Received: 14 July 2022 | Accepted: 17 July 2022

DOI: 10.1002/asmb.2707

RESEARCH ARTICLE

WILEY

Hierarchical optimal designs and modeling for engineering: A case-study in the rail sector

Rossella Berni[1] | **Luciano Cantone**[2] | **Alessandro Magrini**[1] | **Nedka D. Nikiforova**[1]

[1]Department of Statistics Computer Science Applications "G. Parenti", University of Florence, Florence, Italy

[2]Department of Engineering for Enterprise "Mario Lucertini", University of Rome "Tor Vergata", Rome, Italy

Correspondence
Rossella Berni, Department of Statistics Computer Science Applications "G.Parenti", University of Florence, Viale Morgagni, 59, 50134, Florence, Italy.
Email: rossella.berni@unifi.it

Abstract

Complex engineering and technological processes typically generate data with a non-trivial hierarchical structure. To this end, in this article we propose a full procedure for optimizing such processes through optimal experimental designs and modeling. In order to study a hierarchical structure, several types of experimental factors may arise, making the building of the experimental design challenging. Starting from the analysis of a preliminary dataset and a pilot design including nested, branching, and shared experimental factors, as well as a new type of experimental factor called composite-form-factor, we build a hierarchical D-optimal experimental design using genetic algorithms. We apply our proposal to a real case-study in the rail sector aimed at optimizing the payload distribution of freight trains. In this case-study we also achieve the best train configuration by minimizing the in-train forces. The results are very satisfactory and confirm that our full procedure represents a valid method to be successfully applied for solving similar technological problems.

KEYWORDS

D-optimality criterion, genetic algorithm, mixed linear model, random effects

1 | INTRODUCTION

Engineering and technological processes usually involve hierarchical data which make the experimental planning challenging. Proper planning of the experiment requires an in-depth knowledge of the process under study regarding both technical/engineering and statistical aspects. In case of partial knowledge, the experimental planning could be unsuitable, thus undermining the entire study. However, preliminary data are typically available, and an appropriate statistical analysis may be performed to gain valuable quantitative information about the process under study. Such quantitative information can be very useful for improving and validating the experimental planning. In particular, the analysis of preliminary data contributes to better understanding the process by identifying the most relevant variables, checking initial assumptions and the subsequent statistical modeling. The relevance of the information gained through the preliminary analysis may be verified through the building of a pilot design. In this way the pilot design results provide evidence of the validity of the experimental planning, and also help to define the most suitable statistical model. This point is of core importance when considering the optimal design theory framework, given the model-dependent nature of optimal designs.

In this manuscript we define a full procedure for the design and analysis of experiments in complex engineering and technological processes. Our proposal is motivated by a real case-study in the rail sector concerning the risk of derailment and/or disruption of freight trains. The proposed full procedure includes the experimental planning at several hierarchical levels for obtaining a hierarchical D-optimal design. In doing so, one of the main key points is the presence of several types of experimental factors; more precisely, we deal with nested, branching, and shared factors as well as with a new type of factor, which we call composite-form-factor (CFF). By considering the types of factors, a genetic algorithm (GA) is also opportunely implemented for obtaining the final hierarchical D-optimal design. The final step of the full procedure ends with the process optimization.

We apply our proposal to a real case-study in the rail sector where we aim to optimize the payload distribution of freight trains, thus making it possible to minimize the in-train forces, and consequently to protect freight trains from derailment and/or disruption. By applying the suggested procedure, we are able to achieve the best train configuration for minimizing the in-train forces. It must be noted that the responses in our case-study are obtained through computer simulations, and we propose a general framework for the technological field based on the optimal design methodology and GAs, capable of managing complex data structures. The results obtained in the case-study are very satisfactory and confirm that the proposed full procedure represents a valid method to be successfully applied for solving similar technological problems.

The article is organized as follows: Section 2 contains a brief literature review, Section 3 includes an outline of the full procedure and details of the hierarchical optimal experimental design and statistical modeling, Section 4 illustrates the results obtained in the case-study, the discussion follows in Section 5; final remarks conclude the article.

2 LITERATURE REVIEW

Optimal designs are widely used in several research fields, where the design optimality is achieved with respect to a specific design criterion strictly related to the assumed statistical model(s). For instance, it emerges from previous research[1,2] that computer experiments and Kriging modeling play a relevant role in the payload distribution of freight trains; as regards Kriging, optimal designs for prediction or efficient parameter estimation are widely developed in literature.[3-6] In the general framework of this current study, we use linear mixed-effect models. In the context of optimal designs for hierarchical linear mixed-effect models, in Reference 7 the authors study approximate D-optimal designs for fixed effects, while in Reference 8 a systematic approach for finding optimal designs for linear as well as non-linear mixed-effect models with correlated errors is proposed. In Reference 9, sequential D-optimal designs for generalized linear mixed-effect models are investigated. In Reference 10, the authors study optimal designs for both an efficient estimation of the fixed effect parameters and a prediction of the random effects in the context of multi-response linear mixed models. Optimal designs for prediction in hierarchical linear models, based on the integrated mean-squared error criterion, are considered in Reference 11. More recently, in Reference 12 the authors investigate optimal designs for the fixed-effects estimation as well as for prediction of random effects in hierarchical linear models, while in Reference 13 optimal designs for hierarchical random effects models applied in the field of precision medicine are taken into account. Moreover, by specifically considering the design of experiments with a particular type of factors, it is relevant to note that in Reference 14 the design and analysis of computer experiments are addressed in the presence of nested and branching factors, while more recently, in Reference 15 the authors investigate D-optimal designs when nested and/or branching factors are present.

Lastly, when considering the study related to freight trains, previous research on this topic is related to the analysis of the payload distribution of freight trains through computer experiments and Kriging modeling. More precisely, in Reference 1 the payload distribution of freight trains is studied in terms of the overall train mass and length involving a computer experiment with four hundreds trains. In Reference 2, the authors assume that a generic train could be divided into five train sections, each characterized by its own overall mass; a suitable Latin Hypercube design, based on strong orthogonal arrays,[16] is planned for the computer experiment.

3 THEORY

In this section we outline the full procedure (Section 3.1), then we detail the hierarchical framework and the types of experimental factors, as well as the hierarchical D-optimal design, and the related statistical modeling (Section 3.3).

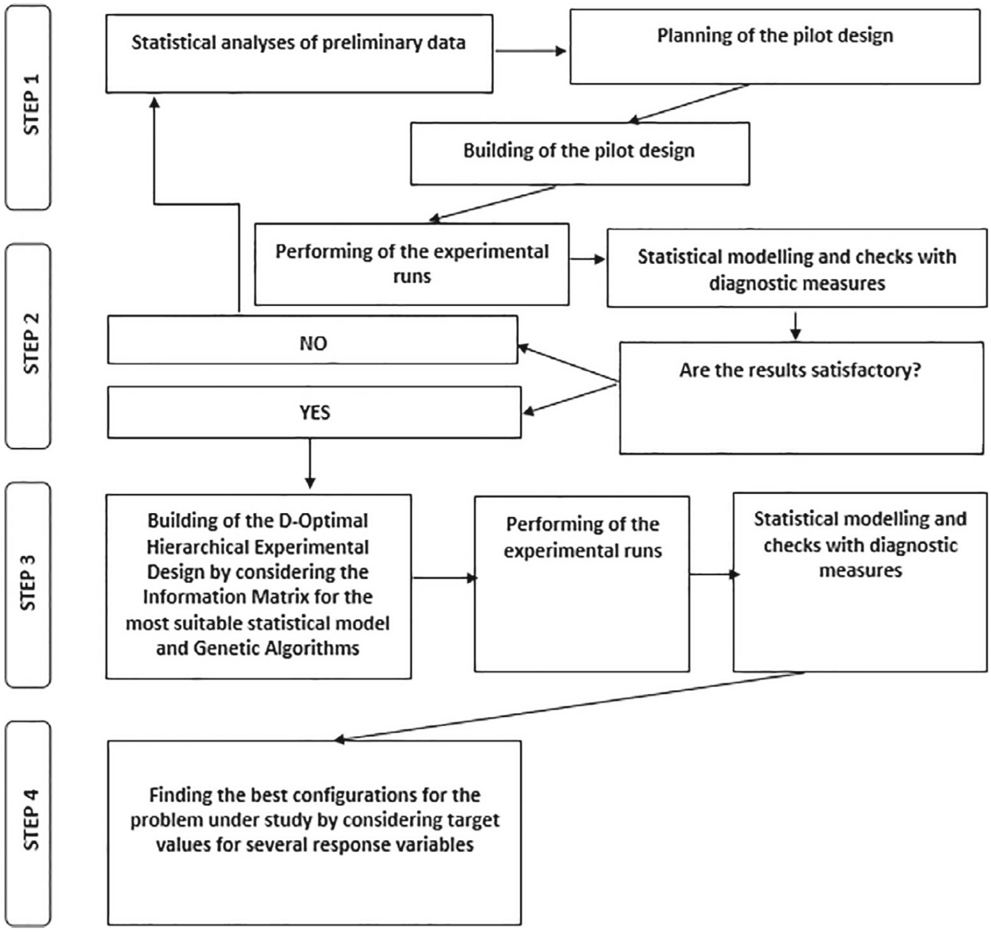

FIGURE 1 Flowchart of the full procedure

3.1 | Outline of the full procedure

The full procedure we propose consists of four main steps, reported in detail in Figure 1. Along with these main steps, we consider as a starting point (Step ♯1): (i) the statistical analysis of preliminary data, and (ii) the building of a pilot design. Both the preliminary data analysis and the pilot design are of key importance since they allow for efficiently planning and validating the experimental design. Step ♯2 consists of performing the experimental runs of the pilot design, and following, the estimation of suitable statistical models is checked through diagnostic measures; this point makes it possible to establish the most suitable statistical model for building the optimal design. In Step ♯3 we consider the D-optimality design criterion, and a GA is opportunely adapted for obtaining the hierarchical optimal design. Depending on the specific situation, other possible design criteria and/or algorithms can also be applied in this step. Lastly, Step ♯4 relates to the achievement of best configurations for the problem under study according to the target values of several response variables.

