Accepted Manuscript submitted to Applied Soft Computing
DOI of the final article: http://dx.doi.org/10.1016/j.asoc.2016.05.047

Multi-fidelity Modeling and Optimization of Biogas Plants

Martin Zaefferer*, Daniel Gaida, Thomas Bartz-Beielstein

Faculty for Computer and Engineering Sciences, TH Köln (University of Applied Sciences), Steinmüllerallee 1, 51643 Gummersbach, Germany

Abstract

An essential task for operation and planning of biogas plants is the optimization of substrate feed mixtures. Optimizing the monetary gain requires the determination of the exact amounts of maize, manure, grass silage, and other substrates. For this purpose, accurate simulation models are mandatory, because the underlying biochemical processes are very slow. The simulation models may be time-consuming to evaluate, hence we show how to use surrogate-model-based approaches to optimize biogas plants efficiently. In detail, a Kriging surrogate is employed. To improve model quality of this surrogate, we integrate cheaply available data into the optimization process. To this end, multi-fidelity modeling methods like Co-Kriging are applied. Furthermore, a two-layered modeling approach is used to avoid deterioration of model quality due to discontinuities in the search space. At the same time, the cheaply available data is shown to be very useful for initialization of the employed optimization algorithms. Overall, we show how biogas plants can be efficiently modeled using data-driven methods, avoiding discontinuities as well as including cheaply available data. The application of the derived surrogate models to an optimization process is only partly successful. Given the same budget of function evaluations, the multi-fidelity approach outperforms the alternatives. However, due to considerable computational requirements, this advantage may not translate into a success with regards to overall computation time.

Keywords: Biogas Plant, Simulation, Optimization, Surrogate Models, Multi-fidelity, Co-Kriging

1. Introduction

Optimizing the operation of biogas plants is and will be one of the main challenges in the field of *anaerobic digestion* (AD) in the near future. Due to a steady decrease in funding and increasing substrate costs only optimal operating biogas plants will be economically advantageous.

The operation of biogas plants is very sensitive to the mixture of the used substrates. Hence, optimizing the mixture is an important task to run or plan such plants efficiently. Due to the very slow processes involved, optimizing the plants in real-time would consume too much time. Models like the *Anaerobic Digestion Model No. 1* (ADM1) allow to compute a good prediction of biogas plant's process variables, based on the used substrates [7]. Thus, ADM1 can be used as a substitute in the optimization process instead of a real plant.

While such models are much cheaper to evaluate than their real-world counter-part, they do take some time to evaluate. Hence, methods that use the smallest amount of evaluations possible are of interest. This situation motivated the central question that will be tackled in this study:

(Q-1) How can the precision of simulation models be improved without increasing the number of evaluations?

Surrogate modeling techniques are therefore a promising choice. Besides the expensive information derived from ADM1, additional performance information is available. A rough performance estimate can be determined based on the biogas potential of the used substrates and their associated costs. This additional knowledge can be integrated into the optimization process, by bolstering the quality of the chosen surrogate-modeling technique. This approach of integrating different levels of granularity or cost has previously been called multi-fidelity optimization [19]. It is worth in-

*Corresponding author. Phone: +49 2261 8196 6327
Email addresses: martin.zaefferer@th-koeln.de (Martin Zaefferer), daniel.gaida@th-koeln.de (Daniel Gaida), thomas.bartz-beielstein@th-koeln.de (Thomas Bartz-Beielstein)

vestigating whether these approaches are applicable to real-world settings. This can be formulated as the second question to be analyzed in this study:

(Q-2) What are the benefits and limitations of multi-fidelity modeling approaches?

In this paper, several multi-fidelity modeling approaches are compared, and the best are tested for their performance in an optimization process.

Section 2 gives an overview of relevant previous work. The specific problem to be solved is introduced in Section 3. In Section 4, methods that were used in this study are described. Section 5 presents experiments, in which various multi-fidelity approaches are tested for their modeling quality, whereas Section 6 tests the best of these for their success in solving the actual optimization problem. A concluding summary of findings as well as an outlook on future research is given in Section 7.

2. Former Research

2.1. Biogas Plant Simulation

Islam et al. [28] analyze the impact of different factors on production of biogas in different biogas plants of Bangladesh. The data was collected from 18 poultry farms. Their analysis is based on collected data from survey, Internet, and other sources. To obtain further insight in the behavior of biogas plants, simulation models such as the ADM1 can be used. ADM1 is very popular and the nowadays most complex mathematical model used to simulate the anaerobic digestion process (for a review see [6]). In several publications it is utilized to dynamically model full-scale agricultural and industrial biogas plants [8, 33, 47]. ADM1 is a structured model incorporating disintegration and hydrolysis, acidogenesis, acetogenesis, and methanogenenesis steps. The ADM1 is implemented as a stiff differential equation system in a MATLAB® toolbox for biogas plant modeling, optimization and control published by Gaida et al. [23]. In this toolbox, a model of a full-scale agricultural biogas plant is developed that is used in the empirical part of this publication. The simulation model of the biogas plant includes the ADM1 and furthermore models of electrical and thermal energy sinks and sources as well as models for performance and stability criteria. Typical criteria include cost versus benefit (with respect to the Renewable Energy Sources Act (EEG 2009) in Germany [9]), stability of substrate degradation processes and operating constraints such as upper and lower pH limits, maximum VFA/TA [52] value, maximum total solids content in the digester, and minimum methane concentration of the biogas.

2.2. Biogas Substrate Feed Optimization

Biogas plant substrate feed mixtures have previously been optimized with a Genetic Algorithm and Particle Swarm Optimization by Wolf et al. [56]. More recently Ziegenhirt et al. [60] used state of the art evolution strategies like *Covariance Matrix Adaption Evolution Strategy* (CMAES) [27, 26] or *Differential Evolution* (DE) [54] to reduce the number of needed simulations. They also used the *Sequential Parameter Optimization Toolbox* (SPOT) [5] to tune the employed algorithms. In our work, we directly use SPOT on the substrate feed optimization problem. That is, we support the optimization procedure with surrogate-models.

Both previous studies used a biogas plant model based on the MATLAB® Simulink® Toolbox SIMBA, developed by ifak system GmbH[1]. The herein presented research on the other hand is based on the MATLAB® Toolbox for Biogas Plant Simulation [23]. In contrast to earlier works by Wolf et al. [56] and Ziegenhirt et al. [60] our approach is not limited to the ADM1. A simple estimate of a substrate mixtures quality is derived from the biogas potential of each ingredient.

2.3. Surrogate Modeling in Optimization

Especially when the evaluation of target functions is expensive, it is a well established approach to exploit surrogate models of the target function to save expensive function evaluations.

A methodical framework for surrogate model based optimization of noisy and deterministic problems is *Sequential Parameter Optimization* (SPO) introduced by Bartz-Beielstein et al. [5]. SPO has been developed for solving expensive algorithm tuning problems but can be directly employed for solving real world engineering problems as well.

One of the most often used surrogate-models is Kriging, which is an especially promising model for continuous, smooth problem landscapes. Besides its prediction performance, it is often employed because it provides an estimator of the local certainty of the model, which can be used to calculate the *Expected Improvement* (EI) of a new sample over the best known sample. Jones et al. [32] introduced this concept to balance exploitation and exploration in expensive optimization, terming it *Efficient Global Optimization* (EGO).

Other models include *Artificial Neural Networks* (ANN) or *Support Vector Regression* (SVR) [14]. Non-continuous problem landscapes, or problems which are not that expensive, may be tackled with approaches like

[1] `www.ifak-system.com`

Random Forest (RF) [11] or *Multivariate Adaptive Regression Splines* (MARS) [20].

A comprehensive overview of surrogate model assisted optimization was provided by Jin [30], focusing on single objective problems.

Extensions of the above concepts to multi-objective problems are available (e.g., multi objective EGO [35, 48, 16] and SPO [58, 59]). Since multi-objective problems are not in the focus of this paper, we refer to the overview by Knowles and Nakayama [36] for further information.

2.4. Multi-fidelity

Multi-fidelity optimization [19] deals with problems where the target function can be evaluated at different levels of fidelity. That is, the actual target function represents the highest level of fidelity, yielding the most accurate but also most expensive fitness estimate. At the same time, one or several cheaper, less accurate estimates can represent the lower fidelity levels. The actual, expensive target function will be referred to as the *fine function*, whereas the cheaper and less accurate function will be referred to as *coarse function*, respectively. Note, that in this study, multi-fidelity will usually refer to the case where in fact at least three levels of fidelity exist: fine function, coarse function and surrogate model. Only the first two are inherent to the problem, the third is learned based on collected data.

Such situations often arise, especially in engineering problems. There, the evaluation of the actual problem may be an expensive real-world evaluation measurement, or a time consuming *Computational Fluid Dynamics* (CFD) simulation. In these cases, a simplified physics-based model may yield an inexpensive but less accurate quality estimate. For some models, fidelity may even be scalable. For instance, simplified meshes with less density can be employed with CFD, or if available pre-converged simulation results may be harnessed.

2.4.1. Multi-fidelity Modeling

To exploit information from different fidelity levels in surrogate modeling, several methods exist, including Co-Kriging. Forrester et al. [19] show how this can be applied to engineering problems. Co-Kriging exploits correlation between coarse and fine function to generate a better surrogate model of the fine function.

More simple, yet often used, is the idea of using some kind of scaling to correct the lower fidelity model. This can be understood as introducing a scaling-model which corrects the error of the lower fidelity model. The most typical approaches are to use multiplicative or additive scaling (or a combination of both). Haftka introduced a multiplicative scaling to correct a global approximation with that of a local one, by using a scaling factor based on the ratio of both approximations and a first order Taylor series expansion [25]. A similar approach that uses additive scaling was first proposed by Lewis and Nash [39]. The original first-order corrections of these types have also been extended to second order [24, 15].

Another approach is *Space Mapping* (SM) [4]. In essence, SM tries to find a transformation, that maps from the parameter space of a fine function into the space of a coarse function or vice versa. This transformation can be used to map a solution found with a coarse function into the fine function space. This also has the advantage that coarse and fine function space need not have the same dimensionality.

While many applications are mostly concerned with deterministic simulations (e.g., using CFD) multi-fidelity techniques have also been applied to stochastic problems. Kuya et al. [38] combine data from simulations (deterministic, low-fidelity) and experiments (non-deterministic, high-fidelity). They studied the influence of experimental design strategies (mainly on the low-fidelity level) on the accuracy of a Co-Kriging model. To deal with systematic error, they also propose to use blocking and randomization to avoid respective bias from the experimental procedure.

2.4.2. Multi-fidelity Optimization

Once a certain multi-fidelity modeling technique is chosen, it has to be decided how to implement the respective models into an optimization framework. One approach is to use model management techniques to switch between different levels of fidelity (which may or may not include surrogate models) [42]. One variety of these techniques are trust-region strategies. Here, a lower fidelity model is used to approximate the higher fidelity model during the one-dimensional search along the direction of the gradient, inside a region of trust (with respect to accuracy of the lower fidelity model) [1, 2]. A rather new approach is to switch between different fidelity methods, as proposed by Mehmani et al. [42]. They initialize with the lowest fidelity model and subsequently make decisions for switching to higher fidelity models. The decision is based on a non-deterministic metric that takes into account the uncertainty of a model as well as the observed improvement in fitness. When the uncertainty is stronger than the improvement, a change of the fidelity level is made. Another model selection technique for a stochastic case is introduced by Mullins and Mahade-

van [44]. They take into account the predictive accuracy of a model as well as the required computational effort (which may have to be approximated) to make a selection.

