Natural Resources Canada Ressources naturelles Canada

Comparing machine learning and linear regression methods for estimating marginal greenhouse gas emission factors of electricity generation with renewables

Prepared by:

Marie Pied, Sophie Pelland, Steven Wong, Dave Turcotte and Vahid Raissi Dehkordi

Project Manager:

Véronique Delisle

Natural Resources Canada (NRCan)

CanmetENERGY – Varennes Research Center

1615 Lionel-Boulet Blvd.

Varennes, Québec J3X 1P7

Date:

November 30, 2021

Canada

Disclaimer

Natural Resources Canada (NRCan) is not responsible for the accuracy or completeness of the information contained in the reproduced material. NRCan shall at all times be indemnified and held harmless against any and all claims whatsoever arising out of negligence or other fault in the use of the information contained in this publication or product.

DOI

The digital object identifier (DOI) for this publication is: https://doi.org/10.4095/329257

Acknowledgements

CanmetENERGY is a federal research laboratory in Canada; financial support for this research work was provided by Natural Resources Canada (NRCan) through the Program on Energy Research and Development (PERD) in the framework of REN-4 Transition to High Renewable Integration on the Grid project. The authors would like to thank Armin Salimi for help in designing the test protocol that was used to compare different methods of estimating marginal emission factors. We also thank Véronique Delisle for project management.

Executive Summary

The electricity sector is key to reducing greenhouse gas emissions: energy supply across sectors is shifting towards electricity, and electricity production in turn is shifting towards low carbon sources. Marginal greenhouse gas emission factors (MEFs) are needed to properly quantify the impact on greenhouse gas emissions of policies and programs that modify electricity demand. We compared different methods of estimating MEFs applied to the Canadian provinces of Ontario and Alberta: a new method using artificial neural networks and existing methods using simple and multiple linear regression. Models were developed using hourly data from the independent electricity system operators of each province for 2017, 2018 and 2019. Models were trained on one year and tested on the following year with respect to bias and root mean square error. Overall, multiple linear regression performed over a full year was the most reliable model: it leads to consistently low bias and can be used to calculate MEFs by hour or by month. While Alberta and Ontario do not provide historical MEFs, indirect comparisons suggest our MEFs were within about 4% and 13% of actual MEFs in Ontario and Alberta, respectively. Finally, we used the Ontario case to explore a rule for determining when wind generators can be marginal. This question has been neglected so far and is becoming important with the ramp-up of renewables worldwide.

Contents

List of Figures

List of Tables

Abbreviations

AEF:	Average Emission Factor
MEF:	Marginal Emission Factor
IESO:	Independent Electricity System Operator (Ontario)
AESO:	Alberta Electricity System Operator
EV :	Electric Vehicle

Chapter 1

Introduction

An important challenge in the fight against climate change is the expected surge in electricity demand with the electrification of transportation and heat, and the growing use of digitally connected devices and air conditioning. Before implementing demand response or electric vehicle (EV) deployment programs, the associated changes in greenhouse gas (GHG) emissions should be quantified. Depending on which electricity generator—wind or gas, for example—responds to a change in load from a given program, the associated impact on emissions can vary significantly.

The generator responding to an incremental change in electricity demand is referred to as the marginal generator, the determination of which is not easy since electricity cannot be traced from a specific generator to a specific consumer. A common practice for estimating emissions linked to the production of electricity is to use average emission factors (AEFs), the total emissions from supplying electricity divided by the total amount of electricity supplied. AEFs are relatively easy to compute but do not reflect the emissions linked to an incremental load change, since a marginal generator's emissions can differ substantially from the AEF. To quantify GHG impacts from programs that impact load, marginal emission factors (MEFs) should be used. MEFs are the incremental changes in GHG emissions divided by the incremental changes in load.

Unfortunately, electricity system operators do not typically provide information about MEFs. A wide range of methods have therefore been developed for estimating MEFs (see e.g. [6] for a review and comparison). There are two broad classes of methods: 1) methods based on historical data and 2) power system optimization methods. While power system optimization methods can be more flexible and powerful than methods based on historical data, they require detailed data concerning grid and generator geometry, characteristics and dispatch mechanisms. In this study, we focus on methods based on historical data that are publicly available. Such methods range from simple linear regressions to more complex approaches using machine learning.

Hawkes [7] uses a simple linear regression between two vectors that represent the changes (between hours t and $t-1$) in the total load of the system and the changes in the total GHG emissions of the system. Siler-Evans et al. [8] extend this method by looking at changes in generation for each generator type vs. change in total generation to obtain the fraction of time that each generator type is marginal. More recently, Gai et al. [4] use multiple linear regression to estimate MEFs in Ontario based on the magnitude of changes in generation and total generation. These authors omit generation from variable renewable sources (such as wind and solar), assuming they are not dispatchable.

There are just a few studies that consider more complex machine learning approaches to estimate MEFs. Tranberg et al. [9] apply a flow tracing algorithm using market, meteorological, generation and consumption data. A support vector machine is used in [10] to estimate MEFs from publicly available data, such as LMP or load data.

In this article, we present and compare estimations of MEFs using simple linear regression, multiple linear regression and artificial neural networks (ANNs) based on historical data for the Canadian provinces of Alberta and Ontario. The multiple linear regression approach is based on the work of Gai et al. [4] while the ANN model is novel. Our work addresses two significant gaps in the literature: First, there have been no comparisons of machine learning approaches and approaches based on linear regression in the literature to date, making it difficult to know which is most appropriate in a given context. Our work evaluates the two types of approaches based on their predictive ability; the models are trained on data from one year and tested on their ability to predict the MEFs of the following year. Second, the literature to date has not developed methods to determine if and when variable renewables (such as solar and wind) can be marginal. This is becoming increasingly relevant as the renewable share of generation steadily increases. In particular, when variable renewables are curtailed, they can respond to increments in load by adjusting their production not only down but also up. In our analysis, we develop a rule for determining when wind energy generators in Ontario are curtailed, and thus can contribute to the marginal generation pool.

The article is structured as follows: In Section 2.1, we describe the data and the filtering process used to remove generators that cannot be marginal from the analysis. Then, in Section 2.2, we describe the different methods, the different partitions of the data, and our performance test. In Sections 3.1 and 3.2, we compare the performance of the different methods and contrast our MEF results with those from the independent system operators (ISOs) of Alberta and Ontario and with results from literature. Finally, in Chapter 4 we offer some concluding remarks and discuss the policy relevance of this work.

Chapter 2

Data and Methodology

2.1 Data

2.1.1 Bulk electric system data

The data used in this study, described below, were obtained or derived from data provided by the ISOs of Ontario ([11]) and Alberta ([12]) for the years 2017, 2018 and 2019.

Generator characteristics and output:

- Generator type, encompassing the fuel used (coal, gas, nuclear, hydro, wind, solar, biofuel, or wasteheat) and, for gas generators, the technology (combined cycle (CC), simple cycle (CT), or steam).

- Generator CO_2 emissions factor (EF), noting that only CO_2 emissions from fuel combustion, i.e. gas and coal, were counted. (Emissions from biofuel were omitted since biofuel generators were never marginal in our analysis, as described in Section 2.1.2).

