CrossMark

Global optimization of general constrained grey-box models: new method and its application to constrained PDEs for pressure swing adsorption

Fani Boukouvala[1,2] · **M. M. Faruque Hasan**[1,2] ·
Christodoulos A. Floudas[1,2]

Abstract This paper introduces a novel methodology for the global optimization of general constrained grey-box problems. A grey-box problem may contain a combination of black-box constraints and constraints with a known functional form. The novel features of this work include (i) the selection of initial samples through a subset selection optimization problem from a large number of faster low-fidelity model samples (when a low-fidelity model is available), (ii) the exploration of a diverse set of interpolating and non-interpolating functional forms for representing the objective function and each of the constraints, (iii) the global optimization of the parameter estimation of surrogate functions and the global optimization of the constrained grey-box formulation, and (iv) the updating of variable bounds based on a clustering technique. The performance of the algorithm is presented for a set of case studies representing an expensive non-linear algebraic partial differential equation simulation of a pressure swing adsorption system for CO_2. We address three significant sources of variability and their effects on the consistency and reliability of the algorithm: (i) the initial sampling variability, (ii) the type of surrogate function, and (iii) global versus local optimization of the surrogate function parameter estimation and overall surrogate constrained grey-box problem. It is shown that globally optimizing the parameters in the parameter estimation model, and globally optimizing the constrained grey-box formulation has a significant impact on the performance. The effect of sampling variability is mitigated by a two-stage sampling approach which exploits information from reduced-order models. Finally, the proposed global optimization approach is compared to existing constrained derivative-free optimization algorithms.

Keywords Derivative-free optimization · Kriging · Quadratic · Constrained optimization · Sampling reduction · Global optimization

✉ Christodoulos A. Floudas
  floudas@tamu.edu

[1] Artie McFerrin Department of Chemical Engineering, Texas A&M University,
  College Station, TX 77843-3122, USA

[2] Texas A&M Energy Institute, Texas A&M University, College Station, TX 77843-3122, USA

🖄 Springer

1 Introduction

Constrained grey-box modeling and optimization have applicability in various fields which primarily rely on expensive simulations and/or input–output data. Many application areas exist ranging from expensive finite-element or partial-differential equation systems, design of separation processes and flowsheet optimization to mechanical engineering, to financial management, to geosciences, to molecular design, to material screening, to supply chain optimization and to pharmaceutical product manufacturing [1–17]. Optimization problems which stem from the aforementioned applications are in their vast majority constrained, with a set of input variables affecting a set of outputs through a complex model, while at the same time the outputs are associated with both the objective function and the constraints of the formulation. We define a constrained grey-box optimization system as a formulation which contains a set of explicitly known equations (constraints or objective function) and a set of explicitly unknown functions (constraints or objective function) for which we only have a black-box computer program (Fig. 1). The direct use of deterministic global optimization methods based on analytical C^2 functions is prohibitive in such applications because of (a) high computational cost of the model which implies that the calculation of derivatives is not viable, (b) noise, discontinuities and multiple local optima in the objective function and/or constraints of the model which leads to unreliable derivative information, and (c) complete lack of the model equations if they are provided in the form of input–output data [18–22].

A promising approach for optimizing grey-box problems is the development of surrogate approximation models for the explicitly unknown equations of the system, which aim to guide the search towards the true optimum of the original model [18,23–36]. Surrogate models serve as analytical approximations to the underlying unknown equations, which allow for the use of derivative-based optimization. However, it has been found that the most efficient surrogate functional forms are multimodal non-convex functions; thus global optimization is necessary in order to solve grey-box optimization problems [37,38]. Existing literature in grey-box modeling and optimization predominantly employs local optimization methods for the optimization of the formed surrogate formulations. Moreover, existing methods typ-
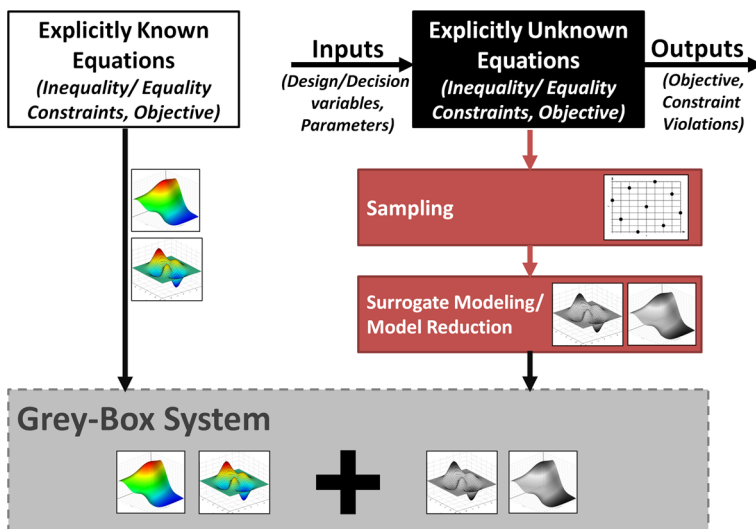

Fig. 1 Definition of a grey-box system

ically treat this class of problems as pure "black-boxes", assuming that no information is available in analytical form, often forcing ignorance of valuable information or limitation of their applicability. Finally, despite the great interest that these methods have attracted in the literature, the majority of the existing methodologies have been developed for unconstrained, box constrained, or known closed-form constrained problems [19,21]. Handling of grey-box constraints is still an open question, while there is scarcity of global optimization approaches for multidimensional general constrained grey-box problems. In this work, constrained grey-box optimization is treated as a compilation of deterministic global optimization subproblems stemming from sampling and design of experiments, parameter estimation and global optimization of surrogate formulations.

Motivated by recent work for the constrained optimization of a complex, non-linear partial differential equation system for pressure swing adsorption [14–16,39,40], this work aims to introduce a general constrained grey-box optimization framework which is suitable for a wide range of applications, and can rigorously handle multiple known and unknown types of constraints. In addition, this work aims to answer three important questions in general constrained grey-box optimization:

(a) How does the variability in the initial sampling set affect the consistency of the grey-box algorithm, and how can this be mitigated?
(b) What is the importance of using deterministic global optimization to solve (i) the surrogate function parameter estimation problems and (ii) the overall constrained grey-box problem?
(c) What is the effect of the surrogate function selection on the performance of the proposed approach?

The following sections are organized as follows. In Sect. 2, the problem formulation and a literature review on the developments in constrained derivative-free optimization are introduced. Section 3 presents the motivating example of a PSA process for CO_2 capture. Section 4, describes the algorithmic components of the proposed methodology, including sampling strategies, surrogate function parameter estimation and validation, global optimization of the overall constrained surrogate model, and bound refinement. Results and comparative analysis are described in Sect. 5 on a set of 11 case studies of PSA based CO_2 capture optimization using different materials.

2 Problem formulation and brief literature review

Grey-box optimization belongs in the category of Derivative-Free Optimization (DFO), which is a broad classification of the methodologies that do not use derivative information of the original model [19]. DFO methods can be divided into subcategories based on the use or not of surrogate models (model-based or direct-search) and on the type of search for samples within the entire input space (global-search) or a local subregion (local-search). The majority of the existing grey-box optimization methods have been developed for box constrained problems limiting their applicability to real-life applications. Extensions to constrained cases have been performed through penalty-type aggregated methods [41–44], filters [45], complex statistical criteria [45–47], aggregated constraint satisfaction functions [25], or surrogate models [26,31,32,48]. A comprehensive analysis and comparison of the existing box-constrained DFO methods available in the literature can be found in two recent reviews, Rios and Sahinidis [21] and Kolda et al. [22].

In this paper we focus on general constrained grey-box problems which are described by the formulation shown in (P1). In (P1) three different types of constraints are present, namely known general closed form constraints, grey-box constraints, and box constraints:

$$\min_{x} f(x) \qquad \qquad \text{(P1)}$$
$$s.t. \quad g_m(x) \leq 0 \quad \forall m \in \{1, \ldots, M\}$$
$$g_k(x) \leq 0 \quad \forall k \in \{M+1, \ldots, K+M\}$$
$$x_i \in \left[x_i^L, x_i^U \right] \quad i = 1, \ldots, n$$
$$x \in \Re^n$$

where sets $k \in \{1, \ldots, K\}$ and $m \in \{1, \ldots, M\}$ represent the constraints with known closed-form and the constraints with unknown form (grey-box), respectively. The set of n continuous independent inputs x have known finite bounds $\left[x_i^L, x_i^U \right]$. The form of the objective function $f(x)$ and the constraints in set M, $g_m(x)$, are not available explicitly.

Significant algorithmic developments were originally developed based on direct-search derivative-free concepts for box-constrained optimization [19,49–52], where the search is driven only based on function evaluations. Local direct-search DFO methods have been extended to constrained using various methods such as filter approaches and penalty or barrier methods [26,42,43,45,53]. However, these methods suffer from high dependence on the initial point and the entrapment within the closest to the initial point local optimum and generally require a large number of samples which may be prohibitive [19,50]. Later on, the idea of using fitted functions (surrogate models) based on the input–output data was found to expedite the search towards optimal solutions. Different methodologies have been proposed which use a surrogate model to approximate $f(x)$ (i.e., kriging, quadratic, linear, or radial-basis functions), as well as the iterative sampling criterion which will lead to optimal solutions using fewer samples within a local-trust region (local model-based methods) [19,54–61], or within the entire region bounded by the upper and lower bounds of the input variables (global model-based methods) [23,24,30,33–35,62].

In this work, we are interested in global-search methods which consider the entire investigated space and do not depend on a single initial point. In all existing global-search surrogate-based optimization methods asymptotic convergence is based on the theorem of Torn and Zilinskas [63] which states that any algorithm may converge to a global optimum of a continuous function within a compact set if "its sequence of iterates is everywhere dense" within the compact set [27,28,30,60]. The key in these types of algorithms is the derivation of search criteria which will reach good solutions faster, by identifying promising regions while retaining a balance between local and global search in order to avoid getting trapped in suboptimal regions. Even though global-search model-based methods have gained great popularity for optimizing expensive models, it is true that there is no guarantee that the final optimum may even be a stationary point [29,30]. Extensions of global-search methods have been proposed using probabilistic criteria, extreme barrier methods, aggregated constraint satisfaction surrogates, and treatment of each constraint using individual surrogate functions [3,31,32,46–48]. However, these approaches require the formulation of optimization problems with multimodal functions which are difficult to solve to global optimality. Extensive sampling methods such as Monte Carlo are proposed to identify the optimal point of the new constrained statistical criteria, which introduce uncertainty into the proposed approaches [46]. The applicability and efficiency of introducing a more complex non-convex search criterion, becomes questionable as the dimensionality and the number of constraints of the problem increases. The performance of all of the above developments has been tested on relatively

small test problems, with a few recent exceptions such as the work of Regis et al. [31,64] who locally optimize radial basis function formulations which represent high dimensional problems. The value of global optimization for black-box optimization was recognized by Jones et al. [30] who proposed a branch-and-bound algorithm to globally optimize the expected improvement function, which was tested on low-dimensional problems.

The work proposed here belongs to the global-search and model-based categories, which are two qualities with many advantages. Firstly, the use of smooth surrogate models allows the interpolation between existing samples, which has been shown to reduce sampling requirements overall. At the same time, development of surrogate approximations enables the use of deterministic global optimization methods to optimize these non-convex surrogate functions. In fact, the use of advances of the deterministic global optimization literature in general constrained grey-box optimization is one of the unexplored and least discussed topics. In the work of Regis et al. it is claimed that asymptotic convergence is assured even if the formed surrogate problem is not globally optimized [32,60]. This is achieved by imposing rules which ensure that any new sample must have a minimum distance from all of the existing samples. However, it is not discussed whether this has an effect on the speed of convergence. This is a valid concern, since the effort which is made into sophisticated sampling methods and development of complicated functional forms, may be impaired if the globally optimal parameters are not used or if a suboptimal solution of the constrained grey-box approximation model is used as the next promising sample.

Secondly, searching the entire investigated region is important because this reduces the high dependence on a single initial point. However, even if a set of initial samples are collected using space-filling techniques in order to build the initial grey-box models, the effect of the initial sampling set may be significant on the final solution. For this reason, most derivative-free optimization studies perform a number of tests, starting from different initial sampling sets, in order to verify the average and variance of the final solution. In a realistic application which may have significant computational requirements, it would be very important to be able to mitigate the effect of the initial sampling, in order to guarantee consistent and reliable performance. Through this work we aim to investigate whether advances in global optimization can lead to improved algorithms with high consistency and reliability.

3 Motivating example

Benchmarking of constrained grey-box methods is extremely difficult due to the diverse nature of the applications which may have different difficulties in terms of computational cost, noise in their output values, dimensionality and the number of constraints and form of the feasible region. In the majority of the cited literature thus far, proposed methods are tested on relatively small test problems, comprised of smoother continuous functions. In a realistic case, where the input–output data depends on a complex simulation, the data may be extremely noisy, while the true underlying function or feasible region can be discontinuous. Another important aspect which can be underestimated when solving benchmark problems is the number of function calls. In fact, in most realistic applications the affordable number of calls to the expensive simulation might direct the importance towards attainment of better solutions with fewer samples, as opposed to asymptotic guarantee of convergence to a global optimum. For this reason, this work has been motivated by a realistic case study, which is a large Non-Linear Algebraic and Partial Differential Equations (NAPDE) system. NAPDE systems have applicability in all fields of engineering for the representation of complex geometries, multiphase flows and reactions.

Specifically, this work has been developed for the optimization of an adsorption-based process for post-combustion CO_2 capture from power plant flue gas which is considered to be a binary mixture of CO_2 (14 %) and N_2 (86 %). The details of the full simulation model are provided in the "Appendix", while the representation of the model as a constrained grey-box problem is provided in (P2). A cyclic PSA process with four-steps (pressurization, adsorption, blowdown and evacuation) [14] is described by a system of complex NAPDE system of equations. The main equations of the simulated model are given in Table 11 of Appendix 1, while the notation, initial and boundary conditions used are provided in "Appendix 2".

The NAPDE model allows for the detailed simulation under different operating conditions and designs (inputs) to obtain the process performance (outputs) at cyclic steady state. More specifically, the PSA process has seven significant input variables: column length (L), blow-down pressure (P_{bd}), evacuation pressure (P_{evac}), adsorption time (t_{ads}), blowdown time (t_{bd}), evacuation time (t_{evac}), and compression pressure (P_H). These input variables affect three outputs of interest which are the total annual cost, the purity and recovery of the outlet CO_2 stream. Both the purity and recovery must be larger than a minimum specified value for the process to be feasible. For a typical CO_2 capture process this is set to be 90 %. Therefore, when designing a cost-effective PSA process, the objective is to minimize the total annual cost (usually per ton of CO_2), and the design constraints are: purity ≥ 0.90, and recovery ≥ 0.90. There is an additional operating constraint to ensure that the blowdown pressure is always higher than the evacuation pressure. All of the design and operating constraints of the problem are given in Table 12 of Appendix 1.

