



MAX-PLANCK-GESELLSCHAFT

Max Planck Institute Magdeburg Preprints

Tobias Keßler Christian Kunde Nick Mertens
Dennis Michaels Achim Kienle

Global Optimization of Distillation Columns using Surrogate Models



Abstract

Surrogate-based optimization of distillation columns using an iterative Kriging approach is investigated. Focus is on deterministic global optimization to avoid suboptimal local minima. The determination of optimal setups and operating conditions for ideal and non-ideal distillation columns, leading to mixed-integer nonlinear programming (MINLP) problems, serve as case studies. It is found that the optimization using the adapted Kriging approach yields similar results compared to the direct global optimization of the original problem in the ideal case, while it leads to a huge improvement compared to a multistart local optimization approach in the non-ideal case.

Contents

1 Introduction	1
2 Kriging models	1
3 Ideal distillation	4
4 Non-ideal distillation	8
5 Conclusion	10

Imprint:

Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg

Publisher:

Max Planck Institute for
Dynamics of Complex Technical Systems

Address:

Max Planck Institute for
Dynamics of Complex Technical Systems
Sandtorstr. 1
39106 Magdeburg

<http://www.mpi-magdeburg.mpg.de/preprints/>

1 Introduction

Rigorous optimization of distillation columns is of major interest in the chemical process industry due to its high economical impact. Due to the presence of discrete and continuous decision variables this leads to mixed-integer nonlinear programs (MINLP). Standard local optimization or stochastic optimization approaches can not guarantee that the optimum found by the optimizer is the global one. Alternatively, deterministic global optimization based on convex relaxations within a branch and bound framework has become an interesting approach for solving such problems, e.g., see the recent textbook by Locatelli and Schoen[1] for an introduction. However, computation times for distillation columns using standard model formulations from first principles are often extremely large. [2, 3, 4]

To overcome this problem, different solution approaches were recently proposed. Quirante et al.[5] suggested to use surrogate models based on Kriging interpolation for optimization of distillation columns to reduce computational complexity. Main emphasis was on local optimization but it was also suggested to use Kriging models to reduce computational complexity in deterministic global optimization. Nallasivam et al.[4] presented an algorithm for calculating minimum energy requirements for thermally coupled distillation column configurations. The algorithm is based on a shortcut model which is only valid for ideal mixtures under minimum reflux conditions. An alternative approach for any reflux based on rigorous tray to tray models was proposed by Ballerstein et al.[2]. It applies to binary ideal mixtures. Illustration was demonstrated for a hybrid distillation crystallization process for isomer separation. The approach is based on monotonicity of the concentration variables in a binary distillation, which can be used to systematically reduce the search space. More recently this strategy could be extended in Mertens et al.[3] to ideal multicomponent distillation processes using a model reformulation strategy, which results in monotonicity of some aggregated concentration variables. However, an extension to non-ideal mixtures is in general not possible as will be argued in the present paper. Therefore, global optimization using Kriging models as proposed by Quirante et al.[5] is further investigated in some detail in this paper and compared to the previous approaches by Mertens et al.[3].

The outline of the paper is as follows: The general concept of Kriging interpolation is briefly explained in Section 2. Section 3 deals with ideal multicomponent distillation, which admits a rigorous global optimization using the reformulation by Mertens et al.[3]. It is shown that very similar results could be obtained with the Kriging approach using iterative refinement. Afterwards a highly non-ideal azeotropic mixture is considered in Section 4. Since rigorous global optimization is currently not possible with standard global optimization software within reasonable time, global optimization with the Kriging approach is compared with local optimization and thereby demonstrating the power of the global Kriging approach for highly non-ideal mixtures.

2 Kriging models

Kriging models can be used to approximate complex mathematical models of real world processes. During the last years they gained increasingly more interest from engineers from different fields, such as chemical engineering, e.g. see Caballero and Grossmann[6] and Quirante

et al.[5]. The accuracy of Kriging models, as well as their complexity, depends on the number of reference points used for their generation. In this section the basic idea of an ordinary Kriging model is briefly sketched. The presentation mainly follows Caballero and Grossmann[6].

