclusion

Our research makes three key contributions. **First, we introduce a deep neural network framework CLEAN to support climate policy-making**. The Cycle Learning Energy-Emission Assessment Network provides a data-driven, transparent policy-making tool by capturing complex, non-linear relationships between energy use and emissions. **Second, we developed the Energy-Emission time series dataset based on the Carbon Monitor project**. Providing a solid foundation for future research in causal inference, data mining, and time series analysis. **Third, the CLEAN model employs the innovative Cycle loss and a new data augmentation method, Densify, to handle sparse training data effectively**. This model outperforms traditional machine learning models like XGBoost, achieving a Mean Absolute Percentage Error of just 5% on unseen countries, demonstrating strong generalization abilities. All data, models, and code from our study are open-source, providing valuable resources for the community. Our research impacts climate policy-making in three ways: **Learning Global Patterns Across Scales**, applicable to national, provincial, city, and sectoral levels; **Supporting Multiple Downstream Tasks**, helpful in predicting emissions in regions without direct carbon monitoring; and **Bidirectional Prediction for Policy Simulation**, enabling the setting of emission reduction targets and identification of optimal energy strategies. Our framework advances climate policy modelling and offers practical tools for more effective and informed policymaking.

References

[1] Christoph Böhringer and Andreas Löschel. Computable general equilibrium models for sustainability impact assessment: Status quo and prospects. *Ecological economics*, 60(1):49–64, 2006.

[2] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[3] William Colglazier. Sustainable development agenda: 2030. *Science*, 349(6252):1048–1050, 2015.

[4] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[5] Rajvikram Madurai Elavarasan, Rishi Pugazhendhi, Taskin Jamal, Joanna Dyduch, Mohammad Taufiqul Arif, Nallapaneni Manoj Kumar, GM Shafiullah, Shauhrat S Chopra, and Mithulananthan Nadarajah. Envisioning the un sustainable development goals (sdgs) through the lens of energy sustainability (sdg 7) in the post-covid-19 world. *Applied Energy*, 292:116665, 2021.

[6] Francesco Fuso Nerini, Benjamin Sovacool, Nick Hughes, Laura Cozzi, Ellie Cosgrave, Mark Howells, Massimo Tavoni, Julia Tomei, Hisham Zerriffi, and Ben Milligan. Connecting climate action with other sustainable development goals. *Nature Sustainability*, 2(8):674–680, 2019.

[7] Maurizio Gargiulo and Brian O Gallachoir. Long-term energy models: Principles, characteristics, focus, and limitations. *Wiley Interdisciplinary Reviews: Energy and Environment*, 2(2):158–177, 2013.

[8] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[9] Niklas Höhne, Takeshi Kuramochi, Carsten Warnecke, Frauke Röser, Hanna Fekete, Markus Hagemann, Thomas Day, Ritika Tewari, Marie Kurdziel, Sebastian Sterl, et al. The paris agreement: resolving the inconsistency between global goals and national contributions. *Climate Policy*, 17(1):16–32, 2017.

[10] Da Huo, Xiaoting Huang, Xinyu Dou, Philippe Ciais, Yun Li, Zhu Deng, Yilong Wang, Duo Cui, Fouzi Benkhelifa, Taochun Sun, et al. Carbon monitor cities near-real-time daily estimates of co2 emissions from 1500 cities worldwide. *Scientific Data*, 9(1):533, 2022.

[11] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[12] Bandy X Lee, Finn Kjaerulf, Shannon Turner, Larry Cohen, Peter D Donnelly, Robert Muggah, Rachel Davis, Anna Realini, Berit Kieselbach, Lori Snyder MacGregor, et al. Transforming our world: implementing the 2030 agenda through sustainable development goal indicators. *Journal of public health policy*, 37:13–31, 2016.

[13] Zhu Liu, Philippe Ciais, Zhu Deng, Steven J Davis, Bo Zheng, Yilong Wang, Duo Cui, Biqing Zhu, Xinyu Dou, Piyu Ke, et al. Carbon monitor, a near-real-time daily dataset of global co2 emission from fossil fuel and cement production. *Scientific data*, 7(1):392, 2020.

[14] Irene Monasterolo, Andrea Roventini, and Tim J Foxon. Uncertainty of climate policies and implications for economics and finance: An evolutionary economics approach. *Ecological Economics*, 163:177–182, 2019.

