IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 4, 2022, accepted April 26, 2022, date of publication April 29, 2022, date of current version May 6, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3171180

Virtual Testing in Automated Driving Systems Certification. A Longitudinal Dynamics Validation Example

RICCARDO DONÀ[1], SÁNDOR VASS[2], KONSTANTINOS MATTAS[2], MARIA CRISTINA GALASSI[2], AND BIAGIO CIUFFO[2], (Member, IEEE)

[1]Uni Systems Italy, 20145 Milano, Italy
[2]European Commission Joint Research Centre (JRC), 21027 Ispra (VA), Italy

Corresponding author: Biagio Ciuffo (biagio.ciuffo@ec.europa.eu)

This work was supported by the European Commission Joint Research Centre.

ABSTRACT The safety validation of Automated Driving Systems (ADSs) needs a combination of tools to ensure testing in a broad range of traffic scenarios. Among the others, virtual testing is expected to play a major role in the future. Differently from other methods, virtual testing allows examining an ADS in complex driving scenarios involving several road users and characterized by any level of criticality in a safe, efficient and effective way. However, before virtual testing can be used in the ADS certification process, proper validation methodologies have to be established to ensure the appropriateness of the simulation-generated evidence to the end of establishing a "*virtual proving ground*" regardless of the specific ADS. In this context, the present paper summarizes the results of the virtual environment validation exercise which involved a Vehicle-Hardware-in-the-Loop (VeHIL) setup against real-world experiments. The analysis only embraces the longitudinal dynamics due to the limitation of the virtual environment. Nonetheless, the methodology presented can be straightforwardly extended to different toolchains to cover more advanced ADS for the sake of virtual certification. The manuscript has a twofold contribution. On one side, it gives a quantitative estimation of the fidelity level achievable using a state-of-the-art VeHIL environment, thus standing apart from validation activities purely based on qualitative comparisons and contributions mostly concerned with the validation of the ADS itself. Secondly, it provides an end-to-end validation procedure that could be generalized to other cases of study or different testing setups. The results achieved are encouraging as they show an overall good match between real-world and simulated data. However, open issues remain in order to define a complete validation framework for virtual testing environments.

INDEX TERMS Automated driving, model validation, simulation, virtual testing.

I. INTRODUCTION

The verification and validation (V&V) of an Autonomous Driving System (ADS) requires new methods for a comprehensive and robust safety assessment. Due to the flexibility and cost-effectiveness that it offers, virtual testing is expected to play a fundamental role in this context [1], [2]. Physical testing only is indeed not a viable option due to the lack of *repeatability* and *scalability*, *dangerousness*, and the extremely *high costs* associated with such a testing procedure. Several works [3], [4] have pointed out how many billions miles might be necessary to validate an ADS depending on its operational domain: by physical testing only, matching such a demand would require traveling for decades on a single vehicle.

On the other side, modeling and simulation tools (M&S) have indeed a well-established history in the automotive industry at the development/testing levels. Simulation offers several advantages, including the unmatched capability of *controlling* the virtual driving scenario, the potentiality of *repeating & replicating* the same scenario on different computing devices, testing *efficiency*, and *safety* increase.

Nonetheless, the adoption of M&S as a *certification* tool is not regulated by any modeling standard given the

The associate editor coordinating the review of this manuscript and approving it for publication was Tamas Tettamanti.

complexity of the toolchains and the level of tailoring that modern ADS simulation solutions exhibit [5]. In order for a virtual testing environment to be certified, the toolchain will have to undergo a proper *validation* procedure first. Hence, the capability of validating complex simulation setups is of paramount importance for the type-approval of future ADSs. Despite the relatively large literature focusing on the validation of ADS using virtual tools, only a few works provide an in-depth validation analysis of the virtual testing tool itself, which are extensively examined in Section II-A.

This work aims at stimulating the scientific discussion on the validation of virtual environments by applying state-of-the-art validation tools to quantify the fidelity level which can be expected from a semi-virtual testing environment with respect to real-world tests. The focus of the present scientific contribution is thus *not* on validating the ADS. Instead, the testing environment is investigated, despite the ADS itself plays a role in the validation process of the tool as it will be discussed. Given the limitations of the virtual testing setup, in this contribution we only study longitudinal maneuvers such as free-flow and car-follow thus limiting to scope of application to lower automation levels. However, the methodology presented can be straightforwardly adopted for more complex maneuvers involving steering actions provided that suitable Key Performance Indicators (KPIs) are used.

II. RELATED WORK

The validation of virtual testing environments for ADS can be fulfilled in two manners: on an ''integrated system'' level, where the overall simulation toolchain is tuned to reproduce specific maneuvers Section II-A, and according to the ''submodels-based'' paradigm Section II-B, where each element of the simulation pipeline is individually validated with respect to its physical counterpart [6]. The methodologies can also be adopted in a sequential cascade, as explained in [5].

A. INTEGRATED-LEVEL SYSTEM

This set of validation methods is concerned with the definition of reference maneuvers and the tuning of a simulation environment to virtually reproduce the driving task. The selection of the KPIs is such that they are representative of the entire (closed-loop) simulation environment. For instance, the vehicle trajectory is typically used, whereas intermediate information, such as KPIs related to the detection of obstacles, is not explicitly accounted for.

Notable contributions in the scientific literature are represented by [7]–[9]. In [7], the correlation of a selection of KPIs is investigated against a Model-in-the-Loop (MIL) and a VeHIL setup for car-following and cut-in scenarios. As an outcome of the European Project Enable-S3, a left-turn maneuver at an unprotected junction is studied in [8] via qualitatively comparing the results obtained in the real-world with simulation-generated evidence. In [9], a method is proposed which firstly checks scenario coverage and then computes the correlation between the simulated and distance to the lane's margin for a lane-keeping application.

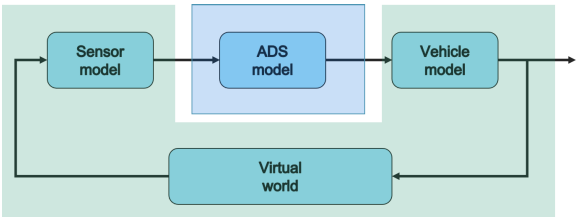


FIGURE 1. Submodels-based validation scheme, [6, Fig. 10].

From a regulation perspective, an approach following this philosophy is currently under approval for the AEBS virtual test [10]. In addition, a similar method is already implemented in the UN/ECE R140 [11] for the Electronic Stability Control (ESC) type approval.

Concerning the potential assets, this approach does not require an in-depth study of the components making up the virtual pipeline. The focus is instead on the ultimate performance of the whole virtual realization. On the downside, this method provides little information on how the simulation model will extrapolate for scenarios different from those used in the validation process.

The cited scientific efforts aim at fulfilling a similar research gap as the current contribution. Moreover, in [7], a setup closely matching ours is used. Nonetheless, our methodology moves a step further by adopting several computational tools, discussing their efficiency, and by performing a *global* assessment of the testing environment as we highlight in Section III.

B. SUBMODELS-BASED APPROACH

An alternative/complementary solution is the subdivision of the virtual pipeline into functional submodules, as graphically represented in Fig. 1, followed by the step-by-step validation of the single modules constituting the toolchain.

The contributions presented hereafter go beyond the level of validation detail provided by our (global) methodology. Nonetheless, we reckon the importance of analyzing separately each virtual ingredient making up the testing pipeline, especially for setups which include a higher degree of virtualization with respect to our VeHIL-based approach. Conversely, more details about the submodels-based approach can be found in our survey [6].

