Markov Regime-Switching Intelligent Driver Model for Interpretable Car-Following Behavior

Chengyuan Zhang[a], **Cathy Wu**[b], **and Lijun Sun**[a,†] († lijun.sun@mcgill.ca)

a. Department of Civil Engineering, McGill University, Montreal, Quebec H3A 0C3, Canada.

b. Laboratory for Information & Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Scan Me

Introduction

Table 1: Comparison of related approaches for modeling driver behavior.

Feature/Model	IDM	Bayesian IDM	GMM	HMM	HMM-GMM	HDP-HMM	NN (LSTM)	FHMM-IDM (Ours)
Model Type	Deterministic	Probabilistic	Probabilistic	Probabilistic	Probabilistic	Probabilistic	Deep Learning	Probabilistic
Adaptivity[1]	✗	✗	✗	✓	✓	✓	✓	✓
Latent Behavior Type[2]	✗	✗	Discrete	Discrete	Discrete	Discrete	Continuous	Discrete
Latent Mode Cardinality[3]	-	-	Fixed	Fixed	Fixed	Infinite	Fixed	Factorial Fixed
Stochasticity	✗	✓	✓	✓	✓	✓	Implicit	✓
Parameter Estimation[4]	Heuristic	MCMC	EM/MCMC	EM/MCMC	EM/MCMC	EM/MCMC	Gradient descent	MCMC
Interpretability[5]	High	High	Moderate	Moderate	Moderate	Moderate	Low	High
Traffic Context Modeling[6]	✗	✗	✓(features)	✗	✓(features)	✓(implicit)	✓(learned)	✓(explicit)
Heterogeneity Handling[7]	Poor	Moderate	Moderate	Moderate	Moderate	Excellent	Excellent	Excellent
Data-driven Flexibility[8]	Low	Moderate	Moderate	Moderate	Moderate	High	High	High
Training Complexity[9]	Low	Moderate	Low	Moderate	Moderate/High	High	High	High

IDM: Punzo et al. [2021], Treiber and Helbing [2003], Treiber et al. [2000, 2006]; **Bayesian IDM**: Zhang and Sun [2024], Zhang et al. [2024b]; **GMM**: Chen et al. [2023], Zhang et al. [2023, 2024a]; **HMM**: Aoude et al. [2012], Gadepally et al. [2013], Sathyanarayana et al. [2008], Vaitkus et al. [2014]; **HMM-GMM**: Wang et al. [2018a,b]; **HDP-HMM**: Taniguchi et al. [2014], Zhang et al. [2021], Zou et al. [2022]; **Neural Networks**: Mo et al. [2021], Wang et al. [2017], Yao et al. [2025], Zhou et al. [2025], Zhu et al. [2018];

[1] *Can the model dynamically adjust to changing behavior?* [2] *Type of latent representation: discrete (mode switches) or continuous (trajectory embeddings).* [3] *Whether the number of latent modes is fixed a priori or inferred.* [4] *How model parameters are estimated: EM, gradient descent, MCMC, etc.* [5] *Can latent states or parameters be interpreted as meaningful driving behavior?* [6] *Whether traffic context (e.g., relative speed, gap) is explicitly used in latent modeling.* [7] *Ability to capture driver-specific variation (e.g., hierarchical priors, class mixture).* [8] *Model's ability to fit and learn from diverse and high-dimensional driving datasets.* [9] *Overall training/inference complexity: data requirements, convergence cost, parallelism.*

Table 2: Clarification of key terminologies used in this study.

Term	Domain	Description
`Latent State`	Latent model variable	Inferred latent states in the HMM that jointly encode traffic scenarios and driving regimes. These are learned from data and do not correspond one-to-one with predefined scenarios or regimes.
`Traffic Scenario`	Traffic context	Inferred external traffic conditions such as free-flow or stop-and-go. Used to contextualize the scenario in which drivers operate.
`Driving Regime`	Driver behavior	A specific, short-term car-following action or behavioral mode (e.g., aggressive gap-closing, cautious following, free-flow cruising). These are transient states, not fixed driver traits.
`Case`	Empirical examples	Real-world driving trajectories selected from the highD dataset to illustrate representative behaviors and validate the model.

