Train speed profiles optimization using a genetic algorithm based on a random-forest model to estimate energy consumption

Ahmed Amrani, Amira Hamida, Tao Liu, Olivier Langlois

VIENNA 2018 TRA
A digital era for transport
solutions for society, economy and environment

Proceedings of 7th Transport Research Arena TRA 2018, April 16-19, 2018, Vienna, Austria

Train speed profiles optimization using a genetic algorithm based on a random-forest model to estimate energy consumption

Ahmed Amrani a*, Amira Ben Hamida a, Tao Liu b, Olivier Langlois b

[a]*IRT SystemX, Paris-Saclay, 8 Avenue de la Vauve, 91120 Palaiseau, France*
[b]*Alstom, 48 Rue Albert Dhalenne, 93400 Saint-Ouen, France*

Abstract

The most important part of the train's energy is consumed by the traction system. The tractive energy depends mainly on the driving behaviour. Improving driving strategies has great potential to enhance the energy efficiency. This paper presents a speed profile optimization approach based on a genetic algorithm. The objective of the genetic algorithm is to find, for each interstation, the best speed profile which minimizes the energy consumption. The optimized profile takes into account both the physical and the operational constraints such as the maximum allowed travel time, the speed limitations per section and the maximum allowed acceleration and jerk. The fitness function is based on a Random Forest model which is built using on-board measurements. The aim of the model is to estimate accurately the energy consumption corresponding to each speed profile. The initial population of genetic algorithm is mainly composed of the best realistic speed profiles extracted from the collected data.

Keywords: energy optimization; genetic algorithm; speed profile; machine learning; random forest.

* Corresponding author. Tel.: +33-1 69 08 06 17.
E-mail address: ahmed.amrani@irt-systemx.fr

1. Introduction

Efficient energy management in the context of smart cities is one of the most crucial challenges for the next years. Due to the increasing costs and needs, the energy management must rhyme with efficiency and sustainability. In addition to the high cost of energy, its exploitation and production have negative effects on the environment. The railway operation considerably contributes to the overall energy consumption. Then, efficient energy management is an important and very complex issue in the railway environment. Though very complex due to the many aspects it includes, efficiently managed energy on-board a train contributes to significant energy and cost reductions.

The electrical energy consumption of urban trains is divided on two parts: the part used by auxiliary system and the part consumed by the traction system. Auxiliary system includes ventilation, air-conditioning and illumination, etc. Consumption of auxiliaries is influenced mainly by the weather and climate conditions. Generally, the most important part of energy (approximatively 80%) is consumed by the traction system Yang et al. (2016) even if air conditioning consumption could sometimes reach a very high level. The tractive energy depends on the driving strategy. Therefore, improving driving strategies has great potential to enhance the energy efficiency of urban railway systems. Furthermore, it does not require any upgrades in infrastructure. The aim of energy-efficient driving strategy is to find the optimal speed profile which minimizes the tractive energy without exceeding a given trip time.

The speed profile optimization methods can be either exact, heuristics or meta-heuristics methods Mohand et al. (2010). The exact methods are not well adapted when the train model is realistic and unsuitable for real-time optimization. Therefore, meta-heuristics methods such as Genetic Algorithm (GA), Simulated Annealing (SA) and Tabu Search (TS) are good alternatives. However, these methods do not guarantee the optimality of the solution. In fact, in an energy optimization context, we do not target an exact solution; it is sufficient to find a high-quality one. Moreover, the advantage of meta-heuristics over exact methods is the capability to take into account new constraints without altering the algorithm.

Improving energy efficiency requires building accurate and adaptable models to study energy consumption. Two types of models are explored to estimate the energy consumption: physical model and data-driven model. On one hand, the physical models are based on the fundamental principle of dynamics such as the motion equations and are used to simulate the behaviour of the train. The construction of physical models requires specific expert knowledge. We have explored this modelling technique in a similar work, and despite its realistic simulation, it was a complex task to build a precise model since it is necessary to estimate a large number of parameters. On the second hand, the data-driven models are built using machine learning algorithms. Building data-driven models does not require any domain knowledge; their accuracy depends mainly on the quality and quantity of the data.

Thus, in this paper, we propose a system that aims at finding efficient speed profiles based on a Genetic Algorithm. The evaluation function of the algorithm estimates energy consumption using a data-driven Random Forest (RF) model. The initial population of genetic algorithm is mainly composed of the best speed profiles extracted from real data.