1) SENSOR-SYSTEM

The individual validation of sensor models is, at the moment of writing, an open topic in the scientific literature given the large amount of modeling options: from white-box (based on the replication of physics underlying the functioning of the sensor) to black-box (based on replicating the I/O characteristic of a sensor) approaches [12], which convey a different level of abstraction information. While white-box sensor models can be validated by directly comparing the generated point cloud with the one gathered using a real

LiDAR/RADAR, black-box models need the detection layer to be included in validation methodology as they only provide obstacle-level information [13]. Moreover, no acceptance threshold has been commonly established [13] in order to deem a model as valid or invalid. Despite the validation of a sensor model has a huge potential to enhance the credibility of a virtual toolchain, we believe that such approaches are not yet mature enough for our purposes. A reason which has led us to step away from the submodels-based approach to resort to the integrated-level solution.

2) VEHICLE-SYSTEM

The validation of virtual vehicle models is a well-accomplished task which is reflected by technical standards supporting the procedures, such as [14] for the Lane Keep Assist application. The standards specify the maneuvers to be carried with the physical vehicle and the corresponding KPIs to be measured to validate the virtual realization. Fidelity levels are also pointed out in the recent standard [15].

In the scientific literature, a comprehensive review on modeling and validation methods for vehicle dynamics is given in [16]. The work emphasizes how virtual models for vehicle dynamics originate from two applications: simulation models supporting the development of production vehicles or artifacts for driving simulator platforms. Within the first class, models typically target a specific use case such as handling or riding studies.

More recently, [17] proposed a framework that accomplishes the validation via fitting the same model on a set of vehicles. The statistical evidence generated ensures the suitable capability of the model to predict correct results for different calibrations.

Validation concepts that go beyond the simulation model's output *vs.* real-world data comparison were introduced in [18]. In particular, the work of Kutluay and Winner surveyed the [16], [17]-identified literature from the point of view of the uncertainty estimation to the end of establishing how validation within the vehicle dynamics modeling field is lagging behind state-of-the-art validation methodologies. Ultimately, statistical methods are identified to alleviate the lack of credibility that characterizes the traditional validation pipeline.

Our scientific effort does not require the individual validation of a vehicle dynamical model given the adoption of a dyno-chassis, which only demands the coast-down curve's coefficients as an input, similarly to [7]. Naturally, the coast-down curve cannot accurately reproduce transient effects as a properly tuned dynamical model would do. That is a known limitation of our setup which limits the domain of application to a selection of scenarios involving mild decelerations as detailed in Section IV-E.

III. METHODOLOGY

A schematic representation of the activity carried out is displayed in Fig. 2. Firstly, a reference validation scenario is defined and executed in the proving ground (PG). In the PG

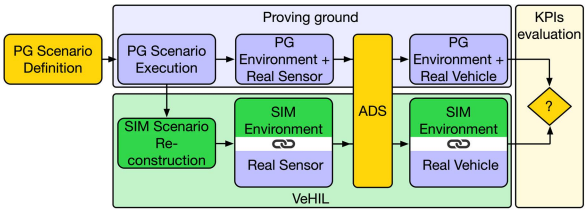


FIGURE 2. Schematic view of the pursued activity.

stream (violet cells in Fig. 2), every component of the testing pipeline is real, and no simulation/data-injection occurs as explained in Section III-A.

Afterward, based on the data collected through the GNSS units, we reconstructed the executed scenario and coded that into a formulation readable by the simulation environment for the VeHIL replication. The realization of the testing platform is explained in Section III-B. A scrupulous fulfillment of such a procedure was necessary to ensure that the very same PG scenario is replicated in the VeHIL setup as any discrepancy in the scenario reconstruction will affect the validation exercise.

The virtual execution, bottom stream in Fig. 2, is characterized by a mixture of simulated and physical components, a combination which is stressed by the green/blue blocks. The detailed explanation of the coupling mechanism is given in Section III-B.

To ensure consistency in the driving behavior, the very same ADS algorithm and controllers' tuning parameters are maintained in both the PG and VeHIL cases. Eventually, the fidelity assessment was carried out at the end of the testing campaign based on the correlation methods detailed in Section III-C.

A. TESTING CAMPAIGN

The vehicle used for the study was a robotized SMART ForTwo provided by BME Automated Drive,[1] which was rented for two weeks and made available at the Ispra (VA, Italy) site of the European Commission Joint Research Center (JRC). The robotized vehicle was tested against several reference scenarios for a total recorded distance of $\approx$ 70 km traveled in about 6.5 h. More specifically, the focus was placed on longitudinal maneuvers such as

free-flow driving : the ego vehicle starts from zero velocity and accelerates to a user-defined speed;

car-follow : the ADS adjusts its traveling speed according to a leader driving policy to maintain a safe distance;

stop & go : similar to car-follow but the leader reduces the velocity up to a full stop and then starts again.

An extensive data-set of maneuver was collected via instrumenting both the robotized vehicle and the supporting vehicles acting as targets with Ublox[2] GNSS units running at 10 Hz with except for the "free-flow" tests where the onboard velocity measuring system was enabled.

[1] https://www.automateddrive.bme.hu/
[2] https://www.u-blox.com/en

The robotized vehicle sensor's setup included a front-facing Mobileye® camera, two side-mounted LiDARs, and a rear-facing RADAR. In our tests, only the camera has been enabled. The choice was due to the later VeHIL replication setup, which relied entirely on the camera stimulation to provide the vehicle with the scenario information. By ruling out the LiDARs and the RADAR, we ensured that no additional effect in the software planning stack could have been brought by sensors that we would not stimulate afterward. The data collection and ADS functionalities were made possible thanks to an embedded dSPACE AutoBox.[3]

The camera-based car-following functionality proved to behave consistently only in the straight portions of the proving ground. In fact, while approaching curves, the camera frequently lost the tracking of the moving leader, thus triggering the free-flow behavior. This resulted in frequent jumps in velocity, which are particularly detectable in the ''Periferica'' portion of the JRC's proving ground as shown in Fig. 3. The discontinuities in the behavior are indeed an undesirable feature for an ADS. Nevertheless, they provide peculiar patterns that can be exploited later to validate the virtual test Fig. 8.

B. VEHIL REPLICATION

The virtual environment for the replication of the experimental campaign is a VeHIL based on the DRIVINGCUBE™ technology, similarly to [7]. Our solution, however, differs from the cited contribution since we used *camera stimulation* rather than *signal injection* and a different ADS. According to our setup, a wide-screen monitor has been placed in front of the robotized vehicle's windscreen, where the camera enabling ADS functionalities is installed as in Fig. 4.

The screen is rendering in real-time the photorealistic simulation provided by the simulation environment Vires VTD®. The simulator was interfaced to the dyno-chassis by means of the DRIVINGCUBE™ technology, which relies, as middleware software, on:

- AVL Model.CONNECT™: a co-simulation tool which enables the duplex communication between the Vires simulator and
- AVL Testbed.CONNECT™: the real-time module controlling the dyno-chassis based on the vehicle velocity.

The simulation point of view was adjusted according to the camera installation position, and the corresponding field-of-view (FOV) calibration procedure was carried out to ensure the distance estimation from the Mobileye camera was correct.

The VeHIL setup allowed the robotized vehicle to freely adjust the velocity on the dyno-chassis based on distance from the vehicle ahead predicted by the camera. However, no automated nor manual steering was possible on the platform. As such, the simulation setup had been adjusted to account for the steering limitation. In particular, in the simulation environment, the robotized vehicle traveled along a path that had been reconstructed from the GNSS units and appropriately converted into the OpenDRIVE® standard. Hence, the lateral dynamics of the vehicle is not part of the validation procedure. Eventually, the targets' trajectories were reconstructed from the GNSS logs and replicated identically in the simulation. Accurate targets' replication resulted to be a crucial aspect to reproduce the velocity oscillations in Fig. 3 given the sensitivity of the ADS system.

