# How can data science models evaluate the effectiveness of government subsidies or carbon tax policies on energy consumption patterns?

**Neil Mehra**


## Abstract

Climate change is one of the biggest challenges of our time, and reducing greenhouse gas emissions is critical. Governments across the world use different policies, such as subsidies for renewable energy and carbon taxes on fossil fuels, to influence energy use. However, it is difficult to measure how effective these policies are using traditional methods, because many other factors—like economic growth, new technologies, or consumer behavior—also affect energy demand and emissions.

This research explores how data science models can improve the evaluation of policy effectiveness. Tools such as time-series models (ARIMA, LSTM), machine learning approaches (random forests, neural networks), and causal inference methods (synthetic control, difference-in-differences) are used to analyze large datasets, including government energy statistics, emission records, and smart meter data. These models can separate real policy impacts from background noise, track long-term changes, and forecast future outcomes under different scenarios.

The study shows that data science offers policymakers stronger evidence, clearer insights, and more reliable predictions to design effective strategies for reducing emissions and promoting sustainable energy use.


**Keywords:** Data science models, Carbon tax, Energy consumption, Renewable energy subsidies, Policy evaluation

**Introduction**

Climate change is a serious problem affecting the entire planet. Human activity, especially the burning of fossil fuels like coal, oil, and gas, has increased the levels of greenhouse gases in the atmosphere. This has led to higher global temperatures and more extreme weather events. In 2023, global greenhouse gas emissions rose again, reaching around 53 billion tonnes of $CO_2$ equivalent—an increase of 1.9% over 2022[1].

July 2025 was recorded as the third-hottest July in history, and over the 12 months from August 2024 to July 2025, global temperatures were about 1.53 °C higher than pre-industrial levels[2]. If current trends continue, the world risks exceeding the Paris Agreement's target of limiting warming to 1.5 °C above pre-industrial levels. Scientists warn that every fraction of a degree matters, and urgent action is required to avoid severe environmental, social, and economic consequences[3].

Governments around the world have introduced various policies to reduce emissions, with two of the most important being subsidies for clean energy and carbon taxes. Subsidies aim to make renewable energy sources more affordable and attractive to businesses and consumers. Carbon taxes, on the other hand, increase the cost of burning fossil fuels, encouraging industries and individuals to switch to cleaner alternatives. While these policies have the potential to reduce emissions, it is often difficult to determine exactly how effective they are[4] . Results can vary widely depending on the country, its economic conditions, and the specific design of the policies. Traditional methods of evaluation sometimes struggle to provide clear evidence or accurately predict the long-term impact of these measures.

This uncertainty makes it essential for policymakers to have reliable, data-driven evidence that can show whether subsidies and carbon taxes are achieving their intended results[5]. Measurable proof not only helps justify public spending but also ensures that future policies are better targeted and more effective. In the fight

---

[1] European Commission, Joint Research Centre (JRC). (2024). *GHG emissions of all world countries: 2024 report*. Emissions Database for Global Atmospheric Research (EDGAR). https://edgar.jrc.ec.europa.eu/report_2024

[2] (Van Campenhout, C. (2025, August 7). *July was Earth's third-hottest on record, included a record for Turkey, EU scientists say*. Reuters. https://www.reuters.com/sustainability/cop/july-was-earths-third-hottest-record-included-record-turkey-eu-scientists-say-2025-08-07/

[3] Associated Press. (2025, August 7). *World experienced third-warmest July on record this year: EU monitor*. NDTV. https://www.ndtv.com/world-news/world-experienced-third-warmest-july-on-record-this-year-eu-monitor-9042107)

[4] (Van der Ploeg, F. (2025). Why green subsidies are preferred to carbon taxes: Climate policy with heightened carbon tax salience. Journal of Environmental Economics and Management, 125, 103129. https://doi.org/10.1016/j.jeem.2025.103129

[5] (United Nations. (n.d.). Renewable energy – Powering a safer future. https://www.un.org/en/climatechange/raising-ambition/renewable-energy)

against climate change, where time is limited, knowing which policies work best—and by how much—can guide faster and smarter decision-making[6].

Data science tools like ARIMA and LSTM (used for time series forecasting), random forests and neural networks (used for machine learning), and causal inference methods such as synthetic controls and difference-in-differences are now being used to study energy policies. These tools help create realistic “what if” scenarios, predict energy demand, and measure the actual impact of subsidies or carbon taxes. For example, LSTM models have been shown to reduce forecasting errors by 15–20% compared to traditional methods, making evaluations more accurate. By combining large datasets with advanced models, researchers can separate real policy effects from other factors like economic growth or technology changes.

This paper explores how data science and machine learning models can be used to evaluate the effectiveness of subsidies and carbon taxes on energy consumption patterns. By combining large datasets from government reports, energy systems, and emissions records with advanced analytical methods, researchers can track changes in energy use and assess the real impact of these policies. Data science enables us to go beyond simple before-and-after comparisons and to create models that capture the complex, real-world dynamics of policy impacts.

The central question this research addresses is: *How can data science models be used to evaluate the effectiveness of government subsidies or carbon tax policies on energy consumption patterns?* In answering this, the paper considers which models are most useful, what types of data are required, and how accurate and reliable these models are in real-world situations. The focus remains on energy-related greenhouse gas emissions and specific policy tools such as subsidies for renewable energy and carbon pricing. It does not cover other climate policies, such as regulations, cap-and-trade systems without carbon pricing, adaptation planning, or emissions from non-energy sectors like agriculture and forestry. The aim is to model and analyse the link between these selected policies and energy consumption patterns, without extending into broader economic or social impacts.

## 2. Background & Literature Review

Data science models are tools that use data, statistics, and computer power to find patterns, make predictions, and test “what if” situations. In simple terms, they help us learn from past data to understand the present and prepare for the future. These models come in many forms—for example, time series models like ARIMA are used to study changes over time, while deep learning models like LSTM can handle complex and non-linear patterns. Machine learning models such as random forests and neural networks are also widely used to make accurate predictions when many different factors are involved.

[6] (Pan, J., Cross, J. L., Zou, X., & Zhang, B. (2024). To tax or to trade? A global review of carbon emissions reduction strategies. Energy Strategy Reviews, 54, 101508. https://doi.org/10.1016/j.esr.2024.101508)

In everyday life, data science is applied in many areas. Doctors use it to predict disease risks, banks use it to detect fraud, and companies use it to forecast sales. Globally, governments and agencies use it to solve large-scale problems. For instance, the International Energy Agency (IEA) applies machine learning to track renewable energy adoption trends, helping policymakers understand whether subsidies are truly encouraging a shift toward clean energy. Similarly, researchers like **Gasmi et al.** (2024) combined ARIMA and LSTM models to predict electricity demand and showed that hybrid models can reduce error by nearly 20% compared to traditional methods.

By handling very large and complex datasets, data science models make it possible to separate policy effects from other influences like economic growth or technology change. This means governments can get a clearer picture of whether carbon taxes or subsidies are working as intended. These examples show how data science not only improves efficiency but also provides stronger evidence for making smarter and fairer policies.

