Rigorous Design of Distillation Columns Using Surrogate Models Based on Kriging Interpolation

Natalia Quirante, Juan Javaloyes, and José A. Caballero

Institute of Chemical Processes Engineering, University of Alicante, PO 99E-03080 Alicante, Spain

The economic design of a distillation column or distillation sequences is a challenging problem that has been addressed by superstructure approaches. However, these methods have not been widely used because they lead to mixed-integer nonlinear programs that are hard to solve, and require complex initialization procedures. In this article, we propose to address this challenging problem by substituting the distillation columns by Kriging-based surrogate models generated via state of the art distillation models. We study different columns with increasing difficulty, and show that it is possible to get accurate Kriging-based surrogate models. The optimization strategy ensures that convergence to a local optimum is guaranteed for numerical noise-free models. For distillation columns (slightly noisy systems), Karush–Kuhn–Tucker optimality conditions cannot be tested directly on the actual model, but still we can guarantee a local minimum in a trust region of the surrogate model that contains the actual local minimum. © 2015 American Institute of Chemical Engineers AIChE J, 61: 2169–2187, 2015
Keywords: simulation, optimization, design (distillation columns), Kriging algorithm, mathematical modeling

Introduction

In recent decades, there has been increasing demand for computationally efficient process models in many engineering applications. In general, the process modeler has to tradeoff model accuracy against computational efficiency. This is especially true when we want to use the models for optimization. Most of these models have a modular structure to which users have limited internal access and they can only see them as "grey or black box models." Some of these models need a notable CPU computational time and/or their derivatives cannot be accurately estimated because they introduce numerical noise.

As the trend in process systems engineering goes toward integrating optimization models with different origins (from home made to complex computer fluid dynamics) involving multiple levels of decision (integration of planning and scheduling or real-time optimization), there is a demand for computationally efficient surrogate models[1] that at the same time ensure an acceptable degree of accuracy.[2]

Biegler et al.,[3] classified surrogate modeling into two categories: model order reduction and data driven modeling. Model order reduction reduces the complexity of a high fidelity model while, at the same time, retaining most of the structure of the original equations. Data driven modeling only uses data which comes from complex models or experiments, but it does not make use of the explicit formulation of any model. This article focuses on this second approach.

The use of simulation data to build models for individual parts of a large system can also be viewed as a multiscale modeling approach,[4] which is widely used in modeling-simulation of complex systems.

Data driven surrogate models have become popular due to their wide applicability.[5–7] In the context of optimization data driven surrogate models—for simplicity «surrogate models» hereinafter – they can be used in the following situations: (1) A mathematical model of the process does not exist but we can draw data from the actual process. (2) There is a model but it is too complex and cannot be reduced to the desired degree using model order reduction techniques. (3) The model can be used for simulation but cannot be included in an optimization environment.

There are different reasons why a simulation model cannot be included in an optimization environment.[7–9] If the user has access to the explicit model formulation the only reason for using surrogate models is that the model is too complex to be directly included in an optimization environment (e.g., computer fluid dynamic models). If the user does not have access to the explicit simulation model it might be possible to use the simulation model in an optimization environment as an "implicit" model; in fact, IPOPT[10] has been used to optimize ASPEN-based simulations.[11] However, this is only possible under some conditions:

• The simulation model requires short execution times. Note that eventually the number of evaluations of the simulation model might be very large, especially if derivative information is obtained from perturbations of independent variables. In this case, the total execution time might eventually become prohibitive.

• The simulation model must not generate numerical noise. Noisy models require large perturbations when

calculating derivative information. As a consequence, the «quality» of the Jacobian matrix is compromised, the behavior of the optimization solver becomes erratic[7,12] and the optimality conditions are difficult (or impossible) to verify.

• The simulation model always converges, or alternatively we have a specialized algorithm designed to recover from simulation convergence failures.

Alternatively, derivative-free optimization (DFO) is a class of algorithms designed to solve optimization problems when derivatives are unavailable, unreliable or prohibitively expensive to evaluate.[13–15] These solvers try to locate an optimal point using a limited number of black-box function calls. Although DFO algorithms can be used in models with costly and/or noisy function evaluations, these methods are often constrained to models in which the number of degrees of freedom does not exceed about 10.[13]

According to the framework of application, a single model of the objective function is usually approximated before optimization; in a few cases, the constraint set is modeled as well but in these cases precautions must be taken to avoid the small but accumulated errors in active constraints and objective function rendering feasible solutions in the surrogate model that could eventually be infeasible in the actual model.[7]

Additionally, instead of considering all the simulation as a black box, some existing techniques disaggregate the simulation model into different blocks and model each block separately before optimization, ensuring that all relevant connectivity variables are modeled.[15] In this way, by disaggregating the process, smaller and more robust models are generated. Moreover, if part of the actual model can be used in its original form (it is fast, reliable, and noise free) the number of surrogate models can be reduced and efficiency increased. These disaggregated process units can be linked via connectivity variables to formulate complex mixed-integer optimization models.[15]

It is worth remarking that not all approximating model approaches have the same accuracy when their parameters are estimated using a reduced number of data sets.[16] Polynomial[17] and Kriging[1] metamodels are often generated from small data sets. Regression by splines[18] usually requires moderately sized data sets, while neural networks[19] typically require quite large data sets. In this context, the work of Cozad and Sahinidis deserves special mention.[15] These authors, using a database of possible basic functions, are able to find the best set of basic functions that accurately reproduces the original model while, at the same time, using a reduced number of sampling points. Overfitting or the use of unnecessary terms is avoided by combining optimization techniques and statistical methods, in particular the Akaike information criteria.[20]

In this article, we focus on Kriging metamodels because they combine computational efficiency with relatively small sampling data. Typically, Kriging models are fitted to data that are obtained for larger experimental areas than the areas used in low order polynomial regression; that is, Kriging models are global (rather than local). These models have been used for both, sensitivity analysis and optimization.[21]

Significant work has been done with Kriging models using the full system approach[9,16,22] or by disaggregating parts of the model.[7] Palmer and Realff[16] studied the optimization of steady-state simulators using Kriging surrogate models. David and Ierapetritou[9] used complete process Kriging models to locate global model solutions and refine them using local response surfaces around the optima. Huang et al.,[22] used Kriging models on the complete processes to address uncertainty in black-box systems. Caballero and Grossmann[7] investigated disaggregated (modular) flow sheet optimization using Kriging models to represent process units with low-level noise. Henao and Maravelias[8] used disaggregated models for each unit in a flow sheet using artificial neural networks. An interesting review of Kriging applications in simulation can be found.[21]

The rigorous design of a distillation column or a column sequence is a challenging problem in chemical process engineering because it involves the simultaneous optimization of continuous decisions related to the operational conditions, and discrete decisions related to the number of trays in each column section; and in the case of column sequences, to column connectivity (i.e., a given column exists or not in the final design). One of the major challenges is to perform a tray by tray optimization in models assuming phase equilibrium.

One of the first approaches to designing a distillation column using mathematical programming was developed by Sargent and Gaminibandara.[23] Here, the objective was to locate the optimal feed location to a column with a fixed number of trays; to this end the column was modeled using the MESH equations (Mass balances, phase Equilibrium, molar fraction Summation equals one in all phases, and Enthalpy balances).

However, the first model that considers the simultaneous optimization of the feed tray position and total number of trays was that of Viswanathan and Grossmann.[24,25] The model takes the form of a Mixed Integer NonLinear Programming (MINLP) Problem and also relies on the MESH equations. A major difficulty with this model is related with nonexisting trays, which could produce numerical problems due to the lack of convergence of equilibrium equations with zero value in the flow of one of the phases. To avoid numerical problems related to the MINLP formulation, Yeomans and Grossmann[26,27] proposed a model based on Generalized Disjunctive Programming (GDP). The advantage of the disjunctive approach is that the MESH equations of the non-existing trays do not have to be converged, and therefore, flows in the column are not forced to take values of zero. Consequently, the convergence of the optimization procedure is more reliable.

Barttfeld et al.,[28] performed a comparative study of MINLP and GDP formulations concluding that GDP formulations are numerically more robust and require less solution time, but are more sensitive to initial values and bounds. To provide a good feasible initial design and good initial values for the rigorous optimization, Barttfeld and Aguirre[29,30] proposed a reversible distillation model that involves minimum reflux as well as minimum entropy generation; but this approach is limited to the so called "preferred separation."

Alternatively, Kraemer et al.,[31] proposed a successive relaxed-MINLP (SR-MINLP) which reformulates the MINLP or GDP problems as pure continuous problems with big-M constraints, where all discrete decisions are represented by continuous variables. The discrete decisions are enforced by nonconvex constraints that make the continuous variables take discrete values. Harward and Marquardt[32] proposed a multistage procedure in which the results of a stage are used as initial values of the following stage. It starts by calculating the minimum energy demand and the concentration profile based on pinch points, then they solve a simplified

model that comprises only component mass balances and equilibrium relations, but not enthalpy balances, then the energy balance is added and the final solution is obtained using an SR-MINLP approach.

Even with all these difficulties, complex problems have been successfully solved, for example Dunnebier and Pantelides[33] solved complex sequences of thermally coupled distillation columns. The synthesis of azeotropic sequences has been addressed by Bauer and Stichlmair,[34,35] or Barttfeld et al.,[36] and the reactive distillation was studied by Ciric and Gu,[37] and Jackson and Grossmann;[38] among others.

While the optimal design of distillation columns using mathematical programming approaches is still an open field of research, the simulation of distillation columns is a mature technology that is robust, reliable and used every day by people ranging from students to experienced professionals. However, distillation columns in commercial process simulators are input-output and slightly noisy black box models. Their use as a systematic design tool is a challenging problem, because some variables cannot be accessed or modified directly by the user; which sometimes introduces nondifferentiability.

Lang and Biegler[39] proposed a Distributed Stream-Tray Optimization method in which the feed flow and the reflux rates can be distributed to a set of candidate trays. To this end they proposed a Differentiable Distribution Function with which the location to the feed, reflux and other side streams can be treated as continuous instead of integer variables.

Caballero et al.,[40] proposed a GDP model, reformulated as an MINLP. The model iterates between an NLP model, which basically corresponds to a column with a fixed number of trays that was optimized with state of the art NLP solvers interacting with the Aspen-Hysys process simulator; and a tailored MILP master problem that was modified by adding to the objective function and to the constraints an extra term that reflects the contributions due to the addition or deletion of trays to each section of the column. The approach proved to be robust; however, the major drawback is the necessity of a tailored master problem, which prevents (or at least complicates) the inclusion of the model in a general flow sheet. Caballero,[41] proposed a new GDP model, solved using a logic based algorithm that avoids the necessity of a modified master problem. In that work it was assumed that a distillation column can be formed by selecting, for each column section, from a set of column sections with a different number of trays. A feasible selection of column sections results in a distillation column (or column sequence) that can be optimized using standard NLP algorithms and the master problems are generated in the usual way.

The above approaches have three major drawbacks. The first is that distillation column models tend to introduce some numerical noise. Consequently it is necessary tighten the convergence parameters (i.e., mass and energy balances) as much as possible and at the same time relax the optimality criteria in the NLP solver. Second, the derivative information must be obtained by perturbations, the perturbation parameter must be adjusted to ensure an accurate derivative and simultaneously reduce the effect of the numerical noise, which could be a very time consuming activity, even with fast convergence of the simulator. And third, lack of convergence of a simulation during the optimization makes the whole procedure fail.

In this article, we study the possibility of substituting distillation columns (single columns or complex arrangements) by surrogate models based on Kriging interpolation, involving both continuous and discrete variables, generated from rigorous models, and using these surrogates to design the distillation columns. In this way, we maintain the rigor of state of the art simulation models, remove most of the numerical problems and increase the reliability of the synthesis algorithms.

In the rest of the article, we first show a brief overview of Kriging interpolation. Then we discuss some aspects of the practical implementation, sampling points and the optimization algorithm. Finally we perform a comprehensive discussion through a set of case studies: the rigorous design of conventional distillation columns; both with fixed trays, and variable number of trays; a divided wall column; an extractive distillation system; we show the results integrating the models in a superstructure for a nonsharp separation; and finally, we study a demethanizer column.

An overview of Kriging interpolation

Kriging was originally developed in geostatistics (also known as spatial statistics) by the South African mining engineer Daniel G. Krige in his Master Thesis.[42]

The Kriging fitting is composed of two parts: a polynomial expression and a deviation from that polynomial

$$y(x) = f(x) + Z(x) \tag{1}$$

where $Z(x)$ is a stochastic Gaussian process that represents the uncertainty about the mean of $y(x)$ with expected value zero. The covariance for two points x_i and x_j is given by a scale factor σ^2 that can be fitted to the data and by a spatial correlation function $R(x_i,x_j)$. Three popular types of correlation functions, for a single input are

$$R(x_i, x_j) = \max\left[1 - \theta(x_i - x_j), 0\right] \quad \text{Linear correlation function}$$

$$R(x_i, x_j) = \exp\left(-\theta(x_i - x_j)\right) \quad \text{Exponential correlation}$$

$$R(x_i, x_j) = \exp\left(-\theta(x_i - x_j)^2\right) \quad \text{Gaussian correlation}$$

where $\theta \geq 0$ is an adjustable parameter.