[3]https://www.dspace.com/en/ltd/home/products/hw/accessories/autobox.cfm#179_25444

C. CORRELATION CRITERIA AND KPIS

The modeling choices adopted when translating the proving ground scenarios to the semi-virtual environment resulted in the VeHIL being a one degree of freedom testing environment. The simple setup enabled us to investigate the environment's fidelity by only checking the traveling velocity V and the longitudinal acceleration a_x as validation KPIs.

1) GOODNESS-OF-FIT ESTIMATION

The definition of KPIs has to be supported by suitable computational tools to provide a quantitative evaluation of the fidelity level. To this end, we adopted as metric the Normalized Root Mean Square Error (NRMSE)

$$\frac{1}{\sigma_{\text{pg}}}\sqrt{\frac{\sum_{i=1}^{N}\left(x_{\text{sim},i}-x_{\text{pg},i}\right)^2}{N}}, \tag{1}$$

the average standard deviation (mainly used as a comparison tool with respect to the work in [7]), the Pearson correlation $r(x_{\text{sim}}, x_{\text{pg}})$

$$\frac{\sum_{i=1}^{N}\left(x_{\text{sim},i}-\bar{x}_{\text{sim}}\right)\left(x_{\text{pg},i}-\bar{x}_{\text{pg}}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_{\text{sim},i}-\bar{x}_{\text{sim}}\right)^2}\sqrt{\sum_{i=1}^{N}\left(x_{\text{pg},i}-\bar{x}_{\text{pg}}\right)^2}}, \tag{2}$$

and the coefficient of determination R^2. In (1)-(2), the x_{sim} is the data sample coming for the VeHIL environment (either velocity V or the acceleration a_x), x_{pg} the data sample recorded on the proving ground, and N the total number of samples in the considered experiment.

In parallel to the widely adopted metrics in (1) and (2), this paper adopts also state-of-the-art techniques to compare time-series in terms of their *phase* difference

$$d_P = \frac{1}{\pi}\arccos\left(\frac{\sum_i^N x_{\text{sim},i}\,x_{\text{pg},i}}{\sqrt{\sum_i^N x_{\text{sim},i}^2\sum_i^N x_{\text{pg},i}^2}}\right), \tag{3}$$

and (integral) magnitude dissimilarity

$$d_M = \sqrt{\frac{\sum_i^N x_{\text{sim},i}^2}{\sum_i^N x_{\text{pg},i}^2}} - 1. \tag{4}$$

The two criteria in (3) and (4) can be combined into a unique indicator

$$d_{SG} = \sqrt{d_M^2 + d_P^2}, \tag{5}$$

in the Sprague-Geers metric [19]. (5) belongs to a set of methodologies comparing time-series known as ''Magnitude

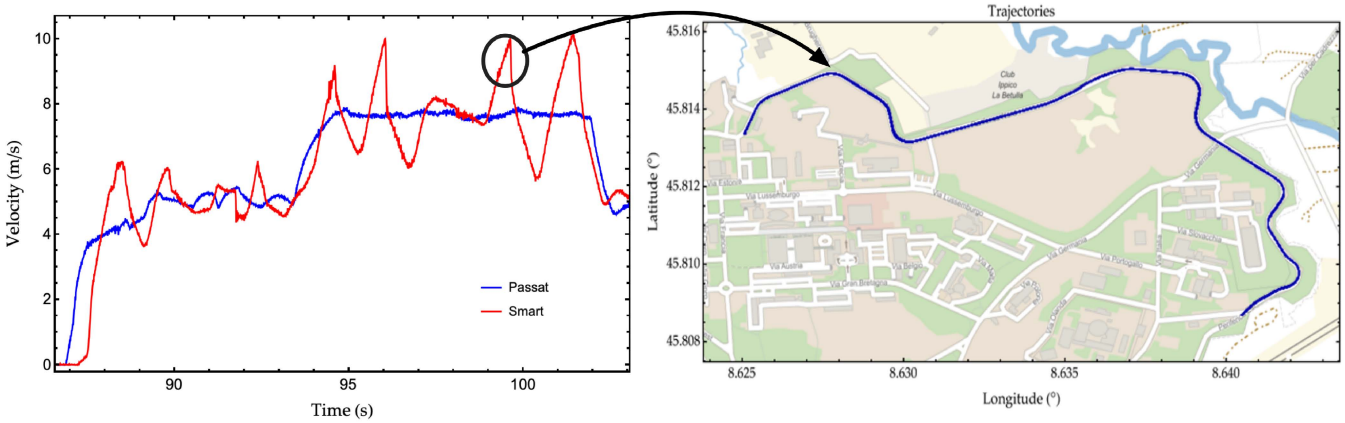


FIGURE 3. Robotized vehicle loss of target tracking and induced velocity jumps in curves (left). Corresponding "Periferica" map on the right. The ego vehicle path is represented by the solid blue line in the right chart.

FIGURE 4. DRIVINGCUBE^{TM} VeHIL testing environment.

Phase Composite" (MPC) and has been widely recommended in several literature works [20]–[22] when the signals to compare are time-dependent such as the case of study. In particular, (4) is not affected by any time-shift difference of the signals as it only evaluates the areas behind the curves. That is practical when measuring the discrepancy of time-series reaching different amplitude (for instance the acceleration signal recorded in a simulated *vs.* laboratory crash experiment as in [20]). On the other side, (3) is nearly insensitive to a magnitude divergence providing instead a quantitative measure of phase-shift, which is a particularly efficient way to discriminate different reaction times of two systems [22]. Similarly to the NRMSE, the optimal agreement is obtained when both d_M and d_P are zero.

2) STATISTICAL HYPOTHESIS TESTING

Model validation can be framed within a statistical hypothesis testing perspective [23]. This is particularly effective when models are affected by stochastic effects such as the VeHIL environment. Statistical testing goes beyond goodness-of-fit by inspecting, based on the outcomes' distributions, whether:

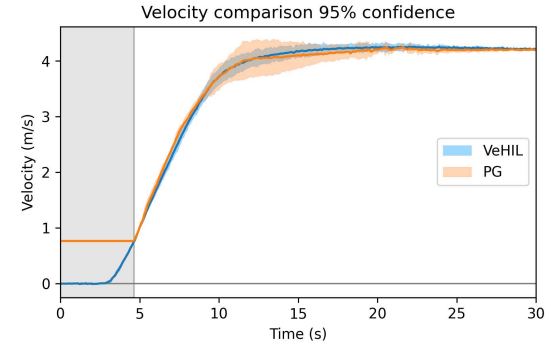
- H_0 the model is an accurate representation of the real process (null hypothesis);
- H_a the model is *not* an accurate representation of the real process (alternative hypothesis).

A widespread solution in literature is the adoption of the two-sample T-test [23]–[25]. Such a test investigates if the two populations' means are equal (H_0) or not (H_a). Alternatively, the Kolmogorov-Smirnov (KS) test [26] can be exploited as a valuable tool to provide a quantitative assessment of the distance between the empirical distribution functions for a similar purpose. In contrast to the T-test, which assumes a normal distribution of the input data, the KS-test is distribution-free. Based on the KS-test, an assessor can judge whether or not to accept the null hypothesis asserting that the two samples derive from the same distributions. In both tests, the computed $p-$ value is a direct measure that allows rejecting H_0 if less than a prescribed significance level. Commonly, a $p-$ value threshold value 0.05% is adopted: a higher figure suggests there is evidence supporting H_0, whereas, a lower value advises rejecting H_0 (not accepting H_a though).

