



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Multi-objective optimisation using surrogate models for the design of VPSA systems

Citation for published version:

Beck, J, Friedrich, D, Brandani, S & Fraga, ES 2015, 'Multi-objective optimisation using surrogate models for the design of VPSA systems', *Computers and Chemical Engineering*, vol. 82, pp. 318-329.
<https://doi.org/10.1016/j.compchemeng.2015.07.009>

Digital Object Identifier (DOI):

[10.1016/j.compchemeng.2015.07.009](https://doi.org/10.1016/j.compchemeng.2015.07.009)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Computers and Chemical Engineering

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Multi-Objective Optimisation using Surrogate Models for the Design of VPSA Systems

Joakim Beck^a, Daniel Friedrich^b, Stefano Brandani^c, Eric S. Fraga^{d,*}

^aDepartment of Statistical Science, UCL (University College London), London WC1E 7HB, UK

^bInstitute for Energy Systems, School of Engineering, The University of Edinburgh, Edinburgh EH9 3DW, UK

^cInstitute for Materials and Processes, School of Engineering, The University of Edinburgh, Edinburgh EH9 3DW, UK

^dCentre for Process Systems Engineering, Department of Chemical Engineering, UCL (University College London), London WC1E 7JE, UK

Abstract

Vacuum/pressure swing adsorption (VPSA) may be an attractive alternative to other separation processes for some applications. In particular, VPSA may be more energy efficient. However, there is often a trade-off between the different objectives in the separation: product purity, product recovery and power consumption. Identifying those trade-offs is possible through use of multi-objective optimisation methods. The use of multi-objective optimisation for the generation of a trade-off curve or surface is computationally challenging due to the size of the search space and the need for high fidelity simulations of the VPSA designs. The latter is necessary due to the inherently dynamic nature of the process. This paper presents the use of surrogate modelling to address the computational requirements of the high fidelity simulations needed to evaluate alternative designs. Surrogate modelling is an approach for defining simpler, lower fidelity models that capture the key behaviour of the high fidelity model that is necessary to select between alternative designs. We present SbNSGA-II ALM, surrogate based NSGA-II. It is a robust and fast multi-objective optimisation method based on kriging surrogate models and NSGA-II with Active Learning MacKay (ALM) design criteria. The SbNSGA-II ALM method is faster than the NSGA-II method while preserving the robustness and diversity of the Pareto front identified. The properties and capabilities of the SbNSGA-II ALM method are evaluated through the application to an industrially relevant cases study. Specifically, we consider a two column six step VPSA system for CO₂/N₂ separation. Using the surrogate modelling approach, we observe a 5 times reduction in the number of high-fidelity VPSA simulations required, when compared with NSGA-II applied directly to the high fidelity model, to generate a similar Pareto front. The surrogate based multi-objective optimisation method may enable the comprehensive optimisation of adsorption based separation processes. This could lead to more efficient separations for applications ranging from carbon capture to portable oxygen concentrators.

Keywords: carbon capture; kriging; multi-objective optimisation; surrogate modeling; vacuum/pressure swing adsorption; dynamic simulation

1. Introduction

The separation of CO₂ from flue gas is a challenging problem which imposes a large energy penalty on the power station or industrial process. Compared with traditional separation processes, such as absorption and membrane processes, the separation with Vacuum/Pressure Swing Adsorption (VPSA) has the potential to lower the energy penalty (Ruthven, 1984). However, to achieve an efficient and cost-competitive VPSA process requires the optimisation of the process conditions to maximise the product purity and recovery at a low power consumption. This optimisation is challenging due to the complex and dynamic nature of VPSA processes. A detailed description of the underlying principles and of how to model pressure swing adsorption processes is provided by Ruthven et al. (1994).

*Corresponding author

Email addresses: joakim.beck@ucl.ac.uk (Joakim Beck), d.friedrich@ed.ac.uk (Daniel Friedrich), s.brandani@ed.ac.uk (Stefano Brandani), e.fraga@ucl.ac.uk (Eric S. Fraga)

10 For carbon capture, the process design is particularly challenging: the adsorption capacity needs to be large because of the large amount of CO_2 emitted but the fact that CO_2 is the more strongly adsorbed component in a dry flue gas makes the regeneration of the adsorbent to achieve high purity the difficult step. Thus, the optimal combination of system and cycle configurations is a compromise between competing requirements. The system configuration can be a single stage multi-column multi-step VPSA system with
15 pressure equalizations to minimise energy consumption or it can be a combination of two VPSA stages, each of could be complex multi-column and multi-step cycles. The cycle configuration depends on the temperatures, flow rates, pressures and step times. The search space for the full optimisation of the system and cycle configuration is therefore large and hence computationally expensive.

Over the years, many publications have considered the optimisation of VPSA systems (Haghpanah et al.,
20 2013; Jiang et al., 2003; Nikoli et al., 2009; Nilchan & Pantelides, 1998; Krutka & Sjoström, 2011; Ishibashi et al., 1996; Krishnamurthy et al., 2014; Ribeiro et al., 2009, 2008; Da Silva et al., 1999). Usually, the governing hyperbolic/parabolic partial differential algebraic equations (PDAE) are simulated to cyclic steady state for each parameter set generated by the optimisation solver. Since the system of PDAEs is usually large and stiff, this approach is computationally expensive. Most optimisation approaches proposed for VPSA
25 optimisation are therefore unfortunately computationally intractable for solving large-scale optimisation problems. Hence, for optimal design, low-fidelity VPSA models are often used or, alternatively, the search for the optimal design is performed over a reduced set of parameters (Biegler et al., 2005).

An alternative approach is to use reduced order modelling techniques (Agarwal et al., 2008). Specifically, surrogate modelling has been considered for VPSA optimisation (Beck et al., 2012; Hasan et al., 2012,
30 2013). A surrogate model is a fast-to-evaluate approximation of a computer model output, which is built from known sample of input-output data points, and can be used to predict the output response at untried points/configurations. However, previous work on surrogate based optimisation for VPSA needs to be extended if it is to be applied to multi-criteria design problems. The implementation of surrogate based optimisation algorithm for multiple criteria is a hard problem (Fiandaca et al., 2009; Liu & Sun, 2013), and
35 may be less efficient than its single objective counterpart if not implemented carefully.

In this paper, we use surrogate models to optimise the cycle configuration for a fixed system configuration. In the future, we will consider the simultaneous identification of the combined system and cycle configurations. We propose an efficient surrogate based multi objective optimisation algorithm for the design of VPSA systems, using NSGA-II (Non-dominated Sorting Genetic Algorithm-II) (Deb et al., 2002).
40 Greater efficiency is achieved by using the kriging surrogate model (Stein, 1999) that has been shown to be advantageous when linked to existing optimisation methods (Forrester & Keane, 2009). In our case, the surrogate model is built from design points (input-output data points) calculated from the VPSA simulator (Friedrich et al., 2013). Kriging based algorithms have been widely used for single-objective optimisation but there are other issues involved in the multi-objective setting (Sóbestor et al., 2014; Voutchkov & Keane,
45 2010). We propose a novel transformation of the purity and recovery outputs in the surrogate model training phase, as well as a procedure for dealing with the possible failure in the simulator runs during the search. The solution obtained by the surrogate based optimisation procedure is validated against the original model and, if necessary, the NSGA-II optimisation procedure is repeated with an updated surrogate model, using the validation simulations with the detailed model to update the surrogate.

50 A case study for the design of a VPSA system for the separation of a CO_2/N_2 mixture will be used to illustrate the potential of a surrogate modelling approach for multi-criteria design. The design problem considers the purity and recovery of CO_2 as criteria and is sufficiently complex to provide a challenging, multi-objective test of the optimisation procedure.

2. Experimental setup of the VPSA system

55 As a case study, we consider the first stage of a two stage VPSA system for the separation of a CO_2/N_2 mixture representative of the flue gas of a coal fired power station Wang et al. (2013b,a); Liu et al. (2012); Wang et al. (2012). With currently available adsorbents, a two stage process is favoured because a single stage process would require very low vacuum pressures. The first stage, also known as the enrichment stage, increases the CO_2 concentration while ensuring a high recovery, ideally above 95%. In the second stage, also
60 known as the purification stage, the CO_2 concentration is increased to the purity target while the combined recovery needs to be above 90%. The configuration and performance of the second stage depend entirely

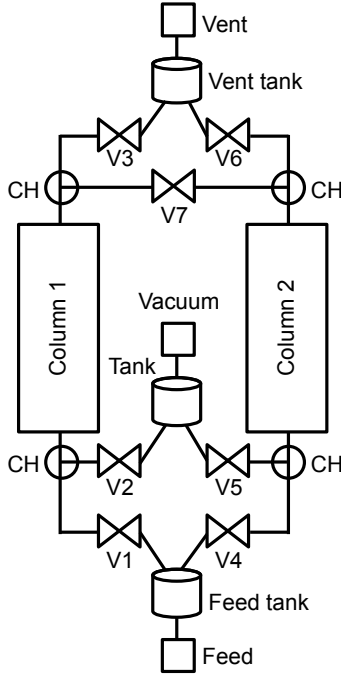


Figure 1: Schematic of the two stage extended Skarstrom cycle VPSA configuration. CH and V stand for column header and valve, respectively.

on the purity and recovery achieved in the first stage. Thus it is important to perform a comprehensive optimisation of the first stage and to identify the Pareto front of purity and recovery trade-offs as a starting point for the optimisation of the second stage.

Specifically, we consider a two column, six step VPSA system as the example case study (Fig. 1). This VPSA system can, depending on the adsorbent and cycle configuration, be used for a wide range of separations. For example, the system can produce the light component in an air separation process (Farooq et al., 1989) or can be used as the first stage in post-combustion carbon capture for the production of the heavy component (Park et al., 2002).

The system is symmetrical with the axis of symmetry going through the feed and vent units. On both sides of each column is a column header which is typically used to ensure a homogeneous flow distribution in the column. The units labeled “Feed”, “Vacuum” and “Vent” provide the boundary conditions for the VPSA system: the Feed unit is an inlet which provides the gas mixture to separate; the Vacuum unit is an outlet which provides vacuum pressure for the purge and blowdown steps; the Vent unit is an outlet at atmospheric pressure and represents the flue gas stack in the power plant. These three units are referred to as *feed* units. The tanks next to the feed units buffer the flows so that the pumps can be operated continuously. The tanks and column headers are connected by valves which control the flow rates in the system and thus the cycle steps.

The operating cycle consists of the following 6 steps: pressurisation, feed, one-sided pressure equalisation, blowdown, purge and one-sided pressure equalisation. This is the original Skarstrom cycle extended by two pressure equalisation steps (Skarstrom, 1966). Fig. 2 shows these 6 steps and the corresponding gas flows and connections between the two columns.

The two sides of the system are cycled through the same steps but with a half cycle offset between them. For example, during the feed step of column 1, the other column performs the purge step. The different cycle steps are defined by opening and closing the 7 valves of the system. This is illustrated by the stem position of the valves: 0 indicates a closed valve, 1 an open valve and any value in between a partially open valve. The stem positions for the 6 cycle steps shown in Fig. 2 are given in Table 1. Note that due to the symmetry in the configuration, s_3 and s_6 will be the same.

In this case study, we consider a lab-scale system with silicalite pellets (HiSiv 3000) supplied by UOP

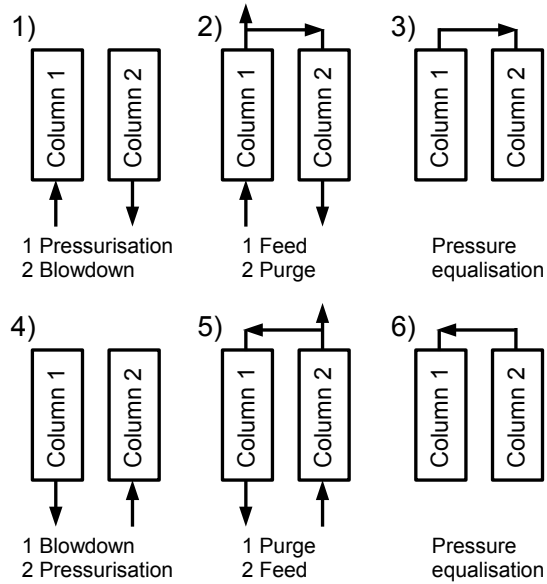


Figure 2: Steps of the Skarstrom cycle with one-sided pressure equalisation.

Table 1: Stem positions for the 7 valves for the 6 cycle steps. The stem positions s_3 ($\equiv s_6$) and s_7 are to be determined by the optimisation procedure.

