



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB



**ICDSAAI 2026**

International Conference on  
**Data Science, Agents  
& AI**

26 & 28 March, 2026



# From GA to RL: Data-Driven Traffic Light Optimization

4th International Conference on Data Science, Agents, and Artificial Intelligence  
ICDSAAI 2026

**Dr. Luciano Alessandro Ipsaro Palesi**

Keynote Speaker

**lucianoalessandro.ipsaropalesi@unifi.it**

Department of Information Engineering (DINFO)

DISIT lab, <https://www.disit.org>, <https://www.snap4city.org>

University of Florence



# Why Traffic Light Optimization Is No Longer a Local Problem

## Network-Wide Congestion

Traffic dynamics span corridors and urban areas, not isolated junctions

## Priority Coordination

Tram and BRT priority must coexist with private traffic flow

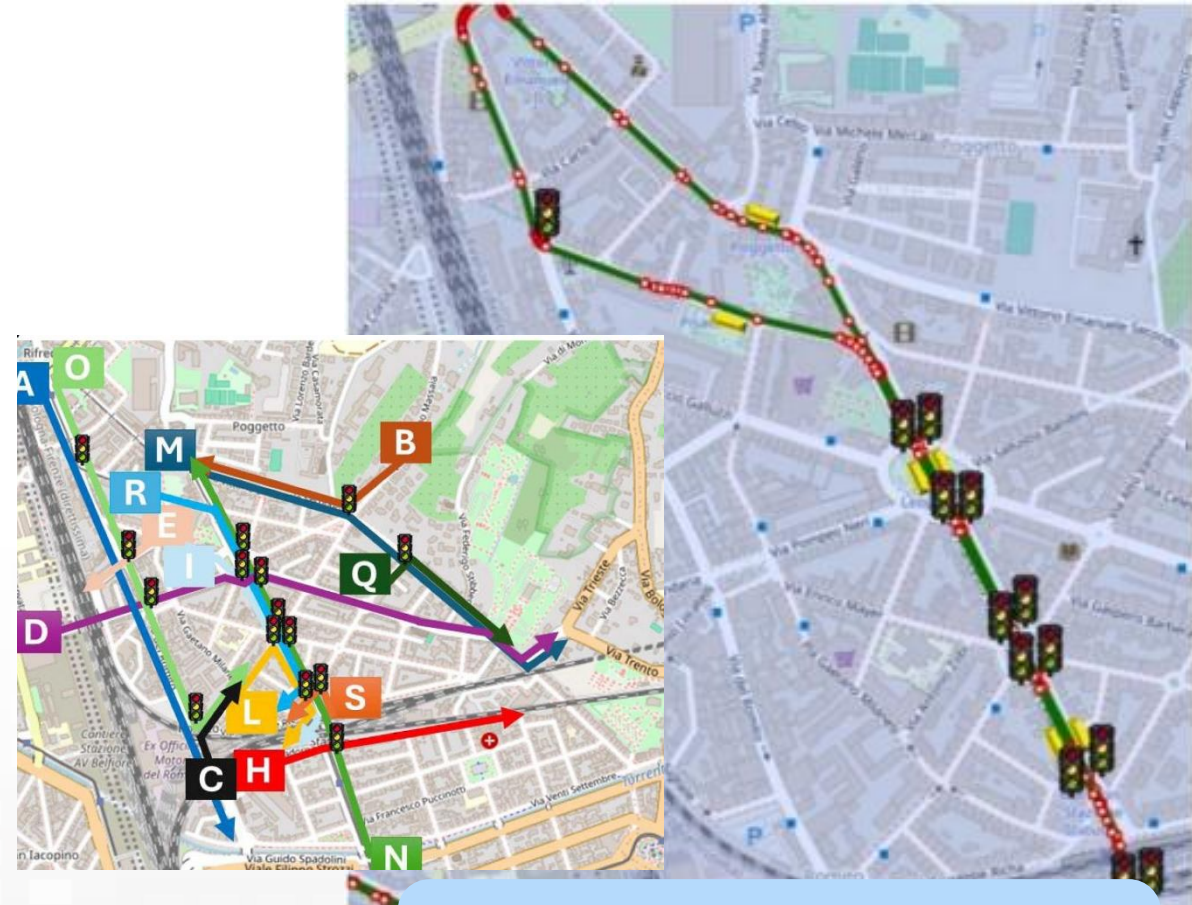
## Traditional Limits

Fixed-cycle and local actuated methods have limited area-wide effectiveness

## System-Level Control

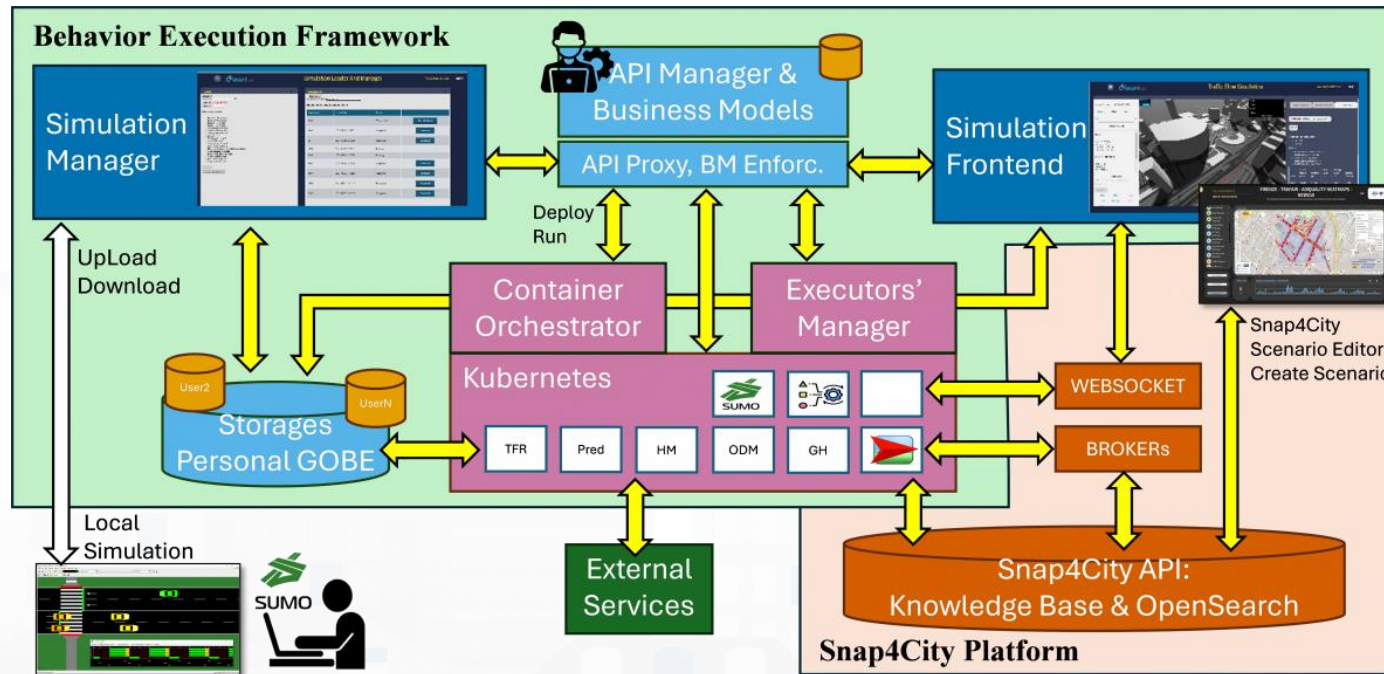
Urban optimization requires coordinated, scalable control strategies