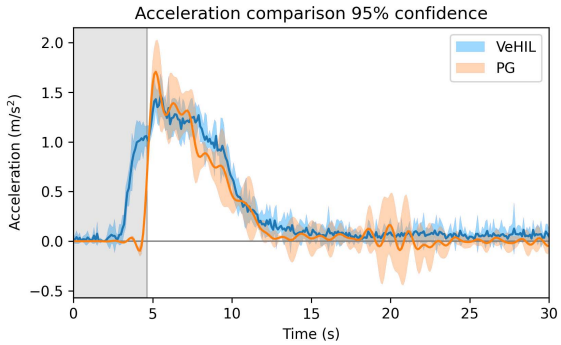
The mentioned statistical tools are typically adopted following the application of an aggregation operator such as the `mean` operator [27]. Data aggregation allows ruling out the time dependency of the signal in a way that a stationary process is obtained and conventional statistical hypothesis testing can be carried out. Additionally, the time-series are strongly autocorrelated as they result from the motion of a causal system. Hence, statistical testing cannot perform at best without any aggregation procedure due to the underlying assumption of independent observations [28].

IV. FIDELITY LEVEL DETERMINATION

This Section applies the methodology outlined in Section III-C to all the concrete driving scenarios recorded. Before any of the computational tool is applied, the data sequences had to be properly time-aligned using the Time of Arrival criterion (ToA) [20], [22]. ToA requires that the signals are synchronized upon reaching for the first time a reference amplitude (a 1 m/s reference velocity was used in this work).



(a) Velocity



(b) Acceleration

FIGURE 5. Free-flow scenario qualitative correlations.

A. FREE-FLOW ANALYSIS

The first scenario had the main purpose of studying the repeatability of the system, both on the proving ground and on the VeHIL. To this end, a simple free-flow (FF) maneuver was performed in which the robotized vehicle traveled from standing still to 4 m/s. 5 repetitions were recorded for both real-world tests and on the dyno-chassis.

The results in terms of longitudinal velocity V and longitudinal acceleration a_x are graphically reported in Fig. 5. Due to the onboard velocity measuring system limitation (the minimum velocity that could be recorded using wheel odometry was $\approx$ 0.8 m/s), the correlation analysis is performed only after the VeHIL system has reached the minimum logged velocity. The dark gray area in Fig. 5 is representative of the data portion excluded from the validation for the explained reason.

In Fig. 5, the solid blue line is the mean velocity profile of the 5 repetitions recorded on the DRIVINGCUBE$^{\text{TM}}$, the light blue area the 95% confidence region for the VeHIL case, the solid orange line the mean velocity of the 5 repetitions recorded on the PG, and the light orange area the 95% confidence interval of the real-world tests. As expected, a larger dispersion of data characterizes the real-world data with respect to the VeHIL case, where the more protected test environment conditions allowed for narrower confidence regions.

Quantitative metrics derived from the time-series depicted in Fig. 5 are instead reported in Table 1. It is worth noticing how the $\bar{\sigma}$ and the Pearson correlation are comparable

TABLE 1. Validation KPIs for the free-flow scenario.

	NRMSE	$\bar{\sigma}$	Pearson	R^2	Sprague-Geers	
					d_M	d_P
V	0.083	0.040 m/s	0.997	0.993	9.00e-4	0.003
a_x	0.290	0.066 m/s^2	0.977	0.916	0.0324	0.081

TABLE 2. Aggregate metrics for the free-flow scenario.

Signal	N (-)	Mean (m/s)	$\bar{\sigma}$ (m/s)
V_{VeHIL}	5	4.00412	0.0167
V_{PG}	5	4.00346	0.0249

with the results in [7, Table 7], where an average standard deviation equal to $\bar{\sigma}_V = 0.0716$ for the velocity and $\bar{\sigma}_{a_x} = 0.0592$ for the acceleration have been reported. This finding is particularly fascinating as the setup in the cited work (albeit a different robotized vehicle was used) is extremely similar to the one here presented with expect for our adoption of sensor stimulation which stands apart from the signal injection in [7]. Nonetheless, the camera stimulation does not play a role in the free-flow tests; hence its contribution to the KPIs is not accounted for. That justifies the very similar fidelity level obtained in the present work with respect to [7] despite the additional element making up the testing toolchain.

The free-flow scenario also afforded the possibility of executing two-sample statistical testing given the availability of multiple repetitions for both the proving ground and the semi-virtual cases following the application of the `mean` operator. Aggregating each experiment using the mean velocity value yields the metrics in Table 2, where the third column represents the mean of each test's average velocity and the fourth column the standard deviation of the means for both the VeHIL (first row) and PG cases. As graphically shown in Fig. 5, the PG tests have a larger dispersion of data which is reflected by the larger standard deviation of PG with respect to the VeHIL in Table 2.

Applying the T- and KS-tests to the aggregated signals yields the first row of Table 3. The H_0 cannot be rejected according to both the T- and KS-tests for the aggregated data case since both the $p-$ values are higher than any reasonable significance level. Hence, the virtual test can be considered *valid*. On the contrary, the second row reports the application of the same tests without any data aggregation procedure. Table 3 clearly shows how a model's validity is dramatically affected by any data conditioning procedures and clustering. In addition, the level of detail of the validation procedure plays a significant role since T-test alone would have promoted the model even without aggregation, whereas the additional KS-test questions the validity of the virtual environment. Similar considerations have also been highlighted in [27] for the validation of a traffic simulation environment.

Another point of concern is related to the fact that averaging the velocity time-series implies weighting more the

TABLE 3. Statistical metrics for the free-flow scenario.

Signal	N	T-test		KS-test	
		$p-$value	T	$p-$value	D
$\bar{V}$	5	0.968	0.042	0.873	0.350
$V(t)$	1100	0.504	0.668	2.108e-10	0.176

TABLE 4. Calibration metrics of (6) for respectively "VeHIL" and "PG" cases.

Signal	N (-)	mean (s)	$\bar{\sigma}$ (s)	V RMSE (m/s)	R^2
τ_{VeHIL}	5	1.48760	0.09176	0.10572	0.95455
τ_{PG}	5	1.51281	0.10357	0.06856	0.97800

TABLE 5. Statistical metrics for the "free-flow" (FF) scenarios rising time.

Signal	N	T-test		KS-test	
		$p-$value	T	$p-$value	D
τ	5	0.743	-0.342	0.873	0.350

steady-state regime (velocity after 15 s) with respect to the transient acceleration in the validation procedure. We argue that we can restore the importance of the transient phase by estimating the rising time τ of each time-series and applying statistical testing on the obtained τ distributions. Given that the system demonstrated no appreciable velocity overshoot for the free-flow tests as in Fig. 5 (a), we can conveniently estimate the rising time with a first-order dynamical system where the velocity is given by

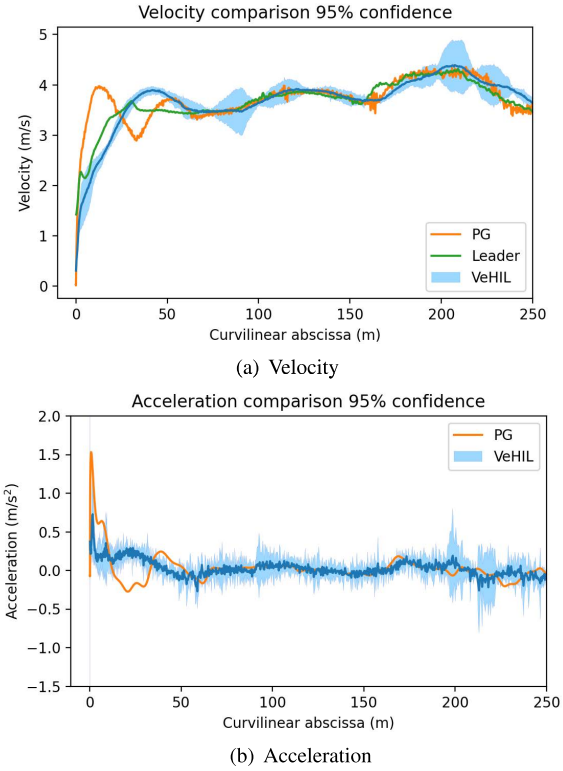
$$v(t) = v_F + (v_0 - v_F)e^{-\frac{t}{\tau}}. \tag{6}$$

In (6), v_0 is the minimum measurable velocity, v_F is the final velocity after the system has converged to the steady-state and τ is the parameter we seek to identify. The calibration of (6) on the recorded time-series yielded the parameters (mean and standard deviation of τ) and residual metrics (RMSE of velocity error predicted by the model $\hat{V}$ and R^2) reported in Table 4. Overall, the residuals are such that the model can be safely adopted to the end of estimating the rising time for the later validation analysis.

