SHIFT-TRUCK: A HIGH-FIDELITY AERODYNAMICS DATASET AND BENCHMARK FOR PICKUP TRUCKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Pickup trucks account for 14% of new light-duty vehicles produced in the United States and yet, are among the least aerodynamic. Their open cargo bed adds a flow that existing automotive aerodynamics datasets such as DrivAerML and SHIFT-SUV do not contain, in which the shear layer leaving the cab roof passes over a recirculating bed flow before separating again at the tailgate. The resulting high aerodynamic drag on pickup trucks decreases fuel efficiency while increasing emissions in internal combustion engines, while also limiting highway range on electric counterparts. Drag reduction through design optimization is therefore crucial, however, the time and compute required to run scale-resolved Computational Fluid Dynamics (CFD) simulations serve as a bottleneck to expansive design exploration. Neural surrogates offer a way to predict flow features at a fraction of that cost, provided they are trained on large-scale, high-fidelity and domain specific data. To bridge this gap, we introduce SHIFT-Truck, the first such dataset for pickup trucks. It comprises 800 Spalart–Allmaras delayed detached-eddy simulations (SA-DDES) of a reference pickup geometry morphed across 17 shape parameters, such as cab geometry, bed dimensions and windshield rake. Every case is computed on a mesh of about 100 million cells at a Reynolds number of 1.4×10^7 and released with time-averaged surface pressure, wall shear stress, volumetric pressure and velocity. The numerical setup is verified by grid refinement and repeated simulations, as well as checked against wind-tunnel measurements. We define geometry-grouped training, validation and test splits and benchmark four recent neural surrogate architectures, DoMINO, GeoTransolver, AB-UPT and SMART, on surface and volume tracks. Beyond in-distribution evaluation, SHIFT-Truck introduces controlled distribution shifts in the operating point, the discretization of the input surface and the vehicle archetype. We find that models with strong in-distribution performance can exhibit substantial degradation under distribution shifts: operating-condition changes expose failures to infer speed dependence, while tessellation and cross-vehicle shifts reveal markedly different robustness across architectures. These controlled evaluations make SHIFT-Truck a benchmark not only for aerodynamic surrogate accuracy, but also for studying generalization of neural surrogates across physical and numerical distributions.

1 INTRODUCTION

Light-duty trucks make up 66% of new light-duty vehicles produced in the United States, and pickup trucks alone account for 14% (U.S. Environmental Protection Agency, 2026). Despite their prevalence, pickup trucks are aerodynamically poor, owing to a bluff front with a large cross-sectional area, a high ground clearance that admits turbulent flow underneath, and an open cargo bed that causes flow separation and a large turbulent wake. At highway speeds aerodynamic drag is the dominant resistive force. Modern sedans have drag coefficients of 0.2 to 0.25 (Hucho, 2013; Heft et al., 2012), while pickup trucks range from 0.45 to 0.5 (Yang & Khalighi, 2005), representing an increase of roughly 60% to 100% in aerodynamic resistance. For Internal Combustion Engine (ICE) fleets, this means higher fuel consumption and emissions, and a reduction of 10 to 15% in drag at highway speed is estimated to save 5 to 7% of fuel (Bellman et al., 2010). For electric vehicles (EVs), the same penalty dictates heavier batteries and limits highway range. For ICE and EV pickups alike, aerodynamic optimization of their design is a critical engineering need.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

To optimize vehicle aerodynamics, engineers need to explore a large, high-dimensional design space with Computational Fluid Dynamics (CFD) over hundreds of design iterations. However, resolving the turbulent flow with scale-resolving approaches such as Detached Eddy Simulation (DES) on massive meshes takes hours to even days per design on high-performance computing clusters, making design exploration infeasible. Steady Reynolds-Averaged Navier-Stokes (RANS) simulations cost less but are less reliable in such flows dominated by separation, recirculation and streamline curvature (Ashton & Revell, 2015; Ashton et al., 2016). In recent years, neural surrogate models such as DoMINO (Ranade et al., 2025) and GeoTransolver (Adams et al., 2025) have emerged as an alternative to traditional CFD. By learning a mapping from raw geometry and operating conditions to output states, they bypass the iterative solver and predict complex flow fields in minutes on a single GPU. Their accuracy, however, is bounded by the scale, fidelity and domain of their training data, restricting widespread practical usage of these models.

While the Scientific Machine Learning community has benefited from several large-scale automotive CFD datasets such as AhmedML (Ashton et al., 2024b), WindsorML (Ashton et al., 2024a), DrivAerML (Ashton et al., 2024c), DrivAerNet (Elrefaie et al., 2025), DrivAerNet++ (Elrefaie et al., 2024), DrivAerStar (Qiu et al., 2025) and SHIFT-SUV (Luminary Cloud, 2025) (Table 11), these datasets have typically focused on closed passenger vehicles. Pickups introduce a distinct aerodynamic challenge through their stepped cab and bed geometry. With a flush bed cover, flow separates behind the cab above the bed, while another wake forms behind the tailgate. Opening the bed exposes a cavity that further changes the recirculating flow and its interaction with the rear wake. Cab height, bed length, and bed configuration therefore influence the flow around the entire rear of the vehicle, making their effects on drag important targets for aerodynamic optimization. Capturing these relationships requires training and evaluation data that span a wide range of pickup geometries and include both open and covered beds.

To address this need, we introduce SHIFT-Truck, to our knowledge the first large-scale, high-fidelity external aerodynamics dataset for pickup trucks. It comprises 800 Spalart–Allmaras Delayed Detached-Eddy Simulations (SA-DDES) of variants of the open Generic Truck Utility (GTU) geometry (Woodiga et al., 2020), morphed across 15 continous design parameters that capture cab geometry, windshield rake, bed dimensions and ride height. Also, the bed cover and chin spoiler serve as two features that change the wetted surface, and can be turned on and off. Every dataset case is computed on a mesh of about 100 million cells at a Reynolds number of 1.4×10^7 and carries time-averaged surface pressure, wall shear stress, volumetric pressure and velocity. The numerical setup is checked against wind-tunnel measurements on the same truck.

To establish SHIFT-Truck as a benchmark, we define geometry-grouped train, validation and test splits and evaluate four recent neural surrogate architectures, DoMINO, GeoTransolver, AB-UPT and SMART, on separate surface and volume tracks. Strong in-distribution accuracy, however, does not show whether a surrogate has learned the physical mapping from geometry to flow or only features tied to its training distribution. SHIFT-Truck makes this distinction measurable through Out-Of-Distribution sets with controlled shifts in the operating point with the geometry held fixed, in the discretization of the input surface with the geometry and the flow held fixed and in the vehicle archetype, from pickup to SUV. On held-out In-Distribution trucks, three of the four models recover drag with R^2 between 0.95 and 0.97. However, under the distribution shifts, their behavior diverges: none of the models captures a change in freestream speed, their sensitivity to surface decimation differs by more than an order of magnitude and all four models underpredict drag on SUVs. In summary, the primary contributions of this work are as follows:

- **Dataset.** A 17-parameter design space (15 continuous morphs of cab, bed, nose and ride height, and 2 discrete parts, a bed cover and a chin spoiler), sampled by a Sobol sequence and solved with SA-DDES, with time-averaged surface and volume fields for every case (§2).
- **CFD assessment.** Grid-refinement and repeatability studies, together with absolute-drag and part-removal comparisons against GTU wind-tunnel measurements, with calibration choices and comparison limits documented (Appendix C).
- **Benchmark.** Geometry-grouped splits with fixed metrics and reference results for SMART, DoMINO, GeoTransolver and AB-UPT on surface and volume tracks, with drag integrated from the predicted fields (§3).
- **Distribution shifts.** Graded out-of-distribution sets along the three axes, released with the dataset, with the four models evaluated on each (§4).

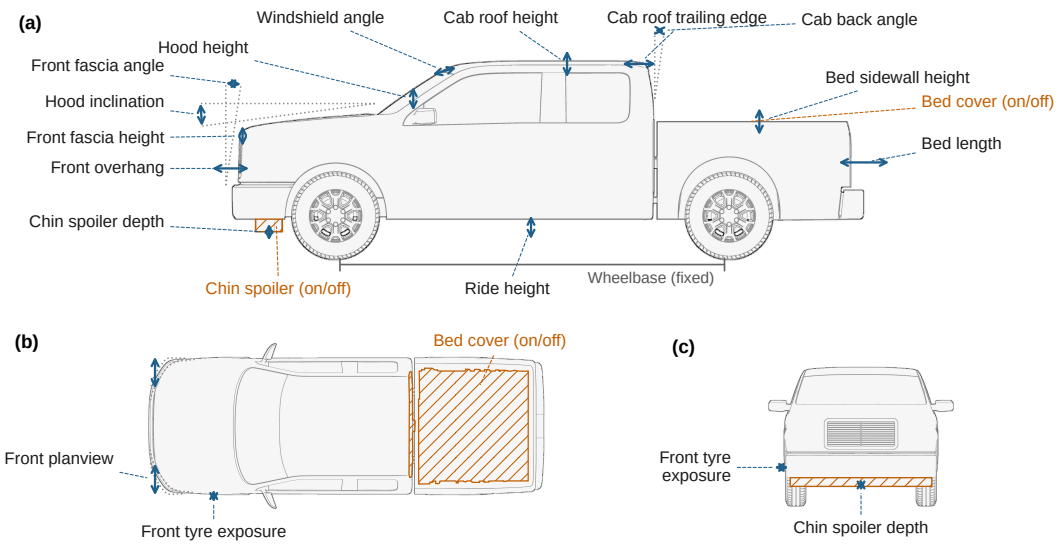

Figure 1: Design parameters of SHIFT-Truck on the GTU baseline. (a) Side, (b) top and (c) front views. Blue arrows mark translational morphs. Dotted line pairs with an arrow across the open end mark angle morphs. Dotted outlines in (b) show the nose at the two extremes of front planview. Hatched regions are the two on/off parts. Ranges are given in Appendix A.

2 The SHIFT-Truck dataset

2.1 Baseline geometry and design space

The baseline geometry for SHIFT-truck is the Generic Truck Utility (GTU), the open reference geometry for pickup and SUV aerodynamics (Woodiga et al., 2020; Howard et al., 2021). We use the 4×2 pickup variant (Fig. 1), with a long cab, a short open bed, exposed mirrors and an airdam under the front fascia. The bed is a cavity between the cab and the tailgate, and the dataset carries both open-bed and closed-bed cases. With the bed open, the flow separates from the roof's trailing edge and the rear of the cab, a low-momentum recirculation fills the bed, and a second separation at the tailgate feeds the base wake (Fig. 2c). A flush cover removes the cavity and leaves a continuous upper surface from the cab to the tailgate. The underbody is open in both states, with exposed wheels and drivetrain. Woodiga et al. (2020) measured wind-tunnel forces for the baseline and for single-part removals, which Appendix C validates against.

Every case is built from this geometry. Fifteen continuous parameters move the front end, the cab, the bed and the ride height (Fig. 1). The body is bound to a deformation cage, so these morphs keep the surface topology and no case needs new CAD. A flush bed cover and a chin spoiler are switched on and off, which adds or removes a part and so changes the topology of the wetted surface. The wheels, underbody and drivetrain are not morphed, the wheelbase is fixed, and ride height moves the body relative to the wheels. Body length varies by about a meter across the dataset and the overall height by 0.4 m. Appendix A gives the cage, the parameter ranges and the dataset composition.

