

Data-Driven Surrogate-Assisted Multi-Objective Evolutionary Optimization of A Trauma System

Handing Wang, *Member, IEEE*, Yaochu Jin, *Fellow, IEEE*, Jan O. Jansen

Abstract—Most existing work on evolutionary optimization assumes that there are analytic functions for evaluating the objectives and constraints. In the real-world, however, the objective or constraint values of many optimization problems can be evaluated solely based on data and solving such optimization problems is often known as data-driven optimization. In this paper, we divide data-driven optimization problems into two categories, i.e., off-line and on-line data-driven optimization, and discuss the main challenges involved therein. An evolutionary algorithm is then presented to optimize the design of a trauma system, which is a typical off-line data-driven multi-objective optimization problem, where the objectives and constraints can be evaluated using incidents only. As each single function evaluation involves large amount of patient data, we develop a multi-fidelity surrogate management strategy to reduce the computation time of the evolutionary optimization. The main idea is to adaptively tune the approximation fidelity by clustering the original data into different numbers of clusters and a regression model is constructed to estimate the required minimum fidelity. Experimental results show that the proposed algorithm is able to save up to 90% of computation time without much sacrifice of the solution quality.

Index Terms—data-driven optimization, multi-objective optimization, evolutionary algorithm, surrogate, trauma system design.

I. INTRODUCTION

Evolutionary algorithms (EAs) are a class of nature-inspired global optimization algorithms. They distinguish themselves with traditional mathematical programming algorithms in that they do not require analytical models of the objective and constraint functions of the problem to be optimized and are less vulnerable to local optimums. For these reasons, EAs have enjoyed great success in solving a wide range of industrial problems [1], [2].

Although they do not require analytical objective or constraint functions, most EAs assume that analytical mathematical functions are available for evaluating the objectives and assess the constraints. Unfortunately, many real-world optimization problems do not have analytic objective or constraints

functions and optimization can be pursued only based on data collected from physical experiments, real events, or complex numerical simulations. Solving optimization problems solely based on data is known as data-driven optimization.

Most existing work on data-driven evolutionary optimization has actually been carried out in the area of surrogate-assisted evolutionary optimization, where fitness evaluations rely on very expensive physical experiments or highly time-consuming computer simulations, such as in drug design [3] and aerodynamic shape optimization [4], [5]. In addition, evolutionary optimization involving subjective human evaluations, which is known as interactive evolutionary optimization [6], [7], also falls in this type of data-driven optimization. We term this type of data-driven optimization *on-line data-driven optimization*. One main feature of on-line data-driven optimization is that during the optimization, it is still possible to acquire new data, though this might be costly or computationally expensive. For on-line data-driven optimisation, most existing surrogate-assisted EAs [8] are applicable.

By contrast, another type of data-driven optimization problems have largely been overlooked so far. In this type of optimization problems, no data can be collected per conducting additional experiments or computer simulations during the optimization, as each data sample may be an event that accidentally occurs. For example, the design of trauma systems can theoretically be based on comprehensive geospatial analysis. In practice, however, optimization of trauma system design can be accomplished using incidents in the previous years as an approximate geospatial guidance, as reported in [9], [10], where a multi-objective evolutionary algorithm (MOEA) has been employed to optimize the trauma system of Scotland. In such problems, the objective values may vary with the number of incidents and a larger number of the incidents can lead to richer geospatial information, and consequently more accurate optimization. Note however, that new records cannot be actively collected during the optimization process. We term such data-driven optimization *off-line data-driven optimization*.

Data-driven optimization usually needs to address multiple conflicting objectives, like in most real-world optimization problems. Such problems are known as multi-objective optimization problems (MOPs) [11]. Over the past two decades, a large number of multi-objective evolutionary algorithms (MOEAs) have been developed [12], which can be largely classified into three groups, i.e., dominance-based, aggregation-based, and performance indicator-based. Dominance-based MOEAs [13], [14] rely mainly on dominance comparisons for selecting parents [15]; aggregation-

This work was supported in part by an EPSRC grant (No. EP/M017869/1) on "Data-driven surrogate-assisted evolutionary fluid dynamic optimisation", in part by the National Natural Science Foundation of China (Nos. 61271301 and 61590922) and in part by the Joint Research Fund for Overseas Chinese, Hong Kong and Macao Scholars of the National Natural Science Foundation of China (No. 61428302).

H. Wang is with the Department of Computer Science, University of Surrey, Guildford GU2 7XH, U.K. (e-mail: wanghanding.patch@gmail.com).

Y. Jin is with the Department of Computer Science, University of Surrey, Guildford GU2 7XH, U.K. (e-mail: yaochu.jin@surrey.ac.uk). He is also affiliated with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, China.

J. O. Jansen is with NHS Grampian, Health Services Research Unit, University of Aberdeen, Aberdeen, AB25 2ZD, U.K. (e-mail: jan.jansen@abdn.ac.uk).

based MOEAs, also known as decomposition-based methods, optimize a set of single-objective sub-problems with a pre-assigned weight set [16], [17]; and indicator-based MOEAs employ a performance indicator as their selection criterion [18], [19]. However, little research work dedicated to addressing challenges in data-driven multi-objective evolutionary optimization has been reported so far, in particular off-line data-driven evolutionary optimization in the presence of huge amount of data.

This paper firstly provides a categorization of data-driven optimization problems and discusses challenges and opportunities in solving these problems using EAs. To the best of our knowledge, an elaborated categorization and discussion of data-driven evolutionary optimization are still missing in the literature. Then, the paper focuses on evolutionary design of a trauma system, a typical off-line data-driven bi-objective optimization problem in the presence of large amount of data [9]. Such optimization problems are computationally very intensive, since the objective as well as the constraint functions are evaluated based on a large number of incidents in the previous year. As it has been shown that the incidents are often distributed in special patterns [20], this work aims to reduce the amount of computation time by grouping the data into clusters and then using the cluster centers as representative data points for function evaluations. Consequently, the fewer clusters the data is grouped into, the faster the calculation will be, and the less accurate the evaluations are. Thus, the key challenge is how to adaptively balance the reduction of computation time and the accuracy of the estimated function values based on the clustered data.

The rest of this paper is organized as follows. We discuss in more detail the challenges in data-driven evolutionary optimization in Section II. The data-driven trauma system design optimization problem is introduced in Section III, followed by a description of an adaptive multi-fidelity surrogate management technique in Section IV. Empirical results are presented and discussed in Section V. Section VI concludes the paper.

II. DATA-DRIVEN EVOLUTIONARY OPTIMIZATION

A generic framework for data-driven evolutionary optimization is in Fig. 1. At each generation, EAs start with a population of candidate solutions (parents), and generate offspring solutions by applying variation operators, such as crossover, mutation, local search to the parent solutions. All offspring solutions are evaluated according to the objective and constraint functions before the parent individuals for the next generation can be selected. For data-driven evolutionary optimization, evaluation of the objectives and constraints are based on data.

As can be seen in Fig. 1, the main difference between data-driven evolutionary optimization and evolutionary optimization using analytical objective and constraint functions lies in function evaluations. In some cases, experimental or simulation data need to be preprocessed before they can be applied to function evaluations. More often than not, acquisition of data is either costly or computationally intensive, seriously limiting

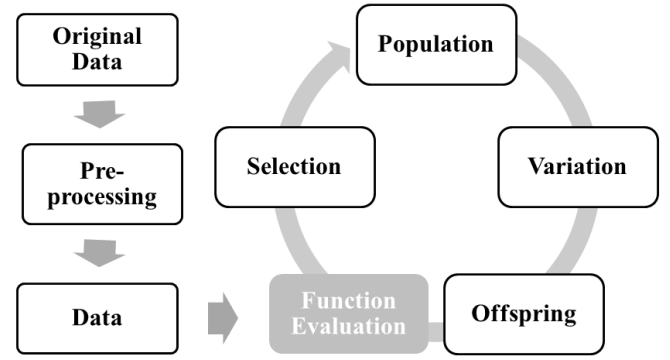


Fig. 1. A generic framework of data-driven evolutionary algorithms.

the number of function evaluations. One widely adopted technique to achieve acceptable solutions using a small number of function evaluations is to use surrogates, also known as metamodels or approximate function evaluations [8], [21] to replace in part the exact function evaluations in optimization. Management of the surrogate, including when to use and update the surrogates, plays a key role in surrogate-assisted optimization [22].

In the following, we elaborate the differences in managing surrogates in off-line and on-line data-driven evolutionary optimization.

A. Off-line and On-line Data-driven Optimization

Because only a very small number of exact function evaluations can be afforded, data-driven EAs usually resort to surrogates to reduce the needed number of expensive function evaluations. Depending on whether it is off-line or on-line data-driven optimization, the surrogate management strategies can vary, as illustrated in Figs. 2 and 3, respectively.

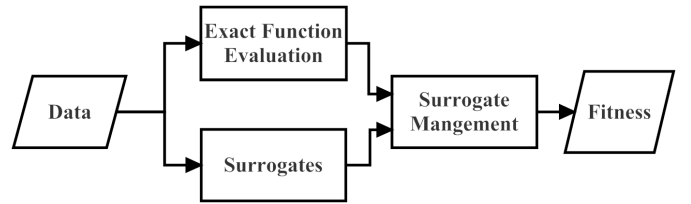


Fig. 2. Surrogate management in off-line data-driven evolutionary optimization.

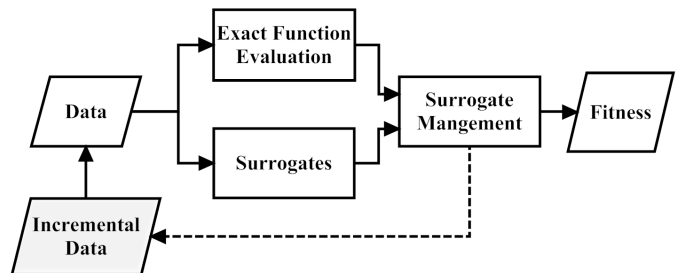


Fig. 3. Surrogate management in on-line data-driven evolutionary optimization.

