

# Surrogate Modeling of Thermodynamic Equilibria: Applications, Sampling and Optimization

Joschka Winz<sup>1,\*</sup>, Corina Nentwich<sup>2</sup>, and Sebastian Engell<sup>1</sup>

DOI: 10.1002/cite.202100092

 This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Models based on first principles are an effective way to model chemical processes. The quality of these depends critically on the accurate description of thermodynamic equilibria. This is provided by modern thermodynamic models, e.g., PC-SAFT, but they come with a high computational cost, which makes process optimization challenging. This can be addressed by using surrogate models to approximate the equilibrium calculations. A high accuracy of the surrogate model can be achieved by carefully choosing the points at which the original function is evaluated to create data for the training of the surrogate models, called sampling. Using a case study, different approaches to sampling are discussed and evaluated with a focus on new approaches to adaptive sampling.

**Keywords:** Machine learning, Process optimization, Surrogate modeling, Thermodynamic equilibria

*Received:* May 31, 2021; *revised:* July 13, 2021; *accepted:* September 06, 2021

## 1 Introduction

Industrial chemical processes can often be described accurately by models based on first principles. A core element of these models are thermodynamic models of phase equilibria. For complex mixtures, advanced thermodynamic models as the perturbed chain statistical associating fluid theory (PC-SAFT) equation of state [1] provide reliable predictions, but this comes at the expense of having to solve the thermodynamic models iteratively. Thus, the complexity of the system of equations that have to be solved by the simulator increases, especially if the thermodynamic model has to be solved many times, e.g., when simulating distillation or extraction columns [2]. This can make the convergence of large flowsheets challenging and lead to long computation times. A solution for this issue is the use of surrogate models. These are generic black-box models that can be used to approximate any functional relationship, given there is enough data available. This contribution deals with parameterizing surrogate models of complex thermodynamic models. Other applications of surrogate models have been described for models that contain partial differential equations [3] and for CFD models [4].

The main steps of applying surrogate models for the approximation of thermodynamic equilibrium calculations are shown in Fig. 1. There are several steps that must be performed to determine a suitable surrogate model for an application. The first important decision is how to choose the input and output variables of the surrogate model such that it can efficiently represent the rigorous equilibrium calculations. Possible options are described in Sect. 2 of this paper. After deciding on the structure of the surrogate model, the next step then is to parameterize or to train the

chosen model, which requires sufficient data. In the case considered here, this data will be generated by the known rigorous model, but as the model calls are computation-intensive the number of data points should be as small as possible. This points to the issue how to choose the data points that are used for training, which is also called sampling, such that not more samples than necessary need to be evaluated. Different sampling methods are discussed in Sect. 3 and 6.

The application of surrogate models for phase equilibria is discussed for the optimization of the process flowsheet for the hydroformylation of 1-dodecene in a thermomorphic solvent system. The case study is described in Sect. 4, and the results are presented in Sect. 5 and 6. Finally, in Sect. 7 the main aspects are summarized and an outlook to further research is given.

## 2 Input/Output Representation of Phase Equilibria for Surrogate Modeling

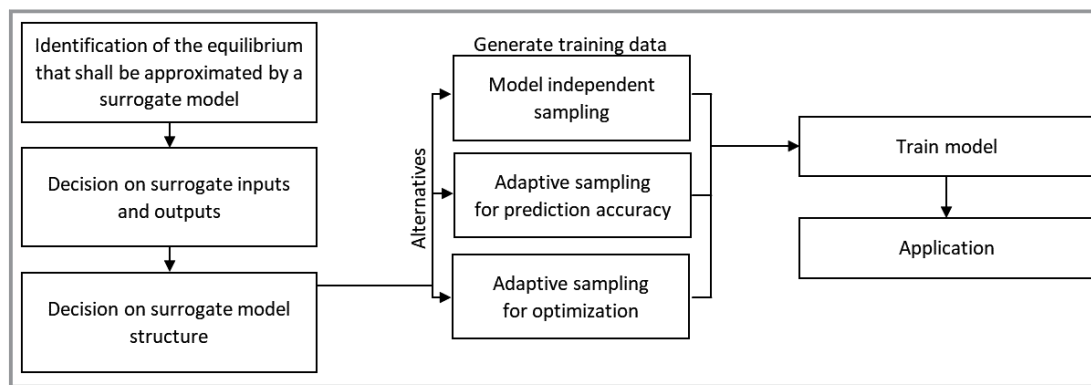
Representing a phase equilibrium in a process model employing surrogate models can, in principle, be done by

<sup>1</sup>Joschka Winz, Sebastian Engell  
joschka.winz@tu-dortmund.de

TU Dortmund, Process Dynamics and Operations Group, Department of Biochemical and Chemical, Emil-Figge-Straße 70, 44227 Dortmund, Germany.

<sup>2</sup>Corina Nentwich

Evonik Operations GmbH, Paul-Baumann-Straße 1, 45772 Marl, Germany.



**Figure 1.** Main steps of applying surrogate modeling for thermodynamic equilibrium computations.

approximating any element in the original equilibrium calculation. For example, the surrogate model can approximate the complete iterative equilibrium computation, i.e., compute the resulting phase compositions as a function of the feed properties. This is referred to as the direct method.

Since the number of phases may not be constant, a classification surrogate to determine the number of phases can be composed with a regression surrogate, which predicts the composition of the phases if and only if two phases exist, according to the classifier as proposed in [5].

As an alternative, one can implement the calculation of the equilibrium in the process model and provide thermodynamic information via auxiliary quantities, e.g., the fugacity coefficients. This is called the indirect method, see [6]. In this case, the process model into which the surrogate is embedded has to include the isofugacity constraints explicitly. This results in less effort to train the surrogate model but a more difficult problem that has to be solved in process simulation and optimization.