The challenge is no longer how to optimize one intersection, but how to coordinate a whole urban traffic system under competing objectives.



**Real-world case:** The Florence urban area — 19 signalized intersections, active tram lines, a major train station, university campus, and hospital — all monitored via city-wide traffic platforms.

# The Executable Digital Twin: Snap4City as Backbone, BEFDIT as Execution Layer



- Snap4City as urban data backbone
- Real data, scenarios, KPIs, and digital twin context
- BEFDIT as execution layer for co-simulation
- Foundation for optimization and adaptive traffic control

# Problem Formulation

## Inputs

- Road network topology
- Signal phase structures
- Observed traffic flows
- Tram schedules & arrivals

## Objectives

- Minimize travel time
- Minimize waiting time
- Reduce stop frequency
- Guarantee tram priority

## Outputs

- Optimized timing plans (*GA*)
- Adaptive control policies (*RL*)

## Constraints

- Safety (conflicting phases)
- Queue length limits
- Cross-intersection coordination
- Operational phase sequencing

This is inherently a **multi-objective, dynamic problem** — not a simple scalar optimization. The interplay between tram priority, network coordination, and time-varying demand makes it a rich testbed for both search-based and learned control strategies.

$S = \{Roads, Intersections, TrafficLights, TramWay, TrafficInformation, From.DateTime, To.DateTime\}$

where:

- *Roads* =  $\{r_i : i = 1, 2, \dots, n\}$  are the roads (road graph all details) with their characteristics in terms of length, number of lanes, starting and ending nodes (respectively, denoted by  $L, N, node_I, node_F$ ). Then,  $r_i = r_i(L_i, N_i, node_{I_i}, node_{F_i})$
- *Intersections* =  $\{nod_{e_j} : j = 1, 2, \dots, k\}$  are the intersections having constraints in terms of incoming roads, outgoing roads and related turning percentage (respectively, denoted by *roadInput*, *roadOutput*, *P*). Then,  $nod_{e_j} = nod_{e_j}(roadInput, roadOutput, P)$ .
- *TrafficLights* =  $\{sig_j : j = 1, 2, \dots, k\}$  are the positions and features of traffic lights at junctions. More precisely, *sig<sub>j</sub>* may contain 0 or 1 traffic light signal. When *sig<sub>j</sub>* has 1 traffic light signal, it may have different phases that are composed by states as described above. The states are related to green/yellow and red time durations for each specific direction.
- *TramWay* =  $\{tr_i : i = 1, 2, \dots, h\}$  is the description of tramway paths/lines passing in a specific intersection *j* with a given time interval.
- *TrafficInformation*,  $TI_{\bar{\Delta}t} = \{(F_i(\bar{\Delta}t), v(\bar{\Delta}t)_i) \quad \forall i \in roads\}$  are the values of traffic flow for each entering traffic flow road segment of the scenario in terms of [# vehicles/h], and mean vehicular speed *v* [km/h] on inflow road in the scenario at specific time interval time (1-hour)( $\bar{\Delta}t$ ).
- *From DateTime to ToDateTime*, that is, the interval time from the start time to the end time in which the scenario *S* is performed.

# Families of Signal Control

## Fixed-Cycle Systems

- Pre-programmed timing sequences
- Consistent cycle lengths throughout day
- Time-of-day based adjustments only
- Minimal hardware requirements

## Actuated Systems

- Sensors detect vehicles and queue conditions
- Local controller adjusts phases in real time
- Reactive to current traffic demand
- Limited network-wide coordination

## Adjusted Systems

- Priority-based green when tram approaches
- Real-time phase adjustments on top of planned cycles
- More adaptive than fixed-cycle
- Not a fully learned policy

# State of the Art: Existing Methods



## Webster Formulas

Analytical baseline for green split computation based on traffic demand ratios



## NSGA-II Methods

Multi-objective evolutionary search applied to limited intersection settings



## Mixed-Integer Models

Network-flow formulations addressing signal coordination across junctions



## Tram Priority Systems

Exist but typically operate on small corridors with few junctions

The literature review reveals a fragmented landscape: some works optimize two objectives on a few intersections, others excel at tram synchronization but on limited cases, while others remain at micro or meso scales. The clear gap is **macroscopic coordination with tram priority, multiple simultaneous objectives, and realistic structural constraints on queues and road capacity** at urban network scale.

# Unresolved Challenges in RL-Based Traffic Control

## Limited Public Transport Priority

Few approaches explicitly incorporate transit vehicle priority into reward structures

## Scalability Questions

Weak empirical evidence on performance at large-scale urban network dimensions

## Fairness Concerns

Often no explicit mechanisms ensuring equity across competing traffic directions

## Problem-Dependent Design

Reward and state formulations remain highly customized, limiting generalizability

## Deployment Trade-offs

Real-time adaptivity requires hardware resources and computational costs affecting practical transferability

# MaMoTLO Solution Families

MaMoTLO organizes its solutions into three base families, each extended by behavioral variants — yielding **9 new solutions** compared against 5 state-of-the-art baselines.