Given a vector-valued function $\mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^d$ and a finite number of *reference points* $\bar{\mathbf{x}}^k \in \mathbb{R}^m$, $k = 1, \dots, N$, a Kriging interpolation is a vector-valued function $\hat{\mathbf{f}} : \mathbb{R}^m \rightarrow \mathbb{R}^d$ with $\hat{\mathbf{f}}(\mathbf{x}) := \mathbf{y}(\mathbf{x}) + \mathbf{Z}(\mathbf{x})$. Here, $\mathbf{y} : \mathbb{R}^m \rightarrow \mathbb{R}^d$ is a vector-valued function consisting of polynomials, and $\mathbf{Z} : \mathbb{R}^m \rightarrow \mathbb{R}^d$ is a vector-valued function used to describe the deviation of $\hat{\mathbf{f}}(\mathbf{x})$ from $\mathbf{y}(\mathbf{x})$. In ordinary Kriging models, as considered in this work, function $\mathbf{y}(\mathbf{x})$ is chosen to be a vector $\boldsymbol{\xi} \in \mathbb{R}^d$ of suitable constants. Although this restriction seems to be rather strong, it does not affect the accuracy of the resulting surrogate model significantly for smooth functions, because most of the information is contained in $\mathbf{Z}(\mathbf{x})$ as noted by Papalambros and Wilde[7]. Function $\mathbf{Z}(\mathbf{x})$ is assumed to be a weighted sum of the deviations at all reference points with certain weights depending on \mathbf{x} that are defined by a weight function $\mathbf{w} : \mathbb{R}^m \rightarrow \mathbb{R}^N$, i.e.

$$\mathbf{Z}(\mathbf{x}) = \sum_{k=1}^N w_k(\mathbf{x}) \cdot (\mathbf{f}(\bar{\mathbf{x}}^k) - \boldsymbol{\xi}).$$

This leads to the following representation for $\hat{\mathbf{f}}$.

$$\hat{\mathbf{f}}^\top(\mathbf{x}) = \boldsymbol{\xi}^\top + \mathbf{w}^\top(\mathbf{x}) \begin{pmatrix} \mathbf{f}^\top(\bar{\mathbf{x}}^1) - \boldsymbol{\xi}^\top \\ \mathbf{f}^\top(\bar{\mathbf{x}}^2) - \boldsymbol{\xi}^\top \\ \vdots \\ \mathbf{f}^\top(\bar{\mathbf{x}}^N) - \boldsymbol{\xi}^\top \end{pmatrix}.$$

It is further postulated that vector $\hat{\mathbf{f}}(\mathbf{x})$ is identical to vector $\mathbf{f}(\mathbf{x})$ for each reference point, i.e.

$$\hat{\mathbf{f}}(\bar{\mathbf{x}}^k) = \mathbf{f}(\bar{\mathbf{x}}^k), \quad \text{for all } k \in \{1, \dots, N\}. \quad (1)$$

In order to ensure the conditions in Equation (1), it is requested that the weight function evaluated at the k th reference point coincides with the k th unit vector $\mathbf{e}_k \in \mathbb{R}^N$, i.e.

$$\mathbf{w}(\bar{\mathbf{x}}^k) = \mathbf{e}_k, \quad \text{for all } k \in \{1, \dots, N\}. \quad (2)$$

A parametrized function $c : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ and a matrix \mathbf{R} are further used to calculate the weights $\mathbf{w}(\mathbf{x})$ as follows.

$$\mathbf{w}^\top(\mathbf{x}) = (c(\mathbf{x}, \bar{\mathbf{x}}^1), \dots, c(\mathbf{x}, \bar{\mathbf{x}}^N)) \mathbf{R}$$

Due to (2), matrix \mathbf{R} needs to be

$$\mathbf{R} = \begin{pmatrix} c(\bar{\mathbf{x}}^1, \bar{\mathbf{x}}^1) & \dots & c(\bar{\mathbf{x}}^1, \bar{\mathbf{x}}^N) \\ \vdots & \ddots & \vdots \\ c(\bar{\mathbf{x}}^N, \bar{\mathbf{x}}^1) & \dots & c(\bar{\mathbf{x}}^N, \bar{\mathbf{x}}^N) \end{pmatrix}^{-1}. \quad (3)$$

Note that function c and the reference points must be chosen such that matrix \mathbf{R} is invertible. In the literature, various strategies for finding a suitable function c and appropriate reference points are available, e.g. see Koehler and Owen[8].