Steps for column 1	Valve 1	Valve 2	Valve 3	Valve 4	Valve 5	Valve 6	Valve 7
Pressurisation	1	0	0	0	1	0	0
Feed	1	0	s_3	0	1	0	s_7
Pressure equalisation	0	0	0	0	0	0	1
Blowdown	0	1	0	1	0	0	0
Purge	0	1	0	1	0	s_6	s_7
Pressure equalisation	0	0	0	0	0	0	1

Table 2: System parameters

Description	Parameter	Value
Column length	L_c	0.12 m
Column radius	R_c	0.0175 m
Column void fraction	ϵ	0.387
Column header length	L_{CH}	0.14 m
Valve coefficient	c_v	0.000171
Pressurisation/blowdown time	t_{pr}	6 s
Pressure equalisation time	t_{pe}	4 s
Pellet radius	R_p	0.000915 m
Pellet void fraction	ϵ_p	0.35
Axial diffusion coefficient	D_i^L	$1.73 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$
Axial thermal conductivity	λ_f^L	$0.37 \text{ W m}^{-1} \text{ K}^{-1}$
Heat transfer coefficient	h_w	$10 \text{ W m}^{-2} \text{ K}^{-1}$
Crystal density	ρ_{cry}	1390 kg m^{-3}
Saturation capacity	q_s	3.13 mol kg^{-1}
Equilibrium parameter CO ₂	b_1	$5.042 \times 10^{-5} \text{ bar}^{-1}$
Equilibrium parameter N ₂	b_2	$1.052 \times 10^{-4} \text{ bar}^{-1}$
Heat of adsorption CO ₂	$-\Delta\tilde{H}_1$	24900 J mol^{-1}
Heat of adsorption N ₂	$-\Delta\tilde{H}_2$	$16010.4 \text{ J mol}^{-1}$
Effective macropore diffusivity CO ₂	$D_{m,1}^e$	$3.7 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$
Effective macropore diffusivity N ₂	$D_{m,2}^e$	$3.9 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$
Feed mole fraction CO ₂	x_1	1/6
Feed mole fraction N ₂	x_2	5/6

(see supplementary information for more details on the lab-scale system). The adsorption isotherm for CO₂ and N₂ on silicalite was taken from (Golden & Sircar, 1994). The isotherm was measured at 31.4 °C and at 68.4 °C and these data were fitted with the multi-component Langmuir isotherm

$$q_i^* = \frac{q_s b_i \exp\left(\frac{-\Delta\tilde{H}_i}{RT}\right) P x_i}{1 + \sum_{j=1}^{N_c} b_j \exp\left(\frac{-\Delta\tilde{H}_j}{RT}\right) P x_j}. \quad (1)$$

We assume that the mass transfer in the silicalite pellets is dominated by macropore diffusion; CO₂ has a kinetic diameter of less than 0.35 nm and silicalite has channels with width greater than 0.5 nm. We use the Linear Driving Force (LDF) model to describe the mass transfer in the macropore and assume that the adsorbed concentration in the micropore is in equilibrium with the macropore concentration. The macropore LDF parameter is calculated from the effective macropore diffusivity by

$$k_i^p = \frac{5D_{m,i}^e}{R_p}.$$

The adsorbent parameters and the system dimensions are collected in Table 2.

90 To fully describe the system, the feed flow rate, feed time, vacuum pressure, feed temperature and the stem positions of the purge valve (V7) and the vent valves (V3 and V6) must be specified. These will define the search space for the design optimisation problem and will be described in Section 5 below. The numerical simulator is described in the next section, and is sufficiently general to design VPSA systems with an arbitrary number and combination of units and cycle steps.

95 3. Dynamic adsorption cycle model

The VPSA system and the associated cycle shown in Fig. 1 and 2 can be simulated in two ways: i) simulating only the columns (Webley & He, 2000) or ii) simulating the columns and the supporting infrastructure,

such as valves and tanks (Liu et al., 2011; Nikolic et al., 2008). In the former approach, the different cycle steps are defined through the boundary conditions of the columns while, in the latter approach, valves are opened and closed according to the schedule of an equivalent experimental apparatus. The former approach has a slightly lower computational complexity while the latter approach is closer to the experiment. Since the increase in computational complexity is manageable and the aim is to simulate an existing experimental system, the second approach is used.

The VPSA system is build from four unit types (column, valve, tank and feed unit) and the simulations are performed with an in-house general adsorption cycle simulator. Briefly, the simulator implements the mass, momentum and energy balances for adsorption columns and auxiliary units such as valves, splitters and tanks. Since the governing equations for these units have been presented previously (Banu et al., 2013; Friedrich et al., 2013), we give here only a list of the used models. However, the governing equations are repeated for completeness in the supplementary information in Appendix A.

The column model uses the axially dispersed plug flow model and the pressure drop is described by the Ergun equation. The adsorption equilibrium is described by the Langmuir model and the mass transfer by the macropore LDF model. The energy balance takes heat transfers between the surroundings, column, gas phase and solid phase into account. The system of partial differential algebraic equations is discretized along the column length with a flux-limited finite volume method. The resulting system of ordinary differential algebraic equations is integrated in time with the backward differentiation formulas implemented in the SUNDIALS solver (Hindmarsh et al., 2005). The system is simulated cycle after cycle until the cyclic steady state is reached.

3.1. Calculation of purity, recovery and molar work

The simulator is used to generate the composition and flow profiles to allow the calculation of the criteria considered for ranking design alternatives. The design criteria are the purity and the recovery of CO₂. These, together with the power requirements, are calculated, for the k -th cycle based on the number of moles passing through the relevant feed units and the work done by those units:

$$\text{Purity}_{\text{CO}_2}^k = \frac{n_{vac, \text{CO}_2}^k - n_{vac, \text{CO}_2}^{k-1}}{\sum_{j=1}^{N_c} (n_{vac, j}^k - n_{vac, j}^{k-1})}, \quad (2)$$

$$\text{Recovery}_{\text{CO}_2}^k = \frac{n_{vac, \text{CO}_2}^k - n_{vac, \text{CO}_2}^{k-1}}{n_{fe, \text{CO}_2}^k - n_{fe, \text{CO}_2}^{k-1}}, \quad (3)$$

$$\text{Molar work}_{\text{CO}_2}^k = \frac{W_{fe}^k - W_{fe}^{k-1} + W_{vac}^k - W_{vac}^{k-1}}{n_{vac, \text{CO}_2}^k - n_{vac, \text{CO}_2}^{k-1}}. \quad (4)$$

The number of moles for any unit k , $n_{j,i}^k$, is a cumulative amount from the start of the process. The work done in the feed units, i.e. W_{fe} and W_{vac} , is calculated from the pump power. Hence, the difference between the amounts for cycle k and cycle $k-1$ give the number of moles and work during the k -th cycle. The purity and recovery of CO₂ will be the objective functions for the subsequent multi-objective optimisation and are always calculated at the 'Vacuum' unit.

4. Surrogate based optimisation with NSGA-II

The need to simulate a dynamic system to provide the data necessary for the evaluation of the design criteria means that the optimisation of the system and cycle configuration is computationally intensive. Traditionally, the computational challenge is addressed through the use of simpler models. However, these models may not capture the detail required to identify the best designs, especially in a multi-objective scenario where potentially conflicting criteria are used to evaluate and compare alternative designs. This section describes a surrogate based optimisation (SBO) procedure developed to address the computational challenge through the automatic and adaptive definition of surrogate models.

A surrogate model is a fast approximation of the computer model output $z(\mathbf{x}) : \mathcal{X} \rightarrow \mathbb{R}$. Here $\mathcal{X} \subset \mathbb{R}^d$ represents the design space and d is the number of design variables. Surrogate models are typically fitted to

n pre-computed input-output data points of $z(\mathbf{x})$, that is, training data $(\mathbf{X}, \mathbf{z}) = (\mathbf{x}_j, z(\mathbf{x}_j))_{j=1}^n$, and can be used to make predictions for untried points in the design space \mathcal{X} . Thus no modification of the original computer code is needed. In the multi-output setting we can assign to each output its own surrogate.

Kriging regression (Jin et al., 2001; Sóbester et al., 2014; Stein, 1999) is a popular surrogate modelling technique for engineering design and is used in this work. Other popular techniques include polynomial response surfaces, radial basis functions (RBF), artificial neural networks (ANN), multivariate adaptive regression splines (MARS), and support vector regression (SVR) (Forrester & Keane, 2009). Kriging’s popularity can be accredited to its ability to easily produce confidence bands for prediction, which has been proven to be a highly attractive feature for exploring the design space.

In the context of optimisation, it is possible to replace the computationally expensive objective function with a surrogate model. However, by doing so, one solves an approximation of the original optimisation problem, which may be considered a naïve approach. Instead, the problem can be approached by solving a sequence of approximations of the original problem. First, employ the surrogate model in lieu of the objective to assist the optimiser in the search for the next design point. Then update the surrogate model using the detailed model evaluation at some point determined by the optimisation procedure. Solve the problem again and iterate. The surrogates are updated as new input-output data points are collected from the detailed model. This procedure is illustrated in Fig. 3.

There are many potential criteria for selecting points with which to update the surrogate, referred to as “surrogate based criteria.” The most common and seemingly natural choice is would be to select the design point \mathbf{x} with the highest $z(\mathbf{x})$ value, as predicted by the surrogate. This, however, may lead to premature convergence (Jones, 2001). The expected improvement (EI) criterion (Jones et al., 1998) is another popular choice, one which guarantees convergence to a global optimum.

4.1. Kriging model

Kriging regression treats the computer model of interest $z : \mathcal{X} \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$ as an unknown deterministic function, except for the inputs that already are known. Following (Sacks et al., 1989), we model $z(\mathbf{x})$ as a Gaussian random field, $Z(\mathbf{x})$, with $E(Z(\mathbf{x})^2) < \infty$ for all $\mathbf{x} \in \mathcal{X}$. The resulting statistical model is known as the kriging model and can be used for prediction. The *ordinary* kriging, used in this work, is characterised by a constant mean, $\beta \in \mathbb{R}$, and a covariance function $\Sigma(\mathbf{x}, \mathbf{x}') = \sigma^2 K(\mathbf{x}, \mathbf{x}')$, for $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$. $K(\mathbf{x}, \mathbf{x}')$ is the correlation function which must be positive semidefinite. A widely used correlation function is the squared exponential correlation function (Rasmussen & Williams, 2006):

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\sum_{i=1}^d \frac{(x_i - x'_i)^2}{2\ell_i}\right), \quad (5)$$

where $\boldsymbol{\ell} = (\ell_1, \ell_2, \dots, \ell_d) \in \mathbb{R}_+^d$ are the correlation length-scales and σ^2 is the process variance.

Let us assume we have some training data $\mathcal{D} = (\mathbf{X}, \mathbf{z})$ where $\mathbf{X} = (\mathbf{x}_j)_{j=1}^n$, $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$, and $z_j = z(\mathbf{x}_j)$. Then, given that $Z(\mathbf{x})$ is a Gaussian Process (GP), we also know that the joint distribution $\hat{\mathbf{Z}} = [Z(\mathbf{x}_1), Z(\mathbf{x}_2), \dots, Z(\mathbf{x}_n)]$ follows a multivariate normal distribution with $1 \times n$ -dimensional mean vector $\mathbf{1}\beta$ and $n \times n$ variance-covariance matrix $\Sigma_{(\sigma^2, \boldsymbol{\ell})}$. Here $\mathbf{1} = (1, 1, \dots, 1)^T$ and the (j, k) th element of $\Sigma_{(\sigma^2, \boldsymbol{\ell})}$ is given by $\Sigma(\mathbf{x}_j, \mathbf{x}_k)$ for $\mathbf{x}_j, \mathbf{x}_k \in \mathbf{X}$. The predictive distribution of $Z(\mathbf{x})|\boldsymbol{\ell}, \mathbf{z}$ is a GP, with mean

$$\hat{z}(\mathbf{x}) = \hat{\beta} + \mathbf{k}(\mathbf{x})\mathbf{K}^{-1}(\mathbf{z} - \mathbf{1}\hat{\beta}), \quad (6)$$

and variance

$$s^2(\mathbf{x}) = \hat{\sigma}^2 \left(1 - \mathbf{k}^T(\mathbf{x})\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}) + \boldsymbol{\gamma}^T(\mathbf{x}) (\mathbf{1}^T\mathbf{K}^{-1}\mathbf{1})^{-1} \boldsymbol{\gamma}(\mathbf{x})\right), \quad (7)$$

where the $m \times 1$ vector $\mathbf{k}(\mathbf{x})$ has entry j given by $K(\mathbf{x}, \mathbf{x}_j)$ for $\mathbf{x}_j \in \mathbf{X}$, $\boldsymbol{\gamma}(\mathbf{x}) = 1 - \mathbf{1}^T\mathbf{K}^{-1}\mathbf{k}(\mathbf{x})$, and the generalized least square solution $\hat{\beta} = (\mathbf{1}^T\mathbf{K}^{-1}\mathbf{1})^{-1}\mathbf{1}^T\mathbf{K}^{-1}\mathbf{z}$. \mathbf{K} is the correlation matrix given by $\mathbf{K} = \sigma^{-2}\boldsymbol{\Sigma}$. Here we estimate the process variance using the unbiased estimate $\hat{\sigma}^2 = \frac{1}{n}(\mathbf{z} - \mathbf{1}\hat{\beta})^T\mathbf{K}^{-1}(\mathbf{z} - \mathbf{1}\hat{\beta})$. Fig. 4 is an illustrative example included for readers unfamiliar with kriging modelling.

