Assessing Mobility Policies by Traffic Simulation and Change Detection*

Felipe Bagni[1], Edoardo Peretti[1], Danilo Macció[2], Cristiano Cervellera[2], and Giacomo Boracchi[1]

[1] Politecnico di Milano, Milano, Italy
felipe.bagni@mail.polimi.it, edoardo.peretti@polimi.it, giacomo.boracchi@polimi.it

[2] National Research Council, Institute of Marine Engineering, Genova, Italy
cristiano.cervellera@cnr.it, danilo.maccio@cnr.it

Abstract. Defining effective policies for managing traffic in large cities, particularly near major logistics hubs such as ports, is a challenging problem due to the critical interaction between mobility and freight flows. In this work, we combine a traffic simulator with a change-detection test to assess a priori whether a specific policy would have an impact on city mobility. More specifically, we propose a general methodology to identify the expected number of days/monitoring samples before gaining evidence that the policy has introduced a detectable change in the traffic data acquired after enforcing the policy. Our experiments, conducted on simulated traffic, focused on the port-city context of Genova, showcase that our proposed methodology can provide outcomes that are consistent with the kind and expected effectiveness of policies under evaluation.

1 Introduction

Traffic management is a critical problem in cities, especially in areas where the urban infrastructure supports both mobility flows and traffic from large logistics hubs. A port located near the city center is a paradigmatic example in which this kind of disruptive interaction occurs, requiring city authorities to define suitable mobility policies to reduce externalities like pollution and congestion [16, 24, 27]. Indeed, policy definition is crucial and involves many stakeholders: citizens, logistics operators, public transport companies, and port authorities.

Policies that can be implemented include diverse strategies like opening or closing roads, changing routing plans and, most commonly, changing traffic light timing [6]. Yet, for any potential new policy, assessing a priori whether it would be effective is crucial to reduce the risk of making wrong or ineffective choices. To this purpose, sophisticated analytics are typically employed, including traffic simulators that can generate realistic traffic conditions over multiple days [4].

* The work was supported by the Project "AI for Sustainable Port-city logistics" funded by the European Union - Next Generation EU Program through the Italian Ministry for University and Research PRIN 2022 PNNR Grant P2022FLLPY.

While the ultimate validation of a policy needs to be conducted by analyzing data acquired on the field, it is important for policy makers to understand a priori the effectiveness of a new policy. In this paper, we propose a general methodology that leverages traffic simulator and change-detection tests for assessing whether a specific policy would be effective, namely whether it has a statistically significant impact on the simulated traffic at specific locations, and how long policymakers should wait before having a confirmation from the field that the policy is effective.

Adopting change detection in urban traffic monitoring is not unheard of, and there is a broad literature in this regard, primarily addressing the detection of accidents or road impairments (see, e.g., [7]). However, using statistical tools to assess how long it will take to conclude that a policy would be effective is challenging for several reasons. First, the input to be analyzed easily becomes high dimensional, as traffic is conveniently monitored at different city locations and selectively for different categories, like the flow and the speed for each type of vehicle (e.g., cars, trucks, etc.). Second, urban networks are complex systems characterized by chaotic and non-linear dynamics, making the distributions of measured quantities unknown. Therefore, running a statistical test to determine whether a high-dimensional stream of mobility data (following an unknown distribution) has changed after policy enforcement is far from trivial, in particular when it comes to granting control over the false alarm rate [11].

Furthermore, the analysis needs to be conducted in a sequential manner, as the final policy assessment is to be conducted on online real data. However, acquiring real data before and after policy enforcement is not trivial, since uncontrolled system dynamics might invalidate the stationarity assumption, and collecting real data requires complex logistics and a waiting time longer than simulated data. Our proposal focuses on assisting policymakers in their planning (the relevant literature is summarized in Section 2). We aim at understanding whether a policy is expected to have an influential impact on traffic. In particular, we combine change-detection with traffic simulation, to make policymakers informed whether the policy would be perceivable in simulated data, and to estimate the delay required to observe an impact on the traffic.

Our methodology (illustrated in Section 4) combines a traffic simulator and a nonparametric change-detection test to assess the expected latency for the policy to induce a detectable change. Traffic simulators can nowadays achieve realistic outcomes, simulating multiple vehicle categories and modeling complex dynamics [8]. Thus, we simulate traffic before and after policy enforcement to replicate mobility data over multiple days. The second component, the change-detection test, is a sequential statistical tool that analyzes multivariate streams. More specifically, we adopt QT-EWMA [11], which does not underpin any parametric model (e.g., a Gaussian distribution) to describe the multivariate distribution of pre-policy enforcement data, and it allows monitoring while controlling the expected number of samples before having a false positive detection.

In our experiments (Section 5), we consider as a case study the port-city context of Genova, one of the main ports in the Mediterranean Sea. In particular, we investigate three different policies and show that our method provides consistent findings with the expected traffic behavior. More specifically, our test can control the expected delay for false positives, and confirms that changes are more perceivable in specific streets. This is an encouraging result, which paves the way for extending the use of QT-EWMA to address real-time traffic monitoring, combining the monitoring scheme with time-series regression methods.