This paper is organized as follows. In section 2, we survey similar works in the literature. In Section 3, we present an overview of our system for speed profiles optimization. Section 4 details the steps of the automatic process of building data-based models of energy consumption. The on-board measurements, the pre-processing steps and the learning procedure are then explained. Section 5 describes the optimization process: the problem representation, the fitness function and optimization algorithm. Finally, some perspectives of the presented work and a conclusion are given in Section 6.

2. Related works

Over the years, several optimization approaches have been applied to speed profile optimization. This section mainly presents approaches using genetic algorithms.

Hoang et al. (1975) studied the energy-efficient driving topic, under certain simplifying assumptions, using heuristic method based on a direct search algorithm. Recently, numerical algorithms are more widely used thanks to the rapid increase in computer performance and the growing success of the distributed and virtualized computing architectures.

In Chang and Sim (1997), the authors proposed a coasting control strategy based on a genetic algorithm. Each gene is composed of a control command associated to a relative position between two stations. The fitness evaluation includes the punctuality, the passenger comfort and the train's energy consumption parameters. The estimation of the energy consumption is based on a physical model. In a similar way, in Han et al. (1999), the authors exploit the genetic algorithms to optimize the tractive energy. Indeed, each solution corresponds to a control strategy determined by the positions of coasting control. Each gene is composed of a position associated to a speed. The number of genes and the speed values depend on the speed limitations per section. The fitness function is computed thanks to a simplified physical model.

A multi-objective genetic algorithm is proposed in Chevrier (2010). The aim of the approach is to minimize the running time and to reduce the energy consumption. Each gene is composed of a position and a speed. The algorithm provides a set of non-dominated solutions. In order to evaluate energy consumption, the fitness function uses a physical model based on the fundamental equation of dynamics.

Dominguez et al. (2011) present a computer-aided procedure for optimal speed profiles selection. The procedure takes into account running times, passenger comfort and energy consumption. To this end, the equations and algorithms that define the train motion and control have been modelled and implemented in a simulator which includes an automatic generator of speed profiles and a graphical assistant for the selection of speed commands. In a further work, the authors, in Dominguez et al. (2012), consider energy savings due to the energy recovered from regenerative braking. A physical model of a train with an on-board energy storage device and a network model is used to estimate the energy recovered by the train. The proposed method could be applied in real time according to the traffic and electrical situation of the line.

Boschetti and Mariscotti (2014) and Brenna et al. (2016) proposed a genetic algorithm approach to search for an optimal control strategy. Each gene is composed of a position associated to a driving command such as decelerating, accelerating and breaking. A physical model is utilized in order to estimate energy consumption. The energy saving was evaluated through the simulations in Matlab-Simulink environment.

Keskin and Karamancioglu (2017) evaluated three metaheuristic algorithms Genetic Simulated Annealing, Firefly, and Big Bang-Big Crunch. The fitness function is based on a physical model of energy consumption. The results are validated using a Matlab simulation. The paper illustrated the efficiency of using metaheuristics to solve the optimal train operation problem.

LEADER (Locomotive Engineer Assist/Display & Event Recorder) described in Nickles et al. (2003) is a global system for freight train management system designed to optimize the train movement in order to reduce fuel consumption. It performs real-time data collection, processing, storage and reporting and provides real-time commands (acceleration or braking) to locomotive engineers; based on the current state, the system performs calculation using a physical models Mosier (1977). The system improves the effectiveness of best train control practices through a more complete understanding of train behavior. To explain the recommended actions, LEADER displays several information such as gradient, track curvature, acceleration, fuel consumption, speed and the pressure of the brake pipe.

The majority of the cited works mostly rely on a combination of meta-heuristics optimization approaches with a physical model which requires specific knowledge and important modelling effort to be built and updated. Recently, in Martinez Fernandez et al. (2016), the authors built a data-driven model to estimate the energy consumption. Our approach targets a fully automated optimization process by using (1) a genetic algorithm to find energy efficient speed profiles and (2) a data-driven model to compute the energy consumption. Considering that each gene is composed of a position and a speed, we minimize the electrical tractive energy consumption. We take into account the operational and physical constraints such as speed limitations per section, maximum allowed acceleration and maximum allowed jerk. The maximum allowed trip time is also considered as a constraint. The fitness function is based on a random-forest model. The aim of the model is to estimate accurately the energy consumption corresponding to each speed profile. The evaluation process is then faster compared to the use of a physical model. In order to enhance the quality of the obtained solution and reduce the convergence time, the initial population of genetic algorithm is mainly composed of the best speed profiles extracted from real data.

