STA457 Time Series Analysis
Summer 2025
Final Project Guidelines

1 Introduction

Modelling environmental indicators is increasingly vital in the face of global climate change. Among these indicators, carbon dioxide (CO_2) emissions play a central role, as they are a major driver of global warming and a key metric for evaluating the effectiveness of climate policies and international agreements. However, forecasting CO_2 emissions is a complex task, influenced by a wide range of factors such as economic activity, energy consumption, technological change, policy interventions, and natural events. These challenges make accurate modeling both difficult and critically important. Understanding and forecasting CO_2 trends is essential for researchers, policymakers, and environmental agencies aiming to set emission targets, design mitigation strategies, and assess progress toward climate goals.

In this project, you will apply time series analysis techniques to model and predict global CO_2 emissions. You will explore statistical methods to analyze historical emission patterns, assess model performance, and interpret results. Through this project, you will gain hands-on experience working with complex environmental time series data and develop analytical skills that are relevant to sustainability research and evidence-based policy planning.

You will have **considerable flexibility** in this project. There are **no restrictions** on the methods you choose—you are encouraged to explore and apply any techniques you find appropriate, even those beyond the scope of our course materials (This is an excellent opportunity for you to engage in self-study and further independent research while deepening your understanding). However, you must provide a clear justification for your chosen approach in your report. Additionally, aside from the data I provide, you may incorporate **any supplementary data** you deem relevant. This is a real-world, open-ended problem with no single correct answer. Your goal is to develop a well-reasoned and data-driven model for forecasting CO_2 emissions and providing meaningful insights.

2 Background Information

Carbon dioxide (CO_2) is a greenhouse gas that plays a central role in regulating the Earth's climate. It is primarily released through the combustion of fossil fuels such as coal, oil, and natural gas, as well as through land-use changes such as deforestation. As the main driver of anthropogenic climate change, monitoring and forecasting CO_2 emissions are essential for assessing environmental impact and informing policy decisions.

2.1 Sources of CO_2 Emissions

Global CO_2 emissions originate from several major sources:

- **Energy Production** – The burning of fossil fuels for electricity and heat is the largest source of CO_2 emissions globally.
- **Transportation** – Cars, trucks, ships, and airplanes emit significant amounts of CO_2, especially in industrialized and rapidly developing nations.
- **Industry** – Processes such as cement production, steelmaking, and chemical manufacturing contribute notably to industrial CO_2 emissions.
- **Residential and Commercial Sectors** – Heating, cooling, and electricity use in buildings also contribute to emissions, particularly in urban areas.
- **Land Use and Forestry** – Deforestation, forest degradation, and land conversion reduce carbon sinks and release stored carbon into the atmosphere.

2.2 Major CO_2-Emitting Regions

CO_2 emissions vary significantly across regions, depending on levels of industrialization, energy consumption, and population size. The **top CO_2-emitting countries and regions** include:

- **China** – Currently the world's largest emitter, driven by coal-based energy production and rapid industrialization.
- **The United States (US)** – Historically the largest emitter, with emissions mainly from transportation, industry, and energy generation.
- **The European Union (EU)** – Although emissions have declined due to policy efforts and renewable energy adoption, the EU remains a major emitter.
- **India** – A rapidly growing economy with increasing energy demand, largely met through coal.
- **Rest of the World** – Emissions from other countries, particularly in Southeast Asia, the Middle East, and parts of Latin America, are rising as these regions industrialize.

2.3 Factors Affecting CO_2 Emissions

CO_2 emissions are influenced by a complex interplay of factors, including:

1. **Economic Activity** – Economic growth is typically associated with increased energy use and emissions, especially in developing economies.
2. **Energy Sources** – The mix of energy sources (e.g., coal vs. renewables) plays a critical role. Transitions to low-carbon energy can significantly reduce emissions.
3. **Government Policies** – Regulations such as carbon pricing, emissions caps, fuel standards, and incentives for renewable energy directly affect emission levels.
4. **Technological Innovation** – Advances in energy efficiency, carbon capture and storage, and clean energy can mitigate emissions.
5. **Climate Events and Natural Variability** – Events like El Niño or large volcanic eruptions can influence short-term emission levels and atmospheric carbon balance.
6. **Population and Urbanization Trends** – Growing urban populations increase energy demand and emissions unless counteracted by sustainable infrastructure and practices.