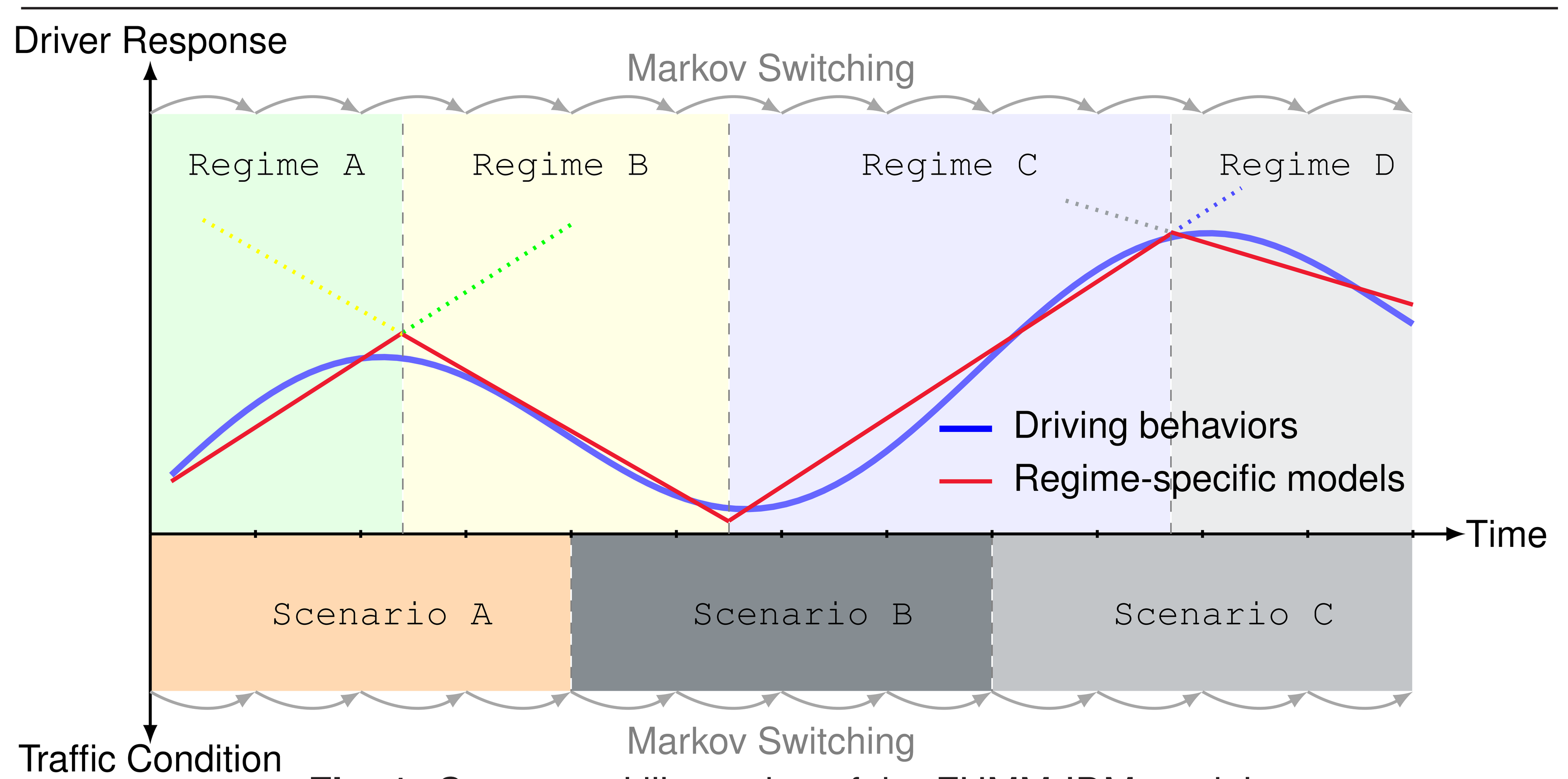

Fig. 1: Conceptual illustration of the FHMM-IDM model.

