Delay and disruption management at ATM: technical details

Federico Malucelli *[1] and Emanuele Tresoldi †[1]

[1]DEIB - Politecnico di Milano
[2]Dipartimento di Informatica - Università degli Studi di Milano

October 10, 2018

1 Introduction

The quality of a local public transport depends on the perceived efficiency and reliability. However, events or delays innate in the system, especially in a urban setting, may generate disruptions that negatively influence this perception. Additionally, disruptions usually increase the operating cost, for instance, involving extra allowances for bus drivers, or penalties to be payed to the municipality that commended the service.

Currently, the daily operations of transit companies are monitored "manually" by an operation central office taking advantage of Automated Vehicle Monitoring (AVM) systems and mobile telecommunication devices. The operation central office runs 18 or 24 hours per day, 7 days per week. Each operator controls the operations of one or more lines, detecting delays or anomalies that may generate disruptions or collecting information about problems on the line such as vehicle breakdowns or accidents. In the presence of a disruption, the operator assists remotely the driver who is facing the disruption, decides the actions to be taken coordinating, in case, also the behavior of the other drivers on the same line or other lines sharing a portion of the network with the disrupted one. In addition, the operator must inform passengers both onboard and waiting at the stops, of the new solution adopted to tackle the disruption. The basic actions that an operator in the central office can evaluate are, for instance, to perform vehicle detours, to delay a vehicle, or to cancel one or more trips. The operator may also decide to use spare resources (drivers or vehicles). However, this is not an option for the ordinary disruption management since, usually, these resources are extremely scarce and are left for tackling exceptional cases only.

*federico.malucelli@polimi.it
†emanuele.tresoldi@unimi.it

In this paper, we report the details of the delay and disruption management system proposed in Malucelli et al. (2018), where the case study of the management of urban surface lines of Azienda Trasporti Milanese (ATM) of Milan is presented. After discussing about the various types of disruptions that can be considered, we will focus on the way the regularity of the service can be assessed. This is one of the most critical points since, from the service provider point of view and also from the municipality or the agency monitoring the service perspective, the regularity of the service should be measured in the simplest and most intuitive way. However, the measure should be also of help when actions, intended to recover the regularity or improve it in the presence of disruptions, must be taken and their definition demanded to a decision support system.

2 Disruptions and measures of regularity

While the term *disruption* is widely used in the transportation literature, the definition of *disruption* as found in a dictionary is of little use for an automated decision support system: for example the Oxford reports: *"disturbance or problems which interrupt an event, activity, or process"*. Therefore, in this section, we classify the different types of disruptions that may occur in public transit, and for each type, we illustrate the effects on the whole transportation network. Later, we will discuss how we can react to each type of disruption, if possible.

A better definition of disruption is given in Clausen et al. (2010), where it is defined as "*a deviation from the original plan sufficiently large to require a substantial change in operations*". In this definition, two important concepts are considered: the **planned service** (or also called the *nominal* plan) produced by the off-line execution of an optimization algorithm, and the **observed service**. When the observed service is too different from the planned one, we say that we have a disruption. However, it is not the case that every deviation from the original plan requires a *substantial* change in operations.

An alternative and complementary term used in case of soft disruptions is *disturbance*. A line has a disturbance if several vehicles operating with limited delays are on duty. Disturbances can, in general, be reduced and solved by carefully holding, speeding up, or slowing down vehicles. Nevertheless, if the disturbance is not controlled, it can easily evolve into severe disruptions giving rise to the so called "snow-ball effect". Cacchiani et al. (2014) presents an overview on disruption problems arising in railway management.

2.1 Disruption identification: Evaluating the regularity of the service

The main challenge in managing disruptions is to be able to distinguish those events that may have a negative impact on the service regularity from those events whose negative effect is limited in time and in a small portion of the network that can be easily managed.

In order to develop an algorithm that proposes a course of actions to mitigate the effect of disruptions or disturbances, we need a precise and formal way to evaluate the regularity of the service along the line. The concept of regularity depends on the type of service offered by the public transport. We distinguish two different cases:

1. **Timetabled service**: At each stop of the line the precise arrival and departure time (timetable) is specified. In this type of service, frequencies are usually low, e.g. less than 6 vehicles per hour.

2. **Frequency based service**: At each stop a frequency of vehicle passages is specified instead of the timetable (e.g., a vehicle every 7 minutes). In this case frequencies are higher and so the headway is smaller. In this case the passenger might estimate the average waiting time at the stop. Note that, in this case, even if generalized delays are present, the service is not perceived as disrupted by passengers, provided that frequencies are regular and aligned with the planned ones.

