Vehicle type distribution and CO_2 emissions in California

Melissa Chen

December 2022

Abstract

In 2022, California announced its intention to move towards making all new car sales zero-emissions vehicles by 2035, a trend more common in European countries but the first of its scale in the United States. We hope to understand the impact that electric vehicles will have on electricity demand and by consequence on the state's carbon emissions. In this project, we use a simple approach for projecting numbers of registered cars by fuel type (gasoline, electric, plug-in hybrid electric...) in the near future based on current growth trends with more or less optimism for EV growth. Using these distributions, we study stacking linear regressors with Random Forest and Gradient Boosting regressors for predicting monthly electricity demand and its associated carbon emissions through energy generation.

1   Introduction

On 25 August, 2022, California Governor Gavin Newsom shared news that the state would take an exceptional step towards mitigating its carbon emissions: starting in 2035, zero-emission vehicles (ZEVs) must account for 100% of all new vehicle sales. Approved by the California Air Resource Board (CARB), this policy is the first of its scale in the United States, with the goal of drastically reducing direct carbon emissions from traditional combustion engine vehicles and eventually hitting zero direct emissions generated by personal transport for the average Californian. It sets intermediary targets of ZEV sales representing 35% of all car sales by 2026 and 68% by 2030, or doubling the 2022 figure of 17.7% within four years and almost quadrupling it within eight. (Office of Governor Gavin Newsom, 2022a,c)

California has already gained a reputation as a home for electric vehicles (EVs), the most popular type of ZEV. More than 40% of all EVs in the U.S. live in California, which topped the state rankings of EV sales in 2021, outnumbering the next 10 states combined. (Office of Governor Gavin Newsom, 2022b) With the proliferation of EVs, Californians now shift emissions from one source to another, from combustion engine vehicles to the energy generation process necessary to fuel every new Tesla that hits the road. We hope therefore to understand this trade-off and how increased electricity demand from EVs will reflect in the state's carbon emissions.

The final process developed follows these steps: (i) choose a projection year and month in the near future and generate sample vehicle distributions for this date, (ii) using a stacked regressor, electric and plug-in hybrid vehicle numbers from these samples, and average maximum temperature and average precipitation for the given month, predict electricity demand for that month, (iii) using a second stacked regressor, use the predicted electricity demand to estimate total carbon emissions from energy generation for that month.

Given the exploratory nature of this project and the current intended audience size (one), this paper may include passages atypical for a formal technical report, such as interjections regarding personal thought process.

2 Background

2.1 Related literature

Accurately predicting regional greenhouse gas (GHG) emissions poses a challenge for many reasons including difficulty gathering precise data on a large-scale, the breadth of contributors to GHG emissions, and identifying each factor's importance and non-linear relation with emissions. These factors, however, make certain machine learning techniques viable tools, for example support vector machines (SVMs) and modified SVMs in the case of Saleh et al. (2016) and Ehteram et al. (2021). Wei et al. (2018) use Random Forest for their influential factor analysis step before generating emissions predictions through an extreme learning machine. In the context of corporate CO_2 emissions, Nguyen et al. (2021) apply a meta-learners over base-learners like Elastic Net for increased prediction accuracy.

Regarding the impact of electric or partially electric vehicles on emissions, Sioshansi and Denholm (2009) estimate that emissions in Texas may be reduced during certain months (May to September or "ozone season") through the use of plug-in electric vehicles, which in addition increase generator efficiency by selling energy directly back to the power grid. Rodrigues Teixeira and Sodré (2018) simulate emissions generated by a gasoline-powered vehicle and an EV and conclude that EV energy consumption produces 10 to 26 times less carbon than a gasoline-powered vehicle's direct emissions, even in the case of a grid powered by high-emissions resources.

For the scope of this paper, we focus only on CO_2 emitted as a result of energy generation and postulate that in California, where the EV population is greatest, vehicle distribution numbers may create a noticeable impact on electricity consumption and emissions. We employ machine learning techniques and explore stacked machine learning model applications.

2.2 Energy and emissions data

We use data from the California Independent System Operator (CAISO), a non-profit independent grid operator that manages about 80% California's electric power system. (California ISO, 2022a). It imports and exports energy through the Western Energy Imbalance Market (WEIM) which connects 19 different power management and generation systems across the western U.S. (Western Energy Imbalance Market, 2022) Note that in CAISO's emissions estimations, it includes emissions produced outside of the CAISO system for imported energy while excluding emissions produced within its system for energy that is eventually exported.