2 Related Work

Most of the literature on policy evaluation in the context of urban traffic and mobility is focused on ex-post assessment, relying on the analysis of real-world data collected after policy enforcement. By this approach, the effectiveness of mobility policies is assessed only after they have been implemented, while policy makers would need to assess the policy effectiveness beforehand. Typically, the assessment involves the definition of ad-hoc indices of performance of the network, often based on user experience surveys [12, 15, 17], rather than directly relying on traffic flow.

A popular approach to evaluate ex-post the impact of policies relies on a regression model trained on the observed traffic data. This translates into analyzing the impact of the model variables affected by the policy on the predicted outcome. For instance, in [13] a simple nonlinear regression model is employed to evaluate the impact of urban transportation policies from various real world datasets, while in [14] a similar approach has been devised to evaluate the impact of driving restrictions in Beijing. In [20] a parametric time series modeling approach based on the so-called basic structural model has been implemented to evaluate the effect of a new mobility policy in Milan, comparing the pollution levels before and after the policy enforcement. In contrast, our approach leverages simulated scenarios and a nonparametric methodology, to investigate a priori whether a policy is effective with statistical guarantees.

A popular and convenient choice to evaluate the impact of policies a priori is to resort to simulation. By simulating the traffic dynamics of urban networks, it is possible to evaluate lines of intervention in a "what-if fashion" before they are actually deployed. Some approaches in the literature adopt a modeling methodology to describe policies at a strategic level, involving several actors and stakeholders of the urban environment (see, for instance, [25, 26] based on a system dynamics approach). However, these papers do not address the problem of determining whether a policy has significantly impacted urban traffic in general, but they limit themselves to particular cases such as strategies to reduce carbon emissions or urban traffic restriction policies.

When the analysis is focused on the actual performance of the urban network, as in our settings, traffic models of the micro-simulation kind are the most suited to capture the effects of changes induced by policies. Micro-simulation [8, 9] is based on the "car following" principle and can represent the behavior of individual vehicles, providing a dynamic and detailed representation of all the relevant features of the network (e.g., traffic lights, junctions, priorities, etc). Thus, traffic simulation models can be used to provide a realistic assessment of the effect of a wide range of policies, both at the strategic and tactical level. Examples of policy evaluation approaches based on traffic micro-simulation can be found in [19] and [22] where the Eclipse/SUMO open source simulator is used, in [10] where the commercial VISSIM software is used, and in [1] using the commercial software AIMSUN. In all these approaches, the simulation tool generates scenarios with different policies, and then suitable performance indices are defined to evaluate the outcomes. Yet, there is no principled statistical methodology to determine whether a new policy would actually affect urban traffic, and how long it takes to confirm it from real measurements. In contrast, our methodology utilizes traffic simulation combined with a sequential change-detection test to estimate the expected time required for a policy to take effect.

Urban traffic contexts characterized by the interaction of mobility and logistics flows are probably among the most critical for policy changes. A port located within a city is the most typical example of these kinds of settings, thus in this paper we take this as the reference scenario for our proposed methodology. The presence of systematic flows from heavy vehicles on top of the general mobility traffic is a source for negative externalities such as pollution and congestion [5,16,28]. These externalities increased the need for policies to efficiently regulate the urban road network where port and mobility traffic interact. In the literature, there are quite a few papers focused on policy evaluation and assessment in port-city environments, but the analysis is generally performed at a very strategic level with a focus on economic and social impact, and is based on ex-post real data analysis (see, e.g., [21,23]). The present paper introduces for the first time a methodology to evaluate the impact of policies in port-city scenarios a priori, focusing on actual changes detectable in the network performance.

3 Problem Formulation

We address the problem of assessing whether a mobility policy significantly impacts an urban network. To this end, we measure traffic data, such as the number and speed of vehicles, at one or more network's points of interest. We denote by $x_t \in \mathbb{R}^d$ a sample of traffic data measured on day t. We assume that the measurements on different days are independent and identically distributed (i.i.d.), and denote by ϕ_0 the unknown distribution of traffic data under a basic scenario.

We define a mobility policy p as a set of actions an authority can enforce on urban roads to modify mobility. For example, a policy may involve altering the timing of traffic lights or regulating access to a street depending on the vehicle class. Each policy p generates a change in the traffic distribution, namely $\phi_0 \xrightarrow{p} \phi_1$, where ϕ_1 is the distribution of the data after the enforcement of the policy. We say that a policy p is *effective* when the distribution ϕ_1 differs from the distribution of traffic data ϕ_0 before policy enforcement.