However, the most common alternative for the spatial correlation function in Kriging models is to use the extended exponential that appears in the influential article by Sacks et al.[43]

$$R(x_i, x_j) = \exp\left(-\sum_{l=1}^{d} \theta_l (x_{i,l} - x_{j,l})^{P_l}\right) = \prod_{l=1}^{d} \exp\left(-\theta_l (x_{i,l} - x_{j,l})^{P_l}\right) \tag{2}$$

where $\theta_l \geq 0$ and $0 \leq P_l \leq 2$ are adjustable parameters. The value of θ_l shows how fast the correlation goes to zero as we move in a lth coordinate direction. The parameter P_l determines the smoothness of the function which is usually fixed to 2 (Gaussian Kriging) in all coordinates.

An important advantage of Kriging models is that the degree of the polynomial $f(x)$ does not significantly affect the fit quality because $Z(x)$ captures the most significant behavior of the function.[1] A constant term μ (Ordinary Kriging) usually suffices in practice.[14,21]

To estimate the values of the parameters σ^2, θ_l, P_l, and μ, it is possible to maximize the likelihood of the obtained data y

$$L(x, \sigma, \mu) = \frac{1}{(2\pi\sigma^2)^{(n/2)}|R|^{(1/2)}} \exp\left(\frac{-(y-1\mu)^T R^{-1}(y-1\mu)}{2\sigma^2}\right)$$
(3)

where y is the vector of obtained responses ($n \times 1$), $\mathbf{1}$ is a vector of ones ($n \times 1$), and n is the number of sampled points.

Instead of maximizing the likelihood of the data y, it is better to maximize the logarithm of the likelihood function

$$\log(L) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\ln(|R|)$$
$$- \left(\frac{-(y-1\mu)^T R^{-1}(y-1\mu)}{2\sigma^2}\right)$$
(4)

The optimal values for μ and σ^2 are obtained by differentiating Eq. 4 with respect to σ^2 and μ, and equating it to zero. After some algebra we get

$$\hat{\mu} = \frac{1^T R^{-1} y}{1^T R^{-1} 1}$$
(5)

$$\hat{\sigma}^2 = \frac{(y-1\hat{\mu})^T R^{-1}(y-1\hat{\mu})}{n}$$
(6)

If we want to interpolate a new point x_{new}, we have to add the point (x_{new}, y_{new}) to the data and compute the augmented likelihood function keeping all the parameters at the previously calculated values. With all the parameters constant, the logarithm of the likelihood function is only a function of y_{new}. Consequently, the predicted value for y_{new} will be the value that maximizes the augmented likelihood function. The following equation gives the final predictor of the Kriging method. A detailed derivation can be found in Sasena[44]

$$\hat{y}(x_{new}) = \hat{\mu} + r^T R^{-1}(y-1\hat{\mu})$$
(7)

where r ($n \times 1$) is the vector of correlations $R(x_{new}, x_i)$ between the sample design points and the point to be correlated.

We can estimate the accuracy of the method with the correlation of the errors. To this end, we can use a full derivation[43] of the formula for the mean-squared error

$$s^2(x_{new}) = \hat{\sigma}^2 \left(1 - r^T R^{-1} r + \frac{(1 - r^T R^{-1} r)^2}{1^T R^{-1} 1}\right)$$
(8)

In this expression, the term $-r^T R^{-1} r$ represents the reduction in prediction error due to the fact that x_{new} is correlated with the sampled points. The term $(-r^T R^{-1} r)^2 / 1^T R^{-1} 1$ represents the uncertainty derived from our lack of knowledge of the exact value of μ. A good review of Kriging interpolation in simulation can be found.[21]

Using Kriging in distillation design

Even though all the considerations for a simulation model also apply to distillation systems, there are some particular details in the implementation that deserve a detailed description.

Jones[45] presented a review of a set of seven different methods for global optimization using Kriging surrogates. In these methods, the search starts with a limited number of sampling points, the Kriging is adjusted and then the mini-

mum of the Kriging interpolator is located. The algorithms differ on how the new sampling points in the next iterations are selected (minimizing a lower bounding function, maximizing the probability of improvement, maximizing the expected improvement, etc.). However, some of these alternatives involve a spatial branch and bound search at each major iteration, which could become numerically expensive. Instead, we focus on algorithms that guarantee a local minimum (Method 2 in Jones' taxonomy).[45]

The algorithm implementation is as follows:

a. Sample N points (see comments further) using a pre-specified space-filling design. The sampled points must be separated enough to ensure that the noise generated by the simulation does not significantly affect the Kriging model.

b. Fit a Kriging model to the I/O simulation data. Because Kriging does not incorporate cross-correlation between the different simulation outputs, univariate Kriging models are fitted.[21]

c. Using cross-validation[14] for a set of test simulations, validate the accuracy of the model. If the error is small enough the Kriging model can be used to substitute the actual one without further considerations. However, in general this is not the case.

d. Substitute the actual model by the Kriging surrogate and perform the (MI)NLP optimization. If the Kriging model is considered a good approximation of the actual model then finish. Otherwise continue with the next point (e).

e. Add the optimal point obtained in the previous step to the set of sampled points. Update Kriging and reoptimize. If in two consecutive iterations there is no improvement then test for optimality by going to step (f).

In "well behaved" simulation models like those for distillation columns, we can finish in step (e). However, Jones[45] pointed out that in the algorithm presented above the stopping criterion should be selected carefully. In particular, if this criterion were «stop when the minimum of the surface is within a distance $\varepsilon > 0$ of a sampled point» the algorithm cannot guarantee a local minimum. Instead, we can sample in a small neighborhood of the tentative solution to force the gradient of the surface to agree with the gradient of the true function. Adding "gradient matching" whenever the minimum of the surface is near a sampled point certainly makes the algorithm more robust. However, even using gradient matching sampling cannot guarantee a local minimum if the stopping rule is based on small improvement in two successive iterations. Instead, local convergence can be ensured by combining gradient matching with a trust region approach.[7,46] With this approach to get the next iterate, the surface is optimized within a trust region around the incumbent solution. If there is no improvement then the region is contracted. Alexandrov et al.,[46] proved that this approach converges to a critical point. Therefore the next step should be:

f. Contract the feasible search region within a trust region around the incumbent solution and repeat steps d to f. In a completely free of error simulation system, this approach can be repeated until we can guarantee that the error in the gradient is below a given tolerance and test the Karush–Kuhn–Tucker optimality conditions. In a noisy system it is not possible to follow this approach, and we must finish with the nonimprovement criterion in a small but large enough region.

As previously noted, the simulation columns produce slightly noisy models. In these systems the gradient matching approach must be done carefully. The sampled points should

Table 1. Common Data for Examples

Distillation Column	
Material of construction	Carbon steel (sieves and tower)
Tray spacing	Ht (m) = 0.6096
Column height	H (m) = 3 + N·Ht
Heat Exchangers	
Condenser	U (W/(m² K) = 800
Reboiler	U (W/(m² K) = 820
Material of construction	Carbon steel (shell and tubes)
Utility Costs	
Low pressure steam (500 kPa, 433 K)	14.05 $/GJ
High pressure steam (4100 kPa, 527 K)	17.70 $/GJ
Cooling water (303–318°C)	0.354 $/GJ
CEPCI cost index 2001	394.3
CEPCI cost index 2013	567.3

Calculation based on 8000 h/year of operation).

be near enough to get an accurate representation of the gradient, but at the same time should be separate enough to avoid "adjust the noise" (separation should be at least one order of magnitude larger than the magnitude of the noise). Also, if it is possible to get tight bounds of the independent variables, and if the number of points is large enough to guarantee that the whole region is approximated by the Kriging model inside some "acceptable error," then we can directly substitute the distillation column by the Kriging interpolator without further iterations.

If we are interested in a good Kriging model with reduced initial error, to avoid resampling and recalibrating the Kriging parameters as much as possible, a correct distribution of sampling points is mandatory. A correct sample must cover all the space defined by the independent variables. Simple random (Monte Carlo) methods result in large confidence intervals and variance. Hence, the randomness for approximating a distribution is not critical,[47] but the error of approximating a function with a finite set of points depends more on the uniformity of the distribution than on its randomness. Therefore, the sampling must be done to preserve uniformity and avoid correlation. Variation reduction techniques like Latin Hypercube Sampling,[48,49] Hammersley,[50] Halton or Sobol sequences,[51] or infill sampling procedures, are better choices.

In this article, we select an *a priori* infill procedure: the maxmin approach; maximize the minimum distance between two sample points. However, instead of distributing N points following the maxmin approach, we fix 2^D points to the D-dimensional vertex of the hypercube that forms the feasible space (i.e., we sample in the bounds of all the independent variables) and then we distribute the other $(N-2^D)$ points following the maxmin approach. In this way we ensure that Kriging does not perform "extrapolations" near the corners of the feasible region.

Examples

Five study cases are presented to verify the effectiveness of the method. The first example consists of the rigorous design of conventional distillation columns, first, a column with a fixed number of trays, and then a column in which we also determine the optimum number of trays. The second example consists of a divided wall column. The third exam-

ple is an extractive distillation system. We show the results for a superstructure for a nonsharp separation in fourth example. Finally, the fifth example consists of a demethanizer column. In all cases, we enumerate the column trays from the top of the column to the bottom. All the examples were solved on a computer with a 2.60 GHz Pentium® Dual-Core Processor and 4 GB of RAM under Windows 7.

The common data used in the examples are shown in Table 1.

Objective Function. The objective function, common to all the examples, consists of minimizing the total annualized cost (TAC) of a chemical process. The objective includes the capital cost of the main items (column shell, trays and heat exchangers) and the most important operating costs (cooling water and vapor steam). The estimations of the capital costs are calculated using the correlations given by Turton et al.[52] Operating costs are calculated from the heat loads of the reboiler and condenser, through the steam and cooling utility costs. The objective function is determined by the following expression

$$\min \text{TAC}\left(\frac{\$}{\text{year}}\right) = C_{op} + F \cdot C_{cap} \qquad (9)$$

where C_{op} is the operating cost per year, F is an annualization factor, and C_{cap} is the capital cost of the equipment, both updated by the global CEPCI cost index of 2013. The annualization factor is calculated by the following equation, recommended by Smith[53]

$$F = \frac{i(1+i)^n}{(1+i)^n - 1} \qquad (10)$$

where i is the fractional interest rate per year and n is the horizon time. In this work, we use a fixed interest rate of 10% and a horizon time of 5 years.

Example 1. Conventional Distillation Column. This example is presented to illustrate the behavior of a modular process simulator and the general procedure followed in more complex examples. Consider that we want to determine the optimal configuration and operating conditions that minimize the TAC of a distillation column for separating a mixture of benzene, toluene and p-xylene (molar fractions: 0.35, 0.35, and 0.30, respectively), simulated in ASPEN-HYS-YS™.[54] We want to recover at least 98.5% of the benzene fed with at least 0.999 in molar fraction.

The data for the problem are shown in Table 2.

Case a. Distillation column (fixed trays): In this case, we want to determine the optimal operating conditions (reflux ratio and boilup ratio) to minimize the TAC of a distillation column with 30 theoretical trays (feed fixed in tray 15).

Table 2. Data for Example 1

Conventional Distillation Column (30 Trays [Feed in Tray 15])	
Feed	100 kmol/h
Composition (mole fraction)	
benzene	0.35
toluene	0.35
p-xylene	0.30
Pressure	101.3 kPa
Thermodynamics (fluid package)	Peng-Robinson
Specifications	
Molar recovery of benzene in distillate	≥98.5 %
Molar fraction benzene in distillate	≥0.999

Table 3. Minimum Cost for Example 1a.

