Automated analysis of train event recorder data to improve micro-simulation models

S. de Fabris, G. Longo & G. Medeossi
Department of Civil and Environmental Engineering, University of Trieste, Italy

Abstract

In recent years, on board digital train event recorders have been developed: these devices allow the collection of very detailed data about train movements and signal status. The new Italian ATC SCMT on board subsystem is combined with the DIS (Driver Information System), which collects both kinetic behaviour and all signal and balises messages. Unfortunately, this large amount of data is normally stored, but not used except for failure and maintenance management. At the same time the use of micro simulation tools has been extended to large scale problems. As is known, a problem exists in the calibration and validation of these models.

In this paper a new tool is presented. This tool allows the analysis of real-life collected data, to perform very detailed analysis of train movements, pointing out speed depending on position and signal aspects, acceleration, braking curves and dwell time graphically and by means of parameters. Train behaviour can also be connected to punctuality, to find out differences between on time and late running. This tool may be very useful for large scale model validation, definition of the stochastic behaviour of the system (travel time, dwell time, initial delay), calibration of braking and acceleration curves for various train types and acceleration percentage depending on different conditions. In other words, it allows a link to be set up between real data and micro simulation models. The tool has been tested in the north-eastern part of Italy. In this case study a significant precision increase in the stochastic simulation results has been reached.

Keywords: railway simulation, railway planning, train event recorder.
1 Introduction

An efficient train operation is a primary success factor for all infrastructure managers since it allows operating a higher number of trains without significant infrastructure investments. In the last few years, railway simulators have become a very powerful instrument to support the different steps of the planning process: from the layout design to capacity investigations, and offer model validations. More recently, the possibility of an automatic import of infrastructure layouts and timetables has widened the application spectrum of micro-simulators to large nodes and to more detailed stochastic stability evaluations.

Stochastic micro-simulators perform very detailed robustness analysis, considering all processes involved in rail traffic and comprehending not only its deterministic aspects, but also human factors and other stochastic phenomena. This is particularly relevant in order to simulate traffic under realistic conditions, considering variability at border (initial delay), various driving styles and stop times. All these parameters have to be calibrated using real-world collected data for single trains or train families, considering their different behaviour in the network and at its border.

Infrastructure operators have developed various tools to analyze these operational data, considering all track circuits or only timetabling points. In both cases trains are checked at discrete points, not allowing a continuous reconstruction of train movements. In this way, on-board collected data represent an ideal input to investigate the behaviour of trains, from acceleration to braking and precise stop times.

Stochastic micro simulation is being widely introduced in Italy by RFI, the infrastructure manager of most lines, to evaluate timetable robustness in future scenarios and during infrastructure improvement works in the busiest nodes, like Rome, Florence, Milan, Venice, Naples and Bologna. To improve the precision of input parameters a new tool has been developed that combines both timetabling point data and the on-board collected data, coming from the new-generation train event recorder recently installed on all trains running on RFI network.

2 The driver information system (DIS)

To improve rail traffic security by a continuous collection of train behaviour data, in 1998 the Italian Ministry of Transportation decided to fund the development of a new generation train event recorder to be installed on all locomotives, trains and driving coaches. Prior requirement of the system was a very high precision in all records, comprehending not only speed and odometer, but also messages from the balises, GPS location and a large amount of data coming from the vehicle-bus. After an exhausting test and validation phase, the DIS has been installed since 2004, and is now fully operating on most lines.

The system is formed by an on-board subsystem that collects data and sends them automatically to the depot subsystem when it passes by one of them, data are stored in a file system archive, accessible from authorised operators all over
Italy. The on board subsystem is formed by a speedometer, an event recorder, a communication computer and an IEEE 802.11 transmission protocol radio and a simplified user interface.

Using data collected from two odometers, the speedometer calculates the speed and the covered distance, taking into account the real wheel diameter. These data are continuously sent to the event recorder. Before starting a new service, drivers must insert their smartcards into the MMI and type train number, speed type, train length and braking weight percentage. These are used for both data collection and braking curve calculation in the ATC subsystem.

Data coming from the speedometer, the GPS, the user Interface and the vehicle MVB Bus, coupled with GMT timing are stored in a solid-state-drive every 100 ms. When a train reaches a given GPS position nearby a depot, the communication computer starts a data link with the server and uploads the past files.

3 The “TRENO” tool

The large amount of data stored is actually only used by Trenitalia as a failure archive and for maintenance, or in the event of accidents. With TRENO (Timetable Reliability and Network Operations analyser), these archives may be used as very precise support to understand rail traffic.