[15] Karen O'Brien. Is the 1.5 c target possible? exploring the three spheres of transformation. *Current opinion in environmental sustainability*, 31:153–160, 2018.

[16] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1:81–106, 1986.

[17] Joeri Rogelj, Michel Den Elzen, Niklas Höhne, Taryn Fransen, Hanna Fekete, Harald Winkler, Roberto Schaeffer, Fu Sha, Keywan Riahi, and Malte Meinshausen. Paris agreement climate proposals need a boost to keep warming well below 2 c. *Nature*, 534(7609):631–639, 2016.

[18] Georgios Savvidis, Kais Siala, Christoph Weissbart, Lukas Schmidt, Frieder Borggrefe, Subhash Kumar, Karen Pittel, Reinhard Madlener, and Kai Hufendiek. The gap between energy policy challenges and model capabilities. *Energy Policy*, 125:503–520, 2019.

[19] S Serban Scrieciu. The inherent dangers of using computable general equilibrium models as a single integrated modelling framework for sustainability impact assessment. a critical note on böhringer and löschel (2006). *Ecological Economics*, 60(4):678–684, 2007.

[20] Leonard A Smith and Nicholas Stern. Uncertainty in science and its role in climate policy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1956):4818–4841, 2011.

[21] Ioannis Tsiropoulos, Wouter Nijs, Dalius Tarvydas, P Ruiz, et al. Towards net-zero emissions in the eu energy system by 2050. *Insights from Scenarios in Line with the*, 2030, 2020.

[22] Thomas Weber, Nicola A Wiseman, and Annette Kock. Global ocean methane emissions dominated by shallow coastal waters. *Nature communications*, 10(1):4584, 2019.

[23] Sabine Weiland, Thomas Hickmann, Markus Lederer, Jens Marquardt, and Sandra Schwindenhammer. The 2030 agenda for sustainable development: transformative change through the sustainable development goals? *Politics and Governance*, 9(1):90–95, 2021.

[24] Ian Sue Wing. Computable general equilibrium models and their use in economy-wide policy analysis. *Technical Note, Joint Program on the Science and Policy of Global Change, MIT*, 2004.

[25] Federica Zennaro, Elisa Furlan, Christian Simeoni, Silvia Torresan, Sinem Aslan, Andrea Critto, and Antonio Marcomini. Exploring machine learning potential for climate change risk assessment. *Earth-Science Reviews*, 220:103752, 2021.

[26] Fan Zhai. Armington meets melitz: Introducing firm heterogeneity in a global cge model of trade. *Journal of Economic Integration*, pages 575–604, 2008.

[27] Ruda Zhang, Patrick Wingo, Rodrigo Duran, Kelly Rose, Jennifer Bauer, and Roger Ghanem. Environmental economics and uncertainty: Review and a machine learning outlook. *arXiv preprint arXiv:2004.11780*, 2020.

[28] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

A Detailed CMEE Data Information

The Carbon Monitor Energy-Emission (CMEE) Time Series Dataset integrates daily CO_2 emission data and daily power generation data from the Carbon Monitor project. Using country and date as common attributes, we merged these datasets to create a structure suitable for machine learning tasks.

A.1 Dataset Composition

The dataset comprises three primary numerical components:

- **Proportion** (p): This vector represents the proportions of 8 different energy sources and 6 sectoral emissions, giving insight into the energy mix and the sources of emissions.
- **Total Energy** (E_n): This indicates the total daily power generation for each country in units of GWh/ day, providing a measure of energy output.
- **Total Emission** (E_m): This represents the total daily CO_2 emissions in units of Mt/ day, capturing the carbon footprint associated with the energy produced and consumed.

Additionally, the dataset includes categorical columns for the country and date, totalling 18 columns. Each row captures the daily emission and energy data for a specific country on a particular day. The CMEE dataset spans 12 countries, each with data from January 1, 2019, to June 30, 2024, resulting in 2,008 days and a total of 24,096 data entries.