We mainly use the following dataset: **estimated CO_2 emissions resulting from energy production in metric tons of CO_2 per hour (mTCO$_2$/h)**, monthly.
CAISO's real-time estimated CO_2 emissions are calculated as the sum of estimated emissions generated from internal ISO dispatches and emissions from imports that serve ISO loads, subtracting emissions from exports. (Hundiwale, 2016) See Table 1.

Though not used in the final process, for initial analyses, we familiarize ourselves with the relations between CO_2 emissions mentioned above, energy generation, and energy demand collected from CAISO, all measurements taken at 5-minute intervals from April 2018 to November 2022: (California ISO, 2022b)
Electricity generated by resource in megawatts (MW): Resources include natural gas, large hydro (produced by facilities with a capacity greater than 30 MW), batteries, nuclear, coal, imports from other states, and other. Renewable resources include solar, wind, geothermal, biomass, biogas, and small hydro (produced by facilities with a capacity at 30 MW or less). These values are used in the below estimation of total CO_2 emissions.
Electricity demand in megawatts: The amount of energy needed on the grid managed by CAISO at a given moment.

$$em_{ISO} = em_{dispatch} + em_{import} - em_{export}$$

em_{ISO}	total emissions produced in order to generate the energy circulated by CAISO.
$em_{dispatch}$	total CO_2 emissions produced through dispatches of ISO internal resources, measured in the above mentioned dataset. For each resource i, we multiply the amount of electricity generated through this resource e_i (MWh), CO_2 emission factor for that resource em_i (mTCO$_2$/MMBTU), and the resource heat rate $heat_i$ (MMBTU/MWh): $em_{dispatch} = \sum_i e_i \times em_i \times heat_i$
em_{import}	emissions generated in the process of producing energy imported through the WEIM. This is approximated as the product of the amount of energy imported and the unspecified emission rate established by CARB of 0.428 mTCO2/MWh.
em_{export}	emissions generated in the process of producing energy within the CAISO system that is exported through the WEIM. Calculated in a similar manner to $em_{dispatch}$

Table 1: CO_2 emissions estimation

The U.S. Energy Information Administration (2022), under the greater U.S. Department of Energy, provides the following dataset: **hourly electricity demand in MWh.** We compress hourly load data to a monthly frequency by summing the values of each timeframe.

2.3 Vehicle data

This paper analyzes light-duty vehicles registrations rather than their sale numbers, the true target of the new ZEV policy, due to data availability and a more salient relation to the quantity of CO_2 ultimately emitted in the state. We look to the California Energy Commission (2022) for data on California's yearly registered vehicle numbers by different fuel types, enumerated in full in Table 2, over the 2010-2021 period.

We estimate monthly electric, plug-in hybrid electric, and gasoline hybrid vehicles registrations by referencing the number of each type of vehicle sold monthly on a national-scale taken from Argonne (2022). For the sake of approximation, we assume that, as California leads all other states in EV sales, state EV registrations grow proportionately to EV sales in the nation. For all other types of vehicles, we approximate monthly registration numbers assuming that each year that number grows linearly.

For context, in 2021, EVs numbered 522,445 or 1.745% of all light-duty vehicle registrations, a surprising number given California's leading position in national EV statistics. In contrast, the country with the world's the highest market penetration per capita, Norway, boasts 22% of all passenger cars being plug-in electric vehicles as of 2021, with plug-ins representing more than 85% of its new car sales. (Kane, 2022) We do note however that Norway is a significantly smaller country both population- and land-wise and likely enjoys a less car-dominated infrastructure than the U.S.

