ENVIRONMENTAL RESEARCH
ENERGY

Validating locational marginal emissions models with wind generation

IOP Publishing Environ. Res.: Energy 1 (2024) 035008 https://doi.org/10.1088/2753-3751/ad72f6

ENVIRONMENTAL RESEARCH
ENERGY

CrossMark

OPEN ACCESS

RECEIVED
25 June 2024

REVISED
19 August 2024

ACCEPTED FOR PUBLICATION
23 August 2024

PUBLISHED
4 September 2024

PAPER

Validating locational marginal emissions models with wind generation

Nat Steinsultz[1,*], **Pierre Christian**[1], **Joel Cofield**[1], **Gavin McCormick**[1] **and Sarah Sofia**[2]

[1] WattTime, Oakland, CA, United States of America
[2] REsurety, Boston, MA, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: nat@watttime.org

Keywords: emissions, power system, energy, electricity emission, causal inference, load shifting

Supplementary material for this article is available online

Abstract

Increasingly large amounts of electric supply and load are being deliberately operated or sited on the basis of marginal emissions factor (MEF) models. Validating and calibrating such models is therefore of growing policy importance. This paper uses a natural experiment involving variation in relative changes in wind generation potential at wind farms in the ERCOT power grid to create a benchmark MEF and examine the relative accuracy of several common classes of short term MEF models. This work focuses on MEFs at the level of a few individual generating nodes, a much smaller geographic scale than the Balancing Authority (BA) or load zone. Additionally, the use of wind generation potential as a regressor allows us to factor in wind curtailment, in contrast to previous work. We evaluate multiple prevalent existing MEF models and find that both dispatch and statistical MEF models have a high degree of agreement with the benchmark MEF, while heat rate and average emissions do not. We also find that the emissions reduction benefits of optimizing electricity with MEFs using a geographically granular model instead of a BA-wide model are 1.4, 1.3 and 1.5 times larger for dispatch, statistical and heat rate models, respectively.

1. Introduction

One strategy to reduce CO_2 emissions from the electricity sector is intentionally shift electricity load away from times when incremental net load will cause more incremental emissions, to times when it causes less. This emissions-based 'load-shifting' strategy can be applied in a variety of sectors including EV charging [1], data centers [2–4] and battery storage dispatch [5]. Since 2017 this approach has begun to be implemented on a wide scale by groups including Apple, Google Nest, Microsoft, Toyota and the California Public Utility Commission (CPUC). The CPUC implemented an MEF-based emissions reduction requirement on incentives for commercial storage systems in 2020, resulting in participating systems after the change reducing emissions by 4.7 kg CO_2 kWh^{-1} instead of increasing emissions by 5.4 kg CO_2 kWh^{-1} [6]. Another strategy for reducing emissions is to intentionally optimize the location of new renewable energy developments to locations when additional generation would displace particularly highly emitting generation [7, 8]. Organizations are beginning to make renewable energy purchase decisions on the basis of such strategies, including Salesforce's purchase of 280 GWh of renewable energy certificates [9], BrightNight partnering with the Nature Conservancy to construct an 800MW solar farm in Kentucky [10] and Amazon's contracting of 2.83 TWh yr^{-1} of renewable energy in India [11].

Both of these strategies require a knowledge about the variation in the change of carbon dioxide (CO_2) emissions in response to changes in load at different times and locations. This variation occurs due to a mix of two factors: different marginal rates of emissions from the operations of the existing power grid, plus different marginal impacts on the structure of that power grid in the future [12]. This study examines the former effect, which is variously known as marginal operating emissions rates, short run marginal emissions, or most commonly the Marginal Emissions Factor (MEF) [13, 14]. Previous studies have estimated the MEF at a regional [15] or balancing authority [16] (BA) level, but understanding MEFs at increasingly smaller geographic resolution is important as local transmission constraints can result in different behaviors within a single BA.

There are multiple distinct approaches to estimating MEFs, with few benchmarks for comparing these models to each other. In the US, the EIA has been directed to publish hourly marginal emissions rates, but is still working on a model design to do so [17]. Given that many organizations are already making load-shifting and renewable energy purchase decisions based on existing MEF models, it is important that these models be validated.

Previous works on measuring the MEF mostly rely on either statistical models (e.g. [15] and [16]), dispatch models based on generator costs and transmission network constraints (e.g. [18, 19] and [20]), or heat rate models based on market prices and fuel costs (e.g. [21]).

Statistical models rely on historical data to measure the statistical relationship between load and emissions. The most common implementation of this strategy is by conditioning the data on potential confounders (e.g. time of day or temperature) and performing linear or non-parametric regressions on the emissions versus the load [16, 22, 23]. The advantage of these models is that they are empirical and can be causal if the confounding variables are controlled for in the model. The main challenge in utilizing these models lies in ensuring that the relationship obtained is causal and not simple correlations; if uncontrolled confounders persist in the final model, the MEF measurement will be biased. Also, because these models rely on statistical regressions, many observations need to be aggregated to get meaningful results. Consequently, most statistical models have limited spatial and temporal granularity compared to the granularity of real world operations.