In this paper we focus on frequency based services. Note that, in this type of service, even though users perceive the service delivered according to frequencies, resources, namely vehicles and drivers, continue to be managed on a timetable base, hence delays or other events which are not not perceived as disruptions by users, may generate disruptions when the resource management must be inevitably adjusted.

There are many ways to estimate the regularity of the service. The most regular service is obviously that is exactly mimicking the planned timetable. Thus the regularity measure should consider the adherence of the provided service with the planned one. In the case of timetabled service the measure will consider the planned timetable, while in the frequency based service this measure can be relaxed and only the headways will be accounted for. In the literature many proposals are present (see for example Barabino et al. (2013) for a brief survey). We can mention the following ones:

- HR Headway Ratio: percentage ratio between the observed headway and the planned one.

- HRI Headway Regularity Index: Gini's coefficient of the observed headways, giving a statistical measure of the dispersion of the headway values.

- SR Service Regularity: percentage of the observed headways below or above with respect to the planned one of a given threshold.

- EWT Excess Waiting Time: it accounts for the waiting times above a given tolerated threshold with respect to the average.

- HSD Headway Standard Deviation: standard deviation of the difference between the observed and the planned headways.

PRDM Percentage Regularity Deviation: percentage of the deviation of the observed headway from the planned one.

HA Headway adherence: it is a classification into subsets of the observed headways considering the standard deviation of the difference between the observed headway and the planned one using the average planned headway to define classes.

It can be observed that, most of the times, these measures are suitable when computed off line, when all data collected during the day are available. This is particularly true when the average, the standard deviation or other cumulative indicators of the observed value are used. However, this is not always possible when the measure has to be computed in real time to support decisions about disruption detection and possible recovery actions. In the following we will propose some alternative functions to be used to measure the regularity in real time.

For the sake of simplicity let us consider a circular line with n total stops. Our definition can be extended to more involved cases. Let n_p be the number of planned stops within a whole day of service, that is the number of stops of the line in both directions multiplied by the number of trips.

2.1.1 Timetable based index of regularity

We count the number of times n_b a vehicle arrives after a given threshold t with respect of the timetable at the stop (e.g., $\rho = 180 seconds$). These cases are informally called the *bad passes*. The index of regularity of the line can be computed as:

$$\left(\frac{n_p - n_b}{n_p} \times 100 \right) \% \qquad \text{(regularity)} \tag{1}$$

and expresses the percentage of regular passes over the total planned passes. Clearly, if the number of bad passes is zero, the regularity index is 100%, meaning that the planned service is always met. Note that this type of index does not consider the early passes which may be perceived even worse than delays in a timetabled service. In order to account for early passes, (1) can be adapted by considering, instead of n_b, n_b' as the number of times a vehicle passes by a stop either more that ρ seconds late or t seconds early. However if all vehicles are late or early of the same amount of time, greater than the threshold t, all passes are bad and the regularity index drops to 0%, even though passengers will not perceive such a dramatic irregularity.

In a real time setting, we can identify the vehicles that originated bad passes on a given line.

2.1.2 Headway based index of regularity: a simple proposal

In order to overcome the drawback emerging from the timetable based index of regularity pointed out above, instead of considering the timetabled passes we consider the headway, that is the inter-arrival time between consecutive vehicles of the same line at a given stop. We count the number of times n_b'' a vehicle arrives more than t minutes late with respect of the planned headway, these are called the *bad passes*. The index of regularity of the line is computed as:

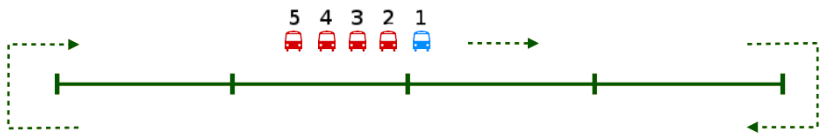
$$I(01) = \left(\frac{n_p - n_b''}{n_p} \times 100 \right) \tag{2}$$

and expressed as percentage of bad passes over the total planned passes. Clearly, if there are not bad passes the regularity is 100%, meaning the planned frequency is always met. This simple index of regularity $I(01)$ is the one currently used by ATM for internal purposes.

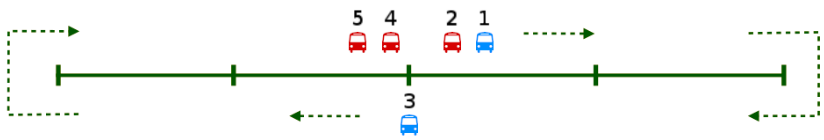
Notice that this type of index does not account for early passes. This feature is intentional, since, considering the actual headway, a vehicle anticipating its pass at a stop, while the other vehicles are maintaining their schedule, will result as an early pass with respect to the previous vehicle, but the following vehicle will result as having a bad pass considering the actual headway. Thus penalizing early passes as well as late ones (of the same amount), would penalize twice the same event. However, this type of index of regularity penalizes in the same way an almost good pass (i.e. a vehicle missing the regularity of only a few seconds) and a very bad one (i.e. a vehicle with a headway much higher that the threshold). While this could have some sense in a purely evaluation environment, without taking any action to improve the regularity, in a setting where the service provider wishes to increase the index of regularity, this type of measure may generate some pathological behaviors, as illustrated by the following example.