From Fig. 2, we can see that in off-line data-driven optimization, data is acquired before evolutionary optimization. During the optimization, the EA can explore the given data as much as possible, but no new data will be made available during the optimization. By contrast, in on-line data-driven evolutionary optimization, new data can be generated during optimization, as shown in Fig. 3. However, the incremental data, those obtained during the on-line data-driven evolutionary optimization, may or may not be controlled by surrogate management. If the surrogate management strategy can actively sample new data at the desired points, the accuracy of the surrogates can be most effectively enhanced. However, in some specific cases, the surrogate management strategy does not have any control on the incremental data. This difference in controlling the incremental data generation is manifested by the dash line connecting the surrogate management block and the incremental data in Fig. 3.

As a whole, data-driven evolutionary optimization can be largely divided into three categories, namely, off-line data-driven optimization where no new data can be generated during optimization, on-line data-driven optimization with uncontrolled incremental data, and on-line data-driven optimization with controlled incremental data.

B. Surrogate Models and Surrogate Management

To reduce the needed number of expensive function evaluations, surrogate-assisted evolutionary algorithms (SAEAs) [22], [8], [21] can be used, where the main idea is to use one [23], [24] or multiple surrogates [25], [26], [27] to approximate the expensive function evaluation globally or locally [28], [29], [30]. Note that an implicit assumption here is that the computational cost for constructing and using the surrogates is much less than that for fitness evaluations using the original expensive function.

Many machine learning models can be used for surrogates, such as linear, nonlinear or polynomial regression models [31], Kriging or Gaussian processes [32], [33], [34], [35], [36], [37], [38], support vector machines (SVMs) [39], radial basis function (RBF) networks [40], [41], [42], and many other neural networks [43], [44], [45], [46], [47]. Several ideas have been proposed for choosing individuals to be re-evaluated using the original objective functions, which is one key issue in surrogate management. These include selecting potentially good solutions [48], [47], [49], selecting representative solutions [25], [50], or selecting solutions with a large amount of uncertainty [51], [52]. As an approximation of the original objective function, surrogate models are subject to errors. It has been indicated that errors introduced by surrogates are acceptable so long as they do not mislead the search [53], and in some cases they can even be exploited [53], [27]. Therefore, how to manage the trade-off between the accuracy and the number of exact function evaluations is the main issue in SAEAs. Metrics for measuring the quality of surrogates have been proposed in [53], [54].

It should be noted that most surrogate management techniques in the literature are proposed for on-line data-driven evolutionary optimization where the model management strategy has control over the incremental data.

C. Challenges

In the following, we discuss the main challenges to data-driven evolutionary optimization from the perspectives of computational cost, data quantity, data quality, and heterogeneity of the data, all of which are related to surrogate construction.

- **Computational cost:** Exact function evaluations in most data-driven optimization problems are expensive due to various reasons. For instance, the problems based on physical experiments or computational simulations, each function evaluation either reply on expensive physical experiments such as wind tunnel tests or crash test, or complex and highly time-consuming numerical simulations such as computational fluid dynamics simulations [5]. For most off-line data-driven optimization problems, such as design of trauma systems [9], [10], it often takes huge amount of time to collect a sufficient and representative amount of clinical and incident location data, to permit a precise evaluation of candidate designs. Evolutionary optimization of such problems will not be able to afford thousands of function evaluations as required by most existing EAs.
- **Data quantity:** Construction of surrogates, either computational models or any estimate of the exact function evaluations, is indispensable in data-driven optimization due to the high computational cost as mentioned above. However, construction of surrogates is often challenged by the amount, quality and distribution. Either a too small size or a very large size of training data samples will create difficulties for training surrogates. Lack of training data will make it impossible to build accurate surrogates when the dimension of the decision space is high [5], while large amounts of training may result in increasing computational cost [55], [56].
- **Data quality:** Quality of the data includes the distribution of the data and the amount of uncertainty in the data. It is very likely that the data is ill-distributed, imbalanced [57], incomplete [58], [59], and is contaminated by noise [60], [61]. In some cases, search for robust optimal solutions will be of great practical significance [62], [63].
- **Heterogeneity of data:** Finally, surrogate construction can be further complicated by the nature of the data. In other words, training data might come from multiple sources, presenting in a combination of heterogeneous forms such as numerical data, texts and images [64]. How to fuse the heterogeneous data may considerably affect the quality of the function evaluations and consequently the quality of the constructed surrogates [65].

To alleviate the difficulties resulting from the quality and heterogeneity of the data, properly pre-processing the data may be helpful. Most machine learning and data mining techniques can be adopted in the pre-processing data, such as data clustering [66], [67] for reducing the size of data, principal component analysis [68] and feature selection [69] for dimension reduction, and regression techniques [70] for learning the relationship in the data.

III. DATA-DRIVEN MULTI-OBJECTIVE OPTIMIZATION OF A TRAUMA SYSTEM

A trauma system is a network of hospitals, stratified by their clinical capability, and supported by emergency medical services [71]. The aim of a trauma system is to reduce death and disability from injury, by matching patients' needs with hospital resources. This concept is sometimes referred to as "right care, right time, right place" [72]. A well-designed trauma system should be both efficient and economical, i.e., providing the most effective treatment with the lowest resource cost [73]. On one hand, every patient should receive optimal care, as close to the point of injury as possible. On the other hand, the resources including trauma centers, surgeons, ambulances, and medicines are finite. Therefore, facilities and locations of trauma centers and the geospatial injury density affect the behavior of the whole trauma system [9]. Thus, the trauma system design problem becomes an optimization problem based on the geospatial information to search for the optimal configuration of trauma centers.

However, the geospatial information is difficult to measure directly. The incidents over a long time period can play that role indirectly, because the distribution of incidents implicitly shows the geospatially injury density. The incidents can be used to simulate the trauma system that is presented by one configuration, which can be evaluated by the simulated output (both clinical and resource) of those incidents. Thus, the function evaluation of the trauma system design problem relies on data, which makes it a data-driven optimization problem.

A. Allocation Algorithm for Patients

The trauma system of Scotland will have three types of trauma centers with different levels of capability [74], including major trauma centres (MTCs), trauma units (TUs), and local emergency hospitals (LEHs). An MTC provides clinical services for patients with major trauma. A TU handles less severely injured patients. An LEH manages patients with minor injuries. In the data, patients are classified into two groups, i.e., patients triaged to MTC or to TU. A patient will be sent to a center with suitable capability according to the patient's injury [75]. The priority of the patient allocation is shown as Fig. 4.

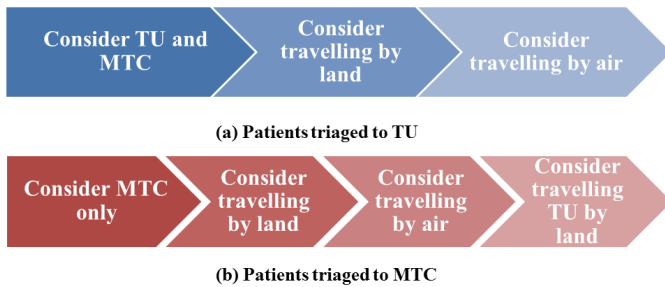


Fig. 4. Priority in the allocation algorithm for patients triaged to TU or MTC.

For a patient who is triaged to a TU, the allocation algorithm first considers travelling by land and chooses the center with the shortest drive time (either a TU or an MTC). If the

time of the shortest land transport is not acceptable, the allocation algorithm considers travelling by air and chooses the center with the shortest flight time (either a TU or an MTC). However, if the flight time is longer than the drive time, the allocation algorithm chooses to travel by land.

For a patient who is triaged to an MTC, the allocation algorithm considers the nearest MTC by land at first. If the drive time to the MTC is not acceptable, it considers travelling to the MTC by air. If the flight time to the MTC is still not acceptable, the allocation algorithm will consider a TU with shorter travel time by land, thus the level of facilities in the TU will be lowered for that patient.

In the data, the location and injury of every patient are recorded. Using the allocation algorithm, patients are assigned to a suitable trauma center. Taking a particular patient (patient i) as an example, the patient is assigned to a matched trauma center according to the accident location and severity of injury. The allocation algorithm will accordingly output information such as the travel time, assigned center, and helicopter usage, as shown in Fig. 5, where T_i is the traveling time from the location of the i -th patient to the assigned center. There is an exceptional situation where the patient should have been triaged to an MTC but had to be sent to a TU because of an unaccepted long travel time to MTC. This is indicated by the i -th bit (L_i) of a binary string L (1 for an MTC exception and 0 for not). The binary string H_i is of length of n_d (the number of helicopter depots) and the j -th ($1 \leq j \leq n_d$) bit of H_i is set to 1 if the i -th patient is assigned to the j -th helicopter depot. In other words, all bits of H_i are set to 0 if patient i is sent by land. The binary string G_i is of length n_{MTC} (the number of MTCs in the configuration), and the l -th ($1 \leq l \leq n_{MTC}$) will be set to 1 if the i -th patient is sent to the l -th MTC, otherwise to 0. Thus, all bits of G_i are set as 0 if the patient is sent to a TU. Table I shows an example of the output of the allocation algorithm for patient i , from which we can see that this patient was sent to the 4-th MTC by air from the 2nd helicopter depot, which took 25 minutes.

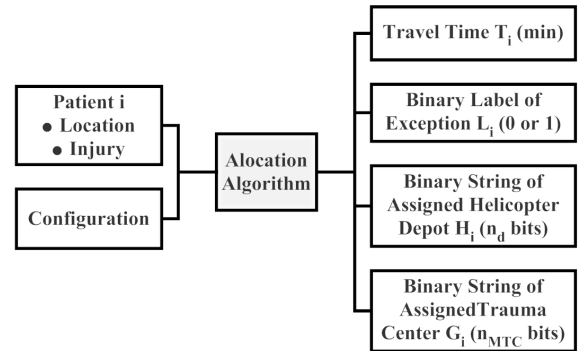


Fig. 5. Outputs of the allocation algorithm for one patient

B. Problem Formulation

The basic aim of a trauma system is to provide the best possible treatment for all patients, as quickly as possible. However, it is impossible to build a large number of trauma centers or send all patients by air given the limited resource.

TABLE I
EXAMPLE OUTPUTS OF THE ALLOCATION ALGORITHM FOR PATIENT i .