Further, a decision has to be taken on how to represent the thermodynamic state as the input for the surrogate model. It is advisable to minimize the number of descriptors by including simple constraints directly into the definition of the inputs. A common representation is using molar fractions for all but one component, together with temperature and pressure.

In the direct method we propose to use a special form of the phase distribution coefficients  $\kappa_i$  as defined in Eq. (1) as outputs. This definition implicitly considers that the concentrations in two phases in equilibrium lie on a line that goes through the feed composition. These should not be confused with the usual distribution coefficients  $K_i$ , which are shown in Eq. (2).

$$\kappa_i = \frac{\dot{n}_i^I}{\dot{n}_i^F} = \frac{x_i^I}{x_i^F} \frac{x_i^F - x_i^{II}}{x_i^I - x_i^{II}} \quad (1)$$

$$K_i = \frac{x_i^I}{x_i^F} \quad (2)$$

For gas-liquid equilibria the output can conveniently be defined as the solubility of the gaseous components for a given pressure, temperature, and composition of the liquid phase.

### 3 Sampling Techniques

In this chapter, different sampling strategies are discussed. Generally, sampling strategies can be divided into explorative and exploitative strategies. Explorative sampling denotes distributing samples such that they cover the input space as well as possible, while exploitation refers to focusing the sampling on regions of interest by an evaluation of the model predictions. As the model error at possible new sampling points is not available before the sampling, techniques for the estimation of the variance of the predictions at these points have to be employed. The relative weight of exploration and exploitation is a key parameter in sampling strategies.

#### 3.1 Explorative Sampling

Exploration is largely independent of the type of surrogate model and the behavior of the original model as it only targets a broad and even distribution of the sample points. Latin hypercube sampling (LHS) is a typical method that builds upon a discretization into levels in the same manner as in a full factorial design, but the number of samples can be chosen arbitrarily [7]. This is achieved by discretizing the input space along each dimension into as many levels as there are samples to be determined. Samples are then placed such that no two samples share a discretization level in any dimension. An example of this design is shown in Fig. 2.

Instead of choosing all samples at the same time, another approach is to repeatedly add samples in a sequential way, e.g., to maximize the distance to the closest existing sample.

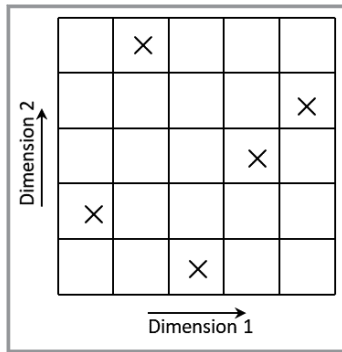


Figure 2. Example of a Latin Hypercube sampling design.

### 3.2 Adaptive Sampling for Better Global Accuracy

In explorative sampling, the objective is to distribute samples as evenly as possible throughout the input space. If the original function shows a larger variability in one region, it can be beneficial to focus the sampling more on the regions with large local variations. This is the goal of the adaptive sampling algorithm presented in [8]. Eason and Cremaschi [9] extended this idea to address the exploration vs. exploitation trade-off, which means finding a compromise between discovering all features of the original function and focusing on already found complex regions. Nentwich et al. [5] further developed this algorithm, having in mind specifically the application to surrogate models for phase equilibria. The flowsheet in Fig. 3 shows the main steps involved in the adaptive sampling algorithm from [9].

As can be seen in Fig. 3, the first step in the adaptive sampling scheme after initialization is to create a set of candidate points, which contains input combinations that are determined using Latin hypercube sampling. From this set, the new samples are later chosen after quantification of how promising the individual candidate points are by using the surrogate model(s) of the current iteration, thus not using expensive original model evaluations here. Care has to be taken in choosing the number of candidate points. If there

are too few points considered, the most promising regions might be missed due to sparse coverage of the input space, while in the case that a high number of candidate points is taken into account, clustering of new samples might occur. As a compromise, the size of this candidate set was chosen to be of the same size as the previous sample set. The set of new candidates is created in every iteration.

For the existing set of samples and the candidate set the nearest neighbor distance (NND)  $d_j$  is calculated for all candidate points  $x_j^{cand}$  to all points in the sample set  $x_j^{sample}$  (see Eq. (3)).

$$d_j = \min_{n \in \{1, \dots, N\}} \|x_j^{cand} - x_n^{sample}\|_2 \quad (3)$$

Here  $N$  represents the number of samples in a given iteration.

The space-filling objective is to maximize the nearest neighbor distance. The exploitation criterion is based on the estimation of prediction error. If Kriging surrogate models are used, the prediction error can be estimated from the surrogate model parameters directly. For other types of surrogate models, the error of the current model at the candidate points is not available. Towards this goal, the so-called jackknife variance is employed, as it showed promising results in previous work [5]. The bootstrap method can also be used as the variance estimation method, Nentwich et al. [6] report similar results compared to the jackknife variance.

First a surrogate model  $\hat{y}^{(0)}$  is trained on the full sample set  $X$ . To calculate the jackknife variance, the sample set  $X$  is divided into  $N_{SS}$  subsets, where the number of subsets  $N_{SS}$  is a hyperparameter of the algorithm. Then surrogate models  $\hat{y}^{(-i)}$  are trained using reduced training sets with the  $i$ th subset left out.