Unlike statistical models, dispatch models do not require multiple observations, making them ideal in environments where the use of statistical models are obstructed by limited data. In particular, dispatch models allow for measurements of the MEF at spatial resolutions down to a nodal level that, due to data quality, are too small for statistical models. Though dispatch models identify the marginal unit for a small change in load (∼1MW), different units or groups of units may be marginal for larger changes in load. It is hitherto unknown whether these models are applicable for larger load changes in the context of emission load-shifting. Furthermore, dispatch models identify marginal generators, but rely on the average emissions intensity of those generators to calculate the MEF, which is a potential source of error as generators can have different emissions rates based on capacity factor or ramping status. [20] These models also need to be calibrated against historical data to ensure values match realistic scenarios and not inaccuracies from numerical methods. Additionally, dispatch models are only possible in grids that are dispatched using security constrained economic dispatch (SCED), which is not widely used outside of the US and Australia.

Heat rate models require only pricing information and assumptions about the available generators in a region and can also be highly spatially granular. Because they are based on simple cost models, they can be easily described and validated by many parties. For this reason, these models have been chosen for government policies, such as those run by the CPUC [21]. However, these models are not empirically based on emissions or generation data and may become inaccurate when multiple generator types are present.

Signals other than the MEF are often used in siting or operating electrical load. Radovanović *et al* [3] uses the average emissions factor (AEF) in optimizing data center operations for carbon emissions. The locational marginal price (LMP) is also often used by entities optimizing for wholesale electricity price. These signals are advantageous because they are easily determined from publicly available data and can be measured in real time. However, the use of these proxy signals for emissions reduction is typically based on assumptions about grid behavior that have not been validated against actual change in emissions. The AEF does not take into account that baseload electricity generators are, in general, not the same as the marginal generators and thus do not typically change their emissions in response to a change of load. And LMP does not always correlate with generator emissions rates and can be heavily influenced by transmission costs which may not have an impact on marginal generation.

With these many different approaches to estimating the MEF, it is important to evaluate the realistic emissions impact of using each to guide decision making. Previous studies have established evaluation criteria [24] and have compared multiple MEF models to each other and analyzed the impact of each model in a simulated application [25]. However, these works did not establish a uniform benchmark as a comparison. Another study was able to compare multiple MEF models to a benchmark using grid simulations [26], but it is not known how these simulation results apply to real world grids. A separate study produced benchmark MEFs by leveraging nuclear outages, but it is limited to the specific time frames during the outages [27].

In this work we leverage wind generation potential as a natural experiment to measure the MEF in a way that is properly identified. Due to the need to balance electricity generation and demand, a decrease in wind generation necessitates an increase in other sources of electricity generation to compensate for the loss of wind power, making changes in wind a quasi-random proxy for local changes in demand. We use wind generation potential instead of actual generation in order to avoid controlling for the causal pathway where wind generation is curtailed due to insufficient load or oversupply.

Previous studies have used wind generation to estimate the MEF [23] and estimate the impact of transmission and congestion on MEFs [28]. However, these studies have been limited in geographic granularity to either the entire balancing authority (BA) or BA-defined load zones which cover a large area. In this study, we use nodal generation data to estimate the MEF at a much smaller and arbitrary geographic granularity, while incorporating the impacts of curtailment. These granular measurements of the MEF allow us to create a benchmark to validate and compare multiple MEF models and understand the impact on total emissions of using these models or proxies.

2. Materials and methods

2.1. Description of data sets

2.1.1. Emissions

For the MEF benchmark model, hourly emissions data are obtained from the Environmental Protection Agency's Clean Air Markets Program Data (CAMPD) [29]. Contained in this dataset are the CO_2 emissions of power generating facilities with more than 25MW of capacity in the continental United States. Because we lacked linking information between power plant identifiers in the CAMPD and ERCOT data, we use generator specific emissions rates provided by ERCOT for calculating the local sub-region average emissions rate [30].

2.1.2. ERCOT data

We obtain load and net generation by fuel type data for ERCOT from the EIA-930 dataset [31]. Our wind and solar generation potential data consists of the High Sustained Limit (HSL) MW input that each individual generator provides to the SCED system in ERCOT. In ERCOT, each generator is provided with a short term power forecast and the HSL for each generator must be set to less than or equal to the short term forecast. By default, the HSL is automatically set to be equal to the short term wind power forecast [32]. The HSL is selected for this paper because it most closely represents the wind generation *potential*, independent of any instructions from the grid operator to curtail some or all of that potential. Under certain grid conditions, actual generation is less than this potential because it is curtailed in response to either ERCOT oversupply or transmission constraints. 5% of all potential wind generation and 9% of all potential solar generation was curtailed in ERCOT in 2022 [33].

2.2. MEF models

2.2.1. Dispatch model

The dispatch model used in this paper was developed by REsurety and estimates marginal emissions at nodal resolution derived from standard calculations of LMPs by the grid operator. [34] This model estimates the required incremental change to the dispatch of the marginal generators to meet an incremental change in load at a given node for each point in time. This is done by solving for the least cost energy subjected to (1) balance of generation and load and (2) preservation of flow over binding constraints, using offers, constraint data and shift factors. The marginal emissions rate is then calculated by multiplying the emissions rate of each marginal generator by its required change in dispatch to meet the load at a given node. Generator emissions rates are calculated based on fuel type and heat rate inferred from reported net generation and fuel consumption data.

2.2.2. Statistical model

The statistical model has two distinct components. The first is a binned regression model produced by WattTime [35], which regresses changes in generation by fuel type on changes in load and which bins by load, LMP and time factors (month, day of week and hour of day). The carbon emissions rate of fossil fuel generators is modeled using a binned regression of changes in CAMPD emissions data against changes in load using the same binning variables. The second component is an an estimated probability of curtailment that is used to de-rate the MEF predicted by the first binned regression model. This second component is calculated by measuring the fraction of nodes in the sub-region with an LMP below $0.