Output	Length	Value	Meaning
T_i	NA	25 min	Travel time equals to 25 min
L_i	1	0	Non-MTC exception
H_i	$n_d = 3$	010	The 2nd depot is used
G_i	$n_{MTC} = 4$	0001	Patient i is sent to the 4th MTC

There are conflicts between the resource and clinical outcomes in the design of a trauma system. From the perspective of a patient, the travel time from the location of the accident to the trauma center is important, the time should be as short as possible. However, it is also recognized that transport directly to a facility capable of providing definitive care, even if associated with slightly longer access times, is associated with better outcomes. Increasing the numbers of TU and MTC decreases the travel time of patients. However, the resources available for a trauma system are limited. In addition, MTC should handle as many patients as possible to gain the institutional expertise to effect improvements in outcome. Finally, to reduce the transport cost and risks, the helicopter usage should also be reduced. From the above discussions, we can see that many targets in designing a trauma system are conflicting with each other.

To date, 18 trauma centers in total have been built in Scotland. The design problem is to search for the optimal configurations of these centers to achieve both good use of resources and clinic outcomes. Different configurations result in different situations for these conflicting targets discussed above and it is usually hard to find a configuration that can simultaneously satisfy all these targets. Essentially, the aim of designing a geospatially optimized trauma system in Scotland is to find an optimized configuration of centers achieving the preferred trade-off between the conflicting targets [9]. In other words, the task is to assign different levels (MTC, TU, and LEH) to the existing centers, which can be seen as a combinatorial MOP [11] as described in the following subsections.

1) *Objectives*: As discussed above, design of a trauma system involves in many conflicting objectives. For simplicity, the problem is formulated as a bi-objective optimization problems, both related to the interest of patients. The first is to minimize the total travel time of all the patients, the other one is to minimize the number of exceptions, where a patient should have been sent to an MTC but was eventually sent to a TU because of the unaccepted long travel time needed to an MTC. The total travel time of all patients in the records can be described by Equation (1), where T_i is the travel time of patient i and N is the number of patient in the data:

$$f_1 = \sum_{i=1}^N T_i \quad (1)$$

The second objective, i.e., the number of exceptions can be calculated by Equation (2), where L_i records whether patient i is an exception, as shown in Fig. 5:

$$f_2 = \sum_{i=1}^N L_i. \quad (2)$$

2) *Constraints*: Although the financial aspects are not as important as patients, they impose the main limits for designing the trauma system. Therefore, helicopter usage, MTC case volume, and TU proximity are set as the constraints of the trauma system design problem.

The capacity of the j -th emergency depot is n_h^j helicopter transfers everyday. Therefore, the constraint of helicopter usage of the j -th helicopter depot should satisfy Equation (3), where D is the number of days and H_i^j is the j -th bit of H_i recording the usage helicopter depot for the i -th patient.

$$h_j = \sum_{i=1}^N H_i^j \leq D n_h^j, 1 \leq j \leq n_d \quad (3)$$

In a trauma system, an MTC has more facilities than a TU and therefore is expected to deal with more cases to gain experiences [76]. The number of patients sent to the l -th MTC everyday in the configuration should be higher than a threshold V . Thus, the constraints of the case volume of the l -th MTC is set as Equation (4), where G_i^l is the l -th bit of G_i recorded for the i -th patient as shown in Fig. 5.

$$g_l = \sum_{i=1}^N G_i^l \geq DV, 1 \leq l \leq n_{MTC}. \quad (4)$$

To control the number of TUs in the trauma system, the TUs should not locate too close to each other. Therefore, the distance between any two TUs should not be smaller than a predefined minimum distance d_{TU} :

$$Dis_{pq} \geq d_{TU}, 1 \leq p \leq n_{TU}, 1 \leq q \leq n_{TU}, p \neq q \quad (5)$$

where Dis_{pq} is the distance between the p -th and q -th TU, $1 \leq p, q \leq n_{TU}$.

In summary, for the configuration with n_{MTC} MTCs and n_{TU} TUs, there would be $n_d + n_{MTC} + C_{n_{TU}}^2$ constraints in total. Of them, $n_d + n_{MTC}$ constraints (h and g) are data-driven.

C. Multi-Objective Evolutionary Algorithm

According to the discussions in Section II, we can see that design of the trauma system is an off-line data-driven bi-objective optimization problem, as no incremental data can be made available during the optimization.

In the following, we present the representation of the configuration of a trauma system for evolutionary optimization.

1) *Genetic Coding*: The 18 candidate hospitals in Scotland are encoded in the following order: Glasgow Royal Infirmary, Aberdeen Royal Infirmary, Ninewells Hospital (Dundee), Dumfries and Galloway Royal Infirmary, Edinburgh Royal Infirmary, Forth Valley hospital, Hairmyres hospital (East Kilbride), Raigmore hospital (Inverness), Crosshouse hospital (Kilmarnock), Wishaw hospital, Royal Alexandra hospital (Paisley), Borders General hospital (Melrose), Southern General hospital (Glasgow), Ayr hospital, Victoria hospital

(Kirkcaldy), Monklands hospital (Airdrie), Inverclyde hospital (Greenock), and Perth Royal Infirmary. We use an integer between 0 and 2 to encode whether a particular center should be set as an MTC, a TU, or an LEH, encoded by "2", "1", or "0", respectively. Since there are 18 centers in the current trauma system, the length of the chromosome (representing one candidate configuration) is 18. For example, the chromosome [111112011100100101] means that Glasgow Royal Infirmary is set to an MTC, Aberdeen Royal Infirmary, Ninewells Hospital (Dundee), Dumfries and Galloway Royal Infirmary, Edinburgh Royal Infirmary, Forth Valley hospital, Hairmyres hospital (East Kilbride), Raigmore hospital (Inverness), Crosshouse hospital (Kilmarnock), Wishaw hospital, Royal Alexandra hospital (Paisley), and Borders General hospital (Melrose) are set to TUs, and Southern General hospital (Glasgow), Ayr hospital, Victoria hospital (Kirkcaldy), Monklands hospital (Airdrie), Inverclyde hospital (Greenock), and Perth Royal Infirmary are set to LEHs.

Based on the above coding for evolutionary design of the trauma system, NSGA-II is adopted for optimizing the bi-objective optimization problem. For reproduction, 3-point crossover with a probability of 1 and point mutations with a probability of 0.2 are employed.

2) *Local Search*: In addition to the genetic variations, variable neighborhood search (VNS) [77], a local search strategy particularly developed for solving combinatorial optimization problems, has also been employed.

The main idea of VNS is to search in a neighborhood starting from a local optimum of another neighborhood. The main reason is that for combinatorial optimization problems, it is easy to get trapped in a local optimum if the local search is performed only in one neighborhood. It has been shown that a large variety of neighborhoods helps improve diversity in local search [78]. Therefore in VNS [79], several different neighborhoods are used. Starting from a local optimum, VNS searches through one neighborhood to obtain a new local optimum, from where it searches within another neighborhood.

VNS is applied only to the non-dominated solutions at every generation. To adapt VNS to trauma system design, three neighborhoods NB_1 , NB_2 , and NB_3 have been defined. Neighborhood NB_1 is defined as all the feasible solutions varied by swapping a fixed gene to other genes. Neighborhood NB_2 is defined as all the feasible solutions varied by changing the value of a fixed gene. Neighborhood NB_3 is defined as all the feasible solutions varied by changing the values of two fixed genes.

Figure 6 provides an illustrative example of VNS for design of a trauma system with four centers, each candidate design consisting of four genes whose value can be "0", "1", or "2". Of the three neighborhoods, NB_1 contains four feasible solutions, NB_2 contains three feasible solutions, and NB_3 contains nine solutions. Assume solution [2011] is a non-dominated solution in the current population and the 2-nd gene of this solution is randomly chosen. Then all the feasible solutions in NB_1 are searched starting from this one. Now assume [2101] dominates [2011], which makes it the starting point to perform local search in NB_2 . In NB_2 , assume the last gene is chosen at random, after performing the local search in

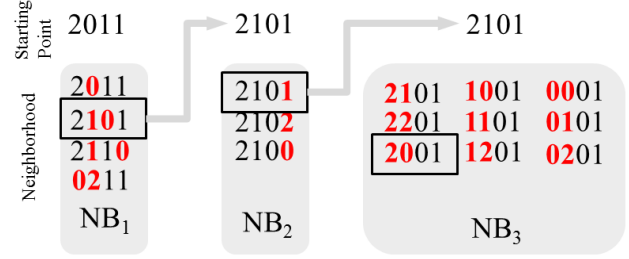


Fig. 6. An illustrative example of VNS for design of a trauma system with four centers.

NB_2 , no solution is able to dominate [2101]. Consequently, [2101] remains the starting point for performing search in NB_3 . According to the local search strategy for NB_3 , two genes, e.g., the 1st and 2nd genes are chosen at random. If among those solutions in NB_3 , solution [2001] dominates all other solutions, it will be the final output of the VNS.

IV. DATA-DRIVEN MULTI-FIDELITY SURROGATES FOR MULTI-OBJECTIVE OPTIMIZATION

Data-driven optimization of the trauma system using NSGA-II has been reported in [10]. In that work, each objective and constraint evaluation involves all patients in the data, which makes the evolutionary optimization very time-consuming. In fact, one evolutionary run using NSGA-II based on the data collected within one year takes about one day and if more data are to be used, the computation time will become prohibitive. In this work, we propose a multi-fidelity surrogate management strategy dedicated to a class of off-line data-driven optimization problems involving huge amount of data to reduce the computation time without degrading the performance of the solutions. Note that by surrogates in this work, we mean estimated function values calculated using part of the data records instead of using all. This is different from most surrogates that are computational models such as artificial neural networks or polynomials.

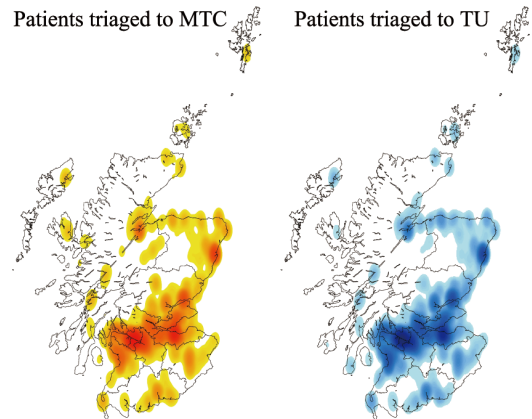


Fig. 7. Density maps of patients triaged to MTC and TU.