We assume a training set of N stationary samples $\widetilde{x}_1, \dots, \widetilde{x}_N \overset{\text{i.i.d.}}{\sim} \phi_0$ from the initial distribution ϕ_0, where each sample corresponds to a distinct day. Then, we enforce a policy p and start collecting samples $x_1, x_2, \dots \overset{\text{i.i.d.}}{\sim} \phi_1$ from the traffic scenario generated by p. Our goal is to determine whether the impact of a policy

p on the network has a statistically significant effect by testing for $\phi_1 \neq \phi_0$. Also, we aim to perform the testing sequentially, meaning that a decision should be taken for every incoming sample x_t, and to measure how many samples after policy enforcement are needed to conclude that the policy is effective. To avoid long data collections, we select candidate policies on simulated data.

4 Proposed Solution

In this section, we describe our solution to assess mobility policies in a road network. As anticipated in Section 1, our solution combines: *(i)* a micro-traffic simulator, and *(ii)* a nonparametric change-detection algorithm. To avoid long acquisition campaigns, we generate realistic traffic data using a traffic simulator. In this way, we can simulate large datasets of traffic data under arbitrary mobility policies, which would otherwise require months of measurements on the real world network. Using simulated data, we model the behavior of the network in a basic scenario with a change-detection algorithm, obtaining an estimate $\widehat{\phi}_0$ of the statistical distribution ϕ_0 of the traffic data in the basic scenario. Each mobility policy p gives rise to a different traffic scenario with data distribution ϕ_1. We monitor a stream of traffic data in search of statistical evidence to assert that $\phi_1 \neq \phi_0$, meaning that p significantly impacts the network's traffic.

4.1 Traffic Simulator

We select candidates policies employing a stochastic simulator to generate traffic data according to a user-defined mobility policy. Thus, we implement the city road network in the Eclipse/SUMO traffic micro-simulator [18], a popular open-source tool that enables a detailed representation of traffic dynamics. Since we are mostly interested in scenarios where mobility and freight traffic interact we define standard vehicle classes (i.e., general mobility and logistic), each characterized by specific size, speed, acceleration, etc. In the simulation, each vehicle represents an equivalent unit of the corresponding class (truck or car).

The simulator models the city road network with several *edges*, which are the basic components of the road system (e.g., a portion of a street or a roundabout). We denote by $\mathcal{S}$ the set of all the network edges. For each edge $s \in \mathcal{S}$, the simulator generates the average values, over intervals of arbitrary length, for the following six variables: *(i)* number of cars (`num_cars`), *(ii)* speed of cars (`speed_cars`), *(iii)* flow (i.e., vehicles that left the edge) of cars (`flow_cars`), *(iv)* number of trucks (`num_trucks`), *(v)* speed of trucks (`speed_trucks`), and *(vi)* flow of trucks (`flow_trucks`). We generate simulations for several days by independent runs of the simulator. Thus, we obtain a sequence of day-indexed traffic data $x_1, x_2, \dots \in \mathbb{R}^d$, where d is the total number of variables to monitor.

4.2 Input preparation

For each simulated day, we generate 6 traffic variables (`num_cars`, `speed_cars`, `flow_cars`, `num_trucks`, `speed_trucks` and `flow_trucks`) for each edge in

$\mathcal{S}$, where the cardinality $|\mathcal{S}|$ is typically in the order of hundreds. Monitoring streams of variables with hundreds of dimensions is challenging. Moreover, most of the edges are strongly correlated due to the simulation dynamics, thus not very statistically informative. Therefore, we select a small subset $\mathcal{S}_m \subset \mathcal{S}$ of edges of interest, namely the monitoring points of the network. We define $\mathcal{S}_m$ in agreement with traffic experts, including edges that are crucial for assessing mobility policies in the road network under study. To facilitate interpretability, we monitor both traffic at all edges and separately at each edge $s \in \mathcal{S}_m$ individually, resulting in $|\mathcal{S}_m|$ distinct monitoring problems with 6-dimensional data.

The raw data provided by the simulator needs to be preprocessed before it can be used in a change-detection algorithm. A mobility policy or the current traffic regulations may restrict the access to certain streets to only cars, resulting in simulated trucks related variables (`num_trucks`, `speed_trucks` and `flow_trucks`) that are constant equal to zero. Since these variables are useless for the monitoring problem, we remove them from the dataset. Moreover, we note that 2 variables, namely `flow_cars` and `flow_trucks`, take on discrete values in any scenario. Since histogram-based methods (including QT-EWMA [11]) struggle with discrete variables, we add a small Gaussian noise to these two variables, exclusively for change-detection purposes.

4.3 Change-detection method

We frame the policy assessment problem as a special change-detection problem where the change point is known and corresponds to the enforcement of a mobility policy, where the stationary distribution ϕ_0 corresponds to the current scenario in the real-world network. We use the training set $\{\widetilde{x}_1, \dots, \widetilde{x}_N\}$ to fit an estimate $\widehat{\phi}_0$ of the distribution ϕ_0 of the basic scenario. Then, given a policy p, we configure the simulator according to p and generate observations $x_1, x_2, \dots \in \mathbb{R}^d$, which we model as i.i.d samples from the unknown distribution ϕ_1 of the traffic scenario generated by the policy p.

