Learning from Visual Demonstrations through Differentiable Nonlinear MPC for Personalized Autonomous Driving: Supplementary Material

Flavia Sofia Acerbo[1,2,3], Jan Swevers[2,4], Tinne Tuytelaars[3] and Tong Duy Son[1]

APPENDIX

In the following sections, we provide supplementary material that we omitted due to space limitations in the main text of the paper "Learning from Visual Demonstrations through Differentiable Nonlinear MPC for Personalized Autonomous Driving," submitted to IROS 2024. This additional material comprises extra plots and metrics aimed at validating DriViDOC models, thereby enhancing the transparency of the study. Firstly, we present the complete set of trained DriViDOC models tested in closed-loop simulations. Secondly, we offer more detailed benchmarking results, including individual driver outcomes for the baselines and the results of statistical tests.

A. Closed-loop Testing

a) State Simulations: In Figures 1, 2, 3, 4, 5, 6 we show relevant states from closed-loop simulations of DriViDOC models, compared with the corresponding driver distribution. Figure 1 is equivalent to the one shown in Figure 3 from the main text. In each plot, we display pairs of models trained on different drivers, except for Figure 6, which exclusively showcases Model12. On the right axis, we indicate the curvature with a dashed line. In the d plots, the lane boundaries are shown as horizontal solid lines.

b) NMPC Parameters: Alongside the state simulations, we also provide the behavior of the NMPC dynamic parameters during the same closed-loop trajectories. These are illustrated in Figures 7,8, 9, 10, 11, and 12, corresponding to the same models depicted in the state simulation figures. Figure 7 mirrors Figure 4 in the main text. These plots offer valuable insights into how the different driving styles are realized by DriViDOC. The offsets $\bar{d}$ and $\bar{v_x}$ are normalized between -1 and 1. On the right axis, we indicate the curvature with a dashed line.

B. Benchmarking

a) Individual drivers results: In the main text, we present benchmarking results with baseline metrics averaged for all drivers in Table II. Here, we provide detailed results for each individual driver, similar to what we provided for DriViDOC in Table I of the main text. For the TRACK baseline, refer to Table I. For SF, refer to Table II. For NOIMG, refer to Table III.

b) Statistical Analysis: We present additional statistical results related to the benchmarking tests. In Figure 13 we show in the form of box-plots the mean Z-scores (MZ) of the relevant states for the TRACK, SF and NOIMG baseline, compared to DriViDOC. As mentioned in the main text, for NOIMG we exclude the models 00, 01, 08 and 09 due to convergence issues. In Figure 14, we show the same kind of representation for the MAE metrics.

Additionally, we assess the statistical significance of our results by conducting one-tailed paired-sample t-tests for each baseline. Specifically, we evaluate whether the baselines yield a statistically significant increase in the MZ and MAE metrics for each state compared to DriViDOC. In Table IV, the statistical results for TRACK are presented. DriViDOC demonstrates statistical superiority for v_x, a_x, and a_y, with notably low p-values, particularly for a_x, corroborating the qualitative findings outlined in the main text. However, the t-test for d is inconclusive, as no statistical significance of increase or decrease can be found. In Table V, the results for SF are presented. In this instance, DriViDOC exhibits superiority over SF for d, v_x, and a_y. However, the t-test for a_x is inconclusive, as no statistical significance of superiority or inferiority can be determined, which is not surprising as we highlighted in the main text that SF showed improvements in a_x for certain models. In Table VI, the results for NOIMG are presented. In this test, data from models 00, 01, 08, and 09 are excluded for both DriViDOC and NOIMG. Again, DriViDOC demonstrates statistical superiority for v_x, a_x, and a_y. The t-test on d is inconclusive as no statistical significance of superiority or inferiority can be found, even though, in Table II of the main text, it was shown that NOIMG was slightly better on average compared to DriViDOC.