Optimization of the trauma system studied in this work relies on about 40,000 incidents (ambulance service patients) collected over a period of one year. Each record includes

information about whether the patient was triaged to MTC or TU care, the location of the incident and the travel time to all the centers by land and air. It previously been shown that the geographical distribution of the incidents shows a high degree of spatial correlation [20], which is related to the population density as shown in Fig. 7. Intuitively, if we cluster the incidents into K groups, where $1 \leq K \leq N$, where N is the number of records (patients) in total. Let C_i and N_i denote the center of the i -th cluster and the number of incidents in the cluster, respectively, where $1 \leq i \leq K$. Once the data is clustered, only one incident closest to C_i , instead of all N_i incidents in the i -th cluster will be used for objective and constraint evaluations. Thus, the number of clusters (K) will determine the fidelity of the objective and constraint evaluations, and the smaller K is, the less accurate the evaluations will be. In this case, the two objectives and the first two constraints in Equations (1)-(4) can be approximated calculated as follows.

$$\hat{f}_1 = \sum_{i=1}^K N_i T_i \quad (6)$$

$$\hat{f}_2 = \sum_{i=1}^K N_i L_i \quad (7)$$

$$\hat{h}_j = \sum_{i=1}^K N_i H_i^j, 1 \leq j \leq n_d \quad (8)$$

$$\hat{g}_l = \sum_{i=1}^K N_i G_i^l, 1 \leq l \leq n_{MTC} \quad (9)$$

A. Hierarchical Clustering for Discrete-Continuous Data

The data for each patient contains information on injury severity, both land and air travel time to all the centers, and air travel time from the helicopter depots to the patient, which means that the data is a mixture of binary and real-valued numbers. As Section III-A shows, the assigned center depends on the injury and location of the patient. The injury severity is a binary value (triaged to MTC or TU), and the location can be presented by the land travel time to all the centers, which is a real vector.

To handle the discrete-continuous data, a hierarchy clustering technique [80] is adopted in this work. In the first hierarchy clustering, only the location of patients are considered. The dimension of the vector of land travel time equals to the number of centers, which is much higher than the dimension of a location. Therefore, principal components analysis (PCA) [68] is employed to reduce the dimension. Then, patients are divided into two classes inside each cluster in the first hierarchy according to the severity of injuries, which forms the second hierarchy of clusters. Finally, the center of every cluster is used as the representative patient.

Fig. 8 is an example with 12 patients to illustrate the hierarchical clustering. We first divide the data into K clusters based on the location, where $K = 2$ in the example. Then, a further classification is applied inside the K clusters based on the injury severity, i.e. patients triaged to MTC are

classified into one sub-cluster, and those triaged to TU are classified into another sub-cluster. The center of every sub-cluster, represented by a pentagram in Fig. 8, will be used as one representative patient in the approximated objective and constraint evaluations using Equations (6)-(9).

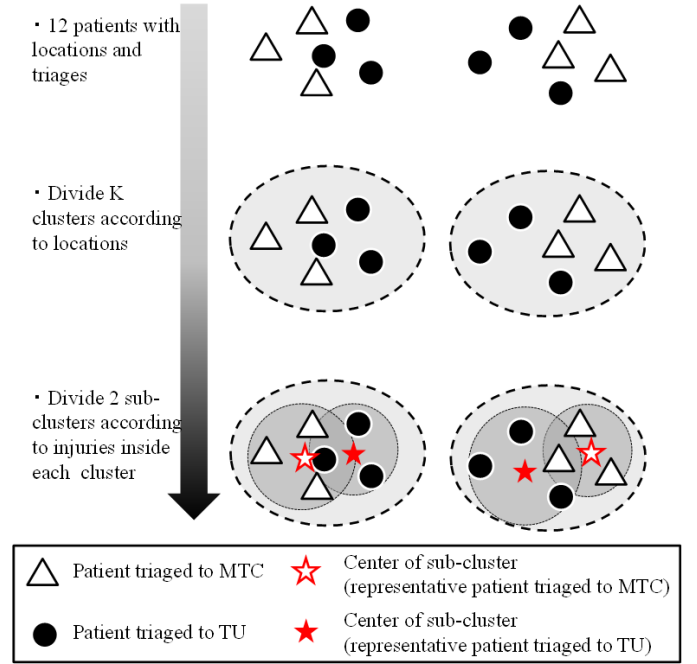


Fig. 8. An illustrative example of hierarchical clustering.

B. Analysis of Surrogates with Various Fidelities

As mentioned above, the geographical distribution of the data shows clearly spatial, which makes it possible to group the data into a number of clusters. Intuitively, the cluster number, K , influences not only the fidelity of the estimated objectives and constraints, but the computational expense as well. In order to better understand the influence of K on environmental selection and the accuracy of the estimated objective and constraint, we generate a reference set by running NSGA-II using exact function evaluations (i.e., using all patients in the data) using Equations (1) - (4) for 100 generations, where the population size is set to 100.

To examine the influence of the cluster number on selection, we compare the parents selected according to the objectives and constraints calculated using Equations (6) to (9) at each generation with those selected in the reference set and calculate the percentage of the correctly selected solutions out of 100 selected parents. This percentage is averaged over 100 generations and is used as an indicator for accuracy of selection. The change of the averaged accuracy of selection over the number of clusters is plotted in the left panel of Fig. 9.

Similarly, we calculate the mean absolute percentage error (MAPE) averaged over 200 individuals (the combined population before selection) at each generation for the two objectives and two constraints, respectively. Finally, these errors are again averaged over 100 generations and plotted in the right panel in

Fig. 9. Since the constraint on TU proximity is not data-driven, it is not considered here.

From the results in Fig. 9, we can see that as K increases, the average MAPE of the objectives and constraints decrease rapidly. Accordingly, the accuracy of selection increases. When K reaches 2000, the approximation errors do not decrease significantly and the selection accuracy no longer improves. This indicates that the EA will not benefit further when the number of clusters is larger than 2000, although the cost of computational expense. Thus, the key issue in saving computation time without sacrificing too much on the solution quality will be to properly tune the number of clusters so that the evolutionary algorithm is still able to find the optimal solutions at a much lower computational cost.

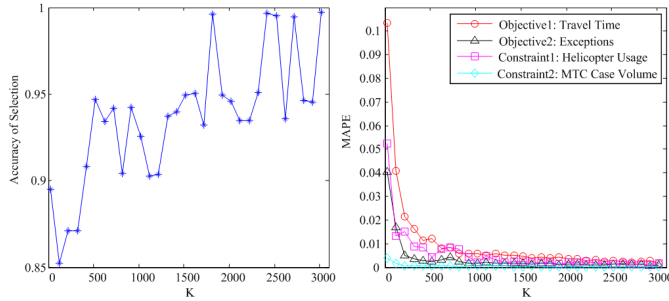


Fig. 9. Accuracy of selection and approximation error of the objectives and constraints over the change of the number of clusters K s.

C. Management of Multi-Fidelity Surrogates

The surrogate management by adjusting the fidelity of the estimated objectives and constraints by tuning the number of clusters K . One intuitive assumption here is the approximation is good enough if the estimated objective functions can ensure that the same individuals are selected as the exact objective functions. For this reason, let us take a look at the relationship between the approximation error on the two objectives and sorted no-dominated fronts. Fig. 10 shows the solutions in the first front (denoted by circles) and the last front (denoted by dots) that will be selected as parents based on the exact function evaluations. Now if a solution on the first front is evaluated using the approximated objectives and the MAPEs on the two objectives are ER_1 and ER_2 , respectively. As a result, the location of this solution evaluated based on the clustered data using Equations in (6) to (9) may locate anywhere in the shaded box defined by $2ER_1 \times 2ER_2$. When there is no overlap between two grey rectangles of the solutions from the first and last fronts, the selection will not be affected by the approximated errors. Therefore, the acceptable maximum error on the two objectives, ER_i^* , $i = 1, 2$, can be calculated by Equation (10), where F_1 is the set of solutions on the first front and F_l is the set of solutions on the last front that will be selected as parent solutions for the next generation.

$$ER_i^* = \frac{1}{2} \min\{f_i^k - f_i^j\}, 1 \leq k \leq |F_l|, 1 \leq j \leq |F_1| \quad (10)$$

Given the acceptable maximum errors ER_1^* and ER_2^* , the desired minimum number of clusters K can be determined if

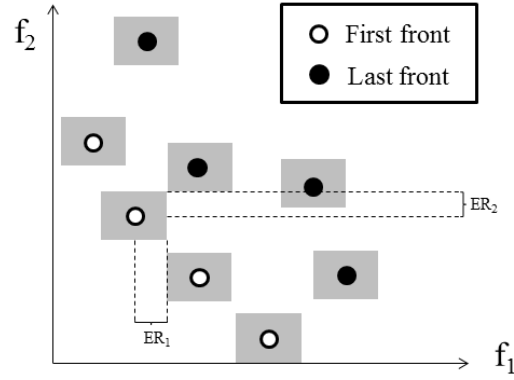


Fig. 10. Influence of the approximation errors on non-dominated sorting in a population.

the relationship between the maximum acceptable errors and the minimum K can be obtained. To achieve this, we construct a regression model with the historical data pairs between the cluster number and the approximation errors. For the trauma system design problem, f_1 is continuous and f_2 is discrete, the error on f_1 is continuous and that on f_2 is discrete. To benefit the regression model and simplify the management, we choose to use the errors on f_1 to estimate the needed number of clusters. In addition, as observed in Fig. 9, the relationship between the error and the number of clusters is nonlinear. Therefore, we use the following model to estimate the error from a given K :

$$ER = \frac{1}{\beta_1 + \beta_2 K}, \quad (11)$$

where β_1 and β_2 are two parameters.

Errors can be evaluated by comparing the approximated objective values with the exact evaluations. To reduce computational cost, we only evaluate the non-dominated solutions at each generation using exact function evaluations and calculate the approximation errors on the first objective. The approximation errors ER and the cluster numbers K are recorded to estimate the parameters in the regression model in Equation (11).