2.2 Simulation

We draw 250 base trucks from a Sobol sequence over the fifteen continuous parameters and simulate each with the four combinations of bed cover and chin spoiler, which gives 1,000 planned cases, of which the release holds 800. Every case uses the same numerical setup, fixed by the study of Appendix C. Turbulence is treated by DDES on the Spalart-Allmaras model (Spalart et al., 2006), with a Vreman subgrid model (Vreman, 2004), so the separations behind the cab and the tailgate are resolved as unsteady turbulence (Fig. 2b) and the attached boundary layers stay modelled. Meshes contain about 100 million cells (Fig. 2a), informed by the grid-refinement study of Appendix C. Each case starts from a converged RANS solution, is time-averaged over the final 15 convective time

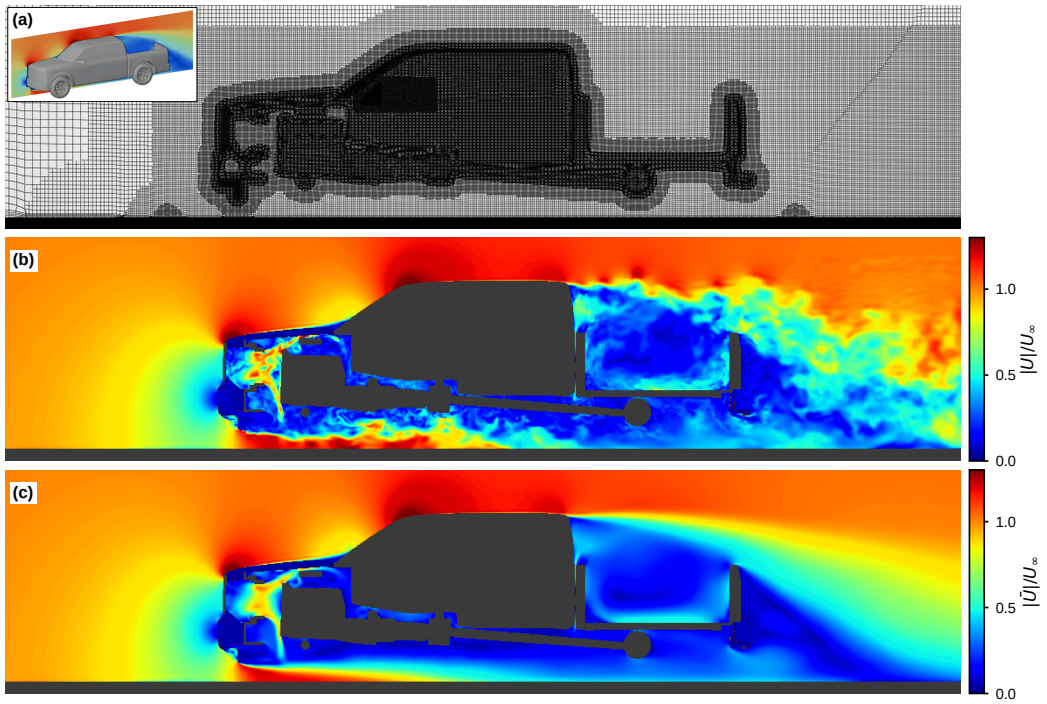

Figure 2: The baseline GTU on the centreplane $y = 0$, the vertical plane through the middle of the truck from nose to tail (inset). (a) Mesh. (b) Instantaneous velocity magnitude. (c) Time-averaged velocity magnitude.

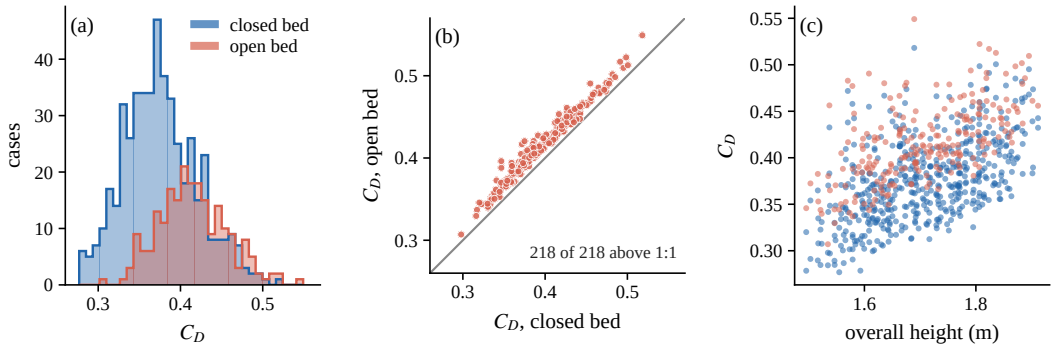

Figure 3: The design space of SHIFT-Truck. (a) Drag distribution by bed state over the 712 cases of the benchmark set. (b) The 218 base trucks carrying both bed states with every other parameter held fixed, each plotted as its closed-bed against its open-bed drag, with the 1:1 line. (c) Drag against overall height, coloured by bed state.

units, and takes 6 h on 4 H100 GPUs. Appendix B gives the solver, the discretization, the mesh construction and the domain. Repeated simulations of the same case agree in drag to within 0.7%. On the reference GTU baseline, simulated C_D exceeds the wind-tunnel measurement of Woodiga et al. (2020) by 0.053. The mirror- and airdam-removal drag changes used during setup calibration differ from measurements by at most 0.007. Appendix C reports the grid study, subsequent changes to the dataset setup, repeatability and these experimental comparisons.

2.3 Dataset Contents

Fig. 3a shows the drag distribution over the dataset. It spans roughly a factor of two, and the closed-bed and open-bed populations occupy different parts of that range with a broad overlap, and overall

Table 1: Benchmark results on the 90 test cases. Error metrics are means over cases and seeds 42, 43 and 44. Drag R^2 entries give the mean $\pm$ the standard deviation across seeds. Params is the trainable parameter count in millions and GPU-h the training cost in H100 hours for 500 epochs in FP32. Inf. is the wall-clock time to predict one case on a surface of 8.8 million faces, on one NVIDIA T4 GPU in float32.

	Surface track							Volume track			
Model	Params	GPU-h	Inf. (s)	C_p MAE	τ_w ε_{L_2}	C_D MAE	C_D R^2	Params	GPU-h	C_p MAE	$\boldsymbol{u}$ ε_{L_2}
GeoTransolver	21.7	91	630	0.021	0.173	0.0069	0.958 ± 0.002	22.2	144	0.029	0.149
SMART	12.6	61	267	0.023	0.175	0.0047	0.971 ± 0.010	12.8	69	0.028	0.114
AB-UPT	16.4	54	883	0.025	0.210	0.0077	0.953 ± 0.010	41.3	89	0.027	0.132
DoMINO	10.1	68	102	0.051	0.605	0.0345	0.034 ± 0.034	18.5	132	0.046	0.163

height is the strongest single driver of the spread (Fig. 3c, Appendix D). Designing the bed is crucial, since bed length, sidewall height, ride height and the cover are all chosen for utility, and each affects the drag. Published pickup aerodynamics studies a handful of configurations at a time (Al-Garni & Bernal, 2010; Yang & Khalighi, 2005; Woodiga et al., 2020), and no open dataset maps these trade-offs across a design space. Because drag force scales with frontal area, each drag count costs more fuel or range on a pickup than on a passenger car.

For machine learning, the dataset poses a harder problem than a single closed body. The target responds to a part appearing and disappearing from the wetted surface, which a continuous morph does not produce. 218 base trucks carry both bed states with every other parameter held fixed, and opening the bed raises drag on every one of them (Fig. 3b), so a surrogate can be scored on a controlled comparison as well as on held-out shapes. The chin spoiler gives a second such comparison on the closed-bed trucks.

Each case is released as a surface in STL and VTK PolyData, carrying time-averaged pressure and wall shear stress, and a volume in VTK UnstructuredGrid, carrying time-averaged pressure and velocity, together with its force coefficients, reference values and a metadata record. The dataset is released under the CC-BY-NC-4.0 license, with its Croissant metadata (Akhtar et al., 2024) and is available for review at `gs://shift-truck-sample/`.

3 BENCHMARK

Models and tracks. We benchmark four neural surrogates that map a raw geometry to field values at arbitrary query points, DoMINO (Ranade et al., 2025), GeoTransolver (Adams et al., 2025), AB-UPT (Alkin et al., 2025) and SMART (Hagnberger & Niepert, 2026). Each is trained for 500 epochs in FP32 on a surface track (pressure and wall shear stress on the body) and a volume track (pressure and velocity in the fluid), using its reference architecture and optimizer. GeoTransolver and DoMINO minimize mean squared error, AB-UPT uses an L_1 loss, and SMART uses relative L_2 loss. Appendix E gives the architectures and training configurations.

Split. The benchmark uses the 712 cases that had completed when the split was frozen. We split by base truck, so that no geometry appears in two sets under any bed or spoiler setting. Validation and test trucks are drawn from the 219 trucks that carry both bed states, each contributing its two closed-bed cases and its open-bed case. They are selected deterministically on a quantile grid over overall height and body length, the two main morph axes, taking in each cell the truck nearest its centre, on a 6×5 grid for test and a 6×4 grid for validation. Test trucks lie as close to the training set in parameter space as training trucks lie to each other (Appendix F.4). The split has 550 training, 72 validation and 90 test cases, and each of the 30 test trucks appears closed with and without the chin spoiler, and open.

Metrics. Fields are scored in non-dimensional form on the solver's own surface polygons and volume cells. All four models use the full volume mesh for evaluation. The pressure coefficient is scored by its mean absolute error, and wall shear stress and velocity by their relative L_2 error. Drag is integrated from the predicted surface fields and no model is trained on forces.

Results. Table 1 gives the results on the test set together with the variation of drag R^2 across training seeds. On the surface, GeoTransolver has the lowest error on both fields, with a pressure error of

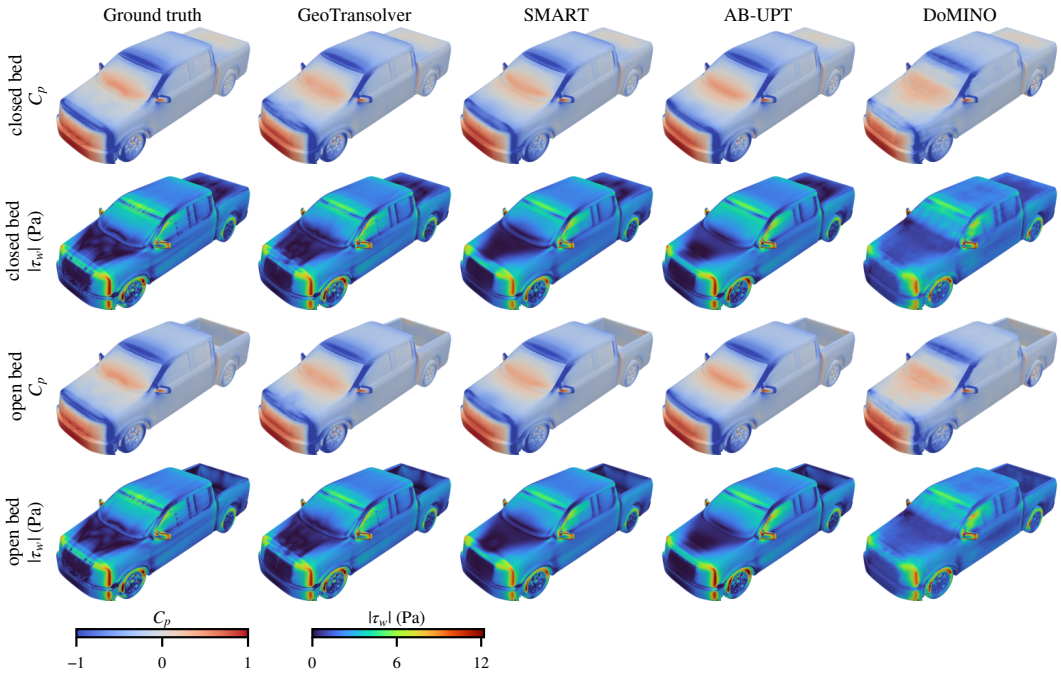

Figure 4: Surface pressure coefficient and wall-shear-stress magnitude for a matched pair of test trucks, the same base geometry with the bed closed (top two rows) and open (bottom two rows). Columns are ground truth and the four surface-track models.

about 2% of the dynamic pressure, and SMART is within 2% of it on shear stress. In the volume, SMART has the lowest velocity error at about half the training cost of GeoTransolver and DoMINO, and the three leading models are within 10% of each other on pressure. A ranking on the body therefore does not carry over to the wake. SMART, GeoTransolver and AB-UPT recover drag with R^2 between 0.95 and 0.97. Across three training seeds, the standard deviation of drag R^2 reaches 0.010, which is comparable to the differences between the three. The lowest drag error, from SMART, is 0.005 in C_D, about 11.4 N at 140 km/h. DoMINO trails on every channel, and its drag R^2 falls to 0.03.

Surface. Figure 4 shows the surface fields of a matched pair (errors in Appendix E.3). The flow stagnates on the grille and fascia, decelerates into the corner at the base of the windshield, and accelerates around the A-pillar and the mirrors. This pressure field is set by the outer flow, which the body shape largely determines, and all four models reproduce it, together with the high wall shear stress along the hood and roof leading edges, the A-pillar and the wheel-arch lips. The models differ where the shear stress depends on the state of the boundary layer. Between the hood leading edge and the windshield the near-wall flow slows under an adverse pressure gradient, and the true field shows a region of low shear on the hood, broken into streamwise streaks. GeoTransolver reproduces the region and its streaks most closely, and SMART and AB-UPT smooth out the streaks. DoMINO fills in every low-shear region: its prediction shows no deceleration ahead of the windshield, no thickening of the boundary layer toward the roof trailing edge, and no slow recirculating flow on the bed floor. Low shear marks where the near-wall flow decelerates, separates or recirculates, and on a bluff body these regions set the pressure drag. DoMINO's pressure also carries a regional offset over the hood and grille, which the drag, a small difference between large forces on the front and rear of the body, inherits.