In case of \mathbf{f} being a Gaussian process, minimizing the variance of the estimation error $\hat{\mathbf{f}}(\mathbf{x}) - \mathbf{f}(\mathbf{x})$ leads to function values $c(\mathbf{x}^1, \mathbf{x}^2)$ being the covariance between the points \mathbf{x}^1 and \mathbf{x}^2 . When interpolating a deterministic function (as considered here), any function $c(\mathbf{x}^1, \mathbf{x}^2)$ may be feasible and the best choice is non-trivial. The choice of the function determines how the model fits the data. Therefore, it usually contains parameters that have to be optimized in order to generate a surrogate model with a good fit. The most commonly used function is an exponential function of the following form[6]

$$c(\mathbf{x}^1, \mathbf{x}^2) = \exp\left(-\sum_{i=1}^m \theta_i \|\mathbf{x}_i^1 - \mathbf{x}_i^2\|^{p_i}\right).$$

For $\mathbf{x}^1 = \mathbf{x}^2$, value $c(\mathbf{x}^1, \mathbf{x}^2)$ equals one, while it tends to zero when the distance between the points \mathbf{x}^1 and \mathbf{x}^2 increases. For every component i , parameter θ_i defines the speed of this tendency and parameter p_i denotes the smoothness of c . In summary, $\mathbf{Z}(\mathbf{x})$ is calculated as

$$\mathbf{Z}^\top(\mathbf{x}) = (c(\mathbf{x}, \bar{\mathbf{x}}^1), \dots, c(\mathbf{x}, \bar{\mathbf{x}}^N)) \cdot \mathbf{F},$$

with constant matrix \mathbf{F} given by

$$\mathbf{F} = \mathbf{R} \cdot \begin{pmatrix} \mathbf{f}^\top(\bar{\mathbf{x}}^1) - \boldsymbol{\xi}^\top \\ \mathbf{f}^\top(\bar{\mathbf{x}}^2) - \boldsymbol{\xi}^\top \\ \vdots \\ \mathbf{f}^\top(\bar{\mathbf{x}}^N) - \boldsymbol{\xi}^\top \end{pmatrix}.$$

Thus, there are some unknown parameters that have to be calculated, namely $\boldsymbol{\xi}$, θ_i and p_i . Following Quirante et al.[5], values of the parameters are obtained by maximizing the logarithmic likelihood function

$$\begin{aligned} \log(L) = & -N/2 (\log(\sigma^2) + \log(2\pi)) \\ & -1/2 \log(\|\mathbf{R}\|) \\ & -1/(2\sigma^2) (\mathbf{Y}(\bar{\mathbf{x}}) - \mathbf{1}\boldsymbol{\xi})^\top \mathbf{R}^{-1} (\mathbf{Y}(\bar{\mathbf{x}}) - \mathbf{1}\boldsymbol{\xi}), \end{aligned}$$

with

$$\begin{aligned} \boldsymbol{\xi} &= (\mathbf{1}^\top \mathbf{R}^{-1} \mathbf{1})^{-1} (\mathbf{1}^\top \mathbf{R}^{-1} \mathbf{Y}(\bar{\mathbf{x}})), \\ \sigma^2 &= 1/N (\mathbf{Y}(\bar{\mathbf{x}}) - \mathbf{1}\boldsymbol{\xi})^\top \mathbf{R}^{-1} (\mathbf{Y}(\bar{\mathbf{x}}) - \mathbf{1}\boldsymbol{\xi}), \\ \mathbf{Y}(\bar{\mathbf{x}}) &= \begin{pmatrix} \mathbf{f}_1(\bar{\mathbf{x}}^1) & \dots & \mathbf{f}_d(\bar{\mathbf{x}}^1) \\ \vdots & \ddots & \vdots \\ \mathbf{f}_1(\bar{\mathbf{x}}^N) & \dots & \mathbf{f}_d(\bar{\mathbf{x}}^N) \end{pmatrix}. \end{aligned}$$