[1]Siemens Digital Industries Software, Leuven, Belgium
[2]Dept. of Mechanical Engineering, KU Leuven, Belgium
[3]Dept. of Electrical Engineering (ESAT), KU Leuven, Belgium
[4]Flanders Make@KU Leuven, Leuven, Belgium
Corresponding author email: flavia.acerbo@siemens.com

TABLE I: TRACK Baseline: performance metrics of the driving models for each driver. Less optimal results (Mean Z-score > 2) are highlighted with a red colour.

	d		v_x		a_x		a_y	
	AE $[m]$	Z-score	AE $[m/s]$	Z-score	AE $[m/s^2]$	Z-score	AE $[m/s^2]$	Z-score
Model00	0.20 ± 0.15	0.96 ± 0.84	0.78 ± 0.40	1.53 ± 0.81	0.20 ± 0.18	4.60 ± 4.38	0.38 ± 0.29	1.29 ± 0.99
Model01	0.39 ± 0.30	1.07 ± 0.99	0.83 ± 0.61	0.95 ± 0.89	0.36 ± 0.28	1.36 ± 1.73	0.39 ± 0.32	0.94 ± 0.89
Model02	0.18 ± 0.14	1.19 ± 0.99	2.05 ± 1.18	1.01 ± 0.66	0.43 ± 0.31	1.76 ± 1.41	0.67 ± 0.55	1.52 ± 1.25
Model03	0.52 ± 0.37	2.32 ± 2.59	0.67 ± 0.55	0.62 ± 0.55	0.40 ± 0.37	3.53 ± 6.59	0.52 ± 0.36	1.36 ± 1.34
Model04	0.41 ± 0.38	1.58 ± 1.67	1.69 ± 0.89	1.88 ± 1.22	0.76 ± 0.60	3.42 ± 2.94	0.61 ± 0.55	1.75 ± 2.16
Model05	0.23 ± 0.18	1.36 ± 1.41	1.03 ± 0.91	1.13 ± 1.03	0.35 ± 0.28	1.13 ± 1.29	0.27 ± 0.23	1.03 ± 0.91
Model06	0.38 ± 0.29	0.94 ± 0.76	1.93 ± 1.25	1.03 ± 0.83	0.38 ± 0.29	1.68 ± 1.71	0.39 ± 0.32	0.96 ± 0.99
Model07	0.49 ± 0.28	2.03 ± 1.47	1.51 ± 1.10	1.09 ± 0.89	0.49 ± 0.39	3.40 ± 5.08	0.36 ± 0.29	0.92 ± 0.82
Model08	0.40 ± 0.30	0.96 ± 0.76	1.31 ± 0.94	1.20 ± 0.95	0.46 ± 0.39	2.01 ± 2.47	0.84 ± 0.65	1.91 ± 1.73
Model09	0.67 ± 0.50	2.29 ± 2.36	1.05 ± 0.69	2.11 ± 1.83	0.46 ± 0.32	3.97 ± 4.27	0.56 ± 0.47	1.94 ± 2.02
Model10	0.31 ± 0.20	1.74 ± 1.56	1.40 ± 1.00	1.17 ± 1.17	0.65 ± 0.40	3.04 ± 3.84	0.54 ± 0.49	1.25 ± 1.49

TABLE II: SF Baseline: performance metrics of the driving models for each driver. Less optimal results (Mean Z-score $>$ 2) are highlighted with a red colour.

	d		v_x		a_x		a_y	
	AE $[m]$	Z-score	AE $[m/s]$	Z-score	AE $[m/s^2]$	Z-score	AE $[m/s^2]$	Z-score
Model00	0.35 ± 0.22	1.73 ± 1.58	0.55 ± 0.36	1.18 ± 0.72	0.04 ± 0.11	0.75 ± 0.71	0.28 ± 0.22	0.99 ± 0.89
Model01	0.59 ± 0.40	1.71 ± 1.68	2.38 ± 2.02	2.91 ± 3.12	0.34 ± 0.27	0.95 ± 0.94	0.72 ± 0.61	1.64 ± 1.61
Model02	0.46 ± 0.33	3.18 ± 2.77	1.17 ± 0.74	0.55 ± 0.31	0.18 ± 0.15	0.71 ± 0.69	0.33 ± 0.26	0.83 ± 1.00
Model03	0.73 ± 0.54	3.14 ± 3.12	1.16 ± 0.89	1.12 ± 0.90	0.29 ± 0.29	2.65 ± 6.66	0.64 ± 0.82	1.45 ± 1.80
Model04	0.44 ± 0.29	1.85 ± 1.45	1.24 ± 1.08	1.28 ± 1.10	0.45 ± 0.49	1.26 ± 1.35	0.51 ± 0.46	1.35 ± 1.25
Model05	0.61 ± 0.37	3.76 ± 3.92	4.26 ± 1.63	3.90 ± 2.05	0.27 ± 0.22	0.78 ± 0.76	0.74 ± 0.53	2.40 ± 1.60
Model06	0.64 ± 0.41	2.11 ± 1.86	4.48 ± 2.48	2.21 ± 1.34	0.25 ± 0.21	1.02 ± 1.04	0.76 ± 0.61	1.50 ± 1.10
Model07	0.55 ± 0.41	2.28 ± 1.97	1.61 ± 1.17	1.17 ± 0.94	0.22 ± 0.21	0.86 ± 0.87	0.46 ± 0.37	1.08 ± 0.94
Model08	0.86 ± 0.53	2.06 ± 1.41	6.85 ± 3.66	7.15 ± 4.91	0.34 ± 0.28	2.63 ± 5.84	1.63 ± 1.08	3.54 ± 2.52
Model09	0.53 ± 0.37	1.87 ± 1.79	0.54 ± 0.34	1.04 ± 0.90	0.09 ± 0.13	0.54 ± 0.46	0.30 ± 0.25	0.95 ± 0.82
Model10	0.71 ± 0.54	3.93 ± 3.67	4.55 ± 2.49	3.44 ± 2.08	0.29 ± 0.23	1.07 ± 1.04	0.96 ± 0.67	1.90 ± 1.20

