

# Digital Twinning From Vehicle Usage Statistics for Customer-Centric Automotive Systems Engineering

KUNXIONG LING <sup>1,2</sup>

<sup>1</sup>Ingolstadt School of Management, Catholic University of Eichstätt-Ingolstadt, 85049 Ingolstadt, Germany

<sup>2</sup>Research and Innovation Center, BMW Group, 80788 Munich, Germany

CORRESPONDING AUTHOR: K. LING (e-mail: kunxiong.ling@bmw.de)

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the "Open Access Publication Funding" Program under Project 512640851.

---

**ABSTRACT** Towards customer-centric automotive systems engineering, it is essential to incorporate physical models and vehicle usage behavior into decision support systems (DSSs). Such DSSs tend to apply digital twin concepts, where simulations are parameterized with fine-grained time-series data acquired from customer fleets. However, logging vast amounts of data from customer fleets is costly and raises privacy concerns. Alternatively, these time-series data can be aggregated into vehicle usage statistics. The feasibility of creating digital twins from these vehicle usage statistics and the corresponding DSSs for systems engineering is yet to be established. This paper aims to demonstrate this feasibility by proposing a DSS framework that integrates four key elements of digital twinning: aggregate usage statistics from customer fleets, logging data from testing fleets, physical models for vehicle simulation, and evaluation models to derive decision support metrics. The digital twinning involves a four-step process: pre-processing, profiling, simulation, and post-processing. Based on a real-world fleet of 57110 vehicles and four evaluation metrics, a proof of concept is conducted. Results show that the digital twin covers the evaluation metrics of 99% of the vehicles and reaches an average fleet twinning accuracy of 91.09%, which indicates the feasibility and plausibility of the proposed DSS framework.

**INDEX TERMS** Customer centricity, decision support system, digital twin, fleet data, systems engineering.

---

## I. INTRODUCTION

ENTERPRISES transform their product-oriented strategies into customer-centric ones, oriented towards the needs of existing or prospect customers [1]. Therefore, the behavior of diverse customers must be synthesized – often coined customer profiling [2]. In business analytics, decision support systems (DSSs) are essential for processing huge amounts of customer data, as those allow for the integration of data capturing, data processing pipelines and analytical techniques for customer profiling and other purposes [3].

While business analytical systems have been primarily focused on business management tasks, a rising demand for decision support for the development of complex

technical systems can be observed. In the automotive industry, Holtmann et al. [4] point out that it is challenging to determine the requirements and the validation targets based on seamless models and customer usage behavior. Regarding the development of automotive systems which usually includes motors, transmissions, batteries etc., there is increased complexity to be understood as of the penetration with electromobility and diverse variants and combinations that have been pushed out, resulting in a high degree of flexibility or complexity of the whole system's components and their usage-dependent developments [5]. According to Nies et al. [6], there is high research interest in determining requirements of customers when using vehicles with complex powertrain systems, e.g., plug-in hybrids which can be individually powered by fuel or electricity. In addition, the cruciality of a DSS for automotive systems engineering is

The review of this article was arranged by Associate Editor Johannes Betz.

highlighted in evaluating validation programs [7], optimizing operating strategies [8], and connected automated driving [9].

Brusa [10] pointed out that for model-based systems engineering, the digital twin is essential to integrate the system design and assessment of reliability-related aspects. To integrate the aspects of models and data in a DSS, therefore, the concept of digital twins, i.e., simulated virtual representations for real physical products with connectivity, has been pioneered in complex product and service design [11]. According to Rasheed et al. [12], the main challenge to build a digital twin is the integration of physical models with data, expert knowledge, and mathematics.

Apart from system integration, it also remains challenging to acquire and feed customer data into physical models with minimum costs and preservation of customer privacy. Since the last few decades, acquiring sensor data from customer fleets has been extensively spread, providing long-term measurement values for customer analytics [13]. With the rapid development of mobile communications technology, sensor data, such as on-board signals, are expected to be logged. For example, Sass et al. [14] demonstrate the potential for predicting component aging with logged signals. However, Wilberg et al. [15] highlight that performing massive data logging for customer fleets may result in privacy violations.

Surprisingly, Esser et al. [16] find that aggregation of time-series data can preserve customer privacy, as the sequential information can remedy privacy concerns. Thereby, a sustainable solution is to return to the accessible long-term statistical data from various control units but with flexible data aggregation using vehicular telemetries [17]. For huge amount of customer fleets throughout their corresponding product lifetime, the costs of extra data loggers and massive storage of time-series data could also be saved [18]. So far, it remains unclear if it is feasible to build digital twins based on vehicle usage statistics aggregated from customer fleets and to industrialize the approaches into a DSS for customer-centric automotive systems engineering.

To address the research gap, this work will:

- develop a DSS framework for customer-centric automotive systems engineering using digital twins.
- build and apply digital twins for multiple customer markets via usage profiling and system simulation using aggregate data from customer fleets and logging data from in-house testing fleets.
- build a proof-of-concept (POC) DSS using the framework for predicting the lifetime metrics of plug-in hybrid engines in three market regions.
- evaluate the feasibility of the POC by comparing the predicted metrics to the real-world metrics acquired from vehicle diagnostics.

Targeting to the intersection of automotive DSSs, digital twins, and customer centricity, this paper has been one of the first attempts to establish a comprehensive decision support framework which provides system simulation inputs from customer fleet usage statistics and delivers quantitative

metrics computed from the simulation results. Based on this framework, the digital twinning method consolidates customer centricity in automotive systems engineering processes. For product development, system or component requirements can be quantitatively derived, to name a few. For aftersales maintenance, individual vehicle lifetime can be predicted. With the help of the twinning performance and scalability, more advanced use cases are presented in this paper, including requirement localization and recall prioritization.

The remainder of this paper is structured as follows. In Section II, the related work in automotive digital twins and DSSs is reviewed. The research gap and the position of this work are identified. Then, the architecture of the proposed DSS and detail on this digital twinning approach is introduced, including pre-processing, profiling, simulation, and post-processing. In Section IV, the system configuration of a POC is presented, followed by a case study for evaluating the plausibility and the practicality of the DSS. Section V highlights the scientific contribution and discuss the implication of this work. Finally, this paper concludes and outlines promising research directions in Section VI.

## II. RELATED WORK

In the context of automotive systems engineering, a digital twin is a digital replica which reflects the operational states of customer fleets by acquiring sensor data from fleets and provides the physical relationship by modeling the vehicle systems [19]. The physical or data-driven models connected to the sensor data could be used for decision support by determining direct or indirect actions back to the products, such as product lifecycle management or developing next generation products.

Digital twins have been widely investigated and applied in the field of systems engineering for the automotive or transportation industry. Undoubtedly, digital twins can be used for design optimization. Gu et al. [20] converted requirements to customize engineering design of elevators. Li et al. [21] applied digital twins to optimize the energy management strategy for plug-in hybrid electric vehicles (PHEVs) while fusing cyber-physical operational data. Also, digital twins can profit operational excellence. González et al. [22] evaluated the operational states of elevators using digital twins with various scale of physical-based models. Sun et al. [23] applied dynamic digital twins to allocate resources in the aerial-assisted Internet of Vehicles. Thonhofer et al. [9] proposed a system architecture of a digital twin to provide decision support for automated driving functions. Furthermore, for product lifecycle management, digital twins provide the connection and interaction between the behavior of physical entities and know-hows using virtual simulation. Ren et al. [24] predicted the abnormal temperature for diesel locomotives using a digital twin driven by machine learning. Qin et al. [25] combined neural networks and physics-based thermal-mechanical fault dynamic model to predict the defect evolution for rolling bearings. Rodríguez et al. [26] estimated

the thermal behavior of inverters in electric vehicles by feeding numerical models with sensor data during test and operation.

Compared to model-based systems engineering processes, Boschert and Rosen [27] summarized the key characteristics of digital twin-based product engineering in two aspects: (i) boundary conditions and simulation inputs come from the real world (customers); (ii) simulation outcomes contribute in operational seamless product lifecycle management. Therefore, in addition to data, models, and twinning approaches to build digital twins, a digital twin requires bi-directional connectivity between real products and virtual models.

Although Rodríguez et al. [26] proposed a co-simulation framework which enables bi-directional connectivity between the inverter and its digital twin, the majority of related research works above focuses on building the digital replica by machine learning [20], [24], optimization [21], [23], simulation [9], [22], [28], or combination of the methods [25]. Considering the research objective, this work focuses on building the digital twin using optimization and simulation.

