Measurements-based constrained control optimization in presence of uncertainties with application to the driver commands for high-speed trains

Julien Nespoulous, Guillaume Perrin, Christine Funfschilling, Christian Soize

Measurements-based constrained control optimization in presence of uncertainties with application to the driver commands for high-speed trains

Julien Nespoulous[a,b], Guillaume Perrin[c,*], Christine Funfschilling[b], Christian Soize[a]

[a] *Universit Gustave Eiffel, MSME UMR 8208, 5 Bd Descartes, 77454 Marne-La-Valle, France*
[b] *SNCF, DTIPG, 1-3 Av Franois Mitterrand, 93574 Saint-Denis, France*
[c] *COSYS, Universit Gustave Eiffel, 14-20 Boulevard Newton, 77447 Marne-la-Valle, France*

Abstract

The railway world is undergoing major changes. The advent of new technologies allows us to rethink the train system and face new challenges, but one must not forget all the ecological constraints that are now accentuated by the increase in energy costs. This paper focuses on the optimization of the driver commands to limit the energy consumption of the trains under punctuality and security constraints. This problem falls within the framework of control optimization problems for nonlinear dynamic mechanical systems in the presence of constraints and uncertainties. A four-step approach is then proposed in this paper to solve this problem: (1) the introduction of simplified and fast-to-evaluate models to model the nonlinear dynamic behavior of the train and its energy consumption; (2) the identification in a Bayesian formalism of the parameters on which these models depend from on-track measurements on commercial trains; (3) the reformulation of the optimization problem so that it integrates the uncertainties related to an imperfect knowledge of these estimated parameters; (4) the resolution of the optimization problem using evolutionary algorithms. The main specificity of this work lies in the fact that not only the objective function to be minimized, here the energy consumed by the train, is impacted by the uncertainties, but also the admissibility constraints of the solution, here punctuality and operating safety. The integration of the uncertainties in the search for the control function is thus not trivial and requires several original adaptations in order to make the final optimization problem well posed.

Keywords: Nonlinear systems, Bayesian estimation, control optimization, dimension reduction

1. Introduction

Simulation is increasingly used to optimize the behavior of more and more systems, whether they are physical, biological, mechanical or economic. In this work, we are interested in the optimization, under constraints, of the control of a nonlinear dynamic mechanical system (whose expression is given in (A.5)). The link between the control and the state vector gathering temporal evolutions of characteristic quantities of the considered system (ex: position and speed) is often achieved thanks to numerical models.

Such models are generally based on three kinds of inputs that must be set to execute the code. First, there is the time evolution of the control function driving the system that one seeks to optimize with respect to a certain criterion. There are also the parameters characterizing the system itself, which are *a priori* constant during the whole simulation, such as its mass or its traction and braking characteristics. Finally, it is necessary to specify environmental conditions, which may change during the circulation (ground topology, wind or humidity, and so on).

*Corresponding author
Email addresses: julien.nespoulous@free.fr (Julien Nespoulous), guillaume.perrin@univ-eiffel.fr (Guillaume Perrin), christine.funfschilling@sncf.fr (Christine Funfschilling), christian.soize@univ-eiffel.fr (Christian Soize)

The difficulties related to the solving of this type of problem are numerous. The first ones come from the fact that the optimal control function is functional, so that it is *a priori* searched in a set of infinite dimension. The computational costs of the direct problem, i.e. the simulation of the temporal evolution of the state vector with a fixed choice of the model parameters, can be long, which can strongly limit the ability to explore the set of possible control functions. Finally, and it is on this difficulty that this paper focuses, the parameters of the system may not be perfectly known and the environmental conditions may be poorly specified [1]. Indeed, whether they are epistemic or random, these uncertainties on the system parameters and environmental conditions lead to the minimization of an uncertain quantity of interest, radically changing the optimization problem of the control function [2]. And these difficulties are all the more important as these uncertainties also impact the constraints that the control function must verify.

The first objective of this work is thus to show to what extent the presence of these uncertainties on both the objective function and the constraints can make the optimization of the control function ill-posed, in the sense that it will admit no solution. Indeed, if the constraints are very sensitive to the uncertainties, it may be impossible to find a deterministic solution that respects the constraints, whatever the realization of the uncertain parameter. The challenge will then to make it well-posed again. This will be done thanks to the introduction of a transformation of the control function making it possible to no longer make the search space of the solution (which includes constraints) depend on the uncertain parameters. The second objective will consist in proposing an approximate method of solving this problem under uncertainties. Several steps will first be proposed to reduce the dimension of the search space in a relevant way. The function to be minimized being *a priori* nonconvex and its gradient being not explicit, an evolutionary-type algorithm will finally be used to solve this problem in reduced dimension. These different methodological developments will then be applied to the case of the minimization of the running consumption of high-speed trains. Indeed, the recent skyrocket of the electrical energy costs encourages the railway companies to reduce their consumption. Three different levers can therefore be activated: modifying the running environment of the train, redesigning the vehicle itself, or adapting the train speed. We propose to focus exclusively on the train speed in this work, not only because it may be the only parameter on which we can play in the short term, but also because it has recently been observed that on the same high-speed line, energy consumption between the different circulations could considerably vary.

Optimizing train speed to reduce energy consumption is not a new problem and several studies can be found in the literature. For instance, the case of systems of several trains has been considered in [3], the importance to take into account uncertainties in the optimization problem is analyzed in [4], and multi-objective formulations can be found in [5]. These optimization problems introduced in these papers are solved with many different methods like evolutionary algorithms [6], dynamic programming [7], pseudo-spectral methods [8], or the maximum principle [9] between others.

The present work differs from these previous works in several aspects. (i) First, the driver commands have been privileged to the train speed as optimization variable. This complicates the problem because the link between driver commands and speed is not trivial, but makes the results of this work much more usable from an operational point of view. (ii) Second, the optimization problem that we introduce relies on models for the train dynamics and the energy consumption that are as close as possible to the train under consideration, while being relatively quick to be evaluated. Note in particular the fine integration of several factors playing on the energy consumption during acceleration phases and the presence of two braking systems, allowing energy recovery during these phases. (iii) Thirdly, a rigorous Bayesian approach has been implemented for the estimation of the parameters, previously introduced in the models. The identification of these parameters was made possible by the deployment of dedicated measurement campaigns on trains operating in commercial conditions. In particular, we will insist on the introduction of two sources of model error, in order to take into account the necessarily uncertain nature of the railway dynamics model, but also of the energy consumption model. (iv) Finally, the uncertainties management is based on a novel formulation of the optimization problem under constraints, allowing to no longer make the constraints depend on the model parameters uncertainties.

The outline of this paper is as follows. Section 2 presents

the general framework proposed for constrained control optimization under uncertainties. Section 3 presents the railway application, and Section 4 concludes the paper.

Remark.. In the rest of the manuscript, in order to facilitate the distinction between deterministic and random versions of the same quantity, we will use capital letters for random quantities and small letters for deterministic quantities.

2. Theoretical framework

2.1. Problem definition

In this work, we are interested in optimizing the nonlinear dynamical behavior of a system of interest (Equation (A.5)) through its control function. Let $0 \leq t \leq 1$ be the time, $\boldsymbol{x} := \{\boldsymbol{x}(t) \in \mathbb{X} \subset \mathbb{R}^p,\ 0 \leq t \leq 1\}$ be the state vector of the system, and $u := \{u(t), 0 \leq t \leq 1\}$ be the system control function. Function u is assumed to belong to $\mathbb{U} := C^0([0,1],[-1,1])$, which is the set of continuous functions defined on $[0,1]$ whose values are in $[-1,1]$. For a given control function $u \in \mathbb{U}$, the state vector $\boldsymbol{x}$ is supposed to verify a deterministic parametric equation, which is written as:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), u(t); \boldsymbol{z}_f),\ \ 0 \leq t \leq 1. \tag{1}$$

Here, the model $\boldsymbol{f}$ is supposed to be known, but the vector of parameters $\boldsymbol{z}_f \in \mathbb{Z}_f$, on which it relies, is *a priori* unknown (or not perfectly known) and needs to be estimated from data. The model $\boldsymbol{f}$ is said to be deterministic in the sense that for given values of u and $\boldsymbol{z}_f$, it leads to a unique value for the state function, which is written

$$\boldsymbol{x}(\cdot; u, \boldsymbol{z}_f) :\ t \in [0,1] \mapsto \boldsymbol{x}(t; u, \boldsymbol{z}_f) \in \mathbb{X}. \tag{2}$$

The system dynamics must also verify a certain number of constraints, which can be written as an envelope $\mathcal{E}$ gathering lower and upper bounds for each component of $\boldsymbol{x}$ at any time t. The lower and upper bounds at a particular time may be infinite when there is no specific constraints, or equal when a component of $\boldsymbol{x}$ has to pass at a specific value at a specific time. In particular, envelop $\mathcal{E}$ includes requirements on the initial and final values of $\boldsymbol{x}$ under the form $\boldsymbol{x}(0) = \boldsymbol{x}_0$ and $\boldsymbol{x}(1) = \boldsymbol{x}_1$. For a given value of $\boldsymbol{z}_f \in \mathbb{Z}_f$, this allows us to define the set of valid control functions, noted $\mathbb{U}^{\text{valid}}(\boldsymbol{z}_f)$, by:

$$\mathbb{U}^{\text{valid}}(\boldsymbol{z}_f) := \left\{u \in \mathbb{U} \mid \boldsymbol{x}(\cdot; u, \boldsymbol{z}_f) \in \mathcal{E}\right\}. \tag{3}$$

The performance of the system of interest is analyzed through another real-valued function h, which also depends on $\boldsymbol{x}$ and u. Function h is again assumed to be known, but may also depend on non-perfectly known parameters, noted $\boldsymbol{z}_h \in \mathbb{Z}_h$. For a given value of $\boldsymbol{z} := (\boldsymbol{z}_f, \boldsymbol{z}_h) \in \mathbb{Z} := \mathbb{Z}_f \times \mathbb{Z}_h$, we are therefore interested in identifying the optimal control, which is defined as the solution of:

$$\min_{u \in \mathbb{U}^{\text{valid}}(\boldsymbol{z}_f)} \int_0^1 h(\boldsymbol{x}(t; u, \boldsymbol{z}_f), u(t); \boldsymbol{z}_h)\, dt. \tag{4}$$

Remarks. In the same manner than for the state vector, there may exist constraints on the values of h under the form of time dependent lower and upper bounds. These constraints are however not considered here for the sake of simplicity. Note also that function h is not necessary positive, as it will be the case for the railway application.