  The heat rate curves of coal and natural gas combined cycle and steam generators were estimated by scaling their respective curves in [13] such that their full-load heat rate values matched those in [14]. The resulting curves were then fitted to quadratic functions whose coefficients are given in Table 2.2. For natural gas simple cycle generators, a flat (constant) heat rate was used, corresponding to their full-load heat rate in [14].

The CO_2 EF of each generator type, in kg/MWh, is given by its heat rate multiplied by its fuel CO_2 content:

$$EF(x) \quad = \quad (\alpha_0 + \alpha_1 x + \alpha_2 x^2) \times CO_2Content \tag{2.1}$$

where $x = output/capacity$. CO_2 content and parameters α_0, α_1 and α_2 are given in Tables 2.1 and 2.2, respectively. The resulting EFs are shown in Figure 2.1.

Fuel	CO_2 Content (kg/MJ)
Gas	0.0503
Coal	0.0882

Table 2.1: CO_2 content of different fuels [5]

Fuel	Technology	α_0	α_1	α_2
Gas	CT	12206	0	0
Gas	CC	12616	-11058	6493
Gas	STEAM	13585	-5132	2570
Coal		12388	-3199	1522

Table 2.2: Heat rate function coefficients (in MJ/MWh)

- Generator production for each hour of the year, extracted from Ontario's "Output and capability report" [11] and Alberta's "Hourly Metered Volumes and Pool Price and AIL data" [12].

Intertie characteristics and flows:

- Intertie GHG EFs associated with electricity imported through provincial interties, extracted from [15] for Canadian provinces and [16] for U.S. states, as summarized in Table 2.3.
- Intertie imports and exports, from [11] for Ontario and [12] for Alberta.

Electricity market prices:

- Average hourly marginal prices, from [11] for Ontario and [12] for Alberta.

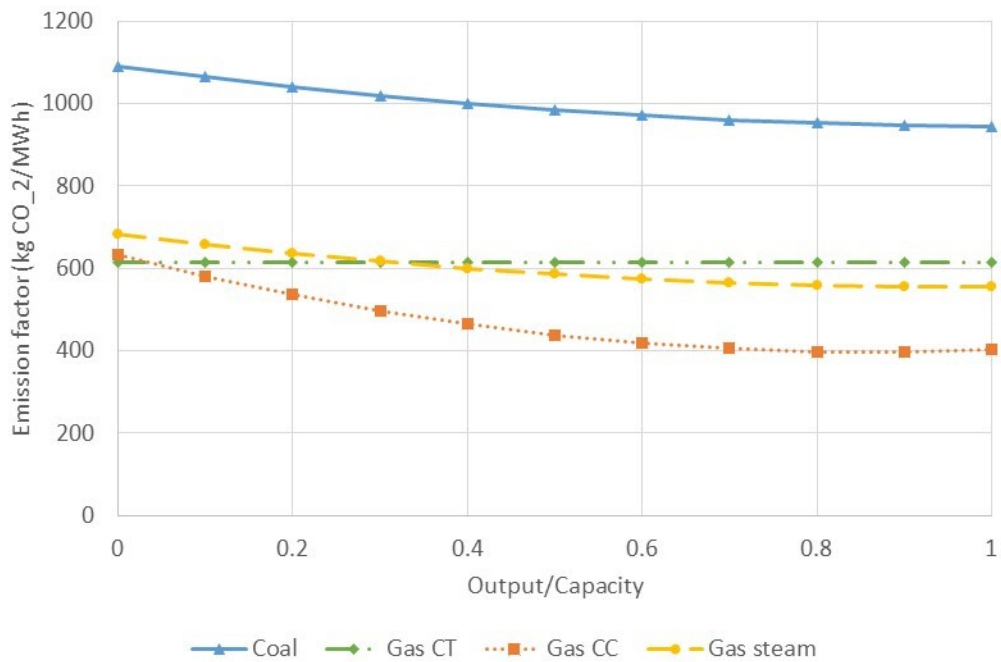


Figure 2.1: Emission factor curves

Intertie	Emission factor (kg CO_2/MWh)
Québec	1.5
Manitoba	2.1
Michigan	525.15
Minnesota	479.7
New York	208.8
British Columbia	9.1
Saskatchewan	710
Montana	548.1

Table 2.3: Intertie emission factors

2.1.2 Output data filtering process

Before estimating MEFs, we filtered hourly output and intertie level data to detect and remove generators that we knew could not be marginal, as they would otherwise add noise to the MEF estimates. This type of filtering has been performed previously in analyses of MEFs for Ontario by Farhat and Ugursal [17] and The Atmospheric Fund [3]. First, generators for which their hourly change in generation was not of the same sign as the change in total generation were removed, since by definition these cannot be marginal (they are not responding to the increment in load). To simplify our analysis, we also excluded generators of types that have been marginal less than 1% of the time, based on data from IESO and AESO as illustrated in Figure 2.2 and Figure 2.3. Based on this process, wasteheat, biofuel and solar generators were filtered out for both provinces and wind generators were filtered out for Alberta.

The treatment of wind generation deserves additional consideration. Wind is a zero or, with financial incentive such as feed-in-tariff, a negative marginal cost generator. Ignoring transmission congestion and other system operational constraints such as stability (complex to consider and beyond the scope of this paper), they will operate at max capacity (as wind resource allows) when market prices are above zero, as it is economically in their best interest. Only when prices fall to zero or below may they constrain their output and be considered the marginal generator.

In Alberta, low/zero marginal cost generators (hydro, wind, solar) make up only a small fraction of generation capacity and thus they have little opportunity to set marginal prices; thus prices are rarely at or near zero and wind is almost never marginal. On the other hand, Ontario has a large fleet of zero/near-zero/negative marginal cost generators; wind (with already a large presence) shares this space with hydro, nuclear, and solar units. There are enough of these generators to provide the capacity Ontario needs during low-demand periods, thus driving market prices to zero or below and giving wind the opportunity to be on the margin, as shown in Figure 2.2.

Given that significant wind curtailment generally occurs only when the marginal price is less than or equal to zero, as seen in Figure 2.4 and discussed in [18] and [19] (p. 81), we set a rule such that a wind generator can be considered marginal only if the average hourly marginal price is zero or negative.