The kriging mean is given by (6), which is our kriging predictor, and the kriging variance is given by (7). The kriging expressions are derived under the assumption that the correlation parameters $\boldsymbol{\ell}$ are known which rarely is the case. The correlation parameters $\boldsymbol{\ell}$ can be estimated by the data using Maximum

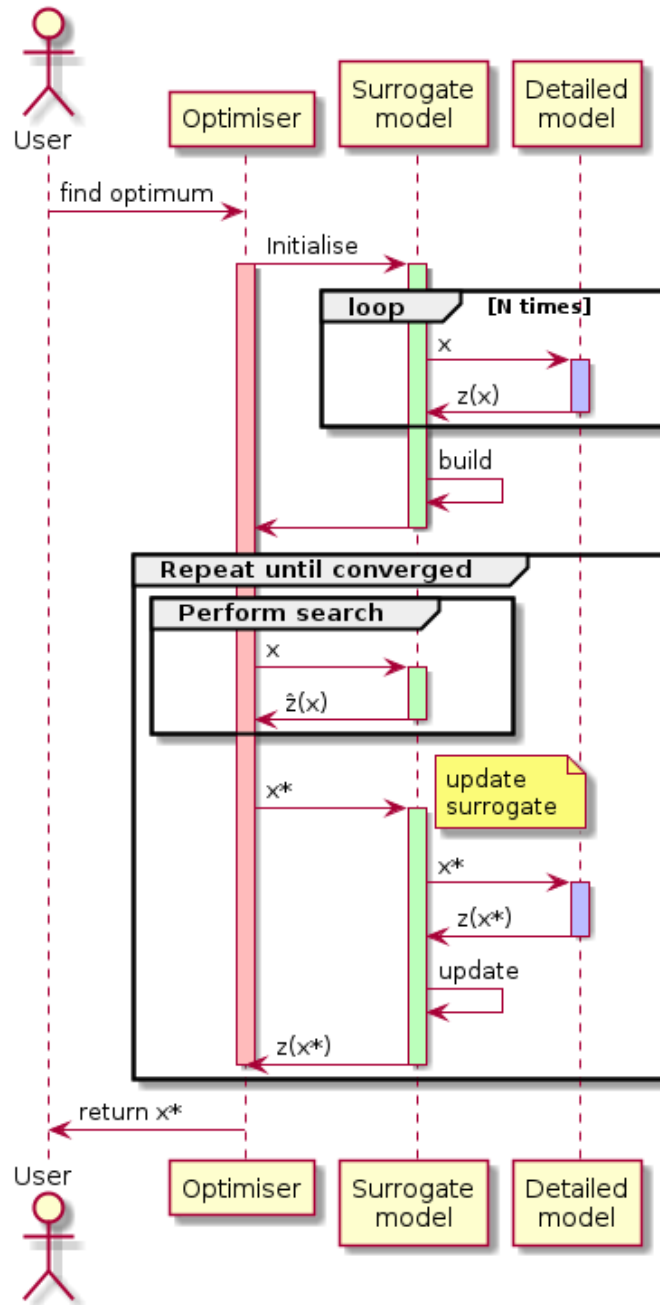


Figure 3: Sequence diagram for a surrogate based optimiser showing when the surrogate is invoked and when the detailed model, or simulation, is invoked. This sequence is based on an optimiser that has some internal search procedure within an outer iteration; an example of such a method would be a genetic algorithm.

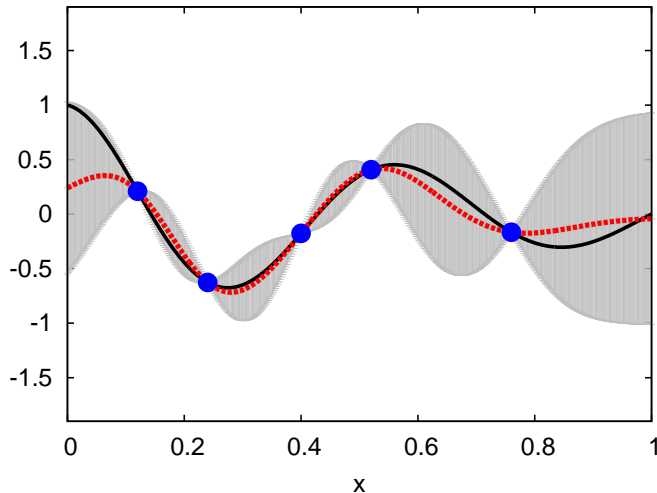


Figure 4: Exemplary illustration of kriging, with function $z(x)$ (black-solid line), known input-output data $\mathcal{D} = (\mathbf{X}, z)$ (blue-solid circles), and the kriging predictor (red-dashed line) with 95%-confidence bands (grey region) produced by the kriging variance.

Likelihood Estimation (MLE). The ML estimates $\hat{\ell}$ can be determined by minimizing the profile log-likelihood $-2 \ln L(\ell) = n \ln \sigma_{\ell}^2 + \ln |\mathbf{K}_{\ell}|$ (Sacks et al., 1989), where $L(\ell)$ is the profile likelihood under the Gaussian assumption. We also add a small regularization parameter $\tau^2 > 0$ (known as the nugget, or the jitter) to the diagonal elements of the correlation matrix \mathbf{K} to improve its condition number before computing its inverse, which is required in (6) and (7). Typically, τ^2 is small. This results in a kriging model (with some slightly perturbed correlation matrix entries) that does not ensure interpolation of the given data \mathcal{D} . Nevertheless, the use of a regularization parameter to improve numerical stability can be advantageous (Gramacy & Lee, 2012). The Cholesky decomposition is used for the matrix inversion step.

4.2. Kriging prediction error analysis

Over the value domains for the decision parameters considered in our case study (described in Section 5) we have assessed the prediction quality of the kriging surrogate model for both the product purity (2) and the product recovery (3). The prediction error is the normalised root-mean-square prediction error (NRMSPE):

$$\text{NRMSPE} = \frac{\sqrt{\left(\frac{1}{M} \sum_{j=1}^M (\hat{z}(\mathbf{x}_j) - z(\mathbf{x}_j))^2\right)}}{\max_j z(\mathbf{x}_j) - \min_j z(\mathbf{x}_j)}, \quad (8)$$

where $z(\mathbf{x})$ is the computer model output and $\hat{z}(\mathbf{x})$ is the kriging predictor. The results of this analysis are presented in Fig. 8. To calculate the NRMSPE, a hold-off set has been used, consisting of 1,000 input-output data samples of the VPSA simulator, and determined by maximin distance Latin-hypercube sampling (LHS). ML estimates are used for the parameters in the squared exponential correlation function (5). The kriging model prediction errors for both objectives are below 10% if the training data set size exceeds 60. As the number of design solutions in the training set increases beyond 60, the rate of the decrease in error slows down, as expected (Loeppky et al., 2009). The conclusion of the error analysis is that the kriging surrogates can be used for well-behaved VPSA model outputs such as purity and recovery.

4.3. Developing a surrogate based optimisation procedure

A standard optimisation problem is to maximise (or minimise) an objective function $z(\mathbf{x}) : \mathcal{X} \rightarrow \mathbb{R}$:

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} z(\mathbf{x}). \quad (9)$$

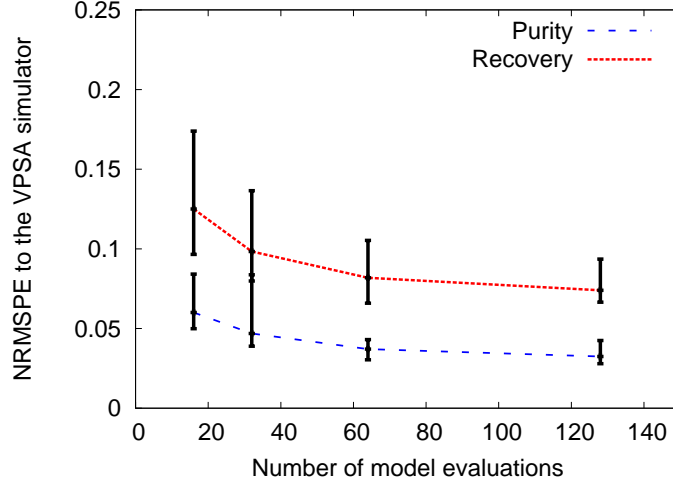


Figure 5: NRMSPE for the kriging model based on the training data consisting of purity and recovery calculations using the detailed VPSA simulator. The error bars represent the variability over ten sets of training data (min-mean-max).

Alternatively, when the objective $z(\mathbf{x})$ is expensive to evaluate, a surrogate model $\hat{z}(\mathbf{x})$ can be used in lieu of the objective, i.e.,

$$\hat{\mathbf{x}}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} \hat{z}(\mathbf{x}). \quad (10)$$

The solution $\hat{\mathbf{x}}^*$ serves as an approximation to the solution of the original problem, \mathbf{x}^* . Indeed, the more accurate the surrogate, the more confidence we will have in the solution $\hat{\mathbf{x}}^*$. On the other hand, even with an accurate surrogate, the difference between the two solutions may be very large. In practice, therefore, the SBO method is used (Viana et al., 2014).

SBO exploits the predictive power of the surrogate model in lieu of the objective. First the surrogate is built on a set of training data. Then, at each iteration of the SBO method, the optimiser of choice is applied for a design criterion evaluated with a surrogate model. At the end of each iteration, the surrogate is updated with the design solution $(\mathbf{x}_k, z(\mathbf{x}_k))$ selected by the optimiser for the chosen design criterion. A common surrogate based design criterion is to just optimise the surrogate response $\hat{z}_{\mathcal{D}_k}(\mathbf{x})$. The procedure is repeated until the stopping criteria are met.

The SBO problem for iteration $k = 1, 2, \dots$ can be written as:

$$\mathbf{x}_k = \arg \max_{\mathbf{x} \in \mathcal{X}} C(\mathbf{x}, \mathcal{D}_k, \hat{z}_{\mathcal{D}_k}, \hat{s}_{\mathcal{D}_k}^2; \cdot), \quad (11)$$

for some surrogate based design criterion $C(\cdot)$, with $\mathcal{D}_k = \mathcal{D}_0 \bigcup_{i=1}^{k-1} (\mathbf{x}_i, z(\mathbf{x}_i))$ being the input-output data known at iteration k and where \mathcal{D}_0 denotes the initial data $(\mathbf{x}_i, z(\mathbf{x}_i))_{i=1}^n$. Here $\hat{z}_{\mathcal{D}_k}(\mathbf{x})$ is the kriging mean (6) and $\hat{s}_{\mathcal{D}_k}^2(\mathbf{x})$ is the kriging variance (7), conditioned on the training data \mathcal{D}_k .

In developing an SBO, the following must be addressed:

- the choice of optimiser;
- the criteria which form the objective functions;
- the initial training for the surrogate;
- how to handle failures in the evaluation of the detailed model; and,
- a parallel implementation, if any.

For the development of the SbNSGA-II SBO method, each of these is discussed in turn.

The choice of optimiser. Most conventional optimisation methods can be used within the SBO framework. We have used the Non-dominated Sorting Genetic Algorithm-II (NSGA-II) (Deb et al., 2002). The NSGA-II method has been extensively used for multi-criteria design problems. It is a genetic algorithm suitable for optimisation where gradient information is not present and where any of the objective function, the constraints or the design space may be non-convex. NSGA-II is able to incorporate infeasible design points into the selection procedure. It also offers diversity control for the evolving population of designs to generate an approximation to the Pareto front that is as broad as possible.

The criteria. The choice of the surrogate based criteria should ensure the surrogates adaptively target the regions of the design space believed to contain good design solutions and should guarantee convergence in the limit to the global optima. With the correct choice, the surrogate should actively learn as the search progresses. Some popular kriging-based design criteria are the *kriging mean*,

$$C(\mathbf{x}, \hat{z}_{\mathcal{D}_k}) = -\hat{z}_{\mathcal{D}_k}(\mathbf{x}), \quad (12)$$

the *kriging variance*,

$$C(\mathbf{x}, \hat{s}_{\mathcal{D}_k}^2) = \hat{s}_{\mathcal{D}_k}^2(\mathbf{x}), \quad (13)$$

and the *expected improvement* (EI) (Jones et al., 1998). The latter selects the design point \mathbf{x}_k with the largest expected improvement relative to the best known design solution:

$$C(\mathbf{x}, \mathbf{z}_{\mathcal{D}_k}, Z(\mathbf{x})) = E_Z(\max\{z_{min} - Z(\mathbf{x}), 0\}), \quad (14)$$

where $z_{min} = \min \mathbf{z}_{\mathcal{D}_k}$. Based on the Gaussian assumption for $Z(\mathbf{x})$, conditioned on the known data \mathcal{D}_k , the EI criterion can be expressed in closed-form:

$$C(\mathbf{x}, \mathbf{z}_{\mathcal{D}_k}, \hat{z}_{\mathcal{D}_k}, \hat{s}_{\mathcal{D}_k}^2) = (z_{min} - \hat{z}_{\mathcal{D}_k}(\mathbf{x}))\Phi\left(\frac{z_{min} - \hat{z}_{\mathcal{D}_k}(\mathbf{x})}{\hat{s}_{\mathcal{D}_k}(\mathbf{x})}\right) + \hat{s}_{\mathcal{D}_k}(\mathbf{x})\phi\left(\frac{z_{min} - \hat{z}_{\mathcal{D}_k}(\mathbf{x})}{\hat{s}_{\mathcal{D}_k}(\mathbf{x})}\right) \quad (15)$$

where $\Phi(\cdot)$ is the cumulative distribution and $\phi(\cdot)$ is the probability density function for the conditional posterior probability distribution of the kriging model. The EI criterion is useful when applied iteratively to improve the global approximation while searching for the optima.

The kriging mean criterion tends to be too optimistic unless the objective function is easy to approximate or convex. Jones (2001) demonstrated that using an optimiser on the kriging mean can cause the method to get trapped at a local optimum and, in turn, the improvement of the surrogate stagnates. The goal of the kriging variance criterion is equivalent to learning the value at the design point of highest uncertainty and, hence, strives to improve prediction accuracy rather than targeting regions of interest in the design space. SBO with the EI criterion is also known as efficient global optimization (EGO) (Jones et al., 1998). EGO has gained popularity as an SBO method but a drawback is that, if the process variance is under-estimated, the EGO search may have limited exploration as it concentrates too early on the current best solution. On the other hand, if the variance is over-estimated, the fact that the location of the currently best solution is known is not exploited enough and the result is slow convergence.

Motivated by Voutchkov & Keane (2010), we use the following hybrid criterion to promote robustness:

$$C(\mathbf{x}, \hat{z}_{\mathcal{D}_k}, \hat{s}_{\mathcal{D}_k}^2) = \frac{(1 - (-1)^k)}{2} \hat{z}_{\mathcal{D}_k}(\mathbf{x}) + \frac{(1 - (-1)^{k-1})}{2} \hat{s}_{\mathcal{D}_k}^2(\mathbf{x}). \quad (16)$$

This criterion is well-balanced between exploration (kriging variance) and exploitation (kriging mean), guarantees global convergence in the limit, and is easy to interpret.