2.2. Specificities, difficulties, and proposed methodology

The optimization problem defined by Eq. (4) may seem relatively standard in the field of control optimization. Nevertheless, the fact that the vector $\boldsymbol{z}$ gathering the model parameters is not perfectly known poses a major difficulty. Indeed, it is necessary to estimate its value from dedicated measurements, i.e. measurements for which the values of $(u, \boldsymbol{x})$ are known at all times. Because of the limited and noisy nature of the available measurements, and the necessary approximate nature of the models $\boldsymbol{f}$ and h, this estimation is likely to be affected by uncertainties. To take into account these uncertainties, we use a Bayesian framework. Vector $\boldsymbol{z}$ is therefore modeled by a random vector with given *a priori* probability distribution and the estimation process amounts at computing its *a posteriori* probability distribution, i.e. its probability distribution conditioned by the information provided by the available measurements. Let $\boldsymbol{Z}$ be the random estimator of vector $\boldsymbol{z}$ obtained after this Bayesian estimation.

In order to obtain a robust control function with respect to the uncertainties of the problem, it is necessary to propagate these uncertainties within the optimization problem defined by Eq. (4). A first way of proceeding consists in solving M deterministic problems associated with

M (independent or not) draws of $\boldsymbol{Z}$, which are denoted by $\boldsymbol{Z}(\omega_1), \ldots, \boldsymbol{Z}(\omega_M)$. In this case, we obtain M control functions, which can be considered as M independent realizations of a particular stochastic process indexed by t in $[0, 1]$. Note that the solution of each deterministic problem is not trivial either: the dimension of the space in which the control function is searched is infinite and the constraints concern *a priori* the complete dynamics of the system (past and future). In order to make this problem numerically tractable, we can replace $\mathbb{U}$ by the subset $\mathbb{U}_d := C^0_d([0, 1], [-1, 1])$ of functions interpolating linearly the d values in $[-1, 1]$ of the control function at the d fixed times $t_1 = 0 < t_2 < \ldots < t_d = 1$ (these times may or not be equally distributed over [0,1]). This would lead to the new optimization problem indexed by $z = (z_f, z_h) \in \mathbb{Z}$:

$$\min_{u \in \mathbb{U}_d^{\text{valid}}(z_f)} \int_0^1 h(\boldsymbol{x}(t; u, z_f), u(t); z_h)\, dt, \tag{5}$$

$$\mathbb{U}_d^{\text{valid}}(z_f) := \left\{ u \in \mathbb{U}_d \mid \boldsymbol{x}(\cdot; u, z_f) \in \mathcal{E} \right\} \subset \mathbb{U}^{\text{valid}}(z_f). \tag{6}$$

Without specifying the problem studied, it is difficult to make a precise statement on the existence and uniqueness of a solution to the former optimization problem. And even if we are interested in a specific problem, which will be the case in Section 3, the large dimension of the research domain, the presence of non-linear constraints, and the a priori non-convex character of the function to be minimized may strongly complicate these convergence analyses. In order to simplify the notations, and to make future developments more readable, without restricting their general character, we assume that for $z \in \mathbb{Z}$, the optimization problem defined by Eq. (5) admits a unique solution, which is denoted by

$$u_d^\star(\cdot; z) : \ t \in [0, 1] \mapsto u_d^\star(t; z) \in [-1, 1]. \tag{7}$$

Even if we work in finite dimension, the comparison and the statistical treatment of different solutions of problem (5) remain complex. Indeed, it is not because each *optimal* control verifies the constraints that any linear transformation of them will also verify the constraints. In particular, if we denote by $u_d^\star(\cdot; \boldsymbol{Z}(\omega_m))$ the solution of (5) associated with the realization $\boldsymbol{Z}(\omega_m)$ of $\boldsymbol{Z}$, it is very likely that the empirical average of these solutions,

$$u_M(t) := \frac{1}{M} \sum_{m=1}^{M} u_d^\star(t; \boldsymbol{Z}(\omega_m)), \ 0 \leq t \leq 1, \tag{8}$$

is not an interesting control function (let us think for instance to the case where the different control functions are slightly shifted in time). Looking for common features between these functions can nevertheless allow us to greatly reduce the dimension of the problem. This is what Karhunen-Loève decomposition naturally allows us to do (see for instance [10, 11, 12, 13, 14, 15] for more details about this method, also known as Principal Component Analysis when working with vectors instead of functions). Indeed, given M solutions of the deterministic problem, it seems interesting to estimate the autocorrelation function of the solutions of (5) by its empirical estimate,

$$r_M(t, t') := \frac{1}{M} \sum_{m=1}^{M} u_d^\star(t, \boldsymbol{Z}(\omega_m)) u_d^\star(t', \boldsymbol{Z}(\omega_m)), \tag{9}$$

and to search the optimal control in $\mathbb{U}_d$ under the form:

$$u_K(t; \boldsymbol{\alpha}) := \sum_{k=1}^{K} \alpha_k v_k(t), \ \boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_K), \tag{10}$$

where $\boldsymbol{\alpha}$ is the vector containing all the scalar values used in the linear combination of the functions $v_1, \ldots, v_K$, which are defined as the $1 \leq K \leq \min(M, d)$ first solutions associated with the K first eigenvalues $\lambda_1, \ldots, \lambda_K$ of the following eigenvalue problem in $\mathbb{U}_d$:

$$\int_0^1 r_M(t, t') v_\ell(t')\, dt' = \lambda_\ell v_\ell(t), \ 1 \leq \ell \leq d, \tag{11}$$

$$\int_0^1 v_\ell(t) v_l(t)\, dt = \delta_{\ell,l}, \ \lambda_1 \geq \lambda_2 \geq \ldots \lambda_d \geq 0, \tag{12}$$

with $\delta_{\ell,l}$ the Kronecker symbol equal to 1 if $\ell = l$ and 0 otherwise. We can notice that since $u_d^\star(\cdot; z)$ is in $\mathbb{U}_d$ for each $z \in \mathbb{Z}$, the function r_M is continuous on a compact set and that its rank is at most $\min(M; d)$ by construction. Under this formalism, the optimal control function can now be written $u_K(\cdot; \boldsymbol{\alpha}^\star(z))$, where for each $z = (z_f, z_h) \in \mathbb{Z}$, $\boldsymbol{\alpha}^\star(z)$ is the solution (whose existence and uniqueness are again assumed) of:

$$\min_{\boldsymbol{\alpha}\in\mathbb{A}_K^{\text{valid}}(\boldsymbol{z}_f)} \int_0^1 h(\boldsymbol{x}(t;u_K(\cdot;\boldsymbol{\alpha}),\boldsymbol{z}_f),u_K(t;\boldsymbol{\alpha});\boldsymbol{z}_h)\,dt, \quad (13)$$

$$\mathbb{A}_K^{\text{valid}}(\boldsymbol{z}_f) := \left\{\boldsymbol{\alpha}\in\mathbb{R}^K \mid u_K(\cdot;\boldsymbol{\alpha})\in\mathbb{U}_d, \boldsymbol{x}(\cdot;u_K(\cdot;\boldsymbol{\alpha}),\boldsymbol{z}_f)\in\mathcal{E}\right\}. \quad (14)$$

The difficult management of uncertainties in the former optimization problem mainly comes from the uncertain nature of the admissible set. To circumvent this problem while still taking into account uncertainties, the optimal control problem defined by Eq. (13) can be reformulated as:

$$\min_{\boldsymbol{\alpha}\in\mathbb{A}_K^{\infty}} \int_0^1 \mathbb{E}_{\boldsymbol{Z}}\left[h(\boldsymbol{x}(t;u_K(\cdot;\boldsymbol{\alpha}),\boldsymbol{Z}_f),u_K(t;\boldsymbol{\alpha});\boldsymbol{Z}_h)\right]\,dt, \quad (15)$$

where $\mathbb{E}[\cdot]$ is the mathematical expectation with respect to probability distribution of $\boldsymbol{Z} = (\boldsymbol{Z}_f, \boldsymbol{Z}_h)$ and $\mathbb{A}_K^{\infty}$ is the limit when M tends to infinity of

$$\bigcap_{m=1}^{M} \mathbb{A}_K^{\text{valid}}(\boldsymbol{Z}_f(\omega_m)). \quad (16)$$

The solution of such a problem no longer depends on $\boldsymbol{Z}$ and minimizes the cost function *on average*, which should lead to reasonable control functions for any value of $\boldsymbol{z}$ in $\mathbb{Z}$. But this problem relies on the strong assumption that $\mathbb{A}_K^{\infty}$ is non-empty, i.e. that there are control functions such that the constraints are verified whatever the value of $\boldsymbol{z}$. This assumption is however often not verified in practice, making this formulation unsuitable as it will be the case in the application proposed in Section 3. Rather than proceeding in this way, it seems to us more judicious to introduce a deterministic transformation on the control function to make it verify the constraints whatever the value of $\boldsymbol{z}$. This transformation being generally case-dependent, we will only assume its existence without discussing on its construction in this part.

Hence, let us suppose that there exists a transformation g such that for all $\boldsymbol{z}_f \in \mathbb{Z}_f$ and all u in $\mathbb{U}_d$, $g(u;\boldsymbol{z}_f)$ can be seen as a pseudo-projection of u into $\mathbb{U}_d^{\text{valid}}(\boldsymbol{z}_f)$, in the sense that $g(u;\boldsymbol{z}_f)$ is in $\mathbb{U}_d^{\text{valid}}(\boldsymbol{z}_f)$, and

$$\int_0^1 |g(u(t);\boldsymbol{z}_f) - u(t)|dt \quad (17)$$

is small. The discretized optimization problem under uncertainties can finally be rewritten as:

$$\min_{\boldsymbol{\alpha}\in\mathbb{A}_K} \int_0^1 \mathbb{E}_{\boldsymbol{Z}}\left[h\begin{pmatrix}\boldsymbol{x}(t;g(u_K(\cdot;\boldsymbol{\alpha});\boldsymbol{Z}_f),\boldsymbol{Z}_f),\\ g(u_K(t;\boldsymbol{\alpha});\boldsymbol{Z}_f);\boldsymbol{Z}_h\end{pmatrix}\right]\,dt, \quad (18)$$

$$\mathbb{A}_K := \left\{\boldsymbol{\alpha}\in\mathbb{R}^K \mid u_K(\cdot;\boldsymbol{\alpha})\in\mathbb{U}_d\right\}. \quad (19)$$

In trying to solve the problem defined by Eq. (18), for which we assume again the existence and uniqueness of a solution noted $\boldsymbol{\alpha}^\star$, we thus look for a deterministic control function $u_K(\cdot;\boldsymbol{\alpha}^\star)$ that does not depend on the model parameters, but depends this time on a transformation g to be defined. And we can say that this control function is robust to uncertainties, in the sense that, whatever the value $\boldsymbol{z}$ taken by $\boldsymbol{Z}$ for the true system, $u_K(\cdot;\boldsymbol{\alpha}^\star)$ is supposed to make the cost function be relatively small (minimization of the integral of h on average), while being not too far from (or even in) the admissible set (up to the distance between $u_K(\cdot;\boldsymbol{\alpha}^\star)$ and $g(u_K(\cdot;\boldsymbol{\alpha}^\star);\boldsymbol{z}_f)$).