The set of the seven input variables can be varied within a predefined region, bounded by their upper and lower bounds based on the operational limits of the process. By changing these input variables within their bounds and by satisfying the known linear constraint, the cost, purity and recovery will be affected in a way which will be dictated by the NAPDE model. However, this model requires significant computational time for one simulation run to reach cyclic steady state, thus attempting to directly solve it using deterministic global optimization methods is prohibitive. The above is due to the fact that the NAPDE system is solved through the finite volume method over a dense grid of spatial discretization elements, dynamically, for a large number of cycles ($\sim$50–200) to ensure steady-state. Depending on the numerical complexity of each specific material, which affects the Langmuir model parameters, the computational time of a single simulation may be about 15 minutes. The grey-box model of the NAPDE model of the PSA process, which is analogous to the representation of (P1), is defined as follows:

$$\min_{\mathbf{x}} \quad TAC(\mathbf{x}) \qquad \text{(P2)}$$

$$s.t. \quad Pu(\mathbf{x}) \geq 0.90$$
$$Re(\mathbf{x}) \geq 0.90$$
$$x_2 \geq x_3$$
$$1 \leq x_1 \leq 5$$
$$0.1 \leq x_2 \leq 1$$
$$0.01 \leq x_3 \leq 1$$
$$20 \leq x_4 \leq 50$$
$$20 \leq x_5 \leq 50$$
$$20 \leq x_6 \leq 100$$
$$1 \leq x_7 \leq 10$$

where, *TAC* is the total cost (Eq. A17), Pu is the purity (Eq. A10), Re is the recovery (Eq. A9), and $\boldsymbol{x}$ is vector representing the seven input variables to the PSA process component x_i ($i = 1, \ldots, 7$) where $x_1 = L$, $x_2 = P_{bd}$, $x_3 = P_{evac}$, $x_4 = t_{ads}$, $x_5 = t_{bd}$, $x_6 = t_{evac}$, $x_7 = P_H$. As a result, the large NAPDE system shown in Tables 11-12 is summarized through (P2), with one grey-box objective, two grey-box constraints ($M = 2$), one known constraint ($K = 1$), and box constraints for variables $x_1 - x_7$.

4 General constrained grey-box global optimization algorithm

The proposed methodology contains multiple novel aspects which are outlined in Fig. 2. The novel features include: (a) optimized selection of the sampling data for model development and updating though optimization of the probabilistic distance of the samples in both the input and the output spaces, (b) the exploration of different types of functional forms such as general quadratic, kriging and signomial functions for approximating the objective function and each of the constraints individually, (c) the global optimization of the parameter estimation of surrogate functions, (d) the global optimization of the overall constrained surrogate problem, and (e) an initial global search strategy followed by partitioning of the investigated space using clustering techniques. An overview of the most important aspects of the methodology follows, while each step is described in detail in the following subsections.

Firstly, it is important to formulate the problem which is equivalent to formulation P1 (equivalently P2 for the motivating example), by identifying the objective, constraints and the significant input variables which affect all of the above. Subsequently, a set of samples are required which will be used to build the input–output mappings for each of the unknown correlations. The selection of this initial set may be performed based on different methods, such as an optimized Latin Hypercube Design (LHD) (Sampling Strategy 1), or through more sophisticated sorting and selection methods which require the availability of a larger prior database (Sampling Strategies 2 and 3).

Based on the obtained samples for which the full simulation is performed, the method requires the selection of the type of surrogate function which will be used to represent each of the constraints and the objective function in order to solve $M + 1$ parameter estimation problems (i.e., one for the parameter estimation of the surrogate function for the objective function, and M for the parameter estimation of the surrogate functions for the M grey–box constraints). A series of different functional forms can be tested and selected, namely general quadratic, kriging (exponential), radial-basis functions and signomial functions. Often the identification of the optimal parameters for each of the fitted functions by the minimization of the sum of least-square differences between the predictions and the observations constitutes a challenging optimization problem, which we aim to solve to global optimality.

Once all of the above have been completed, the formulated constrained surrogate optimization problem, including the surrogate objective $\widetilde{f}(x)$, the surrogate constraints, $\widetilde{g}_m(x)$, and the original constraints, $\widetilde{g}_k(x)$, and variable bounds must be solved to global optimality using deterministic global optimization methods [65–67], collecting a set of possible local solutions and the final global solution. The constrained surrogate formulation does not depend on the expensive simulation, thus it can be solved rigorously using deterministic global optimization methods. Since it is expected that the formulated constrained surrogate formulation will have multiple local solutions, a set of diverse local solutions along with the final global optimum solution are selected as promising future samples. In the sequel, the full simulation must be performed at the set of the unique local and global solutions, in order to obtain their true objective and constraint values. The new samples are incorporated within the sampling set and the
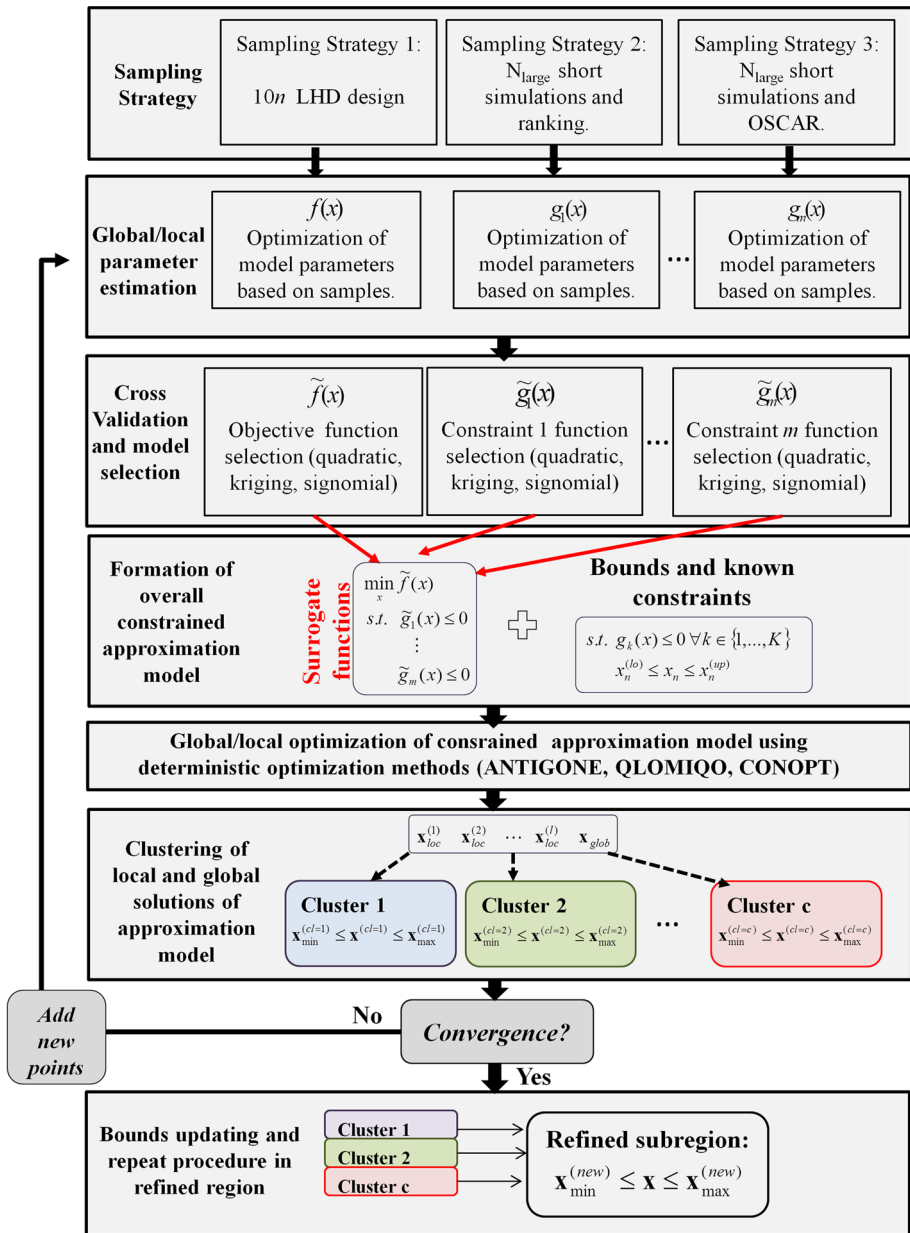
Fig. 2 Building blocks of novel global optimization algorithm for general constrained grey-box model

parameter estimation problem and optimization of the overall constrained grey-box problem are solved iteratively in order to update the model parameters and repeat the aforementioned steps. The procedure is repeated iteratively until certain convergence criteria are met.

During each iteration a clustering procedure assigns each newly obtained sample to a cluster based on its x-space location, objective function value and feasibility when compared to existing samples. The points which are collected often form clusters, since they are a

diverse set of local solutions collected. The developed clustering technique identifies the total number of clusters, the average objective function value and the average feasibility of each cluster and the x-space bounds of the samples contained within each cluster. Once the algorithm converges, the clusters which have been formed are analyzed to provide valuable information about regions of the feasible space which contain promising solutions. This information is used to update the bounds of the search space in such a way so that the clusters which contain the best solutions are incorporated. Details about the clustering technique, as well as the analysis and bound updating are described in Sect. 4.4.

Finally, the algorithm proceeds to a second stage during which the entire procedure is repeated within the space defined by tighter bounds, until the same convergence criteria are met. During this second stage, new samples are collected and new surrogate models are developed for the constraints and the objective function. These models are guaranteed to be more detailed representations of the locally defined search space, while the excluded regions are no longer part of the model.

4.1 Sampling methods and sampling reduction

Selecting the samples for which the full simulation is performed in order to build and update the surrogate function parameters is a critical issue. Unbalanced designs may lead to low accuracy of the predictions in unexplored regions, as well as numerical issues during the parameter estimation due to singularity in the correlation matrices. At the same time, it is desired to keep the sampling requirements to a minimum since the majority of applications of interest rely on simulations which have a high computational cost. For the initial sampling designs, Latin Hypercube Designs are typically used [68], since they have good space-filling abilities with fewer number of samples when compared to full-factorial designs. The popular '10 times the number of input variables' rule of thumb has been used as a baseline in most of the methods developed in literature, and it is denoted as Sampling Strategy 1 in this work (SS1). However, there is a significant amount of variability associated with the initial Latin Hypercube Design, observed and reported by other authors [59]. In other words, by using a slightly different set of $10n$ samples the final result may be significantly different. The variability introduced by the initial sampling depends on the problem itself as well as the total number of samples which are collected. If one can afford to collect a large number of samples, then the effect on the final result will be reduced. However computational expense is often a limiting factor to the total number of collected samples. For example, in the presence of multiple constraints which can significantly limit the feasible space, a simple LHD may not even contain any feasible samples causing more uncertainty to the end result. In addition, the complexity and smoothness of the underlying functions can cause more variability towards the overall performance.

In grey-box optimization problems, where an expensive simulation is available but cannot be directly used for optimization, there are often multiple ways to reduce the complexity of the full simulation in order to obtain a faster model, which serves as another type of surrogate to the original [69–72]. When this information is available, a possibility arises for collecting a large pool of samples which may reveal trends about the final objective and feasible region of the final full simulation. For example, in the motivating example, the discretization of the finite-element method model is coarsened significantly and the number of cycles can be reduced to a small number, resulting to a reduced-order simulation which is solved in 2–3 s as opposed to 15 min. Once the total set of reduced-order samples is collected, the most challenging aspect is the selection of an optimal subset of the large pool of reduced-order

samples for which a full simulation is performed, for subsequent use in surrogate function parameter estimation.

Returning to the motivating example, the objective would be to select a a subset S ($S \subset N$) of samples out of a superset of N reduced order samples, where $card(N) = N_{l\,arg\,e}$, $card(S) = 10n$ and $N_{l\,arg\,e} \gg S$. The reduced order samples have an approximated value for purity, recovery and cost provided by the short simulation described earlier, which we can afford to simulate in abundance. Sampling Strategy 2 (SS2) refers to naive ranking of the fast-simulation samples based on purity and recovery and selecting the first S samples which satisfy the constraints or have the minimum constraint violations. However, this method does not guarantee that the set of S samples is balanced within the investigated space, as opposed to being clustered in several regions leading to surrogate functions which may be highly inaccurate in insufficiently sampled regions. In other words, it is desired to select a subset of samples with promising feasibility as well as objective function values, which also have optimal space-filling properties in the x-space. This set of criteria may be conflicting towards the selection of the sampling subset.

The above observations led to Sampling Strategy 3 (SS3), which employs Mixed Integer Linear Programming (MILP) to perform the optimal selection. Inspired by current work of Li and Floudas [73] it is realized that sampling selection in grey-box optimization is an equivalent problem to scenario reduction in stochastic programming problems. In other words, the reduced-order samples of the grey-box model are treated as different scenarios of an uncertain problem. The full set of reduced order simulations leads to a probability density function, which we want to accurately represent using a subset of the obtained samples. This subset will constitute the set of points for which the full simulation will be performed.

The developed algorithm for Optimal SCenArio Reduction (OSCAR) minimizes the probabilistic distance between the original and reduced input (x) distribution domain, while simultaneously it incorporates objectives in the output domain such as expected, worst and best performance. The MILP formulation used for sampling selection is given in Problem 3:

$$\min \sum_i p_i d_i + \sum_i y_i \left(\sum_j w_j f_{i,j} \right) \tag{P3}$$

$$s.t.$$

$$\sum_{i \in N} y_i = card(S)$$

$$d_i = \sum_{i' \in I} c_{i,i'} v_{i,i'} \quad \forall i \in N$$

$$0 \leq d_i \leq y_i f_{max} \quad \forall i \in N$$

$$\sum_{i' \in N} v_{i,i'} \geq y_i \quad \forall i \in N$$

$$\sum_{i'} v_{i,i'} \leq 1 \quad \forall i \in N$$

$$0 \leq v_{i,i'} \leq 1 - y_i \quad \forall i, i' \in N$$

where $c_{i,i'} = \sum_{k=1}^{K} |x_k^i - x_k^{i'}| + |f_i - f_{i'}|$ is the distance between two samples in $x - y$ space, K is the number of input variables, N is the set of original scenarios, $card(S)$ is the desired
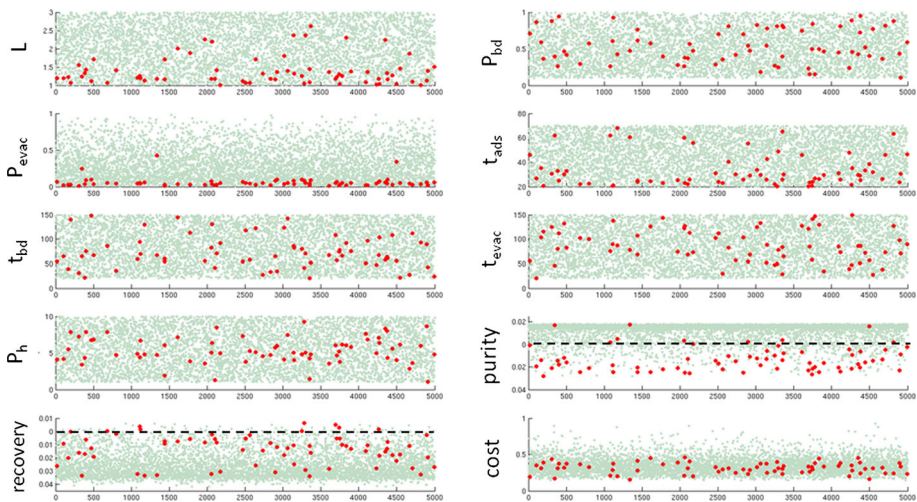
Fig. 3 Example of sampling reduction based on PSA motivating example. *Green* samples (5000) plotted against selected *red* samples (70)

number of scenarios to be selected, p_i is the probability of each scenario, y_i is the binary variable denoting whether a scenario i is removed ($y_i = 1$), d_i is the minimum Kantorovich distance from the selected scenarios to removed scenario i, $v_{i,i'}$ are continuous variables denoting whether scenario i is removed and assigned to scenario i', $f_{i,j}$ correspond to the value of output j for each scenario i and w_j are optional weights assigned to each output depending on its importance towards the final selection. Further details about formulation P3 are provided in [73].

