



Generation of linear-based surrogate models from non-linear functional relationships for use in scheduling formulation

Obermeier, Andreas; Vollmer, Nikolaus; Windmeier, Christoph; Esche, Erik; Repke, Jens-Uwe

Published in:
Computers and Chemical Engineering

Link to article, DOI:
[10.1016/j.compchemeng.2020.107203](https://doi.org/10.1016/j.compchemeng.2020.107203)

Publication date:
2021

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Obermeier, A., Vollmer, N., Windmeier, C., Esche, E., & Repke, J-U. (2021). Generation of linear-based surrogate models from non-linear functional relationships for use in scheduling formulation. *Computers and Chemical Engineering*, 146, Article 107203. <https://doi.org/10.1016/j.compchemeng.2020.107203>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

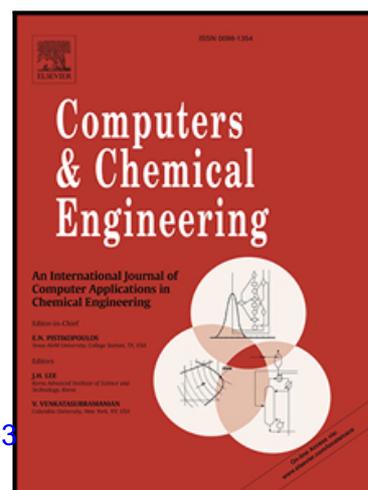
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Journal Pre-proof

Generation of linear-based surrogate models from non-linear functional relationships for use in scheduling formulation

Andreas Obermeier, Nikolaus Vollmer, Christoph Windmeier, Erik Esche, Jens-Uwe Repke

PII: S0098-1354(20)31246-1
DOI: <https://doi.org/10.1016/j.compchemeng.2020.107203>
Reference: CACE 107203



To appear in: *Computers and Chemical Engineering*

Received date: 5 July 2020
Revised date: 30 October 2020
Accepted date: 12 December 2020

Please cite this article as: Andreas Obermeier, Nikolaus Vollmer, Christoph Windmeier, Erik Esche, Jens-Uwe Repke, Generation of linear-based surrogate models from non-linear functional relationships for use in scheduling formulation, *Computers and Chemical Engineering* (2020), doi: <https://doi.org/10.1016/j.compchemeng.2020.107203>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

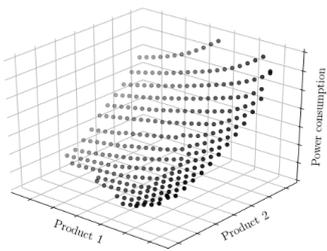
© 2020 Published by Elsevier Ltd.

Highlights:

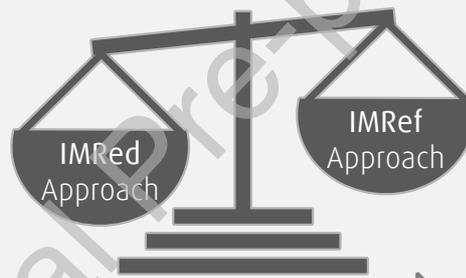
- Generation of data-based piecewise linear surrogate models from non-linear function
- Proposal and comparison of two approaches for generating such surrogate models
- Approach based on iterative edge contraction and error quadric is superior

Journal Pre-proof

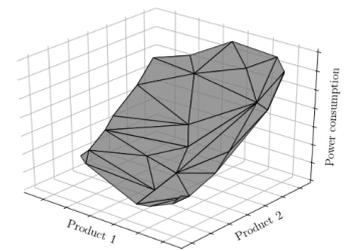
Generation of linear-based surrogate models from non-linear functional relationships for use in scheduling formulation



Data Set
capturing non-linearity and
feasible range



Data Reduction



Surrogate Model
approximation of non-linearity
and feasible range

Generation of linear-based surrogate models from non-linear functional relationships for use in scheduling formulation

Andreas Obermeier^{a,*}, Nikolaus Vollmer^b, Christoph Windmeier^a, Erik Esche^c, Jens-Uwe Repke^c

^aLinde GmbH, Linde Engineering, R&D Process Development, Dr.-Carl-von-Linde-Straße 6-14, D-82049 Pullach im Isartal, Germany

^bPROSYS Research Centre, Technical University of Denmark, Building 229, Søtofts Plads, 2800 Kgs. Lyngby, Denmark

^cProcess Dynamics and Operations Group, Technische Universität Berlin, Sekr. KWT-9, Str. des 17. Juni 135, D-10623 Berlin, Germany

Abstract

Often functional relationships are well known, but they are too complex to be used efficiently in optimization problems like scheduling formulations. Hence the functions are often replaced by data-based surrogate models. Especially, linear models are often used, since they are easier to solve than non-linear ones. The use of piecewise linear surrogate models allows for an improved consideration of nonlinearities. Although, the number of linear elements must be kept small in order not to lose the advantages of a linear-based formulation. In this work, two approaches for generating piecewise linear surrogate models are proposed, whereby the basic idea of both approaches is the determination of a reduced set of data points that provides an appropriate approximation of the original data via multi-dimensional linear interpolation. The approaches differ in their concepts: One is a numerical algorithm, the other an optimization-based technique. In this contribution, these approaches are described and subsequently compared.

Keywords: Data-based surrogate models, Piecewise linear, Data reduction

*Corresponding author

Email address: andreas.obermeier@linde.com (Andreas Obermeier)

1. Introduction

Data-based surrogate models are compact scalable analytic models that predict the continuous outputs of complex systems depending on various independent input variables, whereby the functional relationship is based on a limited set of data, originating from either computationally expensive simulations or real world applications. Therefore, such models represent a simplified description of complex systems and thus are suited to improve the computational performance of large-scale problems while doing parameter studies, sensitivity analyses, and optimization. Due to this ability, surrogate models have been applied in multiple scientific and engineering disciplines for many years. Simultaneously, various approaches have been proposed in literature to generate such simplified models. These include inter alia: i) Linear and polynomial regression (and variations thereof: ridge regression (Tikhonov & Arsenin, 1977), least absolute shrinkage and selection operator (Tibshirani, 1996), and elastic-net regression (Zou & Hastie, 2005)), ii) support vector machine (Vapnik, 1995), iii) gaussian process regression or also referred to as 'Kriging' (Krige, 1951; Bhosekar & Ierapetritou, 2018), iv) multivariate adaptive regression splines (Friedman, 1991), v) radial basis functions (Dyn et al., 1986; Fang & Horstemeyer, 2006), vi) random forest (Hastie et al., 2001; Breiman, 2001), vii) k-nearest neighbor (Hastie et al., 2001; Papadopoulos et al., 2011), viii) artificial neural network (Hastie et al., 2001; Arce-Medina & Paz-Paredes, 2009).

Due to the fact that surrogate modeling approaches are investigated for many years, there are several reviews (e.g., Wang & Shan (2007), Razavi et al. (2012), Yang et al. (2016) and Bhosekar & Ierapetritou (2018)), which give detailed explanations of the aforementioned approaches, as well as reports about current developments in this field. Therefore, this contribution skips a detailed overview of these approaches and refers to respective literature.

Applicability for scheduling formulations

Many of the listed approaches are commonly used to predict functional relationships according to the underlying data points, while having limited or no

knowledge about the detailed mechanisms of the approximated system. However, they can also be employed to generate data-based surrogate models as a simple description of a complex system. For instance, if a functional relationship between input variables \mathbf{x} and output \mathbf{y} is well-known, but too complex to be efficiently used in parametric studies, sensitivity analyses, or optimization formulations. In particular, by formulating optimization problems this kind of approach is often applied to solve the problem in feasible computational time, whereby the selected type of surrogate model mainly determines the kind of optimization problem (e.g., linear programming (LP), quadratic programming (QP), non-linear programming (NLP)). The kind of optimization problem in turn defines the effort to ensure global optimality, since only if the optimization problem is convex, a local optimum is also a global optimum. However, only LP problems are, per definition, always convex, whereas specific requirements have to be fulfilled in case of QP or NLP problems. Otherwise, non-linear optimization problems are likely to have several local optima and thus require greater effort to solve them to global optimality.

Due to the fact that linear models are significantly easier to solve than non-linear ones, linear regression is still a widely used approach to formulate models for optimization problems such as planning and scheduling formulations. This can be seen in a review of Zhang et al. (2016b), who lists several works applying planning and scheduling formulations for industrial demand side management in various energy-intensive industries such as aluminum, cement, chlor-alkali, steel, and air separation. The vast majority of the models reviewed here are formulated as mixed-integer linear programs (MILP) applying a linear description of the respective industrial process. As in this field of research especially the operational planning of cryogenic air separation plants has gathered a high degree of attention, in the following the main focus is placed on this specific sub-field. With this focus, in particular the works of Ierapetritou et al. (2002), Karwan & Kebliis (2007), Mitra et al. (2012), are to be mentioned, who introduced and further developed the concept of operating modes to allow for better modeling of the respective industrial process. This mode-based concept is used in many

works, often in combination with a linear process description (e.g., Mitra et al. (2013, 2014); Zhang et al. (2015)). In order to allow for an improved process description of the industrial process regarding its operational envelope (feasible region) and the performance within this envelope, Zhang et al. (2016a,b) extended this concept by the introduction of so-called Convex Region Surrogate (CRS) models, which allow the approximation of operating modes with non-convex feasible regions by a union of convex subregions. More specifically, the non-convex feasible region – spanned by several operation points – is partitioned into a set of convex polytopes such that the union of these polytopes describe the feasible region, whereby each polytope holds a linear process description. These CRS models represent a kind of piecewise linear description and thus allow for a better approximation of non-linearities in the process behavior, while keeping the computational complexity at an appropriate level. In this context, one of our previous contributions (Obermeier et al., 2019) is to be seen, which also employs a mode-based formulation and proposes a set of constraints to investigate the effect of mechanical fatigue on operational planning. Having the focus on this aspect, the feasible region of the process was considered in a reduced manner (pseudo two-dimensional product space). Nevertheless, this albeit simple process model also provides a piecewise linear approximation to allow for a more precise description of the process behavior.

In a recent review by Tsay & Baldea (2019), the distinction is made between dynamic and static data-driven approaches to generate surrogate models of process dynamics for production scheduling applications. Here, the generation of CRS models is assigned to the class of static approaches as the resulting models capture the steady-state process performance. While this class uses steady-state data, the approaches in the class of dynamic approaches apply transient process data in order to construct reduced-order dynamic models, which capture the process dynamics in scheduling calculations. As Tsay & Baldea (2019) mentioned, such reduced-order dynamic models can also be used in conjunction with multi-parametric programming (Oberdieck et al., 2016) for simultaneous scheduling and control (e.g., (Burnak et al., 2018)). When addressing this topic

with linearly constraint quadratic multi-parametric programming problems, a solution can be generated that describes an optimal scheduling and/or control
95 action by means of an explicit function. This function has again a piecewise affine structure that is valid in polyhedral partitions of the feasible parameter space, also known as critical regions.

In summary, it can be said that in the vast majority of cases linear process models are used in planning formulations since they are much easier to solve
100 than non-linear ones. In the recent years, however, the interest rises to improve the description of the actually non-linear process behavior by applying piecewise linear surrogate models, while keeping the computational complexity manageable. The formulation of CRS models may represent a way to generate such piecewise linear surrogate models for the static approaches. However, the
105 partitioning of the feasible region is only triggered by the availability of non-convex feasible regions and not by the idea of finding subregions that allow a piecewise linear approximation with a specific accuracy. Other approaches in the literature mostly generate non-linear surrogate models and/or are not designed for the use in a mode-based formulation, as the functional region of the
110 surrogate models is not restricted. However, the later is especially important for the description of technical and industrial processes, which have, by their nature, a restricted operational range. Therefore, in the class of static approaches there is a need for an approach for the generation of surrogate models, which have a piecewise linear character and describe an industrial process even more
115 accurately with respect to its operational envelope and performance within this envelope.

Focus of this contribution

This contribution focuses on the generation of data-based piecewise linear surrogate models, which allow non-linear functional relationships, such as a process
120 description, to be considered more accurately in a scheduling formulation than with a simple linear surrogate model. For this purpose two approaches are proposed, which can also be assigned the class of static approaches, because they

are based on steady-state process data. As will be seen later, both approaches also perform a partitioning of the feasible region similar to the CRS model generation. Unlike CRS generation, however, the partitioning is based on the idea of finding subregions that allow for a piecewise linear approximation with a predefined accuracy. This type of partitioning enables a more precise process description and can even be seen as a post-processing step of CRS model generation, which is used to partition the resulting polytopes. Here, the basic idea of both approaches is to interpret a piecewise linear surrogate models as a set of data points forming a mesh consisting of multiple linear elements, whereby each element is represented by the linear interpolation between its associated data points. Within a scheduling formulation – based on a mixed-integer linear programming (MILP) – one of these linear elements can be selected via a binary variable, whereby a disjunctive programming formulation can be used to reduce the combinatorics and to tighten the model relaxation. Therefore, this way of interpreting a piecewise linear surrogate model allows its efficient embedding into an MILP formulation. Another important point of this interpretation is that the set of linear elements simultaneously captures the feasible region and thus no additional equations are required for its description. Moreover, the structure of such surrogate models is compatible with a mode-based formulation.