In the rest of this work, we now propose to apply this formalism (voluntarily theoretical) to the case of the optimization of the driver commands of a high-speed train in order to minimize the consumed energy while respecting safety and punctuality criteria.

3. Railway application

The third part of this paper presents the application of the strategy proposed in Section 2 to the optimization of train speed with respect to energy efficiency criteria, under the constraints of safety, comfort, and punctuality. After a brief introduction of the railway application and its formulation under the form of a constrained control optimization problem, this section explains how the dynamical model and the cost function are constructed, as well as the way the parameters on which they depend are estimated from on-track measurements. It concludes with the approximated solving of this problem and shows the efficiency of the proposed approach on a particular journey.

3.1. General context

Although rail transport is often considered to be environmentally friendly due to its low-carbon emissions, it is one

of the largest electricity consumers. For environmental reasons and because of the recent spike in energy prices, railway companies are encouraged to limit their electricity consumption. To do this, one of the levers is to build speed trajectories that limit energy consumption and/or allow for load shedding during demand peaks. The construction of these optimal speed trajectories is however complex for several reasons.

- First, it is not possible to directly choose the train speed, as this speed is only the (nonlinear) consequence of all the solicitations imposed on the train by its environment, including the driver commands that are in charge of the braking and the motor traction. Thus, the driver commands are the real levers of action of the train dynamics.

- Second, if we want to minimize the energy consumed by the train on a particular route by playing on the driver commands, it is necessary to be able to predict the train dynamics associated with a particular control function, as well as the associated consumed (or recovered) energy. Keeping in mind that these models will have to be called many times during the optimization phase, a compromise between realism and numerical efficiency is necessary for these models. Moreover it is clear that many sources of uncertainty will affect these models [16, 17]. For instance, the mechanical properties of the train may change over time due to wear and are therefore not very well determined [18]: the wind causes always changing aerodynamic loads, humidity alters the adhesion between the wheel and the rail, the number of passengers modifies the mass of the carbodies, the efficiency of the traction and braking chain is poorly known...

- In addition, punctuality, safety, and comfort impose nonlinear constraints (as a result of the nonlinear dynamic behavior) on speed and jerk that are also very sensitive to the uncertain model parameters.

Hence, in agreement with the notations of Section 2, for a particular railway track $\mathcal{T}$ and a particular train $\mathcal{V}$, we can denote by u the time evolution of the driver commands during the train journey. Function u is continuous due to mechanical constraints and takes its values in the interval $[-1, 1]$, where -1 (resp. 1) corresponds to the use of a maximal braking torque (resp. traction torque). The relation between the driver commands and the traction and braking forces applied to the train are given in Equations (A.2) and (A.3). The positions and speeds of each mass body of the train are gathered in the state vector $\boldsymbol{x}$, the model representing the train dynamics is noted $\boldsymbol{f}$, and the parameters on which this model rely are noted $\boldsymbol{z}_f$. It is important to notice that the train dynamics at a given instant t, and therefore its state vector $\boldsymbol{x}(t)$, depends on all the values of u at the previous states. In the same manner, the model of the total electrical power is written h and we denote by $\boldsymbol{z}_h$ the parameters that need to be estimated for this model to be evaluated. Assuming that the duration of the journey must be 1 for punctuality reason (up to a time renormalization), the constraints of maximum allowed speed on the track, departure and arrival at the right place can again be formulated in the form of an envelope $\mathcal{E}$ for $\boldsymbol{x}$ at any time between 0 and 1.
At this stage, the problem of minimizing, under speed limitation and punctuality constraints, the energy consumed by a train $\mathcal{V}$ to go from a particular position at $t = 0$ to another specific position at $t = 1$ along a railway track $\mathcal{T}$ can be formulated exactly as the optimization problem defined by Eq. (4).

3.2. Construction and calibration of approximate quick-to-evaluate models

This section briefly presents the mathematical/physical models that have been constructed in order to represent the train behavior, as well as the way the parameters on which they depend are estimated. Only the most useful elements are presented in this section and more details following the works presented in [19] and [20] can be found in Appendix A.

3.2.1. Longitudinal dynamics of the train

In order to limit the numerical costs, only the longitudinal dynamic behavior of the train is considered and it is modeled with rigid bodies as it is done most often in railway dynamics. The model $\boldsymbol{f}$ aims thus at computing the time evolution of the state vector $\boldsymbol{x}$, which gathers the longitudinal positions and speeds of the centers of gravity of the different elements composing the train. This dynamics mostly depends on three external forces (the inner forces have very little impact on total energy consumption):

- the (nonlinear) resistance (gathering contact and aerodynamic) force f^R modelled according to the Davis law [21],
- the (nonlinear) traction f^T and braking forces f^B, which amplitudes depend on the driver commands, and therefore on control function u,
- the projection f^W of the weight on the longitudinal to the track axis in case of ramp (positive declivity) or slope (negative declivity).

Some of the parameters considered in this model vary from one train to another because of damage, of weather, and of the train loading. A sensitivity analysis made it possible to select only four main parameters for the modeling of the train dynamics: the total mass of the train, m, and the three Davis coefficients noted a, b, and c. All these coefficients are gathered in vector $\boldsymbol{z}_f = (m, a, b, c)$.
Finally, the Newton laws of motion allows us to model the train behavior in the form of a nonlinear system of differential equations, which can be rewritten as in Eq. (1) (see Appendix A for more details about the precise relation between function $\boldsymbol{f}$ and the forces f^R, f^T, f^B, and f^W).
This system of differential equations is solved using a Runge-Kutta scheme and we assume that for a given control function u and for a given parameter $\boldsymbol{z}_f$, it admits a unique solution noted $\boldsymbol{x}(\cdot; u, \boldsymbol{z}_f)$.

3.2.2. Energy consumption of the train

The performance of the train is monitored thanks to the electrical power transmitted by the catenary, which can be decomposed in several terms: p^T, the electrical power used for the traction, p^B, the electrical power recovered by the braking, and p^a, the power transmitted to the auxiliary equipment of the train. This latter is poorly known and is considered constant in this work. In addition, models are introduced for the traction efficiency (resp. the braking efficiency), which depend on new coefficients a_η, b_η (c_η and d_η) that also need to be adapted to the considered train. These five uncertain parameters are gathered in vector $\boldsymbol{z}_h = (p^a, a_\eta, b_\eta, c_\eta, d_\eta)$. The energy consumed by the train is finally given by the integral of the total electrical power during the journey:

$$\int_0^1 h\left(\boldsymbol{x}(t; u, \boldsymbol{z}_f), u(t); \boldsymbol{z}_h\right)\, dt. \tag{20}$$

Function h is thus positive during the traction phases but can be negative in case of electrodynamic braking since power may be restored to the network. Once again, we refer the interested reader to Appendix A for more details on the explicit relation between h, p^T, p^B, and p^a.

3.2.3. Bayesian model calibration

The accuracy of the prediction of the train models constructed in Sections 3.2.1 and 3.2.2 strongly depends on the vehicle parameters. As was proposed in [22], a Bayesian calibration [23, 24, 25, 26] is therefore set up to identify the characteristics of the studied train. It amounts to assuming that the value of $\boldsymbol{z}$ to be estimated can be modeled by a random vector $\boldsymbol{Z}$ to account for its uncertain nature. Estimating the value of the vehicle parameters in a Bayesian formalism requires first to introduce a prior probability distribution for $\boldsymbol{Z}$ and then to characterize as well as possible the statistical distribution of $\boldsymbol{Z}$ given observations of the system.

Definition of the prior

The maximum entropy principle [27, 28, 29] is used to define the prior probability distribution of $\boldsymbol{Z}$. It first leads us to assume that the components of $\boldsymbol{Z}$ are independent. It then leads to uniform prior probability distributions for the mass, the auxiliary power, and the traction and braking efficiencies, and to gamma prior probability distributions for the Davis coefficients. Let $p_{\boldsymbol{Z}}$ be this prior probability distribution of $\boldsymbol{Z}$.

Introduction of two model errors

In order to model the various simplifications of the models, two model errors are introduced (see Appendix B for a precise description of these errors):

- the first term, noted ε^f, is similar in nature to a force and is added in the physical model $\boldsymbol{f}$,
- the second term, noted ε^h, is similar in nature to an electrical power and is added in the energy model h.

In the same manner than for the vehicle parameters, these two errors are assumed to be random and they are modeled by two Gaussian centered white noises indexed by the time. They are assumed to be statistically independent, which make them be characterized by two positive constants characterizing their variance, written σ_f^2 and σ_h^2

respectively. These two constants are *a priori* unknown and also need to be identified from system observations.

Bayesian inference using on-track measurements

To identify $\boldsymbol{\sigma} := (\sigma_f, \sigma_h)$ and $\boldsymbol{Z}$, the time evolution of the train dynamics (train speed and position) and the associated energy consumption have been measured on $N = 30$ different journeys. All these measurements, which characterize the maximum information available for parameter inference, were performed with the same high-speed train (without passengers) on the same high-speed railway track, but on different subsections (i.e. with different starting and ending points), at different speeds and different dates, and therefore for different wind conditions to describe as many different situations as possible. These on-track measurements were collected with high accuracy compared to the modeling errors, so that the measurement errors are assumed to be negligible in the following. Let $\mathcal{D}_N$ be the set gathering all these measurements. Wind forecasts along the train path were also available for these journeys. Given these data, a two-step approach, also called *plug-in* approach (see [30] for further justifications about this approach), is proposed to estimate $\boldsymbol{\sigma}$ and $\boldsymbol{Z}$ because of their different nature. First, we denote by $(z, \boldsymbol{\sigma}) \mapsto L_N(z, \boldsymbol{\sigma})$ the likelihood function, which is introduced to compare measurements and model previsions (see Appendix B for a precise definition of the likelihood function). The maximum likelihood estimator of $(\boldsymbol{Z}, \boldsymbol{\sigma})$, that is to say the value of $(z, \boldsymbol{\sigma})$ that maximizes the likelihood function in $\mathbb{Z}\times]0, +\infty[^2$, is then denoted by $(\boldsymbol{Z}^{\mathrm{MLE}}, \boldsymbol{\sigma}^{\mathrm{MLE}})$. The Basin-hopping algorithm [31] is used to solve this problem, because of its robustness and its ability not to get stuck in local minima.