The objective of the MILP model may be the minimization of the x-space probabilistic distance, which implies that the second term of the objective in P3 is not present, and thus this will result to the optimal space filling design. However, the objective may also be a weighted sum of the x-space distance, the objective function and constraint function values, which will aim to balance between space-filling properties and promising samples. Formulating this selection procedure as an MILP problem has the advantage of the ability to tune the objective based on the final goal in order to choose a different subset of samples. For instance, if feasibility is the most important aspect, a large weight may be placed on the contribution of the predictions of the constraint function values, or else, additional constraints may be added to the MILP problem to make sure only feasible points are selected.

An example of sample selection is provided in Fig. 3 for the PSA case study. In this figure the light green points represent the location of the initial 5000 samples obtained from the fast simulations in both the x and the y space. The 70 selected points are highlighted in red. The selected points clearly have a wide distribution in the x-space, but also the majority is feasible in terms of recovery and purity, and tend to have lower objective function values. Moreover, it is observed that the search space is reduced only in one input variable, the evacuation pressure (P_{evac}). This is a result of the known closed-form constraint of P2 ($P_{bd} \geq P_{evac}$) which is incorporated in the MILP formulation through prior selection of a candidate set which satisfies this constraint. Specifically, when known-constraints are only a function of input variables, the superset of initial samples can be filtered prior to performing the subset selection, simply by removing any samples which do not satisfy the known constraints. Subsequently, subset selection is performed on the filtered data set. At this point, it should be made sure that the initial large design has a sufficient number of feasible samples.

4.2 Selection of functional form for the surrogate model

The proposed framework for general constrained grey-box models is not restricted to a single type of functional form for estimating the underlying unknown objective function and constraints. Several types of functional expressions have been tested, keeping in mind that the complexity and non-linearity of the form will affect the solution times and solution quality of the preceding global optimization problems. The functional forms which have been tested range from linear, general quadratic, kriging, and signomial functions. Moreover, since each of the constraints and the objective function are approximated by a separate surrogate model, there is no restriction in using only one type of functional form for all of the $M + 1$ models, since a combination of surrogate model types may be used to form the final surrogate optimization problem (P1 and P2). This selection may be identified based on prior knowledge or cross-validation techniques.

A key feature of this work is the introduction of deterministic global optimization for the parameter estimation in each selected surrogate function. The final objective of this optimization model is the identification of the mapping $y = f(x)$ which connects a measured output y to a set of input variables of dimensionality n. In certain cases, (i.e., linear, general quadratic) the parameter estimation might have an analytical solution or is a well-posed convex non-linear problem; under the assumption that the samples have good space-filling geometric properties and the training data is adequately scaled. However, in other cases, parameter estimation is a challenging global optimization problem, which has been overlooked in the literature so far. Especially for interpolating techniques such as kriging, where the model predictions are equal to the observed outputs at all the sampled locations, there are usually multiple possible combinations of parameters for which this condition is satisfied, which means that there exist multiple local/global solutions to the parameter estimation problem. In this case, validation techniques are used along with training in order to identify a solution of the problem which best describes the data, while at the same time it does not overfit the data.

4.2.1 Surrogate functions type 1: kriging functions

One of the most commonly used functional forms in surrogate-based optimization is kriging [74,75] which is an interpolating technique with several advantages, but also challenges. Global optimization of the kriging parameters is a demanding task since kriging is by nature a purely interpolating technique, thus cross-validation procedures are usually employed in order to calculate a prediction error for the model. In addition, the functional form of kriging introduces non-linear and non-convex terms. Moreover, the estimation of the optimal kriging parameters requires the use of matrix algebra for the inversion of the covariance matrix of the observations. Formulating the kriging parameter estimation as a global optimization problem leads to relatively large problems, with a number of variables and equations highly dependent on the number of training observations. However, as the dimensionality of the problem increases, it is advisable to collect a larger set of training samples in order to develop an accurate response surface. As more samples are collected, the size of the problem increases and thus globally optimizing the parameters is a very challenging task. By employing a simultaneous parameter estimation and validation technique, and careful tightening of the model parameter bounds, we are able to address this global optimization problem.

Starting from a set of N sample data $\mathbf{X} = \left\{ \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)} \right\}^T$ for which the observed response is $\mathbf{y} = \left\{ y^{(1)}, y^{(2)}, \ldots, y^{(N)} \right\}^T$ the main assumption of kriging is that any pair $i - j$ of the random samples is correlated through Eq. (1), such that:

$$\mathbf{R} = cor\left[y(\mathbf{x}^{(i)}), y(\mathbf{x}^{(j)}) \right] = \exp\left(-\sum_{k=1}^{n} \theta_k \left(x_k^{(i)} - x_k^{(j)} \right)^2 \right) \tag{1}$$

In our approach, a subset $SMB \subset N$ of the samples form the training set, while the rest of the points belong to set SCV, which is the validation data set. This approach is used since the set SMB will be interpolated by the developed kriging function and it is desirable to produce a function which interpolates the best observations. On the other hand, it is not desirable for the training set to be clustered solely around a promising region, since the predictive ability of the model will be very poor in regions outside this space. The remaining validation points of SCV are not purely interpolated, but they are used to calculate the objective function of the parameter estimation model, which is the sum of squared error between the observed y_{SCV} and the one predicted by the kriging model. Based on ideas described in the previous section, OSCAR [73] is used for the selection of the SMB and SCV sets, using 80 % of the samples as interpolation points.

The basic advantage of all interpolating surfaces- compared to non-interpolating functions such as general quadratic functions- is that they are more flexible to approximate non-linear responses with a fewer number of parameters. The basic concept of kriging is that the output of any new point will be more similar to an observation if their Euclidean distance in the $\mathbf{X}$ domain is smaller. This is exactly what the basis function of Eq. (1) aims to capture. The rate at which the change in distance between two points influences the change in the output y in each dimension j is a function of the fitted parameters θ_j. It has been observed that modeling this correlation between pairs of observations is so powerful, that a new point can be predicted by the following Eq. [76]:

$$\hat{y}\left(x^{(new)} \right) = \hat{\mu} + \mathbf{r}^{\mathbf{T}} \mathbf{R}^{-1}(y - \mathbf{1}\hat{\mu}) \tag{2}$$

where $\hat{\mu}$ is an additional parameter of the kriging function which is a mean response throughout the explored region, which is corrected by the second term which is a function of the θ_j parameters. $\mathbf{R}$ is given in Eq. (1) while $\mathbf{r}$ is a vector which contains the correlations of form (1) between the existing samples and the new unknown x. The optimal values of $\hat{\mu}$ and θ_j are found from the maximization of the maximum likelihood function of y_{SMB} subject to the already observed data $\mathbf{X}$, which can be calculated in closed form. The derivation of Eq. (2) is based on the numerical solution of the maximization of the conditional likelihood of $\hat{y}\left(x^{(new)} \right)$ subject to the identified optimal parameters and the already observed data [18]. Based on the kriging properties, a closed form expression of the associated uncertainty of each prediction can also be calculated. However, this is not used in this work.

The optimization problem for the parameter estimation of the kriging surrogate function given a set of $\mathbf{N}$ data points of dimension n is:

$$\begin{aligned}
\min_{\mu,\theta} \ & \sum_{SCV} \left(\hat{y}_{SCV} - y_{SCV}^{(obs)} \right)^2 \Big/ ord(SCV) \\
s.t. \ & y_{SMB} = \hat{\mu} + \sum_{SMB'} \hat{c}_{SMB'} u_{SMB,SMB'} \\
& \hat{y}_{SCV} = \hat{\mu} + \sum_{SMB'} \hat{c}_{SMB'} u_{SCV,SMB'} \\
& \hat{c}_{SMB} = \sum_{SMB'} v_{SMB,SMB'} \left(y_{SMB} - \hat{\mu} \right) \\
& \mathbf{I}_{i,j} = \sum_{SMB'} u_{i,SMB'} v_{SMB',j} \\
& u_{i,j} = \exp\left[-\sum_n \hat{\theta}_n \left(x_n^{(i)} - x_n^{(j)} \right)^2 \right]
\end{aligned} \tag{P4}$$

where the main parameters are $\hat{\theta} \in R^n$ and $\hat{\mu} \in R$. Matrix $\mathbf{U}$ is the square symmetric covariance matrix with elements $u_{i,j}$ defined in P4. Constraints 3–4 of P4 can be omitted since P4 is uniquely defined without them. However, we have found that they are often beneficial for locating feasible solutions. The main limitation of the formulation of the parameter estimation for kriging as a global optimization problem is the inversion of matrix $\mathbf{U}$ which introduces a large number of intermediate variables and equality constraints. In addition, the non-linear form of the terms $u_{i,j}$ increases the non-linear terms of the problem significantly. Finally, for global optimization, it is desirable to provide lower and upper bounds on all of the variables of the problem. The following bounds are provided for the problem:

$$\hat{\theta}_n \in [0.01, 30]$$

$$u_{i,j} \in \left[\exp\left[-\sum_n \hat{\theta}_n^{up} \left(x_n^{(i)} - x_n^{(j)} \right)^2 \right], \exp\left[-\sum_n \hat{\theta}_n^{lo} \left(x_n^{(i)} - x_n^{(j)} \right)^2 \right] \right]$$

The above bounds have physical meaning since parameter $\hat{\theta}_n$ signifies the importance of each input variable and when the input data is scaled, it is not expected for the values to exceed the value of 30 or be less than 0.01 (otherwise the significance of this variable is very small and it should perhaps be removed from the experimental design). Moreover, the elements of matrix $\mathbf{U}$ have a closed form solution which is minimized when $\hat{\theta}_n$ is maximum and is maximized when $\hat{\theta}_n$ is minimum. The bounds for $\mathbf{U}$ can be adaptively tightened by relaxing the value of parameter $\hat{\theta}_n$ obtained from the parameter estimation solved in the prior iteration of the grey-box optimization approach. Providing all of the above bounds has shown to improve the performance of the global optimization of the kriging parameters. However, as the problem size increases the required time for reaching global optimality using the recent global optimization solver ANTIGONE [67] increases significantly. Due to the need to keep a balance between reducing the overall computational cost of the optimization, a time limit is imposed on the solution time for each of the parameter estimation problems. As the iterative grey-box optimization framework proceeds sampling points tend to be clustered in small subregions, and it becomes difficult to locate feasible solutions. This is due to the fact that the value of the off-diagonal elements $u_{i,j}$ can take values closer to 1, and matrix $\mathbf{U}$ becomes near singular. In order to overcome this issue, it is suggested in the kriging literature to add a small positive number (nugget parameter) to the diagonal elements of $\mathbf{U}$. This modification will not affect the solution of the problem, however, it will lead to a kriging model which will not be an exact interpolator of the collected samples. In our work, this is considered as a slack variable, which is a positive small number (≤ 0.1) and the sum of these slack variables is incorporated into the objective function.

4.2.2 Surrogate functions type 2: general quadratic functions

The general quadratic function is a full quadratic with a constant term, linear terms, quadratic terms and bilinear terms to account for the interactions (Eq. 3). This type of response surface is often used for approximating noisy data obtained from physical experiments to describe effects for which there is no physical model available [77].

$$y = \beta_0 + \sum_n \beta_n x_n + \sum_n \sum_{n' \geq n} \beta_{n,n'} x_n x_{n'} \tag{3}$$

The parameter estimation for the $(n + 1)(n + 2)/2$ parameters of the general quadratic function is shown in (P5), and it can be solved to global optimality using the global optimization

solver GLOMIQO [65], which is designed to solve Quadratic Constrained Problems (QCP) and Mixed Integer QCP problems (MIQCP), as well as with ANTIGONE [67].

$$\min_{\beta} \sum_{N} \left(y_{pred}^{(N)} - y_{obs}^{(N)} \right)^2 \Big/ \mathrm{N} \tag{P5}$$

$$y_{pred}^{(N)} = \beta_0 + \sum_{n} \beta_n x_n^{(N)} + \sum_{n} \sum_{n' \geq n} \beta_{n,n'} x_n^{(N)} x_{n'}^{(N)}$$

One of the disadvantages of general quadratic functional forms when compared to kriging functions, is their decreased flexibility in describing non-linearity which deviates from a quadratic form. This is shown in [29] where kriging functions are found to be more suitable for the optimization of highly non-linear global optimization benchmark problems. In addition, the number of parameters of full-quadratic functions increases significantly as the dimensionality of the problem increases, which will have an effect on sampling requirements. On the other hand, using general quadratic forms leads to optimization problems for both the parameter estimation phase, and the constrained grey-box optimization phase which can be solved to global optimality with certainty and less computational cost.

4.2.3 Surrogate functions type 3: signomial functions

Signomials have not been used in the grey-box optimization literature hitherto, despite the fact that they have attracted significant attention in the deterministic global optimization literature [65,66,78]. The latter is a significant characteristic, since if a grey-box model is sufficiently described by a signomial type function, then a global optimization solver such as ANTIGONE [67] may be used to identify the global solution of the surrogate problem. A signomial has the typical form shown in Eq. (4).

$$y(x) = \sum_{i=1}^{s\_ord} \left(c_i \prod_{j}^{n} x_j^{\alpha_{ij}} \right) \quad \text{where } x_j > 0 \tag{4}$$

The parameters of a signomial are $c_i \in R^{s\_ord}$ and $\alpha_{ij} \in R^{[s\_ord \times n]}$, where s_ord is the selected order of the signomial which is defined by the user. Clearly, as the order increases, so does the number of fitted parameters of the function. If all c_i parameters are positive, then the function is called a posynomial, while if all α_{ij} are non-negative integers, the function becomes a polynomial.