As mentioned before, both approaches seek a set of data points, which provides an appropriate approximation of non-linear process description via the formation of several linear elements. Since the process can be described by a well-known functional relationship and thus extracted data are only affected by numerical noise, a linear interpolation between adjacent points of any data set – extracted from this relationship – already yields a relatively accurate piecewise linear approximation of the functional relationship; provided that the given data set sufficiently captures nonlinearities and the operational envelope. Therefore, the determination of a set of data points providing an appropriate approximation can easily be satisfied by a sufficiently large number of data points. However, a surrogate model of this kind has a high number of linear elements, thus requires a high number of binary variables and thus leads to large-scale optimization

problems, which usually suffer from high CPU times. Hence, the actual idea of
155 both approaches is to reduce the number data points and thus the number of
linear elements, and still provide an approximation with sufficient accuracy as
well as the description of the feasible region. Imagining a triangle mesh, which
approximates a 2-dimensional surface function, this means that linear ranges
of the function are captured with large triangles, while non-linear ranges are
160 captured with many small triangles.

The paper is organized as follows: The two approaches are described in detail
in section 2. In the subsequent section 3, both approaches are compared with
regard to several criteria: functionality of approach, computational efficiency,
and compactness of resulting surrogate model. The focus here is on an initial
165 comparison of the proposed approaches and on determining the more promising
approach. Section 4 demonstrates the scope of these approaches by means of an
exemplary application. Finally, the results are summarized in the section 5.

2. Novel approaches for piecewise linear surrogate models

The subsequently described approaches are designed to generate data-based
170 piecewise linear surrogate models for later application in scheduling formulations
based on MILPs. For both approaches, the complex functional relationship
describing a process behavior has to be initially captured in a sufficiently large
amount of data points. Since both approaches generate models, which are based
on linear interpolation, this original set of data points has to cover the relevant
175 range of the functional relationship or – even better – the entire feasible range
of the actual application, in order to prevent subsequent updates of the data
set. Ideally, the total range is spanned by evenly distributed data points and
the data resolution is sufficient to capture any non-linearity within. Therefore,
the complex functional relationship has to be used to generate a high number
180 of data points, whereby the number of data points strongly depends on the
type of relationship and its number of dimensions. Note that the computational
burden of evaluating the non-linear function is moved towards an one-time, off-

line problem and can be skipped during optimization. Apart from that, both approaches are not designed to handle noisy data points, since the focus lies
 185 on the transformation of a well-known functional relationship and not on the processing of noisy data obtained in a real-world application.

As mentioned above, the actual idea of both approaches is to reduce the number of data point of the original set of data in order to determine a reduced set of data points, which still provides an appropriate approximation of the
 190 original data via linear interpolation. In order to determine such a reduced set, this contribution proposes two contrary approaches:

The first one is an adaptation of an algorithm originating from computational geometry and object modeling, whereby a fine mesh – generated with the original data – is iteratively reduced by contracting edges of this mesh. During
 195 this iterative mesh reduction (IMRed) the edges to be contracted are selected using a quadric error metric proposed in Garland & Heckbert (1997, 1998); Garland & Zhou (2005).

In contrast to the first approach, the second one is based on a mixed integer linear programming (MILP) formulation, wherein a mesh – consisting of an *a*
 200 *priori* defined number of linear elements – is fitted to the data points. Or in other words, this formulation selects a reduced set of data points including only those points that are required to represent the convex hull of each linear mesh element. By incrementally increasing the number of elements, this so-called incremental mesh refinement (IMRef) generates a more and more refined mesh,
 205 which gives a piecewise linear approximation of the functional values.

These two approaches are described in the following sections by providing a flow chart for each approach as well as the theoretical fundamentals.

2.1. Iterative mesh reduction

The following section describes the determination of a reduced set of data
 210 points by applying the iterative mesh reduction (IMRed) approach. Initially, the basic concept of IMRed is illustrated followed by an outline of the entire algorithm.

2.1.1. Basic concept of IMRed

The functional relationship for describing the output \mathbf{y} depending on its
 215 input variables $\mathbf{x} = (x_1, x_2, \dots, x_d)$ – with \mathbf{x} as input vector of dimension d –
 is captured via a piecewise linear surface mesh, which is spanned initially by
 the complete set of available data points. Hence, the mesh consists of a set of
 linear elements, each representing a linear interpolation between adjacent data
 points. These elements are usually referred to as simplices and are spanned
 220 by $d + 1$ data points forming a d -dimensional vector space, i.e., depending on
 the dimension d these simplices represent different geometrical objects (e.g.,
 $d = 1$: line, $d = 2$: triangle, $d = 3$: tetrahedra). Furthermore, the data points
 forming such a simplex are usually referred to as vertices and the line between
 two of these vertices as edge of this simplex.

225 The above described surface mesh model represents a highly accurate ap-
 proximation of the functional relationship, provided that a sufficient amount of
 data points is used to generate a high-resolution mesh. However, the computa-
 tional cost of using such a model in an optimization problem is related to the
 mesh resolution, since its modeling requires additional optimization variables
 230 for each vertex and simplex. Consequently, the complexity of the optimization
 problem increases with the resolution of the surface mesh. In order to reduce
 this complexity as well as to maintain a sufficiently accurate approximation of
 the functional relationship, the resolution of the mesh is locally decreased by
 contracting edges of the original mesh. As illustrated in Figure 1 the number
 235 of vertices as well as the number of simplices is decreased by contracting the
 edge $(\mathbf{v}_i, \mathbf{v}_j)$ into one new vertex $\bar{\mathbf{v}}$. Therefore, the iterative contraction of edges
 leads to a surrogate model with reduced complexity, whereby the correct choice
 of edges and the position of the respective new vertex ensures an only small
 decrease in accuracy of the simplified approximation.

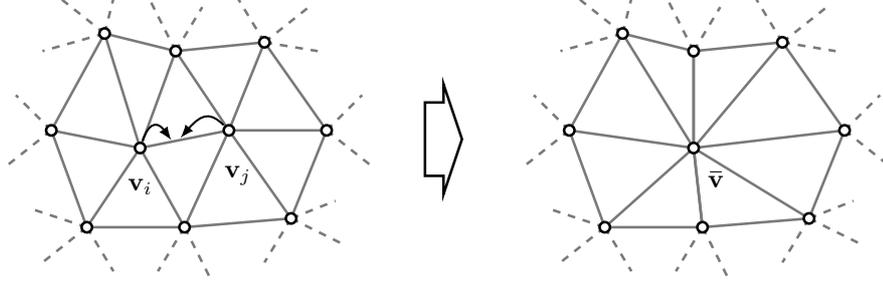


Figure 1: Edge $(\mathbf{v}_i, \mathbf{v}_j)$ is contracted into one vertex $\bar{\mathbf{v}}$, whereby the number of vertices assigned to a triangulated surface mesh ($d = 2$) is reduced by one.

Choice of edges. For choosing the correct edges the cost of contraction is evaluated using a quadric error metric. In its generalized form – as proposed by Garland & Zhou (2005) – the quadric error is defined as

$$Q_{\mathbf{p}}(\mathbf{z}) = \mathbf{z}^{\top} \mathbf{A} \mathbf{z} + 2\mathbf{b}^{\top} \mathbf{z} + c \quad (1)$$

with

$$\mathbf{A} = \mathbf{I} - \sum_{i=1}^d \mathbf{e}_i \mathbf{e}_i^{\top}, \quad (2)$$

$$\mathbf{b} = -\mathbf{A} \mathbf{p}, \quad (3)$$

$$c = \mathbf{p}^{\top} \mathbf{A} \mathbf{p}. \quad (4)$$

240 Considering a d -dimensional surface in an n -dimensional space, at any point \mathbf{p}
 on this surface an orthonormal basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ can be constructed in such a
 way that the tangent space at \mathbf{p} is spanned by $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ and thereby forming
 an affine subspace $\{\mathbf{e}_{d+1}, \dots, \mathbf{e}_n\}$, which is orthogonal to the tangent space at
 \mathbf{p} . Having such a surface, the quadric error $Q_{\mathbf{p}}(\mathbf{z})$ reflects the geometric error
 245 caused by moving the point \mathbf{p} from its current position to a new position \mathbf{z} .
 In its definition, it only measures the squared distance of \mathbf{z} from point \mathbf{p} in
 the affine subspace, i.e., the point \mathbf{p} can be moved in the tangent directions
 $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ without any penalty and only movements in the orthogonal sub-

space $\{\mathbf{e}_{d+1}, \dots, \mathbf{e}_n\}$ are penalized. Note that due to this definition the outer
 250 product $\mathbf{e}_i \mathbf{e}_i^\top$ of each dimension of the tangent space is subtracted from the
 n -dimensional identity matrix \mathbf{I} .

In analogy to Garland & Zhou (2005), the generalized quadric error is applied to construct a fundamental quadric to describe the cost of contracting any edge of the surface mesh. Due to the fact that the mesh represents a complex of various simplices, first, the fundamental quadric of a single simplex has to be defined. Any d -dimensional simplex defines a d -dimensional plane that is characterized by any point \mathbf{p} – contained within this simplex – and a set of d orthonormal tangent vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$. This hyperplane represents the aforementioned tangent space at \mathbf{p} and thus the fundamental quadric $Q_{\mathbf{p}}$ at the point \mathbf{p} can be calculated as specified in equation (1). By defining the fundamental quadric Q_σ of a simplex σ as

$$Q_\sigma(\mathbf{z}) = \int_{\mathbf{p} \in \sigma} Q_{\mathbf{p}}(\mathbf{z}), \quad (5)$$

it can be simply formulated as

$$Q_\sigma = \omega_\sigma Q_{\mathbf{p}}, \quad (6)$$

since all points $\mathbf{p} \in \sigma$ are located in the same hyperplane. Here, ω_σ denotes the d -dimensional content of the simplex σ , which can be described for instance by using the exterior product (cf. Garland & Zhou (2005)). For calculating $Q_{\mathbf{p}}$, any point \mathbf{p} on the interior of this content can be applied. Having a simplex spanned by the set of vertices $V_\sigma = \{\mathbf{v}_0, \dots, \mathbf{v}_d\}$, for instance, the point \mathbf{p} can be defined as barycenter

$$\mathbf{p} = \frac{1}{d+1} \sum_{\mathbf{v} \in V_\sigma} \mathbf{v} \quad (7)$$

of the simplex and the set of d orthonormal tangent vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ can be constructed using a Gram-Schmidt orthogonalization (Strang (2006)) of the edge vectors $\{(\mathbf{v}_1 - \mathbf{v}_0), \dots, (\mathbf{v}_d - \mathbf{v}_0)\}$.

Subsequently, the fundamental quadric Q_σ of a simplex is used to describe the fundamental quadric $Q_{\mathbf{v}}$ of one of its vertices. Since the movement of a vertex affects not only one simplex but each simplex to whom the vertex is assigned, the fundamental quadric $Q_{\mathbf{v}}$ is formulated as

$$Q_{\mathbf{v}}(\mathbf{z}) = \sum_{\sigma \in S_{\mathbf{v}}} \frac{Q_\sigma(\mathbf{z})}{(d+1)}. \quad (8)$$

255 Where the set of simplices $S_{\mathbf{v}} = \{\sigma | \mathbf{v} \in V_\sigma\}$ includes only those simplices to whom the vertex \mathbf{v} is assigned. Note that the quadric Q_σ is normalized by the number of vertices $(d+1)$, to approximate the content ω_σ of each simplex which is closest to the vertex \mathbf{v} .

Since the contraction of an edge represents the movement of its two vertices \mathbf{v}_i and \mathbf{v}_j on a joint new position, the fundamental quadric Q_ε of any edge ε can be formulated as

$$Q_\varepsilon(\mathbf{z}) = \sum_{\mathbf{v} \in V_\varepsilon} Q_{\mathbf{v}}(\mathbf{z}), \quad (9)$$

260 wherein \mathbf{z} denotes the new position and $V_\varepsilon = \{\mathbf{v}_i, \mathbf{v}_j\}$ the set of vertices that are assigned to the edge ε .

This quadric $Q_\varepsilon(\mathbf{z})$ describes the geometric error caused by moving both vertices of any edge on a new joint position \mathbf{z} and thus provides in its definition an option to evaluate edges according to their cost of contraction. However, the new position \mathbf{z} is not yet defined.

