# Combining Data Mining and Evolutionary Computation for Multi-Criteria Optimization of Earthworks

Manuel Parente[1], Paulo Cortez[2] and António Gomes Correia[3]

[1] ISISE Institute for Sustainability and Innovation in Structural Engineering and ALGORITMI Research Centre, University of Minho, Guimarães, Portugal
map@civil.uminho.pt
[2] ALGORITMI Research Centre, Department of Information Systems, University of Minho, Guimarães, Portugal
pcortez@dsi.uminho.pt
[3] ISISE Institute for Sustainability and Innovation in Structural Engineering, University of Minho, Guimarães, Portugal
agc@civil.uminho.pt

**Abstract.** Earthworks tasks aim at levelling the ground surface at a target construction area and precede any kind of structural construction (e.g., road and railway construction). It is comprised of sequential tasks, such as excavation, transportation, spreading and compaction, and it is strongly based on heavy mechanical equipment and repetitive processes. Under this context, it is essential to optimize the usage of all available resources under two key criteria: the costs and duration of earthwork projects. In this paper, we present an integrated system that uses two artificial intelligence based techniques: data mining and evolutionary multi-objective optimization. The former is used to build data-driven models capable of providing realistic estimates of resource productivity, while the latter is used to optimize resource allocation considering the two main earthwork objectives (duration and cost). Experiments held using real-world data, from a construction site, have shown that the proposed system is competitive when compared with current manual earthwork design.

**Keywords:** earthworks; equipment allocation; metaheuristics; data mining

## 1 Introduction

Levelling the ground surface and preparing the required foundation conditions are necessary steps prior to the construction of most Civil Engineering structures. These steps are especially important in the construction of linear structures, as is the case of roads or railways, since they imply the levelling of large extensions of ground surface. In order to achieve this, engineers rely on heavy mechanical equipment, such as

excavators, dumper trucks, bulldozers and compactors, which allow them to handle large amounts of soil or other materials. Usually these tasks include excavating material from areas that are above the target height and transporting them to the areas below target height, where they are spread into layers and compacted, forming an embankment. The tasks associated with the usage of mechanical equipment to excavate, transport, spread and compact material in order to shape the ground surface in order to fulfil a specific purpose are often referred to as earthworks.

The design of earthworks tasks is often performed by a human expert. Such expert often uses her/his experience and intuition as the main criteria for the selection and allocation of resources throughout the construction process, in order to achieve a fixed trade-off between the two key earthwork design objectives, cost and duration. This is not a trivial task. Similarly to a wide range of real-word resource allocation tasks, the two-goal optimization is nonlinear and involves a large search space of design solutions for placing the available equipment in an earthworks project.

Considering such human allocation practice, there is a high potential for reducing costs and duration of earthwork projects by adopting artificial intelligence techniques, such as data mining and Metaheuristics. In effect, data mining techniques have been proposed within this domain, taking advantage of the recent increase of available construction databases to accurately predict the productivity of mechanical equipment given specific site conditions [1–6]. Moreover, several Metaheuristics, such as evolutionary computation, ant colony optimization and swarm intelligence, have been proposed for optimal allocation of resources within the earthworks domain [7–13].

This paper presents a proposal of an intelligent system that uses both data mining and evolutionary computation to tackle the multi-criteria optimization problem associated with resource allocation in earthwork construction. The optimization task addressed in this work is a particular instance of the more general job shop scheduling problem. Considering that the data mining approach has been presented in [18], a stronger emphasis is given towards the evolutionary multi-objective optimization component of the proposed system. The main contributions are associated with the architecture and methodology that comprise the presented optimization system. In terms of optimization methods, previous works either optimize a single objective, such as cost [7, 8] or duration [9], or adopt a weighted approach [10], that optimizes separately three duration-costs weight setups (i.e. 0.8/0.2; 0.7/0.3; and 0.5/0.5). In contrast with these solutions, the system discussed in this paper takes a Pareto front optimization approach, which not only optimizes both objectives simultaneously, but also outputs a set of interesting trade-off solutions. Depending on the budget and deadline restraints, the solution that best adjusts the objectives of the designer can then be selected. Furthermore, regarding productivity estimation, while existent applications lean on the experience of the designer [9, 10], resulting in rough estimation of equipment work rates, others attempt to build computer-demanding simulation models to solve this issue [7, 8, 11]. Contrariwise, the novel system uses data-driven models (fit to real data) to estimate equipment productivity, which allows for a realistic estimation. Finally, proposed the system is validated by experimenting with real-world data from a construction site and comparing the results with those obtained by conventional earthwork design.

The paper is organized as follows. Firstly, the optimization framework for the design of earthworks, where the earthwork problem is described as a series of simultaneous production lines, susceptible to optimization, is presented in Section 2. Then, a brief state of the art description of data mining and Metaheuristics applications to the earthwork domain is described in Section 3. Next, the multi-criteria optimization system is detailed in Section 4, featuring the description of the system and results that were obtained when applying such system with real-world data from a construction site. Finally, closing conclusions and perspectives of future work are presented in Section 5.