Therefore, the optimization process can be carried out offline as well as online. In comparison with the studied approaches, our solution is automatic, faster and continuously updated thanks to the use of data-driven model.

3. Global view of the optimization system

The objective of our system is to find efficient energy speed profiles using on-board collected data. Several transformations are performed in order to clean and restructure the raw data. The pre-treated data is then stored in a database in order to extract realistic speed profiles and to build models for estimating energy consumption. In order to find the best speed profile for each interstation which is defined as the portion of the line between two passenger stations, we use an optimization process based on genetic algorithm. The overall optimization system is described in Fig. 1.

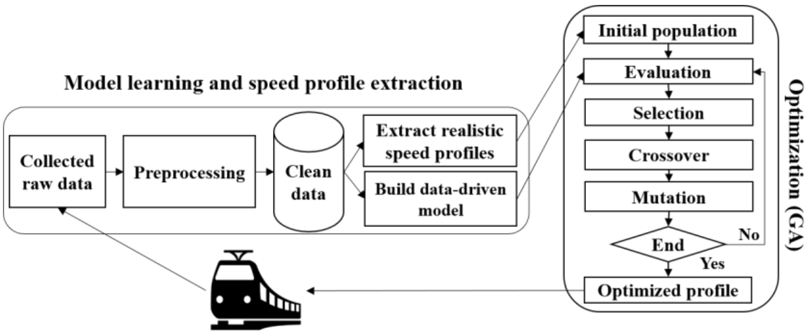

Fig. 1 Optimization system overview

During the optimization process, the fitness function is frequently used to evaluate speed profiles. Then, the quality and execution time of the fitness function have a substantial impact on the convergence time and efficiency of the genetic algorithm. The use of an initial population composed of good quality individuals not only speeds up the convergence process, but also gives the possibility to stop the optimization process at any time with the assurance of having a good solution. Therefore, our optimization algorithm has two important features: the evaluation of speed profiles is based on the data-driven model and the initial population is mainly composed of the best realistic speed profiles. Using a learned model makes our approach potentially automated after a first phase of setting up.

4. Building data-driven models for energy consumption

4.1. Measured data

The train measured dataset is composed of various parameters collected from sensing devices installed on-board an urban tram. Measurements include voltage, current, speed and GPS position. In the used dataset, the sampling frequency is equal to one second. In the deployment of practical systems such as electrical railway system, information has to be collected from heterogeneous sensors. Electrical railway environment is characterized by the presence of high voltage and current with abrupt variation and train movements following urban topography (tunnels, bridges, hills, etc.). These factors disturb the data measurement and transmission. Therefore, the collected data contains errors and inconsistencies.

4.2. Pre-processing

The aim of the pre-processing steps is to reduce errors, filter, clean and transform data. Several pre-processing steps were performed:

- Management of missing values: the time steps where there are missing values are removed.
- Removing invalid points: the trains in depot and maintenance trips are not considered.
- Train route extraction: the route is an ordered list of waypoints where each waypoint represents a turn or significant step of the track.
- Map matching: association of each measured GPS position to a kilometric point in the train route.
- Extract train trips: identification of each distinct trip and its direction.
- Elevation extraction: the elevation of each point in the route is extracted using Google Elevation API.

- Computation of additional parameters: the gradient, the acceleration, the jerk and the electric power.

4.3. Supervised learning (regression and classification)

Based on the preprocessed data, our objective is to obtain the most accurate energy consumption model of the train. We then test several supervised learning algorithms which establish a relationship between a speed profile and its energy consumption. To evaluate energy consumption, in the context of energy management, we use both classification and regression algorithms. Regression is used when the target attribute takes continuous values while classification is used when the target attribute has discrete values. Started from known values of input attributes, the regression model estimates the class value while the classification model tries to assign a label (discrete value) to the class. In order to estimate power consumed for traction, we evaluated three machine learning algorithms which are Probabilistic Neural Networks (PNN), Decision Tree (DT) and Random Forest (RF). These algorithms can be used to solve both classification and regression tasks.