Typically, digital twins are real-time capable [11]. In the context of this paper, utilizing aggregate fleet data means the data acquisition does not keep the time-series signals and accordingly loses the real-time capability. However, according to VanDerHorn and Mahadevan [29], real-time capability does not affect the qualification of digital twins. Despite the definition of digital twinning, its use cases, such as design optimization, operational excellence, and product lifecycle management, usually require no real-time actions and continuous feedback. Therefore, without real-time capability, it is still possible to build digital twins based on aggregate fleet data. Correspondingly, the bi-directional connectivity could be enabled by vehicular telemetries and over-the-air updates, which are the state-of-the-art in the automotive industry [14], [17].

Depending on if the bi-directional connectivity would take place automatically and vice versa, there are two similar concepts: digital model and digital shadow. If both directions of connections occur manually, it is a digital model. If the data flow from real world to the digital replica automatically but backwards manually, then it is a digital shadow [30]. Only with fully automatic capability between the two parts, the replica can be classified as a digital twin. In the context of this paper, digital shadows are to be built by reconstructing the aggregate vehicle usage statistics into time-series profiles. With simulation, evaluation and the DSS, it is feasible to impact the physical fleets automatically and build holistic digital twins.

To deploy digital twin technologies for large-scale industry applications, solution processes require digitalization and proper automation with the help of information systems such as DSSs. Respective DSSs can be broadly classified in data- or model-based. Data-based DSSs have

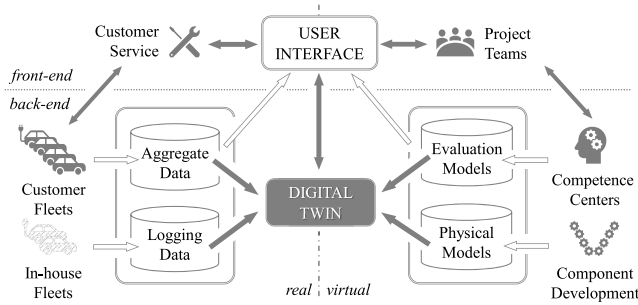
been applied successfully for manufacturing systems [31], service development of business models [32] and product development in wind energy [33]. As the complex physical relationships in automotive systems are hard to be integrated into pure data analytical models, model-based DSSs have been widely used in the automotive industry. Mennenga et al. [34] applied DSSs for lifetime-based fleet planing for electrified vehicles. Kloör et al. [35] repurposed batteries for electric vehicles using a model-based DSS. However, pure model-based DSSs have limitations as no operational data is provided and cannot represent the usage behavior of real fleets. To tackle the problem, Xie et al. [19] developed a DSS using digital twins with real-time capability and applied it to a vehicle body control system. It enables the connection between models and live signal data from the control units, whereas other entities around the vehicles, such as the driver and the environment, remains to be considered. Hereby, Wang et al. [36] developed a mobility digital twin framework as a DSS which integrates humans, vehicles, and traffic environments. Nevertheless, for complex modeling assessments and decision supports that are infeasible to be integrated inside the vehicle control units, novel DSSs have to consider further aspects.

For customer fleets with private cars, car manufacturers suffer from logging and monitoring large amounts of customers throughout the lifetime of vehicles due to cost issues and privacy preservation. To enhance the security of timely DSS, machine learning-driven twinning methods could be distributed on-board. For the security of data exchange, blockchain technologies are investigated and incorporated [37], [38]. Considering the complexity of physical simulation models, however, central twinning solution is more suitable in the context of automotive systems engineering.

Interestingly, Esser et al. [39] proposed a driving cycle synthesis approach for building profiles of customers based on aggregate fleet data such as histograms, whereas (i) the representativeness of customer usage behavior for electromobility remains to be investigated, and (ii) the representativeness of driving cycles on lifetime requires further investigation. Towards both aspects, Ling et al. [18] developed a profiling approach to synthesize usage profiles with respect to parking, charging, and long-term loads on lifetime issues. However, it remains unclear if these synthesized usage profiles could represent digital twins for supporting automotive systems engineering processes. Based on the related work, this paper will propose a DSS solution targeting to digital twinning based on aggregate data and evaluate the feasibility using real-world customer fleet data from three different market regions.

### III. SYSTEM FRAMEWORK

This section introduces the components and their relationships of the proposed decision support system (DSS) in the context of automotive systems engineering.



**FIGURE 1.** An overview of the DSS. Solid arrows indicate connections for direct decision support pipelines for operations configured by users. Hollow arrows refer to indirect maintenance pipelines that clean, verify, and update the data, the models, and simulation outcomes.

### A. OVERVIEW

Typically, a DSS consists of individual sub-systems to coordinate data, models, knowledge, and users [40]. In the proposed DSS, knowledge management is implemented by digital twinning, reflecting estimates of long-term customer vehicle metrics.

Fig. 1 shows an overview of the proposed DSS. In the context of a digital twin, the DSS coordinates real-world data and virtual-world models on the virtual side by a set of pre-defined or configurable pipelines for data processing and simulations. On the front-end side, the user interface provides the capability of configuring data ingestion, models, twinning algorithm, as well as the visualization of customer profiles and decision support metrics. The data processing, profiling, and simulation in digital twins are executed on the back-end side.

The key components of the DSS framework are the data and models. The data subsystem provides the interface between the data aggregated from customer fleets (aggregate data) and the data logged from in-house fleets (logging data). The model subsystem organizes physical and evaluation models for the use of system simulation and metrics computation. With data and models at hand, digital twin-based metrics can be derived for decision support. To allow the simulation outcomes to reflect a digital twin for customer fleets, data processing pipelines are necessary. The pipelines can be executed on-demand or scheduled for forecast purposes.

#### 1) DATA

Aggregate data are long-term statistical measurement values from the onboard control units, which describes the usage behavior of customer fleets. They can be acquired via diagnostic jobs at dealers, or via monthly or on-demand campaign scripts over mobile telemetries [41]. These long-term values are represented in the frequency domain, hence, the size of data remains constant even over a couple of years. Although the temporal sequential information is lost, these data can be easily aggregated, as exemplary shown in TABLE 1. Furthermore, this preserves customer privacy, as what the customer did at what time remains unknown.

**TABLE 1.** An example of aggregate data.  $\Sigma_m$  denotes total mileage. “ $D_v$  in (0,10] km/h” refers to duration when velocity is between 0 and 10 km/h, aggregated in  $h$ . “ $T_s$  in (0,5] km” is the number of trips whose mileages are within 5 km.

Customer vehicle	$\Sigma_m$ in km	$D_v$ in (0,10] km/h	...	$D_v$ in (200,+∞) km/h	$T_s$ in (0,5] km	...	$T_s$ in (400,+∞) km	...
C000001	67345	320	...	0	1450	...	6	...
C000002	85034	431	...	0.2	1267	...	78	...
...	...	...	...	...	...	...	...	...

**TABLE 2.** An example of logging data.  $v$  refers to velocity,  $a$  is the acceleration, and  $T_{ex}$  represents the external temperature.  $\theta_r$  denotes the road slope, which is the inclination angle upwards of the road direction.