Position of new vertices. The geometric error caused during edge contraction essentially depends on the position onto which both vertices of the edge are moved to. Therefore, the position is ideally chosen in a way that causes the lowest possible geometrical error. Since $Q_\varepsilon(\mathbf{z})$ can also be transformed in the quadric form

$$Q_\varepsilon(\mathbf{z}) = \mathbf{z}^\top \mathbf{A}_\varepsilon \mathbf{z} + 2\mathbf{b}_\varepsilon^\top \mathbf{z} + c_\varepsilon, \quad (10)$$

the ideal position \mathbf{z}^* at which $Q_\varepsilon(\mathbf{z})$ is minimal (i.e., $\nabla Q_\varepsilon(\mathbf{z}^*) = 0$) is defined as

$$\mathbf{z}^* = -\mathbf{A}_\varepsilon^{-1}\mathbf{b}_\varepsilon. \quad (11)$$

265 However, if the matrix \mathbf{A}_ε is not invertible, this analytical solution is not possible. An example of this would be if all hyperplanes – considered during the construction $Q_\varepsilon(\mathbf{z})$ are parallel. Then \mathbf{A}_ε becomes singular and thus no unique optimal position exists. In such a case, the position is chosen from a set of predefined positions (here: position of both vertices $\mathbf{v}_i, \mathbf{v}_j$ as well as position of
270 $(\mathbf{v}_i + \mathbf{v}_j)/2$) by evaluating the minimal Q_ε . Finding the new position by using equation 11 or by selecting the best position of a set of predefined positions is hereinafter referred to as the standard method.

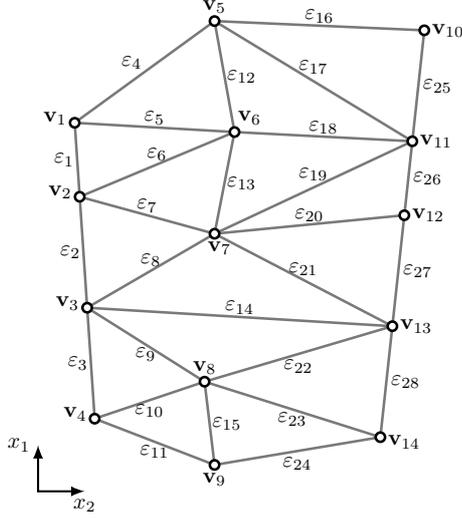
Note that this standard method of choosing the position is suitable for polygonal models possessing a closed surface. However, in case of surface models
275 with open boundaries this will cause degradation of all boundaries (cf. Garland & Zhou (2005)). Unfortunately, the models described in section 2.1.1 represent such open-bounded surface models, since these are an approximation with a limited functional range. Consequently, in this contribution choosing the position \mathbf{z} is modified by introducing a classification. Depending on this classification,
280 the position \mathbf{z} is chosen by either using the standard method or by complying with certain restrictions.

The classification and its restrictions are subsequently described on the basis of the surface mesh depicted in Figure 2. This figure shows the 2-dimensional functional range of a surface mesh consisting of 2-dimensional simplices (triangles).
285

In order to classify the edges of a mesh, the vertices have to be classified initially. For this purpose, the following classes are defined:

- Outer vertices V^O :

Considering the convex hull $\mathbf{Conv}(V)$ of all vertices V , V^O describes the set of all vertices that are assigned to the border of the convex hull



$$\begin{aligned}
 V^O &= \{v_1, v_2, v_3, v_4, v_5, v_9, v_{10}, v_{11}, v_{12}, v_{13}, v_{14}\}, \\
 V^H &= \{v_1, v_4, v_5, v_9, v_{10}, v_{14}\} \subseteq V^O, \\
 V^B &= \{v_2, v_3, v_{11}, v_{12}, v_{13}\} \subseteq V^O, \\
 V^I &= \{v_6, v_7, v_8\}; \\
 E^I &= \{\epsilon_{13}\}, \\
 E^O &= \{\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_{11}, \epsilon_{16}, \epsilon_{24}, \epsilon_{25}, \epsilon_{26}, \epsilon_{27}, \epsilon_{28}\}, \\
 E^H &= \{\epsilon_4, \epsilon_{11}, \epsilon_{16}, \epsilon_{24}\} \subseteq E^O, \\
 E^B &= \{\epsilon_2, \epsilon_{26}, \epsilon_{27}\} \subseteq E^O, \\
 E^C &= \{\epsilon_{14}, \epsilon_{17}\} \subseteq E^O, \\
 E^M &= E \setminus \{E^I, E^H, E^B, E^C\};
 \end{aligned}$$

Figure 2: Functional range of a 2-dimensional surface mesh for illustrating classification.

$\partial \text{Conv}(V)$;

$$V^O = \{v | v \in \partial \text{Conv}(V)\}. \quad (12)$$

- Hull vertices V^H :

V^H denotes a subset of V^O containing vertices that are essential to span the convex hull $\text{Conv}(V)$;

$$V^H = \{v | v \notin \text{Conv}(V \setminus \{v\})\} \subseteq V^O. \quad (13)$$

- Boundary vertices V^B :

Subset of V^O for which is valid:

$$V^B = V^O \setminus V^H \subseteq V^O. \quad (14)$$

- Inner vertices V^I :

All vertices of V that are not included in V^O are referred to as inner vertices

$$V^I = V \setminus V^O. \quad (15)$$

Based on this classes of vertices the set of edges E can be classified into the following classes:

- Inner edges E^I :

$$E^I = \{\varepsilon | V_\varepsilon \subseteq V^I\}. \quad (16)$$

290 By definition inner edges are not located on the boundary of the functional range, thus there are no restrictions during choosing the position \mathbf{z} . Therefore, the best position can be determined by using the standard method as described above.

- Outer edges E^O :

$$E^O = \{\varepsilon | \varepsilon \in \partial \text{Conv}(V)\}. \quad (17)$$

295 Consequently, both vertices of these edges are assigned to the set V^O . As defined above, V^O consists of two subsets and thus the set E^O can be distinguished in three subsets: a subset of E^O that contains edges having i) exclusively hull vertices or ii) exclusively boundary vertices, or iii) both classes of vertices. Subsequently, the first two subsets are described as well as the respective restrictions regarding the choice of \mathbf{z} . The latter subset is outlined as part of a more generalized class at a later point in this section.

- Hull edges E^H :

$$E^H = \{\varepsilon | \varepsilon \in E^O, V_\varepsilon \subseteq V^H\}. \quad (18)$$

305 No contraction of these edges is permitted since a movement of its vertices causes a modified convex hull, generally associated with a reduction of the functional range. For instance the contraction of ε_{16} in Figure 2 always leads to a modified convex hull regardless of the chosen position \mathbf{z} .

- Boundary edges E^B :

$$E^B = \{\varepsilon | \varepsilon \in E^O, V_\varepsilon \subseteq V^B\}. \quad (19)$$

In contrast to hull vertices, boundary vertices can be moved without modifying the convex hull provided that the vertex is moved along the border of the convex hull. With respect to the contraction of a boundary edge, the assumption is made that the optimal position for \mathbf{z} lies on the respective edge. Hence, the position \mathbf{z} – lying on the boundary edge – is chosen by evaluating a set Z of permitted positions; more precisely, a set of three positions, containing the position of each boundary vertex \mathbf{v}_i , \mathbf{v}_j as well as the position of $(\mathbf{v}_i + \mathbf{v}_j)/2$. In case of the boundary edge ε_2 (in Figure 2), this means that $Q_\varepsilon(\mathbf{z})$ is calculated for the positions $\mathbf{z} = \mathbf{v}_2$, $\mathbf{z} = \mathbf{v}_3$ and $\mathbf{z} = (\mathbf{v}_2 + \mathbf{v}_3)/2$. The position with the lowest value for $Q_\varepsilon(\mathbf{z})$ is used to evaluate the edge ε_2 with respect to the cost of contraction.

- Crossing edges E^C :

$$E^C = \{\varepsilon \mid \varepsilon \notin E^O, V_\varepsilon \subseteq V^O\}. \quad (20)$$

This set describes edges that cross the convex hull. The contraction of such edges leads to a constriction of the functional range and thus the contraction of crossing edges is not permitted. In analogy to E^O , set E^C can formally be distinguished into three subsets. However, this distinction is not necessary, since the restriction – no contraction is permitted – is valid for all subsets.

- Mixed edges E^M :

$$E^M = E \setminus \{E^I, E^H, E^B, E^C\}. \quad (21)$$

With respect to the contraction of a mixed edge, it is considered that the vertices show different priorities regarding the preservation of the functional range. Hull vertices have a high, boundary vertices a medium, inner vertices a low priority ($V^H > V^B > V^I$). For instance, considering ε_1 in Figure 2, the new position \mathbf{z} has to be $\mathbf{z} = \mathbf{v}_1$, since the hull vertex \mathbf{v}_1 is essential for the convex hull $\mathbf{Conv}(V)$. Or, in case of ε_6 , the inner vertex \mathbf{v}_6 has to be moved on the position $\mathbf{z} = \mathbf{v}_2$ to prevent a constriction of the

330 functional range. Consequently, the position \mathbf{z} of a mixed edge is defined
 as the vertex \mathbf{v}_i of the respective edge $(\mathbf{v}_i, \mathbf{v}_j)$ with \mathbf{v}_i having a higher
 priority than \mathbf{v}_j ($\mathbf{v}_i > \mathbf{v}_j$).

By applying this classification and the class depending restrictions during the
 choice of \mathbf{z} , the open boundaries of the surface model are preserved. Therefore,
 335 the approach can now be deployed on models as described in section 2.1.1.

2.1.2. Algorithm IMRed

The described basic concept is integrated into an algorithm, which iteratively
 selects and contracts an edge until a defined termination criterion has been
 reached. The algorithm of the IMRed approach is illustrated Figure 3, including
 340 the following main steps:

1. From the inserted set of data points a surface mesh is generated. Consider-
 ing data points in an n -dimensional space representing a d -dimensional
 functional relationship ($d < n$), a d -dimensional surface mesh can be cre-
 ated as combination of various simplices. The set of simplices S can be
 345 determined by performing a *Delaunay triangulation* (Barber et al., 1996)
 within the d -dimensional subspace of data points. This mesh describes the
 $(n - d)$ -dimensional output $\mathbf{y} = (y_{d+1}, \dots, y_n)$ within the d -dimensional
 functional range as linear interpolation between the vertices of the re-
 spective simplex depending from the input $\mathbf{x} = (x_1, x_2, \dots, x_d)$.
2. A reduced set of edges $E^- = E \setminus \{E^H, E^C\}$ is selected, since there are
 edges among the set of all edges E – such as the class of hull and crossing
 edges – for which no contraction is permitted.
3. All edges in the set E^- are evaluated regarding the geometrical error cau-
 sed by contracting the respective edge. For this purpose, the fundamental
 355 quadric $Q_\varepsilon(\mathbf{z})$ of every edge $\varepsilon \in E^-$ is calculated, determining the position
 \mathbf{z} as described in section 2.1.1.
4. The edge showing the lowest value of $Q_\varepsilon(\mathbf{z})$ is contracted by moving its
 vertices $(\mathbf{v}_i, \mathbf{v}_j)$ to the position \mathbf{z} . For this purpose, the vertices \mathbf{v}_i and \mathbf{v}_j

are replaced by the new position \mathbf{z} in each simplex that contains at least
 360 one of these vertices, thereby leading to a reduced number of vertices as
 well as the degeneration of at least one simplex.

5. The removal of a data point results in the generation of a modified surface mesh which likely shows higher deviations in representing the functional relationship. In order to describe this deviation, the output \mathbf{y}_v^d calculated using the modified surface mesh is compared for each vertex \mathbf{v} in the original set of vertices V^* . More precisely, the mean relative deviation

$$\bar{\epsilon} = \frac{\sum_{\mathbf{v} \in V^*} \frac{|\mathbf{y}_v^* - \mathbf{y}_v^d|}{y_v^*}}{|V^*|}. \quad (22)$$

is applied, whereby $|V^*|$ denotes the cardinality of the original set V^* .

Note that from an implementation point of view, $\bar{\epsilon}$ does not need to be
 calculated by evaluating the deviation at each individual vertex \mathbf{v} in the
 365 set of vertices V^* . Instead, this must be done only for those vertices where
 the surface mesh is affected by the edge contraction. For all other vertices,
 this evaluation is superfluous, since a single edge contraction only changes
 the surface mesh locally.

The deviation $\bar{\epsilon}$ is applied as criterion to control the IMRed approach. As long
 370 as $\bar{\epsilon} \leq \check{\epsilon}$ the modified surface mesh is iteratively modified further, whereby the
 steps 2.) to 5.) are repeated cyclically. In doing so, $\bar{\epsilon}$ increases and thus reaches
 the point where $\bar{\epsilon} > \check{\epsilon}$. When this criterion is met, the algorithm returns the
 surface mesh of the previous iteration cycle and stops. Note that $\check{\epsilon}$ denotes an
 upper limit of the mean relative deviation – set by the user prior to the start of
 375 the algorithm – and thus the mesh of the previous iteration is returned in order
 to not violate the upper limit $\check{\epsilon}$. By applying the IMRed approach, a modified
 set V^d is generated showing a reduced number of vertices ($|V^d| < |V^*|$) as well
 as a mean relative deviation $\bar{\epsilon} < \check{\epsilon}$.