Initial training data. Prior to running the SBO algorithm, the design space is sampled to obtain the initial training data for building the kriging model. For kriging regression, a rule of thumb for the moderate size of the training data is $10d$ (Loeppky et al., 2009). Examples of classical experimental design techniques for physical experiments are full-factorial and central composite design (Montgomery, 2007). For designing computer experiments, however, Latin hypercube sampling (McKay et al., 1979) is commonly used as it scales well to high-dimensional design spaces and is easy to implement. Our initial training data are generated through a maximin Latin hypercube design (MmLHD). First we generate a number of latin hypercube

design (LHD) points and the design which has the largest minimum inter-point distance is selected (maximin criterion). MmLHD could be viewed a compromise between the maximin and minimax criterion.

As another option for sampling, although not used in our work, is to sequentially design the training data. Sequential design using entropy-based criteria generally outperform LHDs (Beck & Guillas, 2014), as long as the underlying computer model is computationally expensive. We will be considering this and other alternatives in the future. For the moment, however, we note that LHD works well.

Handling simulation model failures. VPSA process simulations may become unstable for certain design configurations. An unstable design may cause a failure in the evaluation of the design point. In our case, simulator failures are mainly due to failure to solve the set of differential algebraic equations whenever the dynamic behaviour in the adsorbent beds is too challenging. In such a case, even if no output is returned by the simulator, we still include the design point in the training data. This additional information improves the kriging variance (7). Note that the kriging variance does not explicitly depend on the output data, only implicitly via $\hat{\sigma}^2$. The uncertainty can be modelled more accurately by accounting for the failed design points. When the search is guided by the kriging variance, this approach helps us avoid revisiting failed design points, as well as points in close proximity.

Parallel implementation. The parallel implementation of the algorithm is based on the approach of Schonlau (1998) for multi-point selection, known as the *Kriging Believer* (Janusevskis et al., 2012). In the k -th SBO iteration, n design points are selected sequentially but without evaluation until all the points have been selected. We denote the j th design point selected during the k th SBO iteration by \mathbf{x}_{kj} . First we select \mathbf{x}_{k1} by performing the k th SBO iteration algorithm as before. Next, we add $(\mathbf{x}_{k1}, \hat{z}(\mathbf{x}_{k1}))$ to the training data, and the kriging is updated. Observe that we updated the training data with the kriging mean $\hat{z}(\mathbf{x}_{k1})$, instead of the true model output $z(\mathbf{x}_{k1})$. The same steps are performed for \mathbf{x}_{kj} , $j = 2, \dots, n$, sequentially. The computer model is then evaluated in parallel for the n design points, the training data set is corrected with $\{(\mathbf{x}_{kj}, z(\mathbf{x}_{kj}))\}_{j=1}^n$, the kriging model is updated, and the next SBO iteration begins. With this approach, n should be proportional to the number of allocated CPUs. The design selection becomes less efficient for large n , since the training data are updated with the kriging mean at each step j , leading to degeneracy.

4.4. The implementation of SbNSGA-II: a Surrogate based NSGA-II

This section describes the specifics of implementing an SBO based on NSGA-II for the VPSA design problem.

SBO extension to multiple objectives in NSGA-II. A kriging model is used for each objective together with a mixture of surrogate based design criteria (16), which can be written as follows:

$$C(\mathbf{x}, \hat{\mathbf{z}}_{\mathcal{D}_k}, \hat{\mathbf{s}}_{\mathcal{D}_k}^2) = \frac{(1 - (-1)^k)}{2} \hat{\mathbf{z}}_{\mathcal{D}_k}(\mathbf{x}) + \frac{(1 - (-1)^{k-1})}{2} \hat{\mathbf{s}}_{\mathcal{D}_k}^2(\mathbf{x}). \quad (17)$$

where $\hat{\mathbf{z}}_{\mathcal{D}_k}(\mathbf{x}) = [\hat{z}_{\mathcal{D}_{k1}}(\mathbf{x}), \dots, \hat{z}_{\mathcal{D}_{kq}}(\mathbf{x})]$, $\hat{\mathbf{s}}_{\mathcal{D}_k}^2(\mathbf{x}) = [\hat{s}_{\mathcal{D}_{k1}}^2(\mathbf{x}), \dots, \hat{s}_{\mathcal{D}_{k1}}^2(\mathbf{x})]$, and q is the number of objective functions. For the parallel implementation, the index k is replaced by the index j as described above. For example, should eight CPUs be available, four CPUs would be used to compute design points selected by optimising $\hat{\mathbf{z}}_{\mathcal{D}_k}(\mathbf{x})$ and the other four for points selected by maximising $\hat{\mathbf{s}}_{\mathcal{D}_k}^2(\mathbf{x})$.

What makes the NSGA-II environment more challenging with surrogates is that the optimiser is population based and returns non-dominated solutions (in a Pareto sense). First NSGA-II is applied to design criterion (17). The naïve approach would then be to select one (or multiple) non-dominated solutions for further consideration from the Pareto set. Instead, by respecting the nature of the multi-objective setting, the kriging-based Pareto front (from applying NSGA-II on kriging models) is combined with the current Pareto front and one or several non-dominated solutions for model evaluation are selected. This leads to a Pareto front with better diversity, richness, and coverage.

Transformation of purity and recovery. The kriging makes predictions on the real line \mathbb{R} whereas our model outputs, purity and recovery, are ratios in the interval $[0, 1]$. Some predictions made by the kriging will inevitably violate physical constraints, such as concentrations. One could enforce the bounds by using the extreme values 0 and 1 whenever a prediction falls outside the feasible domain $[0, 1]$. This manipulation of the

kriging response may lead to non-differentiable behaviour. We instead suggest the following transformation of the kriging’s training data \mathbf{z} :

$$\mathbf{u} = \log \left(\frac{\mathbf{z}}{(1 + \epsilon)\mathbf{1} - \mathbf{z}} \right), \quad (18)$$

where \mathbf{u} is the transformed data, and $\epsilon > 0$ is a typically very small perturbation. This extends the output space from $[0, 1 + \epsilon]$ to the real line \mathbb{R} . Then, afterwards, in order to make predictions, we transform back using $\hat{z}(\mathbf{x}) = ((1 + \epsilon) \exp(\hat{u}(\mathbf{x}))) / (1 + \exp(\hat{u}(\mathbf{x})))$, where $\hat{u}(\mathbf{x})$ is a transformed kriging prediction. Here ϵ is a small perturbation to allow $z = 1$. The perturbation parameter ϵ may also be necessary to ensure the transformation is valid if any output $z(\mathbf{x})$ is above 1 due to numerical errors, e.g., producing a purity level of 1.0001.

An advantage of this transformation is that, when a multi-objective optimiser is applied, the Pareto ordering of the multi-objective function evaluations is preserved under transformation, i.e., if one response value dominates another in the original space, this response will also dominate the same one in the transformed space. Fig. 6 shows that the proposed transformation improves the algorithm’s robustness for multiple objectives with physical constraints. Other examples of transformation-based kriging models are the log-normal kriging (Diggle & Ribeiro, 2007), trans-Gaussian kriging (De Oliveira et al., 1997) using Box-Cox transformations, and $\mathbf{u} = \log(\mathbf{z} + \mathbf{1})$ for non-negative outputs (Spiller et al., 2014). In fact, this data transformation is an essential ingredient to ensure the kriging-generated Pareto front scales well with the current Pareto front when combined.

Diversity constraint in the design space. We impose a diversity constraint in the design space to avoid clustering: if the distance between the most promising candidate design point $\tilde{\mathbf{x}}$ and the training data $\tilde{\mathbf{X}}$, given by the Euclidean norm,

$$\|\tilde{\mathbf{x}} - \tilde{\mathbf{X}}\|_2 = \min_{\tilde{\mathbf{x}}^* \in \tilde{\mathbf{X}}} \left\{ \sqrt{\left(\sum_j^d (\tilde{x}_j - \tilde{x}_j^*)^2 \right)} \right\} \quad (19)$$

is ϵ -small, the candidate is rejected and the second most promising candidate is tested, and so on, until a viable candidate is found. Here $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{X}}$ are normalised.

Stopping criteria. One of the challenges with evolutionary strategies for optimisation is to decide when to stop. In this work, we focus on the analysis of the methodology and therefore stop the algorithm when a given number of iterations have been completed. A stop criterion could also be designed with respect to some computational budget or with respect to the change in the best design solution found, over a few iterations, either in terms of change in the design space or in the objective space.

5. The case study

As a case study, we use the separation of the flue gas of a coal fired power station by a VPSA process with the adsorbent silicalite. The setup of the VPSA system is described in Section 2 above. In that description, a number of parameters required to fully describe the VPSA process were not defined. These form the design space and are given by the decision parameters presented in Table 3. Out of computational convenience, the decision parameters have all been normalised to $[0, 1]$ for use in the surrogate based optimisation.

While the VPSA process depends on a large number of parameters, these 6 parameters have the largest impact on the process and are those usually adjusted to optimise the process performance. The purge-to-feed ratio controls the amount of gas flow during the purge step. A larger ratio usually results in faster desorption but lowers the resulting purity. The feed time and flow rate influence the productivity and recovery of the process. These parameters are linked if we only consider the purity and recovery. The vacuum pressure controls how much CO_2 is desorbed during the purge step. The lower the vacuum pressure the more CO_2 is desorbed but at the expense of higher energy consumption. The stem positions of the vent valves control the pressure during the feed step and thus the amount of CO_2 adsorbed at equilibrium. The feed temperature

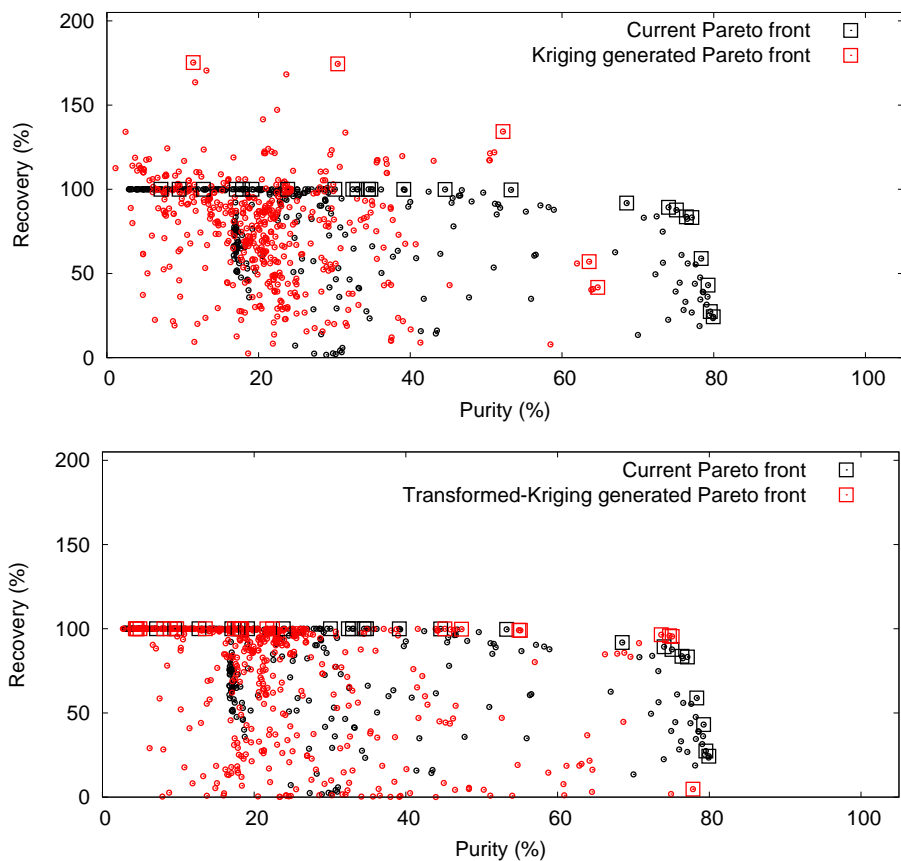


Figure 6: Top: The Pareto front approximation generated by responses from the kriging predictor. Bottom: The Pareto front approximation generated by the transformed-kriging. This illustrates the benefit of using kriging in the transformed space when performing multi-objective VPSA optimisation when there are physical constraints for any of the objective functions.

Table 3: Decision parameters

	Parameter	Range	Units
Purge-to-feed ratio valve stem position	s_7	[0, 1]	-
Feed/purge time	t_{fe}	[10, 200]	s
Feed flow rate	F_{fe}	[0.0005, 0.008]	mol s^{-1}
Vacuum pressure	P_{vac}	[0.02, 0.4]	bar
Adsorption pressure regulator valves stem positions	s_3, s_6	[0.3, 1.0]	-
Feed temperature	T_{fe}	[290, 340]	K

influences the shape of the isotherm and thus the working capacity. At higher temperatures the isotherm becomes more linear and the desorption step might be more efficient.

315 While the general effect each of these parameters individually has on the performance of the VPSA process is known, the interplay between these parameters and the conflicting objectives make the use of optimisation methods mandatory for the design of efficient VPSA processes.

5.1. Optimisation methods: SbNSGA-II and SbNSGA-II ALM

We use our developed SBO method presented in Section 4.3. Ordinary kriging models (see Section 4.1) 320 are surrogate models chosen to approximate the product purity and recovery outputs and are trained on transformed data, as described earlier, to ensure predictions are made in $[0, 1]$. The kriging models assume an empirical squared exponential correlation (5) fitted using maximum likelihood estimation (MLE). A regularization parameter $\tau^2 = 10^{-6}$ is introduced for numerical stability in the kriging computations. The empirical correlation structure is updated when new input-output data arrives. MLE is quite computationally 325 intensive, so we only update the parameter estimates until we have 80 data points. In this case study, we do not expect using more than 80 data points will lead to any significant difference in the estimate values. The initial training data is of size 16 and has been sampled as an MmLHD design.

For the NSGA-II, we have specified the following: population size $N_p = 128$, mutation rate $mr = 1/d$, mutation index $n_m = 20$, crossover rate $cr = 0.9$, and crossover index $n_c = 5$, with real-coded values, 330 simulated binary crossover (SBX) and polynomial mutation. We use $\varepsilon = 10^{-3}$ for the explicit diversity constraint (19) in the design space.

