

特集：DX/AI の活用

02 ビッグデータと生成 AI を活用した市場の燃費分布予測技術**Prediction of Market Fuel Economy Distribution
Using Big Data and Generative AI**波頭 佑哉^{*1} 高原 勇樹^{*2} 久間 裕太郎^{*3}
Yuya Hato Yuki Takahara Yutaro Kyuma**要 約**

マツダでは、燃費性能や空調快適性などの各性能を共立した車両を開発するために、車両を構成する各ユニットを1次元で表した1D車両システムモデルによるシステム全体の最適化のMBD (Model-Based Development) 手法を構築してきた。従来の本モデルによる検討では、認証などで使われる決められた走行モードでの解析を主体としてきたが、今後のカーボンニュートラルの実現に向けては、車両のライフサイクル全体でのCO₂排出量の削減が重要となる。そのため、実際のお客様の車両の使い方 (リアルワールド) を想定した解析により、市場の実用燃費分布を予測することが求められる。そこで、市場のビッグデータに生成AI (Artificial Intelligence) のGAN (Generative Adversarial Networks, 敵対的生成ネットワーク) を応用することで、お客様の使い方の特徴を再現した走行シナリオを生成する手法を構築した。そして一連の予測技術の有用性を確認するために、生成した走行シナリオを車両システムモデルの入力として解析を行い、実測の市場の燃費分布に近い分布をモデルで予測できることを検証した。

Abstract

Mazda has developed an MBD method for comprehensive system optimization to develop a vehicle which combines performances including fuel economy and air-conditioning comfort, utilizing a 1D vehicle system model that represents each unit that makes up the vehicle in one dimension. To achieve carbon neutrality in the future, it is necessary to conduct an analysis of the practical fuel economy distribution in the market, considering the actual usage patterns of customers (real world), although the conventional model study focused on the analysis in a driving mode used for certification. Therefore, by applying Generative Adversarial Networks (GAN) to big data in the market, we built a method to generate driving scenarios that replicate the characteristics of customer usage. We then analyzed the generated driving scenarios as input to the vehicle system model and verified the effectiveness of this method by comparing fuel economy distributions based on actual measurements and model predictions.

Key words : Vehicle development, CAE, Fuel economy, Machine learning, Big data, Real-world**1. はじめに**

マツダでは、2050年のカーボンニュートラルの実現に向けて電動化に取り組んでおり、規制動向や消費者のニーズ、社会インフラなどの変化に柔軟に対応できるよう、段階的に電動化を進める戦略を推進している⁽¹⁾。電動車両では、エンジンやトランスミッションに加えて、LiB (Lithium-ion Battery) やモーターなどの電動ユニット、更には空調ユニットまでの個々の適温領域が異なる

ため⁽²⁾、電動化によるCO₂排出量の削減、すなわち燃費性能の向上には、車両内の熱流れを協調制御する熱マネージメント (以下、熱マネ) が重要である。しかし、ユニットの配置や加熱・冷却方式が多岐にわたるため、各ユニットの単独モデルを用いた個別最適化のMBD手法では最適なシステムの検討に限界がある。そこで、車両を構成する各ユニットの熱、電気、機械の機能を1次元 (以下、1D) で表した車両システムモデル (Fig. 1) を構築し、V字開発プロセスの左バンクでの熱マネシス

*1 MBD 革新部
MBD Innovation Dept.

*2,3 走行・環境性能開発部
Drivability & Environmental Performance
Development Dept.