2.2.3. Heat rate model

The heat rate model is produced using LMP, gas prices and assumptions about the relation between costs and emissions. Daily spot prices at the Henry hub as reported by the EIA [36] are used as the gas price and the mean LMP in the BA or sub-region is used as the electricity market price. The heat rate MEF is then defined as

$$\mathrm{HR}(t) = \frac{\mathrm{LMP}(t) - \mathrm{VOM}}{\mathrm{FP}(t)} \tag{1}$$

$$\mathrm{MEF_{HR}}(t) = \begin{cases} 0, & \text{if } \mathrm{HR}(t) \leqslant 0 \\ \mathrm{HR}(t) * \mathrm{CI}, & \text{if } \mathrm{HR}(t) > 0 \\ & \text{and } \mathrm{HR}(t) \leqslant \mathrm{HR_{max}} \\ \mathrm{HR_{max}} * \mathrm{CI}, & \text{if } \mathrm{HR}(t) > \mathrm{HR_{max}} \end{cases} \tag{2}$$

where $\mathrm{HR}(t)$ is the heat rate in MMBtu/MWh at time t, $\mathrm{LMP}(t)$ and $\mathrm{FP}(t)$ are the LMP and fuel price in \$/MWh and \$/MMBtu respectively, VOM is the variable operating cost in \$/MWh (0.25 \$/MWh, as used by the CPUC [37]), $\mathrm{HR_{max}}$ is the maximum heat rate of the lowest efficiency plants, 15.3 MMBtu/MWh, which is the 90th percentile value of all generators in ERCOT in the 2022 eGRID dataset [38] and CI is the carbon emissions intensity of gas, 53.1 kg CO_2/MMBtu.

2.3. Methodology

2.3.1. Aggregation

Each model in this paper can be calculated at the level of an individual pricing node. In order to increase the statistical power of our regressions, we aggregate multiple nodes with wind generators into sub-regions. For each settlement point associated with a wind generator in ERCOT, we calculate the pair-wise correlation matrix of the dispatch model MEF time series for the year 2022. We then apply a hierarchical clustering analysis to the correlation scores and test multiple different cutoff criterion to select a suitable cutoff score. The resulting clusters are then sorted by aggregate wind HSL inside the cluster and the top 4 clusters are selected and used in the experiment. This approach was selected to create a number of sub-regions with high intra-region correlation, but not require that every settlement point is assigned to a cluster that is used in the experiment. Node clusters are visualized in supplementary figure 2.

2.3.2. Benchmark MEF

We are interested in the extent to which different granular and widely available MEF models can successfully predict the causal effect on total ERCOT emissions of changing net load (which is demand minus variable renewable generation) at a specific place and time. We do not have access to geographically granular load data, so we use nodally resolved wind data to understand geographically resolved changes in net load. This substitution is valid assuming that demand is always matched by generation. We produce a well-identified benchmark MEF derived from a natural experiment by extending the approach used in [23] with nodal data from individual wind generators. Instead of wind generation, we use wind generation *potential* as an independent variable to estimate the local change in net load in the absence of curtailment. The use of wind generation potential allows us to examine an exogenous variable, where actual wind generation may be partly endogenous when wind energy is curtailed as a result of demand. We include solar generation potential as a variable for similar reasons. For the full BA model, we perform this estimation with an OLS regression on the linear model

$$\Delta E = M\Delta W + \beta_1 \Delta L + \beta_2 \Delta S + \vec{\gamma} \cdot \vec{X}\,, \tag{3}$$

where ΔE is the hourly change of total CO_2 emission in ERCOT, ΔL the hourly change of load excluding nuclear and imports, ΔS the hourly change of solar generation potential, M the MEF measurement for the entire BA, ΔW the hourly change of the wind generation potential (wind generation HSL) $\vec{X}$ a vector of controls that contains the time fixed effects of hour of the day and month of the year. The load excluding nuclear and imports is defined as,

$$L \equiv D - G_{\mathrm{N}} + I\,, \tag{4}$$

where D is the electricity demand, G_{N} the nuclear generation and I the interchange for ERCOT.

MEF models measure the coefficient of change in emissions with load, however, we identify the coefficient of change in emission with wind generation potential. We do this in order to achieve sub-regional granularity, using the granular wind generation data. This mirrors the approach of all the MEF models we consider where generation and load are treated symmetrically. While theoretically, M should always equal $-\beta_1$, it is possible that this specification introduces bias in our estimate of the benchmark MEF, as it is not

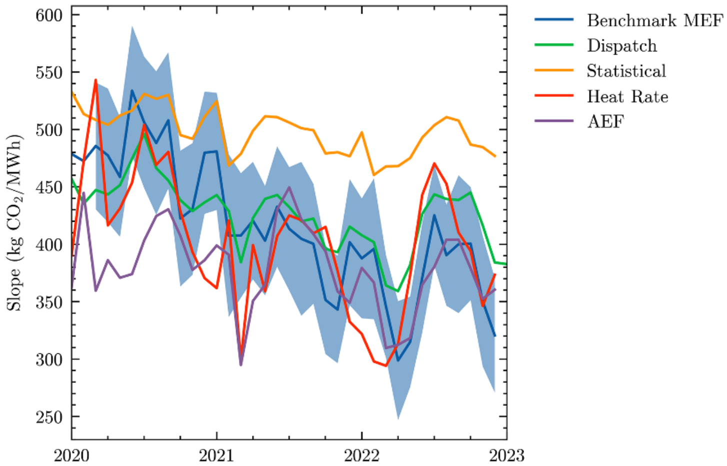

Figure 1. Month by year average MEFs calculated using the benchmark MEF compared with statistical, dispatch and heat rate MEF models. The average emissions factor is also plotted for comparison.

guaranteed the two coefficients are equal under all conditions. We verified that for our experiment, M indeed equals $-\beta_1$ evaluated over the entire experimental time frame.