With this the jackknife variance at the  $j$ th candidate point  $\hat{\sigma}_j^2$  can be calculated as

$$\hat{\sigma}_j^2 = \sum_{i=1}^{N_{SS}} \frac{1}{N(N - N_i)} (\tilde{y}_{ji} - \bar{y}_j)^2 \quad (4)$$

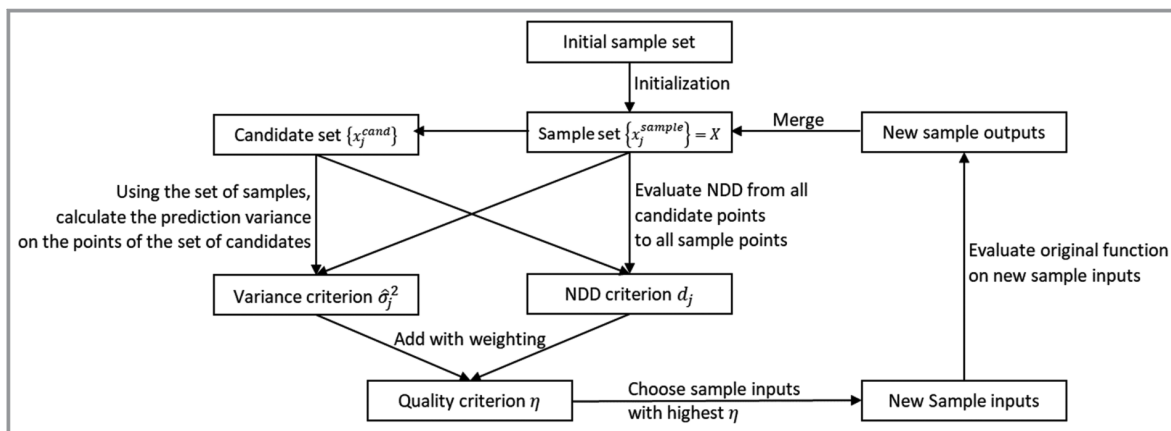


Figure 3. Steps of applying the mixed adaptive sampling algorithm from [9].

with

$$\bar{y}_j = \sum_{i=1}^{N_{ss}} \frac{N - N_i}{\sum_{l=1}^{N_{ss}} N - N_l} \tilde{y}_{ij} \quad (5)$$

and

$$\tilde{y}_{ij} = N\hat{y}^{(0)}(x_j^{cand}) - (N - N_i)\hat{y}^{(-i)}(x_j^{cand}) \quad (6)$$

$N$  denotes the total number of samples, while  $N_i$  refers to the number of samples in subset  $i$ .

With both the explorative and the exploitative objective defined, the overall quality criterion  $\eta_j$ , which describes how promising a candidate point is, can be evaluated as shown in Eq. (7).

$$\eta_j = \frac{d_j}{\max_j d_j} + \frac{\hat{\sigma}_j^2}{\max_j \hat{\sigma}_j^2} \quad (7)$$

The number of candidates  $n_{iter}$  that are added to the sampling set in each iteration is chosen as a multiple,  $\alpha$  of the number of samples in that iteration  $N$ , see Eq. (8).

$$n_{iter} = \alpha N \quad (8)$$

Nentwich et al. [5] extended this method to equilibrium calculations. Here,  $\eta_j$  is defined depending on the prediction of the classification surrogate, which is used to distinguish between the miscible and immiscible cases.  $\eta_j$  is calculated as shown in Eq. (9) for the case where the classification surrogate model predicts only a single stable phase for candidate input  $x_j^{cand}$ .

$$\eta_j = \frac{d_j}{\max_j d_j} + \frac{\hat{\sigma}_{class,j}^2}{\max_j \hat{\sigma}_{class,j}^2} \quad (9)$$

Here  $\hat{\sigma}_{class,j}^2$  is the jackknife variance for the classification surrogate. For the case of predicted immiscibility,  $\eta_j$  is calculated as

$$\eta_j = \frac{d_j}{\max_j d_j} + \frac{1}{2} \frac{\hat{\sigma}_{class,j}^2}{\max_j \hat{\sigma}_{class,j}^2} + \frac{1}{2n_c} \sum_i \frac{\hat{\sigma}_{regr,i,j}^2}{\max_j \hat{\sigma}_{regr,i,j}^2} \quad (10)$$

$\hat{\sigma}_{regr,i,j}^2$  denotes the jackknife variance of the regression model of the  $i$ th component at the  $j$ th candidate point.

## 4 Case Study

The case study that is considered in this work is the process of hydroformylation of 1-dodecene in a thermomorphic solvent system (TMS). The details of this process, the models employed and the cost function for the optimization of the flowsheet are described in the following.

### 4.1 Process Description

Hydroformylation means the reaction of CO and H<sub>2</sub> with an olefin to produce an aldehyde. The most well-known example of such a reaction performed in industrial scale is the Ruhrchemie-Rhone-Poulenc process for the production of butanal. We here consider the reaction of 1-dodecene to  $n$ -tridecanal as shown in Eq. (11). Besides the main reaction, there are unwanted side reactions. These include the reaction of 1-dodecene to one of the *iso*-dodecenes. To improve the speed of reaction and to reduce the side reactions, a selective catalyst is applied, here a homogeneous Rh(acac)(CO)<sub>2</sub> catalyst [10].



The kinetics of this chemical reaction depend on the catalyst concentration as well as on the concentration  $c_i$  of the dissolved gases carbon monoxide (CO) and hydrogen (H<sub>2</sub>) and on temperature. This can be seen from the kinetic equation of the main reaction [11]

$$r = \frac{k_{1,0}(T) c_{\text{1-dodecene}} c_{\text{H}_2} c_{\text{CO}} c_{\text{catalyst}}}{1 + K_{1,1} c_{\text{1-dodecene}} + K_{1,2} c_{\text{n-tridecanal}} + K_{1,3} c_{\text{H}_2}} \quad (12)$$

Here, the reaction rate  $r$  is described as a function of the temperature-dependent reaction rate constant  $k_{1,0}(T)$  and inhibition constants  $K_{1,j}$ .

To minimize the loss of the rhodium-based catalyst, which has a very high cost, the innovative concept of a thermomorphic solvent system [10] is used. The main idea is to use specific solvents, in this case  $n$ -decane and dimethylformamide (DMF), that result in a specific thermodynamic behavior: good miscibility of all components at elevated temperatures, and low miscibility (phase separation) at low temperatures. This leads to the possibility to use a decanter for separation of the products from the recycle stream that contains the catalyst and unreacted feed, see Fig. 4.

