ML approaches for Observation and Control under partial information with applications

Nail Baloul
under the supervision of Amaury Hayat, Thibault Liard, Pierre Lissy[1]

Benasque XI Partial Differential Equations, Optimisation, Numerics, and Machine Learning

September, 1st 2026

École nationale des ponts et chaussées
IP Paris

[1] N. Baloul, A. Hayat, and P. Lissy are with CERMICS, École nationale des ponts et chaussées, France
Thibault Liard is with XLIM, Université de Limoges, France

Table of Contents

- Abstract dynamical system (state $X \in \mathcal{X}$ and control $U \in \mathcal{U}$)[2]

$$\begin{cases} \frac{d}{dt}X(t) & = F\left(X(t), U(t)\right), \quad t \in \mathbb{R}^+, \\ X(0) & = X_0, \end{cases} \tag{1}$$

- Exact controllability consists in driving system from an initital state to a given targen state in a given time, i.e.

$$\begin{aligned} &\forall X_0 \in \mathcal{X}, \quad \forall X_1 \in \mathcal{X}, \quad \forall T > 0, \quad \exists\, (t \in [0, T] \mapsto U(t)) \in \mathcal{U}, \\ &\quad X \text{ solution to (1)}, \quad X(T) = X_1, \end{aligned} \tag{2}$$

- Optimal control consists in driving system while keeping a certain cost minimal

$$\text{Find a minimiser } U_m \text{ of } C(U) := \int_0^T R(X(t), U(t))dt + P(X(T)), \tag{3}$$

among all admissible control stategies

- Asymptotic stabilisation suppose the existence of a couple $(X_e, U_e) \in \mathcal{X} \times \mathcal{U}$ such that $F(X_e, U_e) = 0$ and provide a function $\mathbb{U} : \mathcal{X} \mapsto \mathcal{U}$ such $\mathbb{U}(X_e) = U_e$ for which

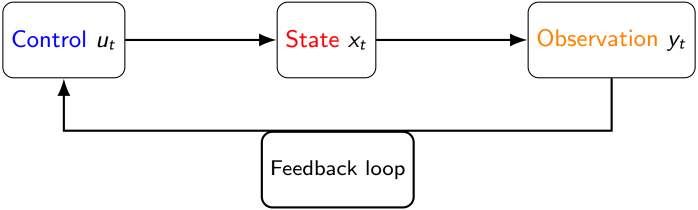
$$\frac{d}{dt}X(t) = F\left(X(t), \mathbb{U}(X(t))\right), \quad t \in \mathbb{R}^+, \quad X(0) = X_0, \tag{4}$$

is asymptotically stable

[2] a key aspect of mathematical analysis consists in selecting suitable spaces $\mathcal{X}, \mathcal{U}$

A central problem: we do not observe everything

- Most real-world dynamical systems have a hidden state: x_t is not fully observable
- Instead, we only observe $y_t = h(x_t) + \varepsilon_t$
- We then choose actions u_t



Our problem

Observe hidden states and **control system** using only *partial and noisy* measurements

A unified learning problem

- Learn from data the unknown dynamics

$$\frac{d}{dt}x = f(x, u)$$

- Estimate hidden variables

$$\hat{x}_t = \mathcal{O}_\alpha(y_{0:t}, u_{0:t})$$

- Choose actions

$$u_t = \pi_\theta(\hat{x}_t)$$

Learn $\longrightarrow$ *Observe* $\longrightarrow$ *Control*

From Supervised Learning to Neural ODE

- Given data $(x_i, y_i)_{i=1,\cdots,N}$ we learn a function $f_\alpha : x \mapsto y$ by minimizing

$$\min_\alpha \sum_{i=1}^{N} \mathcal{L}\left(y_i, f_\alpha(x_i)\right)$$

- Dynamical systems evolve continuously in time

$$\frac{dx(t)}{dt} = f_\alpha(x(t), u(t), t)$$

- Starting from initial data $x(0)$ the predicted state is obtained by integration

$$x_\alpha(t) = x(0) + \int_0^t f_\alpha(x(s), u(s), s)\, ds$$

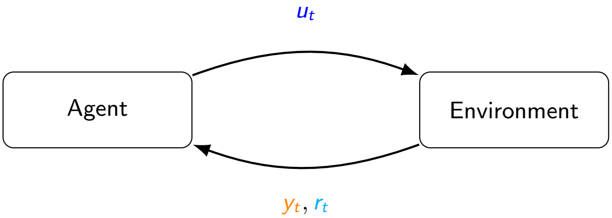
- Training compares predicted and observed trajectories

$$\min_\alpha \sum_k \|x_{\text{data}}(t_k) - x_\alpha(t_k)\|^2$$

Trajectory data → Neural Network f_α → ODE solver → Predicted trajectory

From prediction to decision: Reinforcement Learning

- In Reinforcement Learning, an agent interacts with an environment



- The agent learns a policy $u_t \sim \pi_\theta(\cdot \mid y_t)$ to maximize expected cumulative reward

$$\max_\theta \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t r_t\right]$$

Observation y_t $\longrightarrow$ Action u_t $\longrightarrow$ Reward r_t

Observation and Control: reconstructing hidden state with incomplete data

1. The objective of state reconstruction is to construct an estimator

$$\hat{x}_t = \mathcal{O}_\alpha(y_{0:t}, u_{0:t})$$

Possible ML approaches for state reconstruction

- Recurrent Neural Networks
- Neural ODE
- Physics-Informed Neural Networks
- Hybrid mechanistic + ML models [3]

2. Controlling with incomplete observation via an estimated state

$$u_t = \pi_\theta(\hat{x}_t)$$

Possible ML approaches for control

- POMDP
- recurrent RL
- belief-state methods
- model-based RL
- observer-controller architectures

[3] purely mechanistic models are interpretable, based on physical knowledge but may contain modeling errors while purely data-driven ones are flexible, can capture unknown nonlinearities but require large data amounts

Table of Contents

Traffic congestion

Traffic jam in Beijing

- Traffic congestion is a main contributor of *air pollution and excessive travel time*
- Congestion often appears without any accident: *small perturbations amplify*[4]
- Traffic behaves like a compressible fluid: it *self-organizes* into shocks

[4] Sugiyama, Fukui, Kikuchi, Hasebe, Tanaka, Tomoeda, and Tsuchiya 2008.