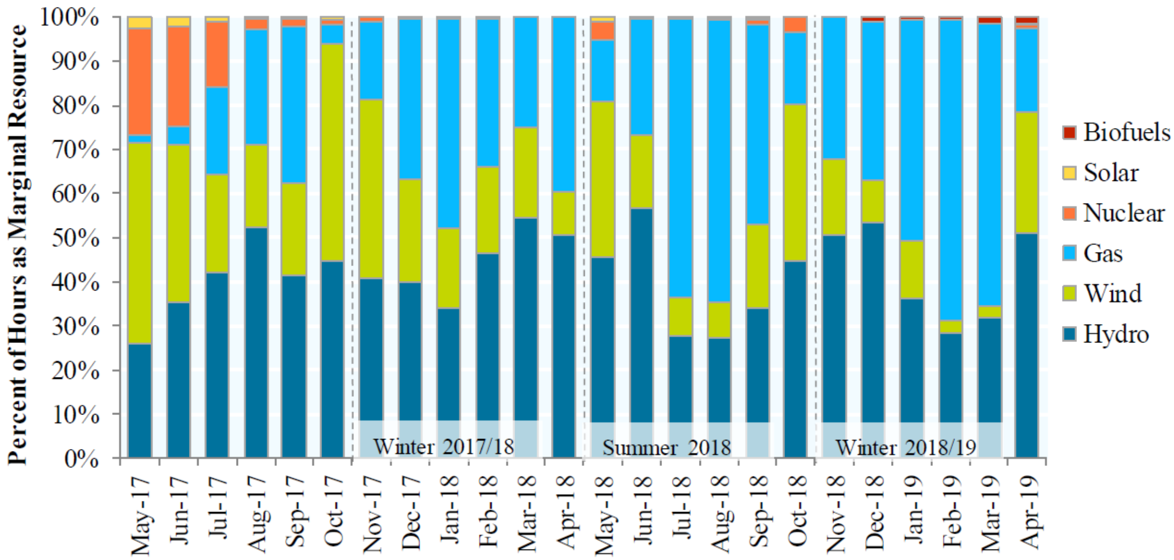


Figure 2.2: Share of generator type setting the market clearing price from May 2017 to April 2019 in Ontario [1].

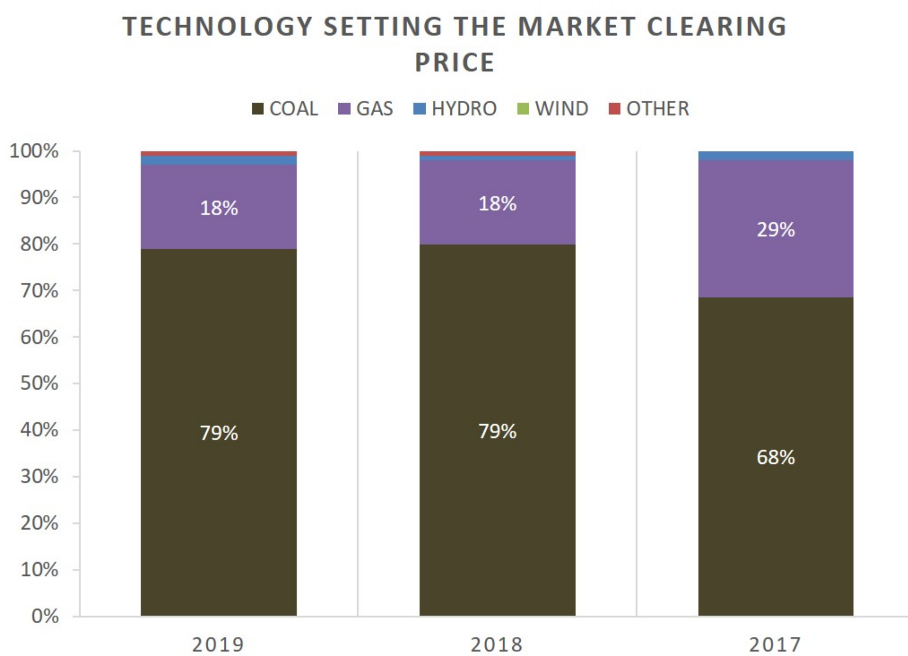


Figure 2.3: AESO 2019 Annual Market Statistics report, page 15 [2]

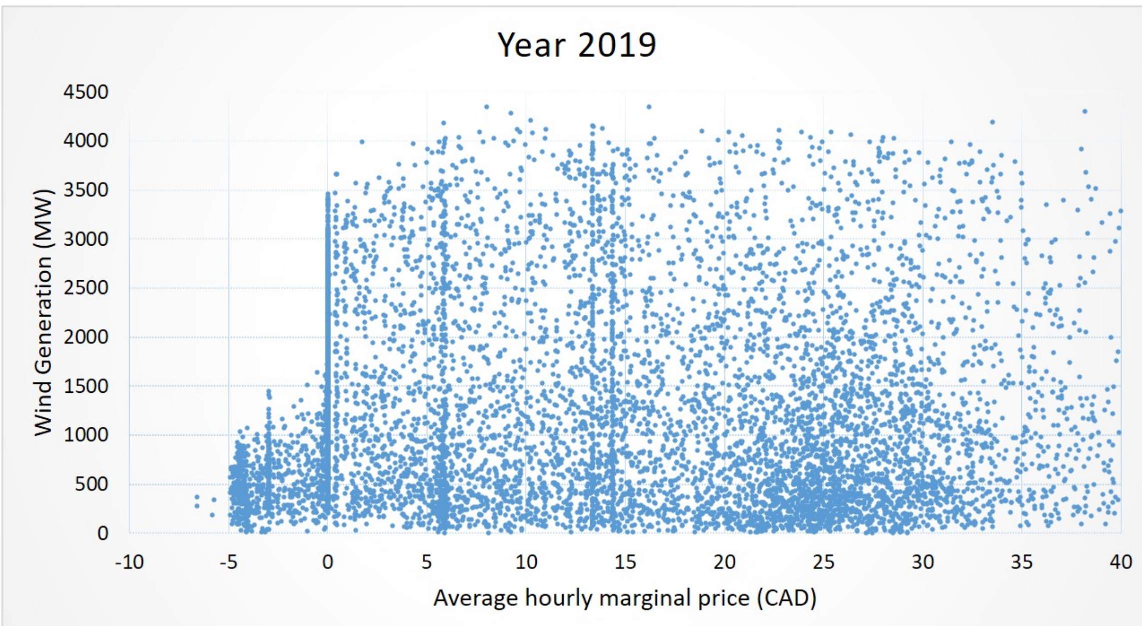


Figure 2.4: Wind generation in MW versus average hourly marginal price in Ontario

2.2 Regression models

2.2.1 Variables used to perform the study

We used a number of variables to estimate MEFs and the share of marginal generation for each generator type. G_t is the total generation in the province at time t (including imports, if part of the analysis) and ΔG_{t-y} is its change over the previous y hours. E_t is the total CO_2 emissions from electricity production at time t and ΔE_{t-y} is its change over the previous y hours. E_t is computed according to (2.2), where $G_{g,t}$ is the output level of the generator g at time t, $G_{i,t}$ is the import level of intertie i at time t, $EF_{g,t}$ is defined in equation (2.1), and AEF_i is the emission factor of imports from intertie i as given in Table 2.3.

$$E_t = \sum_{g \in Generators} EF_{g,t} \times G_{g,t} + \sum_{i \in Interties} AEF_i \times G_{i,t} \tag{2.2}$$

Note that our calculations estimate MEFs associated with electricity generation rather than electricity consumption. To estimate MEFs per MWh consumed, transmission and distribution losses should be included.

To compute shares of marginal generation, we used the variable G_k for generation and ΔG_k for its change over the previous hour for generators of type k.

2.2.2 Different partitions of the data set

We applied regressions to different partitions of our dataset, inspired by [4]. The regression classes are summarized in Table 2.4, where RegLin stands for simple linear regression, MultiRegLin stands for multiple linear regression and ANN designates an artificial neural network approach (explained in Section 2.2.4).