Gasoline	Standard combustion engine vehicle	86.83%
Gasoline hybrid	Fueled by gasoline and electricity generated while driving, e.g., conventional Toyota Prius	4.34%
Flex fuel	Fueled by gasoline, ethanol, or a mix of both	4.04%
Diesel	Fueled by diesel, often a heavy combustion engine vehicle	1.97%
Electric **	All-electric vehicle, e.g., Tesla models	1.74%
Plug-in hybrid *	Fueled by both gasoline and electricity	1.02%
Fuel cell *	Fueled by electricity generated from a fuel cell of compressed hydrogen, e.g., Hyundai NEXO	0.03%
Natural gas	Fueled by compressed or liquefied natural gas	0.03%
Propane	Fueled by liquified petroleum gas, often used in fleet applications, e.g., school buses and police vehicles	0.00%

* Zero-emissions vehicle

* Uses electricity from being plugged in

Table 2: Vehicles by fuel type and the proportion of vehicle registrations in California that they represent in 2021

2.4 Weather data

To elucidate the relationship between electricity-powered vehicles and electricity demand, we take into account weather, known to be a strong predictor for electricity demand, (Taylor and Buizza, 2003; Hor et al., 2005) specifically looking at **monthly maximum temperature in degrees Fahrenheit** and **precipitation in inches** by county of California, (National Centers for Environmental Information, 2022)
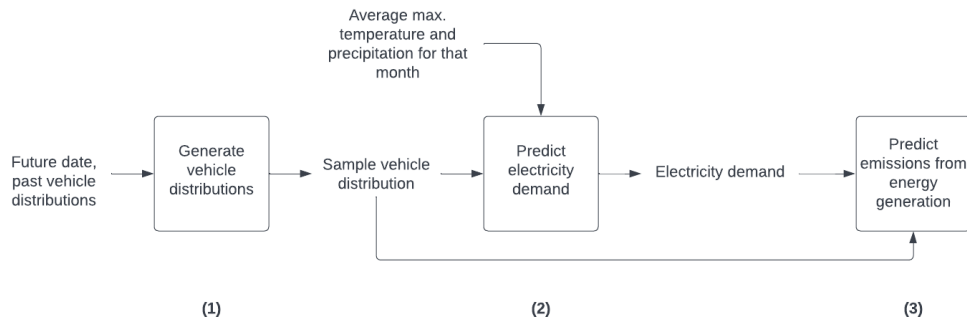
3 Methodology

3.1 Process

Figure 1: Three-step process for generating an estimate for a given month's CO2 emissions from energy generation.

3.2 Models

We apply stacking regressor methods at steps **(2)** and **(3)** illustrated in Figure 1, ensemble techniques using multiple models for more accurate results and to mitigate single model short-
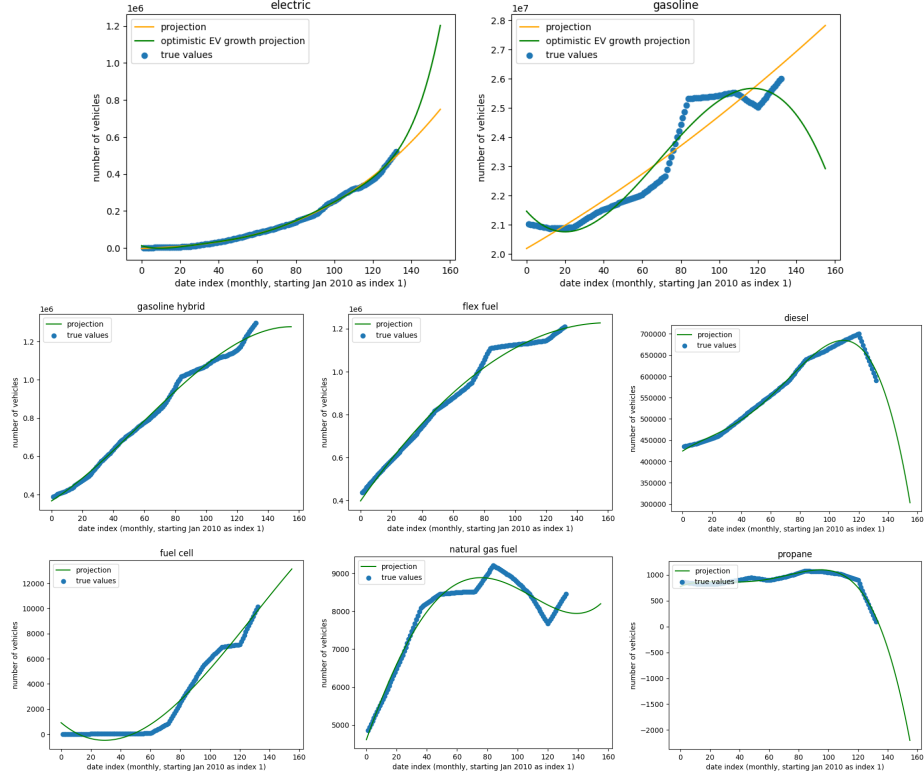
Figure 2: Chosen growth trends for each type of vehicle.

comings. Specifically we try stacking: lasso and Random Forest, linear regression and Random forest, lasso and a gradient boosted regressor, and linear regression and a gradient boosted regressor for both demand prediction and emissions prediction.

