Small Data Driven Evolutionary Multi-objective Optimization of Fused Magnesium Furnaces

Dan Guo, Tianyou Chai, Jinliang Ding
State Key Laboratory of Synthetical Automation for Process Industry
Northeastern University, Shenyang, China, 110004
guodan717@163.com, {tychai, jlding}@mail.neu.edu.cn

Yaochu Jin
Department of Computer Science
University of Surrey, Guildford, Surrey, GU2 7XH, UK
yaochu.jin@surrey.ac.uk

***Abstract*—In most real-world optimization problems, it is very difficult to obtain accurate analytical objective functions derived from process mechanisms. Instead, only approximate objective functions can be built based on sparse historical data. Performance optimization of fused magnesium furnaces is a typical small data-driven optimization problem, where only very limited and noisy data is available. A surrogate-assisted data-driven evolutionary algorithm is proposed in this paper for off-line data-driven optimization of furnaces performance in magnesia production. The multiobjective evolutionary algorithm is assisted by Gaussian process models to search for Pareto optimal solutions. To generate new data samples in surrogate management, a low-order polynomial model is constructed as an approximate mechanism model that can be treated as the real fitness function. To verify the effectiveness of the proposed Gaussian process assisted evolutionary algorithm, it is first tested on nine benchmark problems in comparison with a popular multi-objective evolutionary algorithm and a surrogate-assisted evolutionary algorithm. The algorithm is then applied to a real-world fused magnesium furnaces optimization problem.**

I. Introduction

In many real-world complex optimization problems, it is impractical to build accurate physical or mechanistic models for the objective function, preventing such problems being solved by traditional model-based optimization methods, such as the gradient descent method, Lagrangian relaxation, or dynamic programming. For such optimization problems, data-driven approaches are attracting increasing attention.

Evolutionary algorithms (EAs) have widely been applied to highly nonlinear and uncertain optimization problems in the past decades because of their global search capability by simulating the main mechanisms in natural selection and evolution. EAs have found to be well suited for multi-objective optimization as they are population based search and are considered to be suited for data-driven optimization as they do not require derivative information of the objective functions.

Multi-objective optimization problems in general have the following formulation:

$$\begin{aligned} &\min \ \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_M(\mathbf{x})) \\ &s.t. \quad \mathbf{x} \in D \end{aligned} \tag{1}$$

where $D \subseteq R^N$ is the feasible space of decision variables, and M, N are the dimensions of objectives and decision vectors. Only conflicting objectives are taken into account, and thus there exist multiple Pareto optimal solutions rather than one single global optimum as in single-objective optimization. The Pareto optimal solutions are usually known as Pareto set in decision space and Pareto front in objective space. Numerous multi-objective evolutionary algorithms (MOEAs) have been developed, which roughly fall into three categories, namely, the Pareto dominance based, the decomposition based and the performance indicator based approaches. As a most popular dominance based approach, the elitist nondominated sorting genetic algorithm II (NSGA-II) [1] has been widely employed [2-4].

Most data-driven MOEAs adopt surrogates as a partial substitute for the true objective function. Fitness evaluations (FEs) using the original objective function are also known as expensive FEs because they cost plenty of computational resources or can be obtained only by experiment [5-8]. Many machine learning techniques can be made use of to build surrogates [9-12]. Gaussian process model (GPM) that can predict the mean value and a confidence interval is considered as one of the most effective surrogate methods. Taking both the output optimality and model accuracy into account, several metrics tailored for GPMs have been proposed, such as the expected improvement, probability of improvement, and lower confidence bound [13-15]. Theses metrics have been demonstrated to be able to assist the evolutionary algorithm to quickly find optimal solutions using a limited number of expensive FEs.

Data-driven surrogate-assisted MOEAs can be roughly divided into two types [16]. In the first type, MOEAs use data-driven surrogates and expensive FE methods simultaneously. As suggested in [17], three main methods for combining surrogates and expensive FEs have been proposed, which are known as individual-based, generation-based and population-based. It is also pointed out that a surrogate-assisted EA may converge to a false optimum introduced by the surrogate, therefore it is necessary to combine the surrogate and expensive FEs properly, which is called evolution control or model management [18]. Usually, after new expensive FEs are conducted, the surrogates will be updated to improve its approximation accuracy, which is widely adopted in surrogate-assisted evolutionary optimization [7], [8], [13]. These approaches are called on-line data-driven surrogate-assisted evolutionary optimization [16].

The second type of data-driven optimization can use sur-

rogates only in optimization phase due to the infeasibility of performing new expensive FEs during the optimization. This type of data-driven optimization is known as off-line data-driven optimization [16], which will be extremely challenging in case only limited historical data is available.

The performance optimization of fused magnesium furnaces represents a typical off-line data-driven problem. This optimization is particularly challenging for long-cycle magnesia production since only small data is available for optimization. To the best of our knowledge, no dedicated algorithms have been proposed for off-line data-driven optimization with limited data. To address such problems, this paper suggests to build low-order surrogate models based on the sparse and noisy historical data, which will be less vulnerable to curse of dimension and over-fitting. Then a data-driven surrogate-assisted evolutionary algorithm is used, in which new samples can be generated from the low-order surrogates that replace expensive FEs that are not available in off-line data-driven optimization.

The rest of the paper is organized as follows. Section II introduces the fundamentals for the furnaces performance optimization problem. Details of the surrogate-assisted evolutionary multi-objective algorithm are presented in Section III, including NSGA-II and GPM so that the paper is self-contained. Numerical experiments are conducted and results are presented in Section IV. Section V presents the real-world furnaces optimization problem using the proposed approach. Section VI summarizes the paper.