Summary

- **Markov regime-switching framework:** Separates intrinsic driving regimes from external traffic scenarios, reducing non-identifiability in naturalistic data.
- **FHMM-IDM with Bayesian inference:** Combines an FHMM with IDM; regimes map to distinct IDM parameters, and two latent Markov chains capture intention and context. MCMC enables robust calibration and uncertainty quantification.
- **Interpretable results:** Disentangling behavior and context supports regime-aware analysis. Experiments on HighD recover meaningful regimes and realistic transitions.

arXiv: https://arxiv.org/pdf/2506.14762

Code: The code will be released upon acceptance of the paper.

Methodology

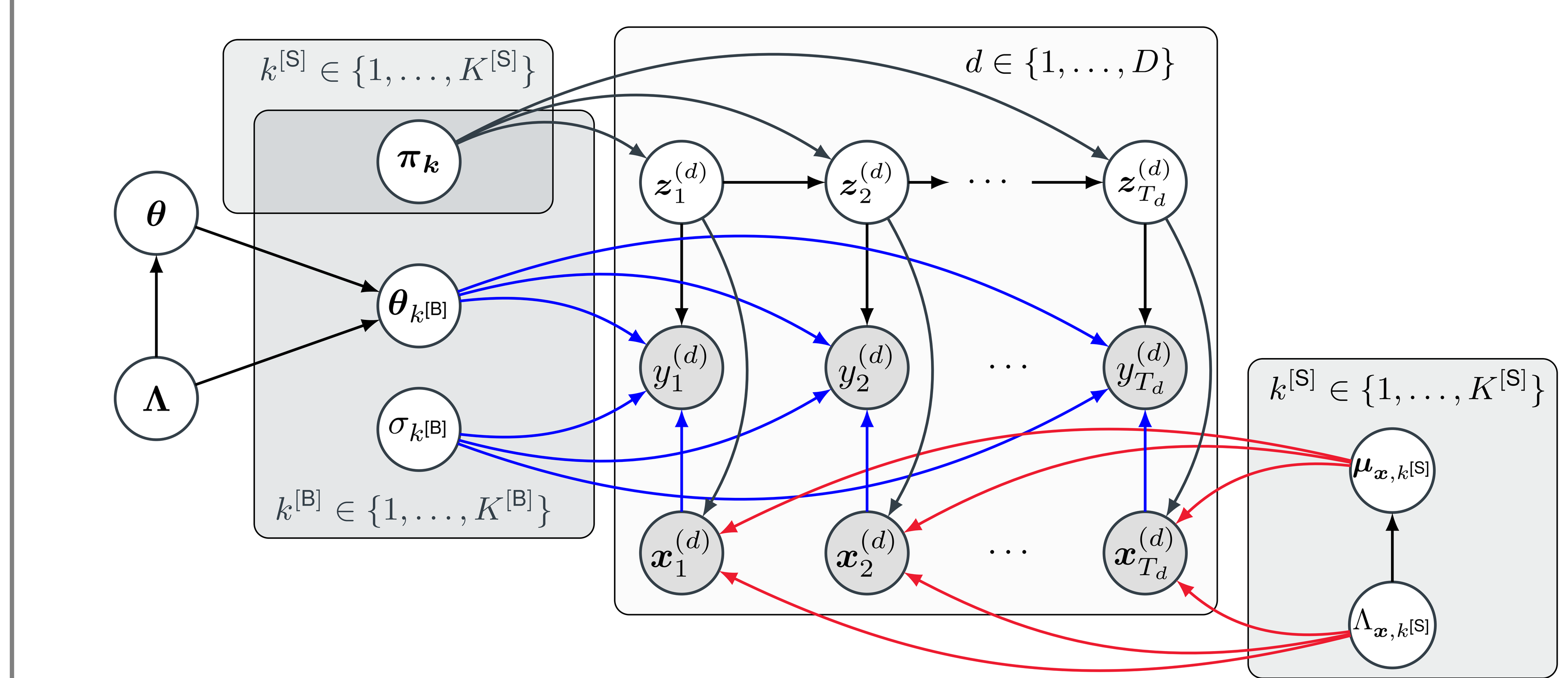

Fig. 2: Probabilistic graphical model of FHMM-IDM.