## 2 An optimization framework for the design of earthworks

Taking into account an optimization point of view, earthwork construction can be described as a number of production lines based on resources and dependency relations between sequential tasks. The resources are the mechanical equipment that is essential for the development of the project, namely excavators, dumper trucks, bulldozers and compactors, while the sequential tasks correspond to the associated processes, specifically excavation, transportation, spreading and compaction, respectively. The speed at which the latter can be completed depends on the amount of the former being allocated into each task. In other words, the work rate (in this case often measured in volume of handled material per hour, $m^3/h$) in each sequential task can be manipulated by increasing or decreasing the amount of associated resources allocated to it. This means that earthworks are strongly susceptible to optimization, which is aimed at minimizing both execution cost and duration. The multi-criteria include conflicting properties: in general, one can decrease execution duration by increasing the amount of allocated resources (mechanical equipment) to a task, but such results in an increase of the associated execution costs and vice-versa. However, it should be noted that the costs related to fuel and machinery maintenance (indirect costs) are substantial. Since these increase accordingly to the duration that mechanical equipment are working, a solution with the least possible amount of allocated resources is not necessarily the least costly. Thus, the optimal balance between the criteria must be established.

The tasks that comprise earthwork projects have a set of specific characteristics in this context, of which the focal point is interdependency. Indeed, earthwork tasks are not only sequential, but also the work rate of each of them is always limited to the work rate of its preceding task. For instance, the dumper trucks cannot undergo the transportation of soil if the latter has yet to be excavated and loaded into them; and bulldozers cannot spread soil into layers so as to allow compaction if the material has not been brought to them by the dumper trucks, and so on. Furthermore, when dealing with sequential and interdependent tasks such as these, the speed at which a single production line can carry out its work is equivalent to the work rate associated with its last task. In this context, maximizing the work rate in the final task (in this case, compaction) would correspond to a solution with minimum execution time for a production line. However, it is noteworthy to emphasize such allocation is limited by the

available equipment and also by the site conditions, such as space restrictions in excavation or compaction areas (usually designated as fronts). To fully take advantage of the available resources, one must guarantee that the allocated compaction equipment is fed enough material so as to allow for constant production. In other words, the work rate in all tasks prior to compaction (excavation, transportation and spreading) must be equal or similar to the work rate obtained in the associated compaction front. Should the work rate of a task fall short of the work rate of succeeding tasks, then the productivity of the whole production line will be limited to the one obtained in that task. This keeps the equipment from reaching its maximum potential in terms of work rate, i.e. by forcing it to idle while waiting for material. Therefore, it is essential to control the work rate in each task within a production line.

Naturally, an earthwork construction is not depicted in a single production line, but rather in several independent production lines working simultaneously. Each of these production lines is associated with a compaction front, since that is the final stage for handling the geomaterials. Moreover, there is one more characteristic specific to these production lines that significantly increases its complexity. As construction ensues in several simultaneous production lines, compaction work will come to completion in one production line at a time. At the point when one production line has completed its assignment, the associated equipment is no longer contributing towards the completion of the earthwork project, thus calling for its reallocation into either an existent or a new production line. However, considering that site conditions have changed since the previous allocation, this reallocation should include all available equipment once again if it is to keep its optimal status. Thus, the whole resource allocation must be reorganized in order to optimally resume the execution of the project. This enhances the problem with a dynamic nonlinear feature, which must always be taken into account in earthworks design.

## 3 Artificial intelligence in earthworks equipment allocation

### 3.1 Data Mining

The quality of an earthworks project design can only be as good as the ability to estimate the associated equipment productivity as close to reality as possible. Nowadays, this parameter estimation is often based on the experience of the designer. In most cases, designers either settle for a somewhat random distribution of equipment, just as long as it is feasible, or attempt to apply a set of standardized teams to every production line, of which an average productivity can roughly be estimated. Obviously these neither guarantee a good design, nor result in optimal executions of the projects. In this context, data mining provides an interesting alternative approach for estimating productivity parameters. Data mining [14] allows the extraction of useful knowledge (e.g. predictive models) from raw data (often based on vast databases and/or with complex relationships), searching for patterns and tendencies in the data. Guided by domain knowledge and under a semi-automated process that uses computational tools, data mining is an iterative and interactive process. Popular predictive

data mining models are based on machine learning techniques such as multiple regression (MR), artificial neural networks (ANN) [15] and support vector machines (SVM) [16]. These techniques are capable of automatically analyzing complex relationships in the data, turning them into knowledge which can be used to predict future values in new environments and for a better understanding of the problem domain variable relationships. Data mining is often framed in the context of a methodology, such as CRISP-DM (Cross Industry Standard Process for Data Mining) [17], which includes six phases (Figure 1) and facilitates the execution of data mining projects in real-world applications.