3.2 | Types of experimental factors

Let's start by considering a non-trivial technological problem with data structured at several hierarchical levels. Consider a categorical factor with T levels ($t = 1, \dots, T$) defined at any hierarchical level. Also, let D and E be two measurable variables such that D takes the same value for each t, say d_t ($t = 1, \dots, T$), and E takes value in an interval which is the same for each t, say $\mathcal{E}_t$ ($t = 1, \dots, T$):

$$\begin{aligned} \Omega_{\mathrm{D}} &= \{d_{k(t)} = d_t : \forall k, k = 1, \dots, K;\ t = 1, \dots, T\}, \\ \Omega_{\mathrm{E}} &= \{e_{k(t)} \in \mathcal{E}_t : k = 1, \dots, K;\ t = 1, \dots, T\}, \end{aligned} \tag{1}$$

where $k(t)$ denotes the kth observation such that the categorical factor takes value t. We define a CFF x_c as follows:

$$x_c = \sum_{k=1}^{K} d_{k(t)} + \sum_{k=1}^{K} e_{k(t)}. \tag{2}$$

Note that a CFF is defined at a higher hierarchical level with respect to the one of the categorical factor. For instance, the total weight of stock consisting of several types of petrol cans is a CFF, where the type of petrol can is the categorical factor. In fact, the tare is constant for each type, while the net weight of each can ranges within an interval of quantitative values which depends on the type. Therefore, the CFF "total stock weight" equates the sum of the tares and the net weights of the petrol cans.

The definition of a CFF can be extended to any number of measured variables. Let's suppose we have Q variables of type D, that is, $D_1, \dots, D_q \dots, D_Q$, taking the same value for each t, and V measurable variables of type E, that is, $E_1, \dots, E_v \dots, E_V$, taking value within an interval which is the same for each t:

$$\begin{aligned} \Omega_{D_q} &= \{d_{q,k(t)} = d_{q,t} : \forall k, k = 1, \dots, K;\ t = 1, \dots, T\} \quad q = 1, \dots, Q, \\ \Omega_{E_v} &= \{e_{v,k(t)} \in \mathscr{E}_{v,t} : k = 1, \dots, K;\ t = 1, \dots, T\} \quad v = 1, \dots, V. \end{aligned} \tag{3}$$

In this general case, a CFF is defined as:

$$x_c = \sum_{q=1}^{Q}\sum_{k=1}^{K} d_{q,k(t)} + \sum_{v=1}^{V}\sum_{k=1}^{K} e_{v,k(t)}. \tag{4}$$

We define the index i for the ith hierarchical level, $i = 1, \dots, I$, while J_i is the number of experimental factors at the ith level. Within each hierarchical level, we may have the following different situations:

- only fixed effects, only random effects or both fixed and random effects, also including first-order interactions between fixed and random effects;
- several sets of experimental factors:

1. $X_c = \{x_{c1}, \dots, x_{cj}, \dots, x_{J_c}\}$ includes the CFFs, where x_{cj} denotes the jth one, $j = 1, \dots, J_c$.
2. $X_s = \{x_{s1}, \dots, x_{sj}, \dots, x_{sJ_s}\}, j = 1, \dots, J_s$ contains the shared factors;[14]
3. $X_b = \{x_{b1}, \dots, x_{bj}, \dots, x_{bJ_b}\}, j = 1, \dots, J_b$ comprises the branching factors;[14]
4. $X_r = \{x_{r1}, \dots, x_{rj}, \dots, x_{rJ_r}\}, j = 1, \dots, J_r$ includes the nested factors.

In our setting, hierarchical variables differ from nested variables. A variable is defined as hierarchical when it refers to a predefined and binding hierarchical structure. Instead, a nested variable in an operational framework relates to a hierarchical structure which is not binding *a priori*. Based on the hierarchical structure defined above, and by indicating with X_i and Z_i the sets for the fixed and random effects respectively at the i th level, we can specify the following general linear mixed-effect model:[17]

$$\boldsymbol{y} = \sum_{i=1}^{I} \boldsymbol{X}_i \boldsymbol{\beta}_i + \sum_{i=1}^{I} \boldsymbol{Z}_i \boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}, \tag{5}$$

where $\boldsymbol{y}$ is the $[n \times 1]$ vector for the dependent variable, $\boldsymbol{X}_i$ is the matrix for fixed effects at the i th hierarchical level with dimension $[n \times p_i]$, and $\boldsymbol{Z}_i$ is the matrix for random effects, with dimension $[n \times p_i]$; $\boldsymbol{\beta}_i$ is the column vector of the unknown coefficients for the fixed effects with dimension $[p_i \times 1]$, while $\boldsymbol{\gamma}_i$ is the vector of the unknown coefficients for the random effects, dimension $[p_i \times 1]$; lastly $\boldsymbol{\varepsilon}$ is the vector $[n \times 1]$ for the random errors.

3.2.1 | Some general examples

The theoretical proposal could be applied to several technological, and similar situations. The generalization is primarily related to a hierarchical structure, in which composite-form, shared, branching, and nested factors could be involved at any hierarchical level. In addition to the case-study, illustrated in details in the following Section 4.2, some examples can be also outlined here. A first example is related to the field of electronic engineering; particularly, we could be interested to study different printed circuit boards-PCB through different combinations of surface finishes, with several component packages, and the correspondent solder joints, including the geometry of the joint. Therefore, in this case, we may have three hierarchical levels: at the 1st level the PCB is defined through the CFF "soldering alloy," where each level is defined as the sum of a basic alloy element plus different alloy elements; the 2nd level is then studied through different electronic components, for example, "component package type." Finally the 3rd level is then determined by the size of pin for each component type. In this example, the CFF is the "soldering alloy," the shared factors are the component package type and the pin size.

A second hierarchical example could be suggested thinking to the detection of harmful gases, and to the study for improving electronic sensors. In this case, the hierarchical study could be articulated at several levels, involving fixed as well as random effects. More specifically, we may have three levels: the 1st level is here characterized by environmental variables, also random, such as humidity, external (environmental) temperature, while the gas concentration is at the 2nd level. The 3rd level is characterized by the chamber, in which several and different material types are involved. Each material type is a chemical powder, formed by different chemical substances, but with a common fixed amount of basic chemical elements, for example "Yttrium plus Cobalt." At the 3rd level an additional factor could be the working temperature, used to analyze a change in response, which is the electrical resistance. In this example, the CFF is the "material type," which is also nested within the chamber. The working temperature is also a nested factor, while the gas concentration is a branching random factor; the others are shared factors.

3.3 | The hierarchical optimal design and modeling

In this section we describe in detail the hierarchical D-optimal design and statistical modeling. Without any loss of generality, we consider three hierarchical levels (i.e., $i = 1, 2, 3$): the highest one (level-1), a middle one (level-2), and the lowest one (level-3). Among the various possible technological situations, we evaluate a general framework including factors with fixed effects at level-1 and level-3 (i.e., $X_1 = \{x_{11}, \dots, x_{1j}, \dots, x_{1J_1}\}$ and $X_3 = \{x_{31}, \dots, x_{3j}, \dots, x_{3J_3}\}$), and random factors at level-2 (i.e., $Z_2 = \{z_{21}, \dots, z_{2j}, \dots, z_{2J_2}\}$), as also reported in Figure 2. To simplify the notation, we omit the subscript 1 identifying the first hierarchical level. Therefore, we consider a linear mixed-effect model[17] defined for a single response variable Y, and for a single observation u , ($u = 1, \dots, n$) as follows:

$$y_u(X_1, X_3, Z_2) = \beta_0 + \sum_{j=1}^{J_c} \beta_{cj} x_{cj_u} + \sum_{j=1}^{J_s} \beta_{sj} x_{sj_u} + \sum_{j=1}^{J_b} \beta_{bj} x_{bj_u} + \sum_{j=1}^{J_r} \beta_{rj} x_{rj_u} + \sum_{j=1}^{J_3} \beta_{3j} x_{3j_u} + \sum_{j=1}^{J_2} \gamma_{2j} z_{2j_u} + \varepsilon_u. \tag{6}$$