Neural networks are computational models based on the underlying structure of biological neural systems. A neural network is composed of neurons which work in parallel to provide an output value based on the values of the input attributes. The learning algorithm modifies iteratively the parameters which regulate the connections between the neurons in order to minimize the error on the training set. The neural networks were used in Martinez Fernandez et al. (2016) in order to compute the energy consumption of electric trains. In this work, we use the Probabilistic Neural Network (PNN) which is trained based on the Dynamic Decay Adjustment method using Constructive Training Berthold and Diamond (1998) as the underlying algorithm.

Decision trees algorithm is a very effective method in supervised learning. It has been introduced by Quinlan (1993). The algorithm takes as input a collection of tagged data, and outputs a tree. Each internal node represents an input attribute where each value corresponds to an edge to children. Each leaf represents a value of the target variable. In order to determine the best splitting attribute, it is possible to use several methods. In this paper, we use gain ratio impurity method to evaluate attributes Quinlan (1993).

The Random Forest model is introduced by Breiman (2001). It is an efficient machine learning model which was used widely for many real world applications. It is an ensemble learning algorithm based on the average prediction of different decision trees. Each tree is fitted on a part of the data, made by two sampling methods: random sampling with replacement of observations which is also known as bootstrap aggregating or bagging method, and random selection of features called feature bagging. The bagging methods and the operation of averaging the results obtained by the different trees allow the Random Forest having better accuracy than a simple decision tree.

4.4. Input attributes

In the used data set, the measurement frequency is one second. Each speed profile between two stations is then represented by the speed as a function of time. To compute the whole energy consumption of a speed profile, a model is built to estimate the traction power in each second. To train and evaluate the different models (PNN, DT and RF) the following input attributes are used: speed, train energy efficiency, acceleration, jerk (the rate of change of acceleration) and gradient (slope of the track). In order to estimate accurately the electrical power for each second, in addition to the information about the current second, we use the information about the previous and the next four seconds. The used attributes to estimate the power consumption in a second *i* of a speed profile, are presented in Table 1:

Table 1. The set of attributes used to estimate **traction power** in a second *i*.

Attributes /Seconds	*i-4*	*i-3*	*i-2*	*i-1*	*i*	*i+1*	*i+2*	*i+3*	*i+4*
Speed	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*
Acceleration	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*
Jerk					*X*	*X*	*X*	*X*	*X*
Energy efficiency					*X*	*X*	*X*	*X*	*X*
Gradient					*X*				

4.5. Model evaluation

To evaluate the quality of traction power estimation (in watt), train consumption models are tested in both classification and regression modes. In order to use classification, the target attribute (traction power) is discretized into *k* intervals of equal size using the equal width discretization method. The data-driven models are trained using a selected set of consumption data measured on-board. In order to measure the differences between the model-estimated values and the measured values, the Root Mean Square Error (RMSE) is used for regression and the accuracy for classification. In our example, the used dataset comprises 179342 lines. To train the model, 70% of the dataset is randomly selected. The evaluation of the model then use the remaining 30% of the dataset. The traction power is a numerical attribute. To test classification models, the attribute is discretized into 15 intervals.

4.6. Case study and results

To determine the most suitable learning algorithm for the estimation of traction power in watts, the three algorithms defined in section 4.3, namely the decision tree (DT), the Probabilistic Neural Network (PNN) and the Random Forest (RF) model are compared using the Knime Analytics platform, a tool able to build the machine learning (ML) workflow. The comparison results are presented in the Table 2.

Table 2. Machine learning algorithms evaluation.

ML algorithms	PNN	DT	RF
Classification (Accuracy)	0.79	0.80	**0.85**
Regression (RMSE)	54000	48308	**33157**

The results show that the RF model is the best for our use case. Indeed, the RF model outperforms the other algorithms in both classification and regression. Fig. 2 compares the measured traction power in watt and the estimated traction power using the RF model, for 330 seconds time-frame.