TABLE III: NOIMG Baselines: performance metrics of the driving models for each driver. Less optimal results (Mean Z-score > 2) are highlighted with a red color. The rows with a grey background correspond to the drivers with convergence to a high BC loss, who are excluded from average computations.

	d		v_x		a_x		a_y	
	AE $[m]$	Z-score	AE $[m/s]$	Z-score	AE $[m/s^2]$	Z-score	AE $[m/s^2]$	Z-score
Model00	1.60 ± 0.39	7.91 ± 4.20	13.67 ± 2.30	30.38 ± 9.85	0.14 ± 0.21	2.69 ± 1.88	2.55 ± 1.67	8.97 ± 7.38
Model01	1.75 ± 0.58	4.71 ± 2.95	12.21 ± 2.26	14.36 ± 7.50	0.34 ± 0.26	1.49 ± 2.47	2.16 ± 1.32	4.62 ± 3.37
Model02	0.15 ± 0.13	1.04 ± 1.01	4.11 ± 0.70	2.04 ± 0.74	0.24 ± 0.16	1.03 ± 0.74	1.00 ± 0.70	1.96 ± 1.27
Model03	0.31 ± 0.25	1.34 ± 1.44	1.73 ± 0.97	1.65 ± 0.97	0.43 ± 0.31	5.26 ± 9.36	0.62 ± 0.47	1.35 ± 0.95
Model04	0.24 ± 0.23	0.93 ± 0.87	1.49 ± 1.38	1.61 ± 1.57	0.57 ± 0.48	2.33 ± 2.19	0.63 ± 0.48	1.64 ± 1.23
Model05	0.52 ± 0.37	3.01 ± 2.80	2.09 ± 1.42	1.86 ± 1.64	0.29 ± 0.23	0.81 ± 0.68	0.35 ± 0.27	1.30 ± 1.09
Model06	0.28 ± 0.27	0.90 ± 1.11	3.27 ± 2.11	1.73 ± 1.42	0.23 ± 0.17	1.00 ± 1.03	0.68 ± 0.67	1.53 ± 1.68
Model07	0.31 ± 0.29	1.31 ± 1.28	3.15 ± 1.62	2.30 ± 1.33	0.33 ± 0.25	2.00 ± 2.62	0.65 ± 0.51	1.50 ± 1.16
Model08	1.48 ± 0.54	3.55 ± 1.95	13.65 ± 2.62	14.14 ± 6.51	0.43 ± 0.25	2.98 ± 3.40	2.56 ± 1.40	5.23 ± 3.46
Model09	1.57 ± 0.58	5.41 ± 4.07	12.53 ± 2.18	24.69 ± 14.27	0.17 ± 0.19	1.43 ± 1.53	2.32 ± 1.40	7.88 ± 6.42
Model10	0.22 ± 0.15	1.30 ± 1.13	1.88 ± 1.24	1.50 ± 1.25	0.43 ± 0.32	1.94 ± 3.80	0.60 ± 0.51	1.30 ± 1.15