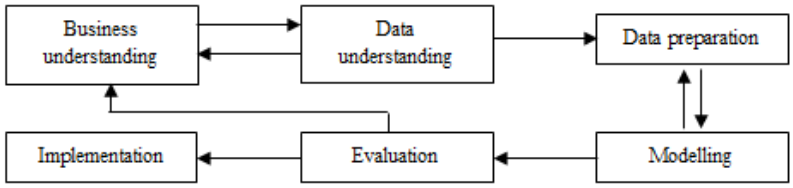


**Fig. 1.** The six phases of the CRISP-DM methodology (adapted from [17])

Most data mining applications to earthworks construction feature the estimation of equipment productivity, namely artificial neural network (ANN) for the estimation of excavation and transport equipment productivity rates [1, 3, 4], execution time and cost in earthwork [5] or productivity of earthwork production lines [6]. Other data mining techniques, such a multiple regression, have been successfully applied to the prediction of excavator cycle time [2]. However, a characteristic shared by all data mining techniques is that the quality of the models is highly dependent on the quality and availability the data used to fit such models. In cases for which databases are not available or do not include the necessary quality or variability to be targeted by data mining applications, it is still possible to use data stemming from technical guides or reports, which is a common practice within the geotechnical field. The practicality of this approach has been demonstrated by [18], in which several data mining models was adjusted for the purpose of estimating the productivity values for compactors in different conditions. In this work, the database consisted of the compaction tables featured in the GTR [19], a widely used empirical compaction guide. It encompassed several variables that were used as inputs for the model and can be summarized into qualitative variables (i.e., material and roller types, as well as compaction energy level) and quantitative variables (i.e., Q/S ratio, layer thickness, roller speed and number of roller passes). These inputs concern the use of soil in embankments and capping layers, representing a thorough description of the compaction conditions in regular construction cases (e.g., number of passes is one of the factors that determines the speed at which the compaction of each layer is completed, while layer thickness is directly related to the volume of material compacted after each set of roller passes). Using the R statistical tool [20] and the rminer library [21], the model that showed the best adjustment to the data was based on an ANN, achieving positive results. The ANN achieved a high coefficient of determination values (e.g., $R^2$=0.99), as well as

low values for root mean squared error and mean absolute error (e.g., RMSE=0.0068, MAE=0.0039), for unseen test data using 20 runs of a 10-fold cross-validation, corresponding to a reliable prediction model. Such predictive model is capable of automatically estimating the compaction conditions, specifically the productivity of compaction equipment for any practical case with high efficiency.

### 3.2 Metaheuristics

Although data mining can be used for estimating parameters with a good adjustment to reality, it cannot, by itself, guarantee an optimal solution in terms of execution costs and durations. Since these criteria are a function of the allocation solution chosen by the designer, optimization becomes a complex task. Considering the non-linear characteristics of the problem and since the solution space includes a large search space (in terms of distribution combinations of equipment throughout the construction site in each phase), conventional Operational Research (e.g. linear programming) and blind search methods are not effective for solving this problem. As such, Metaheuristics are an interesting solution within this domain, since they are capable of searching interesting search space regions under a reasonable use of computational resources. Indeed, several studies have followed this approach by using optimization methods such as Genetic Algorithms (GA) [7, 12, 22] and Swarm Intelligence [9, 10, 13, 23]. Yet, the optimization carried out in most of these systems (e.g., [7, 9, 10, 12, 13, 23]) still requires an estimation of parameters, especially equipment productivity, which is still left to the experience gathered by the designer or attempted to be estimated in theoretical simulation models. Moreover, many of these applications focus on single tasks or partial processes that comprise earthworks, i.e., excavation and hauling [9, 12], in an attempt to deal with the high complexity of the problem. For this reason, these systems lack the advantages of a global optimization of execution durations and costs throughout all construction phases. In terms of optimization objectives, existent systems tend to be limited to single objective optimization, such as cost [7] or duration [9], or attempt to consider both objectives via a weight-based optimization [10]. Although these solutions are considered effective in reducing computation effort requirements, they overlook the advantages of optimizing both objectives simultaneously. Even if it can be looked at as multi-criteria optimization, the weighted-based approach used in [10] only outputs a single trade-off for a particular weight combination (e.g. 0.8 for first criteria and 0.2 for second). However, as one can easily infer in non-trivial multi-criteria optimization problems, often there is not a single optimal trade-off solution, but rather a set of trade-offs with conflicting objectives. Thus, a much natural multi-criteria optimization approach is to optimize a Pareto front of solutions, where each solution is called non-dominated, or Pareto optimal, if none of the objectives can be improved in value without worsening the other. In the context of earthwork optimization, all Pareto-optimal solutions are considered equally good and the main choice criteria for selecting one solution over the other is often decided by the project designer based on the construction final deadline and/or budget. Obviously, secondary criteria may be used to support the final decision, such

as environmental aspects, which can be assessed by the determination of carbon emissions in each solution.