The degrees of freedom when designing this process include the composition of the feed that contains 1-dodecene, the solvents and the catalyst, the operating temperature and pressure, as well as the reactor volume, and the composition of the gas phase (ratio of CO/H<sub>2</sub>). In the decanter, the operating temperature can also be chosen freely. To find the best values for these parameters, a rigorous process optimization can be performed. The cost function is chosen as the

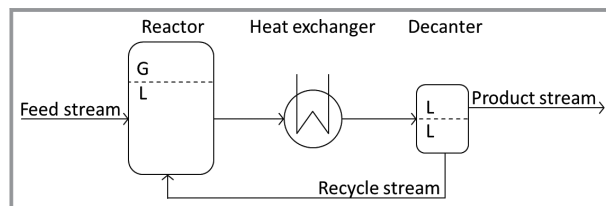


Figure 4. Process flowsheet for the hydroformylation of 1-dodecene in a thermomorphic solvent system.

production cost per ton of *n*-tridecanal produced, with parameters taken from [12] and [13].

In this process, the liquid-liquid equilibrium in the decanter and the gas-liquid equilibrium in the reactor play a central role.

## 4.2 Thermodynamic Modeling and Surrogate Models

The equation of state PC-SAFT [1] is chosen as it provides accurate descriptions of the equilibria [10, 14]. In the models, some components are lumped together based on the assumption that they show similar phase behaviors. This leads to the consideration of a liquid phase that contains four components (decane, DMF, 1-dodecene and *n*-tridecanal), and a gas phase with two components (H<sub>2</sub> and CO).

For the isofugacity criterion PC-SAFT is used in this case to compute the fugacity coefficient  $\varphi_i$  as a function of the temperature  $T$ , the pressure  $p$ , and the molar composition  $x_i$  in each phase. The fugacity coefficient is calculated as a function of the residual Helmholtz energy  $a_{res}$ :

$$\ln(\varphi_i) = a_{res} + \frac{\partial a_{res}}{\partial x_i} - \sum_{j=1}^{n_{comp}} \left( x_j \frac{\partial a_{res}}{\partial x_j} \right) + Z - 1 - \ln(Z) \quad (16)$$

The compressibility factor  $Z$  is calculated by solving the implicit density root problem. This iterative method, together with the iterative solution for the equilibrium conditions leads to a high computational effort for each equilibrium calculation.

To overcome this issue, the equilibrium calculations are approximated by a surrogate model. The surrogate for the LLE equilibrium calculations computes the miscibility, encoded as a discrete binary variable, as well as the phase distribution coefficients  $\kappa_i$  as functions of the mole fractions of the liquid feed to the decanter and the temperature.

The GLE surrogates have as inputs the liquid phase composition and the temperature as well as the pressure and the composition of the synthesis gas. The outputs are the solubilities of CO and H<sub>2</sub> in the liquid. We here present results only for direct models where a classifier is employed in the LLE computation for brevity. In the case study, these surrogate models turned out to be more reliable than indirect models.

The kind of surrogate model used has a large influence on the performance of the approximation. Several choices of surrogate models have shown good performance in previous work [15]. In this work, artificial neural networks (ANN) are considered for both regression and classification problems.

In a pre-study the hyperparameters shown in Tab. 1 were determined.

## 5 Results of the Application of the Sampling Methods

In the following, the results of the application of Latin hypercube sampling (purely explorative) and mixed adaptive sampling considering exploration and exploitation with equal weight are presented, see Eq. (10). Both sampling methods were applied five times as the generation of the samples and the training of the models involves random variables.

The performance of the sampling algorithms is measured by the deviation over a large set of precomputed input output pairs of PC-SAFT. For this set, flash calculations were conducted on approximately 10 000 points for both the LLE and the GLE. The following error metrics are considered.

$$\text{RMSE}(\hat{y}(x)) = \sqrt{\frac{1}{\tilde{n}_{testset}} \sum_{i=1}^{\tilde{n}_{testset}} \left( \hat{y}_{regr}(x^{(i)}) - y_{regr}^{(i)} \right)^2} \quad (17)$$

$$\text{MAX}(\hat{y}(x)) = \max_i \left| \hat{y}_{regr}(x^{(i)}) - y_{regr}^{(i)} \right| \quad (18)$$

$$\text{MisC}(\hat{y}(x)) = \frac{1}{n_{testset}} \sum_{i=1}^{n_{testset}} \left| \hat{y}_{class}(x^{(i)}) - y_{class}^{(i)} \right| \quad (19)$$

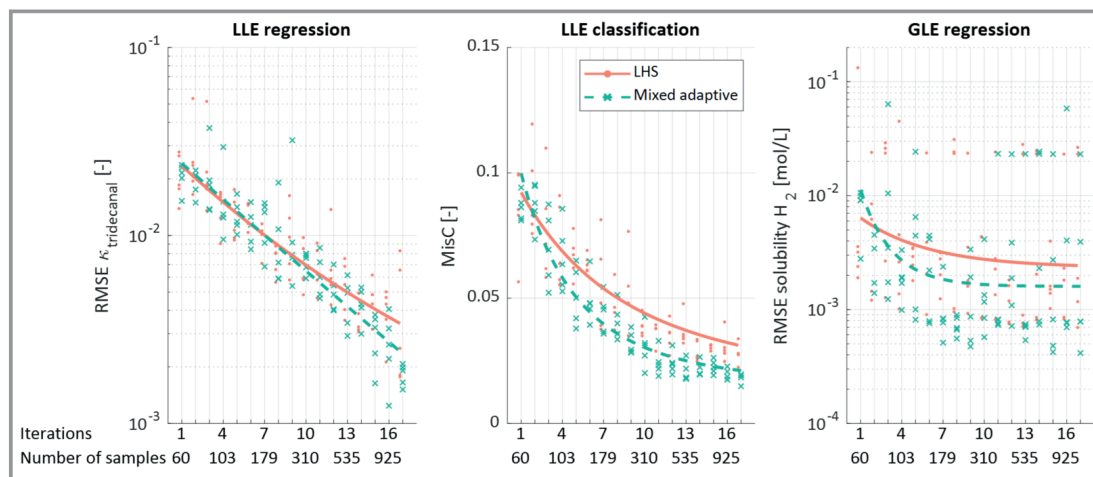
$\hat{y}_{regr}^{(i)}$  denotes the prediction of the  $i$ th test set input  $x^{(i)}$  and  $y_{regr}^{(i)}$  the value provided by PC-SAFT. Only inputs that lie in the two-phase region are considered in the RMSE and in the maximum absolute error MAX.