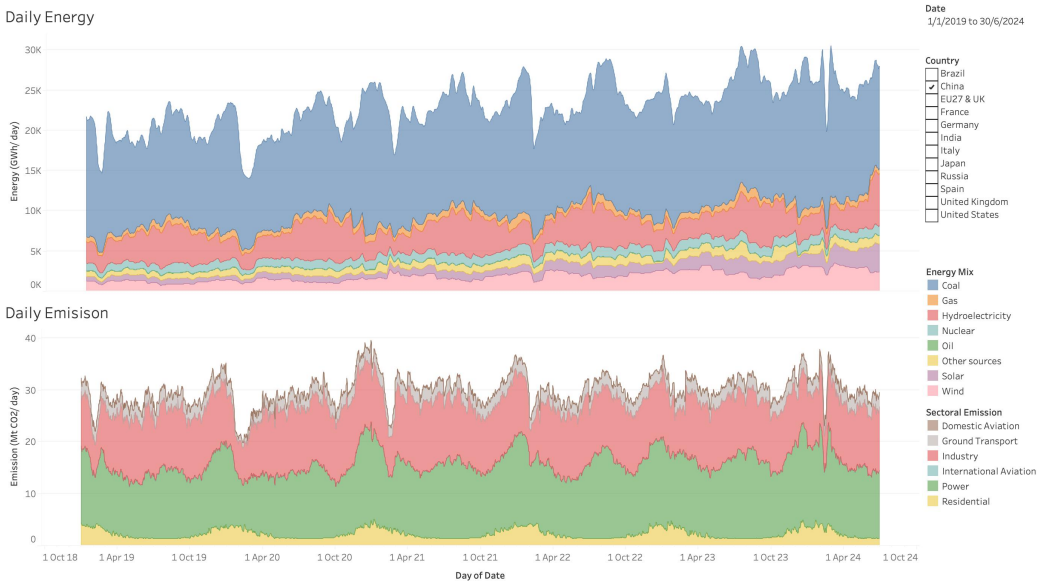
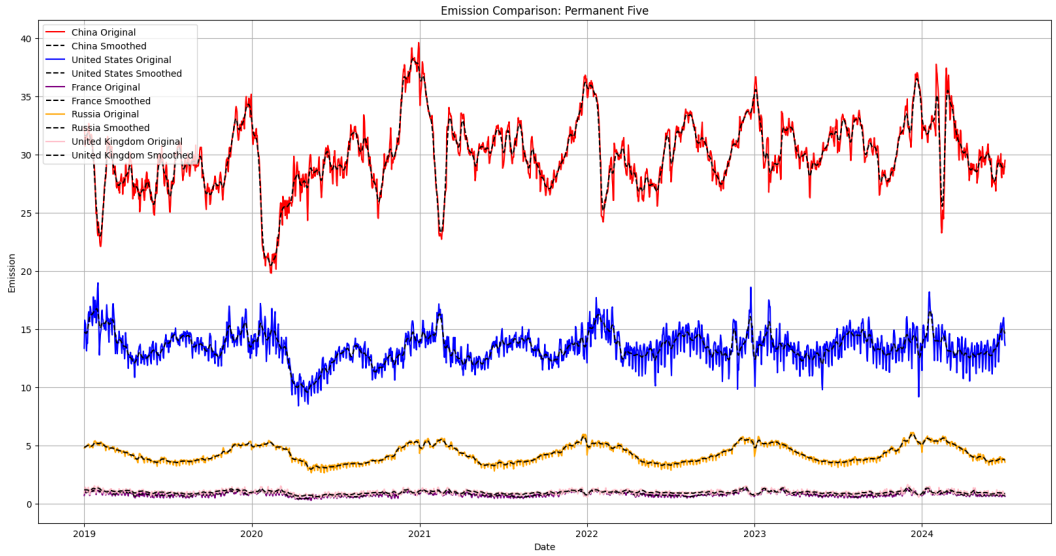


Figure 3: Interactive visualization of aggregated daily emission and power generation data for various countries in the CMEE dataset. The current display shows China's energy mix and sectoral emission structure, highlighting a high reliance on coal and significant industrial emissions. This visualization allows exploration of different countries' data, available at `https://public.tableau.com/app/profile/unknow404/viz/CMEE_Five_Normal_Dashboard/Dashboard1`.