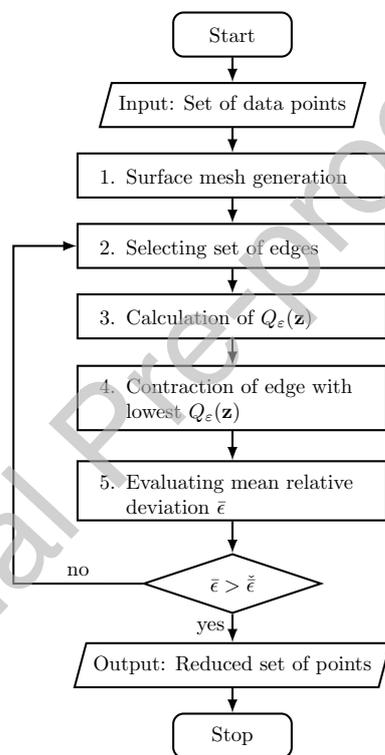


Figure 3: Flowchart of the algorithm IMRed.

2.2. Incremental mesh refinement

380 In the following section the incremental mesh refinement (IMRef) approach is depicted, which also determines a reduced set of data points, but being contrary to the afore-described IMRed approach. At the beginning of the section, the basic concepts of IMRef are outlined followed by a description of the detailed algorithm.

385 2.2.1. Basic concepts of IMRef

The method of incremental mesh refinement is an approach for determining a set of linear elements L whose combination represents a piecewise linear approximation of the original functional relationship satisfying at least a predefined accuracy. In order to satisfy the latter, each element - describing only
390 a part of the functional range - is fitted to the respective data points using a mixed integer linear programming (MILP) formulation. The MILP formulation as well as its basic idea is described as follows.

The underlying concept of the IMRef approach is the partitioning of the original data set V^* into subsets, each capturing the output \mathbf{y} with one linear element λ . In analogy to the first approach, these elements represent different geometrical objects (e.g. $d = 1$: line, $d = 2$: triangle, $d = 3$: tetrahedra) depending on the dimension d of the input $\mathbf{x} = (x_1, x_2, \dots, x_d)$. These linear elements can be created by clustering simplices originating from a surface mesh that is initially generated using the data set V^* . In this sense, the binary variables

$$y_{\sigma,\lambda} = \{0, 1\} \quad \forall \sigma \in S, \forall \lambda \in L \quad (23)$$

are introduced characterizing the assignment of any simplex σ to a linear element λ , whereby a simplex can be assigned to only one linear element. This can be expressed by the following logical formulation:

$$\bigvee_{\lambda \in L} (y_{\sigma,\lambda} = 1) \quad \forall \sigma \in S. \quad (24)$$

This disjunction can be reformulated as the following constraint:

$$\sum_{\lambda \in L} y_{\sigma, \lambda} = 1 \quad \forall \sigma \in S. \quad (25)$$

Furthermore, at least one simplex must be assigned to each linear element to ensure its utilization and thus an effect of the predefined number of elements $l = |L|$ on the resulting approximation. This statement can be written as

$$\bigvee_{\sigma \in S} (y_{\sigma, \lambda} = 1) \quad \forall \lambda \in L, \quad (26)$$

or reformulated as

$$\sum_{\sigma \in S} y_{\sigma, \lambda} \geq 1 \quad \forall \lambda \in L. \quad (27)$$

By assigning each simplex to one of the linear elements in the set L , the functional range is partitioned into l sub-ranges. Consequently, the functional range of each linear element is determined by the binary variables $y_{\sigma, \lambda}$. For capturing the information contained in the vertices of each simplex further binary variables are required. Therefore, the binary variables

$$y_{\mathbf{v}, \lambda} = \{0, 1\} \quad \forall \mathbf{v} \in V^*, \forall \lambda \in L \quad (28)$$

are added to the MILP formulation characterizing the assignment of any vertex \mathbf{v} to a linear element λ . These binary variables are linked with the binary variables $y_{\sigma, \lambda}$, since the assignment of any simplex σ to one element is equivalent to the assignment of all of its vertices $\mathbf{v} \in V_{\sigma}$ to exactly this element and vice versa. This statement can be written as

$$y_{\sigma, \lambda} \Leftrightarrow y_{\mathbf{v}, \lambda} \quad \forall \sigma \in S, \forall \mathbf{v} \in V_{\sigma}, \forall \lambda \in L, \quad (29)$$

or by the following constraints using propositional logic (Raman & Grossmann,

1994):

$$\Rightarrow : y_{\sigma,\lambda} \leq y_{\mathbf{v},\lambda} \quad \forall \sigma \in S, \forall \mathbf{v} \in V_{\sigma}, \forall \lambda \in L; \quad (30)$$

$$\Leftarrow : y_{\sigma,\lambda} \geq \sum_{\mathbf{v} \in V_{\sigma}} y_{\mathbf{v},\lambda} - \underbrace{(|V_{\sigma}| - 1)}_{\equiv d} \quad \forall \sigma \in S, \forall \mathbf{v} \in V_{\sigma}, \forall \lambda \in L. \quad (31)$$

Due to these constraints a vertex is clearly assigned to one specific linear element according to the assignment of its simplex. Despite of this and of having a fixed number of assigned simplices, the number of assigned vertices within one linear element may vary, since one specific vertex \mathbf{v} is probably included in various sets V_{σ} , i.e., adjacent simplices assign fewer vertices to a linear element λ than locally distributed simplices. The clustering of adjacent simplices is thus encouraged by restricting the number of assigned vertices depending on the number of assigned simplices:

$$\underbrace{|V_{\sigma}|}_{\equiv (d+1)} \leq \sum_{\sigma \in S} \sum_{\mathbf{v} \in V_{\sigma}} y_{\mathbf{v},\lambda} \leq |V_{\sigma}| + \sum_{\sigma \in S} y_{\sigma,\lambda} - 1 \quad \forall \lambda \in L. \quad (32)$$

The lower bound is defined by the number of vertices that are assigned to an element when exactly one simplex belongs to it. Consequently, the lower bound is given by $|V_{\sigma}| = d + 1$. For adding adjacent simplices at most one additional vertex is required and thus the upper bound can be formulated as shown in equation (32).

In addition to the previously introduced binary variables, another set of binary variables

$$y_{\mathbf{v},\tilde{\lambda}} = \{0, 1\} \quad \forall \mathbf{v} \in V^*, \forall \tilde{\lambda} \in \tilde{L} \quad (33)$$

is added to the MILP in order to capture all vertices that are assigned to more than one linear element. Note that these points form a set of cutting elements \tilde{L} , whereby each cutting element $\tilde{\lambda}$ has the dimension $(d - 1)$ and lies between two linear elements λ . Therefore, the number of cutting elements $|\tilde{L}|$ is given

by

$$|\tilde{L}| = \binom{l}{2} = \frac{l!}{2!(l-2)!}. \quad (34)$$

Due to the fact that this equation is only valid for $l \geq 2$ and that the modeling of cutting elements is not reasonable for $l < 2$, all formulation in connection with cutting elements are only added to the MILP if $l \geq 2$. The necessity of these cutting elements as well as their binary variables will be explained later on.

For any cutting element $\tilde{\lambda} \in \tilde{L}$ describing the intersection of the respective linear elements λ_i and λ_j , the logical statement

$$y_{\mathbf{v},\tilde{\lambda}} \Leftrightarrow y_{\mathbf{v},\lambda_i} \wedge y_{\mathbf{v},\lambda_j} \quad \forall \mathbf{v} \in V^*, \forall \tilde{\lambda} \in \tilde{L}, \quad (35)$$

applies, which can be reformulated as follows:

$$y_{\mathbf{v},\tilde{\lambda}} \leq y_{\mathbf{v},\lambda_i} \quad \forall \mathbf{v} \in V^*, \forall \tilde{\lambda} \in \tilde{L}; \quad (36)$$

$$y_{\mathbf{v},\tilde{\lambda}} \leq y_{\mathbf{v},\lambda_j} \quad \forall \mathbf{v} \in V^*, \forall \tilde{\lambda} \in \tilde{L}; \quad (37)$$

$$y_{\mathbf{v},\tilde{\lambda}} \geq y_{\mathbf{v},\lambda_i} + y_{\mathbf{v},\lambda_j} - 1 \quad \forall \mathbf{v} \in V^*, \forall \tilde{\lambda} \in \tilde{L}. \quad (38)$$

Due to these constraints a vertex can only be assigned to a cutting element $\tilde{\lambda}$, if this vertex is simultaneously assigned to the two respective linear elements λ_i and λ_j and vice versa. This in turn means that each vertex \mathbf{v} can be assigned exactly to one linear element λ , unless it is also assigned to a cutting element. Hence, the additional equality constraint

$$\sum_{\lambda \in L} y_{\mathbf{v},\lambda} = 1 + \sum_{\tilde{\lambda} \in \tilde{L}} y_{\mathbf{v},\tilde{\lambda}} \quad \forall \mathbf{v} \in V^* \quad (39)$$

may be added to the MILP.

By means of the afore-described constraints, the set V is partitioned into $|L|$ subsets (linear elements) and $|\tilde{L}|$ intersections (cutting elements) based on the assignment of the binary variables $y_{\sigma,\lambda}$, whereby the output \mathbf{y} of each vertex

\mathbf{v} within any subset is approximated using a linear element λ . These elements describe the approximated output $\mathbf{y}_{\mathbf{v},\lambda}^{\triangleleft}$ of any vertex \mathbf{v} in the respective element λ as a linear function of all its input variables $\mathbf{x}_{\mathbf{v}}$ and thus can be generally formulated as

$$\mathbf{y}_{\mathbf{v},\lambda}^{\triangleleft} = \alpha_{\lambda} + \sum_{\delta=1}^d \beta_{\delta,\lambda} \cdot x_{\mathbf{v},\delta} \quad \forall \mathbf{v} \in V^*, \forall \lambda \in L, \quad (40)$$

where α_{λ} and $\beta_{\delta,\lambda}$ are parameters fitting each linear element to the output of its vertices.

The combination of all these linear elements generates a piecewise linear approximation of the set V^* . However, additional restriction regarding the determination of α_{λ} and $\beta_{\delta,\lambda}$ have to be added to obtain a continuous description of the functional range. For this purpose, continuity has to be ensured at the intersection of two linear elements λ_i and λ_j , i.e., the approximated outputs $\mathbf{y}_{\mathbf{v},\lambda_i}^{\triangleleft}$ and $\mathbf{y}_{\mathbf{v},\lambda_j}^{\triangleleft}$ of both elements have to be equal. However, this equality condition has to be valid only if the vertex \mathbf{v} lies in the intersection of two linear elements. This equality condition is, hence, extended by two slack variables $\chi_{\mathbf{v},\tilde{\lambda}}^+$ and $\chi_{\mathbf{v},\tilde{\lambda}}^-$ being assigned to the respective cutting element:

$$\mathbf{y}_{\mathbf{v},\lambda_i}^{\triangleleft} = \mathbf{y}_{\mathbf{v},\lambda_j}^{\triangleleft} + \chi_{\mathbf{v},\tilde{\lambda}}^+ - \chi_{\mathbf{v},\tilde{\lambda}}^- \quad \forall \mathbf{v} \in V^*, \forall \tilde{\lambda} \in \tilde{L}. \quad (41)$$

These slack variables are only forced to be equal to zero, if the vertex \mathbf{v} is assigned to the cutting element $\tilde{\lambda}$. If this requirement is not fulfilled, the slack variables can be greater or equal to zero. This logical expression can be formulated as

$$0 \leq \chi_{\mathbf{v},\tilde{\lambda}}^+ \leq \Psi \cdot (1 - y_{\mathbf{v},\tilde{\lambda}}) \quad \forall \mathbf{v} \in V^*, \forall \tilde{\lambda} \in \tilde{L}, \quad (42)$$

$$0 \leq \chi_{\mathbf{v},\tilde{\lambda}}^- \leq \Psi \cdot (1 - y_{\mathbf{v},\tilde{\lambda}}) \quad \forall \mathbf{v} \in V^*, \forall \tilde{\lambda} \in \tilde{L}, \quad (43)$$

using the afore-defined binary variables $y_{\mathbf{v},\tilde{\lambda}}$ as well as a big-M parameter Ψ with a sufficiently high value.