Model (6) can be expressed in matrix notation according to formula (5) as follows:

$$\boldsymbol{y} = \boldsymbol{X_1 \beta_1} + \boldsymbol{X_3 \beta_3} + \boldsymbol{Z_2 \gamma_2} + \boldsymbol{\varepsilon}, \tag{7}$$

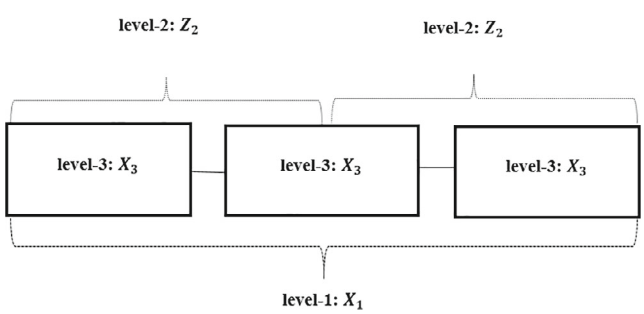

FIGURE 2 A three-level hierarchical data structure and related set of factors following model (6)

where $\boldsymbol{y}$ is the $[n \times 1]$ vector for the dependent variable, $\boldsymbol{X_1}$ is the $[n \times p_1]$ model matrix related to the subset X_1 of composite-form, shared, branching and nested factors (level-1), while $\boldsymbol{\beta_1}$ is the $[p_1 \times 1]$ vector that contains the corresponding unknown coefficients; $\boldsymbol{X_3}$ is the $[n \times p_3]$ model matrix related to the subset X_3 of experimental factors with $\boldsymbol{\beta}_3$ the $[p_3 \times 1]$ vector of unknown coefficients. The matrix $\boldsymbol{Z_2}$ of dimension $[n \times p_2]$ is related to the set of random factors (subset Z_2), while $\boldsymbol{\gamma_2}$ is the corresponding vector of unknown coefficients, dimension $[p_2 \times 1]$. Thus, the total number of unknown coefficients is equal to $p = p_1 + p_2 + p_3$. We assume that both $\boldsymbol{\varepsilon}$ and $\boldsymbol{\gamma_2}$ are i.i.d. Normally distributed with mean zero, variance $\sigma_\varepsilon^2 \boldsymbol{I}_n$ and $\sigma_{\gamma_2}^2 \boldsymbol{I}_{p_2}$, respectively, and $\mathrm{Cov}(\boldsymbol{\gamma_2}, \boldsymbol{\varepsilon}) = \boldsymbol{0}_{p_2 \times n}$.

Therefore, the variance-covariance matrix $\boldsymbol{V}$ of $\boldsymbol{y}$ is defined as follows:

$$\boldsymbol{V} = \boldsymbol{Z_2 G Z_2'} + \boldsymbol{R}, \tag{8}$$

where the first term $\boldsymbol{Z_2 G Z_2'}$ relates to the random effects, and the second term, the matrix $\boldsymbol{R}$, relates to the variance-covariances of the error estimates. The generalized least square (GLS) estimator for the unknown coefficient vector for the fixed effects $[\boldsymbol{\beta_1} : \boldsymbol{\beta_3}]$ is given by:[17]

$$\left[\boldsymbol{\beta_1} : \boldsymbol{\beta_3}\right]_{GLS} = \left[[\boldsymbol{X_1} : \boldsymbol{X_3}]' \boldsymbol{V}^{-1} [\boldsymbol{X_1} : \boldsymbol{X_3}]\right]^{-1} [\boldsymbol{X_1} : \boldsymbol{X_3}]' \boldsymbol{V}^{-1} \boldsymbol{y}, \tag{9}$$

where $[\boldsymbol{\beta_1} : \boldsymbol{\beta_3}]$ denotes the horizontal concatenation of the vectors $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_3}$, while $[\boldsymbol{X_1} : \boldsymbol{X_3}]$ denotes the horizontal concatenation of the matrices $\boldsymbol{X_1}$ and $\boldsymbol{X_3}$.

The D-optimality design criterion is one of the most commonly used for building optimal designs,[18,19] and is also the most versatile in our context. When considering the efficient estimation for the fixed effects $[\boldsymbol{\beta_1} : \boldsymbol{\beta_2}]$ in model (7), the specific D-optimal design maximizes the determinant of the information matrix $\boldsymbol{M}$ defined as:

$$\boldsymbol{M} = [\boldsymbol{X_1} : \boldsymbol{X_3}]' \boldsymbol{V}^{-1} [\boldsymbol{X_1} : \boldsymbol{X_3}], \tag{10}$$

thus minimizing the variance-covariance matrix of the GLS parameter estimates, formula (9).

As regards the variance-covariance matrix $\boldsymbol{V}$ in formula (8), specific structures could be assumed for both $\boldsymbol{G}$ and $\boldsymbol{R}$ matrices, such as variance-components, compound-symmetry, or others. In what follows, we assume a variance-components structure for both $\boldsymbol{G}$ and $\boldsymbol{R}$ matrices. Thus, when considering the case-study it is possible to obtain an equivalent expression for the information matrix $\boldsymbol{M}$ (formula 10), which is particularly useful when computing the hierarchical D-optimal design. Let us define by l_i the generic experimental unit at the i th hierarchical level, $i = 1, 2, 3$, so that we denote by L_i the total number of experimental units within level i . Thus, at level-1 we have a total of L_1 experimental units ($1, \ldots, l_1, \ldots, L_1$). Within each unit l_1 , we deal with a total of L_2 units at level-2, $(1(l_1), \ldots, l_2(l_1), \ldots, L_2(l_1))$, and $m(l_2)$ experimental observations within each unit l_2 ; without any loss of generality, let $m(l_2) = m$. For the case-study, we assume a block-diagonal structure for the matrix $\boldsymbol{V}$ (formula 8), that is, $\boldsymbol{V} = \mathrm{diag}[\boldsymbol{V}_1, \ldots, \boldsymbol{V}_{l_1}, \ldots, \boldsymbol{V}_{L_1}]$. Furthermore, for each matrix $\boldsymbol{V}_{l_1}$, we assume the following structure: $\boldsymbol{V}_{l_1} = \mathrm{diag}[\boldsymbol{V}_{1(l_1)}, \ldots, \boldsymbol{V}_{l_2(l_1)}, \ldots, \boldsymbol{V}_{L_2(l_1)}]$, where each $\boldsymbol{V}_{l_2(l_1)}$, $(1(l_1), \ldots, l_2(l_1), \ldots, L_2(l_1))$, is defined as follows:

$$\boldsymbol{V}_{l_2(l_1)} = \sigma_\varepsilon^2 \boldsymbol{I}_m + \sigma_{\gamma_2}^2 \boldsymbol{1}_m \boldsymbol{1}_m',$$

and $\boldsymbol{1}_m$ is an $m \times 1$ vector of 1's.

According to such assumptions, the equivalent expression of the information matrix $\boldsymbol{M}$ (formula 10) for the case-study, is expressed as follows:

$$\boldsymbol{M} = \frac{1}{\sigma_\varepsilon^2} \left[[\boldsymbol{X_1} : \boldsymbol{X_3}]' [\boldsymbol{X_1} : \boldsymbol{X_3}] - \left(\frac{\sigma_{\gamma_2}^2}{\sigma_\varepsilon^2 + m\sigma_{\gamma_2}^2} \right) \left([\boldsymbol{X_1} : \boldsymbol{X_3}]' \boldsymbol{Z_2} \right) \left([\boldsymbol{X_1} : \boldsymbol{X_3}]' \boldsymbol{Z_2} \right)' \right]. \tag{11}$$

The detailed derivation for the information matrix $\boldsymbol{M}$ (formula 11) is included in the Appendix. Moreover, by considering the full procedure and the corresponding case-study, the building of the hierarchical D-optimal design is carried out through the implementation of a GA.

Several algorithms are available in the literature for building optimal designs. In a first group, we can include well-known and basic algorithms, as for instance the Wynn-Fedorov algorithm,[18,20] the DETMAX algorithm,[21] the KL-exchange algorithm,[22] and the coordinate-exchange algorithm.[23] All allow for dealing with irregular design spaces;

however, they are often too time-consuming since an exhaustive search over all candidate points has to be performed. Some recent improvements, as for example multiplicative-type algorithms,[24] are aimed at overcoming this drawback. In a second group, we can consider meta-heuristic algorithms, which receive considerable attention in research literature. Among the different types of meta-heuristic algorithms, GAs are the most popular. GAs, which trace back to Holland,[25] are very flexible, potential and relatively easy to implement. Two characteristics are relevant for a GA, strictly related to its meta-heuristic nature:[26] (i) the intensification which relates to the local and intensive search around the best solution, and (ii) the diversification which assures that the algorithm explores the design space globally.