We then propose to fix $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma}^{\mathrm{MLE}}$ (hence the name *plug-in*) and then only focus on the posterior probability distribution of $\boldsymbol{Z}$ given $\boldsymbol{\sigma}^{\mathrm{MLE}}$. In that prospect, the Bayes formula tells us that the probability distribution of $\boldsymbol{Z}$ knowing $\boldsymbol{\sigma}^{\mathrm{MLE}}$ and the available measurements, noted $p_{\boldsymbol{Z}|\mathcal{D}_N,\boldsymbol{\sigma}^{\mathrm{MLE}}}$, can be decomposed as:

$$p_{\boldsymbol{Z}|\mathcal{D}_N,\boldsymbol{\sigma}^{\mathrm{MLE}}}(z) \propto p_{\boldsymbol{Z}}(z) L_N(z, \boldsymbol{\sigma}^{\mathrm{MLE}}), \tag{21}$$

here $\propto$ indicates a proportional relationship. The Metropolis-within-Gibbs (MwG) algorithm [32] is finally used to sample the posterior distributions. It combines the single component Metropolis-Hastings (SCMH) algorithm [33] with the Gibbs Sampling (GS) [34], changing component each 5 iterations. The interest is that the components of $\boldsymbol{Z}$ are considered one-by-one so that the effects of the less influential parameters are not hidden by the most influential ones. Here, several iterations of the Metropolis-Hastings algorithm are performed for each iteration of the Gibbs algorithm to ensure a better homogeneity in the number of different points per component. After several numerical tests, it turned out in particular that performing 5 iterations of Metropolis-Hastings offered an interesting compromise between conditional law exploration and full joint probability distribution exploration. To accelerate convergence, the algorithm is moreover initialized in $\boldsymbol{Z}^{\mathrm{MLE}}$ and the likelihood functions for the N different journeys are computed in parallel. There's however no guarantee that this choice of initialization is optimal, and other choices could have been made (see [35] for more details). The proposal is a normal centered distribution, whose standard deviation is constant and determined empirically for each uncertain parameter to reach an acceptation ratio close to 50%. In order to assess the convergence of the method, we estimate the second-order moment $\mathbb{E}\{\|\boldsymbol{Z}\|^2\}$ (see Figure 1). Its value remains relatively stable after $N^{MwG} = 4900$ iterations. Keeping the iterations following the N^{MwG}-th iteration allows us to gather points that are approximately distributed according to the posterior probability distribution of $\boldsymbol{Z}$ (see Figure 2 for a trace plot of the samples obtained after the N^{MwG}-th iteration). Despite all the precautions taken to maximize the relevance of the inference results, it is nevertheless not possible to guarantee that the inference results are not affected by a slight bias linked to the finite character of the number of MCMC iterations.

The posterior probability distribution of $\boldsymbol{Z}$ is finally reconstructed using a kernel-based method (see [36] for more details) and its marginal distributions are presented in Figure 3. In this figure, we notice that the posterior probability distributions of the components of $\boldsymbol{Z}$ associated with the resistance forces, A, B, and C, remain close to their prior ones. This means that the nominal values used for constructing the prior distributions were well adapted. This is not the case for the posterior distributions of the mass M, the auxiliary power P^a, and the four efficiency coefficients A_η, B_η, C_η, and D_η, for which the posterior distributions are relatively different from the prior ones.

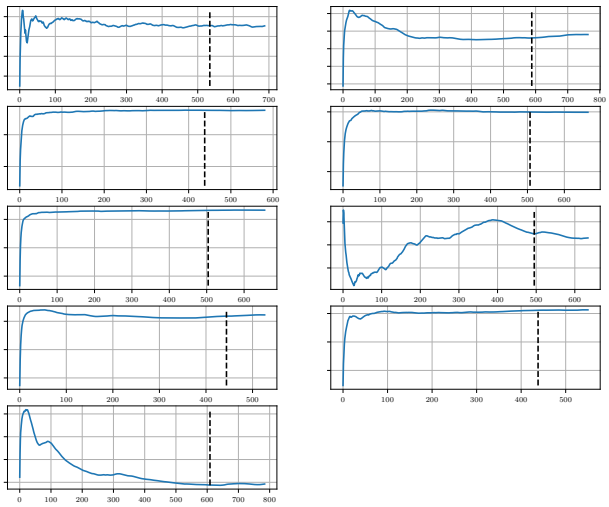

Figure 1: Second-order moment of the uncertain parameters as a function of the number of iterations. From top left to bottom right: ΔM , P^a, A, B, C, A_η, B_η, C_η, and D_η. Iteration number N^{MwG} is represented by a vertical dashed line in each figure.

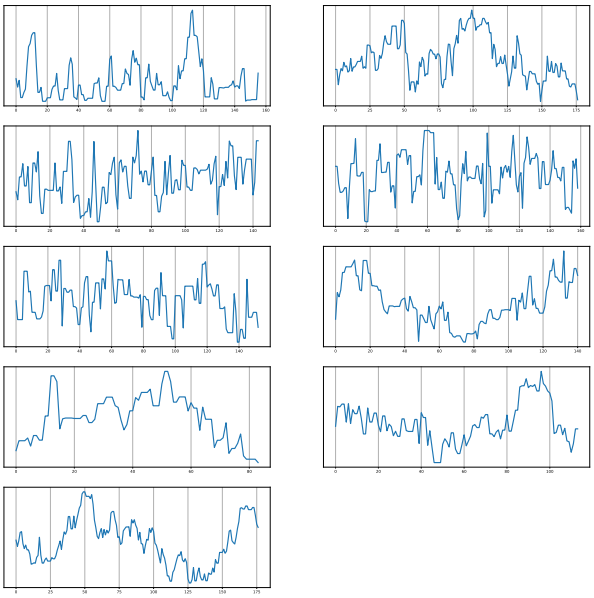

Figure 2: Samples of z obtained after iteration number N^{MwG}. From top left to bottom right: ΔM , P^a, A, B, C, A_η, B_η, C_η, and D_η.

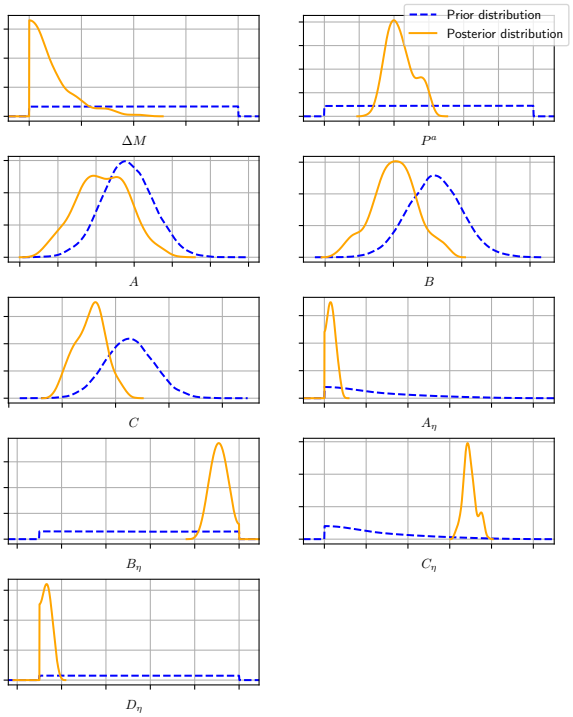

Figure 3: PDFs of the prior (blue dashed line) and posterior (orange solid line) probability distributions for each component of the vector of vehicle parameters $\boldsymbol{Z} = (M, A, B, C, P^a, A_\eta, B_\eta, C_\eta, D_\eta)$. From top left to bottom right (line-by-line): Mass M, auxiliary power P^a, Davis coefficients A, B, C, and the traction and braking efficiency parameters A_η, B_η, C_η, and D_η.

This can be explained by the fact that in those cases, only little information was available for constructing the priors, whereas the energy consumption is sensitive to them.

3.2.4. Model evaluation

In this section, we propose to propagate the uncertainties in the model and to compare in Figures 4 and 5 the measured and simulated energy consumptions. We point out that all these comparisons between model predictions and measurements are associated with the same driver commands, and therefore the same control function u, without taking into account potential operational constraints for the train journey (punctuality, zero speed at arrival, and so on).

Figure 4 corresponds to one of the journeys used for the

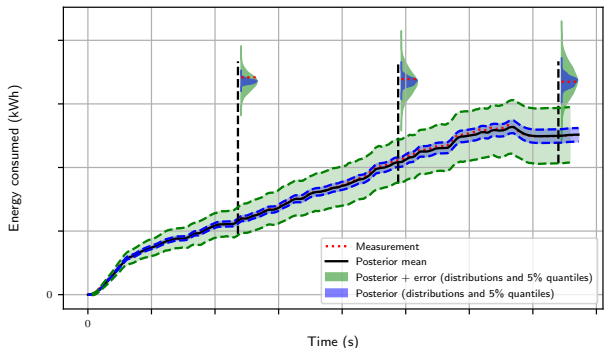

Figure 4: Measurement 1 - (i) Realizations of the random energy consumption generated with the posterior distribution (in blue dashed line), (ii) its mean value (in black solid line), (iii) realizations including the modeling error $\boldsymbol{\varepsilon}^h$ (green curves), and (iv) envelope of the confidence regions for the 95% quantiles. The PDFs of (i) and (iii) are plotted at three given times. Finally, the energy measurements are plotted with a red dotted line

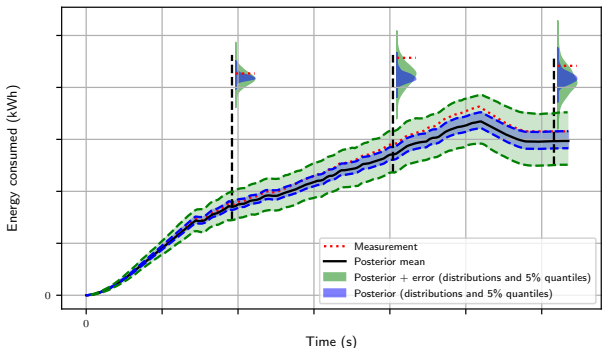

Figure 5: Measurement 2 - (i) Realizations of the random energy consumption generated with the posterior distribution (in blue dashed line), (ii) its mean value (in black solid line), (iii) realizations including the modeling error $\boldsymbol{\varepsilon}^h$ (green curves), and (iv) envelope of the confidence regions for the 95% quantiles. The PDFs of (i) and (iii) are plotted at three given times. Finally, the energy measurements are plotted with a red dotted line

Bayesian inference. In this case the measurements are very close to the mean value estimated with the posterior distribution. In general, most of the energy consumption measurements used in the Bayesian inference are well described by the blue envelope, but some of them are slightly outside (but never outside the green envelope). In other words, the estimated posterior distribution correctly characterizes the variability encountered during all the measured journeys. Note that most of the uncertainty results from the model error since the green envelope (describing the model error) is larger than the blue envelope (describing the parameters' uncertainty).