Eventually, the two-sample T- and KS-tests can be applied to the distributions of τ_{VeHIL} and τ_{PG}. The resultant $p-$ values are reported in Table 5 and show how the rise-time aggregation criterion allows the model to pass the validation tests.

B. SIMPLE CAR-FOLLOWING SCENARIO

The second case-of-study for the VeHIL environment was the replication of a "simple car-following" scenario recorded on a straight portion of the proving ground. Two reference cases are presented here. The first, in Fig. 6, shows a minor inconsistency in the initial conditions of the simulation. The discrepancy results in a mismatch that the ADS compensates while the virtual test evolves until the simulation "converges" to the real-world test. The second, in Fig. 7, shows a better qualitative and quantitative agreement of the



(a) Velocity

(b) Acceleration

FIGURE 6. CF-1 scenario qualitative correlations.

TABLE 6. Validation KPIs for the CF-1 scenario.

	NRMSE	$\bar{\sigma}$	Pearson	R^2	Sprague-Geers	
					d_M	d_P
V	0.537	0.142 m/s	0.911	0.827	8.408e-3	0.013
a_x	0.988	0.062 m/s^2	0.461	0.248	-0.273	0.340

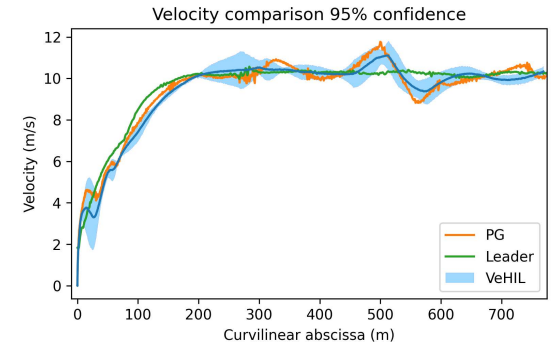
signals throughout the entire duration of the experiment due to the correct initialization of the simulation setup.

Similarly to Section IV-A, several repetitions have been performed for the VeHIL setup, which are emphasized by the 95% confidence interval (light blue areas) in Figs. 6 and 7. However, given the unavailability of a robotized moving target, there was no possibility of identically repeating the tests on the proving ground; hence only one repetition is given for the real-world experiments.

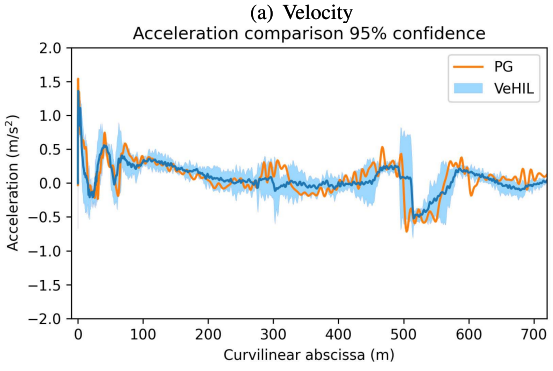
Overall, a larger spread of the VeHIL velocities characterizes Figs. 6 and 7 with respect to Fig. 5, which could be attributed to the higher complexity of the car-following driving scenario triggering some velocity instability phenomena as discussed in Fig. 3.

The quantitative assessment of the fidelity level is reported in Table 6 analogously to Table 1. In this case, given the slight mismatch in the initial conditions and the introduction of sensor stimulation, worse KPIs are obtained.

On the other side, the second car-following scenario presented shows better metrics, as documented in Table 7. Also worthy of consideration is the speed bound at $\approx$ 500 m in Fig. 7. This is due to the tight radius roundabout



(a) Velocity



(b) Acceleration

FIGURE 7. CF-2 scenario qualitative correlations.

TABLE 7. Validation KPIs for the CF-2 scenario.

	NRMSE	$\bar{\sigma}$	Pearson	R^2	Sprague-Geers	
					d_M	d_P
V	0.184	0.361 m/s	0.990	0.966	-5.251e-3	0.012
a_x	0.676	0.166 m/s^2	0.738	0.543	-0.189	0.217

of the alternative track used for the proving ground tests which caused consistent loss of the moving target and the consequent acceleration. As previously pointed out, this behavior might be unacceptable from an ADS developer perspective. Nevertheless, for the sake of this work, this is a welcome feature as it allows to appreciate the capability of the semi-virtual environment to replicate an "unstable" phenomena.

Table 8 presents the results of applying the hypothesis testing tools to the CF-1 (first and second row) and CF-2 (third and fourth row) scenarios. Differently from Table 3, for the aggregated mean velocity cases (first and third row) the one-sample T-test was applied given the lack of a distribution of means for the PG executions. In Table 8, a similar trend as Table 3 emerges, where the aggregated cases show greater evidence in accepting H_0. In particular, given the discussed mismatch in the initial conditions, CF-1 demonstrates a significantly lower $p-$ value with respect to both CF-2 and free-flow cases, which may lead to the decision of rejecting the null hypothesis. Conversely, CF-2 managed to pass the T-test for both the aggregated and non aggregated cases thus demonstrating the goodness of the VeHIL provided that the scenario reconstruction has been performed carefully.

TABLE 8. Statistical metrics for the simple car-following scenarios.

Signal	N	T-test		KS-test	
		$p-$value	T	$p-$value	D
$\bar{V}_{\text{CF-1}}$	10	0.094	2.421		
$V(t)_{\text{CF-1}}$	760	0.046	-1.999	5.154e-5	0.151
$\bar{V}_{\text{CF-2}}$	10	0.702	-0.412		
$V(t)_{\text{CF-2}}$	790	0.395	0.851	8.655e-11	0.173

TABLE 9. Validation KPIs for the "periferica" scenario.

	NRMSE	$\bar{\sigma}$	Pearson	R^2	Sprague-Geers	
					d_M	d_P
V	1.078	0.860 m/s	0.748	0.498	6.436e-3	0.041
a_x	2.759	0.492 m/s^2	0.644	0.269	-0.016	0.285

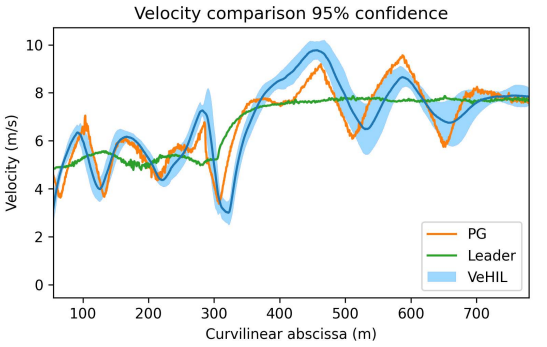
C. "PERIFERICA" ANALYSIS

A more challenging car-following scenario is discussed here where the robotized vehicle was requested to follow a leading car on the curvy road "Periferica". The scenario is particularly interesting as it highlights the importance of the ADS in the integrated test-setup validation. Indeed, a well-performing ADS ready for market introduction shall be able to deal with the temporary loss of the leader's tracking with no unreasonable acceleration. However, this desirable feature from a user perspective might induce an over-reliance of simulation as the damping effect played by the ADS results in reduced discrepancies between virtual and proving ground data. On the other hand, a particularly sensitive system such as the one implemented in our test vehicle allows highlighting subtle differences in the functioning of the two testing environments, thus representing the ideal ADS candidate for validating the environment.