Essentially, even the earlier introduced surrogate modeling techniques (Section 2.3) can be understood as a case of multi-fidelity optimization. They represent a special case where the lower fidelity model is usually a data-driven model. In that sense, some methods (e.g., the trust region strategies) are also classified as surrogate modeling techniques. Surrogate-model based searches, where the design (as well as the model) is sequentially updated are also sometimes called sequential sampling (e.g., the earlier introduced EGO algorithm) [29].

2.4.3. Areas of Application

Multi-fidelity optimization has been applied in several areas. In the following we give a brief overview of some of these areas. One important and very frequent area of application for multi-fidelity optimization is aerodynamic optimization. For example, several of the earlier cited works deal with applications from this field (cf. [2, 24, 19, 38, 42]). Another application of multi-fidelity optimization is laser peening, as reported by Singh and Grandhi [53]. They make use of application-specific models and surrogate models of varying fidelity which they employ in different phases of a particle swarm optimization algorithm. Koziel et al. [37] used Co-Kriging to design electromagnetic antennas [37]. Electro-magnetic design problems are also a very typical application of space mapping, which originated from this area [4].

3. Problem Description

In this paper we deal with a problem where two fidelity levels are available. The optimization objective as well as its two fidelity levels are described in this section.

3.1. The Optimization Problem

Consider a biogas plant fed with $k \in \mathbb{N}$ substrates. Its $m \in \mathbb{N}$ dimensional system state is symbolized by $\boldsymbol{z} : \mathbb{R}_0^+ \rightarrow \mathcal{Z}$ and its substrate feed by $\boldsymbol{x} \in \mathcal{X}$, with $\mathcal{Z} \subseteq \mathbb{R}^m$ and $\mathcal{X} \subseteq \mathbb{R}^k$ denoting the state and input space, respectively.

The objective is to minimize a one-dimensional objective function $f : \mathcal{Z} \times \mathcal{X} \rightarrow \mathbb{R}$, which depends on the state $\boldsymbol{z}$ and the substrate feed $\boldsymbol{x}$ of the biogas plant, approximately modeled by a set of nonlinear differential equations $\dot{\boldsymbol{z}}(t) = \boldsymbol{g}(\boldsymbol{z}(t), \boldsymbol{x})$, called the biogas plant model $\boldsymbol{g} : \mathcal{Z} \times \mathcal{X} \rightarrow \mathbb{R}^m$. The optimization problem is solved by choosing the optimal substrate feed $\boldsymbol{x}$. With the initial state $\boldsymbol{z}_0 \in \mathcal{Z}$ it can be formulated as:

$$\begin{aligned} &\max_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{z}(t), \boldsymbol{x}) \\ \text{s.t. } &\dot{\boldsymbol{z}}(t) = \boldsymbol{g}(\boldsymbol{z}(t), \boldsymbol{x}), \qquad \boldsymbol{z}(0) = \boldsymbol{z}_0, \\ &\boldsymbol{x} \geq \boldsymbol{x}_{lo}, \qquad \boldsymbol{x} \leq \boldsymbol{x}_{up}. \end{aligned} \tag{1}$$

Here, $\boldsymbol{x}_{lo}, \boldsymbol{x}_{up}$ are the lower and upper bounds for the substrate feed $\boldsymbol{x}$.

3.2. The Objective Function

The objective function f represents the daily financial gain of the plant, defined as follows:

$$\begin{aligned} f :=& rev_{\text{elect.}}(\boldsymbol{z}(t), \boldsymbol{x}) \\ &+ rev_{\text{therm.}}(\boldsymbol{z}(t), \boldsymbol{x}) \\ &- cost_{\text{energ.}}(\boldsymbol{z}(t), \boldsymbol{x}) \\ &- cost_{\text{subs.}}(\boldsymbol{x}) \end{aligned} \tag{2}$$

The objective values are given in Euros per day. The decision space is spanned by the amount of each substrate in the mixture which is fed into the plant. In detail. the four terms are:

- revenue from selling electrical energy produced in combined heat and power plants ($rev_{\text{elect.}}$)
- revenue from selling thermal energy produced in combined heat and power plants ($rev_{\text{therm.}}$)
- cost of energy used in plant operation, e.g., stirring the digester content, substrate transportation, heating the digesters ($cost_{\text{energ.}}$)
- cost of substrates ($cost_{\text{subs.}}$).

The revenue is defined by the profit obtained selling the produced electrical and thermal energy, which, in Germany, is determined by the Renewable Energy Sources Act - EEG. These values are obtained at steady state operation of the biogas plant. The different terms are more or less dependent on the specific substrate mixture, and of course dependent on plant-specific parameters which can be assumed to be constant, e.g., size of fermenters or outside temperature.

3.3. The Fine Function

The fine objective function (denoted f_f) is based on the ADM1 simulation of a biogas plant. The modeled biogas plant contains two digesters and produces an electrical power of 500 kW. As mentioned above, the used implementation is the MATLAB® toolbox developed by Gaida et al. [23].

This toolbox is able to yield information about all relevant process variables, as well as calculates the monetary gain for a given setting. Depending on the exact setup, the model will take at least 30 seconds, with an average of about 1 minute to compute the daily monetary gain for a certain substrate mixture in equilibrium state. Note, that simulation time may also depend on the quality of the solution, and the number of substrates used, see Fig. B.13 in the appendix. Simulations may fail, or are stopped if they do not yield a result after 10 minutes. These cases have to be dealt with during optimization, as discussed in Section 4.3.

The volumetric flowrate of each available substrate in the in-feed mixture is varied during the optimization process. That means, the dimension of the decision space depends on the number of available substrate types. The term "available" can refer to physical availability of a substrate at the plant, or the availability of calibration data for that substrate. Only substrates with known parameters can be represented by the ADM1.

In this study, two cases will be tested.

- **Two-dimensional case:** It is assumed that only the substrates maize and pig manure are available. The resulting optimization problem is that of finding the best mixture of both. The low dimensionality allows for visual analysis thus providing an intuitive understanding of the problem.

- **Five-dimensional case:** Here, three additional substrates are available, namely cow manure, grass and corn-cob-mix. This is a realistic scenario for many agricultural biogas plants.

The exact limits of the optimized parameters are summarized in Table 1.

Table 1: Lower and upper boundaries for the optimized parameters, that is the amount of substrates in the mixture.

Substrates	$\boldsymbol{x}_{lo}$ $[\frac{\mathrm{m}^3}{\mathrm{d}}]$	$\boldsymbol{x}_{up}$ $[\frac{\mathrm{m}^3}{\mathrm{d}}]$
Maize	5.00	40.00
Pig Manure	5.00	60.00
Grass Silage	0.00	20.00
Corn-Cob-Mix	0.00	10.00
Cow Manure	0.00	10.00

3.4. The Coarse Function

The coarse, more simple objective function is mostly based on the biomethane potential of each substrate, hence called *biomethane potential* (BMP) model. The BMP can be calculated for each substrate using the Buswell equation [12].

Thus, the BMP model estimates that the amount of produced gas rises linearly with the amount of each substrate feed into the plant. The estimate of produced energy is limited by the maximum load of the block heat and power plant. One evaluation takes two hundred microseconds or less. Hence, ten thousands of coarse function evaluations could be made during one evaluation of the fine objective function.

The BMP model is able to yield basic information like amount of methane gas produced or daily monetary gain. However, it can not yield the complete set of process variables that are available with the ADM1 and is less accurate.

3.5. Advantages of the Coarse Function

In the case of this application, the optimization process can profit in two different ways from the data available in form of the coarse function. First, the low fidelity models optimum can be used to enhance the initial experimental design created by SPOT, or used as a starting guess for non-set-based approaches. Second, the surrogate model of the global landscape can be enhanced by the low-fidelity model, e.g., using Co-Kriging or similar methods.

3.6. Model Implementation, Complexity and Time-Constraints

The Co-Kriging approach taken here is independent of the actual implementation of the fine and crude objective function. Thus, an even more detailed simulation model could be used for the fine objective function. For the sake of this study, we investigate an instance of the ADM1 with medium complexity, which is (relatively) fast to evaluate. This allows us to study a wider range of methods and perform a more in-depth analysis. This can be considered as a benchmark for more complex cases. Various different implementations of the ADM1 do exist and are being developed. A recent implementation [17] contains many more variables, compared to the standard ADM1 that we use. Most models assume that the digester is perfectly mixed, but there are also approaches

that separate the digester into different zones, each of them modeled by one ADM1 instance [46]. But, the ADM1 can also be implemented as a CFD model [21]. Thus, depending on the model complexity one simulation can take several seconds but also several minutes or even hours.

Furthermore, the benchmark in this study does only take the optimization of the equilibrium state into account. In practice, substrate parameters as well as boundary constraints are not constant. Thus, the optimal substrate feed is only optimal for a short time (assuming that parameters and constraints can be approximated as piecewise constant). Therefore, the optimization problem has to be solved repeatedly, e.g., once per hour or once per day. Each time, the optimization problem is solved taking into account the current state of the biogas plant as well as current parameters and constraints. The faster this heavily time-constraint optimization problem can be solved, the smaller is the lag between the present moment and the moment the optimization problem was started. Such a real-time optimization scheme was developed in [22] and it is planned to integrate the methods proposed in this paper into the mentioned work.

4. Methods

4.1. Modeling

4.1.1. Co-Kriging

Kriging is a method for interpolation and regression based on Gaussian process modeling. The following notation is adopted from Forrester et al. [18]. Given a set of n solutions $\mathbf{X} = \{\mathbf{x}^{(i)}\}_{i=1\dots n}$ in a k-dimensional continuous search space with observations $\mathbf{y} = \{y^{(i)}\}_{i=1\dots n}$, Kriging is a method to find an expression for a predicted value at an unknown point by interpreting the observed responses $\mathbf{y}$ as if they are realizations of a stochastic process. The following set of random vectors $\mathbf{Y} = \{Y(\mathbf{x}^{(i)})\}_{i=1\dots n}$ is used to define this stochastic process. The correlation of the random variables $Y(\cdot)$ is modeled as follows [18]:

$$\operatorname{cor}\left[Y(\mathbf{x}^{(i)}), Y(\mathbf{x}^{(l)})\right] = \exp\left(-\sum_{j=1}^{k} \theta_j |x_j^{(i)} - x_j^{(l)}|^{p_j}\right). \quad (3)$$

The matrix that collects correlations of all pairs $\{(i, l)\}$ is called the correlation matrix $\mathbf{\Psi}$. It is used in the Kriging predictor

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \boldsymbol{\psi}^T \mathbf{\Psi}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \quad (4)$$

where $\hat{y}(\mathbf{x})$ is the predicted function value of a new sample $\mathbf{x}$, $\hat{\mu}$ is the maximum likelihood estimate of the mean and $\boldsymbol{\psi}$ is the vector of correlations between training samples $\mathbf{X}$ and the new sample $\mathbf{x}$. The width parameter $\boldsymbol{\theta} = \left(\theta_1, \dots, \theta_j, \dots, \theta_k\right)^T$ determines how far the influence of each sample point $\mathbf{x}$ spreads. The parameter p_j is usually fixed at $p_j = 2$, and defines the shape of the correlation function.