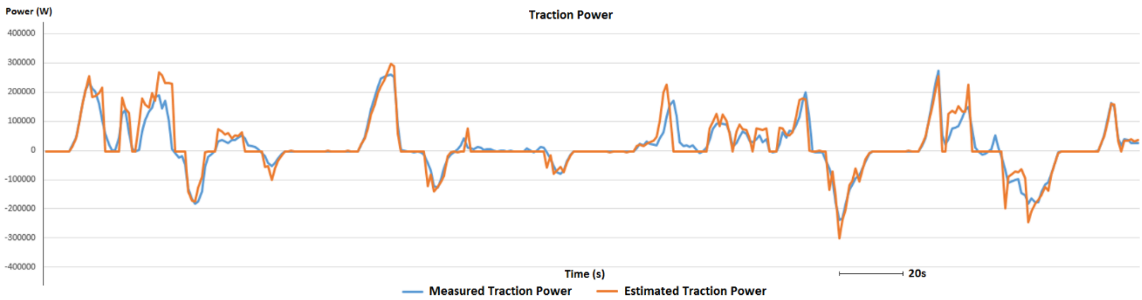

Fig. 2 Measured traction power versus estimated traction power in watt

In order to evaluate the relevance of using the input attributes related to the previous and following seconds, we evaluated the Random Forest algorithm for regression task with several sets of attributes. The Table 3 summarizes the performed tests.

Table 3. The evaluation of the impact of attribute selection on the estimation of traction power in watts.

Input attributes	RF Root Mean Square Error (RMSE)
Using only current second attributes	58070
Using current second and the four previous seconds attributes	53217
Using current second, four previous seconds and the four next seconds attributes	**33157**

The tests show clearly that using attributes related to the next seconds improves significantly the quality of traction power estimation. However, there is still an important room for model improvement. Actually, the used data does not contain any information about passengers mass. Moreover, speed information and GPS information are not very accurate. Increasing the amount of data used to learn also improves the quality of the model. Based on the previous results, we will use the random forest model to optimize speed profiles in the remainder of this paper.

5. Speed profiles optimization

5.1. Genetic algorithm in speed profile optimization

The genetic algorithms is a metaheuristic approach inspired by the process of natural selection. It follows an iterative stochastic process in order to find exact or high-quality solutions by relying on bio-inspired operators such as mutation, crossover and selection. In genetic algorithms, each solution is encoded into strings of digital numbers and is associated to a score with regard to fitness function. The optimization process starts with an initial population. In each generation, the genetic algorithm performs operations like selection, crossover and mutation on individuals with a probability based on their corresponding fitness value. After several generations (iterations), solutions which have better scores are selected as optimal or suboptimal solutions.

The aim of the genetic algorithm is to find, for each interstation and context, the best speed profile which minimizes the energy consumption and takes into account operational constraints. The genetic algorithm starts with an initial population of speed profiles. At each iteration, a new set of speed profiles is generated by performing crossover and mutation operations on a selected existing speed profiles. Iteration after iteration, the quality of the solution is improved. Algorithm ends when the current iteration does not provide significant improvement over previous iterations. It is also possible to stop the algorithm when the maximum optimization time specified is reached.

Speed profiles shall meet train's physical limitations, operational objectives as well as insure the safety and comfort of passengers. Thus, each proposed speed profile must satisfy a set of constraints which are speed limitations per section, the maximum trip time, the maximum acceleration and the maximum jerk.

5.2. Problem coding and initial population

The interstation speed profile is represented by the speed in function of the distance. Each individual is composed of l genes. Each gene is composed of a fixed position and a numerical value corresponding to the speed in this position. The fixed positions are obtained by discretizing the interstation distance into ($l-1$) intervals of equal size. The l positions correspond to the limits of the intervals.

The initial population of speed profiles used by the genetic algorithm contains: a set of the best realistic extracted profiles for the interstation and a set of generated speed profiles. In order to respect all the constraints, we apply the following method: for each interstation, a static speed envelope is pre-calculated taking into account the following constraints: maximum acceleration, maximum jerk, speed limitations, start and end speeds. Then, for each gene i, $0 < i < (l-1)$, we randomly select a corresponding value from the interval of acceptable values given the selected previous speeds.

5.3. Genetic operators

In order to guide the algorithm towards the best speed profile which satisfies operational constraints, we adapt the following genetic operators: mutation and crossover.

The crossover is a process of taking more than one parent individual and producing a child from them. In our implementation, a crossover operation generates a new profile by combining two existing profiles selected using the tournament method. The first part of the generated new profile is extracted from one profile and the second part is extracted from the other profile. The junction position p between the two parts is selected randomly from the interstation. If the speed value at the position p does not respect the constraints, another value is randomly selected from the list of acceptable values.