In order to address the parameter estimation of a signomial function subject to a set of samples, a similar approach to the one for general quadratic functions is followed, where Problem 6 (P6) must be solved using the global optimization solver ANTIGONE [67].

$$\min_{c,\alpha} \quad \sum_{N} \left(y_{pred}^{(N)} - y_{obs}^{(N)} \right)^2 \Big/ \mathrm{N} \tag{P6}$$

$$y_{pred}^{(N)} = \sum_{i=1}^{s\_ord} \left(c_i \prod_{j}^{n} x_j^{\alpha_{ij}} \right)$$

$$x_j > 0 \quad j = 1, \ldots, n$$

4.3 Global optimization of overall constrained grey-box approximation problem and selection of next sampling design points

During this stage of the proposed framework, the overall constrained optimization problem (P2) is solved using the deterministic global optimization solver ANTIGONE [67]. In this formulation, the decision variables are the original variables of the problem. Any grey-box expression which depends on the expensive simulation has been replaced by its surrogate function and any known constraint is incorporated as is. During each iteration, the accuracy of the identified globally optimal solution depends on the accuracy of surrogate functions. This is perhaps one of the most significant steps of every grey-box algorithm, since satisfactory accuracy of the overall constrained surrogate formulation is desired only in promising regions. During this stage, diversity in choosing the next sample is an attribute which most of the competitive methods in the literature strive to achieve, in order to avoid entrapment to false local optima. In the literature, this is achieved by the formulation of complex multimodal expected improvement functions which aim to balance between the prediction quality and uncertainty in order to identify the next sampling location. Optimization of this criterion is one of the disadvantages of these methods, since even if they are cheap to evaluate, they are highly nonlinear and multimodal. Moreover, these methods have been designed to select one sample at a time, while extensions to methods which can identify multiple samples involve complex forms of multiple integrals or multiobjective optimization [18,79].

In the proposed framework, multiple local solutions—which represent upper bounds to the overall constrained surrogate problem—and a global solution of the current formulation are collected during each iteration. The lower bound is obtained by solving the problem to global optimality using the global optimization solver ANTIGONE [67]. The upper bounds are obtained by optimizing the same problem using a local solver (e.g., CONOPT, DICOPT) starting from different initial solutions. This concept is based on the idea that the overall constrained surrogate problem may have multiple local solutions which have a high probability to correspond to regions where true optimal points of the original problem may lie. In other words, the heuristic methods or complex multimodal search criteria which are used in other grey-box methods are replaced by the collection of multiple local/global solutions provided by deterministic optimization algorithms. In addition, local optimization of the overall constrained grey-box problem does not significantly increase the computational cost of the overall method. One could solve as many local problems as the number of obtained solutions up to the current iteration, by using each of the samples as the initial guess. However, as the number of samples increases, this would lead to a large number of optimization problems, without adding new information to the solution pool. This fact is easy to realize since as new samples are collected, they tend to form clusters in promising regions. We perform this selection using OSCAR in order to ensure that we optimally select a diverse set of starting points for local optimization. The criterion for this MILP problem is to select $2n$ samples with maximum diversity in the x-space, in order to increase the probability of obtaining a diverse set of new design points. Feasibility and objective value is also incorporated into the formulation, since it is also important to select starting points with promising predictions. We have found this total number ($2n$) to be a good heuristic rule to balance diversity and computational cost for this specific problem. However, the method can be modified to select any number of starting points with a minimum of 1 (the best point in the sampling set will be selected) and a maximum of the entire set of samples.

Finally, once all of the local and global optimization problems are solved, all of the obtained unique solutions are collected and the full-simulation is performed at those design

points by fixing the input variable values to match the new samples. This approach allows for a variable number of new points to be used at each iteration, based on the complexity of the formulated surrogate problem. After the new points are collected, the algorithm returns to the parameter estimation problem and this approach continues until convergence. The convergence criterion is related to both the accuracy of the surrogate model approximations which should be high, as well as the improvement of the actual final solution within a predefined number of consecutive optimization iterations (coi). A more detailed schematic description of the algorithmic steps is shown in Fig. 4.

4.4 Clustering of obtained samples and updating of x-space bounds

One of the potential disadvantages of surrogate functions is their inability to capture the underlying model with accuracy in the entire domain with a limited number of samples. However, accurate representation of the objective and constraints in the entire space is not necessary, when the end goal is optimization with limited samples. Moreover, it is realized that despite the diversity of obtained local and global solutions during each iteration, samples start to form clusters and the change in the global parameters of the surrogate functions is not significant enough to achieve further improvement in the solution. Thus, it is proposed to repeat the steps of the proposed framework within a subspace of the original x-domain, defined by information obtained by the identified samples.

The challenge presented is the selection of the criteria based on which the bounds of the optimization will be redefined to form subspace(s), within which the optimization procedure will be restricted. These criteria should minimize the probability to discard promising feasible regions. For this purpose, a clustering procedure is incorporated within the proposed framework, which assigns any new obtained samples within clusters during each iteration.
Clustering procedure

1. *Optimization iteration 1 ($opt\_iter = 1$)* Set the number of clusters equal to the number of distinct obtained local and global solutions ($c = l + 1$). Initialize cluster centers and calculate all possible distances between each $i - j$ pair of cluster centers d_{ij}^c. Set radius influence of each cluster i equal to: $r_{in}^c = \frac{1}{a_c}(\min_j d_{ij}^c) j = 1, ..., c$ and $j \neq i$. Parameter a_c is user defined, a default of 2 is used in the proposed framework.
2. *Perform next optimization iteration ($opt\_iter = opt\_iter + 1$)* Collect new set of $l_{new} = l + 1$ local and global solutions.
3. *For $i = 1$ to l_{new}* Calculate the Euclidean distance of sample s to each of the existing clusters and find the nearest cluster. If this distance is less than r_{in}^c, then place this sample in existing cluster and update cluster center. If not, then create new cluster: $c = c + 1$. Recalculate all distances and influence radii d_{ij}^c and r_{in}^c. If all of the new samples belong in existing clusters, the existing set of clusters is updated and no new clusters are created, however, this does not necessarily imply convergence.
4. *If convergence is not met, go to Step 2, otherwise continue to step 5.*
5. *Cluster fathoming* Remove clusters with low number of samples, clusters with no feasible points or clusters with very high objective function values. During this step, the parameters for acceptance or non-acceptance of a cluster must be defined by the user based on the nature of the problem. If locating feasible solutions is difficult, perhaps clusters with lowest feasibility violation should be accepted.
6. *Kept cluster analysis* For any of the remaining clusters, calculate mean ($\mu_c^{(n)}$) and standard deviation ($sd_c^{(n)}$) of the samples in each of the n dimensions. Calculate bounds of each

cluster using the following formula: $\mu_c^{(n)} - \beta_c s d_c^{(n)} \leq x_n \leq \mu_c^{(n)} + \beta_c s d_c^{(n)}$, where β_c is a parameter which should be selected by the user in terms of the tightness of the bounds which are desired for bound updating. In certain cases, it is observed that one of the input variables has zero variability and in this case it is kept constant, leading to an optimization problem with fewer dimensions.

Once the convergence criteria are met, all of the existing samples have been assigned into clusters. At this point, cluster bounds are combined to form a region with bounds which enclose all of the kept clusters. This is a step which is found to be useful in the application tested in this work and it reduces the problem to two repetitions of the proposed global optimization framework, which we will define as: Stage 1 (constrained grey-box optimization within the entire box-constrained region) and Stage 2 (constrained grey-box optimization within the redefined bounds after clustering and cluster bounds combination). There is a possibility for parallel focus in each of the clusters obtained in Step 6, however, this approach has not been explored in this work.

5 Computational studies

The proposed global optimization framework is tested on a set of different instances of the motivating example of the PSA case study. This allows us to form a large number of diverse problems of different complexity by modifying the zeolite which is used for the CO_2 capture inside the adsorption column. It has been found that different materials (i.e., zeolites) exhibit different performance in terms of their total cost, and more importantly in terms of their feasibility [47]. For certain materials the feasible space is very limited- compared to the total investigated operating space, which makes the case studies far more challenging. In addition, different materials lead to simulations of different computational complexity ranging from simulations of 3 min to simulations of 15 min for one computation. In this work, we consider eleven zeolites, namely 13X, AHT, MVY, WEI, ABW, ITW, LTF, NAB, OFF, TON and VNI, which are a collection of popular and promising materials for capture as screened by the most recent literature [40,80]. 13X is the most popular zeolite for CO_2 capture [14,81,82], while AHT, MVY and WEI are the top zeolites for CO_2 capture as screened by the most recent literature [40,80]. Hasan et al. [40,80] identified AHT and MVY to be the top two zeolites for CO_2 capture based on cost, using a hierarchical and computational framework that effectively combines material selection and process optimization. Their top ten materials included WEI, which was also listed as the material requiring the least parasitic energy for CO_2 capture in a separate study by Lin et al. [80].

Out of these zeolites, 13X, AHT, MVY and WEI are selected to perform a series of tests for the efficiency of various steps of the proposed approach and for answering the main questions posed at the beginning of the manuscript. These are: (a) identify the effect of the initial sampling strategy on the final quality and variability of the obtained solutions, (b) reveal the effect of performing global optimization of the parameter estimation of the surrogate functions and the global optimization of the overall constrained surrogate model versus local optimization for different surrogate methods on the quality of the final solution and the overall computational cost, (c) validate the expected importance of clustering and bound updating and finally, (d) test the efficiency of employing a diverse set of surrogate models for the formulation of the surrogate approximation model (P3). All of the runs are performed as single-thread jobs on a linux workstation containing four Intel Core2 2.83 GHz processors.

5.1 Effect of initial sampling strategy on the variability and quality of final solution

In order to assess the effect of the initial sampling on the performance of the proposed framework three different strategies are compared. We set $N_{large} = 5000$ reduced order samples for SS2 and SS3, out of which 70 samples must be selected. More specifically:

(A) *Strategy 1 (SS1)* 70 point Latin Hypercube Design using the function *lhdesign* of MATLAB. This strategy is the most commonly used approach in the literature, assuming that any model is a black-box.
(B) *Strategy 2 (SS2)* 5000 point Latin Hypercube Design using the function *lhdesign* of MATLAB, performance of fast simulation for 5000 points and naive ranking and selection of 70 samples with highest purity and recovery values.
(C) *Strategy 3 (SS3)* 5000 point Latin Hypercube Design using the function *lhdesign* of MATLAB, performance of fast simulations for 5000 points and selection of 70 samples using OSCAR [73], having a mixed objective of the x-space diversity, minimum cost and maximum purity and recovery.

In order to analyze the effects of the above three strategies, 10 repetitions are performed by using a different initial LHD design, for each of the four case studies. In order to eliminate the effects of other parameters of the proposed methodology, general quadratic surrogate functions are used to approximate the cost, purity and recovery. The parameters of the general quadratic functions were identified by solving the parameter estimation problems to global optimality using GLOMIQO [66]. The formed constrained surrogate optimization model (QCP) was also solved to global optimality using GLOMIQO [66]. The mean and standard deviation of the results of the 10 repetitions, the best obtained solution, as well as the average computational cost for each of the zeolites are shown in Table 1.

A common observation of all the case studies, is that SS1 demonstrates inferior performance, in terms of both consistency and quality of solution. Especially for the case of 13X, which is the zeolite for which satisfying feasibility is the most challenging out of the remaining cases, this sampling approach performs poorly. In fact, in three out of the ten runs, the method failed to find a feasible solution overall. As a conclusion, the incorporation of

Table 1 Performance of different sampling strategies

Zeolite	Sampling strategy	Best cost	Average cost	SD	Average CPU time (h)
AHT	1	21.21	23.50	1.9	7.28
AHT	2	21.01	21.90	0.66	7.40
AHT	3	20.91	21.02	0.56	8.34
MVY	1	21.59	22.96	1.54	6.91
MVY	2	20.75	21.38	0.37	6.92
MVY	3	20.72	21.21	0.36	8.30
WEI	1	21.94	23.87	2.13	6.77
WEI	2	21.57	22.74	1.48	7.87
WEI	3	21.39	22.57	0.68	8.97
13X	1	30.57[a]	66.63[a]	59.79[a]	34.26[a]
13X	2	28.05[b]	29.16[b]	0.94[b]	34.85[b]
13X	3	27.16	28.83	1.46	37.77

[a] Three out of the 10 runs lead to no feasible solutions
[b] One out of the 10 runs lead to no feasible solutions

prior short-simulation data followed by sampling reduction is very beneficial in reducing the variability of initial sampling and locating feasible solutions. On the other hand, sampling strategies 2 and 3 are more comparable in terms of average performance and standard deviation of the obtained solutions. With an exception of MVY, for all of the remaining case studies, the best solution is identified when using SS3. It is shown that the algorithm is least dependent on the variability of the initial sampling in the case of SS3, demonstrating that the rigorous selection strategy consistently identifies the optimal and balanced subset of points to use as an initial sampling set. Finally, in the most challenging case study of 13X, SS3 is the only one which does not fail to find a solution at any of the performed runs and it is the method which locates the best solution out of all the remaining methods.

From this analysis it can be concluded that SS3 demonstrates a consistently superior performance when satisfying feasibility is demanding, however, its performance is comparable to SS2 in cases where a larger pool of feasible solutions is available. Overall, SS3 is a rigorous method for selecting a subset of samples when certain a priori data is available.

5.2 Effect of local optimization versus global optimization

In this second study, the aim is to investigate (a) the effect of selection of different functional forms for the objective and constraints and (b) the effect of using local optimization versus global optimization for the parameter estimation of the surrogate functions and the overall constrained surrogate optimization model (P2). For each zeolite, the run which led to the best obtained solution out of the previous study is selected, and the same set of initial samples are used to perform the optimization, in order to alleviate any possible variability caused by the initial sampling. Following, three different types of surrogate functions—general quadratic, kriging and signomials—are used to approximate all of the grey-box functions (cost, purity and recovery). In order to be able to distinguish between the effects of the functional form selection, models of the same type are used for all of the three grey-box models for each of the investigated cases. The runs which are referred to as 'local', correspond to an approach where we have used a local optimization solver for the parameter estimation of the surrogate functions, and a local optimization solver to identify multiple local solutions for Problem P3. In contrast, the 'global' runs employ global optimization solvers to solve both the parameter estimation problems and the general constrained surrogate optimization. In terms of solvers used, CONOPT [83] is used as the local solver, while ANTIGONE [67] is the global solver used. The results of this analysis are summarized in Tables 2, 3, 4 and 5.

For these four case studies, signomial functions often failed to find feasible solutions, or led to significantly suboptimal solutions when compared to those obtained when using general quadratic functions and kriging functions. This result points out the importance of selecting an appropriate functional form, and the results for signomial functions are not presented here.