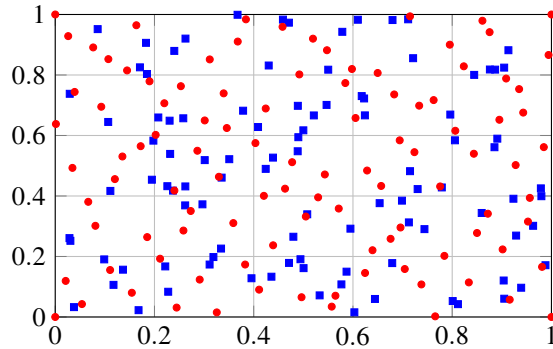


Figure 1: Difference between Halton sequence (red dots) and random number sequence (blue squares) with 100 samples each

In this work the MATLAB function “fmincon” is used for the identification of the parameters, i.e. for the fitting of the Kriging models.

As the choice of the reference points may greatly influence the accuracy, it is important to use a space filling approach instead of using randomly generated data points. Otherwise clustering in unimportant regions of the model may occur and results in a surrogate model of poor quality. In this work, a *Halton sequence*[9] as shown in Figure 1 is used to cover the space evenly. The integer variables are enumerated through the sampling procedure.

After a first Kriging model is generated, it is optimized using the deterministic global optimization software BARON[10]. Following Caballero and Grossmann[6], a second, refined Kriging model is constructed by restricting the sampling region to a certain neighborhood around the found solution of the first Kriging model. The second Kriging model is likely to be more precise in the region of interest. It may happen that one of the variables in the solution of the second Kriging model attains its value at the boundary of its respective domain. In this case, a further Kriging model is generated using a neighborhood around the optimal solution of the second Kriging model as sampling region. Note that the latter neighborhood also covers regions that are not contained in the neighborhood around the optimal solution of the first Kriging model.

All MINLPs are solved using the GAMS 24.6.1 framework with the deterministic global optimization software BARON 15.9.22., Cplex 12.6.3 is used as LP/MIP subsolver and CONOPT 3.17A is utilized as NLP subsolver. The calculations are carried out on a Linux PC with 3.40 GHz Intel Core i7-6700 CPU and 16 GB memory. Master’s at Various estimation $c(\mathbf{x}_1, \mathbf{x}_2)$ a $\mathbf{y}(\mathbf{x})$

3 Ideal distillation

As a first case study an ideal four-component distillation column with variable number of stages and a variable feed stage is chosen. A sketch of such a distillation column is depicted in Figure 2. The distillation process aims to separate the two more volatile components

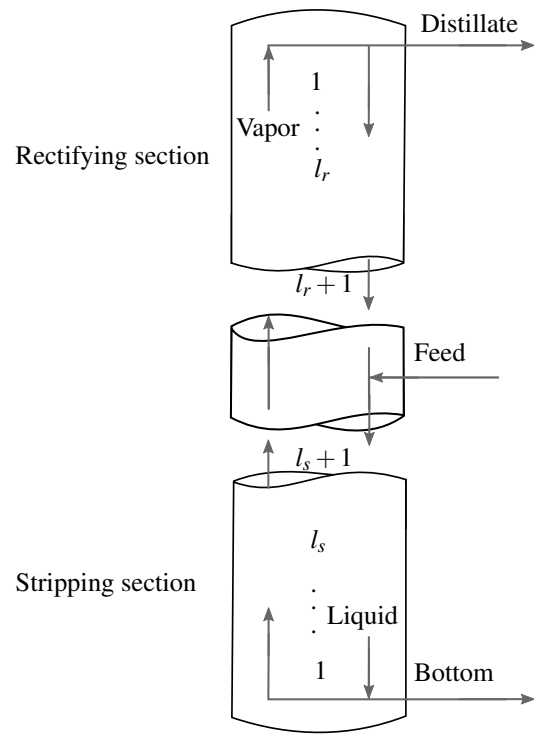


Figure 2: Distillation column scheme with variable number of stages in the upper and lower part of the column and variable feed location. The number of stages is $l_r + l_s + 1$, the feed location is at stage $l_s + 1$

Table 1: Computational results from reference calculations[3] with a feed flow of 1.8 mol/s

