Surrogate based Optimisation for Design of Pressure Swing Adsorption Systems

Ian David Lockhart Bogle and Michael Fairweather (Editors), Proceedings of the 22nd European Symposium on Computer Aided Process Engineering, 17-20 June 2012, London.

Surrogate based Optimisation for Design of Pressure Swing Adsorption Systems

Joakim Beck,[a] Daniel Friedrich,[c] Stefano Brandani,[c] Serge Guillas,[b] Eric S Fraga,[a]

[a] *Centre for Process Systems Engineering, Department of Chemical Engineering, and*
[b] *Department of Statistical Science, UCL (University College London), Gower Street, London, WC1E 6BT, UK*
[c] *Institute for Materials and Processes, The University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh, EH9 3JL, UK*

Abstract

Pressure swing adsorption (PSA) is a cyclic adsorption process for gas separation and purification. PSA offers a broad range of design possibilities influencing the device behaviour. In the last decade much attention has been devoted towards simulation and optimisation of various PSA cycles. The PSA beds are modelled with hyperbolic/parabolic partial differential algebraic equations and the separation performance should be assessed at cyclic steady state (CSS). Detailed mathematical models together with the CSS constraint makes design difficult. We propose a surrogate based optimisation procedure based on kriging for the design of PSA systems. The numerical implementation is tested with a genetic algorithm, with a multi-start sequential quadratic programming method and with an efficient global optimisation algorithm. The case study is the design of a dual piston PSA system for the separation of a binary gas mixture of N_2 and CO_2.

Keywords: dual piston pressure swing adsorption, kriging, surrogate based optimisation

1. Introduction

Pressure swing adsorption (PSA) (Ruthven et al., 1993) is a cyclic adsorption process for gas separation and purification and is a cost efficient alternative to traditional absorption techniques. The main characteristic of a PSA process is to swing between adsorption and desorption by increasing and decreasing the system pressure respectively. For practical applications the PSA process is operated at cyclic steady state (CSS). The performance of PSA systems is a function of the number of adsorption beds, the bed dimensions, the layers, the cycle configuration and the operating conditions. To fully investigate PSA systems by experimental means is an intractable aim in practise; instead, computer simulation is used (Agarwal et al., 2009; Biegler et al., 2005). To address the computational challenges, we use Surrogate Based Optimisation (SBO) methods with kriging models, see e.g. (Jones, 2001). An SBO procedure, with a genetic algorithm (GA), a multi-start sequential quadratic programming method (SQP), and efficient global optimisation (EGO) (Jones et al., 1998), has been applied and analysed for the design of a dual piston PSA system.

2. Modelling and simulation of dual piston pressure swing adsorption

The dual piston PSA (DP-PSA) system is an extension of the conventional PSA process (Ruthven et al., 1993): the movement of the pistons generates fluid flow and pressure variations which induce the separation of the mixture. Operating the pistons out of phase introduces a cyclic variation in pressure and flow which generates adsorption and desorption process steps. The DP-PSA is capable of separating binary mixtures.

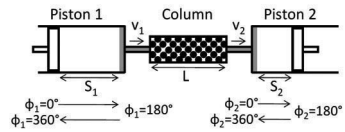


Figure 1: Schematic of the DP-PSA system.

This study considers a binary gas mixture of $80\%\,N_2$ and $20\%\,CO_2$ in a closed DP-PSA system with zeolite 13X packed column. Zeolite 13X has shown potential for the adsorption of CO_2 from flue gas and the adsorption isotherms are reported in the literature (Xiao et al., 2008). The separation performance, i.e. the ratio of the heavy component in piston 1 and 2, is analysed in this study. The DP-PSA model is from (Arvind et al., 2002). The model assumptions are: i) isothermal system; ii) ideal gas law applies; iii) no frictional pressure drop; iv) linear driving force mass transfer; v) Langmuir adsorption isotherm; vi) well-mixed piston chambers. The resulting mathematical model is a system of parabolic partial differential algebraic equations which is solved by the method of lines. The spatial coordinate along the length of the column is discretised with a flux-limiting finite volume scheme with the van Leer flux limiter (Laney, 1998). This discretisation scheme is conservative and guarantees the correct behaviour of the solution; the former is especially important for the simulation of closed systems, such as the DP-PSA, and the latter for the simulation of systems with sharp, moving fronts. The SUNDIALS solver suite (Hindmarsh et al., 2005) is used to solve the resulting system of differential algebraic equations in time. The system is simulated cycle by cycle until CSS is reached.

