Data Analysis in Biological Systems and Traffic Flow

Nour Khoudari
Golomb Visiting Assistant Professor - Purdue University, West Lafayette

PURDUE UNIVERSITY

Data Science: Turning Data into Solutions

Data science is the art and science of turning raw data into meaningful insights. By combining mathematics, statistics, programming, and domain knowledge, data science allows us to uncover patterns, make predictions, and solve complex problems across disciplines. From understanding biological systems to managing traffic flow, data science is a transformative tool:

- In biology, it helps decode hidden structures in genetic data, cellular interactions, and ecological systems.
- In traffic systems, it models patterns of movement, optimizes traffic flow, and enhances urban planning for smarter cities.

By bridging these fields, data science not only uncovers the unseen, but it also empowers innovative solutions to some of today's pressing challenges.

Data Science in Biological Systems: Zebrafish Patterns

Data science in complex biological systems helps uncover patterns and mechanisms underlying certain formation processes. In particular, in zebrafish it reveals insights into developmental dynamics, such as the formation and regulation of their distinctive skin patterns (See figure1).

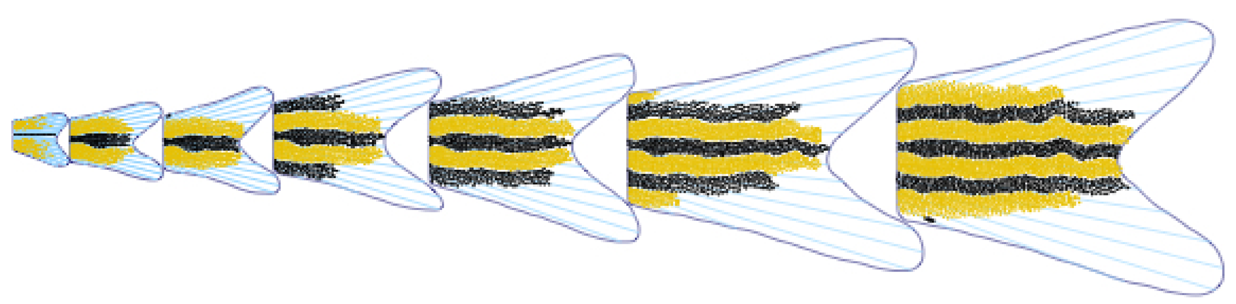

Figure 1. Example picture from [6] of stripe formation on zebrafish tail fins.

Toplogical Data Analysis

Topological Data Analysis (TDA) is an emerging field in data science that uses principles from topology to study the shape and structure of data. Unlike traditional methods, TDA focuses on the geometric and topological properties of data, such as connected components, loops, and voids, to uncover relationships and patterns that are often hidden in complex, high-dimensional datasets. TDA is particularly useful for analyzing data with noise, providing a robust framework for extracting meaningful patterns. For example, it can be used to study the spatial organization of biological systems like zebrafish skin patterns, understand the connectivity of neural networks, or analyze traffic flow to optimize urban planning.

When applied to zebrafish skin patterns [5], topological data analysis techniques can quantify properties related to certain features like stripes and spots, analyzing their consistency, connectivity, and variation (see figure 2). By characterizing these patterns mathematically, we can gain a deeper understanding of biological self-organization, identify deviations linked to genetic mutations, and even draw comparisons across species to study evolutionary traits.