For sub-regional models, we disaggregate the wind potential for each sub-region and use the model

$$\Delta E = \sum_{i=0}^{4} M_i \Delta W_i + \beta_1 \Delta L + \beta_2 \Delta S + \beta_3 \Delta W_e + \vec{\gamma} \cdot \vec{X}\,, \tag{5}$$

where ΔW_i is the hourly change of the wind generation potential for all the nodes in sub-region i, and ΔW_e is the hourly change of wind generation potential for all nodes that are not in any of the four sub-regions. Thus M_i, the benchmark MEF, isolates the effect on emissions of wind potential in a particular subregion. We test our model for multicollinearity and calculate variance inflation factors of 2.1 or less for all variables.

The mean absolute total hourly change of wind generation potential for all nodes (ΔW) across the study period is 823 MW, 90% range: $[54, 2275]$. For the four sub-regions (ΔW_i), the mean absolute value are 182 MW, 90% range: $[8, 555]$, 189 MW, 90% range: $[8, 558]$, 203 MW 90% range: $[7, 627]$, and 103 MW, 90% range: $[4, 317]$.

We use a variation of this regression formulation to estimate the share of marginal generation by fuel type. In place of ΔE, we regress the change in coal generation, gas generation, solar curtailment and wind curtailment (ΔG_C, ΔG_G, ΔC_S and ΔC_W). Wind curtailment represents the amount that wind generation is reduced below its potential through economic dispatch and is defined as:

$$C_{W_i} = \begin{cases} 0, & \text{if } G_{W_i} \geqslant W_i \\ W_i - G_{W_i}, & \text{if } W_i > G_{W_i} \end{cases} \tag{6}$$

where G_{W_i} and C_{W_i} are the wind generation and curtailment at node i respectively. Solar curtailment is defined similarly. We validate our models by performing a placebo test, as shown in supplementary figure 1.

3. Results

3.1. BA-wide results

We first calculate the benchmark MEF for the entire BA for the years 2020-2022 and compare to the MEFs of various models. We plot the month-by-year average benchmark MEF from the regression model alongside several other models in figure 1. We find good agreement in the overall trend from both the statistical model and the dispatch model, with high Spearman correlation scores and low mean bias. We report the correlation scores and mean bias for all models in table 1

3.2. Local results

We next characterize five different signal types based on nodal data to investigate the significance of granular sub-BA MEFs. We examine three MEF models: heat rate, dispatch and statistical, along with two proxy signals: local average emissions and LMP. These two proxy signals do not predict an MEF and cannot be directly compared to MEF values. However, they are frequently used as control signals in load dispatch, so we

Table 1. Summary of mean differences in kg CO_2/MW between MEFs during time of high and low values, with standard deviation denoted with parentheses, as determined by each model or signal.

model	Heat Rate	Dispatch	Statistical	Average	LMP
Monthly Correlation	0.561	0.833	0.787	0.446	—
Monthly Mean Bias	−15.0	10.2	81.4	−37.2	—
Full ISO Model	259.2 (14.5)	265.3 (14.2)	306.5 (15.4)	−6.2 (14.4)	258.5 (14.9)
Sub−regional Model	389.5 (65.1)	382.0 (87.6)	396.2 (70.3)	−31.7 (77.5)	388.8 (93.3)
Geographic	400.8 (27.2)	371.7 (70.3)	331.9 (202.4)	—	—

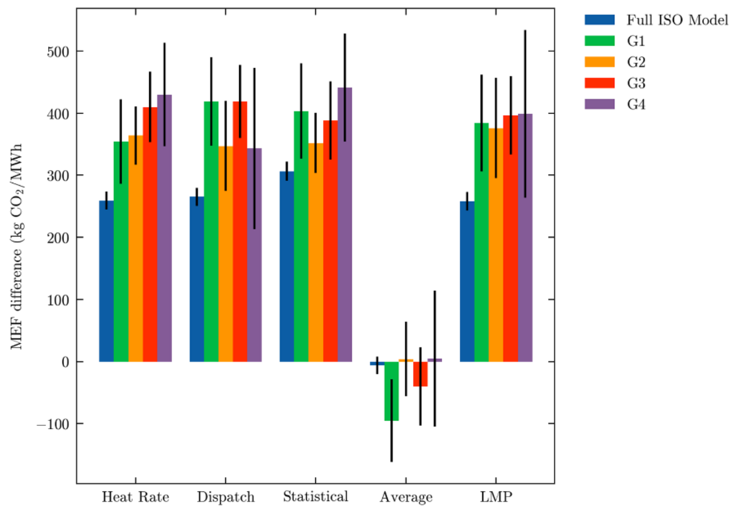

Figure 2. Difference in the benchmark MEFs between times when the evaluated signal is above or below the 10th percentile value for that signal. Full ISO is the model applied to all wind generators in ERCOT and G1-4 are localclusters of generators. Error bars are the 95% confidence interval.