The following partitions were used:

1. Partition 1: by year

2. Partition 2: by year and month (12 data subsets per year)

3. Partition 3: by year and hour of day (24 data subsets per year)

ID	Type	Data partition
ANN	ANN	Full Year
LR1	RegLin	Full Year
LR2	RegLin	Monthly
LR3	RegLin	Hourly
LR4	RegLin	GenBins
LR5	RegLin	Hour,Month
LR6	RegLin	GenBins,Month
MLR1	MultiRegLin	Full Year
MLR2	MultiRegLin	Monthly
MLR3	MultiRegLin	Hourly
MLR4	MultiRegLin	GenBins
MLR5	MultiRegLin	Hour,Month
MLR6	MultiRegLin	GenBins,Month

Table 2.4: Different classes of regressions used and associated partitioning of the data

4. Partition 4: by year and level of generation. For each year, we computed the maximum and minimum generation levels and sorted the data into 20 bins of increasing generation levels with equal numbers of points per bin, as in Figure 2.5.

5. Partition 5: by year, month and hour of day ($12\times24=288$ data subsets per year)

6. Partition 6: by year, month and level of generation ($12\times20=240$ data subsets per year, corresponding to 20 generation level bins as in Figure 2.5).

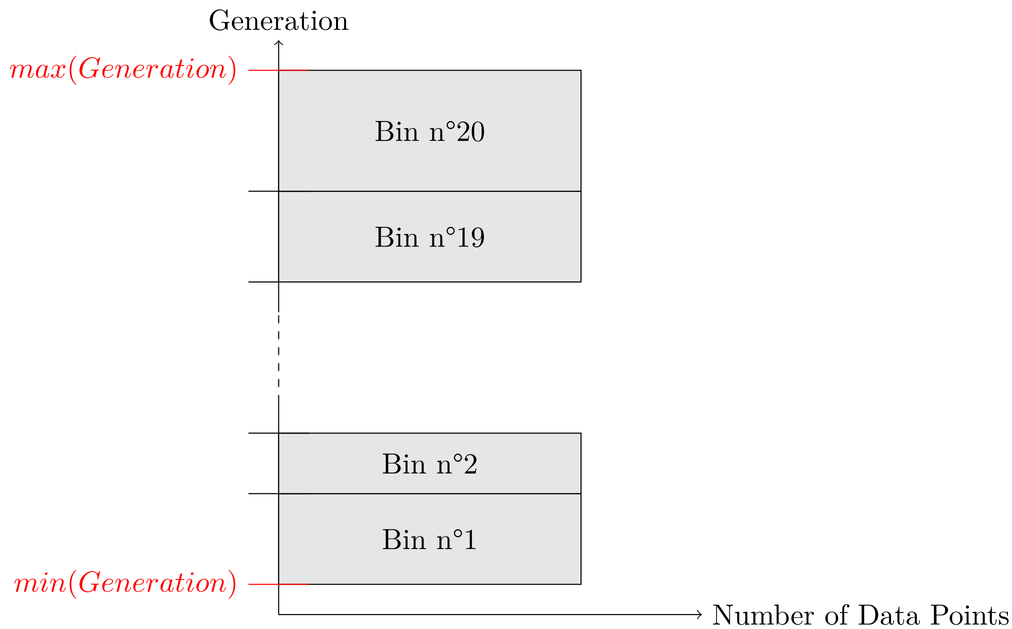


Figure 2.5: Diagram of the disaggregation into bins with increasing generation levels, where each bin contains the same number of data points

2.2.3 Regression models

We used various regression models to estimate MEFs: simple and multiple linear regressions are described in this section, and ANNs in the next. We also computed the fraction of time, γ_k, for which generators of type k operated as marginal by performing regressions predicting the change in generation of type k generators from the total change in generation. The following equations were used to estimate MEFs and γ_k:

Simple linear regression:

$$\Delta E = MEF \times \Delta G + \beta_0 \tag{2.3a}$$

$$\Delta G_k = \gamma_k \times \Delta G + \epsilon_{0,k} \tag{2.3b}$$

Multiple linear regression :

$$\Delta E = \overbrace{(\beta_1 + \beta_2 G + \beta_3 \Delta G)}^{MEF} \times \Delta G + \beta_0 \tag{2.4a}$$

$$\Delta G_k = \overbrace{(\epsilon_{1,k} + \epsilon_{2,k} G + \epsilon_{3,k} \Delta G)}^{\gamma_k} \times \Delta G + \epsilon_{0,k} \tag{2.4b}$$

The expression for MEFs in the multiple linear regression was developed by Gai et al. [4] in their study of MEFs in Ontario. The only difference is the inclusion of the constant term β_0.

2.2.4 Artificial neural network regression

Regressions for predicting ΔE were also developed using ANNs to test whether using additional predictors or non-linear regressions could improve model performance. The ANN regression problem was set up to predict $\Delta E_{t+1} = E_{t+1} - E_t$ given $\Delta G_{t+1} = G_{t+1} - G_t$ and variables available at times up to and including t. Note that ΔG_{t+1} is allowed since, by definition, we are interested in predicting the change in emissions for a given change in generation. We considered all province-wide variables from t to $t-2$:

- $\Delta G_{t+1}, \Delta G_t, \Delta G_{t-1}$
- G_t, G_{t-1}, G_{t-2}
- $\Delta G_{k,t}, \Delta G_{k,t-1}, \Delta G_{k,t-2}$
- $G_{k,t}, G_{k,t-1}, G_{k,t-2}$
- $Price_t, Price_{t-1}, Price_{t-2}$
- E_t, E_{t-1}, E_{t-2}
- $\Delta E_t, \Delta E_{t-1}, \Delta E_{t-2}$
- $TimeOfDay_t, TimeOfDay_{t-1}, TimeOfDay_{t-2}$
- $TimeOfYear_t, TimeOfYear_{t-1}, TimeOfYear_{t-2}$

$Price$ is the average marginal hourly price and $TimeOfDay$ and $TimeOfYear$ are proxies for the hour of day and month and take into account the periodic nature of these variables. They are defined as

$$TimeOfDay = \cos\left(\frac{2\pi(mod(Hour, 24) + 1)}{24}\right) \tag{2.5a}$$

$$TimeOfYear = \cos\left(\frac{2\pi Hour}{8760}\right), \tag{2.5b}$$

where $Hour$ runs from 1 to 8760 for each year.

The Pearson correlation coefficients between each variable and the output variable ΔE_{t+1} were computed for 2017 and used to select input features to the ANN regressions. Table 2.5 shows the features that were selected for each province, namely those features with Pearson correlation coefficients of 0.3 or more for Ontario and 0.5 or more for Alberta. Note that the correlations between ΔE_{t+1} and ΔG_{t+1} are much stronger in Alberta than in Ontario.

Ontario		Alberta	
Feature	Correlation coefficient	Feature	Correlation coefficient
ΔG_{t+1}	0.6	ΔG_{t+1}	0.98
$\Delta G_{gas,t}$	0.5	ΔG_t	0.6
ΔE_t	0.5	ΔE_t	0.6
ΔG_t	0.4	$\Delta G_{coal,t}$	0.5
$TimeOfDay_t$	0.3	$TimeOfDay_t$	0.5
E_{t-2}	0.3		
$\Delta G_{gas,t-1}$	0.3		
$G_{gas,t-2}$	0.3		
ΔE_{t-1}	0.3		
$\Delta G_{hydro,t}$	0.3		
$G_{hydro,t-2}$	0.3		
$TimeOfDay_{t-1}$	0.3		
G_{t-2}	0.3		

Table 2.5: Pearson correlation coefficients with ΔE_{t+1} in Ontario and Alberta

The `train` function in MATLAB R2019a was used to train the networks. Default parameters were used, including assignment of 70%, 15% and 15% of the data randomly to training, validation and test subsets, respectively. Single hidden layer networks were considered and the number of neurons in the hidden layer was optimized to yield the best performance (lowest mean square error) on the test subset of the training data.