As an extension of Kriging, Co-Kriging may include information of a coarse function into the model. To that end, Co-Kriging exploits correlation between the different fidelity levels. According to Forrester et al. [19], Co-Kriging can be understood to regress the coarse function while coinciding with the fine function. We now have two vectors with n_f samples from the fine function and n_c samples from the coarse function, i.e., $\mathbf{X}_f = \{\mathbf{x}_f^{(j)}\}_{j=1\dots n_f}$ and $\mathbf{X}_c = \{\mathbf{x}_c^{(i)}\}_{i=1\dots n_c}$ in a k-dimensional continuous search space with observations $\mathbf{y}_{cf} = \{y_c^{(i)}, y_f^{(j)}\}_{i=1\dots n_c, j=1\dots n_f}$. Accordingly, the stochastic process can now be defined by the set of random vectors $\mathbf{Y}_{cf} = \{Y_c(\mathbf{x}_c^{(i)}), Y_f(\mathbf{x}_f^{(j)})\}_{i=1\dots n_c, j=1\dots n_f}$. Then, we obtain the covariance matrix $\mathbf{C} =$

$$\begin{pmatrix} \sigma_c^2 \mathbf{\Psi}_c(\mathbf{X}_c, \mathbf{X}_c) & \rho\sigma_c^2 \mathbf{\Psi}_c(\mathbf{X}_c, \mathbf{X}_f) \\ \rho\sigma_c^2 \mathbf{\Psi}_c(\mathbf{X}_f, \mathbf{X}_c) & \rho^2\sigma_c^2 \mathbf{\Psi}_c(\mathbf{X}_f, \mathbf{X}_f) + \sigma_d^2 \mathbf{\Psi}_d(\mathbf{X}_f, \mathbf{X}_f) \end{pmatrix}, \quad (5)$$

where we have the same correlation function, but with two sets of model parameters for $\mathbf{\Psi}_d$ and $\mathbf{\Psi}_c$ respectively. An additional parameter ρ is introduced as a constant scaling factor. While $\mathbf{\Psi}_c$ does represent the correlation structure in the coarse function, $\mathbf{\Psi}_d$ captures the difference between the Gaussian process representing the cheap function (scaled by ρ) and the unknown Gaussian process representing the fine function. The Co-Kriging predictor for the fine function is

$$\hat{y}_f(\mathbf{x}) = \hat{\mu} + \mathbf{c}^T \mathbf{C}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \quad (6)$$

where $\mathbf{c}$ is the vector of covariances between the known solutions (fine and coarse) $\mathbf{X}_{cf} = \begin{pmatrix} \mathbf{X}_f \\ \mathbf{X}_c \end{pmatrix}$ and the solution to be predicted $\mathbf{x}$ and $\mathbf{1}$ denotes a vector of ones. We refer to Forrester et al. [19] for further information. The model we use in this work is a re-implementation in R based on MATLAB® code of Forrester et al. [18].

Two experimental designs are evaluated for Co-Kriging, one is a large experimental design which covers the design space of the coarse function. Hence, all points in this design are evaluated with the coarse function. The second design is the smaller set of points evaluated on the fine function. It is nested in the larger design, that means, each point evaluated on the fine function is also evaluated on the coarse function. Both designs should respect some criterion of space-fillingness. In this work, we create Latin Hypercube Designs (LHDs) [40] that maximize the minimum distance

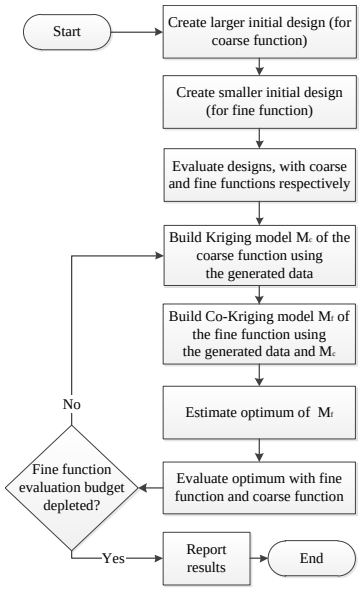

Figure 1: A flow chart to visualize the algorithmic procedure of optimization with a Co-Kriging surrogate model.

between the samples. That means, the search space is divided into subsections with a regular grid. Sample points are placed inside grid cells, such that only one sample is in each row and column of the grid. Inside a grid segment, sample points are placed at random, using a uniform distribution. This procedure is repeated h times, hence generating h LHDs. The minimum pairwise distance of each LHD is calculated, and the LHD with the largest distance is chosen. This helps to avoid the off chance of placing sample points too closely, in case they are in neighboring grid cells.

A flow chart of model based optimization with Co-Kriging is depicted in Fig. 1.

4.1.2. Alternative multi-fidelity Models

Several simplified alternatives to Co-Kriging can be used to integrate information from both coarse and fine function into the modeling process. The following methods are all compared to Co-Kriging in a preliminary investigation of model quality.

Diff. Model A very intuitive idea is to assume that the coarse function is able to model the general structure correctly. The remaining error can then simply be corrected by modeling the difference between coarse and fine function (f_c, f_f). The residuals of the coarse function are used as training data for a data-driven model. That is, the surrogate model $\widehat{\mathcal{M}}$ is built with design $\mathbf{X}$ and observations $f_f(\mathbf{X}) = \mathbf{y}_f, f_c(\mathbf{X}) = \mathbf{y}_c$ and

$$\widehat{\mathcal{M}}_{diff} = f_{\mathcal{M}}\left(\mathbf{X}, \mathbf{y}_f - \mathbf{y}_c\right). \tag{7}$$

Here, $\mathcal{M}(.)$ is function that builds the respective surrogate model. A new prediction is then always based on the result of the model, as well as the result of the coarse function:

$$\hat{y}_f = \hat{f}_f(\mathbf{x}) = \hat{f}_{diff}(\mathbf{x}) + f_c(\mathbf{x}), \tag{8}$$

where $\hat{y}_f$ is the predicted fine function value and $\hat{f}_{diff}$ represents the prediction of the surrogate model $\widehat{\mathcal{M}}_{diff}$. This model will be referred to as the *Difference Model* (Diff. Model). It is essentially the same as the additive scaling approach mentioned in Sec. 2.4.1.

Ratio Model The Ratio Model works in a very similar way. Instead of differences, i.e., residuals, the ratio between fine and coarse function is modeled.

$$\widehat{\mathcal{M}}_{ratio} = f_{\mathcal{M}}\left(\mathbf{X}, \frac{\mathbf{y}_f}{\mathbf{y}_c}\right) \tag{9}$$

The prediction of the Ratio Model $\widehat{\mathcal{M}}_{ratio}$ is given as

$$\hat{y}_f = \hat{f}_f(\mathbf{x}) = \hat{f}_{ratio}(\mathbf{x}) f_c(\mathbf{x}). \tag{10}$$

It is essentially the same as the multiplicative scaling approach mentioned in Sec. 2.4.1.

Input Model The input model takes a slightly different approach. The response of the coarse target function is used as an additional input parameter of the model, e.g., Kriging.

$$\widehat{\mathcal{M}}_{input} = f_{\mathcal{M}}(\{\mathbf{X}, \mathbf{y}_c\}, \mathbf{y}_f), \tag{11}$$

with prediction:

$$\hat{y}_f = \hat{f}_f(\mathbf{x}) = \hat{f}_{input}(\{\mathbf{x}, f_c(\mathbf{x})\}). \tag{12}$$

These three simple approaches can all be applied to arbitrary models, e.g., Neural Networks or Support Vector Machines. They require the coarse function during prediction. Of course, if the coarse function itself is somewhat costly, although cheaper than the fine function, it can again be replaced by a separate surrogate model.

4.1.3. Two-layer Modeling

One problem in surrogate modeling of biogas plants is that the modeled landscape is not continuous, as illustrated in Fig. 2. In this example, the *actual gain* function has a saltus at $x = 30\%$. In fact, the optimum is often in the vicinity of a saltus in decision space, which can also be seen in Fig. 3. This behavior is caused by the so called manure bonus, which is a fixed bonus paid to biogas producers. This bonus is paid, if more than 30% of the substrate contains specific manures [9].

Models like Kriging are best suited for continuous landscapes. To some extent, they are able to deal with discontinuities, at least globally. Still, a Kriging model will always deteriorate in regions close to the discontinuities. This is especially problematic due to the fact that the optimum may often be close to the 30 percent bonus limit. The model quality would therefore be deteriorated in the vicinity of the optimum. To avoid this problem, two approaches are eligible for this application.

1. While the modeled landscape is that of the actual monetary gain, the simulation does provide additional information. This could be exploited by modeling the exact amount of produced gas (e.g., with Co-Kriging), and calculating the monetary gain on-the-fly during prediction. The amount of produced gas would be continuous over the whole design space, thus yielding a reasonable surrogate model. The drawback is, that at least two models would need to be trained: the first, which models the amount of produced methane gas, and the second, which models further results from the ADM1 simulation which affect the gain of the plant. Also, the on-the-fly calculation of the monetary gain would be added on top of the effort of the prediction during the surrogate-optimization process.

2. The alternative is, to create two Kriging models for the monetary gain, one approximating monetary gain without manure bonus and one with the manure bonus. During surrogate-optimization the optimizer will switch between the two models, depending on whether the 30 percent bonus limit is reached. The two layers are illustrated in Fig. 2.

The former approach would be more time consuming, since it needs to calculate the monetary gain for each predicted sample. Also, each model would have to predict each sample, because both values are needed for the monetary gain calculation. The latter approach would only require to take the 30 percent bonus limit into account, to switch between models, without any further calculations. The drawback would be the loss of information, since the former approach is able to give an estimate of the produced amount of gas to the interested user. In this study, it was decided to take the less informative but more efficient approach two. We refer to this as the Two-layer approach, due to the two different models of the monetary gain. The difference in model quality will be investigated in a preliminary study in Sec. 5.

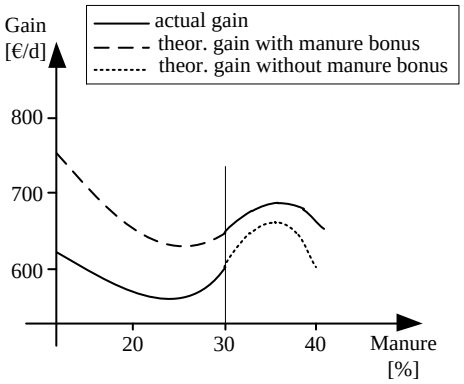

Figure 2: Illustration of the two different modeling layers in the Two-layer surrogate model. Here, only the percentage of manure for a fixed amount of other substrates is assumed to vary. The discontinuity in the curve arises at exactly 30 percent manure.

Essentially, this approach is comparable to the combined meta-modeling of discrete (logic) and continuous functions proposed by Meckesheimer et al. [41]. Our approach would be best represented by the case where the meta-modeling only encompasses the continuous responses, whereas the logic/decision function does not need to be approximated due to its low evaluation cost.

Please note, that the earlier described bump in the decision space is not a constant offset. The manure bonus which causes this discontinuity affects the revenue from sold gas, thus having a multiplicative influence on a single part of the objective value calculation. The impact of gas pricing on the overall gain varies significantly in the given search space.

4.2. Error Measure

Two error measures will be used in the model quality experiments. The *Mean Squared Error* (MSE) of the vector of n observations $\mathbf{y} = (y_1, \ldots, y_i, \ldots, y_n)^T$ and the vector of corresponding predictions $\hat{\mathbf{y}} = (\hat{y}_1, \ldots, \hat{y}_i, \ldots, \hat{y}_n)^T$

$$\mathrm{MSE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{13}$$

The second error measure is the *Scaled MSE* (SMSE) as introduced by Keijzer [34]. Keijzer [34] defines the SMSE as follows:

$$\begin{aligned} \mathrm{SMSE}(\mathbf{y}, \hat{\mathbf{y}}) &= \mathrm{MSE}(\mathbf{y}, \mathbf{1}a + b\hat{\mathbf{y}}) = \\ &\overset{(13)}{=} \frac{1}{n}\sum_{i=1}^{n}(y_i - (a + b\hat{y}_i))^2 \\ \text{where} \quad b &= \frac{\mathrm{cov}(\mathbf{y}, \hat{\mathbf{y}})}{\mathrm{var}(\hat{\mathbf{y}})} \quad \text{and} \quad a = \bar{\hat{\mathbf{y}}} - b\bar{\mathbf{y}} \end{aligned} \tag{14}$$

Here, $\bar{\hat{\mathbf{y}}}$ and $\bar{\mathbf{y}}$ indicate the respective mean values. The SMSE can be understood to evaluate differences between two vectors after scaling them to a common range. That allows to ignore errors that are irrelevant to the optimization procedure. A simple example would be a prediction $\hat{\mathbf{y}}$ that differs from the observations $\mathbf{y}$ only by a constant offset. While this prediction would receive a comparatively large MSE value, the location of the optimum would be perfectly accurate. SMSE is considered as the adequate error measure in this case. Section 5 will further motivate this choice for the biogas application, showing preliminary results where SMSE is clearly the more reasonable indicator.