II. Furnaces Performance Optimization Problem

Magnesia is a kind of highly purified MgO crystal, which has high melting points, oxidation resistance, structural integrity, and good insulation. Magnesia is widely used in metallurgy, chemical industry and construction. In the submerged arc furnace, caustic calcined magnesia will be heated to above 2800 degrees Celsius, and final magnesia product can be obtained after natural cooling and crystallization of melt [19], [20].

The furnaces manufacture process is illustrated in Fig. 1. The whole process takes about 10 hours, which can be roughly divided into three stages, namely, starting, smelting and ending. In the smelting stage, certain amount of caustic calcined magnesia powder will be added into furnaces every 8-12 minutes and each furnace makes a frequent transformation within different operation conditions. Many electric arcs will be developed among three-phase electrodes and furnace burden, and electric energy is continuously converted into thermal energy. A molten pool is formed and the pool surface is gradually raised along with charging and fusion of burden. When the surface approaches to furnace mouth, power supply is cut off and one production batch is over.

Nowadays human users heuristically set the target value of electricity consumption for a ton of magnesia (ECT), then the operational control and optimization system decides the positive-negative rotating of motors to adjust the electrode height on the basis of real current, ECT, material ingredient and granularity, which is beneficial to stabilize the current [21]. The overall performance indices of furnaces are total output, high-quality rate and electricity consumption, whose optimization can be realized by suitable ECT setpoint of every furnace. There are a few reasons why the relationship between ECT and performance indices cannot be made clear by theoretical analysis. Above all, a blend of solid, liquid and gas in the high temperature leads to the failure of various sensors. In addition, magnesia production is composed of many physical and chemical processes such as fusing, impurity precipitation and crystallization. Finally dynamics of control loops are also complex [22], [23]. Thus historical production data is applied to build function models, and the performance optimization of furnaces is an off-line problem.

The crystal whose MgO content is more than 98% is regarded as of high-quality. The goal of furnaces performance optimization is to reduce the total energy consumption E while increasing the total output Y and maintaining a high-quality rate Q. If the ECT set point and performance indices of the i-th furnace are denoted as r_i, y_i, q_i, and e_i, and there are N furnaces in total, the optimization problem can be formulated as follows:

$$\begin{aligned} &\min -Y, \min -Q, \min E \\ &s.t. \\ &Y = \sum_{i=1}^{N} y_i, Q = \frac{\sum_{i=1}^{N} y_i \times q_i}{\sum_{i=1}^{N} y_i}, E = \sum_{i=1}^{N} e_i, \\ &y_i = h_{1,i}(r_i), q_i = h_{2,i}(r_i), e_i = h_{3,i}(r_i), \\ &r_{i,min} \leq r_i \leq r_{i,max}, i = 1, 2, ..., N. \end{aligned} \tag{2}$$

where $h_{1,i}$, $h_{2,i}$ and $h_{3,i}$ denote the relationship between ECT and each performance index of the i-th furnace, respectively. Note that a prior information that the performance indices of the i-th furnace have nothing to do with ECT set points of the other furnaces is applied to the above equation.

III. Proposed Algorithm

As mentioned in Introduction, this work aims to propose a surrogate-assisted data-driven optimization method where no analytic objective functions are available for accurate calculation of the objective functions and only a limited amount of historical data is available. This type of data-driven optimization problems are extremely challenging, as it is very hard to perform effective model management without new accurate fitness evaluations (FEs). To address this difficulty, one way is to derive simplified analytical models for calculating the objective functions so that new samples can be obtained for surrogate management. In case where such simplified analytical models are unavailable, like the furnaces performance optimization problem to be solved in this work, we propose to build a low-order model based on the historical production data for model management.

This work adopts the elitist non-dominated sorting genetic algorithm (NSGA-II) [1] as the basic evolutionary optimization algorithm and the Gaussian process model (GPM) as the surrogate model. A brief description of the key component

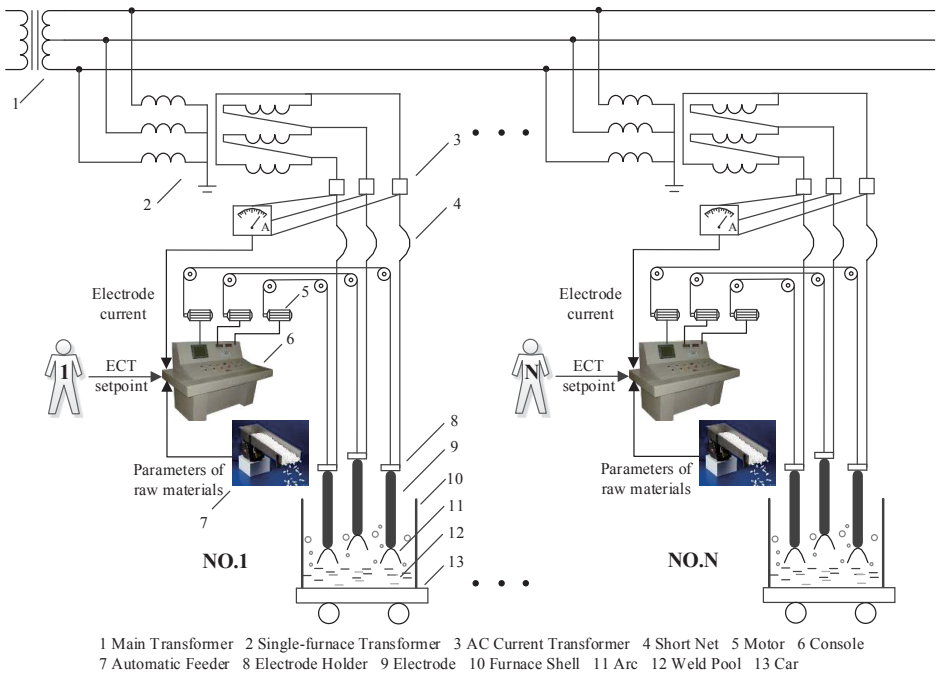

Fig. 1. Diagram of the production process of fused magnesium furnaces.

of NSGA-II is shown in Algorithm 1. The main steps for GPM are listed in Algorithm 2. Note that matrix inversion operations are needed in GPM, where the matrix dimension is equal to the size of the training set. Therefore, use of Gaussian processes for surrogate-assisted evolutionary optimization is very computationally intensive, in particular when the number of training data is large and the GPM needs to be updated in each generation.