Optimal Solution (Distillation Column (30 Trays [Feed in Tray 15]))	
TAC (M$/year)	0.5538
Operating Cost (M$/year)	0.4685
Capital Cost (M$/year)	0.0853
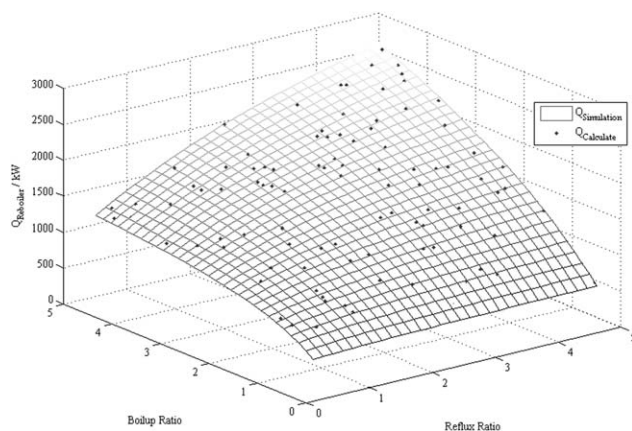
Figure 1. Comparison of actual and interpolated values.
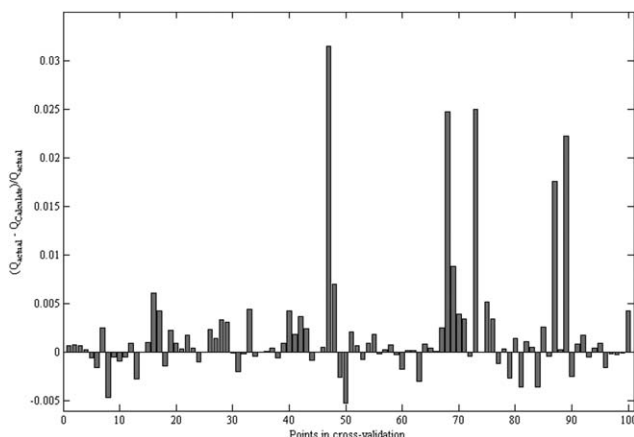
Figure 2. Relative errors in the sampled points.

If we assume that the nominal pressure in the column remains constant at 101.3 Pa there are two degrees of freedom. It would be possible to select as degrees of freedom the purity and molar flow of the benzene in distillate, and we would not need optimization at all, but as a general rule it is better to choose a set of independent variables that facilitates the convergence as much as possible (reflux ratio and boilup

ratio are «easy to converge» specifications). In this small test problem, even though we have no convergence problems, we follow the general approach to illustrate the procedure.

So the model we want to solve can be conceptually written as follows

$$\min : f = TAC$$

$$\text{s.t.} \quad \left[x_{\text{Benzene}}^{\text{Distillate}}, D_{\text{Benzene}}^{\text{Distillate}}, Q_{\text{reb}}, Q_{\text{Cond}}, D\right] = Kr(RR, BR)$$

$$x_{\text{Benzene}}^{\text{Distillate}} \geq 0.999$$

$$F_{\text{Benzene}}^{\text{Distillate}} \geq 0.985 \, F_{\text{Benzene}}^{\text{Feed}}$$

$$0.5 \leq RR \leq 5.0$$

$$0.5 \leq BR \leq 5.0$$

$$(11)$$

where $x_{\text{Benzene}}^{\text{Distillate}}$ refers to the molar fraction of benzene in distillate, and F refers to molar flows, Q_{Reb} and Q_{Cond} are the heat flows, and D is the column diameter. The function Kr refers to the Kriging interpolators. RR and BR are the reflux ratio and boilup ratio, which are used as degrees of freedom.

Using 100 sampling points, we calibrate the Kriging parameters with the values of dependent variables for each simulation. Next, the column is replaced by the surrogate model and, using state-of-the-art optimizers (CONOPT[55] and SNOPT[56]) available in TOMLAB-MATLAB,[57] the column can be optimized.

The solution to this problem is $x = [2.7353, 1.7818]$ (reflux and boilup ratios, respectively) and the objective function is 0.5538 M$/year. In Table 3, the TAC is summarized.

To validate the effectiveness of the Kriging metamodel, we have to evaluate the error of the surrogate model. To this end, we use a set of extra random simulations and compare the results of the simulation with the results obtained with Kriging.

In Figure 1a, comparison of the reboiler heat flow in the simulated points (surface) with the calculated points through the Kriging metamodel (random points) is shown.

The Kriging metamodel is adequate in the entire interval. In Figure 2, we can see, with more accuracy, the relative error in the sampled points.

The largest relative error is around 3%. Therefore, we can replace the entire column by the Kriging metamodel obtaining good results.

The results obtained with the Kriging metamodel and the optimal values substituted in the simulator are shown in Table 4. As we can see, the results obtained with the Kriging metamodel are equal to the data obtained with the simulator, inside the uncertainty produced by the numerical noise.

The total CPU time used in the optimization (including sampling, Kriging calibration and model optimization) was around 19 s.

Table 4. Summary of Optimal Solution Obtained with Different Methods for Example 1a

	Reflux Ratio	Boilup Ratio	Condenser Heat Flow (kW)	Reboiler Heat Flow (kW)	Diameter (m)	TAC (M$/year)
KRIGING	2.7353	1.7818	1092.7	1130.3	1.0059	0.5538
HYSYS	2.7353	1.7818	1092.9	1130.5	1.0060	0.5539

Number of sampled points = 100.
CPU time sampling = 9.37 s
CPU optimization Kriging parameters = 8.28 s
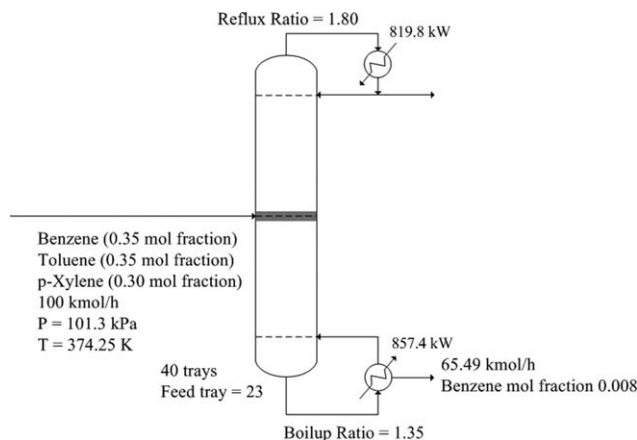CPU optimization model = 1.12 s.

Figure 3. Optimal solution for example 1b.

Table 5. Minimum Cost for Example 1b

Optimal Solution (Distillation Column (40 Trays [Feed in Tray 23]))	
TAC (M\$/year)	0.4397
Operating Cost (M\$/year)	0.3553
Capital Cost (M\$/year)	0.0844

Case b. Distillation column (variable number of trays): In this case we are interested in determining the optimal configuration (total number of trays and feed location) besides the operating conditions, to minimize the TAC of a distillation column for separating the same mixture. With the nominal pressure fixed ($P = 101.3$ kPa) there are four degrees of freedom. We choose two more independent variables for the rigorous design of the conventional distillation column: number of trays in the rectifying section and number of trays in the stripping section.

The model can be conceptually written as follows

$$\min : f = TAC$$

$$\text{s.t.} \ \left[x_{\text{Benzene}}^{\text{Distillate}}, D_{\text{Benzene}}^{\text{Distillate}}, Q_{\text{reb}}, Q_{\text{Cond}}, D \right] = \text{Kr}(RR, BR, N_R, N_S)$$

$$x_{\text{Benzene}}^{\text{Distillate}} \geq 0.999$$

$$F_{\text{Benzene}}^{\text{Distillate}} \geq 0.985 \, F_{\text{Benzene}}^{\text{Feed}}$$

$$15 \leq N_R \leq 30$$

$$15 \leq N_S \leq 30$$

$$[N_R, N_S] \in \text{ Integers}$$

$$(12)$$

The previous problem is a nonconvex MINLP problem.

The solution to this problem is $x = [1.80, 1.35, 22, 17]$ (Reflux ratio, Boilup ratio, number of trays in the rectifying

section, and number of trays in the stripping section, respectively) and the objective function is 0.4397 M\$/year. Figure 3 shows the optimal solution for example 1b.

Costs to the optimal solution are shown in Table 5.

The results obtained with the Kriging metamodel and the simulator are shown in Table 6. Again the errors are inside the noise of the model generated by the process simulator.

The total CPU time used in the optimization (including sampling, Kriging optimization, and model optimization) was around 27 s.

In previous examples, Kriging models were used in an implicit form. In other words, the equations of the Kriging are not directly accessed by the solver. The Kriging interpolation is maintained as an external function in Matlab that receives from the optimization solver the point to be interpolated and returns the interpolated value. From the point of view of the optimization solver it is a black box model. In this way we get a small MINLP model and at the same time take advantage of the effective matrix manipulation of MATLAB,[TM,58] where Kriging was interpolated. The price we have to pay is that we can only ensure a local optimum. Alternatively, if the Kriging surrogate is an accurate representation of the model, we can explicitly introduce the Kriging equations in a deterministic global optimizer like BARON,[59] we used the GAMS-BARON modeling system.[60] With this approach the optimization solver "can see" explicitly the interpolation equations and we can ensure a global optimum. Of course, it is the global optimum of the surrogate model, but if the surrogate model accurately reproduces the actual model, then we can expect that the solution is also the global optimum of the actual model.

Explicitly include the equations is only practical for systems involving a reduced number of Kriging models with a relatively low number of sampling points to avoid very large matrices. However, for a single distillation column it is a feasible approach.

Therefore, we repeat example 1b, explicitly introducing the equations in the solver. The global optimal solution was the same as with the implicit model, confirming that we obtained the global optimal solution. Some statistics of this approach are shown in Table 7.

Example 2. Divided wall column. In this example we want to optimize a Divided Wall Column (DWC) for separating a mixture of benzene, toluene and xylene (molar fractions: 0.35, 0.35, and 0.30, respectively).

Fully (or partially) thermally coupled distillations are gaining more acceptance due to their lower energy consumption compared with conventional distillation column sequences.[61–65] For an N-component mixture, a fully thermally coupled configuration (Petlyuk[66] in a three component mixture) needs $N-1$ columns with a single reboiler and a single condenser.[67–74] For a ternary separation, we need two columns to obtain the separation with three pure products.

Table 6. Summary of Optimal Solution Obtained with Different Methods for Example 1b

	Number of Trays Rectifying Section	Number of Trays Stripping Section	Condenser Heat Flow (kW)	Reboiler Heat Flow (kW)	Diameter (m)	TAC (M\$/year)
KRIGING	22	17	819.80	857.38	0.8679	0.4397
HYSYS	22	17	819.82	857.38	0.8679	0.4397

Number of sampled points = 100.
CPU time sampling = 14.17 s.
CPU optimization Kriging parameters = 11.61 s.
CPU optimization model = 1.19 s.

Table 7. Summary of Optimal Solution Obtained with Implicit and Explicit Method for Example 1b

	Number of Trays Rectifying Section	Number of Trays Stripping Section	Condenser Heat Flow (kW)	Reboiler Heat Flow (kW)	Diameter (m)	TAC (M$/year)
Implicit	22	17	819.80	857.38	0.8679	0.4397
Explicit	22	17	819.80	857.38	0.8679	0.4397

Number of sampled points = 100.
CPU time sampling = 14.17 s.
CPU optimization Kriging parameters = 11.61 s.
CPU optimization model = 369 s.
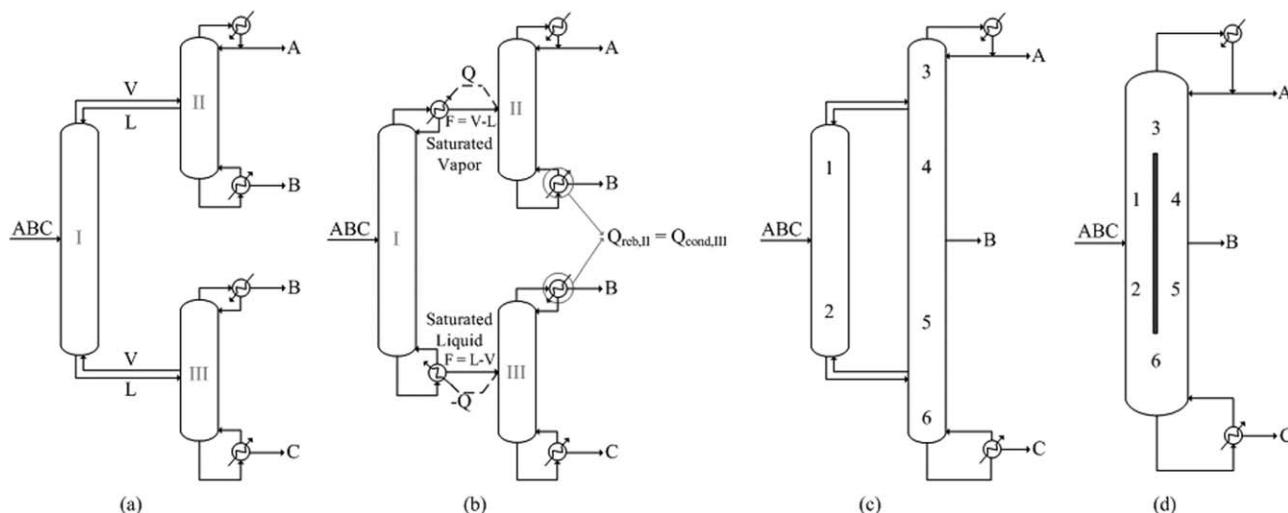
Figure 4. Equivalent configurations.