Volume. Figure 5 shows the centreplane velocity for the same pair (errors and pressure in Appendix E.4). Every model reproduces the acceleration over the hood and roof, which the body shape sets through the outer flow. Every model also separates the flow at the roof trailing edge and, on the open truck, fills the bed with a recirculation. The geometry fixes both, since the flow leaves a sharp edge at a fixed point and a cavity of this depth traps a recirculation whatever the state of the incoming boundary layer. The largest errors sit in the shear layer leaving the cab roof and in the near wake

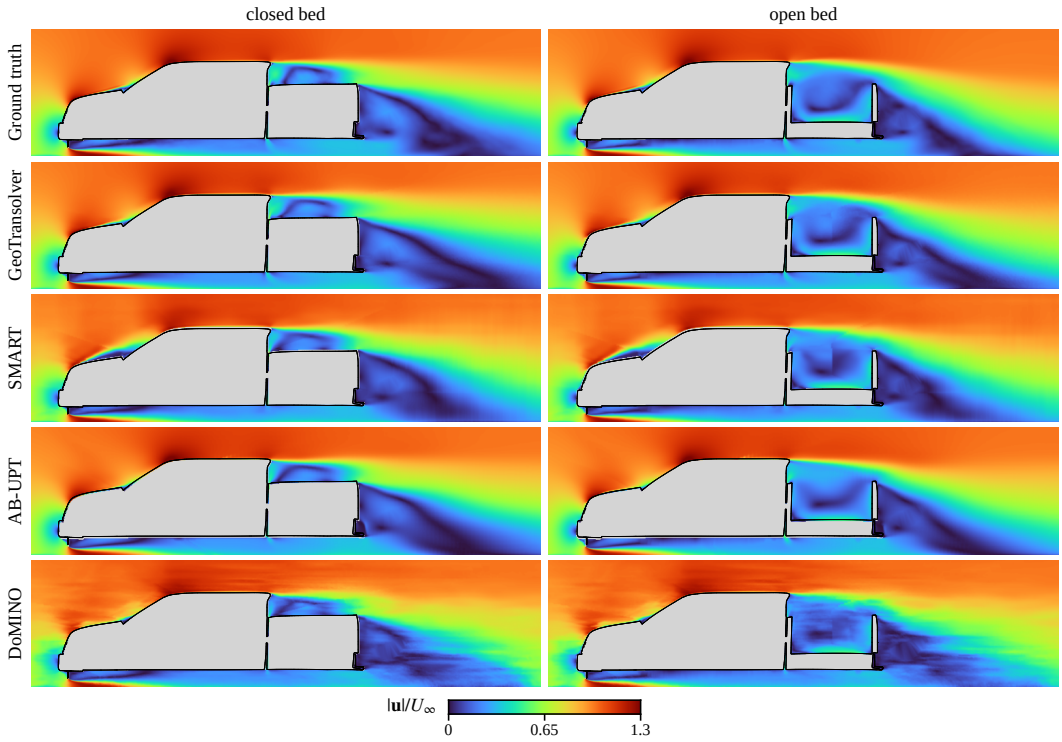

Figure 5: Time-averaged velocity magnitude on the centerplane $y = 0$ for the pair of Fig. 4, with the bed closed (left) and open (right). Rows are ground truth and the four volume-track models.

behind the tailgate, the regions of turbulent mixing that the geometry does not fix. The spreading of the shear layer sets how much momentum it exchanges with the bed and the wake, and the length of the near wake is tied to the base pressure and so to the pressure drag. Every model predicts the velocity less accurately with the bed open, and the added error sits in the bed recirculation and the shear layer above it, which exchange momentum. SMART, GeoTransolver and DoMINO also show a step in velocity halfway along the bed with no counterpart in the true flow, possibly due to a signed distance function as model input. DoMINO recovers the gross wake, but its field is noisy in the outer flow, where the true field is smooth, and its bed recirculation is blurred, matching its filled-in shear stress on the bed floor.

4 Out-of-distribution evaluation

Distribution shifts in Scientific Machine Learning arise not only from changes in the physical problem, such as the geometry or the operating conditions, but also from changes in its numerical representation, as the same field may be represented on meshes of different resolution or connectivity. This discretization shift is particularly important for neural operators, whose intended advantage is to approximate mappings between function spaces rather than mappings tied to a particular discretization.

We evaluate the four trained models of §3 along each axis. For the operating point, a training-set truck is re-simulated at 100, 110, 120 and 130 km/h on the same mesh and with the same solver settings, while all training data are at 140 km/h. For the discretization, the input surface of the same truck at 140 km/h is decimated by quadric-error-metric edge collapse (Garland & Heckbert, 1997) to 50, 30, 20 and 10% of its 8.8 million faces, and the prediction is compared with the solver drag of the full-resolution case. At 10% of the faces the decimated surface lies within 7.6 mm of the original, and integrating the solver fields over it changes C_D by 1 count. The shift therefore changes how the geometry is represented but leaves the geometry and the flow essentially unchanged. For the vehicle archetype, the truck-trained models predict ten SHIFT-SUV cases (Luminary Cloud, 2025), and a model trained on SHIFT-SUV gives the in-distribution reference. Every error is the absolute

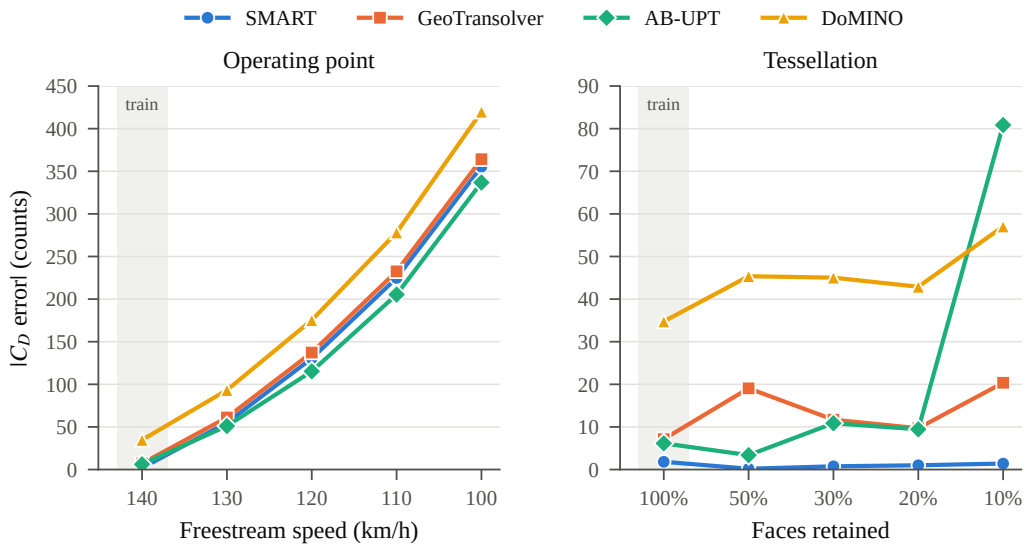

Figure 6: Absolute drag error of the four truck-trained models against freestream speed (left) and the fraction of input surface faces retained (right). The shaded column is the training condition, a training-set truck at 140 km/h on its full surface. One count is 10^{-3} in C_D.

drag-coefficient error in counts, where one count is 10^{-3} in C_D. These OOD sets are released with the dataset.

Operating point. All four models receive freestream speed as an input, but it is constant during training. Without physical rescaling, drag error grows with the speed offset and reaches 337 to 420 counts at 100 km/h (Fig. 6, left), while the solver C_D changes by less than 7 counts. Predicted drag force is unchanged for SMART, GeoTransolver and DoMINO, and varies by less than 5% for AB-UPT. At fixed reference density and area, constant predicted force gives $\widehat{C}_D(U) = \widehat{C}_D(U_{\text{train}})(U_{\text{train}}/U)^2$, which accounts for nearly all of the error growth. Multiplying a speed-invariant force prediction by $(U/U_{\text{train}})^2$ would instead keep its predicted C_D at the training-speed value; the remaining variation in its error would follow the solver's C_D variation. This sweep therefore diagnoses missing dynamic-pressure scaling in the uncorrected predictions of these single-speed-trained models and motivates the speed correction below.

Discretization. The error of SMART stays below 2 counts at every decimation level (Fig. 6, right). GeoTransolver lies between 9.7 and 20.4 counts, above its full-surface error of 7 counts at every level. AB-UPT stays within 11 counts down to 20% of the faces and reaches 81 counts at 10%. DoMINO rises from 35 counts on the full surface to 57 counts at 10%.

Vehicle archetype. The SUV cases are at 31.3 m/s (112.7 km/h), so the truck-trained predictions are rescaled by $(U/U_{\text{train}})^2$ before scoring. After rescaling, every truck-trained model underpredicts C_D on all ten SUVs (Table 2). Mean errors range from 67 counts for GeoTransolver to 251 counts for SMART. The same models are within 2 to 35 counts on the training-set truck. The model trained on SHIFT-SUV predicts the ten cases with a mean error of 1.5 counts. Without rescaling, the speed offset adds an overprediction that partly cancels the geometry error, and the raw column of Table 2 carries both effects.

5 Conclusion

Pickup trucks are among the highest-selling and yet, least aerodynamic vehicles in the United States and the move to electric powertrains turns their drag into a question of battery size and range. We introduced SHIFT-Truck, to our knowledge the first open, high-fidelity aerodynamics dataset for this vehicle class, with 800 SA-DDES simulations of a parametric pickup built on the GTU, each on a mesh of about 100 million cells. The setup is verified on a 5-level grid and reproduces measured

Table 2: Absolute drag error on ten SHIFT-SUV cases in counts (10^{-3} in C_D), averaged over the cases. Truck-trained predictions are rescaled by $(U/U_{\mathrm{train}})^2$ to the SUV speed in the rescaled column and left as predicted in the raw column. The truck column is each model's error on the training-set truck at 140 km/h. The reference model is trained on SHIFT-SUV.

Model	Truck	SUV, rescaled	SUV, raw
SMART	1.8	251	214
GeoTransolver	7.1	67	81
AB-UPT	6.1	241	199
DoMINO	34.8	147	54
SHIFT-SUV reference	–	1.5	1.5

drag changes on the same truck to within 0.007. Generating the dataset took about 19,200 H100 GPU-hours.

In our benchmark of four recent neural surrogates, every model reproduces the stagnation on the front, the separation at the roof trailing edge and the recirculation in the open bed, all of which are dictated by the body shape. The errors concentrate in the shear layer leaving the cab roof and in the near wake, the regions of turbulent mixing that set the base pressure and therefore the pressure drag. Every model is less accurate with the bed open and a ranking on the surface does not carry over to the volume. Under the controlled distribution shifts, no model responds to a change in freestream speed, the sensitivity to surface decimation differs by more than an order of magnitude between models and every model underpredicts drag on SUVs. The open-bed recirculation and the parts that change the wetted surface make SHIFT-Truck a demanding test for surrogate architectures and for distribution shift in physics-based machine learning, as well as a basis for data-driven design of this vehicle class.

6 Limitations and future work

SHIFT-Truck describes the time-averaged flow, as a result of which the data cannot train models of unsteady loads or of the vortex shedding behind the tailgate. Every geometry is morphed from one GTU configuration, so the dataset spans variation within a single vehicle archetype and not the range of production pickup trucks. The benchmark scores surrogates against the solver, and its targets carry the modelling assumptions of DDES with wall functions. Training uses one freestream speed at zero yaw. The speed sweep of §4 checks deployment of these fixed models on one training geometry, where the solver drag coefficient varies only weakly with speed. Crosswind remains outside the dataset. The experimental comparison covers the GTU baseline and two part removals used during setup calibration, with a baseline drag offset of 0.053 against measurement (Appendix C). The benchmark evaluates point predictions, and seed variation is reported for drag only.

We continue to add steady-state cases to the current design space. Future work will extend SHIFT-Truck with transient simulations, so that unsteady loads and wake dynamics can be learned alongside the mean flow. The errors of current surrogates in the shear layer and the near wake call for architectures that represent turbulent mixing more faithfully. Uncertainty quantification for field surrogates and their use in data-driven design optimization of pickup trucks are further natural directions.

AI use statement

We used large language model (LLM) tools to obtain feedback on the manuscript, primarily to identify opportunities for improving clarity and presentation. The feedback was used in accordance with the ICLR author guidelines to improve the manuscript. LLMs were not used to develop the scientific methodology, formulate hypotheses, design or conduct experiments, analyze or interpret scientific results, generate scientific claims or produce the technical contributions presented in this work. The authors independently conducted and verified all scientific work and are fully responsible for the content, claims and conclusions of this paper.

Ethics statement

This work releases simulation data of a publicly available reference vehicle geometry and involves no human subjects, personal data or proprietary vehicle designs. The dataset is released under a non-commercial license (CC-BY-NC-4.0). Surrogate models trained on it are intended for early-stage design exploration and are not a substitute for simulation or testing in safety-relevant decisions. Section 6 states the validation scope within which the data should be used.