While data science models are powerful, they also have some drawbacks. Many models act like a "black box," making it hard for policymakers to understand how the results were produced. They also rely on good quality data, but in many countries, energy and emissions data are incomplete or delayed, which reduces accuracy. In some cases, models may overfit the data—meaning they perform well in the past but fail to predict the future correctly. Finally, using these tools often requires high computing power and skilled experts, which may not always be available to governments.

Climate change is happening faster than ever. Over the past decade (2014–2023), global warming accelerated to a record rate of about 0.26 °C per decade, making 2023 the hottest year on record with an average temperature rise of 1.19 °C compared to pre-industrial times[7],[8]. Human activities—especially the burning of coal, oil, and gas—have pushed atmospheric carbon dioxide ($CO_2$) levels to around 427 ppm in 2024, which is about 50 percent higher than before the Industrial Revolution (IPCC, 2023). Energy-related emissions alone reached a record 37.4 gigatonnes of $CO_2$ in 2024[9].

The main sources of these emissions are well known. Power generation, transportation, cement production, and industrial processes burn large amounts of fossil fuels, releasing $CO_2$ and other greenhouse gases into the atmosphere[10]. The link between energy consumption and greenhouse gas emissions is direct—using more

---

[7] (World Meteorological Organization. (2024, March 19). *Climate change indicators reached record levels in 2023: WMO*. https://wmo.int/news/media-centre/climate-change-indicators-reached-record-levels-2023-wmo)(Garric, A. (2024, June 5)

[8] Climate change is accelerating faster than ever. *Le Monde*. https://www.lemonde.fr/en/environment/article/2024/06/05/climate-change-is-accelerating-faster-than-ever_6673802_114.html)

[9] (Lindsey, R., & Miller, J. (2025, May 21). *Climate change: Atmospheric carbon dioxide*. Climate.gov. https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide)

[10] United States Environmental Protection Agency. (n.d.). *Sources of greenhouse gas emissions*. https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

coal, oil, or gas leads to higher $CO_2$ levels. Electricity demand, road transport, and industrial energy use are the biggest contributors.

### 2.1. Policy Mechanisms

Governments use different tools to reduce emissions, with **subsidies for clean energy** and **carbon pricing** being two of the most common. Subsidies make renewable energy like solar, wind, and hydropower cheaper, encouraging households, industries, and power companies to adopt them. For example, the United States, European Union countries, and India have offered incentives for solar and wind projects, which have led to significant growth in renewable energy capacity[11].

**Carbon taxes** work in the opposite way—they make fossil fuel use more expensive by charging a fee for every tonne of $CO_2$ emitted. Sweden, for example, introduced a carbon tax in the 1990s, which has been credited with reducing emissions while maintaining economic growth (World Bank, 2023). Canada and Singapore also use carbon pricing systems to shift energy demand toward cleaner sources. The main aim of these policies is to cut fossil fuel use, reduce greenhouse gas emissions, and promote cleaner energy alternatives[12].

Although these policies are widely used, their actual effectiveness is not always clear. Traditional evaluation methods—such as cost-benefit analysis or simple regression models—compare emissions before and after a policy or look for correlations between policy changes and energy use. However, these methods have limitations. They often cannot handle massive datasets or account for non-linear patterns in the data (IEA, 2023). They also struggle to predict long-term impacts, especially when many factors—such as economic growth, technological changes, and global energy prices—affect energy consumption at the same time.

### Evaluation Problem and Data Science Approaches

Measuring how well government subsidies or carbon taxes change energy consumption patterns is surprisingly difficult using traditional methods. Simple before-and-after comparisons or basic cost-benefit analyses often fail to account for the many overlapping influences that can obscure a policy's true effect. For instance, one study found that a $10 increase in carbon tax per ton of $CO_2$ reduced emissions per capita by only 1.3% in the short run and 4.6% in the long term, even after controlling for economic development, urbanization, and energy mix. But determining whether this drop was truly driven by the tax—and not by technological improvements or economic shifts—is not easy using simple trends alone. Traditional econometric models such as linear regression assume stable and linear relationships, yet in reality, energy consumption is affected

---

[11] Qadir, S. A., Al-Motairi, H., Tahir, F., & Al-Fagih, L. (2021). Incentives and strategies for financing the renewable energy transition: A review. *Energy Reports*, *7*, 3590–3606. https://doi.org/10.1016/j.egyr.2021.06.041).

[12] Government of Sweden. (2020, December). Sweden's long-term strategy for reducing greenhouse gas emissions. United Nations Framework Convention on Climate Change (UNFCCC). https://unfccc.int/sites/default/files/resource/LTS1_Sweden.pdf

by multiple non-linear and delayed responses. A recession, for example, can reduce emissions even without policy change, while economic growth can increase overall energy demand and hide the impact of taxes or subsidies. Similarly, technological advances—such as cheaper solar panels or energy-efficient appliances—can independently drive reductions in fossil fuel use, and consumer behavior often changes due to social norms or crises like pandemics. These overlapping effects make it very hard to isolate the real impact of subsidies or carbon taxes with old statistical tools.

This is where data science provides new possibilities. Advanced time series models, machine learning, and causal inference techniques are now being used to address these gaps. Time series models like ARIMA (AutoRegressive Integrated Moving Average) can track energy demand or emissions before and after a policy is introduced. They work well with stable trends but often struggle when policies create sudden or non-linear changes. That is why deep learning models such as LSTM (Long Short-Term Memory networks) are increasingly applied, as they can capture long-term dependencies and complex seasonal patterns. For example, Gasmi et al. (2024) tested a hybrid ARIMA–LSTM model to forecast electricity demand and showed that LSTM reduced prediction error by nearly 15–20% compared to ARIMA alone. Similarly, a study in Poland (Manowska, 2020) applied LSTM to forecast energy demand across industrial, residential, and commercial sectors, achieving accuracy levels within 1–3% error. These improvements matter because they allow policymakers to build more realistic "no-policy" baselines and then compare them with real-world data to estimate the true effects of subsidies or carbon taxes.

Machine learning approaches such as random forests and neural networks go even further. They can process very large datasets that combine fuel prices, household income, population growth, renewable energy adoption, and even weather data. By doing this, they uncover complex, non-linear relationships that traditional regression methods miss. Bhatt et al. (2023) applied machine learning to forecast global $CO_2$ emissions and found that hybrid models reduced error rates by more than 20% compared to simple regression. Likewise, Jin and Sharifi (2025) reviewed urban greenhouse gas prediction studies and showed that ensemble models like random forests consistently outperformed traditional econometric approaches. These methods are particularly useful for evaluating carbon taxes, where policy impacts interact with fuel prices, consumer choices, and industrial adaptation at the same time. A challenge, however, is that some models, especially deep neural networks, act as "black boxes." Policymakers may struggle to interpret why the model predicts a certain impact, which raises questions of transparency and accountability in decision-making.

Finally, causal inference tools like difference-in-differences combined with machine learning, or synthetic control methods, offer powerful ways to test counterfactuals. For example, researchers can construct a "synthetic" version of a country that did not introduce a subsidy and compare it with the real country that did. By doing so, they isolate the net effect of the policy from other confounding influences. Recent studies in Europe have used these techniques to measure the impact of renewable energy subsidies on emissions reduction, showing that subsidies accounted for almost 40% of the observed shift toward wind and solar between 2010 and 2020.