The Random Forest regression algorithm (RF) and gradient boosting regression (GB) themselves already involve ensemble learning, for the first by combining regression trees, where each represents a hierarchy of conditions from root to leaves, (Breiman, 2001) and the second through application of an ensemble of decision trees as a weak learner. Though efficient over large datasets with many features, given their structures as combinations of decision trees, both RF and GB fare poorly with inputs outside of a previously seen range. Additionally, because we must work at the monthly level, our dataset is small, indicating a risk of overfitting. In order to address this, we attempt stacking these models with linear models to "remember" projections outside of the input range. We use linear regression and the lasso (Least Absolute Shrinkage and Selection Operator) regression model, which is also linear.

3.3   Generating vehicle distribution samples

In order to pass realistic test vehicle distribution data and demand values into our models, for each vehicle type, we project into the future the number of vehicles over time given current growth trends through the application of polynomial regression.

For all vehicle types, we assume a slow growth, aside from electric and gasoline vehicles, for which we generate different distributions using more or less optimistic growth rates for each, illustrated in figure 2.

4 Results

Assuming that the methods and resources used to generate energy in California remain relatively the same, we predict monthly emissions from energy generation for August of 2026 for two different vehicle distribution predictions, the first with neutral gasoline and electric vehicle registration growths, the second with an optimistic EV registration growth rate and a slow declining gasoline-vehicle growth rate.

4.1 Stacked regressor performances

We use a train-test split of 8-2, with random search over hyperparameters, using R-2 scoring for training and testing scores.

Model	Train score	Test score
RF	0.93	0.83
Linear regression & RF	0.92	0.93
Lasso & RF	0.95	0.95
GB	0.99	0.84
Linear regression & GB	0.98	0.98
Lasso & GB	0.98	0.96

Table 3: Accuracy scoring for demand prediction using electric vehicle, plug-in hybrid electric, maximum temperature, and average precipitation data.

Scoring is roughly comparable for all methods, with slightly increased score for stacking linear regression and GB for demand prediction. We use this method in the below results.

Model	Train score	Test score
RF	0.83	0.64
Linear regression & RF	0.69	0.75
Lasso & RF	0.79	0.79
GB	0.91	0.71
Linear regression & GB	0.82	0.81
Lasso & GB	0.93	0.89

Table 4: Accuracy scoring for emissions prediction using vehicle distribution and electricity demand data.

Stacking lasso and GB performs best among tested models for predicting emissions. We use this method in the below results.

4.2 Neutral projected vehicle growths

We generate the following vehicle distribution:

vehicle type	projected number of registrations	% of all registrations
electric	699,835	2.22%
plug-in hybrid electric	312,033	0.99%
fuel cell	12,573	0.04%
diesel	379,989	1.21%
flex fuel	1,224,642	3.89%
gasoline	27,575,770	87.57%
gasoline hybrid	1,276,945	4.06%
natural gas	8,080	0.03%
propane	0	0%

Table 5: Projected number of registrations by vehicle type for August 2026

This would be an increase of about 51% (237,584 EVs) from August of 2021.

Using these projections, the average August maximum temperature, of 88° F, and average August precipitation of 0.1 inches, we estimate a demand of 27,869,201.23 MWh for the month of August 2026, compared to 27,525,448 MWh from August 2021.

In turn, we project for that month 5,624,180.8 mTCO2 emissions from energy generation, compared to the August 2021 quantity of 5,607,233.68 mTCO2.