By applying these constraints, the combination of all linear elements forms a continuous piecewise linear surrogate model approximating the set V^* . In order to generate a surrogate model representing the best possible approximation while using a predefined number of linear elements, an appropriate objective function has to be formulated for the MILP. Here, the objective function is based on the relative deviation $\epsilon_{\mathbf{v},\lambda}$, which can be defined as

$$\epsilon_{\mathbf{v},\lambda} = \frac{|(\mathbf{y}_{\mathbf{v}} - \mathbf{y}_{\mathbf{v},\lambda}^{\triangleleft})|}{\mathbf{y}_{\mathbf{v}}} \quad \forall \mathbf{v} \in V^*, \forall \lambda \in L \quad (44)$$

for each $\mathbf{v} \in V^*$ regarding the respective linear element λ . However, this formulation has to be modified in order to be used in the MILP formulation. Instead of the norm, two linear inequality constraints are applied, avoiding any non-differentiabilities:

$$\epsilon_{\mathbf{v},\lambda} \geq \frac{(\mathbf{y}_{\mathbf{v}} - \mathbf{y}_{\mathbf{v},\lambda}^{\triangleleft})}{\mathbf{y}_{\mathbf{v}}} \quad \forall \mathbf{v} \in V^*, \forall \lambda \in L, \quad (45)$$

$$\epsilon_{\mathbf{v},\lambda} \geq \frac{(\mathbf{y}_{\mathbf{v}} - \mathbf{y}_{\mathbf{v},\lambda}^{\triangleleft})}{\mathbf{y}_{\mathbf{v}}} \quad \forall \mathbf{v} \in V^*, \forall \lambda \in L. \quad (46)$$

In the objective function only those deviations $\epsilon_{\mathbf{v},\lambda}$ are required, for which the vertex \mathbf{v} is assigned to the linear element λ . This may be achieved by defining a modified relative deviation expressed by the formulation

$$\check{\epsilon}_{\mathbf{v},\lambda} = \epsilon_{\mathbf{v},\lambda} \cdot y_{\mathbf{v},\lambda} \quad \forall \mathbf{v} \in V^*, \forall \lambda \in L. \quad (47)$$

For integrating this non-linear description in the MILP formulation, the two cases

$$y_{\mathbf{v},\lambda} = 1 \Rightarrow \check{\epsilon}_{\mathbf{v},\lambda} = \epsilon_{\mathbf{v},\lambda}:$$

$$\check{\epsilon}_{\mathbf{v},\lambda} \leq \epsilon_{\mathbf{v},\lambda} \quad \forall \mathbf{v} \in V^*, \forall \lambda \in L, \quad (48)$$

$$\check{\epsilon}_{\mathbf{v},\lambda} \geq \epsilon_{\mathbf{v},\lambda} - \Omega(1 - y_{\mathbf{v},\lambda}) \quad \forall \mathbf{v} \in V^*, \forall \lambda \in L; \quad (49)$$

$$y_{\mathbf{v},\lambda} = 0 \Rightarrow \check{\epsilon}_{\mathbf{v},\lambda} = 0:$$

$$\check{\epsilon}_{\mathbf{v},\lambda} \leq \Omega \cdot y_{\mathbf{v},\lambda} \quad \forall \mathbf{v} \in V^*, \forall \lambda \in L, \quad (50)$$

$$\check{\epsilon}_{\mathbf{v},\lambda} \geq 0 \quad \forall \mathbf{v} \in V^*, \forall \lambda \in L; \quad (51)$$

are distinguished by applying another big-M parameter Ω with a sufficiently high value. Using this modified relative deviation, an appropriate objective function for the MILP can be formulated as

$$\min_{\substack{y_{\sigma,\lambda}, y_{\mathbf{v},\lambda}, y_{\mathbf{v},\bar{\lambda}}, \\ \alpha_{\lambda}, \beta_{\delta,\lambda}}} \quad Obj = \sum_{\lambda \in L} \sum_{\mathbf{v} \in V^*} \check{\epsilon}_{\mathbf{v},\lambda}, \quad (52)$$

410 wherein the binary variables $y_{\sigma,\lambda}, y_{\mathbf{v},\lambda}, y_{\mathbf{v},\bar{\lambda}}$ and the continuous variables $\alpha_{\lambda}, \beta_{\delta,\lambda}$ are adjusted to minimize the sum over all modified relative deviations $\check{\epsilon}_{\mathbf{v},\lambda}$. Note that due to the formulation of the objective function all relative derivations of vertices – being assigned to a cutting element – are considered several times and thus have a higher weight in the objective function. This is, however, negligible
415 as long as the following two points are satisfied: i) The set V^* contains a high number of vertices and ii) the surrogate model is created applying a low number of linear elements.

By having specified the objective function as well as the afore-described constraints, the MILP is complete and can be solved to determine a piecewise
420 linear surrogate model using an *a priori* defined number of linear elements.

2.2.2. Algorithm IMRef

The described basic concept is integrated into an algorithm which incrementally increases the number of linear elements, this way, generating a more and more refined mesh. By means of Figure 4 the algorithm of the IMRef approach
425 is described, including the following main steps:

1. The input set of data V^* is used to generated a surface mesh. This process step is done analogous to the mesh generation step in the IMRed algorithm

(cf. section 2.1.2).

2. Based on the generated mesh, the optimization model of the previous section is created, initially using only one linear element ($l = 1$) as well as starting values for the big-M parameters Ψ and Ω .
430
3. On solving the MILP, the binary variables $y_{\sigma,\lambda}, y_{\mathbf{v},\lambda}, y_{\mathbf{v},\bar{\lambda}}$ ($y_{\mathbf{v},\bar{\lambda}}$ only if $l > 1$) and the continuous variables $\alpha_\lambda, \beta_{\delta,\lambda}$ are adjusted to minimize the objective function, thereby leading to a surrogate model that represents a piecewise linear approximation of the input set of data.
435
4. The quality of this surrogate model is primarily influenced by the number of linear elements l , but also by the values of the big-M parameters Ψ and Ω . For this reason, both parameters Ψ and Ω are evaluated with respect to the obtained solution of the optimization model. If the parameters affect this solution, the parameters have to be adjusted and step 3.) has to be repeated subsequently.
440
5. The mean relative deviation $\bar{\epsilon}$ is calculated according to equation (22) to evaluate the surrogate model.

As with the IMRed approach, $\bar{\epsilon}$ is applied as convergence criterion to control the IMRef approach's accuracy. As long as $\bar{\epsilon} > \check{\epsilon}$, the number of linear elements l is increased by 1 ($l = l + 1$). Thereon the steps 2.) to 5.) are repeated. In doing so, the mean relative deviation $\bar{\epsilon}$ decreases, since the MILP formulation applies more and more linear elements to capture non-linearities in the data set. Therefore, the algorithm at some point meets the criterion $\bar{\epsilon} \leq \check{\epsilon}$. At this point, the algorithm returns a set containing the hull vertices of each linear element, i.e., only such vertices of each linear element that are relevant to describe the piecewise linear approximation of the original data set V^* . In this way the IMRef approach also generates a set V^\triangleleft showing a reduced number of vertices ($|V^\triangleleft| < |V^*|$) as well as a mean relative deviation $\bar{\epsilon} < \check{\epsilon}$.
450

Note that the a-priori specification of the number of linear elements l – using an algorithm – allows for a compact formulation of the optimization problem
455

having the objective to minimize $bar{\epsilon}$. Whereas, alternative formulations, e.g. where $\bar{\epsilon}$ is constraint to less than $\check{\epsilon}$ while minimizing the number of linear elements l , result in larger and more complex formulations, which also require an algorithmic approach. The latter is due to the fact that the underlying MILP formulation assigns simplices to linear elements by using binary variables and that these variables must be defined previous to the model construction. Hence, such an alternative formulation requires an a-priori defined maximum of linear elements, combined with an algorithm for controlling this maximum and large model formulations that lack in the tightness of their model relaxation.

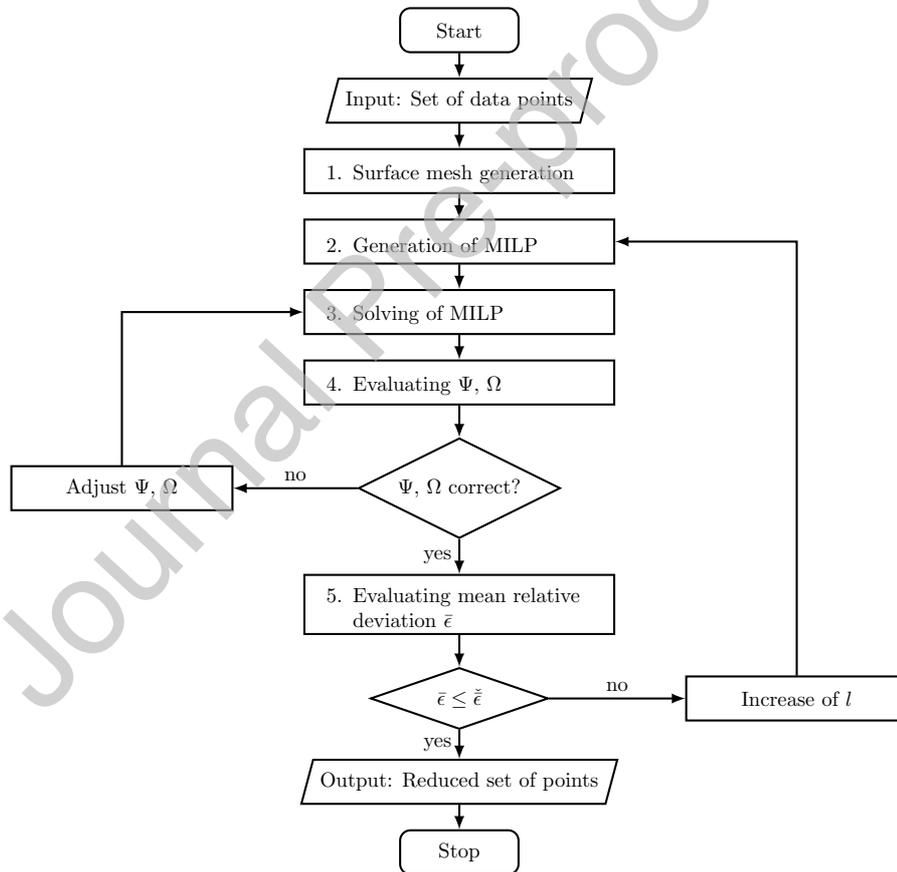


Figure 4: Flowchart of the algorithm IMRef.

3. Functional check and comparison of approaches

In this chapter, the developed approaches are applied to several generic data sets V^* describing various functional relationships. In doing so, both approaches are compared and assessed based on the following aspects:

470 1. Functionality of approach:

The functionality of each approach is evaluated by applying it to data sets, where the resulting surrogate model is known in advance, i.e., these data sets originate from piecewise linear functions. Thus the resulting surrogate model must be congruent with the respective piecewise linear
475 function in order to guarantee functionality.

2. Computational efficiency:

In order to evaluate the performance of each approach, it is applied to data sets originating from piecewise linear functions but with different cardinality $|V^*|$. Therefore, the performance of an approach can be evaluated
480 by investigating the rise of the computation time as a function of $|V^*|$.

3. Compactness of the obtained surrogate models:

Here, the compactness of the resulting surrogate model is characterized by the number of data points that are required to describe it. This number depends on i) the desired accuracy of the surrogate model (i.e., the specified upper limit of $\tilde{\epsilon}$), ii) the non-linearity of the original functional
485 relationship, and iii) the efficiency of the chosen approach. Hence, this aspect is evaluated for each approach by investigating the accuracy of the respective surrogate model as a function of the number of assigned data points, while having specified the same generic data set for both approaches.
490

Both approaches are implemented in Python 2.7. Pyomo (Hart et al., 2011, 2017) was applied as modeling environment for the optimization model of the IMRef approach and the commercial solver Gurobi 7.5 has been used to solve

the MILPs. All calculations have been performed on an Intel®Core(TM) i5-
 495 4300M machine at 2.60 GHz with four cores and 8 GB RAM running Windows
 7 Enterprise. All MILP formulations of the IMRef approach have been solved
 to a relative optimality gap of 1×10^{-4} .

3.1. Piecewise linear functional relationships

For generating the data sets V^* , two piecewise linear functional relationships
 (PLFs) are applied, differing in the dimension d of the input variable $\mathbf{x} =$
 (x_1, x_2, \dots, x_d) . The first function (PLF-1) represents a part of a trapezoid
 function of dimension $d = 1$ and is defined in the interval $100 \leq x_1 \leq 400$ as:

$$y(x_1) = \begin{cases} 200 & \text{for } 100 \leq x_1 \leq 200, \\ -x_1 + 400 & \text{for } 200 < x_1 < 300, \\ 100 & \text{for } 300 \leq x_1 \leq 400. \end{cases} \quad (53)$$

The second functional relationship (PLF-2) is a constant extension of the function
 PLF-1 in a second dimension. Being specified in the interval $100 \leq x_1 \leq 400$,
 $100 \leq x_2 \leq 200$, PLF-2 can be defined as:

$$y(x_1, x_2) = \begin{cases} 200 & \text{for } 100 \leq x_1 \leq 200, 100 \leq x_2 \leq 200, \\ -x_1 + 400 & \text{for } 200 < x_1 < 300, 100 \leq x_2 \leq 200, \\ 100 & \text{for } 300 \leq x_1 \leq 400, 100 \leq x_2 \leq 200. \end{cases} \quad (54)$$

These PLFs are discretized using various step sizes, in order to generate
 500 several data sets V^* . For instance, the discretization of PLF-1 using a step size
 $\mu = 50$ results in a data set V^* with seven equidistantly distributed data points
 at the position $x_1 \in \{100, 150, \dots, 350, 400\}$ and thus a cardinality of $|V^*| = 7$.
 Table 1 lists for both PLFs the cardinality $|V^*|$ of the generated data sets $|V^*|$
 depending on various step sizes μ .