Together, these approaches address the key weaknesses of traditional evaluation. Data science allows researchers to use large, high-resolution datasets, model complex and delayed effects, and construct reliable counterfactuals. While data gaps and model biases remain, and while explainability is still a challenge, these tools provide clearer and stronger evidence of whether subsidies and carbon taxes are working as intended. By combining statistical rigor with computational power, data science helps policymakers separate real policy signals from background noise, ensuring that public money and carbon pricing strategies are not only ambitious but also effective in changing energy consumption patterns.

For example, hybrid ARIMA-LSTM models have been used to forecast hourly electricity demand with accuracy rates of over 97 percent by combining statistical time-series analysis with neural networks (IEA, 2023). In Denmark, machine learning models have been used to predict $CO_2$ emission intensity by analysing factors like electricity demand, production type, imports, and weather conditions. These predictions were then combined with ARIMA models to improve forecast reliability[13].

Beyond forecasting, simulation tools like ElecSim—a Monte Carlo, agent-based model—have been used to show how a carbon tax of £40 per tonne of $CO_2$ could lead to a 70 percent renewable electricity mix by 2050[14].

In China, machine learning models have been applied to provincial carbon emissions data from 1997 to 2021 to identify key drivers of emissions and predict future trends[15].

Machine learning has also been used at a global scale to forecast $CO_2$ emissions for major economies with accuracy rates above 96 percent, showing its potential to assess and compare policy impacts across countries[16].

[13] (Laid, G., Salheddine, K., Laiche, N., & Nichani, R. (2024). Time series forecasting using deep learning hybrid model (ARIMA-LSTM). *Studies in Engineering and Exact Sciences, 5*(2), e6976. https://doi.org/10.54021/seesv5n2-125)

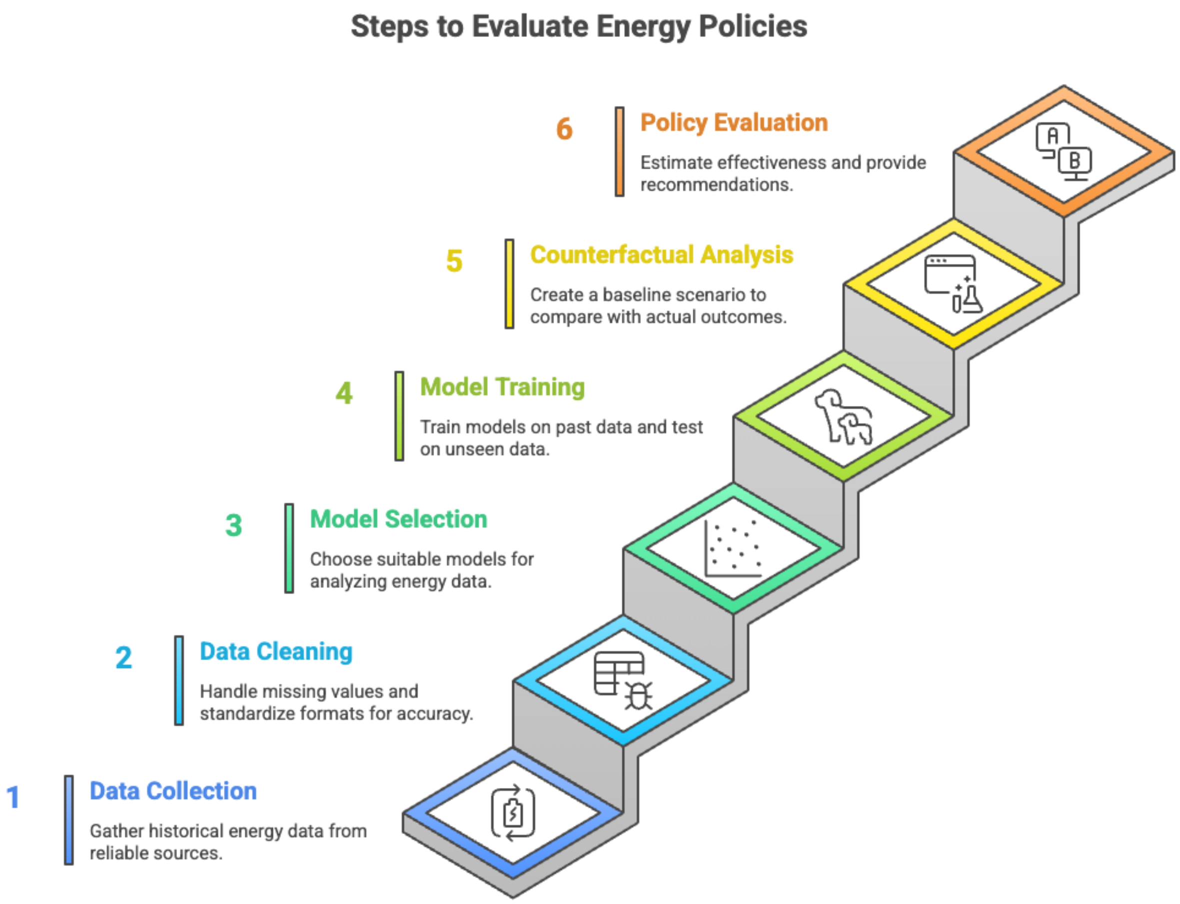
[14] (Hu, Z., Gao, Y., Sun, L., & Mae, M. (2025). A novel attention-enhanced LLM approach for accurate power demand and generation forecasting. *Renewable Energy, 242*, 123465. https://doi.org/10.1016/j.renene.2025.123465

[15] (Bhatt, H., Davawala, M., Joshi, T., Shah, M., & Unnarkat, A. (2023). Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques. *Cleaner and Circular Economy, 4*, 100095. https://doi.org/10.1016/j.clce.2023.100095)

[16] (Jin, Y., & Sharifi, A. (2025). Machine learning for predicting urban greenhouse gas emissions: A systematic literature review. *Renewable and Sustainable Energy Reviews, 188*, 115625. https://doi.org/10.1016/j.rser.2025.115625).

### 3. Data Science Framework for Policy Evaluation

Evaluating the effectiveness of government subsidies or carbon tax policies on energy consumption patterns requires a structured, step-by-step approach. Data science provides the tools to collect, clean, analyse, and model large and complex datasets, making it possible to understand the true impact of these policies. The framework below outlines how this process can be carried out.




### Data Collection

The first step in any data science project is **gathering the right data**. Without accurate and relevant data, even the most advanced models will produce misleading results. For policy evaluation, three main categories of data are needed:

1. **Emissions Data** – This includes carbon dioxide ($CO_2$) and methane ($CH_4$) emissions. Government agencies such as the **U.S. Energy Information Administration (EIA)**, the **International Energy Agency (IEA)**, and the **Intergovernmental Panel on Climate Change (IPCC)** regularly publish detailed datasets. For example, the IEA tracks national and sector-level $CO_2$ emissions, while the EIA provides annual, monthly, and real-time statistics for the U.S. and other countries.

2. **Energy Mix Data** – This describes how much of a country's energy comes from coal, oil, natural gas, nuclear, and renewable sources. Changes in the energy mix often indicate shifts driven by policies. If subsidies are effective, renewable energy's share should rise while fossil fuel use declines.