Example 1 *Consider the situation in Figure 1a where we have all vehicles grouped in a pack: the first (blue) bus is late, while the following four (red) buses are ahead of schedule, and give rise to early passes at all stops. With the current index, only the blue bus originates a bad pass, and the regularity of service will be around 80%, which, in general, is not bad. Nevertheless, the situation depicted in Figure 1a is clearly unacceptable. In addition, if the operators try to improve the regularity by detouring a (red) vehicle of the pack, as shown in Figure 1b, the average waiting time of users decreases, but the index of regularity worsen since the detoured vehicle would generate new bad passes, which is clearly a paradox. Hence, the use of such an index in an automated system that seeks to improve the index of regularity, in the presence of disturbances, would make the service converge towards situations as that depicted in Figure 1a, and from where it is impossible to escape by applying single detour actions.*

Another anomaly can come from the fact that (2) relies upon the planned headway. Although it can be extremely rare, suppose that the line experiences a generalized delay, sufficient to increase the actual headway of an amount of time slightly greater that the threshold t. In the case of a line having a roundtrip time of 60 minutes and 5 vehicles, giving rise to a planned headway of 12 minutes. Considering a threshold $t = 180$ seconds, an increased roundtrip time of 15 minutes would imply that even if all vehicles are evenly spaced, all headways are



(a) The first bus is late; the other four buses are travelling ahead of schedule.

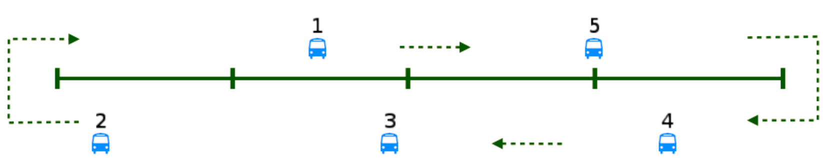


(b) The current index of regularity will decrease if the 3-rd bus were to turn at mid-line.

Figure 1: Anomalies in the index of regularity

above the threshold. The condition of the line becomes that depicted in Figure 2a, where every (blue) bus is equally late of about three minutes and gives rise to a bad pass. Clearly, from the passenger perspective, there is little difference with the regular service, nevertheless, the current index of regularity would be equal to 0. Moreover, note that, as mentioned above, from this situation any detour that moves a vehicle just behind another one, thus spoiling the equally spaced setting, would increase the index of regularity.

In order to overcome the anomaly coming from a generalized delay, we could modify the way a bad pass is defined by considering, instead of the planned headway, the expected one computed as the average of the observed headways or as the average roundtrip time divided by the number of vehicles operating the line.



(a) All five buses are equally late, but the frequency at the bus stops is regular.

Figure 2: Generalized delay

2.1.3 Piecewise linear function for evaluating the index of regularity

In order to overcome the drawbacks emerging from the previously described methods for evaluating the index of regularity, we can consider a function $f(x)$

of the gap x between the planned headway and the observed one. Denoting by h_p the planned headway and by h_o the observed one, the gap can be considered either in an absolute way:

$$x = h_o - h_p \tag{3}$$

or in a relative way:

$$x = \frac{h_o - h_p}{h_p} \tag{4}$$

In this second way, a gap on a line with high frequency will have more impact than the same gap on a low frequency line. Notice that in both cases a negative value of x means an early pass, while a positive value means a delay.

The function used to define the index of regularity can be described as follows:

$$f(x) = \begin{cases} -\alpha x & \text{if } x < -\theta_1 \\ 0 & \text{if } -\theta_1 \leq x < \theta_2 \\ \beta x & \text{if } \theta_2 \leq x < \theta_3 \\ \gamma x + \delta & \text{if } \theta_3 \leq x \end{cases} \tag{5}$$

where $\theta_1, \theta_2, \theta_3 (> \theta_2)$ and $\alpha, \beta, \gamma (> \beta), \delta = \beta\theta_2$ are suitable parameters. If we want to ignore the contribution of earliness on the index of regularity it suffices to set $\alpha = 0$. A negative value of $x(q)$ means an early pass, while a positive value means a delay.