- Follow-the-Leader (FtL), **microscopic** first order model
  $\Rightarrow$ dynamics of *each vehicle depend on vehicle immediately in front*

$$\begin{cases} \frac{d}{dt}x_N^N(t) = v_{\max}, & t > 0, \\ \frac{d}{dt}x_i^N(t) = v\left(\frac{L}{N(x_{i+1}^N(t)-x_i^N(t))}\right), & t > 0, \quad i = 0, \cdots, N-1 \\ x_i^N(0) = \bar{x}_i^N, & i = 0, \cdots, N \end{cases} \tag{5}$$

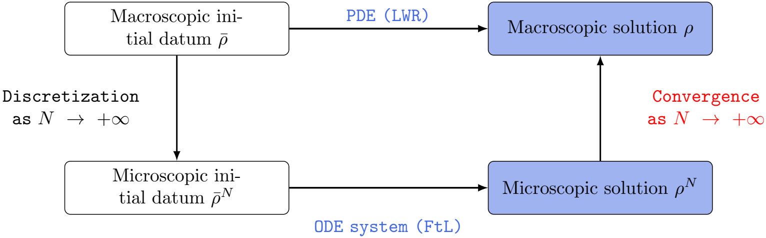
  ✓ encodes individual movements
  × requires more data

- Lighthill-William-Richards (LWR), **macroscopic** traffic flow model
  $\Rightarrow$ vehicles treated as a *continuous medium similar to particles in fluid*

$$\begin{cases} \frac{\partial}{\partial t}\rho(t,x) + \frac{\partial}{\partial x}f(\rho)(t,x) = 0, & x \in \mathbb{R}, \quad t > 0, \\ \rho(0,x) = \rho_0(x), & x \in \mathbb{R} \end{cases} \tag{6}$$

  ✓ less data-intensive
  × oversimplifies traffic phenomena

- **Convergence analysis of FtL approximation [5] scheme towards LWR model[6]**
- Solution of PDE (6) can be recovered as many particle limit[7] of ODE system (**??**)



Coupled Resolution of a Microscopic ODE System and a Macroscopic PDE

[5] Link between FtL and LWR based on atomisation of initial density ρ_0

$$\bar{x}_{i+1}^N := \sup \left\{ x \in \mathbb{R} : \int_{\bar{x}_i^N}^{x} \rho_0(y)dy = \frac{L}{N} \right\}, \quad i = 0, \cdots, N-1 \tag{7}$$

[6] Holden and Risebro 2019.
[7] Di Francesco and Rosini 2017.

- **Sparse probe data**: $n \ll N$ and only initial and final positions are collected
- **Approximate density reconstruction**[8] $\Rightarrow$ find optimal interaction parameter α

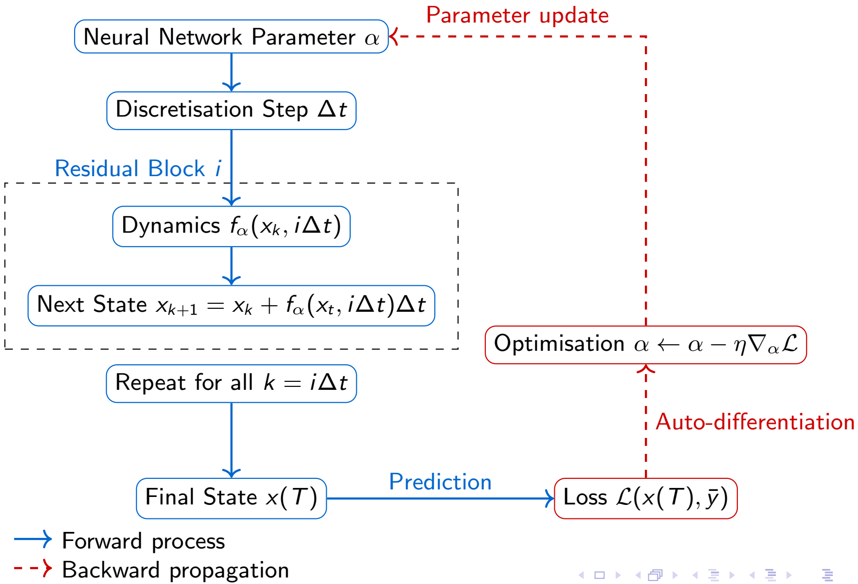
$$\begin{aligned} &\underset{\alpha}{\text{minimise}} \quad \frac{1}{2}\sum_{i=0}^{n} |x_i(T) - \bar{y}_i|^2 \\ \text{s.t.} \quad & \tfrac{d}{dt}x(t) = V\left(W_\alpha x(t) + b_\alpha(t)\right) \\ & x_i(0) = \bar{x}_i \\ & \alpha \in \mathcal{A}^N \end{aligned} \tag{8}$$

- Existence of solutions guaranteed by assumptions on $V := v \circ \frac{1}{\cdot}$ and $\mathcal{A}^N := \left\{\alpha \in \mathbb{R}^n, \quad \alpha_i \in \left[1, min\left\{\left(\bar{x}^N_{i+1} - \bar{x}^N_i\right)/L, N\left(\bar{y}^N_{i+1} - \bar{y}^N_i\right)/L\right\}\right], \quad \sum_{i=0}^{n-1} \alpha_i = N\right\}$
- No uniqueness (a priori) since nonlinear dynamics can lead to multiple minima

[8] Baloul, Hayat, Liard, and Lissy 2025.

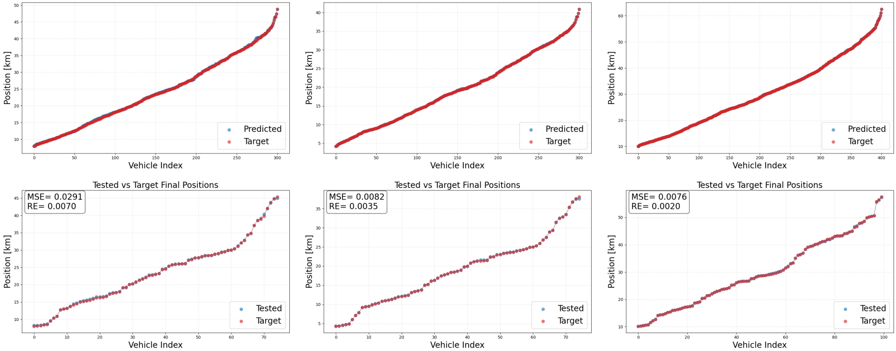
Neural Network for Constrained Optimisation

Learning Architecture



Stop-and-go wave characterized by alternating regions of congestion and free flow

Sampling such 10% **of total fleet serve as PVs for training and** 2.5% **for testing**