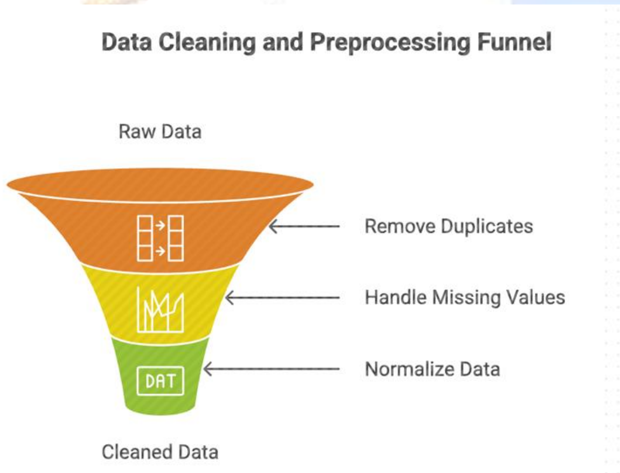
3. **Policy Records** – These include the dates policies were introduced, the amount of subsidies given, tax rates applied, and any policy changes over time. For example, Sweden's carbon tax started at €24 per tonne in 1991 and has increased to over €110 per tonne today. Tracking such details is critical for linking policies to changes in energy use.

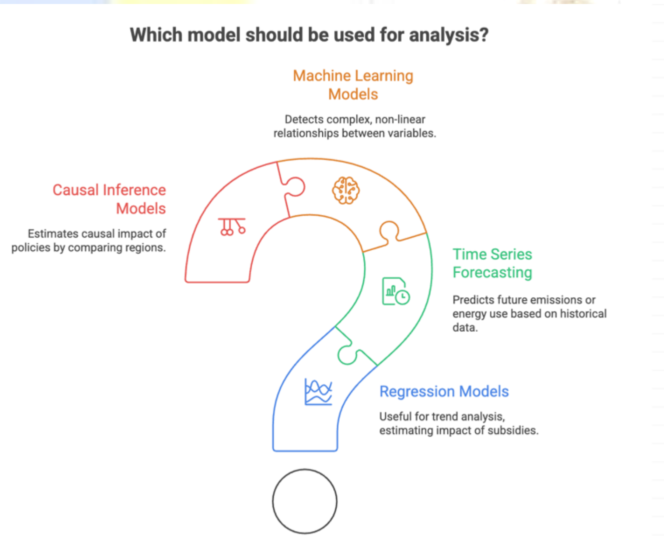
Other useful data sources include:

**Satellite data** on emissions and air quality (e.g., NASA OCO-2 satellite for $CO_2$ levels).
**Smart meter data** from households and businesses to capture detailed electricity use patterns.
**Transportation datasets** on fuel consumption, vehicle registrations, and distances travelled.

### Data Preparation and Model Selection

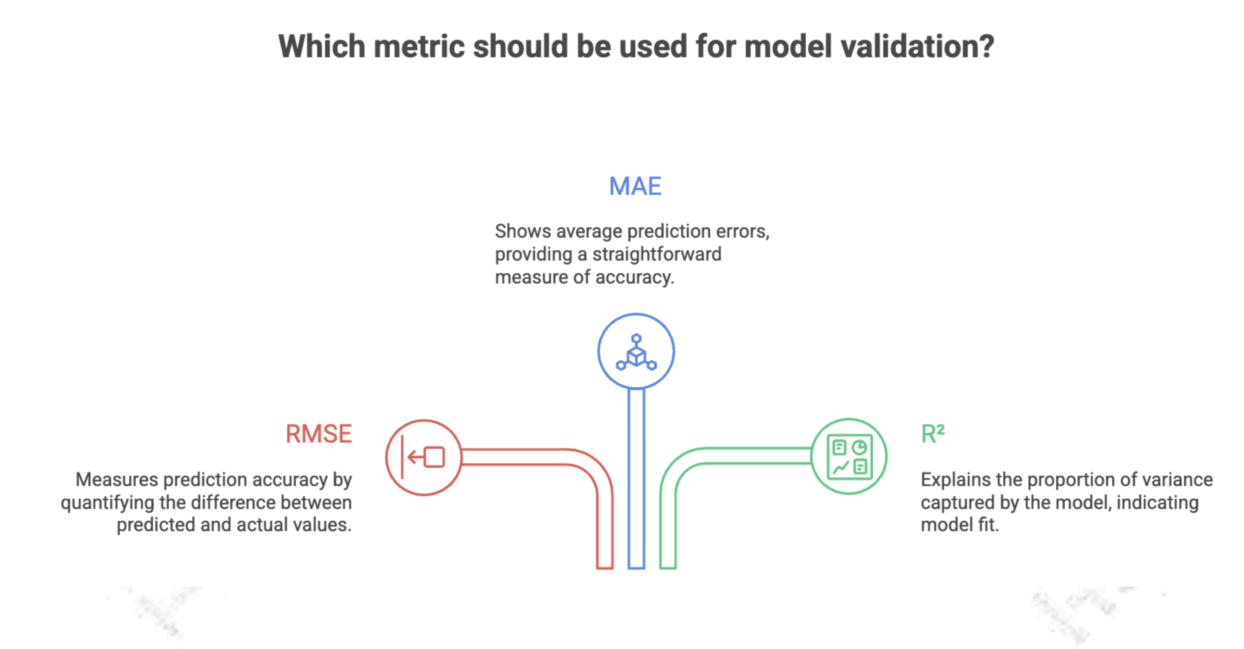
Before analyzing the effectiveness of subsidies or carbon taxes, data must be properly prepared. Raw data is often messy, with missing values, inconsistent formats, or errors that can distort results. Data cleaning helps ensure accuracy by removing duplicates, filling missing values with averages or estimates, and standardizing units like tonnes of $CO_2$, kilowatt-hours, or dollars for better comparison. Tools such as Python (with libraries like *pandas* and *NumPy*) or R are commonly used in this stage, as they make it easier to detect errors and restructure data consistently. Without careful cleaning, incorrect conclusions may be drawn about whether a policy is effective.



Once the data is ready, the next step is model selection. Different models suit different types of analysis. For instance, regression models are useful for identifying trends—for example, estimating how much emissions might fall for every $1 million invested in renewable energy subsidies. Time series forecasting methods like ARIMA or Prophet can predict future energy use and emissions under different policy scenarios, such as testing what would happen if a carbon tax increased by 20%. Machine learning models, including Random Forests, Gradient Boosting, and Neural Networks, are especially powerful at capturing complex, non-linear relationships between variables such as energy prices, weather, and consumer behavior. Causal inference models, like Difference-in-Differences and Synthetic Control, help establish whether a policy directly caused changes by comparing regions with and without interventions. A well-known example is the use of synthetic control to estimate how Sweden's emissions would have evolved without its carbon tax, compared against the actual observed outcomes.

Finally, no model is complete without validation. Reliable validation ensures that findings are not based on random noise but on solid evidence. Common validation metrics include RMSE (Root Mean Squared Error) to measure prediction gaps, MAE (Mean Absolute Error) for average errors, and $R^2$ to show how much variation is explained by the model. Cross-validation techniques, where data is split into training and testing sets, also help test accuracy. This step ensures that conclusions—such as "a subsidy reduced fossil fuel use by 10% "—are credible and reliable for policymakers.