Case	TAC	D	V	l_s	l_r
1a	23566	0.9	0.9408	3	1
1b	33597	0.9	1.2796	4	3
1c	41177	0.9	1.4034	6	6
2a	25419	0.9	0.99	4	1
2b	37109	0.9	1.35198	5	4
2c	46727	0.9	1.5416	7	7
3a	27993	0.9	1.0719	4	2
3b	42633	0.9	1.499	6	5
3c	55629	0.9	1.7518	9	8

from the two less volatile components under given purity requirements, while minimizing the total annualized cost (TAC) of the process. The mathematical model description as well as the numerical test instances are taken from Mertens et al. [3]. Despite the fact that the vapor-liquid-equilibrium (VLE) of the column is modeled as ideal, i.e. with constant relative volatilities, finding a global optimum of the corresponding MINLPs is rather challenging using standard model formulations. It was however demonstrated in Mertens et al.[3] that a model reformulation in combination with tailor-made optimization techniques can reduce the computational effort considerably. Thus, globally optimal solutions for different test instances are available and can be used as references to investigate the accuracy of the surrogate based approach presented in this work.

The case study deals with three different mixtures that are separated with three different product specifications each. This leads to a total number of nine test instances. The mixtures are labeled by numbers 1, 2 and 3 where the difficulty of the separation task increases from mixture 1 to 3. The product specifications are given by the purity requirements on the distillate, i.e. on the mixture consisting of the two more volatile components, and on the bottom product, i.e. the mixture consisting of the two less volatile components. Different product specifications are labeled by letters a , b and c , and become more restrictive for the separation task from a to c . The concrete parameter setting defining each test instance is provided in Mertens et al.[3]

Note that feasible solutions to the considered MINLPs represent feasible distillation column designs of the corresponding separation tasks. A characterization of each such designs is given by the length l_r of the rectifying section, i.e. the part above the feed stage (see Figure 2), by the length l_s of the stripping section, i.e. the part below the feed stage, by the distillate flow rate D in mol/s and by the vapor flow rate V in mol/s. For the known optimal solutions to our test instances, these characteristic properties are summarized in Table 1.

the flow of

The approaches described in Section 2 are applied to a model of the distillation column implemented in MATLAB. The resulting Kriging models are implemented and optimized in GAMS/BARON[10]. The results obtained for all test cases can be found in Table 2. The results for the cases labeled with subscript 1 refer to the results obtained with the first Kriging model, while the results for the cases labeled with subsequent subscripts are obtained by

Table 2: Computational results using surrogate model. Values with a subscript higher than 1 are calculated using an adaptive sampling technique

Case	TAC	D	V	l_s	l_r
1a ₁ ^{krig}	22197	1	0.85	3	2
1a ₂ ^{krig}	24318	0.9	0.9424	3	2
1b ₁ ^{krig}	37002	0.9	1.3104	6	4
1b ₂ ^{krig}	34563	0.9	1.2831	4	4
1c ₁ ^{krig}	59179	0.9	2.0610	5	8
1c ₂ ^{krig}	44362	0.9	1.5281	5	7
2a ₁ ^{krig}	24184	1	0.85	5	3
2a ₂ ^{krig}	26216	0.9	0.9942	4	2
2b ₁ ^{krig}	38383	0.9	1.3296	6	5
2b ₂ ^{krig}	37992	0.9	1.3505	5	5
2c ₁ ^{krig}	58178	0.9	2.022	7	6
2c ₂ ^{krig}	50639	0.9	1.731	5	8
3a ₁ ^{krig}	30142	0.9	1.0642	5	4
3a ₂ ^{krig}	28785	0.9	1.1067	4	2
3b ₁ ^{krig}	44871	0.9	1.5884	7	4
3b ₂ ^{krig}	43703	0.9	1.5418	6	5
3c ₁ ^{krig}	59683	0.9	1.9854	8	7
3c ₂ ^{krig}	56986	0.9	1.722	10	9

Kriging models generated with the iterative sampling approach presented in Section 2. The new sampling region for the degrees of freedom lies within a 25% range of the found optimum for the continuous variables, V and D , and ± 2 around the optimum of the integer variables, l_s and l_r . It can be seen that the iterative Kriging optimization approach is able to calculate a solution lying in the neighborhood of an actual global optimum. However, it is important to note that the iterative Kriging approach may lead to solutions that are not close to a global optimal one. This may be due to a possibly high inaccuracy of the first Kriging model that may yield initial solutions that are already far away from a global optimum. If the first Kriging model is then refined around a solution of poor quality, it may not be possible to arrive at an actual global optimum as it is no longer captured by the refined sampling region of the subsequent Kriging model.