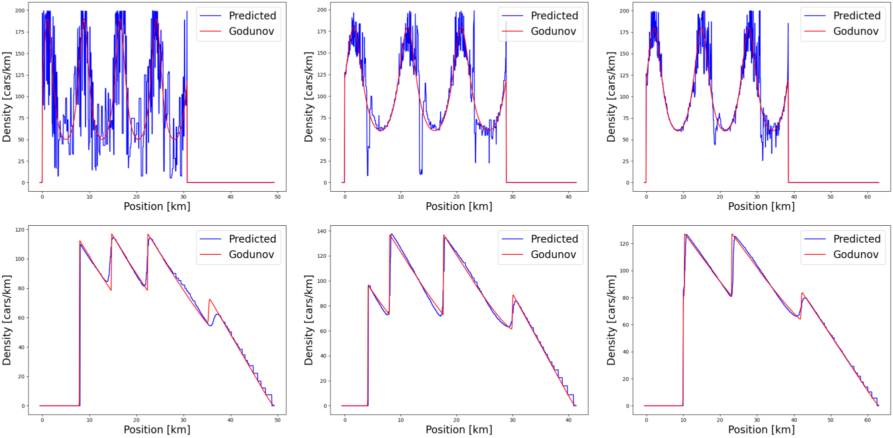
(a) $N = 2000$ (b) $N = 3000$ (c) $N = 4000$

Comparison of predicted and target final PV positions

Top Results from **training** procedure

Bottom Results on **test** sounds

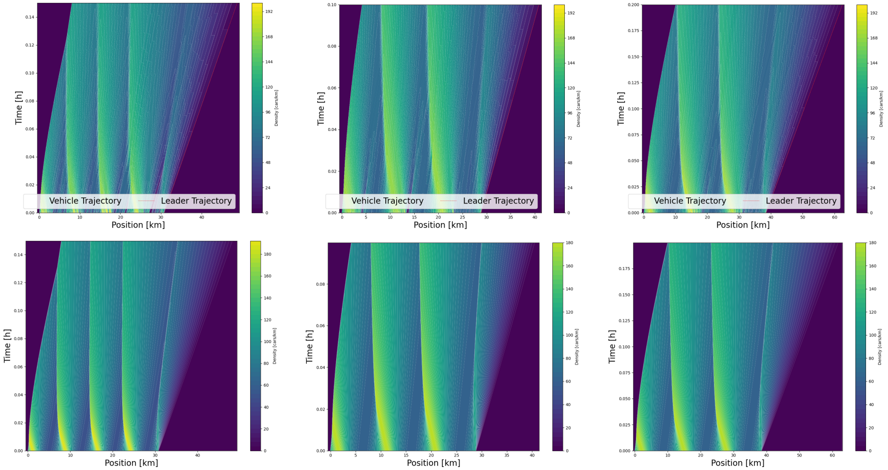
Numerical experiments



(a) $N = 2000$ (b) $N = 3000$ (c) $N = 4000$

Comparison of reconstructed and macroscopic densities
Top Initial densities
Bottom Final densities

Numerical experiments



(a) $N = 2000$ (b) $N = 3000$ (c) $N = 4000$

Comparison of **reconstructed** and **macroscopic** densities
Top Reconstructed density from learning-based optimisation
Bottom Macroscopic density from LWR PDE (Godunov scheme)

Table of Contents

- *Aedes aegypti* [a]
- Mosquito del dengue
- Mosquito momia
- Mosquito de la fiebre amarilla

REINO	Animalia
FILO	Arthropoda
CLASE	Insecta
ORDEN	Diptera
FAMILIA	Culicidae
GÉNERO	Aedes
ESPECIE	*Aedes aegypti*

[a] primary vector of dengue and also transmits Zika and chikungunya, a serious and recurrent public-health burden in Cuba, as across much of the tropics.

[a]

[a] no specific antiviral treatment and no broadly deployed vaccine: the only way to reduce transmission is to reduce the mosquito population itself.

Sterile Insect Technique (SIT)

Joint work with Luis Almeida, Jean-Michel Coron, René Gato Armas and Amaury Hayat

- **Release sterile males** into a wild mosquito population
- They compete for mating; matings with sterile males yield no viable offspring
- Used *against disease* vectors (dengue, Zika, malaria)

Four-compartment system of ODE: eggs E, females F, wild males M, sterile males M_s

$$\dot{E} = \beta_E F \left(1 - \frac{E}{K}\right) \frac{M}{M + \gamma M_s} - (\nu_E + \delta_E) E \tag{9}$$

$$\dot{F} = \nu \, \nu_E E - \delta_F F \tag{10}$$

$$\dot{M} = (1 - \nu) \, \nu_E E - \delta_M M \tag{11}$$

$$\dot{M}_s = u(t) - \delta_s M_s \tag{12}$$

Control question

Design a release rate $u(t) \geq 0$ that drives the wild population to (and keeps it at) zero, at minimal cost

Operational question

Determine *when* and *how many* sterile males to release over time

- Our goal is to learn a release policy $u(t)$ that meets simultaneously
  - **(C1)** drive the female population to near zero
  - **(C2)** stop releasing by the end of the campaign
  - **(C3)** leave no residual stock of sterile males
- Crucially, the controller **does not observe the true population** state directly
  $\Rightarrow$ only partial, periodic field measurements (traps): this a *partially observed* problem

Observation

- The controller sees the egg signal (an "ovitrap" integral)

$$E_n := \int_{T_{n-1}}^{T_n} \beta_E \, F(t) \left(1 - \frac{E(t)}{K}\right) \mathrm{d}t \tag{13}$$

  and/or an adult-trap signal $M + M_s$ (wild and sterile males lumped together)
- Sampling at their own periods, not continuously
  $\Rightarrow$ a key modelling distinction separates the *measurement* frequency T_{obs} from the *release* frequency T_{rel}

Reward

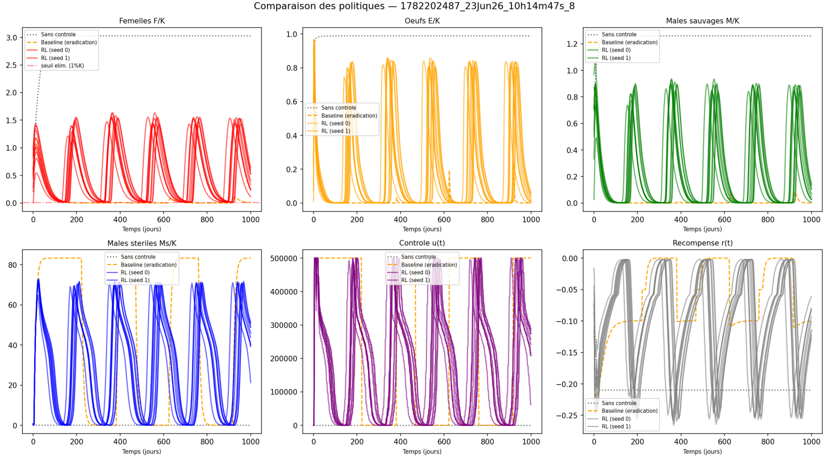
- Reward is modular: a penalty on females (C1), a penalty on residual sterile males (C3), a quadratic release cost (C2)
- Two refinements proved important: the *scale of the release cost*, and *normalizing state penalties by the endemic equilibrium* rather than by carrying capacity K.