**Cross-validation** is often used, where the dataset is split into training and testing parts. The model is trained on one part and tested on the other to see how well it predicts outcomes.

Reliable validation ensures that the conclusions drawn—such as "this subsidy reduced fossil fuel use by 10%"—are based on solid evidence and not random chance.

## Case Studies & Applications

Case studies help show how data science models can be used to evaluate the effectiveness of subsidies and carbon tax policies in different countries. By looking at real-world examples, we can see how these policies change energy use patterns, how much they reduce emissions, and how well predictions match reality. The three examples below—Sweden, India, and a hybrid policy scenario—illustrate this.

### Country Example: Sweden's Carbon Tax

Sweden introduced a carbon tax in 1991, starting at about €24 per tonne of $CO_2$ and steadily increasing to over €110 per tonne by 2024 (World Bank, 2023). This makes Sweden one of the countries with the highest carbon prices in the world.

To evaluate its effectiveness, researchers have used **data science models** that track changes in fossil fuel use and emissions over more than 30 years. Historical emissions data from the IEA and national statistics show that Sweden's per capita $CO_2$ emissions fell by about **27% between 1991 and 2022**, even as the economy grew.

| Policy Type | Country | Year | CO2 Emissions Reduction (%) | Renewable Energy Share (%) |
|---|---|---|---|---|
| **Carbon Tax** | Sweden | 1991 | 25 | 54 |
| **Renewable Subsidy** | Sweden | 2020 | 15 | 60 |
| **Carbon Tax** | India | 2010 | 5 | 17 |
| **Renewable Subsidy** | India | 2015 | 12 | 22 |
| **Carbon Tax** | Canada | 2008 | 15 | 19 |
| **Renewable Subsidy** | Canada | 2017 | 18 | 25 |

**How data science helps:**

**Data collection:** Annual emissions by sector (transportation, industry, power generation), fuel consumption data, and tax rates.

**Exploratory analysis:** Visualising trends shows a steady decline in oil and coal use, especially in heating, where fossil fuels have been replaced by biofuels and district heating.

**Modeling:** Using regression and **difference-in-differences** models, analysts compare Sweden's emissions to a synthetic "control country" without a carbon tax. The results indicate that without the tax, emissions could have been **12–15% higher** than actual levels.

**Forecasting:** Time-series models such as ARIMA have been applied to project future emissions, showing that maintaining or increasing the tax could help Sweden meet its 2030 climate targets.

The Swedish case shows that a well-designed and steadily increasing carbon tax can significantly cut fossil fuel use without harming economic growth, and data science methods make these impacts measurable.

**Case Study: Forecasting Electricity Demand in Poland Using LSTM**

A recent study by Manowska (2020) demonstrates how Long Short-Term Memory (LSTM) networks can significantly improve electricity demand forecasting in Poland, a country heavily reliant on coal. Poland faces mounting pressure from the European Union to reduce emissions and increase renewable energy use, while also ensuring energy security and affordability. These challenges make accurate long-term electricity demand forecasting essential for planning subsidies, capacity markets, and investments in renewable energy.

Traditional statistical tools such as ARIMA often struggle with Poland's energy data because it is non-stationary and influenced by many sudden shifts, such as EU environmental regulations, rising household incomes, or technological adoption. The study applied LSTM models across multiple sectors — industry, residential, transport, commercial services, and agriculture — using limited statistical datasets. Results showed that LSTM achieved remarkable accuracy, with mean absolute percentage error (MAPE) as low as **1–3%**, far outperforming conventional models. For example, forecasts revealed that electricity demand in commercial services could rise by nearly **19% by 2040**, while household demand would grow steadily by **7%**, reflecting income growth and electrification trends.

These findings highlight how LSTM can capture long-term, non-linear patterns that standard time series models cannot. Importantly, by accurately predicting demand across sectors, LSTM enables policymakers to better evaluate the effectiveness of subsidies or carbon taxes. For instance, if subsidies increase solar adoption or carbon taxes reduce coal dependency, LSTM can track these shifts in consumption more reliably than older models. In the context of global climate policy, the Poland case demonstrates that deep learning tools are not only academically robust but also practically essential for guiding energy transitions and ensuring the efficient use of public resources[17].

A recent study by Jailani et al. (2023) explored the use of Long Short-Term Memory (LSTM) models for forecasting solar energy and highlighted why these models are powerful for evaluating energy systems. Solar energy is highly variable because of weather changes, cloud cover, and seasonal shifts, which makes it difficult to forecast using traditional statistical tools. The authors compared standalone LSTM models with hybrid models that combine LSTM with other deep learning techniques like Convolutional Neural Networks (CNN). Their findings show that while standalone LSTM performs better than traditional machine learning methods in capturing time-series patterns, hybrid models such as CNN–LSTM consistently achieved the lowest error rates, even though they required longer training times. In some case studies, hybrid LSTM models reduced

---

[17] Manowska, A. (2020). Using the LSTM network to forecast the demand for electricity in Poland. *Applied Sciences*, *10*(23), 8451. https://doi.org/10.3390/app10238451

forecasting errors like Root Mean Square Error (RMSE) by more than 20% compared with standard models, proving their accuracy and reliability.

This research is relevant to evaluating the impact of subsidies and carbon taxes because accurate energy forecasting is crucial for understanding how policies change consumption patterns and renewable energy adoption. For instance, if subsidies increase solar adoption, an LSTM model can forecast the expected power generation more precisely than linear models, making it easier to measure whether the subsidy truly reduced fossil fuel dependence. Similarly, reliable energy demand forecasts can test how carbon taxes shift consumption toward renewables. By showing that LSTM-based approaches can outperform traditional models, this study strengthens the argument that advanced data science tools are essential for robust policy evaluation[18].

**Country Example: India's Renewable Subsidies**

India is one of the fastest-growing energy consumers in the world, and coal still makes up over 70% of its electricity generation[19]. To change this, the Indian government has introduced several renewable energy subsidies, including capital cost subsidies for solar farms, feed-in tariffs for wind power, and incentives for rooftop solar installation.

Between 2015 and 2023, India's installed solar capacity grew from less than 5 GW to over 66 GW, and wind capacity increased from 25 GW to more than 43 GW[20] (MNRE, 2024). This is a huge leap, but the question remains: how much did subsidies contribute compared to other factors like falling technology costs?

**How data science helps:**

**Data collection:** Records of annual subsidy amounts, renewable capacity installed, electricity generation mix, and coal consumption.

**Exploratory analysis:** Comparing periods before and after major subsidy programs, visualising growth in renewables alongside changes in coal use.

**Modeling:** Using **machine learning models** such as Random Forest and Gradient Boosting to predict renewable energy growth based on inputs like subsidy size, technology cost trends, and electricity demand. The models show that subsidies account for roughly **35–40% of the growth in solar capacity** in the past decade.