include them in this analysis to understand the impact of using them in optimization. In order to estimate the potential impact on emissions of flexible load, we identify the times when each MEF model and signal predicts the least amount of emissions (below the 10th percentile) and estimate the benchmark MEF during those times for both the entire BA and for four granular sub-regions. We then estimate the benchmark MEF during the remaining times, when the signal is higher (above the 10th percentile). The impact of moving load from one time to another is the difference in marginal emissions between those two times. We plot the difference of the benchmark in figure 2 and summarize the mean values in table 1. For the average emissions signal, we find that the difference for BA-wide models is smaller than that of the other signals and the sub-regional models have a difference that is not statistically significantly different from 0. By contrast, variation in the dispatch, statistical, heat rate and LMP signals all successfully predict significant variation in the benchmark MEF, with a larger difference in the local models than in the BA-wide model.

These results indicate that local granular models are predictive of real MEF values and offer a greater opportunity for emissions reduction than BA-wide models. Transmission constraints within the BA are the primary driver of renewable energy curtailment in ERCOT, which could explain why different sub-regions have different MEF values at the same time, when one region has access to curtailed wind generation, but another does not. This is an important result because causal regression models have not previously been applied to geographic sub-regions this small. Models applied to an entire BA region will average together multiple marginal emissions rates from different sub-regions during times of transmission constraint, with a smaller opportunity for load-shifting to reduce emissions. Using the mean values reported in table 1, we find that the potential impact of using sub-regional models over BA-wide models are 1.4, 1.3 and 1.5 times larger for dispatch, statistical and heat rate models, respectively. The one notable exception to this are average emissions models, which perform worse on sub-regions, with a mean impact near 0. The emissions reductions potential from optimizing load using AEFs decreases, rather than increases, as geographic granularity increases.

To get a better understanding of the signals, we plot both the MEFs and the share of marginal fuel generation for the largest granular sub-region (G1) with values binned by percentile bucket of the model signal in figure 3. Both the dispatch and the statistical model show good agreement with the benchmark MEF

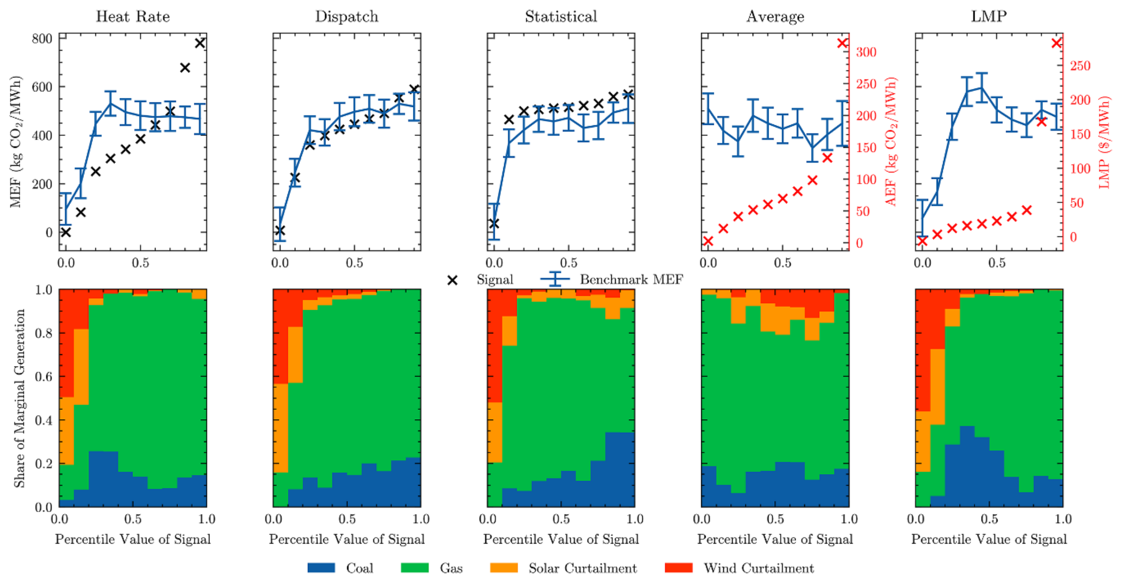

Figure 3. Top row: MEFs for the largest granular sub-region (G1) binned by percentile value of the signal (heat rate, dispatch, statistical, average or lmp), along with the mean value of the signal in that percentile bin. Because Average and LMP are proxy signals and do not produce an MEF estimate, the signal value is plotted on against the right axis (red). Bottom row: share of total marginal generation by fuel type for G1 binned by percentile value of the signal.

in both rank-ordering and absolute value. However, the dispatch model has some negative bias at low values, while the statistical model has some positive bias at high values.

The heat rate model and LMP signal are both effective at identifying times of low benchmark MEF in ERCOT, due to the correlation between low LMP and wind curtailment. However, both have a non-monotonic relation between signal and MEF, with both having a small peak in MEFs around the 40th percentile bin. This peak in MEFs corresponds to a peak in marginal coal generation. This is due to coal typically having lower marginal operating costs, making it a marginal fuel at lower LMPs than gas. This result may be unique to the fuel mix in ERCOT, particularly the large amount of renewable curtailment. Previous studies have shown optimizing only for price can increase total emissions [39, 40]. Even in ERCOT, while low LMP values correspond to low MEF values, because of the larger MEFs around the 40th percentile bin, it is not clear what the net effects on emissions of optimizing only on LMP would be. In other grids which do not have as much renewable curtailment, it is possible that heat rate models and LMP would perform significantly worse at predicting benchmark MEFs.