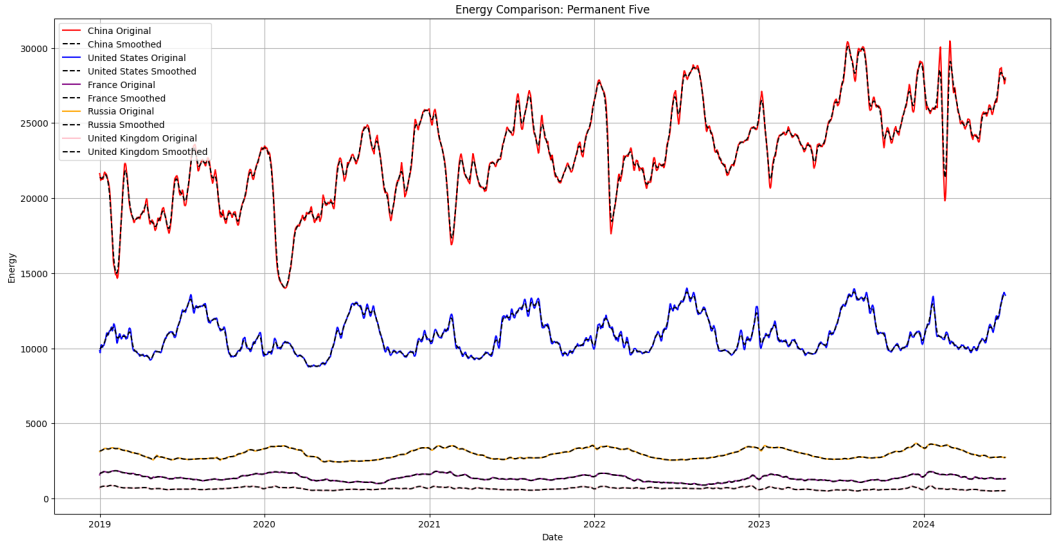
A.2 Data Visualization and Analysis

Figure 3 provides an interactive visualization of aggregated daily emissions and power generation data for various countries. The current display focuses on China's energy mix and sectoral emissions, revealing a high reliance on coal and a significant industrial sector. This pattern underscores the challenges China faces in transitioning to a low-carbon economy.

However, the raw data contains considerable noise, particularly in the emission data, which can hinder the model's ability to learn generalizable patterns. To address this, we applied Gaussian filtering to smooth the data. The following figures illustrate the time series before and after smoothing:



(a) Comparison of emission time series before and after smoothing. The solid lines represent the original data, while the dashed lines show the smoothed data.



(b) Comparison of energy time series before and after smoothing. The solid lines represent the original data, while the dashed lines show the smoothed data.

Figure 4: Smoothing the data using Gaussian filtering reduces noise, enhancing the machine learning model's ability to capture underlying trends and patterns without being distracted by random fluctuations.