2.2.5 Model testing protocol

We adopted the following testing protocol to evaluate the performance of all models in deriving MEFs:

- Step 1: Train the model with data from year y to predict hourly changes in emissions (ΔE) for each hour from the model features (inputs).
- Step 2: Predict ΔE values for year $y+1$ with the model trained in Step 1.
- Step 3: Compare this prediction with the actual ΔE values of year $y+1$.

Two train-test pairs were used: 1) Models were trained on 2017 data and tested on 2018 data and 2) Models were trained on 2018 data and tested on 2019 data. We used two metrics to evaluate performance: root mean square error (RMSE) and mean bias error (Bias). These two metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\Delta E_i - \widehat{\Delta E}_i)^2} \tag{2.6a}$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(\Delta E_i - \widehat{\Delta E}_i) \tag{2.6b}$$

where n is the number of data points, ΔE_i is the actual value of the change in emissions at time i and $\widehat{\Delta E}_i$ is the predicted corresponding value. In the next section, we present the RMSE and bias as percentages of the mean of the absolute value of the change in emissions, i.e. $\frac{1}{n}\sum_{i=1}^{n}|\Delta E_i|$.

Chapter 3

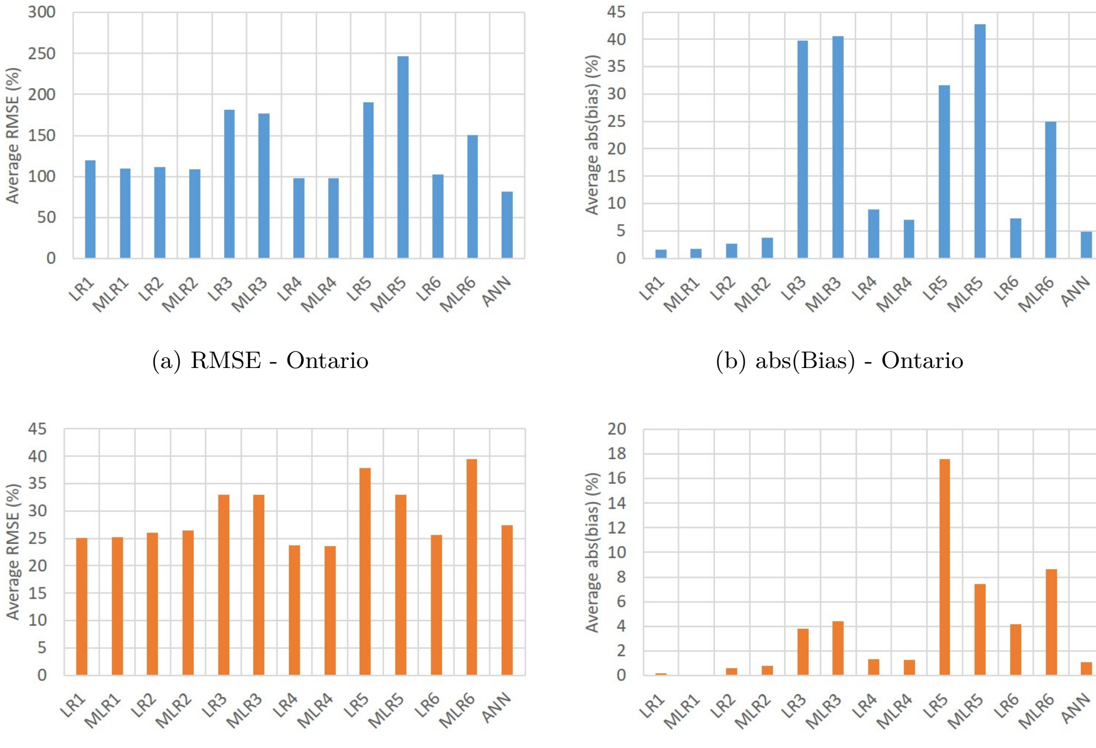
Results and Discussion

3.1 Comparison of methods for determining marginal emission factors

Figure 3.1 shows the RMSE and the absolute value of the bias of the predicted hourly changes in GHG emissions, averaged over the test years 2018 and 2019, as a percentage of the mean of the absolute value of the hourly change in emissions. The simple and multiple regressions with data partitions 3, 5 and 6 tend to do comparatively poorly in terms of bias and, to a lesser extent, RMSE. Data partitions 3 and 5 both involve binning the data by hour; their poor performance suggests that hour of day is a poor predictor of MEFs. Data partition 6 involves binning by month and load level; since data partitions 2 (binning by month) and 4 (binning by load level) perform relatively well, data partition 6 may simply be overfitting the data.

Relative performance of the ANN and the linear regression methods using data partitions 1, 2 and 4 varies by metric and province. The linear regressions LR1 and MLR1, where MEFs are computed over the full year, do best in terms of bias, whereas the ANN approach has the lowest RMSE for Ontario and MLR4 has the lowest RMSE for Alberta. The lowest biases for both provinces are less than 2%. Meanwhile, RMSEs for Alberta, with a minimum of 24%, are much lower than for Ontario, with a minimum of 81%.

The ANN method was included in part to investigate whether adding more predictors or



(a) RMSE - Ontario

(b) abs(Bias) - Ontario

(c) RMSE - Alberta

(d) abs(Bias) - Alberta

Figure 3.1: RMSE and absolute value of bias of the different methods tested averaged over 2018 and 2019

non-linear regressions could help reduce the RMSEs. This method does indeed have lower RMSEs for Ontario (81% vs. 98% for the next best performer), but this is associated with a higher bias of about 5% averaged over both years (and 7% for 2017). During model training for Ontario, we found that the bias varied considerably as the model parameters were varied, suggesting that the ANN approach will not consistently lead to low bias. Meanwhile, for Alberta, the ANN method had a higher RMSE than the best linear regressions, possibly due to overfitting since ΔG_{t+1} is very strongly correlated with ΔE_{t+1} (see Table 2.5).