Based on the results of Tables 2, 3, 4 and 5, it becomes evident that there is an effect both in terms of the type of optimization (i.e., local vs. global) and the type of functions used. Specifically, for all of the case studies, when using general quadratic functions the optimal cost is better when using global optimization with a fewer number of required samples for convergence. The only exceptions to this observation are 13X and WEI, where the global quadratic optimization renders the same solution as the local. However, this is achieved with the same or fewer number of function calls. Moreover, there is a clear indication that when starting from exactly the same initial sampling set, using global optimization is meaningful for all cases and surrogate functions. There are only two exceptions, those of MVY and WEI where the global optimization run of kriging does not locate a better solution when compared to the local runs, because it converges at a lower number of samples. The superiority of the

Table 2 Local versus global optimization for quadratic and kriging models for AHT

Sampling strategy	Model	Optimization	Cost	Samples	CPU (h)
SS1	Quadratic	Local	24.66	470	6.93
SS1	Quadratic	Global	21.21	577	10.02
SS1	Kriging	Local	22.45	578	11.12
SS1	Kriging	Global	22.06	608	12.31
SS2	Quadratic	Local	23.51	345	4.51
SS2	Quadratic	Global	21.01	523	7.59
SS2	Kriging	Local	21.00	638	11.00
SS2	Kriging	Global	20.91	636	11.54
SS3	Quadratic	Local	21.37	254	5.38
SS3	Quadratic	Global	20.91	303	8.34
SS3	Kriging	Local	20.93	624	11.15
SS3	Kriging	Global	20.65	631	12.44

Table 3 Local versus global optimization for quadratic and kriging models for MVY

Zeolite	Model	Optimization	Cost	Samples	CPU (h)
SS1	Quadratic	Local	22.82	316	5.78
SS1	Quadratic	Global	21.59	334	5.81
SS1	Kriging	Local	21.15	627	10.63
SS1	Kriging	Global	20.65	544	12.06
SS2	Quadratic	Local	21.79	332	5.89
SS2	Quadratic	Global	20.75	333	6.19
SS2	Kriging	Local	21.23	635	10.11
SS2	Kriging	Global	20.19	669	12.74
SS3	Quadratic	Local	21.88	395	6.21
SS3	Quadratic	Global	20.72	354	8.30
SS3	Kriging	Local	20.48	637	12.90
SS3	Kriging	Global	20.58	549	9.82

Table 4 Local versus global optimization for quadratic, signomial and kriging models for WEI

Sampling strategy	Model	Optimization	Cost	Samples	CPU (h)
SS1	Quadratic	Local	23.73	428	6.51
SS1	Quadratic	Global	23.11	411	6.74
SS1	Kriging	Local	21.38	627	11.11
SS1	Kriging	Global	20.99	576	10.20
SS2	Quadratic	Local	23.95	271	4.50
SS2	Quadratic	Global	23.95	271	4.55
SS2	Kriging	Local	20.86	614	10.83
SS2	Kriging	Global	20.90	610	12.13
SS3	Quadratic	Local	22.30	468	6.78
SS3	Quadratic	Global	21.39	411	8.11
SS3	Kriging	Local	21.36	559	8.50
SS3	Kriging	Global	21.23	619	11.11

Table 5 Local versus global optimization for quadratic and kriging models for 13X

Zeolite	Surrogate model type	Optimization	Cost	Samples	CPU (h)
SS1	Quadratic	Local	31.75	231	3.63
SS1	Quadratic	Global	31.36	231	4.043
SS1	Kriging	Local	29.60	567	12.21
SS1	Kriging	Global	28.66	567	9.34
SS2	Quadratic	Local	27.77	173	3.45
SS2	Quadratic	Global	27.70	173	3.55
SS2	Kriging	Local	28.05	283	5.78
SS2	Kriging	Global	27.60	283	5.66
SS3	Quadratic	Local	27.16	314	4.92
SS3	Quadratic	Global	27.16	254	18.81
SS3	Kriging	Local	27.36	561	9.80
SS3	Kriging	Global	26.85	588	11.94

global optimization approach is associated with an increase in the total CPU time required to reach the final solution. Based on the results, however, it can be verified that this increase in CPU time is meaningful due to the attainment of better solutions. More interestingly, there are several instances where the global optimization runs render improved solutions, when using the same or less amount of function calls to the expensive simulation. Consequently, this creates a trade-off since the computational cost associated with global optimization may become negligible as the cost of the simulation increases.

Finally, comparing the performance of the proposed framework when using different types of surrogate functions, it is clear that kriging performs better for all of the case studies. On the other hand, kriging does require an increased number of samples and computational time to reach convergence. This fact can be explained due to the nature of kriging, which is purely interpolating and this usually leads to multiple local optima in each iteration. The global optimization using kriging functions is associated with high computational cost, since both the parameter estimation and the optimization of the overall constrained surrogate model (P2) are large non-convex problems with multiple exponential terms.

Comparison of different surrogate functions is useful in order to demonstrate the importance of selecting the best possible function to approximate each of the grey-box functions. Ideally, selecting the appropriate type of surrogate function for each of the unknown correlations individually, coupled with global optimization of the parameters and the surrogate formulation (P2), would be the most powerful approach.

Finally, based on the results of this study, similar conclusions are drawn for the effect of the sampling strategy. For all of the zeolites, SS1 exhibits the worse performance in terms of the best solution obtained, while sampling strategies 2 and 3 are competing in terms of the best solution for the different case studies.

5.3 Importance of clustering and bounds refinement

The performance of clustering depends on the location and the quantity of local solutions obtained throughout the iterations. Specifically for this problem $2n$ (14) local problems and one global problem are solved in each iteration, leading to multiple new design points that are sampled in the next iteration. Out of these 15 total optimization problems, we observe that an average of 2–4 new distinct solutions are obtained when using quadratic functions, while

8–11 new solutions are obtained when using kriging functions. The diversity of solutions when using kriging leads to overall increased sampling requirements, but also improved overall performance. Lastly multiplicity of local solutions when using kriging can be explained by the increased complexity of kriging based surrogate models when compared to the smoother quadratic models.

In order to investigate the significance of bound refinement and iterative nature of the proposed framework within a smaller subregion, the best obtained solution at the end of the first stage is compared to the final obtained solution for all the test problems reported in the previous section. A typical performance of the algorithm during the two stages is shown in Fig. 5. This example is for zeolite 13X, using general quadratic functions, sampling strategy 3 and global optimization. In the first stage of the algorithm, no feasible solutions
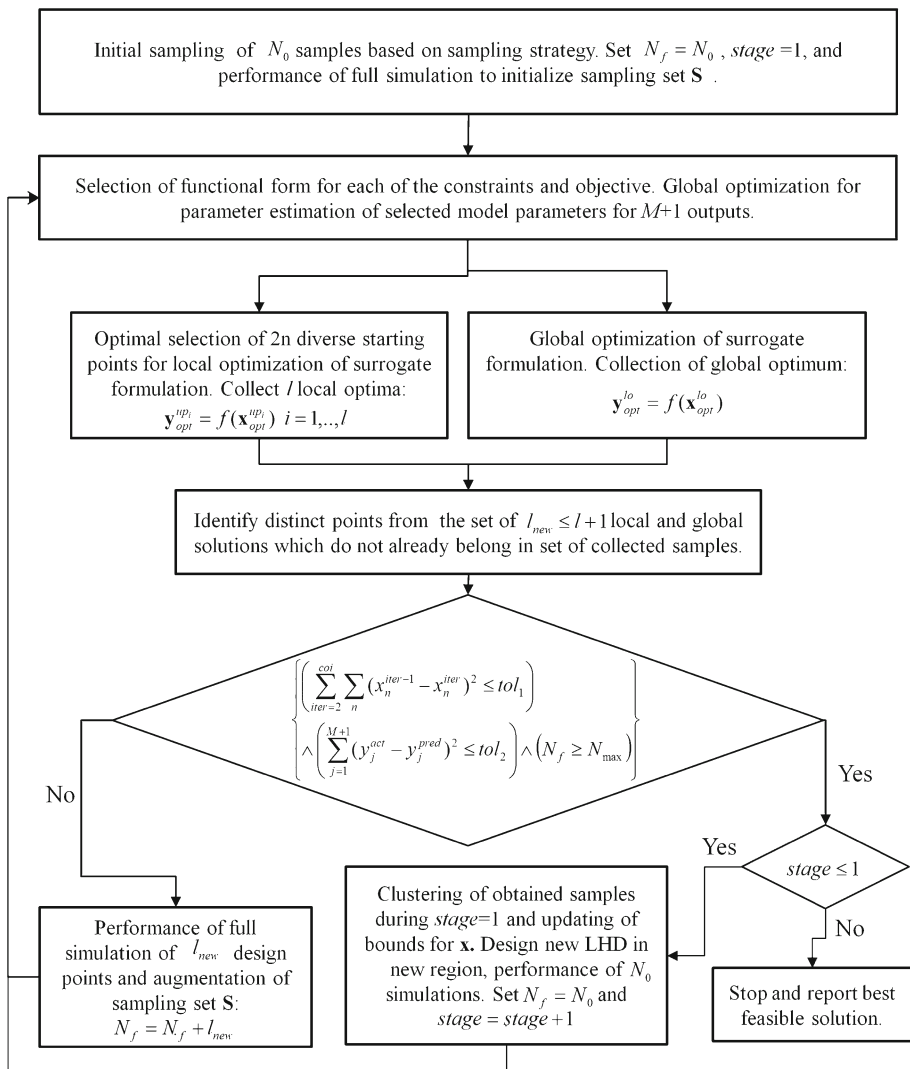

Fig. 4 Algorithmic steps of constrained global optimization for grey-box models

Table 6 Percentage of improvement between first stage and second stage

Zeolite	Model	Average improvement (%)
AHT	Quadratic	15
AHT	Kriging	12
MVY	Quadratic	25
MVY	Kriging	11
WEI	Quadratic	29
WEI	Kriging	13
13X	Quadratic	50
13X	Kriging	16

are obtained until iteration 10. However, there is no improvement in the incumbent solution for the following iterations, and thus the first stage of the algorithm converges and the bounds are updated based on the clustering algorithm. Specifically, the bounds of variables x_1 and x_3 have been significantly reduced, while the bounds for variables x_4 and x_5 remain unchanged, and the remaining variable bounds are moderately altered. This refinement has a significant effect on the incumbent solution, which is reduced by 19 % when compared to the result obtained during the first stage.

A collective average percentage of improvement for each zeolite and methodology is summarized in Table 6. Average improvement is reported across the different sampling strategies and optimization methods since it is found that these aspects have no significant correlation to the performance of the second stage. In the case where no feasible solution is obtained during the first stage, the percentage of improvement is considered to be 100 %. Results for different surrogate functions are reported since it was found that the second stage is important for the general quadratic case, than for the kriging functions. This result can be explained by the fact that general quadratic functions are less sensitive to change when new samples are added during the first stage during their non-interpolating form. However, once the bounds are updated, the general quadratic functions change drastically and lead to improved solutions. Overall, there is improvement observed during the second stage, which is attributed to the fact that when the search is limited within a smaller region, the surrogate functions are significantly updated to represent the samples, and new directions for optimization are generated. This result is also a verification of the efficiency of the clustering approach, since if the bounds were not updated efficiently there would be a great risk of excluding promising regions.

5.4 Comparison of proposed method with publically available solvers

The performance of the developed method is compared with the performance of the NOMAD software, an in-house implementation of the EGO algorithm [30] for constrained problems following an extension proposed by Sasena et al. [47], the ssmGO algorithm [5,84] and a version of COBYLA [61] implemented in [85]. The implementation of the constrained EGO uses kriging for approximating the cost, purity and recovery of the problem, while new sampling locations are identified by maximizing the Expected Improvement function [30] subject to the kriging-based constraints. The only differentiation to solving formulation P3, is the fact that the minimization of cost as an objective, is replaced by the maximization

of the expected improvement function. The only commercial version of constrained EGO can be found in TOMLAB [86], however, for the types of problems with expensive grey-box constraints, the only way for their incorporation is the augmentation of the objective by the summation of the constraint violation as a penalty term, which does not guarantee satisfaction of the grey-box constraints. SsmGO is a MATLAB based scatter search algorithm originally developed for bioprocess optimization. It is a population based meta-heuristic approach which follows and circulates amongst a series of steps; diversification generation; improvement; reference set update; subset generation and solution combination. Constraints are handled using a static penalty function. NOMAD is a mesh adaptive direct search algorithm which uses surrogate models to assist the direct search, while COBYLA uses linear approximations of both the objective and constraints in a trust-region framework.

We perform 10 runs for each of the above algorithms, in order to compare the average performance, consistency, best solution, sampling requirements and CPU time. NOMAD and COBYLA are local search optimization methods, which require a single initial point in order to proceed, while ssmGO and conEGO develop their own space filling design as part of their algorithms. Random initial space filling designs are generated for the global-search methods (ssmGO, NOMAD and the proposed methodology), while random single initial points are generated for the local-search methods (COBYLA and NOMAD). It should be mentioned that the local-search methods are expected to have an improved and reliable performance when started at a good initial point. This point can either be available based on prior knowledge or it can be obtained by an initial "global search" which would increase the sampling requirements of the method significantly. We have not performed this for this comparative analysis which aims to show the average performance of the methods when they are randomly initialized. We perform this analysis on the full set of zeolites which have the best performance for CO_2 capture identified in [40]. The proposed framework for general constrained grey-box optimization is run globally, using SS3, and kriging functions are used for approximating cost, purity and recovery. In Tables 7, 8, 9 and 10, the best cost, mean cost, standard deviation of cost, average number of samples and average CPU time are reported for each of the algorithms.

The results in Table 7 show that ssmGO has a consistent performance, however, it requires a significant number of samples, in order to report the final solution. In two of the cases, LTF

Table 7 Performance of constrained ssmGO	Zeolite	Best cost	Average cost	SD cost	Average samples	Average CPU (h)
	ABW	25.76	26.66	0.80	1490	25.94
	AHT	20.75	21.77	0.85	1083	18.33
	ITW	27.17	29.22	1.70	1088	18.65
	LTF[a]	30.78	35.10	4.60	4890	59.39
	MVY	20.75	21.55	0.74	1069	19.55
	NAB	22.66	23.18	0.48	1066	18.85
	OFF	26.86	30.67	3.02	1868	66.16
	TON	25.58	27.19	1.40	1595	28.16
	VNI	26.68	33.76	5.83	1146	40.12
	WEI	21.30	21.95	0.58	1075	18.48
[a] Average of 6 feasible runs [b] Average of 3 feasible runs	13X[b]	27.98	29.28	1.84	4883	75.29

Table 8 Performance of NOMAD

Zeolite	Best cost	Average cost	SD cost	Average samples	Average CPU (h)
ABW	29.41	30.94	2.17	700	25.79
AHT	20.64	22.48	2.60	700	27.82
ITW	24.87	26.34	1.86	700	28.59
LTF	30.81	34.23	3.87	700	32.81
MVY	20.55	21.08	0.52	700	26.68
NAB[a]	22.25	23.07	1.16	700	25.31
OFF	26.44	28.17	1.50	700	32.67
TON	27.69	28.42	0.67	700	28.36
VNI	26.65	30.41	4.85	700	30.49
WEI	20.60	21.08	0.48	700	23.21
13X[b]	26.71	29.71	3.08	700	108.61

[a] Average of 8 feasible runs
[b] Average of 4 feasible runs

Table 9 Performance of constrained conEGO

Zeolite	Best cost	Average cost	SD cost	Average samples	Average CPU (h)
ABW	24.05	25.1	1.05	149	5.59
AHT	20.98	21.84	0.9	268	6.93
ITW	25.01	25.55	0.48	177	4.92
LTF	29.3	34.61	8.61	132	4.73
MVY	20.67	21.71	1.29	491	12.98
NAB	21.97	23.07	0.99	495	12.69
OFF	26.56	27.89	1.44	219	5.51
TON	25.27	26.37	1.28	208	4.46
VNI	26.8	27.87	0.78	222	5.09
WEI	21.11	21.66	0.54	421	9.33
13X[a]	26.17	28.55	2.3	172	10.07

[a] Average of 8 runs

and 13X ssmGO did not locate feasible solutions for all of the 10 runs. The performance of NOMAD is reported in Table 8. It is observed that the algorithm reaches the maximum number of samples in all cases, while it has a relatively high variability in performance. This can be attributed to the random selection of the initial point for each of the 10 runs. In two of the cases, NAB and 13X NOMAD was not able to locate a feasible solution for two and six of the runs respectively.