Reproducibility statement

The dataset is released under the CC-BY-NC-4.0 license, and is available for review at `gs://shift-truck-sample/`. Appendix A gives the parameterization and the sampling, Appendix B the numerical setup, Appendix C the verification and validation, Appendix E the benchmark models and Appendix F the files, the fields, the split and a datasheet.

References

Corey Adams, Rishikesh Ranade, Ram Cherukuri, and Sanjay Choudhry. GeoTransolver: Learning physics on irregular domains using multi-scale geometry aware physics attention transformer, 2025. URL https://arxiv.org/abs/2512.20399.

Mubashara Akhtar, Omar Benjelloun, Costanza Conforti, Luca Foschini, Pieter Gijsbers, Joan Giner-Miguelez, Sujata Goswami, Nitisha Jain, Michalis Karamousadakis, Satyapriya Krishna, Michael Kuchnik, et al. Croissant: A metadata format for ML-ready datasets. In *Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, volume 37, pp. 82133–82148, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/9547b09b722f2948ff3ddb5d86002bc0-Abstract-Datasets_and_Benchmarks_Track.html.

Abdullah M. Al-Garni and Luis P. Bernal. Experimental study of a pickup truck near wake. *Journal of Wind Engineering and Industrial Aerodynamics*, 98(2):100–112, 2010. doi: 10.1016/j.jweia.2009.10.001.

Benedikt Alkin, Richard Kurle, Louis Serrano, Dennis Just, and Johannes Brandstetter. AB-UPT for automotive and aerospace applications, 2025. URL https://arxiv.org/abs/2510.15808.

Neil Ashton and Alistair Revell. Key factors in the use of DDES for the flow around a simplified car. *International Journal of Heat and Fluid Flow*, 54:236–249, 2015. doi: 10.1016/j.ijheatfluidflow.2015.06.002.

Neil Ashton, Alastair West, Sylvain Lardeau, and Alistair Revell. Assessment of RANS and DES methods for realistic automotive models. *Computers & Fluids*, 128:1–15, 2016. doi: 10.1016/j.compfluid.2016.01.008.

Neil Ashton, Jordan B. Angel, Aditya S. Ghate, Gaetan K. W. Kenway, Man Long Wong, Cetin Kiris, Astrid Walle, Danielle C. Maddix, and Gary Page. WindsorML: High-fidelity computational fluid dynamics dataset for automotive aerodynamics. In *Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, volume 37, pp. 37823–37835, 2024a. doi: 10.52202/079017-1194.

Neil Ashton, Danielle C. Maddix, Samuel Gundry, and Parisa M. Shabestari. AhmedML: High-fidelity computational fluid dynamics dataset for incompressible, low-speed bluff body aerodynamics, 2024b. URL https://arxiv.org/abs/2407.20801.

Neil Ashton, Charles Mockett, Marian Fuchs, Louis Fliessbach, Hendrik Hetmann, Thilo Knacke, Norbert Schönwald, Vangelis Skaperdas, Grigoris Fotiadis, Astrid Walle, Burkhard Hupertz, and Danielle C. Maddix. DrivAerML: High-fidelity computational fluid dynamics dataset for road-car external aerodynamics, 2024c. URL https://arxiv.org/abs/2408.11969.

Miles Bellman, Ramesh Agarwal, Jonathan Naber, and Lee Chusak. Reducing energy consumption of ground vehicles by active flow control. In *Proceedings of the ASME 2010 4th International Conference on Energy Sustainability*, volume 1, pp. 785–793. ASME, 2010. doi: 10.1115/ES2010-90363.

Thomas D. Economon, Darren Engwirda, Jonathan Ho, Michael Mara, and Raghav Mansharamani. Luminary contribution to the AutoCFD4 workshop. 4th Automotive CFD Prediction Workshop, 2024. URL https://autocfd.org/autocfd4/.

Mohamed Elrefaie, Florin Morar, Angela Dai, and Faez Ahmed. DrivAerNet++: A large-scale multimodal car dataset with computational fluid dynamics simulations and deep learning benchmarks. In *Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, volume 37, pp. 499–536, 2024. doi: 10.52202/079017-0016.

Mohamed Elrefaie, Angela Dai, and Faez Ahmed. DrivAerNet: A parametric car dataset for data-driven aerodynamic design and prediction. *Journal of Mechanical Design*, 147(4):041712, 2025. doi: 10.1115/1.4068104.

Michael Garland and Paul S. Heckbert. Surface simplification using quadric error metrics. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*, pp. 209–216. ACM, 1997. doi: 10.1145/258734.258849.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92, 2021. doi: 10.1145/3458723.

Jan Hagnberger and Mathias Niepert. SMART: Scalable mesh-free aerodynamic simulations from raw geometries using a transformer-based surrogate model. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, 2026. URL https://arxiv.org/abs/2601.18707.

Angelina I. Heft, Thomas Indinger, and Nikolaus A. Adams. Introduction of a new realistic generic car model for aerodynamic investigations. SAE Technical Paper 2012-01-0168, SAE International, 2012. URL https://doi.org/10.4271/2012-01-0168.

Kevin Howard, Paul Norman, Sudesh Woodiga, Neil Lewington, and Burkhard Hupertz. A detailed aerodynamics investigation of three variants of the Generic Truck Utility. SAE Technical Paper 2021-01-0950, SAE International, 2021. URL https://doi.org/10.4271/2021-01-0950.

Kevin Howard, Sheng Chen, Sayan Dobronsky, Zachary Kochanek, and Shaun Skinner. Introduction of the eGTU – an electric version of the Generic Truck Utility aerodynamic research model. SAE Technical Paper 2024-01-2273, SAE International, 2024. URL https://doi.org/10.4271/2024-01-2273.

Wolf-Heinrich Hucho (ed.). *Aerodynamics of Road Vehicles: From Fluid Mechanics to Vehicle Engineering*. Butterworth-Heinemann, 2013. ISBN 9781483108414. Reprint of the 1987 edition.

Burkhard Hupertz, Neil Lewington, Charles Mockett, Neil Ashton, and Lian Duan. Towards a standardized assessment of automotive aerodynamic CFD prediction capability – AutoCFD 2: Ford DrivAer test case summary. SAE Technical Paper 2022-01-0886, SAE International, 2022. URL https://doi.org/10.4271/2022-01-0886.

Joshua Krakos, Pedro Gomes, Jonathan Ho, Michael Mara, Gonzalo Sáez, Adrien Loseille, Thomas Economon, and Juan J. Alonso. GPU-based and adaptive solution technology for the 5th AIAA High Lift Prediction Workshop. In *AIAA SciTech 2025 Forum*, pp. 0496. AIAA, 2025. doi: 10.2514/6.2025-0496.

Luminary Cloud. SHIFT-SUV: High-fidelity computational fluid dynamics dataset for SUV external aerodynamics. Hugging Face dataset, CC-BY-NC-4.0, 2025. URL https://huggingface.co/datasets/luminary-shift/SUV.

Jiyan Qiu, Lyulin Kuang, Guan Wang, Yichen Xu, Leiyao Cui, Shaotong Fu, Yixin Zhu, and Ruihua Zhang. DrivAerStar: An industrial-grade CFD dataset for vehicle aerodynamic optimization. In *Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, 2025. URL https://openreview.net/forum?id=GVB1M5K0ch.

Rishikesh Ranade, Mohammad Amin Nabian, Kaustubh Tangsali, Alexey Kamenev, Oliver Hennigh, Ram Cherukuri, and Sanjay Choudhry. DoMINO: A decomposable multi-scale iterative neural operator for modeling large scale engineering simulations, 2025. URL https://arxiv.org/abs/2501.13350.

I. M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967. doi: 10.1016/0041-5553(67)90144-9.

P. R. Spalart, S. Deck, M. L. Shur, K. D. Squires, M. Kh. Strelets, and A. Travin. A new version of detached-eddy simulation, resistant to ambiguous grid densities. *Theoretical and Computational Fluid Dynamics*, 20(3):181–195, 2006. doi: 10.1007/s00162-006-0015-0.

U.S. Environmental Protection Agency. The 2025 EPA Automotive Trends Report: Fuel economy and technology since 1975. Report EPA-420-R-26-001, Office of Transportation and Air Quality, U.S. Environmental Protection Agency, February 2026. URL https://www.epa.gov/automotive-trends/download-automotive-trends-report.

A. W. Vreman. An eddy-viscosity subgrid-scale model for turbulent shear flow: Algebraic theory and applications. *Physics of Fluids*, 16(10):3670–3681, 2004. doi: 10.1063/1.1785131.

H. G. Weller, G. Tabor, H. Jasak, and C. Fureby. A tensorial approach to computational continuum mechanics using object-oriented techniques. *Computers in Physics*, 12(6):620–631, 1998. doi: 10.1063/1.168744.

Sudesh Woodiga, Kevin Howard, Paul Norman, Neil Lewington, Robert Carstairs, Burkhard Hupertz, and Karel Chalupa. The GTU: A new realistic generic pickup truck and SUV model. *SAE International Journal of Advances and Current Practices in Mobility*, 2(4):1986–1998, 2020. doi: 10.4271/2020-01-0664.

Zhigang Yang and Bahram Khalighi. CFD simulations for flow over pickup trucks. SAE Technical Paper 2005-01-0547, SAE International, 2005. URL https://doi.org/10.4271/2005-01-0547.

APPENDIX CONTENTS

A Geometry and parameterization **14**
A.1 Baseline . . . 14
A.2 Deformation cage . . . 14
A.3 Parameters . . . 14
A.4 Discrete parts . . . 14
A.5 Sampling . . . 14

B Numerical setup **15**
B.1 Solver . . . 15
B.2 Turbulence . . . 15
B.3 Time integration and averaging . . . 16
B.4 Mesh . . . 16
B.5 Domain and boundary conditions . . . 16
B.6 Operating point and reference quantities . . . 17

C Verification and validation **17**
C.1 Grid refinement . . . 17
C.2 From the grid family to the dataset setup . . . 18
C.3 Repeatability . . . 19
C.4 Comparison with experiment and reference CFD . . . 19

D Aerodynamics of the design space **20**
D.1 Baseline flow . . . 20
D.2 Drag distribution . . . 20
D.3 Bed state . . . 20
D.4 Chin spoiler . . . 20

E Benchmark models and predictions **21**
E.1 Architectures . . . 21
E.2 Training configuration . . . 21
E.3 Surface predictions . . . 21
E.4 Volume predictions . . . 22

F Release, files and datasheet **25**
F.1 Access and license . . . 25
F.2 Files . . . 25
F.3 Fields . . . 25
F.4 Benchmark split . . . 25
F.5 Maintenance . . . 25
F.6 Datasheet . . . 28

G Existing datasets **28**

A GEOMETRY AND PARAMETERIZATION

A.1 BASELINE

The Generic Truck Utility (GTU) is a family of open reference geometries for pickups and SUVs (Woodiga et al., 2020; Howard et al., 2021), distributed by the European Car Aerodynamic Research Association (ECARA). Howard et al. (2024) describe an electrified variant. We morph configuration 7 of the first release, a 4×2 pickup with a long cab, a short bed and a wheelbase of 3.27 m. It has an open cargo bed closed by a tailgate, exposed mirrors, an airdam below the front fascia, exposed wheels and a detailed underbody with drivetrain. Woodiga et al. (2020) measured the forces on this configuration in a wind tunnel, with and without single parts, and simulated the same configurations. Appendix C validates our setup against both.

A.2 DEFORMATION CAGE

The body is bound to a deformation cage of 126 vertices with the harmonic-coordinate mesh-deform binding of Blender. The cage carries the cab, the bed, the hood, the underhood, the bed cover and the mirrors. The wheels, the underbody and the drivetrain sit outside the cage and keep their baseline shape. The wheelbase is fixed as a result, and the ride-height parameter moves the body relative to the wheels. A second cage of eight vertices carries the chin spoiler. Its depth is an independent parameter. Three further shape keys on this cage follow front overhang, front tyre exposure and ride height, and they keep the spoiler attached to the fascia as the body morphs.

A.3 PARAMETERS

Table 3 lists the 15 continuous parameters. Each is a shape key with its own slider range, and a slider value of zero gives the baseline geometry. The ranges are asymmetric about the baseline. Nine keys run from -0.5 to $+1.0$ and four from -0.5 to $+0.5$. Front overhang runs from -0.5 to $+0.75$ and bed sidewall height from -1.0 to $+1.0$. The sampler draws a normalized value in $[-1, 1]$ for each parameter and maps it linearly onto the slider range. In normalized coordinates the baseline sits at -0.33 for the nine keys that extend to $+1.0$, at -0.20 for front overhang and at zero for the five symmetric keys. The centre of the design space is a truck with a longer bed, a higher roof and a more swept nose than the baseline. Front planview moves the outboard corners of the nose inboard with the centreline held fixed, which changes the plan-view taper of the front end.