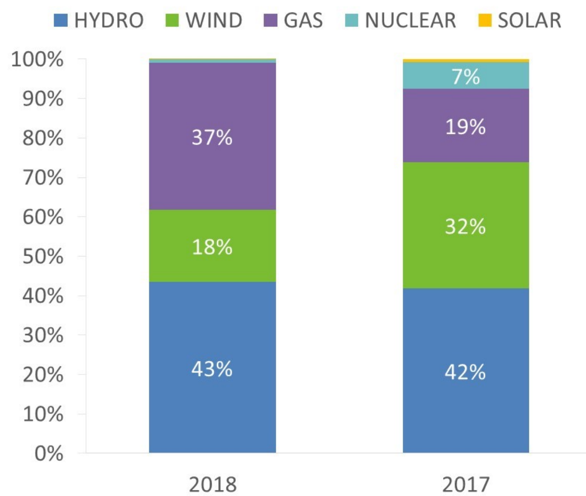
One open question was whether the intercepts in (2.3) and (2.4) should be set to zero, i.e. whether change in emissions should be zero when change in generation is zero. In principle, if the data filtering described in Section 2.1.2 were perfect, the intercept should be zero since the filtering is supposed to remove all emission changes that are not related to a change in total generation. In practice, allowing a non-zero intercept generally leads to slightly lower biases, so we used non-zero intercepts. However, setting the intercept of the linear regression to 0 should not significantly impact the results; when it did, this acted as a red flag that a method may be inappropriate. For instance, the goodness of fit of regressions with partitions 3 and 5 was sometimes quite sensitive to whether or not the intercept was forced to 0, even leading to incoherent results (negative MEFs) in some cases unless the intercept was forced to 0.

For use cases where changes in annual emissions are the main interest, bias is key, so methods LR1 and MLR1 may be the most appropriate. We selected MLR1 for the remainder of our analysis since its MEFs depend on generation level, which allows it to generate MEF profiles by hour and by month (for instance), while LR1 provides a single value for the entire year. The coefficients of MLR1 for Ontario and Alberta for 2019, along with the calculated MEFs by month and hour, are provided in Appendix A. It is important to note, however, that recent inter-comparisons of methods to compute MEFs of electricity generation in Alberta suggest that MLR1 strongly underestimates the variability of MEFs across hours and months, at least for that province [20].

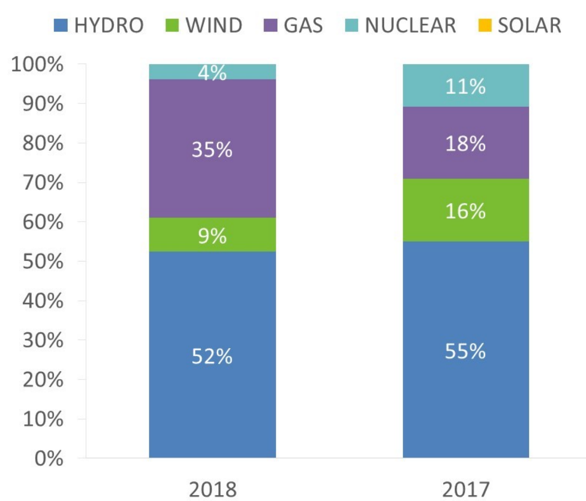
3.2 Benchmarking results against ISO data and literature

As noted earlier, the ISOs of Ontario and Alberta report how often each generator type sets the market clearing price ([19], [1], [2]). We compare this to our results using method MLR1 in Figures 3.2 and 3.3. In Ontario, our method estimates the fraction of time that gas is marginal quite accurately, but it underestimates the marginal contribution of wind and overestimates that of hydroelectric generators. For Alberta, our analysis underestimates the contribution of coal to marginal generation and overestimates the contributions of gas and hydroelectric generation, as shown in Figure 3.3. The differences observed with the IESO and AESO data reflect the limitations of the data filtering methods we used to try to isolate the marginal generators for each hour. More sophisticated methods that more closely reflect how generators are dispatched may be needed to improve the accuracy of the MEF estimates.

Table 3.1 shows the annual AEFs and MEFs for both provinces in 2017, 2018, 2019, with

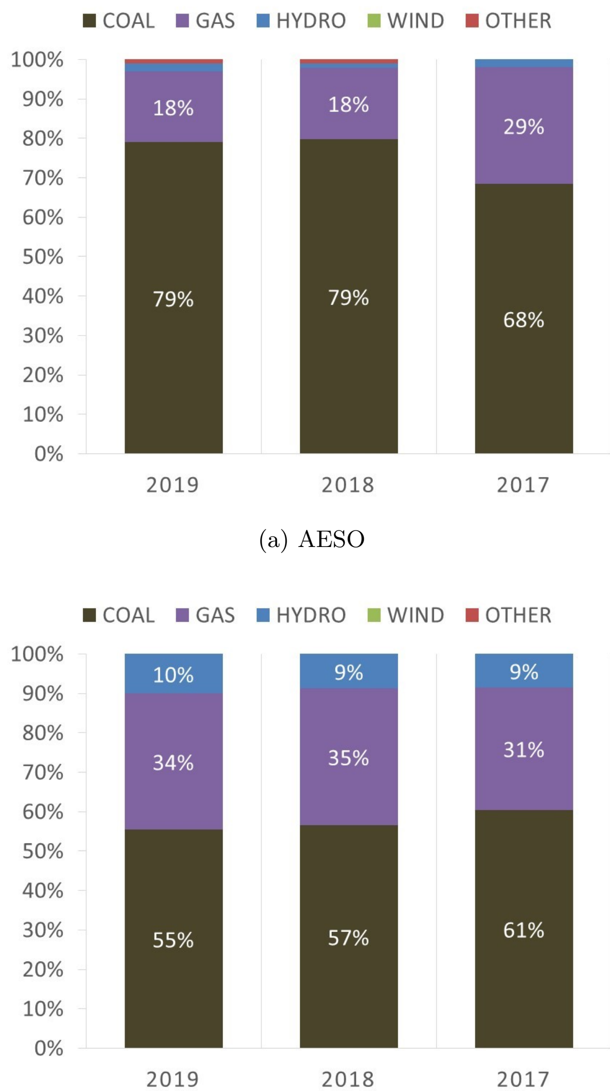


(a) IESO



(b) Our analysis

Figure 3.2: Fraction of time that different generator types are marginal based on IESO data (a) and our analysis (b)



(a) AESO

(b) Our analysis

Figure 3.3: Fraction of time that different generator types are marginal based on AESO data (a) and our analysis (b)

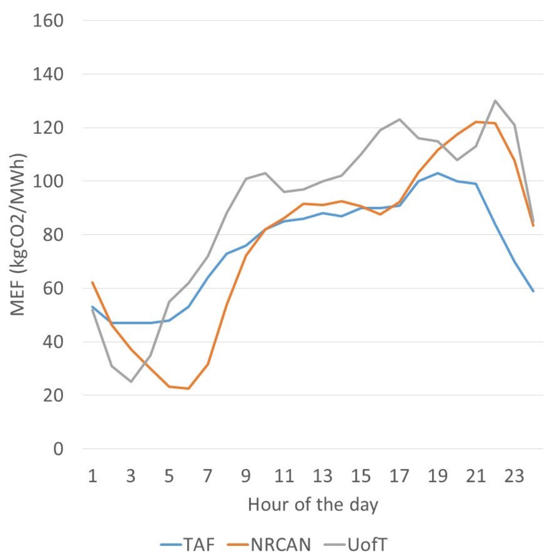
MEFs computed using the MLR1 method. Since IESO and AESO do not provide historical MEFs, the "MEF-AESO" and "MEF-IESO" columns in Table 3.1 were estimated instead by multiplying the fraction of time different generator types are marginal according to AESO and IESO with average emission factors by generator type from our analysis (note that the needed IESO data for all of 2019 were not available for our analysis, so this year was left out). On average, our estimated MEFs are 4% lower in Ontario and 13% lower in Alberta than those estimated from ISO data.