Algorithm 1 The NSGA-II Algorithm for Multi-objective Optimization

Input: Population size N_{pop} and maximum iterations T;
Output: Final population P_T;
1: Randomly initialize a population P_0 of size N_{pop};
2: Assign fitness for P_0 based on nondominated sorting and crowding distance;
3: **for** $t = 1$ to T **do**
4: Generate parent population from P_t by tournament selection;
5: Simulated binary crossover and polynomial mutation operate on parent to generate offspring population;
6: Intermediate population is formed by the combination of parent and offspring, and then fitness is assigned;
7: Individuals of P_{t+1} are selected from intermediate population;
8: **end for**

A Gaussian process assisted NSGA-II, termed NSGA-II_GP for short, is proposed in this work, where a low-order model constructed using small noisy data is adopted to replace the accurate fitness evaluation in surrogate management. The details

Algorithm 2 Gaussian Process Modelling

Input: Training data in parameter space $X = [\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^{N1}]^T$ and in objective space $F = [f^1, f^2, ..., f^{N1}]^T$;
A new untested point $\mathbf{x}^{new}$;
1: Choose a correlation function c which reflects the relationship between objective errors through parameter distance;
2: X is used to calculate correlation matrix,whose element is $c(\mathbf{x}^{n_1}, \mathbf{x}^{n_2})$, $n_1, n_2 = 1, ..., N1$;
3: Hyperparameters in function c are estimated by maximizing the likelihood of the sample;
4: A new correlation matrix between $\mathbf{x}^{new}$ and each of X is calculated based on the identified hyperparameters;
5: Predicted mean and variance value $\hat{f}(\mathbf{x}^{new})$ and $s^2(\mathbf{x}^{new})$ are output.

of the proposed algorithm are described in the following.

A. Gaussian Process Assisted NSGA-II

$$EI_j(\mathbf{x}) = (f_{jmin} - \hat{f}_j(\mathbf{x}))\Phi(\frac{f_{jmin} - \hat{f}_j(\mathbf{x})}{s_j^2(\mathbf{x})}) + s_j^2(\mathbf{x})\varphi(\frac{f_{jmin} - \hat{f}_j(\mathbf{x})}{s_j^2(\mathbf{x})}) \quad (3)$$

Expected improvement (EI) [13] considers both optimal and most uncertain solutions predicted by the surrogate as the most promising ones. Suppose the objective value for the training data used in the j-th ($j = 1, ..., M$) GPM is F_j, of which the minimum is f_{jmin}. $\hat{f}_j(\mathbf{x})$ and $s_j^2(\mathbf{x})$ are the predicted mean and variance of j-th GPM at point $\mathbf{x}$, then the corresponding

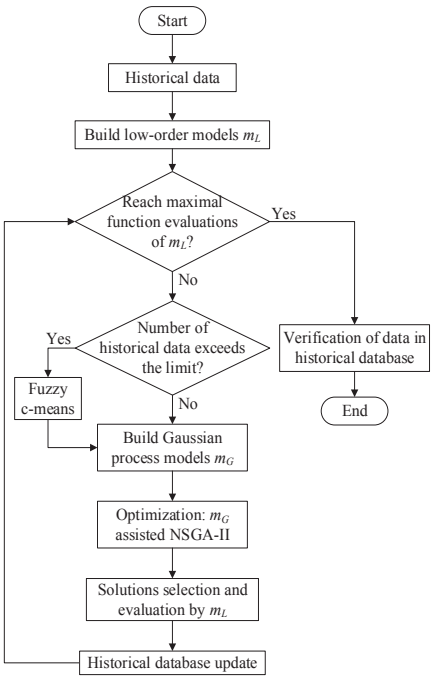

Fig. 2. Flowchart of NSGA-II_GP.

EI value can be calculated as expression (3), where Φ and φ represent standard normal distribution and probability density function, respectively.

NSGA-II_GP iterates to optimize EIs. Fig. 2 plots the flow chart of the algorithm and its main steps are as follows:

Step 1 Limited and noisy historical data is collected.

Step 2 Low-order models building. Second-order polynomial models m_L are built based on the historical data, and these models are applied to approximate the accurate FEs later.

Step 3 Stopping criterion and solutions verification. If the maximal FEs of m_L have been reached, solutions in the historical data set are verified, otherwise go on to Step 4.

Step 4 Data screening. Fuzzy c-means clustering is applied to screen L data from the historical set. If the number of historical data is less than the limit L, all are used to build GPM.

Step 5 GPMs building. A GPM m_G for each objective is built using the selected data, and the corresponding EI metrics are defined.

Step 6 Optimization. Use NSGA-II assisted by GPMs to find solutions maximizing EIs of all objectives.

Step 7 Solutions selection and evaluation. K are chosen from solutions in the last step by k-means clustering, then these solutions are re-evaluated by low-order models m_L built in Step 2.

Step 8 Update of historical data set. The K solutions are combined with original historical data. Go to Step 3 to a new round of optimization.