The mutation operator alters one or more speed values of an individual. In our algorithm, the mutation can be applied to modify a speed value in a selected random position or for all the positions starting from a position selected randomly. Several types of mutation are applied:

- Modify randomly a speed value.
- Replace a speed value by the maximum or the minimum speed value. For individuals whose travel time exceeds the maximum allowed running time, the probability of selecting the maximum speed value is increased.
- Replace a speed value by the same speed value of the previous position.

The search space of the genetic algorithm is limited by the constraints of the system. Indeed, for all operations performed, the new speed value is always selected from the range of values that satisfies all constraints. Moreover, to ensure that the profile generated by the genetic operators satisfies all the constraints, its integrity is systematically verified.

5.4. Fitness Function

In order to evaluate the energy consumption of each generated profile, the fitness function of each new profile is computed. The fitness function takes a speed profile as input and produces as output how "good" the profile is with respect to the energy consumption and all constraints. The fitness function is evaluated repeatedly in genetic algorithm, therefore it should be fast to compute. In the current work, the fitness function is based on the computation of the traction power of the train. To compute the traction power from a speed profile, the above mentioned random forest regression model is used. Using a data-driven model, the fitness function is then accurate and fast to compute. As described in Section 4, the model computes the traction power for each second. The total energy consumption of the profile is estimated by summing all traction powers corresponding to every second. Profiles that do not respect the constraints have bad scores even if they are energetically efficient. The fitness function (*f*) of speed profile *sp* with running time *T* (in seconds) is defined as follows:

$$f(sp) = (\sum_{t=0}^{T} TracPower(t)) + ConstSatisfaction(sp),$$

where *TracPower(t)* is the traction power estimation at second *t* using the Random Forest model and *ConstSatisfaction(sp)* is a function that returns 0 if the speed profile *sp* respects the constraints (see Section 5.1), otherwise it returns a high enough positive value to penalize the profile. The objective of the genetic algorithm is to minimize this fitness function.

5.5. Validation upon a realistic case study

Many speed profiles are possible for a given mission between two passenger stations. Fig. 3 illustrates various profiles extracted from measurements on a tram line.

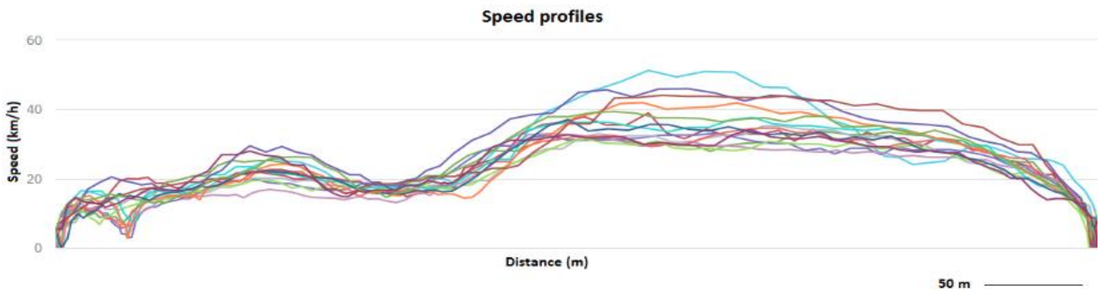

Fig. 3 Interstation speed profiles

To evaluate the energy gain, we applied the genetic algorithm to an interstation. Before starting the optimization process, all system constraints must be determined. The maximum values of speed, travel time, acceleration and jerk are fixed in order to strictly respect the material and operational constraints and to guarantee a good passenger comfort. In order to produce realistic profiles, the maximum speed at each point of the line is also limited by the maximum speed recorded in the measurements. As described in section 4, the model is capable of estimating the power required for traction and power generated during braking. In fact, in a multi-train context, the power generated is potentially reusable by other trains. However, when evaluating the proposed optimization algorithm, we only consider the mono-train use case. Therefore, the objective of the algorithm is to minimize the energy required for traction of a single train.

For the experimental validation, an interstation of 530 meters long is considered. To encode each speed profile for the genetic algorithm, the interstation distance is discretized into 69 intervals (*l=69*) of equal size. The other parameters are set as follows: the value of jerk is lower than 0.3 m/s^3, the maximum value of acceleration is equal to 1 m/s^2, and the maximum value of travel time is equal to 106 seconds. Experimental results show that the genetic algorithm produces energy-efficient speed profiles. Fig. 4 shows a speed profile obtained through the optimization process and the maximum speeds envelope extracted from the measurements.