テムの全体最適化の MBD 手法を開発してきた⁽³⁾。

これまでの本モデルによる検討では、WLTC (Worldwide harmonized Light vehicles Test Cycle) などの既存の走行モードでの解析を主体としてきた。一方で、前述のカーボンニュートラルの実現には、車両のライフサイクル全体での CO₂ 排出量の削減が重要となるため、走行シーンにおいては実際のお客様の使い方（本稿では、リアルワールドと定義）を想定した解析により、市場での実用燃費のばらつき度合いを予測することが求められる。しかし、お客様の車両の使い方は多様なため、人の手によるリアルワールドの走行シナリオの設定が困難である。そこで、市場の車両から入手可能なビッグデータに生成 AI を適用して、お客様の使い方の特徴を再現した走行シナリオを生成する手法を構築した。そして、生成したシナリオを車両システムモデルの入力として解析を行い、実測の燃費分布と比較することで一連の予測技術の有用性を検証した。

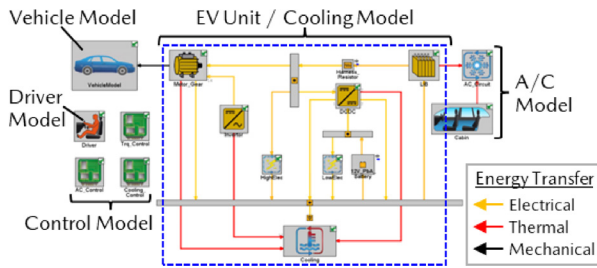


Fig. 1 1D Vehicle System Simulation Model

2. リアルワールドの走行シナリオ生成

2.1 市場ビッグデータ

マツダでは、Fig. 2 のように個車の使われ方を表す燃費を始め、車速やエンジンの使用領域、空調の作動頻度などの記録（時系列ではなく、頻度や平均値などの統計情報）を蓄積した市場車両のデータベースを構築している。個車の記録は車両がディーラーに入庫した際に収集を行い、このデータベースから市場の燃費分布などをモニターする。今回は既存アセットの有効活用をコンセプトに、このデータベースの中で走行シナリオと相関性の高い、車速と加速度の使用領域を表した Fig. 3 のヒストグラムを元にシナリオ生成を行う。なおヒストグラムは、横軸が車速、縦軸が加速度となり、高頻度の領域が右上に集まるほど高車速・急加速の多い走行を経験したことを示す。

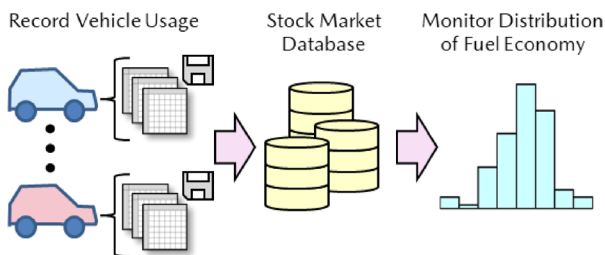


Fig. 2 Monitoring Process of Fuel Economy Distribution

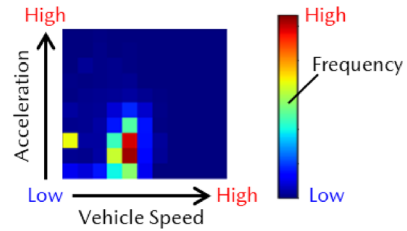


Fig. 3 Histogram of Vehicle Speed and Acceleration

2.2 走行シナリオ生成の課題

車両システムモデルに入力する走行シナリオは、車速推移を時間ごとに示した時系列形式である。前述の車速と加速度のヒストグラムは、個車の市場走行時の車速を時々刻々と集計した結果であるため、このヒストグラムから時系列の車速パターンの逆推定が可能となれば、リアルワールドを再現した走行シナリオのモデルへの入力を実現する。

一方で Fig. 4 のように、このヒストグラムに集計する際は車速と加速度の使用域を時間ごとに積算するため出力が一意に定まるが、ヒストグラムから車速パターンに変換する際は分布を成す過程が無数にあるため出力が一意に定まらない課題がある。そこで、生成 AI の GAN⁽⁴⁾ を用いて、本課題の解決を行う。