A.3 Correlation Analysis

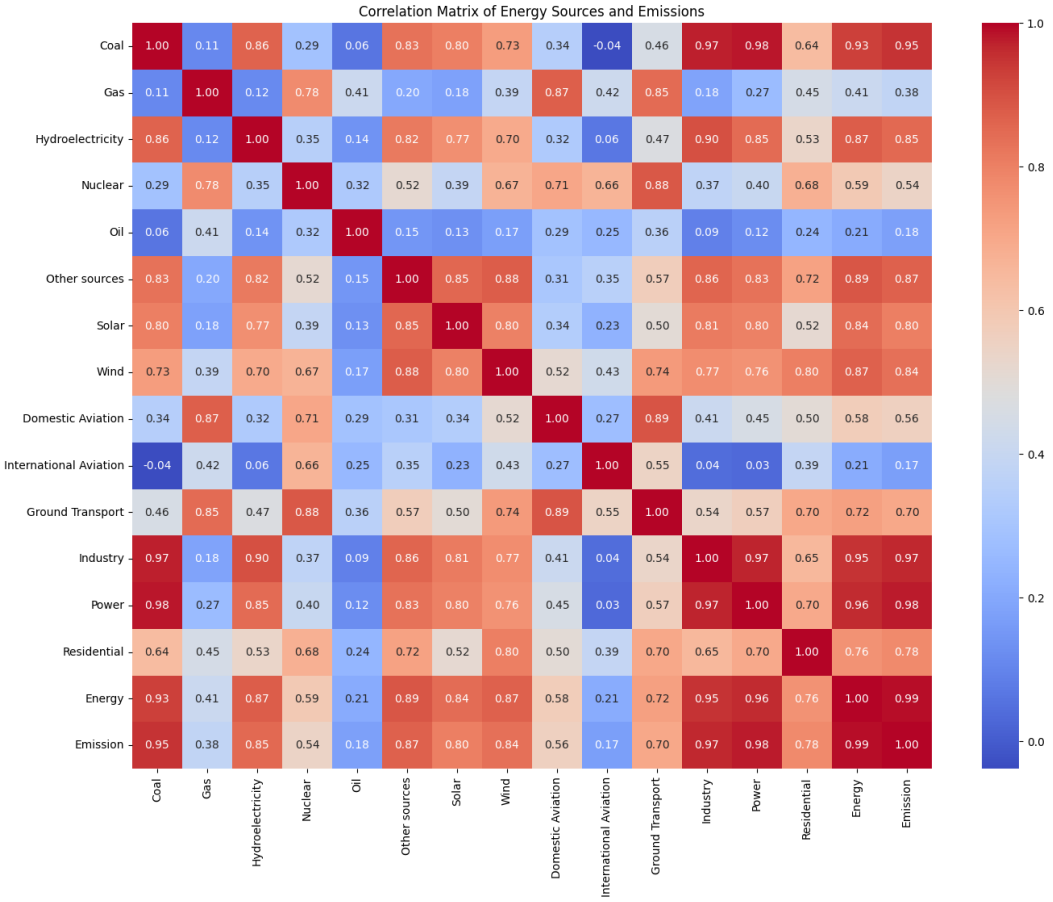


Figure 5: Correlation matrix of various energy sources and sectoral emissions, illustrating the relationships between different types of energy production and their associated emissions. High correlations between certain energy types, such as coal and total emissions, indicate significant impacts on CO_2 levels.

The correlation matrix in Figure 5 shows the relationships between various energy sources and emissions. High correlations between certain energy types, such as coal and total emissions, suggest that changes in coal usage significantly impact CO_2 emissions. Understanding these correlations helps in designing policies that target specific energy transitions to achieve emission reduction goals effectively.

By cleaning and structuring this data, we create a robust foundation for developing and applying machine learning models to predict energy demand and emissions, enabling more informed climate policy-making decisions.

B Model Selection & CLEAN Design

Traditional machine learning models like Linear Regression, K-nearest Neighbors (KNN), Decision Trees, and XGBoost are effective for discrete data but struggle with capturing continuous patterns necessary for regression tasks. Linear models can handle continuous data but are limited by their linear assumptions, making them inadequate for capturing non-linear relationships. Our deep learning-based model, the Cycle Learning Energy-Emission Assessment Network (CLEAN), addresses these limitations by offering three main advantages. **Capturing Non-linear Relationships**: Neural networks like Multi-Layer Perceptrons (MLPs) are capable of modellinging complex, non-linear relationships.**Learning Continuous Features**: Unlike tree-based models, neural networks can effectively learn from continuous data, making them more suitable for regression. **Flexibility in Model Design**: The CLEAN model's architecture allows for end-to-end learning with two encoders (MLP or CNN), each designed for specific tasks—Energy-to-Emission and Emission-to-Energy mapping. Inspired by CycleGAN, we use a Cycle loss function to reinforce learning between these tasks, enhancing model robustness, illustrated in Figure 6.