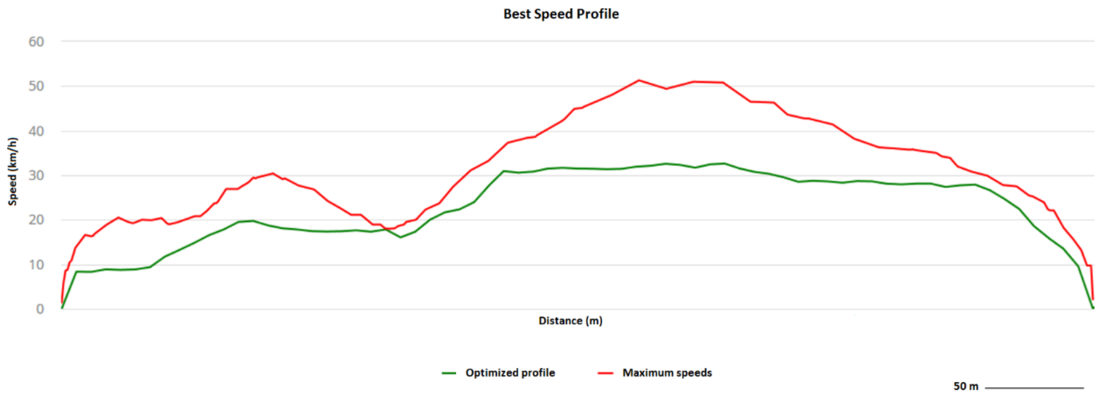

Fig. 4 The optimized speed profile

The Table 4 presents a comparison between the optimized profile and a set of speed profiles extracted from real measurements, the traction energy in watt-hours (Wh) and the time in seconds (s).

Table 4. Speed profiles comparison.

Profile	Traction energy (Wh)	Time (s)	Constraints satisfaction
Optimized profile	***679***	***105***	***Yes***
Real profile 1	793	105	Yes
Real profile 2	813	105	Yes
Real profile 3	733	111	No
Real profile 4	745	119	No

For the test case, the use of an optimized profile allows, in addition to respecting the constraints, to reduce energy consumption by 14% compared to the use of the best profile extracted from measured data. Thanks to the use of a data-driven model, the fitness function is fast to compute. As a result, the genetic algorithm can explore a larger space in a limited searching time.

The proposed approach is generic and adaptable. Indeed, the genetic algorithm allows to easily add new constraints on the speed profiles. In addition, the number of discretization points of the speed profiles is variable (Section 5.2). The approach can therefore be applied to manual operation or in the framework of Automatic Train Operation.

6. Conclusion and perspectives

In this paper, we present an innovative solution for energy trains optimization. Our approach exploits the Data Analytics techniques to find energy efficient speed profiles. The prerequisite is to collect data coming from sensors measuring the speed, the acceleration, the electric current, the voltage, the passengers' mass and the GPS positions. In fact, we are able to model the train's behavior and its related energy consumption based on on-board measurements. The performed experimentations show that the Random Forest algorithm is adapted to power estimation and that using attributes related to the next seconds improves significantly the quality of the model.

Optimal train driving strategies are developed using a genetic algorithm based on a Random Forest model to estimate the energy consumption. In our approach, the genetic algorithm is initialized by a population composed mainly by the best real speed profiles. The optimization process can then be stopped at any time with the assurance of having a good solution. For a realistic case study, the experimental validation showed that the profile proposed by our approach significantly reduces energy consumption compared to a set of interstation profiles extracted from the measurements.

In the literature, similar works are based on physical model which requires specific knowledge and important modelling effort to be built and updated. In this paper, we propose to use a data-driven model and to completely automatize the optimization process. A generic and automatic solution is then obtained. The data-driven model can be updated automatically on a regular basis, as the on-board data is uploaded. As a result, the quality of the model is improved progressively with a growing learning set. In addition, the model automatically adapts to the evolution of the rolling stock and the infrastructure. Indeed, the consumption of each train evolves when aging and after maintenance updates; the proposed speed profile will be adapted consequently to each train. Based on this approach, we can build models even for unknown train types, which is particularly useful for operators when managing different trains on the same line. Furthermore, our approach is easily adaptable to any type of railway systems such as metros, suburban and high-speed lines.