GAs are highly powerful for building optimal designs where some kinds of irregularities exist in the design space, for example, constraints when dealing with mixture experiments and similar. Since the GA does not need to specify an explicit candidate-set of points, it is particularly suitable for design problems with medium-large numbers of factors in constrained regions.[26-29] In our setting, we select the GA by specifically considering the non-trivial hierarchical structure, the types of experimental factors defined, and the specific engineering problem under study. More specifically, the choice to use a GA arises from the presence of the CFF, the new type of experimental factor (Section 3.2). In building the optimal design through a GA, we are able to efficiently deal with the CFF, due to the peculiar characteristics of this type of experimental factor, and by also considering its similarities with mixture variables; for instance, the two examples for CFFs reported in Section 3.2.1, for example, the soldering alloy and the material type. Even though there is no guarantee of converging to the true optimum, GAs represent a powerful alternative for building optimal designs, since they have been proven to be effective in solving complex problems in several engineering and technological areas.[27,29] A detailed review on the application of GAs for optimal designs can be found in References 30 and 29. The specific GA implemented here is detailed in Section 4.3 in relation to the case-study.

4 | THE CASE-STUDY

In this section we describe the application of the full procedure to the case-study in the rail sector. In particular, we discuss engineering issues (Section 4.1) and provide the outline of the full procedure (Section 4.2). Afterwards, we report the GA for building the hierarchical D-optimal design (Section 4.3), the models results (Section 4.4), and the best train configuration minimizing the in-train forces (Section 4.5).

4.1 | Engineering issues in the rail freight sector

One of the most important challenges to be addressed, when a new type of trainset is put in service, is the determination of risk in case of train derailment or disruption due to excessive in-train forces during braking. Such forces arise because train braking is not homogeneous for classic freight trains: it starts from the wagons closest to the traction units, and propagates to the wagons further away. The reason for this behavior is the progressive venting of the brake pipe running along the train, which results in progressive filling of the brake cylinders and, therefore, in nonsynchronous braking of the wagons. During braking, this lack of synchronicity results in a longitudinal oscillation of the wagons along the train, which in turns creates in-train forces between two consecutive wagons. If excessive in-train compressive forces are exerted on a wagon running a curve, the wagon can derail (i.e., it exits from the track); if excessive in-train tensile forces occur, the draw gears can fail, and the train breaks into two parts. Both events must be avoided: the first is dangerous for transported goods, and the railway track, the second causes service interruptions and delays. To this end, specific simulators are used by railway undertakings to determine the in-train forces, since physical experimentation is time consuming, very expensive, and thus unfeasible. TrainDy* is one of such simulators[31] that can be considered state-of-the-art software for the computation of in-train forces, in accordance with the International Union of Railways (UIC).[32]

In-train forces depend on train mass and length, train operation (emergency braking, service braking and so on), train braking regime (time to reach the maximum air pressure in brake cylinders), type of wagons used (tare, length, mechanical characteristics of coupling devices, friction characteristics of braking devices, relationship between braked weight and mass), railway track, and wagon arrangements. In general, higher in-train forces arise when the initial speed is between 20 and 40 km/h. In simulating the in-train forces, we consider the speed of 30 km/h, since it is the customary

*TrainDy-International Union of Railways (UIC)

TABLE 1 Description of the experimental factors and variables

Variables	Symbol	Coded levels
Subset X_1 (level-1)		
Train mass	x_{c1}	$[-1, 1]$
Train length	x_{s1}	$[-1, 1]$
Mass shape	x_{b1_f}	$f = 1, 2, 3, 4$
Wagon arrangement for mass shape 1	$x_{r1_{a(1)}}$	$a = 1, 2, 3$
Wagon arrangement for mass shape 2	$x_{r1_{a(2)}}$	$a = 1, 2, 3$
Wagon arrangement for mass shape 3	$x_{r1_{a(3)}}$	$a = 1, 2, 3$
Wagon arrangement for mass shape 4	$x_{r1_{a(4)}}$	$a = 1, 2, 3$
Subset Z_2 (level-2)		
Pairs of wagons	z_{21_h}	$h = 1, \dots, 35$
Subset X_3 (level-3)		
Type of wagon	x_{1_t}	$t = 1, \dots, 7$

speed in many programs of in-train force assessment, and the track radius of horizontal curvature, because it affects the permissible longitudinal compressive force. Moreover, even though several environmental conditions may have an effect on longitudinal train dynamics, such as the degraded performance of braking devices, they are not involved in the case-study due to being beyond the scope.

In some cases, railway undertakings cannot change any of the characteristics listed before; therefore, they must decide whether to allow a train to operate (computing the in-train forces or checking observance of established rules for the specific train set). In other cases, they can manage the position of wagons in the trainset, adjusting the mass along the train to reduce the in-train forces. In fact, among all the factors affecting the in-train forces, this is the parameter that can be most easily managed. The problem under study is particularly challenging, especially when considering the planning of the experiment. More precisely, each train is composed of a large number of wagons; each type of wagon has its own mass, it can be located in any position along the train, and each wagon allocation impacts the in-train forces in a different way. It follows that there is an extremely high number of possible wagon arrangements, each having a different impact on the in-train forces. This particularly challenging point can be successfully handled through the full procedure we propose.

4.2 | Outline of the full procedure for the case-study

According to our full procedure (Figure 1), we start with the analysis of existing data containing information about a large number of trains which are used for building the pilot design. Each train in the pilot design is composed of $K = 36$ wagons because most trains consist of this number in the preliminary dataset. Once the pilot design runs are performed, suitable linear mixed-effect models are estimated and checked with diagnostic measures. The satisfactory results obtained confirm the validity of the experimental planning. More precisely, we consider the following three hierarchical levels: (i) level-1 related to the entire train, (ii) level-2 related to two consecutive wagons (in the rest of the article, we will also use pairs of wagons), and (iii) level-3 related to each individual wagon. A description of the experimental factors for each hierarchical level is reported in Table 1. Namely, at level-3 the subset X_3 is composed of one categorical factor, that is, type of wagon x_{1_t} at seven levels, $t = 1, \dots, 7$. At level-2, the subset Z_2 consists of one categorical factor z_{21_h}, $h = 1, \dots, 35$, that is the pairs of wagons. It must be noted that in our case-study, the number of levels H for the categorical factor z_{21_h} is equivalent to L_2 (i.e., $H = L_2$). Rather, if there are two or more random factors, we may have $H \neq L_2$; for instance, as in the second general example related to the study for improving electronic sensors (Section 3.2.1), in which we deal with two random factors. At level-1, the subset X_1 is composed of:

- one CFF: x_{c1} representing the total train mass, with one variable of type D, the tare of wagons, and one type-E variable, the mass of wagons;

TABLE 2 Mass shapes and examples of wagon arrangements

Type of mass shape	Wagon arrangement for each mass shape
$f = 1$: ascending scalene triangle	$a = 1$: (1, 1, 3, 1, 1, 1, 1, 4, 1, …, 6, 7, 4, 7, 7, 7, 7, 2, 1)
	$a = 2$: (1, 2, 1, 1, 1, 1, 1, 4, 1, …, 3, 7, 7, 1, 7, 7, 7, 7, 1)
	$a = 3$: (2, 1, 4, 1, 1, 1, 1, 1, 4, …, 7, 7, 4, 7, 6, 7, 7, 7, 7)
$f = 2$: descending scalene triangle	$a = 1$: (1, 7, 1, 6, 7, 4, 7, 7, 7, …, 1, 1, 1, 1, 4, 1, 1, 1, 1)
	$a = 2$: (3, 4, 7, 3, 7, 7, 1, 7, 7, …, 1, 1, 1, 1, 4, 1, 1, 3, 1)
	$a = 3$: (4, 1, 3, 7, 7, 4, 7, 6, 7, …, 1, 1, 1, 1, 1, 4, 1, 1, 1)
$f = 3$: trapezoidal	$a = 1$: (1, 4, 4, 4, 1, 1, 1, 4, 3, …, 4, 1, 1, 1, 4, 3, 4, 7, 1)
	$a = 2$: (4, 1, 7, 1, 1, 2, 4, 1, 5, …, 1, 1, 2, 4, 1, 5, 2, 4, 2)
	$a = 3$: (2, 3, 1, 3, 2, 4, 7, 1, 1, …, 3, 2, 4, 7, 1, 1, 1, 1, 2)
$f = 4$: uniform	$a = 1$: (1, 1, 1, 4, 2, 7, 4, 7, 1, …, 4, 2, 7, 4, 7, 1, 4, 5, 4)
	$a = 2$: (7, 1, 7, 1, 7, 4, 2, 1, 1, …, 1, 7, 4, 2, 1, 1, 1, 4, 1)
	$a = 3$: (6, 2, 2, 7, 1, 1, 7, 4, 4, …, 7, 1, 1, 7, 4, 4, 1, 5, 3)

TABLE 3 Assignment of tare, length, and mass to a wagon, depending on its type

Type	Tare (tons)	Length (m)	Mass (tons)
1	12.500	14.61	(1, 2, … , 12)
2	13.000	14.61	(13, 14, 15)
3	13.800	14.02	(16, 17, 18, 19)
4	14.400	14.22	(20, 21, 22, 24, … , 28)
5	14.932	15.50	23
6	17.800	17.25	(29, 30, 31, 32)
7	20.000	17.00	(33, 34, … , 70)

Note: The mass of a wagon must be randomly selected from among all the values included in the set.