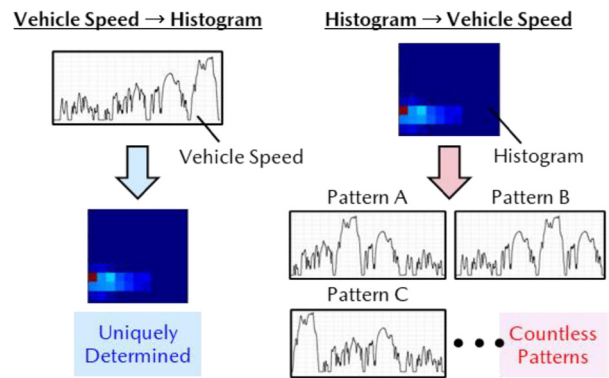


Fig. 4 Conversion between Vehicle Speed and Histogram

2.3 GAN による走行シナリオ生成手法

初めに GAN とは、Fig. 5 に示す生成器 (Generator) と識別器 (Discriminator) の 2 つのニューラルネットワーク (以下、NN) を用いる生成 AI の一種であり、画像生成などに使われている。生成器ではランダムノイズ (0~1 の乱数ベクトル) を入力として画像などの生成対象を出力し、識別器では生成器から出力した偽物のデータに対して真偽を予測する。識別器は本物のデータと生成器の偽物のデータを用いて、本物と偽物を正しく識別できるように学習を行い、一方、生成器は生成した偽物のデータを識別器が誤って本物と識別するように学習する。このように生成器と識別器が互いに競い合って学習することで、生成器が本物と同等のクオリティのデータを生成可能となる。

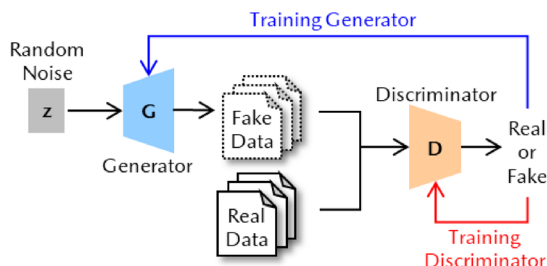


Fig. 5 Concept of GAN

以上の GAN の学習は式(1)に示す二人のプレイヤーによるミニマックス法(想定される最大の損失が最小になる選択を探索)の最適化問題となる。ここで x は本物のデータ, $p_{data}(x)$ は本物のデータの確率分布, D は識別器, z はランダムノイズ, $p_z(z)$ はランダムノイズの確率分布, G は生成器である。

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

今回はこの GAN を発展させた, CGAN (Conditional GAN, 条件付き GAN)⁽⁵⁾ を用いた車速パターンの生成手法を提案する。

CGAN では, 生成器から特定の属性の出力を可能とするために, ランダムノイズに属性を指定する条件ベクトルを結合して生成器へ入力する。よって, CGAN の学習の最適化問題は, 元の GAN の式(1)から条件ベクトル y を追加した式(2)となる。

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x|y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z|y)))] \quad (2)$$

この条件付与の特徴を用いて, Fig. 6 のようにランダムノイズに結合する条件ベクトルを前述のヒストグラムとする。これにより, ランダム性とヒストグラムの特徴を併せもつ情報を生成器へ入力することで, 特定のヒストグラムの車速パターンへの一意性のない変換を実現する。本手法を以下の手順①~⑤に示す。

① 実路を走行しながら車速を計測したデータ(2.4節)と, そのデータをヒストグラムに変換した条件ベクトルを準備する。なお, 加速度の情報は車速の微分により算出する。

② 生成器にランダムノイズとヒストグラムのベクトルを与えて, そのヒストグラムの時の車速パターンを生成する。

③ 識別器は実測した車速パターンと生成した車速パターンを用いて, 本物と偽物を識別できるように学習を行う。

④ 生成器は識別器が誤って実測の車速パターンと識別するように学習を行う。(※②~④を学習完了まで繰り返す。)