Considering P as the set of all passes that occurred in the observed period, the index of regularity based on the piece-wise function ($I(PW)$) is defined as follows:

$$I(PW) = \sum_{q \in P} f(x(q)). \tag{6}$$

2.1.4 Non linear function for evaluating the index of regularity

In order to account for lateness (and earliness) in the index of regularity more than linearly, so that a large lateness with respect to the planned headway will be penalize more than the equivalent sum of small delays, we can consider a quadratic penalty function as follows:

$$f'(x) = \begin{cases} \eta' x^2 & \text{if } x < \delta' \\ \eta'' x^2 & \text{if } x \geq \delta'' \end{cases} \tag{7}$$

The distance $|\delta' - \delta''|$ defines a tolerance zone in which vehicles that travel not later than δ'' and not earlier than δ' are not penalized. Parameters $\eta', \eta'', \delta', \delta''$ must be selected in such a way that the resulting index of regularity measure is as close as possible to the current measure on the same data, and this can be done by standard data fitting on the historical data. The index of regularity based on the quadratic function ($I(QA)$) can be defined as follows:

$$I(QA) = \sum_{q \in P} f'(x(q)) \tag{8}$$

2.2 Parameters setting

The proposed regularity indices based on piecewise linear (PW) or quadratic (QA) functions have several parameters that need to be carefully set up. In order to find a suitable parameter setting, a preliminary two-step test phase has been conducted. In the first step, the parameters affecting the slopes of the functions, namely α, β and γ for PW and η^1 and η^2 for QA are taken into account. In the second phase, once the slopes have been selected, the interval parameters are tested, namely θ_1, θ_2 and θ_3 for PW and δ_1 and δ_2 for QA. In the first step, six different parameters configurations (from A to F) have been taken into account and are detailed in tables 1 and 2, for PW and QA respectively. In both tables settings A have been suggested by ATM.

Table 1: PW Settings

Conf.	α	β	γ	δ
A	0.002	0.007	0.002	0.400
B	0.010	0.001	0.001	0.400
C	0.002	0.025	0.002	0.800
D	0.050	0.002	0.080	1.000
E	0.002	0.002	0.100	1.500
F	0.001	0.010	0.020	2.000

Table 2: QA Settings

Conf.	η'	η''
A	10^{-6}	30^{-6}
B	10^{-5}	30^{-6}
C	10^{-6}	10^{-5}
D	10^{-7}	10^{-5}
E	10^{-5}	10^{-7}
F	30^{-6}	10^{-5}

The average results obtained on all available instances are reported in tables 3, 4, 5 and 6. The percentage improvements $\%I$ cannot be directly compared since the numerical value of both IR_0 and IR^* strictly depends on the value of the parameters. However, it is worth noting that large improvements in the quality of the service are achieved with all parameter settings. Columns $Avg\%I$ represent an indication of the overall quality of the solution found. In particular, if $Avg\%I$ is negative then the quality of the solution found is questionable since at least one index of regularity is lower in the optimized solution than in the starting situation. Based on these considerations, settings B for the PW and E for the QA penalty function can be already discarded since they generate poor quality solutions. These two settings are also characterized by a lower than average number of non improving solutions and a higher than average number of actions required per solution. This is due to the fact that with these settings, situations characterized by an already good initial quality of the service are unnecessarily optimized using a lot of actions to achieve negligible improvements. Regarding the cost in the implementation of the proposed solutions, three main components have to be taken into account: KM-Ac, KM-D* and, to a smaller extent, $Extra^*$. Small values of these parameters characterize solutions with a low impact on the implementation cost. The setting C for the PW penalty function and setting B for the QA can then be discarded due to the high implementation cost. Finally, columns %D-D* and %M-D* give and indication of the difficulty of implementing a solution and the degree of acceptance of drivers. All remaining parameter settings offer a similar level of performance. However, in order to carry out the test phase, setting E for the PW and setting A for the QA penalty functions are selected since they offer a good compromise between quality and cost of the proposed solutions.

Table 3: PW Settings - Average results on all instances

Conf.	%I	Avg %I	%No	N Ac	KM-Ac	V T
A	9.506	2.984	5.581	5.811	6.746	13.779
B	58.787	0.806	3.256	6.225	14.960	14.013
C	119.430	1.800	5.116	5.912	10.218	13.815
D	46.106	3.586	6.046	5.893	6.125	13.568
E	126.696	4.667	6.046	5.781	4.669	13.641
F	224.386	4.168	5.581	6.065	4.141	13.719

Table 4: PW Settings - Average results on all instances

Conf.	KM-D*	%D-D*	%M-D*	Extra*	D T
A	3.343	0.991	7.109	882.988	17.993
B	4.977	1.270	8.387	1007.786	23.875
C	4.829	1.436	9.900	1066.847	22.504
D	2.827	0.818	6.206	779.054	21.320
E	2.385	0.826	6.645	888.580	16.720
F	2.953	0.841	7.067	936.124	23.349