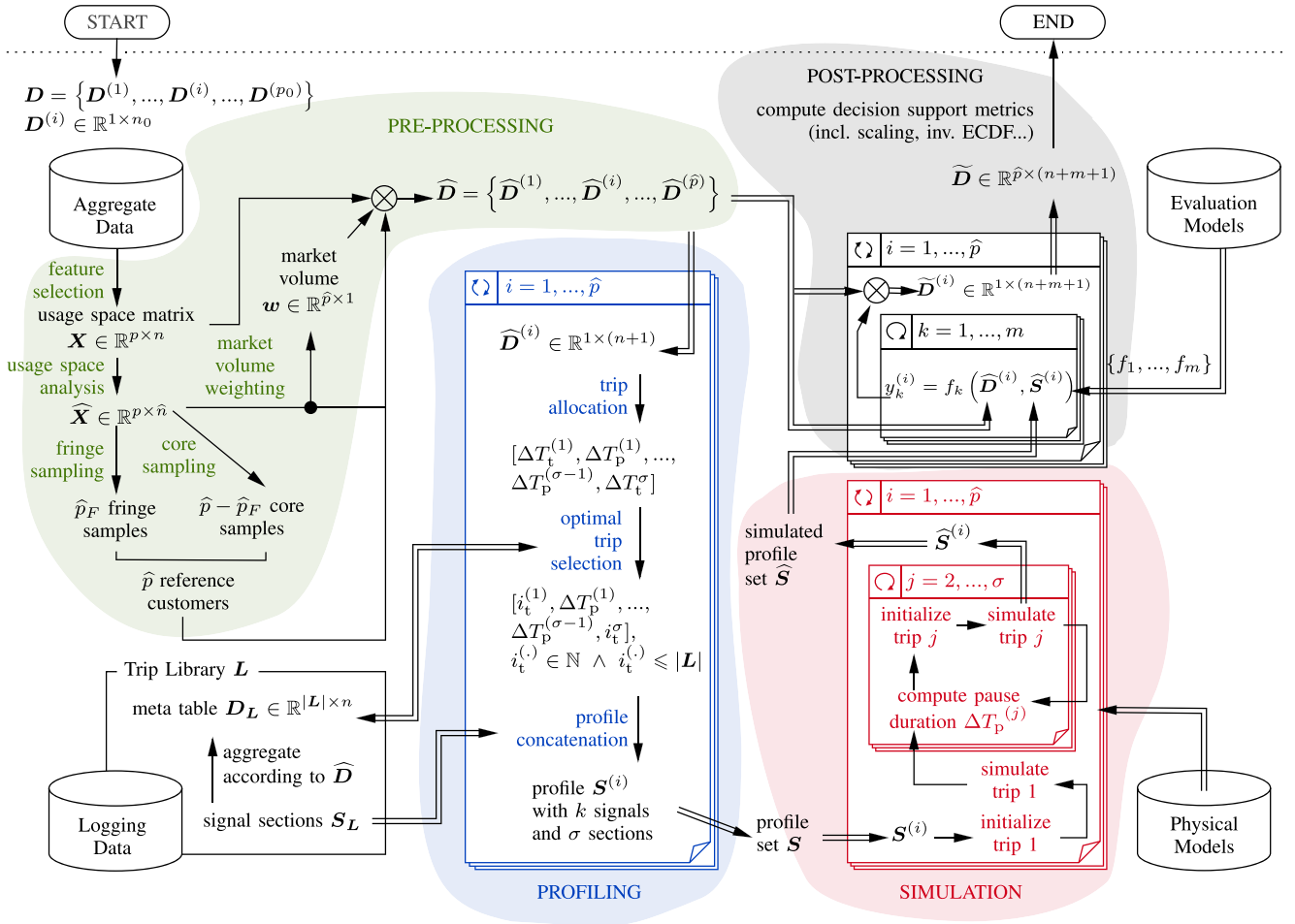
Testing vehicle	Timestamp	$v$ in km/h	$a$ in m/s <sup>2</sup>	$\theta_r$ in °	$T_{ex}$ in °C	...
T000001	2021-03-05 13:58:00	33.2	0.720	2.25	26.2	...
T000001	2021-03-05 13:58:01	35.8	0.315	1.55	26.2	...
...	...	...	...	...	...	...

The data is prepared by extracting necessary attributes from those source data for the digital twin.

Logging data are time-series signal traces from external data loggers installed in in-house fleets. These fleets could be all kinds of testing vehicles on the road. As the number of testing vehicles is far less than the customer fleets, it is insufficient to represent the usage behavior of a concrete customer with a single testing vehicle. However, the logging data are finely resolved over a long time (usually over one year) and contain temporal information of sensors, which could be regarded as the inputs for simulations. As the frame rates for those signals are not identically flowing in the control units, they should be down-sampled uniformly to, e.g., 1 Hz before feeding into the simulation models, as presented in TABLE 2. In the digital twin, logging data are used as the “raw material” to synthesize usage profiles for aggregate customer data.

#### 2) MODELS

Provided by diverse departments which are in charge of developing automotive components, physical models could include all essential vehicle components such as engines, transmissions, electric machines, batteries, vehicle chassis, auxiliaries, etc. They should be parameterized to provide sufficient inputs for evaluation models after simulation. For instance, in order to count the number of cold starts, the output signals of engine model should be parameterized with speed, torque, and cooling temperature. Technically, the component models could be zero- or one-dimension models that resolve loads of components. This ensures that the system model could exceed further from real-time capabilities, allowing digital twin simulations for a group of customer fleets within feasible time. Having the component models available, pre-defined configurations of the components, their parameters, the functional relationship between the components, as well as operating strategies are



**FIGURE 2.** The flow chart of digital twinning pipeline including pre-processing, profiling, simulation and post-processing. Boxes with double circular arrows are parallelizable loops. Boxes with a single circular arrow represent non-parallelizable loops.

integrated. The operating strategies are especially important for hybrid powertrains, where the combustion engine and electric machines are working together. This is also an important motivation why this work uses the simulation for estimating the component states of customer fleets instead of directly using in-house fleet signals. As a whole, the inputs of system models are signals such as velocity, external temperature, and road slope. Then the simulation outputs various loads of components such as engine coolant temperature and battery charging states. In this paper, we use an in-house developed system model to simulate the dynamic loads of vehicle components. The detail of this model can be found in [42], [43].

Provided by various competence centers with focus on features such as lifetime, evaluation models are usually analytical models with the inputs from simulation outcome, outputting decision support metrics. These metrics are, in most cases, scalars. A simple example could be, according to the engine loads, fuel consumption. With the development of the physics of failure models like bearing wear, thermal aging, etc., customer-centric reliability evaluation is possible in the proposed DSS framework.

## B. DIGITAL TWINNING

As illustrated in Fig. 2, each digital twinning pipeline consists of a core pipeline and its corresponding constraints, explained in the following four steps in detail, i.e., pre-processing, profiling, simulation, and post-processing.

### 1) PRE-PROCESSING

Suppose that the aggregate customer fleet data is summarized in a tabular dataset  $\mathbf{D}$  with  $p_0$  tuples (rows) and  $n_0$  attributes (columns). In  $\mathbf{D}$ , each tuple represents a customer vehicle with the latest ingestion of their measurement statistics and other meta information such as region, dealer, vehicle, and powertrain configurations. Referring from those meta-information, the pipeline firstly drills the target customers down, yielding  $p$  customer vehicles. Considering the discrete nature of aggregate usage attributes, all attributes are selected that describe the usage behavior and thus getting  $n$  features. The extracted dataset is concatenated into a matrix  $\mathbf{X} \in \mathbb{R}^{p \times n}$ , namely the usage space matrix.

For analyzing huge amounts of customer vehicles, the computational costs in simulations later on could be over budget if the twinning is performed one by one. To reduce

the costs, the number of digital twins could be reduced by selecting reference customer vehicles using sampling. Given the sample set cardinality  $\hat{p}$ , a speed-up of profiling and simulation reaches a factor of  $p/\hat{p}$  as they are with an algorithmic complexity of  $O(n)$  for various customer vehicles, i.e., those samples are computational independent. Typically, random sampling should reach acceptable representativeness for distributions. Concerning product lifetime, however, the coverage of fringe customers whose usage pattern could potentially lead to extreme product loads that may cause reliability issues, which are usually strongly non-linear correlated to the usage behavior. From the geometric perspective, Ling et al. [44] developed a usage space sampling method which is particularly suitable for sampling tasks focusing on such latent fringe coverage. Hence, the method is utilized to pre-process  $\mathbf{D}$ , detailed in the following steps.

First,  $n$ -dimensional matrix  $\mathbf{X}$  is compressed into  $\hat{\mathbf{X}} \in \mathbb{R}^{p \times \hat{n}}$  with  $\hat{n}$  principal components by performing  $z$ -score normalization and truncated singular value decomposition, namely Principal Component Analysis (PCA). This is because (i) it is one of the most widely used dimensionality reduction technique, (ii) the resulted principal components are linear combinations from the original axes and thus deliver reproducible geometrical properties to the next sampling steps. In [7] and [44], this step is referred to as usage space analysis. Starting from the geometry of reduced vector space  $\hat{\mathbf{X}}$ , a convex hull [45] is then constructed and find  $\hat{p}_F$  geometric samples on the boundary, i.e., fringe sampling. A complement to the fringe samples to the cardinality target  $\hat{p}$  is to select  $\hat{p} - \hat{p}_F$  samples from the rest via Halton sequence, a quasi-random point set [46], i.e., core sampling. For the selected  $\hat{p}$  samples, their importance ought to be estimated, as some of them may be so rare that almost no customer in the population is similar to them. For customer vehicle  $i = 1, \dots, \hat{p}$ , the weight  $w_i$  is computed by counting their neighbors segmented by Voronoi tessellation based on Euclidean distances. The weights indicate their market volumes from the population, hence the step is called market volume weighting.

Overall, this sampling process is deterministic, reproducible and proven to be robust for various datasets and particularly suitable for sampling that focuses on capturing latent fringe customers. For further information, please refer to [44]. So far, the pre-processed dataset  $\hat{\mathbf{D}}$  has  $\hat{p}$  rows and  $n+1$  features. The additional feature is the market volume  $\mathbf{w}$ .