## Behavioral Variants

### P — Penalty

Encodes tram priority directly into the optimization objective via a penalty term.

### A — Adjust on Demand

Dynamically adapts green/red phases when a tram is detected approaching.

### P-A — Penalty + Adjust

Combines structured offline penalty with real-time local adaptation.

## Structural Constraints

Constraints on queues and flows prevent unrealistic solutions

## Penalty Constraint

Protects tram passage by penalizing competing directions when needed

## Post-Synchronization

Adds coordination across different junctions

## Base Families

### NTNS-MWD

Optimizes number of stops and waiting delay — no directional priority bias.

### 2TWD-NTNS-MWD

Extends the problem to two priority traffic directions.

### 4TWD-NTNS-MWD

Adds four priority directions for richer, more balanced network control.

# Before Optimization: Reconstructing the Traffic State



## Sparse Sensing

Raw sensor data is insufficient alone



## Traffic Flow Reconstruction

Dense network-wide traffic state derived from data



## Scenario Workloads

Simulation inputs derived from reconstructed traffic



## Optimization Start

Control based on realistic reconstruction, not assumptions

Traffic signal optimization begins not with algorithms, but with quality of traffic-state representation. Traffic Flow Reconstruction derives dense network states and typical time trends for consistent simulation workloads.

## Data-Driven Foundation

Traffic Flow Reconstruction provides realistic network-wide state for optimization

# Multi-Objective Structure: What the GA Optimizes

01

## Number of Stops

Minimizing vehicle stops across network

02

## Mean Waiting Delay

Average delay experienced by vehicles

03

## Directional Waiting Delay

Waiting time on prioritized tram corridors

04

## Synchronization

Coordinated timing across junctions

05

## Tram-Aware Constraints

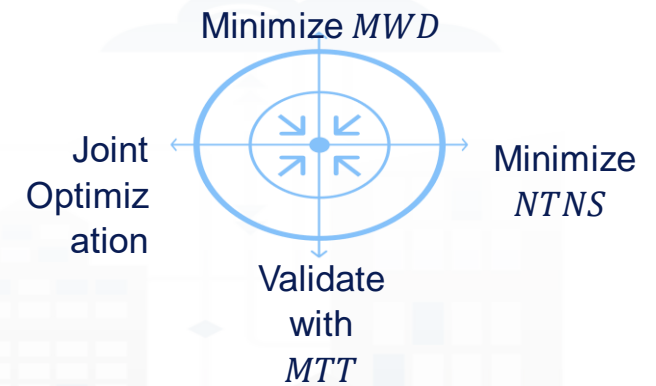
Protection mechanisms for tram priority

## Macroscopic Multi-Objective Traffic Light Optimization

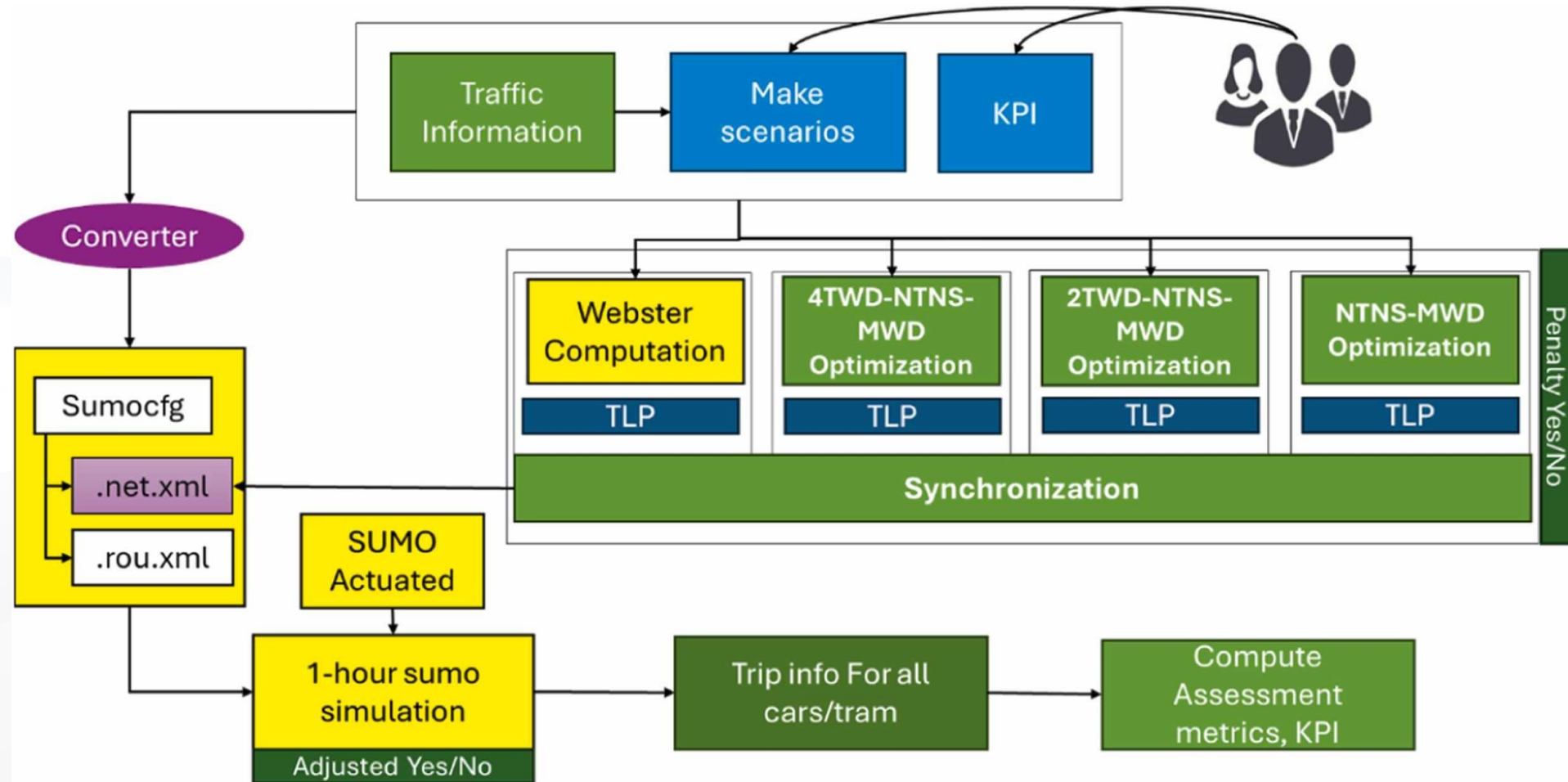
Framework designed for urban scenarios where trams receive priority without sacrificing private traffic quality