505 Note that the step sizes are chosen in a way that each edge of the PLFs
 is explicitly captured by a data point in the generated data set. In doing so,

Table 1: Cardinality $|V^*|$ for all data sets V^* – generated via discretization of both PLFs applying various step sizes μ .

| Step size μ | $ V^* $ using PLF-1 | $ V^* $ using PLF-2 |
|-----------------|---------------------|---------------------|
| 100 | 4 | 8 |
| 50 | 7 | 21 |
| 25 | 13 | 65 |
| 20 | 16 | 96 |
| 10 | 31 | 341 |
| 5 | 61 | 1281 |
| 2 | 151 | 7701 |
| 1 | 301 | 30401 |

the linear interpolation between adjacent data points within any data set V^* is equivalent to the original PLF. In addition, all generated data sets V^* – independent of their cardinality – can be reduced to an altered data set V^∇ , which contains only those data points, which are at least required to represent the PLF via linear interpolation. This altered data set V^∇ is equivalent to the data set V^* of each PLF that was generated by applying a step size of $\mu = 100$. Figure 5 illustrates the reduced data set V^∇ of both piecewise linear functional relationships PLF-1 and PLF-2.

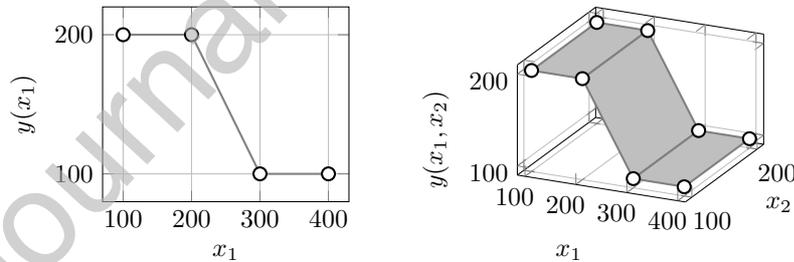


Figure 5: Illustration of the data sets V^∇ for PLF-1 (left) and PLF-2 (right).

By applying the proposed approaches (IMRed or IMRef) to the data sets V^* , a reduced data set V^∇ is returned, which is identical to the respective data set V^∇ , independently of the chosen approach and data set V^* . This functional check provides evidence that both approaches perform as they were designed and thus allows for further evaluations.

520 For comparing the computational efficiency of both approaches, the computation time until the set V^\triangleleft is returned is investigated as a function of the cardinality $|V^*|$. Figure 6 shows the rise of the computation time, applying the approaches to the data sets V^* generated with PLF-1.

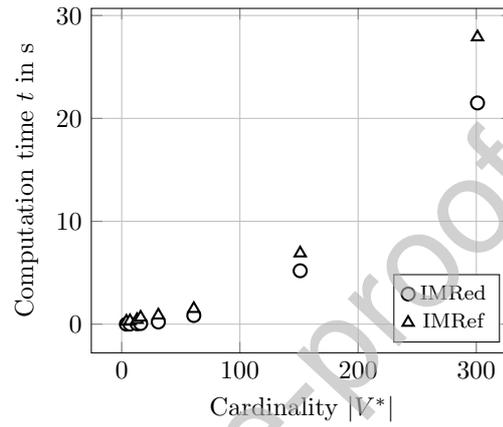


Figure 6: Computation time of IMRed approach (○) and IMRef approach (△) plotted over the cardinality $|V^*|$ of all data sets V^* generated via discretization of the functional relationship PLF-1.

525 As can be seen here, both approaches show an exponential increase of the computation time t with increasing cardinality $|V^*|$. However, the increase is more pronounced in the IMRef approach. Consequently, the IMRed approach shows a higher computational efficiency considering the results of the functional relationship PLF-1.

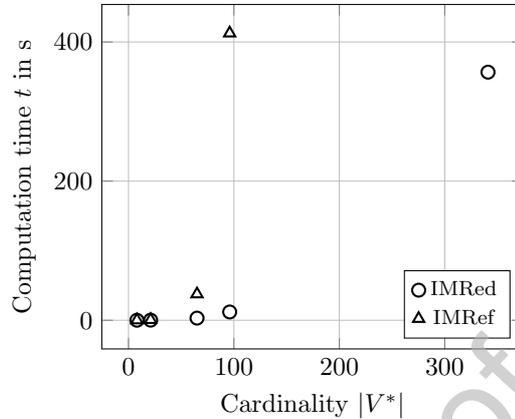


Figure 7: Computation time of IMRed approach (○) and IMRef approach (△) plotted over the cardinality $|V^*|$ of all data sets V^* generated via discretization of the functional relationship PLF-2.

Figure 7 shows a similar behavior of the computational efficiency for the two-dimensional functional relationship PLF-2. However, the rise of the computation time t with increasing cardinality $|V^*|$ is significantly stronger, especially looking at the results of the IMRef approach. This is due to the fact, that in higher dimensions d the number of simplices $|S|$ – generated during the surface mesh generation – increases non-linearly with the cardinality $|V^*|$. This in turn causes a higher number of edges and associated calculation steps in terms of the IMRed approach as well as a higher number of variables and equations in terms of the IMRef approach. Particularly in the IMRef approach, this results in a significant increase in the computation time t .

This effect is intensified by the fact that the size of the optimization problem increases additionally with each linear element L needed to approximate the functional relationship. Additional dimensions are usually associated with non-linearities and thus the need of more linear elements for their approximation. The piecewise linear approximation of data describing high-dimensional or highly non-linear functional relationships, therefore, requires a high number of linear elements and thus leads to the generation of large-scale MILPs in case of the IMRef approach. Based on the computational performance observed in the

two previous examples (PFL-1 and PLF-2), it can already be seen that there is only a very limited applicability of the IMRef approach to real-world applications, as these usually have a higher complexity in terms of dimensionality and non-linearity.

It can be concluded from the afore-described results that in terms of computational efficiency, the IMRed approach is superior to the IMRef approach, since the computation time of the latter approach increases more exponentially depending on various parameters such as cardinality $|V^*|$ and dimension d . Although, the IMRed approach shows a superior behavior, the results – shown in Figure 6 and Figure 7 – obviously show that both approaches suffer from the phenomenon – commonly referred to as the *curse of dimensionality* (Bellman, 2015). This phenomenon describes that the number of data points required for appropriately capturing a functional relationship increases exponentially with the number of dimensions d . The exponential increase of data points in turn reinforces the rise in the number of simplices $|S|$ generated during the surface mesh generation. As described above, in higher dimensions this leads to very high number of edges and associated calculations steps in terms of the IMRed approach and high number of variables and equation in terms of the IMRef approach. Therefore, the performance of both approaches obviously drops with increasing dimensionality. This is largely due to the *curse of dimensionality* that affects all data-based approaches and thus also the IMRed and IMRef approach.

3.2. Non-linear functional relationships

In analogy to section 3.1, two non-linear functional relationships (NLFs) – differing in their dimension d – are used to generate data sets V^* capturing a non-linear behavior. The first function (NLF-1) represents a segment of a circular arc in the dimension $d = 1$ and is defined in the interval $100 \leq x_1 \leq 200$ as:

$$y(x_1) = 100 + \sqrt{10000 - (x_1 - 100)^2} \quad \text{for} \quad 100 \leq x_1 \leq 200. \quad (55)$$

The second functional relationship (NLF-2) is once again a constant extension of the one-dimensional function in a second dimension. Being specified in the interval $100 \leq x_1 \leq 200$, $10 \leq x_2 \leq 30$, NLF-2 can be defined as:

$$y(x_1, x_2) = 100 + \sqrt{10000 - (x_1 - 100)^2} \quad \text{for} \quad 100 \leq x_1 \leq 200, 10 \leq x_2 \leq 30. \quad (56)$$

570 The NLFs are illustrated in Figure 8.

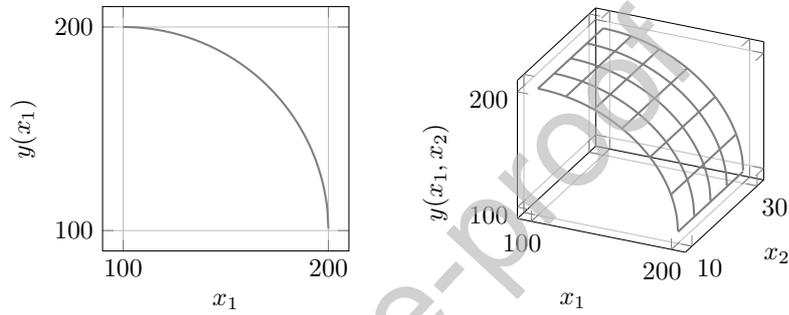


Figure 8: Illustration of the both NLFs; NLF-1 (left) and NLF-2 (right).

In order to generate exemplary data sets V^* – capturing the non-linearity of each NLFs with equidistantly distributed data points – the NLFs are discretized using a homogeneous step size of $\mu = 10$: This results in case of NLF-1 in a data set with cardinality $|V^*| = 11$ and in case of NLF-2 in a second set with
 575 $|V^*| = 11 \times 3 = 33$.

Obviously, the more linear elements the approximating surrogate model consists of, the better the approximation will be. The application of an additional linear element and thus of more data points in the data set V^\triangleleft leads to a discrete reduction of the smallest achievable mean relative deviation $\bar{\epsilon}_\dagger$. In order
 580 to compare both approaches regarding the compactness of their surrogate models, this smallest achievable relative deviation $\bar{\epsilon}_\dagger$ is investigated as a function of the number of points that are assigned to the set V^\triangleleft , i.e., as a function of the cardinality $|V^\triangleleft|$.

Figure 9 shows the decrease of $\bar{\epsilon}_\dagger$ with increasing cardinality $|V^\triangleleft|$ for the
 585 one-dimensional function NLF-1. Note that a surrogate model with $|V^\triangleleft| = 11$

can contain all points in the set V^* and thus can accomplish a mean relative deviation of $\bar{\epsilon} = 0$. Although both approaches generate such a surrogate model with $\bar{\epsilon} = 0$, the resulting point can not be depicted in Figure 9. Hence, $\bar{\epsilon}_\dagger$ is only depicted for the discrete values $|V^\triangleleft| \in \{2, 3, \dots, 9, 10\}$. As can be seen here, the
 590 IMRef approach shows lower values for $\bar{\epsilon}_\dagger$ when the surrogate model has a low number of data points (i.e., $|V^\triangleleft| \ll |V^*|$), whereas the IMRed approach shows better results when a high number of data points are assigned to the surrogate model (i.e., $|V^\triangleleft| < |V^*|$). This is due to the general design of each approach and can be best explained by looking at the surrogate models with $|V^\triangleleft| = 2$
 595 and $|V^\triangleleft| = 10$.

For $|V^\triangleleft| = 2$ the IMRed approach is forced to maintain the two hull vertices V^H due to the introduced classification, i.e., the surrogate model represents only a linear interpolation between the vertices $(x_1, y) = (100, 200)$ and $(x_1, y) = (200, 100)$. Whereas the IMRef approach determines two optimal
 600 points (at the position $x_1 = 100$ and $x_1 = 200$ with optimal y -values), whose linear interpolation forms a straight line showing the lowest $\bar{\epsilon}$ regarding the points in set V^* .

When looking at the surrogate model with $|V^\triangleleft| = 10$ the drawback of the IMRef approach becomes evident: The more linear elements are used, the more
 605 the optimization is affected by the discretization initially applied to the NLF, since the position of each point has to be one of the vertices $\mathbf{v} \in V^*$. Here, the equidistant discretization is not ideal for capturing the functional relationship of a segment of a circular arc and thus the compactness of the resulting surrogate model decreases with increasing cardinality $|V^\triangleleft|$. By contrast, the IMRed
 610 approach can contract two vertices on a position that is not assigned to one vertex $\mathbf{v} \in V^*$ and, therefore, can counter the discretization-related problem. Consequently, the IMRed approach is superior with regard to the model compactness when it comes to surrogate models showing a very low mean relative deviation $\bar{\epsilon}$.