The tool is based on a POSTGRES SQL DBMS; it integrates track occupation data coming from the Infrastructure Manager RFI, the planned timetable of different scenarios and the DIS recordings.

This integration allows a macroscopic analysis of train traffic and a microscopic focus on selected trains or stations; the planned timetable of the whole network is used to select different days and lines and as a basis for the export to Opentrack micro-simulator.
Figure 2: Data flows to and from TRENNO.

RailML stochastic simulation files can also be imported in the tool as new scenarios that can be stored and be analysed with the Macro-Analyser, in order to evaluate the robustness of different timetables, or the advantages of infrastructural improvements. Figure 2 shows TRENNO’s structure.

The “Planned” module allows one to import or create planned timetable scenarios. RFI data containing whole Italian network timetable can be coupled with user-defined scenarios. A simple Infrastructure model is also contained to support corridor analysis and to allow running time calculations. Timetables can be exported as RailML file for any line section and day/time.

In the “Macro” module, conventional departure, travel time and stop time distributions can be plotted and analysed. Moreover their parameters may be calculated. Graphical timetable with various display options is another possible output. Simpler parameters like mean values, piecewise linear distributions or percentage areas are also shown to be used in micro simulation.

Mean registered stop-time supplement to the planned one is automatically added to the RailML planned timetable file for each stop of the single train to be imported in the micro-simulator. Maximum supplement for the train family is estimated, since in the used micro-simulator stop time variability has at the moment no other inputs option than the stochastic combination of these parameters.

The module is completed by a critical element search function that allows one to select trains, stations and line parts on the basis of given delay parameters. The “micro” module is widely presented in the next paragraph.

4 Microscopic analysis

The real life registered files can be used for a very wide range of detailed analysis of train movement, also considering signal status and system failures.
The very high precision of stored data and the large amount of additional information require a filtering process to point out failures and traffic perturbations that have influenced a train movement. Moreover, the 100ms interval between measurements has to be reduced to 1 second in order to obtain faster processing and easier data aggregation.

This filtering process is performed automatically when files are imported in the database; failures are also stored in different tables as well as balises messages reporting signal status; speed limits and ATC-calculated braking curves are also archived to enable analysis in the event of maintenance speed reductions. ATC balises messages enable considering primary and secondary delays separately, since restrictive signal status are pointed out. At the moment, these have not been used to perform delay propagation analysis, since the study has focused on the behaviour of single trains or train families and not on traffic as a whole. In other words, excluding trains with failures or with non-free trajectories in the focus area is a basic principle to extract train trajectories as allowed by infrastructure and timetable and as required to calibrate motion equation parameters.

The module can analyze stop times at each station, acceleration, full speed and braking phases separately. For each focus various two and three dimensional diagrams are provided, including statistic parameters representation or fitted density curves.

4.1 Stop times

The use of event recorder data allows a precise definition of stop time from the speed profile; door opening and closing is also stored to estimate the train preparation time. A separation between punctual and delayed trains is also possible, to point out both minimum dwell time and departure imprecision, often significant in Italy. All distributions can be viewed in diagrams with optional aggregation intervals. Statistic parameters are calculated, and negative exponential, Weibull, log-normal and gamma density fits are performed.

4.2 Acceleration and full speed behaviour

Acceleration is investigated by choosing a station, a trainset and a time/date interval and then by loading relative data on a diagram. Time speed, distance, time and acceleration can be selected on x- and y-axis; each single day train running is shown. To represent distributions better, these are shown in three-dimensional diagrams.

Since an automatic fitting has not been provided yet, the software is endowed with a running time calculator that integrates motion equation calculating the best acceleration curve based on the maximum tractive effort $F$ for the train. This is reduced to its measured values by two coefficients, obtaining a new $F_r$:

$$F_r = F \times f_1 \times f_2$$  \hspace{1cm} (1)
The curve based on $F$ can be graphically fitted as upper bound of real ones, obtaining the best performance rate $f_1$. Once this first value has been fixed, its variability as piecewise linear probability function is calculated, obtaining a second performance factor $f_2$, on the basis of running time distribution to cover a distance D. D and $f_2$ intervals are defined by the user, the corresponding time values $t_i$ are calculated automatically. The resulting parameters and their distribution are stored in a table and compared for all stations. Acceleration analyser window is shown in Figure 3.

![Acceleration analysis graph](image)

**Figure 3:** Acceleration analysis.

Full speed behaviour is also investigated in a similar module. In this case, two stations or kilometre points must be chosen; this allows representing longer line sections. Realistic trajectories are obtained with $f_1f$ and $f_2f$ parameters to scale the maximum speed. Coefficients are calculated as in the acceleration module; and saved in the same table with acceleration factors to support the choice of parameters to be used in micro-simulation.