- one shared quantitative factor: x_{s1}, that is the train length;
- one branching factor: x_{b1_f}, that is, the mass shape (form) with four levels, $f = 1, \ldots, 4$;
- one nested factor: $x_{r1_{a(f)}}$ which is the type of wagon arrangement for each mass shape. For each mass shape, we define three types of wagon arrangement, that is, $a = 1, 2, 3$.

In Table 2, we report in detail the four types of mass shapes, that is, ascending scalene triangle, descending scalene triangle, trapezoidal, and uniform. For each mass shape we also report the three types of wagon arrangement; for the sake of brevity we only report the first nine and last nine wagons for each arrangement (Table 2). We also point out that the wagon arrangement for each mass shape, defined in terms of type of wagon, is identified by considering the analysis of the preliminary data, and the pilot design; as regards the characteristics of each type of wagon (e.g., mass, tare, and length) reference must be made to Table 3. Following, we proceed to build the D-optimal hierarchical design through a GA, described in detail in the following section.

4.3 | The genetic algorithm for building the hierarchical D-optimal design

The GA defined for building the hierarchical D-optimal design consists of four main steps: generation of the initial population, selection, crossover, and mutation. Also, it requires the following inputs: (i) the number of designs in the initial population g, set to 100; (ii) the crossover probability π_c, set to 0.8; (iii) the mutation probability π_m, set to 0.3; (iv) the number of iterations R, set to 1000. In what follows the steps of the algorithm are detailed.

4.3.1 | Generation of the initial population

The algorithm begins by randomly generating an initial population of g experimental designs, say $\mathcal{P} = \{\mathbb{D}_1, \dots, \mathbb{D}_g\}$. Each of these g designs has a fixed structure consisting of 72 trains, given the previous study on this topic,[2] with 36 wagons each. A starting design matrix composed of 72 trains is defined by considering the experimental factors previously described (Section 4.2 and Table 1). For a given mass shape, each wagon's arrangement is represented in terms of types of wagon, as established in Step no.1 of the full procedure, and validated through the pilot design. Thus, tare, length, and mass of each wagon are assigned on the basis of the type, as indicated in Table 3. Following, the CFF, that is, the train mass, is obtained as the sum of tares and mass for each wagon. Once the initial population of designs is generated, the fitness of each design is computed as the determinant of the information matrix in formula (11).

4.3.2 | Selection step

Once the initial population of designs $\mathcal{P}$ has been generated, a group of fittest designs is selected with probability proportional to the fitness. These fittest designs act as chromosomes designed to generate the next generation. The selected designs constitute the population of good parents $\mathcal{P}_0$, which will undergo crossover and mutation processes in the next steps to create a new offspring.

4.3.3 | Crossover step

Designs in the population of good parents $\mathcal{P}_0$ are selected with probability π_c as candidates for crossover. The selected designs are paired and one crossover process is carried out for each pair. Let $\mathbb{D}$ be a design with matrix including the trains $\mathcal{T}_1, \dots, \mathcal{T}_{72}$, and $\mathbb{D}'$ another design with matrix including the trains $\mathcal{T}_1', \dots, \mathcal{T}_{72}'$. The crossover process between designs $\mathbb{D}$ and $\mathbb{D}'$ consists of randomly selecting an integer number $l_1 \in [1, 72]$ and of generating two new designs, one with matrix including the trains $\{\mathcal{T}_1, \dots, \mathcal{T}_{l_1}, \mathcal{T}_{l_1+1}', \dots, \mathcal{T}_{72}'\}$, and the other including the trains $\{\mathcal{T}_1', \dots, \mathcal{T}_{l_1}', \mathcal{T}_{l_1+1}', \dots, \mathcal{T}_{72}'\}$. In practice, the trains act as genes, thus the two designs (chromosomes) undergoing crossover exchange two portions of genes. All the new designs generated by the crossover processes are included in the initial population $\mathcal{P}$. Afterwards, a new population $\mathcal{P}^*$ is obtained by selecting the g designs in $\mathcal{P}$ with the highest fitness.

4.3.4 | Mutation step

Each design in the population $\mathcal{P}^*$ undergoes a mutation process with probability π_m. A mutation process generates a new design equal to the original one, except that one of its trains (genes) is randomly selected and replaced by another randomly selected train (gene) of a randomly selected design (chromosome) in $\mathcal{P}^*$. All the new designs generated by mutation processes are included in the population $\mathcal{P}^*$. Afterwards, a new population $\mathcal{P}^{**}$ is obtained by selecting the g designs with the highest fitness. The new population $\mathcal{P}^{**}$ acts as a new initial population and the algorithm is repeated from the selection step. When the algorithm has been repeated R times, population $\mathcal{P}^{**}$ is returned as the set of designs.

The GA, detailed above, is employed as a design construction algorithm for searching for the hierarchical D-optimal design. After performing 1000 iterations, the GA returns the final hierarchical D-optimal design which maximizes the determinant of the information matrix $\boldsymbol{M}$ (formula 11).

4.4 | Model results

The main aim of the case-study is to improve the payload distribution of freight trains for guaranteeing emergency braking with the minimum in-train compressive and tensile forces among wagons (Section 4.1). To this end, we consider the

TABLE 4 Description of the response variables

Variables	Symbol	Target τ_l
Compressive forces at 10 m	y_1	$\tau_1 < 400$ (kN)
Tensile forces at 2 m	y_2	$\tau_2 < 550$ (kN)

following two quantitative response variables, also reported in Table 4: (i) compressive forces at 10 m (y_1), and (ii) tensile forces at 2 m (y_2). By taking into account the obtained hierarchical D-optimal design (Section 4.3), the true values of compressive and tensile forces are simulated through the TrainDy software,[31] which is considered state-of-the-art software for this type of computation by the UIC. The linear mixed-effect model for the three subsets of experimental factors X_1, Z_2, and X_3, for a single response variable Y, and for a single observation u ($u = 1, \ldots, n$) is defined as follows:

$$y_u(X_1, X_3, Z_2) = \beta_0 + \beta_{c1} x_{c1_u} + \beta_{s1} x_{s1_u} + \sum_{f=1}^{F-1} \beta_{b1_f} x_{b1_{fu}} + \sum_{f=1}^{F} \sum_{a=1}^{A-1} \beta_{r1_{a(f)}} x_{r1_{a(f)u}} + \sum_{t=1}^{T-1} \beta_{31_t} x_{31_{t_u}} + \sum_{h=1}^{H} \gamma_h z_{21_{h_u}} + \varepsilon_u. \quad (12)$$

In formula (12), x_{c1} is the CFF (i.e., the train mass); x_{s1} is the shared factor (i.e., the train length), while x_{b1_f} and $x_{r1_{a(f)}}$ are the branching and the nested factors respectively. There are a total of $L_1 = 72$ trains, each composed of $K = 36$ wagons. Thus, in our setting $H = L_2 = 35$ pairs of wagons with $m = 2$ (total number of observations $n = L_1 \times L_2 \times m = 72 \times 35 \times 2$). When considering Section 3.3 and model (7), we have the following structures:

- the matrix $\boldsymbol{X_1}$ is related to level-1, with a dimension of:

$$[n \times p_1] = [(L_1 \times L_2 \times m) \times p_1] = [(72 \times 35 \times 2) \times 14];$$

- at level-2, the matrix $\boldsymbol{Z_2}$ related to the random effects has a dimension of:

$$[n \times \gamma_2] = [(L_1 \times L_2 \times m) \times p_2] = [(72 \times 35 \times 2) \times 35];$$

- the matrix $\boldsymbol{X_3}$ is related to level-3 with the following dimension:

$$[n \times p_3] = [(L_1 \times L_2 \times m) \times p_1] = [(72 \times 35 \times 2) \times 6];$$

- the vectors $\boldsymbol{\beta_1}$, $\boldsymbol{\gamma_2}$, and $\boldsymbol{\beta_3}$, related to the unknown coefficients for level-1, level-2, and level-3, have dimensions: $[p_1 \times 1] = [14 \times 1]$, $[p_2 \times 1] = [35 \times 1]$, and $[p_3 \times 1] = [6 \times 1]$, respectively.

Model (12) is estimated for each response variable, by applying the GLIMMIX procedure of the SAS software (version 9.4, Windows Platform). A variance-components structure is assumed for both the $\boldsymbol{G}$ and $\boldsymbol{R}$ matrices in formula (8). In Table 5 we report the GLS estimates of model (12) for each response variable. By observing Table 5 we note that large numbers of the estimated coefficients are significant and/or highly significant, and for all of them very small standard errors occur. The CFF x_{c1} is highly significant for both models, with the lowest standard error compared to all the other estimates. We do not report the estimates for the random effects. We only point out that similar to the estimates for the fixed effects, most of the random coefficient estimates are highly significant, with standard errors close to zero.