## Domain-Informed Model

Not generic optimizer applied to urban problem, but model built with traffic engineering knowledge



# Integrated data flows



# GA Strengths: Advantages/Limitations

## Interpretable Multi-Objective Trade-offs

Explicit objectives and constraints make optimization decisions transparent

## Robust Performance Across Traffic Loads

Consistent results regardless of varying traffic conditions and demand

## Fairness Across Directions

Prevents severe imbalance between competing traffic flows and directions

## Strong Real Urban Performance

Selected solutions outperform reference methods by 10%+ in real-world conditions

**Key Insight:** The GA block is already scientifically strong and operationally convincing—precisely why the move to RL must be carefully justified, not assumed.

## Plan-Oriented Optimization

Produces optimized plans for scenarios rather than learning direct policies

## Adaptation Added, Not Learned

Adjustments are extensions of plans, not learned sequential control

## Scenarios vs. Continuous Feedback

Strong on scenario optimization, weaker on continuous environment interaction

## Scaling Control Dynamics

Scaling search is easier than scaling distributed, real-time control dynamics

**Key Insight:** When traffic deviates from expected workload or distributed control with fast local reaction is needed, the problem changes nature. This naturally leads to: if GA structures the problem correctly, what is the right way to make it adaptive?

# From MaMoTLO to RL: A Natural Evolution

## GA Planning Foundation

Multi-objective problem structuring with interpretability, fairness guarantees, and robust plans

## RL Policy Learning

Sequential adaptation, decentralized control, and dynamic responsiveness to traffic conditions

## Digital Twin Requirement

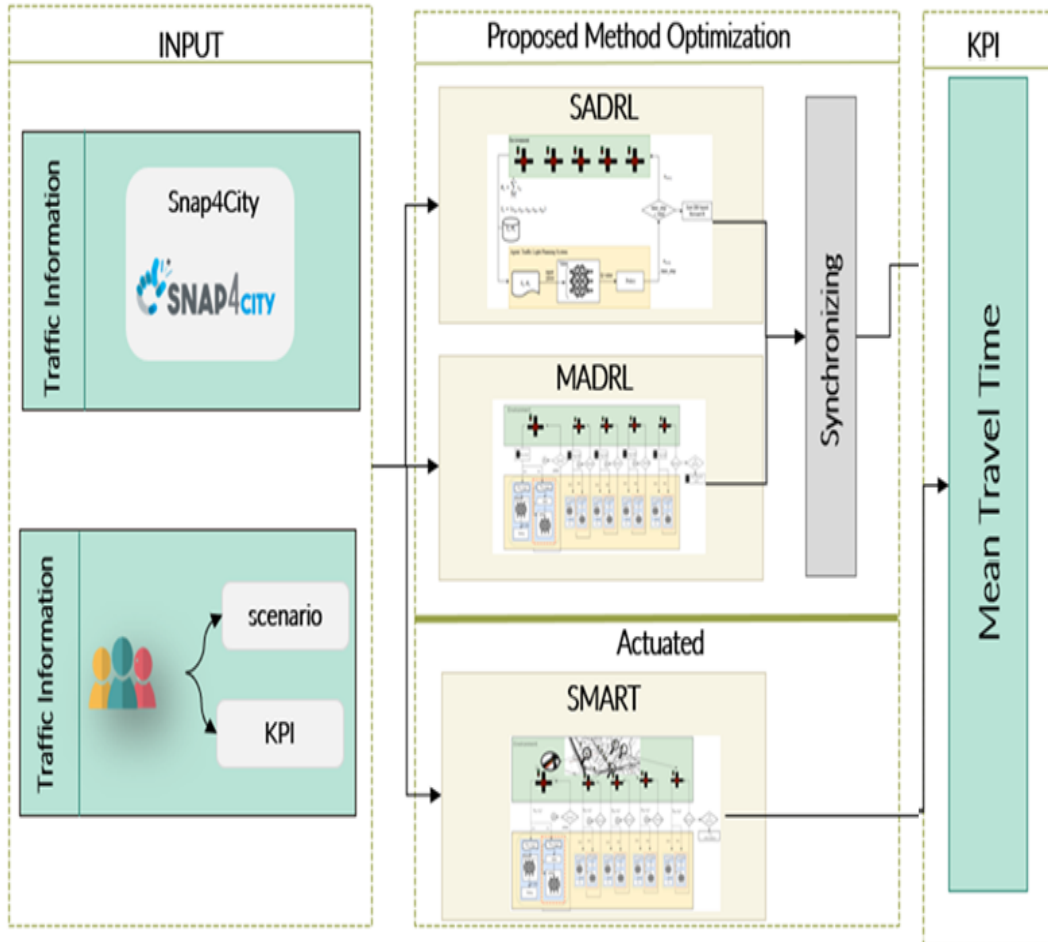
Simulation environment essential for policy learning and validation before deployment

## Same Problem, New Paradigm

RL extends rather than replaces GA. The GA teaches how to structure objectives, constraints, priorities, and synchronization properly. RL takes that same problem and reformulates it as sequential decision-making.