For each new incoming sample x_t, an online change-detection algorithm assesses whether the sequence $\{x_1, \dots, x_t\}$ provides sufficient statistical evidence of a change in the data-generating process, namely whether the policy is effective. Specifically, at each time instant t, the algorithm computes a statistic T_t, based on the estimated initial distribution $\widehat{\phi}_0$, and applies a decision rule comparing it to a time-dependent threshold h_t. The detection time t^* is defined as the first instant of time when the statistic T_t exceeds the threshold h_t, that is:

$$t^* := \min\{t \in \mathbb{N} : T_t > h_t\}. \tag{1}$$

We recall that a policy is always enforced at the beginning of the data simulation process. Thus, when p induces a real change $\phi_0 \to \phi_1$, such that $\phi_0 \neq \phi_1$, the detection time t^* corresponds to the *detection delay*, since the policy is enforced at the beginning of the monitored stream.

A fundamental metric of any online change-detection algorithm is the expected number of time steps before a false detection, the Average Run Length (ARL_0) [2] is defined as:

$$\mathrm{ARL}_0 := \mathbb{E}_{x\sim\phi_0}[t^*],$$

where the expectation is with respect to the training distribution ϕ_0, namely the data stream comes from the basic traffic scenario. Our goal is to estimate the average detection delay t^*, controlling the ARL_0, namely obtaining an *empirical* ARL_0 close to a user-specified target ARL_0. We note that controlling the ARL_0 is of utmost importance, since it is the sequential counterpart of the Type I error probability in offline hypothesis testing. An ARL_0 that is too small would result in numerous false alarms, causing us to incorrectly conclude that a policy is effective when it is not, making the detection delay meaningless.

We detect changes in traffic data using the QuantTree Exponentially Weighted Moving Average (QT-EWMA) algorithm [11]. During training, the algorithm uses the training set $\widetilde{x}_1, \dots, \widetilde{x}_N \overset{\text{i.i.d.}}{\sim} \phi_0$ to approximate the unknown distribution ϕ_0 of the basic scenario, constructing a histogram with K bins, which are created by partitioning $\mathbb{R}^d$ along random directions (a QuantTree [3]). The splitting points are defined so that each bin contains a user-specified quantile of the training set, resulting in a uniform partition of the data into bins. During monitoring, for each new sample x_t, QT-EWMA updates a statistic T_t in an incremental manner and compares it against a time-dependent threshold h_t. The details of the QT-EWMA algorithm can be found in [11].

We choose QT-EWMA since it is a state-of-the-art online nonparametric change-detection method suitable for high-dimensional data with arbitrary distribution. Remarkably, QT-EWMA provides theoretical guarantees on the ARL_0, allowing one to control the false alarm performance of the algorithm. The thresholds h_t do not depend on the stationary distribution ϕ_0 or the number d of dimensions. They depend exclusively on the size of the training set and the number of histogram bins, and need to be precomputed using Monte Carlo simulations. In our experiments, we use the publicly available thresholds that can be found in the official repository[3].

5 Experiments

5.1 The Genova Port-City Scenario

We choose as case study the port-city context of Genova, a town in north-west Italy hosting one of the most important container ports in the Mediterranean Sea. The area considered is depicted in Figure 1, and is characterized by the presence of port gates and highway toll booths in close proximity, thus it is an area of intense interaction between logistics and mobility traffic. In the figure, the 'T' signs denote a traffic entry/exit point. In particular, T1 corresponds to mobility vehicles to/from western parts of the town through Lungomare Canepa, and trucks to/from the Aeroporto highway toll booth. T2 is the direction to/from the Genova Ovest highway toll booth. T3 is a direction to/from western parts of

[3] https://github.com/diegocarrera89/quantTree

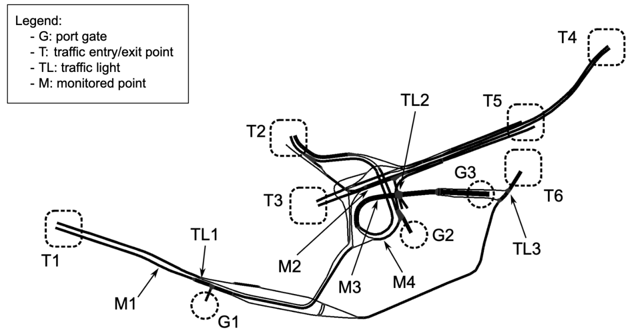

Fig. 1. The Genova's port-city network under study.

the town interested mainly by mobility traffic and ferry passengers. T4, T5 and T6 are various entry/exit points for eastern parts of the town, interested only by mobility traffic. The points denoted by 'G' are port gates. Specifically, G1 and G2 are the Etiopia and San Benigno gates, respectively, devoted mainly to freight traffic, while G3 is the Alberatzzi gate, mainly devoted to ferry passengers. The 'TL' signs mark the position of traffic lights that will be used to define changes of policies as will be detailed below. The points at which the variables of interest are monitored are indicated by 'M', and are respectively called `Lungomare Canepa`, `Via di Francia`, `Elicoidale Upstream` and `Elicoidale Downstream`. These four edges of interest represent edges in which general mobility traffic mixes with logistics (M1, M3 and M4) and ferry traffic, giving rise to possible congestion.