Taking into account that Pareto front multi-optimization requires the tracking of a population of solutions, population based Metaheuristics such as evolutionary computation, have become a natural and popular solution. Evolutionary computation is inspired in natural evolution and selection processes. Several computational variants have been proposed, such as GA [24], which are quite used within the earthwork construction domain. Evolutionary computation methods often start with a random population of possible solutions (or individuals), which are evaluated according to their fitness in a given situation. Then, the best fitted solutions are most likely to produce offspring, in the form of a new set of solutions that include characteristics from the individuals that originated then. This way, the initial set of solutions is improved in each iteration (or generation), ultimately coming to an optimal or near-optimal set of solutions.

Several evolutionary computation methods have been proposed for Pareto front optimization. In this work, we adopt the Non-dominated Sorting Genetic Algorithm-II (NSGA-II) [25] due two main reasons. Firstly, NSAGA-II is a popular and standard method for multi-criteria evolutionary optimization. Secondly, NSGA-II is easily available for a computational use in the R statistical tool [20] via the package mco [26], which is the same tool adopted for the development of our integrated optimization system. The R tool was selected since the data mining models (i.e., ANNs) were also fit using this computational environment, thus allowing an easier of integration of both data mining and NSGA-II methods. Moreover, the R tool includes several conventional optimization methods, such as the Linear Programming (LP) method that is used for individual fitness calculation (see Section 4.1).

# 4 Multi-criteria optimization of earthworks

## 4.1 System overview

The developed system is comprised of a data mining module integrated into the multi-objective optimization module of equipment distribution in earthworks. The first module takes care of estimating equipment productivity, while the second module carries out its optimal allocation. The system architecture falls into the framework proposed in [6]. In this work, the data mining module was applied to the GTR guide, aiming to determine compactor productivity given the material and site conditions.

The algorithmic flow for the multi-criteria evolutionary optimization method and its associated fitness function is shown in Figure 2. By interpreting the problem as a series of production lines, it becomes possible to focus the NSGA-II allocation of resources to the compaction task (last task of the production lines), which sets the work rate target value for each production line. Each solution is represents the compaction equipment for all necessary construction phases. For a particular construction phase, the solution is composed of a sequence of $C$ integer genes: $g_1\ g_2\ g_3 \ldots g_C$, where $g_i$ denotes the position of the $i$-th compactor (or roller) in terms of its compaction

front and $C$ represents the total number of compactors. The genes can take a value that ranges from 0 (not used) to the maximum number of compaction fronts $F$ that have to be completed. The whole individual (or chromosome) includes all construction phase gene sequences, thus the total number of genes corresponds to the number of available compactors times the number of necessary construction phases: $C$x$F$. For demonstration purposes, Figure 3 exemplifies a particular case where there are $C$=2 rollers and $F$=2, thus individuals are represented using four genes.