A.4 DISCRETE PARTS

The bed cover is a flush tonneau panel over the cargo bed. With the cover in place the bed is a closed box. Without it the bed is an open cavity between the cab back, the sidewalls and the tailgate. The chin spoiler extends the lower fascia downward and restricts the flow into the underbody. Both parts are added to or removed from the simulated assembly, and each change alters the topology of the wetted surface. The chin spoiler is a separate part from the GTU airdam removed in the validation of Appendix C.

A.5 SAMPLING

The continuous parameters are sampled at 250 points of a scrambled Sobol sequence (Sobol', 1967), generated with the `qmc.Sobol` sampler of SciPy at seed 42. Each point defines a base truck. Every base truck is simulated with all four combinations of bed cover and chin spoiler, for 1,000 planned cases. Case ids 0000–0249 carry both parts, 0250–0499 the cover alone, 0500–0749 the spoiler alone and 0750–0999 neither. The ids i, $i + 250$, $i + 500$ and $i + 750$ belong to the same base truck. Each case is meshed and simulated independently. The benchmark of §3 uses the 712 cases that had completed at the training freeze, and the public release holds 800 cases. Table 4 gives the composition of both sets.

Table 3: The fifteen continuous design parameters. Each is a shape key on the deformation cage. The sampler draws a normalized value in $[-1, 1]$ and maps it linearly onto the slider range. Displacement is the largest cage-vertex displacement at each end of the range, and the last column gives the normalized value of the baseline geometry.

Parameter	Direction	Slider range	Displacement (m)	Baseline
Front overhang	x, forward	-0.50 to $+0.75$	-0.154 to $+0.231$	-0.20
Hood height	z	-0.5 to $+1.0$	-0.046 to $+0.092$	-0.33
Hood inclination	z at hood front (rotation about the cowl)	-0.5 to $+0.5$	-0.192 to $+0.192$	$+0.00$
Front fascia height	z	-0.5 to $+0.5$	-0.125 to $+0.125$	$+0.00$
Front fascia angle	x at fascia top (rotation about its base)	-0.5 to $+1.0$	-0.188 to $+0.375$	-0.33
Front planview	y, corners inboard	-0.5 to $+1.0$	-0.225 to $+0.451$	-0.33
Front tyre exposure	y, outboard	-0.5 to $+0.5$	-0.050 to $+0.050$	$+0.00$
Windshield angle	x at windshield top	-0.5 to $+1.0$	-0.098 to $+0.196$	-0.33
Cab roof height	z	-0.5 to $+1.0$	-0.111 to $+0.222$	-0.33
Cab roof trailing edge	x, rearward	-0.5 to $+1.0$	-0.152 to $+0.305$	-0.33
Cab back angle	x at cab-back top	-0.5 to $+1.0$	-0.114 to $+0.227$	-0.33
Bed length	x, rearward	-0.5 to $+1.0$	-0.242 to $+0.484$	-0.33
Bed sidewall height	z	-1.0 to $+1.0$	-0.139 to $+0.139$	$+0.00$
Ride height	z, body relative to wheels (down)	-0.5 to $+0.5$	-0.061 to $+0.061$	$+0.00$
Chin-spoiler depth	z, downward	-0.5 to $+1.0$	-0.048 to $+0.096$	-0.33
Bed cover	on/off	—	—	—
Chin spoiler	on/off	—	—	—

Table 4: Composition of the benchmark set, the 712 cases complete at the training freeze, and of the release.

Cases at the freeze	712
closed bed / open bed	492 / 220
Distinct base trucks	248
appearing three, two, one times	218 / 28 / 2
carrying both bed states	219
carrying both bed states, chin spoiler held fixed	218
carrying both spoiler states, bed closed	245
Body length (m)	4.90 to 5.87
Overall height (m)	1.49 to 1.90
Wetted area (m^2)	41.5 to 53.7
Cases in the public release	800
Cases planned	1,000

B Numerical setup

B.1 Solver

The flow is solved with a GPU-native, cell-centred finite-volume code on unstructured polyhedral meshes (Krakos et al., 2025; Economon et al., 2024). The gas is ideal, and the viscosity follows Sutherland's law. The convective flux uses a hybrid upwind–centred scheme with low dissipation in the resolved regions of the flow and stable behaviour in the coarse far field. Face states are reconstructed from cell-centred values and gradients to second order in space, without a limiter. Time integration is second-order implicit (BDF2) with five inner iterations per step. Every operation runs on the GPU in double precision.

B.2 Turbulence

Turbulence is treated by delayed detached-eddy simulation on the Spalart–Allmaras model (Spalart et al., 2006), with the Vreman subgrid model (Vreman, 2004). The delayed formulation shields the attached boundary layer, which stays modelled. This prevents modelled-stress depletion and the premature separation it would cause.

B.3 TIME INTEGRATION AND AVERAGING

Each case starts from a converged RANS solution of 4,000 iterations. The time step is ramped down between iterations 1,000 and 1,500 of the transient run and then held at $\Delta t = 2.0 \times 10^{-4}$ s. A case runs 20,000 iterations, about 6.25 s of physical time. Fields and forces are averaged over the final 10,000 steps. This window lasts about 2.0 s, or 15 convective time units of $L/U_\infty = 0.135$ s. The validation runs of Appendix C average over a longer window of 18,000 steps.

B.4 MESH

Meshes are generated with snappyHexMesh (Weller et al., 1998) following the practice of the Automotive CFD Prediction Workshop (Hupertz et al., 2022; Economon et al., 2024). Refinement follows the body through distance-based shells around the surface, with additional refinement at feature edges. The surface refinement level is 8 on the body panels, 9 on the wheels, 10 on the chin spoiler and 11 on the mirrors. Dedicated boxes refine the mirror wakes, which distance-based sizing alone leaves under-resolved. The near-wall mesh has 5 prism layers at an expansion ratio of 1.2, and every vehicle surface uses wall functions. The body panels sit at y^+ of 20 to 35. The mirrors and the chin spoiler carry the finest surface refinement and sit at 2 to 4 (Fig. 7). The baseline GTU mesh has 122 million cells. The morphed geometries have about 100 million, and `variant_0000` has 97.6 million cells and 103.2 million points.

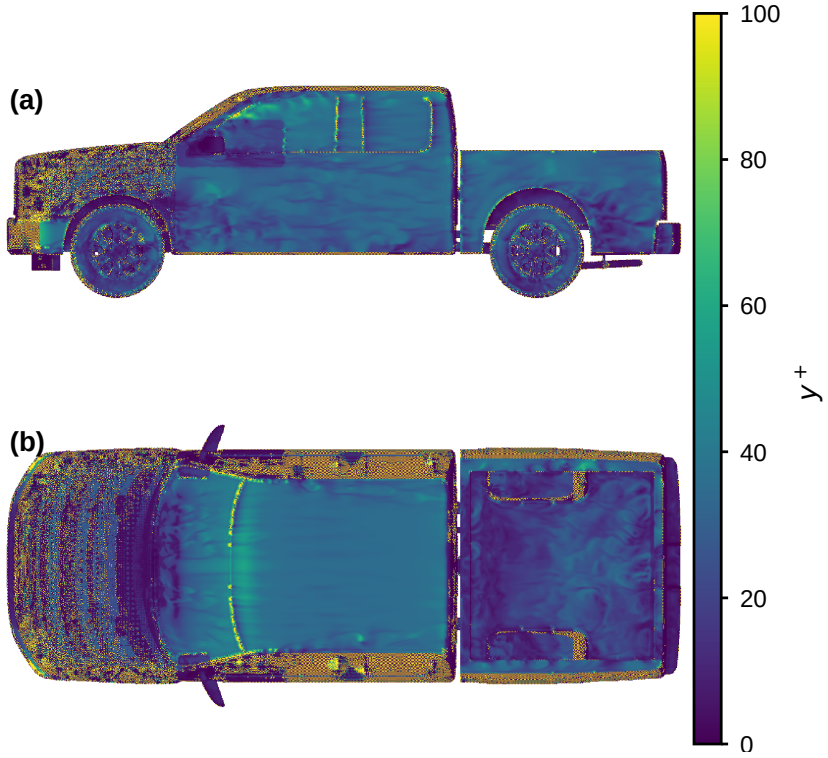

Figure 7: Wall y^+ on the baseline GTU with the dataset setup, (a) side and (b) underside. The body panels are wall-modelled at y^+ of 20 to 35. The mirrors, the chin spoiler and the wheel edges resolve below 5.

B.5 DOMAIN AND BOUNDARY CONDITIONS

The domain is a closed box 59 m long, 50 m wide and 30 m high, the size of the CFD domain of Woodiga et al. (2020). The vehicle sits about 21 m downstream of the inlet and 32 m upstream of the outlet. The inlet is a far-field boundary and the outlet a pressure outlet. The ceiling, the side walls and the floor ahead of the vehicle are slip walls. Under and behind the vehicle the floor is no-slip. It contains 5 belts moving at the freestream speed, one under each wheel and a main belt between

them, which reproduce the rolling road of the experiment. The wheels rotate at 105.76 rad s^{-1}, which matches the belt speed at their radius of 0.3677 m. The released surface excludes the tunnel boundaries. The vehicle surface ends where the tyres meet the belts, and the contact patches are open.

B.6 Operating point and reference quantities

Every case runs at a single operating point of 140 km h^{-1} at zero yaw, with a freestream static temperature of 293.15 K (Table 5). The force coefficients and the pressure coefficient are normalized with constant reference values of the baseline geometry, which the release stores in `params.json`. These include a reference temperature of 288.15 K. The Reynolds number and the blockage ratio are derived from these values. Woodiga et al. (2020) state neither for the experiment. Each case runs on 4 H100 GPUs in about 6 h. At the validation settings the baseline GTU takes 7 h 02 min on 8 GPUs and 1 h 27 min on 56.

Table 5: Operating point, reference quantities and domain, common to every case.

Quantity	Symbol	Value
Freestream speed	U_∞	38.89 m s^{-1}
Freestream static temperature	T_∞	293.15 K
Freestream density (derived)	ρ_∞	1.204 kg m^{-3}
Reynolds number $U_\infty L/\nu$ (derived)	Re	1.4×10^7
Reference length (body length)	L	5.25 m
Reference area (frontal)	A_{ref}	2.65 m^2
Reference pressure	p_{ref}	101,325 Pa
Reference temperature	T_{ref}	288.15 K
Reference density (derived)	ρ_{ref}	1.225 kg m^{-3}
Convective time unit	L/U_∞	0.135 s
Time step	Δt	2.0×10^{-4} s
Domain, streamwise $\times$ spanwise $\times$ vertical	—	$59 \times 50 \times 30$ m
Blockage ratio (derived)	—	0.18%

C Verification and validation

Table 6: Validation against the wind-tunnel experiment and the CFD of Woodiga et al. (2020) on the GTU baseline and two single-part removals. Deltas are referenced to each source's own baseline. The reference CFD is reported as deltas only.

	Absolute C_D		ΔC_D from baseline		
Configuration	Ours	Experiment	Ours	Experiment	Reference CFD
4×2 baseline	0.446	0.393	0	0	0
No mirrors	0.435	0.381	-0.0106	-0.012	-0.010
No airdam	0.481	0.421	$+0.0349$	$+0.028$	$+0.035$

C.1 Grid refinement

The numerical setup was fixed in two steps on the baseline GTU, both run on the three configurations of Table 6. A grid refinement study set the cell budget and the surface resolution of the body panels (Fig. 8, Table 7). The near-field meshing and the numerics were then tuned at that budget, with the two measured part-removal deltas as the acceptance test. The grid family has 5 meshes of 26, 80, 107, 113 and 304 million cells. They are built on a structured background mesh with volume refinement boxes and three prism layers, and they use the default limiter and blending of the solver. Surface refinement rises by one level from the first mesh to the second and stays fixed after that, so the third to fifth meshes differ only in volume refinement. Baseline drag falls from 0.443 on the coarsest mesh to 0.427 on the finest, and it changes by less than 0.003 from the third mesh onward.

The configuration deltas converge on coarser meshes than the absolute level. The airdam delta lies between $+0.0445$ and $+0.0450$ on every mesh from 80 million cells. The mirror delta has the wrong sign on the coarsest mesh and settles between -0.0070 and -0.0082 from 107 million cells. The mean velocity on the centreplane (Fig. 8c–e) is indistinguishable between meshes from 80 million cells upward.