Table 9 summarizes the performance of the constrained EGO (conEGO) implementation. This algorithm performs well in locating good solutions with a small number of samples and total CPU times. However, in certain cases the standard deviation of the optimal solution found for ten random runs is significantly high. This variability may be a result of the random initial LHD sample, or the local optimization of the multimodal EI criterion. The algorithm fails to find a solution for two out of the ten runs for the hardest case study of 13X.

Table 10 Performance of the proposed framework for constrained global optimization for grey-box models

Zeolite	Best Cost	Average cost	SD cost	Average samples	Average CPU (h)
ABW	23.11	24.11	0.6	513	11.42
AHT	20.6	20.92	0.29	533	11.63
ITW	24.7	25.37	0.31	506	12.00
LTF	28.11	28.68	0.47	547	11.09
MVY	20.19	20.53	0.27	534	12.55
NAB	21.67	22.16	0.23	527	13.03
OFF	26.2	26.79	0.33	374	12.39
TON	25.11	25.47	0.29	521	13.21
VNI	26.24	26.64	0.63	443	11.30
WEI	20.59	21.19	0.3	422	15.59
13X	25.96	26.82	0.64	336	29.14

Finally, the performance of the proposed framework is provided in Table 10. It is shown that overall the algorithm results in improved solutions with increased consistency. The sampling requirements are lower than ssmGO and NOMAD, while the algorithm requires more samples than the conEGO implementation for convergence. The increased CPU time is attributed to the sampling requirements, the global optimization components, and the initial reduced order sampling procedure. However, this cost does lead to improvements in the final solution and consistency in the performance of the method. Specifically for the case of 13X, which is the case study with the highest computational cost and reduced feasible space, the pay-off for employing the proposed methodology becomes evident since the method always manages to find a feasible solution, while the CPU time is less when compared to that of other methods which require more samples. Results for COBYLA are not presented here, since it was found to have significant difficulty locating feasible solutions for all case studies. The average performance and the variance of each method for all zeolites are shown in Fig. 6, where the edges of the box plot represent the 25th and 75th percentiles of the data, the middle red mark corresponds to the median and the whiskers represent the minimum and maximum cost values obtained from each method. From Fig. 6 it becomes evident that the proposed framework for constrained global optimization of grey-box models has an overall consistent and reliable performance.

6 Conclusions

A method for general constrained grey-box global optimization is proposed which uses multiple novel features regarding sampling designs and sampling selection, global optimization for the parameter estimation of various surrogate functions, global optimization of the overall constrained grey-box model, and clustering for updating variable bounds. The methodology is developed for expensive simulations which cannot be directly used for optimization, such as large NAPDE systems. The motivating example which is optimized in this work is a pressure swing adsorption system for CO_2 capture. The proposed framework is based on identifying appropriate surrogate functions for any unknown constraint functions and the objective function, which are then used to form an overall constrained grey-box model which
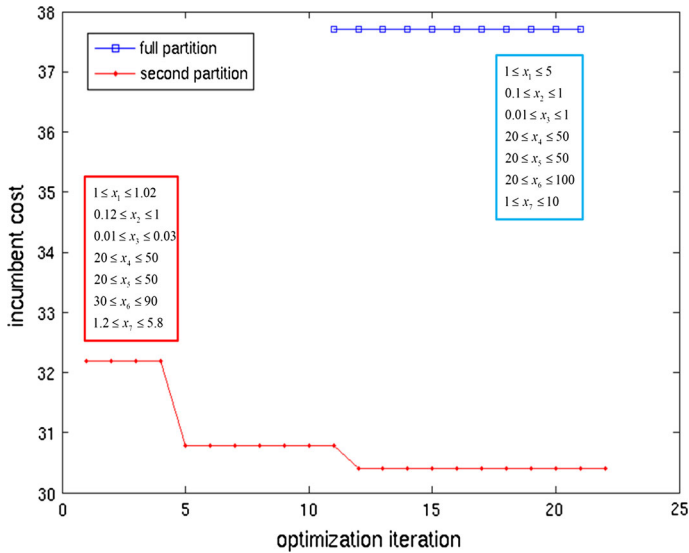
Fig. 5 Incumbent solutions for 13X, sampling strategy 3, quadratic, global optimization for stage 1 and stage 2

is optimized in an iterative framework. There is no complex search criterion introduced in this work, while diversity of sampling is achieved by collecting local solutions from the global optimization framework.

We have demonstrated that the variability in the initial sampling design can have significant impact on the consistency and quality of the optimal solution and we propose a sampling strategy which increases the consistency of the framework in obtaining improved solutions. We achieve this through a two-stage sampling approach which initially requires data obtained from a faster reduced-order NAPDE model and then uses Mixed Integer Linear optimization for selecting a reduced sampling set which is balanced throughout promising regions of the decision variable space.

The proposed framework has the capability of using different surrogate functions in order to approximate each of the constraints and objective of the original model, such as general quadratic functions, kriging functions and signomial functions. We have shown that the selection of the surrogate function plays an important role towards the performance of the method. Specifically kriging functions lead to consistently improved solutions for the case studies presented in this work. Finally, we have studied the importance of using deterministic global optimization for the parameter estimation of the surrogate functions, as well as the optimization of the overall constrained grey-box model. The results presented in this work illustrate that solutions with a lower objective function value can be obtained when using global optimization as opposed to local optimization. In certain cases, improved solutions are obtained with fewer function calls to the expensive simulation. Finally, we compare the performance of the proposed method with four available algorithms for constrained derivative-free optimization. It is demonstrated that the proposed framework is promising in terms of consistency and value of the optimal solution.

The performance of the proposed methodology has been tested on several instances of the NAPDE system for CO_2 capture which have different characteristics and levels of complexity. The consideration of various types of surrogate functions according to optimal fitting criteria,
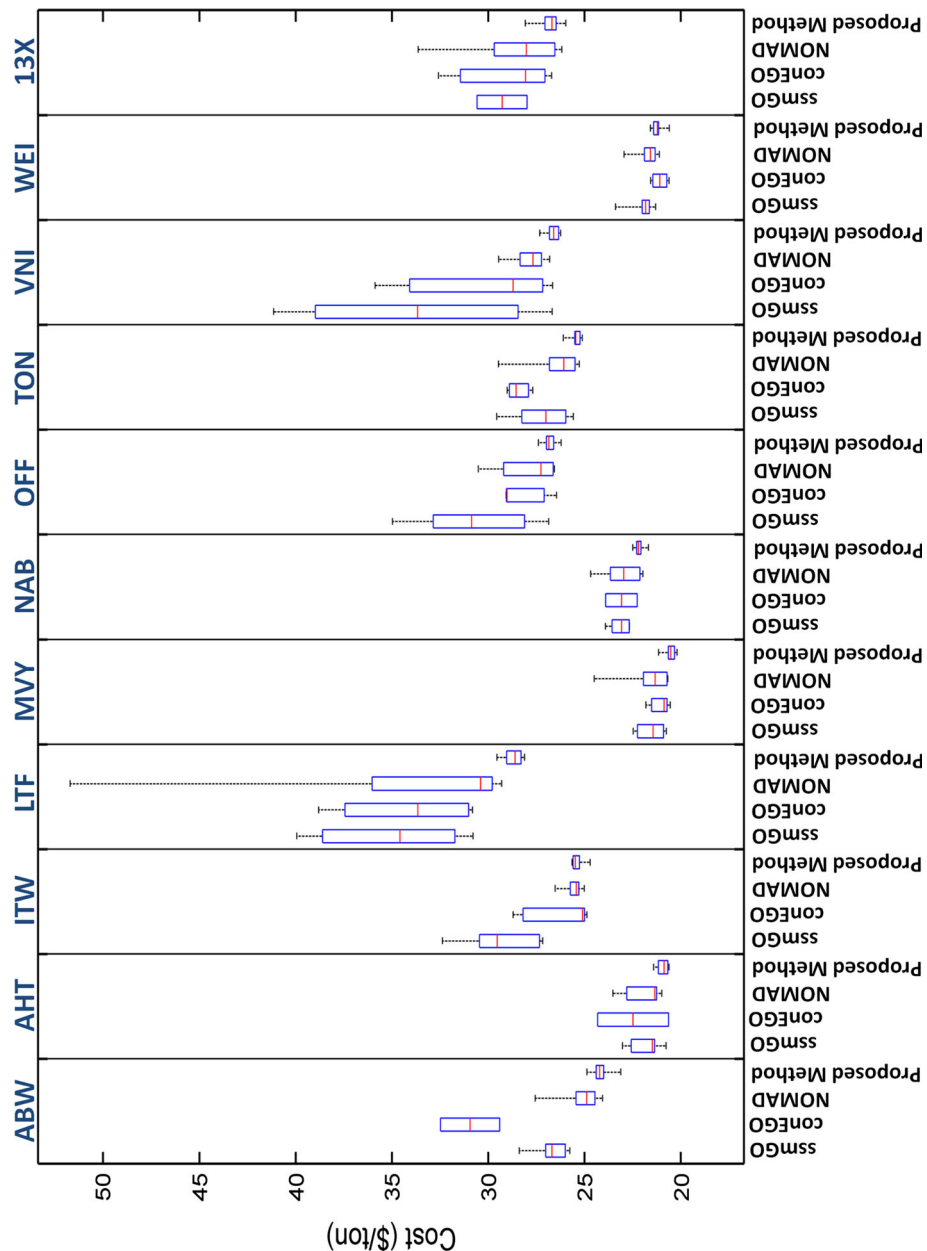
Fig. 6 Comparison of average optimal cost (out of 10 runs) for ssmGO, conEGO, NOMAD and constrained grey-box global optimization algorithm for all zeolites

and the ability of the method to handle any number of known and unknown constraints make it applicable to any problem which follows the grey-box formulation described in this work. We expect the method to have a wide range of applicability in many different scientific fields. However, the aspect of sampling reduction which is one of the reasons we can achieve improved performance, can only be applied only if a reduced-order form of the model is

available. We plan to further generalize our method and perform thorough testing on a large set of constrained optimization problems with increased dimensionality and increased number of constraints.

Appendix 1: Model equations

See Tables 11 and 12.

Table 11 NAPDE model equations for PSA process

Description	Equation	
CO_2 mole fraction in gas phase (Mass balance)	$$\frac{\partial y_{CO_2}}{\partial t} = \frac{1}{Pe}\left(\frac{\partial^2 y_{CO_2}}{\partial Z^2} + \frac{1}{P}\frac{\partial y_{CO_2}}{\partial Z}\frac{\partial P}{\partial Z} - \frac{1}{T}\frac{\partial y_{CO_2}}{\partial Z}\frac{\partial \overline{T}}{\partial Z}\right)$$ $$- \overline{u}_z\frac{\partial y_{CO_2}}{\partial Z} + \frac{\psi\overline{T}}{P}\left((y_{CO_2}-1)\frac{\partial x_{CO_2}}{\partial \tau} + y_{CO_2}\frac{\partial x_{N_2}}{\partial \tau}\right)$$ where $\psi = \left(\frac{1-\varepsilon}{\varepsilon}\frac{RT_0 q_s}{P_{ads}}\right)$	(A1)
Component balance	$y_{CO_2} + y_{N_2} = 1$	(A2)
Overall mass balance	$$\frac{\partial \overline{P}}{\partial \tau} - \frac{\overline{P}}{\overline{T}}\frac{\partial \overline{T}}{\partial \tau} = \left(-\overline{P}\frac{\partial \overline{u}_z}{\partial Z} - \overline{u}_z\frac{\partial \overline{P}}{\partial Z} + \frac{\overline{P}}{\overline{T}}\overline{u}_z\frac{\partial \overline{T}}{\partial Z}\right) - \overline{T}\psi\frac{\partial}{\partial \tau}\left(\sum_i x_i\right)$$	(A3)
Heat transfer across column wall	$$\frac{\partial \overline{T}_w}{\partial \tau} = \pi_1\frac{\partial^2 \overline{T}_w}{\partial Z^2} + \pi_2\left(\overline{T}-\overline{T}_w\right) - \pi_3\left(\overline{T}_w-\overline{T}_a\right)$$ where $\pi_1 = \frac{K_w}{\rho_w C_{pw} u_0 L}$, $\pi_2 = \frac{2r_1 h_{in} L}{\rho_w C_{pw} u_0\left(r_0^2 - r_{in}^2\right)}$, $\pi_3 = \frac{2r_0 h_0 L}{\rho_w C_{pw} u_0\left(r_0^2 - r_{in}^2\right)}$	(A4)
Energy balance	$$\frac{\partial \overline{T}}{\partial \tau} = \pi_4\frac{\partial^2 \overline{T}}{\partial Z^2} - \pi_5\overline{u}_z\frac{\partial \overline{T}}{\partial Z} + \sum_{i=1}^{N}\left(\pi_{6i}+\pi_7\overline{T}\right)\frac{\partial x_i}{\partial \tau} - \pi_8\left(\overline{T}-\overline{T}_w\right)$$ where $$\pi_4 = \frac{K_z}{\varepsilon\left(\rho_g C_{pg}+\frac{1-\varepsilon}{\varepsilon}(\rho_s C_{ps}+C_{p\alpha}q_s)\right)u_0 L}, \quad \pi_5 = \frac{\rho_g C_{pg}}{\left(\rho_g C_{pg}+\frac{1-\varepsilon}{\varepsilon}(\rho_s C_{ps}+C_{p\alpha}q_s)\right)},$$ $$\pi_{6i} = \frac{\frac{1-\varepsilon}{\varepsilon}(C_{pg}-C_{p\alpha i})T_0 q_s}{T_0\left(\rho_g C_{pg}+\frac{1-\varepsilon}{\varepsilon}(\rho_s C_{ps}+C_{p\alpha}q_s)\right)u_0 L}, \quad \Delta H_i = \Delta U_i - RT,$$ $$\pi_7 = \frac{\frac{1-\varepsilon}{\varepsilon}(C_{pg}-C_{p\alpha i})T_0 q_s}{T_0\left(\rho_g C_{pg}+\frac{1-\varepsilon}{\varepsilon}(\rho_s C_{ps}+C_{p\alpha}q_s)\right)}, \quad \pi_8 = \frac{2h_{in}L}{\varepsilon T_i u_0\left(\rho_g C_{pg}+\frac{1-\varepsilon}{\varepsilon}(\rho_s C_{ps}+C_{p\alpha}q_s)\right)}$$	(A5)
Linear driving force	$$\frac{\partial x_i}{\partial \tau} = \frac{k_i L}{u_0}\left(x_i^* - x_i\right) \quad i = \{CO_2, N_2\}$$	(A6)
Dual-site Langmuir model with temperature dependent parameters	$$x_i^* = \sum_{s=1}^{2}\frac{\left(\frac{q_{i,s}^{sat}}{q_s}\right)b_{i,s}c_i}{1+\sum_i b_{i,s}c_i} \quad i = \{CO_2, N_2\}$$ where $b_{i,s} = \frac{b_{i,s}^0 e^{-\frac{\Delta U_{i,s}}{RT}}}{RT} \quad i = \{CO_2, N_2\}, s = 1,...,2$	(A7)