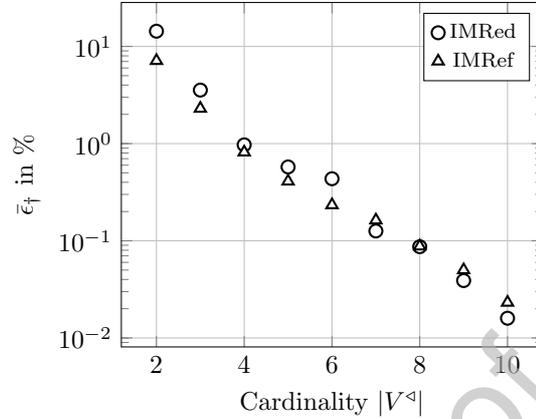


Figure 9: Smallest achievable mean relative deviation $\bar{\epsilon}_\dagger$ using IMRed approach (O) or IMRef approach (Δ) plotted over the number of data points ($|V^\triangleleft|$) assigned to the reduced set V^\triangleleft for the functional relationship NLF-1.

615 In analogy to the one-dimensional functional relationship, the NLF-2 is used to compare both approaches regarding the compactness of their surrogate models applying them to a two-dimensional function. The results are depicted in Figure 10.

The data set V^* of NLF-2 represents a threefold of the set of NLF-1, only 620 differing in the second dimension. Therefore, the middle set can be reproduced by a linear interpolation between the other two sets and thus a surrogate model with $|V^\triangleleft| = 22$ can describe the V^* of NLF-2 with a mean relative deviation of $\bar{\epsilon} = 0$. Both approaches produce such a surrogate model with $\bar{\epsilon} = 0$. On the other hand, at least four vertices are necessary to span the total functional 625 range of NLF-2. Hence, $\bar{\epsilon}_\dagger$ is depicted for discrete values in the range of $|V^\triangleleft| \in \{4, 5, \dots, 20, 21\}$.

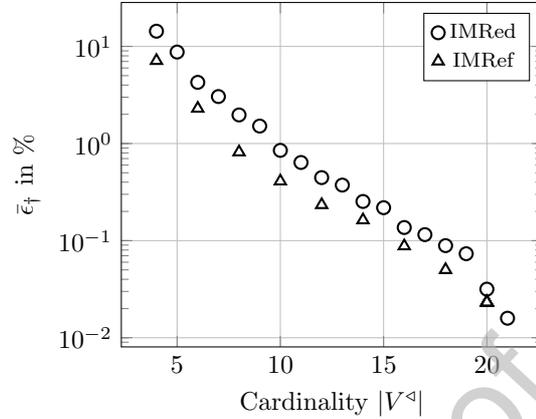


Figure 10: Smallest achievable mean relative deviation $\bar{\epsilon}_\dagger$ using IMRed approach (○) or IMRef approach (△) plotted over the number of data points ($|V^\dagger|$) assigned to the reduced set V^\dagger for the functional relationship NLF-2.

Here two aspects stand out: The IMRef approach only generates surrogate models with even cardinality $|V^\dagger|$ and these models always show a lower $\bar{\epsilon}_\dagger$ compared with the respective models of the IMRed approach. The former aspect
 630 arises from the fact that the IMRef approach is an optimization based approach. Therefore, this approach recognizes the fact, that the best way to describe the constant extension of NLF-2 is the addition of linear elements showing a constant behavior in the respective dimension. This behavior, however, can only be achieved by simultaneously adding two vertices to the set $|V^\dagger|$. Whereas the
 635 step-by-step contraction of vertices in the IMRed approach removes vertices one by one and thus also generates surrogate models with uneven cardinality $|V^\dagger|$.

In contrast to IMRef, the IMRed approach does not generate linear elements showing a constant behavior in dimension x_2 , since each contraction is based on a locally defined quadratic error metric. Instead, multiple elements are used to
 640 approximate the constant behavior leading to a higher mean relative deviations $\bar{\epsilon}$. This becomes clearer when looking at Figure 11, which shows surrogate models of both approaches having a cardinality $|V^\dagger| = 12$.

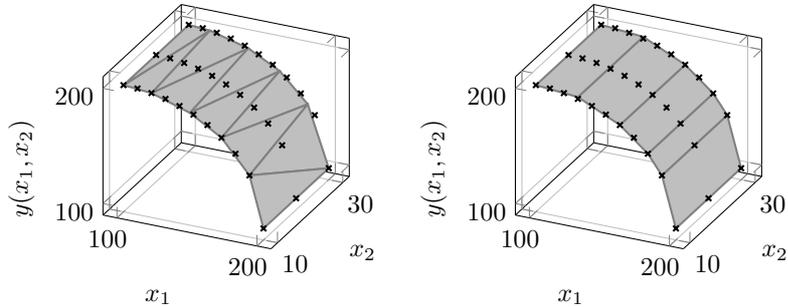


Figure 11: Illustration of $V^*(\star)$ and surrogate models with $|V^{\triangleleft}| = 12$ describing NLF-2; generated with IMRed approach (left) and IMRef approach NLF-2 (right).

Although Figure 10 might give the impression that the IMRef approach might be superior in terms of compactness of the resulting surrogate model for high-dimensional applications, this is not the case per se. As observed with the NLF-1, the IMRef approach is strongly influenced by the discretization initially applied to the NLF, in order to generate the data set V^* . In case of the NLF-2, the IMRef approach benefits both from the orientation of the points in the data set and from the fact that they are evenly distributed. This allows for the generation of cutting elements λ forming straight lines (in case of $d = 2$) through the functional range and thus the generation of linear elements λ , which differ in their fitting parameters α_λ and $\beta_{\delta,\lambda}$.

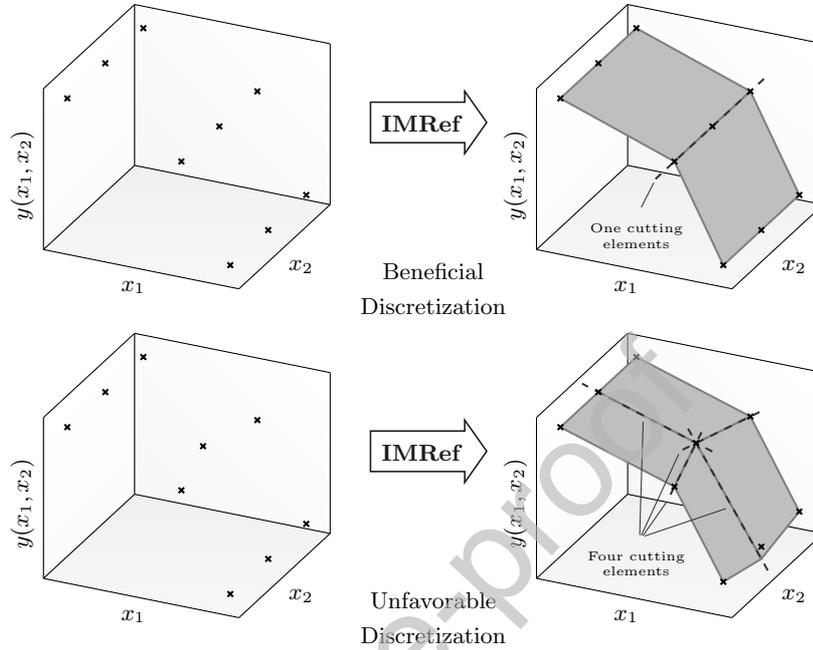


Figure 12: Sketch illustrating the dependency of the surrogate model generated by IMRef approach on discretization; Surrogate model resulting from beneficial discretization (top); Surrogate model resulting from unfavorable discretization (bottom).

To make this dependency on the discretization more evident, Figure 12 sketches the results of applying the IMRef approach to sets of 9 data points. These points describe in a minimalistic and simple way a two-dimensional piecewise linear function consisting of two linear elements. At the top of this figure a case with a beneficial discretization is shown, i.e., the points are evenly distributed and the orientation matches the shape of the piecewise linear function. This enables the IMRef approach to generate a surrogate model consisting of two linear elements and one cutting element. At the bottom of Figure 12 another case is shown where only the data point in the middle is slightly shifted, while the orientation is basically the same. As a consequence, the data points previously formed the cutting element cannot form a straight line. This in turn means that two linear elements must have the same fitting parameters α_λ and $\beta_{\delta,\lambda}$, since the equations (41) to (43) enforce continuity. Therefore, a larger

set of cutting elements and thus of linear elements is needed to allow for linear elements with different fitting parameters. The latter is necessary to describe the piecewise linear function. So the resulting surrogate model and its compactness strongly depends on the discretization, as with the IMRef approach, each vertex in V^\triangleleft must match the position of a data point in the original data set V^* . This is clearly a major disadvantage of the IMRef approach, since with more complex functional relationships (i.e., higher dimensionality and degree of non-linearities) the generation of a beneficial data set V^* becomes the problem itself.

This leads to the conclusion that the IMRef approach is superior with respect to the compactness of the resulting surrogate models only if a small number of linear elements is required to approximate the functional relationship and/or the generated data set V^* is beneficial regarding the position of its data points.

3.3. Summary and applicability

In the previous sections the approaches were applied to simple and understandable examples in order to present the basic concept of both approaches in a comprehensible way and to enable a comparison. Despite their simplicity, these examples allow the following conclusions:

The IMRef approach is an optimization-based approach that considers all data points and can therefore provide more accurate surrogate models compared to the locally based IMRed approach when it comes to a rough estimation employing only a few linear elements. However, the performance of the IMRef approach slows down exponentially with increasing dimensionality and increasing degree of non-linearity, as more and more linear elements are needed for the approximation, causing time-consuming solving of numerous large MILPs. The latter is reinforced by its strong dependency on the discretization used to generate the data set V^* . Especially for higher dimensional problems, this can lead to surrogate models with low compactness, which in turn require the solving of numerous large-scale MILPs. As a consequence, the IMRef application is only limitedly applicable to real-world applications, which generally show

a higher complexity in terms of dimensionality and degree of non-linearities.

In comparison, the IMRed approach has a higher applicability to real-world applications, as its performance is less strongly dependent on the complexity of the problem. Although higher complexity leads to more evaluation and calculation steps, the complexity of each of these steps does not increase. Furthermore, the IMRed approach is limited to just retaining some of the vertices in V^* and is therefore less dependent on the initial data set. Although the error quadric is valid for any dimensional problem, the presented classification in its current form is primarily designed to allow for an initial comparison of both approaches, with the focus on the application to one- and two-dimensional problems. Therefore, the classification probably needs to be generalized for use with higher dimensional problems.

In summary, the IMRed approach represents the more promising approach. Therefore, in the following section an exemplary application of this approach is given, although it is limited to a two-dimensional problem.

4. Exemplary application

Subsequently, the IMRed approach is applied to an industrial example – described in Obermeier et al. (2019) – to make its scope of application more evident. Here, a mixed-integer linear programming based discrete-time scheduling model for power-intensive processes is proposed, which allows for the operational planning of a single air separation unit (ASU) while considering mechanical fatigue. While Obermeier et al. (2019) considers the feasible region of the plant in a reduced manner to keep the optimization model simple, the IMRed approach provide a way to generate a more precise description of the feasible region. As the IMRed approach is data-based, a set of data points – or more precisely, operation points (OPs) – is first of all required to capture the feasible region of the respective ASU. For this purpose a highly detailed simulation model, which is implemented in Linde’s in-house simulator OPTISIM[®], is used to generate numerous, efficiency-optimized, steady-state OPs. As des-

725 cribed in Obermeier et al. (2019), this simulation model does not only contain
equations for capturing the basic heat and mass balance, but also detailed equa-
tions for capturing the realistic behavior of the installed equipment in off-design
operation. Consequently, a set of OPs – generated with this simulation model
– can capture the realistic behavior of the respective ASU (steady-state) within
730 its operational envelope (feasible region) as well as its boundaries. To ensure
this, a reasonably large set of data points is created, which is evenly distributed
over the entire feasible range and captures any non-linearity of the simulation
model sufficiently (as recommended at the beginning of section 2). Note that
such a simulation model must be replaced by a surrogate model, as it is too
735 complex to be used directly in such a scheduling formulation.

For this exemplary application, the OPs are regarded in a simplified manner
as three-dimensional data points, which are located in a two-dimensional pro-
duct space and have a corresponding power consumption each, i.e., $n = 3$ and
 $d = 2$. Figure 13 shows the plant's performance within this two-dimensional
740 product space in the form of the generated data (278 OPs). These data points
capture the mode of regular operation (reg), which is considered in the mode-
based formulation of Obermeier et al. (2019) by only 4 OPs. In this previous
contribution, these 4 OPs are used to form a piecewise linear surrogate model
consisting of three one-dimensional linear elements, which form a path through
745 the product space of the reg mode. This path of lines represents the product
space only pseudo two-dimensionally. In order to enable a real two-dimensional
representation, the lines are replaced by triangles, i.e., by two-dimensional linear
elements. Consequently, the subregions of the reg mode are spanned via inter-
polation between three instead of two points, whereby each of these subregions
750 is characterized by one binary variable in analogy to the scheduling formulation
proposed in Obermeier et al. (2019). The application of a *Delaunay triangulation*
on the 278 OPs results in a highly accurate piecewise-linear surrogate
model consisting of 511 linear elements. These elements could be used in the
scheduling formulation, as the combination of these already forms an accurate
755 and continuous description of the entire product space. In this form, however,

embedding the surrogate model into the scheduling formulation requires a high number of binary variables and thus leads to a large-scale optimization problem with high computational costs. That is due to the fact that the scheduling formulation is based on a disjunctive programming formulation (cf. Obermeier et al. (2019)) and thus a binary variable must exist for each linear element and time step to specify the active linear element at any time step. In contrast, the surrogate model resulting from the IMRed approach requires only 35 elements (cf. Figure 13 (right)), while having specified $\check{\epsilon} = 0.3\%$ as upper limit of the mean relative deviation. Consequently, the scheduling formulation is more compact and thus shows smaller computational costs. Plus, these 35 linear elements also entirely cover the operational envelope, which was originally spanned by the 278 OPs, and thus allowing for a complete description of the product space.