- We train a **PPO agent** [9]
- A run produces a policy network mapping observations to a release rate
- Assess elimination rate, final female level, release cost, and time-to-elimination [10]

[9] Proximal Policy Optimisation, Stable-Baselines3 on a cluster (e.g. 20 million environment steps per run)
[10] Evaluation is done offline over 200 random initial conditions

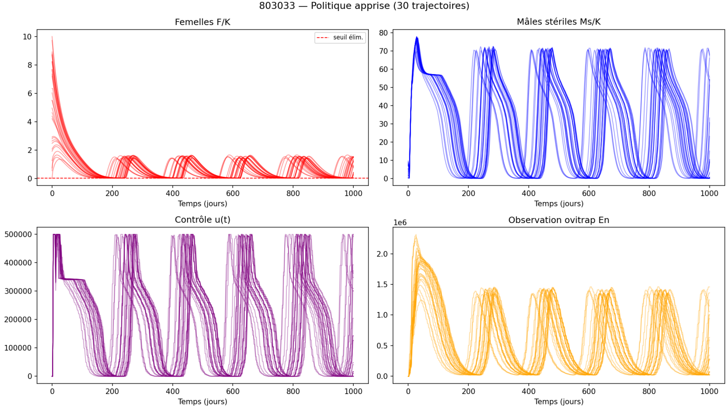
Numerical experiments

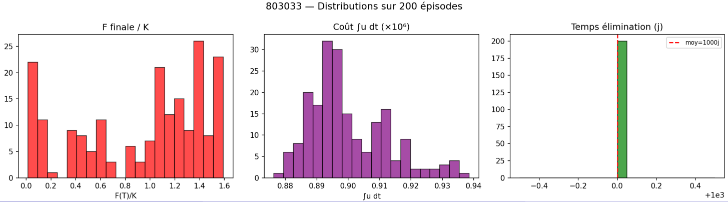
Run	$M + Ms$	Élim.	$F(T)/K$	Coût $\int u\,dt$	Récomp.	t_{elim}
S1	×	0 %	0.931	8.86×10^7	-50.32	1000



S1

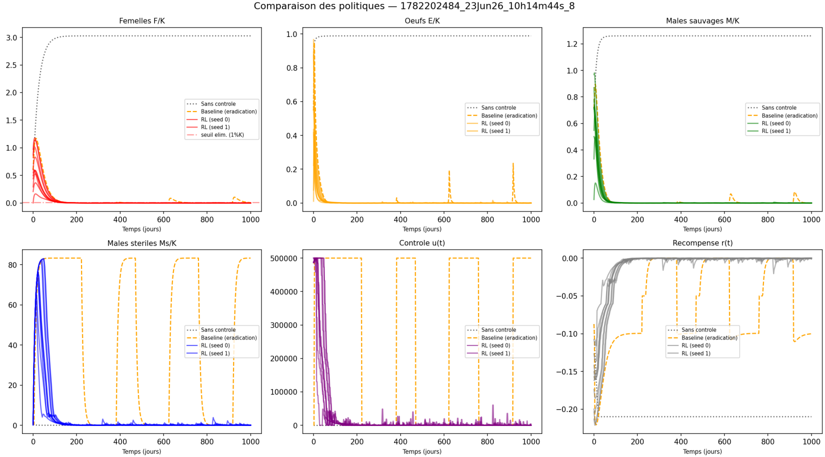
Numerical experiments





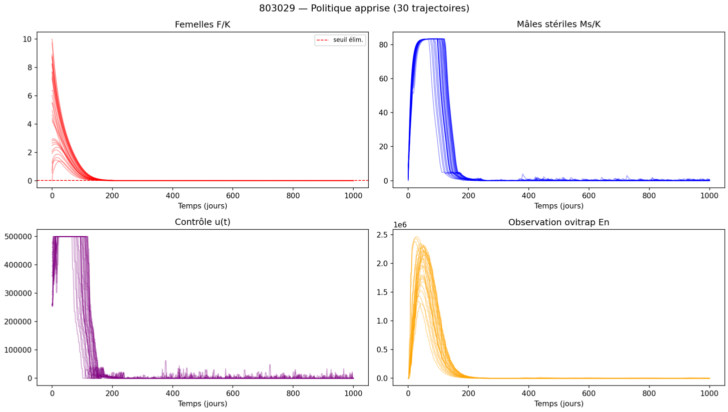
Numerical experiments

Run	$M + Ms$	Élim.	$F(T)/K$	Coût $\int u\,dt$	Récomp.	t_{elim}
S2	✓	100 %	0.001	1.11×10^7	-4.91	199



S2

Numerical experiments



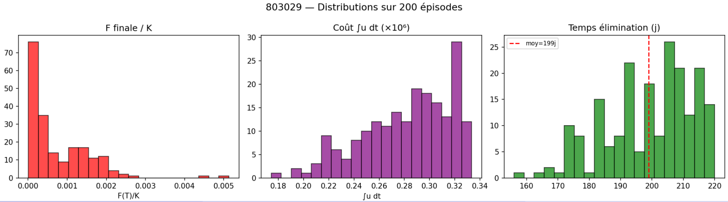
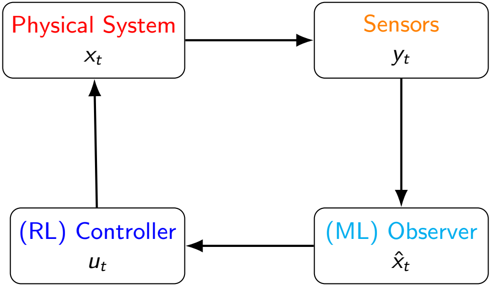


Table of Contents

Two applications, one similar problem

	Traffic	**Mosquitoes**
Hidden state	Traffic density	Population stages
Observations	Sensors / cameras	Traps / environmental data
Dynamics	Traffic flow PDE/ODE	Population dynamics
Control	Lights / speed / ramps	Interventions
Objective	*Reduce congestion/pollution*	*Reduce population/cost*

A general architecture



Stabilisation of solutions to a toll-gate traffic model

Joint work with Kala Agbo Bidi, Thibault Liard and Pierre Lissy

- Segment $[0, G]$, $G > 0$, with a *toll gate at each end* governed by

$$\begin{aligned} \partial_t \rho(t,x) + \partial_x f(\rho)(t,x) &= 0, \quad x \in (0,G), \quad t \in \mathbb{R}_+, \\ f(\rho(t,0)) &\le F_1(t,\rho), \quad t \in \mathbb{R}_+, \\ f(\rho(t,G)) &\le F_2(t,\rho), \quad t \in \mathbb{R}_+, \\ \rho(0,x) &= \rho_0(x), \quad x \in (0,G), \end{aligned} \tag{14}$$

- Choose F_1, F_2 so $\rho(t,\cdot)$ is driven (and stays) at **prescribed target profile** $w(\cdot)$
- Feedback law may need[11] $m(t) = \int_0^G (\rho(t,x) - w(x))\, dx$ at every instant
- In practice, only a handful of sensors: situation from reconstruction part
- **Goal**: replace $\rho(t,\cdot)$ in $m(t)$ by the reconstructed density $\rho^N(t,\cdot)$
  $\Rightarrow$ assess whether the *same convergence* to $w(\cdot)$ still holds
  $\Rightarrow$ deploy RL method to obtain suitable controllers

[11] Liard, Marx, and Perrollaz 2023.