Fig. 11 illustrates one scenario of the proposed surrogate management strategy based on the regression model. Given the current cluster number K_1 , the maximum error caused by the approximate functions is larger than the acceptable error ER^1 , which means that K_1 is too small and should be increased. Based on the regression model, a new cluster number, K_2 is obtained. For K_2 , the new approximation error is calculated, which is still larger than the acceptable error ER^2 . This procedure continues until a cluster number K_4 is adopted, where results in an approximation error smaller than the acceptable error. Therefore, K_4 will be adopted as the cluster number for further evolutionary optimization.

The main steps of the algorithm for adapting K are presented in Algorithm 1. In the algorithm, K starts with a small number, typically the number of trauma centers. For a given K , the evolutionary optimization will be run for a number of generations no improvement can be achieved. This is re-

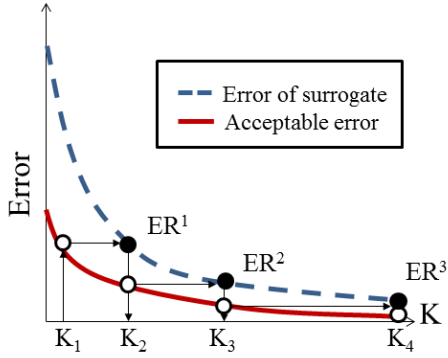


Fig. 11. An example of the surrogate management, where K is changed from K_1 to K_4 . When the approximation error for the current K is larger than the acceptable error, a new K will be calculated according to the regression model in an effort to reduce the approximation error to an accepted level.

Algorithm 1 Adaptive multi-fidelity surrogate using a regression model.

Input: K -the number of clusters, K_{max} -maximal K , ER^* -the acceptable error of f_1 , S -historical pairs of (K, ER) .

Output: K_n .

```

1: if no improvement in the non-dominated set then
2:   Evaluate the non-dominated solutions using all records
     (exact evaluations).
3:   Calculate  $ER$ .
4:   Add the pair  $(K, ER)$  to  $S$ .
5:   if  $|S| < 2$  then
6:      $K_n = 2K$ .
7:   else
8:     Regression on  $S$  by Equation (11).
9:     Calculate  $ER^*$  by Equation (10).
10:    if  $ER^* < ER/2$  then
11:       $ER^* = ER/2$ .
12:    end if
13:     $K_n = \frac{1}{\beta_2}(\frac{1}{ER^*} - \beta_1)$ .
14:  end if
15:  if  $K_n > K_{max}$  then
16:     $K_n = K_{max}$ .
17:  end if
18: else
19:    $K_n = K$ .
20: end if

```

alised by comparing the obtained non-dominated sets between consecutive generations. When there is no improvement, the algorithm considers the current K too small and then adapt the number K using the regression model.

A few special situations need to be taken into account. For example, if the historical data pairs (K, ER) are insufficient for building the regression model, K is simply doubled to add a new data pair. In addition, dramatic increase in K should be avoided, which is realized by limiting the maximum decrease in the acceptable error to the half of the current error, as described in lines 10-12 in Algorithm 1. Finally, we set the maximum of K to K_{max} to limit the computational cost.

V. EXPERIMENTAL RESULTS

A. Experimental Settings

To evaluate effectiveness of the proposed multi-fidelity surrogate management strategy, we compare the performance of three variants of the NSGA-II for the trauma system design, i.e., NSGA-II assisted with the adaptive multi-fidelity surrogate management strategy, NSGA-II with a fixed cluster number with K being fixed to 20, and NSGA-II using exact function evaluations. The population size of the three compared algorithms is set to 100 and the maximum of generations is 200. 20 independent runs are performed for the two variants using clustered data while five independent runs are performed for the NSGA-II using exact functions evaluations. Note that each run of the NSGA-II using exact function evaluations (i.e., using all 40,000 records) takes about 45 hours. The non-dominated solution sets obtained from the five independent runs of the NSGA-II using exact fitness evaluations are combined and the non-dominated solutions of these combined solutions are used as the reference set for calculating performance indicators such as the inverted generational distance (IGD) [81]. Note that we normalize the objectives by the extreme solutions in the reference set because of the different scale of two objectives. Furthermore, we use the number of calls to the patient allocation algorithm to evaluate the computational expense. For each exact function evaluations, the number of calls equals to the number of patients, while for clustered data, the number of calls equals to the number of data clusters.

B. Non-Adaptive Surrogate Management

We first examine the performance of the NSGA-II using clustered data with the cluster number K being fixed during the evolutionary optimization. To investigate the influence of the cluster number K on the performance, K has been set to from 100 to 2000 with an increment of 100. For each fixed K , 20 independent runs are performed. The average IGD value and the number of allocated patients over the change of K are shown in Fig. 12.

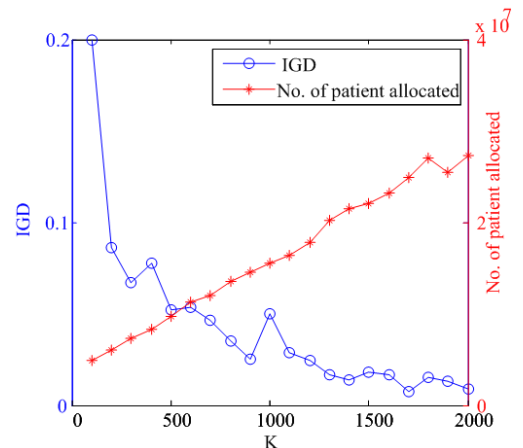


Fig. 12. Average IGD values and the number of allocated patients for different K s.

From Fig. 12, we can see that the number of allocated patients linearly increases with the growing K , which represents the computation time. As K increases, more patients (represented by the cluster centers) need to be allocated in objective evaluations, leading to increases in the computation time. At the same time, the approximation accuracy of the surrogates as the cluster number increases resulting in a better IGD. However, the improvement of the IGD is minor when K is larger than 1500 in comparison to the increases in the computational cost.

C. Adaptive Surrogate Management

1) *Maximal Number of Clusters*: For the surrogate management algorithm with an adaptive cluster number, the lower and upper bounds of K need to be predefined. As we mentioned before, the lower bound of K is usually set to the number of trauma centers (18 in this example), while the upper bound K_{max} is a parameter to be specified. In the following, we first test the sensitivity of the performance to different specifications of K_{max} .

For this purpose, we run the proposed algorithm with K_{max} being set to 500, 1000, 1500, and 2000, respectively, and each case is run for 20 independent times. The trade-off between the performance and computation time is plotted in Fig. 13, where purity (the rate of non-dominated solutions in the union set of the compared algorithms) [82] is used as the performance indicator.

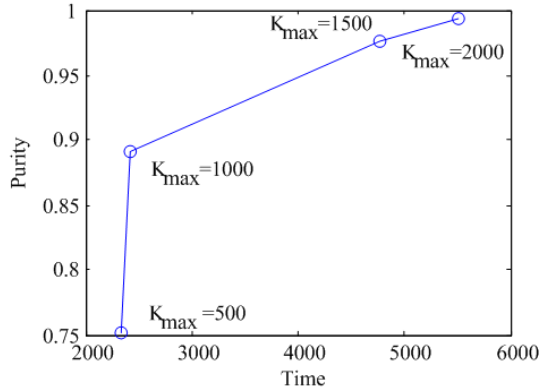


Fig. 13. Trade-off between the performance in terms of purity and computation time in the the proposed multi-fidelity surrogate management scheme.

From Fig. 13, it is clear that a larger K_{max} can bring about better performance but also increases computation time. Interestingly, we notice there is a knee point in the trade-off between purity and computational cost when $K_{max} = 1000$.

2) *Adaptive Number of Clusters Based on Regression*: To better understand the behavior the proposed algorithm, we perform additional runs of the proposed algorithm with $K_{max} = 1000$ for 20 independent times in order to take a look into the details in adaptive surrogate management, including the changes of the acceptable error, the parameters in the regression model, and the number of clusters K during the whole evolutionary optimization.

The average ER (the real approximation error of the surrogate) and ER^* (the maximum acceptable error) as K adapts are presented in Fig. 14. We can see that both ER and ER^* decrease as K increases, which is very similar to the changes shown in Fig. 9. However, the decrease of ER^* is not as significant ER . ER is larger than ER^* , which is the reason why K keeps being changed. Once ER is smaller than ER^* , K stops changing. Another observation we can make is that at the end of the evolutionary run, the difference between ER and ER^* becomes very small as K approaches to $K_{max} = 1000$.

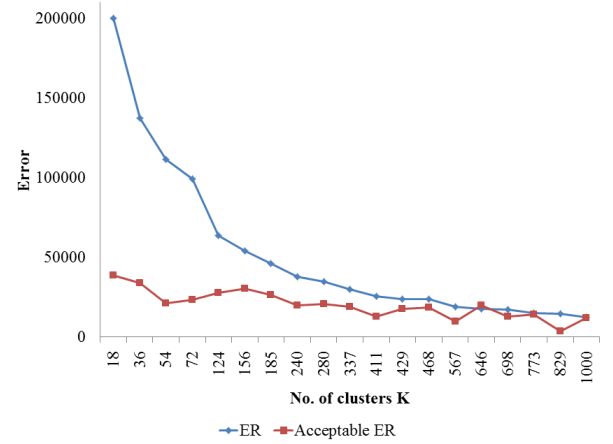


Fig. 14. Average ER and ER^* over different K s.

To check whether the estimation of the parameters in the regression model described in Equation (11) converge, Fig. 15 records the change of the parameters β_1 and β_2 over the change of K during the evolutionary optimization. The results are averaged over 20 runs. From these results we can see that the estimation of the parameters in the regression model is able to converge.

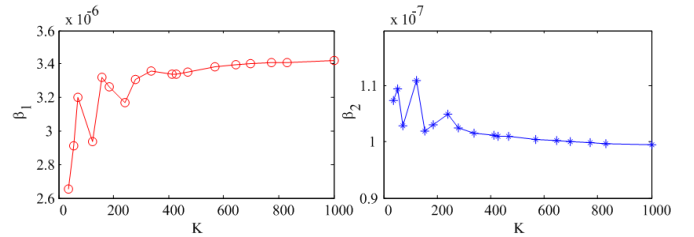


Fig. 15. The convergence profile of β_1 and β_2 as the cluster number K increases during the evolutionary optimization. Results are averaged over 20 runs.