4.3 Optimistic electric vehicle growths

For the same date, using a more optimistic EV registration growth rate and an eventually declining gasoline-powered vehicle registration rate, we generate the following vehicle distribution:

vehicle type	projected number of registrations	% of all registrations
electric	1,009,995	3.64%
plug-in hybrid electric	312,033	1.12%
fuel cell	12,573	0.05%
diesel	379,989	1.37%
flex fuel	1,224,642	4.41%
gasoline	23,517,676	84.77%
gasoline hybrid	1,276,945	4.6%
natural gas	8,080	0.03%
propane	0	0%

Table 6: Projected number of registrations by vehicle type for August 2026 with optimistic EV growth

This would be an increase of about 118% (547,744 EVs) from August of 2021 and a decline of about 8% for gasoline-fueled vehicles (2,155,196 cars).

Using these projections and the same average August maximum temperature and precipitation values as previously, we estimate a demand of 28,027,557.14 MWh for the month of August 2026, compared to 27,525,448 MWh from August 2021.

In turn, we project for that month 5,640,539.45 mTCO2 emissions, compared to the August 2021 quantity of 5,607,233.68 mTCO2.

4.4 Comparison

In the case of Section 4.2, which predicts an increase in EV registrations by 237,584 between August 2021 and August 2026, California would experience 16,947.12 mTCO2 greater emissions that month than in August of 2021.

EV registration increase	emissions increase	(1)	(2)
237,584	16,947.12	98,281.92	38,013.44
547,744	33,305.77	208,142.72	87,639.04

Table 7: Projected EV registration increases and monthly emissions increases based on model results, compared to **(1)** estimated emissions from the same number of gasoline vehicles and **(2)** estimated emissions generated from EV electricity usage using EPA averages.

For Section 4.3, projections indicate an increase in EV registrations by 547,744 in that timeframe and 16,947.12 mTCO2 more emissions than month than in August of 2021.

In table 7, we compare these values with the estimated emissions from the same number of combustion engine cars in a month, assuming that a typical passenger car emits 4.6 mTCO2 a year, or roughly 0.38 $mTCO_2$ per month. (U.S. Environmental Protection Agency, 2022a)

Additionally, as we cannot verify emissions projections, we look at another method of estimation. According to one source, (Marsh, 2022) the average EV consumes 406.5 kWh per month, and the U.S. Environmental Protection Agency (2022b) calculates an average of 884.2 lbs CO_2 per MWh consumed in the United States—we assume roughly 0.16 $mTCO_2$ per month per EV. We calculate in the fourth column of table 7, based on these values, the estimated quantity of emissions produced by the number of additional EV registrations.

In both cases, the emissions due to increased electricity demand represent fractions of those generated by the typical gasoline-powered vehicle. We also notice major discrepancies between this project's estimated emissions and those calculated as the product of the number of EVs and average CO_2 emissions generated per MWh according to the EPA. We may attribute this to two factors: our modeling approach is simply not a good fit and the fact that the EPA average takes into account the entire U.S., where each region releases carbon emissions during energy generation at different rates.

5 Conclusions

5.1 Reflection

I hoped in my original proposal and approach to develop a more accurate, or at least informed, method of approximating energy consumption as electricity-powered vehicles increase than simply multiplying averages as we did in section 4.4. Ultimately, however, this technique may suffice for ballpark approximations of potential emissions and is more understandable.

One main concern regarding the robustness of my approach was the tenuous relation between vehicle distribution numbers and energy demand based on current data, considering the relatively small number of electric and partially electric vehicles in California, which surprised me. I tried to mitigate this with other highly correlated at such as weather, but I fear my understanding of contributors to electricity demand remains shallow and the data too few.

Indeed, the second concern would be data, as I was warned, particularly for vehicle registration, since energy and motor vehicle departments usually only release this data on a yearly basis, which constrains the applicability of machine learning approaches. I extrapolated monthly information, but we would be hard-strapped to find or measure this data on a finer level, such as weekly or daily.

This project could certainly be refined with more specificity for other data like county-level electricity demand and generation and a better understanding of the car market and sales trends. (I do not, for example, have a good sense of what sorts of vehicles use propane and if their number will continue to dwindle or if there exists a purpose for propane-fueled vehicles that precludes their disappearance.)

Though all thrown together and glossed-over in this single paper, each step in this process merits its own in-depth study: vehicle distribution prediction, electricity demand prediction, and emissions from energy generation.

I believe I got lost in the sauce.