- Conservation law with moving bottleneck[12] (autonomous vehicle)

$$\begin{cases} \text{LWR PDE (6) with} \\ f\left(\rho(t,y(t))\right) - \dot{y}(t)\rho(t,y(t)) \leq \frac{\alpha\rho_{\max}}{4v_{\max}}\left(v_{\max} - \dot{y}(t)\right)^2, & t>0, \\ \dot{y}(t) = \omega\left(\rho(t,y(t)_+)\right), & t>0, \\ y(0) = y_0 \end{cases} \tag{15}$$

- Network with a junction[13] J and N incoming roads and M outgoing ones

$$\begin{cases} \partial_t \rho_l(t,x) + \partial_x\left(f(\rho_l(t,x))\right) = 0, & t>0, \quad x \in I_l, \quad l = 1,\dots,N+M \\ \rho_l(0,x) = \rho_{0,l}(x), & x \in I_l = [a_l, b_l], \quad l = 1,\dots,N+M \end{cases} \tag{16}$$

$\Rightarrow \sum_{i=1}^{N} f\left(\rho_i(t,(b_i)_-)\right) = \sum_{j=N+1}^{N+M} f\left(\rho_j(t,(a_j)_+)\right)$ (Rankine Hugoniot)

$\Rightarrow \sum_{i=1}^{N} f\left(\rho_i(t,(b_i)_-)\right)$ is maximized[14] s.t. $f\left(\rho_j(\cdot,(a_j)_+)\right) = \sum_{i=1}^{N} a_{j,i} f\left(\rho_i(\cdot,(b_i)_-)\right)$

[12] Liard and Piccoli 2021.

[13] Coclite, Piccoli, and Garavello 2005.

[14] Garavello and Piccoli 2006.

Perspectives for Control of Mosquito Populations

- Only a few countries (e.g., Cuba, China, Greece, Singapore, USA) currently deploy SIT at field scale. Optimal, adaptive control of these campaigns is still to be built
- Multi-site and spatiality (control on graphs, reaction–diffusion PDE, optimal budget allocation across sites)
- Robustness and uncertainty: poorly known biological parameters, seasonality, variable sterile-male competitiveness, robust and uncertainty-aware control

References I

Baloul, N., A. Hayat, T. Liard, and P. Lissy (2025). "Traffic Flow Reconstruction from Limited Collected Data". In: *64th IEEE Conference on Decision and Control (CDC 2025).*

Coclite, G. M., B. Piccoli, and M. Garavello (2005). "Traffic Flow on a Road Network". In: *SIAM Journal on Mathematical Analysis* 36.6, pp. 1862–1886.

Colombo, R. M. and P. Goatin (2007). "A Well-Posed Conservation Law with a Variable Unilateral Constraint". In: *J. Differential Equations* 234.2, pp. 654–675.

Di Francesco, M. and M. D. Rosini (2017). "Rigorous Derivation of Nonlinear Scalar Conservation Laws from Follow-the-Leader Type Models via Many Particle Limit". In: *Arch. Rational Mech. Anal.* 225, pp. 129–166.

Garavello, M. and B. Piccoli (2006). *Traffic Flow on Networks*. Vol. 1. Applied Mathematics. American Institute of Mathematical Sciences.

Holden, H. and N. H. Risebro (2019). "The continuum limit of follow-the-leader models — A short proof". In: *SIAM J. Math. Anal.* 51.2, pp. 1771–1786.

Liard, T., S. Marx, and V. Perrollaz (2023). "Stabilization of conservation laws for G-solutions with application to traffic flow". In: *hal-04099345.*

Liard, T. and B. Piccoli (2021). "On entropic solutions to conservation laws coupled with moving bottlenecks". In: *Communications in Mathematical Sciences* 19.4, pp. 1041–1068.

Monneau, R (2024). "Structure of Riemann solvers on networks (preliminary version)". In: *hal preprint hal-04764513v1*.

Sugiyama, Yuki, Minoru Fukui, Macoto Kikuchi, Kazuhiro Hasebe, Katsuhiro Tanaka, Akihiro Tomoeda, and Satoshi Tsuchiya (2008). "Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam". In: *New Journal of Physics* 10.3, p. 033001. DOI: 10.1088/1367-2630/10/3/033001.

Learning Method

- Dataset consists of **artificial data** based on simulated (classical) FtL dynamics (5)
- Sampling of PVs yielding a **balanced representation** of overall traffic
- Neural network architecture designed to **understand dynamics of traffic**
- Residual network (ResNet) where each block corresponds to a single time step
- Input $\bar{x}$ and state $x(.)$ is propagated by mirroring Euler discretisation

$$x(t+\Delta t) = x(t) + V(Wx(t)+b)\Delta t \tag{17}$$

- Weights and biases W, b are functions of α

$$\begin{cases} W_{i,i} & := -\frac{N}{\alpha_i L}, i = 0, \ldots, n-1, \\ W_{i,i+1} & := \frac{N}{\alpha_i L}, i = 1, \ldots, n-2, \\ W_{i,j} & := 0, \text{otherwise}, \end{cases} \tag{18}$$

$$b_i(t) =:= \delta_{i,n} \frac{N}{\alpha_{n-1} L} \left(v_{\max} t + \bar{x}_n^N \right), \quad t \in [0, T] \tag{19}$$

- Nonlinear dynamic map V acts as physics grounded activation function
- Backpropagation to minimise predictions errors $\frac{1}{n} \sum_{j=0}^{n} |x_j^{\alpha}(T) - \bar{y}_j^N|^2$

Model validation

- Through predicted parameter $\bar{\alpha}$, training yields piecewise constant discrete density

$$\rho^N(t,x) = \sum_{i=0}^{n-1} \frac{\bar{\alpha}_i L}{N(x_{i+1}^N(t) - x_i^N(t))} \chi_{[x_i^N(t), x_{i+1}^N(t))}(x), \quad x \in \mathbb{R}, \quad t \in [0,T], \tag{20}$$

- Simulation on **test data** by solving ODE system

$$\begin{cases} \dot{x}_i^N(t) = v\left(\rho^N(t, x_i(t)^+)\right), & t \in (0,T], \\ x_i^N(0) = \bar{x}_i^N & i = 0, \dots, n_{\text{test}} \end{cases} \tag{21}$$