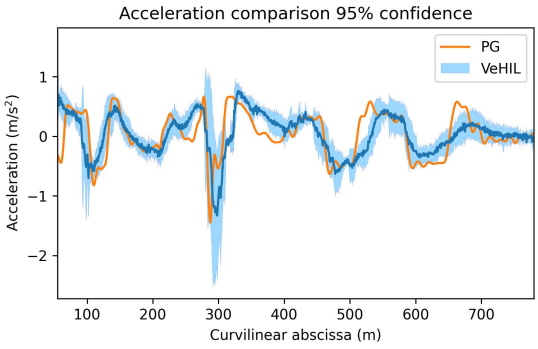
A qualitative assessment of the obtained correlations using the VeHIL environment is provided in Fig. 8. Despite the large dispersion of data (the confidence intervals in Fig. 8 are considerably larger than the intervals in Fig. 6 and Fig. 7), the average VeHIL's trend (solid blue line) follows the proving ground's velocity and acceleration time-series closely. Every jump in the speed profile is grasped by the virtual system, albeit the actual amplitude differs in every repetition due to the instability of the phenomenon.

The corresponding numerical KPIs are reported in Table 9. The more challenging driving scenario in terms of virtual replication impacts quite significantly the validation metrics. In particular, the R^2 metric shows a substantial reduction. On the other side, the Pearson correlation of velocity still returns an acceptable value.

It is also worth noticing how minor velocity deviations at the beginning of the experiments accumulate in the continuation of the tests. When dealing with longer time-series, loss of information synchronization between virtual and physically derived data may require advanced techniques such as Dynamic Time Warping (DTW) [29].



(a) Velocity



(b) Acceleration

FIGURE 8. "Periferica" scenario qualitative correlations.

TABLE 10. Statistical metrics for the "periferica" scenario.

Signal	N	T-test		KS-test	
		$p-$value	T	$p-$value	D
$\bar{V}$	10	0.501	0.7053		
$V(t)$	650	0.052	-1.999	5.801e-5	0.125

The statistical metrics for "periferica" (PF) scenario are reported in Table 10. The $p-$ value associated to the aggregation case (first row) allows accepting the null hypothesis whereas providing the full data sequence to the T-test questions the validity of H_0. Similarly to Table 8, the one-sample T-test was used to compute the first row of Table 10.

In addition to time domain metrics, frequency domain techniques are reported to be valuable validation tools especially when vehicle dynamics is in action [16]. To this end, we computed the power spectral density (PSD) of the velocities recorded on VeHIL and PG and displayed in Fig. 9. From Fig. 9, the string-unstable behavior that characterizes the VuT [30] can be clearly grasped from the resonance pick. Moreover, the signals show a clear qualitative agreement up to the string un-stable resonance pick before the impact of windowing and conditioning overcomes the signals' power.

D. STOP-AND-GO ANALYSIS

The last virtual replication discussed is concerned with the stop-and-go (S&G) scenario. The criticality of the S&G lies in the stronger leader's deceleration applied in this

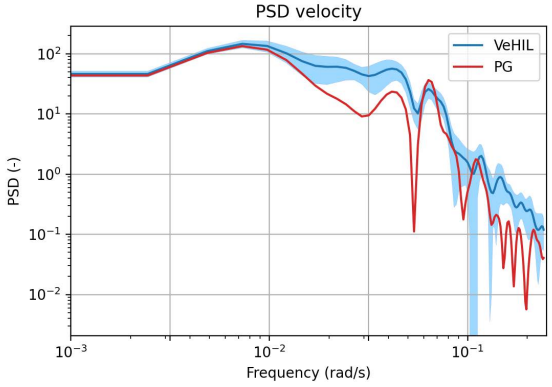
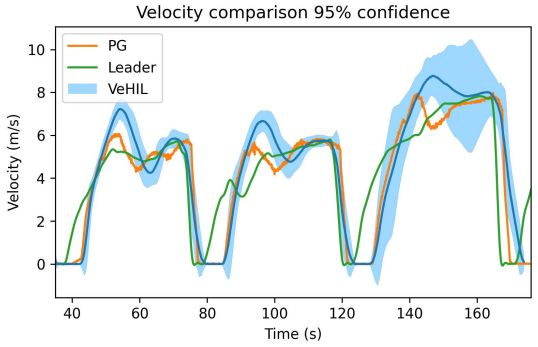
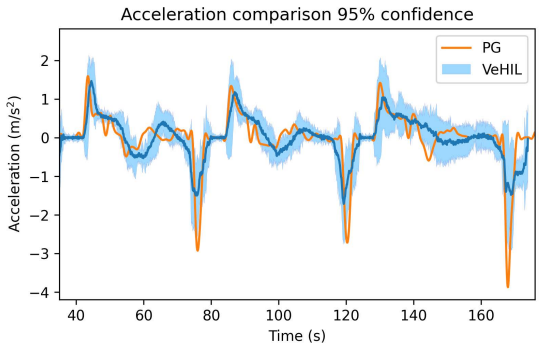


FIGURE 9. Velocities PSD for the "Periferica" scenario.



(a) Velocity



(b) Acceleration

FIGURE 10. Stop & Go scenario qualitative correlations.

test compared to the other tests and in the frequent starts from stop where friction and clutch engagement play a significant role.

A graphical correlation for the S&G scenario is proposed in Fig. 10. The charts show how an initial mismatch is accumulated after every standing still departure due to the lower deceleration achieved on the dyno-chassis with respect to the proving ground, which causes the ego vehicle to stop closer to the leader.

The quantitative metrics for the S&G scenario are given in Table 11. The NRMSE reported in Table 11 exceeds by far the equivalent metric obtained for the rest of the driving scenarios described. Given the large discrepancy in the absolute magnitude of the signals, no statistical evaluation has been carried out.

TABLE 11. Validation KPIs for the S&G scenario.

	NRMSE	$\bar{\sigma}$	Pearson	R^2	Sprague-Geers	
					d_M	d_P
V	1.523	0.860 m/s	0.951	0.895	0.066	0.051
a_x	2.015	0.448 m/s^2	0.752	0.275	-0.225	0.229

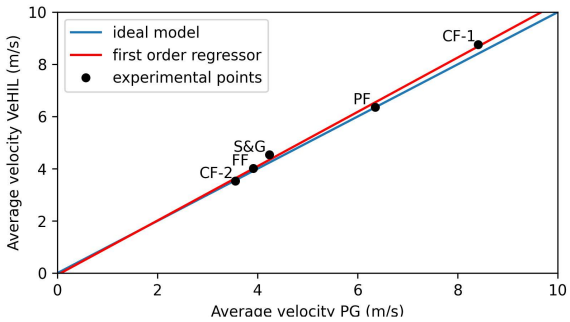


FIGURE 11. First-order regressor average velocity.