TABLE IV: Results of the one-tailed paired-sample t-test, validating increase on MAE and MZ for TRACK compared to DriViDOC. Statistically significant increases are highlighted in bold.

	d	v_x	a_x	a_y
MAE	p=0.514, stat=-0.04	**p=0.017, stat=2.45**	**p=0.000, stat=7.10**	**p=0.011, stat=2.70**
MZ	p=0.618, stat=-0.31	**p=0.026, stat=2.21**	**p=0.000, stat=5.43**	**p=0.015, stat=2.54**

TABLE V: Results of the one-tailed paired-sample t-test, validating increase on MAE and MZ for SF compared to DriViDOC. Statistically significant increases are highlighted in bold.

	d	v_x	a_x	a_y
MAE	**p=0.001, stat=3.97**	**p=0.012, stat=2.65**	p=0.158, stat=1.05	**p=0.018, stat=2.43**
MZ	**p=0.002, stat=3.65**	**p=0.017, stat=2.46**	p=0.464, stat=0.09	**p=0.023, stat=2.28**

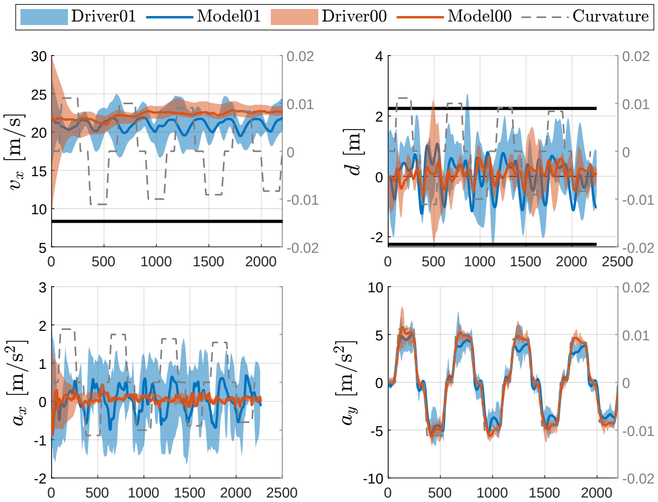

Fig. 1: DriViDOC models 00 and 01, compared with the corresponding driver distribution.

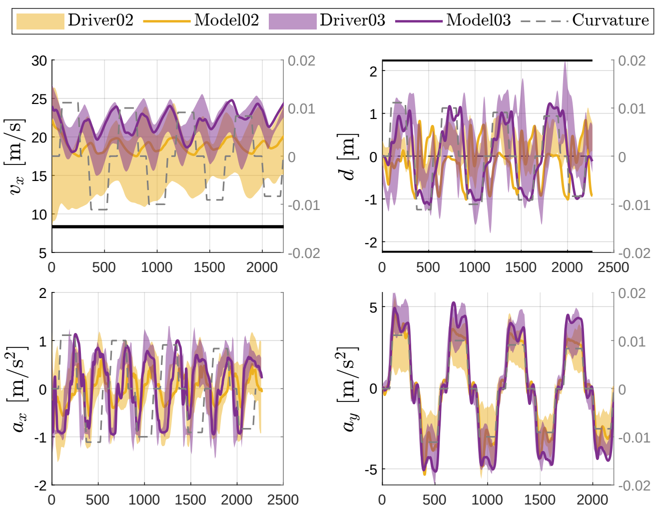

Fig. 2: DriViDOC models 02 and 03, compared with the corresponding driver distribution.

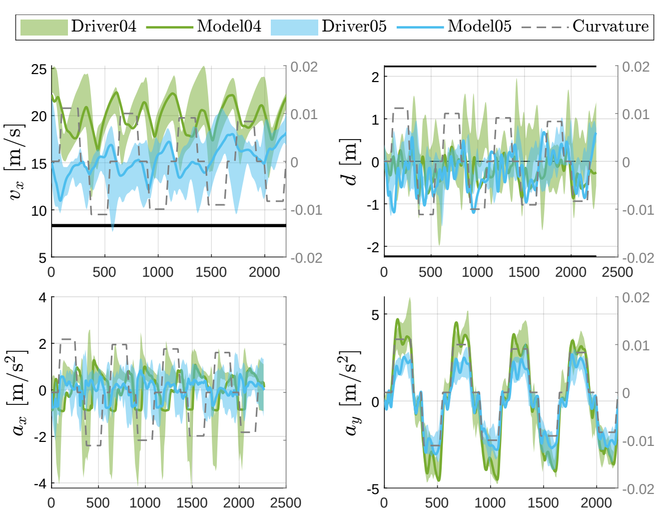

Fig. 3: DriViDOC models 04 and 05, compared with the corresponding driver distribution.