Model validation

- Through predicted parameter $\bar{\alpha}$, training yields piecewise constant discrete density

$$\rho^N(t,x) = \sum_{i=0}^{n-1} \frac{\bar{\alpha}_i L}{N(x_{i+1}^N(t) - x_i^N(t))} \chi_{[x_i^N(t), x_{i+1}^N(t))}(x), \quad x \in \mathbb{R}, \quad t \in [0,T], \tag{20}$$

- Simulation on **test data** by solving ODE system

$$\begin{cases} \dot{x}_i^N(t) = v\left(\rho^N(t, x_i(t)^+)\right), & t \in (0,T], \\ x_i^N(0) = \bar{x}_i^N & i = 0, \dots, n_{\text{test}} \end{cases} \tag{21}$$

Model validation

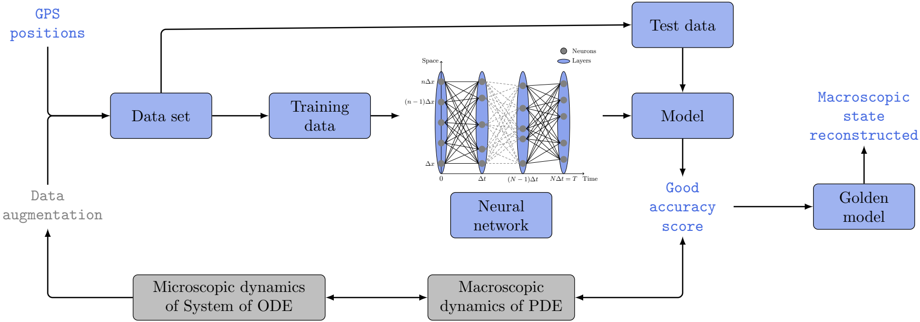
- Through predicted parameter $\bar{\alpha}$, training yields piecewise constant discrete density

$$\rho^N(t,x) = \sum_{i=0}^{n-1} \frac{\bar{\alpha}_i L}{N(x_{i+1}^N(t) - x_i^N(t))} \chi_{[x_i^N(t), x_{i+1}^N(t))}(x), \quad x \in \mathbb{R}, \quad t \in [0,T], \tag{20}$$

- Simulation on **test data** by solving ODE system

$$\begin{cases} \dot{x}_i^N(t) = v\left(\rho^N(t, x_i(t)^+)\right), & t \in (0,T], \\ x_i^N(0) = \bar{x}_i^N & i = 0, \dots, n_{\text{test}} \end{cases} \tag{21}$$

- Assess model's performance by measuring test error $\frac{1}{n_{\text{test}}} \sum_{j=0}^{n_{\text{test}}} |x_j(T) - \bar{y}_i^N|^2$



Traffic Flow Reconstruction Pipeline

Outline of Convergence Analysis

- Prove that, **if only using data from dynamical systems**, approximate density ρ^N predicted by our machine learning model converges to solution of the LWR model (6) when the number of vehicles approaches infinity
- Main challenge lies in imposing a condition on distribution of α which would guarantee convergence
- Demonstrate that discrete initial density $\rho^N(0,\cdot)$ converges to the initial condition ρ_0 in the LWR model (6) under this additional assumption
- In addition to Euler discrete density (**??**), consider empirical discrete density

$$\widehat{\rho}^N(t,\cdot) := \frac{L}{N}\sum_{i=0}^{n-1} \alpha_i^N \delta_{x_i(t)}(\cdot), \quad t \in [0,T]. \tag{22}$$

- **Important observation**: by construction, initial traffic density must satisfy

$$\bar{x}_{i+1} = \sup\left\{x \in \mathbb{R} : \int_{\bar{x}_i}^{x} \rho_0(y)dy \leq \frac{\alpha_i L}{N}\right\}, \quad i = 0, \dots, n-1 \tag{23}$$

$\Rightarrow$ although no access to ground-truth initial car density ρ_0, initial positions $\bar{x}_i$ of probe vehicles verify (23) (i.e. number of unobserved vehicles between $[x_i, x_{i+1})$ is given by α_i)

- Weak solution of (6) is entropy admissible if it satisfies **Kruzhkov entropy condition**

$$\int_0^T \int_{\mathbb{R}} |u-k| \frac{\partial \phi}{\partial t} + \operatorname{sign}(u-k)(f(u)-f(k)) \frac{\partial \phi}{\partial x} dxdt \geq 0, \quad \forall k \in \mathbb{R} \tag{24}$$

Theorem Convergence of approximate density to solution of LWR

Under some assumptions, piecewise-constant density

$$\rho^N(t,x) = \sum_{i=0}^{n-1} \frac{\bar{\alpha}_i^N L}{N(x_{i+1}^N(t) - x_i^N(t))} \chi_{[x_i^N(t), x_{i+1}^N(t))}(x), \quad x \in \mathbb{R}, \quad t \in [0,T], \tag{25}$$

where $\bar{\alpha}_i^N \in \mathcal{A}^N$ is a solution to (8) converges to **unique entropy** solution ρ of

$$\begin{aligned} &\frac{\partial \rho}{\partial t}(t,x) + \frac{\partial f(\rho)}{\partial x}(t,x) = 0, \quad x \in \mathbb{R}, \quad t \in [0,T], \\ &\rho(0,x) = \rho_0(x), \quad x \in \mathbb{R} \end{aligned} \tag{26}$$

- Impose a condition that ensures controlled growth of α_N

- Ensures discretisation aligns consistently with true initial density when $N \to \infty$
- Notations: for $t \in [0, T]$, $\rho(t) := \rho(t, \cdot)$ and $\widehat{\rho}(t) := \widehat{\rho}(t, \cdot)$.
  In particular, at $t = 0$, $\rho(0) := \rho(0, \cdot)$ and $\widehat{\rho}(0) := \widehat{\rho}(0, \cdot)$
- Use **Wasserstein distance** defined in Di Francesco and Rosini 2017 by

$$\mathrm{W}_{L,1}(f, g) = \| f(]-\infty, \cdot]) - g(]-\infty, \cdot]\|_{L^1(\mathbb{R},\mathbb{R})} \tag{27}$$

Proposition

Let ρ_0 satisfy (23) and assume that

$$\max_{i=0,\dots,n-1} \alpha_i^N = o(N) \tag{28}$$

Then, both sequences $(\rho^N(0))_{n\in\mathbb{N}}$ and $(\widehat{\rho}^N(0))_{n\in\mathbb{N}}$ converge to ρ_0 in the sense of the $\mathrm{W}_{L,1}$ -Wasserstein distance in (27)

Remark

A particular case of assumption (28) is when $\max_{i=0,\dots,n-1} \alpha_i^N \leq \frac{CN}{\log(N)}$ for some $C > 0$

- Using $l_N := L/N$, $\mathrm{W}_{L,1}$- distance and discrete density in (??)