⑤ ランダムノイズと市場データベース(2.1節)から選択したヒストグラムのベクトルを, 学習が完了した生成器に与えて市場を想定した車速パターンを生成する。

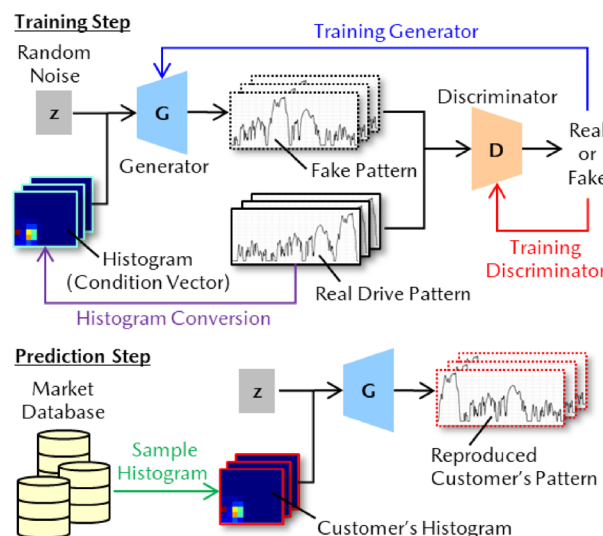


Fig. 6 Concept of Driving Scenario Generation Based on CGAN

2.4 学習データ

本手法における CGAN の学習には, 2.3 節で述べたように実路を走行した時系列の車速データが必要となる。今回は Table 1 に示す, 北米地域を対象に市街地や高速道路などの実路を走行時に計測した 45 ケースの車速データを使用する。

Table 1 Condition of Real-World Driving Data

Region	North America
Road Type	City, Highway, Suburban, Hill
Vehicle Speed Range	0~171km/h
Duration Range	2594~12283s
Distance Range	12~329km
Number of Data	45

CGAN では, 一律の長さのデータを学習してその長さの車速パターン(今回は WLTC などの走行モードと同程度の 1600 秒)を生成するが, この実路データではケースごとに長さが異なっている。また, 学習データの増強と学習する車速域に偏りを生じさせないために, Fig. 7 のように各実路データの分割を行い, それらの分割データを結合してねらいの時間長さ及び車速域の学習データを作成する。一方で, 結合するデータはランダムで選択するため, 結合箇所で車速が急変するなど, 実路に則さない学習データが生じる可能性がある。そこで, 各区間のデータをランダムに選択する際に制約を付与し, 制約を満たすまでその区間のデータの選択を繰り返し行う。

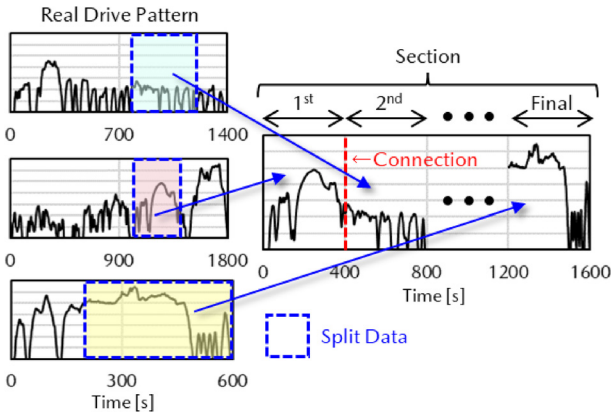


Fig. 7 Driving Pattern Creation Process for Train Data

ここで、制約の具体例を示すと、第1区間の初期値と最終区間の最終値が0であることや、第2区間以降は前区間の最終値と今回区間の初期値の差が指定した車速差の範囲内であること、などが挙げられる。以上により結合された学習データについて、最後に全区間での最高車速の確認を行う。Table 2には作成した学習データの一例と内訳を示しており、最高車速が低速から高速のデータまでそれぞれ均一な割合となるように学習データを準備した。なお、表中の v_{max} は各パターンにおける最高車速を示し、学習データは全部で500パターンとなる。

Table 2 Train Data Composition

Condition [km/h]	Example Pattern (Time vs Vehicle Speed)	Number of Data
$v_{max} < 50$		100
$50 \leq v_{max} < 70$		100
$70 \leq v_{max} < 90$		100
$90 \leq v_{max} < 110$		100
$110 \leq v_{max}$		100
Total Pattern of Train Data		500