# RL Design Space: Three Approaches



SADRL

**Single-Agent DRL:** Centralized fixed-cycle learning. One agent optimizes globally.



MADRL

**Multi-Agent DRL:** Distributed fixed-cycle learning. Multiple agents learn in parallel.



SMART

**Actuated Real-Time Control:** Each agent modifies green time via SUMO/TraCI interaction.

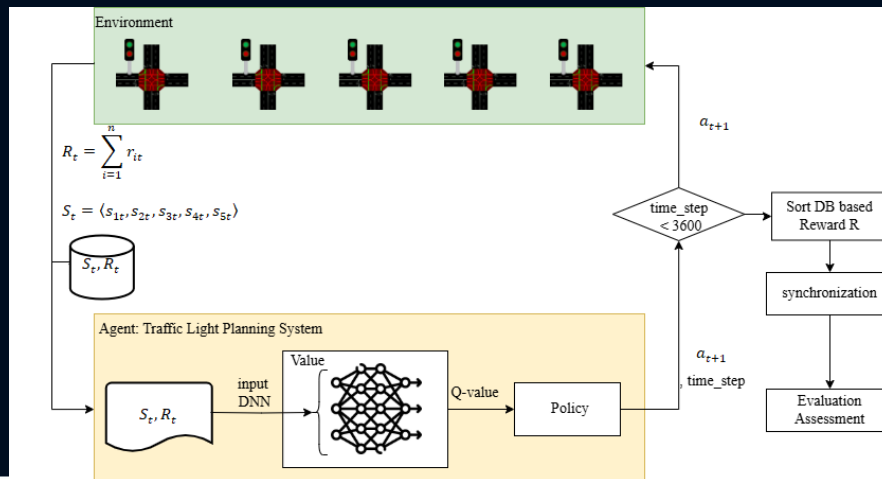
DQN and PPO explored across traffic loads—algorithm choice matters depending on load conditions.

# SADRL: Single-Agent Deep Reinforcement Learning

## Architecture

A single agent controls the entire traffic network, representing the full joint state and issuing a unified action over all regulated intersections.

In the fixed-cycle variant, the cycle duration (e.g., 60 seconds) is predetermined, and the agent's task is to optimally distribute green time across phases — maintaining compatibility with existing infrastructure.



## Trade-offs at a Glance

**Centralized Coordination**

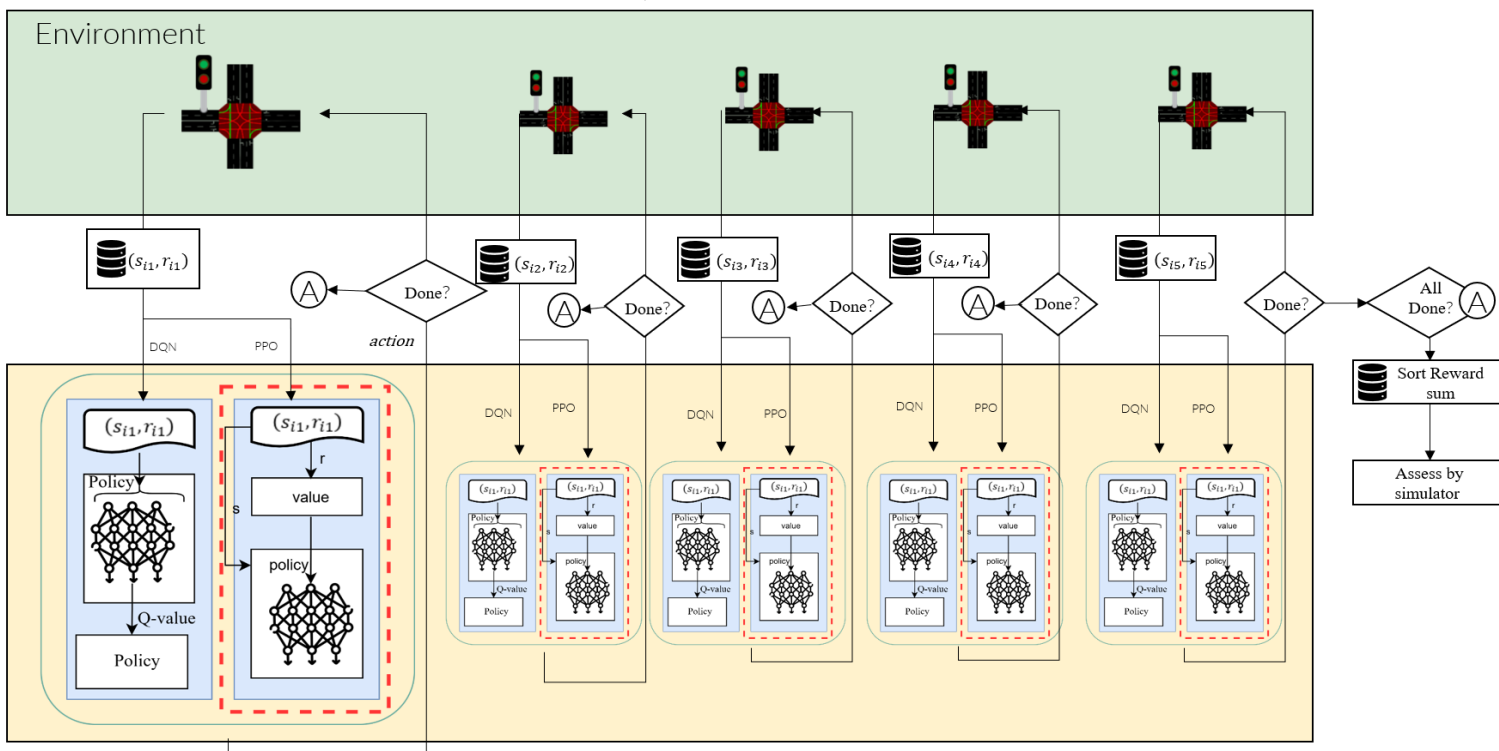
Global visibility enables theoretically optimal system-wide decisions

**Scalability Limits**

State and action spaces grow sharply with intersection count

**Training Complexity**

High-dimensional joint spaces make convergence slower and computationally expensive



# MADRL: Multi-Agent Deep Reinforcement Learning

Rather than a single centralized controller, MADRL assigns an independent agent to each intersection or local cluster. Each agent observes **local traffic conditions** and reacts rapidly, producing a naturally scalable architecture well-suited to complex urban networks.