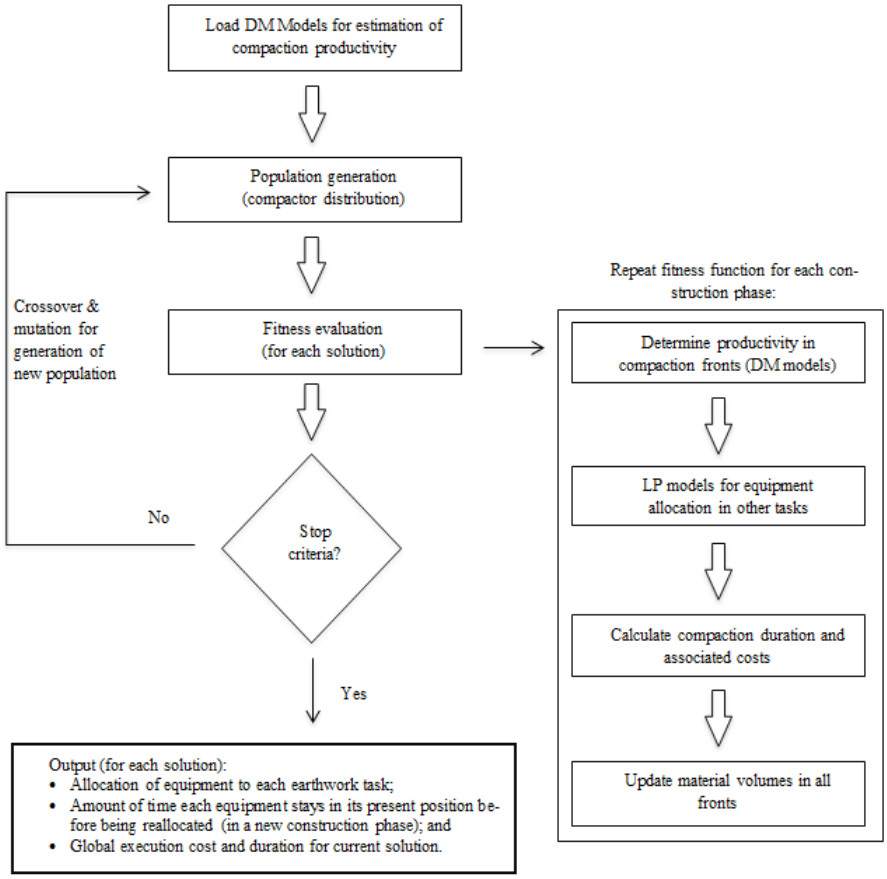


**Fig. 2.** Algorithmic flow of multi-criteria optimization system

At the start of the fitness evaluation procedure, only the genes that correspond to the first construction phase are selected and a repair strategy is implemented to ensure that all solutions are feasible. Given that the NSGAII implementation of the R tool only works with real values, the first step of the repair strategy is to round each chromosome to the nearest integer. Then, the work rate in each compaction front is then estimated by the data mining module. Under these conditions, the equipment for the remaining tasks (excavation, transportation, spreading) of each production line (asso-

ciated with each compaction front) is then distributed by using LP optimization models. Essentially, there is one LP model per equipment type (or per task type) which is responsible for distributing the associated equipment according to the work rate obtained in the last task of the each production line. Each model targets the minimization of the total cost of the allocated equipment, while ensuring that total productivity is as close as possible to the productivity estimated in the associated compaction front.

With the allocation process completed for each production line (or compaction front), it is then possible to ascertain which one will be completed first. The information regarding the duration and costs of the current phase (up to the point the compaction in one of the fronts is completed), as well as the completed front, is saved into memory. Then, the remaining material volumes in every other active compaction and excavation fronts are updated. As the genes for the next construction phase are selected and the previously described process is executed, a second step of the repair strategy is added, which verifies which compaction fronts have been completed. The second step assures that any compactor in the current construction phase that is allocated to a completed front is: a) allocated to another front, if possible; or b) not allocated (by changing the gene to zero). For executing step a), the gene is iteratively changed according to the rule: $g_i = (g_i+1) \bmod (F+1)$, until a feasible front value is found. This way the available equipment is reorganized throughout the construction fronts at the beginning of each construction phase, while assuring that work fronts that have already been completed are excluded from future allocations, as exemplified in the left of Figure 3. In each solution, this process is repeated for each construction phase, resulting in a determination of global costs and durations for the initial distribution of compaction equipment. The best solutions are then subjected to NSGA-II genetic operators, namely crossover and mutation, generating new solutions which are evaluated using the same methodology.


**Fig. 3.** Example of an initial chromosome for 2 compactors and 2 compaction fronts (left) and the final chromosome after the execution of the repair strategy (right)

### 4.2 Results

The proposed system was tested using real-word data from a construction site. A subset from a database that has been previously used during the development of the data mining models [6] was used as a reference. The available data includes the daily allocation of earthwork equipment throughout a road construction site, including information on available equipment, material volumes and types in excavation and compaction fronts and distances between fronts. The selected subset includes five production lines working simultaneously ($F$=5), to which equipment was originally allocated by conventional design methodologies. Since there is a total of C=5 availa-

ble compactors, the resulting individuals in the optimization system will be comprised of 5x5=25 genes each, defining the search space for this problem.

**Pareto results.** Using the methodology described in Section 4.1, the system outputs a Pareto-optimal set of solutions and their associated global costs and durations. The Pareto line represents several potential allocations of equipment throughout the construction site, in each construction phase, and this information can be accessed for each solution. Such Pareto-optimal set of solutions is useful for the designer or engineer, as she/he might want to choose different solutions depending on the available budget and the required deadlines. As earthwork construction is inherently a dynamic and unpredictable environment, different cost-duration solutions might become better adjusted to the ever-changing site conditions as construction develops.

The default parameterization of NSGA-II method, as implemented in the R tool, was adopted, namely: population size of 100, stop after 100 generations, crossover probability of 0.7 and mutation probability of 0.2. The rationale is to focus more on assessing and validating the capabilities of the proposed integrated system when compared with current human design, rather than calibrating the optimization algorithm. We note that in preliminary tests, smaller population sizes (e.g., 20) were explored, but the obtained results were worse than the default population size of 100. Also, the fitness evaluation is computationally costly, as it requires several data mining model estimations and LP optimizations (for each front), thus a population size much larger than 100 individuals would increase the computational effort. In effect, with the default NSGA-II population size, the method was executed with 3 runs on an Intel Core 2 Duo 2.66 GHz processor, and it required from 24 to 28 hours to complete each single run. The total computational time for the computational experiment (all 3 runs) was approximately 76 hours, even though this could be easily reduced using parallel computation (e.g., server with several multicore processors). The average Pareto-optimal front (over all 3 runs) is shown in Figure 4a, while a comparison between the conventional human design and those obtained by the proposed optimization system is illustrated in Figure 4b. The Pareto-optimal front (Figure 4a) is the result of a vertical averaging (i.e., according to the Duration objective) of the Pareto curves outputted by each run, using the averaging algorithm proposed by Fawcett [27] for vertical averaging of ROC curves. Since these are mean values, indication of the 95% confidence interval according to a t-student distribution was also included.