Table 12 PSA process constrains, operating constraints and cost objective

Description	Equation	
Pressure constraints	$P_{ads} \geq P_{bd} \geq P_{evac}$	(A8)
Recovery constraint	$Re = \dfrac{\int_0^{t_{pr}} \bar{u}_z y_{CO_2} \frac{P}{RT} dt + \int_0^{t_{ads}} u_z y_{CO_2} \frac{P}{RT} dt}{\int_0^{t_{evac}} \bar{u}_z y_{CO_2} \frac{P}{RT} dt} \geq 0.9$	(A9)
Purity constraint	$Pu = \dfrac{\int_0^{t_{evac}} \bar{u}_z y_{CO_2} \frac{P}{RT} dt}{\int_0^{t_{evac}} \bar{u}_z y_{CO_2} \frac{P}{RT} dt + \int_0^{t_{evac}} \bar{u}_z y_{N_2} \frac{P}{RT} dt} \geq 0.9$	(A10)
Total/Maximum time	$t_c = t_{pr} + t_{ads} + t_{bd} + t_{evac}, \; t_{max} = \max\left[t_{pr}, t_{ads}, t_{bd}, t_{evac}\right]$	(A11)
Number of columns	$N = \frac{t_c}{t_{max}}$	(A12)
Column diameter	$D = \sqrt{\frac{4}{\pi} \times \left(\frac{P_{atm}}{P_{ads}}\right)^{\beta_c} \times \frac{T_0}{273.15} \times 22.4 \times \left(\frac{F \times t_c}{u_0 \times t_{ads} \times N}\right)}$	(A13)
	$TIC = \alpha_c \left(\bar{W}_{ads}\right)^{\beta_c} + \alpha_v \left(\bar{W}_{bd}\right)^{\beta_v} + \alpha_v \left(\bar{W}_{evac}\right)^{\beta_v} + \alpha_c \left(\bar{W}_{seq}\right)^{\beta_c} + \alpha_{he} \left(\bar{W}_{cool}\right)^{\beta_{he}} + \sum_n PC_n$	
	where $\bar{W}_j = \frac{W_j}{t_j \times 746} j = \{ads, bd, evac\}$	
	$\bar{W}_{seq} = \frac{6}{\eta} \bar{M}_{evac} \times 8314 \times 746 \times T_{atm} \times \frac{\gamma}{\gamma-1} \left[2.3^{\frac{\gamma-1}{\gamma}} - 1\right]$	
	$\bar{W}_{cool} = 5 \bar{M}_{evac} \times C_{pg} \left[2.3^{\frac{\gamma-1}{\gamma}} - 1\right] \times T_{atm}$	
Total investment cost (TIC)	$W_{ads} = \frac{1}{\eta} \varepsilon \pi D^2 u_0 P_{ads} \frac{\gamma}{\gamma-1} \int_0^{t_{ads}} \left(\bar{u}_z \bar{P}\big|_{Z=0}\right) \left[\left(\frac{P_{ads} \bar{P}\big|_{Z=0}}{P_f}\right)^{\frac{\gamma-1}{\gamma}} - 1\right] dt$	(A14)
	$W_{bd} = \frac{1}{\eta} \varepsilon \pi D^2 u_0 P_{ads} \frac{\gamma}{\gamma-1} \int_0^{t_{bd}} \left(\bar{u}_z \bar{P}\big|_{Z=0}\right) \left[\left(\frac{P_{atm}}{P_{ads} \bar{P}\big|_{Z=0}}\right)^{\frac{\gamma-1}{\gamma}} - 1\right] dt$	
	$W_{evac} = \frac{1}{\eta} \varepsilon \pi D^2 u_0 P_{ads} \frac{\gamma}{\gamma-1} \int_0^{t_{evac}} \left(\bar{u}_z \bar{P}\big|_{Z=0}\right) \left[\left(\frac{P_{atm}}{P_{ads} \bar{P}\big|_{Z=0}}\right)^{\frac{\gamma-1}{\gamma}} - 1\right] dt$	

Table 12 continued

Description	Equation	
Purchase cost	$PC_n = F_M C_{V,n} + C_{PL,n} + V_{P,n} C_{PK} + C_{DR,n}$ where	
	$C_{V,n} = \exp\left\{7.2756 + 0.28255\ln(W_n) + 0.02297\left[\ln(W_n)^2\right]\right\}$ $W_n = \pi(39.37 D_n + t_s)\left[39.37\left(0.8 D_n + L_n\right)\right]t_S\rho$ $C_{PL,n} = 300.9\,(3.281 D_n)^{0.63316}\,(3.281 L_n)^{0.80161}$ $V_{P,n} = \frac{\pi}{4}\,(3.281 D_n)^2\,(3.281 L_n)$ $C_{DR,n} = 3 \times 125 \times \frac{\pi}{4}\,(3.281 D_n)^2$	(A15)
Annual operating cost	$AOC = \left[8000 \times 3600 \sum_j UC_j W_j + 8000 \times 3600 \sum_k UC_k W_k + UCM \times \frac{\pi}{4} D^2 L N \rho_s\right]$	(A16)
Total annual cost (TAC)	$TAC = \phi TIC + AOC$	(A17)

Appendix 2: Nomenclature

Indices

i	Component
n	Packed column

Sets

The set of all components is given as follows: $i \in I = \{CO_2, N_2\}$

Parameters

The following parameters are defined:

F	Feed flow rate
$y_{i,f}$	Mole fraction of component i in the feed gas
T_o	Feed temperature
P_f	Feed flow rate
u_o	Feed velocity
T_a	Outside air temperature
$y_{i,0}$	Initial mole fraction of component i in the bed
P_{atm}	Atmospheric pressure
K_z	Effective heat conductivity
K_w	Thermal heat conductivity of column wall
h_{in}	Heat transfer coefficient inside the column
h_o	Heat transfer coefficient for outside the column
C_{pw}	Specific heat capacity of column wall
C_{pg}	Specific heat capacity of gas mixture
C_{pa}	Specific heat capacity of adsorbed gas in solid
C_{ps}	Specific heat capacity of adsorbent
R	Gas constant
q_s	Saturation capacity
b_i^o	Isotherm parameter for component i
ΔU_i	Heat of adsorption for component i
τ_p	Tortuosity factor
D_M	Molecular diffusivity of $CO_2 - N_2\, mixture$
D_p	Macropore diffusivity, $D_p = \frac{D_M}{\tau_p}$
D_L	Axial dispersion coefficient defined by the correlation: $D_L = 0.7 D_M + 0.5 u_o d_p$
UC	Unit operating cost
d_p	Particle diameter
r_p	Particle radius, $\frac{d_p}{2}$
ρ_w	Density of column
ρ_s	Density of solid particle
μ	Gas viscosity
ε_p	Particle porosity
ε	Bed porosity
ϕ	Annualization factor
α	Unit capital cost
β	Exponent of power cost
k_p	Bed permeability defined by the correlation: $k_p = \frac{d_p^2}{150}\left(\frac{\varepsilon}{1-\varepsilon}\right)^2$

Variables

We consider the variation both in time and along the column length. To denote this, we define:

l Bed length
t Time

The following system variables are defined to describe the system:

$P(l, t)$	Pressure
$T(l, t)$	Temperature
$T_W(l, t)$	Column wall temperature
$u_z(l, t)$	Velocity
$c_i(l, t)$	Concentration of component i in the gas phase
$y_i(l, t)$	Mole fraction of component i in the gas phase
$x_i(l, t)$	Fractional loading of component i in the solid phase
$x_i^*(l, t)$	Equilibrium fractional loading of component i in the solid phase
$k_i(l, t)$	Mass transfer coefficient of component i
$b_i(l, t)$	Temperature dependent parameter for component i
$B_i(l, t)$	An isotherm parameter for component i
$\Delta H_i(l, t)$	Change in internal energy for component i
$\rho_g(l, t)$	Density of gas obtained from the ideal gas law

The decision variables for process optimization are as follows:

N	Number of columns
P_{ads}	Highest pressure of the system
L	Column length
D	Column diameter
P_{bd}	Blowdown pressure
P_{evac}	Evacuation pressure
t_{pr}	Duration of the pressurization step, 20
st_{ads}	Duration of the adsorption step
t_{bd}	Duration of the blowdown step
t_{evac}	Duration of the evacuation step

Variable scaling

The model is described in terms of these dimensionless variables:

$$\overline{P} = \frac{P}{P_{ads}}; \quad \overline{P}_{evac} = \frac{P_{evac}}{P_{ads}}; \quad \overline{P}_{bd} = \frac{P_{bd}}{P_{ads}}; \quad \overline{T} = \frac{T}{T_0}; \quad \overline{T}_w = \frac{T_w}{T_0}; \quad \overline{T}_a = \frac{T_a}{T_0};$$

$$x_i = \frac{\overline{q}_i}{q_s}; \quad x_i^* = \frac{q_i^*}{q_s}; \quad \overline{u}_z = \frac{u_z}{u_0}; \quad Z = \frac{l}{L}; \quad \tau = \frac{t u_o}{L}$$

We also need appropriate boundary and initial conditions to describe the four steps, since they are different for different steps. These are defined subsequently.

Initial and boundary conditions

Pressurization The initial conditions for the first pressurization step are given by the following equation:

$$y_i = y_{i,0}$$
$$\overline{P} = \overline{P}_{evac}$$
$$\overline{T} = 1$$
$$x_i = x_i^*\big|_{y_{i,0}}$$

In a 4-step cyclic process, the initial conditions for each subsequent step are the final conditions of the previous step, so initial conditions are not specified for the remaining steps.

Let $Z = 0^+$ and $Z = 1^-$ be the two ends of an adsorption column where the boundary conditions must be applied. The boundary conditions at $Z = 0^+$ for the pressurization step are as follows:

$$\frac{1}{Pe}\frac{\partial y_i}{\partial Z} = -\overline{u}_z(y_{i,f} - y_i)$$
$$\frac{1}{Pe_H}\frac{\partial \overline{T}}{\partial Z} = -\overline{u}_z(1 - \overline{T})$$
$$\overline{T}_w = \overline{T}_a$$
$$\overline{P} = f(\tau); \quad \overline{P}_{evac} \to 1$$

The boundary conditions at $Z = 1^-$ are:

$$\frac{\partial y_i}{\partial Z} = 0$$
$$\frac{\partial \overline{T}}{\partial Z} = 0$$
$$\overline{T}_w = \overline{T}_a$$
$$\frac{\partial \overline{P}}{\partial Z} = 0$$

Adsorption The boundary conditions at $Z = 0^+$ and $Z = 1^-$ for the adsorption step are the following:

$$\frac{\partial y_i}{\partial Z} = -Pe(y_{i,f} - y_i)$$
$$\frac{1}{Pe_H}\frac{\partial \overline{T}}{\partial Z} = -(1 - \overline{T})$$
$$\overline{T}_w = \overline{T}_a$$
$$\overline{P} = 1 + \frac{\Delta}{P_{ads}}$$
$$\frac{\partial y_i}{\partial Z} = 0$$
$$\frac{\partial \overline{T}}{\partial Z} = 0$$
$$\overline{T}_w = \overline{T}_a$$
$$\overline{P} = 1$$

where $Pe_H = \frac{\varepsilon u_o L \rho_g C_{pg}}{K_z}$

Blowdown The boundary conditions at $Z = 0^+$ and $Z = 1^-$ for the blowdown step are given by:

$$\frac{\partial y_i}{\partial Z} = 0$$

$$\frac{\partial \overline{T}}{\partial Z} = 0$$

$$\overline{T}_w = \overline{T}_a$$

$$\frac{\partial \overline{P}}{\partial Z} = 0$$

$$\frac{\partial y_i}{\partial Z} = 0$$

$$\frac{\partial \overline{T}}{\partial Z} = 0$$

$$\overline{T}_w = \overline{T}_a$$

$$\overline{P} = f(\tau): 1 \rightarrow \overline{P}_{bd}$$

Evacuation The boundary conditions at $Z = 0^+$ and $Z = 1^-$ for the evacuation step are:

$$\frac{\partial y_i}{\partial Z} = 0$$

$$\frac{\partial \overline{T}}{\partial Z} = 0$$

$$\overline{T}_w = \overline{T}_a$$

$$\frac{\partial \overline{P}}{\partial Z} = 0$$

$$\frac{\partial y_i}{\partial Z} = 0$$

$$\frac{\partial \overline{T}}{\partial Z} = 0$$

$$\overline{T}_w = \overline{T}_a$$

$$\overline{P} = f(\tau) : \overline{P}_{bd} \rightarrow \overline{P}_{evac}$$

References

1. Audet, C., Bechard, V., Chaouki, J.: Spent potliner treatment process optimization using a MADS algorithm. Optim. Eng. **9**(2), 143–160 (2008)
2. Bartholomew-Biggs, M.C., Parkhurst, S.C., Wilson, S.P.: Using DIRECT to solve an aircraft routing problem. Comput. Optim. Appl. **21**(3), 311–323 (2002)
3. Boukouvala, F., Ierapetritou, M.G.: Surrogate-based optimization of expensive flowsheet modeling for continuous pharmaceutical manufacturing. J. Pharm. Innov. **8**(2), 131–145 (2013)
4. Caballero, J.A., Grossmann, I.E.: An algorithm for the use of surrogate models in modular flowsheet optimization. AIChE J. **54**(10), 2633–2650 (2008)
5. Egea, J.A., Rodriguez-Fernandez, M., Banga, J.R., Marti, R.: Scatter search for chemical and bio-process optimization. J. Glob. Optim. **37**(3), 481–503 (2007)
6. Fahmi, I., Cremaschi, S.: Process synthesis of biodiesel production plant using artificial neural networks as the surrogate models. Comput. Chem. Eng. **46**, 105–123 (2012)
7. Fowler, K.R., Reese, J.P., Kees, C.E., Dennis Jr, J.E., Kelley, C.T., Miller, C.T., Audet, C., Booker, A.J., Couture, G., Darwin, R.W., Farthing, M.W., Finkel, D.E., Gablonsky, J.M., Gray, G., Kolda, T.G.: Com-

parison of derivative-free optimization methods for groundwater supply and hydraulic capture community problems. Adv. Water Resour. **31**(5), 743–757 (2008)