	Alberta			Ontario		
	AEF	MEF	MEF-AESO	AEF	MEF	MEF-IESO
2017	710	697	803	18	78	83
2018	654	693	853	27	150	153
2019	657	682	853	26	144	N/A

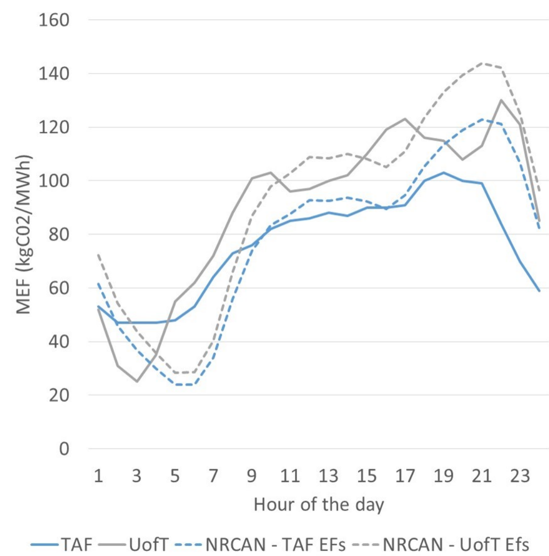
Table 3.1: Annual AEF and MEF for Ontario and Alberta in kg of CO_2 per MWh

Comparing the two provinces, both MEFs and AEFs are much lower for Ontario than for Alberta, reflecting their different generation fleets. Alberta's fleet is dominated by fossil fuel based generation both on average and at the margin, while Ontario's is dominated by clean and renewable generation. For Alberta, AEFs and MEFs are comparable, while for Ontario, MEFs are significantly higher than AEFs. This reflects the fact that for the latter, gas meets 6% of annual electricity demand but provides 20% of marginal generation (based on 2019 data).

We also compared our MEFs for Ontario to those of The Atmospheric Fund (TAF) ([3]) and the University of Toronto (UofT) ([4]) for 2017, as shown in Figure 3.4. We compared our results directly, Figure 3.4a, and also re-ran our analysis using the generation type GHG EFs used by TAF and UofT, Figure 3.4b. On an annual basis, the NRCan and TAF values are within 2% both for the original data and for the data with rescaled EFs. Meanwhile, the annual MEFs for UofT are 16% above our values, but the difference reduces to about 2% when the same EFs are used. The diurnal patterns of the three methods differ to some extent, but they all show lower values in the morning followed by a period of increase and a final dip at the end of the day. For the case of Alberta, we did not find other MEF estimates covering the same years as in this study.



(a) NRCAN method using EFs from Table 2.2



(b) NRCAN method using EFs from The Atmospheric Fund or The University of Toronto

Figure 3.4: Results of our method compared to The Atmospheric Fund [3] and The University of Toronto [4]

Chapter 4

Conclusion

We considered a number of options for estimating marginal emission factors (MEFs) of electricity generation based on historical data. Our results show promise with just a slight underestimate, as evidenced through indirect comparisons with data from Ontario (IESO) and Alberta (AESO). The comparisons between simple linear regressions, multiple linear regressions and artificial neural network approaches suggest that the multiple linear regression model based on a full year of data (MLR1) is the most reliable model. In particular, this model has low bias, which means that it can be used reliably to predict the impacts of policies and programs that modify electricity loads on annual greenhouse gas emissions. In addition, the MLR1 model allows MEFs to be estimated for any time frame of interest, such as by time of day and by month. This being said, recent inter-comparisons of methods for deriving MEFs [20] indicate that MLR1 strongly underestimates the variability of MEFs across hours and months, at least in the case of Alberta.

While the artificial neural networks had the lowest root mean square errors in Ontario, this was not the case in Alberta, and in Ontario it was achieved at the expense of significantly higher bias. This suggests that using more features or non-linear regressions is not key to improving MEF estimates. Rather, it seems that the limiting factor in these approaches is the simple set of rules used to exclude generators at each hour from the marginal generation pool. More complex models based on power system optimization may be needed to further improve MEF estimates. Such models are also more suitable to estimating MEFs decades ahead by taking into account power system planning.

Along these lines, it is worth noting that electricity system operators have the most comprehensive data available about their own systems and are therefore in the best position to provide information about MEFs. Such a practice among ISOs would certainly facilitate policy and program development in line with greenhouse gas emissions reduction targets. For instance, the IESO in Ontario has now begun providing MEF forecasts for the next twenty years as part of its Annual Planning Outlook [21].

APPENDIX

Appendix A

MLR1 Results

The distributions of MEFs by hour and by month according to MLR1 are provided below for Ontario and Alberta, along with the coefficients of the multiple linear regressions. However, as mentioned in Section 3.1, recent inter-comparisons of methods to compute MEFs of electricity generation in Alberta suggest that MLR1 strongly underestimates the variability of MEFs across hours and months, at least for that province [20].

A.1 Ontario

MEF	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec	Full Year
1	170	152	118	93	97	121	145	107	86	90	122	148	121
2	155	133	107	84	85	103	127	94	81	83	109	133	108
3	144	134	103	84	78	92	118	93	77	74	103	127	102
4	143	126	101	79	68	82	111	90	73	71	98	121	97
5	139	122	100	79	70	81	108	85	70	70	93	117	95
6	135	120	103	87	75	81	108	90	76	76	95	118	97
7	144	129	120	99	90	90	114	94	90	94	103	126	108
8	162	152	146	118	106	113	139	113	103	116	128	142	128
9	190	178	160	125	120	134	157	127	112	129	152	166	146
10	201	183	156	127	126	143	171	141	120	130	160	176	153
11	200	183	156	125	126	146	187	154	125	128	163	182	156
12	204	192	150	127	127	150	198	158	130	130	163	188	160
13	203	185	146	123	124	154	210	165	131	130	161	184	160
14	199	179	137	120	124	158	218	175	134	127	158	185	159
15	198	171	132	110	121	157	221	177	134	128	157	177	157
16	195	177	135	110	124	156	225	179	135	122	155	178	158
17	194	175	130	116	124	165	225	183	138	132	152	175	159
18	207	177	138	126	135	171	243	195	151	139	165	189	170
19	236	198	147	123	131	172	245	200	148	145	191	220	180
20	244	212	169	128	136	177	237	192	155	156	191	217	184
21	242	215	176	141	140	179	230	194	158	156	187	211	186
22	237	209	172	141	137	178	231	186	144	144	178	202	180
23	221	199	155	123	129	171	213	165	122	124	163	186	164
24	194	172	133	107	112	142	176	129	102	106	142	171	140
All hours	190	170	137	112	113	138	182	145	116	117	145	168	144

Figure A.1: MEFs obtained with the MLR1 method for 2019 in Ontario

	MLR1
β_0	6769.689928 kg
β_1	-291.8180402 kg/MWh
β_2	0.025920053 kg/MWh^2
β_3	-0.035565223 kg/MWh^2