(a) The thermal couple can be substituted by a superheated stream (distillation section) or subcooled stream (stripping section) [75] and (b) or by a saturated stream and a heat stream,[76] with negligible error. (c) Petlyuk configuration. (d) Thermodynamically equivalent divided wall column.
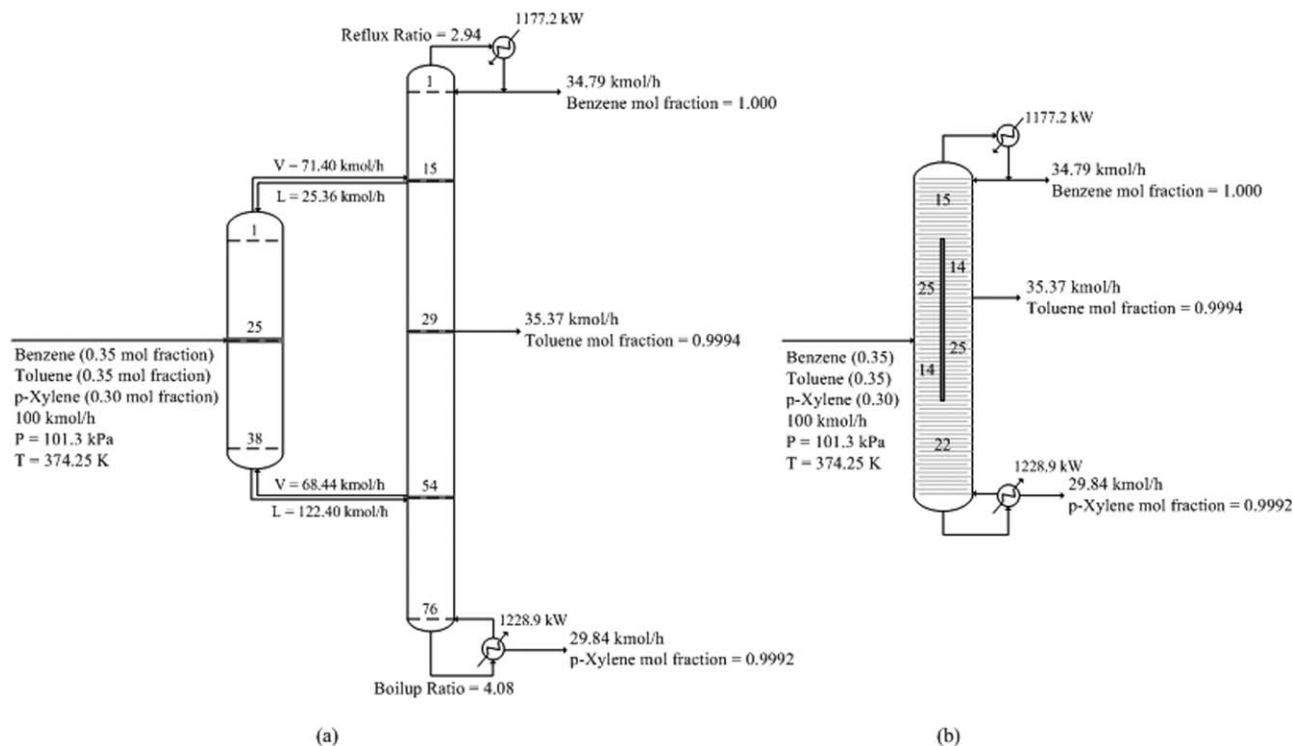
Figure 5. Optimal solution. (a) Simulation of the column Petlyuk. (b) Divided wall column.

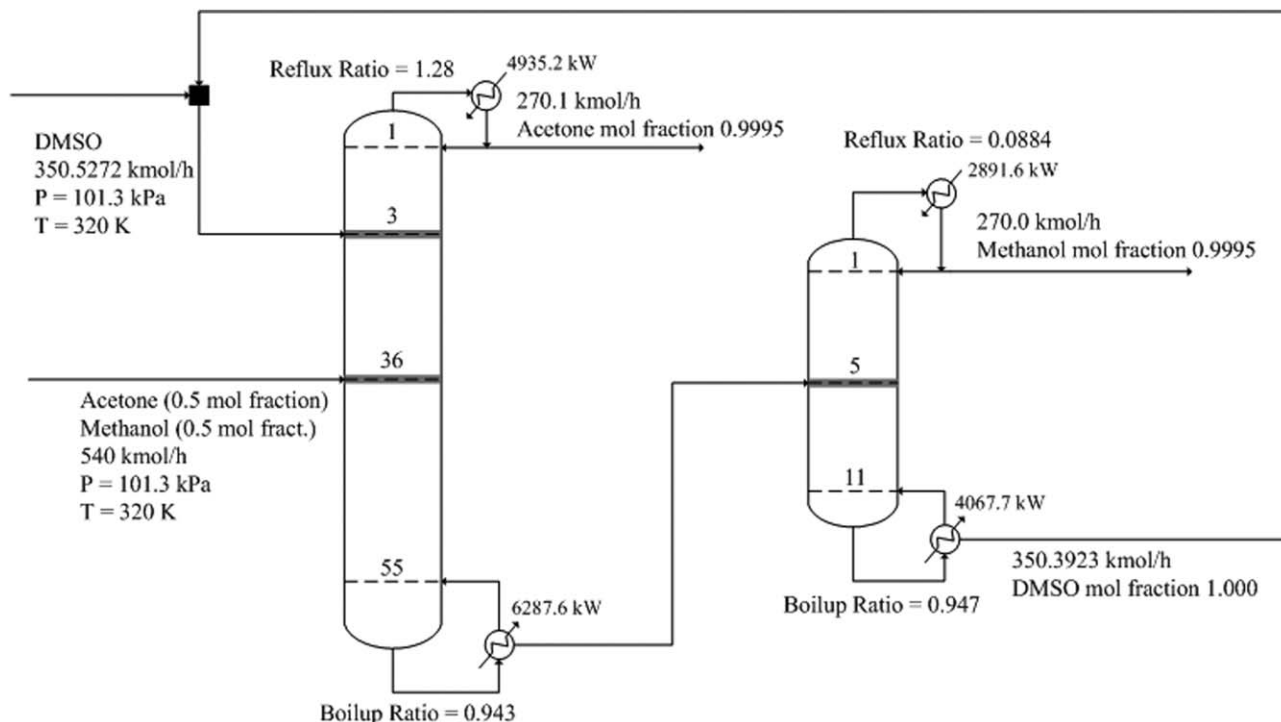**Figure 6. Optimal solution for example 3.**

Table 11. Minimum Cost for Example 3

Optimal Solution (Extractive Distillation System)	
TAC (M$/year)	5.056
Operating Cost (M$/year)	4.698
Capital Cost (M$/year)	0.358

equivalent to the energy added if we include a reboiler which provides vapor to the first column and the material stream is liquid at its bubble point (Figure 4b).

We want to determine the optimal configuration (total number of trays and feed location of the three columns) to minimize the TAC of the divided wall column. With the nominal pressure fixed ($P = 101.3$ kPa) there are six degrees of freedom. As design specification we fix the recovery of each component to be at least at least 99.9% of the component fed. The independent variables chosen for the rigorous design of the conventional distillation column are the number of trays in the rectifying section and number of trays in the stripping section for each column.

Conceptually the model can be written as follows

$$\min : f = \text{TAC}$$

$$\text{s.t. } [D_1] = \text{Kr}_1(N_{R1}, N_{S1})$$

$$\left[x_i^{\text{Distillate,C2}}, F_i^{\text{Distillate,C2}}, x_i^{\text{Bottoms,C2}}, F_i^{\text{Bottoms,C2}}, Q_{\text{Cond}}, D_2 \right]$$
$$= \text{Kr}_2(N_{R2}, N_{S2})$$

$$\left[x_i^{\text{Distillate,C3}}, F_i^{\text{Distillate,C3}}, x_i^{\text{Bottoms,C3}}, F_i^{\text{Bottoms,C3}}, Q_{\text{Reb}}, D_3 \right]$$
$$= \text{Kr}_3(N_{R3}, N_{S3})$$

$$14 \leq N_{R1} \leq 35$$
$$14 \leq N_{S1} \leq 35$$
$$14 \leq N_{R2} \leq 35$$
$$14 \leq N_{S2} \leq 35$$
$$14 \leq N_{R3} \leq 35$$
$$14 \leq N_{S3} \leq 35$$
$$[N_{R1}, N_{S1}, N_{R2}, N_{S2}, N_{R3}, N_{S3}] \in \text{Integers}$$

$$(13)$$

Although we simulate two columns separately, all the column sections are actually located inside a single shell. For this, in the mathematical model we force the number of trays in the stripping section of the column II plus the number of

trays in the rectifying section of the column III to equal the number of trays of the first column.

The equivalent divided wall column is obtained by integrating the two columns of the Petlyuk configuration in a single shell. The divided wall column has an interior wall that divides the column shell. At the top of the column, the number of trays corresponds with the number of trays in the rectifying section of column II. At the bottom of the column, the number of trays corresponds with the number of trays in the stripping section of column III. Furthermore, on one side of the divided area the number of trays corresponds with the number of trays of the first column and, on the other side of the divided area the number of trays corresponds with the sum of the number of trays in the stripping section of column II and the number of trays in the rectifying section of column III. Figure 5 shows the optimal solution.

The minimum objective function (TAC = 0.6893 M$/year) is obtained with a column of 76 trays, with the feed on the 40th tray. Benzene and p-xylene are extracted on the top and the bottom of the column, respectively, and toluene is extracted on the 29th tray. The costs of the optimal solution are shown in Table 8.

The results obtained with the Kriging metamodel and the simulator are shown in Table 9. As we can see, the results obtained with the Kriging metamodel only have a 5% error with respect to the data obtained with the simulator. As expected, it is larger than in the first example due to the larger number of independent variables but it is still a good result.

The total CPU time used in the optimization (including sampling, Kriging optimization and model optimization) was around 8 min.

Example 3. Extractive distillation system. The fundamental principle of the distillation process is the difference in volatility of the components forming the mixture. However, there are cases in which mixtures of components have close-boiling points or homogeneous azeotropes. When this occurs, conventional distillation may not be able to promote a separation with the desired degree of purity. Extractive distillation is considered as an alternative distillation process in these cases.

Therefore, the third example studied was the optimization of an extractive distillation system. This kind of distillation is used to separate mixtures by the addition of an entrainer, a component with a high boiling point.

The study case is adapted from Luyben.[83] The objective is the separation of an isomolar mixture of methanol and acetone using dimethyl sulfur oxide (DMSO) as the entrainer.

The extractive distillation system consists of two distillation columns. The first column, extractive column, has two

Table 12. Summary of Optimal Solution Obtained With Different Methods for Example 3.

	Extractive Column			Recovery Column			
	Condenser Heat Flow (kW)	Reboiler Heat Flow (kW)	Diameter (m)	Condenser Heat Flow (kW)	Reboiler Heat Flow (kW)	Diameter (m)	TAC (M$/year)
KRIGING	4935.2	6287.6	1.9943	2891.6	4067.7	1.7019	5.056
HYSYS	4989.1	6302.9	2.0021	2896.0	4083.4	1.7053	5.071

Extractive Column:
  Number of sampled points = 196.
  Sampling time = 118.33 s.
  CPU optimization Kriging parameters = 33.51 s.
Recovery Column:
  Number of sampled points = 99.
  Sampling time = 23.90 s.
  CPU optimization Kriging parameters = 5.55 s.
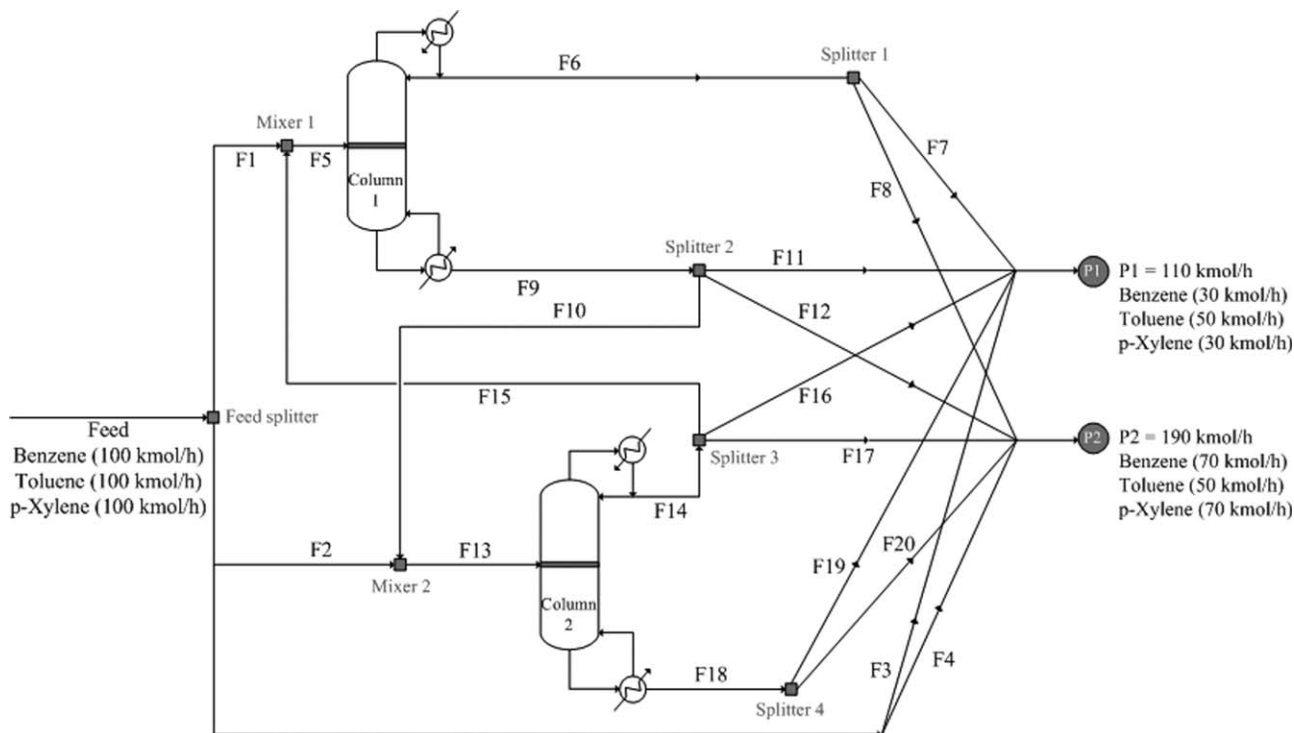CPU optimization model = 3.76 s.