Understanding these drivers is essential for effectively modeling CO_2 emissions. These factors introduce various trends, cycles, structural changes, and potential external shocks that must be accounted for in time series forecasting models.

3 Data

For this project, the data are sourced from the Our World in Data™: CO_2 and Greenhouse Gas Emissions project, curated by Hannah Ritchie, Pablo Rosado, and Max Roser (2023) [1]. The project is published online at OurWorldinData.org, and the data used in this analysis were retrieved from the following resource: Ritchie, H., Rosado, P., & Roser, M. (2023). "CO_2 and Greenhouse Gas Emissions." Our World in Data™. Available at: `https://ourworldindata.org/co2-and-greenhouse-gas-emissions` [1]

The underlying data are based on the Global Carbon Budget (2024) and population estimates compiled from various sources (2024), with substantial processing and integration conducted by the Our World in Data™ team. A comprehensive description of the project methodology is available at: `https://ourworldindata.org/co2-and-greenhouse-gas-emissions`.

The dataset is provided in two files: `owid-co2-data.csv`, which contains the main data, and `owid-co2-codebook.csv`, which includes the corresponding codebook detailing variable definitions and data sources. Full documentation is available at `https://github.com/owid/co2-data`.

Please ensure appropriate attribution when using or redistributing the data, as per the terms specified by Our World in Data™ and their source providers.

While the aforementioned datasets form the core of the analysis, you are encouraged to **incorporate any additional data sources** that you find relevant. And the additional data should be properly documented and justified in your report. For example:

- **Global and per-country fossil fuel consumption:** Data on coal, oil, and natural gas consumption by country over time can help capture trends in CO_2 emissions from energy use. Such data are available from the BP Statistical Review of World Energy, the International Energy Agency (IEA), and Our World in Data™.
- **Global and per-country clean energy usage:** This includes the proportion and total energy consumption from renewable sources such as solar, wind, and hydroelectric. It can help assess decarbonization progress. Data are available from the International Renewable Energy Agency (IRENA), the World Bank, and Our World in Data.
- **Global and per-country population and urbanization rate:** Demographic factors are critical drivers of CO_2 emissions. Larger or more urbanized populations typically correlate with higher energy consumption. Population and urbanization data can be found from the United Nations, the World Bank, and Our World in Data.

4 Output Requirements

Your final output for this project should be a well-structured report that clearly presents your approach to modelling CO_2 emissions. The report should be written in a formal style and include appropriate explanations, justifications, and interpretations of results. Below is a suggested structure for your report:

1. **Introduction** The introduction should provide an overview of the project, including a brief description of the study, the motivation of the study and its real-world significance. You may also briefly outline the objectives of your analysis and the key challenges involved.
2. **Literature Review** This section should summarize relevant research on the topic, including time series modelling and forecasting. Highlight how your approach builds upon or differs from existing methods.
3. **Methodology** Clearly describe the forecasting methods you choose to apply. As mentioned earlier, this may include classical time series models covered in this course, or more sophisticated, e.g., machine learning approaches. Justify your choice of models based on theoretical considerations, past research and/or the pattern of the data. Explain in detail any steps, including preprocessing steps such as handling missing data, stationarity checks, transformation or feature engineering. A minimum of two methods is required for the project.
4. **Data** Provide a detailed description of the datasets used in the analysis. For any external data, please indicate the source of the data. You may also include Summary statistics and visualizations that help illustrate the characteristics of the data.
5. **Forecasting and Results** Present the results of your forecasting models, including:

- Model training and validation process.
- Performance evaluation using appropriate metrics.
- Forecasted values and any observed patterns.
- Graphical representations of predictions versus actual CO2 emissions.

6. **Discussion and Conclusion** Interpret your results and discuss their implications. You may also consider including the limitations of your approach, the challenges you face and how can your analysis be improved or extended.

7. **Appendix**

   The appendix should include any supplementary materials that support your analysis but are not suitable for the main body of the report. The full code used for the project **must** be provided in the Appendix. The appendix may also include: Additional elaborations, derivations, or mathematical justifications and extended statistical analyses, tables, or figures that complement the main discussion.

8. **References** Include a properly formatted list of references for any external sources cited in your report. Follow a standard citation style (e.g., APA, IEEE, or any format specified in the submission guidelines). All literature, datasets, and tools used in your analysis should be appropriately credited.