Table A.1: Regression coefficients (see equation (2.4)) for the MLR1 method

A.2 Alberta

MEF	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec	Full Year
1	689	684	688	693	698	697	695	695	695	692	691	689	692
2	685	680	684	690	695	693	691	692	691	689	688	687	689
3	685	679	682	690	693	692	691	691	690	688	687	685	688
4	683	678	682	689	691	691	689	690	689	687	685	684	687
5	681	677	680	685	689	687	687	687	685	684	683	681	684
6	678	676	678	681	686	687	684	682	681	679	679	678	681
7	673	671	673	677	683	682	680	679	676	674	676	674	677
8	668	671	672	675	678	676	673	674	672	670	673	670	673
9	672	673	675	676	679	675	673	674	676	677	677	672	675
10	678	675	677	679	682	679	673	674	676	680	678	676	677
11	674	673	678	682	683	680	674	677	677	679	679	677	678
12	676	676	677	684	683	679	676	677	679	680	681	680	679
13	678	678	680	685	684	682	679	680	681	682	683	681	681
14	679	677	680	683	685	680	680	680	680	682	681	680	681
15	680	678	681	684	686	684	681	681	681	682	682	681	682
16	678	676	680	683	684	682	680	680	680	683	680	678	680
17	675	675	679	681	683	682	681	679	677	678	675	672	678
18	670	672	679	683	684	684	679	678	679	678	675	673	678
19	679	672	679	687	688	688	685	685	685	682	685	682	683
20	679	676	680	684	688	686	684	685	683	680	682	681	682
21	680	676	677	681	687	686	685	685	679	685	682	679	682
22	683	680	683	681	686	689	685	683	690	688	685	685	685
23	688	683	687	691	691	690	686	690	692	691	688	688	689
24	688	681	688	693	696	695	693	694	696	693	692	690	692
All hours	679	677	680	684	687	685	683	683	683	683	682	680	682

Figure A.2: MEFs obtained with the MLR1 method for 2019 in Alberta

	MLR1
β_0	1062.090109 kg
β_1	717.3722402 kg/MWh
β_2	-0.005051087 kg/MWh^2
β_3	-0.025865943 kg/MWh^2

Table A.2: Regression coefficients (see equation (2.4)) for the MLR1 method for Alberta in 2019

References

[1] Ontario Energy Board Market Surveillance Panel, “Monitoring Report on the IESO-Administered Electricity Markets,” Ontario Energy Board, 2021, accessed: 2021-03-26. [Online]. Available: https://www.oeb.ca/utility-performance-and-monitoring/electricity-market-surveillance/panel-reports

[2] AESO, “Annual Market Statistic Report,” 2019, accessed: 2020-05-25. [Online]. Available: https://www.aeso.ca/market/market-and-system-reporting/annual-market-statistic-reports/

[3] J. Sotes, “A Clearer View on Ontario’s Emissions,” The Atmospheric Fund, 2019, accessed: 2020-05-25. [Online]. Available: https://taf.ca/wp-content/uploads/2019/06/A-Clearer-View-on-Ontarios-Emissions-June-2019.pdf

[4] Y. Gai, A. Wang, L. Pereira, M. Hatzopoulou, and I. D. Posen, “Marginal greenhouse gas emissions of ontario’s electricity system and the implications of electric vehicle charging,” *Environmental science & technology*, vol. 53, no. 13, pp. 7903–7912, 2019.

[5] U.S. Energy Information Administration (EIA), “CO2 content of different fuels,” 2020, accessed: March 2021. [Online]. Available: https://www.eia.gov/tools/faqs/faq.php?id=73&t=11/

[6] N. A. Ryan, J. X. Johnson, and G. A. Keoleian, “Comparative assessment of models and methods to calculate grid electricity emissions,” *Environmental science & technology*, vol. 50, no. 17, pp. 8937–8953, 2016.

[7] A. D. Hawkes, "Estimating marginal co2 emissions rates for national electricity systems," *Energy Policy*, vol. 38, no. 10, pp. 5977–5987, 2010.

[8] K. Siler-Evans, I. L. Azevedo, and M. G. Morgan, "Marginal emissions factors for the us electricity system," *Environmental science & technology*, vol. 46, no. 9, pp. 4742–4748, 2012.

[9] B. Tranberg, O. Corradi, B. Lajoie, T. Gibon, I. Staffell, and G. B. Andresen, "Real-time carbon accounting method for the european electricity markets," *Energy Strategy Reviews*, vol. 26, p. 100367, 2019.

[10] C. Wang, Y. Wang, C. J. Miller, and J. Lin, "Estimating hourly marginal emission in real time for pjm market area using a machine learning approach," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, 2016, pp. 1–5.

[11] IESO, "Power Data - Data Directory," 2020, accessed: April 2020. [Online]. Available: http://www.ieso.ca/en/Power-Data/Data-Directory

[12] AESO, "Data Request," 2020, accessed: April 2020. [Online]. Available: https://www.aeso.ca/market/market-and-system-reporting/data-requests/

[13] D. Lew, G. Brinkman, N. Kumar, P. Besuner, D. Agan, and S. Lefton, "Impacts of wind and solar on fossil-fueled generators," in *Presented at IEEE Power and Energy Society General Meeting (preprint)*. NREL, 2012.

[14] U.S. Energy Information Administration (EIA), "Electric power annual report, 2020, Table 8.2, Year 2011 data," 2020, accessed: March 2021. [Online]. Available: https://www.eia.gov/electricity/annual/html/epa_08_02.html/

[15] Environment and Climate Change Canada, "Canada - 2019 National Inventory Report," 2019, Part 3, Annex 13 - accessed: April 2020. [Online]. Available: https://unfccc.int/documents/194925

[16] U.S Energy Information Administration, "State Electricity Profiles," 2020, Table 7 - accessed: April 2020. [Online]. Available: https://www.eia.gov/electricity/state/

[17] A. A. Farhat and V. I. Ugursal, “Greenhouse gas emission intensity factors for marginal electricity generation in canada,” *International Journal of Energy Research*, vol. 34, no. 15, pp. 1309–1327, 2010.

[18] T. B. Group, “Negative pricing in wholesale energy markets,” 2018, accessed: 2020-05-25. [Online]. Available: http://www.ieso.ca/-/media/Files/IESO/Document-Library/working-group/market-renewal/nersc/NERSC-20181130-Negative-Pricing-in-Wholesale-Electricity-Markets.pdf?la=en

[19] Ontario Energy Board Market Surveillance Panel, “Monitoring Report on the IESO-Administered Electricity Markets,” Ontario Energy Board, 2019, accessed: 2020-05-25. [Online]. Available: https://www.oeb.ca/utility-performance-and-monitoring/electricity-market-surveillance/panel-reports

[20] V. Raissi Dehkordi, I. Das, V. Delisle, S. Pelland, M. Hansen, M. Hundal, and F. Bensebaa, “Open and Common Approaches for Evaluating Marginal Emission Factors: A Case Study of the Alberta Electric Grid,” 2021, accessed: 2021-06-24. [Online]. Available: https://emi-ime.ca/wp-content/uploads/2021/03/EMI-2020-Delisle_report_Open-and-Common-Approaches-for-Evaluating-Marginal-Emission-Factors-2.pdf

[21] IESO, “Annual Planning Outlook,” 2020, accessed: 2021-02-10. [Online]. Available: https://www.ieso.ca/en/Sector-Participants/Planning-and-Forecasting/Annual-Planning-Outlook/