Furthermore, we record in Fig. 16 the changes of K as the evolution proceeds and compare the difference in the change of K with and without the smoothness control strategy as implemented in lines 10-12 in Algorithm 1. From these results can see that without the smoothness control, K increased dramatically and altogether 11 changes are needed, whereas with the change smoothness control, K increases more smoothly and only 10 changes are needed. Also, we will show in the next section that the smoothness control in change of K results in slightly better performance.

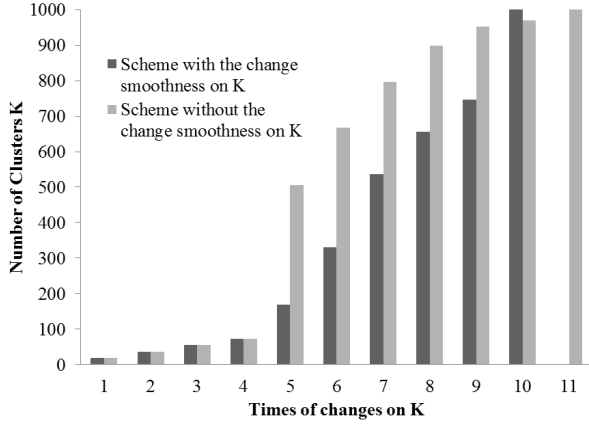


Fig. 16. Changes of K in the adaptive surrogate management schemes with and without the change smoothness.

D. Comparison of Computational Cost

In this subsection, we compare the computational cost of the NSGA-II using proposed adaptive surrogate management scheme with the NSGA-II using exact function evaluations (NSGA-II for short). Here, we also consider two minor variants of the proposed algorithm, one with smoothness control in change of K , termed SA-NSGA-II with CS, one without, termed SA-NSGA-II without CS. The three compared algorithms are all tun for 100 generations, however, NSGA-II is repeated for five times due to its high computational cost, while the two variants of SA-NSGA-II are repeated for 20 times. Here, we use the non-dominated set of the combined results from the five runs of the NSGA-II as the reference set for calculating the IGD. The IGD values and computation times (in seconds) of these three algorithms are presented in Table II. From these results, the IGD values of both SA-NSGA-II variants are very close to 0, which means that the solution sets they obtained are very similar to the reference set. However, the computation time needed for the two variants of SA-NSGA-II is much shorter than NSGA-II and with the smoothness control of changes in K , the computation time is further reduced with even slightly better performance. SA-NSGA-II without CS takes longer time than SA-NSGA-II with CS, because the former algorithm calls the hierarchy clustering approach more times than SA-NSGA-II with CS as shown in Fig. 16, where the hierarchy clustering approach costs extra computational time. However, the consuming time by the hierarchy clustering approach is much shorter than its saving time from a large number of evaluations, which can be observed by the large difference between the time of SA-NSGA-II and NSGA-II.

TABLE II
COMPARISON OF PERFORMANCE AND COMPUTATIONAL COST

	IGD	Time (s)
SA-NSGA-II with CS	$2.47\text{e-}02 \pm 1.96\text{e-}02$	$6.54\text{e}+03 \pm 3.57\text{e}+03$
SA-NSGA-II without CS	$2.51\text{e-}02 \pm 2.79\text{e-}02$	$10.31\text{e}+03 \pm 5.00\text{e}+03$
NSGA-II	$0.00\text{e}+00 \pm 0.00\text{e}+00$	$8.14\text{e}+04 \pm 5.36\text{e}+02$

While the above comparison is based on the same number of generations, it is also of interest to compare the performance

of the algorithms using the computation time as the baseline. For this purpose, we set the stopping criterion to be 1 hour for each of the three compared and then calculate the IGD value of the solution set using the same reference set as the above experiment. The results averaged over 20 independent runs are presented in Table III. By using the Wilcoxon signed-rank test [83], it has been found that the IGD values of the two SA-NSGA-II variants are significantly better than that of the NSGA-II. In addition, we can see that the IGD value of SA-NSGA-II with CS is again better than SA-NSGA-II without CS control. The non-dominated solution sets obtained the three algorithms in the run with the median IGD value are plotted in Fig. 17.

TABLE III
THE IGD VALUES OF THE SA-NSGA-II VARIANTS AND NSGA-II USING 1 HOUR RUNTIME AS THE STOPPING CRITERION.

	IGD
SA-NSGA-II with CS	$2.69\text{e-}02 \pm 1.93\text{e-}02$
SA-NSGA-II without CS	$4.31\text{e-}02 \pm 1.01\text{e-}01$
NSGA-II	$9.57\text{e-}02 \pm 5.08\text{e-}02$

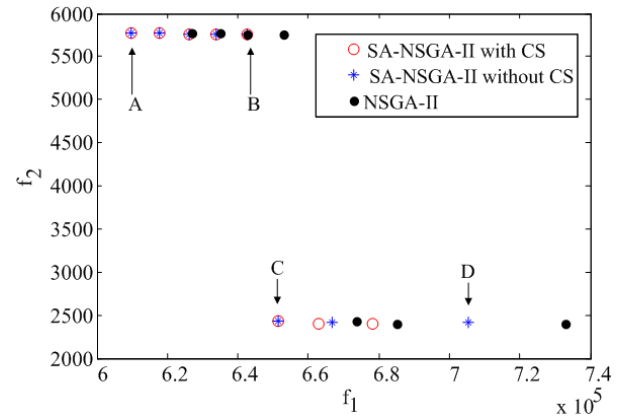


Fig. 17. Non-dominated solution sets of the SA-NSGA-II variants and NSGA-II obtained by stopping the algorithms after 1 hour.

It is of interest to take a look at a few designs selected from the non-dominated solution set to understand the optimal designs obtained by SA-NSGA-II with CS. The resulting configurations of the four solutions, as indicated by A, B, C, and D in Fig. 17, are shown in the map in Fig. 18. It is clear that those solutions can be classified into two categories, where in solutions A and B one MTC is established in Aberdeen, and in solutions C and D one MTC in Glasgow. These two different choices for MTC actually match the patient density in Fig. 7, where Aberdeen and Glasgow are two centers of high incidence areas. Since Aberdeen is located closer to the middle of Scotland than Glasgow, the total traveling time of solutions A and B is less than that of C and D. However, in A and B, patients triaged to MTC in the south of Scotland cannot be timely sent to the MTC in Aberdeen, which increases the number of MTC exceptions. Therefore, C and D are better than A and B in terms of MTC exceptions. Compared with A and C, B and D have a smaller number of TUs, which significantly increases the traveling time. It is also noticed that the obtained



Fig. 18. Different solutions (A-D) on the map, where triangles are MTCs, circles are TUs, and plus signs are LEHs.

solutions requires only one MTC, which is smaller than the currently proposed configuration in Scotland, which has four MTCs.

The currently proposed configuration of the network in Scotland is not efficient, it cannot satisfy the constraint in Equation (4) because the case volume of several of the MTCs would be too low. Therefore, the currently proposed configuration cannot be compared directly with the configurations obtained by the proposed algorithm. To show the effectiveness of the proposed algorithm, we relax the constraints until the currently proposed configuration becomes feasible, even though the new constraints are not clinically reasonable. Then, with the relaxed constraints, we perform 20 independent runs, using SA-NSGA-II with CS. In each run, SA-NSGA-II stops after 1 hour. The obtained solutions and the current configuration are shown in Fig. 19. Some of the obtained solutions dominate the currently proposed configuration. Also, other solutions in the obtained solution set can be options for decision markers. Therefore, our proposed algorithm can improve the clinical and resource output of the trauma system within an acceptable time.

VI. CONCLUDING REMARKS

This paper discusses data-driven optimization problems and categorizes them into two large groups, one termed off-line data-driven optimization problems where no incremental data

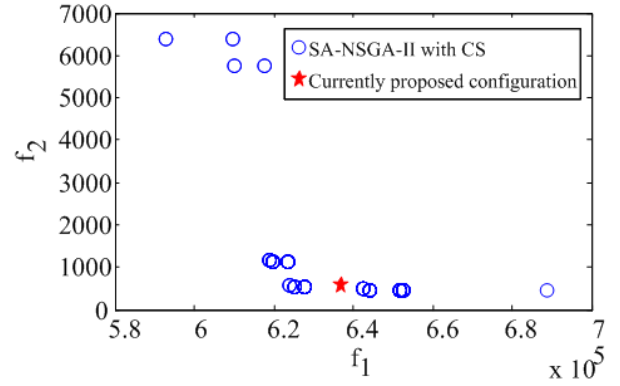


Fig. 19. Non-dominated solution sets of the SA-NSGA-II obtained by stopping the algorithm after 1 hour.

is available during optimization, and the other on-line data-driven optimization problems where new data can be collected during the optimization. The latter can further be divided into two types of problems, one of which can control the position of the newly sampled data, while the other cannot. One of the principal challenges in data-driven evolutionary optimization, in addition to the high cost of collecting the data, lies in the quantity, quality and other nature of the data. As data-driven evolutionary optimization typically needs to be assisted by surrogates, most challenges in surrogate-assisted evolutionary optimization will also need to be addressed. However, not much work on surrogate-assisted evolutionary optimization has considered difficulties in the presence of huge amount of data, or imbalanced, ill-distributed, or heterogeneous data.

In this work, we have addressed the challenges inherent in designing a trauma systems, which is a real-world off-line data-driven evolutionary optimization that involving very large amounts of data. The main idea is to group the data into a number of clusters using a hierarchical clustering algorithm, thereby reducing the amount of computation time. To achieve an optimal balance between computation time and the solution quality, we propose a surrogate management scheme by establishing a regression model that can estimate the number of clusters required based on the maximum acceptable approximation error. The experimental results demonstrate that the proposed algorithm is able to considerably reduce computational time of the data-driven EA using all data for function evaluations.

Although the proposed multi-fidelity surrogate management strategy was specifically developed for the design of trauma systems, the main ideas in this algorithm are likely to be applicable to other data-driven optimization problems relying on huge amount of data for function evaluations. However, the clustering algorithm used and the regression model for estimating the needed number of clusters may need to be tailored. Future efforts should also address the other challenges imposed by the particular nature of the data, e.g., by combining advanced learning and evolutionary techniques for dealing with big data [56], [84].