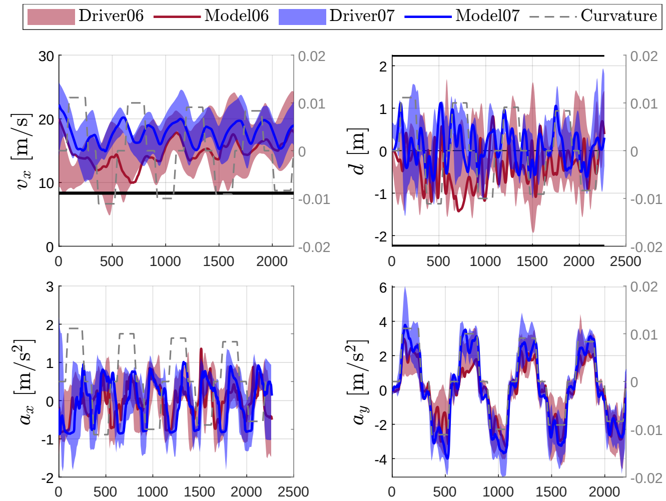

Fig. 4: DriViDOC models 06 and 07, compared with the corresponding driver distribution.

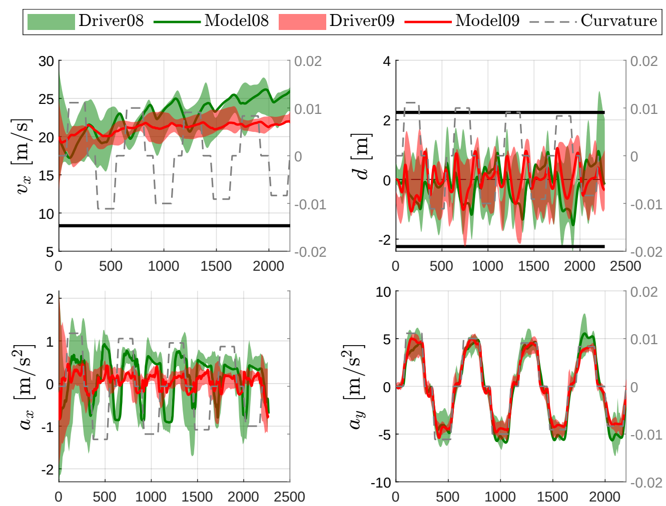

Fig. 5: DriViDOC models 08 and 09, compared with the corresponding driver distribution.

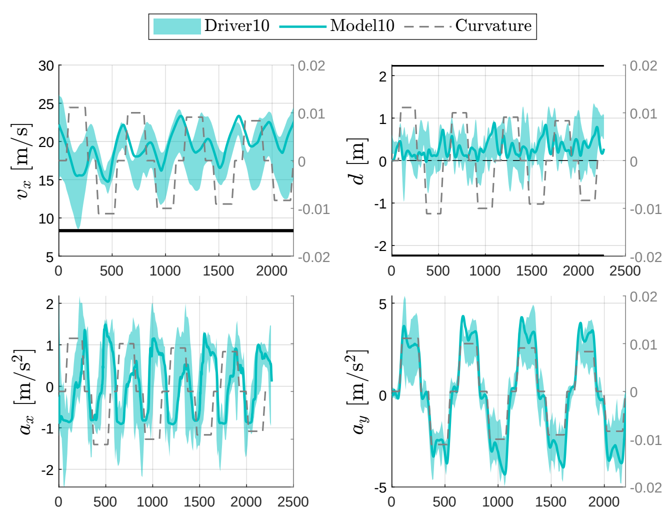

Fig. 6: DriViDOC model 12, compared with the corresponding driver distribution.