Two implementations have been developed, one using the NSGA-II along with the kriging mean criterion (SbNSGA-II) and one which introduces Active Learning MacKay (ALM) to construct a mixture of criteria (17) that broaden the approximation to the Pareto front (SbNSGA-II ALM). The implementations make 335 full use of parallel computing with the Kriging believer approach and have been tested on an 8 CPU core system, yielding close to linear speed up. Eight design points are selected per SBO iteration, and for those design configurations we run the VPSA model in parallel over the eight CPU cores.

5.2. Optimisation results

Two objectives are considered, the purity and the recovery, and these are maximised simultaneously in 340 a Pareto sense. For comparison, we have also included results for the stand-alone NSGA-II, given a budget of 1600 model evaluations of the VPSA simulator. We have used five separate sets of training data for SbNSGA-II to show the method’s robustness with respect to the training data.

Fig. 7 shows the optimisation results of the stand-alone NSGA-II, SbNSGA-II, and SbNSGA-II ALM, 345 at different numbers of evaluations of the VPSA simulator. Although there is only a small difference in performance between SbNSGA-II and SbNSGA-II ALM for 256 points, our criterion (17) is more robust as the optimiser can escape local optima and enables active learning for the kriging model. In Fig. 7, we can see that, with some initial training data, the SbNSGA-II can be substantially worse than SbNSGA-II ALM. The benefit with SbNSGA-II is its fast convergence when close to the optima. The Pareto front for the SBO methods using 256 design points overlaps with the Pareto front for the stand-alone NSGA-II using 350 a normally excessive budget of 1600 design points (see bottom left in Fig. 8). The SBO methods reach a reliable Pareto solution. A trade-off between the SBO methods can be achieved by modifying the mixture criterion of the SbNSGA-II ALM to use the predictive mean criterion more frequently than the predictive variance criterion.

The ML estimates of the parameters in the squared exponential correlation function (5) are $\ell_{purity} =$ 355 $(1.2, 4.7, 1.6, 15, 14.4, 15)^T$, and $\ell_{recovery} = (3.2, 1.8, 1.7, 5.6, 8.5, 15)^T$. The parameter order in the vectors is the same order as introduced in Table 3. These correlation parameters represent the importance of the different parameters on the respective outputs and can be used to guide the analysis of the optimised design. A small value of a correlation parameter indicates a strong nonlinear relationship between the objective and the design parameter. Thus, parameters with lower values have a large influence on the objective. On the 360 other hand, for a large value of a correlation parameter, the objective function is smooth with respect to the corresponding design parameter. While this indicates that the relationship has no significant nonlinear behaviour, it does not indicate that the design parameter has no influence at all on the objective. The possible linear relationship can be deduced from the high level visualization given in Fig. 9.

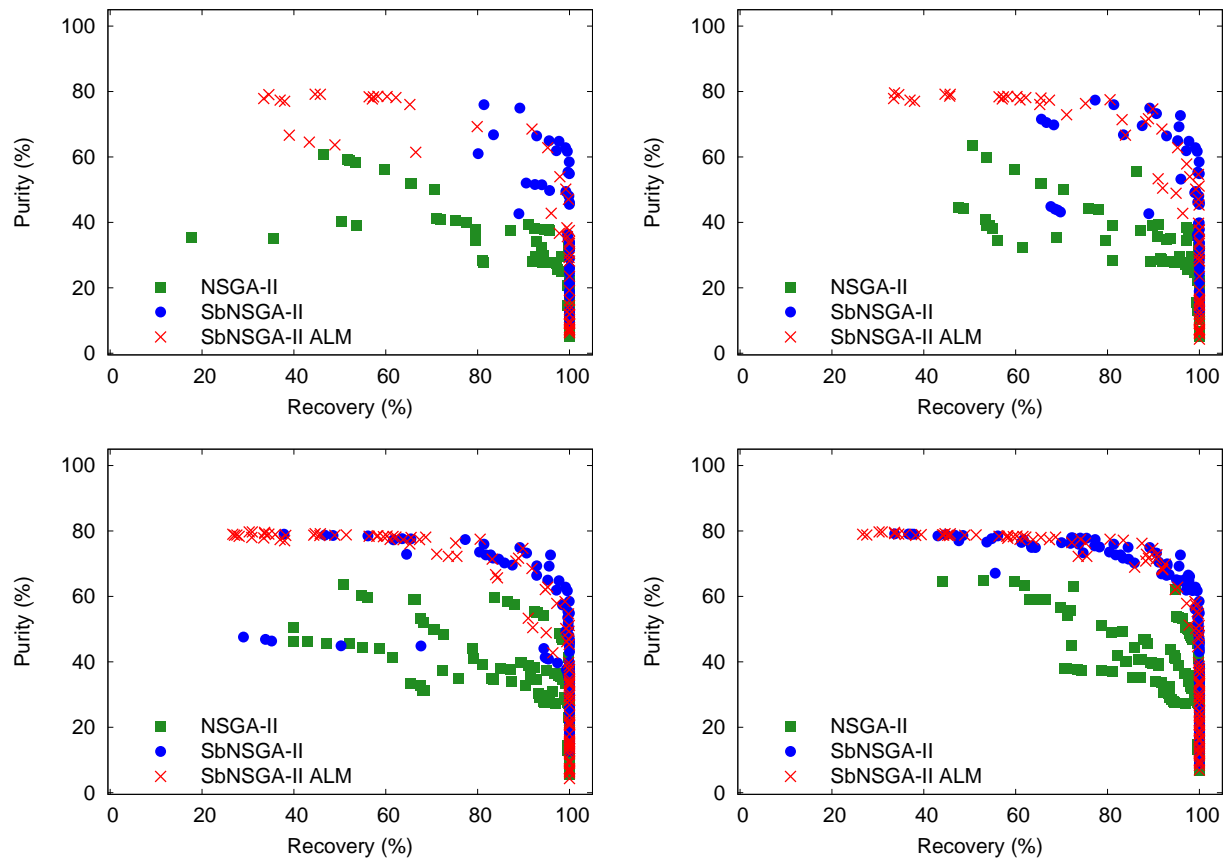


Figure 7: Results using different optimisation strategies. Top left: 64 evaluations of the high-fidelity simulation, right: 96. Bottom left: 176, right: 256. The results for the different strategies are produced by combining five Pareto fronts resulting from using five different initial training data.

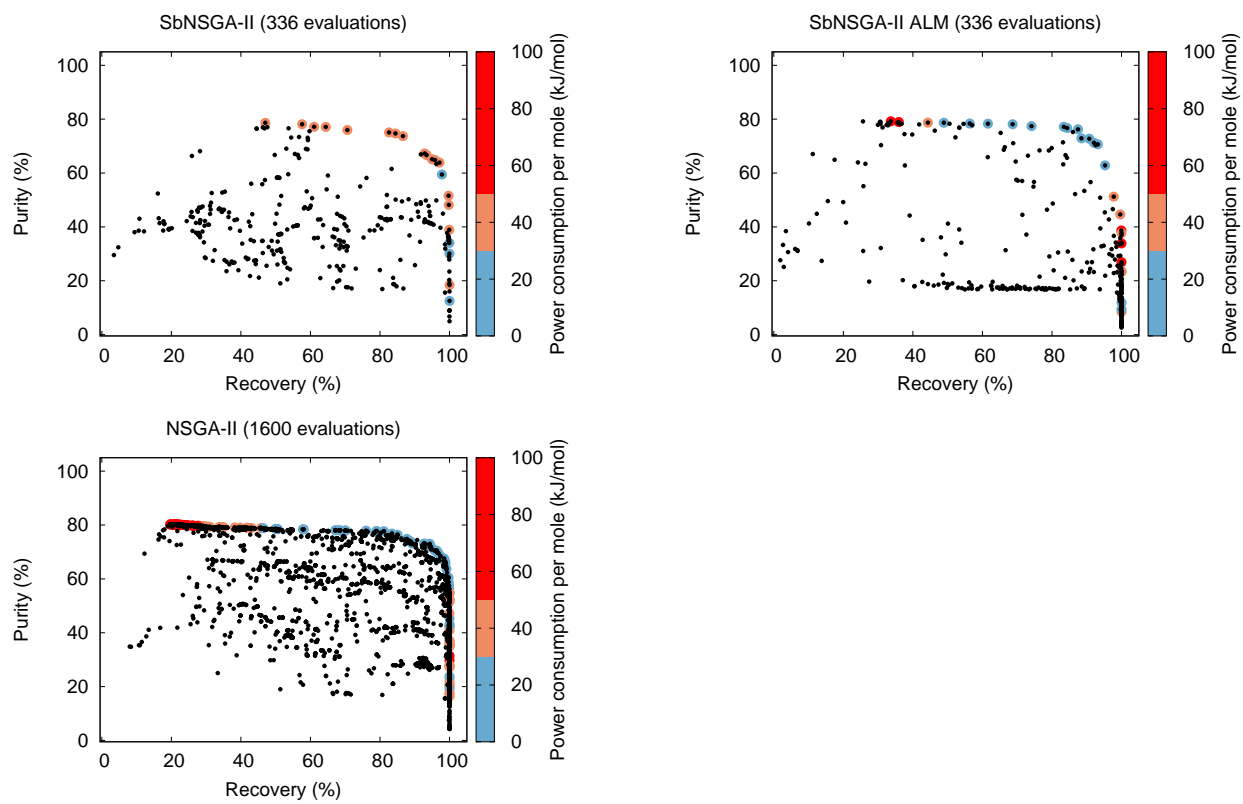


Figure 8: Power consumption per mole of CO₂ captured.

While the optimisation performed above only considered the purity and recovery, the power consumption is essential for the design of an effective VPSA system. For the design solutions, we also calculated the system's power consumption defined by (4). Fig. 8 shows on the Pareto fronts the power consumption per mole highlighted by the color range. The power consumption for our SbNSGA-II ALM and NSGA-II agree along the whole Pareto front and both show a minimum at the bend of the Pareto front. This is to be expected because at the bend the system recovers most of the CO₂ at a high purity. Interestingly, the SbNSGA-II optimisation does not show a clear minimum at the bend of the Pareto front. The reason for this higher power consumption of the solutions is not obvious but is probably due to the limited exploration of the design space of SbNSGA-II.

5.3. Discussion

The low values of the correlation parameters ℓ_{purity} and $\ell_{recovery}$ tell us that the purge-to-feed ratio has a large influence on the purity and that the feed/purge time and the feed flowrate have a large influence on the recovery. The purge flow controlled by the purge-to-feed ratio consists mostly of the light component and thus the purity of the heavy component is reduced by a large value of this ratio. If the product of the feed/purge time and the feed flowrate are too large, a part of the heavy component will break through during the feed step which will reduce the recovery. Thus these relationships are expected and it is more interesting to investigate why the other parameters have a lower influence.

To investigate the influence of the design parameters with higher values of the correlation parameters, we need to also consider the high level visualisation presented in Fig. 9. The link between the different Pareto front segments highlighted with colours (cyan, red, green, and blue) and the design solutions is revealed in Fig. 9. This link enables us to better understand the underlying process. For example, in the regime of high purity, we can see that the purge valve should be closed during the feed/purge step. This is reasonable as the CO₂ is the heavy product taken out during the feed step and to purge the second bed with a gas mixture of high N₂ is likely to deteriorate the CO₂ level of the outgoing product.

The two correlation parameters for the feed temperature are at the upper limits and thus it seems that the feed temperature has little influence on the outputs. Since the CO₂ isotherm on silicalite is nonlinear at the lower temperature and becomes more linear towards the upper boundary of the feed temperature, we would have assumed that the feed temperature has at least some influence on the purity and recovery. However, the high level visualisation in Fig. 9 shows that the feed temperature varies without any discernible trend, confirming that the feed temperature has very little influence on either the purity or recovery. This very low influence can be explained by the large column headers in the CCSi VPSA unit (see Table 2). The column headers act like heat exchangers so that, even for the highest feed temperature, the gas entering the adsorption column is cooled to almost room temperature. Thus the feed temperature has almost no influence on the processes in the adsorption column and this validates the use of the estimated correlation parameters in combination with Fig. 9 as a sensitivity analysis tool.

According to the correlation parameters, the vacuum and adsorption pressure have some influence on the recovery. The two pressure values influence the working capacity of the system. For a larger pressure range, i.e. the difference between vacuum and adsorption pressure, the working capacity of the adsorbent is larger and thus the amount of CO₂ adsorbed in each cycle increases. This increases the recovery because the losses of CO₂ in each cycle to the vent are not increasing accordingly. On the other hand, the correlation parameter for the purity is at the upper limit, indicating, similar to the feed temperature, that the pressures have little influence on the purity. While Fig. 9 confirms that there is no trend in the feed pressure, this is not true for the vacuum pressure which is at the lower limit except for the design points with very low purity, i.e. at or close to the feed concentration. It is intuitively clear that a sufficiently low vacuum pressure is required to efficiently remove the adsorbed CO₂ from the bed. Thus the best performance in terms of CO₂ purity is achieved when the vacuum pressure is at its lowest point in the specified range. This shows that for high values of the correlation parameters it is essential to also look at the high level visualisation.