2.5 学習結果

本手法におけるCGANの学習結果を以下に述べる。Fig. 8では、生成器に低車速と高車速走行時のヒストグラムを入力した際に出力された車速パターンについて、学習の経過による変化を示している。図中の上部には条件ベクトルとして与えた、北米市場のデータベースからサン

プリングした低車速域と高車速域の分布をもつ個車のヒストグラム（すなわち未学習の分布）を掲載している。まず学習初期には、NNの重みの調整ができていないため、各ヒストグラムで車速パターンとは認識できない類似した結果を出力した。また学習の中盤では、ノイズが含まれるが車速のようなパターンを出力することを確認した。ただしこの時点においては、双方に明確な車速差の特徴が出ていない。一方で学習完了時には、各ヒストグラムの特徴をとらえたノイズのない本物と同様の車速パターンを出力することができた。なお、生成器に入力するランダムノイズについて、図中の検証時は学習を通して同じ値を使用している。

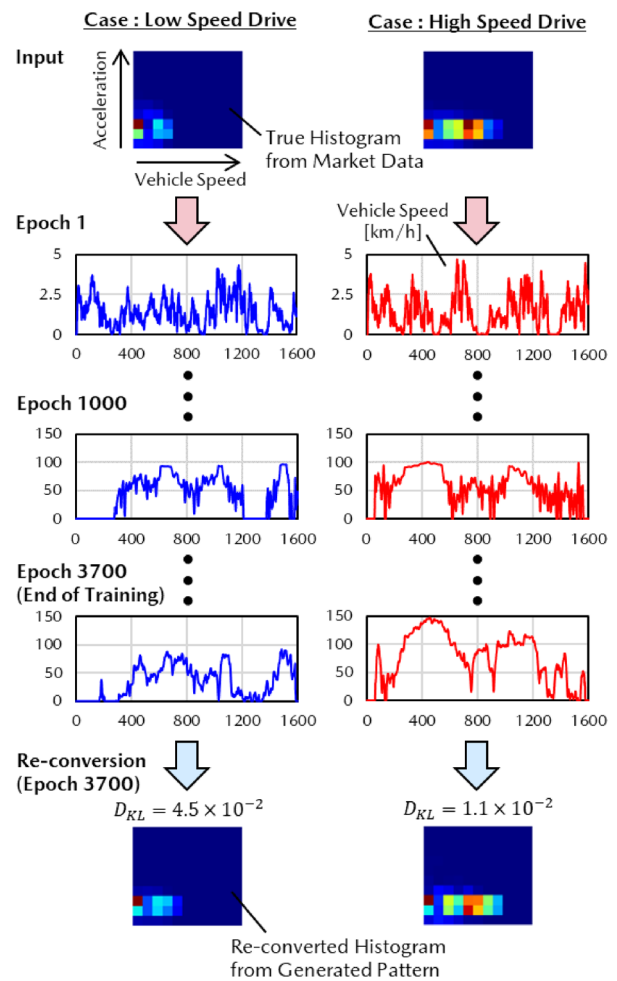


Fig. 8 Training Result of Driving Scenario Generation

最後に図中の下部では、生成した車速パターンの確度を検証するために、この車速パターンを再度ヒストグラムに変換を行い、入力に用いたヒストグラムと比較した結果を示している。目視においても各ヒストグラムを再現できていることが分かるが、分布の一致度を定量的に比較するために各々のKL (Kullback-Leibler) ダイバージェンスを算出する。KL ダイバージェンスは式(3)で求めることができ、0に近いほど分布が一致していることを表す。ここで、 $p_{true}(x)$ は真の確率分布、 $q_{est}(x)$ は推定

した確率分布である。

$$D_{KL}(p_{true} \parallel q_{est}) = \sum_{i=1}^N p_{true}(x_i) \log \frac{p_{true}(x_i)}{q_{est}(x_i)} \quad (3)$$