3.3. Cross-cluster comparison

In order to test each model's predictive capabilities to distinguish between regions at a given time, we compare sub-region groups and estimate the benchmark MEF during times when the MEF model predicts a difference of more than 181 kg CO_2 MWh^{-1} in the compared groups. This allows us to filter out the majority of the times when the model MEF is relatively equal across the investigated sub-regions. For each pair-wise comparison of sub-regions, we estimate the benchmark MEF in each group and then discard any comparisons where the range of the 95% confidence intervals of the difference are larger than the 181 kg CO_2 MWh^{-1} difference. This was done to remove some sub-region comparisons with only a few observations when the model MEF difference is larger than the threshold, leading to estimates with high error.

We report the mean difference for all significant cross-cluster comparisons in table 1. All three models show a mean difference greater than the 181 kg CO_2 MWh^{-1} threshold, indicating the capability to predict where MEFs are lower and that their definitions of BA sub-regions provide meaningful information about where MEFs diverge.

Because transmission constraints are the major driver of curtailment and variations in MEFs, it is important for an MEF model to be able to predict those impacts. By comparing the benchmark MEFs during times when a model predicts significantly different values at two different locations, we can validate a models ability to not only predict when MEFs are low, but also where. We find that all three MEF models perform well at this task with a mean benchmark MEF difference of 368 kg CO_2 MWh^{-1}.

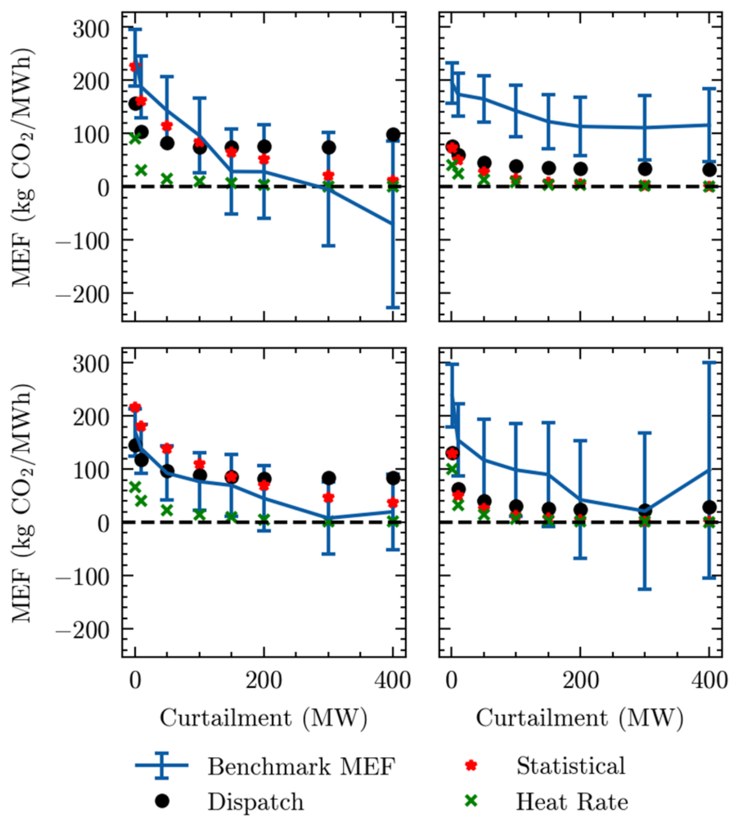

Figure 4. Estimations of the benchmark MEF during times when the total curtailment is above the varying threshold value for the four sub-regions, G1-4.

3.4. Curtailment

We perform a sensitivity test by estimating the benchmark MEF during times when the total amount of curtailed wind and solar generation is above a threshold value, which we vary from 10-400MW. We plot these results along with the predictions from the three geographically granular MEF models in figure 4. We observe that all sub-regions have a benchmark MEF above 150 kg CO_2 MWh^{-1} at small amounts of curtailment, with the benchmark MEF trending towards 0 as the amount of curtailment increases. The sensitivity to the amount of curtailment could be because when there is limited curtailment, the change in net load may exceed that small amount of curtailment, thus requiring other generators to provide at least some additional generation in response, generating some emissions despite the presence of curtailment.

There are several potential reasons why the benchmark MEF is non-zero, even at higher levels of total curtailment. It is possible that by using static definitions of sub-regional clusters, we are including times when there is a transmission constraint within the sub-region, where the benchmark MEF may be 0 at some generators but not at others. Because our measure of curtailment is defined as the difference between the forecasted and actual generation, it is possible we are also including times when the forecast over-predicts in our times of curtailment, even though no additional wind generation may be dispatchable. While these are limits of our current experimental design, it is also possible that the benchmark MEF is higher during those times because of additional constraints, such as fossil generator ramping and capacity, which prevent grid operators from fully dispatching curtailed energy. Further study into this question is needed, as curtailment is a significant source of variation in marginal emissions factor, that grows in importance as total renewable generation increases.