4.3. Handling Evaluation Failures

Simulations of the expensive, fine target function may fail and lead to unreasonable or missing results. These failures are consistent, i.e., a repeated evaluation of a failed configuration will fail again. Such points can not always be ignored. Ignoring them could lead to a situation were the optimization process would repeatedly suggest an instable configuration. This would possibly prevent the optimization progress. Instead, only failed simulations in the initial design are removed.

Failed simulations in the optimization process itself are replaced by the imputation method suggested by Forrester et al. [18]. That means, instead of the missing simulation result $\hat{y}(\mathbf{x}) + \hat{s}(\mathbf{x})$ is imputed. Here, $\hat{y}(\mathbf{x})$ is the prediction of the (Co-)Kriging model and $\hat{s}(\mathbf{x})$ is the uncertainty estimate, which is used as a penalty. In a sense, we impute some upper confidence bound, as estimated by the Kriging model.

Naturally, such a method is only possible where a variance estimate of the model is available. All non-model-supported approaches use a fixed penalty value. That means, they receive a constant value that represents bad quality whenever the simulation fails.

4.4. Optimization Algorithms

Three different optimization approaches are employed in this study, either to optimize the fine/coarse function directly, or to optimize the corresponding surrogate models.

- **Downhill Simplex Method** (Simplex) is a classic, derivative free, local optimization method developed by Nelder and Mead [45]. For the experiments in this paper a bound constrained Simplex [10] implementation from the NLopt library [31] is used, interfaced by the R-package *nloptr* [57].
- **Differential Evolution** (DE) [49, 50] is a state-of-the-art derivative free optimization method based on the principles of evolution. Due to being population-based and due to its stochastic nature it has the capability (although no guarantee) to leave local optima. The R-implementation in the package *DEoptim* [3, 43] was used in the experiments.
- **Latin Hypercube Sampling** (LHS) is a very simple optimization strategy, where a space filling Latin Hypercube Design (LHD) of experiment is evaluated in the decision space of the optimization problem. The best found solution in this design is the estimated optimum. As with all LHD instances in this study, the design creation uses $h = 100$ repeats for the selection of a design with largest minimum pairwise distance.

5. Preliminary Study: Model Quality

A first study was performed to analyze how well the different surrogate-models represent the problem landscape, i.e., testing for modeling error. To get a simple and understandable example, we restrict this Pre-Study to the two dimensional case. Only the substrates pig manure and maize are assumed to be available.

Three questions are of interest:

1. What error measure should be used?
2. Does the Two-layer modeling approach improve model quality?
3. Which (multi-fidelity-) modeling method, e.g.., Kriging, Co-Kriging, diff or ratio model, works best?

Please note, that approximately five percent of randomly chosen substrate-mixtures may yield simulation failures due to numerical instability. In this preliminary study, such points are ignored, i.e., manually removed. A more complicated imputation as described in Sec. 4.3 is only employed during the later optimization experiments in Sec. 6.

5.1. Experimental Setup

We perform four sets of experiments, where each represents a different size of the experimental design. In each set, two design types are created and evaluated. Type one is a smaller LHD (size 5, 10, 15 and 20 points) evaluated with the fine function f_f. Type two is a very large LHD (always 100 points) evaluated with the coarse function f_c. The derived information is used to build surrogate models of the fine function, i.e., $\hat{f}_f$. In both cases, $h = 100$ LHDs are created, and the one with the largest minimum pairwise distance chosen.

To evaluate performance, we will look at the earlier introduced error measures (MSE, SMSE). The considered surrogate models are standard Kriging, Co-Kriging and a Neural Network approach (Quantile Regression Neural Network QRNN [13]). QRNN and Kriging are also tested with the earlier introduced simple multi-fidelity approaches, that is, Input, Ratio and Difference modeling (see Section 4.1.2). QRNN is introduced as an alternative approach to determine whether certain observations are actually linked to the Multi-fidelity modeling approach, or rather to the employed model type.

It has to be noted that Co-Kriging is the most time-consuming method. Since the model building takes less than a second in any case, this is not significant in comparison to the cost of evaluation f_f. However, in the later optimization experiments runtime deserves more attention, since the time-consumption will sum up over all sequential optimization steps. Higher search space dimension will also play an important role.

For each combination of error measure, design size and chosen surrogate model, 20 repeats are performed. The modeling error is estimated based on data from a larger Latin Hypercube Design, consisting of 400 design points. This data set is not available during model training. This validation is only performed for the purpose of this research study and is not proposed to be performed under real-world time-restrictions. This split-sample approach with a large test-set allows to easily detect differences that are only present in small regions of the search space (e.g., the discontinuity in the search space or the vicinity of the optimum). This is the main reason why other popular model-independent error evaluation methods like cross-validation or predictive error measures [51, 55, 42] are not used here. These may be used for model selection where a large test set is not available, as they may not require additional function evaluations.

5.2. Results and Discussion

As a first result, Fig. 3 shows filled contour plots representing the ADM1 (f_f) and BMP (f_c) target functions,

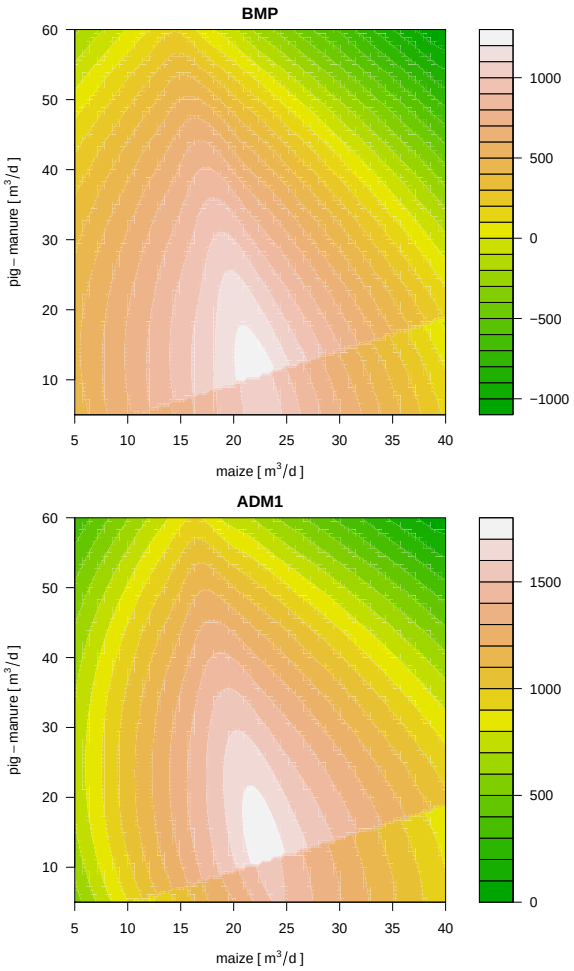

Figure 3: Contour plots of ADM1 and BMP model, based on 400 samples from a LHD. The two contour plots are both interpolated with Two-layer Kriging, using data from 400 samples generated with LHS. The depicted values are monetary daily gain in Euro, thus larger values are better.

respectively.

Here, both are interpolated with Two-layer Kriging, using data from 400 samples created with LHS. In the given region of interest, both show similar behavior. While the BMP has only a slightly different shape, a strong offset can be observed. Still, the optima of both functions are not to far from each other. They are also close to the discontinuity, hence the need for the Two-layer approach.

To show the influence of the five different substrates in the 5D optimization case, a sensitivity analysis is performed. One simple indicator of the sensitivity to a certain variable is the parameter θ_j (with $j = 1, ..., k$, see Eq. (3)), which specifies the width of the correlation function. This parameter can be interpreted to deter-

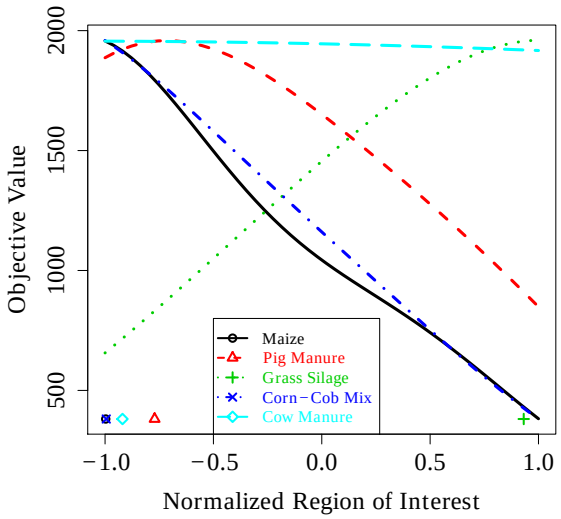

Figure 4: Sensitivity of the objective function value (gain in euro) to changes in each parameter (substrate) over the normalized parameter range. Values at −1 are the lower bound and 1 the upper bound of the respective parameter. When one parameter is varied, all other parameters are fixed to the best solution observed. The optimal substrate values are represented by the respective symbols close to the x-axis.

mine how active the modeled landscape is in the respective dimension. Table 2 shows estimated θ_j values for each substrate.

Table 2: Estimated θ_j values for a model based on 400 evaluations of the fine function (ADM1). Larger values indicate stronger activity of the respective variable.

Maize	Pig Manure	Grass Silage	Corn-Cob Mix	Cow Manure
3.85	1.72	1.58	2.29	0.02

It can be observed that Maize has the strongest impact on the objective value, whereas Cow Manure seems to have the least impact. As a further indication of the parameters sensitivity, Fig. 4 presents the sensitivity along each dimension (biogas substrate). Here, all but one parameters are fixed at the value of the best known solution. The single remaining parameter is varied, and the respective function values are depicted in the plot. The plot verifies the results from the Table 2. For example, cow manure has little influence, as the objective value line is nearly horizontal in the plot. Maize and corn-cob mix on the other hand may account for very large changes, if they are varied.

The different scale of objective values visible in the

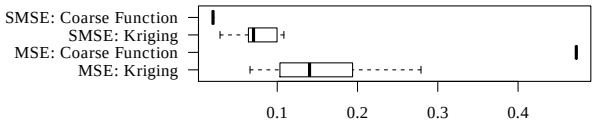

Figure 5: Depicted are SMSE and MSE of a Kriging model of the fine function as well as the BMP model (coarse function) evaluated by two different error measures. The Kriging model was built based on a LHD of size 5. The process of creating the LHD design was repeated 20 times. Smaller values are better.