For classification surrogates the fraction of misclassified points MisC is calculated by Eq. (19).  $y_{class}$  and  $\hat{y}_{class}$  are binary values. The resulting RMSE plot for the phase distribution coefficient of the product tridecanal,  $\kappa_{tridecanal}$ , is shown in Fig. 5. Additionally, the fraction of misclassified points MisC and the RMSE for the hydrogen solubility in the GLE is visualized.

The number of additional samples for each method is the same, increasing by 20 % in every iteration, from initially 60 samples to 1110 samples. The test set error for the individual components is shown in Tab. 2 for the regression surrogate models for the last iteration as a mean value over the five runs of the method. MisC is 0.0305 for LHS and 0.0194

**Table 1.** Hyperparameters of the surrogate models.

|                           | LLE regression                               | LLE classification | GLE regression |
|---------------------------|--|--------------------|----------------|
| Surrogate model type      | ANN  | ANN                | ANN            |
| Number of hidden layers   | 2  | 2                  | 1              |
| Number of nodes per layer | 15   | 15                 | 30             |
| Training algorithm        | Levenberg-Marquardt [16]                     |                    |                |
| Validation data           | 10 % of training dataset                     |                    |                |
| Stopping criterion        | Validation error not decreasing for 6 epochs |                    |                |



**Figure 5.** Test set errors of surrogate models trained on sample sets generated by applying LHS (orange) and mixed adaptive sampling (green). Left hand side: RMSE of the phase distribution coefficient of tridecanal, middle: fraction of misclassified points MisC, right hand side: RMSE of the solubility of hydrogen. Dots: results for five runs, lines: fit line. Iterations refers to sequential sampling iterations.

**Table 2.** Test set errors of different sampling methods in the last iteration of the sequential sampling method using 1110 samples for LLE and GLE regression surrogate outputs.

|   | LHS    | LHS    | Mixed adaptive sampling |        |
|---|--------|--------|-------------------------|--------|
|   | RMSE   | MAX    | RMSE                    | MAX    |
| $\kappa_{\text{decane}}$ [-]                    | 0.0050 | 0.1199 | 0.0025                  | 0.0554 |
| $\kappa_{\text{DMF}}$ [-]                       | 0.0035 | 0.0749 | 0.0017                  | 0.0264 |
| $\kappa_{1\text{-dodecene}}$ [-]                | 0.0048 | 0.1143 | 0.0024                  | 0.0483 |
| $\kappa_{\text{tridecanal}}$ [-]                | 0.0042 | 0.0910 | 0.0018                  | 0.0303 |
| Solubility $\text{H}_2$ [ $\text{mol L}^{-1}$ ] | 0.0107 | 0.0315 | 0.0058                  | 0.0199 |
| Solubility CO [ $\text{mol L}^{-1}$ ]           | 0.0182 | 0.0541 | 0.0099                  | 0.0339 |

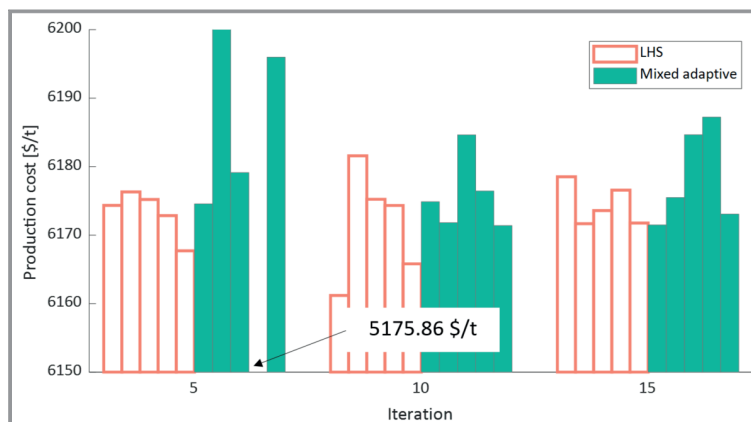
for mixed adaptive sampling. In general, the adaptive sampling reduces the RMSE and MAX errors to about 50% compared to LHS sampling.

The MAX error of the LLE regression is significantly higher than the corresponding RMSE for both methods. Obviously, the surrogate response exhibits a complex behaviour for some parts of the response surface while being almost flat in others. The RMSE and MisC values indicate a good average approximation over the ranges of interest. To investigate whether the surrogates are also suited for process optimization, the minimization of the cost described in Sect. 4.1 is performed using IPOPT [17] with a multistart heuristic using six different initial points to find the best local minima. This optimization is applied using the surrogate models

from different iterations of the sampling procedure. The results are shown in Fig. 6.

At iteration 10, the adaptive sampling performs better than LHS, but at iteration 15, the variability of the optimum for the adaptive sampling method is still significant. While the adaptive sampling leads to a better overall approximation of the equilibria, the behaviour near the optimum introduces errors in the optimization. In order to further investigate whether the predicted optimum at  $5175.86 \text{ \$ t}^{-1}$  that largely differs from the other values corresponds to a realistic operating point, the surrogate model outputs are compared to PC-SAFT predictions at the different optima for the mixed adaptive sampling method at iteration 5. The results are shown in Tab. 3.