NSGA-II_GP will be applied to the real-world furnaces performance optimization problem, after its effectiveness is verified on standard test problems. It is important to note that there are different treatments in the first three steps of the algorithm when it is applied to two kinds of problems above. First of all, decision variables of historical data in standard tests are generated by Latin hypercube sampling, then real fitness functions are used to evaluate these points, and stochastic noise is added to the function values to generate outputs of the historical data. Since data collected from the actual magnesia production process are originally noisy, they can be directly applied to build the low-order models, but the decision variables of these data are not distributed as uniformly as those of tests. In the next place, a low-order model is built for each objective in standard tests. However, there are $M \times N$ low-order models in the furnaces problem, where M and N represent the number of overall performance indices and furnaces, respectively. The reason is that the performance indices of a furnace are only related to ECT setpoint of itself, and this prior information made use of in low-order models building is partly beneficial to accurate approximation of unknown analytic functions of furnaces performance. Eventually, solutions verification in standard tests is firstly re-evaluating all data in historical set by real functions, then the non-dominated solutions are regarded as Pareto front approximation. While the non-dominated performance indices of furnaces in the historical data set are directly output as optimal solutions as real furnaces performance functions are unavailable.

Some details of NSGA-II_GP are described following. In Step 1 the stochastic noise added to a j-th ($j = 1, ..., M$) function value is generated according to

$$noise = (f_{jmax} - f_{jmin}) \times rand \tag{4}$$

where f_{jmax} and f_{jmin} are the maximum and minimum of real function values in the j-th objective respectively, and $rand$ is a random number in the domain of [-0.1,0.1]. Parameters of second-order models in Step 2 can be identified by multiple linear regression. Limit L in Step 4 is to restict the time taken to build GPM as the evolution proceeds, and the details of fuzzy c-means clustering can be found in [8]. To ensure the diversity of the training data, solutions different from the training data are chosen with priority in Step 7. A threshold based on the Euclidean distance is defined to indicate how different two solutions are in the decision space, which is set to 10^{-5} in this work. Then the solutions are further screened by the k-means clustering in the search space. It should be made clear that GPMs building is based on noisy data and the data re-evaluated by low-order models, so GPMs do not entirely approximate the low-order models. The number of GPM in all problems is equal to the number of problem objectives.

IV. Numerical Experiments

In this section, the performance of NSGA-II_GP is compared with that of the original NSGA-II as well as ParEGO, one popular surrogate-assisted multi-objective evolutionary algorithm [7]. Benchmark problems selected from three widely used test suites, ZDT [24], DTLZ [25] and WFG [26] are used for comparison, which are illustrated in Table I. Three different cases for all test functions are considered, where the number of decision variables is set to 5(6), 10, and 16, respectively. Note

TABLE I
BENCHMARK PROBLEMS

Problems	Objective number M	Variable dimension N	The range of $\mathbf{x}$ or $\mathbf{z}$	Objective functions	Optima	Characters
DTLZ1	3	5,10,16	[0,1]	$f_1(\mathbf{x}) = 0.5x_1x_2(1+g(\mathbf{x}))$ $f_2(\mathbf{x}) = 0.5x_1(1-x_2)(1+g(\mathbf{x}))$ $f_3(\mathbf{x}) = 0.5(1-x_1)(1+g(\mathbf{x}))$ $g(\mathbf{x}) = 100[N-M+1+$ $\sum_{i=3}^{N}(x_i-0.5)^2 - \cos(20\pi(x_i-0.5))]$	$x_{1,2} \in [0,1]$, $x_i = 0.5$, $i = 3,...,N$.	multifrontal
DTLZ4	3	5,10,16	[0,1]	$f_1(\mathbf{x}) = (1+g(\mathbf{x}))\cos(x_1^\alpha\pi/2)\cos(x_2^\alpha\pi/2)$ $f_2(\mathbf{x}) = (1+g(\mathbf{x}))\cos(x_1^\alpha\pi/2)\sin(x_2^\alpha\pi/2)$ $f_3(\mathbf{x}) = (1+g(\mathbf{x}))\sin(x_1^\alpha\pi/2)$ $g(\mathbf{x}) = \sum_{i=3}^{N}(x_i-0.5)^2$	$x_{1,2} \in [0,1]$, $x_i = 0.5$, $i = 3,...,N$.	biased
DTLZ5	3	5,10,16	[0,1]	$f_1(\mathbf{x}) = (1+g(\mathbf{x}))\cos(\theta_1\pi/2)\cos(\theta_2\pi/2)$ $f_2(\mathbf{x}) = (1+g(\mathbf{x}))\cos(\theta_1\pi/2)\sin(\theta_2\pi/2)$ $f_3(\mathbf{x}) = (1+g(\mathbf{x}))\sin(\theta_1\pi/2)$ $\theta_i = \frac{1+2g(\mathbf{x})x_i}{2(1+g(\mathbf{x}))}, i = 2,...,M-1$, the rest $\theta_i = x_i$ $g(\mathbf{x}) = \sum_{i=3}^{N}(x_i-0.5)^2$	$x_{1,2} \in [0,1]$, $x_i = 0.5$, $i = 3,...,N$.	degenerate
ZDT3	2	5,10,16	[0,1]	$f_1(\mathbf{x}) = x_1$ $f_2(\mathbf{x}) = g(\mathbf{x})[1-\sqrt{\frac{x_1}{g(\mathbf{x})}} - \frac{x_1}{g(\mathbf{x})}\sin(10\pi x_1)]$ $g(\mathbf{x}) = 1+9(\sum_{i=2}^{N}x_i)/(N-1)$	$x_1 \in [0,1]$, $x_i = 0$, $i = 2,...,N$.	disconnected
ZDT6	2	5,10,16	[0,1]	$f_1(\mathbf{x}) = 1-exp(-4x_1)\sin^6(6\pi x_1)$ $f_2(\mathbf{x}) = g(\mathbf{x})[1-(f_1(\mathbf{x})/g(\mathbf{x}))^2]$ $g(\mathbf{x}) = 1+9[(\sum_{i=2}^{N}x_i)/(N-1)]^{0.25}$	$x_1 \in [0,1]$, $x_i = 0$, $i = 2,...,N$.	biased
WFG2 WFG3 WFG4 WFG5	3	6,10,16 and $k = 4$, $l = 2,6,12$	$[0,2i]$, $i = 1,...,N$	Details see [26]	$z_p \in [0,2p]$, $p = 1,...,k$. $z_q = 0.7q$, $q = k+1,...,N$.	1 disconnected non-separable 2 degenerate non-separable 3 multimodal 4 deceptive

that GPM is implemented in MATLAB toolbox downloadable from http://www.gaussianprocess.org/.