Table 5: QA Settings - Average results on all instances

Conf.	%I	Avg %I	%No	N Ac	KM-Ac	V T
A	3.739	4.689	4.651	6.111	5.251	13.517
B	19.381	4.047	3.256	6.052	9.322	13.568
C	10.055	4.643	4.651	6.281	4.760	13.644
D	9.401	4.572	4.651	6.386	4.772	13.639
E	7.297	-0.446	0.465	7.128	21.900	13.900
F	13.232	4.694	4.651	6.029	5.234	13.609

Table 6: QA Settings - Average results on all instances

Conf.	KM-D*	%D-D*	%M-D*	Extra*	D T
A	2.006	1.018	7.154	857.339	15.711
B	3.661	1.048	7.780	892.724	22.828
C	2.216	1.035	6.934	860.696	14.828
D	2.298	0.968	6.721	897.152	19.018
E	6.289	1.386	10.986	1214.478	29.644
F	2.187	1.027	7.423	854.649	16.980

Table 7: PW Settings with configuration E (seconds)

Conf.	θ_1	θ_2	θ_3
E_0	120	150	210
E_1	120	120	180
E_2	60	180	300
E_3	120	120	300
E_4	120	60	240
E_5	180	120	240
E_6	60	60	150

Table 8: QA Settings with configuration A (seconds)

Conf.	δ_1	δ_2
A_0	120	120
A_1	60	60
A_2	180	180
A_3	60	120
A_4	120	60
A_5	180	120
A_6	120	180

In the second step, seven different configurations (from 0 to 6, tables 7 and 8) for the parameters defining inflecting points of the regularity functions are taken into account. The average results obtained on all available instances are reported in tables 9, 10, 11 and 12. Following the same evaluation structure as in the first step the configurations E_3 and A_0 are selected as best parameter configurations to be used for the computational analysis and the comparison with other penalty functions.

Table 9: PW Settings with configuration E - average on all instances

Conf.	%I	Avg %I	%No	N Ac	KM-Ac	V T
E_0	126.696	4.667	6.047	5.781	4.669	13.641
E_1	241.984	3.469	4.186	6.316	4.942	13.658
E_2	69.515	3.523	3.721	6.382	4.584	15.616
E_3	56.478	3.092	3.256	6.356	4.351	13.874
E_4	908.719	3.010	0.465	7.950	3.744	13.990
E_5	282.558	3.172	4.186	6.314	4.494	13.868
E_6	103.821	3.091	0.930	8.078	4.385	14.036

Table 10: PW Settings with configuration E - average on all instances

Conf.	KM-D*	%D-D*	%M-D*	Extra*	D T
E_0	2.385	0.826	6.645	888.580	16.720
E_1	2.228	0.904	6.962	879.275	14.829
E_2	3.347	0.841	8.101	913.191	21.338
E_3	2.017	0.970	7.016	888.914	17.299
E_4	2.044	0.879	8.802	967.772	20.237
E_5	2.436	0.889	6.903	934.006	21.836
E_6	2.855	0.947	9.764	960.972	24.857

Table 11: QA Settings with configuration A - average on all instances

Conf.	%I	Avg %I	%No	N Ac	KM-Ac	V T
A_0	3.739	4.689	4.651	6.111	5.251	13.517
A_1	3.527	4.506	0.465	8.061	4.756	14.052
A_2	4.247	4.706	13.023	5.183	6.131	13.600
A_3	3.634	4.624	1.860	6.938	4.904	14.066
A_4	3.536	4.496	0.930	7.821	4.849	13.893
A_5	4.092	4.659	9.767	5.888	5.925	13.902
A_6	3.817	4.738	6.512	5.754	5.006	13.577

Table 12: QA Settings with configuration A - average on all instances

Conf.	KM-D*	%D-D*	%M-D*	Extra*	D T
A_0	2.006	1.018	7.154	857.339	15.711
A_1	2.411	0.867	9.473	971.089	25.964
A_2	2.843	0.897	5.979	894.810	19.749
A_3	2.972	0.965	8.539	882.740	22.452
A_4	2.559	0.927	8.852	918.883	24.404
A_5	3.050	0.855	6.743	896.831	22.816
A_6	1.970	1.043	7.450	891.659	15.549

3 Disruption-delay management framework

Currently the operations central officers take their decisions about recovering the regular service relying upon their expertise. The actions that they can enforce to drivers are basically unplanned detours and shortening (or skipping) breaks at terminals. In certain locations along the line, it is also possible to hold the vehicle, if it is on duty. The effects of these actions on the service regularity are evaluated only according to their experience and intuition, and it is almost impossible to asses the impact of alternatives.