## 2) PROFILING

To infer the usage behavior of a customer vehicle, time-series driving profiles are synthesized by reconstructing logging data towards the aggregate data, i.e., profiling. Assuming that the aggregate data contain adequate information to characterize the usage behavior, it is feasible to perform individual profiling from the logging data towards the aggregate statistical information considering comprehensive usage [18]. The core profiling mechanism is to find a suitable

combination of logging signal sections to represent the usage statistics of customer fleets, so that customer behavior could be modeled based on the trip sections.

To ensure the synthesis process, the logging data from various testing fleets is firstly decomposed and reorganized to establish a trip library  $\mathbf{L}$ , as described in [18]. These logging data includes  $k$  time-series variables such as velocity, road slope, and external temperature, and can be directly imported into physical models for system simulation.

Given the pre-processed dataset  $\hat{\mathbf{D}}$ , denote customer vehicle (row)  $i$  as  $\hat{\mathbf{D}}^{(i)} \in \mathbb{R}^{1 \times (n+1)}$ . For  $i = 1, \dots, \hat{p}$ , the objective of profiling is to synthesize a multivariate time-series profile  $\mathbf{S}^{(i)}$ . The profile represents the long-term aggregate usage information from  $\hat{\mathbf{D}}^{(i)}$ , which spans roundabout a week. The profiling procedure includes following steps.

First, allocate a timetable of  $\sigma$  trips and  $(\sigma - 1)$  pauses using inverse transform sampling [47]. For trip  $j = 1, \dots, \sigma$ , denote  $\Delta T_t^{(j)}$  and  $\Delta T_p^{(j)}$  as the trip duration and the pause duration exactly after trip  $j$ . Respectively, the timetable can be expressed by  $[\Delta T_t^{(1)}, \Delta T_p^{(1)}, \dots, \Delta T_p^{(\sigma-1)}, \Delta T_t^{(\sigma)}]$ . Note that there is no pause after the last trip written in profile  $\mathbf{S}^{(i)}$ . Second, search representative trips from trip library  $\mathbf{L}$  by comparing the aggregate features of combined trips and that of target customer  $\hat{\mathbf{D}}^{(i)}$  using meta-heuristics [48]. For trip  $j$ , denote the indexes of the trip found in  $\mathbf{L}$  as  $i_t^{(j)}$ . As indexes are natural numbers and act as the interface to  $\mathbf{L}$ , the selected index subjects to  $i_t^{(j)} \in \mathbb{N} \wedge i_t^{(j)} \leq |\mathbf{L}|$ , where  $|\mathbf{L}|$  is the cardinality, namely the number of trips contained, in the trip library. Thus, the time table is extended by assigning trip indexes, yielding  $[i_t^{(1)}, \Delta T_p^{(1)}, \dots, \Delta T_p^{(\sigma-1)}, i_t^{(\sigma)}]$ . Third, build a realistic week profile  $\mathbf{S}^{(i)}$  by rearranging the trips from  $\mathbf{S}_L$  and pauses alternatively and concatenating them. The last step of profiling from [18], i.e., integration, is extended and applied in Section III-B.3.

After conducting the profiling procedure for all  $\hat{p}$  customer vehicles, the usage profile set  $\mathbf{S}$  with  $\hat{p}$  signal profiles serve as the inputs for system simulation.

## 3) SIMULATION

Depending on the requirements from utilized physical models, profile  $\mathbf{S}^{(i)}$  includes  $k$  time-series signals and  $\sigma^{(i)}$  trips selected from  $\mathbf{L}$ . If signals related to pause or charging are not available in  $\mathbf{L}$ , or no dynamic simulation during pauses is considered, the pauses can be clipped out of the signals to save the storage capacity and computational expense. The pause information is not completely lost, but implicitly kept inside the timestamps  $\mathbf{S}^{(i, T)}$ .

For the start of trip  $j = 2, \dots, \sigma^{(i)}$ , the corresponding timestamp is denoted as  $T_j$ . Between trip  $j$  and trip  $(j - 1)$ , the pause duration is then

$$\Delta T_P^{(j)} = \mathbf{S}^{(i, T)}(T_j) - \mathbf{S}^{(i, T)}(T_j - 1), \quad (1)$$

which supports the initialization of boundary conditions (such as the battery state of charge or engine temperatures which are relevant to pause durations) before simulating

trip  $j$ . As initial conditions of trip  $j = 2, \dots, \sigma^{(i)}$  are derived after simulating their previous trips, the simulation for customer  $i$  itself could not be paralleled. If there are any opening boundary conditions, e.g., charging habit, they could be configured as scenarios for what-if analysis.

After system simulation, profile  $\mathcal{S}^{(i)}$  with  $r$  signals is extended with simulation outputs, yielding  $\widehat{\mathcal{S}}^{(i)}$  with  $\widehat{r}$  signals, where  $\widehat{r} > r$ . Profile  $\widehat{\mathcal{S}}^{(i)}$  includes various vehicle loads such as the rotational speed, moment, and coolant temperature of engines, providing necessary inputs as estimating indicators for decision support.

#### 4) POST-PROCESSING

Given the  $\widehat{p}$  simulated profiles as set  $\widehat{\mathcal{S}}$  and their meta-information with market volumes  $w$  as set  $\widehat{\mathcal{D}}$ , it is ready to perform decision analytics by computing evaluation models followed by computing decision support metrics.

Assume that there are  $m$  evaluation models, and each of them generates a scalar output. For customer  $i = 1, \dots, \widehat{p}$ , its corresponding evaluation model  $k = 1, \dots, m$  could be expressed as a function  $y_k^{(i)} = f_k(\widehat{\mathcal{D}}^{(i)}, \widehat{\mathcal{S}}^{(i)})$ . The  $m$  outputs  $y_1^{(i)}, \dots, y_m^{(i)}$  extends the meta-information of customer vehicles, yielding  $\widetilde{\mathcal{D}}^{(i)}$  with  $n + m + 1$  features, where the additional one feature indicates the market volume  $w_i$ . Bringing all the metadata of  $\widehat{p}$  customer vehicles together, the prepared dataset  $\widetilde{\mathcal{D}} \in \mathbb{R}^{\widehat{p} \times (n+m+1)}$  is the last input of computing metrics for decision support.

Relative metrics, such as percentages or time share, are dimensionless quantities and are thus directly comparable from the evaluation functions. However, absolute metrics, such as counts, cumulative operating duration, cumulative mileages, and other functional expressions, are to be scaled towards defined targets, such as their lifetime.

In the automotive industry, lifetime targets are typically defined by  $z_a$  years,  $z_m$  mileage (kilometers), and/or  $z_h$  operating hours of corresponding components. For customer  $i$ , the synthesized profiles usually cannot reach those targets. Corresponding to the lifetime targets, from their  $\mathcal{S}^{(i)}$  profile time span  $\widetilde{z}_a^{(i)}$  can be computed as years equivalent, mileage  $\widetilde{z}_m^{(i)}$  as kilometers, and  $\widetilde{z}_h^{(i)}$  as operating hours. Typically, reaching one of the targets is sufficient for vehicle lifetime, i.e., whatever first. The scaling factor for customer  $i$  is then

$$z_c^{(i)} = \min \left\{ \frac{z_a}{\widetilde{z}_a^{(i)}}, \frac{z_m}{\widetilde{z}_m^{(i)}}, \frac{z_h}{\widetilde{z}_h^{(i)}} \right\}, \quad (2)$$

which is the minimum of the multiplier of all targets.

After computing all required evaluation models and considering lifetime scaling factor as multipliers for all customer vehicles, metrics are to be computed for decision support. A typical type of metric is the critical scaled evaluation output which covers the majority of the customer vehicles. As only a small number of reference customer vehicles are twinned, the coverage should consider other customer vehicles that are not sampled, i.e., their market

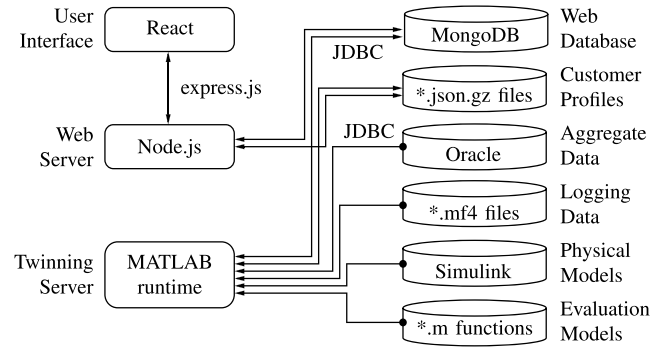


FIGURE 3. System configuration for the proof of concept (POC) of digital twinning.

volumes. Before deriving the metric, the weighted empirical cumulative density function (ECDF) of the output of model  $k$  ought to be computed over the market, i.e.,

$$F_k(\theta) = \frac{1}{\sum_i w_i} \sum_{i=1}^{\widehat{p}} w_i \mathbf{1}_{z_c^{(i)} y_k^{(i)} \leq \theta}, \quad (3)$$

where  $\theta$  represents the functional output value of model  $k$ , and indicator function  $\mathbf{1}$  is activated only when the model output for customer  $i$  is smaller than or equals (covers)  $\theta$ . The metric is then the inverse ECDF value with a given coverage. For example,  $F_k^{-1}(0.99)$  indicates the 99% quantile of critical indicator of model  $k$ , which covers over 99% of the customers in the market.