In order to define the traffic demand between origins and destinations, we assume different behavior for trucks and mobility vehicles. In particular, we assume that trucks follow fixed routes between gates and highway toll booths. We considered 6 truck routes, namely (G2 $\rightarrow$ T2), (T2 $\rightarrow$ G2), (T1 $\rightarrow$ G1), (T1 $\rightarrow$ G2), (G1 $\rightarrow$ T1) and (G1 $\rightarrow$ T2). General vehicles and ferry passengers instead are assumed to move freely among the origins and destinations, adapting the route depending on current traffic conditions. For this kind of traffic we considered a total of 19 origin/destination pairs, i.e., (T1 $\rightarrow$ T4), (T1 $\rightarrow$ T5), (T1 $\rightarrow$ T6), (T1 $\rightarrow$ T2), (T3 $\rightarrow$ T2), (T3 $\rightarrow$ T4), (T3 $\rightarrow$ T5), (T3 $\rightarrow$ G3), (T2 $\rightarrow$ T1), (T2 $\rightarrow$ G3), (T2 $\rightarrow$ T4), (T2 $\rightarrow$ T5), (T5 $\rightarrow$ T3), (T4 $\rightarrow$ T3), (T4 $\rightarrow$ T2), (T4 $\rightarrow$ T1), (G3 $\rightarrow$ T2), (T6 $\rightarrow$ T1) and (T6 $\rightarrow$ T2). Overall, these correspond to a 25-dimensional vector of inflow rates F used to drive the simulation scenario.

5.2 Simulator Configuration and Datasets Generation

To test our methodology, we define two basic scenarios by setting Poissonian distributions as input inflows of the simulator and we then record the 25-dimensional vector of hourly inflow rates F. Adopting Poissonian inflows is a standard in queueing and traffic systems, and yield realistic discrete values. The first basic scenario (denoted by A) is intended to simulate the traffic during working days,

		Average Empirical ARL_0		
N	ARL_0	ϕ_0^A	ϕ_0^B	ϕ_0^{AB}
64	500	490.4	482.6	487.3
	1000	998.0	984.4	984.5
128	500	502.8	489.5	509.0
	1000	1017.6	1003.1	1006.5
256	500	495.0	510.9	496.0
	1000	986.4	1006.8	1017.5

Table 1. Empirical ARL_0 with respect to training set size, target ARL_0 and basic scenario distribution. The empirical ARL_0 is close to the target ARL_0 in all cases.

while the second (denoted by B) is for weekends, which are characterized by less traffic. We additionally consider a mixture basic scenario (denoted by AB) by sampling traffic data from scenario A and scenario B, with a 70-30 proportion. We denote the data distributions under these three basic scenarios by ϕ_0^A, ϕ_0^B and ϕ_0^{AB}, respectively. For each basic scenario, we consider the following three policies with increasing impact:

- *Policy 1* (p_1): little shifting of a single traffic light's phase (TL2).
- *Policy 2* (p_2): shifting the phase of three traffic lights (TL1, TL2, TL3).
- *Policy 3* (p_3): two traffic lights (TL1, TL2) are now dynamic, adjusting the phases based on current traffic instead of using fixed phase programs.

To generate the datasets, for each scenario, we simulate 1000 days with 1 hour of simulation time. More specifically, we considered a simulated time of 4200 seconds per day, the first 600 seconds are considered as warm up to populate the empty network, and then discarded. Then, to generate the six traffic data variables of interest described in Section 4.1, we aggregate the values from the simulator. Notice that, given the short time interval we consider every day, the traffic at the gates and the highway booths contain all the required information for the logistic flows, while other port-related factors such as incoming ships/trains or yard storages would act on longer time frames.

5.3 Results

In this section, we present the results obtained from the experiments described above. We start by measuring empirical ARL_0, then show the ability of our methodology to assess the effectiveness of policies.