8. Graciano, J.E.A., Roux, G.A.C.L.: Improvements in surrogate models for process synthesis. Application to water network system design. Comput. Chem. Eng. **59**, 197–210 (2013)

9. Hemker, T., Fowler, K., Farthing, M., Stryk, O.: A mixed-integer simulation-based optimization approach with surrogate functions in water resources management. Optim. Eng. **9**(4), 341–360 (2008)

10. Henao, C.A., Maravelias, C.T.: Surrogate-based superstructure optimization framework. AIChE J. **57**(5), 1216–1232 (2011)

11. Kleijnen, J.P.C., van Beers, W., van Nieuwenhuyse, I.: Constrained optimization in expensive simulation: Novel approach. Eur. J. Oper. Res. **202**(1), 164–174 (2010)

12. Wan, X.T., Pekny, J.F., Reklaitis, G.V.: Simulation-based optimization with surrogate models—application to supply chain management. Comput. Chem. Eng. **29**(6), 1317–1328 (2005)

13. Espinet, A., Shoemaker, C., Doughty, C.: Estimation of plume distribution for carbon sequestration using parameter estimation with limited monitoring data. Water Resour. Res. **49**(7), 4442–4464 (2013)

14. Hasan, M.M.F., Baliban, R.C., Elia, J.A., Floudas, C.A.: Modeling, simulation, and pptimization of postcombustion CO_2 capture for variable feed concentration and flow rate. 2. Pressure swing adsorption and vacuum swing adsorption processes. Ind. Eng. Chem. Res. **51**(48), 15665–15682 (2013)

15. Hasan, M.M.F., Baliban, R.C., Elia, J.A., Floudas, C.A.: Modeling, simulation, and optimization of postcombustion CO_2 capture for variable feed concentration and flow rate. 1. Chemical absorption and membrane processes. Ind. Eng. Chem. Res. **51**(48), 15642–15664 (2013)

16. Hasan, M.M.F., Boukouvala, F., First, E.L., Floudas, C.A.: Nationwide, regional, and statewide CO_2 capture, utilization, and sequestration supply chain network optimization. Ind. Eng. Chem. Res. **53**(18), 7489–7506 (2014)

17. Li, S., Feng, L., Benner, P., Seidel-Morgenstern, A.: Using surrogate models for efficient optimization of simulated moving bed chromatography. Comput. Chem. Eng. **67**, 121–132 (2014)

18. Forrester, A.I.J., Sóbester, A., Keane, A.J.: Engineering Design Via Surrogate Modelling—A Practical Guide. Wiley, Chichester (2008)

19. Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimization. MPS-SIAM Series on Optimization, vol. 8. SIAM, Philadelphia (2009)

20. Martelli, E., Amaldi, E.: PGS-COM: a hybrid method for constrained non-smooth black-box optimization problems: brief review, novel algorithm and comparative evaluation. Comput. Chem. Eng. **63**, 108–139 (2014)

21. Rios, L.M., Sahinidis, N.V.: Derivative-free optimization: a review of algorithms and comparison of software implementations. J. Glob. Optim. **56**(3), 1247–1293 (2013)

22. Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by direct search: new perspectives on some classical and modern methods. SIAM Rev. **45**(3), 385–482 (2003)

23. Bjorkman, M., Holmstrom, K.: Global optimization of costly nonconvex functions using radial basis functions. Optim. Eng. **1**(4), 373–397 (2000)

24. Booker, A.J., Dennis, J.E., Frank, P.D., Serafini, D.B., Torczon, V., Trosset, M.W.: A rigorous framework for optimization of expensive functions by surrogates. Struct. Multidiscip. Optim. **17**(1), 1–13 (1999)

25. Boukouvala, F., Ierapetritou, M.G.: Derivative-free optimization for expensive constrained problems using a novel expected improvement objective function. AIChE J. **60**(7), 2462–2474 (2014)

26. Conn, A.R., Le Digabel, S.: Use of quadratic models with mesh-adaptive direct search for constrained black box optimization. Optim. Methods Softw. **28**(1), 139–158 (2013)

27. Forrester, A.I.J., Keane, A.J.: Recent advances in surrogate-based optimization. Prog. Aerosp. Sci. **45**(1), 50–79 (2009)

28. Jakobsson, S., Patriksson, M., Rudholm, J., Wojciechowski, A.: A method for simulation based optimization using radial basis functions. Optim. Engi. **11**(4), 501–532 (2010)

29. Jones, D.R.: A taxonomy of global optimization methods based on response surfaces. J. Global Optim. **21**(4), 345–383 (2001)

30. Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. J. Global Optim. **13**(4), 455–492 (1998)

31. Regis, R.G.: Constrained optimization by radial basis function interpolation for high-dimensional expensive black-box problems with infeasible initial points. Eng. Optim. **46**(2), 218–243 (2014)

32. Regis, R.G., Shoemaker, C.A.: Constrained global optimization of expensive black box functions using radial basis functions. J. Glob. Optim. **31**(1), 153–171 (2005)

33. Yao, W., Chen, X.Q., Huang, Y.Y., van Tooren, M.: A surrogate-based optimization method with RBF neural network enhanced by linear interpolation and hybrid infill strategy. Optim. Methods Softw. **29**(2), 406–429 (2014)

34. Muller, J., Shoemaker, C.A.: Influence ensemble surrogate models and sampling strategy on the solution quality of algorithms for computationally expensive black-box global optimization methods. J. Glob. Optim. **60**(2), 123–144 (2014)
35. Viana, F.A.C., Haftka, R.T., Watson, L.T.: Efficient global optimization algorithm assisted by multiple surrogate techniques. J. Glob. Optim. **56**(2), 669–689 (2013)
36. Davis, E., Ierapetritou, M.: A kriging method for the solution of nonlinear programs with black-box functions. AIChE J. **53**(8), 2001–2012 (2007)
37. Floudas, C.A.: Deterministic Global Optimization, vol. 37. Springer, Berlin (1999)
38. Floudas, C.A., Gounaris, C.E.: A review of recent advances in global optimization. J. Glob. Optim. **45**(1), 3–38 (2009)
39. First, E.L., Hasan, M.M.F., Floudas, C.A.: Discovery of novel zeolites for natural gas purification through combined material screening and process optimization. AICHE J. **60**(5), 1767–1785 (2014)
40. Hasan, M.M.F., First, E.L., Floudas, C.A.: Cost-effective CO_2 capture based on in silico screening of zeolites and process optimization. Phys. Chem. Chem. Phys. **15**(40), 17601–17618 (2013)
41. Abramson, M.: Pattern Search Algorithms for Mixed Variable General Constrained Optimization Problems. Rice University, Houston (2002)
42. Audet, C., Dennis, J.: Mesh adaptive direct search algorithms for constrained optimization. SIAM J. Optim. **17**(1), 188–217 (2006)
43. Audet, C., Dennis Jr, J.E.: A progressive barrier for derivative-free nonlinear programming. SIAM J. Optim. **20**(1), 445–472 (2009)
44. Holmstrom, K., Quttineh, N.-H., Edvall, M.M.: An adaptive radial basis algorithm (ARBF) for expensive black-box mixed-integer constrained global optimization. Optim. Eng. **9**(4), 311–339 (2008)
45. Audet, C., Dennis Jr, J.E.: A pattern search filter method for nonlinear programming without derivatives. SIAM J. Optim. **14**(4), 980–1010 (2004)
46. Parr, J.M., Keane, A.J., Forrester, A.I.J., Holden, C.M.E.: Infill sampling criteria for surrogate-based optimization with constraint handling. Eng. Optim. **44**(10), 1147–1166 (2012)
47. Sasena, M.J., Papalambros, P., Goovaerts, P.: Exploration of metamodeling sampling criteria for constrained global optimization. Eng. Optim. **34**(3), 263–278 (2002)
48. Regis, R.G.: Stochastic radial basis function algorithms for large-scale optimization involving expensive black-box objective and constraint functions. Comput. Oper. Res. **38**(5), 837–853 (2011)
49. Abramson, M., Audet, C., Dennis, J., Digabel, S.: OrthoMADS: a deterministic MADS instance with orthogonal directions. SIAM J. Optim. **20**(2), 948–966 (2009)
50. Audet, C., Bechard, V., Digabel, S.: Nonsmooth optimization through mesh adaptive direct search and variable neighborhood search. J. Glob. Optim. **41**(2), 299–318 (2008)
51. Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by direct search: new perspectives on some classical and modern methods. SIAM Rev. **45**(3), 385–482 (2003)
52. Wild, S.M., Regis, R.G., Shoemaker, C.A.: Orbit: optimization by radial basis function interpolation in trust-regions. SIAM J. Sci. Comput. **30**(6), 3197–3219 (2008)
53. Sankaran, S., Audet, C., Marsden, A.L.: A method for stochastic constrained optimization using derivative-free surrogate pattern search and collocation. J. Comput. Phys. **229**(12), 4664–4682 (2010)
54. Le Digabel, S.: Algorithm 909: NOMAD: nonlinear optimization with the MADS algorithm. ACM Trans. Math. Softw. (TOMS) **37**(4), 1–15 (2011)
55. Oeuvray, R.: Trust-region methods based on radial basis functions with application to biomedical imaging. PhD in Mathematics, Ecole Polytechnique Federale de Lausanne (2005)
56. Powell, M.J.D.: The BOBYQA Algorithm for Bound Constrained Optimization Without Derivatives. Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge (2009)
57. Regis, R.G., Shoemaker, C.A.: Parallel radial basis function methods for the global optimization of expensive functions. Eur. J. Oper. Res. **182**(2), 514–535 (2007)
58. Regis, R.G., Shoemaker, C.A.: A stochastic radial basis function method for the global optimization of expensive functions. INFORMS J. Comput. **19**(4), 497–509 (2007)
59. Regis, R.G., Shoemaker, C.A.: Improved strategies for radial basis function methods for global optimization. J. Glob. Optim. **37**(1), 113–135 (2007)
60. Regis, R.G., Shoemaker, C.A.: A quasi-multistart framework for global optimization of expensive functions using response surface models. J. Glob. Optim. **56**(4), 1719–1753 (2013)
61. Powell, M.J.D.: A direct search optimization method that models the objective and constraint functions by linear interpolation. In: Gomez, S., Hennart, J.-P. (eds.) Advances in Optimization and Numerical Analysis: Mathematics and its Applications, vol. 275. Mathematics and Its Applications, pp. 51–67. Springer, Berlin (1994)

62. Quan, N., Yin, J., Ng, S.H., Lee, L.H.: Simulation optimization via kriging: a sequential search using expected improvement with computing budget constraints. IIE Trans. **45**(7), 763–780 (2013)
63. Torn, A., Zilinskas, A.: Global optimization. In: Lecture Notes in Computer Science, vol. 350. Springer, Berlin (1989)
64. Regis, R.G.: Evolutionary programming for high-dimensional constrained expensive black-box optimization using radial basis functions. IEEE Trans. Evol. Comput. **18**(3), 326–347 (2014). doi:10.1109/TEVC.2013.2262111
65. Misener, R., Floudas, C.A.: Global optimization of mixed-integer models with quadratic and signomial functions: a review. Appl. Comput. Math. **11**, 317–336 (2012)
66. Misener, R., Floudas, C.A.: GloMIQO: global mixed-integer quadratic optimizer. J. Glob. Optim. **57**(1), 3–50 (2013)
67. Misener, R., Floudas, C.: ANTIGONE: algorithms for continuous/integer global optimization of nonlinear equations. J. Glob. Optim. **59**(2–3), 503–526 (2014)
68. Mckay, M.D., Beckman, R.J., Conover, W.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **42**(1), 55–61 (2000)
69. Kahrs, O., Marquardt, W.: The validity domain of hybrid models and its application in process optimization. Chem. Eng. Process. **46**(11), 1054–1066 (2007)
70. Sobieszczanski-Sobieski, J., Haftka, R.T.: Multidisciplinary aerospace design optimization: survey of recent developments. Struct. Optim. **14**(1), 1–23 (1997). doi:10.1007/BF01197554
71. Willcox, K., Peraire, J.: Balanced model reduction via the proper orthogonal decomposition. AIAA J. **40**(11), 2323–2330 (2002)
72. Lucia, D.J., Beran, P.S., Silva, W.A.: Reduced-order modeling: new approaches for computational physics. Prog. Aerosp. Sci. **40**(1–2), 51–117 (2004)
73. Li, Z., Floudas, C.A.: Optimal scenario reduction framework based on distance of uncertainty distribution and output performance: I. Single reduction via mixed integer linear optimization. Comput. Chem. Eng. **70**, 50–65 (2014)
74. Cressie, N.: Statistics for Spatial Data. Wiley Series in Probability and Statistics. Wiley-Interscience, New York (1993)
75. Kleijnen, J.P.C.: Kriging metamodeling in simulation: a review. Eur. J. Oper. Res. **192**(3), 707–716 (2009)
76. Sacks, J., Welch, W.J., Toby, J.M., Wynn, H.P.: Design and analysis of computer experiments. Stat. Sci. **4**(4), 409–423 (1989)
77. Myers, R.H., Montgomery, D.C.: Response Surface Methodology: Process and Product in Optimization Using Designed Experiments. Wiley, New York (1995)
78. Bjork, K.-M., Lindberg, P.O., Westerlund, T.: Some convexifications in global optimization of problems containing signomial terms. Comput. Chem. Eng. **27**(5), 669–679 (2003)
79. Gramacy, R.B., Lee, H.K.H.: Optimization Under Unknown Constraints. University of Cambridge, Cambridge (2010)
80. Lin, L.-C., Berger, A., Martin, R., Kim, J., Swisher, J., Jariwala, K., Rycroft, C., Bhown, A., Deem, M., Haranczyk, M., Smit, B.: In silico screening of carbon-capture materials. Nat. Mater. **11**(7), 633–641 (2012)
81. Siriwardane, R.V., Shen, M.-S., Fisher, E.P., Poston, J.A.: Adsorption of CO_2 on molecular sieves and activated carbon. Energy Fuels **15**(2), 279–284 (2001)
82. Zhang, J., Webley, P.A., Xiao, P.: Effect of process parameters on power requirements of vacuum swing adsorption technology for CO_2 capture from flue gas. Energy Convers. Manag. **49**(2), 346–356 (2008)
83. Drud, A.: CONOPT—a large-scale GRG code. ORSA J. Comput. **6**, 207–216 (1992)
84. Egea, J.A., Martí, R., Banga, J.R.: An evolutionary method for complex-process optimization. Comput. Oper. Res. **37**(2), 315–324 (2010)
85. Johnson, S.G.: The NLopt nonlinear-optimization package. http://ab-initio.mit.edu/nlopt
86. Holmstrom, K., Goran, A.O., Edvall, M.M.: Users Guide for TOMLAB CGO. http://tomopt.com/docs/TOMLAB_CGO.pdf (2008)