We also propose to assess the quality of the posterior distribution by plotting the predicted energy consumption for measurements that have not been used during the Bayesian inference. This is shown in Figure 5, using the same conventions as the ones used in Figure 4. Once again, the blue envelope is sufficient to describe the measurements, but in a slightly higher number of journeys, we observe measurements that are outside this blue envelope while remaining inside the green one. All these results justify the importance of the two model errors and seem to indicate a good estimate of the posterior distribution of $\boldsymbol{Z}$, as the model makes it possible to reproduce the measurement with a controlled confidence (once the uncertainties are propagated). Looking at Figures 4 and 5, we can notice that the impact of model errors and uncertainties on $\boldsymbol{Z}$ is far from negligible on the consumed energy, which justifies their use in the optimization problem.

3.3. Optimization of the driver commands to save energy

As explained at the beginning of Section 3, for a given railway journey and a given train, the objective of this work is to identify a deterministic driver command allowing to minimize the energy consumed by the train on average. This command can be used as a guide for current drivers, while preparing the arrival of autonomous trains. Taking into account the operational constraints related to the fact that the train departing from a given point stops at the right place at the right time without exceeding the speed limits is then essential in this research. Without these constraints, which can be written as an envelop $\mathcal{E}$ for the state vector $\boldsymbol{x}$ gathering the position and the speed of the train at each time t, the optimization problem is as trivial as it is useless, as the absence of train motion clearly minimizes energy consumption. The problem is thus complex and the transformation and reduction methods proposed in section 2 are introduced for the solving. The search for the optimal control function $u^\star$ finally requires the solving of the optimization problem defined in Eq. (18):

$$u^\star(t) = u_K(t;\boldsymbol{\alpha}^\star) = \sum_{k=1}^{K} \alpha_k^\star v_k(t),\ 0 \le t \le 1, \qquad (22)$$

$$\boldsymbol{\alpha}^\star := \underset{\boldsymbol{\alpha}\in\mathbb{A}_K}{\arg\min} \int_0^1 \mathbb{E}_{\boldsymbol{Z}}\left[h\begin{pmatrix}\boldsymbol{x}(t;g(u_K(\cdot;\boldsymbol{\alpha});\boldsymbol{Z}_f),\boldsymbol{Z}_f),\\ g(u_K(t;\boldsymbol{\alpha});\boldsymbol{Z}_f);\boldsymbol{Z}_h\end{pmatrix}\right] dt. \tag{23}$$

Here, models $\boldsymbol{f}$ and h are the train dynamics and energy consumption models presented in Sections 3.2.1 and 3.2.2, $\boldsymbol{Z} = (\boldsymbol{Z}_f, \boldsymbol{Z}_h)$ is the random vector gathering the parameters on which these models depend, whose distribution has been estimated in Section 3.2.3. Nevertheless, we note the absence of the model errors in this formulation. Indeed, as these errors play in a linear way on the model outputs and as they have been assumed to be centered (see Appendix B), their impact is cancelled out when we turn to the expectation of the consumed energy. The search set,

$$\mathbb{A}_K := \left\{\boldsymbol{\alpha} \in \mathbb{R}^K \mid u_K(\cdot;\boldsymbol{\alpha}) \in \mathbb{U}_d\right\}, \tag{24}$$

is nevertheless non-trivial, since it assumes the prior identification of a relevant projection basis of dimension K, $\{v_1, \ldots, v_K\}$, while allowing the reconstructed control function to take values in $[-1, 1]$. Note also in this formalism the presence of the function g (described in section 3.3.2), whose role is to allow the control function to verify the operational constraints whatever the values of $\boldsymbol{Z}$.

3.3.1. Definition of the search set

As explained in Section 2.2, the choice of the search set relies on a three-step procedure. Each control function is characterized by its value on a given d-dimensional time grid to limit the dimension of the search space for the optimization. In this work, the discretization step is chosen constant in time and corresponds to the case where the driver can modify the commands every $4\,s$ (which is a reasonable value for the driver), which leads to $d = 200$ (for a journey of about $800\,s$). Given this fixed time discretization, the (deterministic) constrained optimization problem defined by Eq. (5) is solved for $M = 100$ values of $\boldsymbol{Z}$ picked at random according to its posterior probability distribution. Let us denote by $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(M)}$ these sampled values of $\boldsymbol{Z}$. To solve these problems, we focus on the Covariance Matrix Adaptation - Evolution Strategy (CMA-ES), when using an adaptive augmented Lagrangian method to handle the constraints. In brief, at each step, the algorithm randomly draws a small population of points (18 in the considered application), according to a multivariate Gaussian distribution. For each of these points, the algorithm evaluates the cost function, possibly penalized by the constraints. From these evaluations, it updates the center point and the covariance matrix of the Gaussian distribution, which will be used at the next iteration. More details about the general method are given in [37, 38, 39], and indications on how to adapt the hyperparameters of the algorithm, such as the size of the population, the adaptation coefficient or the number of draws selected are given in [40]. Let $u^{(1)}, \ldots, u^{(M)}$ be the M optimal control functions associated with the M chosen values of $\boldsymbol{Z}$. In addition to these solutions, 30 000 control functions close to verify the constraints are extracted from all the control functions tested during these optimizations, from which an empirical estimate of the control function covariance was constructed. We finally keep the $K = 60$ first eigenfunctions associated with this estimate to define the reduced basis $\{v_1, \ldots, v_K\}$, corresponding to a conservation on average of 99% of the information contained in each control function.

3.3.2. Definition of a transformation of the control function

The introduction of the function g is mainly motivated by the fact that, for the considered application, it is not possible for a single control function u to be valid from the constraints point of view when associated with different values of $\boldsymbol{Z}$. As an illustration, Figure 6 shows the predicted arrival positions (using model $\boldsymbol{f}$) of a train when associating the same control function u (the same kind of conclusions would have been obtained for another control function) with different realizations of $\boldsymbol{Z}$ chosen according to its posterior probability distribution. We thus note that even if on average the train arrives at the right place (distance 0 with respect to the target position), the uncertainties on $\boldsymbol{Z}$ can lead to an arrival position that can be $750\,m$ away from the target position, which would be unacceptable.

As explained in Section 2.2, the choice of g strongly depends on the studied problem. For the railway application presented in this work, this function is associated with two distinct modifications of the control function, corresponding respectively to the verification of the speed limit

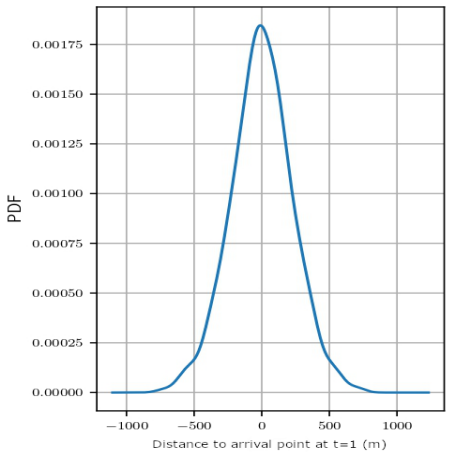

Figure 6: Distribution of the distances to the arrival point at arrival time $t = 1$ obtained for one choice of u and 100 values of $\boldsymbol{Z}$.

and the punctuality. First, for a given value z of $\boldsymbol{Z}$, if the control function does not respect the speed limitation at a given time, we propose to directly modify it so that the train speed reaches the speed limitation without overtaking it (saturation of the constraint). This can be done explicitly using the train dynamics equations when playing on the braking and traction forces. If it is necessary to modify the control command at several time steps, the algorithm sets the value of the control function to -1 at these time steps (braking with maximum amplitude) until the train speed is below the speed limitation. As braking is equivalent to a loss of energy, these candidates will not be selected by the algorithm.

Regarding the punctuality constraints, we then propose to multiply the whole control function by two constants: $c_T(z)$ is a factor applied on the traction parts of the journey and $c_B(z)$ is a factor applied on the braking parts only. Dichotomy approaches can finally be applied to find these values of $c_T(z)$ and $c_B(z)$ in very few iterations.

3.3.3. Solving the final constrained optimization problem

Once search set $\mathbb{A}_K$ and function g have been defined, we can turn to the solving of the optimization problem defined by Eq. (23). First, since the computation of the expectation in $\boldsymbol{Z}$ cannot be performed explicitly, we propose to replace it by an empirical estimate based on 30 independent and identically distributed realizations of $\boldsymbol{Z}$ renewed at each step of the algorithm. Then, this problem being *a priori* nonlinear and nonconvex, we propose to apply again the CMA-ES algorithm to solve this problem. As an evolutionary strategy, this algorithm makes a small population of points evolve in $\mathbb{A}_K$ following a multivariate Gaussian distribution, whose mean vector and covariance matrix are re-estimated at each iteration. To accelerate the convergence of this algorithm, the initial value of the mean vector is chosen equal to the empirical mean of the projection coefficients obtained when projecting the solutions of the constrained optimization problem defined by Eq. (5) on $\{v_1, \ldots, v_K\}$. The initial value of the covariance matrix is defined as the identity matrix, not to favor a specific direction. The size of the population of new search points is $N^{pop} = 16$ points, as proposed by the semi-empirical population size, which is equal to $4 + 3\log(K)$. The other hyperparameters are chosen in order that the algorithm explores a relatively large range of possible solutions.

The evolution of the cost function with respect to the iteration number i of the CMA-ES algorithm is shown in Figure 7. During the first iterations, we see that the mean value decreases rapidly with i and that the convergence seems relatively well stabilized after $i = 4\,000$. This can be explained from the sample approximation of the expectation value of the cost function. Indeed, for each point of the population, the mean sample of the expectation value will be slightly deviated from its true value. During the first iterations of CMA-ES algorithm, the deviation will be small compared to the gain of the cost function, so the algorithm will manage to find an interesting direction to explore. But when the algorithm is close to convergence, this deviation is no longer negligible and the algorithm shows difficulties to converge. Note that it is expected that the width of the grey envelope can be reduced by further discretizing the expectation in $\boldsymbol{Z}$, but at the cost of a higher numerical cost. For the optimal value, we propose to take the mean of the last 500 iterations.