$$\mathrm{W}_{L,1}(\rho^N(0), \widehat{\rho}^N(0)) = \sum_{i=0}^{n-1} \int_{\bar{x}_i}^{\bar{x}_{i+1}^N} \left(\alpha_i^N l_N - \rho_i^N(t)\,(x - \bar{x}_i) \right) dx$$
$$= \sum_{i=0}^{n-1} \alpha_i^N l_N \int_{\bar{x}_i}^{\bar{x}_{i+1}} \left(1 - \frac{x - \bar{x}_i}{\bar{x}_{i+1} - \bar{x}_i} \right) dx \qquad (29)$$
$$\leq \max_{i=0,\dots,n} \{\alpha_i^N\} l_N \,(\bar{x}_n - \bar{x}_0)$$

$\Rightarrow$ it suffices to prove that $(\widehat{\rho}^N(0))_{n\in\mathbb{N}}$ converges to ρ_0 wrt $\mathrm{W}_{L,1}$- distance

- Using expressions of both Euler (??) and empirical (22) discrete densities

$$\mathrm{W}_{L,1}(\widehat{\rho}^N(0), \rho_0)$$
$$= \sum_{i=0}^{n-2} \int_{\bar{x}_i}^{\bar{x}_{i+1}} \left(\sum_{j=0}^{i} \alpha_j^N l_N - \int_{-\infty}^{x} \rho_0(y)dy \right) dx + \int_{\bar{x}_{n-1}}^{\bar{x}_n} \left(L - \int_{-\infty}^{x} \rho_0(y)dy \right) dx$$
$$= \sum_{i=0}^{n-2} \int_{\bar{x}_i}^{\bar{x}_{i+1}} \left(\left(\sum_{j=0}^{i-1} \alpha_j^N l_N - \int_{\bar{x}_0}^{\bar{x}_i} \rho_0(y)dy \right) + \left(\alpha_i^N l_N - \int_{\bar{x}_i}^{x} \rho_0(y)dy \right) \right) dx$$
$$+ \int_{\bar{x}_{n-1}}^{\bar{x}_n} \left(\sum_{j=0}^{n-2} \alpha_j^N l_N - \int_{\bar{x}_0}^{\bar{x}_{n-1}} \rho_0(y)dy \right) + \left(\alpha_{n-1}^N l_N - \int_{\bar{x}_{n-1}}^{x} \rho_0(y)dy \right) dx$$

Sketch of proof

- From atomisation of initial density (23), deduce

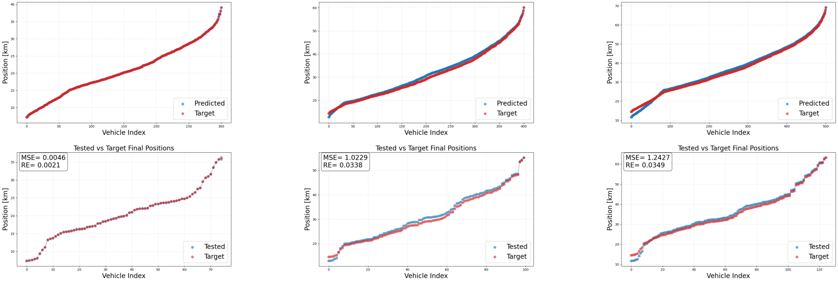
$$
\begin{aligned}
& \mathrm{W}_{L,1}(\widehat{\rho}^N(0), \rho_0) \\
& \leq \sum_{i=0}^{n-2} \int_{\bar{x}_i}^{\bar{x}_{i+1}} \left(\alpha_i^N l_N - \int_{\bar{x}_i}^{x} \rho_0(y) dy \right) dx + \int_{\bar{x}_{n-1}}^{\bar{x}_n} \left(\alpha_{n-1}^N l_N - \int_{\bar{x}_{n-1}}^{x} \rho_0(y) dy \right) dx \\
& = \sum_{i=0}^{n-1} \alpha_i^N l_N \int_{\bar{x}_i}^{\bar{x}_{i+1}} \left(1 - \frac{1}{\alpha_i^N l_N} \int_{\bar{x}_i}^{x} \rho_0(y) dy \right) dx \\
& \leq \max_{i=0,\dots,n} \left\{ \alpha_i^N \right\} l_N (\bar{x}_n - \bar{x}_0)
\end{aligned}
$$

- From assumption (28) and estimate (29), conclude that $(\rho^N(0))_{n\in\mathbb{N}}$ converges to ρ_0 in sense of $\mathrm{W}_{L,1}$-Wasserstein distance
- Moreover, by leveraging expression of discrete density (**??**), generalize convergence to unique entropy solution of conservation law (6), referring to Di Francesco and Rosini 2017, Theorem 3 $\Rightarrow$ require only minor modifications to original arguments

Numerical experiments

- Parameters
  - Maximum traffic speed $v_{\max} = 120$ km/h
  - Maximum traffic density $\rho_{\max} = 200$ cars/km
  - Greenshields velocity $v(\rho) = v_{\max} \max\left\{1 - \frac{\rho}{\rho_{max}}, 0\right\}, \quad \rho \in [0, \rho_{max}]$
  - Final time horizon $T = 0.1$ h or $T = 0.2$
- Sampling such 10% **of total fleet serve as PVs for training and** 2.5% **for testing**
- Two traffic scenarii modelled
  1. Shock wave represents an abrupt transition in traffic conditions
  2. Stop-and-go wave characterized by alternating regions of congestion and free flow

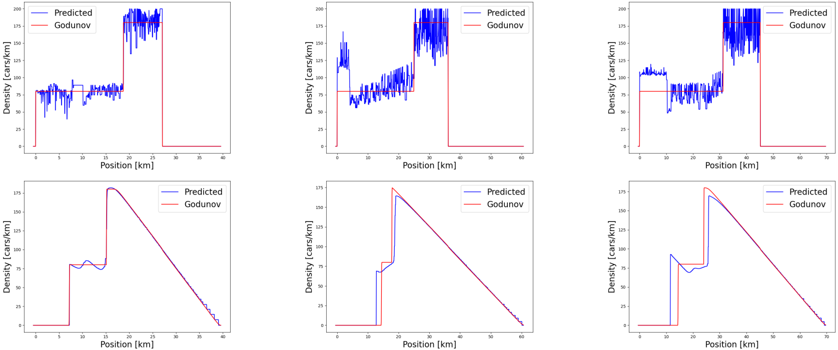
Shock wave scenario



(a) $N = 3000$ (b) $N = 4000$ (c) $N = 5000$

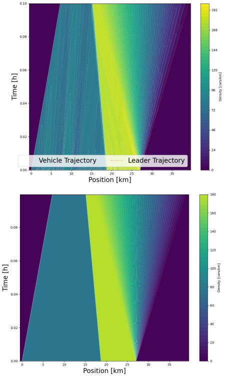
Comparison of predicted and target final PV positions
Top Results from **training** procedure
Bottom Results on **test** sounds