However, once a set of indices to evaluate the regularity of the service has been defined, with the support of an automated system is possible to estimate the effect of each action using a stochastic simulation based on the empirical distributions obtained with an analysis of the historical data. The objective of this simple localized (in space and time) simulation is to evaluate the effect of detours or other changes in the service.

In order to feed the simulator with the complete information, all possible actions must be identified for each line, together with the time needed to perform each of them, in particular for the first type of action.

The simulation is intended to evaluate the effect of the actions on the service within a sufficiently wide interval of time. To this aim, we compare the regularity index output by the simulation when some actions are made with that coming from the simulation when no action is carried out. Since the setting is stochastic, the simulations are repeated several times in both cases, in order to have a reliable evaluation. A set of actions should be discarded if in every simulation did not give an improvement on the regularity index. It could be conditionally accepted, if it gave an improvement on a significant number of simulations; while it is accepted if it always gave an improvement. In any case, the final decision whether to accept or not actions is left to the officer.

Once one or more actions that increase the regularity of service in the simulation have been identified and accepted by the officer, the driver scheduling is adapted. This is done by solving an optimization problem with the objective of minimizing the changes with respect to the planned duties and the minimizing uncovered service. The description of the overall working framework of the delay/disruption management system can be found in Carosi et al. (2015).

3.1 Simple discrete event simulator

The simulator routine is based on a set of events stored in a priority queue Q. The events are identified by triplets (vehicle, position, time). The routine considers a simulation interval Δ starting from the current moment t_0. Initially Q contains the last AVM observation for each vehicle currently on duty. In addition, the events corresponding to new vehicles entering the service within the simulation interval are inserted into Q. Their position is the entering stop and their time is the planned one. The routine iteratively extracts form Q the oldest event and generates a new event with the same vehicle, its new position and its new time. Position and time are randomly generated by using the travel time distribution as discussed in section 4. If the time of the new event is within the simulation interval, it is inserted in Q, otherwise it is discarded. The routine stops when Q is empty.

Within the simulation routine the indices of service regularity are updated according to the functions defined in the previous sections.

3.2 Selection of the optimal actions via a tabu-search procedure

Ideally, the set of actions maximizing the improvement of the regularity index should be identified. Since the evaluation of this objective requires a time demanding simulation phase, we decided to approach the problem by means of a tabu-search procedure Glover and Laguna (1997). The tabu-search starts from the current vehicle and crew scheduling (V and C) and explores neighboring solutions by trying all possible actions selecting them from a set $\mathcal{A}$.

The output of the tabu-search is a subset $\mathcal{W}^* \subset \mathcal{A}$ of actions, which, if applied, will allow with high probability to improve the index of regularity obtained without applying any action (IR_0).

The procedure is summarized in Algorithm 1. It starts with an initialization phase (lines from 2 to 5) where both the set of proposed actions $\mathcal{W}$ and the tabu list $\mathcal{F}$ are initialized with an empty set. Then the expected index of regularity without applying any action IR_0 is computed (line 4) using the event simulator procedure described in section 3.1. The simulation is carried out considering the current status (queue Q). The parameter $\mathcal{A}$ represents the set of actions that must be applied. In the main part of the algorithm (lines from 6 to 18), the impact of all available actions is assessed exploiting the same event simulator. In particular, in each iteration $i \in \{1, \ldots, k\}$, starting from the current set of actions $\mathcal{W}$, the complete neighborhood is explored (lines from 9 to 13). The neighborhood space is obtained removing or adding a single action to $\mathcal{W}$. The new $\mathcal{W}$ is equal to $\{\mathcal{W} \cup a\} \setminus \{\mathcal{W} \cap a\} \forall a \in \mathcal{A} \setminus \mathcal{F}$. At the end of the ith iteration the action a_i^* leading to the best solution in the neighborhood is stored and both the tabu list $\mathcal{F}$ and the current status $\mathcal{W}$ are updated (lines 14 and 15). The action a_i^* is added to $\mathcal{F}$ and all elements in $\mathcal{F}$ by more than f_{max} iterations are removed while $\mathcal{W}$ becomes $\{\mathcal{W} \cup a_i^*\} \setminus \{\mathcal{W} \cap a_i^*\}$. If the index of regularity IR_i^* associated with the new set is better than the best index of regularity found so