To summarize, let us take the decision support pipeline as a whole. The physical twin for customer vehicle  $i$  is aggregated by  $n$  usage statistical indicators, i.e.,  $\widehat{\mathcal{D}}^{(i)}$ . After digital twinning,  $\widehat{\mathcal{D}}^{(i)}$  is expanded with  $m$  post-processed evaluation metrics and becomes  $\widetilde{\mathcal{D}}^{(i)}$ . In addition, profile  $\widehat{\mathcal{S}}^{(i)}$  includes representative time-series signals and thus enables detailed inspection and analysis of the usage behavior. Hence, a digital twin for  $i$  can be represented by  $\{\widetilde{\mathcal{D}}^{(i)}, \widehat{\mathcal{S}}^{(i)}\}$ . From the market perspective, reference vehicle  $i$  represents  $w_i$  customers in the market. To support decision-making in automotive systems engineering processes, a market digital twin is built by bringing the reference vehicles together, represented by  $\{\widetilde{\mathcal{D}}, \widehat{\mathcal{S}}\}$ .

## IV. EVALUATION

To demonstrate the feasibility of the framework and evaluate the prediction accuracy of twinning pipeline, a proof of concept (POC) is conducted based on in-house real-world customer fleet data. In this section, the configuration of this DSS is introduced. According to various indicators related to requirement engineering in automotive development, the evaluation results of twinned customer profiles are presented.

### A. SYSTEM CONFIGURATION

The architecture of DSS (introduced in Section III) is carried out as illustrated in Fig. 3. All data and models were established and available. The aggregate data of customer

fleets are stored in an Oracle database. The logging data of testing fleets, which are originally mf4 (measurement data format) files, are significantly larger and more versatile as a lot of signal channel names are not unified or not available for some vehicles. Hence, they are saved in shared file storage via in-house intranet. The physical and evaluation models based on Simulink models and MATLAB m-functions are also placed in ordinary file systems without using databases.

The software infrastructure of the other components is established for this POC on premise, including a full-stack Web application and a twinning server instance. Considering current working load and needs of POC, a Linux virtual machine is used from an in-house cloud cluster, equipped with 8 CPUs and 16 GB RAM.

The Web application is developed using MERN, a popular JavaScript-based technology stack. As the front-end, the user interface is developed in React. The back-end Web server used Node.js and interact with React by Express.js. As the data to be presented on the front-end are highly unstructured, MongoDB, a document-based database, is used to save the metadata of data and models available for configuring twinning, and to save the shortcuts or interface information of customer profiles as digital twins.

The twinning server instance is a running program based on MATLAB R2021b runtime. It monitors and synchronizes the status and metadata of twinning cases from the front-end to the MongoDB via a java database connectivity (JDBC) driver. When a new twinning request is submitted, it starts a batch instance which submits the data pipeline with configured parameters, including which customer groups, which physical model, which evaluation models, and the variables to control the twinning pipeline. Depending on the work load of running instances and the requirement of parallel computing, the instance will run on the local server or be submitted to a high-performance cluster. In each instance, aggregate data are queried by MATLAB via JDBC. The output profile signals are stored in gzip-compressed JavaScript object data format (JSON) to save storage space and operating costs of database, and to provide simple data access to the Web server.

Furthermore, it is worthwhile to declare that this technical configuration of the DSS framework is a possible example based on the existing IT infrastructure and resources. There is no strict restriction of selecting the hardware and software infrastructure.

## B. CASE DESCRIPTION

Based on the hardware and software infrastructure of the POC, a case study is conducted using the following data sources, models, and decision support metrics.

The aggregate data consists of three market regions with 34711, 9211, and 13188 customer vehicles, 57100 vehicles in total. These customer fleets are mid-size sedans with plug-in hybrid technologies. In the pre-processing of digital

twinning, 100 samples are selected for each market region. These samples build a reference customer fleet.

The logging data are acquired from 823 in-house vehicles with various vehicle variants and powertrain configurations, including test vehicles and volunteer customer vehicles for logging. It is worthwhile to mention that the majority of such testing fleets are from the market region I. The time-series signals from those testing fleets are decomposed into 657909 trips, which includes driving signals of 135600 h and 6.009 million km in total.

The physical model is developed in-house and simulates the total vehicle system dynamics with a focus on powertrain systems [42], [43]. Its configuration and parameters are identical to that of the customer fleets selected in pre-processing, which has been used for optimizing energy management strategy [43] and building an engine-in-the-loop environment [49].

To enable the evaluation of twinning accuracy, only the evaluation models and metrics are chosen which are available from aggregate data on real-world customer fleets. Hence, complex damage models, such as thermal aging, wear intensity, or thermal-mechanical fatigue, are not evaluated here. Instead, four evaluation metrics are chosen in this study: (i) average velocity in km/h to compare the profiling quality independent from physical simulation, (ii) number of engine starts to evaluate the prediction quality of long-term engine reliability, (iii) average fuel consumption rescaled by the range from real-world customer fleets, and (iv) time fraction of recuperation in which the electric motor utilizes the breaking energy to charge the battery.

As most of the metrics are average values or fractions, no lifetime scaling is necessary, except the number of engine starts. The lifetime targets are typically defined by regulations and manufacturers. This paper assumes a scaling target of 30 years, or 450000 km, or 18000 operating hours, whatever first. Approximately, this scaling target represents an annual mileage of 15000 km. For each customer, the scaling factor is calculated according to (2).

In profiling, each reference customer vehicle is twinned by an one-week profile. As no charging behavior is aggregated in the customer fleets, the uncertainty is quantified by simulating two extreme scenarios, i.e., always charge (S1) or never charge (S2) the battery between the trips. Here, no repetition is needed for quantifying the uncertainty of the whole twinning process, as the sampling, simulation for S1 and S2, as well as evaluation are deterministic. In addition, the random seeds for meta-heuristics in profiling are also initialized using identical set of pseudo-random numbers. Hence, it appears to be random but is deterministic. At the very beginning of each customer profile, following initial conditions are specified: battery state of charge as 50%, coolant and oil temperatures of engine and electric motor equal the corresponding ambient temperature.

The twinning accuracy could be quantified by relative error  $e$ , which is the relative fraction of real-world reference metric  $y_{ref}$  uncovered by the interval between scenarios S1



$y_{S1}$  and  $S2$   $y_{S2}$ , i.e.,

$$e(y_{\text{ref}}, y_{S1}, y_{S2}) = \frac{\max\{0, \min\{y_{S1}, y_{S2}\} - y_{\text{ref}}\} + \max\{0, y_{\text{ref}} - \max\{y_{S1}, y_{S2}\}\}}{y_{\text{ref}}}. \quad (4)$$

For metric  $\mu$  from the four indicators presented above and quantile  $q \in [0, 1]$ , according to equation (3), its real-world metric and predicted metric from scenarios S1 and S2 by twinning are denoted as  $F_{\mu, \text{ref}}^{-1}(q)$ ,  $F_{\mu, S1}^{-1}(q)$ , and  $F_{\mu, S2}^{-1}(q)$ . The quantile relative error (QRE) is then the relative fraction of real-world quantile metric uncovered by the interval between S1 and S2, i.e.,

$$\text{QRE}(\mu, q) = e\left(F_{\mu, \text{ref}}^{-1}(q), F_{\mu, S1}^{-1}(q), F_{\mu, S2}^{-1}(q)\right). \quad (5)$$

This work compares the QRE of median, 99% and max quantiles, i.e.,  $q = 0.5, 0.99$ , or  $1$ .