We define the observation model in FHMM-IDM with separate emission functions for the two factors:

1. **Driving-Behavior Local Evidence** $\psi_t^{[B]} \in \mathbb{R}^{k^{[B]}}$: The observed output y_t is independently influenced by the latent states $z_t^{[B]}$ and the covariates $\boldsymbol{x}_t$. The emission is modeled as:

$$y_t \mid \boldsymbol{x}_t, \boldsymbol{\Theta}, z_t^{[B]} \sim \mathcal{N}\left(\mathrm{IDM}\left(\boldsymbol{x}_t; \boldsymbol{\theta}_{z_t^{[B]}}\right), \sigma^2_{z_t^{[B]}}\right), \tag{1}$$

where $\mathrm{IDM}\left(\boldsymbol{x}_t; \boldsymbol{\theta}_{z_t^{[B]}}\right)$ is the predicted output based on the IDM, and $\sigma^2_{z_t^{[B]}}$ is the variance of the noise for state $z_t^{[B]}$.

2. **Traffic-Scenario Local Evidence** $\psi_t^{[S]} \in \mathbb{R}^{K^{[S]}}$: For the traffic scenario, we model the relationship between the covariates $\boldsymbol{x}_t$ and the latent state $z_t^{[S]}$ as

$$\boldsymbol{x}_t \mid z_t^{[S]}, \boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Lambda}_{\boldsymbol{x}} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{x}, z_t^{[S]}}, \boldsymbol{\Lambda}^{-1}_{\boldsymbol{x}, z_t^{[S]}}\right), \tag{2}$$

where $\boldsymbol{\mu}_{\boldsymbol{x}, z_t^{[S]}}$ and $\boldsymbol{\Lambda}_{z_t^{[S]}}$ are the mean and precision matrix of the scenario-driven input. We represent the collections of these parameters by $\boldsymbol{\mu}_{\boldsymbol{x}} = \{\boldsymbol{\mu}_{\boldsymbol{x},k^{[S]}}\}_{k^{[S]}=1}^{K^{[S]}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{x}} = \{\boldsymbol{\Lambda}_{\boldsymbol{x},k^{[S]}}\}_{k^{[S]}=1}^{K^{[S]}}$.

Therefore, the joint local evidence is given as

$$p(y_t, \boldsymbol{x}_t \mid \boldsymbol{z}_t, \boldsymbol{\Theta}, \boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Lambda}_{\boldsymbol{x}}) = p\left(y_t \mid \boldsymbol{x}_t, \boldsymbol{\Theta}, z_t^{[B]}\right) \cdot p\left(\boldsymbol{x}_t \mid z_t^{[S]}, \boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Lambda}_{\boldsymbol{x}}\right) \tag{3a}$$

$$= \mathcal{N}\left(y_t; \mathrm{IDM}\left(\boldsymbol{x}_t; \boldsymbol{\theta}_{z_t^{[B]}}\right), \sigma^2_{z_t^{[B]}}\right) \cdot \mathcal{N}\left(\boldsymbol{x}_t; \boldsymbol{\mu}_{\boldsymbol{x}, z_t^{[S]}}, \boldsymbol{\Lambda}^{-1}_{\boldsymbol{x}, z_t^{[S]}}\right). \tag{3b}$$