Table 1 shows that the empirical ARL_0 consistently aligns with the target across all training set sizes N, target ARL_0 values, and data distributions, confirming the reliability of our method. The agreement improves when N increases, with $N = 256$ showing the best match. Slight deviations, particularly in the ϕ_0^B setting at $\text{ARL}_0 = 1000$, suggest some sensitivity to data characteristics. Overall, the method controls ARL_0 under all conditions, even small N yielding reliable results. Tables 2, 3, and 4 report the detection delays for the three basic scenarios, namely weekdays (ϕ_0^A), weekends (ϕ_0^B) and mixture (ϕ_0^{AB}). We present

Policy	N	ARL_0	Average Detection Delay				
			`Elicoidale Upstream`	`Elicoidale Downstream`	`Lungomare Canepa`	`Via di Francia`	`All 4 edges`
p_1	64	500	478.9	238.9	556.9	49.3	251.9
		1000	935.3	497.2	1050.1	73.7	405.1
	128	500	488.2	202.6	615.3	35.5	170.2
		1000	876.5	358.3	1068.3	46.4	255.4
	256	500	471.8	158.8	701.9	30.6	130.7
		1000	922.0	268.6	1207.5	35.8	186.1
p_2	64	500	489.8	35.7	132.5	16.0	42.1
		1000	982.9	50.1	223.5	19.2	55.4
	128	500	531.8	26.5	79.8	13.9	26.5
		1000	926.5	32.7	112.7	16.0	38.7
	256	500	520.3	22.4	59.7	12.2	22.7
		1000	950.0	27.6	74.4	13.7	27.7
p_3	64	500	323.4	5.0	22.2	8.7	7.7
		1000	622.6	5.0	26.0	9.8	8.2
	128	500	266.1	5.0	16.4	7.8	6.6
		1000	455.5	5.0	18.7	8.6	7.3
	256	500	194.9	5.0	14.4	7.5	6.4
		1000	335.0	5.0	16.1	8.2	6.9

Table 2. Average detection delays (in days) during weekdays. Policy p_1 is weakly effective since delays are large in all segments, policy p_2 is more effective in few edges, while policy p_3 is effective at network level and in all edges except `Elicoidale Upstream`.

the detection delays for each of the three different policies (p_1, p_2, and p_3), as a function of the training set size N and target ARL_0. In particular, we report in each column the detection delays for each monitored edge individually $s \in \mathcal{S}_m$, as well as for their aggregation $\mathcal{S}_m$. The results reveal a strong dependency between the type of policy change and the timeliness of the detection of alterations in the traffic flow.

In the weekday-only scenario (Table 2), detection delays are longest under p_1, which involves a minor phase shift at a single traffic light. This is most evident in `Lungomare Canepa`, where delays consistently exceed the target ARL_0, indicating that p_1 only produces subtle traffic changes and is weakly effective. Increasing the training set size N yields limited improvement, as the underlying change remains minimal. In contrast, p_2, which modifies three signals, shows shorter delays, especially in `Elicoidale Downstream`, where detection occurs in less than 50 days, confirming that broader interventions improve responsiveness. p_3, which implements dynamic signal control, further reduces delays: in `Elicoidale Downstream`, detection occurs in just five days, the minimum for QT-EWMA. Even `Lungomare Canepa`, previously slow to respond, shows a significant improvement under p_3, confirming the effectiveness of system-wide adaptive changes, whereas `Via di Francia` demonstrates consistently short detection delays across all policies, indicating a strong and immediate sensitivity

Policy	N	ARL_0	Average Detection Delay: `Elicoidale Upstream`	`Elicoidale Downstream`	`Lungomare Canepa`	`Via di Francia`	`All 4 edges`
p_1	64	500	515.1	353.0	558.1	537.7	441.3
		1000	936.7	671.0	1027.7	1034.6	906.7
	128	500	555.5	301.9	623.2	594.3	486.0
		1000	1016.6	608.8	1079.7	1079.3	923.0
	256	500	604.3	288.7	712.4	664.9	512.0
		1000	1079.2	517.9	1130.6	1191.5	923.6
p_2	64	500	511.2	194.8	409.4	530.7	328.0
		1000	983.7	366.9	749.3	1053.4	646.4
	128	500	525.4	144.4	408.8	590.6	323.5
		1000	1024.8	268.3	736.6	1020.5	563.6
	256	500	577.7	121.1	425.7	676.6	290.2
		1000	1089.9	187.3	685.2	1146.1	535.0
p_3	64	500	505.9	14.1	198.0	555.9	131.7
		1000	926.2	16.5	310.9	1022.6	228.8
	128	500	539.7	12.5	149.4	594.1	99.7
		1000	1026.5	14.4	228.8	1075.5	147.8
	256	500	565.7	12.0	112.5	687.6	83.0
		1000	1038.5	13.6	153.2	1199.7	126.8

Table 3. Average detection delays (in days) during weekends. All policies appear less effective because of reduced traffic on weekends.

to all 3 policies. However, detection remains slow in `Elicoidale Upstream`, suggesting limited local impact. Across all the policies, higher ARL_0 values increase the detection delay, reflecting the expected trade-off between false alarms and responsiveness. This is most pronounced in p_1 due to its weak signals, while p_3 maintains rapid detection even at higher ARL_0, confirming the distribution change is very apparent.

The weekend-only (Table 3) and mixed scenarios (Table 4) show similar behaviors, but with added complexity. Detection is less timely during weekends, probably due to lower traffic volumes. Without congestion, variations in vehicle counts and speeds are minimal, limiting the effectiveness of change detection. In near free-flow conditions, signal timing changes have little influence on overall dynamics, so even structural interventions like p_3 may go largely undetected. This highlights a key dependency: while high congestion amplifies the effects of policy changes, low traffic volumes reduce their detectability.