Figure 7. Superstructure for a three-component system.

feeds, the mixture to be separated and the entrainer. The entrainer is fed above the mixture. In this column, one of the components is obtained at the top of the column, while the other is extracted together with the entrainer at the bottom of the column. The second column, recovery column. is fed with the bottom of the extractive column. In this column, the entrainer is separated from the component. Finally, the entrainer can be recirculated to the first column.

Assume we want to recover acetone at the top of the first column with at least 0.9995 in molar fraction, methanol at the top of the second column with at least 0.9995 in molar fraction and the entrainer with at least 0.9999 in molar fraction at the bottom of the second column. We want to determine the optimal configuration (total number of trays and feed location of the two columns) to minimize the TAC of the system. If we fix the nominal pressure ($P = 101.3$ kPa) and all the recoveries in both columns (equivalent to fix the compositions) then there are seven degrees of freedom (the number of trays in each column section and the entrainer molar flow). The data for this example are shown in Table 10.

Conceptually, the model can be written as follows

$$\min : f = \text{TAC}$$
$$\text{s.t.} \; [D^{C1}, Q_{\text{Reb}}^{C1}, Q_{\text{Cond}}^{C1}] = \text{Kr}_1(N_{R1}, N_{M1}, \; N_{S1}, \; E)$$
$$[D^{C2}, Q_{\text{Reb}}^{C2}, Q_{\text{Cond}}^{C2}] = \text{Kr}_2(N_{R2}, \; N_{S2}, \; E)$$
$$1 \leq N_{R1} \leq 8$$
$$25 \leq N_{M1} \leq 34 \tag{14}$$
$$15 \leq N_{S1} \leq 24$$
$$2 \leq N_{R2} \leq 10$$
$$4 \leq N_{S2} \leq 10$$
$$[N_{R1}, N_{M1}, N_{S1}, N_{R2}, N_{S2}] \in \text{Integers}$$

The flow of entrainer has both upper and lower bounds for feasible separations;[83] therefore, *a priori* sensitivity analysis

was performed to determine feasible values. Conservative bounds were used in the optimization to avoid convergence problems in the sampling step. As the optimal value was not at its bound no further action was required, however, if the optimal value had lain at a bound, it would have been necessary to resample to ensure we got the optimal solution.

In Figure 6, the optimal solution of the extractive distillation system studied is shown.

The minimum objective function (TAC = 5.056 M\$/year) is obtained with an extractive column of 55 trays, with the entrainer on the 3rd tray and the feed on the 36th tray, and a recovery column of 11 trays, with the feed on the 5th tray. The costs of the optimal solution are shown in Table 11.

The results obtained with the Kriging metamodel and the simulator are shown in Table 12. As we can see, the results obtained with the Kriging metamodel are very close to the data obtained with the simulator.

The total CPU time used in the optimization (including sampling, Kriging optimization, and model optimization) was around 3 min 5 s.

Example 4. Distillation sequences. Nonsharp separations. In this example, we want to separate a given multicomponent feed stream into several desired multicomponent product streams. The superstructure is proposed by Aggarwal and Floudas[84] and contains all possible alternatives to get the

Table 13. Data for Example 4

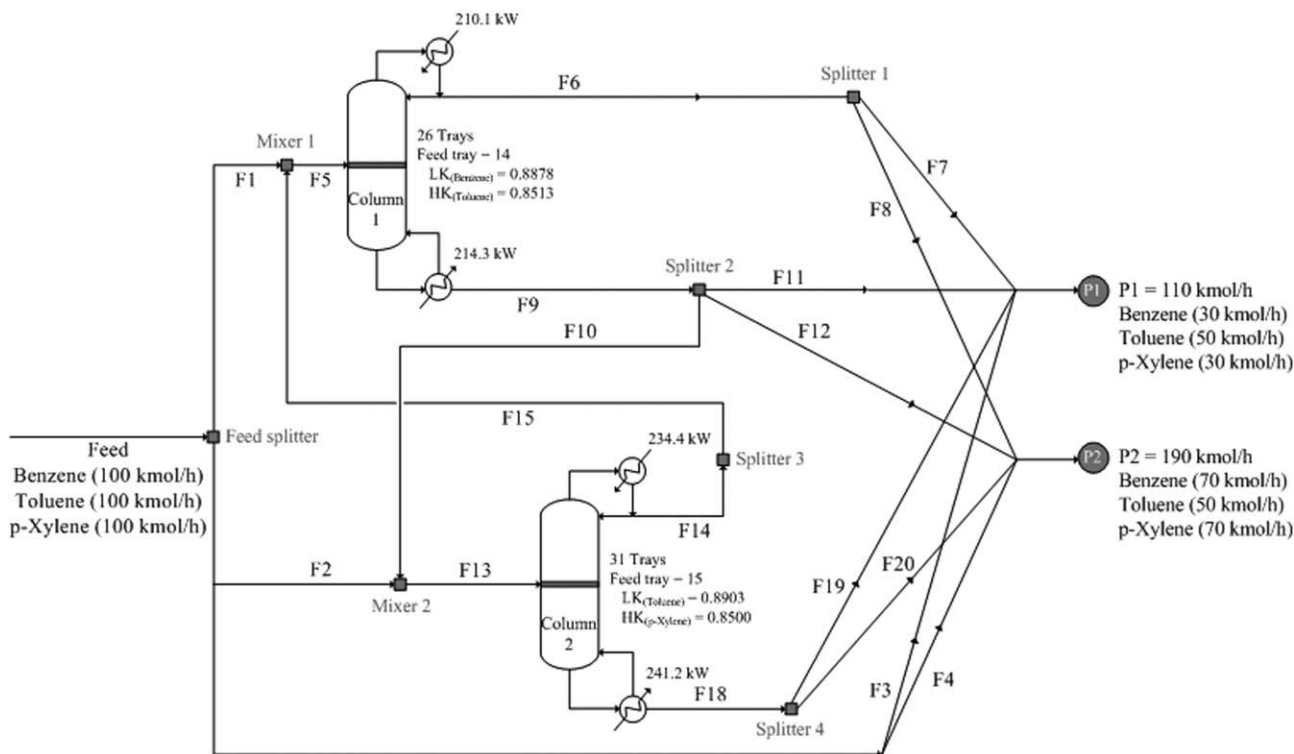| Distillation Sequence. Nonsharp Separation	
Feed	300 kmol/h
Composition (Molar Flow)	
Benzene	100 kmol/h
Toluene	100 kmol/h
p-Xylene	100 kmol/h
Pressure	101.3 kPa
Thermodynamics (fluid package)	Peng-Robinson

Figure 8. Optimal solution for example 4.

desired products (Figure 7). Consider that we want to determine the optimal configuration (sequence and number of trays of the columns) to minimize the TAC for separating an isomolar mixture of benzene, toluene and p-xylene.

In the original problem Aggarwal and Floudas[84] indicated that the lower bounds on key component recoveries to avoid the distribution of nonkey components must be fixed no lower than 0.85. So that is the value we use as the lower bound.

Data for this example are shown in Table 13. Assume that the pressure is fixed ($P = 101.3$ kPa). The independent variables chosen for the rigorous design of the conventional distillation columns are the number of trays in the rectifying section, number of trays in the stripping section, and the recoveries of key components for each column.

The model can be written as follows:

The objective function consists of minimizing the TAC. It is composed of the operational and fixed costs of both columns. Mixers and splitters costs are neglected

$$\min : \ \text{TAC} = \text{TAC}_{\text{Column1}} + \text{TAC}_{\text{Column2}} \quad (15)$$

Mass balance in the inlet (feed) splitter. The index "i" refers to components

$$\left. \begin{array}{l} \text{Feed}_i = F1_i + F2_i + F3_i + F4_i \\[2mm] F1_i = \left(\sum_i F1_i \right) z_i \\[2mm] F2_i = \left(\sum_i F2_i \right) z_i \\[2mm] F3_i = \left(\sum_i F3_i \right) z_i \\[2mm] F4_i = \left(\sum_i F4_i \right) z_i \end{array} \right\} \text{Feed splitter} \quad (16)$$

In previous equations "z" refers to the feed molar fraction, which is known and therefore all those equations are linear.

Mass balance in mixers 1, 2, and final products $P1$ and $P2$

$$F1_i + F15_i = F5_i \} \text{Mixer 1}$$

$$F2_i + F10_i = F13_i \} \quad \text{Mixer 2} \quad (17)$$

$$F7_i + F11_i + F16_i + F19_i + F3_i = P1_i$$

$$F8_i + F12_i + F17_i + F20_i + F4_i = P2_i$$

Mass balances in splitters 1, 2, 3, and 4

$$\left. \begin{array}{l} F7_i = \alpha_1^1 F6_i \\[1mm] F8_i = \alpha_2^1 F6_i \\[1mm] \alpha_1^1 + \alpha_2^1 = 1 \end{array} \right\} \text{splitter 1}$$

$$\left. \begin{array}{l} F10_i = \alpha_1^2 F9_i \\[1mm] F11_i = \alpha_2^2 F9_i \\[1mm] F12_i = \alpha_3^2 F9_i \\[1mm] \alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1 \end{array} \right\} \text{splitter 2}$$

$$\left. \begin{array}{l} F15_i = \alpha_1^3 F14_i \\[1mm] F16_i = \alpha_2^3 F14_i \\[1mm] F17_i = \alpha_3^3 F14_i \\[1mm] \alpha_1^3 + \alpha_2^3 + \alpha_3^3 = 1 \end{array} \right\} \text{splitter 3} \qquad (18)$$

$$\left. \begin{array}{l} F19_i = \alpha_1^4 F18_i \\[1mm] F20_i = \alpha_2^4 F18_i \\[1mm] \alpha_1^4 + \alpha_2^4 = 1 \end{array} \right\} \text{splitter 4}$$

Disjunctions related with the existence or not of a given distillation column

$$\begin{bmatrix} Y_1 \\ \left[Q_{\text{Reb}}^{\text{C1}}, Q_{\text{Cond}}^{\text{C1}}, D^{\text{C1}}\right] = \text{Kr}_1\left(N_{\text{R1}}^{\text{C1}}, N_{\text{S1}}^{\text{C1}}, f_{\text{LK}}^{\text{C1}}, f_{\text{HK}}^{\text{C1}}, F5_i\right)_1 \\ \text{TAC}_1 = \Omega\left(Q_{\text{Reb}}^{\text{C1}}, Q_{\text{Cond}}^{\text{C1}}, D^{\text{C1}}\right) \end{bmatrix} \veebar \begin{bmatrix} Y_1 \\ \text{TAC}_1 = 0 \\ F5_i = F6_i = F9_i = 0 \ \forall i \end{bmatrix}$$

$$\begin{bmatrix} Y_2 \\ \left[Q_{\text{Reb}}^{\text{C2}}, Q_{\text{Cond}}^{\text{C2}}, D^{\text{C2}}\right] = \text{Kr}_2\left(N_{\text{R1}}^{\text{C2}}, N_{\text{S2}}^{\text{C2}}, f_{\text{LK}}^{\text{C2}}, f_{\text{HK}}^{\text{C2}}, F13_i\right)_2 \\ \text{TAC}_2 = \Omega\left(Q_{\text{Reb}}^{\text{C2}}, Q_{\text{Cond}}^{\text{C2}}, D^{\text{C2}}\right) \end{bmatrix} \veebar \begin{bmatrix} Y_2 \\ \text{TAC}_2 = 0 \\ F13_i = F14_i = F18_i = 0 \ \forall i \end{bmatrix} \quad (19)$$

$$8 \leq N_{\text{R1}} \leq 35; \qquad 8 \leq N_{\text{S1}} \leq 35; \qquad [N_{\text{R1}}, N_{\text{S1}}] \in \mathbb{Z}$$

$$0.85 \leq f_{\text{LK1}} \leq 1.00; \qquad 0.85 \leq f_{\text{HK1}} \leq 1.00;$$

$$11.176 \leq F5_{\text{Benzene}} \leq 100; \qquad 9.5 \leq F5_{\text{Toluene}} \leq 150; \qquad 0 \leq F5_{\text{p-Xylene}} \leq 117,65;$$

$$8 \leq N_{\text{R2}} \leq 35; \qquad 8 \leq N_{\text{S2}} \leq 35; \qquad [N_{\text{R1}}, N_{\text{S1}}] \in \mathbb{Z}$$

$$0.85 \leq f_{\text{LK2}} \leq 1.00; \qquad 0.85 \leq f_{\text{HK2}} \leq 1.00;$$

$$1.676 \leq F13_{\text{Benzene}} \leq 117.65; \qquad 9.5 \leq F13_{\text{Toluene}} \leq 150; \qquad 11.176 \leq F13_{\text{p-Xylene}} \leq 117.65;$$

$Y1$ and $Y2$ are Boolean variables that take the value of True if the corresponding column is selected and False otherwise. The symbol "f" refers to the key components recoveries.