### 4.3 Braking behaviour

In both running time and ATC braking curve calculation algorithms, only the BWP is used as input. In TRENO, braking behaviour of a train type can be aggregated and fitted to a linear function depending on BWP. To reproduce the real train behaviour better, a braking rate table can be defined.

Braking actions can be analysed in a window similar to the ones used for acceleration and full speed; $f_{1b}$ and $f_{2b}$ parameters similar to $f_1$ and $f_2$ are calculated as braking performance parameters. Otherwise, using $f_1$ and $f_2$, BWP
or the braking rates are calculated. In this case, $f_1$ and $f_2$ can be used as performance factor for the complete motion modifying train braking formula.

5 A case study: Trieste – Venice

TRENO has been used to analyze the traffic in north-eastern Italy, on the Trieste – Venice line. The double track, electrified line is about 130 km long and plays an important role in regional transport and as a freight corridor between eastern Europe and Italy. Figure 4 shows a simple layout of the line; the analyzed network was extended to approximately 30 km on each branch line, obtaining totally 280 km line.

After performing the macroscopic analysis, the study has focused on evaluating behaviour of trains in non-disturbed conditions, to estimate a standard performance percentage and a piecewise linear braking function for every train type. Performance variability has also been defined as piecewise linear by fitting it to real distributions, as well as initial delay for every train family. Since DIS data were not available before starting the study, on-board registered GPS speed-time data have been used, for testing purposes. Stop time, represented by a mean time and a variability range, as defined in RailML standard, has been automatically exported for every train and station.

![Figure 4: Simple topology of the Venice – Trieste line layout.](image)

Table 1: Comparison between stochastic simulations. Calibration parameters source: Scen. 1-4 timetabling points, Scen. 5 Collected data.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Stochastic Parameters</th>
<th>DIS calibrated Braking</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial Delay</td>
<td>Running Time</td>
<td>Stop Time</td>
</tr>
<tr>
<td>1</td>
<td>Yes (neg. exp)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>Yes (neg. exp)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Yes (neg. exp)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>Yes (pie.lin.)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Yes (pie.lin.)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
The results of 100 stochastic simulations have been analyzed with the Macro tool and compared real data referred to 100 days. Different simulation scenarios have been considered with growing calibration precision to point out the importance of various distributions and of micro and macroscopic analysis. In the first scenario only departure delay as negative exponential function per train family have been considered, then, in the second scenario, performance variability as empirically evaluated by timetable planners has been added. In a third phase stop time variability has come and in a fourth one departure delays have been modelled using a piecewise-linear function.

Scenario results have been compared using three simple punctuality parameters, similar to quality indicators commonly used in performance measurements. For each train, the difference between the percentage arriving at each station within 1, within 3 minutes and the mean delay in real and simulated data has been calculated; the mean differences on all trains and for each indicator are shown in table 1. A significant precision increase has been obtained in each scenario, showing that:

- Negative exponential initial delay can be used only as first step;
- Departure variability at all stations is the most important parameter for stopping trains; since it’s often very different for various services and train types the automatic import function is very useful;
- Running time variability is important for long distance and freight trains;

On-board collected data effectively allow a better estimation of stop type and train performance parameters; even better results could be obtained by a re-calibration of some more specific resistance factors. In fact trains accelerate with lower performances than expected and then they run very close to the speed limit, forcing to choose mean parameters that do not represent both situations at best. Moreover, some trains accelerate faster than estimated with the running time calculator, but brake with no more than 50% of their theoretic maximum. (Fig. 5, 6)

![Figure 5: Recorded speed v(s) profile at departure compared to calculated curves with different performance factors (smooth curves).](image-url)
Figure 6: Recorded speed profile $v(t)$ at arrival compared to calculated curves with different performance factors (straight lines).

6 Conclusions and further research

TRENO has been developed to perform both macro and microscopic analysis of train traffic, as system investigation support and as semi-automatic calibration tool for many simulation parameters. On-board recorded data are the key element to increase precision of traffic studies, pointing out the behaviour of drivers and trains under real conditions.

The case study has shown a significant improvement in the stochastic micro simulation output, demonstrating the importance of a precise parameter calibration especially regarding train braking and acceleration. A further improvement will come from the calibration of the motion equation parameters to fit the DIS data.

The tool is being continuously upgraded, to make it easier to use and to improve statistic analysis on DIS data and to analyse complete train trips. New functions will also be inserted to enlarge the amount of used data. This regards both driving parameters coming from the vehicle bus (e.g. throttle, brake use) and signal status, to study driver behaviour in presence of restrictive signals and to calculate block occupancy.

References