In this optimization attempt, the equipment available for the conventional design allocation was kept fixed. In other words, the presented results stem from a simple reorganization of the available equipment throughout the construction fronts, without the addition of any other piece of equipment. Bearing in mind Figure 4b, it is easy to infer how the solution obtained by conventional design is far from optimal. In fact, should this system be implemented for this construction project, a high impact could be achieved, with an estimated reduction of around 50% to 70% of both costs and durations. Still, the system does not take into account the occurrence of unpredictable events during construction (i.e., equipment malfunction). However, this could be mitigated by rerunning a new optimization procedure that included new restraints, which would result in a new set of optimal solutions for current site conditions.

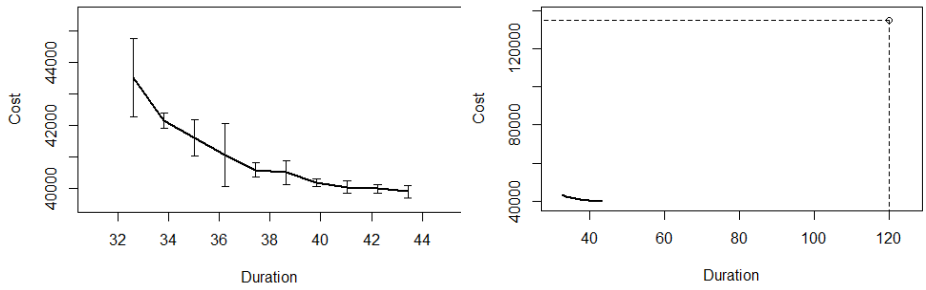


**Fig. 4.** Optimization results: a) vertically averaged Pareto-optimal front; b) comparison between optimized Pareto front (line) and the real-world human based allocation solution (dot); in both graphs, the Cost objective, *y*-axis, is presented in Euro, while the Duration objective, *x*-axis, is presented in hours

**Allocation analysis.** From the original solution regarding equipment allocation throughout the five production lines, a general improvement could be observed. In most cases, a reasonable reduction in both durations and costs was attained by the optimization system, which is the case of the second example described in Table 1 (production line 2). One production line featured a significant increase in global work rate without increasing costs, while in other cases considerable reductions in costs were achieved without a relevant increase in total duration. Regarding the latter, a comparison between the resource allocation in the original human based solution and the one obtained by the proposed optimization system is shown in Table 1 (production line 1). This also corresponds to the production line where the highest amount of material volume was handled. The same methodology was used to determine costs and durations for both the original and the optimized setup.

It is easy to infer that, for both production lines depicted in Table 1, the work rates in each task of the original setup are not homogeneous, as opposed to the work rates of the optimized solution. For both cases, the whole production line is limited by the work rate of excavators in the original setup, which means that the other tasks have to wait for material to be excavated in order to allow for its transport, spreading and finally compaction. This incurs in equipment idle time while waiting for material to be ready for handling, which represents wastes in terms of resources (since these do not work at full efficiency) and fuel (contributing to unnecessary costs), as well as an increase on unnecessary carbon emissions. As a result, the total work rate of these production lines cannot be considered superior to that of the minimum work rate obtained in the production line tasks, in this case excavation (1080 $m^3$/h in production line 1 and 394 $m^3$/h in production line 2). In contrast, the work rates obtained in the proposed optimized solutions for each task that comprises the production line are as homogeneous as possible, given the available equipment. As such, a constant flow of material throughout tasks can be achieved, using the allocated resources to their full potential and efficiency. It is noteworthy to emphasize that, besides optimizing the whole allocation in terms of costs and durations, the developed system is expected to always keep the allocated equipment working at full efficiency. This is done by min-

imizing equipment idle time as much as possible, which will also result in minimization of unnecessary carbon emissions. This is very challenging to achieve by conventional design methodologies.