Product specifications

$$P1_{\text{Benzene}} = 30 \, \text{kmol/h}; \quad P1_{\text{Toluene}} = 50 \, \text{kmol/h};$$

$$P1_{\text{p-Xylene}} = 30 \, \text{kmol/h}$$

$$P2_{\text{Benzene}} = 70 \, \text{kmol/h}; \quad P2_{\text{Toluene}} = 50 \, \text{kmol/h};$$

$$P2_{\text{p-Xylene}} = 70 \, \text{kmol/h}$$

The problem can be solved as an MINLP by reformulating the disjunctions using a Big-M approach.[85]

Figure 8 shows the optimal solution of the system studied.

The minimum objective function (TAC = 0.3137 M$/year) is obtained with the sequence shown (Figure 8). Column 1 has 26 trays and it is fed on the 14th tray. Column 2 has 31 trays and it is fed on the 15th tray.

The costs of the optimal solution are shown in Table 14.

A summary of results obtained for example 4 is shown in Table 15. In this case the initial Kriging is not accurate enough to ensure that the optimal solution obtained by the Kriging matches the actual model. Therefore, we continue with the algorithm previously described by adding to the Kriging metamodel the optimal point obtained in the previous stage, and repeat this procedure until there is no improvement. As noted before, this point is not necessarily the optimum. It is necessary to contract the dominion around the best obtained solution, resampling and recali-

brating the Kriging. We performed a contraction step around the optimal solution (10% around the optimal point). The contraction step confirmed that we are in an optimal solution.

A summary for all the streams obtained is shown in Table 16.

Example 5. Demethanizer column. In this example a demethanizer column is optimized. The process to recovery the valuable heavier hydrocarbons from natural gas are carried out in a series of distillation columns. The first column of the process is the demethanizer column. This is a cryogenic high-pressure column in which the methane is separated overhead and the ethane and the others heavier hydrocarbons are taken as a bottom (NGL). The demethanizer flow sheet is proposed by Luyben.[86]

To save energy the feed is precooled, before entering the column, by a complex heat exchanger network that includes external heat exchangers and two trays in the distillation column (See Figure 9). A fraction of the feed is condensed and used as reflux load in the column. The design of the heat exchanger network is out of the scope of this article so we maintain the topology of the heat exchanger network and the distillation column and optimize the operating conditions.

We want to minimize the TAC of a demethanizer column for separating a mixture of methane, ethane, propane and i-butane (molar fractions: 0.783, 0.134, 0.056, and 0.027, respectively) to recover at least 99.8% of the methane fed with at least 0.96 in molar fraction. This includes the cold and hot utility costs that are the main costs in the system and the investment cost. Data for this example are shown in Table 17.

Assume that the pressure in the column varies between 2533 kPa (reboiler) and 2543 kPa (condenser). In this example we suppose that the feed enters into the column in the trays specified by Luyben.[86] The objective in this case is not to optimize the feed tray position, but we want to determine the optimal heat exchanged to precool the feed and the flux that is sent overhead. The model we want to solve can be conceptually written as follows

Table 14. Minimun Cost for Example 4

	Optimal Solution (Distillation Sequence)		
	Column 1	Column 2	Total
TAC (M$/year)	0.1493	0.1634	0.3127
Operating Cost (M$/year)	0.0888	0.1000	0.1888
Capital Cost (M$/year)	0.0605	0.0634	0.1239

Table 15. Summary of Results Obtained for Example 4

Iteration 1
Number of sampled points, Column 1 = 200
  Sampling time = 32.61 s
  CPU optimization Kriging parameters = 22.47 s
TAC = 0.2592 M$/year

Number of sampled points, Column 2 = 200
Sampling time = 33.11 s
CPU optimization Kriging parameters = 22.96 s

Optimal values

$F5_{,benzene}$ (kmol/h)	13.1490	$F13_{,benzene}$ (kmol/h)	4.3396
$F5_{,toluene}$ (kmol/h)	11.4914	$F13_{,toluene}$ (kmol/h)	9.5000
$F5_{,p\text{-}Xyene}$ (kmol/h)	3.1538	$F13_{,p\text{-}Xyene}$ (kmol/h)	11.1760
Number of trays in the rectifying section, Column 1	17	Number of trays in the rectifying section, Column 2	14
Number of trays in the stripping section, Column 1	17	Number of trays in the stripping section, Column 2	16
$Recovery_{(LK, Column 1)}$	0.8500	$Recovery_{(LK, Column 2)}$	0.8740
$Recovery_{(HK, Column 1)}$	0.8500	$Recovery_{(HK, Column 2)}$	0.8500

Iteration 2
Number of sampled points, Column 1 = 201
TAC = 0.3002 M$/year

Number of sampled points, Column 2 = 201

Optimal values

$F5_{,benzene}$ (kmol/h)	13.1490	$F13_{,benzene}$ (kmol/h)	4.6010
$F5_{,toluene}$ (kmol/h)	11.5476	$F13_{,toluene}$ (kmol/h)	10.0828
$F5_{,p\text{-}Xyene}$ (kmol/h)	2.8603	$F13_{,p\text{-}Xyene}$ (kmol/h)	11.1760
Number of trays in the rectifying section, Column 1	17	Number of trays in the rectifying section, Column 2	15
Number of trays in the stripping section, Column 1	17	Number of trays in the stripping section, Column 2	18
$Recovery_{(LK, Column 1)}$	0.8500	$Recovery_{(LK, Column 2)}$	0.8740
$Recovery_{(HK, Column 1)}$	0.8500	$Recovery_{(HK, Column 2)}$	0.8509

Iteration 3
Number of sampled points, Column 1 = 202
TAC = 0.2640 M$/year

Number of sampled points, Column 2 = 202

Optimal values

$F5_{,benzene}$ (kmol/h)	13.1490	$F13_{,benzene}$ (kmol/h)	1.7978
$F5_{,toluene}$ (kmol/h)	12.0738	$F13_{,toluene}$ (kmol/h)	9.5000
$F5_{,p\text{-}Xyene}$ (kmol/h)	2.3910	$F13_{,p\text{-}Xyene}$ (kmol/h)	11.1760
Number of trays in the rectifying section, Column 1	17	Number of trays in the rectifying section, Column 2	8
Number of trays in the stripping section, Column 1	17	Number of trays in the stripping section, Column 2	20
$Recovery_{(LK, Column 1)}$	0.8500	$Recovery_{(LK, Column 2)}$	0.8740
$Recovery_{(HK, Column 1)}$	0.8500	$Recovery_{(HK, Column 2)}$	0.8500

Iteration 4
Number of sampled points, Column 1 = 203
TAC = 0.2790 M$/year

Number of sampled points, Column 2 = 203

Optimal values

$F5_{,benzene}$ (kmol/h)	13.1490	$F13_{,benzene}$ (kmol/h)	1.9723
$F5_{,toluene}$ (kmol/h)	11.1760	$F13_{,toluene}$ (kmol/h)	9.5621
$F5_{,p\text{-}Xyene}$ (kmol/h)	2.8603	$F13_{,p\text{-}Xyene}$ (kmol/h)	11.1760
Number of trays in the rectifying section, Column 1	17	Number of trays in the rectifying section, Column 2	8
Number of trays in the stripping section, Column 1	17	Number of trays in the stripping section, Column 2	17
$Recovery_{(LK, Column 1)}$	0.8500	$Recovery_{(LK, Column 2)}$	0.8500
$Recovery_{(HK, Column 1)}$	0.8500	$Recovery_{(HK, Column 2)}$	0.8500

Iteration 5
Number of sampled points, Column 1 = 204
TAC = 0.3245 M$/year

Number of sampled points, Column 2 = 204

Optimal values

$F5_{,benzene}$ (kmol/h)	13.1490	$F13_{,benzene}$ (kmol/h)	4.5828
$F5_{,toluene}$ (kmol/h)	12.4471	$F13_{,toluene}$ (kmol/h)	10.6642
$F5_{,p\text{-}Xyene}$ (kmol/h)	2.5016	$F13_{,p\text{-}Xyene}$ (kmol/h)	11.1760
Number of trays in the rectifying section, Column 1	18	Number of trays in the rectifying section, Column 2	16
Number of trays in the stripping section, Column 1	19	Number of trays in the stripping section, Column 2	18
$Recovery_{(LK, Column 1)}$	0.8500	$Recovery_{(LK, Column 2)}$	0.8775
$Recovery_{(HK, Column 1)}$	0.8500	$Recovery_{(HK, Column 2)}$	0.8760

Iteration 6
Number of sampled points, Column 1 = 205
TAC = 0.3195 M$/year

Number of sampled points, Column 2 = 205

	Optimal values		
$F5_{,benzene}$ (kmol/h)	13.1490	$F13_{,benzene}$ (kmol/h)	4.8304
$F5_{,toluene}$ (kmol/h)	13.1271	$F13_{,toluene}$ (kmol/h)	11.2046
$F5_{,p-Xyene}$ (kmol/h)	2.4929	$F13_{,p-Xyene}$ (kmol/h)	11.1760
Number of trays in the rectifying section, Column 1	18	Number of trays in the rectifying section, Column 2	16
Number of trays in the stripping section, Column 1	19	Number of trays in the stripping section, Column 2	18
Recovery$_{(LK, Column 1)}$	0.8500	Recovery$_{(LK, Column 2)}$	0.8740
Recovery$_{(HK, Column 1)}$	0.8500	Recovery$_{(HK, Column 2)}$	0.8500

Iteration 7
Number of sampled points, Column 1 = 200
Sampling time = 26.40 s
CPU optimization Kriging parameters = 19.53 s
TAC = 0.3064 M$/year

Number of sampled points, Column 2 = 200
Sampling time = 26.35 s
CPU optimization Kriging parameters = 50.91 s

	Optimal values (after refining stage)		
$F5_{,benzene}$ (kmol/h)	13.1490	$F13_{,benzene}$ (kmol/h)	3.6504
$F5_{,toluene}$ (kmol/h)	11.1760	$F13_{,toluene}$ (kmol/h)	9.5000
$F5_{,p-Xyene}$ (kmol/h)	3.4128	$F13_{,p-Xyene}$ (kmol/h)	12.6595
Number of trays in the rectifying section, Column 1	12	Number of trays in the rectifying section, Column 2	12
Number of trays in the stripping section, Column 1	12	Number of trays in the stripping section, Column 2	16
Recovery$_{(LK, Column 1)}$	0.8740	Recovery$_{(LK, Column 2)}$	0.8903
Recovery$_{(HK, Column 1)}$	0.8505	Recovery$_{(HK, Column 2)}$	0.8500

Iteration 8
Number of sampled points, Column 1 = 201
TAC = 0.3237 M$/year

Number of sampled points, Column 2 = 201

	Optimal values (after refining stage)		
$F5_{,benzene}$ (kmol/h)	13.1490	$F13_{,benzene}$ (kmol/h)	3.6504
$F5_{,toluene}$ (kmol/h)	11.2809	$F13_{,toluene}$ (kmol/h)	9.5000
$F5_{,p-Xyene}$ (kmol/h)	3.4167	$F13_{,p-Xyene}$ (kmol/h)	12.6594
Number of trays in the rectifying section, Column 1	13	Number of trays in the rectifying section, Column 2	14
Number of trays in the stripping section, Column 1	12	Number of trays in the stripping section, Column 2	16
Recovery$_{(LK, Column 1)}$	0.8880	Recovery$_{(LK, Column 2)}$	0.8903
Recovery$_{(HK, Column 1)}$	0.8512	Recovery$_{(HK, Column 2)}$	0.8500