Remark. Even if evaluating the performance of a particular control function only takes 4 seconds for the considered journey, doing it for 30 values of $\boldsymbol{Z}$, for all the $N^{pop} = 16$ points, and for a large number of CMA-ES steps can be very long. In practice, two different termination criteria were selected regarding the dispersion of the new search points and the cost function decrease between

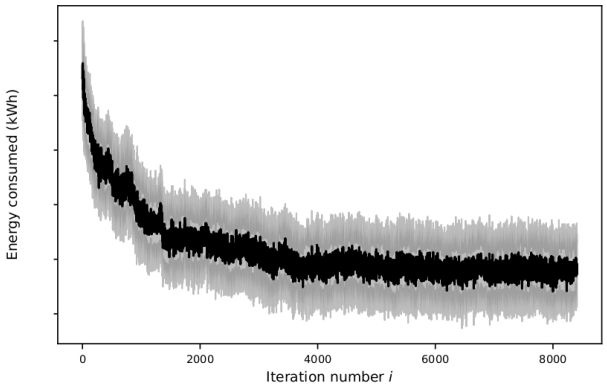

Figure 7: Convergence of the cost function depending on the iteration number i: mean value over the population (black solid line), standard deviation envelope of the population (grey envelope).

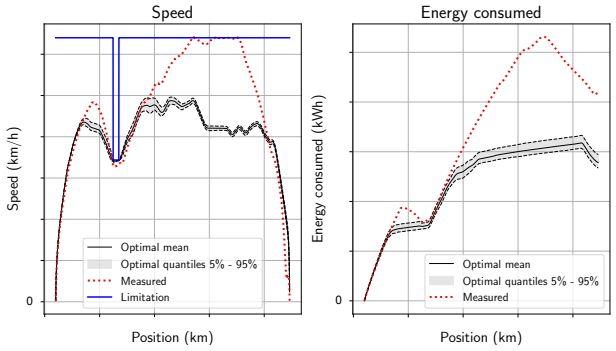

Figure 8: Speed profile (left) and energy consumption (right) associated with the transformed optimal driver commands $g(u_K(\cdot;\boldsymbol{\alpha}^\star);\mathbf{Z}_f)$. The black lines represent the mean value and the envelopes stand for the quantile intervals empirically estimated from realizations of $\mathbf{Z}$. The measurements are plotted in red and the speed limitation is in blue.

two steps, leading to approximately 8 000 CMA-ES steps. With a 30 cores parallelization of the model evaluations, this has represented about 6 days of computation for the estimation of optimal coefficients $\boldsymbol{\alpha}^\star$. Working on methods to speed up these evaluation times is a very important perspective for this work.

3.3.4. Analysis of the results

The efficiency of the proposed method is now evaluated for a particular railway journey, for which on-track measurements of the train speed and energy consumption were available. Hence, Figure 8 compares to the measurements what could be the train speed profile and the associated energy consumption, if the optimal control function was applied. We note strong differences in speed, which translate into a significant and very promising gain in energy for the proposed method (about 25% of energy saved).

Let us notice that we are representing in this figure not a single speed profile (and the associated energy consumption), but the 95% interquantiles and the average of the speeds (and consumed energies) when varying the values of $\boldsymbol{Z}$ according to its posterior probability distribution. For each value of $\boldsymbol{Z} = (\boldsymbol{Z}_f, \boldsymbol{Z}_h)$ tested, it is not directly $u^\star = u_K(\cdot;\boldsymbol{\alpha}^\star)$ that is applied, but $g(u_K(\cdot;\boldsymbol{\alpha}^\star);\boldsymbol{Z}_f)$ so that the constraints are well respected. This can explain that, in spite of the variability of the train configurations, the optimal speed profile has a variance close to 0 (The function g attenuating the influence of $\boldsymbol{Z}_f$).

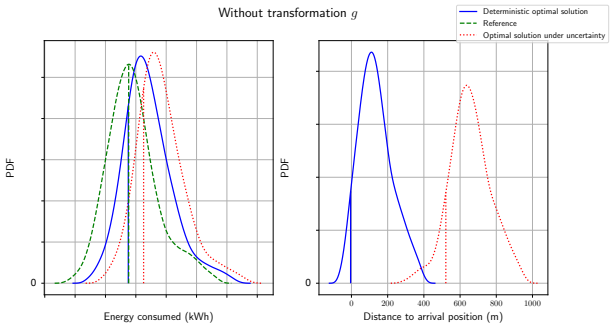

Figure 9: Normalized PDFs of the consumed energy (left) and of the distance to arrival position (right) for the optimal solution of the deterministic problem u^{mode} (in blue), the family of optimal solutions of the deterministic problems $u^{(1)},\dots,u^{(M)}$ (in green), and the optimal solution under uncertainty $u_K(\cdot;\boldsymbol{\alpha}^\star)$ (in red) empirically estimated from realizations $z^{(1)},\dots,z^{(M)}$ of $\mathbf{Z}$. z^{mode} is represented with a vertical dotted line.

We now propose to evaluate the relevance of the probabilistic formulation of the problem, compared to a more classical deterministic formulation. To this end, we denote by z^{mode} the (deterministic) most likely value of $\boldsymbol{Z}$ and by u^{mode} the optimal control function associated with z^{mode}, that is to say the solution of problem (13) for $\boldsymbol{Z} = z^{\text{mode}}$. We then compare in Figures 9 and 10 the energy consumptions (computed using models $\boldsymbol{f}$ and h) that would be obtained if u^{mode} or $u_K(\cdot;\boldsymbol{\alpha}^\star)$ was chosen as control function for the $M = 100$ values of $\boldsymbol{Z}$ introduced in Section 3.3.1 and noted $z^{(1)},\dots,z^{(M)}$. A reference is

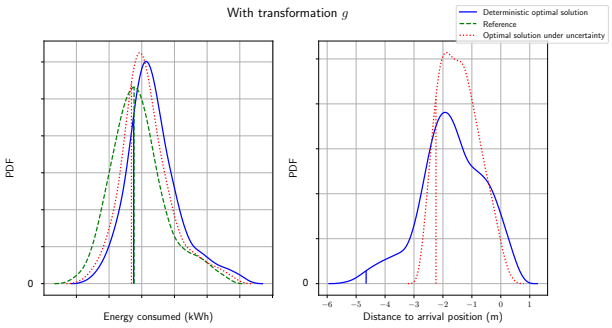

Figure 10: Normalized PDFs of the consumed energy (left) and of the distance to arrival position (right) for the transformed optimal solution of the deterministic problem $g(u^{\text{mode}}; \boldsymbol{Z}_f)$ (in blue), the family of optimal solutions of the deterministic problems $u^{(1)}, \ldots, u^{(M)}$ (in green), and the transformed optimal solution under uncertainty $g(u_K(\cdot; \boldsymbol{\alpha}^\star); \boldsymbol{Z}_f)$ (in red) empirically estimated from realizations $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(M)}$ of $\boldsymbol{Z}$. $\boldsymbol{z}^{\text{mode}}$ is represented with a vertical dotted line.

added to this graph, which corresponds to the consumed energy when applying for each $\boldsymbol{z}^{(k)}$ the optimal control function $u^{(k)}$ obtained in Section 3.3.1. These values are named as reference because they are optimal in terms of energy saving while verifying the constraints for each realization $\boldsymbol{z}^{(k)}$. Nevertheless, they have no practical reality as they suppose that the train characteristics are perfectly known in advance, which is not the case. The consumed energy (and the associated distance to target position) for the particular case $\boldsymbol{Z} = \boldsymbol{z}^{\text{mode}}$ is represented with a vertical line.

Focusing first on Figure 9, we see that u^{mode} allows to further reduce the consumed energy than $u_K(\cdot; \boldsymbol{\alpha}^\star)$ (while being less efficient than the reference as expected), but all this at the cost of a poor respect of the constraints (the arrival position is in average several hundreds meters away from the target position). We also observe that these final position constraints are even less verified for the values associated with $u_K(\cdot; \boldsymbol{\alpha}^\star)$. This just means that the $u_K(\cdot; \boldsymbol{\alpha}^\star)$ function relies heavily on the transformation g to verify the constraints, and therefore has no real meaning taken alone.

Focusing on Figure 10, we see that the constraints are correctly taken into account (up to the numerical tolerance of the minimization algorithm that was $\pm 5\,m$ for the target distance) when applying the transformation g on u^{mode} and $u_K(\cdot; \boldsymbol{\alpha}^\star)$. We also notice that the performance of the solution in terms of energy saving is reduced as the PDFs are shifted to the right (higher energy consumption). As expected, the reference (in green) is still the best solution, but we see that this time, the solution of the problem under uncertainty (in red) is closer to the reference than the solution of the deterministic one (in blue). Since the deterministic solution is already particularly efficient for this particular journey, this gain is nevertheless smaller than the one obtained when compared to the measurement. Comparing the vertical bars, we also observe that the solution under uncertainties behaves slightly better than the deterministic solution for $\boldsymbol{Z} = \boldsymbol{z}^{\text{mode}}$. This result was not expected *a priori*. But it can be explained *a posteriori* by the fact that the deterministic problem being solved in a space of dimension $d = 200$ may have more difficulties to identify the global optimum than the problem under uncertainties, which is defined on a space of dimension $K = 60$.

4. Conclusions

This paper focuses on the optimization of driver commands to minimize the energy consumed by high-speed trains, under punctuality and safety constraints, using simplified physical models for the train dynamics and its energy consumption. This problem can be seen as a particular instance of a constrained control optimization in presence of uncertainties. Indeed, the first difficulty of the presented problem comes from the unknown nature of several parameters characterizing the models on which it depends. A Bayesian formalism exploiting on-track measurements made on commercial trains is thus presented to identify them. But when the residual uncertainties associated with these parameters are propagated in the optimization problem, not only the objective function becomes random, but also the search set through the imposed constraints. This difficulty led us to propose a reformulation of the optimization problem, so that it can correctly integrate these uncertainties. The solving of the reformulated problem poses another series of difficulties. Indeed, the new optimization problem has no reason to be convex, the gradient of the cost function is most of the time not accessible, and the solution has *a priori* to be searched in a space of infinite dimension. This motivated the development of a novel method to approach the solution of this problem in reasonable time. This method is

mostly based on dimension reduction techniques and the use of evolutionary algorithms.