A. Experimental settings

The hypervolume (HV) [27] is adopted as the performance indicator, which can capture both diversity and convergence of solutions. A higher HV indicates better quality of the solution set. The reference point is determined by equation $b_j = max_j + \delta(max_j - min_j)$, $j = 1,...,M$ [16], where max_j and min_j are the maximization and minimization of all results to be assessed in the j-th objective, and δ takes 0.01 here. In addition, all HV values will be normalized by dividing $\prod_{j=1}^{M} b_j$.A detailed description of the experimental settings is given below:

1) Ten independent runs are performed for each experiment. The Wilcoxon rank sum is introduced to compare the results obtained by the proposed algorithm and the compared algorithms at a confidence level of 5%. Symbol '+' indicates that the proposed algorithm is much better than the compared one, while '−' indicates the compared algorithm is significantly better, and '=' means there is no notable difference between the results obtained by the two algorithms.
2) In each run, initial historical data are newly generated, and the number of these data is set to $11N-1$, where N is the dimension of decision space. The initial data are the same for all algorithms in comparative experiments.
3) In NSGA-II_GP, the upper limit of historical data number L are 80, 130 and 200 for N = 5(6), 10, and 16, respectively.
4) In NSGA-II_GP and ParEGO, the population size and the iteration number are both set to 50, and other parameters in the algorithm are the same as in [1], [7]. The population number of NSGA-II is also 50.
5) Taking the computation time into account, we set the maximal FEs of low-order polynomials to 250.
6) In each generation, the number of re-evaluated solutions K is 5.

B. Comparative experiments among NSGA-II_GP, ParEGO and NSGA-II

Solutions obtained by three algorithms need to be re-evaluated by real fitness function of benchmark problems,

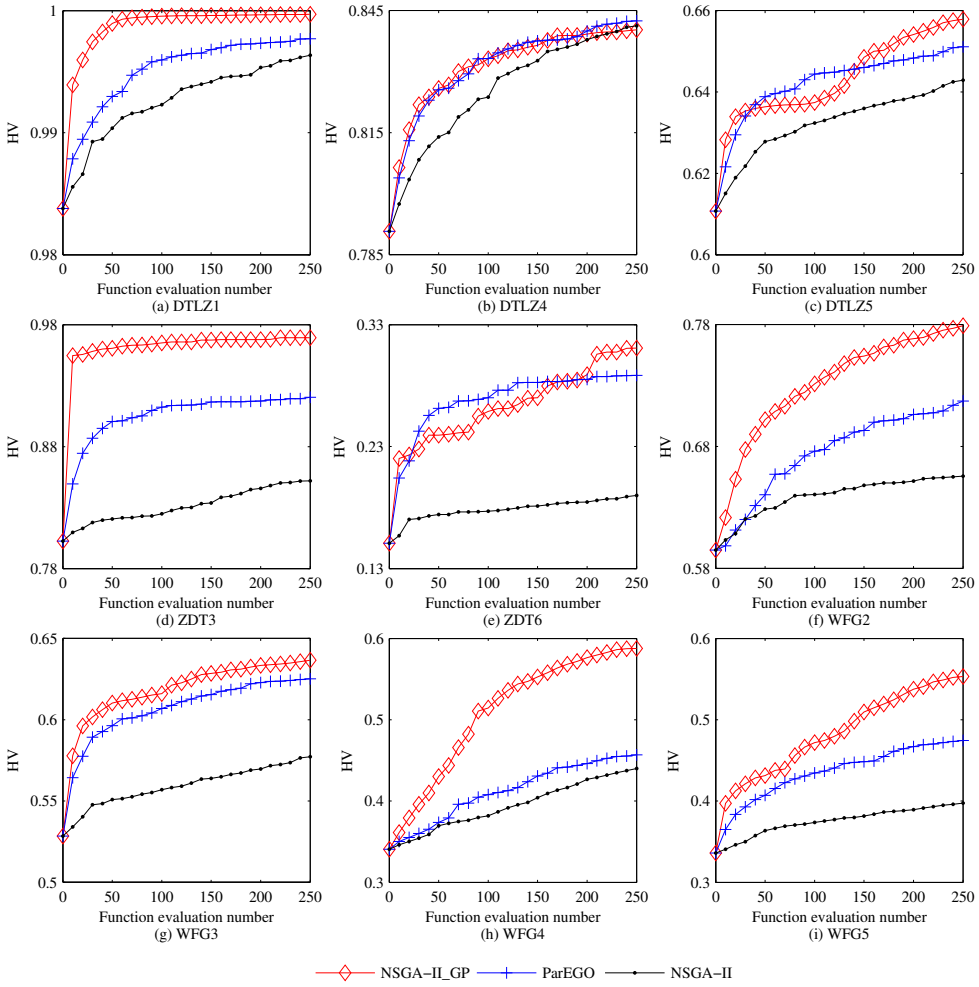

Fig. 3. Mean value of HV calculated by real fitness function versus the function evaluation number of the low-order models when dimension of decision variables is 5 or 6.

then the non-dominated solutions are chosen to calculate HV. Assessment results for all cases are listed in Table II, where the best results are highlighted. It is plotted in Fig. 3 that the change of mean HV over the accumulated FE number of the low-order models in the case of variable dimension N = 5(6).