E. GLOBAL ASSESSMENT

Considering the traveling speed as the most relevant quantity, one can cluster the mean velocities recorded for each scenario for both testing environments and report the obtained quantities on a scatter chart. The data aggregation procedure relies once again on the `mean` operator coherently with the rest of the Section. An ideal simulation model should provide that a first-order fitting of the data scatter is characterized by a unit slope and zero intercept straight line. Such a perfect virtual replication is depicted in the blue line in Fig. 11. In the same figure, the black dots are the actual data samples obtained for the considered scenarios, whereas the red line the actual setup first-order fitting.

The fitted straight line in Fig. 11 has a slope $m = 1.04121$ and intercept $y_0 = -0.06966$. The small deviation between the actual fitted line and theoretical optimum is another confirmation of the VeHIL goodness in replicating the real-world tests. Unfortunately, no other work is available in the scientific literature where the same setup as ours is used. Hence the global metrics obtained cannot be benchmarked against independent data.

The final step in the validation procedure is establishing the *domain* of validity as absolute validation is not attainable [31]. Given the restricted dimensionality of the case-of-study, one can formulate the validation domain in the velocity-acceleration plane as in Fig. 12. Fig. 12 displays all the $(V(t_i), a_x(t_i))$ data samples collected in the testing campaign and assigns a color to each point depending on the corresponding absolute velocity error based on the VeHIL replication.

In Fig. 12, the higher speed data samples show an absolute velocity error < 1 m/s. Vice versa, a larger error is recorded while performing stopping maneuvers as of the dyno-chassis's limitations. One can thus deem the VeHIL not an appropriate tool for the virtual validation of emergency braking-like functionalities and restrict the validity domain

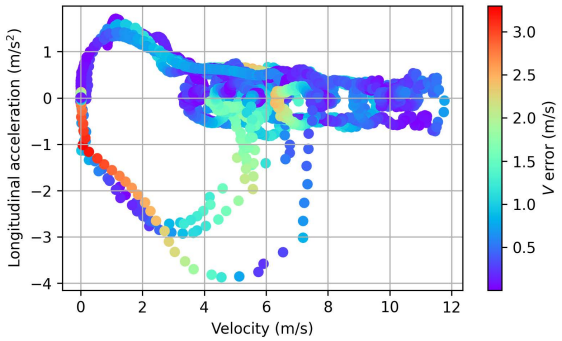


FIGURE 12. Velocity absolute error over the validation domain.

to a portion of Fig. 12 which satisfies the velocity error threshold.

V. CONCLUSION

In the present paper, the authors have tried to define the fidelity level that can be achieved with state-of-the-art commercial tools for virtual testing of low-automation level ADSs. The assessment was carried out by replicating a number of real-world driving scenarios on the VeHIL test-bench.

Validating the described VeHIL environment has proven a challenging activity due to the sensitivity of the ADS to small discrepancies in the formulation of the virtual scenario, confirming what is reported in the available literature [32].

Another point of concern was identified in the selection of the statistical tools to perform hypothesis testing and the impact that aggregation operators play on the validation results. In Section IV-A, we presented two solutions to aggregate data dealing with different aspects of the system: the `mean` operator, which is mostly concerned with the steady-state value reached at the end of the test, and the estimation of the rising-times distribution as an alternative aggregation operation to better convey the transients portion of the signals.

Considering the scenarios that did not allow exact replication of the proving ground, we studied the VeHIL system's repeatability and computed the confidence intervals. Next, we carried out the validation via computing the NRMSE, σ, Pearson, R^2, and one sample statistical testing. Overall, the agreement between virtual testing and proving ground generated data was extremely high for all the scenarios but the S&G one. This is also in line with the general observed degradation of fidelity metrics with the increase of the driving scenario complexity. Nonetheless, the exact degree of fidelity needed to validate a solution will likely depend on the criticality of technology involved, thus limiting the general consideration that this paper can deliver since no claim of absolute validation is possible [31].

Moreover, the considerations here reported are only valid for the given setup as several factors, including the tuning of the camera stimulation system and the coast-down curve's parameters for the dyno-roller bench, ultimately affect the

validation exercise. Thus, a true characterization of VeHIL's fidelity level can only be carried out by repeating a similar testing campaign with other vehicles and, possibly, another team setting the system to rule out any human/technological factor which might have potentially impacted on the experiments.

ACKNOWLEDGMENT

The authors are grateful to the JRC colleagues who supported the experimental campaign: R. Suarez Bertoa, V. Padovan, and F. Re, and to the colleagues of the Vehicle Emissions Laboratory (VeLA-8) who supported the setup of VeHIL environment M. Otura, M. Centurelli, and C. Ferrarese.

REFERENCES

[1] United Nations Economic Commission for Europe. (2020). *Proposal for a New un Regulation on Uniform Provisions Concerning the Approval of Vehicles With Regards to Automated Lane Keeping System (ECE/TRANS/WP.29/2020/81)*. [Online]. Available: https://unece.org/sites/default/files/2021-01/ECE-TRANS-WP29-2021-017e_%0.pdf

[2] United Nations Economic Commission for Europe. (2020). *Proposal for a New un Regulation on Uniform Provisions Concerning the Approval of Vehicles With Regards to Automated Lane Keeping System (ECE/TRANS/WP.29/2020/81)*. [Online]. Available: https://unece.org/sites/default/files/2021-01/ECE-TRANS-WP29-2021-017e_%0.pdf

[3] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transp. Res. A, Policy Pract.*, vol. 94, pp. 182–193, Dec. 2016.

[4] H. Abdellatif and C. Gnandt, "Use of simulation for the homologation of automated driving functions," *ATZelectronics worldwide*, vol. 14, no. 12, pp. 68–71, Dec. 2019.

[5] T. Dueser and C. Gutenkuns, "A comprehensive approach for the validation of virtual testing toolchains," IAMTS, Chennai, India, Tech. Rep. n. IAMTS0001202104, Apr. 2021. [Online]. Available: https://iamts.sae-itc.com/binaries/content/assets/itc/content/iamts/iamts0001202104.pdf

[6] R. Dona and B. Ciuffo, "Virtual testing of automated driving systems. A survey on validation methods," *IEEE Access*, vol. 10, pp. 24349–24367, 2022.

[7] S. Riedmaier, J. Nesensohn, C. Gutenkunst, T. Düser, B. Schick, and H. Abdellatif, "Validation of X-in-the-loop approaches for virtual homologation of automated driving functions," in *Proc. 11th Graz Symp. Virtual Vehicle*, 2018, pp. pp. 1–12.

[8] A. Leitner and M. Paulweber. (2019). *Enable-S3 Summary of Results*. [Online]. Available: https://drive.google.com/file/d/15c1Oe69dpvW5dma8-uS8hev17x-6V3zU/view

[9] S. Riedmaier, D. Schneider, D. Watzenig, F. Diermeyer, and B. Schick, "Model validation and scenario selection for virtual-based homologation of automated vehicles," *Appl. Sci.*, vol. 11, no. 1, p. 35, Dec. 2020.

[10] IWG AEBS UTAC. *Validation Method: Virtual Testing*. Accessed: May 2, 2021. [Online]. Available: https://wiki.unece.org/download/attachments/101554586/AEBS-12-07%20%2%8UTAC%29%20Virtual%20testing%20AEBS.pdf?api=v2

[11] United Nations Economic Commission for Europe. (2017). *Uniform Provisions Concerning the Approval of Passenger Cars With Regard to Electronic Stability Control (ESC) Systems*. [Online]. Available: https://unece.org/fileadmin/DAM/trans/main/wp29/wp29regs/2017/R140e.pdf%

[12] P. Cao, W. Wachenfeld, and H. Winner, "Perception sensor modeling for virtual validation of automated driving," *IT Inf. Technol.*, vol. 57, no. 4, pp. 243–251, Aug. 2015.