双方の KL ダイバージェンスを算出すると、ともに 0.05 以下となり、定量的にも各ヒストグラムの再現度が高いことを確認した。以上のことから本手法によるリアルワールドを再現した走行シナリオの生成が妥当であると推察する。

なお学習対象は識別器の真偽となるため、誤差関数は二値分類に用いる式(4)の BCE (Binary Cross Entropy) 損失とし、最適化アルゴリズムには Adam⁽⁶⁾ を使用する。ここで、 Y は正解のラベル (0 または 1)、 \hat{Y} は予測のラベル (0~1 の連続値) である。

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^N (Y_i \log \hat{Y}_i + (1 - Y_i) \log (1 - \hat{Y}_i)) \quad (4)$$

GAN は 2 つの NN を学習させる特性上、通常の NN より計算が発散しやすいため、識別器の学習を 5 回実施ごとに生成器の学習を 1 回行うことで、学習の安定性を向上させた。

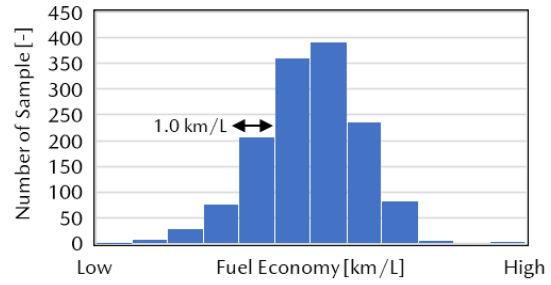
3. 1D モデルによる市場の燃費分布解析

3.1 検証データ

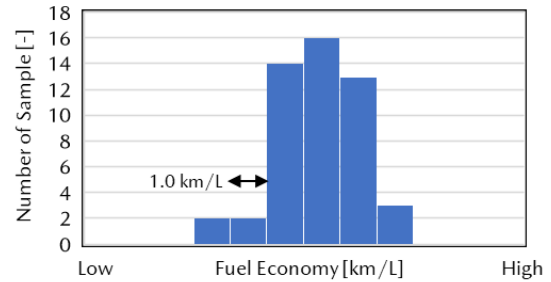
本章では、第 2 章で検討したリアルワールドの走行シナリオの生成手法を用いて、車両システムモデルでの市場の実用燃費分布解析を行う。Table 3 では、本解析に用いる検証データを示しており、2.4 節で述べたように今回は北米地域が対象のため、同地域の車両から収集したデータベースを使用する。対象車両は CX-30 の SKYACTIV-G 2.5 を搭載したモデルの AT (Automatic Transmission) 仕様とし、2022 年の 2 月~10 月に取得した全 1406 サンプルのデータから母集団と類似した燃費分布になるようにランダムで選択した 50 サンプルのデータを用いる。Fig. 9(a) は、母集団の全データにおける燃費分布であり、正規分布の傾向を示している。そして Fig. 9(b) は、選択した 50 サンプルのデータの燃費分布であり、この分布のデータと紐づく各車速と加速度のヒストグラムから車速パターンの生成を行い、モデルに入力する。

Table 3 Validation Target of Fuel Economy Distribution Prediction

Product	CX-30
Specification	SKYACTIV-G 2.5 / AT
Region	North America
Period	February ~ October, 2022
Sample Size	50 (Sampled)/1406 (Population)