From the results in Tab. 3 it becomes clear that the surrogate models show relatively low deviations from PC-SAFT



**Figure 6.** Optimized production cost using the surrogate models trained on sample sets generated by applying LHS (orange) and mixed adaptive sampling (green). The bar charts correspond to five different runs of the sampling and surrogate computation methods and optimization for the resulting models.

**Table 3.** Comparison of surrogate model predictions (marked with ^ to PC-SAFT outputs for different sampling runs in iteration 5 of the mixed adaptive sampling method, evaluated at the predicted process optima. The indices 1 to 4 denote the components decane, DMF, 1-dodecene and tridecanal.

| Production cost [ $\text{\$ t}^{-1}$ ] | Phase distribution coefficients (LLE) [-] |            |                  |            |                  |            |                  |            | Solubilities (GLE) [ $\text{mol L}^{-1}$ ] |                  |                       |                 |
|--|---|------------|------------------|------------|------------------|------------|------------------|------------|--|------------------|-----------------------|-----------------|
|  | $\hat{\kappa}_1$                          | $\kappa_1$ | $\hat{\kappa}_2$ | $\kappa_2$ | $\hat{\kappa}_3$ | $\kappa_3$ | $\hat{\kappa}_4$ | $\kappa_4$ | $\hat{c}_{\text{H}_2}$                     | $c_{\text{H}_2}$ | $\hat{c}_{\text{CO}}$ | $c_{\text{CO}}$ |
| 6174.56                                | 0.96                                      | 0.96       | 0.35             | 0.35       | 0.96             | 0.96       | 0.83             | 0.83       | 0.10                                       | 0.10             | 0.21                  | 0.21            |
| 6200.35                                | 0.97                                      | 0.97       | 0.41             | 0.40       | 0.97             | 0.97       | 0.87             | 0.87       | 0.10                                       | 0.10             | 0.21                  | 0.21            |
| 6179.14                                | 0.96                                      | 0.96       | 0.30             | 0.30       | 0.95             | 0.95       | 0.81             | 0.81       | 0.10                                       | 0.10             | 0.21                  | 0.22            |
| 5175.86                                | 0.77                                      | 0.83       | 0.00             | 0.06       | 0.75             | 0.81       | 0.40             | 0.45       | 0.00                                       | 0.00             | 0.36                  | 0.36            |
| 6195.98                                | 0.96                                      | 0.96       | 0.35             | 0.35       | 0.95             | 0.95       | 0.82             | 0.82       | 0.10                                       | 0.10             | 0.21                  | 0.21            |

in all cases except for the outlier with a predicted operating cost of  $5175.86 \text{ \$ t}^{-1}$ . This leads to the conclusion that this optimum is not physically feasible. Therefore, for the goal of process optimization, another approach is proposed that favors samples close to the predicted optimum rather than in regions of large variations of the original function.

## 6 Sampling for Optimization

The problem of optimizing an objective function that is costly to evaluate has attracted much research effort over the years. One approach to this problem is Bayesian optimization, see [18] for an early application. Here, the objective function is evaluated at the sample points and a surrogate model is used to predict the objective function between the samples. Sampling is conducted by optimizing an acquisition function that depends on a surrogate model of the objective function. The assumption in Bayesian optimization in its original form is that the objective function is described directly by the surrogate model. This however is not the case in chemical process optimization when submodels are replaced by surrogates and the surrogates have internal model variables as inputs. In other words, what we are addressing here is an optimization problem with a gray-box model, in which the surrogate models can also be seen as constraints that are added to the remaining, first-principles based model that is part of the optimization problem.

The approach proposed in [19] extends the upper confidence bound (UCB) function to gray-box models. The original formulation of the acquisition function  $f^{UCB}$  proposed by Cox and John [20] is:

$$f^{UCB}(x) = \hat{y}(x) + b\sqrt{\hat{\sigma}^2(x)} \quad (20)$$

Here,  $\hat{y}$  denotes the surrogate prediction and  $\hat{\sigma}^2$  the estimated variance of the prediction. The scaling  $b$  is used to address the exploration vs. exploitation trade-off. It is set to a negative value in the case that the optimization involves minimization. This acquisition function can be understood as optimistically assuming that the expected surrogate error leads to a better objective function value. Therefore, the

result of optimizing  $f^{UCB}$  leads to promising new sampled inputs.

This idea was extended in [19] to gray-box models. Here, the sensitivity of the optimum to the deviation of the prediction by the surrogate model,  $g_i$ , is approximated by using Lagrange multipliers. The resulting acquisition function  $f^{UCB-GB}$  can be represented as shown below:

$$f^{UCB-GB}(x) = f^{GB}(x, \hat{y}(x)) + b \sum_i^{n_{surr}} |g_i| \sqrt{\hat{\sigma}_i^2(x)} + \Phi(x) \quad (21)$$

$f^{GB}$  represents the first-principles based model equations, which depend on optimization variables  $x$  as well as surrogate model outputs  $\hat{y}(x)$ .  $\Phi(x)$  is a penalty term that is used to generate multiple samples in an iteration.  $n_{surr}$  describes the number of surrogate model outputs that are part of the process model.