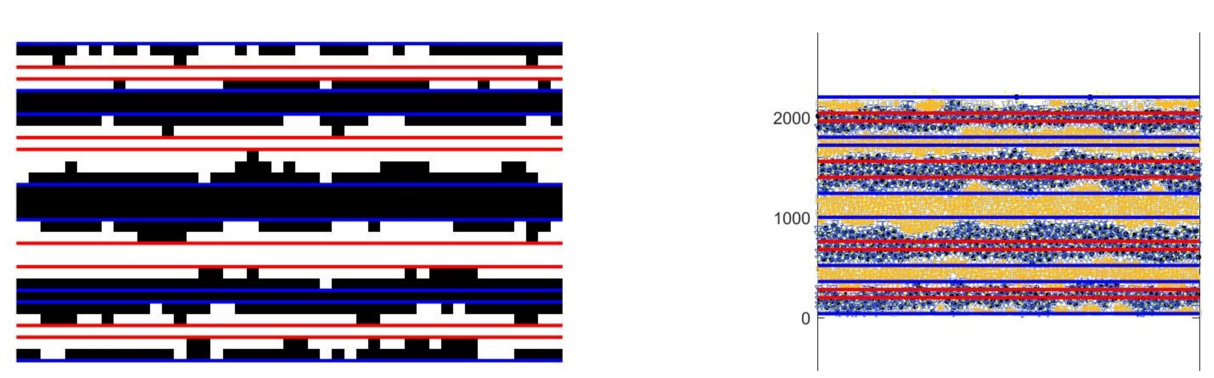

Figure 2. Example of quantifying stripe width using TDA on a pixel binned image (left) vs raw data image (right). Red lines denote the minimum estimated stripe width and the blue lines denote the maximum estimated stripe width.

Data Science in Traffic Flow: A Field Experiment with Autonomous Vehicles

Data science in traffic systems uncovers patterns in movement, congestion, and flow, enabling smarter transportation solutions. It plays a key role in field experiments, specifically with autonomous vehicles optimizing traffic efficiency. An example of such experiments is the CIRCLES project [1]. In this project, we aimed to reduce instabilities in traffic flow, called "phantom jams" occurring naturally due to human driving behavior. These instabilities are a significant source of wasted energy. Our last field experiment (2022) on I24 in Nashville leveraged a heterogeneous fleet of 100 longitudinally-controlled vehicles as Lagrangian traffic actuators [4]. An example of the data collected during this experiment can be seen in figure 4.