In every subfigure of Fig.3, the starting point is the mean HV of ten groups of initial historical data. It can be seen that HV of all three algorithms constantly increase as the evolution proceeds although real fitness functions are not known in the optimization, indicating that the low-order surrogate models are able to accelerate the convergence. NSGA-II_GP and ParEGO have clear advantages over NSGA-II, and this should be attributed to GPMs, which are built based on both the initial noisy data and the data sampled from low-order models.

The results in Table II demonstrate that the performance difference of these algorithms is reduced as the dimension of the decision space increases. In two biased problems, DTLZ4 and ZDT6, NSGA-II_GP and ParEGO have almost equal performance. Overall, we can see that NSGA-II assisted by the Gaussian process based surrogate significantly outperforms NSGA-II and ParEGO on almost all test problems for the considered search dimensions.

V. Performance Optimization of Furnaces

The furnaces performance optimization problem considered in this work is based on data collected from a factory in Liaoning Province, China. The fused magnesium furnaces have a designed production capacity of 30 ton. In this application example, five furnaces are connected to one transformer, that is, N in Eq. (2) equals 5.

Since only one group of performance indices can be measured in one day, data collection for this optimization problem is very time-consuming. In this work, we gathered 60 groups of performance indices for the five furnaces, including output, high-quality rate, energy consumption, and ECT. The data of one furnace is illustrated in Fig. 4 (solid lines), from which we can see that industrial data is very noisy. Such noisy data can easily be over-fitted if a highly complex model is adopted. For this reason, we employ low order polynomials to fit the production data and the results are shown by the dashed lines in Fig. 4. All resulting approximate models are presented

TABLE II
MEAN (STANDARD DEVIATION) VALUES OF HV (CALCULATED BY REAL FUNCTIONS) OBTAINED BY NSGA-II_GP, PAREGO AND NSGA-II

DTLZ1	$N = 5$	$N = 10$	$N = 16$
NSGA-II_GP	**0.9997**(0.0001)	**0.9967**(0.0012)	**0.9953**(0.0008)
ParEGO	0.9977(0.0006)$^+$	0.9827(0.0025)$^+$	0.9854(0.0019)$^+$
NSGA-II	0.9964(0.0014)$^+$	0.9753(0.0026)$^+$	0.9803(0.0031)$^+$
DTLZ4	$N = 5$	$N = 10$	$N = 16$
NSGA-II_GP	0.8403(0.0091)	0.9314(0.0079)	0.9270(0.0041)
ParEGO	**0.8425**(0.0059)$^=$	**0.9328**(0.0059)$^=$	0.9318(0.0030)$^-$
NSGA-II	0.8413(0.0079)$^=$	0.9306(0.0083)$^=$	**0.9334**(0.0047)$^-$
DTLZ5	$N = 5$	$N = 10$	$N = 16$
NSGA-II_GP	**0.6579**(0.0044)	**0.8242**(0.0105)	0.8090(0.0038)
ParEGO	0.6511(0.0051)$^+$	0.8096(0.0079)$^+$	**0.8148**(0.0030)$^-$
NSGA-II	0.6429(0.0094)$^+$	0.7999(0.0110)$^+$	0.8098(0.0052)$^=$
ZDT3	$N = 5$	$N = 10$	$N = 16$
NSGA-II_GP	**0.9693**(0.0100)	**0.9092**(0.0278)	**0.8587**(0.0172)
ParEGO	0.9204(0.0116)$^+$	0.8288(0.0171)$^+$	0.7987(0.0167)$^+$
NSGA-II	0.8519(0.0189)$^+$	0.7608(0.0158)$^+$	0.7459(0.0192)$^+$
ZDT6	$N = 5$	$N = 10$	$N = 16$
NSGA-II_GP	**0.3107**(0.0996)	**0.2386**(0.0784)	**0.2015**(0.0388)
ParEGO	0.2885(0.0255)$^=$	0.2035(0.0141)$^=$	0.1917(0.0133)$^=$
NSGA-II	0.1900(0.0221)$^+$	0.1455(0.0116)$^+$	0.1491(0.0134)$^+$
WFG2	$N = 6$	$N = 10$	$N = 16$
NSGA-II_GP	**0.7792**(0.0389)	**0.7426**(0.0351)	0.6743(0.0354)
ParEGO	0.7173(0.0334)$^+$	0.6949(0.0228)$^+$	0.6759(0.0358)$^=$
NSGA-II	0.6557(0.0261)$^+$	0.6935(0.0482)$^+$	**0.6798**(0.0265)$^=$
WFG3	$N = 6$	$N = 10$	$N = 16$
NSGA-II_GP	**0.6366**(0.0099)	**0.6379**(0.0128)	**0.5993**(0.0129)
ParEGO	0.6251(0.0076)$^+$	0.6258(0.0084)$^+$	0.5898(0.0160)$^=$
NSGA-II	0.5773(0.0094)$^+$	0.5774(0.0191)$^+$	0.5790(0.0208)$^+$
WFG4	$N = 6$	$N = 10$	$N = 16$
NSGA-II_GP	**0.5877**(0.0110)	**0.4196**(0.0120)	0.3756(0.0161)
ParEGO	0.4569(0.0163)$^+$	0.3689(0.0200)$^+$	0.3796(0.0163)$^=$
NSGA-II	0.4401(0.0469)$^+$	0.3871(0.0147)$^+$	**0.3918**(0.0221)$^=$
WFG5	$N = 6$	$N = 10$	$N = 16$
NSGA-II_GP	**0.5533**(0.0281)	**0.4305**(0.0157)	**0.4260**(0.0443)
ParEGO	0.4745(0.0181)$^+$	0.3705(0.0126)$^+$	0.3589(0.0108)$^+$
NSGA-II	0.3974(0.0244)$^+$	0.3102(0.0117)$^+$	0.3460(0.0114)$^+$