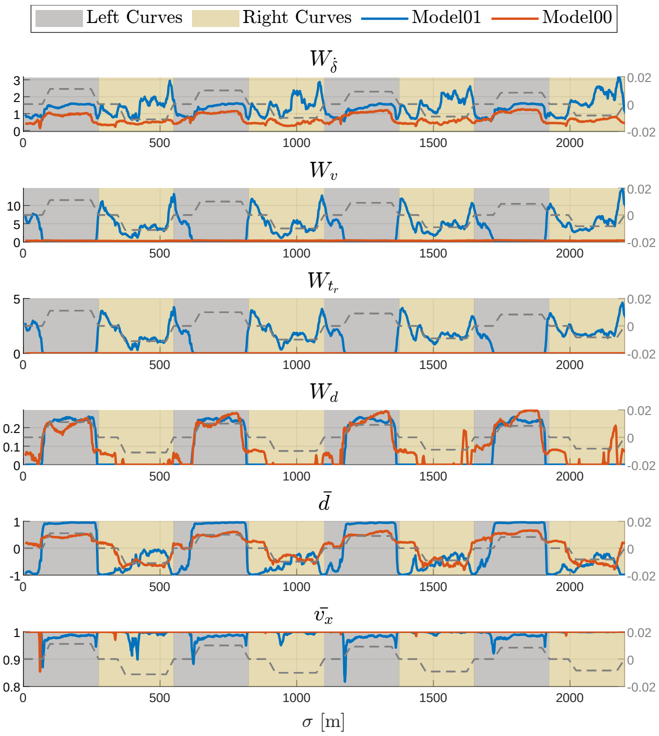

Fig. 7: Parameters $\mathbf{p}(t)$ values from closed-loop simulations of DriViDOC models 00 and 01.

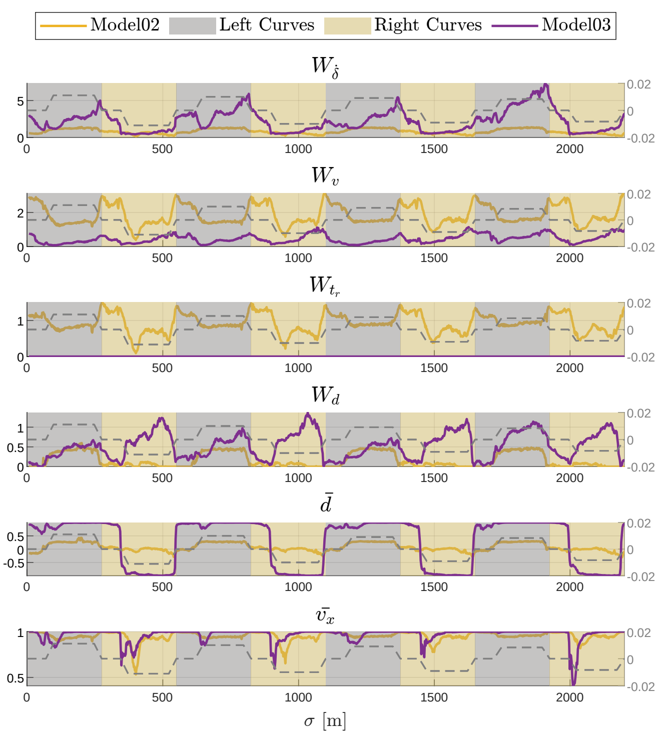

Fig. 8: Parameters $\mathbf{p}(t)$ values from closed-loop simulations of DriViDOC models 02 and 03.

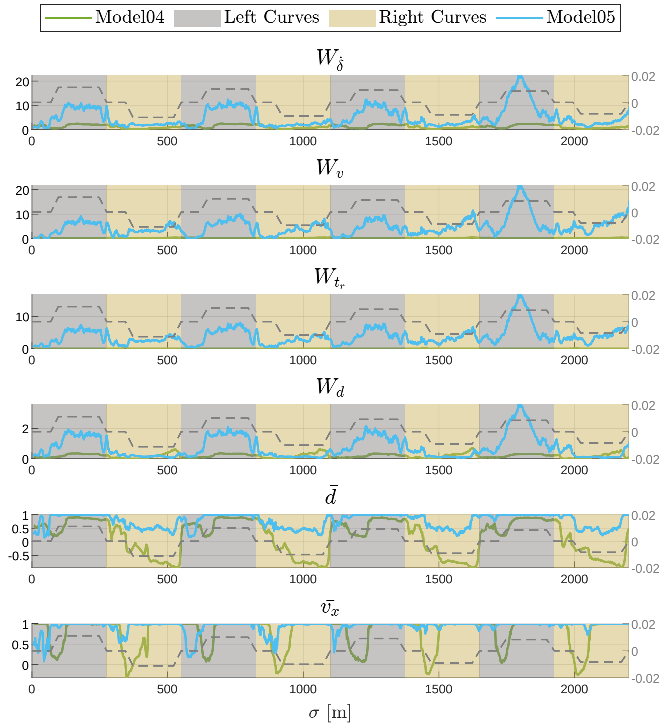

Fig. 9: Parameters $\mathbf{p}(t)$ values from closed-loop simulations of DriViDOC models 04 and 05.