Apart from quantile distributions, the twinning quality of individual vehicle is also worthwhile to be evaluated. For selected customer sample  $i = 1, \dots, 100$ , its real-world metric is denoted as  $y_{\mu, \text{ref}}^{(i)}$ , predicted metric from scenarios S1 and S2 as  $y_{\mu, S1}^{(i)}$  and  $y_{\mu, S2}^{(i)}$ . The individual relative error (IRE) for customer  $i$  is, therefore, represented by

$$\text{IRE}(\mu, i) = e\left(y_{\mu, \text{ref}}^{(i)}, y_{\mu, S1}^{(i)}, y_{\mu, S2}^{(i)}\right). \quad (6)$$

One IRE value represents one customer sample but without considering the market volume. To consider the impact of  $w_i$  for reference customer vehicle  $i$ , therefore, it is worthwhile to count the fraction of customer fleets whose real-world metrics are covered by S1 and S2 to all the customer fleets. Given a tolerance of  $\delta$  for exceeding the S1 and S2 bands, the fraction is defined as fleet twinning accuracy (FTA) for metric  $\mu$ , i.e.,

$$\text{FTA}(\mu) = \frac{1}{\sum_i w_i} \sum_{i=1}^{100} w_i \mathbf{1}_{\text{IRE}(\mu, i) \leq \delta}. \quad (7)$$

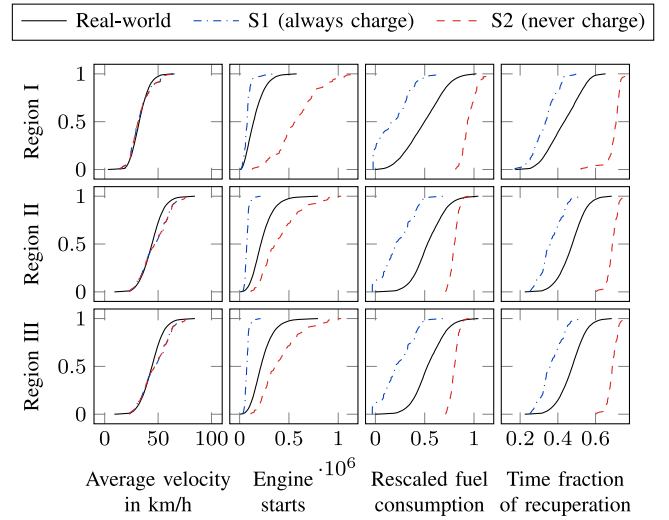
In this case study, zero tolerance is allowed for all metrics with scenario difference, but 10% of IRE for those without a difference, i.e.,

$$\delta = \begin{cases} 0.1 & \text{if } y_{\mu, S1}^{(i)} = y_{\mu, S2}^{(i)}, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

### C. RESULTS

The twinning case, including 300 reference customer vehicles from three market regions in total, takes less than 2.8 h using four cores, or roundly 5.2 h elapsed, including database query of aggregate and logging data. The most time consuming parts are profiling which takes  $36.2 \pm 33.7$  s per customer, and simulation which takes  $4.2 \pm 2.9$  s per customer per scenario.

Qualitatively, Fig. 4 shows the ECDFs of real-world reference case and both scenarios for four metrics and three regions. The first column of sub-figures show the distribution of average velocity among three market regions. As the velocity profiles are simulation inputs, their predicted



**FIGURE 4.** Cumulative distributions of evaluation metrics for real-world and twinned customer fleets under two scenarios, grouped by market regions. The horizontal axes are the metrics. The vertical axes are quantiles ranging from 0 to 1.

**TABLE 3.** Comparison of performance indicators of digital twinning grouped by evaluation metrics and market regions. Respectively, QRE and FTA represent the quantile relative error and fleet twinning accuracy. Generally, lower QREs or higher FTAs indicate better performance.

Metric	Market Region	QRE			FTA
		median	99%	max	
Average velocity in km/h	I	2.93%	9.53%	9.01%	97.08%
	II	0.31%	0.58%	23.27%	93.38%
	III	5.32%	6.54%	17.75%	99.62%
Number of engine starts	I	0	0	10.93%	84.61%
	II	0	0	0	76.09%
	III	0	0	2.41%	85.96%
Rescaled fuel consumption	I	0	0	91.53%	99.31%
	II	0	0	24.46%	92.01%
	III	0	0	15.82%	79.47%
Time fraction of recuperation	I	0	0	19.93%	95.55%
	II	0	0	10.21%	95.26%
	III	0	0	0.85%	94.76%

distributions under S1 and S2 are identical. Generally, the twinned reference customer profiles are capable of representing the average velocities of real-world customer fleets. In region III, the reference profiles are generally slightly faster than real-world values, but with an acceptable range, as only a minority of trip recordings in the trip library come from this region. Furthermore, the other metrics are calculated based on the simulation outcomes. Generally, all the real-world ECDF curves are between the predicted distribution of S1 and that of S2.

Quantitatively, TABLE 3 summarizes the QREs and FTAs. As the prediction uncertainty of charging behavior is quantified by S1 and S2, the performance indicators QRE and FTA already consider the uncertainty by computing the

coverage error. As discussed above, the average velocity is independent from the charging scenarios. On the one hand, their QREs are generally within 10% for median and 99% quantile, which can be, to some extent, compensated by considering a safety band on the prediction ECDFs, such as  $\pm 10\%$ . On the other hand, their FTAs indicate that the average velocity of over 93% of the individual customer fleets are correctly covered by the prediction of two scenarios. In terms of 99% quantile for the other evaluation metrics, all 300 reference customer profiles have an QRE of zero. This indicates that long-term behavior of engine starts, fuel consumption and recuperation ratio are covered by the reference customer fleet for all three markets. However, the maximum QRE for rescaled fuel consumption in region I is significantly larger than other regions. A possible cause could be the calibration data of engine fuel consumption used in the physical model mainly refers region II and III, which are closer to each other. Regarding their FTA, however, individual prediction coverage could be less than 80%, especially the engine starts in region II and fuel consumption in region III. In reality, a transient deep press of gas pedal could trigger an extra start, which is not considered in the simulation. This effect could lead to larger prediction errors for individual customer vehicles.

Regarding the digital twins as a whole, the derived requirements from these digital twin reference profiles can cover 99% of the real-world customer fleets. It is implied that if relevant system requirements are derived from the 99% quantile of metrics, the powertrain system can cover 99% of the usage scenarios of target customers throughout the product lifetime. When predicting the metrics for individual customer vehicles based on their usage statistics, on average, 91.09% of individual customer vehicles are well covered by S1 and S2. This implies that 91.09% of the vehicles could be represented by the digital twins in terms of the evaluation metrics.

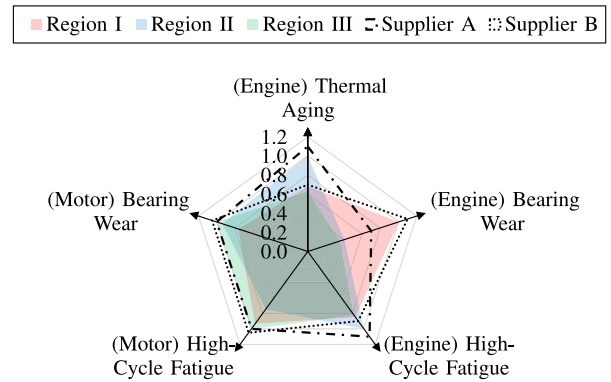
Generally, the target is to cover the distribution of real-world customer fleets by that of the predicted metrics under S1 and S2. Hence, the twinned reference customer fleet is feasible of representing all the four metrics for all three market regions.

#### D. EXEMPLARY USE CASES

Using the twinned reference customer profiles, the DSS is capable of predicting complex engineering objectives such as lifetime indicators. With the focus on customers, the digital twin-driven DSS proposed in this paper could support versatile decision-making tasks towards customer-centric automotive systems engineering. This section presents two potential use cases, including requirement localization and recall prioritization.

##### 1) REQUIREMENT LOCALIZATION

To select component suppliers for various market regions, known as localization, it is necessary to specify lifetime requirements corresponding to the regions. The case study



**FIGURE 5.** An example for localizing requirements of plug-in hybrid electric vehicles. The engine refers to the combustion engine. The motor refers to the electric motor. All metrics are rescaled according to the minimal and 99% quantile of customer fleet metrics.

above has three market regions. Suppose that there are two suppliers A and B. The relevant lifetime metrics are: thermal aging for the combustion engine [50], bearing wear [51] and high-cycle fatigue [52] for both of the combustion engine and the electric motor. These metrics are computed based on in-house evaluation models. An exemplary radar chart for the decision support task is visualized in Fig. 5.