To optimize the gray-box acquisition function  $f^{UCB-GB}$  and to estimate the sensitivities  $g_i$  from the Lagrange multipliers  $\lambda^{UCB-GB}$ , the following constrained optimization problem is solved:

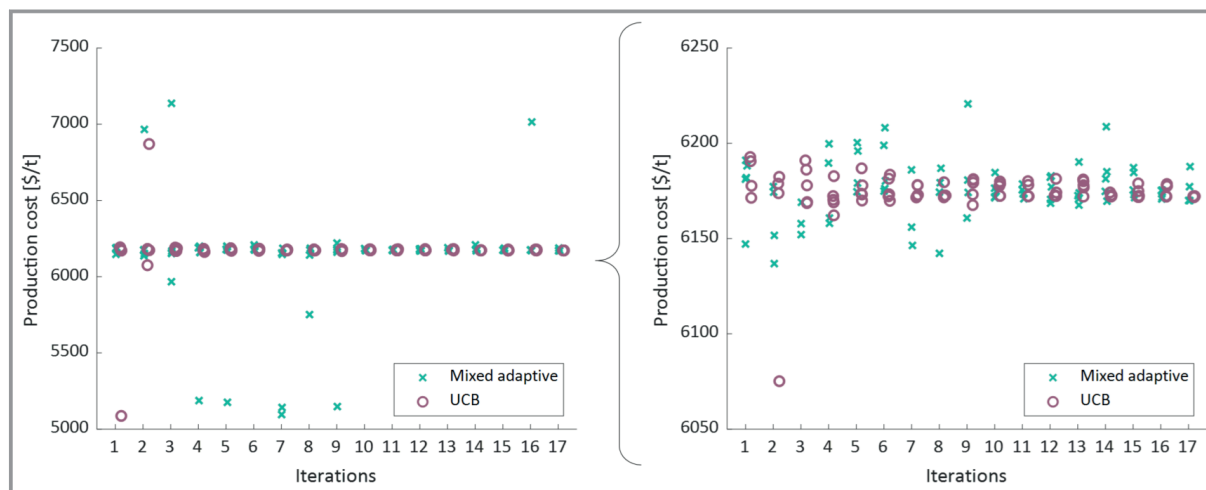
$$\begin{aligned} \min_{x,z} f^{GB}(x, z) + b \sum_i^{n_{surr}} |g_i| \sqrt{\hat{\sigma}_i^2(x)} + \Phi(x) \\ \text{s.t. } h(x, z) < 0 \\ \hat{y}(x) - z = 0 \quad |\lambda^{UCB-GB} \end{aligned} \quad (22)$$

The auxiliary decision variables  $z$  are constrained to the surrogate model output by means of the Lagrange multipliers  $\lambda^{UCB-GB}$ .  $h(x, z)$  denotes constraints that result from the first-principles based model.

The optimization problem shown in Eq.(22) is solved iteratively updating the sensitivities with the Lagrange multipliers. Further details can be found in [19].

This sampling method, denoted as UCB, is applied to the case study described before with the  $b$  parameter set to  $-1.96$ . The results of this sampling, training and optimization procedure are shown in Fig. 7.

There is some inherent noise in these optimum values due to the reinitialization of the ANN weights and bias parameters in each iteration because the training procedure



**Figure 7.** Result of the optimization of the production cost using surrogate models that were trained on sample sets generated by applying mixed adaptive sampling (green) and UCB (purple). Left: full range, right: close-up near the optimum.

incorporates a nonconvex optimization problem that is susceptible to local minima. In general, it can be seen that in later iterations with a higher sample size the optimum is predicted to be around  $6175 \text{ \$ t}^{-1}$ . As shown before, optima of low surrogate approximation error are found in this production cost range. Therefore, it can be assumed that the true optimum lies there. For the UCB sampling, the deviations in the predicted optimum are small after six iterations (149 samples), so that the minimum cost is determined very accurately already for a small number of samples (note that the rigorous model is also not perfectly describing reality).

## 7 Conclusion and Outlook

Surrogate models are a suitable tool to reduce the computational effort in process simulation and optimization when thermodynamic properties are computed by complex models that require iterative solutions of nonlinear equations. Their application involves the choice of the model inputs and outputs, which should incorporate as much knowledge as possible like, e.g., the constraint of the molar balance over the phases. A fundamental choice is between direct or indirect representations. In our case, the direct representation turned out to provide a better accuracy.

We demonstrated that for the goal of achieving a uniformly high accuracy of the model, adaptive sampling methods are preferable over space filling methods. However, this does not necessarily lead to models that are also well suited to determine optimal designs and operating conditions. For this, a new method that proposes candidates with an improvement potential was proposed and demonstrated to be efficient.

There is an interplay between the sampling strategy and the computation or training of the surrogate models and the choice or possibly also adaptation of their structure and

the training parameters. This was not investigated deeply here; surrogate models of a fixed structure and a fixed training method were used. To investigate this interplay more deeply is an interesting topic for further research.

This research has been supported by the project “KI-Inkubator-Labore in der Prozessindustrie – KEEN”, funded by the Bundesministerium für Wirtschaft und Energie (BMWi) under grant number 01MK20014T. This support is gratefully acknowledged. Open access funding enabled and organized by Projekt DEAL.

### Symbols used

|             |   |  |
|-------------|---|--|
| $b$         | [-]   | Weighting parameter                            |
| $c$         | [mol L <sup>-1</sup> ]                              | Concentration                                  |
| $d$         | [-]   | Distance                                       |
| $f$         | [-]   | Objective function                             |
| $g$         | [-]   | Sensitivity                                    |
| $h$         | [-]   | Constraints from gray-box model                |
| $K$         | [-; L mol <sup>-1</sup> ]                           | Distribution coefficient / inhibition constant |
| $k$         | [L <sup>4</sup> mol <sup>-3</sup> s <sup>-1</sup> ] | Reaction rate constant                         |
| $\dot{n}$   | [mol s <sup>-1</sup> ]                              | Mole flow                                      |
| $n$         | [-]   | Number   |
| $r$         | [mol s <sup>-1</sup> ]                              | Reaction rate                                  |
| $x$         | [-]   | Mole frac. sample inputs                       |
| $y$         | [-]   | Function output                                |
| $\hat{y}$   | [-]   | Surrogate prediction                           |
| $\tilde{y}$ | [-]   | Jackknife pseudo value                         |
| $\bar{y}$   | [-]   | Mean jackknife pseudo value                    |
| $Z$         | [-]   | Compressibility factor                         |
| $z$         | [-]   | Auxiliary decision variables                   |