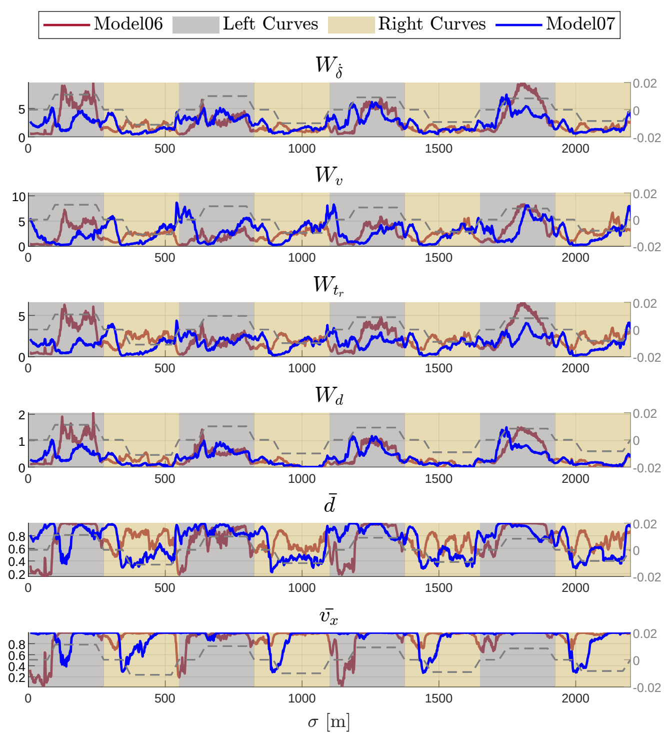

Fig. 10: Parameters $\mathbf{p}(t)$ values from closed-loop simulations of DriViDOC models 06 and 07.

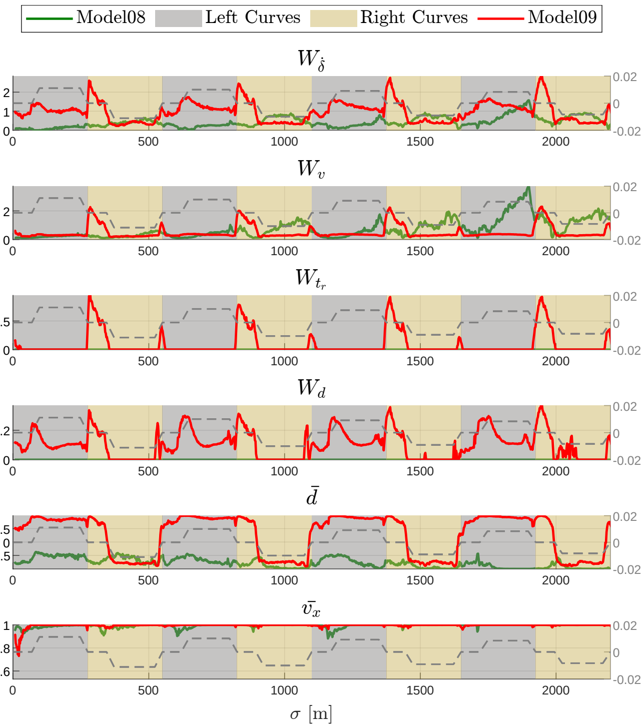

Fig. 11: Parameters $\mathbf{p}(t)$ values from closed-loop simulations of DriViDOC models 08 and 09.