## Distributed Observation

Each agent perceives only its local environment — queue lengths, speeds, and signal state — enabling fast, reactive decisions without global communication overhead.

## Improved Scalability

Adding intersections adds agents rather than expanding a monolithic state space. PPO-based MADRL configurations show strong adaptability, especially under **heavy traffic loads**.



## Coordination Challenge

Purely local optimization can produce globally suboptimal behavior. Agents that ignore neighboring conditions risk creating bottlenecks at the network level.

# SMART: Real-Time Adaptive Signal Control

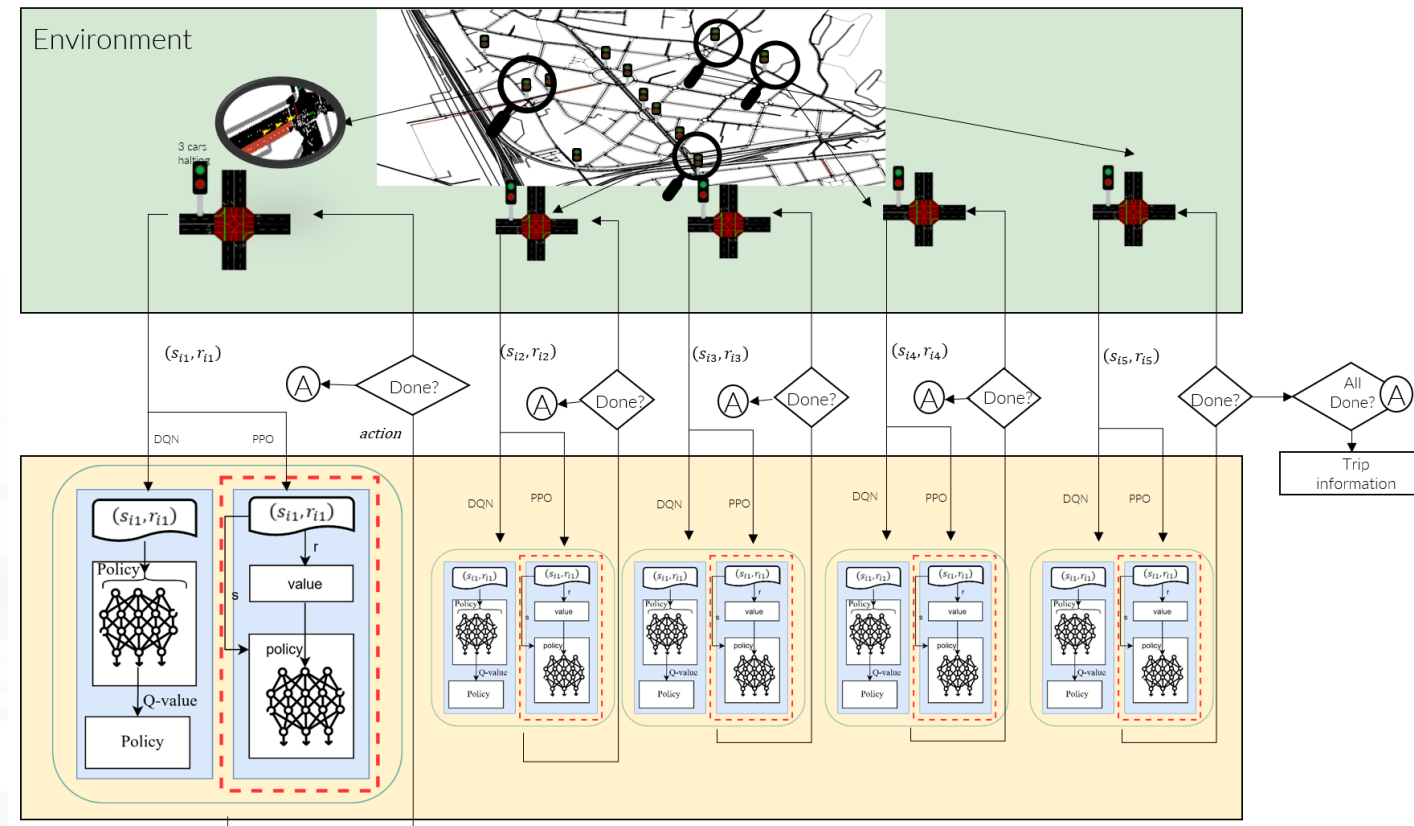
## What Makes SMART Different

SMART is the most dynamic system in the RL paper benchmark. Rather than distributing green time within a fixed cycle, it applies a **learned, actuated policy** that responds in real time to current traffic conditions

### Performance vs. Cost

**Strong performance** — SMART A consistently achieves top results on Mean Travel Time and tram prioritization across traffic conditions.

**High computational cost** — A single one-hour simulation can require **over 24 hours** of wall-clock computation, limiting practical deployment at scale.



# GA vs RL: Not Opposition, But Progression

## Genetic Algorithms

- Structured, interpretable planning
- Multi-objective optimization
- Explicit objectives and constraints
- Urban-readable formulations
- Robust scenario solutions
- Strong fairness guarantees

### Adaptive Response to Changing Conditions

Not just searching for a good plan, but learning how to act while traffic evolves in real time

### Natural Fit for Distributed Multi-Agent Settings

More consistent with large urban networks and local coordination requirements

RL adds sequential adaptation, better naturality in distributed control, and greater responsiveness—but GA remains strong on fairness, interpretability, and plan robustness.

## Reinforcement Learning

- Adaptive, sequential control
- Policy-based behaviors
- Continuous learning
- Dynamic adaptation to load
- Natural distributed logic
- High responsiveness

### Better Sequential Control Under Medium/Heavy Load

MADRL with PPO emerges as particularly effective when traffic demand increases

### Strong Real-Time Responsiveness with SMART

Actuated control modifying green times based on current observations through TraCI

# Case Study and Evaluation Setup

The experimental validation is grounded in a real, non-synthetic urban scenario — a portion of the city of Florence featuring heavy private traffic and active tram lines.

01

## Real Urban Network — Florence

Intersection topology, tram corridors, and traffic demand are drawn from real-world operational data, not stylized test environments.