---

[18] Jailani, N. L. M., Dhanasegaran, J. K., Alkawsi, G., Alkahtani, A. A., Phing, C. C., Baashar, Y., Capretz, L. F., Al-Shetwi, A. Q., & Tiong, S. K. (2023). Investigating the power of LSTM-based models in solar energy forecasting. *Processes*, *11*(5), 1382. https://doi.org/10.3390/pr11051382

[19] International Energy Agency. (2023). *India: Energy system overview*. https://www.iea.org/countries/india).

[20] (Press Information Bureau. (2025, June 22). India's energy landscape: Powering growth with sustainable energy. https://www.pib.gov.in/PressNoteDetails.aspx?id=154717&NoteId=154717&ModuleId=3

**Scenario testing:** Simulation models project that if current subsidies are maintained until 2030, India's coal share in electricity generation could drop from 70% to about 55%, helping to cut annual $CO_2$ emissions by an estimated 200 million tonnes.

This case shows that subsidies are a powerful tool for accelerating renewable adoption, but their effectiveness can be clearly measured only when data science separates their impact from other influencing factors.

**Case Study: Using the GLIMPSE model to holistically assess the impacts of policies.**

Another example of using data science models to forecast the impacts of policies is the GLIMPSE tool provided by the United States Environmental Protection Agency (GLIMPSE – A computational framework for supporting state-level environmental and energy planning | US EPA). GLIMPSE is a decisionmaking tool that uses the GCAM (Global Change Analysis Model) simulation model over a more traditional regression or classification. GCAM takes into account energy systems, land and water use, climate dynamics, and economic activity, creating scenario-based simulations instead of using direct statistical inference. Instead of historical data, GCAM predictions are shaped by user defined assumptions about the future, such as population growth rates and technological advances, which output emissions, energy mixes, and climate impacts.

The GLIMPSE software has been used in tandem with other frameworks to create sets of parameters that show how climate, the economy, and energy mix change given different socioeconomic scenarios. The IPCC, which utilised GLIMPSE, created 5 sets of SSPs (Shared Socioeconomic Pathways) and a similar amount of RCPs (Representative Concentration Pathways) which are used together by policymakers to review pre-existing policies and inform the creation and tuning of new ones. Thus, these findings show how scenario models such as those provided by GLIMPSE are able to simulate long periods of time across multiple sectors that traditional economic models are not able to do.

**Combined Policy Impacts**

Some countries are experimenting with **both subsidies and carbon pricing** at the same time. A combined approach can offer stronger results by pushing users away from fossil fuels through higher costs while pulling them toward renewables through lower costs.

For example, simulations using **agent-based modeling** and **system dynamics** have been applied to test how a country could achieve its 2030 climate goals with a mix of policies. In these models, a carbon tax of $50 per tonne combined with renewable energy subsidies covering 20% of project costs led to:

**Faster coal phase-out** compared to carbon tax alone.
**Higher renewable adoption rates**, reaching 65–70% of total generation by 2030.
**Greater emissions reductions**, with up to 25% more $CO_2$ avoided than with a single policy approach.

**How data science helps here:**

**Integration of datasets:** Combining energy mix data, emissions data, and economic data into a single simulation framework.

**Scenario analysis:** Running "what-if" simulations to compare outcomes of different policy mixes.

**Visualisation:** Using dashboards and charts to communicate results to policymakers, making it easy to see the trade-offs between costs, emissions, and energy security.

Studies also show that hybrid policies can make the transition fairer. For example, part of the revenue from carbon taxes can be used to finance renewable subsidies, reducing the financial burden on households while still cutting emissions.

By collecting and analysing detailed data, building predictive and causal models, and running scenario tests, data science enables governments to make informed decisions about how to design and adjust policies for maximum effectiveness. Without these tools, it would be much harder to separate policy effects from other changes in the energy market, making it difficult to know which strategies truly work in the fight against climate change.

**Discussion**

The results from case studies and applications show that **data science models** can play a major role in evaluating the effectiveness of government subsidies and carbon tax policies on energy consumption patterns. By combining large amounts of data, advanced statistical methods, and machine learning techniques, these models give policymakers a much clearer picture of how policies work in the real world.

**Effectiveness of Data Science Models in Real-World Policy Evaluation**

One of the biggest challenges in climate policy evaluation is separating the impact of a specific policy from other influencing factors, such as global energy prices, economic growth, or new technology developments. Data science models can help overcome this problem.

For example, in Sweden, **difference-in-differences** models have shown that the carbon tax directly contributed to lowering fossil fuel use compared to similar countries without the tax. In India, **machine learning models** have been able to estimate how much of the growth in renewable energy was due to subsidies, rather than falling technology costs. These examples prove that data science can produce **quantifiable, evidence-based results** rather than relying on assumptions or general trends(Dwivedi, Y. K., Hughes, L., et al. (2021). Climate change and COP26: Are digital technologies and information management part of the problem or the solution? An editorial reflection and call to action[21].

[21] International Journal of Information Management, 63, 102456. https://doi.org/10.1016/j.ijinfomgt.2021.102456) .

Data science is also effective because it can track changes over time and across different sectors. A carbon tax may affect transportation differently from power generation, and subsidies may have a bigger impact on solar power than wind. By processing detailed datasets, models can identify these sector-specific effects and guide targeted policy adjustments.

**Strengths of Data Science Models**

**1. Handles Large and Complex Data**

Modern energy systems produce massive amounts of data—from satellite observations of emissions to smart meter readings in households. Data science tools can handle millions of data points from multiple sources without losing accuracy. For example, the International Energy Agency (IEA) collects global energy and emissions data from more than 150 countries, and models can integrate this with local policy and economic data to form a complete picture.

**2. Predictive Power**

Machine learning models, such as **Random Forests** or **Neural Networks**, can identify patterns and predict future energy use under different policy scenarios. This predictive ability allows policymakers to test the potential effects of changing subsidy levels or increasing carbon tax rates before implementing them. For instance, simulations show that combining a $50 per tonne carbon tax with renewable subsidies could achieve 25% greater $CO_2$ reductions by 2030 compared to a single policy approach.

**3. Scenario Testing and Simulation**

Data science allows for “what-if” analysis. Governments can simulate different futures:

- What happens if oil prices rise sharply?
- What if renewable energy technology costs drop faster than expected?
- How would energy use change if subsidies are reduced?

This flexibility is valuable because climate policy decisions often involve uncertainty. Scenario testing helps choose policies that work well across a range of possible futures, reducing the risk of policy failure.

Challenges and Limitations **of Data Science Models**

Evaluating the effectiveness of government subsidies and carbon tax policies with data science tools is promising, but it is not without challenges. While models like ARIMA, LSTM, random forests, and synthetic controls can provide useful insights, their accuracy and reliability depend heavily on data quality, modeling choices, and external conditions. This section explains the main challenges and limitations.

The biggest limitation in policy evaluation is data. Many countries still lack detailed, high-frequency data on energy consumption and emissions. Even in developed countries, data from different agencies might use different definitions or formats, requiring careful cleaning. Missing or poor-quality data can lead to inaccurate results.According to the World Bank (2024), over 60% of developing countries report their emissions with a lag of more than two years. This delay makes it hard to use real-time models like LSTM for policy analysis. Even in advanced economies, data often comes from multiple sources (energy agencies, tax records, company reports), and the formats or definitions may not match. For example, electricity use may be reported monthly in one country and annually in another, making cross-country comparisons unreliable. Poor or incomplete data leads to errors in models and can result in misleading conclusions about whether subsidies or taxes truly worked. Furthermore Machine learning models are only as good as the data they are trained on. If the data reflects historical biases—such as underestimating renewable energy growth in low-income regions—the model might produce biased predictions. This is especially important when using models to guide funding or tax policy, as biases could unfairly disadvantage certain groups or regions.