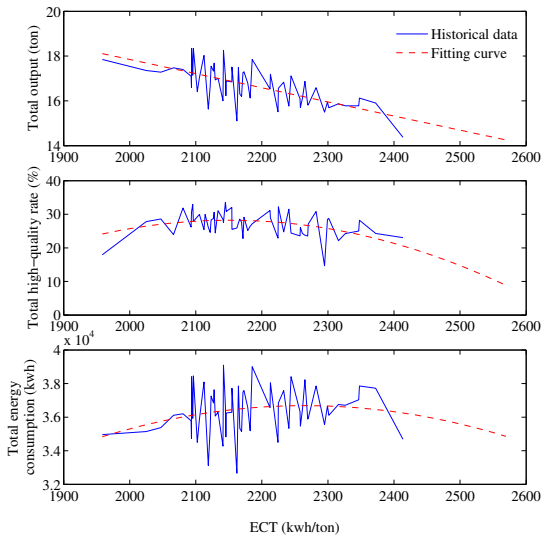

Fig. 4. An example for fitting the performance indices of one furnace.

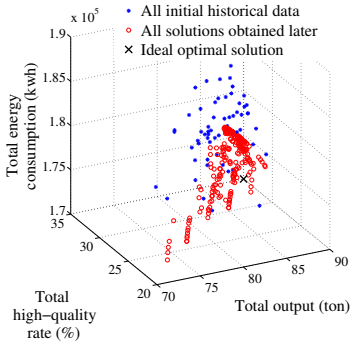

Fig. 5. Optimization results on the furnaces performance problem.

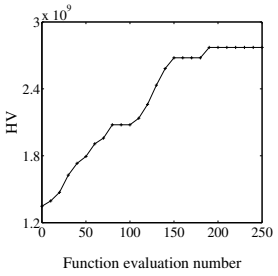

Fig. 6. The mean HV value over the function evaluation number of the low-order models in the performance optimization of furnaces.

in Table III. Worker intervention and equipment depreciation lead to the difference among the approximate performance functions of furnaces in Table III, so final parameters of low-order models takes the mean value of those of all furnaces equations. The range of ECT is determined by the maximum and minimum of collected data. As a result, the optimization problem in Eq. (2) can be re-formulated as follows:

$$\begin{aligned}
&\min -Y, \min -Q, \min E\\
&s.t.\\
&Y = \sum_{i=1}^{5} y_i, Q = \frac{\sum_{i=1}^{5} y_i \times q_i}{\sum_{i=1}^{5} y_i}, E = \sum_{i=1}^{5} e_i,\\
&y_i = -5942.1 r_i + 29521,\\
&q_i = -40.7548 r_i^2 + 170.1776 r_i - 148.1144,\\
&e_i = -17353 r_i^2 + 78201 r_i - 51871,\\
&1958.5 \le r_i \le 2569.5, i = 1, 2, ..., 5.
\end{aligned} \tag{5}$$

Simulation settings are the same with experiments on the test problems. In this application example, the low-order polynomial models built from the data are used as the real fitness functions. Fig. 5 plots all performance indices, where a black '$\times$' indicates the ideal optimal solution. Note that although not all production indices resulting from the optimization are better than historical data, the HV mean value in Fig. 6 averaged over 10 runs increases constantly.

VI. CONCLUSIONS

This paper proposes a data-driven optimization method for the off-line problems, which are very commonly seen

TABLE III
LOW-ORDER MODELS OF FIVE FURNACES

Furnace number	Function of output	Function of high-quality rate	Function of energy consumption
1	$y_1 = -7234.8r_1 + 32185$	$q_1 = -9.4773r_1^2 + 29.0256r_1 + 10.6462$	$e_1 = 5378.5r_1^2 - 23704r_1 + 61734$
2	$y_2 = -2891.0r_2 + 22761$	$q_2 = -39.0944r_2^2 + 161.5855r_2 - 137.7960$	$e_2 = -55697r_2^2 + 248230r_2 - 240420$
3	$y_3 = -6312.0r_3 + 30471$	$q_3 = -110.6837r_3^2 + 476.1242r_3 - 483.8129$	$e_3 = -19726r_3^2 + 89357r_3 - 64514$
4	$y_4 = -7126.7r_4 + 32168$	$q_4 = -74.7296r_4^2 + 321.4685r_4 - 314.8097$	$e_4 = -7664.1r_4^2 + 34548r_4 - 2628.8$
5	$y_5 = -6146.1r_5 + 30019$	$q_5 = 30.2109r_5^2 - 137.3158r_5 + 185.2007$	$e_5 = -9056.1r_5^2 + 42577r_5 - 13526$

in process industry. Since no accurate analytical models are available for the objective functions, low-order polynomial models are built using the limited and highly noisy data, which will be treated as the real fitness functions in data-driven surrogate-assisted evolutionary optimization. A multi-objective evolutionary algorithm assisted by a Gaussian process is proposed to solve a performance optimization problem of fused magnesium furnaces, where the low-order polynomials are used for surrogate management. With a limited computation budget, the effectiveness of the proposed algorithm is verified on nine standard tests in different search dimensions.

The use of low-order polynomials to replace the real objective function is the first step towards solving off-line data-driven optimization problems with very limited data. Our next step will be to improve the validation model for performance evaluation in the furnaces optimization problem by combining parameterized physical model with a data-driven approach. In addition, it might be necessary to adapt the number of solutions to be re-evaluated during the optimization.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China Projects under Grant 61525302 and Grant 61590922, and in part by the Projects of Liaoning Province under Grant 2014020021 and Grant LR2015021.