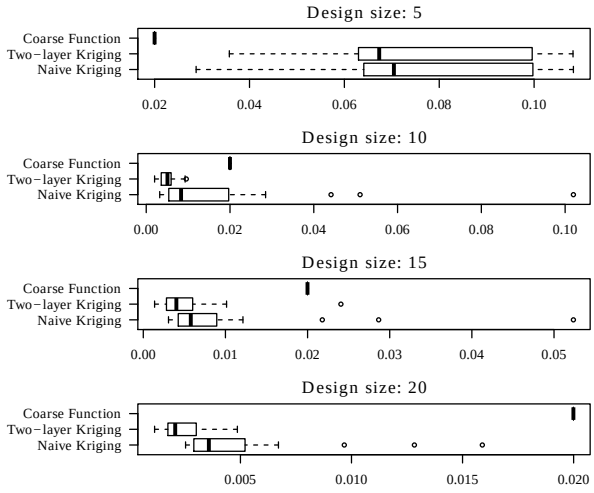

Figure 6: Again, all plots are based on LHD designs, repeated 20 times. Design sizes are in each header. The two-layer approach models values with and without manure bonus separately. Smaller values are better.

contour plots indicates that SMSE should be preferred over MSE. An unscaled error measure would not be a fair comparison, as the optimum of the coarse function is very close to the one of the fine one. Figure 5 shows how this choice affects the estimation of quality, comparing MSE and SMSE of the coarse function to a Kriging model of the fine function.

SMSE is better suited to evaluate the usefulness of the model for optimization purposes. As shown in Fig. 3, bad MSE values are caused by the saltus, although the location of the optimum is very well approximated.

Figure 6 shows how SMSE results vary depending on whether or not the Two-layer approach is used.

As expected, the model profits from using the Two-layer approach, as it avoids the discontinuity introduced by the manure bonus. The exception is the smallest design size of just five points. Here, no advantage is visible. Two reasons can be given. Firstly, the very sparse design of five points will yield such a poor model that the discontinuity becomes irrelevant. Secondly, it becomes unlikely that any of the five points is in the small

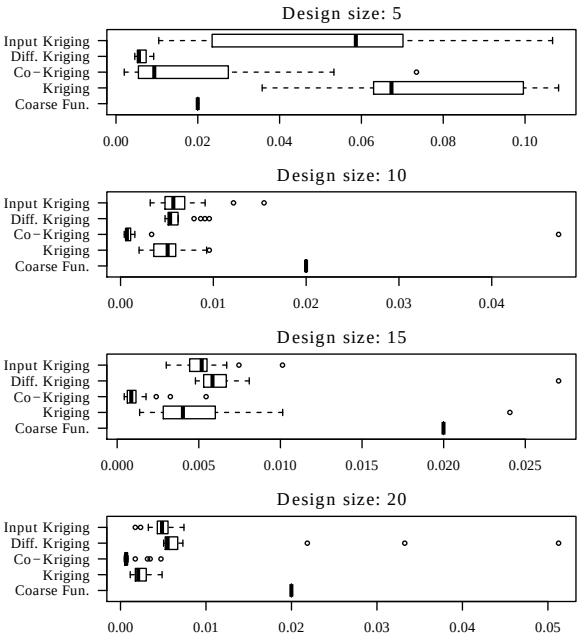

Figure 7: These plots show how different multi-fidelity methods, based on Kriging, perform in comparison to the coarse function and the single-fidelity Kriging model. Diff. Kriging models the differences between coarse and fine function. Input Kriging uses the coarse function values as an additional input variable. Smaller values for the SMSE are better.

area of the region of interest, where manure bonus is paid.

Figure 7 visualizes SMSE results of the different multi-fidelity, i.e., Two-layer, modeling approaches.

In all cases, Co-Kriging outperforms the standard Kriging model. Results for the Ratio model are not shown. Their results were worst, and including them in the figure would have made them hard to read. The Difference- and Input-based models seem to work for the smallest design size, but are later outperformed even by standard Kriging. This is despite the fact that the Difference-based model is the best for the smallest design size.

One peculiar observation in this context is, that the Difference-based model does not seem to profit from larger design sizes, whereas all other methods do. This can be seen more clearly in Fig. 8. The reason for this behavior is currently unclear and is subject of further research.

As a final comparison in this preliminary study, Fig. 9 shows how choosing a different model type would affect results.

For the smallest design size, Input QRNN outperforms Input Kriging, whereas the standard models show

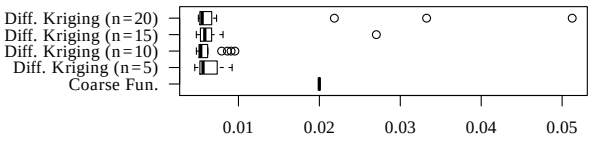

Figure 8: This plot uses the same data as Fig. 7 but for the Difference-based multi-fidelity approach only. Smaller values for the SMSE are better.

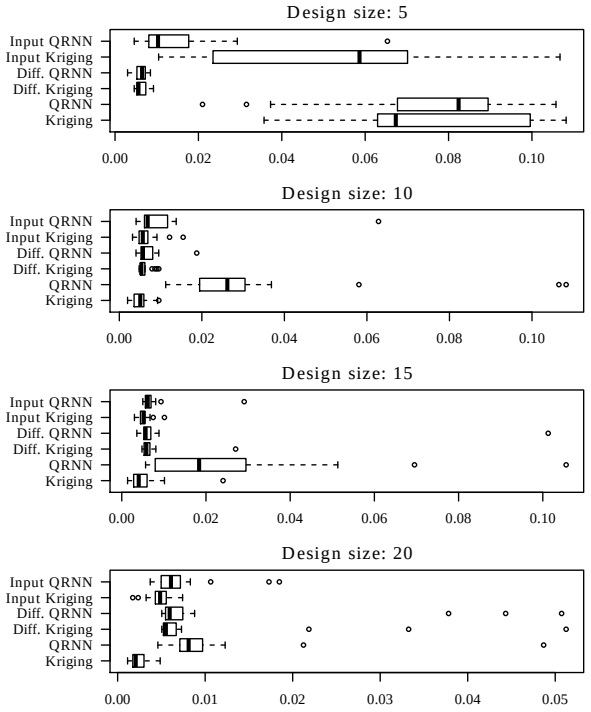

Figure 9: This plot compares how choosing a different model type (QRNN instead of single-fidelity Kriging) would affect results. Smaller values for the SMSE are better.

no striking difference. For all larger design sizes, standard Kriging clearly works best. The previous observation that the Difference-based approach does not profit from larger design sizes can be confirmed for the QRNN model, too.

6. Main Study: Optimization Performance

6.1. Experimental Setup

To run the optimization experiments, the R-package SPOT is used. Choosing parameters for the model-based optimization process in SPOT is a hard optimization problem. Tuning a model-based optimization algorithm requires large computational effort. In case of an expensive optimization problem this effort becomes too

large to warrant the potential benefit. Therefore, we restrict the choice of parameters to one fixed set for all experiments. Since the model-based approaches all share several of these parameters, it can at least be expected that the bias due to lack of tuning is not exceedingly large.

In the following list, k is the number of optimized substrates, that is, the dimension of the optimization problem. The list contains the important parameters which affect the optimization performance.

- **Choice of surrogate model:** The different models are Kriging and Co-Kriging. The implementations are taken from the R-package SPOT and are based on the MATLAB® code by Forrester et al. [18].
- **Surrogate optimizer:** While the objective landscape in the Pre-Study looked rather well-natured and uni-modal, there is no guarantee that this holds for any other scenario. Adding further substrates, or changing the costs of certain substrates, may easily introduce local optima. Furthermore, there is no guarantee that a surrogate model of the unimodal landscape is unimodal. Due to modeling errors, additional optima might be introduced. Therefore, it was decided to use DE to optimize the surrogate landscape. DE is well suited to solve multi-modal optimization problems. The used implementation is the DEoptim Package in R.
- **Surrogate optimizer population size:** Population size of the surrogate-optimizer (DEoptim) is chosen to be $10k$, which is the minimal suggested value according to the DEoptim package documentation.
- **Surrogate optimizer budget:** $500k$ evaluations of the surrogate model are allowed for each sequential step.
- **Number of Function evaluations (fine):** The number of evaluations of the fine function (ADM1) is limited to $5k$.
- **Design creation:** Experimental designs are created with LHS.
- **Initial Design size (fine):** From the $5k$ budget, $3k$ evaluations are used for the initial design.
- **Design Size (coarse):** $50k$ points are evaluated on the coarse function only. In addition, the coarse function is evaluated at every point where the fine function is evaluated, leading to an overall number of $55k$ evaluations of the coarse function at the end of an optimization run.
- **Infill criterion:** The chosen infill criterion is the mean predicted by the Kriging models. To make best use of the very small budget, expected improvement (EI) is not used, thus the optimization may focus on exploitation only. This results in a greedy strategy, i.e., exploitation dominates exploration of the search space. The benefit of this choice was validated in pre-experimental studies.

Based on the results from the preliminary study, it was chosen to test only Co-Kriging as the best performing multi-fidelity method. To get an estimate of the performance improvement this yields, Co-Kriging is compared against the basic Kriging approach. Both are tested once with and once without a fixed point in the initial design. That point is the optimum of the coarse model, which is cheap to determine.

Furthermore, two other methods were chosen as baselines in the comparison. Due to the strict budget limitations, DE was not considered here. Instead, Simplex and LHS are compared to the model-based approaches. The former is a good local optimizer, while the latter is more globally oriented. LHS can be viewed as an approach that has to be beaten by any reasonable search strategy, because it simply selects the best out of a set of randomly generated solutions. Simplex may be unbeatable by the competing approaches if the target function is of a sufficiently simple structure. Compared to the surrogate-supported approaches, LHS and Simplex have little overhead. To make use of the coarse function, LHS and Simplex use its optimum as a starting point.

As Simplex and LHS do not employ models, failures of the fine function evaluations will be compensated by using a constant penalty value of 5,000 €. Otherwise the imputation method described in Sec. 4.3 is used.

6.2. Analysis

Table 3 summarizes the parameters of the best solutions found in any experiment. In both 2D and 5D optimization, the optimal amount of manure is chosen to barely reach the required level for the manure bonus. When two substrates are optimized, the amount of maize is thus the largest part of the mixture, yielding a daily gain of roughly 1,770 € per day. Optimizing five substrates at the same time yields a better result of 1,987 €. Here, maize is mostly replaced with grass silage which is not that expensive and therefore has a better cost-benefit ratio. Nevertheless, one of the disadvantages of using a high amount of grass silage is

Table 3: Overall optimal solutions found.

Parameter	2D	5D
Gain [€/d]	1,770	1,987
Maize [m^3/d]	22.85	5.22
Pig Manure [m^3/d]	11.85	11.48
Grass Silage [m^3/d]	0	18.98
Corn-Cob-Mix [m^3/d]	0	0.01
Cow Manure [m^3/d]	0	0.08
Manure Bonus	yes	yes
Ammonia Digester [mg/l]	163.4	216.6
Ammonia Post-digester [mg/l]	291.0	464.2

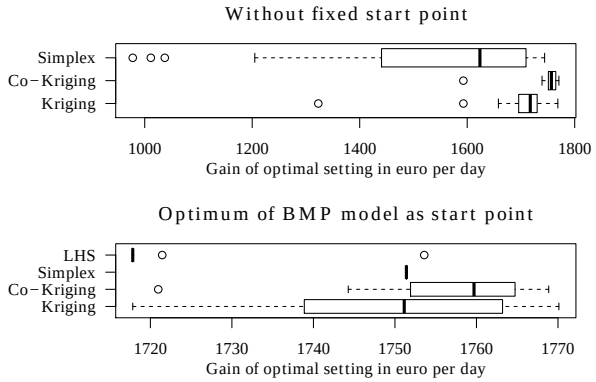

Figure 10: Final results of 2D optimization. Larger values are better. Results are based on 20 repeats of each optimization run. The upper plot shows results where the initial design (or starting guess) was created in a random fashion. For the lower plot, the initial guess (or at least one point in the initial design) was set to the optimum of the coarse objective function.

its high nitrogen content. In the digester the nitrogen is released and becomes ammonium and ammonia. As ammonia is toxic for the bacteria a too high concentration in the digester must be avoided. While this effect is respected by the ADM1, plant operators will still prefer mixtures resulting in lower ammonia concentrations. Hence, future studies may be conducted that consider ammonia as a second objective or as a constraint. A coarse model to calculate ammonia concentration would be to use the extended Buswell equation.