From a methodological point of view, the approaches proposed in this paper to solve the optimization problem are presented in a general framework and their interest is not limited to the railway application motivating this work. From an application point of view, this work seems to make it possible to envisage significant gains in terms of energy consumption. Testing these identified driver command profiles during future measurement campaigns would indeed be very interesting. These optimized controls could thus serve as a guide for drivers, but also as instructions for future trains that are expected to be increasingly autonomous.

Many directions might be explored to extend this work. For instance, we can think of configurations managing several trains, or taking into account traffic delays or extreme weather events such as a violent wind or snow.

Appendix A. Definition of the train dynamics and electrical power models

This appendix details the physical models introduced to model the train behavior. The first section focuses on the description of the longitudinal dynamics model. The second section presents the energy consumption model of the train. More details about these models can also be found in [19] and [20].

To begin, let us consider a train $\mathcal{V}$, moving in an environment $\mathcal{T}$ (including the railway track, the wind, ...). The driver commands are denoted with the time-dependent function u, which takes its values in the interval $[-1, 1]$: -1 corresponds to the position of the manipulator during maximal braking and 1 stands for the position of the manipulator during maximum traction. At each time $t \in [0, 1]$, we note $y(t) := y(t; u, \mathcal{T}, \mathcal{V})$ the longitudinal position of the train, $\dot{y}(t) := \dot{y}(t; u, \mathcal{T}, \mathcal{V})$ its longitudinal speed, and $\ddot{y}(t) := \ddot{y}(t; u, \mathcal{T}, \mathcal{V})$ its longitudinal acceleration.

Appendix A.1. Definition of the train dynamics model

A first approach is to consider the train as a set of rigid bodies interacting with each other as it is done in [41]. This approach describes the complete train with about 50 rigid bodies (wheelsets, bogies, and carbodies). And for each of them, we have to solve 6 equations (one for each degree of freedom in rotation and translation). In practice, solving the train dynamics by this approach approximately takes 2 minutes for a 5 km journey using commercial softwares such as Vampire [42]. It seems obvious that this approach is not applicable in the context of our study (journey of 100 km and several tens of thousands of evaluations to solve one deterministic problem (5)).

For this reason, a second approach consists in considering only the longitudinal direction of the problem and considering the train as a single element. This approach reduces the number of equations to be solved to only one at the cost of simplifications whose validity have to be verified. The passage from the first to the second approach is carried out in detail in [20].

The dynamics of the train mainly depends on three external forces. First, the resistance force (also known as Davis force) f^R approaches the contact and the aerodynamic forces projected on the longitudinal axis. It depends on the train speed $\dot{y}(t)$ but also on the mean speed of the wind v^w projected on the track at the train position $y(t)$. The resistance force is detailed in [21] and is given by:

$$\begin{aligned} f^R\left(\dot{y}(t), v^w(y(t), \mathcal{T}), \mathcal{V}\right) &= a + b\dot{y}(t) \\ &+ c\left(\dot{y}(t) - v^w(y(t), \mathcal{T})\right)\left|\dot{y}(t) - v^w(y(t), \mathcal{T})\right|. \end{aligned} \tag{A.1}$$

Coefficient c refers to an aerodynamic coefficient, coefficient b is associated with the dynamic friction, and coefficient a refers to the static friction. Thus, a and b correspond to the longitudinal contact force and c to the longitudinal aerodynamic force applied to the train. A corrective term is often added in curve, where the friction force is naturally higher.

The traction and braking forces f^T and f^B are then calculated from the driver commands and the maximum capacity of the motors or brakes $f^{T,max}$ and $f^{B,max}$ (which depends on the train speed and the vehicle) as

$$f^T(\dot{y}(t), u(t), \mathcal{V}) = \begin{cases} u(t) f^{T,max}(\dot{y}(t), \mathcal{V}) & \text{if } u(t) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{A.2}$$

$$f^B(\dot{y}(t), u(t), \mathcal{V}) = \begin{cases} 0 & \text{if } u(t) > 0, \\ u(t) f^{B,max}(\dot{y}(t), \mathcal{V}) & \text{otherwise.} \end{cases} \tag{A.3}$$

Empirical models based on experiments are chosen for

these maximal capacities. For confidentiality reasons, these models are however not provided in details here.

Finally, the weight affects the dynamics of the train when the declivity of the track is different from zero. Its projection on the track axis is expressed as a function of the train mass m, the gravity g, and the track gradient θ at the train position $y(t)$ as

$$f^W(y(t), \mathcal{T}, \mathcal{V}) = -mg\theta(y(t), \mathcal{T}) . \quad (A.4)$$

The dynamics of the train is described by the following Newton law:

$$\begin{aligned} mk^{rot}\ddot{y}(t) = f^T(\dot{y}(t), u(t), \mathcal{V}) - f^B(\dot{y}(t), u(t), \mathcal{V}) \\ - f^R(\dot{y}(t), v^w(y(t), \mathcal{T}), \mathcal{V}) + f^W(y(t), \mathcal{T}, \mathcal{V}) , \end{aligned} \quad (A.5)$$

for $t \in [0, 1]$ and with initial Cauchy conditions. It is assumed that for given u, $\mathcal{T}$, and $\mathcal{V}$, and without constraints, this nonlinear differential equation has a unique solution. The left-hand side of Eq. (A.5) represents the inertial term, including the rotation of the wheels thanks to the corrective factor k^{rot}. In the literature, we often find $k^{rot} = 1.04$.

To match the notations of the first section, the state vector is therefore written:

$$\boldsymbol{x}(t) = \begin{pmatrix} y(t) \\ \dot{y}(t) \end{pmatrix} . \quad (A.6)$$

Vector z_f contains all the parameters of the dynamic model, that is to say $z_f = (\mathcal{T}, \mathcal{V})$. We propose to only keep the most influential model parameters to simplify the notations (the others are fixed at their nominal value). In this case, we have $z_f = (m, a, b, c)$.

Finally, using Eq. (A.5), it is possible to explicit the nonlinear function $\boldsymbol{f}$ introduced in Eq. (1).

The elapsed time for one evaluation justifies the use of this second approach consisting in a longitudinal model. Indeed, it is much shorter than the one obtained with the rigid body model (4 seconds are enough for a 100 km journey). This reduction of the computation time results from several simplifications and approximations (longitudinal dynamics, Davis forces, ...) introducing a source of error in our model. This source of error needs to be monitored and will justify the introduction of an error model (see Section 3.2.3).

Appendix A.2. Definition of the electrical power model

To estimate the energy consumed by the train, we need to build a model for the efficiency of the traction chain η^T and for the braking energy recovery η^B. As explained in [20], the following models can be considered:

$$\eta^T(\dot{y}(t), u(t), \mathcal{V}) = a_\eta f^T(\dot{y}(t), u(t), \mathcal{V})\dot{y}(t) + b_\eta , \quad (A.7)$$

$$\eta^B(\dot{y}(t), u(t), \mathcal{V}) = c_\eta f^B(\dot{y}(t), u(t), \mathcal{V})\dot{y}(t) + d_\eta , \quad (A.8)$$

with a_η, b_η, c_η, and d_η four coefficients that need to be identified (as it is done in Section 3.2.3). From these efficiency models, we can deduce the electrical power consumed by the traction p^T and recovered by the braking p^B, which are given by:

$$\eta^T(\dot{y}(t), u(t), \mathcal{V})\, p^T(\dot{y}(t), u(t), \mathcal{V}) = f^T(\dot{y}(t), u(t), \mathcal{V})\dot{y}(t) , \quad (A.9)$$

$$p^B(\dot{y}(t), u(t), \mathcal{V}) = \eta^B(\dot{y}(t), u(t), \mathcal{V})\, f^B(\dot{y}(t), u(t), \mathcal{V})\dot{y}(t) . \quad (A.10)$$

Note here that the traction and braking efficiencies are defined from two different points of view. On the one hand, the traction efficiency describes the quantity of electrical traction power converted into mechanical power. On the other hand, the braking efficiency is defined in inverse, that is to say the quantity of mechanical power that is injected in the catenary. This choice of notation allows the efficiencies to be always in $[0, 1]$.

Finally, the total electrical power p^E is calculated from the power consumed by the traction chain p^T, the power recovered during braking p^B, and the auxiliary power p^a, i.e. the power transmitted to the elements of the train that do not contribute directly to the traction (air conditioning, ...). This relation is given by:

$$p^E(\dot{y}(t), u(t), \mathcal{V}) = p^T(\dot{y}(t), u(t), \mathcal{V}) - p^B(\dot{y}(t), u(t), \mathcal{V}) + p^a . \quad (A.11)$$

Since we are trying to minimize the consumed energy, i.e. the integral of the electrical power, the function h introduced in Section 2 can be chosen equal to the electrical power p^E defined in Eq. (A.11), with $z_h = (p^a, a_\eta, b_\eta, c_\eta, d_\eta)$.

Once again, the simplifications made to build this energy consumption model (simplified energy efficiency model, constant auxiliary power, ...) has the advantage of being fast to evaluate (less than one second). Nevertheless, these simplifications may also introduce some errors in the model, which have to be quantified (see Section 3.2.3).

Appendix B. Definition of the likelihood function

To define the likelihood function, we need to quantify the distance of a simulation with respect to reality. In particular, this implies being able to quantify the error of the model (presented in Appendix A) and the error linked to the measurements. In our case, we have access to energy consumption and dynamics (position and speed) measurements carried out on a commercial train. These measurements are associated with an error whose variance is very low. Thus, we assume that the measurement error can be neglected and we only integrate the model error source in the representation.

Another difficulty lies in the incompleteness of the measurements. Indeed, solving the inverse problem in order to identify the *a posteriori* distributions of the uncertain parameters requires to have dynamic and energy consumption measurements knowing the train environment, as well as the driver commands used during the measurements. In our case, we do not have access to these commands. We therefore propose to use the dynamic measurements to identify in inverse the "experimental" commands before identifying the uncertain parameters from the energy measurements (see [19] for more details).