Your report should be well-organized and written in a clear and concise manner. The use of formulas, tables, figures, and references is encouraged to support your analysis and they should be well labelled. Ensure that your submission adheres to proper citation standards. The main body of the report should be between 15 and 20 pages in length, excluding the appendix.

5 Marking Rubric

As mentioned in the course syllabus, the assessment of the group project is divided into two components: 20 marks for the overall quality of the project and 6 marks for individual contributions. For the first 20 marks, all group members will receive the same score, and a detailed rubric will be provided as follows. The remaining 6 marks will be awarded based on peer evaluations, reflecting each individual's contribution to the project.

The first 20 marks for the overall quality of the project will be assessed based on three key criteria: Writing and Organization, Method Correctness, and Creativity. Below is a breakdown of how each section will be evaluated.

1. **Writing and Organization (5 marks, 25%)** - Clear and concise writing that effectively communicates ideas. - Well-organized structure with appropriate use of headings, subheadings, and sections. - Logical flow of content with smooth transitions between sections. - Proper spelling, grammar, and formatting throughout the report. - Appropriate use of tables, figures, and visualizations to support the analysis.

2. **Method Correctness (5 marks, 25%)** - Appropriateness of the chosen methods and models for the problem with accurate application. - Justification of model choices, with clear explanations of why certain methods were used. - Proper handling of data, including necessary preprocessing steps. - Clear presentation of the models.

3. **Prediction Performance (5 marks, 25%)** - Accuracy of the forecasts based on relevant performance metrics (e.g., RMSE, MAE, MAPE). - Interpretation of forecast results and discussion of their reliability.

4. **Creativity (5 marks, 25%)** - Originality and creativity in the approach to the problem. - Exploration of techniques beyond the course material, including any self-driven research. - Innovative solutions or novel insights derived from the data and analysis. - Effective use of external data or additional sources to enhance the model or analysis. - Thoughtfulness in addressing the limitations of the chosen methods and suggesting improvements or future work.

6 Policy on Generative AI Use

All written components of this project—including the report, analysis, and code explanations — must be completed entirely by the student and reflect their own original work. The use of generative AI tools is prohibited. Submissions that show signs of AI-generated content may be subject to investigation, and we reserve the right to deduct marks or impose academic penalties accordingly. This guideline is also in adherence with the University's academic integrity policy and is in place to ensure fairness and to promote the development of your own critical thinking, analytical, and communication skills.

7 Individual Contribution

The individual contribution mark is calculated as follows:

First, each group member will assign a contribution score between 0 and 6 to every other member in their group, reflecting their perception of each member's effort and involvement in the project.

Then, each student's final individual contribution mark will be computed using the following formula:

$$\begin{aligned}&\text{Individual Contribution Mark}\\ =&6 \times \text{number of students in the group} \times \left(\frac{\text{Total Marks Received by the individual}}{\text{Total Marks Assigned by Group Members}}\right)\end{aligned}$$

For example, consider a group of four students. The following table represents their peer evaluation results, where each row corresponds to the marks assigned by an individual, and each column represents the marks received by each student.

		Receiver				Total
		A	B	C	D	
Grader	A		n_{AB}	n_{AC}	n_{AD}	
	B	n_{BA}		n_{BC}	n_{BD}	
	C	n_{CA}	n_{CB}		n_{CD}	
	D	n_{DA}	n_{DB}	n_{DC}		
Total		n_A	n_B	n_C	n_D	n_{total}

Table 1: A table elaborating the calculation of individual contribution mark

Then, the marks received by Student A is $24 \times \frac{n_A}{n_{total}}$. Under this marking scheme, a student may receive an individual contribution mark greater than 6. Any excess marks will be treated as bonus points and included in the final evaluation for this course.

This adjustment ensures that students who contribute more significantly to the project receive appropriate recognition, while those who contribute less will have a lower individual contribution mark.

8 Deadline

The deadline of the final project is **Sunday, June 22nd, 2025 23:59 PM**.

References

[1] H. Ritchie, P. Rosado, and M. Roser. CO2 and greenhouse gas emissions. *Our World in Data*, 2023. https://ourworldindata.org/co2-and-greenhouse-gas-emissions.