Separating the effect of a subsidy or carbon tax from other factors is extremely difficult. Economic growth, recessions, new technologies, and consumer behavior all influence energy use. For instance, during the COVID-19 pandemic, global $CO_2$ emissions fell by 5.4% in 2020 (IEA, 2021), but this was not because of carbon pricing—it was due to reduced travel and industrial slowdown. A traditional model might wrongly attribute such a drop to policy. Even advanced data science tools can struggle to fully disentangle these overlapping factors, especially when policies are introduced alongside other measures such as renewable energy targets or efficiency regulations.

Data science models are powerful, but they are not perfect. Time-series models like ARIMA assume stability and may fail when policies trigger sudden shifts in behavior. LSTM models can capture non-linear changes, but they require huge amounts of training data and are sensitive to overfitting, where the model learns past data too well and fails to predict future trends. A study by Gasmi et al. (2024) showed that hybrid ARIMA–LSTM models reduced electricity demand forecasting errors by 15–20%, but errors still remained, especially during unexpected shocks like extreme weather. Similarly, machine learning models such as random forests can identify complex relationships, but their “black box” nature makes it hard for policymakers to understand why the model gives a certain prediction. This lack of transparency can reduce trust in the results.

Governments rarely use a single policy. Subsidies often come with tax incentives, efficiency standards, or renewable energy targets. This makes it hard to say which policy produced which effect. For example, Sweden introduced both a carbon tax in 1991 and renewable energy subsidies in the same decade. Studies later found that emissions fell by nearly 25% between 1990 and 2019 (EDGAR, 2024), but separating how much was due to the tax versus the subsidies is extremely difficult. Another limitation is the rebound effect, where efficiency gains lower costs and encourages higher consumption. A subsidy for electric vehicles may reduce oil demand, but if cheaper electricity leads people to drive more, the total energy savings may be smaller than expected.

Data science models face political and ethical challenges. Governments may not share complete data, fearing negative results. For instance, some carbon pricing experiments in Asia underreported energy use reductions because industries lobbied against strict reporting. Moreover, while complex AI models can show high accuracy, their use in policymaking requires transparency, accountability, and fairness. Policymakers need to understand and trust the model, not just rely on a "black box" output.

When using data science in policymaking, **transparency** is critical. Policymakers need to understand how a model works, what data it uses, and what assumptions it makes. This is where **explainable AI** becomes important. Models should not be "black boxes" where decisions are made without clear reasoning. Instead, results should be presented in a way that both experts and the public can understand.

Another ethical issue is **data privacy**. Using household energy consumption data from smart meters can reveal personal habits. This data should be anonymised and protected to ensure individual privacy while still allowing for meaningful policy analysis.

Finally, there is the question of **equity**. Policies evaluated by data science models must consider not just efficiency, but fairness. For example, a carbon tax might reduce emissions effectively, but if it places a heavy burden on low-income households, policymakers should consider using part of the revenue for rebates or subsidies to offset these costs.

**Recommendations**

The findings of this paper show that data science models can be a powerful tool for evaluating the effectiveness of government subsidies and carbon tax policies. However, to make the best use of these models, policymakers, researchers, and other stakeholders need to improve how data is collected, shared, and integrated. One important step is to improve policy design by using real-time data monitoring. Many current evaluations rely on yearly or quarterly data, which can delay the detection of problems or successes. Governments should invest in systems such as smart meters for electricity use, satellite-based emissions monitoring, and automated reporting from power plants. By linking these systems directly to data science models, it becomes possible to track the immediate effects of subsidies or carbon taxes[22].If a subsidy for solar installations results in a sudden rise in renewable generation, this can be confirmed within weeks instead of years, allowing for faster policy adjustments before large sums are spent with little impact[23].

[22] (Carl, J., & Fedor, D. (2016). Tracking global carbon revenues: A survey of carbon taxes versus cap-and-trade in the real world. *Energy Policy, 96,* 50–77. https://doi.org/10.1016/j.enpol.2016.05.023).

[23] (Pan, J., Cross, J. L., Zou, X., & Zhang, B. (2024). To tax or to trade? A global review of carbon emissions reduction strategies. *Energy Strategy Reviews, 54,* 101508. https://doi.org/10.1016/j.esr.2024.101508)

Another key recommendation is to encourage open data sharing between governments and researchers. In many cases, governments collect high-quality energy and emissions data but restrict access to internal agencies. Opening these datasets to universities, research institutes, and independent analysts would lead to more thorough evaluations and a wider range of perspectives. Creating national and international open data platforms—similar to those maintained by the International Energy Agency—would allow independent verification of government results and build public trust by showing the evidence behind policy decisions[24].

It is also important to integrate economic, environmental, and social datasets so that models can capture the full range of policy impacts. A carbon tax may reduce emissions but also influence fuel prices, household budgets, and industrial competitiveness. Likewise, subsidies for renewable energy may create new jobs in the clean energy sector while reducing employment in fossil fuel industries. By combining economic data such as energy prices and employment figures with environmental data like emissions and air quality, as well as social indicators such as household energy access and poverty levels, models can identify trade-offs and help design policies that are both effective and fair.

Finally, governments should build capacity and skills within their own agencies to make better use of data science. This means investing in training programs, forming partnerships with universities, and recruiting skilled data scientists to strengthen the ability of institutions to make evidence-based decisions. By focusing on real-time monitoring, open data sharing, integrated datasets, and strong institutional capacity, governments can ensure that subsidies and carbon taxes deliver the maximum possible environmental, economic, and social benefits.

## Conclusion

The purpose of this research was to explore how data science models can be used to evaluate the effectiveness of government subsidies and carbon tax policies on energy consumption patterns. Climate change, driven largely by greenhouse gas emissions from fossil fuel use, remains one of the most urgent challenges of our time. Governments have introduced policies like subsidies for renewable energy and carbon pricing to encourage cleaner energy use, but the real challenge lies in measuring whether these policies are actually working. This paper has shown that data science offers powerful tools for answering this question.

The findings from case studies in countries like Sweden and India, as well as simulations of combined policy approaches, demonstrate that data science models can process vast amounts of information, detect patterns, and link policy actions to measurable outcomes. In Sweden, long-term emissions data and causal models showed that a steadily increasing carbon tax reduced fossil fuel use without slowing economic growth. In India, machine learning models revealed that subsidies played a significant role in driving the rapid expansion

[24] Finley, A., He, W., Huang, H., & Hon, C. (2024). Analyzing the effectiveness of carbon pricing instruments in reducing carbon emissions in major Asian economies. *Sustainability, 16*(23), 10542. https://doi.org/10.3390/su162310542.

of solar and wind capacity, reducing dependence on coal. In combined policy simulations, integrating taxes with subsidies led to greater emissions reductions than either policy alone.