Shock wave scenario

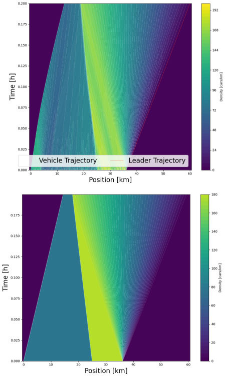


(a) $N = 3000$ (b) $N = 4000$ (c) $N = 5000$

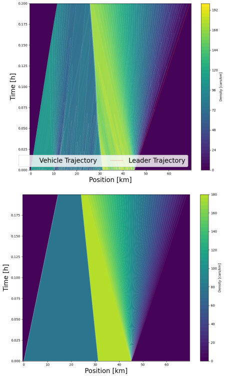
Comparison of reconstructed and macroscopic densities
Top Initial densities
Bottom Final densities



(a) $N = 3000$



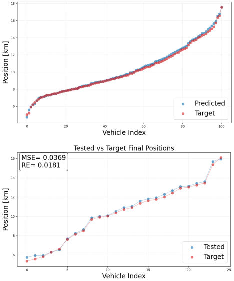
(b) $N = 4000$



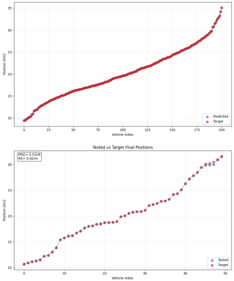
(c) $N = 5000$

Comparison of **reconstructed** and **macroscopic** densities
Top Reconstructed density from learning-based optimisation
Bottom Macroscopic density from LWR PDE (Godunov scheme)

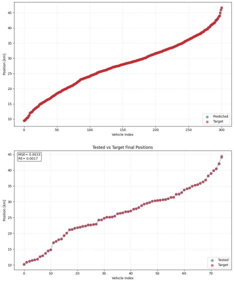
Rarefaction wave scenario



(a) $N = 1000$



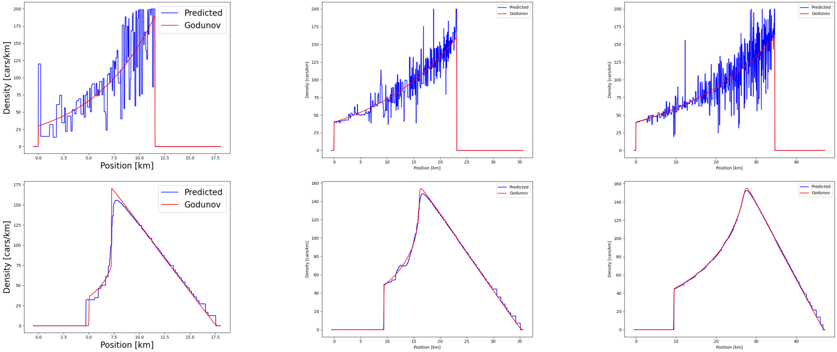
(b) $N = 2000$



(c) $N = 3000$

Comparison of predicted and target final PV positions
Top Results from **training** procedure
Bottom Results on **test** sounds

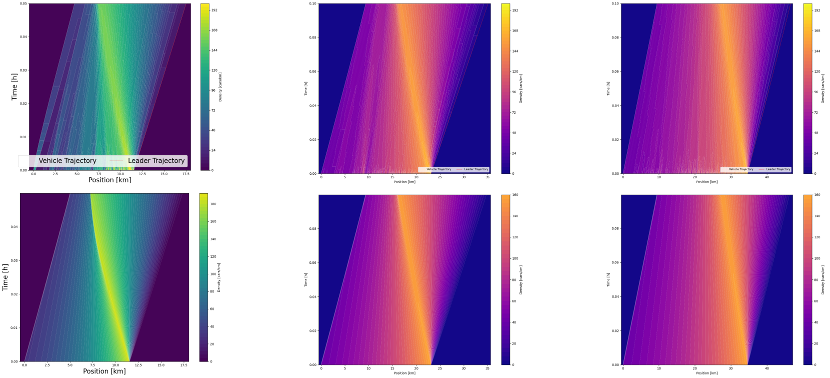
Rarefaction wave scenario



(a) $N = 1000$ (b) $N = 2000$ (c) $N = 3000$

Comparison of reconstructed and macroscopic densities
Top Initial densities
Bottom Final densities

Rarefaction wave scenario



(a) $N = 1000$ (b) $N = 2000$ (c) $N = 3000$

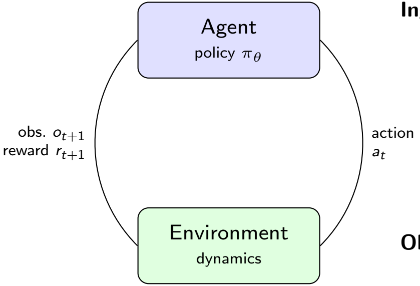
Comparison of **reconstructed** and **macroscopic** densities
Top Reconstructed density from learning-based optimisation
Bottom Macroscopic density from LWR PDE (Godunov scheme)

- **The adult-trap signal is necessary**: (as it stands) observing the sterile-male stock is not a refinement, it is a condition for control
- **Control needs a "target" signal and a "lever" signal**: no single signal suffices: the agent needs both information about the *target* (the female population, via F or its egg proxy E) *and* about the *lever* (the sterile-male stock, via $M + M_s$). [15]
- **The information bottleneck is near extinction**: the agent drives females very low but cannot *finish* [16] because, blind to the sterile-male stock, it cannot fine-tune the terminal phase; this is a genuine observability limitation

[15] notably, a direct female signal F is a *more informative* target than eggs E (which are only an upstream precursor), though neither suffices alone

[16] it stabilizes just above the elimination threshold

Reinforcement Learning: structure



The agent acts, the environment returns an observation and a reward; the loop repeats.

Ingredients

- **State** s_t — true system state (E, M, F, M_s)
- **Observation** o_t — what is measured (traps); partial
- **Action** a_t — the decision (release rate u)
- **Reward** r_t — scalar score (eliminate F, limit cost)
- **Policy** $\pi_\theta(a \mid o)$ — learned map obs. $\rightarrow$ action

Objective

$$\max_\theta \ \mathbb{E}\left[\sum_{t\geq 0} \gamma^t \, r_t\right]$$

maximise the expected discounted sum of rewards ($\gamma \in (0, 1)$: weights the future).

How it learns (PPO). No model is assumed: the agent *samples* trajectories by interacting, estimates which actions led to higher return, and nudges π_θ that way — repeatedly, over millions of steps. Here the observation is *partial* (traps only), so the policy maps measurements, not the true state, to releases.