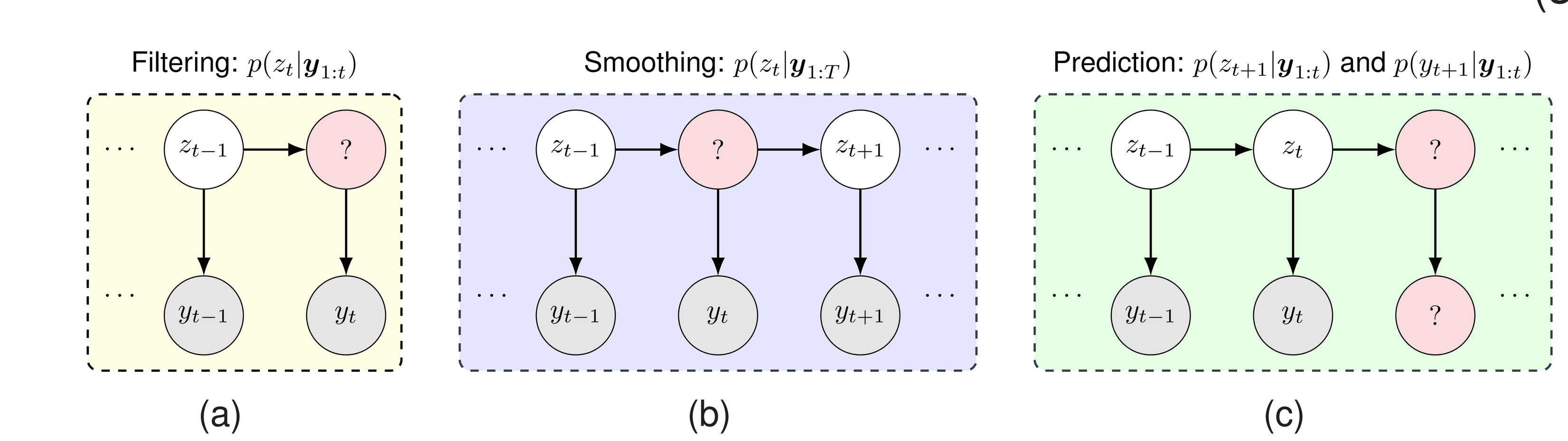

Fig. 3: Illustration of the filtering, smoothing, and prediction problem in HMM.

Identification of Interpretable Driving Regimes

Table 3: Learned IDM parameters ($\boldsymbol{\theta}_k$) and noise standard deviation (σ_k) for each regime.

Total states	State No.	$\boldsymbol{\theta}_k = [v_f, s_0, T, a_{\max}, b]$	$\boldsymbol{\sigma_k}$	Driving Regime
$K^{[B]} = 1$	#1	[38.57, 2.71, 0.86, 0.14, 1.33]	0.40	Averaged Behavior
$K^{[B]} = 2$	#1	[39.26, 1.80, 0.59, 0.30, 1.40]	0.47	High-Speed Seeking
	#2	[9.84, 5.10, 1.29, 0.08, 0.50]	0.15	Congested Cruising
$K^{[B]} = 5$	#1	[31.51, 4.32, 1.60, 0.13, 1.42]	0.11	Cautious Following
	#2	[34.10, 1.17, 0.44, 0.90, 4.45]	0.33	Aggressive Following
	#3	[11.26, 10.20, 3.09, 0.07, 1.51]	0.23	Congested Cruising
	#4	[33.11, 2.15, 0.90, 0.37, 1.51]	0.08	Steady-State Following
	#5	[42.11, 1.15, 0.71, 0.62, 1.72]	0.11	High-Speed Seeking

Table 4: Learned parameters of each traffic scenario latent state. Each scenario is characterized by the mean speed μ_v, relative speed $\mu_{\Delta v}$, and spacing μ_s, forming the mean vector $\boldsymbol{\mu}_{\boldsymbol{x},k^{[S]}}$. The interpretation column describes the typical traffic condition reflected by each state, inferred from statistical patterns and their behavioral context.

Total states	Scenario No.	$\boldsymbol{\mu}_{\boldsymbol{x},k^{[S]}} = [\mu_v, \mu_{\Delta v}, \mu_s]$	Interpretation
$K^{[S]} = 1$	#1	[5.73, 0.00, 14.32]	Averaged Traffic
$K^{[S]} = 2$	#1	[3.95, -0.08, 9.12]	Congested and Dense Traffic
	#2	[8.30, 0.10, 21.82]	High-Speed Cruising
$K^{[S]} = 5$	#1	[5.71, 0.73, 19.04]	Approaching (Stop-and-Go)
	#2	[6.20, -0.34, 38.96]	Gradual Dissipation
	#3	[4.89, 0.02, 12.67]	Steady-State Following
	#4	[3.66, -0.20, 6.90]	Congested and Dense Traffic
	#5	[10.22, -0.17, 16.54]	High-Speed Cruising