The performance of the different model-based and model-free approaches is compared for the 2D-case in Fig. 10. It can be seen that Co-Kriging works much better in comparison to the standard Kriging model. After the specified budget is exhausted, both outperform Simplex in the case where no start point is derived from the coarse function optimum. In this case, Simplex shows a very large variance caused by the random start point.

When the optimum of the coarse function is used to initialize the different methods, Simplex becomes completely deterministic. All methods perform better than LHS. The median of the plain Kriging approach is approximately the same as the Simplex performance. Co-Kriging is better than both.

Figures B.14 and B.15 in the appendix show the same results for the 2D case, but include results after each single evaluation. That way, the convergence progress can be observed. Note, that in this case, the LHS approach evaluates the promising candidate solution determined with the BMP model last. Similarly, the model based approaches evaluate the same candidate solution always after the non-deterministic candidates in the initial design. This is the reason for the respective lack of variance after ten (LHS with BMP) and six (Kriging/Co-Kriging with BMP) evaluations. Kriging requires about 9 and Co-Kriging 8 solutions to be competitive with the Simplex (all with BMP start point). While choosing a start point with BMP is essential for quicker convergence with the basic Kriging model, Co-Kriging seems to be less reliant on that. This can intuitively be attributed to the fact that Co-Kriging already has other ways of exploiting the BMP model information.

At a first glance, the situation does seem to be very similar for the larger search space of five parameters. The results of the 5D-case in Fig. 11 seem to show the same behavior as the 2D runs. Again, more detailed figures B.17 and B.16 in the appendix compare objective function values against number of evaluations. For this case, the advantage of choosing a start point with the BMP model seems to be more even more advantageous. All methods with random start points are not competitive. In contrast to the 2D-case the Co-Kriging based optimization now also profits more clearly from the start point determined by BMP. Compared to the basic Kriging model, the Co-Kriging model requires two expensive simulations less to receive similar results to the deterministic Simplex approach.

It is important to note, that Fig. 10 and Fig. 11 show results after an equal number of function evaluations, i.e., after the budget is expended. This does not take runtime into consideration. In case of the 2D-optimization there is nearly no difference in runtime between the different model-based approaches. This observation can not be made for the 5D case. Here, due to the much larger coarse design sizes, the model building becomes expensive. This can be seen in more detail from Tab. 4, which compares the net time used by the optimization algorithm runs in the 2D and 5D case. In fact, build-

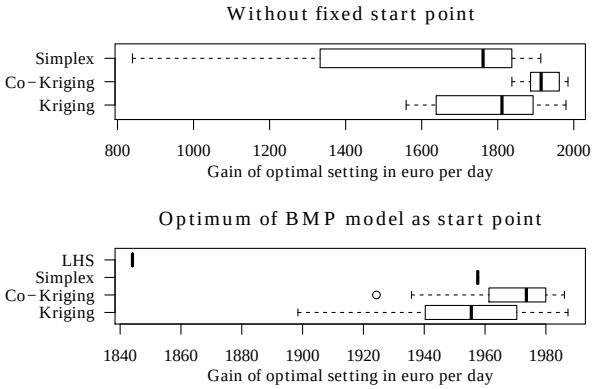

Figure 11: Boxplot of final results for the 5D optimization **with equal number of fine objective function evaluations**. Larger values are better. Results are based on 20 repeats of each optimization run. The upper plot shows results where the initial design (or starting guess) was created in a random fashion. For the lower plot, the initial guess (or at least one point in the initial design) was set to the optimum of the coarse objective function.

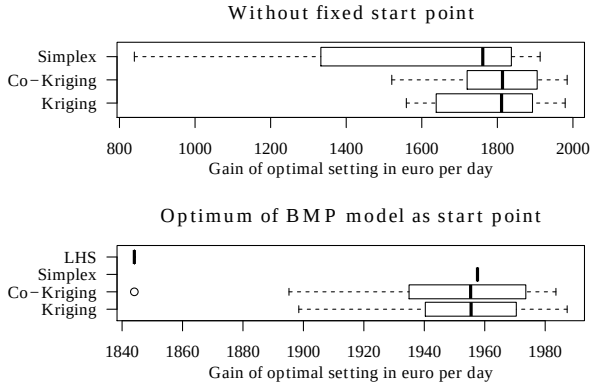

Figure 12: Boxplot of final results for the 5D optimization **with equal runtime**. Larger values are better. Results are based on 20 repeats of each optimization run. The upper plot shows results where the initial design (or starting guess) were created in a random fashion. For the lower plot, the initial guess (or at least one point in the initial design) was set to the optimum of the coarse objective-function.

ing and optimizing the Co-Kriging model now takes a similar effort as single evaluation of the fine function (ADM1).

Hence, Fig. 12 shows a comparison which is based on results after an equal runtime of roughly 25 minutes. While this does not affect the model-free optimization algorithms, Co-Kriging is not significantly different from the standard Kriging model any more.

The main reason for the difference in time-consumption is the large computational effort spend on building a model with data of 250 (initial design: $50k$, Section 6.1) evaluations of the coarse function, or more. This is not done with Kriging, which still remains comparatively cheap. Co-Kriging however suffers from a strong increase in time consumption. Careful tuning of parameters (e.g., design sizes) may avoid this issue, but would be extremely expensive and thus undesirable.

Finally, Table 4 summarizes the attained best objective function values (gain) and required gross and net times taken by the optimization algorithms as well as the number of failed simulations. Co-Kriging provides the best median performance for the 2D and 5D case. In the 2D case, it is also the fastest with respect to gross time, while it clearly has the worst net time. This contradiction can be explained by the fact that good substrate settings will usually require less simulation time[2].

In some cases, a single simulation of an inappropriate substrate mixture may take up to 10 minutes. Clearly, avoiding these cases seems to be worth the invested net time, at least in case of the 2D runs. At the same time, it has to be considered that the values of the Simplex runs with deterministic start point are based on a single run. Hence, they may even be outliers. This issue is stressed by the fact, that the deterministic Simplex run used much more gross time than the one with random start point (2D case). This single, deterministic Simplex run had one failed simulation (aborted after 600 seconds), which accounts for the difference to the non-deterministic runs. These do also have such outliers, but the depicted median value is not that much affected by those. On the other hand, the deterministic Simplex run in the 5D case did not encounter a failed simulation, hence yielding a gross time of barely more than 25 minutes.

The net time required is larger for the model based approaches, as expected. As stated earlier, the net time required by the Co-Kriging model is larger, but differences in case of the 2D problem are negligible. Still, the detrimental explosion of net time required by the Co-Kriging based 5D optimization runs becomes evident from the data in Table 4.

Overall, the results show both the advantages as well as the disadvantages of using multi-fidelity information. Essentially, the use of the coarse function is shown to be useful for deriving a reasonable start-point for the optimization. Also, integrating the coarse function by the means of Co-Kriging does increase performance, if measured on a per-evaluation basis. Despite this success, evaluating the results based on actual runtime re-

[2]The variance in simulation time is mostly due to the substrate mixture. Repetition of a simulation usually yields very similar simulation times. Simulation time and function values (gain) are negatively correlated, although only weakly.

Table 4: Results of all optimization approaches. Depicted are median values attained after 25 fine function evaluations, in 20 optimization experiments for each method. Only the `F` column is not a median, but the sum of all simulation failures over all experiments. In the method column, `BMP` means a deterministic start point (optimum of the BMP model) is used, whereas `ran` means a random start point. k is the dimensionality of the problem. The objective value Gain is in Euro per day. GT is the gross time taken by the optimization run (duration of complete run) in seconds. NT is the net time used by the optimization algorithm (gross time minus the time taken by all required ADM1 simulations) in seconds. Bold numbers are the respective best median for the 2D and 5D case. Note that Simplex (BMP) is deterministic, hence all values are based on a single optimization run.

method	k	Gain	GT	NT	F
Simplex (ran)	2	1554	898.24	0.37	1
Simplex (BMP)	2	1751	1451.61	0.41	0
LHS (BMP)	2	1553	975.49	**0.32**	1
Kriging (ran)	2	1702	909.30	10.68	2
Kriging (BMP)	2	1755	895.94	10.66	0
Co-Krig. (ran)	2	1757	861.50	15.20	2
Co-Krig. (BMP)	2	**1760**	**805.76**	15.08	0
Simplex (ran)	5	1760	2779.87	0.92	7
Simplex (BMP)	5	1958	**1594.88**	0.98	0
LHS (BMP)	5	1620	3156.73	**0.88**	0
Kriging (ran)	5	1808	2763.89	53.16	2
Kriging (BMP)	5	1963	2615.60	53.51	4
Co-Krig. (ran)	5	1917	2980.53	293.60	4
Co-Krig. (BMP)	5	**1968**	2877.09	288.97	1

veals an important issue. Co-Kriging becomes very slow, especially for higher dimensions and number of (coarse or fine) observations. Since the presented application of biogas substrate feed optimization is not extremely expensive, model and target function runtime may have the same order of magnitude. Hence, the resulting performance gain due to the Co-Kriging model is nearly negligible. In essence, such a trade-off between model and objective function runtimes can not be avoided with multi-fidelity approaches. Clearly, the whole purpose of using multi-fidelity models is to increase performance by use of additional information. In turn, this will always lead to increased complexity of the derived model. Whether or not such a model is useful will necessarily depend on the runtime of the objective function.

7. Summary and Outlook

The experiments showed that the employed approaches can be successful in building cheaper yet quite accurate models of the concerned biogas plant optimization problem. Therefore, Question (Q-1) can be affirmed. In detail, Co-Kriging based on cheap evaluations from a basic biomethane potential model was shown to improve the model quality, compared to a standard Kriging model based on evaluations of an accurate ADM1 based simulation model only. Furthermore, model quality could be improved by using a Two-layer approach, thus avoiding discontinuities in the searched landscape.

Besides these successes in improving surrogate model quality, their application in an optimization process proved to be more difficult. In detail, time consumption of Co-Kriging proved to be too large in case of a 5-Dimensional optimization problem. On the other hand, Co-Kriging was still successful in the case of a 2-Dimensional problem formulation. The core issue here is the trade-off between computational effort of the objective function (ADM1) and the surrogate-model optimization procedure. Should the optimization be subject to tighter time-constraints or a more expensive objective function, Co-Kriging based optimization may be profitable even in the higher-dimensional case. This depends on the kind of control problem to be solved in practice.

Apart from that, the integration of cheaply available data was shown to be very profitable. That is, all tested methods benefited from using a coarse representation of the target function to generate a promising initial solution. Using this initialization method, a simple Downhill Simplex method was shown to be as efficient as the more complex model based approaches. This comparison does however require to be investigated in more detail. Since Downhill Simplex, with a fixed start point, is completely deterministic, while the model-based approaches are not, the comparison does not provide any information on statistical significance. Without a fixed starting point, the surrogate-model-based approaches proved to be much more efficient than the Simplex method. Summarizing, detailed answers to Question (Q-2) were presented in this study.