Data science models stand out because they can handle large and complex datasets, predict future trends, and test "what-if" scenarios to guide policy adjustments before large resources are committed. However, these models also have limitations. Gaps in data, bias in training datasets, and uncertainties in long-term forecasts mean that results must be interpreted carefully. Ethical considerations, such as ensuring transparency in model design and protecting personal data, are also essential if these tools are to be trusted in policymaking.

When looking between models, different approaches are best suited for different contexts. For example: short-term policy evaluation, time-series models like ARIMA or hybrid ARIM-LSTM perform most ideally at forecasting near-term energy demand shifts, whilst longer-term structural changes are better captured by deep learning models such as LSTM or CNN-LSTM hybrids, and by integrated assessment tools like GLIMPSE or GCAM. When considering attribution (identifying whether a specific policy directly caused observed changes), causal inference methods such as difference-in-differences and synthetic controls are strongest. Contrastingly, when prediction and scenario testing are the priority, machine learning approaches such as random forests and neural networks excel, especially in data-rich environments. Policymakers must also balance interpretability against complexity: regression-based and causal models offer clearer explanations, while deep learning and ensemble models provide higher accuracy but can function as "black boxes." Finally, single-policy impacts are best measured with time-series and causal models, while combined policies often require system dynamics or agent-based modeling to capture interactions across decades.

Looking ahead, future research should focus on AI-powered real-time monitoring systems that can detect policy impacts almost immediately. Combining smart meter data, satellite emissions tracking, and automated data feeds could make policy evaluation far more responsive. Another important area is the development of integrated climate-economic models that combine environmental, economic, and social datasets, allowing for a more complete understanding of policy impacts, including fairness and equity considerations.

## References

[1] European Commission, Joint Research Centre (JRC). (2024). *GHG emissions of all world countries: 2024 report*. Emissions Database for Global Atmospheric Research (EDGAR). https://edgar.jrc.ec.europa.eu/report_2024

[2] (Van Campenhout, C. (2025, August 7). *July was Earth's third-hottest on record, included a record for Turkey, EU scientists say*. Reuters. https://www.reuters.com/sustainability/cop/july-was-earths-third-hottest-record-included-record-turkey-eu-scientists-say-2025-08-07/

[3] Associated Press. (2025, August 7). *World experienced third-warmest July on record this year: EU monitor*. NDTV. https://www.ndtv.com/world-news/world-experienced-third-warmest-july-on-record-this-year-eu-monitor-9042107)

[4] (Van der Ploeg, F. (2025). Why green subsidies are preferred to carbon taxes: Climate policy with heightened carbon tax salience. Journal of Environmental Economics and Management, 125, 103129. https://doi.org/10.1016/j.jeem.2025.103129

[5] (United Nations. (n.d.). Renewable energy – Powering a safer future. https://www.un.org/en/climatechange/raising-ambition/renewable-energy)

[6] (Pan, J., Cross, J. L., Zou, X., & Zhang, B. (2024). To tax or to trade? A global review of carbon emissions reduction strategies. Energy Strategy Reviews, 54, 101508. https://doi.org/10.1016/j.esr.2024.101508)

[7] (World Meteorological Organization. (2024, March 19). *Climate change indicators reached record levels in 2023: WMO*. https://wmo.int/news/media-centre/climate-change-indicators-reached-record-levels-2023-wmo)(Garric, A. (2024, June 5)

[8] Climate change is accelerating faster than ever. *Le Monde*. https://www.lemonde.fr/en/environment/article/2024/06/05/climate-change-is-accelerating-faster-than-ever_6673802_114.html)

[9] (Lindsey, R., & Miller, J. (2025, May 21). *Climate change: Atmospheric carbon dioxide*. Climate.gov. https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide)

[10] United States Environmental Protection Agency. (n.d.). *Sources of greenhouse gas emissions*. https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

[11] Qadir, S. A., Al-Motairi, H., Tahir, F., & Al-Fagih, L. (2021). Incentives and strategies for financing the renewable energy transition: A review. *Energy Reports*, *7*, 3590–3606. https://doi.org/10.1016/j.egyr.2021.06.041).

[12] Government of Sweden. (2020, December). Sweden's long-term strategy for reducing greenhouse gas emissions. United Nations Framework Convention on Climate Change (UNFCCC). https://unfccc.int/sites/default/files/resource/LTS1_Sweden.pdf

[13] (Laid, G., Salheddine, K., Laiche, N., & Nichani, R. (2024). Time series forecasting using deep learning hybrid model (ARIMA-LSTM). *Studies in Engineering and Exact Sciences*, *5*(2), e6976. https://doi.org/10.54021/seesv5n2-125)

[14] (Hu, Z., Gao, Y., Sun, L., & Mae, M. (2025). A novel attention-enhanced LLM approach for accurate power demand and generation forecasting. *Renewable Energy, 242*, 123465. https://doi.org/10.1016/j.renene.2025.123465

[15] (Bhatt, H., Davawala, M., Joshi, T., Shah, M., & Unnarkat, A. (2023). Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques. *Cleaner and Circular Economy, 4*, 100095. https://doi.org/10.1016/j.clce.2023.100095)

[16] (Jin, Y., & Sharifi, A. (2025). Machine learning for predicting urban greenhouse gas emissions: A systematic literature review. *Renewable and Sustainable Energy Reviews, 188,* 115625. https://doi.org/10.1016/j.rser.2025.115625).

[17] Manowska, A. (2020). Using the LSTM network to forecast the demand for electricity in Poland. *Applied Sciences*, *10*(23), 8451. https://doi.org/10.3390/app10238451

[18] Jailani, N. L. M., Dhanasegaran, J. K., Alkawsi, G., Alkahtani, A. A., Phing, C. C., Baashar, Y., Capretz, L. F., Al-Shetwi, A. Q., & Tiong, S. K. (2023). Investigating the power of LSTM-based models in solar energy forecasting. *Processes*, *11*(5), 1382. https://doi.org/10.3390/pr11051382

[19] International Energy Agency. (2023). *India: Energy system overview*. https://www.iea.org/countries/india).

[20] (Press Information Bureau. (2025, June 22). India's energy landscape: Powering growth with sustainable energy. https://www.pib.gov.in/PressNoteDetails.aspx?id=154717&NoteId=154717&ModuleId=3

[21] International Journal of Information Management, 63, 102456. https://doi.org/10.1016/j.ijinfomgt.2021.102456) .

[22] (Carl, J., & Fedor, D. (2016). Tracking global carbon revenues: A survey of carbon taxes versus cap-and-trade in the real world. *Energy Policy*, *96,* 50–77. https://doi.org/10.1016/j.enpol.2016.05.023).

[23] (Pan, J., Cross, J. L., Zou, X., & Zhang, B. (2024). To tax or to trade? A global review of carbon emissions reduction strategies. *Energy Strategy Reviews*, *54,* 101508. https://doi.org/10.1016/j.esr.2024.101508)

[24]Finley, A., He, W., Huang, H., & Hon, C. (2024). Analyzing the effectiveness of carbon pricing instruments in reducing carbon emissions in major Asian economies. *Sustainability*, *16*(23), 10542. https://doi.org/10.3390/su162310542.