Although the total work rate of the original setup is still slightly superior to the one obtained in its optimized counterpart in production line 1, the human based allocation solution features several pieces of equipment which are not necessary for its progress, as is the case of the six 40 ton dumpers that have been originally selected, for instance. In this case, the optimization system allocated five considerably smaller trucks (lower capacity, but lower fuel consumption and, thus, lower operation costs) to fulfil this role instead. As a result, the optimized setup for this case resulted in a decrease of 75% in total costs, while not incurring in any significant increase in duration (the actual increase in total duration is less than 1 hour of work). In the case of production line 2, besides solving the problem of work rate bottlenecks when compared to the original solution, the proposed optimized solution also features the allocation of higher productivity equipment. Consequently, a substantial decrease, over 50% in both cost and duration objectives, is obtained when comparing the integrated system optimized setup with the original human based solution. These results emphasize the importance of using intelligent computational tools for optimizing this type of construction works, also revealing how conventional human design allocation methodologies can be relatively counter-productive in some situations.

**Table 1.** Comparison between the conventional allocation the optimized allocation for two different production lines

| | Production line 1 | | Production line 2 | |
|---|---|---|---|---|
| Parameter | Original solution | Optimized solution | Original solution | Optimized solution |
| Average distance to excavation fronts (m) | 700 | | 175 | |
| *C* - Number of Compactors | 2 | 2 | 1 | 1 |
| Compactor work rate ($m^3$/h) | 1831 | 1008 | 614 | 1055 |
| Number of spreaders | 2 | 2 | 1 | 2 |
| Spreader work rate ($m^3$/h) | 1500 | 1088 | 413 | 1239 |
| Number of dumper trucks | 6 | 5 | 2 | 2 |
| Dumper truck work rate ($m^3$/h) | 2228 | 1009 | 2960 | 1600 |
| Number of excavators | 2 | 2 | 1 | 2 |
| Excavator work rate ($m^3$/h) | 1080 | 1080 | 394 | 1080 |

Finally, is important to note that the results associated with Figure 4 and Table 1 were obtained using an efficiency factor for mechanical equipment, k, of 0.75. This efficiency factor is related to the amount of time that the mechanical equipment spends in actual production. According to earthwork technical guides [19], actual "on-the-job" productivity is commonly influenced by factors such as operator skill, personal delays, job layout and other delays. Since one of the main focuses of this system is to maximize productivity of all the allocated mechanical equipment, it makes sense to consider the maximum value commonly suggested in earthwork technical guides (k=0.75). However, it is very hard to achieve the same efficiency factor in practice by means of conventional design, especially taking into account the fact that, as previously mentioned, it is mostly based on the experience of each designer. Additionally, unforeseen delays due to unpredictable situations that can occur in a real environment often have a significant impact on the actual efficiency factor of equipment in a construction site. In the present case, the available data indicates that the average actual efficiency factor for the mechanical equipment was just over k=0.3, which is not uncommon in this type of construction. As such, this large gap between these efficiency factors must be taken into account when analyzing the apparent discrepancy between the optimized results and the ones obtained by conventional design.

## 5 Conclusions and future work

Earthwork tasks are resource-dependant processes which aim to level target ground areas so as to allow for the construction of structures or infra-structures. Considering that these tasks represent a significant percentage of total execution durations and costs of road and railway projects, optimizing the resources involved is essential. However, the fact that conventional design methodologies lack the tools for optimal resource allocation can significantly hinder the durations and costs associated with the obtained solutions. Moreover, these methodologies are not prepared to keep up with the recent increasing demands regarding higher productivities and environmental aspects, such as minimizing carbon emissions.

In this work, a Pareto approach based on Non-dominated Sorting Genetic Algorithm-II (NSGA-II) was chosen as a basis for the development of an earthworks optimization system. The proposed system integrates several technologies, including artificial intelligence, in the form of evolutionary computation and data mining methods, and linear programming optimization, in an attempt to adjust to the complex reality associated with these types of constructions. The aim is to optimize the available resource allocation (represented by mechanical equipment) throughout the sequential tasks (namely excavation, transportation, spreading and compaction of geomaterials) that comprise the earthworks process. In this framework, the data mining technology supports the optimization techniques by providing realistic estimates to the productivity of the available equipment given site conditions.

Experiments have been carried out, using real-word data from a construction site and focusing on the assessment of the capabilities of the integrated system when compared with human allocation design. Competitive results were achieved by the

proposed system, stressing the importance of using intelligent optimization tools in the design of earthworks. Also, some limitations of conventional human allocation design were shown, in particular where the production line equipment is either significantly above the required work rate requirements (incurring in unnecessary costs) or below it (resulting in idle times and low efficiency ratios). Moreover, it was possible to verify the capability of the proposed system to distribute equipment in a relatively homogeneous way (when compared to conventional design), while minimizing costs and durations, which was the goal of this research.

Future work should include the addition of features which should allow a better adjustment to reality by the system, as is the case of better grasping of space restriction conditions in the construction site. The determination of carbon emissions, either to be used as secondary criteria or a minimization objective, also fits the future work category. Furthermore, the exploration of different NSGA-II parameterization (e.g. crossover and mutation probability), as well as other multi-objective optimization methods, such as Strength Pareto Evolutionary Algorithm 2 (SPEA-2) or S-Metric Selection Evolutionary Multi-objective Optimization Algorithm (SMS-EMOA), will be addressed in future work.