Algorithm 1 Tabu-search procedure

1: **procedure** TABU-SEARCH($\mathcal{A}, V, C, Q$)
2: $\quad \mathcal{W} \leftarrow \emptyset$ ▷ Initialization phase
3: $\quad \mathcal{F} \leftarrow \emptyset$ ▷ Forbidden moves
4: $\quad IR_0 \leftarrow simulate(V, C, Q, \emptyset)$ ▷ Simulation without applying any action
5: $\quad IR^* \leftarrow IR_0$
6: $\quad$ **for all** $i \in \{1, \ldots, k\}$ **do** ▷ Main loop
7: $\qquad IR_i^* \leftarrow \infty$
8: $\qquad a_i^* \leftarrow \emptyset$
9: $\qquad$ **for all** $a \in \mathcal{A} \setminus \mathcal{F}$ **do** ▷ Neighborhood exploration
10: $\qquad\quad IR_i^a \leftarrow simulate(V, C, Q, \{\mathcal{W} \cup a\} \setminus \{\mathcal{W} \cap a\})$ ▷ Simulation applying actions
11: $\qquad\quad$ **if** $IR_i^a < IR_i^*$ **then** ▷ Local best solution
12: $\qquad\qquad IR_i^* \leftarrow IR_i^a$
13: $\qquad\qquad a_i^* \leftarrow a$
14: $\qquad \mathcal{F} \leftarrow \mathcal{F} \cup a_i^* \setminus \{a_j : a_j \in \mathcal{F} \wedge i - j > f_{max}\}$ ▷ Update tabu list
15: $\qquad \mathcal{W} \leftarrow \{\mathcal{W} \cup a_i^*\} \setminus \{\mathcal{W} \cap a_i^*\}$ ▷ Update the current set of actions
16: $\qquad$ **if** $IR_i^* < IR^*$ **then** ▷ Global improving solution found
17: $\qquad\quad IR^* \leftarrow IR_i^*$
18: $\qquad\quad \mathcal{W}^* \leftarrow \mathcal{W}$
19: $\quad$ **return** $IR^*, \mathcal{W}^*$ ▷ Best subset of actions

far IR^* then the best solution $\mathcal{W}^*$ is updated as well (lines from 16 to 18). At the end of the procedure the best subset of improving actions W^* is returned (line 19).

4 Historical data analysis

In order to set up the simulation tool, a thorough analysis of the historical and real-time data has been carried out. In particular, we analyzed the observed travel times and their relation with headways in comparison with the planned ones. To give an idea we report here a few plots that illustrate the possible deviations of the real service from its nominal plan. The data that we report below refer to the travel times between two stops of Line 92, direction Bovisa (first stop is viale dei Mille, while the second stop is viale Abruzzi).

Figure 3 shows the distribution of the travel times of the 3718 trips whose planned time is 5 minutes (300 seconds). While the overall average travel time is quite close to the planned time (the average is 320 seconds, equal to 5 minutes and 20 seconds), the distribution shows that in practice there is a number of trips that take 4 minutes or less, and that there exists also a large number of trips that last more than 6 minutes. In addition, the standard deviation is also a poor indicator of the service, since these data have a standard deviation of 51 seconds, but the distribution is clearly asymmetric.

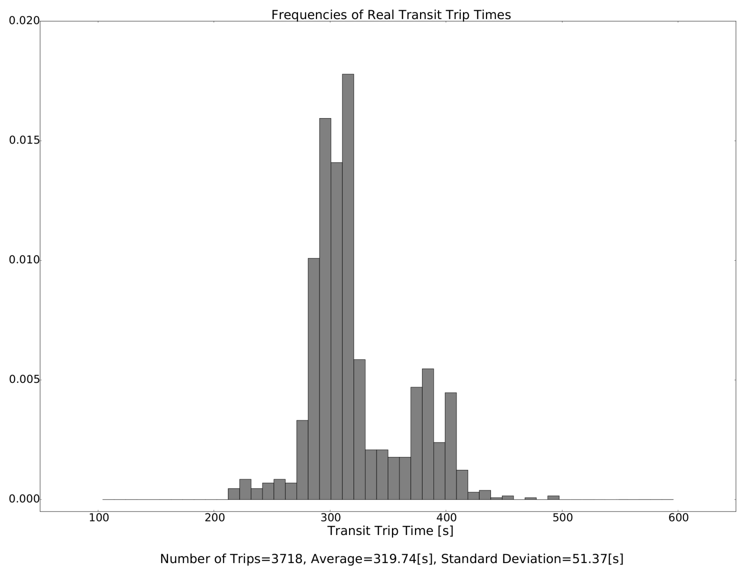


Figure 3: Distribution of the travel times (ATM, Line 92)

Similar asymmetric behavior can be observed for tram lines, which however is more regular since usually trams travel on dedicated lanes for most of their itinerary.

In our *simulation-based evaluation algorithm*, we have decided to use directly the empirical distribution that uses a mix of the distribution of historical data (same weekday, same time of the day and same whether conditions) and of real-time data, to better capture possible variations.