Note that in most cases the first solutions obtained do not meet the constraints due to the inaccuracy of the first Kriging model. However, after the refinement of the sampling region the constraints are met by the found optimal solutions.

Strict comparison of computation times of the different approaches is not possible due to different hardware and optimization software configurations. However, it was observed that the reference model typically could not be solved within 10 h, whereas the reformulated model with some tailor made bound tightening strategies was most of the time solved in less than an

hour. For the detailed statistics we refer to Mertens et al.[3] In each iteration, optimization times of the Kriging models are in the same order of magnitude using standard software (BARON). About the same time is required for reference simulations and fitting of the Kriging models. done results

4 Non-ideal distillation

In the previous section focus was on separation of ideal mixtures with constant relative volatilities. Next, a highly non-ideal azeotropic mixture with variable volatilities is considered.

From Section 3 it is clear that the global optimization of the reformulated model is preferable to the optimization of the Kriging model, because the global optimality of the result obtained with the Kriging approach can not be guaranteed. However, global optimization of the rigorous model is computationally even much more expensive compared to the previous example. Further, a monotonic reformulation $\hat{x}_i(z)$ of the molar fractions $x_i(z)$ like in the previous ideal case is also not possible anymore, as will be shown in the following.

Take the transformed states \hat{x}_i as functions of the molar fractions x_i

$$\hat{x}_1 = x_1, \quad (4)$$

$$\hat{x}_2 = x_1 + x_2, \quad (5)$$

$$\hat{x}_3 = 1, \quad (6)$$

and their derivatives with respect to the spatial coordinate z

$$\frac{\partial \hat{x}_1}{\partial z} = \frac{\partial x_1}{\partial z} \geq 0, \quad (7)$$

$$\frac{\partial \hat{x}_2}{\partial z} = \frac{\partial (x_1 + x_2)}{\partial z} = \frac{\partial (x_1 + 1 - x_3 - x_1)}{\partial z} \quad (8)$$

$$= -\frac{\partial x_3}{\partial z} \geq 0, \quad (9)$$

$$\frac{\partial \hat{x}_3}{\partial z} = 0, \quad (10)$$

for the proof.

The assumption of positive derivatives in (7) and (9) always holds for the distillation of ideal mixtures, but it cannot be guaranteed for non-ideal mixtures because of the occurring azeotropes.

One example for such a mixture is the distillation of Toluene, Methanol and Methylbutyrate[11], which is investigated here. It is easy to see from the residue curve map, representing the column profiles at total reboil and reflux, depicted in Figure 3, that the above condition that the column profiles have to be monotonic for at least two of the components is not met.

To model non-ideal phase behavior, several activity coefficient models, such as UNIQUAC and Wilson[12], may be used. In the present case study a Wilson approach is chosen. The column is assumed to be operated at a constant pressure of 1.013 hPa, the feed rate f is set to 1.8 mol/s and the feed composition mole fraction x_{Feed} is [0.2806, 0.6566, 0.0628], where the

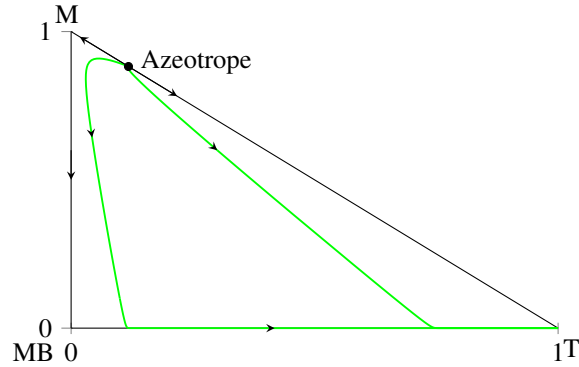


Figure 3: Residue curve map of the mixture Methanol(M)-Methylbutyrate(MB)-Toluene(T). Arrows indicate the phase behavior for increasing temperature, i.e. from the top to the bottom of the column.

