NeuralMOVES: Extracting and Learning Surrogates for Diverse Vehicle Emission Models

E. Ramirez-Sanchez[a,*], C. Tang[a], Y. Xu[b], N. Renganathan[a], V. Jayawardana[a] and C. Wu[a]

[a] Massachusetts Institute of Technology, Cambridge, MA, USA
edgarrs@mit.edu, cattang@mit.edu, npreng@mit.edu, vindula@mit.edu, cathywu@mit.edu,
[b] Harvard University, Cambridge, USA
yaoshengxu@g.harvard.edu
[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 30, 2024

Keywords: Surrogate modeling, deep learning, vehicle emission models, MOVES, transportation

1 INTRODUCTION

The transportation sector is the largest contributor to greenhouse gas (GHG) emissions in the US, accounting for a 28% of total CO_2 emissions (EPA, 2023). As such, it presents a critical challenge for climate change mitigation. The sector's transformation through electrification, automation, and intelligent infrastructure offers promising avenues for substantial emissions reductions (Sciarretta *et al.*, 2020, International Energy Agency, 2023, McKinsey Center for Future Mobility , 2023). However, the success of these innovations critically depends on the availability of accurate and comprehensive emission models to guide the design and deployment of new technologies.

The landscape of emission modeling is vast (Mądziel, 2023), but the Motor Vehicle Emission Simulation (MOVES) (USEPA, 2022), provided and maintained by the Environmental Protection Agency (EPA), serves as the official and state-of-the-science emission model in the U.S. Despite its established use, MOVES is primarily tailored for specific governmental applications, such as State Implementation Plans and Conformity Analyses. Its complexity, steep learning curve, and high computational demands pose significant challenges for users outside of its intended governmental context and trained practitioners. Furthermore, MOVES operates on a macroscopic level, making it unsuitable for the microscopic analyses required by many Intelligent Transportation Systems (ITS) applications.

Multiple ITS technologies, include eco-driving (Mintsis *et al.*, 2020), stop-and-go wave mitigation, variable speed limit optimization (Zegeye *et al.*, 2010), among many others, necessitate microscopic emission models that can compute, in real-time, emissions from a single action taken at each time step for a specific vehicle and environment. Existing adaptations of MOVES have been developed aiming to bridge MOVES' emission modeling with applications that require faster, programmatic and microscopic processing, such as MOVES-Matrix (Liu *et al.*, 2016) and MOVEStar (Wang *et al.*, 2020). These efforts have attempted to address these needs by either pre-calculating emissions data or simplifying the MOVES framework. However, these variants either require extensive storage space and MOVES expertise, or sacrifice accuracy and comprehensiveness for accessibility and speed, leaving a gap for ITS applications that demand both detailed environmental data and user-friendly modeling capabilities.

In response, this paper presents NeuralMOVES, a new generation of MOVES surrogate as an effort to get microscopic models that are diverse enough (i.e. with comprehensive set of scenario parameters) to capture real-world conditions, accurate enough to serve as a valid MOVES substitutes, and lightweight enough to run in real time and be accessible for all types of users and use cases. By employing reverse engineering to extract detailed emissions data from MOVES and applying machine learning techniques, NeuralMOVES offers a surrogate model that combines the comprehensive scenario parameters of MOVES with the real-time, microscopic modeling capabilities required by ITS technologies.

The implications of our work are twofold: our models simplify GHG emission evaluation in transportation-related analyses by providing a faster, programmatic alternative to MOVES and enable control and optimization approaches by offering microscopic and environment feature-rich models compared to alternative models.

2 METHODOLOGY

Our methodology encompasses three primary phases: data collection through reverse engineering, surrogate model development, and validation.

Data Collection: We devised a reverse-engineering approach to MOVES, generating a comprehensive dataset by extracting instantaneous emissions data. This was achieved by designing custom trajectory inputs into MOVES to isolate emissions attributable to specific vehicular actions.

Model Development: We then constructed surrogate models using machine learning techniques to fit the instantaneous emissions data. including neural networks and decision trees. These models were designed to capture the complex relationships between vehicle dynamics and environmental conditions, thereby enabling precise emission predictions at a microscopic level.

Validation: The surrogate models were validated against MOVES using a set of diverse driving trajectories. This process involved comparing the emissions computed by our models with those obtained from MOVES across various vehicle types and environmental conditions. The validation helped in refining the models and confirming their accuracy and reliability.

3 RESULTS

The key results from the study are as follows:

Instantaneous Emission Dataset: Through reverse engineering moves, extensive instantaneous emission data were extracted. The study successfully generated a dataset with over 121 million data points mapping various vehicle and environmental parameters to instantaneous emissions. This data is the backbone and ground truth used, and the surrogate models are function approximations to replicate and interpolate the data.