(a) Distribution of Population



(b) Distribution of Sampled Data

Fig. 9 Fuel Economy Distribution of Validation Data

3.2 解析モデル

本解析に使用する車両システムモデルを Fig. 10 に示す。まずハードウェアの機能では、(1) 車両の加速度と車速を計算する車両モデル、(2) エンジンの燃料消費量を計算するエンジンモデル、(3) ギア段に応じて駆動力を変換するトランスミッションモデルの構成となる。そしてソフトウェアの機能では、(4) 目標車速の追従に必要な駆動力とペダルポジションを計算するドライバーモデル、(5) ギアの変速とエンジンのスロットルを制御する制御モデルの構成となる。また、今回の目的は車両システム内の詳細な熱挙動の解析ではなく、走行シナリオの違いによる燃費分布の再現性の検証である。そのためサーマルモデルの連成は行わず、空調は非作動、エンジンの水温及び油温とトランスミッションの油温は暖機後の温度を境界条件として設定する。なお本解析には、1D シミュレーションツールの GT-SUITE を使用する。GT-SUITE は、Gamma Technologies, LLC により開発されたソフトウェアである。

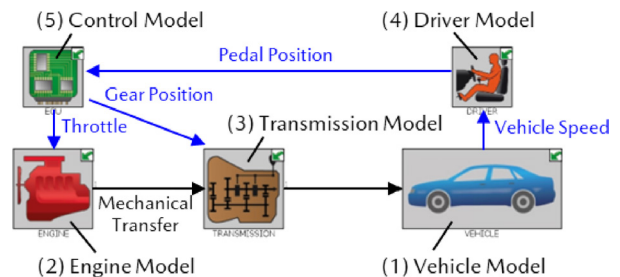


Fig. 10 1D Simulation Model for Fuel Economy Distribution Prediction

Fig. 11 では、北米の認証モードの一つである UDDS (Urban Dynamometer Driving Schedule) におけるモデルの解析精度を示している。エンジン負荷に影響するギアの変速制御について実測の挙動を再現できており、今回の解析目標である燃料消費量を精度よく予測可能なことを確認した。

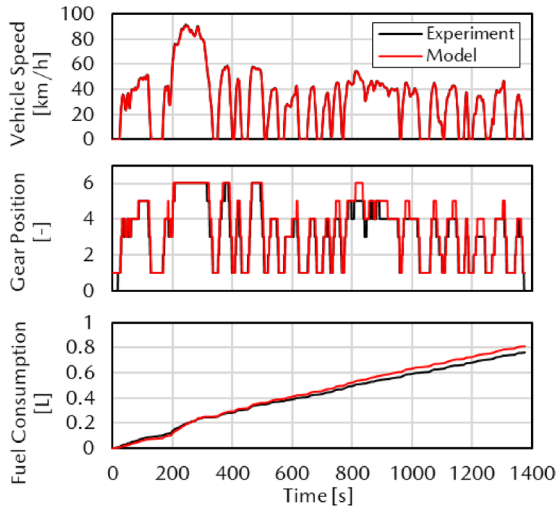


Fig. 11 Validation Result of Mode Drive Situation

3.3 解析結果

今回検討した手法を用いた実用燃費分布の解析結果を以下に示す。3.1 節で述べたように、50 サンプルのヒストグラムから生成した車速パターンをモデルに入力して、実用燃費分布の解析を行った。まず Fig. 12 では、各サンプルにおけるモデルで予測した燃費と実際の燃費を比較した結果を示す。図中の青の点線は、モデルの予測値を±10%オフセットした範囲を表しており、各サンプルの実測がこの範囲の近傍にあることが分かる。モデルでの燃費が実際の傾向をとらえられ、平均で±10%以内の誤差で予測できることを確認した。

そして Fig. 13 では、モデルで予測した燃費分布と実際の母集団の燃費分布を比較した結果を示している。燃費の平均値（分布の頂点）が0.5km/L以内の誤差と、モデルでの分布が実際の分布に近い形状となっており、本手法による市場の燃費分布の机上予測が有用であることを確認した。なお、分布の分散についてはモデルの方が小さく、実際の分布より燃費が平均値付近に集中していることが分かった。これは、今回のモデルが3.2 節で述べたようにハードの温度条件が一定で、また、空調についても非作動で統一しているため、車速パターンの違いのみが燃費の変動因子となることが理由と推察する。