first entry refers to Toluene, the second entry to Methanol and the third entry to Methylbutyrate. The aim of the optimization is to find optimal operating conditions such that at least 80 % of component Toluene contained in the initial feed flow rate is gained at the bottom of the column with a purity of at least 90 %. This gives rise to the two constraints

$$x_{T,Bott} \geq 0.95, \quad (11a)$$

$$x_{T,Bott} \cdot B \geq 0.8 \cdot f \cdot x_{Toluene,Feed} = 0.4041, \quad (11b)$$

where B is the bottom flow rate in mol/s. The same objective function as in the previous example is used. The degrees of freedom are the distillate flow rate $D \in [1.1, 1.7]$ mol/s, the vapor flow rate $V \in [2, 13]$ mol/s, as well as the number of stages $l_r + l_s + 1$ and the feed location $l_s + 1$ with $l_r, l_s \in [1, 24]$.

Since global optimization with the rigorous model or its reformulation like in the ideal case is not possible within reasonable time anymore, local optimization is applied as a reference. For this, the local multi-start heuristic provided by BARON is applied with 1000 starting points. The results are displayed in the second row of Table 3 labeled with “Local”.

The Kriging models have been fitted using 1190 sampling points. A cross validation of the first Kriging model (case “Surrogate₁” in Table 3) was done with 100 points and is depicted in the upper part of Figure 4. The first Kriging model is rather inaccurate, especially in the occurring discontinuity of $x_{T,Bott}$, that is roughly contained in the interval (0.72, 0.85). This is due to the large sampling region and due to the complexity of the problem. As a result, the operating conditions obtained in the first optimization are violating the constraints with $x_{T,Bott} = 0.9034$ and $x_{T,Bott} \cdot B = 0.3820$ mol/s.

Since the operating conditions found during the reference optimization “Local” are far away from the optimal operating conditions for “Surrogate₁”, the sampling region for the adaptive Kriging approach is chosen to be larger than in the ideal case, with ± 3 for the integer variables, $\pm 25\%$ for D and $+3$ for V . After sampling around the obtained optimum and generating the second Kriging model (“Surrogate₂” in Table 3) a second cross validation with 100 points was

Table 3: Computational results for non-ideal distillation. Cases with numbered a subscript higher than 1 are calculated using adaptive sampling techniques, cases labeled with M are calculated using MATLAB

Case	TAC	D	V	l_s	l_r
Local	153840	1.3747	5.0424	3	17
Surrogate ₁	71759	1.3772	2	8	17
Surrogate ₂	103355	1.3747	2.7036	11	20
Surrogate ₃	94255	1.3747	2.4403	13	18
Local _{M1}	93189	1.3747	2.4095	13	18
Local _{M2}	86604	1.3747	2.3380	17	11

done. The results are shown in the lower part of Figure 4. The second Kriging model is much more accurate and is able to model the discontinuity quite well. It turns out, that the new optimum obtained for Surrogate₂ satisfies the desired conditions stated in Equations (11).

Note that the values for l_r and l_s given by the computed optimal solution of the second Kriging model lie on the boundary of their respective sampling region. Hence, a third Kriging model (Surrogate₃) is generated around this optimum with $l_r, l_s \pm 3$, D within a 25% range and $V \in [2, 4]$. In the third optimization the objective value could be lowered further. Comparing the result of the local optimization with the iterative global optimization of the Kriging models, the objective value is finally lowered by 38.73%.

The purity of Toluene achieved with the operating conditions of “Surrogate₃” is higher than the required specification of 95%. To decrease the objective function value further, a local search using the obtained operating conditions as initial conditions is done in MATLAB and thereby reducing the purity to 95%. The results of that optimization can be found in Table 3 (see row “Local_M”).

This solution can be further improved by additional local optimization runs in the neighborhood of Local_M, leading to the solution that is shown in row “Local_{M2}” of Table 3, which is the best local optimum we found through our computations. However, the improvement achieved by applying “Local_{M2}” compared to the use of “Surrogate₃” is minor with respect to the improvement that is achieved by applying “Surrogate₃” compared to the use of “Local”.

5 Conclusion

In this work, two case studies concerning the distillation of multi-component mixtures have been conducted, where global optimization techniques have been applied to surrogate models of the distillation columns investigated.