Another issue addressed in our analysis is the relation between travel times and the gap in the expected headway, due to the possible increased dwell time.

Figures 4 and 5 show the values observed on line 92. The first plot is obtained with the same dataset used for the analysis of the travel time distribution; the Pearson's coefficient computed on this dataset is equal to 0.02 and no clear correlation between travel times and headway variation can be observed, indeed the majority of the observations are concentrated around 300[s] independently of the headway. The second plot is obtained taking into account all pairs of consecutive stops of line 92 and uses relative values: on X-axis the ratio between actual and expected headway is reported while the ratio between actual and expected traveling time is used on Y-axis. It shows a similar behavior as in the Figure 4 (Pearson's coefficient equal to -0.01).

From the empirical analysis of the plots and the correlation indices we can conclude that in our cases there is no relationship between the two values and in our simulation tool we decided to omit the effect of the dwell time.

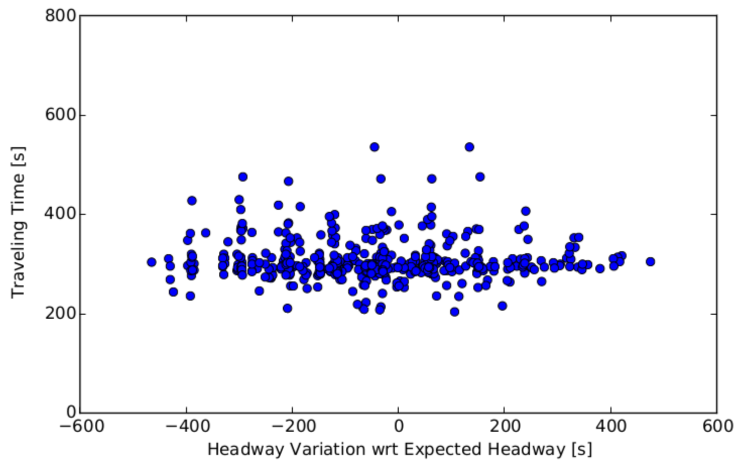


Figure 4: Dwell time impact, two stops (ATM, Line 92)

References

B. Barabino, M. Di Francesco, and S. Mozzoni. Regularity diagnosis by automatic vehicle location raw data. *Public Transport*, 4:187–208, 2013.

V. Cacchiani, D. Huisman, M. Kidd, L. Kroon, P. Toth, L. Veelenturf, and J. Wagenaar. An overview of recovery models and algorithms for real-time railway rescheduling. *Transportation Res. B*, 63:15–37, 2014.

S. Carosi, F. Malucelli, , and E. Tresoldi. Delay management in public transportation: Service regularity issues and crew re-scheduling. *Transportation Research Procedia*, 10:483 – 492, 2015.

J. Clausen, A. Larsen, J. Larsen, and N. J. Rezanova. Disruption management in the airline industryÑconcepts, models and methods. *Computers & Operations Research*, 37(5):809–821, 2010.

F. Glover and M. Laguna. *Tabu Search*. Kluwer Academic Publishers, 1997.

F. Malucelli, , and E. Tresoldi. Delay and disruption management in local public transportation via real-time vehicle and crew re-scheduling: a case study. *Public Transport*, 2018. Manuscript submitted for publication.

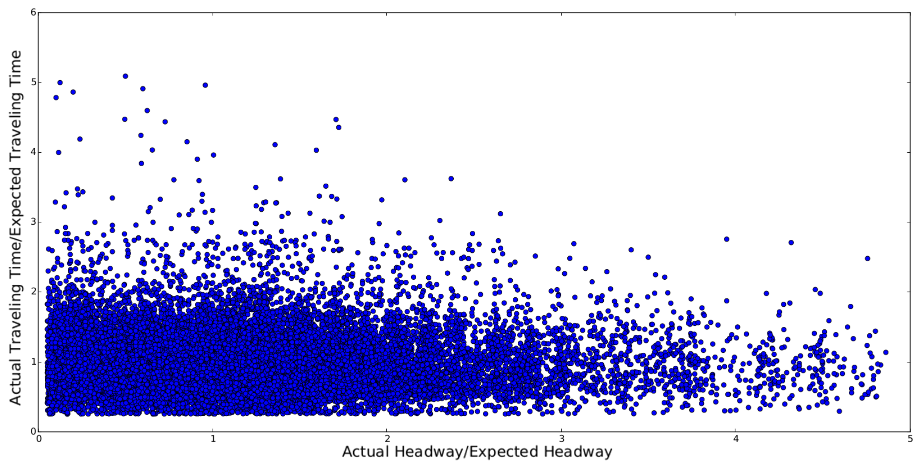


Figure 5: Dwell time impact, all stops (ATM, Line 92)