For future research, it is therefore commendable to use test cases with larger variety. The herein presented research is based on a simulation for a specific combination of plant parameters, substrate costs, substrate parameters and environmental parameters. For other plants, the initialization with the cheaper BMP model may not work that well, or the searched landscape can become more complex.

As mentioned previously, the larger amount of ammonia in the sludge may make the obtained optimum for the 5D optimization undesirable to plant operators. The amount of ammonia can therefore be included as a constraint or even as a secondary objective. In both cases,

it may again be modeled with data-driven approaches like Kriging, to cheapen the evaluation of the constraint or secondary objective, respectively. In the same way, there are other details of additional parameters which may require to be included. In previous studies [56, 60] these were included in a weighted sum of objectives. Since the weights are hard to set, a multi-objective approach may make more sense. Of course, the computational effort of a surrogate-model-based approach will increase, since a separate model has to be built for each objective. At the same time, classical approaches like the Simplex method are not applicable anymore.

Acknowledgments. This work has been kindly supported by the Federal Ministry of Education and Research (BMBF) under the grants MCIOP (FKZ 17N0311) and CIMO (FKZ 17002X11).

Appendix A. List of Symbols

$cost_{energ.}$ cost of energy used in plant operation (e.g., stirring of digester content, substrate transportation, heating of digester)

$cost_{subs.}$ cost of substrates

$\boldsymbol{C}$ covariance matrix used in the Co-Kriging model

$\boldsymbol{c}$ vector of covariances of a new sample with the existing samples known to the Co-Kriging model

f objective function

f_f fine (accurate, expensive) simulation of the objective function

f_c coarse (inaccurate, cheap) simulation of the objective function

h Parameter of the Latin Hypercube Sampling approach. Number of times Latin Hypercube designs are randomly created, before choosing the one with the least pair-wise distance of samples.

k dimensionality of the decision space, i.e., the number of biogas substrates

n number of observations / number of data points

m dimensionality of the system state space vector

$\widehat{\mathcal{M}}$ fit of a surrogate model

$f_{\mathcal{M}}$ function used to learn/build a surrogate model

$\boldsymbol{\psi}$ vector of correlations of a new sample with the existing samples known to the Kriging model

$\boldsymbol{\Psi}$ correlation matrix used in the Kriging model

$rev_{\textbf{elect.}}$ revenue from selling electrical energy produced in combined heat and power plants

$rev_{\textbf{therm.}}$ revenue from selling thermal energy produced in combined heat and power plants

ρ scaling parameter used in the Co-Kriging model

t time step

$\boldsymbol{\theta}$ parameter vector of the Gaussian correlation function

$\boldsymbol{p}$ parameter vector of the Gaussian correlation function

$\boldsymbol{x}$ input vector of length k, i.e., the optimized substrate feed values

$\boldsymbol{x}_{lo}$ lower boundary of the decision variables (substrate feed values)

$\boldsymbol{x}_{up}$ upper boundary of the decision variables (substrate feed values)

y observed objective function value

$\boldsymbol{y}$ vector of observed objective function values of length n

z state space vector of length n_z, i.e., state of the biogas plant

z_0 initial state of the biogas plant system

$\mathcal{Z}$ system state space

$\mathcal{X}$ input parameter space

Appendix B. Additional Figures and Tables

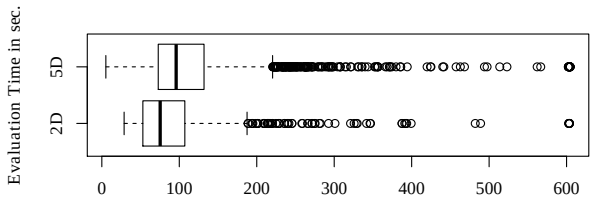

Figure B.13: Boxplots depicting the evaluation time of the ADM1 simulation (fine function) in seconds, split by the dimensionality of the problem (i.e., number of substrates).

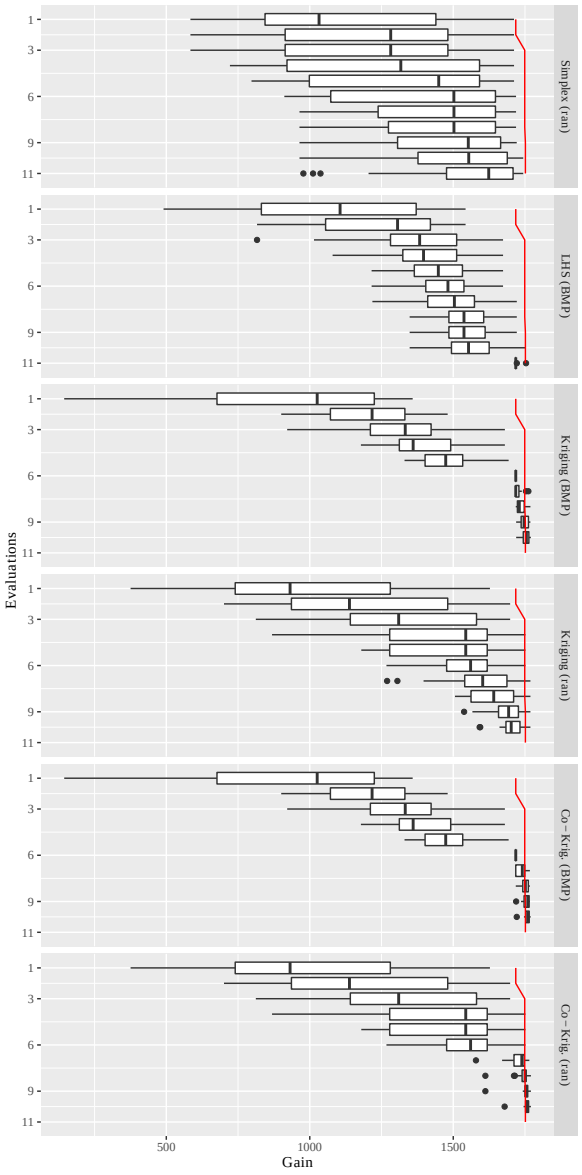

Figure B.14: Boxplot of 2D optimization performance after each evaluation performed. Larger values are better. Results are based on 20 repeats of each optimization run. The red line presents the performance of the simplex run with deterministic start point.

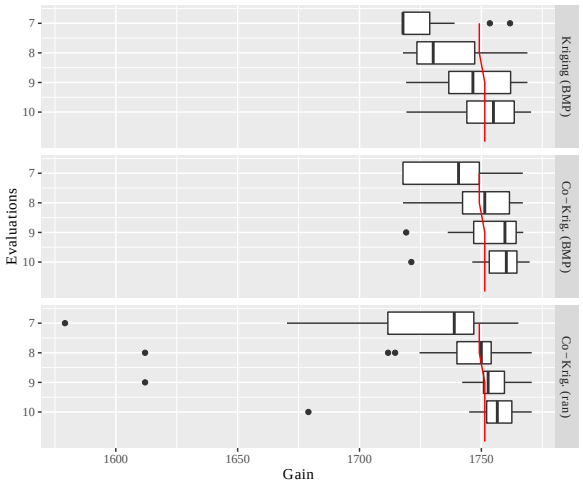

Figure B.15: Boxplot of 2D optimization performance after each evaluation performed, focusing on best performin methods after the initial 6 evaluations. Larger values are better. Results are based on 20 repeats of each optimization run. The red line presents the performance of the simplex run with deterministic start point.

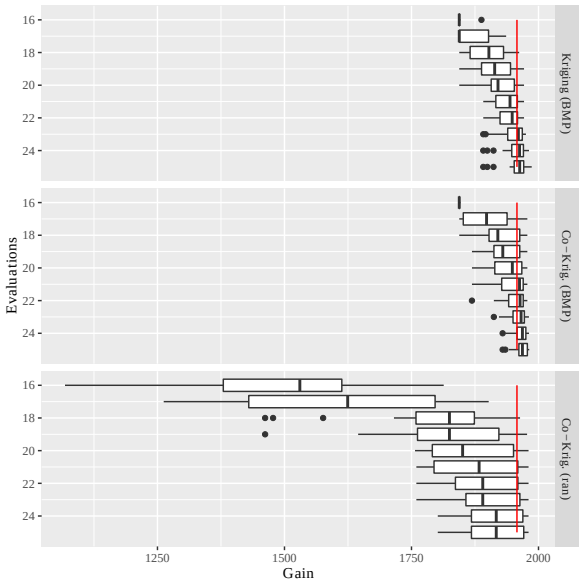

Figure B.16: Boxplot of 5D optimization performance after each evaluation performed, focusing on best performing methods after the initial 15 evaluations. Larger values are better. Results are based on 20 repeats of each optimization run. The red line presents the performance of the simplex run with deterministic start point.

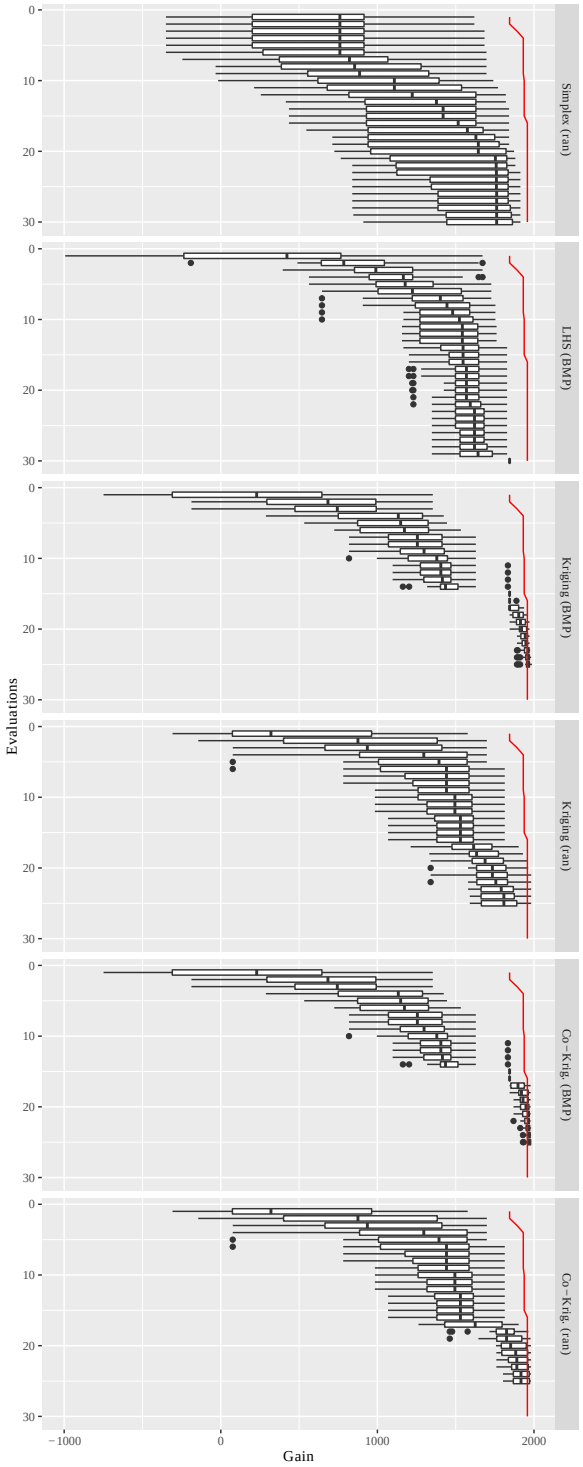

Figure B.17: Boxplot of 5D optimization performance after each evaluation performed. Larger values are better. Results are based on 20 repeats of each optimization run. The red line presents the performance of the simplex run with deterministic start point.

References

[1] Alexandrov, N. M., Dennis Jr, J. E., Lewis, R. M., Torczon, V., 1998. A trust-region framework for managing the use of approximation models in optimization. Structural Optimization 15 (1), 16–23.