03

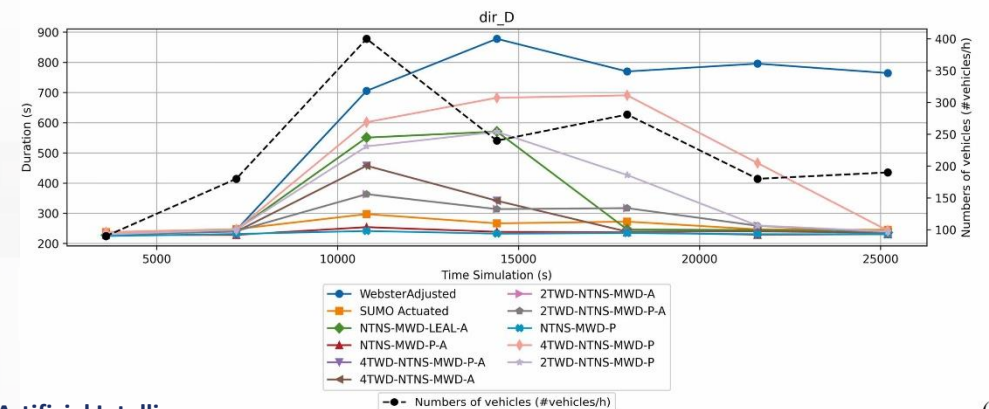
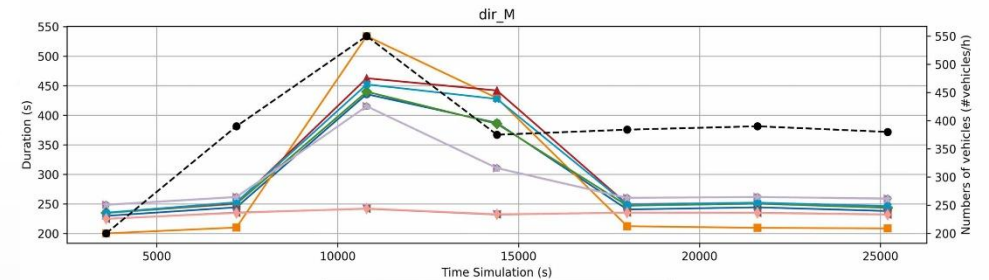
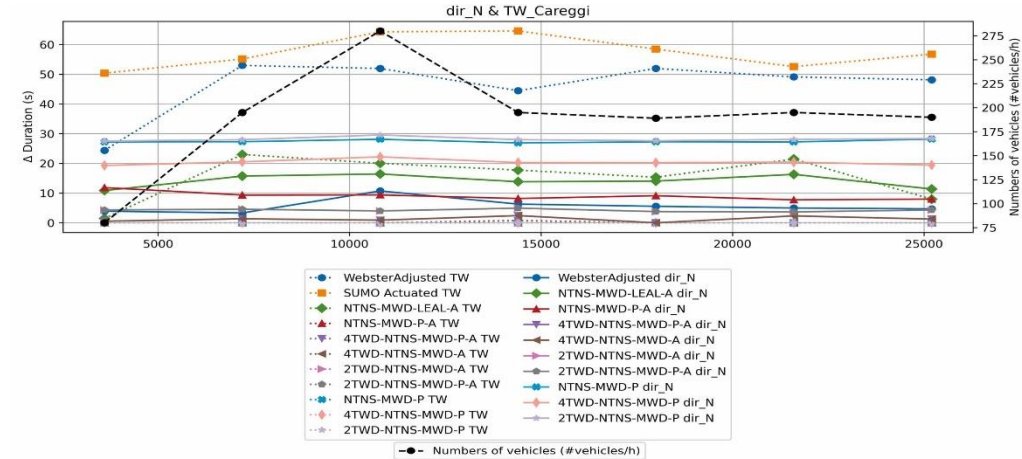
## Strong Baseline Comparisons

Benchmarks include Webster timing, SUMO actuated control, and other multi-objective or genetic variants already established in the literature — not trivial strawmen.

02

## Macroscopic Planning, Microscopic Evaluation

Timing plans are designed at the network level, then evaluated in high-fidelity microsimulation using **SUMO** and the **TraCI** API — enabling global design with detailed testing.



# Mean Travel Time (MTT)

Traffic Load	4TWD-NTNS-MWD-A	SUMO Actuated	Webster
1.0	3,013.85	2,935.41 ☑	5,188.87 ✗
1.5	3,242.71 ☑	3,409.13	6,474.95 ✗
2.0	3,457.86 ☑	4,666.32 ✗	7,636.76 ✗

## GA Dominates

The GA method achieves the lowest MTT, maintaining a stable trajectory even as congestion doubles.

## Webster Degrades Sharply

Webster's fixed timing fails under increasing demand — travel time increases sharply from load 1.0 to 2.0, exposing its core limitation.

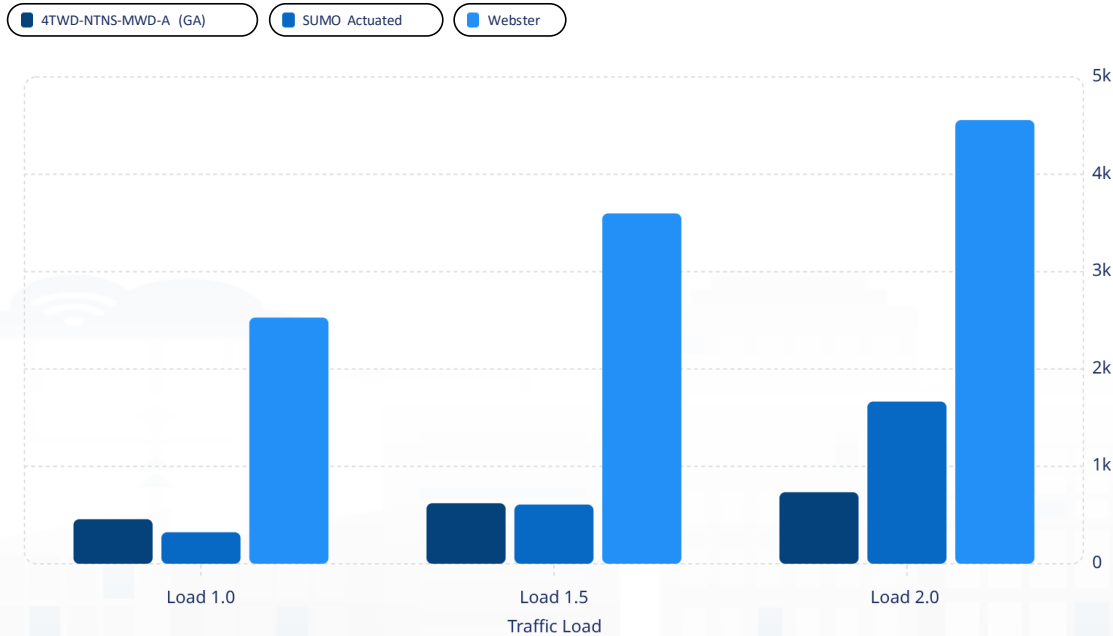
Values in seconds. Lower is better.