Iteration 9
Number of sampled points, Column 1 = 202
TAC = 0.3129 M$/year

Number of sampled points, Column 2 = 202

	Optimal values (after refining stage)		
$F5_{,benzene}$ (kmol/h)	13.1490	$F13_{,benzene}$ (kmol/h)	3.6504
$F5_{,toluene}$ (kmol/h)	11.2260	$F13_{,toluene}$ (kmol/h)	9.5000
$F5_{,p-Xyene}$ (kmol/h)	3.4158	$F13_{,p-Xyene}$ (kmol/h)	12.6589
Number of trays in the rectifying section, Column 1	13	Number of trays in the rectifying section, Column 2	14
Number of trays in the stripping section, Column 1	12	Number of trays in the stripping section, Column 2	16
Recovery$_{(LK, Column 1)}$	0.8878	Recovery$_{(LK, Column 2)}$	0.8903
Recovery$_{(HK, Column 1)}$	0.8513	Recovery$_{(HK, Column 2)}$	0.8500

$$\min : f = \text{TAC}$$
$$\text{s.t. } [D, Q_{cooler}, \text{TAC}] = \text{Kr}(\text{MF}, Q_1, Q_2)$$
$$1100 \leq \text{MF} \leq 1600 \quad\quad (20)$$
$$1300 \leq Q_1 \leq 2000$$
$$1100 \leq Q_2 \leq 1600$$

where D is the diameter of the column, Q_{cooler} is the heat flow removed by the refrigerant and *TAC* is the TAC. The function Kr makes reference to the Kriging interpolators. In this example, the degrees of freedom used are the following: MF makes reference to the flow that is sent overhead, Q_1 and Q_2 are the heat flows used to precool the fed.

We calibrate the Kriging parameters with the values of dependent variables for each simulation and then, we replace the demethanizer column by the surrogate model. The column is optimized using CONOPT.[55]

The solution to this problem is $x = [1310.863, 1927.165, 1585.490]$ (Molar flow sent overhead, Q_1 and Q_2, respectively) and the objective function is 0.9269 M$/year. In Table 18, the TAC is summarized.

The results obtained with the Kriging metamodel and the simulator are shown in Table 19.

The total CPU time used in the optimization (including sampling, Kriging optimization, and model optimization) was around 110 s. The results obtained with the Kriging metamodel have an error below 5% in comparison to the data obtained with the simulator. Although we have an error, as in other examples, we can expect a good result. Figure 9 shows the optimal solution of the demethanizer column.

Conclusions

In this work, an algorithm for constrained optimization including continuous and discrete values was presented and

Table 16. Summary of Optimal Solution Obtained for Example 4

	Optimal Values (kmol/h)		
$F1_{,benzene}$	9.4986	$F11_{,benzene}$	0.0005
$F1_{,toluene}$	2.7683	$F11_{,toluene}$	0.0031
$F1_{,p-Xyene}$	1.5170	$F11_{,p-Xyene}$	0.0011
$F2_{,benzene}$	2.1833	$F12_{,benzene}$	0.0083
$F2_{,toluene}$	0.0002	$F12_{,toluene}$	0.0537
$F2_{,p-Xyene}$	9.2634	$F12_{,p-Xyene}$	0.0192
$F3_{,benzene}$	24.1064	$F13_{,benzene}$	3.6504
$F3_{,toluene}$	48.8044	$F13_{,toluene}$	9.5000
$F3_{,p-Xyene}$	24.4424	$F13_{,p-Xyene}$	12.6589
$F4_{,benzene}$	64.2117	$F14_{,benzene}$	3.6504
$F4_{,toluene}$	48.4271	$F14_{,toluene}$	8.4577
$F4_{,p-Xyene}$	64.7772	$F14_{,p-Xyene}$	1.8988
$F5_{,benzene}$	13.1490	$F15_{,benzene}$	3.6504
$F5_{,toluene}$	11.2260	$F15_{,toluene}$	8.4577
$F5_{,p-Xyene}$	3.4158	$F15_{,p-Xyene}$	1.8988
$F6_{,benzene}$	11.6732	$F16_{,benzene}$	0.0000
$F6_{,toluene}$	1.6694	$F16_{,toluene}$	0.0000
$F6_{,p-Xyene}$	0.0000	$F16_{,p-Xyene}$	0.0000
$F7_{,benzene}$	6.0931	$F17_{,benzene}$	0.0000
$F7_{,toluene}$	0.8714	$F17_{,toluene}$	0.0000
$F7_{,p-Xyene}$	0.0000	$F17_{,p-Xyene}$	0.0000
$F8_{,benzene}$	5.5800	$F18_{,benzene}$	1.0423
$F8_{,toluene}$	0.7980	$F18_{,toluene}$	10.7601
$F8_{,p-Xyene}$	0.0000	$F18_{,p-Xyene}$	0.0000
$F9_{,benzene}$	1.4758	$F19_{,benzene}$	0.5212
$F9_{,toluene}$	9.5566	$F19_{,toluene}$	5.3800
$F9_{,p-Xyene}$	3.4158	$F19_{,p-Xyene}$	0.0000
$F10_{,benzene}$	1.4671	$F20_{,benzene}$	0.5212
$F10_{,toluene}$	9.4998	$F20_{,toluene}$	5.3800
$F10_{,p-Xyene}$	3.3955	$F20_{,p-Xyene}$	0.0000

Table 17. Data for Example 5

Demethanizer Column	
Feed	5269 kmol/h
Composition (mole fraction)	
methane	0.783
ethane	0.134
propane	0.056
i-butane	0.027
Pressure	5978,175 kPa
Thermodynamics (fluid package)	Peng-Robinson

Table 18. Minimum Cost for Example 5

Optimal Solution (Demethanizer Column)	
TAC (M$/year)	0.9269
Operating Cost (M$/year)	0.8387
Capital Cost (M$/year)	0.0882

specifically applied to the rigorous design of distillation columns or distillation sequences. This optimization contains implicit black box functions which are substituted by Kriging metamodels. The surrogate model presented has proved to be robust, reliable and also allows a fast interpolation of new values.

It is interesting to remark that even though the final model includes discrete variables (number of trays in a given section) the Kriging surrogate can be generated without taking this into account. Of course we must ensure that the sampling points always correspond with integer values. But we
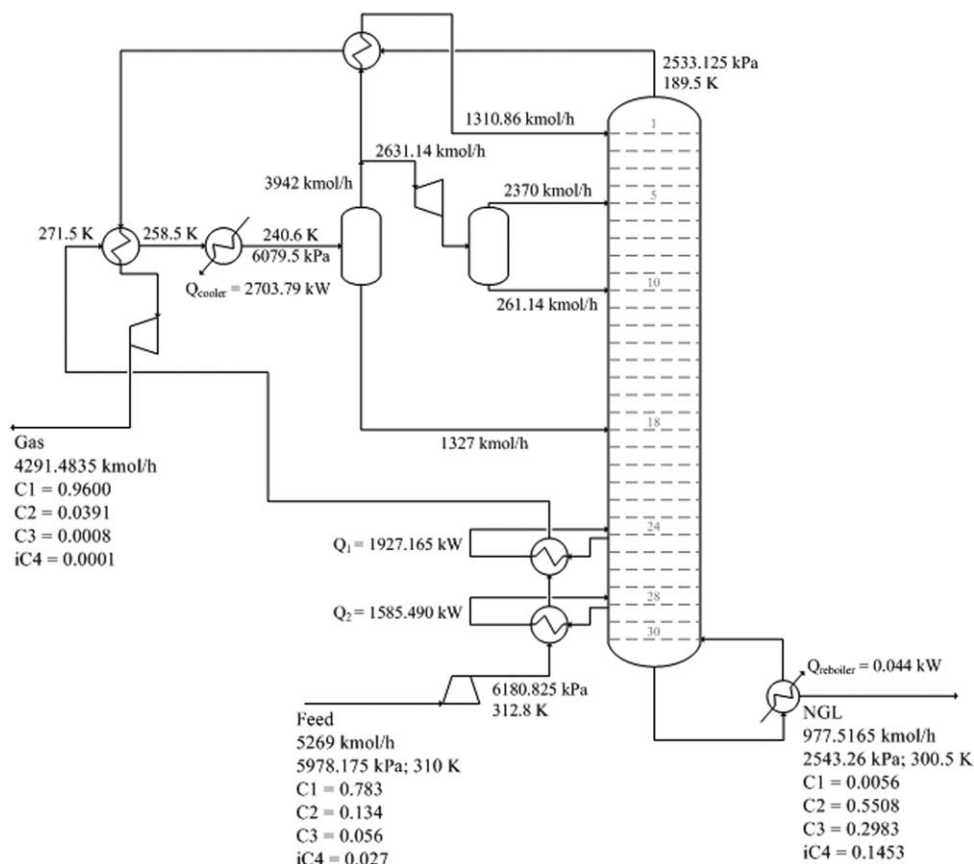

Figure 9. Optimal solution for example 5.

Table 19. Summary of Optimal Solution Obtained for Example 5

	Cooler Heat Flow (kW)	Reboiler Heat Flow (kW)	Distillate (kmol/h)	Bottom (kmol/h)	Diameter Column (m)	TAC (M$/year)
KRIGING	2703.79	0.044	4291.48	977.52	1.8954	0.9269
HYSYS	2857.19	0.045	4314.54	954.46	1.8966	0.9745

Number of sampled points = 78.
CPU time sampling = 84.36 s.
CPU optimization Kriging parameters = 19.63 s.
CPU optimization model = 6.99 s.

have checked that this is not a major problem. The possible correlation that is included in the sampling due to forcing some of the variables to only take integer values has no numerical effects, maybe because in the final optimization these variables must also be integer.

The different examples, with increasing difficulty, showed that a global approach is usually good enough and further contraction steps simply confirm that we got an optimal solution. But it is worth pointing out that we can only ensure a critical point if either the surrogate model accurately reproduces the model in all the points of the domain or we use contraction step(s) within a trust region.

A critical issue is the stopping criterion. The ideal situation would be to be able to evaluate the derivatives, from the actual model, in the optimal point obtained from the Kriging interpolator and test the Karush–Kuhn–Tucker optimality conditions. However, this is not possible in noisy systems. Alternatively, the nonimproving criterion in two successive iterations with contraction of the feasible domain in a trust region, with the objective of obtaining a set of sampling points that accurately reproduce both the actual function and its derivative could also eventually guarantee a local optimum. But the contraction and separation of sampling points is limited by the noise of the system, that is, sampling points must be separated enough to avoid adjusting for noise.

The implementation presented in this article only guarantees a local optimum. However, for the particular case of a single distillation column, where the surrogate models accurately reproduce the actual column model and the number of sampling points is relatively small –the number of basis functions is not too large- it is possible to explicitly include the equations in a deterministic global optimizer (we used GAMS-BARON) to obtain the global optimal solution. It is worth remarking that we are obtaining the global solution of the surrogate model and not of the actual model. The "quality" of that optimum will depend on how accurately the surrogate model represents the actual model. The location of the global optimum in complex flow sheets using surrogate models is under current research.

In the case of distillation columns and column sequences of different complexity we have checked that it is possible to get accurate global surrogate models with up to 7 degrees of freedom, at the expense of increasing the number of sampling points. There is a tradeoff between the number of sampling points and the iterations in the optimization algorithm, that is, with large sets of sampling points it could be possible to get an accurate model that globally reproduces the actual model, but at the price of large CPU times for calibrating the Kriging. A reduced set of sampling points reduces the CPU time for calibrating the Kriging and interpolation, but it is likely that we would need more contraction and resampling (and recalibrating) stages.