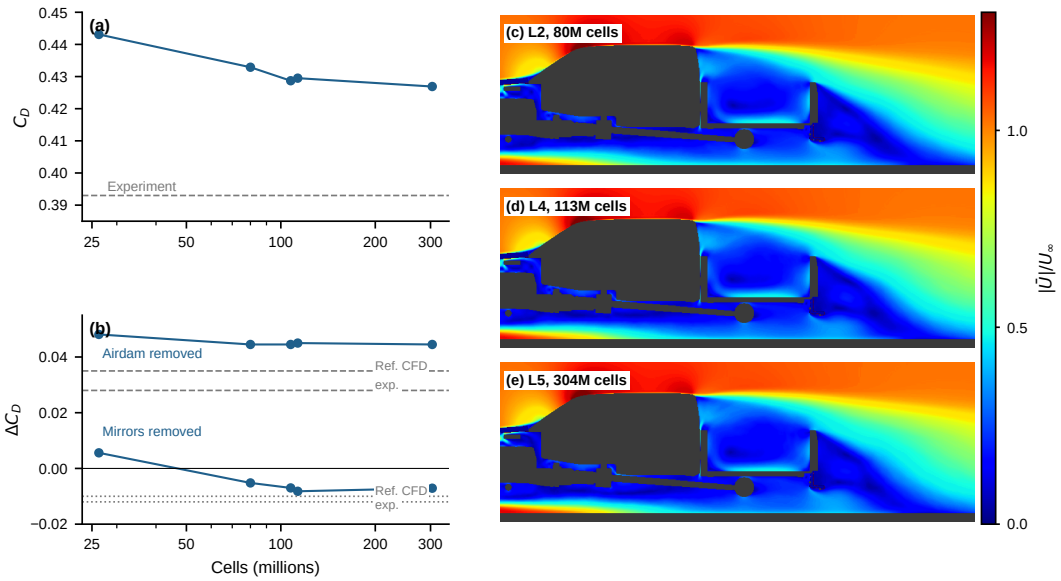

Figure 8: Grid refinement on the baseline GTU. (a) Drag coefficient and (b) configuration deltas against cell count for the five-mesh family. Dashed and dotted lines are the experiment and the CFD of Woodiga et al. (2020). (c)–(e) Time-averaged velocity magnitude on the centreplane for the 80-, 113- and 304-million-cell meshes.

Table 7: Grid refinement study on the baseline GTU. Deltas are referenced to each mesh's own baseline.

Mesh	Cells (M)	C_D	ΔC_D mirrors	ΔC_D airdam
L1	26	0.4431	$+0.0056$	$+0.0481$
L2	80	0.4329	-0.0052	$+0.0445$
L3	107	0.4287	-0.0070	$+0.0445$
L4	113	0.4295	-0.0082	$+0.0450$
L5	304	0.4269	-0.0071	$+0.0445$
Experiment (Woodiga et al., 2020)	—	0.3930	-0.0120	$+0.0280$
CFD (Woodiga et al., 2020)	—	—	-0.0100	$+0.0350$

C.2 From the grid family to the dataset setup

The cell budget was set at about 100 million cells, with the body panels at surface level 8. The near-field meshing follows the practice of the Automotive CFD Prediction Workshop (Hupertz et al., 2022; Economon et al., 2024). Body-conformal refinement shells and feature-edge refinement replace the volume boxes of the grid family. Dedicated boxes refine the mirror wakes, and the mirrors and the chin spoiler receive finer surface refinement (Appendix B). Five prism layers replace three, which removed a premature separation that the thinner stack had produced on the body panels. The slope limiter was removed and the low-dissipation blending of the convective flux was raised to suit the finer near field. A sweep of the blending over its usable range moves the mirror delta by about 0.003 and the airdam delta by less than 0.001. On the baseline GTU the dataset setup gives 122 million cells and the two deltas of Table 6, both within 0.001 of the reference CFD. The absolute drag responds more strongly to these choices than the deltas. It moves by 0.02 between the grid family and the

dataset setup, while the deltas move by 0.003 to 0.010. Neither study removes the offset in absolute drag from the experiment. The limiter and blending were set with the two experimental deltas in view, and the sweep above bounds how far that choice moves each delta.

C.3 Repeatability

Some cases of the dataset were simulated a second time with unchanged shape parameters. Across these repeated pairs, drag agrees to within 0.07 to 0.7%.

C.4 Comparison with experiment and reference CFD

We compare the GTU baseline and two single-part removals, the mirrors and the airdam, with the wind-tunnel measurements and the CFD of Woodiga et al. (2020) for the same configurations (Fig. 9, Table 6). The measurements were taken in a three-quarter open-jet wind tunnel with a five-belt rolling road at 140 km h^{-1}. The coefficients use the same frontal area of 2.65 m^2 as ours. The reference CFD is an SA-DDES on an unstructured mesh of about 100 million cells at y^+ of about 30. It uses a time step of 2×10^{-4} s, averages over the final 5.5 s of a 6 s run and uses a closed domain of the same size as ours. It is a simulation of comparable resolution and modelling with an independent solver. The source reports the reference CFD as deltas only, and we compare the absolute level with the experiment alone.

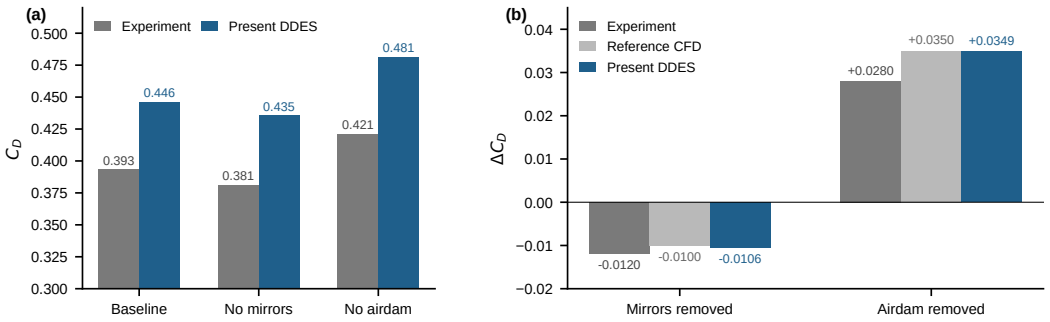

Figure 9: Validation on the baseline GTU. (a) Absolute drag coefficient for the three configurations, from the experiment of Woodiga et al. (2020) and the present DDES. (b) Drag change on removing the mirrors and the airdam, from the experiment and CFD of Woodiga et al. (2020) and the present DDES.

The configuration deltas agree with experiment on both removals. Removing the mirrors reduces drag by 0.0106 in the present simulations, against 0.012 measured and 0.010 in the reference CFD. Removing the airdam raises drag by 0.0349, against 0.028 measured and 0.035 in the reference CFD. Both signs are correct. The mirror delta lies between the two reference values. The airdam delta matches the reference CFD to within 0.0001 and exceeds the measurement by 0.007. The two removals are independent and of opposite sign. The mirrors are small appendages in locally attached flow, and the airdam controls the flow into the underbody. Agreement on two such different effects makes a single cancellation of errors unlikely. We keep the label “no airdam” of Woodiga et al. (2020) for the second removal. The part is the baseline airdam of the GTU and is distinct from the chin spoiler of Appendix A.

The absolute level agrees less well. The present baseline drag coefficient is 0.446 against a measured 0.393, an offset of 0.053. An offset of the same sign appears on every mesh and every setting of the grid study, including the finest mesh, so resolution does not explain it. The experiment used an open-jet tunnel with a nozzle area of 22.45 m^2, a nozzle blockage of 12%. Woodiga et al. (2020) state no blockage correction, Reynolds number or reference length. Both simulations, ours and the reference CFD, use a closed domain with a blockage of 0.18%. The absolute level of the reference CFD is not reported, and we cannot tell whether it carries a similar offset. The open jet against the closed domain, the reference convention and the wall-modelled near-wall treatment are the candidate contributions, and the present data cannot separate them. Woodiga et al. (2020) use a domain of the

same kind, note that it omits the nozzle, supports and collector of the tunnel, and expect differences from the experiment for that reason. Their simulation over-predicts the airdam delta by the same amount as ours. Comparisons of absolute drag from this dataset with tunnel values should carry this offset.

D AERODYNAMICS OF THE DESIGN SPACE



Figure 10: Time-averaged pressure coefficient on the centreplane $y = 0$ of the baseline GTU, the same plane and case as Fig. 2.

D.1 BASELINE FLOW

The flow accelerates over the hood leading edge and again over the roof leading edge. A thin shear layer leaves the A-pillar and the roof and stays attached along the cab. Behind the cab the flow separates from the roof trailing edge and the cab back. A low-momentum recirculation fills the cargo bed between the cab back, the bed floor and the tailgate. The tailgate sheds a second separation into the base wake, and the bed flow feeds this wake from above. Below the vehicle the flow accelerates between the floor and the underbody into a jet that leaves behind the rear axle and bounds the base wake from below. The mean pressure (Fig. 10) shows stagnation on the front face and the airdam, suction over the hood and roof leading edges, and a nearly uniform base pressure over the tailgate and the bed. In the instantaneous field (Fig. 2b) every separated region is resolved as unsteady turbulence, and the modelled boundary layer is confined to the attached surfaces.

D.2 DRAG DISTRIBUTION

Over the 712 cases of the benchmark set, drag spans 0.277 to 0.549 with a standard deviation of 0.047 (Fig. 3a). The median drag is 0.372 for closed-bed cases and 0.412 for open-bed cases, with standard deviations of 0.045 and 0.041. Between the two bed states the median shifts by about one standard deviation, and the spread stays about the same. Drag rank-correlates most strongly with overall height, with a Spearman coefficient of 0.53 for closed beds and 0.60 for open beds. Its rank correlation with body length is weak, at -0.14 and -0.05.

D.3 BED STATE

218 base trucks carry both bed states with the chin spoiler present in both. Opening the bed raises drag on every one of them. The median increase is $+0.018$, and the increases range from $+0.005$ to $+0.050$ (Fig. 3b).

D.4 CHIN SPOILER

245 base trucks carry both spoiler states with the bed closed. Every open-bed case carries the spoiler. Adding the spoiler raises drag on every one of them, with a median increase of 0.036.

E BENCHMARK MODELS AND PREDICTIONS

E.1 ARCHITECTURES

Each model maps a query point on the surface or in the volume, together with a representation of the geometry, to the field values at that point. All four were trained from scratch with the architecture and optimizer of their reference implementations. The training settings and model-specific losses are given below.

DoMINO. DoMINO (Ranade et al., 2025) is a point-cloud model. It predicts the fields at each output point from the geometry in a local region around that point, combined with a coarse encoding of the whole body. Working on local regions lets it run at the full resolution of the mesh.

GeoTransolver. GeoTransolver (Adams et al., 2025) extends the Transolver backbone. Self-attention acts on learned slices of the flow state, and every block cross-attends to a shared context built from multi-scale ball queries on the geometry and from the operating parameters. The geometry enters every layer of the network through this context.

AB-UPT. AB-UPT (Alkin et al., 2025) is an anchored-branched universal physics transformer. A fixed set of anchor tokens on the surface and in the volume carries the attention, and each query point is decoded from the anchors by cross-attention. The cost of attention is independent of the number of query points at inference.

SMART. SMART (Hagnberger & Niepert, 2026) is a transformer that encodes a point cloud of the geometry and the simulation parameters into a shared latent space. A physics decoder maps query points to field values by cross-attending to the intermediate latents of the encoder. The geometric features and the predicted field are updated together through the decoder.

E.2 TRAINING CONFIGURATION

Every model is trained for 500 epochs in FP32 in both tracks. Table 8 gives the optimizer, learning-rate schedule, geometry and query sampling budgets, target normalization and loss for each model. The surface-track runs use eight NVIDIA H100 GPUs each. GeoTransolver uses Muon for two-dimensional weight tensors and AdamW for the remaining parameters. Lion uses $\beta = (0.9, 0.99)$ for AB-UPT and SMART, and Adam uses $\beta = (0.9, 0.999)$ for DoMINO. AB-UPT excludes normalization parameters and biases from weight decay. Its warmup lasts 5% of training, or 25 epochs. DoMINO redraws its surface training queries each epoch. AB-UPT pools its geometry points to 16,384 supernodes.

GeoTransolver and DoMINO use mean squared error (MSE), AB-UPT uses L_1 loss, and SMART uses relative L_2 loss. For volume-track training, GeoTransolver and SMART average the surface and volume loss terms. SMART additionally weights near-wall and wake regions in its volume loss during training. The common evaluation metrics are those of §3, and every model is evaluated on the full volume mesh. The SMART volume model has 12.8 million trainable parameters, as reported in Table 1.

E.3 SURFACE PREDICTIONS

Figures 11 and 12 give the surface pressure coefficient and the wall-shear-stress magnitude for the matched pair of Fig. 4, in the front-high view of the main text and in a side view. Each prediction is shown beside its error, the prediction minus the ground truth. The error colour scale is symmetric at the 99th percentile of the absolute error over the four models. The two bed states give the same value to within 1.5%, and both share one scale.

Pressure. The pressure error of every model sits on the front of the truck. It is largest on the grille and the front fascia, along the hood leading edge, and in a band across the hood where the pressure recovers toward the windshield base. The wheels and wheel arches carry a second concentration of error. The cab sides, the roof and the bed sides carry little error. GeoTransolver has the weakest band

Table 8: Training configurations. All models use 500 epochs in FP32. Query counts are per case during training; volume evaluation uses the full mesh for every model. The volume-track rows give the surface and volume query counts used together in that track.