# Mean Number of Stops (MNS)

Traffic Load	4TWD-NTNS-MWD-A	SUMO Actuated	Webster
1.0	21.99 ✓	38.55 ✗	43.85 ✗
1.5	32.07 ✓	56.08 ✗	72.14 ✗
2.0	41.61 ✓	78.68 ✗	101.10 ✗

- The GA method reduces vehicle stops by up to **47% vs. SUMO** and **59% vs. Webster** at high traffic load — a dramatic reduction in both emissions and driver frustration.

# Mean Waiting Time (MWT)



**Low Load (1.0)**  
SUMO Actuated edges out GA slightly (323 vs. 457 sec). Both vastly outperform Webster (2,528 sec).

**Medium–High Load (1.5–2.0)**  
GA takes the lead decisively. SUMO degrades rapidly, approaching 1,666 sec at load 2.0.

**Webster — Consistent Worst**  
Waiting times balloon to over 4,500 sec at peak load. Fixed-time control is clearly insufficient under dynamic demand.

# MTT Results: Full Comparison Across Traffic Loads

Mean Travel Time (in seconds) measured at three traffic load levels. Lower is better. **Green** = best-in-class performance.

Traffic Load	SMART (RL)	MADRL (RL)	SADRL (RL)	MamoTLO (GA)	SUMO	Webster
1.0×	2599.13	2910.76	2760.12	3013.85	2935.41	5188.87
1.5×	2277.58	3229.54	3189.84	3242.71	3409.13	6474.95
2.0×	3443.96	3301.19	3490.87	3457.86	4666.32	7636.76

**Key Takeaway:** SMART (RL) achieves the lowest MTT at all load levels. At 2.0× load, MADRL closes the gap significantly — suggesting strong scalability under heavy congestion.

# Method Comparison: Performance, Scalability, and Cost

No single method dominates on all three dimensions. The right choice depends on deployment constraints — available compute, network size, and acceptable latency.

Method	Performance	Scalability	Cost
SMART (RL)	● ● Top	●	● ● ●
MADRL (RL)	● ●	● ●	●
SADRL (RL)	●	●	●
MamoTLO (GA)	● ●	● ●	●
SUMO	●	●	●
Webster	●	●	●

## Research Context

Prioritize SMART (RL) for benchmarking upper bounds

## Urban Deployment

MADRL or MamoTLO offer the best real-world balance

# Interpreting the Results: RL vs. GA

Each method occupies a distinct niche — understanding the trade-offs is essential for practical deployment.

## SMART (RL Actuated)

Best global MTT across all load levels. However, its **extremely high computational cost** limits real-world applicability. Ideal as a performance upper bound in research settings.

## MADRL (Multi-Agent RL)

Excellent performance under high traffic. Scales gracefully across intersections, making it the **most realistic RL candidate** for deployment in urban networks.

## SADRL (Single-Agent RL)

Competitive at low traffic loads, but **degrades under high congestion**. Single-agent architecture limits coordination in complex, multi-intersection scenarios.

## MamoTLO (GA)

Consistently stable across all load levels. Offers a **strong performance-to-cost ratio** — highly robust without the training overhead of deep RL approaches.

# Direct Comparison: GA vs. DRL

Neither paradigm universally dominates. The choice depends on whether the priority is **interpretable constraint satisfaction** or **state-dependent adaptive control**.

Aspect	GA / NSGA	DRL
Nature	Offline multi-objective optimization	Sequential policy learning from interaction
Main strength	Explicit objectives, constraints, and interpretable trade-offs	High online adaptivity and state-dependent control
Main limitation	Plan-oriented; adaptation is added, not learned	Lower interpretability; reward design and training matter
Best use case	Robust scenario optimization with fairness and constraint control	Distributed, real-time adaptive control under changing traffic
Interpretability	Higher — Pareto fronts are inspectable	Lower — neural policy is a black box
Computational cost	High offline search cost	High training/runtime cost, especially for SMART

## Conclusion



### Data-Driven Control Needs Digital Twins

No robust traffic signal optimization without data, reconstruction, and executable digital twins



### GA Is the Strong Structuring Block

Multi-objective optimization makes priorities, constraints, and trade-offs explicit



### RL Matters When Adaptation Becomes Central

The transition is natural when control shifts from plans to policies



### The Real Contribution Is the Full Pipeline

Data, reconstruction, simulation, GA, and RL reinforce each other as one methodology



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DINFO**  
DIPARTIMENTO DI  
INGEGNERIA  
DELL'INFORMAZIONE

**DISIT**  
DISTRIBUTED SYSTEMS  
AND INTERNET  
TECHNOLOGIES LAB



# From GA to RL: Data-Driven Traffic Light Optimization

4th International Conference on Data Science, Agents, and Artificial Intelligence  
ICDSAAI 2026

---

**Dr. Luciano Alessandro Ipsaro Palesi**

Keynote Speaker

[lucianoalessandro.ipsaropalesi@unifi.it](mailto:lucianoalessandro.ipsaropalesi@unifi.it)

Department of Information Engineering (DINFO)

DISIT lab, <https://www.disit.org>, <https://www.snap4city.org>

University of Florence

Thank you for your  
attention

Questions and  
discussion