Literature Cited

1. Papalambros PY, Wilde DJ. *Principles of Optimal Design: Modeling and Computation*, 2nd ed. New York: Cambridge University Press, 2000:390.
2. Chung PS, Jhon MS, Biegler LT. The Holistic Strategy in Multi-Scale Modeling, *Advances in Chemical Engineering*. 2011;40:59–118.
3. Biegler LT, Lang Y, Lin W. Multi-Scale Optimization for Process Systems Engineering. *Comput Chem Eng*. 2014;60(10):17–30.
4. Lucia A. Multi-scale methods and complex processes: a survey and look ahead. In: *Proceedings of the Seventh International Conference on the Foundations of Computer-aided Process Design*. Breckenridge, CO: CRC Press, 2009.
5. Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Kevin Tucker P. Surrogate-based analysis and optimization. *Prog Aerospace Sci*. 2005;41(1):1–28.
6. Wang G, Shan S. Review of Metamodeling Techniques in Support of Engineering Design Optimization. *J Mech Des*. 2006. 129: p. 370–380.
7. Caballero JA, Grossmann IE. An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE J*. 2008;54(10): 2633–2650.
8. Henao CA, Maravelias CT. Surrogate-based superstructure optimization framework. *AIChE J*. 2011;57(5): 1216–1232.
9. Davis E, Ierapetritou M. A kriging method for the solution of nonlinear programs with black-box functions. *AIChE J*. 2007;53(8): 2001–2012.
10. Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program*. 2006;106(1): 25–57.
11. Chen W, Shao Z, Qian J. Interfacing IPOPT with Aspen open solvers and CAPE-OPEN, In: *International Symposium on Process Systems Engineering*. Amsterdam, The Netherlands: Elsevier. 2009:201–206.
12. Caballero JA, Grossmann IE. Rigorous flowsheet optimization using process simulators and surrogate models. In: Bertrand B, Xavier J, editors. *Computer Aided Chemical Engineering*. Amsterdam, The Netherlands: Elsevier, 2008:551–556.
13. Rios L, Sahinidis N. Derivative-free optimization: a review of algorithms and comparison of software implementations. *J Glob Optim*. 2013;56(3): 1247–1293.
14. Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. *J Glob Optim*. 1998;13(4): 455–492.
15. Cozad A, Sahinidis NV, Miller DC. Learning surrogate models for simulation-based optimization. *AIChE J*. 2014;60(6): 2211–2227.
16. Palmer K, Realff M. Metamodeling approach to optimization of steady-state flowsheet simulations: model generation. *Chem Eng Res Des*. 2002;80(7): 760–772.
17. Draper NR, Smith H. *Applied Regression Analysis*. New York: Wiley, 1981.
18. Friedman JH. Multivariate adaptive regression splines. *Ann Stat*. 1991;19(1):1–141.
19. Himmelblau DM. Applications of artificial neural networks in chemical engineering. *Korean J Chem Eng*. 2000;17(4):373–392.
20. Akaike H. A new look at the statistical model identification. IEEE Trans Automatic Control 1974 (reprinted 2003);19(6):716–723.

21. Kleijnen JPC. Kriging metamodeling in simulation: a review. *Eur J Oper Res.* 2009;192(3):707–716.

22. Huang D, Allen TT, Notz WI, Zeng N. Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models. *J Glob Optim.* 2006;34(3):441–466.

23. Sargent RWH, Gaminibandara K. Optimum design of plate distillation columns. In: Dixon LWC, Editor. *Optimization in Action*. London: Academic Press, 1976:267.

24. Viswanathan J, Grossmann IE. Optimal feed locations and number of trays for distillation-columns with multiple feeds. *Ind Eng Chem Res.* 1993;32(11):2942–2949.

25. Viswanathan J, Grossmann IE. An alternate MINLP model for finding the number of trays required for a specified separation objective. *Comput Chem Eng.* 1993;17(9):949–955.

26. Yeomans H, Grossmann IE. Optimal design of complex distillation columns using rigorous tray-by-tray disjunctive programming models. *Ind Eng Chem Res.* 2000;39(11):4326–4335.

27. Yeomans H, Grossmann IE. Disjunctive programming models for the optimal design of distillation columns and separation sequences. *Ind Eng Chem Res.* 2000;39(6):1637–1648.

28. Barttfeld M, Aguirre PA, Grossmann IE. Alternative representations and formulations for the economic optimization of multicomponent distillation columns. *Comput Chem Eng.* 2003;27(3):363–383.

29. Barttfeld M, Aguirre PA. Optimal synthesis of multicomponent zeotropic distillation processes. 1. Preprocessing phase and rigorous optimization for a single unit. *Ind Eng Chem Res.* 2002;41(21):5298–5307.

30. Barttfeld M, Aguirre PA. Optimal synthesis of multicomponent zeotropic distillation processes. 2. Preprocessing phase and rigorous optimization of efficient sequences. *Ind Eng Chem Res.* 2003;42(14):3441–3457.

31. Kraemer K, Kossack S, Marquardt W. An efficient solution method for the MINLP optimization of chemical processes. *Comput Aided Chem Eng.* 2007;24:105–110.

32. Harwardt A, Marquardt W. Heat-integrated distillation columns: vapor recompression or internal heat integration? *AIChE J.* 2012;58(12):3740–3750.

33. Dunnebier G, Pantelides CC. Optimal design of thermally coupled distillation columns. *Ind Eng Chem Res.* 1999;38(1):162–176.

34. Bauer MH, Stichlmair J. Superstructures for the mixed integer optimization of nonideal and azeotropic distillation processes. *Comput Chem Eng.* 1996;20:S25–S30.

35. Bauer MH, Stichlmair J. Design and economic optimization of azeotropic distillation processes using mixed-integer nonlinear programming. *Comput Chem Eng.* 1998;22(9):1271–1286.

36. Barttfield M, Aguirre PA, Grossmann IE. A decomposition method for synthesizing complex column configurations using tray-by-tray GDP models. *Comput Chem Eng.* 2004;28(11):2165–2188.

37. Ciric AR, Gu D. Synthesis of nonequilibrium reactive distillation processes by MINLP optimization. *AIChE J.* 1994;40(9):1479–1487.

38. Jackson JR, Grossmann IE. A disjunctive programming approach for the optimal design of reactive distillation columns. *Comput Chem Eng.* 2001;25(11–12):1661–1673.

39. Lang YD, Biegler LT. Distributed stream method for tray optimization. *AIChE J.* 2002;48(3):582–595.

40. Caballero JA, Milán-Yañez D, Grossmann IE. Rigorous design of distillation columns: integration of disjunctive programming and process simulators. *Ind Eng Chem Res.* 2005;44(17):6760–6775.

41. Caballero JA. Logic hybrid simulation-optimization algorithm for distillation design. *Compu Chem Engi.* 2015;72:284–299.

42. Krige DG. *A Statistical Approach to Some Mine Valuations and Allied Problems at the Witwatersrand*. South Africa: University of Witwatersrand, 1951.

43. Sacks J, Welch WJ, Toby JM, Wynn HP. Design and analysis of computer experiments. *Stat Sci.* 1989;4(4):409–423.

44. Sasena MJ. *Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations*. University of Michigan: Michigan, 2002:237.

45. Jones DR. A taxonomy of global optimization methods based on response surfaces. *J Glob Optim.* 2001;21(4):345–383.

46. Alexandrov NM, et al. *Optimization with variable-fidelity models applied to wing design in 38th Aerospaces Sciences Meeting & Exibit*. AIAA Paper 2000-0841. 2000.

47. Diwekar U. Introduction to applied optimization. In: Pardalos PM, Hearn DW, editors. *Applied Optimization*. Dordrecht: Kluwer Academic Publishers. 2003:335.

48. Iman RL, Helton JC, Camobell JE. An approach to sensitivity analysis of computer models, Part 1. Introduction, input variable selection and preliminary variable assestment. *J Quality Technol.* 1981;13(3):174–183.

49. McKay MD, Conover WJ, Beckman RJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 1979;21(2):239–245.

50. Kalagnanan JR, Diwekar U. An efficient sampling technique for offline quality control. *Technometrics.* 1997;21(2):239.

51. Press WH, et al. *Numerical Recipies in Fortran*. Cambridge: Cambridge University Press, 1992.

52. Turton R, et al. *Analysis, Synthesis and Design of Chemical Processes*. 4th ed. Upper Saddle River, NJ, USA: Pearson Education, Inc., 2013.

53. Smith R. *Chemical Process Design and Integration*. Chichester: Wiley, 2005:687.

54. Hyprotech Ltd. *HYSYS*. Hyprotech Ltd, 1995–2002.

55. Drud AS. *CONOPT: A System for Large Scale Nonlinear Optimization. Reference Manual*. Bagsvaerd, Denmark: ARKI Consulting and Development A/S, 1996:69.

56. Gill WPE, Murray W, Saunders MA. SNOPT: an SQP algorithm for large-scale constrained optimization. *SIAM J Optim.* 2002;47(1):979–1006.

57. Holmström K. The Tomlab Optimization Environment in Matlab. *Adv Model Optim.* 1999;1:47–69.

58. The Mathworks, Inc. *Matlab 8.3*. Natick, MA: The Mathworks, Inc. null, 2014.

59. Sahinidis NV. BARON: A general purpose global optimization software package. *J Glob Optim.* 1996;8(2):201–205.

60. Rosenthal RE. *GAMS—A User's Guide*. Washington, DC: GAMS Development Corporation, 2012.

61. Agrawal R, Fidkowski ZT. More operable arrangements of fully thermally coupled distillation columns. *AIChE J.* 1998;44(11):2565–2568.

62. Fidkowski ZT, Agrawal R. Multicomponent thermally coupled systems of distillation columns at minimum reflux. *AIChE J.* 2001;47(12):2713–2724.

63. Halvorsen IJ, Skogestad S. Minimum energy consumption in multicomponent distillation. 2. Three-product Petlyuk arrangements. *Ind Eng Chem Res.* 2003;42(3):605–615.

64. Halvorsen IJ, Skogestad S. Minimum energy consumption in multicomponent distillation. 3. More than three products and generalized Petlyuk arrangements. *Ind Eng Chem Res.* 2003;42(3):616–629.

65. Wolff EA, Skogestad S. Operation of integrated 3-product (petlyuk) distillation-columns. *Ind Eng Chem Res.* 1995;34(6):2094–2103.

66. Petlyuk FB, Platonov VM, Slavinsk DM. Thermodynamically optimal method for separating multicomponent mixtures. *Int Chem Eng.* 1965;5(3):555.

67. Agrawal R. Synthesis of distillation column configurations for a multicomponent separation. *Ind Eng Chem Res.* 1996;35(4):1059–1071.

68. Giridhar A, Agrawal R. Synthesis of distillation configurations: I. Characteristics of a good search space. *Comput Chem Eng.* 2010;34(1):73.

69. Shah VH, Agrawal R. A matrix method for multicomponent distillation sequences. *AIChE J.* 2010;56(7):1759–1775.

70. Caballero JA, Grossmann IE. Generalized disjunctive programming model for the optimal synthesis of thermally linked distillation columns. *Ind Eng Chem Res.* 2001;40(10):2260–2274.

71. Caballero JA, Grossmann IE. Design of distillation sequences: from conventional to fully thermally coupled distillation systems. *Comput Chem Eng.* 2004;28(11):2307–2329.

72. Caballero JA, Grossmann IE. Logic-sequential approach to the synthesis of complex thermally coupled distillation systems. In: Pistikopoulos MCGEN, Kokossis AC, editors. *Computer Aided Chemical Engineering*. Elsevier, 2011:211–215.

73. Caballero JA, Grossmann IE. Synthesis of complex thermally coupled distillation systems including divided wall columns. *AIChE J.* 2013;59(4):1139–1159.

74. Caballero JA, Grossmann IE. Optimal synthesis of thermally coupled distillation sequences using a novel MILP approach. *Comput Chem Eng.* 2014;61:118–135.

75. Carlberg NA, Westerberg AW. Temperature heat diagrams for complex columns .3. Underwoods method for the petlyuk configuration. *Ind Eng Chem Res.* 1989;28(9):1386–1397.

76. Navarro MA, et al. Strategies for the robust simulation of thermally coupled distillation sequences. *Comput Chem Eng.* 2012;36(1):149–159.

77. Fidkowski Z, Krolikowski L. Minimum energy-requirements of thermally coupled distillation systems. *AIChE J*. 1987;33(4):643–653.

78. Wright RO, Inventor. Fractionation Apparatus. US Patent 2471134 A, 1949.

79. Triantafyllou C, Smith R. The design and optimization of fully thermally coupled distillation-columns. *Chem Eng Res Des*. 1992;70(2):118–132.

80. Amminudin KA, et al. Design and optimization of fully thermally coupled distillation columns part 1: Preliminary design and optimization methodology. *Chem Eng Res Des*. 2001;79(A7):701–715.

81. Muralikrishna K, Madhavan KP, Shah SS. Development of dividing wall distillation column design space for a specified separation. *Chem Eng Res Des*. 2002;80(A2):155–166.

82. Kim YH. Rigorous design of extended fully thermally coupled distillation columns. *Chem Eng J*. 2002;89(1–3):89.

83. Luyben WL. Effect of solvent on controllability in extractive distillation. *Ind Eng Chem Res*. 2008;47(13):4425–4439.

84. Aggarwal A, Floudas CA. Synthesis of general distillation sequences-nonsharp separations. *Comput Chem Eng*. 1990;14(6):631–653.

85. Grossmann IE, Ruiz JP. Generalized disjunctive programming: a framework for formulation and alternative algorithms for MINLP optimization. In: Lee J, Leyffer S, editors. *Mixed Integer Nonlinear Programming*. New York: Springer, 2012:93–115.

86. Luyben WL. NGL demethanizer control. *Ind Eng Chem Res*. 2013; 52(33):11626–11638.