## Greek letters

|                  |     |  |
|------------------|-----|--|
| $\alpha$         | [-] | Factor determining number of new samples |
| $\eta$           | [-] | Candidate criterion                      |
| $\kappa_i$       | [-] | Phase distribution coefficient           |
| $\lambda$        | [-] | Lagrange multiplier                      |
| $\hat{\sigma}^2$ | [-] | Estimated variance                       |
| $\varphi$        | [-] | Fugacity coefficient                     |
| $\Phi$           | [-] | Parallelization penalty                  |

## Sub- and Superscripts

|        |  |
|--------|--|
| (-0)   | No subset left out                                 |
| cand   | Candidate  |
| class  | Classification                                     |
| comp   | Components   |
| F      | Feed   |
| GB     | Gray-box   |
| I      | Phase 1  |
| II     | Phase 2  |
| iter   | Iteration  |
| (-i)   | <i>i</i> th subset left out                        |
| regr   | Regression   |
| sample | Sample   |
| SS     | Subset   |
| surr   | Surrogate  |
| UCB    | Upper confidence bound                             |
| UCB-GB | Upper confidence bound extended to gray-box models |

## Abbreviations

|         |  |
|---------|--|
| CFD     | Computational fluid dynamics                         |
| DMF     | Dimethylformamide                                    |
| G       | Gas  |
| GLE     | Gas liquid equilibrium                               |
| L       | Liquid   |
| LLE     | Liquid liquid equilibrium                            |
| MAX     | Maximum error  |
| MisC    | Fraction of misclassified points                     |
| NND     | Nearest neighbor distance                            |
| PC-SAFT | Perturbed chain statistical associating fluid theory |
| RMSE    | Root mean squared error                              |

## References

- [1] J. Gross, G. Sadowski, *Ind. Eng. Chem. Res.* **2001**, *40* (4), 1244–1260. DOI: <https://doi.org/10.1021/ie0003887>
- [2] P. Schäfer, A. Caspari, K. Kleinhans, A. Mhamdi, A. Mitsos, *AIChE J.* **2019**, *65* (5), e16568. DOI: <https://doi.org/10.1002/aic.16568>
- [3] J. Sirignano, K. Spiliopoulos, *J. Comput. Phys.* **2018**, *375*, 1339–1364. DOI: <https://doi.org/10.1016/j.jcp.2018.08.029>
- [4] A. I. J. Forrester, N. W. Bressloff, A. J. Keane, *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2006**, *462* (2071), 2177–2204. DOI: <https://doi.org/10.1098/rspa.2006.1679>
- [5] C. Nentwich, S. Engell, *Comput. Chem. Eng.* **2019**, *126*, 204–217. DOI: <https://doi.org/10.1016/j.compchemeng.2019.04.006>
- [6] C. Nentwich, J. Winz, S. Engell, *Ind. Eng. Chem. Res.* **2019**, *58* (40), 18703–18716. DOI: <https://doi.org/10.1021/acs.iecr.9b02758>
- [7] M. D. McKay, R. J. Beckman, W. J. Conover, *Technometrics* **1979**, *21* (2), 239–245. DOI: <https://doi.org/10.1080/00401706.1979.10489755>
- [8] J. P. C. Kleijnen, W. C. M. Van Beers, *J. Oper. Res. Soc.* **2004**, *55* (8), 876–883. DOI: <https://doi.org/10.1057/palgrave.jors.2601747>
- [9] J. Eason, S. Cremaschi, *Comput. Chem. Eng.* **2014**, *68*, 220–232. DOI: <https://doi.org/10.1016/j.compchemeng.2014.05.021>
- [10] E. Schäfer, Y. Brunsch, G. Sadowski, A. Behr, *Ind. Eng. Chem. Res.* **2012**, *51* (31), 10296–10306. DOI: <https://doi.org/10.1021/ie300484q>
- [11] G. Kiedorf, D. M. Hoang, A. Müller, A. Jörke, J. Markert, H. Arellano-Garcia, A. Seidel-Morgenstern, C. Hamel, *Chem. Eng. Sci.* **2014**, *115*, 31–48. DOI: <https://doi.org/10.1016/j.ces.2013.06.027>
- [12] R. Turton, R. Bailie, W. Whiting, J. Shaeiwitz, *Analysis, Synthesis and Design of Chemical Processes*, Prentice Hall, Englewood Cliffs, NJ **2008**.
- [13] B. Hentschel, A. Peschel, H. Freund, K. Sundmacher, *Chem. Eng. Sci.* **2014**, *115*, 69–87. DOI: <https://doi.org/10.1016/j.ces.2013.09.046>
- [14] C. Vogelpohl, C. Brandenbusch, G. Sadowski, *J. Supercrit. Fluids* **2014**, *88*, 74–84. DOI: <https://doi.org/10.1016/j.supflu.2014.01.017>
- [15] C. Nentwich, C. Varela, S. Engell, in *Proc. Int. Jt. Conf. Neural Networks*, Vol. 2019 July, Institute Of Electrical And Electronics Engineers Inc., Piscataway, NJ **2019**.
- [16] M. T. Hagan, M. B. Menhaj, *IEEE Trans. Neural Networks* **1994**, *5* (6), 989–993. DOI: <https://doi.org/10.1109/72.329697>
- [17] A. Wächter, L. T. Biegler, *Math. Program.* **2006**, *106* (1), 25–57. DOI: <https://doi.org/10.1007/s10107-004-0559-y>
- [18] D. R. Jones, M. Schonlau, W. J. Welch, *J. Glob. Optim.* **1998**, *13* (4), 455–492. DOI: <https://doi.org/10.1023/A:1008306431147>
- [19] J. Winz, S. Engell, in *31st European Symposium on Computer Aided Process Engineering* (Eds: M. Türkay, R. Gani), Elsevier, Amsterdam **2021**.
- [20] D. D. Cox, S. John, in *Proc. of 1992 IEEE Int. Conf. on Systems, Man and Cybernetics*, IEEE, Piscataway, NJ **1992**.