3. Surrogate based optimisation

The optimal design of a PSA system is represented by

$$\begin{array}{ll} \max\limits_{\mathbf{x}\in\Omega} & \mathrm{y}(\mathbf{q}(t,\mathbf{x})) \\ \text{subject to} & \mathbf{F}(\mathbf{q}(t,\mathbf{x}))=0, \\ & \mathbf{g}(\mathbf{q}(t,\mathbf{x}))\geq 0, \\ & \text{CSS conditions}, \end{array} \tag{1}$$

where $\mathbf{F}$ are model equations, $\mathbf{g}$ are design constraints, $\mathbf{q}$ is the state variable vector, $\mathbf{x}=(x_1,x_2,\ldots,x_d)$ is the design variable vector, $\Omega\subset\mathbb{R}^d$ is the design space, and $\mathrm{y}:\Omega\to\mathbb{R}$ is a scalar response which we wish to optimise. For simplicity, we refer to $\mathrm{y}(\mathbf{q}(t,\mathbf{x}))$ by $\mathrm{y}(\mathbf{x})$. When the objective function $y(\mathbf{x})$ is computationally expensive, surrogate models, which can be much faster to evaluate, could be introduced to replace $y(\mathbf{x})$, partially or in full, to guide the search more efficiently. Surrogate models (also known as response surfaces, metamodels, or emulators) are approximate functions of input-response mapping. A surrogate model $\hat{y}:\Omega\to\mathbb{R}$ is a predictor of response y fitted to a generated input-response dataset $(\mathbf{X},\mathbf{Y})$ of y on Ω, where $\mathbf{X}=(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m)$, $\mathbf{x}_j=(x_1,x_2,\ldots,x_d)\in\Omega\subset\mathbb{R}^d$, and $\mathbf{Y}=(\mathrm{y}_1,\mathrm{y}_2,\ldots,\mathrm{y}_m)$, $\mathrm{y}_j\in\mathbb{R}$. The cost to evaluate a surrogate model rapidly increases as the dimension of the input data, d, increases. Multivariate polynomial regression, artificial neural networks, support vector regression, radial basis function interpolation, and kriging are popular surrogate modelling techniques (Forrester and Keane, 2009). Recently a kriging based SBO method using DACE Matlab Kriging Toolbox (Lophaven et al., 2002) was applied for the design of a vaccum swing adsorption system (Faruque Hasan et al., 2011). We consider kriging: a response $\mathrm{y}(\mathbf{x})$ can be understood as a realisation of a stochastic process, $\mathrm{Y}(\mathbf{x})$, described by a linear spatial regression model

$$\mathrm{Y}(\mathbf{x})=\mu(\mathbf{x})+\varepsilon(\mathbf{x}),\quad \mathbb{E}[\varepsilon(\mathbf{x})]=0, \tag{2}$$

where $\mathbb{E}[\cdot]$ is the expected value operator, the stochastic model $\mathrm{Y}(\mathbf{x})$ consists of a regression model, $\mu(\mathbf{x})$, representing the drift or mean of $y(\mathbf{x})$, and a centred residual stochastic process, $\varepsilon(\mathbf{x})$, with covariance $\operatorname{cov}(\varepsilon(\mathbf{x}_i), \varepsilon(\mathbf{x}_j)) = \sigma^2 \operatorname{corr}(\varepsilon(\mathbf{x}_i), \varepsilon(\mathbf{x}_j))$, for $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$, where $\operatorname{corr}(\cdot,\cdot)$ is the spatial correlation function and σ^2 represents the global variation of y. The preferred choice of spatial correlation function (Sacks et al., 1989) is

$$\operatorname{corr}(\varepsilon(\mathbf{x}_i), \varepsilon(\mathbf{x}_j)) = \prod_{\ell=1}^{d} \exp\left(-\theta_j |x_{i,\ell} - x_{j,\ell}|^{p_\ell}\right), \tag{3}$$