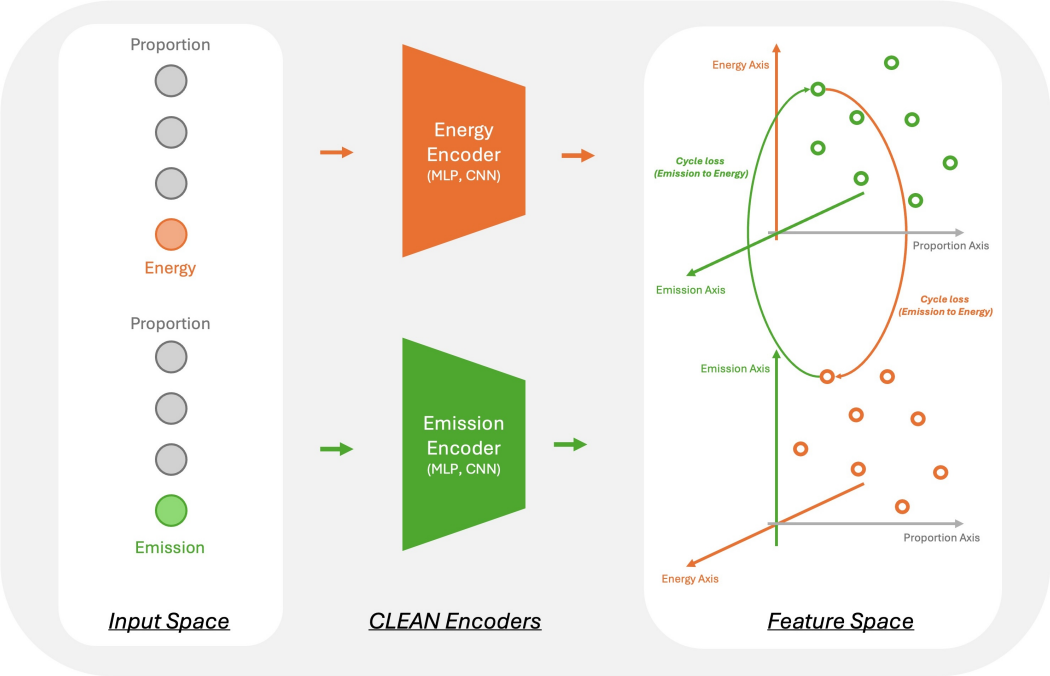


Figure 6: The CLEAN model architecture with two encoders for Energy-to-Emission and Emission-to-Energy mappings. The Cycle loss ensures consistent feature spaces, enhancing performance and generalization.

The Cycle loss in the CLEAN model acts as a regularization term, promoting consistency between the Energy-to-Emission and Emission-to-Energy feature spaces, thus improving generalization. The mathematical formulation for the Cycle loss is:

$$\mathcal{L}_{cycle} = \mathcal{L}_{E_n \to E_m} + \mathcal{L}_{E_m \to E_n} + \lambda \left(\mathcal{L}_{E_m \to E_n \to E_m} + \mathcal{L}_{E_n \to E_m \to E_n} \right)$$

Where $\mathcal{L}_{E_n \to E_m}$ and $\mathcal{L}_{E_m \to E_n}$ is the standard prediction losses using mean square error from predicting emissions based on energy inputs, and from predicting energy based on emissions inputs. Regularised term $\mathcal{L}_{E_m \to E_n \to E_m}$ and $\mathcal{L}_{E_n \to E_m \to E_n}$ ensures that when predicting emissions, the subsequent prediction of energy based on those emissions returns to the original energy input, and mapping back. λ is a hyperparameter that controls the weight of the cycle consistency loss, balancing it with the standard prediction losses.

C Densify Augmentation

We propose a novel data generation method called Densify to address the inherent sparsity issues in tabular data. In our context, the CMEE dataset provides energy and industrial structure data for 12 different countries. However, each country is an independent and heterogeneous entity with significant differences from one another, leading to what we define as data sparsity. Learning continuous and dense features from such highly sparse data is challenging. Deep Neural Networks (DNNs) rely heavily on a continuous and dense sample space, which is why they perform well in modalities like NLP and computer vision (CV), where training datasets (such as text and images) are typically large and dense. In contrast, DNNs often underperform on tabular data because training on sparse datasets can easily result in overfitting.

To mitigate this issue, we introduce the Densify method. This approach involves randomly sampling m countries from the existing dataset and assigning each country a sample weight $\lambda_i \sim \mathcal{U}(0, 1)$, with the constraint that the sum of all weights equals one, i.e., $\sum_{i=1}^{m} \lambda_i = 1$. The augmented or "synced" country data, $\mathbf{c}_{\text{synced}}$, is then generated as:

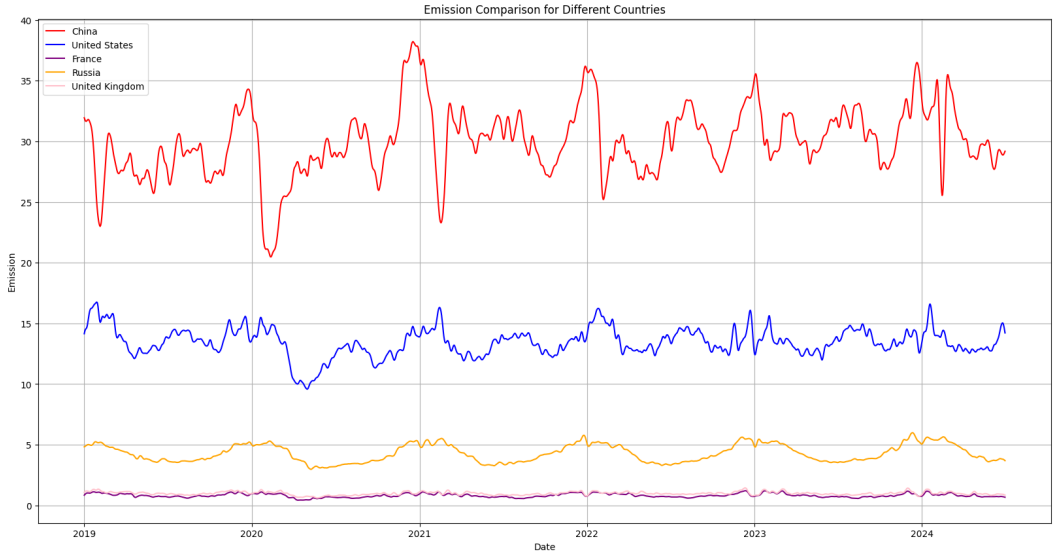
$$\mathbf{c}_{\text{synced}} = \sum_{i=1}^{m} \lambda_i \cdot \mathbf{c}_i, \quad \text{where } \mathbf{c}_i \in C. \tag{1}$$