The worst case scenario from S1 and S2 are plotted for 99% of the customer fleets in Region I, II, and III, which are significantly different in thermal aging and bearing wear. Assuming that the maximum thresholds for such metrics of the reliability tests from supplier A and B are reached, as shown in Fig. 5, supplier A could well cover region II and III, whereas the engine bearing wear from region I could not be covered. However, supplier B could cover all metrics of region I. In terms of reliability coverage, it is recommended to select supplier A for markets II and III, but supplier B for market I.

##### 2) RECALL PRIORITIZATION

Suppose that a small number of customer vehicles have engine failures. Their dealers report the issues to the car manufacturer. After analyzing those issues, the aftersales specialists observe that all those engines have severe thermal aging. However, the usage patterns of those vehicles are different. This could indicate that all vehicles with similar severity of thermal aging could have engine failures. To possibly eliminate the risk of vehicle failures, vehicles with higher risks should be firstly identified and then suggested to undergo maintenance.

The recall prioritization could be supported by digital twinning with the relevant vehicle and powertrain variants filtered. From the evaluation models, all relevant thermal aging ones are chosen as the decision support metrics. Taking all those metrics into account, it is possible to find which customer vehicles are possible to have those reported failures. Here, all customer vehicles with similar thermal aging damages could be identified. Then, customer services could send notifications to those vehicles for deeper

inspection as soon as possible. In this way, large amounts of breakdown issues could be well prevented.

## V. DISCUSSION

The core innovations of this work can be summarized from multiple perspectives, with significant implications for the digital twins, engineering practices, and customer centricity in the automotive industry.

In the scope of digital twins, this framework allows physics-based system simulation from customer vehicle usage statistics, representing a notable advancement in digital twins technology. The integration of physics-based simulations with customer data not only provides a realistic representation of automotive systems but also enables scalability from individual customer vehicles to the entire market.

In the field of automotive DSSs, this twinning pipeline delivers inputs for quantitative and fine-grained automotive systems engineering purposes, thereby influencing engineering practices. Engineers can utilize digital twins not only for individual vehicles but also for market-wide analysis, contributing to the development and optimization of automotive systems. The integration of data-driven customer profiling and physics-based simulations into the process of business analytics via digital twinning aligns with contemporary trends. This suggests that the automotive industry can leverage the framework not only for engineering purposes but also for strategic decision-making, influencing business analytics practices.

From the perspective of customer centricity, this solution enables digital twinning from customer data with low cost and privacy preservation. The customer-centric approach, including the anonymization of aggregated vehicle sensor statistics, addresses privacy concerns and enhances customer satisfaction and trust. Addressing privacy and ethical concerns, the framework incorporates robust measures for aggregate data and prioritizes privacy through on-board aggregation. This ensures the provision of anonymous aggregated vehicle sensor statistics without personal information or location data, emphasizing ethical considerations in data-driven technologies.

Although the evaluation results indicate the feasibility of this DSS, it is worthwhile to discuss the limitations of the digital twinning method and the case study. The QREs for maximum values are not yet well-covered by digital twins, sometimes leading to a larger error of over 20%. Hence, for systems engineering processes where a 99% quantile is insufficient for market requirements, it is inappropriate to perform the twinning method for individual customer vehicles. Furthermore, the twinning performance strongly depends on the data quality for aggregate and logging data.

In fact, a small number of relatively large individual IREs are observed. This could be caused by a lack of representative trips in the testing fleets, as well as the quality of aggregate data from customer vehicles. This may lead to a propagation of uncertainties for profiling,

resulting in unrepresentative profiles. Potentially, this issue could be improved by increasing the number of samples in pre-processing, increasing the profile length to a month, enlarging the trip library considering diversity of usage maneuvers, and tightening the convergence criteria of the profiling algorithm.

In conclusion, the proposed framework has the potential to reshape how the automotive industry approaches customer-centric automotive systems engineering. By addressing limitations and continually improving the framework, there is an opportunity for positive impacts on the efficiency of engineering practices, the satisfaction of end-users, and the ethical implementation of data-driven technologies in the automotive industry.

## VI. CONCLUSION

Based on digital twinning, a DSS was introduced which connects aggregate data from customer fleets, logging data from in-house testing fleets, physical models, and evaluation models for providing quantitative guidance in the process of customer-centric automotive systems engineering. The digital twins are built by a twinning pipeline, including pre-processing, profiling, simulation, and post-processing. To ensure privacy preservation for customers and reduce costs, no signal logging is necessary from customer fleets. In a case study, digital twins are built for a plug-in hybrid customer fleet from three market regions and compared four decision support metrics that are available from the real-world customer fleets.

Results indicate that the proposed digital twinning method is plausible and capable of providing accurate predictions for engineering requirements. This DSS framework is also feasible for solving such large-scale market prediction without losing technical details in systems engineering. Furthermore, two real-world decision support use cases are discussed, including the localization of lifetime requirements, as well as finding fleets with high risks of failures for potential recalls. The case study indicates that this DSS has potentials of improving the pace of customer-centric automotive systems engineering.

Despite providing quantitative results for the use cases presented, various aspects of the digital twinning could be further investigated. For vehicle usage statistics, currently one tuple represents one customer vehicle. It remains promising to make use of every acquisition of data from each customer, which are multivariate, high-dimensional, histogram-valued, and with sequential patterns. Furthermore, the model management requires effective management of data related to boundary conditions, applicability, and simulation results from simulation and evaluation models. These data are sourced from multiple departments and suppliers. The complexity and diversity of these data necessitate significant optimization of the data logistics process and corresponding information systems. Another important aspect is that, as mentioned in related work, the bi-directional connectivity could be improved by enabling the twinning

algorithm for cloud edge computing. The real-time capability could also be realized by hybrid digital twin frameworks which could incorporate cloud-based systems simulation for basic twinning and sensor signal processing for updating the twinning results on vehicle board computers. Moreover, further use cases in automotive systems engineering could be investigated and verified, e.g., the optimization of vehicle testing programs, sustainability evaluation, and predictive maintenance.

## ACKNOWLEDGMENT

The author would like to thank Thomas Setzer, Jan Thiele, Zihan Kong, Benjamin Esterl, and Maximilian Spannagl for their continuous support and valuable discussion throughout the research.