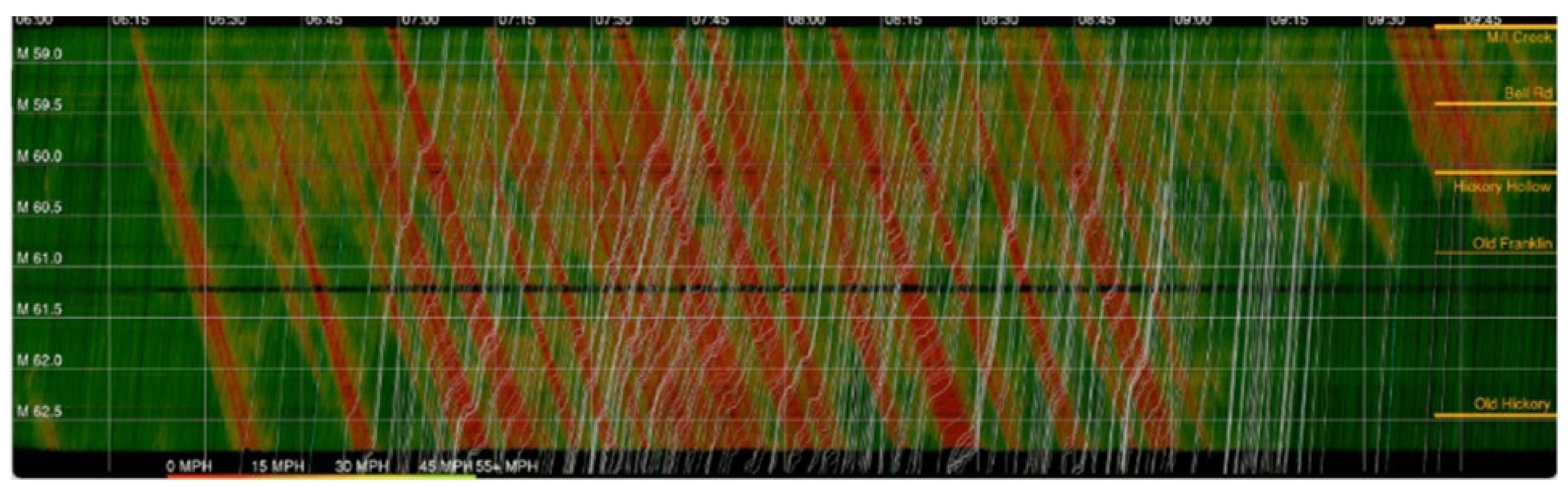

Figure 3. Time (horizontal axis) space (vertical axis) diagram generated by I-24 MOTION [2] and associated visualization library during the experiment. Vehicle trajectories are colored based on the speed of vehicles (green: freeflow to red:congested). AV trajectories in the experiment are overlaid in white.

The quantification of the energy demand of the vehicles on the road, given their trajectories, requires vehicle-specific energy models that take as an input the velocity profile $v(t)$ and the road grade profile $\theta(t)$ and output the resulting energy/fuel consumption rate. This project required the quantification of the energy demand of traffic flow at large, composed of many vehicles; and also the use of reinforcement learning and optimization techniques that minimize (under certain constraints) the energy demand of traffic. Thus, the energy models used had to accurately represent different vehicle types on the road without averaging out any local non-convexity behavior due to gear switching, to avoid trapping the optimizer in local minima. For that purpose, we used energy models described below.

Fuel Consumption Models

We develop energy models to quantify vehicle fuel consumption by fitting some high-fidelity fuel models into simplified models [3]. Those models have a simple polynomial structure that can easily be integrated into optimization and control problems, yet they are highly accurate. The fuel consumption rate function is

$$f(v,a,\theta) = \max\{\beta, (c_0+c_1v+c_2v^2+c_3v^3)+(p_0+p_1v+p_2v^2)a+(q_0+q_1v)a^2++(z_0+z_1v+z_2v^2)\theta\}.$$

To capture the diversity and prevalence of different vehicle types on US roads we select a representative group of vehicle classes on which we apply the model-reduction process to derive their corresponding simplified energy models. Those vehicle classes are divided into two categories: (1) light-duty vehicles: compact size sedan, midsize sedan, midsize SUV, and Pickup, and (2) heavy-duty vehicles: Class4PND (Pickup and Delivery) and Class8Tractor. The models are validated, for all different vehicle types, against Autonomie models as the ground truth on standard EPA drive cycles for flat roads and constant road grades drive cycles, and the results showed that the models are highly accurate (errors within 4% for zero road grades).

Macroscopic Fields from Microscopic Data

We construct macroscopically meaningful fields in time–space that helps analyze the energy efficiency/inefficiency of traffic at large on the I-24 highway segment. This can be achieved by applying Edie's method on boxes of size $h_x \times h_t$ to the I-24 MOTION trajectories to construct the following fields:

- vehicle density $\rho(t,x)$, as the total vehicle time spent in each box, divided by the size of the box, $h_t \cdot h_x$.
- flow rate $q(t,x)$, as the total distance traveled in each box, divided by $h_t \cdot h_x$.
- fuel rate density $f(t,x)$, as the total fuel consumed in each box, divided by $h_t \cdot h_x$.
- bulk velocity field $u(t,x) = q(t,x)/\rho(t,x)$.
- bulk fuel rate $\Phi(t,x) = f(t,x)/\rho(t,x)$.
- bulk fuel consumption $\Psi(t,x) = f(t,x)/q(t,x)$.

The figure below is an example of one of the macroscopic quantities plotted in time-space. This clearly shows the traffic waves traveling backwards along the highway, as well as the increased fuel consumption incurred in these waves. A similar study could be carried in local regions to show how AVs can decrease energy waste and reduce traffic congestion (see [4]).