Setting	GeoTransolver	AB-UPT	DoMINO	SMART
Optimizer	Muon / AdamW	Lion	Adam	Lion
Base learning rate	10^{-3}	5×10^{-5}	10^{-3}	5×10^{-5}
Weight decay	10^{-4}	0.05	0	0
Learning-rate schedule	Halve every 100 epochs; no warmup	5% linear warmup; cosine to 10^{-6}	Halve at epochs 50, 100, 200, 250, 300, 350, 400, 450	Cosine to 0
Gradient clipping	None	Norm 2.0	Norm 10	Norm 2.0
Geometry points	200,000	200,000	300,000	200,000
Target normalization	Mean/std	Mean/std	Min–max	Mean/std
Field loss	MSE	L_1	MSE	Relative L_2
Surface-track queries	60,000	16,384	32,000	16,000
Volume-track surface queries	16,000	16,384	16,000	16,000
Volume-track volume queries	80,000	16,384	32,000	80,000

across the hood. SMART and AB-UPT carry a band of alternating sign across the hood, where the pressure rises toward the windshield. DoMINO over-predicts the pressure over the grille, the fascia and the hood leading edge, and under-predicts it on the rear wheel. Opening the bed leaves the error on the front half of the truck unchanged.

Shear stress. The shear-stress error concentrates on the hood, the grille and the wheels. GeoTransolver, SMART and AB-UPT under-predict the shear over the centre of the hood, where the true field carries streaks of low shear. The under-prediction is weakest for GeoTransolver and strongest for SMART. DoMINO over-predicts the shear where the true shear is low, on the hood, the cab sides and the bed floor. Its side view also shows a band of under-predicted shear along the lower body side.

E.4 VOLUME PREDICTIONS

Figures 13 and 14 give the velocity magnitude and the pressure coefficient on the centreplane $y = 0$ for the same pair. The ground truth is the released field at the native cell centres that lie within a thin slab around the plane. Each model is queried at the same cell centres, mixed with a random background drawn from the rest of the volume, and only the slab points are read out. The points are triangulated in the plane and interpolated to a 5 mm lattice for display. DoMINO returns no value outside its bounding box, and those points are left blank. The error colour scale is symmetric and common to both bed states, and it covers the 99th percentile of the absolute error over the four models.

Velocity. The velocity error sits in the shear layer that leaves the roof trailing edge, in the bed and in the near wake. With the bed closed, GeoTransolver and AB-UPT carry small errors at the roof trailing edge and in the upper wake. SMART under-predicts the velocity in a thin layer over the hood and the windshield. DoMINO over-predicts the velocity in the lower wake, and its error is noisy through the outer flow. With the bed open, every model carries error inside the bed recirculation and along the shear layer above it. GeoTransolver and AB-UPT under-predict the velocity through the near and far wake.

Pressure. The pressure error is largest around the front bumper and at the base of the windshield, where SMART under-predicts the pressure. GeoTransolver, SMART and AB-UPT predict a smooth field with a small negative offset through the wake. DoMINO carries a noisy error through the outer

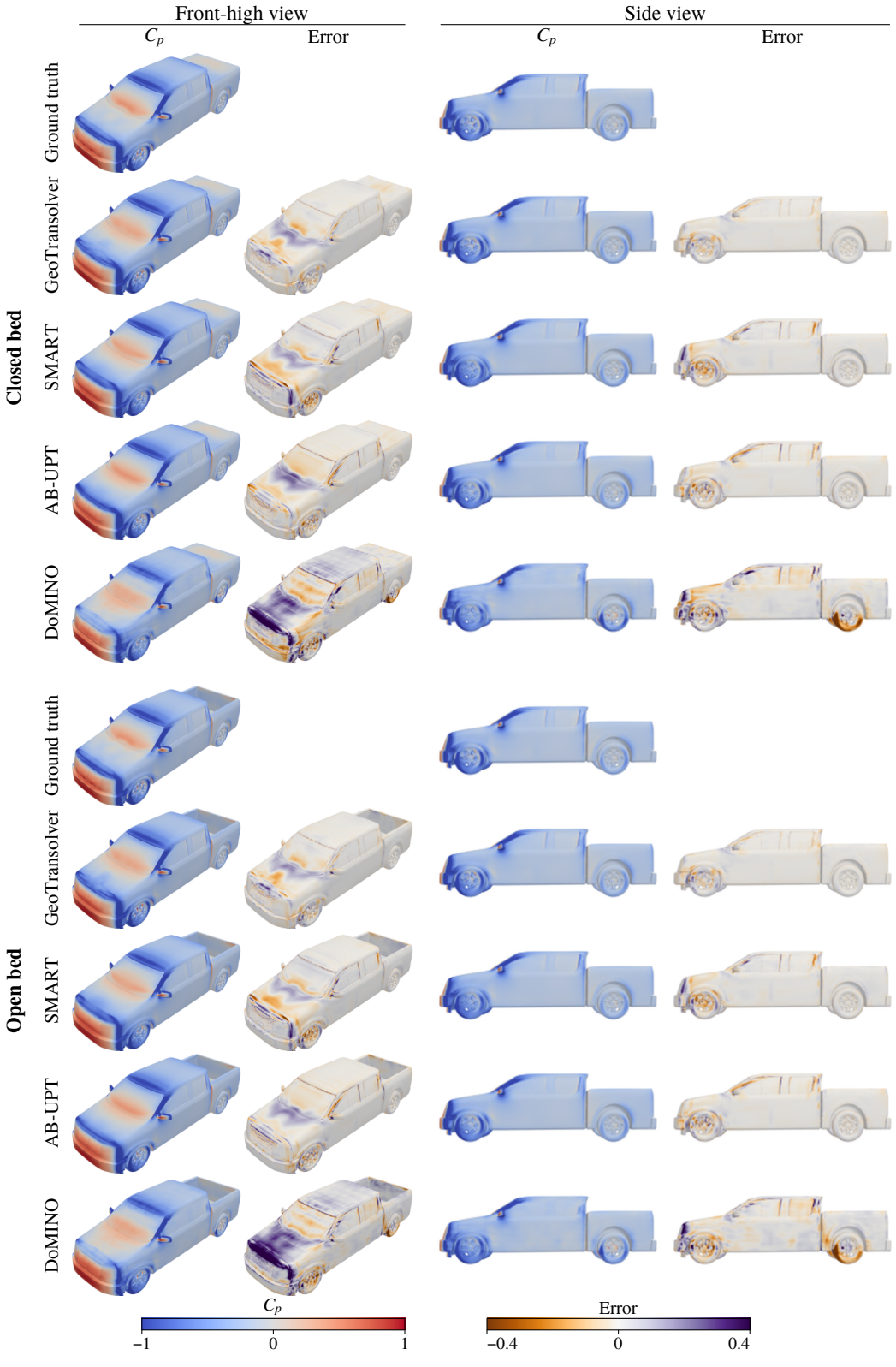

Figure 11: Surface pressure coefficient for the matched pair of Fig. 4, with the bed closed (top) and open (bottom), in the front-high and side views. Each prediction is shown beside its error, the prediction minus the ground truth. Both bed states share one error scale.

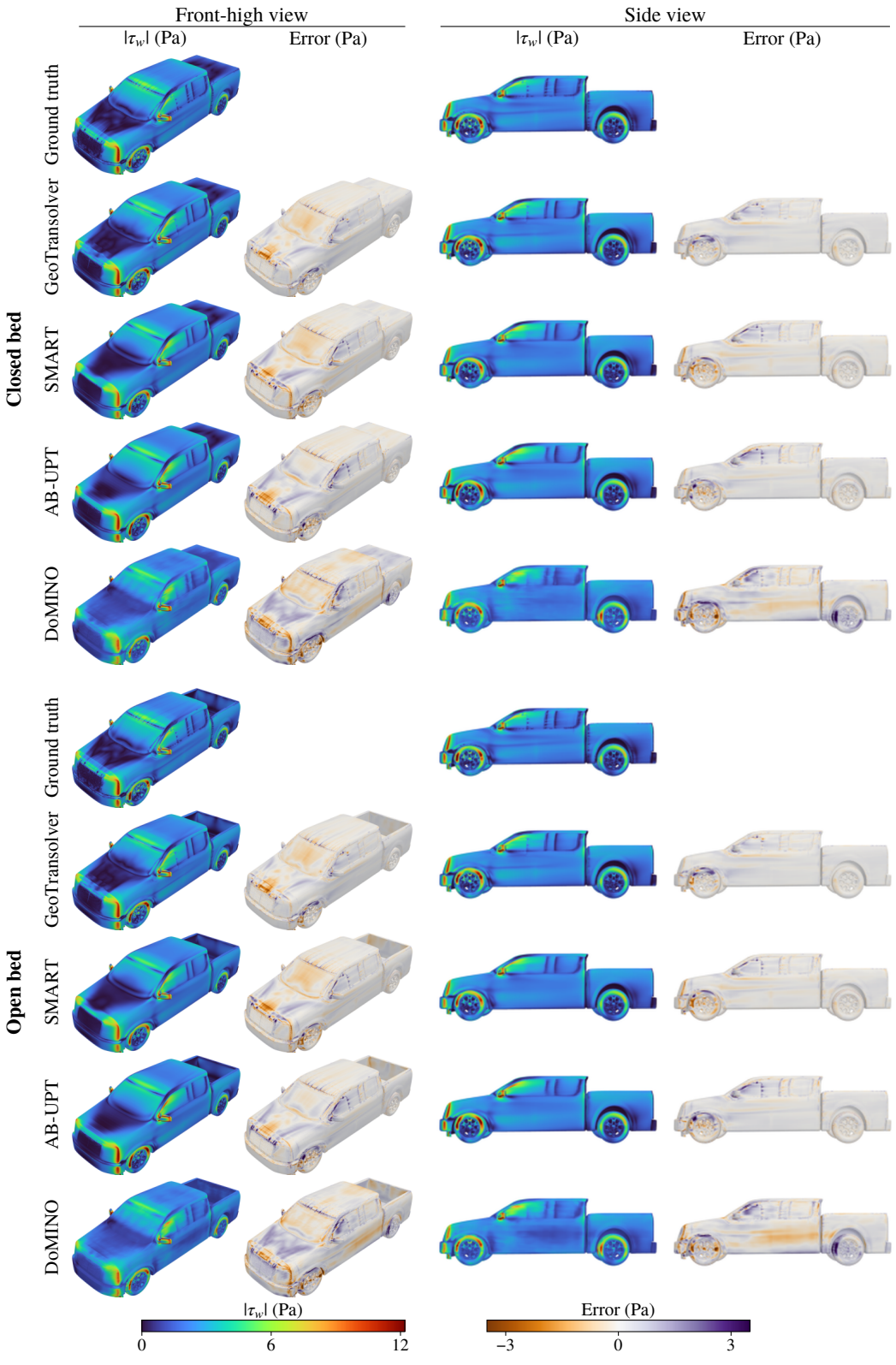

Figure 12: Wall-shear-stress magnitude for the matched pair of Fig. 4, laid out as in Fig. 11.

flow, where the true field is smooth. DoMINO also over-predicts the pressure in the near wake of the closed truck.

F RELEASE, FILES AND DATASHEET

F.1 ACCESS AND LICENSE

The dataset is released under the CC-BY-NC-4.0 license, and is available with its Croissant metadata (Akhtar et al., 2024) at `gs://shift-truck-sample/`. The GTU baseline remains under the terms of its distributor, ECARA, and work that uses it should cite Woodiga et al. (2020).

F.2 FILES

Each case is one directory of six files (Table 9). The root of the release holds `doe_parameters.csv`, with the design parameters, force coefficients and provenance of every case. It also holds the license and three utility scripts, which check the release, integrate the surface fields to forces and visualize a case. A case takes about 781 MB without its volume file and 11.9 GB with it. The full release takes about 9.5 TB. The surface, force and metadata files of all cases take about 0.6 TB and can be downloaded without the volume files, which suffices for the surface track.

Table 9: Files per case, with sizes for `variant_0000`.

File	Format	Contents	Size
`merged_surfaces.stl`	binary STL	vehicle surface, triangulated	386 MB
`merged_surfaces.vtp`	VTK PolyData	time-averaged pressure and wall shear stress, point and cell data	395 MB
`merged_volumes.vtu`	VTK UnstructuredGrid	time-averaged pressure and velocity on the full mesh	11.1 GB
`forces.json`	JSON	C_D, C_L, length of the averaging window	—
`metadata.json`	JSON	solver run provenance, iteration count, physical time, removed patches	—
`params.json`	JSON	reference values of Table 5	—

F.3 FIELDS

The surface file carries `Pressure Average (Pa)`, the time-averaged absolute static pressure, and the three components of `Wall Shear Stress Average (N/m`^{2}`)`. Both are stored as point data and as cell data. The file also carries the unit normals and the index of each point and cell in the full simulation surface. The volume file carries `Pressure Average (Pa)` and the three components of `Velocity Average (m/s)`. Gauge pressure follows from subtracting the reference pressure of Table 5. The contact patches under the wheels are open, and the absolute pressure level does not cancel in surface integrals. The surface covers the vehicle alone. The tunnel walls, the inlet, the outlet and the floor patches are removed, and `metadata.json` lists them. Every field and force coefficient is a time average over the final 10,000 steps, and the release contains no instantaneous fields or force histories. Each case has its own mesh, with 3.6 to 4.0 million surface points and about 98 million volume cells.