## Acknowledgement

The authors wish to thank FCT for the financial support under the doctoral Grant SFRH/BD/71501/2010.

## References

1. Shi, J.J.: A neural network based system for predicting earthmoving production. Constr. Manag. Econ. 17, 463–471 (1999).
2. Edwards, D.J., Griffiths, I.J.: Artificial intelligence approach to calculation of hydraulic excavator cycle time and output. Min. Technol. 109, 23–29 (2000).
3. Tam, C.M., Tong, T., Tse, S.: Artificial neural networks model for predicting excavator productivity. J. Eng. Constr. Archit. Manag. 9, 446–452 (2002).
4. Schabowicz, K., Hoła, B.: Application of artificial neural networks in predicting earthmoving machinery effectiveness ratios. Arch. Civ. Mech. Eng. 8, 73–84 (2008).
5. Hola, B., Schabowicz, K.: Estimation of earthworks execution time cost by means of artificial neural networks. Autom. Constr. 19, 570–579 (2010).
6. Parente, M., Gomes Correia, A., Cortez, P.: Artificial Neural Networks Applied to an Earthwork Construction Database. In: Toll, D., Zhu, H., Osman, A., Coombs, W., Li, X., and Rouainia, M. (eds.) Advances in Soil Mechanics and Geotechnical Engineering. pp. 200–205. IOS Press, Durham, UK (2014).
7. Marzouk, M., Moselhi, O.: Selecting Earthmoving Equipment Fleets Using Genetic Algorithms. In: Yucesan, E., Chen, C.-H., Snowdon, J.L., and Charnes, J.M. (eds.) Proceedings of the 2002 Winter Simulation Conference. pp. 1789–1796. , Montreal, Canada (2002).
8. Cheng, T., Feng, C., Chen, Y.: A hybrid mechanism for optimizing construction simulation models. Autom. Constr. 14, 85–98 (2005).

9. Kataria, S., Samdani, S.A., Singh, A.K.: Ant Colony Optimization in Earthwork Allocation. Int. Conf. Intell. Syst. 1–9 (2005).
10. Zhang, H.: Multi-objective simulation-optimization for earthmoving operations. Autom. Constr. 18, 79–86 (2008).
11. Cheng, F., Wang, Y., Ling, X.: Multi-Objective Dynamic Simulation-Optimization for Equipment Allocation of Earthmoving Operations. Constr. Res. Congr. 328–338 (2010).
12. Xu, Y., Wang, L., Xia, G.: Research on the optimization algorithm for machinery allocation of materials transportation based on evolutionary strategy. Procedia Eng. 15, 4205–4210 (2011).
13. Nassar, K., Hosny, O.: Solving the Least-Cost Route Cut and Fill Sequencing Problem Using Particle Swarm. J. Constr. Eng. Manag. 138, 931–942 (2012).
14. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery in Databases. Am. Assoc. Artif. Intell. 17, 1–18 (1996).
15. Haykin, S.: Neural Networks – A Compreensive Foundation. Prentice Hall (1999).
16. Hearst, M.A.: Support vector machines. IEEE Intell. Syst. 13, 18–28 (1998).
17. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 Step-by-step data mining guide. (2000).
18. Marques, R., Gomes Correia, A., Cortez, P.: Data Mining Applied to Compaction of Geomaterials. Eight International Conference on the Bearing Capacity of Roads, Railways and Airfields (BCR2A'09), Montreal, Canada, pp. 597-605, Taylor & Francis (2009).
19. SETRA, LCPC: Guide des Terrassements Routiers - Réalisation des Semblais et des Couches de Forme, (2000).
20. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2011).
21. Cortez, P.: Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool, In P. Perner (Ed.), Advances in Data Mining - Applications and Theoretical Aspects 10th Industrial Conference on Data Mining (ICDM 2010), LNAI 6171, pp. 572-583, Berlin, Germany, July, 2010, Springer (2010).
22. Moselhi, O., Alshibani, A.: Crew optimization in planning and control of earthmoving operations using spatial technologies. J. Inf. Technol. Constr. 12, 1–17 (2007).
23. Miao, K., Sun, X., Li, L.: A roadbed earthwork allocation model based on ACO algorithm. Appl. Mech. Mater. 44-47, 3483–3486 (2011).
24. Holland, J.H.: Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, Michigan (1975).
25. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6, 182–197 (2002).
26. Mersmann, O., Trautmann, H., Steuer, D., Bischl, B., Deb, K.: Package "mco": Multiple Criteria Optimization Algorithms and Related Functions, http://git.p-value.net/p/mco.git, (2014).
27. Fawcett, T.: An introduction to ROC analysis. Pattern Recognit. Lett. 27, 861–874 (2006).