for any $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$, with parameters $\theta_\ell \geq 0$ and $2 \geq p_\ell > 0$. The regression term $\mu(\mathbf{x})$ is a linear combination of a family of multivariate polynomials $(f_k(\mathbf{x}))_{k\in\mathbb{N}:k\leq n}$ for some $n \in \mathbb{N}_0$. Let us introduce the following notation in order to formulate the kriging predictor. Let $\Sigma \in \mathbb{R}^{m\times m}$ denote the covariance matrix, where the elements (i,j) are given by $\operatorname{cov}(\varepsilon(\mathbf{x}_i), \varepsilon(\mathbf{x}_j))$ for $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$; $\mathbf{c}(\mathbf{x}) = (\operatorname{cov}(\mathbf{x}, \mathbf{x}_i))_{\mathbf{x}_i\in\mathbf{X}} \in \mathbb{R}^m$; $\mathbf{f}(\mathbf{x}) = (f_k(\mathbf{x}))_{k\in\mathbb{N}:k\leq n} \in \mathbb{R}^n$ for $n \in \mathbb{N}_0$, and $\mathbf{F} = (f_k(\mathbf{x}_i))_{k\in\mathbb{N},\mathbf{x}_i\in\mathbf{X}} \in \mathbb{R}^{n\times m}$. Then, $\mu(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T\delta$ for mean parameter δ. Here $\mathbf{A}^T$ denotes the transpose of a matrix $\mathbf{A}$.
By assuming $(\mathbf{X}, \mathbf{Y})$ has been generated from a stochastic model, eq. (2), with correlation between the residuals, eq. (3), the *kriging predictor* $\hat{y}$ is the best unbiased linear predictor of y (Gaetan and Guyon, 2010) and can be written as

$$\hat{y}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T\hat{\delta} + \mathbf{c}(\mathbf{x})^T\Sigma^{-1}(\mathbf{Y} - \mathbf{F}\hat{\delta}), \tag{4}$$

where $\hat{\delta} = (\mathbf{F}^T\Sigma^{-1}\mathbf{F})^{-1}\mathbf{F}^T\Sigma^{-1}\mathbf{Y}$.
The values of parameters θ_ℓ and p_ℓ in eq. (3) are often unknown but can be estimated to maximise the likelihood function of $(\mathbf{X}, \mathbf{Y})$. $\mu(\mathbf{x}) = \alpha$, where α is a constant, and $p_\ell = 2$, are common assumptions. When $p_\ell = 2$, eq. (3) is the Gaussian kernel. Let $\theta_\ell \in [0.001, 10]$. An attractive feature with kriging is that the sample variance of the prediction error (MSE), $\mathrm{s}^2(\mathbf{x})$, can be easily computed, see (Gaetan and Guyon, 2010). A consequence of $\varepsilon(\mathbf{x})$ being a Gaussian process, the distribution of the kriging predictor at an unobserved point $\mathbf{x}$ is normally distributed with mean $\hat{\mathrm{y}}(\mathbf{x})$ and variance $\mathrm{s}^2(\mathbf{x})$. We propose a SBO based on kriging, see Algorithm 1 where $\|\cdot\|$ is the Euclidean norm and $\mathbb{E}_{\mu,\sigma}[\mathrm{f}(\mathbf{Y})]$ is the expected value operator of a function $\mathrm{f}(\mathbf{Y})$ with $\mathbf{Y}$ being a Gaussian process with mean μ and variance σ^2.

Algorithm 1 Surrogate based optimisation procedure based on kriging.

1: Generate dataset $(\mathbf{X}, \mathbf{Y})$
2: Let $i = 1$ and $i_{\max} \in \mathbb{N}$
3: **while** $i \leq i_{\max}$ **do**
4: Fit kriging model $\hat{\mathrm{y}}(\mathbf{x})$, $\mathbf{x} \in \Omega$, to $(\mathbf{X}, \mathbf{Y})$
5: **solve** $\mathbf{x}^* = \underset{\mathbf{x}\in\Omega:\, \frac{\|\mathbf{x}-\bar{\mathbf{x}}\|}{\|\bar{\mathbf{x}}\|}\geq\epsilon\ \forall\bar{\mathbf{x}}\in\mathbf{X}}{\arg\max}\ \hat{\mathrm{y}}(\mathbf{x})$
6: **or** $\mathbf{x}^* = \underset{\mathbf{x}\in\Omega:\, \frac{\|\mathbf{x}-\bar{\mathbf{x}}\|}{\|\bar{\mathbf{x}}\|}\geq\epsilon\ \forall\bar{\mathbf{x}}\in\mathbf{X}}{\arg\max}\ \mathbb{E}_{\hat{\mathrm{y}}(\mathbf{x}),s(\mathbf{x})}\left[\max\{\hat{\mathrm{y}}(\mathbf{x}) - \max(\mathbf{Y}), 0\}\right]$ (EGO (Jones et al., 1998))
7: $\mathbf{X} \leftarrow \mathbf{x}^*$
8: $\mathbf{Y} \leftarrow \mathrm{y}(\mathbf{x}^*)$
9: $i \leftarrow i + 1$
10: **end while**