F.4 BENCHMARK SPLIT

The three case lists of the split of §3 are released with the dataset. In the 15-dimensional parameter space, with each parameter normalized to $[-1, 1]$, the median distance from a test truck to its nearest training truck is 1.87, against 1.89 from each training truck to its nearest other training truck, and the two distributions do not differ (two-sample Kolmogorov–Smirnov test, $p = 0.86$). None of the 15 parameter distributions differs between the training and test sets ($p \geq 0.06$). Test drag spans the range of the benchmark set, from 0.277 to 0.549, with a median of 0.400 against 0.383 in training.

F.5 MAINTENANCE

Cases are added to the release as their simulations complete. Each addition is recorded in a change log distributed with the release.

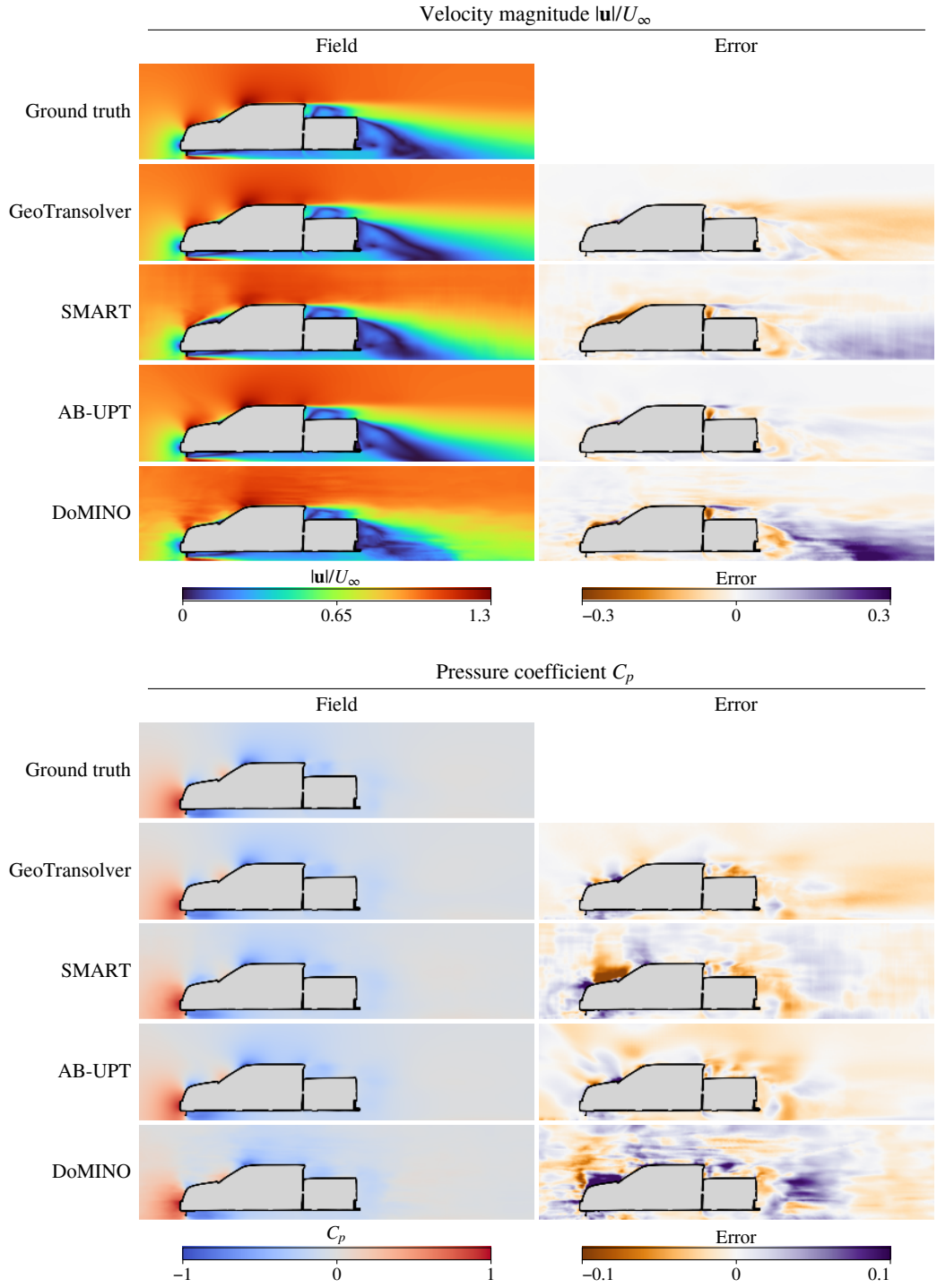

Figure 13: Velocity magnitude (top) and pressure coefficient (bottom) on the centreplane $y = 0$ for the closed-bed truck of the matched pair of Fig. 4. Each prediction is shown beside its error, the prediction minus the ground truth. Figure 14 uses the same colour scales.

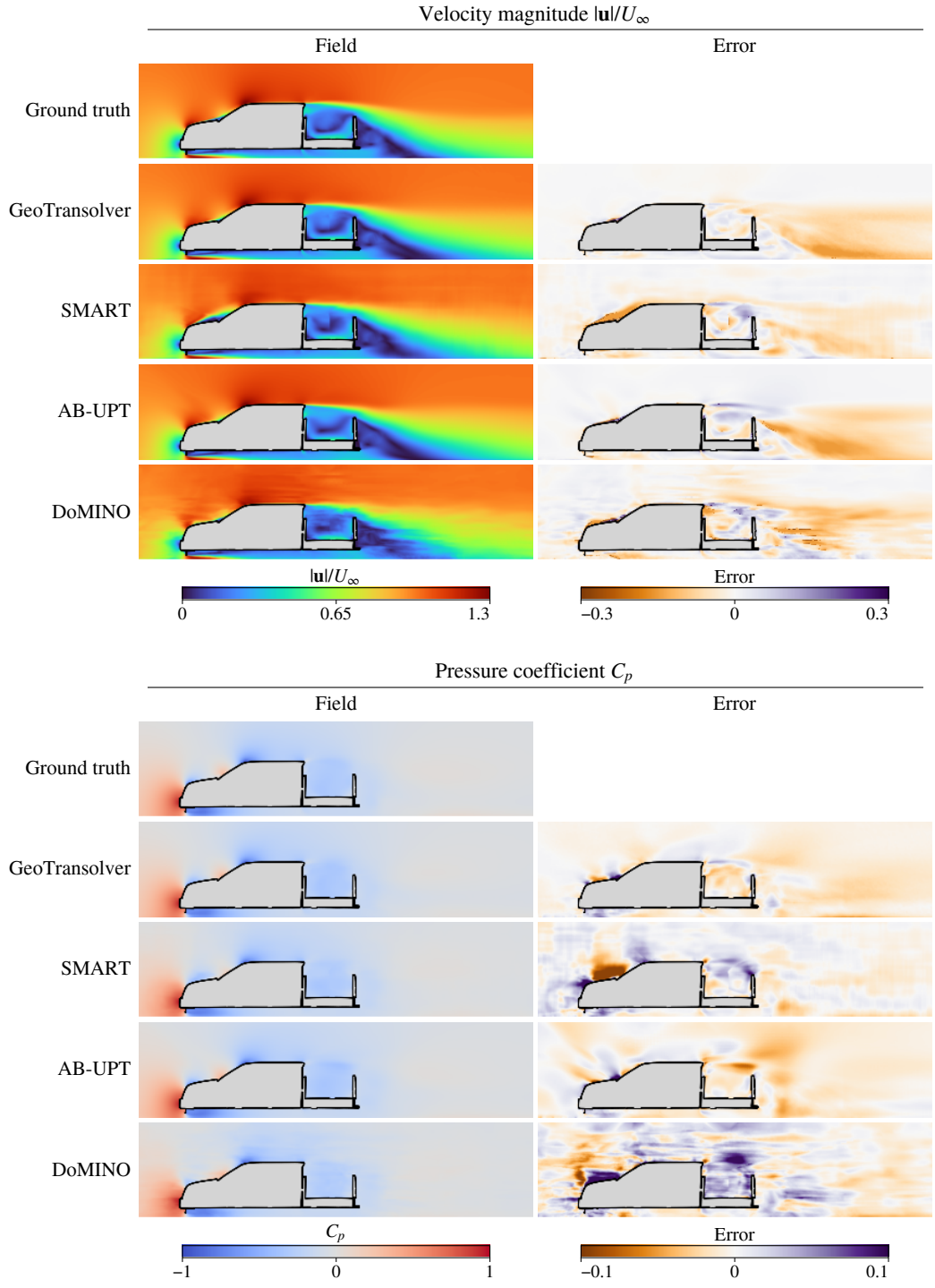

Figure 14: Velocity magnitude (top) and pressure coefficient (bottom) on the centreplane $y = 0$ for the open-bed truck of the matched pair, laid out as in Fig. 13.

Table 10: Datasheet.

Motivation	Training and benchmarking of neural surrogates for the external aerodynamics of pickup trucks, a vehicle class absent from open datasets.
Composition	800 cases, each a morphed GTU pickup at one operating point, with 15 continuous shape parameters and two parts switched on and off. Each case has a surface with time-averaged pressure and wall shear stress, a volume with time-averaged pressure and velocity, and its drag and lift coefficients. The data contain no personal information.
Collection	Geometries morphed from GTU configuration 7 with a deformation cage (Appendix A), then meshed and simulated with DDES (Appendix B) on 4 H100 GPUs per case. Appendix C validates the setup.
Preprocessing	Fields and forces averaged over the final 10,000 steps. Surface restricted to the vehicle. Instantaneous data discarded. Forces of two cases recovered by integration.
Uses	Surrogate training and benchmarking, studies of shape and drag, and transfer between vehicle classes. Absolute drag carries an offset against the wind tunnel (Appendix C). The data do not replace simulation or testing in safety-relevant decisions.
Distribution	Anonymized sample for review in a public cloud bucket, full release linked in the final version. CC-BY-NC-4.0, with Croissant metadata.
Maintenance	Cases added as they complete and recorded in a change log distributed with the release.

F.6 DATASHEET

Table 10 summarizes the dataset under the headings of Gebru et al. (2021).

G EXISTING DATASETS

Table 11 lists the open high-fidelity datasets for external vehicle aerodynamics. AhmedML, WindsorML and DrivAerML (Ashton et al., 2024b;a;c) publish scale-resolving solutions, from detached-eddy or wall-modelled large-eddy simulation, on reference bodies, with a few hundred cases each. DrivAerNet, DrivAerNet++ (Elrefaie et al., 2025; 2024) and DrivAerStar (Qiu et al., 2025) extend a DrivAer-derived design space to thousands of cases with steady RANS. SHIFT-SUV (Luminary Cloud, 2025) applies detached-eddy simulation to parametric variants of the AeroSUV. SHIFT-Truck is the only entry that describes a pickup truck. Every dataset validates on one to four baseline geometries and none on its morphed variants.

Table 11: Open high-fidelity datasets for external vehicle aerodynamics. *n/s*: not stated in the source. Case counts are those of each release as of September 2026. Validation gives the drag error against the wind tunnel on the baseline geometries; where a source gives coefficients only, the error is computed from them.

Dataset	Class	Cases	Method	Surface	Volume	Experimental validation
AhmedML	bluff body	500	SA-DDES	yes	yes	C_D 4.7%, C_L 0.6% (baseline)
WindsorML	bluff body	355	WMLES	yes	yes	C_D 0.34 vs. 0.33 (baseline); PIV
DrivAerML	passenger car	500	SA-DDES	yes	yes	C_D 7.5% (baseline)
DrivAerNet	passenger car	4,000	RANS, k–ω SST	yes	yes	C_D 0.81% (baseline)
DrivAerNet++	passenger car	8,000	RANS, k–ω SST	yes	yes	C_D 2.2–4.9% (4 baselines)
DrivAerStar	passenger car	12,000	RANS, k–ω SST	yes	yes	C_D 1.04% (mean, 3 baselines)
SHIFT-SUV	SUV	2,000	DDES	yes	yes	n/s (FKFS tunnel)
SHIFT-Truck	**pickup truck**	800	SA-DDES	yes	yes	C_D 13.5% (baseline); ΔC_D within 0.007