In Table 6, three diagnostic measures (−2 residual log-likelihood, Akaike information criterion-AIC, Bayesian information criterion-BIC) are reported. The residual diagnostics for both the estimated models are evaluated by considering three types of residuals: raw, studentized, and Pearson. For each estimated model, the results are very satisfactory

TABLE 5 GLS estimates for fixed effects of model (12) for each response

$\hat{y}_1$				$\hat{y}_2$			
Coefficient	**Estimate**	**Std. Err.**	***p*-value**	**Coefficient**	**Estimate**	**Std.Err.**	***p*-value**
β_0	−0.4055	0.0437	< 0.0001	β_0	0.2510	0.0375	< 0.0001
β_{c1}	−0.0236	0.0019	< 0.0001	β_{c1}	0.0726	0.0035	< 0.0001
β_{s1}	−0.0674	0.0301	0.0253	β_{s1}	0.0436	0.0549	0.4266
β_{b1_1}	−0.0184	0.0179	0.3037	β_{b1_1}	−0.1416	0.0325	< 0.0001
β_{b1_2}	0.0140	0.0179	0.4346	β_{b1_2}	0.0301	0.0325	0.3537
β_{b1_3}	0.0649	0.0076	< 0.0001	β_{b1_3}	0.0476	0.0138	0.0006
β_{b1_4}	0	-	-	β_{b1_4}	0	-	-
$\beta_{r1_{1(1)}}$	0.0053	0.0233	0.8200	$\beta_{r1_{1(1)}}$	−0.0017	0.0425	0.9685
$\beta_{r1_{2(1)}}$	−0.0042	0.0194	0.8295	$\beta_{r1_{2(1)}}$	0.1050	0.0352	0.0029
$\beta_{r1_{3(1)}}$	0	-	-	$\beta_{r1_{3(1)}}$	0	-	-
$\beta_{r1_{1(2)}}$	−0.0603	0.0235	0.0104	$\beta_{r1_{1(2)}}$	0.0427	0.0428	0.3188
$\beta_{r1_{2(2)}}$	0.1050	0.0194	< 0.0001	$\beta_{r1_{2(2)}}$	0.1004	0.0352	0.0044
$\beta_{r1_{3(2)}}$	-	-	-	$\beta_{r1_{3(2)}}$	0	-	-
$\beta_{r1_{1(3)}}$	−0.0847	0.0180	< 0.0001	$\beta_{r1_{1(3)}}$	−0.0917	0.0327	0.0050
$\beta_{r1_{2(3)}}$	−0.0290	0.0063	< 0.0001	$\beta_{r1_{2(3)}}$	0.0597	0.0115	< 0.0001
$\beta_{r1_{3(3)}}$	0	-	-	$\beta_{r1_{3(3)}}$	0	-	-
$\beta_{r1_{1(4)}}$	−0.1696	0.0439	0.0001	$\beta_{r1_{1(4)}}$	0.1291	0.0799	0.1061
$\beta_{r1_{2(4)}}$	0.1140	0.0090	< 0.0001	$\beta_{r1_{2(4)}}$	0.2069	0.0180	< 0.0001
$\beta_{r1_{3(4)}}$	0	-	-	$\beta_{r1_{3(4)}}$	0	-	-
β_{31_1}	0.0255	0.0029	< 0.0001	β_{31_1}	−0.0373	0.0053	< 0.0001
β_{31_2}	0.0141	0.0043	0.0010	β_{31_2}	−0.0381	0.0078	< 0.0001
β_{31_3}	0.0119	0.0051	0.0201	β_{31_3}	0.0023	0.0093	0.8055
β_{31_4}	0.0156	0.0035	< 0.0001	β_{31_4}	−0.0189	0.0063	0.0028
β_{31_5}	0.0180	0.0069	0.0087	β_{31_5}	−0.0109	0.0125	0.3820
β_{31_6}	0.0201	0.0073	0.0061	β_{31_6}	−0.0086	0.0133	0.5168
β_{31_7}	0	-	-	β_{31_7}	0	-	-

TABLE 6 Diagnostic statistics for model (12), for each response

Diagnostic statistic	$\hat{y}_1$	$\hat{y}_2$
−2 residual log-likelihood	−11172.2	−5222.72
Akaike information criterion-AIC	−11168.2	−5218.72
Bayesian information criterion-BIC	−11165.1	−5215.61

especially in view of the histogram of residuals with the overlay of the Normal density curve, the Q-Q plot, and the box-plot of the corresponding residuals.

4.5 | The best train configuration for minimizing the in-train forces

Referring to Figure 1, the final step ♯4 corresponds to finding the best train configuration by considering the experimental factors and variables involved (reported in Table 1), and allowing for minimizing the in-train forces in order to decrease, or better, to avoid the risk of derailment and disruption. To this end, the two response variables, for example, the in-train forces (compressive at 10 m, tensile at 2 m), must be minimized taking the roles of both forces into account, also considering the target values, for example, $\tau_1 < 400$ (kN) and $\tau_2 < 550$ (kN), according to the STB-smaller the better situation. As technically detailed in Section 4, the simultaneous involvement of both forces during the train braking requires the definition of an objective function able to evaluate all the technical parameters (e.g., train length, mass, shape, and wagon arrangements), also including the random effects. Moreover, the relative importance of the compressive forces with respect to the tensile forces has to be considered when finding the optimal setting for the train configuration. In other words, when computing the optimal setting, the predominant importance of compressive forces with respect to the tensile ones must be fixed.

Undoubtedly, the best train configuration can primarily be settled with respect to the mass and length of a train, conditioned to the mass shape and wagon arrangement (categorical variables), and also to the random effects related to the pair of wagons. In this respect, and accounting for all the previously mentioned issues, the objective function expression to be minimized is defined as the following:

$$\begin{aligned} \min_{\chi} F_{obj} &= (\boldsymbol{Y}|X_1, Z_2, X_3, w_{y_1}, w_{y_2}, \tau) \\ \boldsymbol{Y} &= (\boldsymbol{y}_1, \boldsymbol{y}_2) \\ \tau &= (\tau_1, \tau_2) \\ w_{y_1} + w_{y_2} &= 1. \end{aligned} \tag{13}$$

By considering the two estimated models, and the specific independent variables (fixed and random) involved, the objective function becomes:

$$\begin{aligned} \min_{\chi} F_{obj} &= (\hat{\boldsymbol{y}}_{\boldsymbol{1}}, \hat{\boldsymbol{y}}_{\boldsymbol{2}}|x_{c_1}, x_{s_1}, \gamma, x_{b1_f} = f, x_{r1_{a(f)}} = a, w_{y_1}, w_{y_2}, \tau_1, \tau_2) \\ &= [w_{y_1}(\hat{y}_1 - \tau_1)^2 + w_{y_2}(\hat{y}_2 - \tau_2)^2 | \boldsymbol{\beta}, \boldsymbol{\gamma}]; \\ &\text{and} \\ w_{y_1} + w_{y_2} &= 1 \\ \boldsymbol{\gamma} &= \boldsymbol{\gamma}_{y_1}, \boldsymbol{\gamma}_{y_2}. \end{aligned} \tag{14}$$

In Tables 7 and 8 we illustrate several scenarios, according to the train mass and length, at pre-specified values for the mass shape and arrangement (the latter nested within the mass shape), and the random effects related to the pairs of wagons. The results are satisfactory when considering the choice of defining a set of variable weights, for example, w_{y_1}, w_{y_2}, without specifying a-priori weight values. As we can observe, the weights obtained make it possible to satisfy the technical requirements, generally achieving greater importance for the in-train force at 10 m with respect to the in-train force at 2 m. Moreover, the minimization of the objective function, as defined in formulas (13) and (14), achieves good performances for the global scenarios, also considering the simultaneous minimization of both responses. In the following section, the scenarios will be discussed by emphasizing the technical and engineering aspects, and highlighting the notable achievements.

5 | DISCUSSION

In this article a general procedure for building hierarchical optimal designs for engineering and technological processes is suggested, and applied to the rail freight sector. The main aim of our case-study is to minimize the in-train forces, for example, compressive forces at 10 m and tensile forces at 2 m, that are the main cause of risk for derailment and disruption respectively.

The results reported in Tables 7 and 8 are related to some of the possible scenarios considering the random and fixed effects involved in the computer experiment. We select the specific scenarios that can be exhaustive for the technical discussion. In fact, as it can be observed, we obtain very satisfactory results, also taking into account the weight values achieved for the mass shape ♯2 and ♯3, even though for the mass shape ♯3 and arrangement $a = 3$ we do not achieve the desired weighting for the in-train force at 10 m.