Diversity Importance: Figure 1 shows the relationship between emissions and factors such as vehicle age, type, road grade, temperature, and humidity. The analysis revealed that the diversity of emission model is paramount, as parameters that are not usually considered in reduced-order models affect emissions greatly. Newer vehicles tend to emit less, and road grade significantly affects emissions, with some grades leading to emissions four times higher than others. Weather conditions also influenced emissions, although the maximum variation observed was about 10%.

Surrogate Model Accuracy: The surrogate models achieved a mean absolute percentage error (MAPE) of 6.013% when compared to MOVES across more than 2 million scenarios. Figure 2 shows a detailed breakdown of error distributions across different model dimensions.

Model Architecture: Various machine learning models were explored (Table 1), with neural networks showing the best performance. The optimal neural network architecture consisted of three layers with a hyperbolic tangent activation function and a hidden dimension of 64.

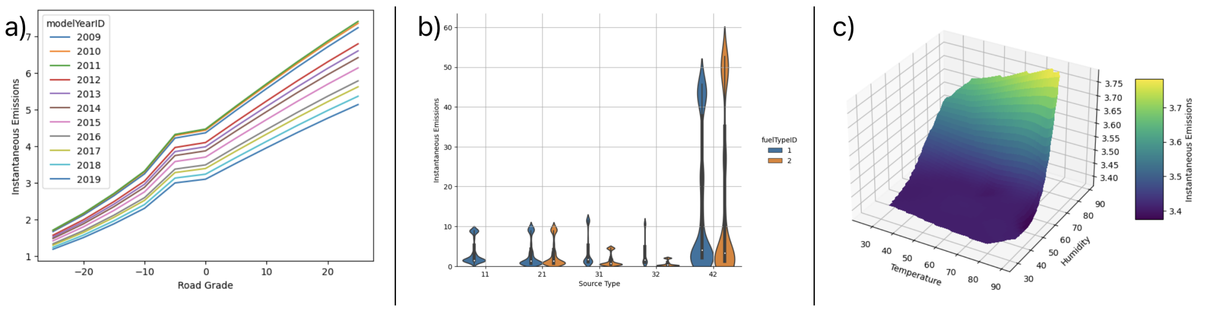

Figure 1 – *Variations of raw instantaneous emissions extracted by reverse-engineering MOVES across a) Vehicle Age and Road Grade; b) Type of Vehicle (Source Type) and Fuel Type; and c) Temperature and Humidity. Results highlight the significant impact on parameters that reduce-order emission model tend to ignore. Road grade, in particular, shows a variability of around 500% across different road grades. On the other hand, temperature and humidity shows a modes 10% variability in emissions, suggesting that emissions estimations can be extended across different regions with small correction factors.*