REFERENCES

- [1] P. J. Fleming and R. C. Purshouse, "Evolutionary algorithms in control systems engineering: a survey," *Control Engineering Practice*, vol. 10, no. 11, pp. 1223–1241, 2002.
- [2] D. Dasgupta and Z. Michalewicz, *Evolutionary algorithms in engineering applications*. Springer Science & Business Media, 2013.
- [3] K. De Grave, J. Ramon, and L. De Raedt, "Active learning for high throughput screening," in *Discovery Science*. Springer, 2008, pp. 185–196.
- [4] K. Giannakoglou, "Design of optimal aerodynamic shapes using stochastic optimization methods and computational intelligence," *Progress in Aerospace Sciences*, vol. 38, no. 1, pp. 43–76, 2002.
- [5] Y. Jin and B. Sendhoff, "A systems approach to evolutionary multiobjective structural optimization and beyond," *IEEE Computational Intelligence Magazine*, vol. 4, no. 3, pp. 62–76, 2009.
- [6] H. Takagi, "Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation," *Proceedings of the IEEE*, vol. 89, no. 9, pp. 1275–1296, 2001.
- [7] A. Eiben and J. Smith, "Interactive evolutionary algorithms," in *Introduction to Evolutionary Computing*. Springer, 2015, pp. 215–222.
- [8] Y. Jin, "Surrogate-assisted evolutionary computation: recent advances and future challenges," *Swarm and Evolutionary Computation*, vol. 1, no. 2, pp. 61–70, 2011.
- [9] J. O. Jansen, J. J. Morrison, H. Wang, R. Lawrenson, G. Egan, S. He, and M. K. Campbell, "Optimizing trauma system design: the GEOS (geospatial evaluation of systems of trauma care) approach," *Journal of Trauma and Acute Care Surgery*, vol. 76, no. 4, pp. 1035–1040, 2014.
- [10] J. O. Jansen, J. J. Morrison, H. Wang, S. He, R. Lawrenson, J. D. Hutchison, and M. K. Campbell, "Access to specialist care: optimizing the geographic configuration of trauma systems," *Journal of Trauma and Acute Care Surgery*, vol. 79, no. 5, pp. 756–765, 2015.
- [11] K. Miettinen, *Nonlinear multiobjective optimization*. Springer, 1999.
- [12] A. Zhou, B. Qu, H. Li, S. Zhao, P. Suganthan, and Q. Zhang, "Multiobjective evolutionary algorithms: a survey of the state of the art," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 32–49, 2011.
- [13] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [14] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.
- [15] H. Wang and X. Yao, "Corner sort for Pareto-based many-objective optimization," *IEEE Transactions on Cybernetics*, vol. 44, no. 1, pp. 92–102, 2014.
- [16] Q. Zhang and H. Li, "MOEA/D: a multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.
- [17] K. Li, Q. Zhang, S. Kwong, M. Li, and R. Wang, "Stable matching-based selection in evolutionary multiobjective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 6, pp. 909–923, 2014.
- [18] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *Parallel Problem Solving from Nature-PPSN VIII*. Springer, 2004, pp. 832–842.
- [19] H. Wang, L. Jiao, and X. Yao, "Two_Arch2: An improved two-archive algorithm for many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 4, pp. 524–541, 2015.
- [20] J. O. Jansen, J. J. Morrison, H. Wang, S. He, R. Lawrenson, M. K. Campbell, and D. R. Green, "Feasibility and utility of population-level geospatial injury profiling: prospective, national cohort study," *Journal of Trauma and Acute Care Surgery*, vol. 78, no. 5, pp. 962–969, 2015.
- [21] M. Tabatabaei, J. Hakanen, M. Hartikainen, K. Miettinen, and K. Sindhya, "A survey on handling computationally expensive multiobjective optimization problems using surrogates: non-nature inspired methods," *Structural and Multidisciplinary Optimization*, vol. 52, no. 1, pp. 1–25, 2015.
- [22] Y. Jin, "A comprehensive survey of fitness approximation in evolutionary computation," *Soft Computing*, vol. 9, no. 1, pp. 3–12, 2005.
- [23] J. Knowles and H. Nakayama, "Meta-modeling in multiobjective optimization," in *Multiobjective Optimization*. Springer, 2008, pp. 245–284.
- [24] I. Loshchilov, M. Schoenauer, and M. Sebag, "A mono surrogate for multiobjective optimization," in *Proceeding of the 12th Annual Conference on Genetic and Evolutionary Computation Conference*. ACM, 2010, pp. 471–478.
- [25] Y. Jin and B. Sendhoff, "Reducing fitness evaluations using clustering techniques and neural network ensembles," in *Proceeding of the 6th Annual Conference on Genetic and Evolutionary Computation Conference*. Springer, 2004, pp. 688–699.
- [26] D. Lim, Y.-S. Ong, Y. Jin, and B. Sendhoff, "A study on metamodeling techniques, ensembles, and multi-surrogates in evolutionary computation," in *Proceeding of the 9th Annual Conference on Genetic and Evolutionary Computation Conference*. ACM, 2007, pp. 1288–1295.
- [27] D. Lim, Y. Jin, Y.-S. Ong, and B. Sendhoff, "Generalizing surrogate-assisted evolutionary computation," *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 3, pp. 329–355, 2010.
- [28] E. Rigoni and A. Turco, "Metamodels for fast multi-objective optimization: trading off global exploration and local exploitation," in *Simulated Evolution and Learning*. Springer, 2010, pp. 523–532.
- [29] Z. Zhou, Y. S. Ong, P. B. Nair, A. J. Keane, and K. Y. Lum, "Combining global and local surrogate models to accelerate evolutionary optimization," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 1, pp. 66–76, 2007.
- [30] Y. Tenne and S. W. Armfield, "A framework for memetic optimization using variable global and local surrogate models," *Soft Computing*, vol. 13, no. 8–9, pp. 781–793, 2009.
- [31] Z. Zhou, Y. S. Ong, M. H. Nguyen, and D. Lim, "A study on polynomial regression and Gaussian process global surrogate model in hierarchical surrogate-assisted evolutionary algorithm," in *IEEE Congress on Evolutionary Computation*, vol. 3. IEEE, 2005, pp. 2832–2839.
- [32] D. Büche, N. N. Schraudolph, and P. Koumoutsakos, "Accelerating evolutionary algorithms with Gaussian process fitness function models," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 35, no. 2, pp. 183–194, 2005.
- [33] R. Cheng, Y. Jin, K. Narukawa, and B. Sendhoff, "A multiobjective evolutionary algorithm using Gaussian processbased inverse modeling," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 6, pp. 838–856, 2015.
- [34] M. Emmerich, "Single-and multi-objective evolutionary design optimization assisted by Gaussian random field metamodels," Ph.D. dissertation, 2005.
- [35] B. Liu, Q. Zhang, and G. G. Gielen, "A Gaussian process surrogate model assisted evolutionary algorithm for medium scale expensive optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 2, pp. 180–192, 2014.
- [36] J. Lu, B. Li, and Y. Jin, "An evolution strategy assisted by an ensemble of local Gaussian process models," in *Proceeding of the 15th Annual Conference on Genetic and Evolutionary Computation Conference*. ACM, 2013, pp. 447–454.
- [37] L. Willmes, T. Baeck, Y. Jin, and B. Sendhoff, "Comparing neural networks and Kriging in fitness approximation in evolutionary optimization," in *Proceedings of the IEEE Congress on Evolutionary Computation*, vol. 1. IEEE, 2003, pp. 663–670.
- [38] Q. Zhang, W. Liu, E. Tsang, and B. Virginas, "Expensive multiobjective optimization by MOEA/D with Gaussian process model," *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 3, pp. 456–474, 2010.
- [39] I. Loshchilov, M. Schoenauer, and M. Sebag, "Comparison-based optimizers need comparison-based surrogates," in *Parallel Problem Solving from Nature, PPSN XI*. Springer, 2010, pp. 364–373.
- [40] L. Bajer and M. Holeňa, "Surrogate model for continuous and discrete genetic optimization based on RBF networks," in *Intelligent Data Engineering and Automated Learning-IDEAL 2010*. Springer, 2010, pp. 251–258.
- [41] R. G. Regis, "Evolutionary programming for high-dimensional constrained expensive black-box optimization using radial basis functions," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 3, pp. 326–347, 2014.
- [42] S. Zapotecas Martínez and C. A. Coello Coello, "MOEA/D assisted by RBF networks for expensive multi-objective optimization problems," in *Proceeding of the 15th Annual Conference on Genetic and Evolutionary Computation Conference*. ACM, 2013, pp. 1405–1412.
- [43] N. Azzouz, S. Bechikh, and L. Ben Said, "Steady state IBEA assisted by MLP neural networks for expensive multi-objective optimization problems," in *Proceeding of the 16th Annual Conference on Genetic and Evolutionary Computation Conference*. ACM, 2014, pp. 581–588.
- [44] A. E. Brownlee, J. McCall, Q. Zhang *et al.*, "Fitness modeling with Markov networks," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 6, pp. 862–879, 2013.
- [45] A. Gaspar-Cunha and A. Vieira, "A multi-objective evolutionary algorithm using neural networks to approximate fitness evaluations," *International Journal of Computers, Systems and Signals*, vol. 6, no. 1, pp. 18–36, 2005.