The location of affected road edges highlights the localized impact of policy changes. In the weekday-only scenario, `Elicoidale Downstream` is most impacted, followed by `Via di Francia`, `Lungomare Canepa`, and lastly `Elicoidale Upstream`. This suggests that an edge's response depends not only on detection timing but also on its proximity to the modified signals and local traffic patterns. Edges closer to the changes react more clearly and quickly, while distant ones like

Policy	N	ARL_0	Average Detection Delay: `Elicoidale Upstream`	`Elicoidale Downstream`	`Lungomare Canepa`	`Via di Francia`	`All 4 edges`
p_1	64	500	502.4	295.2	555.1	109.6	314.5
		1000	959.3	541.1	1078.1	169.0	538.1
	128	500	520.4	255.6	623.7	76.6	293.0
		1000	967.5	454.2	1114.2	109.7	486.5
	256	500	558.1	222.5	691.0	59.2	256.1
		1000	1025.5	377.9	1215.9	75.4	436.1
p_2	64	500	523.3	79.2	225.2	37.5	123.7
		1000	983.1	120.6	383.7	48.4	187.8
	128	500	523.4	53.0	146.3	29.2	87.4
		1000	971.2	73.1	232.1	35.5	127.7
	256	500	570.8	42.7	106.8	24.2	61.0
		1000	1054.9	52.6	145.9	28.8	87.4
p_3	64	500	381.5	5.5	38.8	27.3	19.2
		1000	756.8	5.9	51.3	33.5	24.7
	128	500	536.1	5.4	26.0	19.8	14.1
		1000	619.8	5.6	31.0	23.7	16.3
	256	500	323.3	5.3	21.3	17.7	12.1
		1000	520.1	5.6	23.7	20.1	14.0

Table 4. Average detection delays (in days) on the mixed scenario. All policies appear less effective because data are influenced by weekend samples.

`Elicoidale Upstream` show weaker effects. This spatial variation underscores the need to consider localized traffic dynamics when evaluating mobility policies.

Our experiments confirm that the magnitude of policy interventions significantly influences detection performance, thus the time needed to confirm a policy is effective. Small traffic light phase shifts in p_1 are hard to detect, while broader interventions—such as modifying multiple intersections (p_2) or implementing adaptive signals (p_3)—substantially reduce detection delays. Detection is also affected by the choice of target ARL_0, with higher values increasing delay, and by traffic flow conditions, with weekday congestion enhancing detectability and weekend free-flow hindering it. Our analysis demonstrates that our methodology can effectively differentiate between weak, moderate, and strong interventions based on detection latency, offering a statistically rigorous improvement over traditional ex-post evaluations, which often miss short-term or spatially localized effects [13,20]. Notably, p_3 consistently yielded rapid detections, even under strict false alarm thresholds, supporting prior findings that dynamic, coordinated policies outperform isolated changes [6]. In contrast, the limited impact of p_1 reflects the ineffectiveness of marginal measures, echoing real-world challenges like those observed in Milan [20]. Effectiveness also varied by time and location—detections were slower during low-demand periods and differed across road segments—aligning with insights from [19,21].

6 Conclusions

We presented a general methodology for assessing mobility policies in an urban road network, with a focus on areas characterized by the competition of mobility and logistics flows, such as port-adjacent city centers. Using a change-detection algorithm in a simulated environment, our approach assesses a priori the impact of candidate policies without relying on extensive real-world data collection. Through online change-detection, the algorithm can promptly identify changes in the data-generating process, allowing for timely and data-driven assessment of policies. Applied to the port-city context of Genova, our experiments showed how the methodology is able to promptly identify effective policies, namely those requiring few days of monitoring to confirm the change. In future work, we will apply QT-EWMA to the residuals of surrogate models (e.g., neural networks) predicting traffic at each location. We believe that this approach may enhance the sensitivity to subtle deviations while filtering out expected variability.