[2] Alexandrov, N. M., Lewis, R. M., Gumbert, C. R., Green, L. L., Newman, P. A., 1999. Optimization with variable-fidelity models applied to wing design. Tech. rep.

[3] Ardia, D., Mullen, K. M., Peterson, B. G., Ulrich, J., 2013. DEoptim: Differential Evolution in R. Version 2.2-2.

[4] Bandler, J. W., Biernacki, R. M., Chen, S. H., Grobelny, P., Hemmers, R. H., et al., 1994. Space mapping technique for electromagnetic optimization. Microwave Theory and Techniques, IEEE Transactions on 42 (12), 2536–2544.

[5] Bartz-Beielstein, T., Parsopoulos, K. E., Vrahatis, M. N., 2004. Design and analysis of optimization algorithms using computational statistics. Applied Numerical Analysis and Computational Mathematics (ANACM) 1 (2), 413–433.

[6] Batstone, D., Keller, J., Steyer, J., Aug. 2006. A review of ADM1 extensions, applications, and analysis: 2002-2005. Water Sci Technol 54 (4), 1.

[7] Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, S. V., Pavlostathis, S. G., Rozzi, A., Sanders, W., Siegrist, H., Vavilin, V. A., 2002. Anaerobic Digestion Model No. 1, Scientific and technical report no.13. Tech. rep., IWA Task Group for Mathematic Modelling of Anaerobic Digestion Processes, London.

[8] Biernacki, P., Steinigeweg, S., Borchert, A., Uhlenhut, F., Brehm, A., 2013. Application of Anaerobic Digestion Model No. 1 for describing an existing biogas power plant. Biomass and Bioenergy.

[9] BMU, 2009. Gesetz für den Vorrang Erneuerbarer Energien (Erneuerbare-Energien-Gesetz - EEG).

[10] Box, M. J., Apr 1965. A new method of constrained optimization and a comparison with other methods. The Computer Journal 8 (1), 42–52.

[11] Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5 –32.

[12] Buswell, A. M., Mueller, H. F., Mar. 1952. Mechanism of Methane Fermentation. Industrial & Engineering Chemistry 44 (3), 550–552.

[13] Cannon, A. J., 2011. Quantile regression neural networks: Implementation in r and application to precipitation downscaling. Computers & Geosciences 37 (9), 1277–1284.

[14] Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., Vapnik, V., 1996. Support vector regression machines. In: Mozer, M., Jordan, M. I., Petsche, T. (Eds.), NIPS. MIT Press, pp. 155–161.

[15] Eldred, M., Giunta, A., Collis, S., Alexandrov, N., Lewis, R., 2004. Second-order corrections for surrogate-based optimization with model hierarchies. In: Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, NY,, Aug. pp. 2013–2014.

[16] Emmerich, M., Deutz, A., Klinkenberg, J., 2011. Hypervolume-based expected improvement: Monotonicity properties and exact computation. In: Evolutionary Computation (CEC), 2011 IEEE Congress on. IEEE, pp. 2147–2154.

[17] Flores-Alsina, X., Kazadi Mbamba, C., Solon, K., Vrecko, D., Tait, S., Batstone, D., Jeppsson, U., Gernaey, K., 2015. A plant-wide aqueous phase chemistry module describing pH variations and ion speciation/pairing in wastewater treatment process models. Water Research 85, 255–265.

[18] Forrester, A., Sobester, A., Keane, A., 2008. Engineering Design via Surrogate Modelling. Wiley.

[19] Forrester, A. I., Sóbester, A., Keane, A. J., Dec 2007. Multi-fidelity optimization via surrogate modelling. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 463 (2088), 3251–3269.

[20] Friedman, J. H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19 (1), 1–141.

[21] Gaden, D. L. F., 2013. Modelling Anaerobic Digesters in Three Dimensions: Integration of Biochemistry with Computational Fluid Dynamics. Doctor, The University of Manitoba.

[22] Gaida, D., 2014. Dynamic Real-Time Substrate Feed Optimization of Anaerobic Co-Digestion Plants. Doctoral thesis, Leiden University.

[23] Gaida, D., Wolf, C., Bongards, M., Bäck, T., April 2011. Matlab toolbox for biogas plant modelling and optimization. In: Progress in Biogas II - Biogas production from agricultural biomass and organic residues. Vol. 2. Stuttgart, pp. 67–70.

[24] Gano, S. E., Pérez, V. M., Renaud, J. E., Batill, S. M., Sanders, B., 2004. Multilevel variable fidelity optimization of a morphing unmanned aerial vehicle. In: AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference, Palm Springs, CA.

[25] Haftka, R. T., 1991. Combining global and local approximations. AIAA journal 29 (9), 1523–1525.

[26] Hansen, N., 2006. The CMA evolution strategy: a comparing review. In: Lozano, J., Larranaga, P., Inza, I., Bengoetxea, E. (Eds.), Towards a new evolutionary computation. Advances on estimation of distribution algorithms. Springer, pp. 75–102.

[27] Hansen, N., Ostermeier, A., 1996. Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. In: Evolutionary Computation, 1996., Proceedings of IEEE International Conference on. pp. 312–317.

[28] Islam, M. S., Islam, A., Islam, M., May 2014. Variation of biogas production with different factors in poultry farms of bangladesh. In: 3rd International Conference on the Developments in Renewable Energy Technology (ICDRET). pp. 1–6.

[29] Jin, R., Chen, W., Sudjianto, A., 2002. On sequential sampling for global metamodeling in engineering design. In: ASME 2002 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers, pp. 539–548.

[30] Jin, Y., 2005. A comprehensive survey of fitness approximation in evolutionary computation. Soft Computing 9 (1), 3–12.

[31] Johnson, S. G., 2010. The nlopt nonlinear-optimization package. `http://ab-initio.mit.edu/nlopt`, accessed: 2013-11-20.

[32] Jones, D., Schonlau, M., Welch, W., 1998. Efficient global optimization of expensive black-box functions. Journal of Global Optimization 13, 455–492.

[33] Južnič-Zonta, Z., Kocijan, J., Flotats, X., Vrečko, D., Nov. 2012. Multi-criteria analyses of wastewater treatment bio-processes under an uncertainty and a multiplicity of steady states. Water research 46 (18), 6121–31.

[34] Keijzer, M., 2004. Scaled symbolic regression. Genetic Programming and Evolvable Machines 5 (3), 259–269.

[35] Knowles, J., January 2006. Parego: A hybrid algorithm with online landscape approximation for expensive multiobjective optimization problems. IEEE Transactions on Evolutionary Computation 10 (1), 50–66.

[36] Knowles, J. D., Nakayama, H., 2008. Meta-modeling in multiobjective optimization. In: Multiobjective Optimization. Springer, pp. 245–284.

[37] Koziel, S., Ogurtsov, S., Couckuyt, I., Dhaene, T., November 2012. Cost-efficient electromagnetic-simulation-driven antenna design using co-kriging. Microwaves, Antennas Propagation, IET 6 (14), 1521–1528.

[38] Kuya, Y., Takeda, K., Zhang, X., J. Forrester, A. I., 2011. Multifidelity surrogate modeling of experimental and computational

aerodynamic data sets. AIAA journal 49 (2), 289–298.

[39] Lewis, R. M., Nash, S. G., 2000. A multigrid approach to the optimization of systems governed by differential equations. AIAA paper 4890, 2000.

[40] McKay, M. D., Beckman, R. J., Conover, W. J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.

[41] Meckesheimer, M., Barton, R., Simpson, T., Limayen, F., Yannou, B., 2001. Metamodeling of combined discrete/continuous responses. AIAA Journal 39 (10), 1950–1959, cited By 0.

[42] Mehmani, A., Chowdhury, S., Messac, A., 2015. Predictive quantification of surrogate model fidelity based on modal variations with sample density. Structural and Multidisciplinary Optimization 52 (2), 353–373.

[43] Mullen, K., Ardia, D., Gil, D., Windover, D., Cline, J., 2011. DEoptim: An R package for global optimization by differential evolution. Journal of Statistical Software 40 (6), 1–26.

[44] Mullins, J., Mahadevan, S., 2014. Variable-fidelity model selection for stochastic simulation. Reliability Engineering & System Safety 131, 40 – 52.

[45] Nelder, J. A., Mead, R., Jan 1965. A simplex method for function minimization. The Computer Journal 7 (4), 308–313.

[46] Ogurek, M., Alex, J., 2013. Überlegungen zur Modellierung mit dem ADM1. In: ifak Workshop: Steuerung, Regelung und Simulation von Biogasanlagen. ifak e.V., Leipzig.

[47] Page, D. I., Hickey, K. L., Narula, R., Main, A. L., Grimberg, S. J., Jan. 2008. Modeling anaerobic digestion of dairy manure using the IWA Anaerobic Digestion Model no. 1 (ADM1). Water science and technology : a journal of the International Association on Water Pollution Research 58 (3), 689–95.

[48] Ponweiser, W., Wagner, T., Biermann, D., Vincze, M., 2008. Multiobjective optimization on a limited budget of evaluations using model-assisted s-metric selection. In: PPSN. pp. 784–794.

[49] Price, K., 1996. Differential evolution: A fast and simple numerical optimizer. In: Proceedings NAFIPS'96. pp. 524–525.

[50] Price, K. V., Storn, R. M., Lampinen, J. A., January 2006. Differential Evolution - A Practical Approach to Global Optimization. Natural Computing. Springer-Verlag, iSBN 540209506.

[51] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., Tucker, P. K., 2005. Surrogate-based analysis and optimization. Progress in Aerospace Sciences 41 (1), 1 – 28.

[52] Schoen, M. A., Sperl, D., Gadermaier, M., Goberna, M., Franke-Whittle, I., Insam, H., Ablinger, J., Wett, B., Dec. 2009. Population dynamics at digester overload conditions. Bioresource technology 100 (23), 5648–55.

[53] Singh, G., Grandhi, R. V., 2010. Mixed-variable optimization strategy employing multifidelity simulation and surrogate models. AIAA journal 48 (1), 215–223.

[54] Storn, R., Price, K., 1997. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization 11 (4), 341–359.

[55] Viana, F. A. C., Haftka, R. T., Steffen, V., 2009. Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. Structural and Multidisciplinary Optimization 39 (4), 439–457.

[56] Wolf, C., McLoone, S., Bongards, M., 2008. Biogas plant optimization using genetic algorithms and particle swarm optimization. In: IET Irish Signals and Systems Conference (ISSC 2008). Institution of Electrical Engineers, pp. 244–249.

[57] Ypma, J., 2013. nloptr vers-0.9.6: R interface to nlopt. `http://cran.r-project.org/package=nloptr`, accessed: 2013-11-20.

[58] Zaefferer, M., Bartz-Beielstein, T., Friese, M., Naujoks, B., Flasch, O., July 2012. Multi-criteria optimization for hard problems under limited budgets. In: Soule, T., et al. (Eds.), GECCO Companion '12: Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference companion. ACM, Philadelphia, Pennsylvania, USA, pp. 1451–1452.

[59] Zaefferer, M., Bartz-Beielstein, T., Naujoks, B., Wagner, T., Emmerich, M., 2013. A case study on multi-criteria optimization of an event detection software under limited budgets. In: Evolutionary Multi-Criterion Optimization. Springer, pp. 756–770.

[60] Ziegenhirt, J., Bartz-Beielstein, T., Flasch, O., Konen, W., Zaefferer, M., Jul 2010. Optimization of biogas production with computational intelligence a comparative study. In: IEEE Congress on Evolutionary Computation. Institute of Electrical and Electronics Engineers, pp. 1–8.