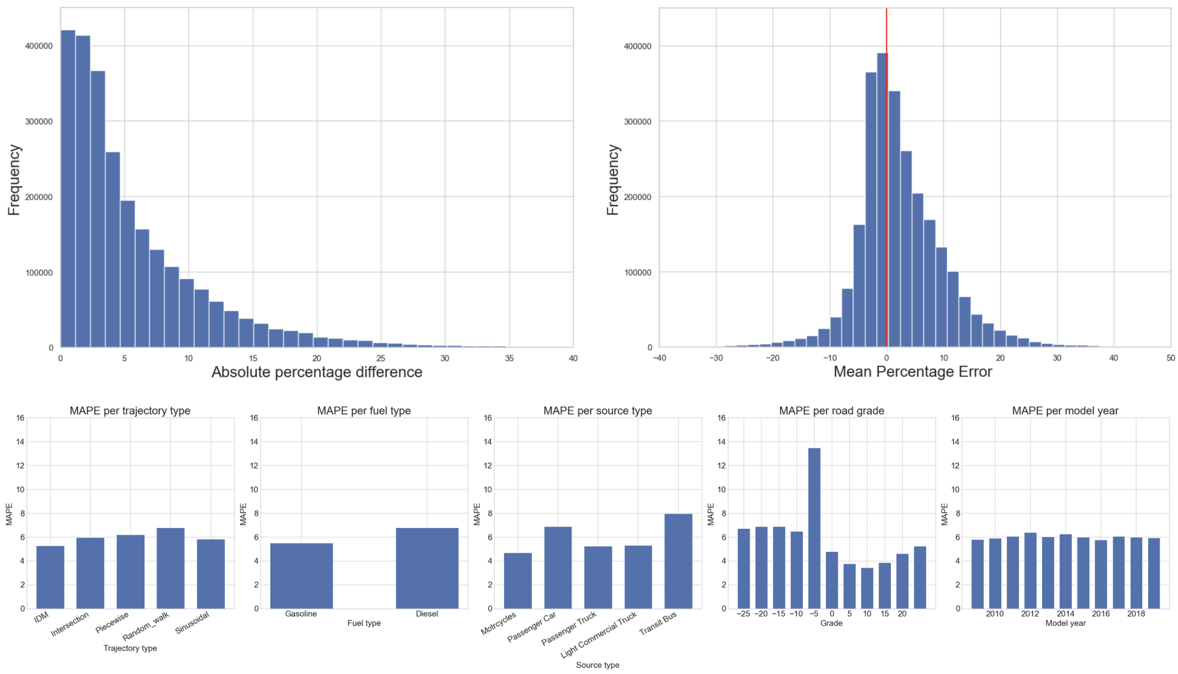

Figure 2 – *Distribution of Mean Absolute Percentage Error (MAPE) across over 2 million tests, evaluating the performance of surrogate models against the MOVES standard for representative trajectories and diverse environments. The MAPE, averaging 6.013%, illustrates the deviation between the surrogate models' emissions estimates and those calculated by MOVES. The error distribution, centered around zero, indicates a high precision and consistent accuracy of the surrogate models in estimating emissions across diverse scenarios and trajectories. This figure highlights the models' robustness and reliability as substitutes for MOVES, suitable for both micro and macro-scale environmental analysis.*

Table 1 – *Surrogate model architectures and ablations with end-to-end error statistics.*

Model type	Architecture	Training	MAPE (%)	MPE (%)	MdPE (%)	StdPE
Polynomial	3rd order		31.04	8.9	5.22	50.16
Decision Tree	Depth 50		6.17	5.37	3.78	**7.75**
NN	2 layers, dim 5	11 epochs	149.73	124.43	65.64	165.17
NN	3 layers, dim 64	11 epochs	11.54	10.91	8.63	11.56
NN	2 layers, dim 5	300 epochs	7.85	6.04	4.81	9.44
NN	2 layers, dim 5 0.97 scaling	300 epochs	**6.01**	**2.46**	**1.22**	8.90

Trajectory Validation: As the final step of our validation analysis, we conducted a detailed examination of individual trajectories to gain insights into specific driving cycle properties that may influence the surrogate models' performance.

4 DISCUSSION

With a 6% mean absolute percentage end- to-end error, the surrogate models effectively capture the essence of MOVES and can serve as reliable substitutes for a wide range of applications. The lightweight and user-friendly nature of the surrogate models empowers transportation professionals and stakeholders, enabling them to reliably conduct microscopic and macroscopic (like control-based approaches and eco-driving) analyses with ease and efficiency. This level of precision signifies a substantial advancement in the field, providing a more accessible and computationally efficient alternative to the industry-standard emission model. Moreover the surrogate models' ability to accurately capture the diverse emission profiles across a wide spectrum of vehicle types, fuels, ages, road grades, and weather conditions, with over 22,000 unique profiles, enables more realistic modeling and marks a significant improvement over existing reduced-order models. **Future Directions and Considerations** include the integration of these models into traffic simulation tools and the replication of this methodology to get models for other pollutants.

References

2023. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2021.*

International Energy Agency. 2023. *Global EV Outlook 2023.*

Liu, Haobing, Xu, Yanzhi Ann, Rodgers, Michael O, Akanser, Alper, & Guensler, Randall. 2016. Improved energy and emissions modeling for project evaluation (MOVES-Matrix).

Mądziel, Maksymilian. 2023. Vehicle emission models and traffic simulators:a review. *Energies*, **16**(9), 3941.

McKinsey Center for Future Mobility . 2023. *Autonomous driving's future: Convenient and connected.* data retrieved from https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/autonomous-drivings-future-convenient-and-connected#/,.

Mintsis, Evangelos, Vlahogianni, Eleni I, & Mitsakis, Evangelos. 2020. Dynamic eco-driving near signalized intersections: Systematic review and future research directions. *Journal of Transportation Engineering, Part A: Systems*, **146**(4), 04020018.

Sciarretta, Antonio, Sciarretta, Antonio, Vahidi, Ardalan, Vahidi, Ardalan, & Vahidi, Ardalan. 2020. Energy Saving Potentials of CAVs. *null.*

USEPA. 2022 (8). *Motor Vehicle Emission Simulator: MOVES3.*

Wang, Ziran, Wu, Guoyuan, & Scora, George. 2020. *MOVESTAR: An Open-Source Vehicle Fuel and Emission Model based on USEPA MOVES.*

Zegeye, S. K., De Schutter, B., Hellendoorn, J., & Breunesse, E. A. 2010. Variable speed limits for area-wide reduction of emissions. *Pages 507–512 of: 13th International IEEE Conference on Intelligent Transportation Systems.*