[13] P. Rosenberger, J. T. Wendler, M. F. Holder, C. Linnhoff, M. Berghöfer, H. Winner, and M. Maurer, "Towards a generally accepted validation methodology for sensor models—Challenges, metrics, and first results," in *Proc. Event Title: Grazer Symp. Virtuelles Fahrzeug*, May 2019, pp. 1–13. [Online]. Available: http://tuprints.ulb.tu-darmstadt.de/8653/

[14] *Passenger Cars Vehicle Dynamic Simulation and Validation Steady-State Circular Driving Behaviour*, document ISO 19364, ISO, Tech. Rep., 2016.

[15] *Passenger Cars Simulation model Classification—Part 1: Vehicle Dynamics*, document ISO/DIS11010-1, ISO, Tech. Rep., 2022.

[16] E. Kutluay and H. Winner, "Validation of vehicle dynamics simulation models—A review," *Vehicle Syst. Dyn.*, vol. 52, no. 2, pp. 186–200, 2014.

[17] M. Viehof and H. Winner, "Research methodology for a new validation concept in vehicle dynamics," *Automot. Engine Technol.*, vol. 3, nos. 1–2, pp. 21–27, Aug. 2018.

[18] B. Danquah, S. Riedmaier, and M. Lienkamp, "Potential of statistical model verification, validation and uncertainty quantification in automotive vehicle dynamics simulations: A review," *Vehicle Syst. Dyn.*, vol. 60, pp. 1–30, Dec. 2020.

[19] M. A. Sprague and T. L. Geers, "A spectral-element method for modelling cavitation in transient fluid–structure interaction," *Int. J. Numer. Methods Eng.*, vol. 60, no. 15, pp. 2467–2499, Aug. 2004.

[20] M. H. Ray, M. Mongiardini, and C. Plaxico, "Quantitative methods for assessing similarity between computational results and full-scale crash tests," in *Proc. 91th Annu. Meeting Transp. Res. Board*, 2012, pp. 1–21.

[21] K. A. Maupin, L. P. Swiler, and N. W. Porter, "Validation metrics for deterministic and probabilistic data," *J. Verification, Validation Uncertainty Quantification*, vol. 3, no. 3, pp. 1–12, Sep. 2018.

[22] L. E. Schwer, "Validation metrics for response histories: Perspectives and case studies," *Eng. With Comput.*, vol. 23, no. 4, pp. 295–309, Oct. 2007.

[23] O. Balci and R. G. Sargent, "Some examples of simulation model validation using hypothesis testing," in *Proc. 14th Conf. Winter Simulatio*, vol. 2, Dec. 1982, pp. 621–629.

[24] G. W. Snedecor and W. G. Cochran, *Statistical Methods*, 8th ed. Ames: Iowa State Univ. Press, 1989.

[25] J. P. Kleijnen, "Validation of models: Statistical techniques and data availability," in *Proc. 31st Conf. Winter Simulation Simulation Bridge Future*, vol. 1, 1999, pp. 647–654.

[26] F. J. Massey, Jr., "The Kolmogorov–Smirnov test for goodness of fit," *J. Amer. Stat. Assoc.*, vol. 46, no. 253, pp. 68–78, Mar. 1951.

[27] L. Rao, L. Owen, and D. Goldsman, "Development and application of a validation framework for traffic simulation models," in *Proc. Simulation Conf. Winter*, vol. 2, Dec. 1998, pp. 1079–1086.

[28] D. Murray_smith, "Methods for the external validation of contiuous system simulation models: A review," *Math. Comput. Model. Dyn. Syst.*, vol. 4, no. 1, pp. 5–31, 1998.

[29] L. Rabiner, *Fundamentals of Speech Recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[30] M. Makridis, K. Mattas, B. Ciuffo, F. Re, A. Kriston, F. Minarini, and G. Rognelund, "Empirical study on the properties of adaptive cruise control systems and their impact on traffic flow and string stability," *Transp. Res. Rec.*, vol. 2674, no. 4, pp. 471–484, 2020.

[31] W. L. Oberkampf and T. G. Trucano, "Verification and validation in computational fluid dynamics," *Prog. Aerosp. Sci.*, vol. 38, no. 3, pp. 209–272, 2002.

[32] K. Groh, S. Wagner, T. Kuehbeck, and A. Knoll, "Simulation and its contribution to evaluate highly automated driving functions," *SAE Int. J. Adv. Current Practices Mobility*, vol. 1, pp. 539–549, Apr. 2019.

RICCARDO DONÀ received the B.Sc. and M.Sc. degrees in mechatronics engineering and the Ph.D. degree in materials, mechatronics and systems engineering (as part of the EU research project Dreams4Cars) from the University of Trento, in 2014, 2017, and 2021, respectively. Currently, he is an External Consultant with Uni Systems Italy, where he serves as a Virtual Testing Expert for the Joint Research Centre for the European Commission. In particular, he supports the development of the traffic models and virtual testing. His scientific research interests include simulation and control algorithms for autonomous driving vehicles.

SÁNDOR VASS received the M.Sc. degree in mechanical engineering and the Ph.D. degree in mechanical and automotive engineering from the Department of Automotive Technologies, Budapest University of Technology and Economics, in 2011 and 2020, respectively. Since then, he has been working on research and development of automotive systems. In 2020, he joined the European Commission–Joint Research Centre, Italy, where he supports the research made on testing and validation processes for automated vehicles. While in the past, he was working on internal combustion engines and transmission systems development. His research interests include automated driving systems and functions and testing.

KONSTANTINOS MATTAS received the B.Eng. degree in civil engineering (specialized in transportation), the M.Sc. degree in applied mathematics, and the Ph.D. degree in fuzzy logic applications in transportation engineering from the Democritus University of Thrace, Xanthi, Greece, in 2014, 2017, and 2021, respectively. Since 2017, he has been working with the European Commission–Joint Research Centre, Ispra, Italy. His research interests include automated driving systems, intelligent transportation systems, simulation of vehicle dynamics and driver behavior, microscopic simulation of traffic networks and network control, optimization, and traffic safety.

MARIA CRISTINA GALASSI received the degree in aerospace engineering and the Ph.D. degree in nuclear and industrial safety from the University of Pisa. She is currently a Scientific Project Officer with the European Commission Joint Research Centre (JRC). At present, she is leading JRC research activities on the safety assessment of connected and automated vehicles, supporting the development of the new EU and global regulatory framework for the approval of automated driving systems. She is also responsible for the JRC RICAM project, covering the broader scope of requirements and implications of connected and automated mobility.

BIAGIO CIUFFO (Member, IEEE) received the Ph.D. degree in transportation engineering from the Department of Transportation Engineering, University of Napoli Federico II, in 2008. He held a three-year postdoctoral position at the European Commission Joint Research Centre (JRC), Ispra, Italy, working on the sustainability assessment of traffic and transport-related measures and policies. He is currently an Official of the European Commission, working for the Directorate for Energy, Transport, and Climate of the JRC. In the past, he has led different projects concerning the analysis of the environmental and economic impacts of different transport policies. He is currently leading the JRC project, focusing on the wide implications of a connected and automated mobility. He has published more than 100 scientific papers in peer-reviewed journals and conference proceedings in transportation and traffic engineering. He is also one of the main authors of the *The Future of Road Transport* (JRC Report), which analyzes the wide implications of a connected, automated, low-carbon, and shared mobility. He has been awarded the 2012 Greenshields Prize from the Traffic Flow Theory and Characteristics Committee and the 2013 and 2020 Prizes of the SimSub Committee of the Transportation Research Board of the U.S. National Academy of Science, for his research activities on traffic simulation. He is an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and serves as a reviewer for the most important journals in the transportation field.