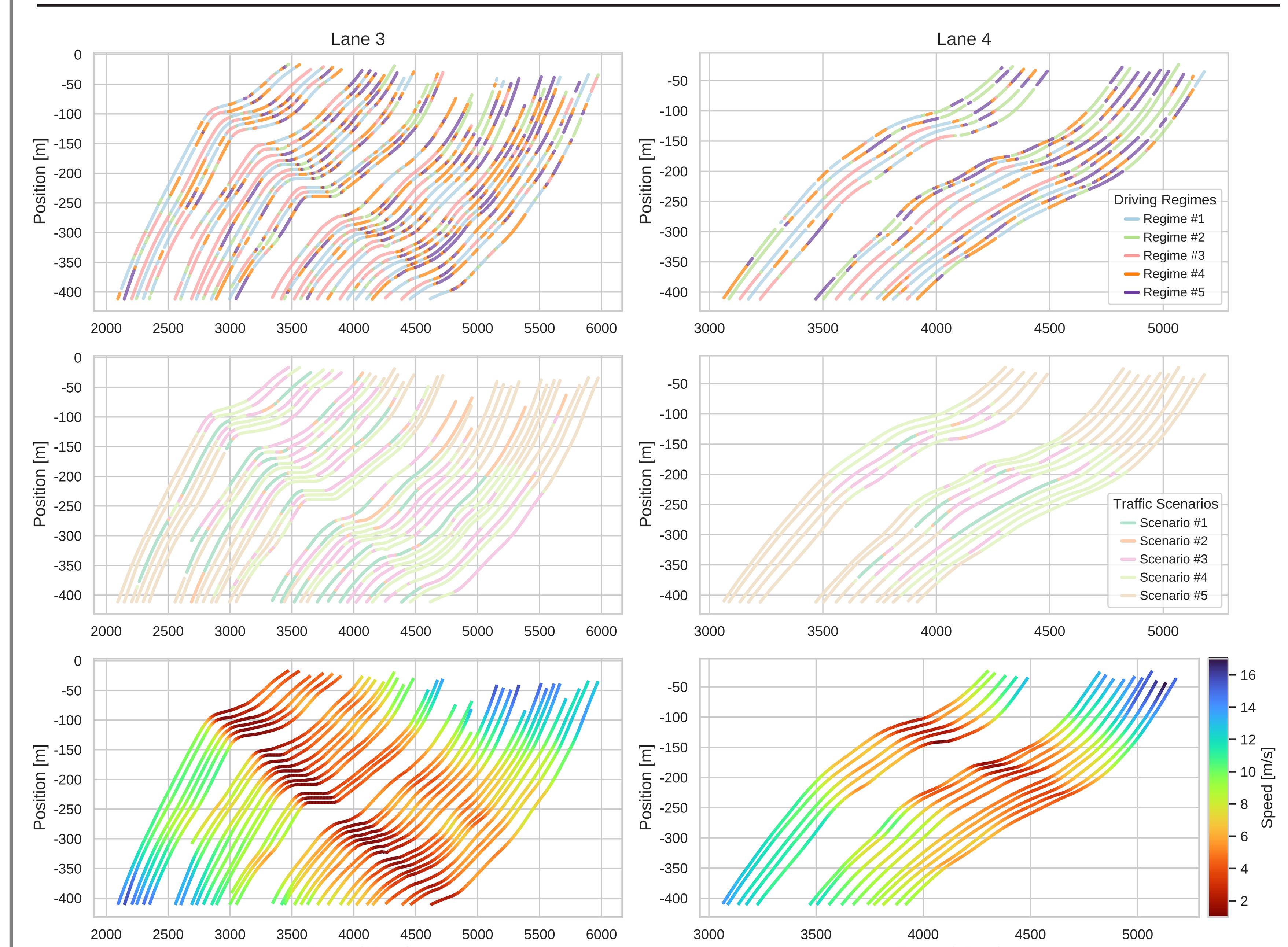

Fig. 4: Samples of time–space trajectories from the HighD dataset. *First row*: driving regime coloring. *Second row*: scenario coloring. *Third row*: Speed Coloring. Patterns: `Scenario #4` often aligns with lower-speed segments, while `Scenario #5` consistently aligns with higher-speed segments; `Scenario #1` often co-occurs with `Regime #3` in the approaching situations.