TABLE 7 Best train configuration and force results: Some scenarios

Variables	Symbol	Optimal setting
Scenario ♯1: mass shape ♯1, $a = 2$		
Compressive forces at 10 m	y_1	−77.15 (kN)
Tensile forces at 2 m	y_2	98.95 [kN]
Train mass	x_{c1}	1421.98 (t)
Train length	x_{s1}	549.69 (m)
Weight	$w_{y_1} = 0.6460$	$w_{y_2} = 0.3540$
Scenario ♯2: mass shape ♯1, $a = 3$		
Compressive forces at 10 m	y_1	−71.64 [kN]
Tensile forces at 2 m	y_2	52.05 [kN]
Train mass	x_{c1}	1505.90 (t)
Train length	x_{s1}	546.03 (m)
Weight	$w_{y_1} = 0.3342$	$w_{y_2} = 0.6658$
Scenario ♯3: mass shape ♯2, $a = 1$		
Compressive forces at 10 m	y_1	−78.60 [kN]
Tensile forces at 2 m	y_2	113.54 [kN]
Train mass	x_{c1}	1414.78 (t)
Train length	x_{s1}	550.11 (m)
Weight	$w_{y_1} = 0.8024$	$w_{y_2} = 0.1976$

Furthermore, when considering and comparing the results (Tables 7 and 8), we can confirm the experience of railway undertakings for classic freight trains: uniform payload distributions (i.e., mass shape ♯4) are not able to give the best train configuration, and therefore, they are not even reported. On the contrary, the mass shape ♯2, corresponding to a descending scalene triangle, is able to provide low values for the longitudinal compressive forces, similar to the mass shape ♯3, for example, the trapezoidal mass distribution. The ascending scalene triangle (mass shape ♯1) is more suitable for low longitudinal tensile forces. Specific mass arrangements can change this behavior, as the results show, emphasizing the need for an analysis able to discriminate between the reciprocal position of different wagons: some arrangements are better than others and keep the same mass shape. The fact that the refined method suggested here is able to replicate some of the results already known is encouraging for other more complex applications in which there is no (or little) prior knowledge, for example, the trains obtained by coupling two or more trains and communicating by radio or wire.

However, it must be noted that each scenario depends exclusively by the arrangement involved, for example, by the wagon types used to make up the freight train, and therefore each train configuration is strictly related to empirical (and real) constraints.

The analysis shows that further best train compositions can be obtained by changing the relative weight of compressive and tensile forces. This makes the method suitable for general applications. For freight trains, it is usually preferable to find the best train configuration by considering a higher weight for the compressive forces than for the tensile forces, but this can be a useful feature when the type of railway track, train operation, or train braking regime causes very high tensile forces that could be of concern: in this case the two weights can be very similar. Therefore, the best configuration with respect to in-train forces may depend by the arrangement involved, for example, by the wagon types used to make up the freight train, but also by the specific setting of all the technical features (length, mass, mass shape), according to the operators' requirements.

6 | FINAL REMARKS

In this manuscript we illustrate a general procedure for design and analysis of experiments in complex technological processes. A core part of the proposal is the building of a hierarchical D-optimal design in a general framework of

TABLE 8 Best train configuration and force results: Some scenarios

Variables	Symbol	Optimal setting
Scenario ♯4: mass shape ♯2, $a = 2$		
Compressive forces at 10 m	y_1	−32.51 please note that in Table 7 the symbols and numbers are reported in different fonts.
Tensile forces at 2 m	y_2	279.44 [kN]
Train mass	x_{c1}	1478.86 [t]
Train length	x_{s1}	548.07 (m)
Weight	$w_{y_1} = 0.8066$	$w_{y_2} = 0.1934$
Scenario ♯5: mass shape ♯3, $a = 1$		
Compressive forces at 10 m	y_1	−55.21 [kN]
Tensile forces at 2 m	y_2	103.85 [kN]
Train mass	x_{c1}	1398.93 (t)
Train length	x_{s1}	549.68 (m)
Weight	$w_{y_1} = 0.8621$	$w_{y_2} = 0.1379$
Scenario ♯6: mass shape ♯3, $a = 3$		
Compressive forces at 10 m	y_1	−39.06 [kN]
Tensile forces at 2 m	y_2	120.75 [kN]
Train mass	x_{c1}	1505.79 [t]
Train length	x_{s1}	546.22 (m)
Weight	$w_{y_1} = 0.4008$	$w_{y_2} = 0.5992$

complex data structures. The suggested procedure may be applied by also considering other possible design criteria and/or statistical models, according to the specific problem under study.

The method is applied in this specific case-study by considering the wagon mass as a governing factor for the optimization. In any case, the wagon mass determines the wagon's braked weight as well as the corresponding percentage of braked weight, via its specific braking devices. The best train configuration, in terms of longitudinal forces, can also be found by considering the percentage of braked mass, which is a quantity commonly used in railway applications. Another possible further development of the method, for application in the railway sector, could be a search for the best train configuration by optimizing the ratios between the longitudinal forces, and their admissible counterparts (immediately known from the type of wagon and railway track). Moreover, the proposed case-study assumes a fixed number of wagons, shapes, and arrangements, but the methodology we propose can be applied to trains with general characteristics, including trains with multiple sections.

ACKNOWLEDGMENT

Our thanks to Ms. Susan Mary Cadby for her revision of the English language aspects of the article.

Rossella Berni contributed to this research also as University of Florence Vice-Coordinator for the Competence Center ARTES4.0. Open Access Funding provided by Universita degli Studi di Firenze within the CRUI-CARE Agreement.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Rossella Berni https://orcid.org/0000-0002-7782-1777
Luciano Cantone https://orcid.org/0000-0001-7416-1040
Alessandro Magrini https://orcid.org/0000-0002-7278-5332
Nedka D. Nikiforova https://orcid.org/0000-0002-5948-5917

REFERENCES

1. Arcidiacono G, Berni R, Cantone L, Placidoli P. Kriging models for payload-distribution optimization of freight trains. *Int J Prod Res*. 2017;55:4878-4890.
2. Nikiforova ND, Berni R, Arcidiacono G, Cantone L, Placidoli P. Latin hypercube designs based on strong orthogonal arrays and Kriging modelling to improve the payload distribution of trains. *J Appl Stat*. 2021;48:498-516.
3. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. *Stat Sci*. 1989;4:409-423.
4. Baldi Antognini A, Zagoraiou M. Exact optimal designs for computer experiments via kriging metamodelling. *J Stat Plan Infer*. 2010;140:2607-2617.
5. Müller WG, Pronzato L, Rendas J, Helmut W. Efficient prediction designs for random fields. *Appl Stoch Model Bus Ind*. 2015;31:178-194.
6. Jones B, Silvestrini RT, Montgomery DC, Steinberg DM. Bridge designs for modeling systems with low noise. *Technometrics*. 2015;57:155-163.
7. Schmelter T. Considerations on group-wise identical designs for linear mixed models. *J Stat Plan Infer*. 2007;137:4003-4010.
8. Dette H, Pepelyshev A, Holland-Letz T. Optimal designs for random effect models with correlated errors with applications in population pharmacokinetics. *Ann Appl Stat*. 2010;4:1430-1450.
9. Sinha SK, Xu X. Sequential D-optimal designs for generalized linear mixed models. *J Stat PlanInference*. 2011;141:1394-1402.
10. Liu X, Yue R-X, Wong WK. D-optimal designs for multi-response linear mixed models. *Metrika*. 2019;82:87-98.
11. Prus M, Schwabe R. Optimal designs for the prediction of individual parameters in hierarchical models. *J Royal Stat Soc Ser B*. 2016;78:175-191.
12. He L, Yue R-X. D-optimal designs for hierarchical linear models with intraclass covariance structure. *Stat Pap*. 2021;62:1349-1361.
13. Prus M, Norbert B, Schwabe R. Optimal design in hierarchical random effect models for individual prediction with application in precision medicine. *J Stat Theory Pract*. 2020;14:24-36.
14. Hung Y, Joseph VR, Melkote SN. Design and analysis of computer experiments with branching and nested factors. *Technometrics*. 2009;51:354-365.
15. Goos P, Jones B. Optimal experimental design in the presence of nested factors. *Technometrics*. 2019;61:533-544.
16. He Y, Tang B. Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika*. 2013;100:254-260.
17. Searle SR, Casella G, McCulloch CE. *Variance Components*. John Wiley & Sons; 1992.
18. Fedorov VV. *Theory of Optimal Experiments*. Academic Press; 1972.
19. Atkinson AC, Donev AN, Tobias RD. *Optimum Experimental Designs, with SAS*. Oxford University Press; 2007.
20. Wynn HP. The sequential generation of D-optimal experimental designs. *Ann Math Stat*. 1970;41:1055-1064.
21. Mitchell T. An algorithm for construction of D-optimal experimental designs. *Technometrics*. 1974;16:203-210.
22. Atkinson A, Donev A. The construction of exact D-optimum experimental designs with application to blocking response surface designs. *Biometrika*. 1989;3:515-526.
23. Meyer RK, Nachtsheim CJ. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*. 1995;37:60-69.
24. Torsney B, Martín-Martín R. Multiplicative algorithms for computing optimum designs. *J Stat Plan Infer*. 2009;139:3947-3961.
25. Holland JH. *Adaptation in Natural and Artificial Systems*. MIT Press; 1975.
26. Martín-Martín R, García Camacha I, Torsney B. Efficient algorithms for constructing D- and I-optimal exact designs for linear and non-linear models in mixture experiments. *Stat Oper Res Trans*. 2019;43:163-190.
27. Heredia-Langer A, Carlyle W, Montgomery D, Borror C, Runger G. Genetic algorithms for the construction of D-optimal designs. *J Qual Technol*. 2003;35:28-46.
28. Limmun W, Borkowski JJ, Chomtee B. Using a genetic algorithm to generate D-optimal designs for mixture experiments. *Qual Reliab Eng Int*. 2013;29:1055-1068.
29. Lin CD, Anderson-Cook CM, Hamada MS, Moore LM, Sitter RR. Using genetic algorithms to design experiments: a review. *Qual Reliab Eng Int*. 2015;31:155-167.
30. Pradubsri W, Chomtee B, Borkowski JJ. Using a genetic algorithm to generate D-optimal designs for mixture-process variable experiments. *Qual Reliab Eng Int*. 2019;35:2657-2676.
31. Cantone L. TrainDy: the new Union internationale des Chemins de Fer software for freight train interoperability. *J Rail Rapid Trans*. 2011;225:57-70.
32. IRS 40421. Rules for the consist and braking of international freight trains: new IRS 40421; 2021.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Berni R, Cantone L, Magrini A, Nikiforova ND. Hierarchical optimal designs and modeling for engineering: A case-study in the rail sector. *Appl Stochastic Models Bus Ind*. 2022;1-18. doi: 10.1002/asmb.2707