4. Discussion

This work is the first time that causal inference has been applied to nodal emissions data, creating benchmark MEF estimates for arbitrary sub-regions that can be used to benchmark the predictions of geographically granular models. This benchmark technique cannot be used in every BA given current limitations on data availability from grid operators other than ERCOT. However, the results can be used to benchmark the accuracy of different MEF models that can be more widely applied. Results find clear evidence that both the dispatch and statistical MEF models considered can accurately predict the changes in emissions that are caused by changes in net load for times and places where data exist to compare. Further, prior research estimating the potential impact of optimizing load in response to MEFs has focused on BA-level or even larger regional MEF models [7, 16, 23, 41]. We find that the emissions reduction potential of nodal MEF models is larger than BA-level models, which suggests that prior estimates of the impact potential of load shifting based on an MEF signal are most likely an underestimate. Further impact is possible through the use of more geographically granular MEF models. However, the same is not true for more geographically granular AEF models.

Some limitations of the current study are that it is limited to one BA and is only able to estimate benchmark MEFs at nodes with significant wind generation. Repeating this study in other BAs is possible if disaggregated wind generation data is made available. However, generalizing to nodes without wind generation will require further development of a new method to obtain a well-identified regression. Additionally, this study was performed in a BA where electricity interchange is minimal. To apply this method to BA with marginal interchange, additional data and controls will be required.

This study primarily focused on the timing of MEF signals and their ability to predict specific times or locations where the MEF is significantly lower, primarily due to curtailment. This study is primarily relevant to applications where electricity consumption can be varied in time, such as flexible loads or grid connected batteries. Due to the fact that this study was limited to nodes with wind generation in ERCOT, the mean MEF over multiple years did not significantly vary between locations. This means we were not able to thoroughly evaluate the impact of location on siting new generation or load. A further study incorporating more locations with a wider variation in annual average MEF would be needed to better understand these models suitability for load or generation siting decisions. This future study should also consider the long term marginal impacts, which the current one does not.

This methodology is based on publicly available data from ERCOT and can serve as a check on validity for any MEF model, particularly those with sub-regional predictions, regardless of method. This is important as a good estimate of the MEF can drive significant emissions reduction in a variety of applications, such as renewable energy siting, grid battery storage, industrial load-shifting and hydrogen production.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work is part of the VERACI-T collaboration's efforts for validating emission rates and was funded in part by Meta and Microsoft. The authors would like to thank the members of VERACI-T, particularly Valentin Verdier and Avi Allison, for their feedback on early drafts of this work.

Conflict of interest

N S, P C, J C and G M are all employees of WattTime, a 501(c)(3) non-profit which developed the statistical MEF model evaluated in this work. S S is an employee of REsurety a company which developed the dispatch MEF model evaluated in this work. The method used in this work to produce a causally clean estimate of MEFs using wind generation, is independent from the models and methods used by either organization.

ORCID iDs

Nat Steinsultz https://orcid.org/0009-0008-8198-4246
Pierre Christian https://orcid.org/0000-0001-6820-9941
Joel Cofield https://orcid.org/0009-0004-2092-6748
Gavin McCormick https://orcid.org/0000-0002-0570-2824
Sarah Sofia https://orcid.org/0000-0001-7471-7133

References

[1] Daniels L, Gross B, Lewis C and Stone L 2022 *More EVs, Fewer Emissions* (Rocky Mountain Institute) (available at: https://rmi.org/insight/more-evs-fewer-emissions)
[2] Xing J, Acun B, Sundarrajan A, Brooks D, Chakkaravarthy M, Avila N, Wu C-J and Lee B C 2023 Carbon responder: coordinating demand response for the datacenter fleet (arXiv:2311.08589)