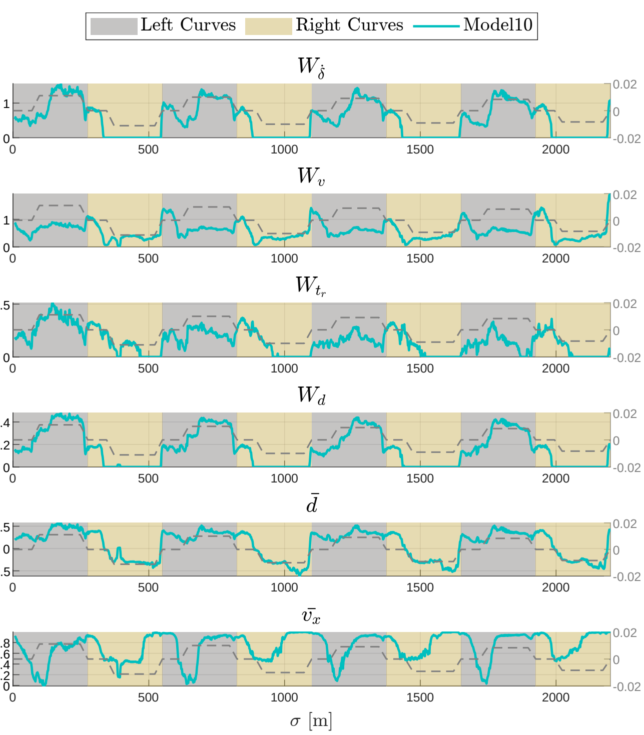

Fig. 12: Parameters $\mathbf{p}(t)$ values from closed-loop simulations of DriViDOC model 12.

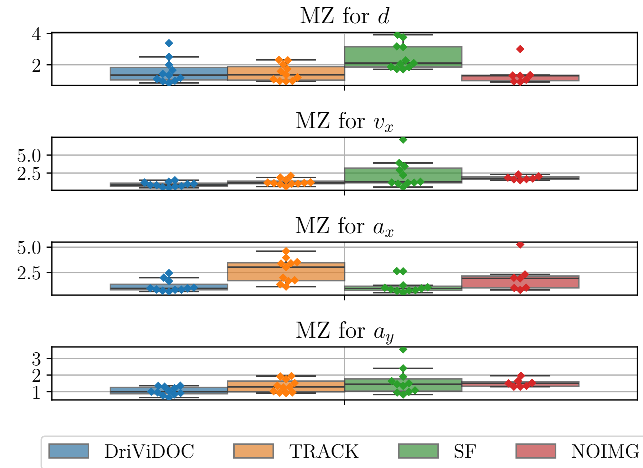

Fig. 13: DriViDOC benchmarking against TRACK, SF and NOIMG baselines. Box plots of the Mean Z-score (MZ), computed with respect to the driver's distributions, for states d, v_x, a_x and a_y.

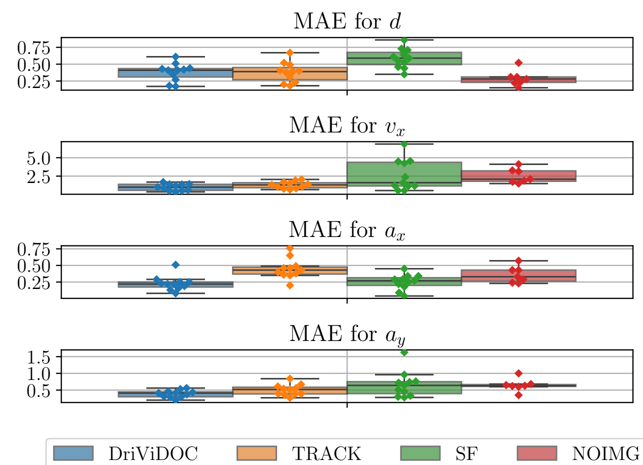

Fig. 14: DriViDOC benchmarking against TRACK, SF and NOIMG baselines. Box plots of the Mean Absolute Error (MAE), computed with respect to the driver's distributions, for states d, v_x, a_x and a_y.

TABLE VI: Results of the one-tailed paired-sample t-test, validating increase on MAE and MZ for NOIMG compared to DriViDOC. Models trained on drivers 00,03,10 and 11 have been excluded from the test (see main text). Statistically significant increases are highlighted in bold.

	d	v_x	a_x	a_y
MAE	p=0.897, stat=-1.42	**p=0.006, stat=3.53**	**p=0.011, stat=3.07**	**p=0.004, stat=3.89**
MZ	p=0.837, stat=-1.07	**p=0.005, stat=3.64**	**p=0.037, stat=2.15**	**p=0.007, stat=3.48**