The likelihood function is thus defined from two distinct sources of model error:

- A first source of error is associated with the dynamic model. It is represented by a centered additive Gaussian process ε^f whose covariance matrix (once discretized in time) is diagonal, with constant coefficients equal to σ_f^2, which is added in the expression of the traction and braking forces (Eq. (A.2) and (A.3)):

$$\begin{cases} F^{T,mod}\Big(\dot{y}(t), u(t), \boldsymbol{Z}, \sigma_f\Big) = f^T(\dot{y}(t), u(t), \boldsymbol{Z}) + \varepsilon^f(\sigma_f), \\ F^{B,mod}\Big(\dot{y}(t), u(t), \boldsymbol{Z}, \sigma_f\Big) = f^B(\dot{y}(t), u(t), \boldsymbol{Z}) + \varepsilon^f(\sigma_f). \end{cases} \tag{B.1}$$

- A second centered Gaussian process ε^h is added in the calculation of the electrical power (Eq. (A.11)). Its covariance matrix is also assumed diagonal (once discretized in time) with coefficients constant and equal to σ_h^2. It comes:

$$\begin{aligned} P^{E,mod}(\dot{y}(t), u(t), \boldsymbol{Z}, \boldsymbol{\sigma}) = P^{T,mod}\Big(\dot{y}(t), u(t), \boldsymbol{Z}, \sigma_f\Big) \\ - P^{B,mod}\Big(\dot{y}(t), u(t), \boldsymbol{Z}, \sigma_f\Big) + P^a + \varepsilon^h(\sigma_h), \end{aligned} \tag{B.2}$$

with $\boldsymbol{\sigma} := (\sigma_f, \sigma_h)$.

Once the structure of the model error has been defined, we can propagate it in the models to quantify its impact on the energy consumption. It is described by the random process ε^{mod} such that

$$\mathcal{F}^{mod}(\dot{y}(t), u(t), \boldsymbol{Z}, \boldsymbol{\sigma}) = \mathcal{F}(\dot{y}(t), u(t), \boldsymbol{Z}) + \varepsilon^{mod}(\boldsymbol{Z}, \boldsymbol{\sigma}). \tag{B.3}$$

It can easily be verified that ε^{mod} is also a centered Gaussian process (from the propagation of ε^f and ε^h). Once discretized in time, it is described by a random vector $\boldsymbol{\varepsilon}^{mod}$ whose covariance matrix can be expressed using σ_f and σ_h (this development is realized in [20]).

Finally, the local likelihood function L_i for a given environment $\mathcal{T}_i$ is directly deduced from $\boldsymbol{\varepsilon}^{mod}$. Since the train speed $\dot{y}$ depends on the realization $\boldsymbol{z}$ of $\boldsymbol{Z}$, the discretized driver commands $\boldsymbol{u}$, and the environment $\mathcal{T}_i$, we propose to write the local likelihood function L_i with its original dependencies as presented below:

$$L_i(\boldsymbol{e}; \boldsymbol{z}, \boldsymbol{u}, \mathcal{T}_i, \boldsymbol{\sigma}) = g_{\boldsymbol{\varepsilon}^{mod}(\boldsymbol{z}, \boldsymbol{\sigma})}(\boldsymbol{e} - \boldsymbol{\mathcal{F}}(\boldsymbol{z}, \boldsymbol{u}, \mathcal{T}_i)), \tag{B.4}$$

in which $g_{\boldsymbol{\varepsilon}^{mod}(\boldsymbol{z}, \boldsymbol{\sigma})}$ is the multivariate Gaussian probability density function with zero mean vector and with covariance matrix $[C^{mod}]$ (it can be verified that this covariance matrix is invertible for all $\boldsymbol{z}$ and $\mathcal{T}_i$). The complete likelihood function L_N is the product of all the local likelihood function L_i (as they are supposed independent to each other), such that

$$L_N(\boldsymbol{e}; \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{\mathcal{T}}, \boldsymbol{\sigma}) = \prod_{i=1}^{N} L_i(\boldsymbol{e}; \boldsymbol{z}, \boldsymbol{u}, \mathcal{T}_i, \boldsymbol{\sigma}). \tag{B.5}$$

As explained in the body of the paper, this likelihood function will be optimized to identify values for σ_f and σ_h and will be used in the MwG algorithm once applied to the measurements. Thus, one evaluation of the likelihood function is equivalent to N identifications in inverse of the "experimental" driver commands, as well as N direct evaluations of the models $\boldsymbol{f}$ and h to estimate the energy consumption. This computational capacity remains acceptable thanks to the simplified models presented in Appendix A, especially since the evaluations of the different local likelihood functions can be parallelized for each of the N journeys.

[1] K. M. Hanson, A framework for assessing uncertainties in simulation predictions, Physica D: Nonlinear Phenomena 133 (1999) 179–188.

[2] T. R. Bewley, R. Temam, M. Ziane, A general framework for robust control in fluid mechanics, Physica D: Nonlinear Phenomena 138 (2000) 360–392.

[3] P. Wang, R. M. Goverde, Two-train trajectory optimization with a green-wave policy, Transportation Research Record 2546 (2016) 112–120.

[4] P. Wang, A. Trivella, R. M. Goverde, F. Corman, Train trajectory optimization for improved on-time arrival under parametric uncertainty, Transportation Research Part C: Emerging Technologies 119 (2020) 102680.

[5] R. Chevrier, P. Pellegrini, J. Rodriguez, Energy saving in railway timetabling: A bi-objective evolutionary approach for computing alternative running times, Transportation Research Part C: Emerging Technologies 37 (2013) 20–41.

[6] Z. Michalewicz, D. Dasgupta, R. G. Le Riche, M. Schoenauer, Evolutionary algorithms for constrained engineering problems, Computers & Industrial Engineering 30 (1996) 851–870.

[7] H. Ko, T. Koseki, M. Miyatake, Application of dynamic programming to the optimization of the running profile of a train, WIT Transactions on The Built Environment 74 (2004).

[8] D. Dutykh, A brief introduction to pseudo-spectral methods: application to diffusion problems, arXiv preprint arXiv:1606.05432 (2016).

[9] R. R. Liu, I. M. Golovitcher, Energy-efficient operation of rail vehicles, Transportation Research Part A: Policy and Practice 37 (2003) 917–932.

[10] H. Stark, J. W. Woods, Probability, random processes, and estimation theory for engineers, Prentice-Hall, Inc., 1986.

[11] R. Ghanem, P. Spanos, Polynomial chaos in stochastic finite elements, Journal of Applied Mechanics Transactions of teh ASME 57 (1990) 197–202.

[12] M. Loeve, Probability theory, Courier Dover Publications, 2017.

[13] M. Williams, The eigenfunctions of the karhunen-loeve integral equation for a spherical system, Propabilistic Engineering Mechanics 26 (2011) 202–207.

[14] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, A posteriori error and optimal reduced basis for stochastic processes defined by a finite set of realizations, SIAM/ASA J. Uncertainty Quantification 2 (2014) 745–762.

[15] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Karhunen–loève expansion revisited for vector-valued random fields: Scaling, errors and optimal basis., Journal of Computational Physics 242 (2013) 607–622.

[16] C. Funfschilling, G. Perrin, S. Kraft, Propagation of variability in railway dynamic simulations: application to virtual homologation, Vehicle System Dynamics 50 (2012) 245–261.

[17] C. Funfschilling, G. Perrin, Uncertainty quantification in vehicle dynamics, Vehicle system dynamics 57 (2019) 1062–1086.

[18] D. Lebel, C. Soize, C. Funfschilling, G. Perrin, High-speed train suspension health monitoring using computational dynamics and acceleration measurements, Vehicle System Dynamics 58 (2020) 911–932.

[19] J. Nespoulous, C. Soize, C. Funfschilling, G. Perrin, Optimisation of train speed to limit energy consumption, Vehicle System Dynamics 60 (2022) 3540–3557.

[20] J. Nespoulous, Constrained optimization under uncertainty of the driver's command for energy saving of high-speed trains using computational stochastic nonlinear dynamics and statistics, Ph.D. thesis, Universit Gustave Eiffel, 2022.

[21] W. J. Davis, The Tractive Resistance of Electric Locomotives and Cars, volume 29, General Electric, 1926.

[22] D. Lebel, C. Soize, C. Fünfschilling, G. Perrin, Statistical inverse identification for nonlinear train dynamics using a surrogate model in a bayesian framework, Journal of Sound and Vibration 458 (2019) 158–176.

[23] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, volume 160, Springer Science & Business Media, 2006.

[24] P. Congdon, Bayesian Statistical Modelling, John Wiley & Sons, 2007.

[25] C. Soize, R. G. Ghanem, C. Desceliers, Sampling of bayesian posteriors with a non-gaussian probabilistic learning on manifolds from a small dataset, Statistics and Computing 30 (2020) 1433–1457.

[26] P. Lu, P. F. Lermusiaux, Bayesian learning of stochastic dynamical models, Physica D: Nonlinear Phenomena 427 (2021) 133003.

[27] R. M. Gray, Entropy and Information Theory, Springer Science & Business Media, 2011.

[28] J. N. Kapur, H. K. Kesavan, Entropy optimization principles and their applications, in: Entropy and Energy Dissipation in Water Resources, Springer, 1992, pp. 3–20.

[29] C. Soize, Maximum entropy approach for modeling random uncertainties in transient elastodynamics, The Journal of the Acoustical Society of America 109 (2001) 1979–1996.

[30] J. Fessler, Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood) : Applications to tomography, IEEE transactions on image processing : a publication of the IEEE Signal Processing Society 5 (1996) 493–506.

[31] D. J. Wales, J. P. Doye, Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms, The Journal of Physical Chemistry A 101 (1997) 5111–5116.

[32] D. Van Ravenzwaaij, P. Cassey, S. D. Brown, A simple introduction to markov chain monte–carlo sampling, Psychonomic Bulletin & Review 25 (2018) 143–154.

[33] H. Haario, E. Saksman, J. Tamminen, Componentwise adaptation for high dimensional mcmc, Computational Statistics 20 (2005) 265–273.

[34] S. Geman, D. Geman, Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6 (1984) 721–741.

[35] M. Betancourt, A conceptual introduction to hamiltonian monte carlo, arXiv preprint arXiv:1701.02434 (2017).

[36] G. Perrin, C. Soize, N. Ouhbi, Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints, Journal of Computational Statistics and Data Analysis 119 (2018) 139–154.

[37] N. Hansen, The cma evolution strategy: a comparing review, Towards a new evolutionary computation: Advances in the estimation of distribution algorithms (2006) 75–102.

[38] K. Nishida, Y. Akimoto, Population size adaptation for the cma-es based on the estimation accuracy of the natural gradient, in: Proceedings of the Genetic and Evolutionary Computation Conference 2016, 2016, pp. 237–244.

[39] P. Dufossé, N. Hansen, Augmented lagrangian, penalty techniques and surrogate modeling for constrained optimization with cma-es, in: Proceedings of the Genetic and Evolutionary Computation Conference, 2021, pp. 519–527.

[40] N. Hansen, The CMA evolution strategy: A tutorial, arXiv preprint arXiv:1604.00772 (2016).

[41] J. Wittenburg, Dynamics of Systems of Rigid Bodies, volume 33, Springer-Verlag, 2013.

[42] D. G. Ltd, VAMPIRE Pro User ManualV 5.02, Derby, UK, 2006.