REFERENCES

[1] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Trans. Evol. Comput., vol. 6, no. 2, pp. 182-197, April 2002.

[2] M. Abido, "Multiobjective evolutionary algorithms for electric power dispatch problem," IEEE Trans. Evol. Comput., vol. 10, no. 3, pp. 315-329, Jun. 2006.

[3] G. Yu, T. Chai, and X. Luo, "Multiobjective production planning optimization using hybrid evolutionary algorithms for mineral processing," IEEE Trans. Evol. Comput., vol. 15, no. 4, pp. 487-514, Aug. 2011.

[4] R. B. Kasat and S. K. Gupta, "Multi-objective optimization of an industrial fluidized-bed catalytic cracking unit (FCCU) using genetic algorithm (GA) with the jumping genes operator," Comput. Chem. Eng., vol. 27, no. 12, pp. 1785-1800, Dec. 2003.

[5] D. Horn, T. Wagner, D. Biermann, C. Weihs, and B. Bischl, "Model-based multi-objective optimization: taxonomy, multi-point proposal, toolbox and benchmark," in Proc. 8th International Conference on EMO, Portugal, pp. 64-78, 2015.

[6] D. Lim, Y. Jin, Y. S. Ong, and B. Sendhoff, "Generalizing surrogate-assisted evolutionary computation," IEEE Trans. Evol. Comput., vol.14, no.3, pp. 329-355, Jun. 2010.

[7] J. Knowles, "ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," IEEE Trans. Evol. Comput., vol. 10, no. 1, pp. 5066, Feb. 2006.

[8] Q. Zhang, W. Liu, E. Tsang, and B. Virginas, "Expensive multiobjective optimization by MOEA/D with Gaussian Process Model," IEEE Trans. Evol. Comput., vol. 14, no. 3, pp. 456-474, June 2010.

[9] C. Sun, Y. Jin, J. Zeng, and Y. Yu, "A two-layer surrogate-assisted particle swarm optimization algorithm," Soft Computing, vol.19, no.6, pp. 1461-1475, June 2015.

[10] Y. Jin and B. Sendhoff, "Reducing fitness evaluations using clustering techniques and neural network ensembles," in Proc. GECCO, pp. 688699, 2004.

[11] L. Pavelski, M. Delgado, and C. Almeida, "Extreme learning surrogate models in multi-objective optimization based on decomposition," NEUROCOMPUTING, vol. 180, pp. 55-67, March 2016.

[12] D. Huang, T. T. Allen, W. I. Notz and R. A. Miller, "Sequential kriging optimization using multiple-fidelity evaluations," Struct. Multidisc. Optim., vol. 32, no. 5, pp. 369-382, Nov. 2006.

[13] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," J. Global Optim., vol. 13, no. 4, pp. 455492, Dec. 1998.

[14] M. Emmerich, K. Giannakoglou, and B. Naujoks, "Single- and multi-objective evolutionary optimization assisted by gaussian random field metamodels," IEEE Trans. Evol. Comput., vol. 10, no. 4, pp. 421-439, Aug. 2006.

[15] N. Beume, B. Naujoks, and M. Emmerich, "SMS-EMOA: multiobjective selection based on dominated hypervolume," European Journal of Operational Research, vol. 181, no. 3, pp. 1653-1669, Sep. 2007.

[16] H. Wang, Y. Jin, and J. Jansen, "Data-driven surrogate-assisted multi-objective evolutionary optimization of a trauma system," IEEE Trans. Evol. Comput., 2016 (accepted).

[17] Y. Jin, "Surrogate-assisted evolutionary computation: recent advances and future challenges," Swarm and Evolutionary Computation, vol. 1, no. 2, pp. 61-70, June 2011.

[18] Y. Jin, "A comprehensive survey of fitness approximation in evolutionary computation," Soft Comput. vol. 9, no. 1, pp. 3-12, Jan. 2005.

[19] L. Gu, Special Refractory Matter (in Chinese). Beijing, Metallurgical Industry Press, 2000.

[20] Q. Hu, Manufacture and Application of Magnesium Compound (in Chinese). Beijing, Chemical Industry Press, 2004.

[21] Y. J. Wu, L. Zhang, H. Yue and T. Y. Chai, "Intelligent optimal control based on CBR for fused magnesia production," Journal of Chemical Industry and Engineering (China), vol. 59, no. 7, pp. 1686-1690, 2008.

[22] Z. W. Wu, T. Y. Chai, J. Fu and J. Sun, "Hybrid intelligent optimal control of fused magnesium furnace," In: 49th IEEE Conference on Decision and Control, Atlanta, USA, pp. 3313-3318, Dec. 2010.

[23] Z. W. Wu, T. Y. Chai, J. Fu and Z. W. Yan, "Intelligent setpoints control of smelting process of fused magnesium furnace," Control and Decision (China), vol. 26, no. 9, pp. 1417-1420, 2011.

[24] E. Zitzler, K. Deb, and L. Thiele, "Comparison of Multiobjective Evolutionary Algorithms: Empirical Results," Evol. Comput., vol. 8, no. 2, pp. 173-195, Summer 2000.

[25] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable Test Problems for Evolutionary Multi-Objective Optimization," In Proc. CEC, Piscataway, New Jersey, May 2002, vol. 1, pp. 825-830.

[26] S. Huband, P. Hingston, L. Barone, and L. While, "A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit," IEEE Trans. Evol. Comput., vol. 10, no. 5, pp. 477-506, Oct. 2006.

[27] L. While, P. Hingston, L. Barone, and S. Huband, "A Faster Algorithm for Calculating Hypervolume," IEEE Trans. Evol. Comput., vol. 10, no. 1, pp. 29-38, Feb. 2006.