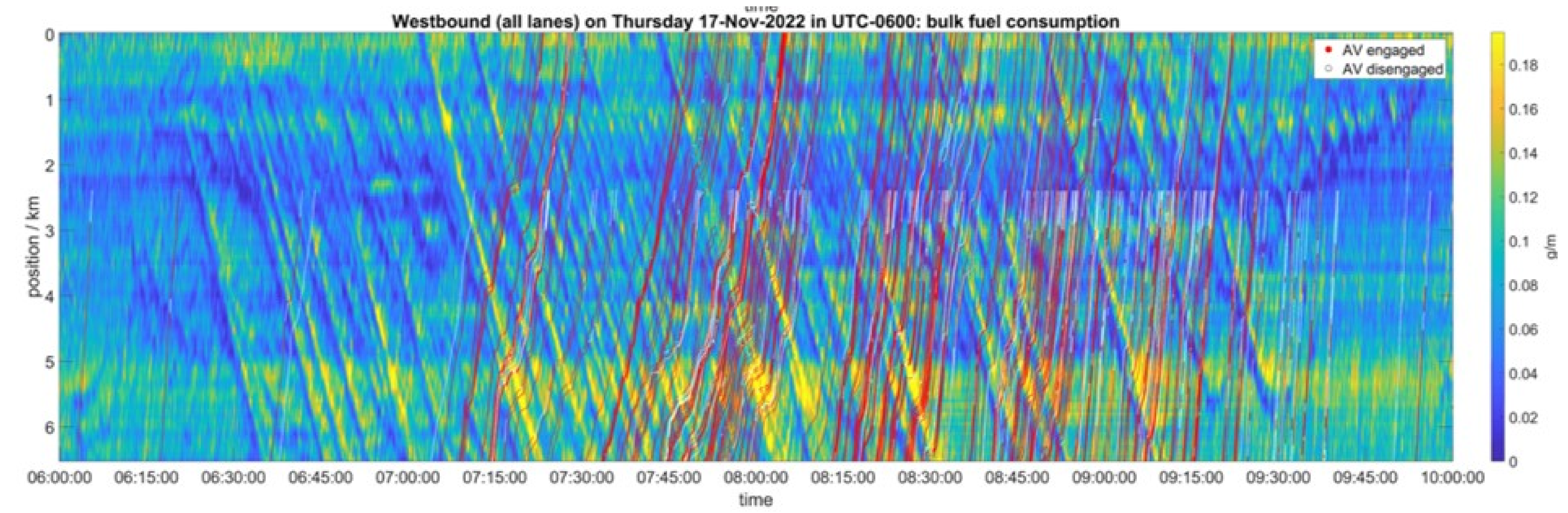

Figure 4. Bulk fuel consumption heatmap in time (horizontal axis) and space (vertical axis), based on I-24 MOTION [2] of all vehicles (aggregated over all lanes) driving in the Westbound direction. AV trajectories are overlaid (white: controller not engaged; red: controller engaged).

Conclusions

Data science bridges the gap between complexity and clarity, allowing us to understand and solve real-world problems. Its applications are vast and interdisciplinary, proving invaluable in advancing our understanding of life, society, and technology. Whether in biological systems or traffic flow, its tools and insights pave the way for smarter, more efficient solutions to today's challenges.

References

[1] *CIRCLES* web page, `https://circles-consortium.github.io`, 2020.

[2] D. Gloudemans, et al. *I-24 MOTION: An instrument for freeway traffic science*. Transportation Research Part C: Emerging Technologies. 155:104311, 2023.

[3] N. Khoudari, et al. *Reducing Detailed Vehicle Energy Dynamics to Physics-Like Models*. arxiv: `https://arxiv.org/abs/2310.06297`, 2023.

[4] J. W. Lee, et al. *Traffic Control via Connected and Automated Vehicles: An Open-Road Field Experiment with 100 CAVs*. IEEE Control Systems CSM, 2024.

[5] MR McGuirl, A Volkening, B Sandstede. *Topological data analysis of zebrafish patterns.*Proceedings of the National Academy of Sciences of the USA. 117(10):5113-5124, 2020.

[6] A. Volkening, MR Abbott, N Chandra, B Dubois, F Lim, D Sexton, B Sandstede.*Modeling stripe formation on growing zebrafish tailfins*. Bulletin of Mathematical Biology, 82:56, 2020.