This first study and achievements widely open the door to further improvements at several layers. First, an important room for model improvement is still possible. Several solutions are conceivable, such as increasing the size of the dataset, implementing better accuracy sensors, improving the pre-processing steps, using models based on deep learning and including domain knowledge into the learning process. Second, the proposed profiles can be improved and adapted by taking into account other constraints. The use of a physical simulation and field-test campaign will allow us to better calibrate the constraints according to the operational context.

Acknowledgements

This research work has been carried out under the leader-ship of the Institute of Technological Research SystemX, within the scope of a joint collaboration between the In2Rail European H2020 Project and the Smart City Energy Analytics French research project.

7. References

Martinez Fernandez, P., Garcia Roman, C., Ricardo, F., 2016. Modelling Electric Trains Energy Consumption Using Neural Networks. Transportation Research Procedia. 18. 59-65. 10.1016/j.trpro.2016.12.008.

Berthold, Michael R.., Diamond, J., 1998. Constructive training of probabilistic neural networks. Neurocomputing 19 (1998): 167-183.

Yang, X., Li, X., Ning, B., Tang, T., 2016. A Survey on Energy-Efficient Train Operation for Urban Rail Transit. In IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 1, pp. 2-13.

Boschetti G., Mariscotti A., 2014. Optimizing the Energy Efficiency of Electric Transportation Systems Operation Using a Genetic Algorithm. International Review of Electrical Engineering 9(4), 783–791.

Breiman, L., 2001. Random forests, Machine learning, vol. 45, no. 1, pp. 5–32.

Quinlan, J.R.. 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann 1993, ISBN 1-55860-238-0.

Hoang, H., Polis, M., Haurie, A. 1975. Reducing energy consumption through trajectory optimization for a metro network. IEEE Trans. Automat. Control, vol. 20, no. 5, pp. 590–595.

Mezmaz, M., Tuyttens, D., Fei, H., Jalwan, J. 2010. Towards the State-of-the-Art of the Use of Genetic Algorithms to Solve Energy-Conscious Train Control Problem. International Conference on Metaheuristics and Nature Inspired Computing, Djerba, Tunisia.

Chevrier, R., 2010. An Evolutionary Multi-objective Approach for Speed Tuning Optimization with Energy Saving in Railway Management, Evaluation des Systèmes de Transports Guid's et leur Sécurité – ESTAS- INRETS, Mars.

Keskin, K., Karamancioglu, A., 2017. Energy-Efficient Train Operation Using Nature-Inspired Algorithms. Journal of Advanced Transportation. Volume 2017, Article ID 6173795

Brenna, M., Foiadelli, F., Longo, M., 2016. Application of Genetic Algorithms for Driverless Subway Train Energy Optimization. In International Journal of Vehicular Technology; 2016; 1-14; International Journal of Vehicular Technology.

Han, S.H., Byen, Y.S., Baek, J.H., An, T.K., Lee, S.G., Park, H.J., 1999. An optimal automatic train operation (ATO) control using genetic algorithms (GA), TENCON 99. Proceedings of the IEEE Region 10 Conference, vol. 1, 1999.

Chang, C., Sim, S., 1997. Optimising train movements through coast control using genetic algorithms. Proc. Inst. Elect. Eng.—Elect. Power Appl., vol. 144, no. 1, pp. 65–73.

Nickles, S.K., Hawthorne, M.J., Haley, J.E. 2003. Method of optimizing train operation and training, US Patent 6,587,764, Google Patents https://www.google.ch/patents/US6587764.

Mosier, J.E., Railway train control simulator and method. 1977. https://www.google.ch/patents/US4041283, 1977, Google Patents, US Patent 4,041,283.

Dominguez, M., Fernandez-Cardador, A., Cucala A. P., Pecharroman, R. R., 2012. Energy Savings in Metropolitan Railway Substations Through Regenerative Energy Recovery and Optimal Design of ATO Speed Profiles. In IEEE Transactions on Automation Science and Engineering, vol. 9, no. 3, pp. 496-504.

Dominguez, M., Fernandez-Cardador, A., Cucala, A. P., Lukaszewicz, P., 2011. Optimal design of metro automatic train operation speed profiles for reducing energy consumption. Proc. Inst. Mech. Eng., Part F: J. Rail Rapid Transit, vol. 225, no. 5, pp. 463–474.