The product of feed/purge time and feed flowrate is crucial to achieve high purity and high recovery. If the product is large, the adsorption front will break through at the vent end of the column. This will lead to high purity CO₂ streams in the purge step but will yield a low recovery. If the product is small, none of the

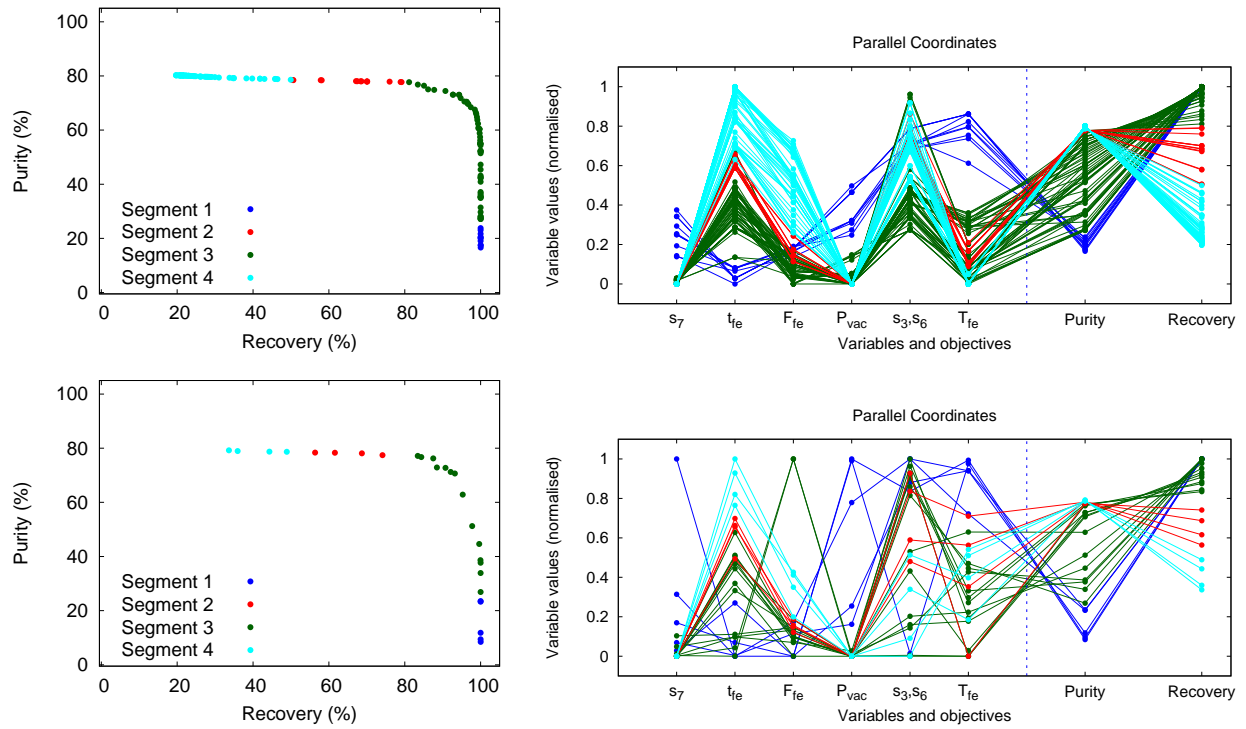


Figure 9: Pareto front divided into segments (Left) by colours. Parallel coordinates visualisation (Right) shows that the design solutions associated with different Pareto front segments are clustered. The top figures are for the NSGA-II (1600 evaluations), and the bottom ones are for the SbNSGA-II ALM (320 evaluations). The SbNSGA-II ALM, with only 320 evaluations, provides a sufficient approximation of the Pareto curve as well as the associated design solutions.

415 CO₂ will leave the column at the vent end and thus the system will have a high recovery but a low purity. To reach a system with a high purity and high recovery the product of feed/purge time and feed flowrate needs to be an intermediate value which ensures that the adsorption front almost reaches the vent end of the adsorption column. As we move the adsorption front closer to the vent side of the column, a small amount of CO₂ is breaking through at the vent side. By increasing the purge-to-feed ratio, a larger percentage of this CO₂ is recovered which explains the influence of the purge-to-feed ratio on the recovery.

420 It is interesting to note that the power consumption per mole of CO₂ shown in Fig. 8 increases from the bend for decreasing purity but goes back to the minimum value for very low purities. These low purities are similar to the feed mole fraction and could be achieved without any separation at the vent side of the system. However, the purity and recovery are always calculated at the vacuum side. The high level visualisation shows that, for these cases, both the feed/purge time and feed flowrate are at very low values. This indicates that the flow through the system is low and that only a small amount of the feed gas leaves the system at the vent side.

430 While the power consumption and productivity are not as important for a demonstration system, these are critical in large scale applications. For example, the low vacuum pressure of 0.02 bar, while achievable in a small lab-scale or demonstration unit, would be difficult to reach in large scale industrial applications and would lead to large power consumption. Furthermore, a high productivity of the process is crucial for small and light systems. This is important for small scale portable units, e.g. portable oxygen concentrators, as well as for large scale industrial units, e.g. carbon capture units for fossil fuel power station.

435 This discussion shows that the correlation parameters together with the high level visualisation in Fig. 9 are a valuable tool in the analysis of the optimisation results. The high level visualisation, in particular, helps to understand the interplay between different decision variables and the trade-off between purity and recovery. Multi-objective visualization methods can be a valuable tool to explore the relationship between competing objectives for the promising VPSA designs (Žilinskas et al., 2015). For example, the lines connecting the feed/purge time and feed flowrate for each segment of the Pareto front, i.e. lines of the same colour, are almost all crossing. This indicates that the amount of feed gas ($t_{fe} \times F_{fe}$) for each segment of the Pareto front is almost constant. This shows that the feed/purge time and feed flowrate are linked with respect to purity and recovery, i.e. purity and recovery depend only on the amount of feed gas. However, the two design parameters are independent with respect to power consumption and productivity and this will be investigated in a future publication.

445 6. Conclusion

In this work, we propose the first multi-objective optimisation method that involves the use of kriging surrogate models for providing fast solutions to the design of VPSA systems. We have demonstrated its efficiency against the standard NSGA-II and verified its robustness by analysing the performance using five different initial training data. It is well-known that surrogate based optimisation involving multiple objectives is challenging. We propose a transformation of the objective space during the training phase of the surrogate models (for the purity and recovery) to constrain their response space to $[0, 1]$ (i.e., to obey the physical constraints). This modification leads to better predictions of the Pareto front.

455 We have developed two implementations, SbNSGA-II and SbNSGA-II ALM, which both outperform the standard NSGA-II. SbNSGA-II ALM has been shown to be more robust than SbNSGA-II, although SbNSGA-II displays a stronger convergence when close to the optima. Both SBO methods achieved a reduction in the number of VPSA model simulations by a factor between 2 and 5, depending on the desired resolution of the Pareto curve approximation. This reduction in the number of high-fidelity simulations directly translates to a reduction in the computational time and thus enables the efficient optimisation of VPSA systems. Furthermore, the power consumption of the solutions of SbNSGA-II ALM is comparable with the solutions of NSGA-II, while the solutions of SbNSGA-II show a significantly higher power consumption in the high purity and high recovery region. While the inclusion of power consumption and productivity as objectives is beyond the scope of this contribution, the multi-objective SBO framework is capable of handling more than two objectives and this will be investigated in the future.

465 A further advantage of the SBO methods is the calculation of the correlation parameters. Together with the high level visualisation these parameters offer significant information about the interplay between the

decision variables and objectives. This insight can be used to analyse the optimisation results and to guide further optimisation studies.

Appendix A. Supplementary information

Appendix A.1. Carbon Capture and Storage Interactive Unit

470 The Carbon Capture and Storage interactive (CCSi) unit was built by the University of Edinburgh in collaboration with SciFun (<http://www.scifun.ed.ac.uk>). The rig's primary function is to demonstrate the idea of carbon capture using vacuum/pressure swing adsorption (VPSA) at science and engineering fairs. Once it is not used for this purpose anymore, it will be refitted and recalibrated for scientific experiments. The rig is shown in Fig. A.10.

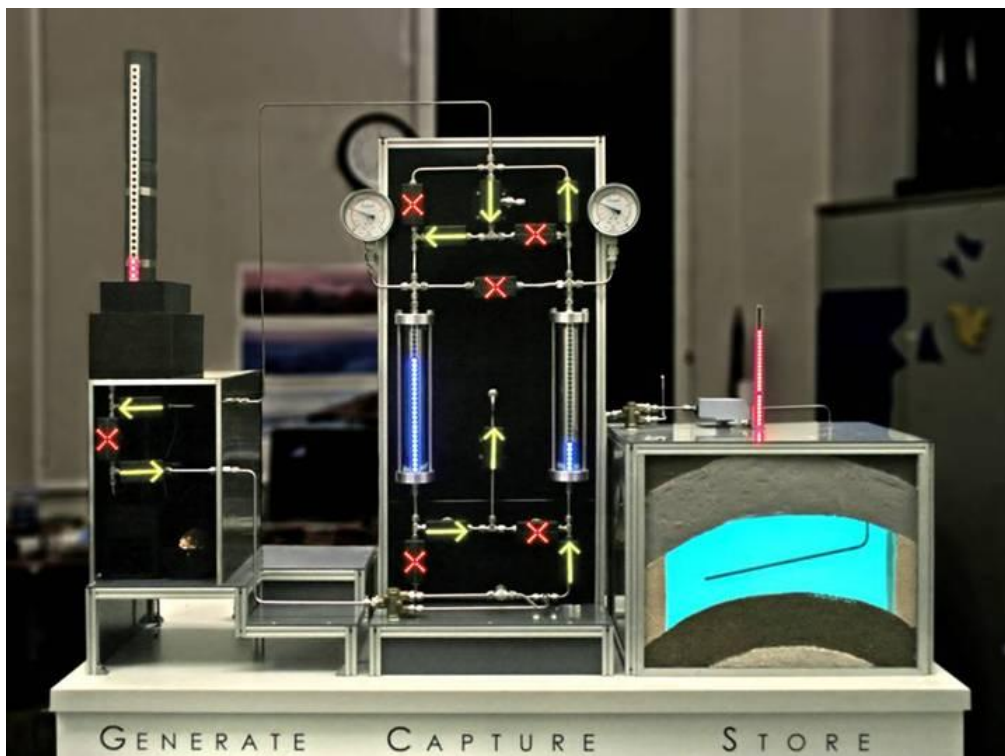


Figure A.10: Picture of the CCSi unit

475 The rig is split into three sections: generation, capture and storage. Of these the generation and storage sections are only representations of the actual process while the capture unit is a fully functioning VPSA unit. The generation section represents a typical power plant in which electricity is generated by the burning of fossil fuels. The resulting flue gas is routed to the capture section. Since the combustion is only simulated, compressed air and carbon dioxide cylinders are used to generate a simulated flue gas of 1/6 CO_2 with the remainder air.

480 As shown in the diagram the rig uses LED lights which show the valve configurations. In Fig. A.10, the valves are set to divert the flue gas to the capture unit which separates CO_2 from the flue gas. The capture unit makes use of a six step, two column VSA process. The flue gas enters at a feed flowrate of 3 L/min and a pressure of 1.3 bar. The raffinate leaves the top of the column at atmospheric pressure through the power station chimney stack. The chimney stack has a sensor with inbuilt LED indicators which measures the concentration of CO_2 leaving in the raffinate stream. The extract (CO_2 rich) stream leaves the process under a vacuum of 0.1 bar. The columns are 0.3 m long and have an inner diameter of 0.035 m. The two columns are filled with approximately 0.08 kg of silicalite pellets (HiSiv 3000) supplied by UoP. This amount of silicalite fills about 0.12 m of each column. The remainder (evenly split between the top and bottom of the columns) is filled with solid beads.

485

490

Once the CO₂ has been separated, the gas flows to the storage section. The storage section shown in Fig. A.10 illustrates the use of a geological structure for the storage of CO₂.

Appendix A.2. Governing equations

495 The following paragraphs present the governing mass, energy and momentum balance equations for the column, CSTR, valve and feed unit.

Column. The flow through the packed bed is described by the axial dispersed plug flow model. More detailed models, e.g. including radial dispersion, are generally not necessary (Ruthven, 1984) while the ideal plug flow can be approximated by letting the dispersion coefficient go to zero. For the plug flow model the mass and energy balance in the column are given by the following equations.

$$\frac{\partial c_i}{\partial t} = -\frac{1-\epsilon}{\epsilon} \frac{d\bar{Q}_i}{dt} - \frac{\partial(v \cdot c_i)}{\partial z} - \frac{\partial J_i}{\partial z}, \quad (\text{A.1})$$

$$\frac{d\bar{Q}_i}{dt} = \epsilon_p \frac{dc_i^m}{dt} + (1-\epsilon_p) \frac{d\bar{q}_i}{dt} = k_i^p \frac{A_p}{V_p} (c_i - c_i^m), \quad (\text{A.2})$$

$$\frac{d\bar{q}_i}{dt} = k_i^{cr} \frac{3}{R_{cr}} (q_i^* - \bar{q}_i), \quad (\text{A.3})$$

$$\begin{aligned} \epsilon \frac{\partial \check{U}_f}{\partial t} = & -(1-\epsilon) \frac{\partial \check{U}_p}{\partial t} - \epsilon \frac{\partial(\check{H}_f \cdot v)}{\partial z} - \frac{\partial J_T}{\partial z} \\ & - \sum_{i=1}^{N_c} \frac{\partial(J_i \check{H}_i)}{\partial z} - h_w \frac{A_c}{V_c} (T_f - T_w), \end{aligned} \quad (\text{A.4})$$

$$\frac{\partial \check{U}_p}{\partial t} = \epsilon_p \frac{\partial \check{U}_{p,f}}{\partial t} + (1-\epsilon_p) \frac{\partial \check{U}_{p,s}}{\partial t} = h_p \frac{A_p}{V_p} (T_f - T_p), \quad (\text{A.5})$$

$$\rho_w \hat{c}_{P,w} \frac{\partial T_w}{\partial t} = -h_w \frac{A_c}{V_w} (T_w - T_f) - U \frac{A_{co}}{V_w} (T_w - T_\infty). \quad (\text{A.6})$$

Due to the large heat capacity of the wall the wall temperature calculated by Eq. A.6 is almost constant; thus in this contribution the wall is assumed to be isothermal, i.e. Eq. A.6 is not used.