We consider a GA and a multi-start SQP for the inner optimisation at line 5 in Algorithm 1, henceforth referred to as SbGA and SbSQP ("Sb" means "Surrogate based"). The optimisation problem at line 6 is known as *Efficient Global Optimization* (EGO) (Jones et al., 1998), where the expected improvement rather than the kriging predictor $\hat{y}$ is maximised, here through the use of a GA. GA is also used at line 4 for the maximum likelihood estimation when fitting the kriging model. The numerical implementation of kriging

closely follows the DACE Matlab Kriging Toolbox (Lophaven et al., 2002). The GA is an in-house implementation and the configurations have been chosen carefully based on experience (Fiandaca et al., 2009). The SQP is fmincon provided in MATLAB 2010b.

4. Numerical results

For the optimisation problem, eq. 1, with the DP-PSA model described, the response of interest, y, is the purity of CO_2 in piston chamber 1 at CSS. The design space, Ω, is a six-dimensional hyperrectangle domain defined by the decision variables: cycle time $\mathrm{t_c}$ $\in$ [1, 20] (s), offset angles of piston 1 and 2, ϕ_1 and $\phi_2 \in [0, 2\pi]$, volumes of piston chamber 1 and 2, V_1 and $\mathrm{V}_2 \in$ [0.5, 15.0] (m^3), and temperature $\mathrm{T} \in$ [15, 70] ($^\circ C$). We have restricted the computation to 50 full evaluations of y, of which 20 are initially generated with latin hypercube sampling and used as the initial dataset $(\mathbf{X}, \mathbf{Y})$ for the SBO methods (SbGA, SbSQP, and EGO) and as the initial population for the GA. For the standalone GA, we allowed 150 evaluations. The SBO methods were configured to make similar numbers of calls to the kriging predictor. Design point diversity is enforced by $\epsilon = 10^{-4}$. Here follows a comparison between GA and the SBO methods followed by a visualisation of the best DP-PSA designs found.

	Full evaluations	Elapsed time (s)
GA	150	12400
SbGA	50	5900
SbSQP	50	6000
EGO	50	6300

The SBO methods all outperform GA in elapsed time to reach same quality of solution, see table above and Figure 2. The performance is similar between the different SBO methods in terms of quality of solution, although SbSQP has marginally better performance as the number of full evaluations increases. The kriging predictor has smooth and infinitely differentiable responses, which suits SQP as it is a local optimiser guided by gradient and hessian information. Even if the underlying objective is multi-modal, with good initial guesses, which could be obtained with the multi-start algorithm, the SbSQP exhibits a rapid local convergence behaviour and ensures global exploration of the kriging response surface.

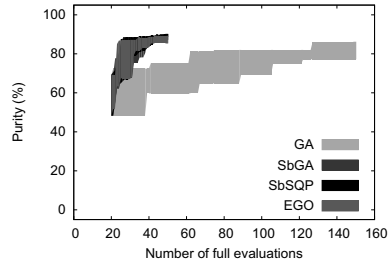


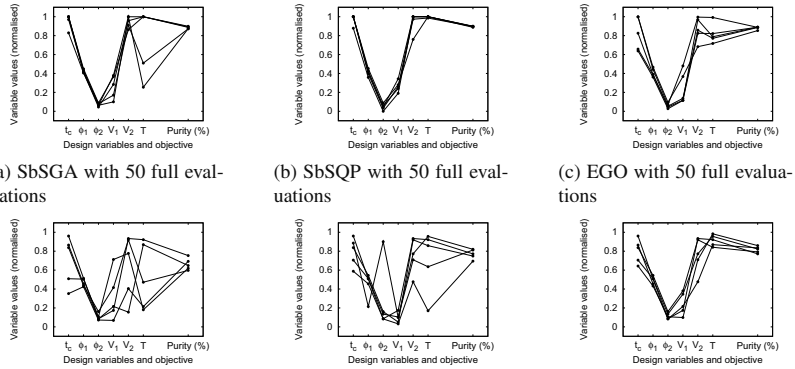
Figure 2: CO_2 purity obtained using 5 different initial datasets for each method, where the range between the best and worse performance curves is filled.