個車の実用燃費と全体の燃費分布を十分に再現できており、従来は車両の量産後にしか分からなかった市場での燃費分布を、量産前の段階から把握して車両開発に活用することを実現した。今後の更なる再現性の向上には前述の温度や空調要素に加えて、乗員数及び積み荷によ

る車重の違いや、走行する路面の状況（摩擦や勾配）の考慮が考えられる。

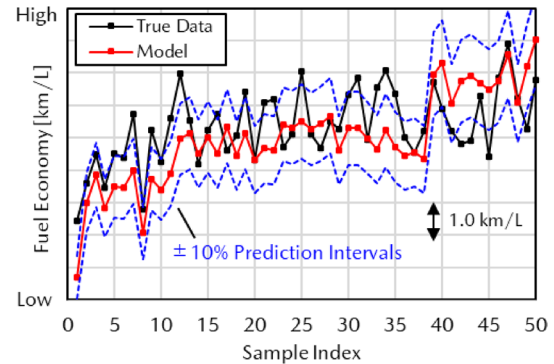


Fig. 12 Validation Result of Each Vehicle Fuel Economy

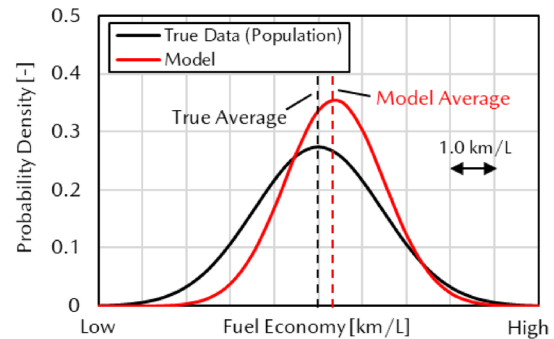


Fig. 13 Validation Result of Fuel Economy Distribution

4. おわりに

マツダでは、車両を構成する各ユニットを1次元で表した1D車両システムモデルによるシステム全体の最適化のMBD手法を構築してきた。従来の本モデルによる検討では、決められた走行モードでの解析を主体としてきたが、今後のカーボンニュートラルの実現に向けては、実際のお客様の使い方を想定した解析により、市場の実用燃費分布を予測することが求められる。そこで本稿では、以下について紹介した。

a. 市場の車両から入手可能なビッグデータに生成AIのGANの発展系であるCGANを適用することで、リアルワールドを再現した走行シナリオを生成する手法を構築した。CGANの条件ベクトルにビッグデータ内の車速と加速度のヒストグラムを与えることで、お客様の使い方を表すヒストグラムから車速パターンの生成を可能とした。

b. 市場のビッグデータのヒストグラムから生成した車速パターンを車両システムモデルの入力として解析を行い、実測と机上予測による燃費分布を比較した。個車の実用燃費と全体の燃費分布の再現性を確認し、ビッグデータと生成AI、モデルを用いた一連の予測技術によって、車両の量産前の段階から市場での燃費分布を把握することを実現した。

今後、車速パターン以外の市場での燃費の変動因子を再現する手法を構築することで本予測技術の開発を更に進め、マツダの魅力ある車両をお客様に届けながらカーボンニュートラルの実現にも貢献していきたい。

参考文献

- (1) MAZDA INTEGRATED REPORT 2023
- (2) 渋谷ほか：BEV 急速充電用次世代サーマルフルードの紹介，自動車技術，Vol.76，No.11，pp.86-92 (2022)
- (3) 波頭ほか：1D 車両全体モデルを活用した EV 熱マネジメントシステムの検討，[マツダ技報](#)，No.39，pp.206-212 (2022)
- (4) I. Goodfellow et al.: Generative Adversarial Nets, In Conference on Neural Information Processing Systems (2014)
- (5) M. Mirza et al.: Conditional Generative Adversarial Nets, arXiv preprint arXiv, 1411.1784 (2014)
- (6) D. Kingma et al.: Adam: A method for stochastic optimization, In International Conference for Learning Representations (2015)

■著者■



波頭 佑哉



高原 勇樹



久間 裕太郎