It was shown that the reformulation developed earlier by the authors is not always applicable in the case of a distillation of a mixture with non-ideal VLE, rendering the problems unsolvable within a reasonable amount of time for standard deterministic global optimization software. In these cases, the iterative global optimization of surrogate models is a good alternative, which yields better optima than ordinary local solver and in some cases comes close to the global optimum. It can, however, not be guaranteed that the solution obtained by the optimization of

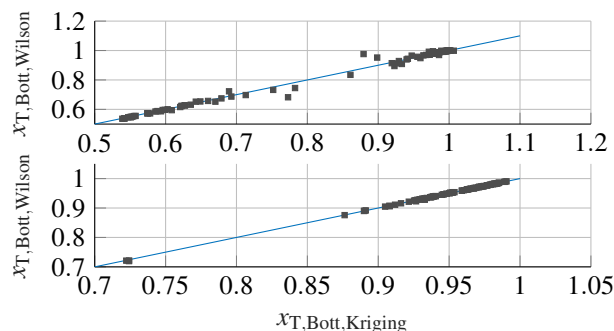


Figure 4: Cross validation of non-ideal distillation column in terms of $x_{T,Bott}$, the mole fraction of Toluene in the Bottom product. The upper part shows the results for the first Kriging model, the lower part shows the results for the second Kriging model.

these surrogate models is the actual global optimum or lies in a close neighborhood of it.

Further work will be concerned with more advanced adaptive sampling methods and algorithms for the generation of global models.

Acknowledgement: This work is part of the Collaborative Research Center “Integrated Chemical Processes in Liquid Multiphase Systems - InPROMPT”. Financial support by the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged through SFB/TRR 63.

References

- [1] Locatelli, M. and Schoen, F. *Global Optimization. Theory, Algorithms, and Applications*. MOS-SIAM Series on Optimization. SIAM, (2013).
- [2] Ballerstein, M., Kienle, A., Kunde, C., and Michaels, D. Deterministic global optimization of binary hybrid distillation/melt-crystallization processes based on relaxed MINLP formulations. *Optimization and Engineering* **16**, 409–440 (2015).
- [3] Mertens, N., Kunde, C., Kienle, A., and Michaels, D. A reformulation strategy for deterministic global optimization of ideal multi-component distillation processes. In *Proceedings of the 26th European Symposium on Computer Aided Process Engineering*, 691–696, (2016).
- [4] Nallasivam, U., Shah, V. H., Shenvi, A. A., Huff, J., Tawarmalani, M., and Agrawal, R. Global optimization of multicomponent distillation configurations: 2. enumeration based global minimization algorithm. *American Institute of Chemical Engineers Journal* **62**(6), 2071–2086 (2016).

- [5] Quirante, N., Javaloyes, J., and Caballero, J. A. Rigorous design of distillation columns using surrogate models based on Kriging interpolation. *American Institute of Chemical Engineers Journal* **61**(7), 2169–2187 (2015).
- [6] Caballero, J. A. and Grossmann, I. E. An algorithm for the use of surrogate models in modular flowsheet optimization. *American Institute of Chemical Engineers Journal* **54**(10), 2633–2650 (2008).
- [7] Papalambros, P. Y. and Wilde, D. J. *Principles of Optimal Design: Modeling and Computation*. Cambridge University Press, 2nd edition, (2000).
- [8] Koehler, J. R. and Owen, A. B. *Computer Experiments*, chapter 9, 261–308. Elsevier Science (1996).
- [9] Kocis, L. and Whiten, W. J. Computational investigations of low-discrepancy sequences. *ACM Transactions on Mathematical Software* **23**(2), 266–294 (1997).
- [10] Sahinidis, N. V. BARON: A general purpose global optimization software package. *Journal of Global Optimization* **8**, 201–205 (1996).
- [11] Dorn, C., Güttinger, T. E., Wells, G. J., Morari, M., Kienle, A., Klein, E., and Gilles, E. D. Stabilization of an unstable distillation column. *Industrial & Engineering Chemistry Research* **37**, 506–515 (1998).
- [12] Poling, B. E., Prausnitz, J. M., and O’Connell, J. P. *The Properties of Gases and Liquids*. McGraw-Hill Professional, 5th revised edition, (2000).