The mass and heat transfer into the pellet depend on the mass and heat transfer coefficient and the ratio of the pellet surface area to pellet volume. In this contribution the pellets were assumed to be spherical so that $A_p/V_p = 3/R_p$. The mass transfer into the adsorbent crystals assumes that the crystals are spherical with radius R_{cr} . The pellet void fraction, ϵ_p , is the ratio of the macropore volume over the total pellet volume. The macropore LDF parameter is calculated from the effective macropore diffusivity

$$D_{m,i}^e = \frac{\epsilon_p}{\tau} \left(\frac{1}{D_i^m} + \frac{1}{D_i^K} \right)^{-1} \quad (\text{A.7})$$

where the individual diffusion coefficient are given by

$$\begin{aligned} D_i^m &= 1.60 \times 10^{-7} \sqrt{\frac{T^3 (MW_1^{-1} + MW_2^{-1})}{P \sigma_{12}^2 \Omega_{12}}} \\ D_i^K &= 97 r_{pore} \sqrt{\frac{T}{MW_i}} \end{aligned}$$

500 Here MW_i is the molecular weight in g mol⁻¹, σ_{12} is the collision diameter from the Lennard-Jones potential in Å, Ω_{12} is a function depending on the Lennard-Jones force constant and temperature and r_{pore} the mean pore radius in m.

The thermal diffusive flux J_T and the diffusive flux of component i in the fluid phase J_i are given by

$$\begin{aligned} J_T &= -\lambda_f^L \epsilon \frac{\partial T_f}{\partial z} - \lambda_p^L (1-\epsilon) \frac{\partial T_p}{\partial z}, \\ J_i &= -D_i^L c_T \frac{\partial x_i}{\partial z}. \end{aligned} \quad (\text{A.8})$$

Here the axial thermal conductivity in the fluid and pellet are given by λ_f^L and λ_p^L , respectively, and the axial dispersion coefficient by D_i^L .

The boundary conditions for the gas phase concentrations and the enthalpy are given by the Danckwerts boundary conditions for flow into the column and the no diffusive flux for flow out of the column. With the conventions that the positive flow directions is from 0 to L_c these can be written in a combined form as

$$J_T|_{z=0} = \frac{v + |v|}{2} (\check{H}_{f,0^-} - \check{H}_{f,0}) \quad (\text{A.9})$$

$$J_T|_{z=L_c} = \frac{v - |v|}{2} (\check{H}_{f,L_c^+} - \check{H}_{f,L_c}) \quad (\text{A.10})$$

$$J_i|_{z=0} = \frac{v + |v|}{2} (c_{i,0^-} - c_{i,0}) \quad (\text{A.11})$$

$$J_i|_{z=L_c} = \frac{v - |v|}{2} (c_{i,L_c^+} - c_{i,L_c}) \quad (\text{A.12})$$

where the superscripts $-$ and $+$ indicate the concentration values to the left and right of the boundary, respectively.

The constitutive equations for the energy terms are given by

$$\begin{aligned} \check{U}_f &= c_T \check{U}_{f,ref} + c_T \int_{T_{ref}}^{T_f} \check{c}_V dT, \\ \check{H}_f &= c_T \check{H}_{f,ref} + c_T \int_{T_{ref}}^{T_f} \check{c}_P dT, \\ \check{U}_p &= \epsilon_p \check{U}_{p,f} + (1 - \epsilon_p) \check{U}_{p,s} \\ \check{U}_{p,f} &= c_T^m \check{U}_{p,f,ref} + c_T^m \int_{T_{ref}}^{T_p} \check{c}_V^m dT, \\ \check{U}_{p,s} &= \check{U}_{sol} + \check{U}_{ads}, \\ \check{U}_{sol} &= \rho_{cry} \check{U}_{sol,ref} + \rho_{cry} \int_{T_{ref}}^{T_p} \hat{c}_{P,sol} dT, \\ \check{U}_{ads} &= \check{H}_{ads} = q_T \check{H}_{ads,ref} + q_T \int_{T_{ref}}^{T_p} \check{c}_{P,ads} dT - (-\Delta \check{H}_{ads})_{T_p}, \\ (-\Delta \check{H}_{ads})_{T_p} &= \sum_{i=1}^{N_c} \int_0^{q_i} (-\Delta \check{H}_i)_{T_p, q_j \neq i} dq, \\ \check{H}_i &= \check{H}_{i,ref} + \int_{T_{ref}}^{T_f} \check{c}_{P,i} dT. \end{aligned}$$

505 \check{U}_{sol} and \check{U}_{ads} are the internal energy per unit volume in the adsorbent and the adsorbed phase, respectively. Here the tilde indicates the molar quantities and the subscript “ref” indicates the reference value at T_{ref} and P_{ref} . The total concentration in the fluid phase and in the macropore are given by c_T and c_T^m , respectively. ρ_{cry} is the crystal density, q_T is the total adsorbed concentration in the micropore and $\hat{c}_{P,sol}$ is the specific heat capacity at constant pressure in the solid phase. The total heat of adsorption per unit volume $\Delta \check{H}_{ads}$ is calculated from the component heat of adsorptions $\Delta \check{H}_i$.

The molar heat capacities at constant volume \check{c}_V in the fluid phase and at constant pressure in the fluid \check{c}_P and in the adsorbed phase $\check{c}_{P,ads}$ are calculated from the respective component heat capacities by

$$\begin{aligned} \check{c}_V &= \sum_{i=1}^{N_c} x_i \check{c}_{V,i}, & \check{c}_P &= \sum_{i=1}^{N_c} x_i \check{c}_{P,i}, \\ \check{c}_{P,ads} &= \sum_{i=1}^{N_c} \frac{q_i}{q_T} \check{c}_{P,i}, & \check{c}_V^m &= \sum_{i=1}^{N_c} x_i^m \check{c}_{V,i}. \end{aligned}$$

The superscript m indicates the macropore concentrations and heat capacities and x_i are the component mole fractions.

The pressure drop along the column is given by the Ergun equation

$$-\frac{dP}{dz} = \frac{150\mu(1-\epsilon)^2}{\epsilon^2(2R_p)^2}v + \frac{1.75(1-\epsilon)\rho_f}{\epsilon 2R_p}v|v|. \quad (\text{A.13})$$

Continuously stirred tank. The column header and the tanks are modelled as continuously stirred tanks with an arbitrary number of connections N_I . The governing equations are given by

$$F_j = -F_{n_j}, \quad j = 1, \dots, N_I, \quad (\text{A.14})$$

$$V \frac{dc_i}{dt} = \sum_{j=1}^{N_I} \left(\frac{F_j + |F_j|}{2} \frac{c_{i,n_j}}{c_{T,n_j}} + \frac{F_j - |F_j|}{2} \frac{c_i}{c_T} \right), \quad (\text{A.15})$$

$$\frac{dU_f}{dt} = \sum_{j=1}^{N_I} \left(\frac{F_j + |F_j|}{2} \tilde{H}_{f,n_j} + \frac{F_j - |F_j|}{2} \tilde{H}_f \right) + h_w A (T_w - T_f). \quad (\text{A.16})$$

The pressure in the tank is calculated from the temperature and total gas concentration by the ideal gas law. Equation A.14 requires that the flow rate is provided by the neighbouring units. For the inlets connected to either a valve or feed unit, this is fulfilled since these units calculate the flow rate from the composition and temperature. However, for the inlet connected to the column, the flow rate needs to be provided. Here the flow rate is calculated by

$$F = k(P_n - P). \quad (\text{A.17})$$

With a sufficiently large value of the coefficient k , the pressure drop between the column header and column will be negligible.

Valve. The flow rate through the valve is calculated by

$$F = s_j c_v c_T \sqrt{\frac{|P_0 - P_1|}{\rho_f}} \quad (\text{A.18})$$

515 where c_v is the valve coefficient, P_0 and P_1 are the pressures at the two inlets, respectively, c_T is the total concentration and ρ_f is the fluid density. The stem position, s_j , of the valve indicates if the valve is open or closed. Here 0 indicates a closed valve, 0.5 a half open valve and 1 a fully open valve. The simulator allows the opening and closing of valves at specified times. Through the coordinated operation of the valves, different steps in the cycle, e.g. blowdown and pressure equalisation, are performed (see Table 1). Pressure, 520 temperature and gas composition are passed through the valve to the neighbouring units.

Feed unit. Pressure, temperature and feed composition are kept constant at the initial conditions. The flow rate is set according to

$$F = \begin{cases} F_i, & \left| \frac{P - P_n}{P} \right| > 0.01, \\ F_i, 0.5 (1 + \cos(100\pi (0.01 - \left| \frac{P - P_n}{P} \right|))), & \text{otherwise.} \end{cases} \quad (\text{A.19})$$

Here, F_i is the set flow rate and P and P_n are the pressures in the feed unit and the neighbouring unit, respectively.

The feed unit also calculates the number of moles of each component to pass through the outlet and the power consumption of the blower or vacuum pump. The number of moles is calculated through the following ODEs

$$\frac{dn_i}{dt} = \frac{F + |F|}{2} \frac{c_{i,n}}{c_{T,n}} + \frac{F - |F|}{2} \frac{c_i}{c_T}, \quad (\text{A.20})$$

with initial values of $n_i = 0$. Since the sign on the right hand side of equation A.20 changes with the flow direction, this equation gives a positive result if more molecules have entered the feed unit than have left the feed unit and a negative results otherwise.

The power consumption is calculated by

$$Pow = \begin{cases} \frac{\gamma}{\gamma-1} RT F \left[\left(\frac{P_n}{P_{atm}} \right)^{\frac{\gamma-1}{\gamma}} - 1 \right], & \text{if } P_n > P_{atm} \text{ and } F < 0, \\ \frac{\gamma_n}{\gamma_n-1} RT_n F \left[\left(\frac{P_{atm}}{P_n} \right)^{\frac{\gamma_n-1}{\gamma_n}} - 1 \right], & \text{if } P_n < P_{atm} \text{ and } F > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.21})$$

$$\frac{dW}{dt} = Pow, \quad W(0) = 0 \quad (\text{A.22})$$

Here $\gamma = \frac{\tilde{c}_P}{\tilde{c}_V}$ is the ratio of the molar heat capacities at constant pressure and constant volume.

Table A.4: Nomenclature

Parameter	Description	Units
A	Surface area	m^2
A_c	Inner surface area of the column	m^2
A_{co}	Outer surface area of the column	m^2
A_p	Pellet surface area	m^2
b_i	Equilibrium parameter of comp. i	bar^{-1}
c_i	Concentration of component i	mol m^{-3}
\hat{c}_P	Specific heat capacity	$\text{J kg}^{-1} \text{K}^{-1}$
\tilde{c}_P	Molar heat capacity at constant pressure	$\text{J mol}^{-1} \text{K}^{-1}$
c_T	Total concentration	mol m^{-3}
c_v	Valve coefficient	$\text{m}^2 \sqrt{10^{-5}}$
\tilde{c}_V	Molar heat capacity at constant volume	$\text{J mol}^{-1} \text{K}^{-1}$
D_i^K	Knudsen diffusion coefficient of comp. i	$\text{m}^2 \text{s}^{-1}$
D_i^m	Molecular diffusion coefficient of comp. i	$\text{m}^2 \text{s}^{-1}$
$D_{m,i}^e$	Effective macropore diffusivity of comp. i	$\text{m}^2 \text{s}^{-1}$
D_i^L	Effective axial dispersion coefficient of comp. i	$\text{m}^2 \text{s}^{-1}$
F	Molar flow rate	mol s^{-1}
F_{fe}	Feed flow rate	mol s^{-1}
h_p	Pellet-to-fluid heat transfer coefficient	$\text{W m}^{-2} \text{K}^{-1}$
h_w	Fluid-to-wall heat transfer coefficient	$\text{W m}^{-2} \text{K}^{-1}$
$-\Delta \tilde{H}_i$	Heat of adsorption of comp. i	J mol^{-1}
\tilde{H}	Enthalpy per unit volume	J m^{-3}
\tilde{H}	Molar enthalpy	J mol^{-1}
J_i	Diffusive flux of comp. i in the fluid phase	$\text{mol m}^{-2} \text{s}^{-1}$
J_T	Thermal diffusive flux in the fluid phase	$\text{J m}^{-2} \text{s}^{-1}$
k	Flow rate coefficient	$\text{mol s}^{-1} \text{bar}^{-1}$
k_i^{cr}	LDF coefficient of comp. i in the adsorbent crystal	m s^{-1}
k_i^p	LDF coefficient of comp. i in the pellet	m s^{-1}
L_c	Column length	m
n_i	Number of moles of comp. i	mol
$n_{j,i}^k$	Cumulative number of moles of comp. i through unit j at the end of the k -th cycle	
P	Pressure	bar
P_{vac}	Vacuum pressure	bar
\bar{q}_i	Sorbate concentration of comp. i	mol m^{-3}
q_s	Saturation capacity	mol m^{-3}
q_i^*	Sorbate concentration of comp. i at equilibrium	mol m^{-3}
q_T	Total adsorbed concentration	mol m^{-3}

Table A.4: Nomenclature

Parameter	Description	Units
Q_i	Concentration of comp. i in the pellet	mol m^{-3}
R	Gas constant	$\text{J mol}^{-1} \text{K}^{-1}$
R_c	Column radius	m
R_{cr}	Crystal radius	m
R_p	Pellet radius	m
s_j	Stem position of valve j	-
t_{fe}	Feed/purge time	s
T	Temperature	K
T_{fe}	Feed temperature	K
U	Internal energy	J
\check{U}	Internal energy per unit volume	J m^{-3}
v	Interstitial velocity	m s^{-1}
V	Volume	m^3
V_c	Column volume	m^3
V_p	Pellet volume	m^3
V_w	Column wall volume	m^3
W	Work	J
x_i	Mole fraction of comp. i	
γ	Ratio of heat capacities at constant pressure and volume	
δ_w	Column wall thickness	m
ϵ	Column void fraction	-
ϵ_p	Pellet void fraction	-
λ^L	Effective axial thermal dispersion coefficient	$\text{W m}^{-1} \text{K}^{-1}$
μ	Fluid viscosity	Pas
ρ	Density	kg m^{-3}
τ	Pore tortuosity of the pellet	-
Subscript		
ads	Adsorbed phase	
atm	Atmospheric	
c	Column	
f	Fluid phase	
n	Neighbouring unit	
p	Pellet	
ref	Reference state	
s	Solid phase	
sol	Adsorbent phase	
w	Column wall	
∞	Surroundings	
Superscript		
cr	Crystal	
L	Axial	
m	Macropore	
p	Pellet	

Acknowledgements

The authors gratefully acknowledge the support provided by the UK Engineering and Physical Sciences Research Council (EPSRC) through the funds provided to the EP/G062129/1 and EP/G02037X/1 projects.