Here, $\mathbf{c}_i$ represents a sample from the existing country dataset C. This data augmentation technique helps create a more continuous and dense feature space, allowing deep neural networks to learn more effectively from tabular data. Below is the pseudo-code of the Densify method:

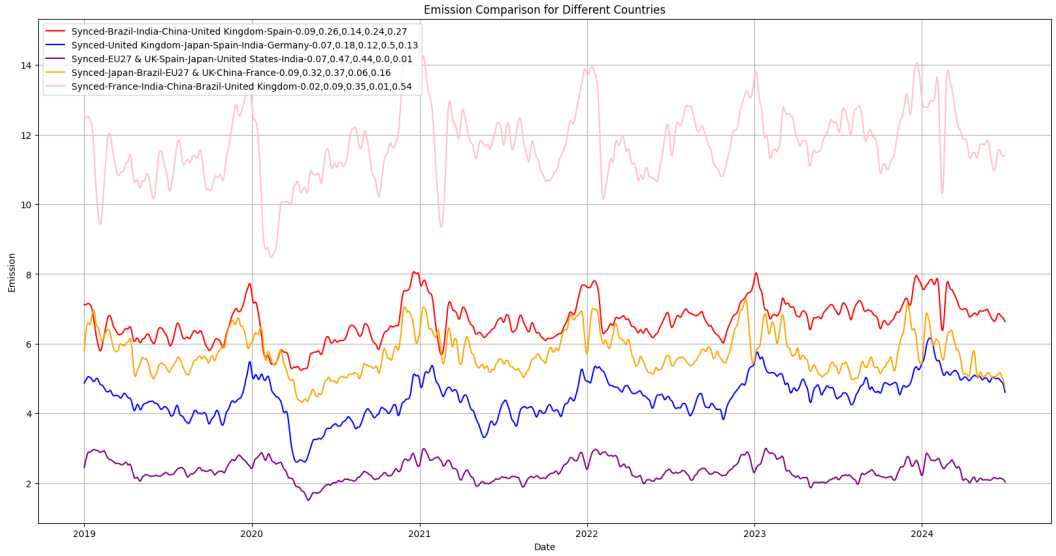
Algorithm 1 Densify Pseudo-code

Input: n ▷ Number of synced countries to generate
Input: m ▷ Number of countries to sample per synced country
Input: C ▷ Set of existing countries' data
for $i = 1$ to n **do**
  $C_s = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m\}$ ▷ Sample m countries from C
  $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_m\} \sim \mathcal{U}(0, 1)$ ▷ Generate weights such that $\sum_{j=1}^{m} \lambda_j = 1$
  $\mathbf{c}_{\text{synced}} = \sum_{j=1}^{m} \lambda_j \cdot \mathbf{c}_j$ ▷ Compute synced country data
  Store $\mathbf{c}_{\text{synced}}$
end for
Output: Synced country data $\mathbf{c}_{\text{synced}}$ for N countries

We plot the emission time series of five normal countries from CMEE data and compare them with synced data shown in Figure 7:



(a) Emission time series of five normal countries from CMEE data.



(b) Emission time series of the synced countries after data augmentation using the Densify method. Here, m is set to 5, indicating that each synced country is generated by randomly sampling and weighting five original countries. This augmentation introduces more diverse patterns into the original CMEE data, enabling the model to learn a denser and more continuous feature space.

Figure 7: Comparison of carbon emission time series for five normal countries and their synthesized counterparts.