## REFERENCES

- [1] P. Fader, *Customer Centricity: Focus on the Right Customers for Strategic Advantage*. Philadelphia PA, USA: Wharton Digit. Press, 2020.
- [2] A. Albers et al., "Product profiles: Modelling customer benefits as a foundation to bring inventions to innovations," *Procedia CIRP*, vol. 70, pp. 253–258, May 2018.
- [3] J. Cardoso, H. Fromm, S. Nickel, G. Satzger, R. Studer, and C. Weinhardt, *Fundamentals of Service Systems*. Cham, Switzerland: Springer, 2015.
- [4] J. Holtmann, J. Meyer, and M. Meyer, "A seamless model-based development process for automotive systems," in *Proc. Softw. Eng. Workshopband*, 2011, pp. 79–88.
- [5] J. Weber, *Automotive Development Processes: Processes for Successful Customer Oriented Vehicle Development*. Berlin, Germany, Springer-Verlag, 2009.
- [6] B. J. Nies, O. Robert, and U. Lindemann, "Customer focused requirement engineering and system design for plug-in hybrid electric vehicles (PHEV)," in *Proc. 19th Int. Conf. Eng. Design (ICED13)*, vol. 1, 2013, pp. 329–338.
- [7] N. Haselgruber, K. Mautner, and J. Thiele, "Usage space analysis for reliability testing," *Qual. Rel. Eng. Int.*, vol. 26, no. 8, pp. 877–885, 2010.
- [8] J. Schoch, J. Gaertner, A. Schuller, and T. Setzer, "Enhancing electric vehicle sustainability through battery life optimal charging," *Transp. Res. B, Methodol.*, vol. 112, pp. 1–18, Jun. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0191261517305337>
- [9] E. Thonhofer et al., "Infrastructure-based digital twins for cooperative, connected, automated driving and smart road services," *IEEE Open J. Intell. Transp. Syst.*, vol. 4, pp. 311–324, 2023.
- [10] E. Brusa, "Digital twin: Toward the integration between system design and RAMS assessment through the model-based systems engineering," *IEEE Syst. J.*, vol. 15, no. 3, pp. 3549–3560, Sep. 2021.
- [11] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019.
- [12] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 8, pp. 21980–22012, 2020.
- [13] A. Deicke, "The electrical/electronic diagnostic concept of the new 7 series," SAE, Warrendale, PA, USA, SAE Rep. 2002-21-0022, 2002.
- [14] A. U. Sass, E. Esatbeyoglu, and T. Fischer, "Monitoring of powertrain component aging using in-vehicle signals," *Diagnose Mechatronischen Fahrzeugsystemen XIII: Neue Verfahren für Test, Prüfung und Diagnose von E/E-Systemen im Kfz*. Norderstedt, Germany: Books on Demand, 2019, pp. 15–28.
- [15] J. Wilberg, F. Schäfer, P. Kandlbinder, C. Hollauer, M. Omer, and U. Lindemann, "Data analytics in product development: Implications from expert interviews," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage (IEEM)*, 2017, pp. 818–822.
- [16] A. Esser, F. Kohnhäuser, N. Ostern, K. Engleson, and S. Rinderknecht, "Enabling a privacy-preserving synthesis of representative driving cycles from fleet data using data aggregation," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, 2018, pp. 1384–1389.
- [17] B. Martens and F. Mueller-Langer, "Access to digital car data and competition in aftermarket maintenance services," *J. Compet. Law Econ.*, vol. 16, no. 1, pp. 116–141, 2020.
- [18] K. Ling, N. Shah, and J. Thiele, "Customer-centric vehicle usage profiling considering driving, parking, and charging behavior," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, 2020, pp. 146–151.
- [19] G. Xie, K. Yang, C. Xu, R. Li, and S. Hu, "Digital twinning based adaptive development environment for automotive Cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 1387–1396, Feb. 2022.
- [20] Y. Gu, S. Zhang, and L. Qiu, "Digital twin driven requirement conversion in smart customized design," *IEEE Access*, vol. 9, pp. 64414–64426, 2021.
- [21] J. Li, Q. Zhou, H. Williams, H. Xu, and C. Du, "Cyber-physical data fusion in surrogate- assisted strength Pareto evolutionary algorithm for PHEV energy management optimization," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 4107–4117, Jun. 2022.
- [22] M. González, O. Salgado, J. Croes, B. Pluymers, and W. Desmet, "A digital twin for operational evaluation of vertical transportation systems," *IEEE Access*, vol. 8, pp. 114389–114400, 2020.
- [23] W. Sun, P. Wang, N. Xu, G. Wang, and Y. Zhang, "Dynamic digital twin and distributed incentives for resource allocation in aerial-assisted Internet of Vehicles," *IEEE Internet Things J.*, vol. 9, no. 8, pp. 5839–5852, Apr. 2022.
- [24] Z. Ren, J. Wan, and P. Deng, "Machine-learning-driven digital twin for lifecycle management of complex equipment," *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 1, pp. 9–22, Jan.–Mar. 2022.
- [25] Y. Qin, X. Wu, and J. Luo, "Data-model combined driven digital twin of life-cycle rolling bearing," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1530–1540, Mar. 2022.
- [26] B. Rodríguez, E. Sanjurjo, M. Tranchero, C. Romano, and F. González, "Thermal parameter and state estimation for digital twins of E-powertrain components," *IEEE Access*, vol. 9, pp. 97384–97400, 2021.
- [27] S. Boschert and R. Rosen, "Digital twin—The simulation aspect," in *Mechatronic Futures*. Cham, Switzerland: Springer, 2016, pp. 59–74.
- [28] K. Kuru, "Conceptualisation of human-on-the-loop haptic teleoperation with fully autonomous self-driving vehicles in the urban environment," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 448–469, 2021.
- [29] E. VanDerHorn and S. Mahadevan, "Digital twin: Generalization, characterization and implementation," *Decis. Support Syst.*, vol. 145, Jun. 2021, Art. no. 113524.
- [30] M. Riesener, G. Schuh, C. Dölle, and C. Tönnies, "The digital shadow as enabler for data analytics in product life cycle management," *Procedia CIRP*, vol. 80, pp. 729–734, May 2019.
- [31] E. Elghoneimy and W. A. Gruver, "Agent-based decision support and simulation for wood products manufacturing," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 1656–1668, Nov. 2012.
- [32] M. Clemente, M. P. Fanti, G. Iacobellis, M. Nolich, and W. Ukovich, "A decision support system for user-based vehicle relocation in car sharing systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 8, pp. 1283–1296, Aug. 2018.
- [33] R. D. Labati, A. Genovese, V. Piuri, F. Scotti, and G. Sforza, "A decision support system for wind power production," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 1, pp. 290–304, Jan. 2020.
- [34] M. Mennenga, S. Thiede, and C. Herrmann, "Life-cycle oriented decision support for the planning of fleets with alternative powertrain vehicles," in *Re-Engineering Manufacturing for Sustainability*. Singapore: Springer, 2013, pp. 75–80.
- [35] B. Kloör, M. Monhof, D. Beverungen, and S. Braäer, "Design and evaluation of a model-driven decision support system for repurposing electric vehicle batteries," *Eur. J. Inf. Syst.*, vol. 27, no. 2, pp. 171–188, 2018.
- [36] Z. Wang et al., "Mobility digital twin: Concept, architecture, case study, and future challenges," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 17452–17467, Sep. 2022.
- [37] J. Leng et al., "Blockchain-secured smart manufacturing in industry 4.0: A survey," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 237–252, Jan. 2021.

- [38] R. A. Mallah, D. López, and B. Farooq, "Cyber-security risk assessment framework for blockchains in smart mobility," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 294–311, 2021.
- [39] A. Esser, M. Zeller, S. Foulard, and S. Rinderknecht, "Stochastic synthesis of representative and multidimensional driving cycles," *SAE Int. J. Alternat. Powertrains*, vol. 7, no. 3, pp. 263–272, 2018.
- [40] E. Turban, J. Aronson, and T. Liang, *Decision Support Systems and Intelligent Systems.*, 7th ed. Hoboken, NJ, USA: Prentice-Hall, 2005.
- [41] B. Schlegel, *Off-Board Car Diagnostics Based on Heterogeneous, Highly Imbalanced and High-Dimensional Data Using Machine Learning Techniques*, vol. 14. Kassel, Germany: Kassel Univ. Press, 2019.
- [42] A. Sittig, G. Mumelter, and F. Rabenstein, "Development and calibration of hybrid electric vehicle operating strategies using simulations and a hybrid power train test bench," in *Proc. IEEE Veh. Power Propul. Conf. (VPPC)*, 2011, pp. 1–5.
- [43] A. Sittig, "Optimierung und Applikation von Betriebsstrategien in Hybridfahrzeugen," Ph.D. Dissertation, Technischen Universität München zur Erlangung des akademischen Grades eines Technical Univ. Munich, Munich, Germany, 2014.
- [44] K. Ling, J. Thiele, and T. Setzer, "Usage space sampling for fringe customer identification," in *Proc. 54th Hawaii Int. Conf. Syst. Sci.*, 2021, pp. 1748–1757.
- [45] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, 1996.
- [46] L. Kocis and W. J. Whiten, "Computational investigations of low-discrepancy sequences," *ACM Trans. Math. Softw.*, vol. 23, no. 2, pp. 266–294, 1997.
- [47] L. Devroye, *Non-Uniform Random Variate Generation*. New York, NY, USA: Springer, 1986.
- [48] E.-G. Talbi, *Metaheuristics: From Design to Implementation*. Oxford, U.K.: Wiley, 2009.
- [49] M. Dietrich, K. Ling, R. Schmid, Z. Song, and C. Beidl, "Design and evaluation of an engine-in-the-loop environment for developing plug-in hybrid electric vehicle operating strategies at conventional test benches," *Autom. Eng. Technol.*, vol. 6, no. 3–4, pp. 247–259, Sep. 2021.
- [50] R. L. Feller, *Accelerated Aging: Photochemical and Thermal Aspects*. Los Angeles, CA, USA: Getty Publ., 1995.
- [51] T. Nicholas, *High Cycle Fatigue: A Mechanics of Materials Perspective*. Amsterdam, The Netherlands: Elsevier, 2006.
- [52] H. Karimaei and H. Chamani, "Effects of bearing steps on probable wear damage of main bearings of an IC engine," *Int. J. Autom. Mech. Eng.*, vol. 18, no. 2, pp. 8714–8727, 2021.



**KUNXIONG LING** received the M.Sc. degree in energy science and engineering from the Darmstadt University of Technology, Germany, in 2018, and the Ph.D. degree in business informatics from the Ingolstadt School of Management, Catholic University of Eichstätt-Ingolstadt, Germany, in 2022. He is currently working with BMW Group, Munich, Germany, focusing on data aggregation and sampling procedures for customer-centric automotive systems engineering.