APPENDIX

Derivation of the information matrix $\boldsymbol{M}$ (formula 11)

As already defined in Section 3.3, for the case-study we assume a block-diagonal structure for the matrix $\boldsymbol{V}$ (formula 8), that is, $\boldsymbol{V} = \text{diag}[\boldsymbol{V}_1, \dots, \boldsymbol{V}_{l_1}, \dots, \boldsymbol{V}_{L_1}]$. Furthermore, for each matrix $\boldsymbol{V}_{l_1}$, $(1, \dots, l_1, \dots, L_1)$, we assume the following structure: $\boldsymbol{V}_{l_1} = \text{diag}[\boldsymbol{V}_{1(l_1)}, \dots, \boldsymbol{V}_{l_2(l_1)}, \dots, \boldsymbol{V}_{L_2(l_1)}]$, where each $\boldsymbol{V}_{l_2(l_1)}$, $(1(l_1), \dots, l_2(l_1), \dots, L_2(l_1))$, is given by the following:

$$\boldsymbol{V}_{l_2(l_1)} = \sigma_\epsilon^2 \boldsymbol{I}_m + \sigma_{\gamma_2}^2 \mathbf{1}_m \mathbf{1}_m', \tag{A1}$$

and $\mathbf{1}_m$ is an $m \times 1$ vector of 1's.

Recall that for two nonzero constants a and b, $\boldsymbol{I}$ being an $n \times n$ identity matrix, and $\boldsymbol{J}$ being an $n \times n$ matrix having every element unity, the following general result holds[17(p. 443)]:

$$(a\boldsymbol{I} + b\boldsymbol{J})^{-1} = \frac{1}{a}\left\{\boldsymbol{I} - \left(\frac{b}{a + nb}\right)\boldsymbol{J}\right\}.$$

Thus, with $a = \sigma_\epsilon^2$, $b = \sigma_{\gamma_2}^2$, $\boldsymbol{I} = \boldsymbol{I}_m$, and $\boldsymbol{J} = \mathbf{1}_m\mathbf{1}_m'$, it follows that:

$$\boldsymbol{V}_{l_2(l_1)}^{-1} = \frac{1}{\sigma_\epsilon^2}\left\{\boldsymbol{I}_m - \left(\frac{\sigma_{\gamma_2}^2}{\sigma_\epsilon^2 + m\sigma_{\gamma_2}^2}\right)\mathbf{1}_m\mathbf{1}_m'\right\}. \tag{A2}$$

From formula (A2), since $\boldsymbol{V}_{l_1}^{-1} = \text{diag}[\boldsymbol{V}_{1(l_1)}^{-1}, \dots, \boldsymbol{V}_{l_2(l_1)}^{-1}, \dots, \boldsymbol{V}_{L_2(l_1)}^{-1}]$, it follows that:

$$\boldsymbol{V}_{l_1}^{-1} = \frac{1}{\sigma_\epsilon^2}\left\{\boldsymbol{I}_{mL_2} - \left(\frac{\sigma_{\gamma_2}^2}{\sigma_\epsilon^2 + m\sigma_{\gamma_2}^2}\right)\boldsymbol{Z}_{2_{l_1}}\boldsymbol{Z}_{2_{l_1}}'\right\}, \tag{A3}$$

where $\boldsymbol{Z}_{2_{l_1}} = \text{diag}[\mathbf{1}_m, \dots \mathbf{1}_m, \dots, \mathbf{1}_m]$, with dimension $[(mL_2) \times L_2]$. Furthermore, since $\boldsymbol{V}^{-1} = \text{diag}[\boldsymbol{V}_1^{-1}, \dots, \boldsymbol{V}_{l_1}^{-1}, \dots, \boldsymbol{V}_{L_1}^{-1}]$ and given formula (A3), it follows that:

$$\boldsymbol{V}^{-1} = \frac{1}{\sigma_\epsilon^2}\left\{\boldsymbol{I}_n - \left(\frac{\sigma_{\gamma_2}^2}{\sigma_\epsilon^2 + m\sigma_{\gamma_2}^2}\right)\boldsymbol{Z}_2\boldsymbol{Z}_2'\right\},$$

where $\boldsymbol{Z}_2 = \text{diag}[\boldsymbol{Z}_{2_1}, \dots \boldsymbol{Z}_{2_{l_1}}, \dots, \boldsymbol{Z}_{2_{L_1}}]$, and $n = L_1 \times L_2 \times m$.

Without any loss of generality, let $[\boldsymbol{X}_1 : \boldsymbol{X}_3] = \boldsymbol{X}$. Thus, starting from formula (10):

$$\begin{aligned}
\boldsymbol{M} &= [\boldsymbol{X}_1 : \boldsymbol{X}_3]'\boldsymbol{V}^{-1}[\boldsymbol{X}_1 : \boldsymbol{X}_3] \\
&= \boldsymbol{X}'\left\{\frac{1}{\sigma_\epsilon^2}\left[\boldsymbol{I}_n - \left(\frac{\sigma_{\gamma_2}^2}{\sigma_\epsilon^2 + m\sigma_{\gamma_2}^2}\right)\boldsymbol{Z}_2\boldsymbol{Z}_2'\right]\right\}\boldsymbol{X} \\
&= \frac{1}{\sigma_\epsilon^2}\left\{\boldsymbol{X}'\left[\boldsymbol{I}_n - \left(\frac{\sigma_{\gamma_2}^2}{\sigma_\epsilon^2 + m\sigma_{\gamma_2}^2}\right)\boldsymbol{Z}_2\boldsymbol{Z}_2'\right]\boldsymbol{X}\right\} \\
&= \frac{1}{\sigma_\epsilon^2}\left\{\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}'\left[\left(\frac{\sigma_{\gamma_2}^2}{\sigma_\epsilon^2 + m\sigma_{\gamma_2}^2}\right)\boldsymbol{Z}_2\boldsymbol{Z}_2'\right]\boldsymbol{X}\right\} \\
&= \frac{1}{\sigma_\epsilon^2}\left\{\boldsymbol{X}'\boldsymbol{X} - \frac{\sigma_{\gamma_2}^2}{\sigma_\epsilon^2 + m\sigma_{\gamma_2}^2}\left[\boldsymbol{X}'\left(\boldsymbol{Z}_2\boldsymbol{Z}_2'\right)\boldsymbol{X}\right]\right\}
\end{aligned}$$

$$= \frac{1}{\sigma_{\varepsilon}^{2}} \left\{ \boldsymbol{X}'\boldsymbol{X} - \frac{\sigma_{\gamma_2}^{2}}{\sigma_{\varepsilon}^{2} + m\sigma_{\gamma_2}^{2}} \left(\boldsymbol{X}'\boldsymbol{Z}_2\right)\left(\boldsymbol{X}'\boldsymbol{Z}_2\right)' \right\}.$$

Given that $[\boldsymbol{X}_1 : \boldsymbol{X}_3] = \boldsymbol{X}$, it follows that:

$$\boldsymbol{M} = \frac{1}{\sigma_{\varepsilon}^{2}} \left[[\boldsymbol{X}_1 : \boldsymbol{X}_3]'[\boldsymbol{X}_1 : \boldsymbol{X}_3] - \left(\frac{\sigma_{\gamma_2}^{2}}{\sigma_{\varepsilon}^{2} + m\sigma_{\gamma_2}^{2}} \right) \left([\boldsymbol{X}_1 : \boldsymbol{X}_3]'\boldsymbol{Z}_2\right)\left([\boldsymbol{X}_1 : \boldsymbol{X}_3]'\boldsymbol{Z}_2\right)' \right].$$