The SBO methods show qualitative agreement in optimum when reached 50 full evaluations, whereas the GA, for some initial populations, does not, see Figures 3 (a-d). See the evolution of the optimum found with GA in Figures 3 (d-f).

5. Conclusions

Conventional PSA cycles are often more complex and demand longer cycle times than the single column DP-PSA system and for this reason also more computationally demanding. When designing these complex systems we are usually limited to a moderate number of full evaluations, and the numerical results therefore motivate the use of SBO to achieve a faster convergence towards regions of quality designs. The proposed SBO procedure with kriging (see Algorithm 1) has here been shown to suffice with GA and multi-start

SQP, while also allowing the use of other optimisation methods. The numerical results presented in Figures 3 (a-c) indicate that a DP-PSA system should be designed as follows: use longer cycle times to operate close to equilibrium, intermediate offset between piston 1 and 2, the piston chamber volume should be small and large for piston 1 and 2, respectively, and the column should be at a high temperature. These design guidelines are reached by optimising the purity alone. The influence of further objectives such as recovery and energy requirements will be investigated in a future publication.



(a) SbSGA with 50 full evaluations

(b) SbSQP with 50 full evaluations

(c) EGO with 50 full evaluations

(d) GA with 50 full evaluations

(e) GA with 100 full evaluations

(f) GA with 150 full evaluations

Figure 3: High dimensional visualisation of the collection of best design point found with SbGA, SbSQP, EGO and GA, respectively, using 5 different initial datasets.

References

Agarwal, A., Biegler, L. T., and Zitney, S. E. (2009). Simulation and optimization of pressure swing adsorption systems using reduced-order modeling. *Industrial & Engineering Chemistry Research*, 48(5):2327–2343.

Arvind, R., Farooq, S., and Ruthven, D. M. (2002). Analysis of a piston PSA process for air separation. *Chemical Engineering Science*, 57(3):419 – 433.

Biegler, L. T., Jiang, L., and Fox, V. G. (2005). Recent advances in simulation and optimal design of pressure swing adsorption systems. *Separation & Purification Reviews*, 33:1–39.

Faruque Hasan, M., Karimi, I., Farooq, S., Rajendran, A., and Amanullah, M. (2011). Surrogate-based VSA process optimization for post-combustion CO_2 capture. *21st European Symposium On Computer Aided Process Engineering*, 29:402–406.

Fiandaca, G., Fraga, E. S., and Brandani, S. (2009). A multi-objective genetic algorithm for the design of pressure swing adsorption. *Engineering Optimization*, 41(9):833–854.

Forrester, A. and Keane, A. (2009). Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3):50–79.

Gaetan, C. and Guyon, X. (2010). *Spatial Statistics and Modeling (Springer Series in Statistics)*. Springer.

Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E., and Woodward, C. S. (2005). Sundials: Suite of Nonlinear and Differential/Algebraic Equation Solvers. *ACM Transactions on Mathematical Software*, 31(3):363–396.

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

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

Laney, C. B. (1998). *Computational Gasdynamics*. Cambridge University Press, New York.

Lophaven, S., Nielsen, H., and Søondergaard, J. (2002). DACE: A MATLAB Kriging toolbox version 2.0. Technical Report IMM-REP-2002-12, Informatics and Mathematical Modelling, Technical University of Denmark.

Ruthven, D. M., Faroog, S., and Knaebel, K. S. (1993). *Pressure Swing Adsorption*. Wiley-VCH.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and Analysis of Computer Experiments. *Statistical Science*, 4(4):409–423.

Xiao, P., Zhang, J., Webley, P., Li, G., Singh, R., and Todd, R. (2008). Capture of CO_2 from flue gas streams with zeolite 13X by vacuum pressure swing adsorption. *Adsorption*, 14:575–582.