References

1. Aditjandra, P.T., Galatioto, F., Bell, M.C., Zunder, T.H.: Evaluating the impacts of urban freight traffic: application of micro-simulation at a large establishment. European Journal of Transport and Infrastructure Research **16**(1) (2016)
2. Basseville, M.: Detection of abrupt changes: Theory and application. Prentice-Hall google schola **2**, 3–11 (1993)
3. Boracchi, G., Carrera, D., Cervellera, C., Macciò, D.: Quanttree: Histograms for change detection in multivariate data streams. In: International Conference on Machine Learning. pp. 639–648. PMLR (2018)
4. Cervellera, C., Macciò, D., Rebora, F.: Copula-based scenario generation for urban traffic models. Expert Systems with Applications **210**, 118389 (2022)
5. Cervellera, C., Macciò, D., Rebora, F.: Model predictive control of port-city traffic interactions over shared urban infrastructure. IEEE Trans. on Control Systems and Technology **32**, 688–695 (2024)
6. de Souza, F., Minatto Saucedo, R., Mousavizadeh, O., Castelan Carlson, R., Keyvan-Ekbatani, M.: On the evaluation and selection of network-level traffic control policies: Perimeter control, tuc, and their combination. Transportation Research Part A: Policy and Practice **186**, 104161 (2024)
7. Djenouri, Y., Belhadi, A., Lin, J.C.W., Djenouri, D., Cano, A.: A survey on urban traffic anomalies detection algorithms. IEEE Access **7**, 12192–12205 (2019)
8. Dorokhin, S., Artemov, A., Likhachev, D., Novikov, A., Starkov, E.: Traffic simulation: an analytical review. In: IOP Conference Series: Materials Science and Engineering. vol. 918, p. 012058. IOP Publishing (2020)
9. Ejercito, P.M., Nebrija, K.G.E., Feria, R.P., Lara-Figueroa, L.L.: Traffic simulation software review. In: 2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA). pp. 1–4. IEEE (2017)
10. Fontes, T., Fernandes, P., Rodrigues, H., Bandeira, J.M., Pereira, S.R., Khattak, A.J., Coelho, M.C.: Are hov/eco-lanes a sustainable option to reducing emissions in a medium-sized european city? Transportation Research Part A: Policy and Practice **63**, 93–106 (2014)

11. Frittoli, L., Carrera, D., Boracchi, G.: Nonparametric and online change detection in multivariate datastreams using quanttree. IEEE Transactions on Knowledge and Data Engineering **35**(8), 8328–8342 (2022)
12. Gao, S., Ran, Q., Su, Z., Wang, L., Ma, W., Hao, R.: Evaluation system for urban traffic intelligence based on travel experiences: A sentiment analysis approach. Transportation Research Part A: Policy and Practice **187**, 104170 (2024)
13. de Grange, L., Troncoso, R., González, F.: An empirical evaluation of the impact of three urban transportation policies on transit use. Transport Policy **22**, 11–19 (2012)
14. Gu, Y., Deakin, E., Long, Y.: The effects of driving restrictions on travel behavior evidence from beijing. Journal of Urban Economics **102**, 106–122 (2017)
15. Karjalainen, L.E., Juhola, S.: Urban transportation sustainability assessments: a systematic review of literature. Transport reviews **41**(5), 659–684 (2021)
16. Kong, Y., Liu, J.: Sustainable port cities with coupling coordination and environmental efficiency. Ocean & coastal management **205**, 105534 (2021)
17. Liu, X., Dijk, M.: The role of data in sustainability assessment of urban mobility policies. Data & Policy **4**, e2 (2022)
18. Lopez, P.A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., Wießner, E.: Microscopic traffic simulation using sumo. In: 2018 21st international conference on intelligent transportation systems (ITSC). pp. 2575–2582. IEEE (2018)
19. Lu, Q.L., Qurashi, M., Antoniou, C.: Simulation-based policy analysis: the case of urban speed limits. Transportation research part A: policy and practice **175**, 103754 (2023)
20. Maranzano, P., Fassò, A., Pelagatti, M., Mudelsee, M.: Statistical modeling of the early-stage impact of a new traffic policy in milan, italy. International journal of environmental research and public health **17**(3), 1088 (2020)
21. Merk, O., Dang, T.T.: The effectiveness of port-city policies: A comparative approach (2013)
22. Pais, F.V., Nogueira, B., Pinheiro, R.G.: Performance evaluation of urban traffic using simulation: A case study in brazil. IEEE Latin America Transactions **21**(12), 1275–1281 (2023)
23. Schipper, C., Vreugdenhil, H., De Jong, M.: A sustainability assessment of ports and port-city plans: Comparing ambitions with achievements. Transportation Research Part D: Transport and Environment **57**, 84–111 (2017)
24. Shankar, R., Choudhary, D., Jharkharia, S.: An integrated risk assessment model: A case of sustainable freight transportation systems. Transportation Research Part D **63**, 662 – 676 (2018)
25. Shi, J., Guo, X., Hu, X.: Engaging stakeholders in urban traffic restriction policy assessment using system dynamics: The case study of xi'an city, china. Sustainability **11**(14), 3930 (2019)
26. Wen, L., Bai, L.: System dynamics modeling and policy simulation for urban traffic: a case study in beijing. Environmental Modeling & Assessment **22**, 363–378 (2017)
27. Zheng, Y., Zhao, J., Shao, G.: Port city sustainability: A review of its research trends. Sustainability **12**(20), 8355 (2020)
28. Zhou, Y., Zhang, Y., Ma, D., Lu, J., Luo, W., Fu, Y., Li, S., Feng, J., Huang, C., Ge, W., Zhu, H.: Port-related emissions, environmental impacts and their implication on green traffic policy in Shanghai. Sustainability **12**(4162), 1–17 (2020)