[3] Radovanović A *et al* 2023 Carbon-aware computing for datacenters *IEEE Trans. Power Syst.* **38** 1270–80
[4] Dodge J Prewitt T Tachet Des Combes R Odmark E Schwartz R Strubell E Luccioni A S Smith N A DeCario N Buchanan W 2022 Measuring the carbon intensity of AI in cloud instances *2022 ACM Conf. on Fairness, Accountability and Transparency* (Seoul Republic of Korea) pp 1877–94
[5] Jha R Lee S Iyengar S Hajiesmaili M H Irwin D Shenoy P 2020 Emission-aware energy storage scheduling for a greener grid *Proc. 11th ACM Int. Conf. on Future Energy Systems* (Virtual Event Australia) pp 363–73
[6] 2021-2022 SGIP impact evaluation 2024 (available at: www.cpuc.ca.gov/-/media/cpuc-website/divisions/energy-division/documents/self-generation-incentive-program/sgip-2021-2022-impact-evaluation.pdf)
[7] Fell H Johnson J X 2020 Regional disparities in emissions reduction and net trade from renewables *Nat. Sustain.* **4** 358–65
[8] Brown P R O'Sullivan F M 2020 Spatial and temporal variation in the value of solar power across united states electricity markets *Renew. Sustain. Energy Rev.* **121** 109594
[9] 2023 Salesforce to buy 280,000 MWh of renewable energy to increase clean power access in emerging markets (available at: www.salesforce.com/news/stories/salesforce-drec-energy-investment/)
[10] Rivian and the nature conservancy unite to transform starfire coal mine into kentucky's largest renewable power project (available at: www.nature.org/en-us/newsroom/brightnight-rivian-tnc-starfire/)
[11] Oster J Putting carbon matching into practice 2023 (available at: www.emissionsfirst.com/post/putting-carbon-matching-into-practicewe)
[12] Gagnon P J Cole W 2022 Planning for the evolution of the electric grid with a long-run marginal emission rate *iScience* **25** 103915
[13] Hawkes A D 2014 Long-run marginal CO_2 emissions factors in national electricity systems *Appl. Energy* **125** 197–205
[14] Deetjen T A Azevedo I L 2019 Reduced-order dispatch model for simulating marginal emissions factors for the united states power sector *Environ. Sci. Technol.* **53** 10506–13
[15] Siler-Evans K Azevedo I L Morgan M G 2012 Marginal emissions factors for the U.S. electricity system *Environ. Sci. Technol.* **46** 4742–8
[16] Callaway D Fowlie N McCormick G 2017 Location: the variable value of renewable energy and demand-side efficiency resources *J. Assoc. Envir. Resour. Econ.* **5** 160–208
[17] Palmer K, Prest B, Iler S and Villanueva S Options for EIA to publish CO_2 emissions rates for electricity
[18] Rudkevich A, Ruiz P and Carroll R 2011 Locational carbon footprint and renewable portfolio policies: a theory and its implications for the eastern interconnection of the US *44th Hawaii Int. Conf. on System Sciences* (*Kauai, HI, USA, 4–7 January 2011)* pp 1–12
[19] Wang Y Wang C Miller C McElmurry S Miller S Rogers M 2014 Locational marginal emissions: analysis of pollutant emission reduction through spatial management of load distribution *Appl. Energy* **119** 141–50
[20] 2021 PJM interconnection, marginal emission rate–a primer (available at: https://pjm.com/-/media/etools/data-miner-2/marginal-emissions-primer.ashx) (Accessed 31 October 2023)
[21] Evans M C *SGIP GHG signal working group final report*
[22] Zohrabian A Mayes S Sanders K T 2023 A data-driven framework for quantifying consumption-based monthly and hourly marginal emissions factors *J. Cleaner Prod.* **396** 136296
[23] Kaffine D T McBee B J Lieskovsky J 2013 Emissions savings from wind power generation in texas *Energy J.* **34** 155–76
[24] Koebrich S, Cofield J, McCormick G, Saraswat I and Steinsultz N 2023 Towards objective evaluation of the accuracy of marginal emissions factors SSRN 4631565 (available at: https://papers.ssrn.com/abstract = 4631565) (Accessed 6 June 2024)
[25] Ryan N A, Johnson J X and Keoleian G A 2016 Comparative assessment of models and methods to calculate grid electricity emissions *Environ. Sci. Technol.* **50** 8937–53
[26] Elenes A G N Williams E Hittinger E Goteti N S 2022 How well do emission factors approximate emission changes from electricity system models? *Environ. Sci. Technol.* **56** 14701–12
[27] Christian P, Cofield J, Koebrich S and McCormick G 2024 Causal emissions factor benchmarks using nuclear outages *SSRN* 4748952 (available at: https://ssrn.com/abstract = 4748952) (Accessed 6 June 2024)
[28] Fell H, Kaffine D T and Novan K E 2021 Emissions, transmission, and the environmental value of renewable energy *Am. Econ. J.: Econ. Policy* **13** 241–72
[29] United States Environmental Protection Agency 2023 Power sector emissions data (v1.2.126) (Office of Atmospheric Protection, Clean Air Markets Division) (available at: https://campd.epa.gov/) (posted online 15 August 2023)
[30] Labeling of electricity (available at: www.ercot.com/mktinfo/retail/electric)
[31] Real-time Operating Grid - U.S. Energy Information Administration (EIA) (available at: www.eia.gov/electricity/gridmonitor/index.php)
[32] ERCOT, Current Protocols - Nodal 2024 (available at: www.ercot.com/mktrules/nprotocols/current) (Accessed 17 January 2024)
[33] EIA 2023 A case study of transmission limits on renewables growth in texas (available at: www.eia.gov/electricity/markets/quarterly/archive/2023/transmission_limits_07_2023.pdf) (Accessed 17 January 2024)
[34] Oates D L and Spees K 2021 Locational marginal emissions: a force multiplier for the carbon impact of clean energy programs (available at: https://resurety.com/locational-marginal-emissions/) (Accessed 19 April 2024)
[35] Methodology + Validation (available at: https://watttime.org/data-science/methodology-validation/)
[36] Henry hub natural gas spot price (Dollars per Million Btu) (available at: www.eia.gov/dnav/ng/hist/rngwhhdm.htm)
[37] 2019-2020 IRP events and materials (available at: www.cpuc.ca.gov/industries-and-topics/electrical-energy/electric-power-procurement/long-term-procurement-planning/2019-20-irp-events-and-materials)
[38] United States Environmental Protection Agency, Emissions & Generation Resource Integrated Database (eGRID) 2021 (Office of Atmospheric Protection, Clean Air Markets Division) (available at: www.epa.gov/egrid)
[39] Carson R T and Novan K 2013 The private and social economics of bulk electricity storage *J. Envir. Econ. Manage.* **66** 404–23
[40] Holland S P and Mansur E T 2007 Is real-time pricing green? The environmental impacts of electricity demand variance (https://doi.org/10.3386/w13508)
[41] Holland S P Kotchen M J Mansur E T Yates A J 2022 Why marginal emissions are not decreasing for US electricity: estimates and implications for climate policy *Proc. Natl Acad. Sci.* **119** e2116632119