5.2 Final thoughts

Though the data and approach used cast a pall of doubt on the real merit of the models applied and certainly on their results, they do seem to reaffirm that emissions engendered by increased energy generation for powering EVs represent a fraction of what would be released directly into the air by an equivalent number of traditional combustion engine vehicles. Not only does this reduce total state emissions in California, but it also reduces air pollution and associated health problems in areas with urban areas with high car densities.

Though the energy generation process will continue to shoulder more of the total carbon emissions for the state, there remains in it great room for improvement, notably through the integration of renewable energies and energy grid connections to more efficiently meet real-time needs with less energy waste, such as through the TransWest Express system, which seeks to bring wind-generated energy from Wyoming to California. (Transwest Express LLC, 2022)

Acknowledgment

References

Argonne (2022). Monthly sales data for electric vehicles. https://www.anl.gov/esia/light-duty-electric-drive-vehicles-monthly-sales-updates.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

California Energy Commission (2022). Light-duty vehicle population in california. https://www.energy.ca.gov/data-reports/energy-almanac/zero-emission-vehicle-and-infrastructure-statistics/light-duty-vehicle.

California ISO (2022a). The ISO grid. http://www.caiso.com/about/Pages/OurBusiness/The-ISO-grid.aspx.

California ISO (2022b). Today's outlook. https://www.caiso.com/TodaysOutlook/Pages/index.html.

Ehteram, M., Sammen, S. S., Fatemeh, P., and Sidek, L. M. (2021). A hybrid novel svm model for predicting co_2 emissions using multiobjective seagull optimization. *Environmental Science and Pollution Research*, 28(46):66171–66192.

Hor, C.-L., Watson, S. J., and Majithia, S. (2005). Analyzing the impact of weather variables on monthly electricity demand. *EEE Transactions on Power Systems*, 20(4):2078–2085.

Hundiwale, A. (2016). Greenhouse gas emission tracking methodology. Technical report, California ISO.

Kane, M. (2022). Norway sets plug-in car sales record for the end of the year 2021. https://insideevs.com/news/558447/norway-plugin-car-sales-december2021/.

Marsh, J. (2022). How many watts does an electric car charger use? https://news.energysage.com/how-many-watts-does-an-electric-car-charger-use/.

National Centers for Environmental Information (2022). County mapping. https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/county/mapping/4/pcp/201902/1/value.

Nguyen, Q., Diaz-Rainey, I., and Kuruppuarachchi, D. (2021). Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. *Energy Economics*, 95:105129.

Office of Governor Gavin Newsom (2022a). California enacts world-leading plan to achieve 100 percent zero-emission vehicles by 2035, cut pollution.

Office of Governor Gavin Newsom (2022b). California leads the nation's zev market, surpassing 1 million electric vehicles sold.

Office of Governor Gavin Newsom (2022c). California zev sales near 18% of all new car sales in 2022.

Rodrigues Teixeira, A. C. and Sodré, J. R. (2018). Impacts of replacement of engine powered vehicles by electric vehicles on energy consumption and co2 emissions. *Environmental Science Technology*, 59:375–384.

Saleh, C., Dzakiyullah, N. R., and Nugroho, J. B. (2016). Carbon dioxide emissions prediction using support vector marchine. *IOP Conf. series: Materials Science and Engineering*, 114(012148).

Sioshansi, R. and Denholm, P. (2009). Emissions impacts and benefits of plug-in hybrid electric vehicles and vehicle-to-grid services. *Environmental Science Technology*, 43:28985–28997.

Taylor, J. W. and Buizza, R. (2003). Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, 19(1):57–70.

Transwest Express LLC (2022). Critical grid infrastructure to connect the west. https://www.transwestexpress.net.

U.S. Energy Information Administration (2022). California electricity data and map. https://www.eia.gov/beta/states/states/ca/data/dashboard/electricity.

U.S. Environmental Protection Agency (2022a). Greenhouse gas emissions from a typical passenger vehicle. https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle.

U.S. Environmental Protection Agency (2022b). Greenhouse gases equivalencies calculator - calculations and references. https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references.

Wei, S., Wang, Y., and Zhang, C. (2018). Forecasting co2 emissions in hebei, china, through moth-flame optimization based on the random forest and extreme learning machine. *Environmental Science and Pollution Research*, 25:28985–28997.

Western Energy Imbalance Market (2022). About. https://www.westerneim.com/Pages/About/default.aspx.