⁵³⁰ Agarwal, A., Biegler, L., & Zitney, S. (2008). Simulation and Optimization of Pressure Swing Adsorption Systems Using Reduced-Order Modeling. *Ind. Eng. Chem. Res.*, 48, 2327–2343.

- Banu, A., Friedrich, D., Brandani, S., & Düren, T. (2013). A Multiscale Study of MOFs as Adsorbents in H₂ PSA Purification. *Ind. Eng. Chem. Res.*, *52*, 9946–9957.
- Beck, J., Friedrich, D., Brandani, S., Guillas, S., & Fraga, E. S. (2012). Surrogate based Optimisation for Design of Pressure Swing Adsorption Systems. *Proc. ESCAPE22*, (pp. 1217–1221).
- Beck, J., & Guillas, S. (2014). Sequential Design with Mutual Information for Computer Experiments (MICE): Emulation of a Tsunami Model. *ArXiv e-prints*, . arXiv:1410.0215.
- Biegler, L., Jiang, L., & Fox, V. (2005). Recent Advances in Simulation and Optimal Design of Pressure Swing Adsorption Systems. *Sep. Purif. Rev.*, *33*, 1–39.
- Da Silva, F. A., Silva, J. A., & Rodrigues, A. E. (1999). A General Package for the Simulation of Cyclic Adsorption Processes. *Adsorption*, *5*, 229–244. URL: <http://link.springer.com/article/10.1023/A%3A1008974908427>. doi:10.1023/A:1008974908427.
- De Oliveira, V., Kedem, B., & Short, D. (1997). Bayesian Prediction of Transformed Gaussian Random Fields. *J. Am. Stat. Assoc.*, *92*, 1422–1433.
- Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.*, *6*, 182–197.
- Diggle, P., & Ribeiro, P. (2007). *Model-based Geostatistics*. Springer.
- Farooq, S., Ruthven, D., & Boniface, H. (1989). Numerical simulation of a pressure swing adsorption oxygen unit. *Chem. Eng. Sci.*, *44*, 2809–2816.
- Fiandaca, G., Fraga, E. S., & Brandani, S. (2009). A multi-objective genetic algorithm for the design of pressure swing adsorption. *Eng. Optimiz.*, *41*, 833–854.
- Forrester, A. I., & Keane, A. J. (2009). Recent advances in surrogate-based optimization. *Prog. Aerosp. Sci.*, *45*, 50–79.
- Friedrich, D., Ferrari, M., & Brandani, S. (2013). Efficient Simulation and Acceleration of Convergence for a Dual Piston Pressure Swing Adsorption System. *Ind. Eng. Chem. Res.*, *52*, 8897–8905.
- Golden, T., & Sircar, S. (1994). Gas Adsorption on Silicalite. *J. Colloid Interface Sci.*, *162*, 182–188.
- Gramacy, R., & Lee, H. (2012). Cases for the nugget in modeling computer experiments. *Stat. Comput.*, *22*, 713–722.
- Haghpanah, R., Majumder, A., Nilam, R., Rajendran, A., Farooq, S., Karimi, I., & Amanullah, M. (2013). Multiobjective Optimization of a Four-Step Adsorption Process for Postcombustion CO₂ Capture Via Finite Volume Simulation. *Ind. Eng. Chem. Res.*, *52*, 4249–4265.
- Hasan, M., Baliban, R., Elia, J., & Floudas, C. (2012). Modeling, Simulation, and Optimization of Post-combustion CO₂ Capture for Variable Feed Concentration and Flow Rate. 2. Pressure Swing Adsorption and Vacuum Swing Adsorption Processes. *Ind. Eng. Chem. Res.*, *51*, 15665–15682.
- Hasan, M., First, E., & Floudas, C. (2013). Cost-effective CO₂ capture based on in silico screening of zeolites and process optimization. *Phys. Chem. Chem. Phys.*, *15*, 17601–17618.
- Hindmarsh, A., Brown, P., Grant, K., Lee, S., Serban, R., Shumaker, D., & Woodward, C. (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM T. M. Software*, *31*, 363–396.
- Ishibashi, M., Ota, H., Akutsu, N., Umeda, S., Tajika, M., Izumi, J., Yasutake, A., Kabata, T., & Kageyama, Y. (1996). Technology for removing carbon dioxide from power plant flue gas by the physical adsorption method. *Energy Conversion and Management*, *37*, 929–933. URL: <http://www.sciencedirect.com/science/article/pii/0196890495002790>. doi:10.1016/0196-8904(95)00279-0.

- Janusevskis, J., Le Riche, R., Ginsbourger, D., & Girdziusas, R. (2012). Expected Improvements for the Asynchronous Parallel Global Optimization of Expensive Functions: Potentials and Challenges. In *Lect. Notes Comput. Sc.* (pp. 413–418). Springer.
- Jiang, L., Biegler, L., & Fox, V. (2003). Simulation and optimization of pressure-swing adsorption systems for air separation. *AIChE Journal*, *49*, 1140–1157.
- Jin, R., Chen, W., & Simpson, T. (2001). Comparative studies of metamodelling techniques under multiple modelling criteria. *Struct. Multidiscip. O.*, *23*, 1–13.
- Jones, D. (2001). A Taxonomy of Global Optimization Methods Based on Response Surfaces. *J. Global Optim.*, *21*, 345–383.
- Jones, D., Schonlau, M., & Welch, W. (1998). Efficient Global Optimization of Expensive Black-Box Functions. *J. Global Optim.*, *13*, 455–492.
- Krishnamurthy, S., Rao, V. R., Guntuka, S., Sharratt, P., Haghpanah, R., Rajendran, A., Amanullah, M., Karimi, I. A., & Farooq, S. (2014). CO₂ capture from dry flue gas by vacuum swing adsorption: A pilot plant study. *AIChE Journal*, *60*, 1830–1842. URL: <http://onlinelibrary.wiley.com/doi/10.1002/aic.14435/abstract>. doi:10.1002/aic.14435.
- Krutka, H., & Sjostrom, S. (2011). *Evaluation of solid sorbents as a retrofit technology for CO₂ capture from coal-fired power plants – Final technical report*. Technical Report Report number 05649FR01 ADA Environmental Solutions.
- Liu, Y., & Sun, F. (2013). Parameter estimation of a pressure swing adsorption model for air separation using multi-objective optimisation and support vector regression model. *Expert. syst. appl.*, *40*, 4496–4502.
- Liu, Z., Grande, C., Li, P., Yu, J., & Rodrigues, A. (2011). Multi-bed Vacuum Pressure Swing Adsorption for carbon dioxide capture from flue gas. *Sep. Purif. Technol.*, *81*, 307–317.
- Liu, Z., Wang, L., Kong, X., Li, P., Yu, J., & Rodrigues, A. E. (2012). Onsite CO₂ Capture from Flue Gas by an Adsorption Process in a Coal-Fired Power Plant. *Industrial & Engineering Chemistry Research*, *51*, 7355–7363. URL: <http://dx.doi.org/10.1021/ie3005308>. doi:10.1021/ie3005308.
- Loeppky, J., Sacks, J., & Welch, W. (2009). Choosing the Sample Size of a Computer Experiment: A Practical Guide. *Technometrics*, *51*.
- McKay, M., Beckman, R., & Conover, W. (1979). Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, *21*, 239–245.
- Montgomery, D. (2007). *Design and Analysis of Experiments*. (6th ed.). John Wiley & Sons, Ltd.
- Nikolic, D., Giovanoglou, A., Georgiadis, M. C., & Kikkinides, E. S. (2008). Generic Modeling Framework for Gas Separations Using Multibed Pressure Swing Adsorption Processes. *Industrial & Engineering Chemistry Research*, *47*, 3156–3169. URL: <http://dx.doi.org/10.1021/ie0712582>. doi:10.1021/ie0712582.
- Nikoli, D., Kikkinides, E. S., & Georgiadis, M. C. (2009). Optimization of Multibed Pressure Swing Adsorption Processes. *Industrial & Engineering Chemistry Research*, *48*, 5388–5398. URL: <http://dx.doi.org/10.1021/ie801357a>. doi:10.1021/ie801357a.
- Nilchan, S., & Pantelides, C. (1998). On the Optimisation of Periodic Adsorption Processes. *Adsorption*, *4*, 113–127.
- Park, J., Beum, H., Kim, J., & Cho, S. (2002). Numerical Analysis on the Power Consumption of the PSA Process for Recovering CO₂ from Flue Gas. *Ind. Eng. Chem. Res.*, *41*, 4122–4131.
- Rasmussen, C., & Williams, C. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

- Ribeiro, A. M., Grande, C. A., Lopes, F. V. S., Loureiro, J. M., & Rodrigues, A. E. (2008). A
615 parametric study of layered bed PSA for hydrogen purification. *Chemical Engineering Science*, *63*,
5258–5273. URL: <http://www.sciencedirect.com/science/article/pii/S0009250908003837>. doi:10.
1016/j.ces.2008.07.017.
- Ribeiro, A. M., Grande, C. A., Lopes, F. V. S., Loureiro, J. M., & Rodrigues, A. E. (2009). Four beds
620 pressure swing adsorption for hydrogen purification: Case of humid feed and activated carbon beds. *AIChE*
Journal, *55*, 2292–2302. URL: <http://onlinelibrary.wiley.com/doi/10.1002/aic.11886/abstract>.
doi:10.1002/aic.11886.
- Ruthven, D. (1984). *Principles of adsorption and adsorption processes*. John Wiley & Sons.
- Ruthven, D. M., Farooq, S., & Knaebel, K. S. (1994). *Pressure Swing Adsorption*. New York: Wiley.
- Sacks, J., Welch, W., Mitchell, T., & Wynn, H. (1989). Design and Analysis of Computer Experiments.
625 *Stat. Sci.*, *4*, 409–423.
- Schonlau, M. (1998). *Computer Experiments and Global Optimization*. Ph.D. thesis University of Waterloo.
- Skarstrom, C. (1966). Oxygen concentration process. US Patent 3,237,377.
- Sóbestor, A., Forrester, A., Toal, D., Tresidder, E., & Tucker, S. (2014). Engineering design applications of
surrogate-assisted optimization techniques. *Optim. Eng.*, *15*, 243–265.
- 630 Spiller, E., Bayarri, M., Berger, J., Calder, E., Patra, A., Pitman, E., & Wolpert, R. (2014). Automating
Emulator Construction for Geophysical Hazard Maps. *SIAM ASA J. Uncertain. Quantif.*, *2*, 126–152.
- Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.
- Žilinskas, A., Fraga, E. S., Beck, J., & Varoneckas, A. (2015). Visualization of multi-objective de-
cisions for the optimal design of a pressure swing adsorption system. *Chemometrics and Intelli-
635 gent Laboratory Systems*, *142*, 151–158. URL: [http://www.sciencedirect.com/science/article/pii/
S0169743915000040](http://www.sciencedirect.com/science/article/pii/S0169743915000040). doi:10.1016/j.chemolab.2015.01.002.
- Viana, F., Simpson, T., Balabanov, V., & Toropov, V. (2014). Metamodeling in Multidisciplinary Design
Optimization: How Far Have We Really Come? *AIAA Journal*, *52*, 670–690.
- Voutchkov, I., & Keane, A. (2010). Multi-Objective Optimization Using Surrogates. In *Computational*
640 *Intelligence in Optimization* (pp. 155–175). Springer.
- Wang, L., Liu, Z., Li, P., Yu, J., & Rodrigues, A. E. (2012). Experimental and modeling investigation
on post-combustion carbon dioxide capture using zeolite 13x-APG by hybrid VTSA process. *Chemical*
Engineering Journal, *197*, 151–161. URL: [http://www.sciencedirect.com/science/article/pii/
S138589471200589X](http://www.sciencedirect.com/science/article/pii/S138589471200589X). doi:10.1016/j.cej.2012.05.017.
- 645 Wang, L., Yang, Y., Shen, W., Kong, X., Li, P., Yu, J., & Rodrigues, A. E. (2013a). CO₂ Capture from
Flue Gas in an Existing Coal-Fired Power Plant by Two Successive Pilot-Scale VPSA Units. *Industrial &*
Engineering Chemistry Research, *52*, 7947–7955. URL: <http://dx.doi.org/10.1021/ie4009716>. doi:10.
1021/ie4009716.
- Wang, L., Yang, Y., Shen, W., Kong, X., Li, P., Yu, J., & Rodrigues, A. E. (2013b). Experimental evaluation
650 of adsorption technology for CO₂ capture from flue gas in an existing coal-fired power plant. *Chemical*
Engineering Science, *101*, 615–619. URL: [http://www.sciencedirect.com/science/article/pii/
S0009250913005216](http://www.sciencedirect.com/science/article/pii/S0009250913005216). doi:10.1016/j.ces.2013.07.028.
- Webley, P., & He, J. (2000). Fast solution-adaptive finite volume method for PSA/VSA cycle simulation; 1
single step simulation. *Comput. Chem. Eng.*, *23*, 1701–1712.