- [46] A. Giotis, K. Giannakoglou, and J. Périaux, "A reduced-cost multi-objective optimization method based on the Pareto front technique, neural networks and PVM," in *Proceedings of the ECCOMAS*, 2000.
- [47] Y. Jin, M. Olhofer, and B. Sendhoff, "A framework for evolutionary optimization with approximate fitness functions," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 5, pp. 481–494, 2002.
- [48] —, "On evolutionary optimization with approximate fitness functions," in *Proceeding of the 2nd Annual Conference on Genetic and Evolutionary Computation Conference*, 2000, pp. 786–793.
- [49] W. Gong, A. Zhou, and Z. Cai, "A multioperator search strategy based on cheap surrogate models for evolutionary optimization," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 5, pp. 746–758, 2015.
- [50] L. Gräning, Y. Jin, and B. Sendhoff, "Efficient evolutionary optimization using individual-based evolution control and neural networks: A comparative study," in *European Symposium on Artificial Neural Networks*, 2005, pp. 273–278.
- [51] H. Ulmer, F. Streichert, and A. Zell, "Evolution strategies assisted by Gaussian processes with improved preselection criterion," in *IEEE Congress on Evolutionary Computation*, vol. 1. IEEE, 2003, pp. 692–699.
- [52] J. Branke and C. Schmidt, "Faster convergence by means of fitness estimation," *Soft Computing*, vol. 9, no. 1, pp. 13–20, 2005.
- [53] Y. Jin, M. Hüsken, and B. Sendhoff, "Quality measures for approximate models in evolutionary computation," in *Proceeding of the 5th Annual Conference on Genetic and Evolutionary Computation Conference*, 2003, pp. 170–173.
- [54] M. Hüsken, Y. Jin, and B. Sendhoff, "Structure optimization of neural networks for evolutionary design optimization," *Soft Computing*, vol. 9, no. 1, pp. 21–28, 2005.
- [55] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. Coello Coello, "A survey of multiobjective evolutionary algorithms for data mining: Part I," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 1, pp. 4–19, 2014.
- [56] S. Thomas and Y. Jin, "Reconstructing gene regulatory networks: Where optimization meets big data," *Evolutionary Intelligence*, vol. 7, no. 1, pp. 29–47, 2014.
- [57] S. Wang and X. Yao, "Using class imbalance learning for software defect prediction," *IEEE Transactions on Reliability*, vol. 62, no. 2, pp. 434–443, 2013.
- [58] J. L. Arbuckle, G. A. Marcoulides, and R. E. Schumacker, "Full information estimation in the presence of incomplete data," *Advanced Structural Equation Modeling: Issues and Techniques*, vol. 243, pp. 243–277, 1996.
- [59] Y. Liu, F. Shang, L. Jiao, J. Cheng, and H. Cheng, "Trace norm regularized CANDECOMP/PARAFAC decomposition with missing data," *IEEE Transactions on Cybernetics*, vol. 45, no. 11, pp. 2437–2448, 2015.
- [60] Y. Jin and J. Branke, "Evolutionary optimization in uncertain environments—a survey," *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 3, pp. 303–317, 2005.
- [61] H. Wang, Q. Zhang, L. Jiao, and X. Yao, "Regularity model for noisy multiobjective optimization," *IEEE Transactions on Cybernetics*, vol. PP, no. 99, pp. 1–1, 2015, doi=10.1109/TCYB.2015.2459137.
- [62] Y. Jin, K. Tang, X. Yu, B. Sendhoff, and X. Yao, "A framework for finding robust optimal solutions over time," *Memetic Computing*, vol. 5, no. 1, pp. 3–18, 2013.
- [63] R. R. Chan and S. D. Sudhoff, "An evolutionary computing approach to robust design in the presence of uncertainties," *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 6, pp. 900–912, 2010.
- [64] S. Castano and V. De Antonellis, "Global viewing of heterogeneous data sources," *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 2, pp. 277–297, 2001.
- [65] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97–107, 2014.
- [66] P. Arabie, L. J. Hubert, and G. De Soete, *Clustering and classification*. World Scientific, 1996.
- [67] J. Luo, L. Jiao, and J. Lozano, "A sparse spectral clustering framework via multi-objective evolutionary algorithm," *IEEE Transactions on Evolutionary Computation*, vol. PP, no. 99, pp. 1–1, 2015, doi=10.1109/TEVC.2015.2476359.
- [68] G. H. Duntzman, *Principal components analysis*. Sage, 1989, no. 69.
- [69] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [70] F. Mosteller and J. W. Tukey, "Data analysis and regression: a second course in statistics." Addison-Wesley Series in Behavioral Science: Quantitative Methods, 1977.
- [71] C. Mock, *Guidelines for essential trauma care*. World Health Organization, 2004.
- [72] B. J. Gabbe, G. D. Biostat, P. M. Simpson, A. M. Sutherland, G. Dip, R. Wolfe, M. C. Fitzgerald, R. Judson, and P. A. Cameron, "Improved functional outcomes for major trauma patients in a regionalized, inclusive trauma system," *Annals of Surgery*, vol. 255, no. 6, pp. 1009–1015, 2012.
- [73] E. J. MacKenzie, S. Weir, F. P. Rivara, G. J. Jurkovich, A. B. Nathens, W. Wang, D. O. Scharfstein, and D. S. Salkever, "The value of trauma center care," *Journal of Trauma-Injury, Infection, and Critical Care*, vol. 69, no. 1, pp. 1–10, 2010.
- [74] A. C. of Surgeons Committee on Trauma *et al.*, "Resources for optimal care of the trauma patient," *American College of Surgeons, Chicago, IL*, 2006.
- [75] J. O. Jansen and M. K. Campbell, "The GEOS study: Designing a geospatially optimised trauma system for scotland," *The Surgeon*, vol. 12, no. 2, pp. 61–63, 2014.
- [76] C. A. Macias, M. R. Rosengart, J.-C. Puyana, W. T. Linde-Zwirble, W. Smith, A. B. Peitzman, and D. C. Angus, "The effects of trauma center care, admission volume, and surgical volume on paralysis following traumatic spinal cord injury," *Annals of Surgery*, vol. 249, no. 1, pp. 10–17, 2009.
- [77] N. Mladenović and P. Hansen, "Variable neighborhood search," *Computers & Operations Research*, vol. 24, no. 11, pp. 1097–1100, 1997.
- [78] P. Hansen and N. Mladenović, "Variable neighborhood search: Principles and applications," *European Journal of Operational Research*, vol. 130, no. 3, pp. 449–467, 2001.
- [79] E. K. Burke, A. J. Eckersley, B. McCollum, S. Petrovic, and R. Qu, "Hybrid variable neighbourhood approaches to university exam timetabling," *European Journal of Operational Research*, vol. 206, pp. 46–53, 2010.
- [80] G. W. Milligan and M. C. Cooper, "A study of the comparability of external criteria for hierarchical cluster analysis," *Multivariate Behavioral Research*, vol. 21, no. 4, pp. 441–458, 1986.
- [81] Q. Zhang, A. Zhou, S. Zhao, P. Suganthan, W. Liu, and S. Tiwari, "Multiobjective optimization test instances for the CEC 2009 special session and competition," *University of Essex, Colchester, UK and Nanyang Technological University, Singapore, Special Session on Performance Assessment of Multi-Objective Optimization Algorithms, Technical Report*, pp. 1–30, 2008.
- [82] S. Bandyopadhyay, S. Pal, and B. Aruna, "Multiobjective GAs, quantitative indices, and pattern classification," *IEEE Transactions on Cybernetics*, vol. 34, no. 5, pp. 2088–2099, 2004.
- [83] M. Hollander and D. Wolfe, *Nonparametric statistical methods*. Wiley-Interscience, 1999.
- [84] Z.-H. Zhou, N. V. Chawla, Y. Jin, and G. J. Williams, "Big data opportunities and challenges: discussions from data analytics perspectives," *IEEE Computational Intelligence Magazine*, vol. 9, no. 4, pp. 62–74, 2014.



Handing Wang (S'10-M'16) received the B.Eng. and Ph.D. degrees from Xidian University, Xi'an, China, in 2010 and 2015, respectively.

She is currently a research fellow with the Department of Computer Science, University of Surrey, Guildford, UK.

Dr. Wang is a member of IEEE Computational Intelligence Society. Her research interests include nature-inspired computation, multiobjective optimization, multiple criteria decision making, surrogate-assisted evolutionary optimization,

and real-world problems.

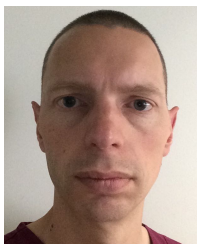


Yaochu Jin (M'98-SM'02-F'16) received the B.Sc., M.Sc., and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 1988, 1991, and 1996 respectively, and the Dr.-Ing. degree from Ruhr University Bochum, Germany, in 2001.

He is a Professor of Computational Intelligence, Department of Computer Science, University of Surrey, Guildford, U.K., where he heads the Nature Inspired Computing and Engineering Group. He is also a Finland Distinguished Professor funded by the Finnish Agency for Innovation (Tekes) and a Changjiang Distinguished Visiting Professor appointed by the Ministry of Education, China. His science-driven research interests lie in the interdisciplinary areas that bridge the gap between computational intelligence, computational neuroscience, and computational systems biology. He is also particularly interested in nature-inspired, real-world driven problem-solving. He has (co)authored over 200 peer-reviewed journal and conference papers and been granted eight patents on evolutionary optimization. His current research is funded by EC FP7, UK EPSRC and industry. He has delivered over 20 invited keynote speeches at international conferences.

He is the Editor-in-Chief of the IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS and Complex & Intelligent Systems. He is also an Associate Editor or Editorial Board Member of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON NANOBIO-SCIENCE, Evolutionary Computation, BioSystems, Soft Computing, and Natural Computing.

Dr Jin was an IEEE Distinguished Lecturer (2013-2015) and Vice President for Technical Activities of the IEEE Computational Intelligence Society (2014-2015). He was the recipient of the Best Paper Award of the 2010 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology and the 2014 IEEE Computational Intelligence Magazine Outstanding Paper Award. He is a Fellow of IEEE.



Jan O. Jansen received a Bachelor of Science degree in Anatomy from University College, London, in 1994, and his medical degree from the United Medical and Dental School of Guy's and St Thomas's, in 1997. He also holds a Diploma in Medical Education, from the University of Dundee, awarded in 2006, and a Doctorate of Medicine, from the University of Berlin, awarded in 2010.

He is a consultant general/trauma surgeon and intensivist, in Aberdeen and London. He is also a Fellow of the Royal College of Surgeons of England (FRCS) and a Fellow of the Faculty of Intensive Care Medicine (FFICM) of the Royal College of Anaesthetists. He is an honorary clinical senior lecturer at the University of Aberdeen.

In 2009, he was awarded a Queen's Commendation for Valuable Service, for his work in Afghanistan. His research interests include the organisation of trauma services, with particular reference to the effects of geography, and trauma resuscitation. He has published more than 70 peer-reviewed articles.