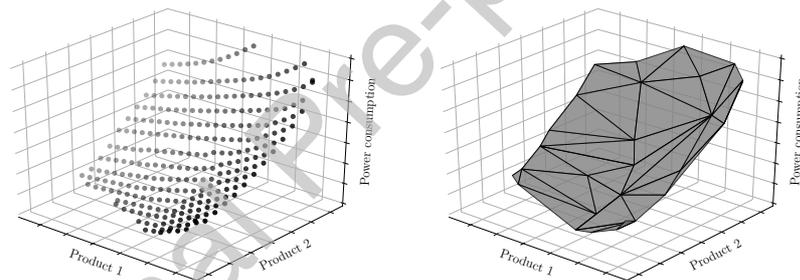


Figure 13: Performance of ASU in two-dimensional product space; set of OPs generated with detailed simulation model (left), and surrogate model resulting from IMRed approach (right).

In case this surrogate model is used in the aforementioned scheduling formulation to evaluate the CASE A-w/o defined in Obermeier et al. (2019), the operational profile depicted in Figure 14 is obtained. This case represents a scenario in which operational planning is carried out with the objective of reducing operational expenditures, without taking into account any additional constraints regarding mechanical fatigue. For further details on CASE A-w/o, please refer to the case definition in our previous contribution. However, in deviation from this definition, in this contribution the plant is operated at one operating mode only (reg mode), in order to keep the focus on the generated surrogate model.

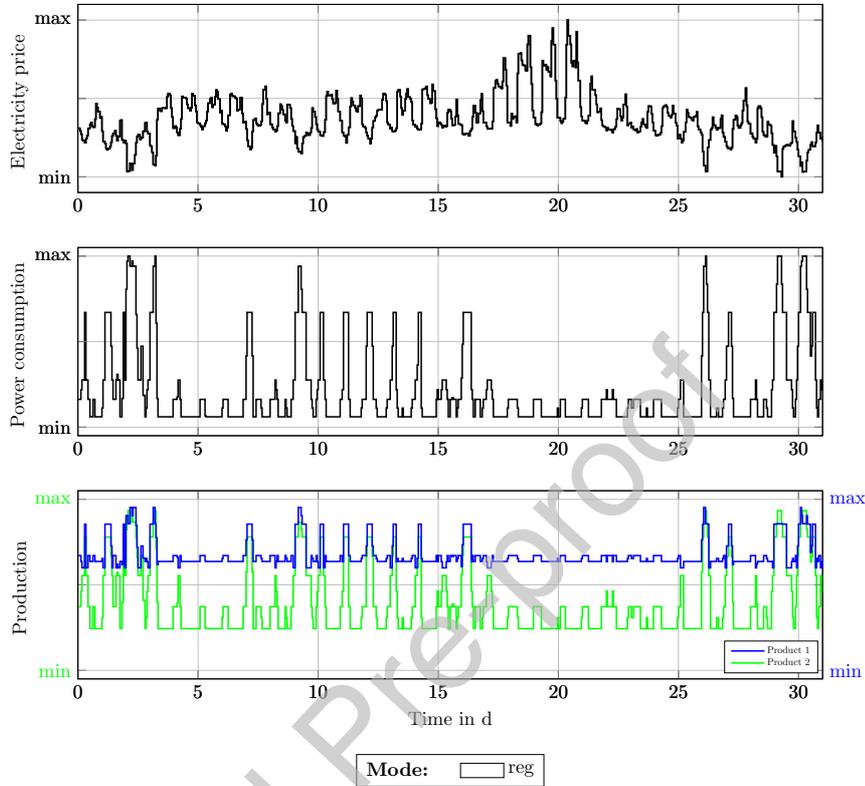


Figure 14: Optimized operational profile based on CASE A-w/o defined in Obermeier et al. (2019). With the exception that the plant is only operated in its regular operating mode (reg). The electricity price as well as the optimized power consumption and production rate is presented for the observation period (January, 2016, European Energy Exchange).

As expected, production rates and thus energy consumption are reduced in times of higher energy prices, while they are increased in times of lower energy prices (cf. Figure 14). Furthermore, the operational planning is no longer limited to a pseudo two-dimensional description of the product space, i.e., a wide variety of OPs can be used. Nevertheless, only OPs are applied, which can be described by the 4 linear elements highlighted in Figure 15. This is mainly due to the fact that in CASE A-w/o a steady sales volume (red marked point in Figure 15) is assumed to keep the example simple. Hence, only those linear elements are used with which, on average, a sufficient production can be guaranteed and at the same time the most cost-effective operation can

be achieved. For instance, the OP with minimal power consumption (green marked point in Figure 15) and the respective linear element is not used even when energy prices are high. This is because the power consumption does not reduce linearly with the production volume. Therefore this OP has the minimum power consumption, but its specific power consumption in terms of product volume is higher compared to other OPs. This becomes obvious when looking at Figure 14, as OPs with almost minimal power consumption are used here, which nevertheless still have relatively high production volumes. Consequently, the green marked OPs represents an inefficient OP and thus is not used. In contrast, in times of low energy prices, the OP with maximum power consumption (blue marked point in Figure 15) is used, since also high production rates can be achieved with it.

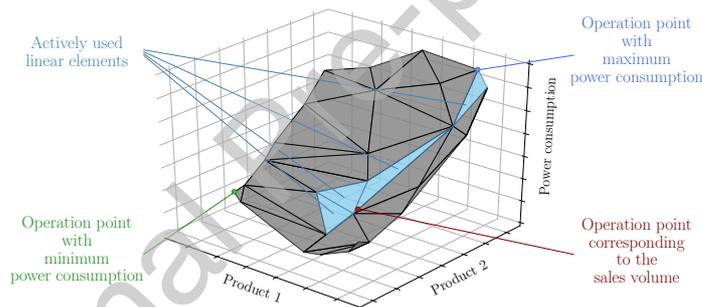


Figure 15: Surrogate model resulting from IMRed approach with depiction of specific operation points and highlighting of those linear elements which are actively used in the optimized operational profile (CASE A-w/o).

As this example illustrates, the IMRed approach represents a way to generate piecewise linear surrogate models. These provided a more accurate representation of the functional behavior than using a linear approximation (which is commonly applied in scheduling and planning formulations, cf. section 1), while keeping the computational complexity at an appropriate level by using as few linear elements as necessary. This is proven by Table 2, which lists characteristics of three surrogate models and their associated optimization problems. Before taking a closer look at this table, note that the figure in this table are averages, as the scheduling problem is decomposed into smaller optimization problems

(cf. Obermeier et al. (2019)). The first line of this table shows that using a simple linear approximation as a surrogate model leads to optimization problems of small size. These MILPs can be solve in less than one second. However, the mean relative deviation $\tilde{\epsilon}$ of a surrogate model – applying only one element to approximate the non-linear behavior of the ASU – is 2.4%. Replacing this linear approximation with a surrogate model consisting of 35 elements, reduces the mean relative deviation $\tilde{\epsilon}$ by 2.1% and increases the solution times by an average of 4.5s due to the larger size of the optimization problems. However, a further reduction of $\tilde{\epsilon}$ by 0.3% requires the use of 511 elements and thus leads to an disproportional rise of the solution time. Consequently, the surrogate model generated with the IMRed approach (35 linear elements) allows for higher accuracy while keeping the solution time at a still acceptable level.

Table 2: Characteristics of a surrogate model and the associated optimization problem using 1, 35, or 511 linear elements.

| Number of linear elements | $\tilde{\epsilon}$ in % | Number of variables ¹ | Number of constraints ¹ | Solution time ² in s |
|---------------------------|-------------------------|----------------------------------|------------------------------------|---------------------------------|
| 1 | 2.4 | 2000 | 3500 | 0.5 |
| 35 | 0.3 | 30000 | 26000 | 5 |
| 511 | 0.0 | 430000 | 245000 | 500 |

¹ Rounded number in the respective optimization problem;

² The averaged over all optimization problems resulting from the moving horizon approach.

5. Conclusions

In this contribution two approaches have been proposed for transforming well-known, but complex non-linear functional relationship into a piecewise linear approximation that can be embedded in linear programming optimization problem (e.g., a scheduling problem). This allows for a more precise approximation of the functional relationship than the commonly used linear approximation, while keeping the computational complexity of the resulting optimization problem on an appropriate level.

The basic idea of both approaches is the determination of a reduced set of
830 data points that provides an appropriate approximation of the original data via
linear interpolation. In this form the approximation is ideally suited for linear
based optimization problems using disjunctive programming formulations. Both
approaches differ significantly in their underlying concepts: The first one is an
adaptation of a numeric algorithm originating from computational geometry
835 and object modeling, whereby a fine mesh is iteratively reduced by contracting
edges of this mesh. Here referred to as iterative mesh reduction (IMRed). In
contrast to the first approach, the second one is based on a mixed integer linear
programming formulation. A mesh – consisting of an *a priori* defined number of
convex linear elements – is fitted to the data points. By incrementally increasing
840 the number of elements, this so-called incremental mesh refinement (IMRef)
generates an approximation with steadily increasing accuracy.

Both approaches were applied to generic data sets and consequently compared
assessing each approach in terms of the aspects: i) functionality, ii) computa-
tional efficiency, and iii) compactness of resulting surrogate models. Regarding
845 the first aspect it was shown that both approaches generate valid surrogate
models. In order to evaluate the second aspect, each approach was applied to
several data sets showing an increasing cardinality, i.e., a rising number of data
points. Here, it was found that the IMRef approach shows a stronger exponen-
tial rise of the computation time with increasing cardinality, especially in data
850 sets having more independent input variables. It was concluded that in terms
of computational efficiency the IMRed approach appears superior to the IMRef
approach. In this contribution, the third aspect is characterized by the num-
ber of data points, which are assigned to the resulting surrogate model. This
number depends on i) the desired accuracy of the surrogate model, ii) on the
855 non-linearity captured by the data set, and iii) the efficiency of the respective
approach. Therefore, the third aspect was evaluated for each approach by in-
vestigating the accuracy of the surrogate model as a function of the number of
assigned data points, while having specified a fixed generic data set. Here it
was concluded that the IMRef approach may be superior with respect to the

compactness of the resulting surrogate model. Unfortunately, this is only the case if the approximation consists of a very small number of linear elements and/or the distribution of the data points in the original data set is beneficial for the approach, since the IMRef approach is strongly dependent on the position of each data point. As result of this assessment, it can be concluded that the IMRed approach is the more promising approach when it comes to use in real-world applications.

The IMRed approach uses a classification of the edges in order to prevent a degradation of open boundaries. However, in its current form this classification is primarily intended to allow for an initial evaluation of the proposed approaches. In future work, this classification shall be more generalized to render it applicable to higher dimensional data sets. Besides, the approach shall be applied to data of a complex non-linear functional relationship to generate a piecewise linear approximation being used in a linear programming optimization problem. In doing so, it can be evaluated how the results as well as the computational performance of the optimization problem is affected by the piecewise linear surrogate model and its chosen accuracy.

6. Acknowledgement

The project on which this report is based was supported by funds from the German Federal Ministry of Education and Research, funding code 03SFK3X1. Responsibility for the content of this publication is assumed by the author.

Nomenclature**General — Vectors**

| | |
|-----------------------------|--|
| \mathbf{v} | Vertex/data point of simplex |
| \mathbf{x} | Vector of input values of functional relationship |
| ⁸⁸⁵ \mathbf{y} | Vector of output values of functional relationship |

General — Indices

| | |
|-----|---|
| i | Index of first vertex \mathbf{v} or linear element λ |
| j | Index of second vertex \mathbf{v} or linear element λ |

General — Sets

| | |
|-----------------------------|--|
| ⁸⁹⁰ S | Set of all simplices σ |
| V | Set of all vertices \mathbf{v} |
| V^* | Original set of all vertices |
| V^{\triangleright} | Set of vertices required to represent PLFs |
| V^{\triangleleft} | Resulting set of all vertices showing a reduced number of vertices |
| ⁸⁹⁵ V_{σ} | Set of vertices of simplex σ |

General — Parameters

| | |
|----------------------------|---|
| $\bar{\epsilon}$ | Mean relative deviation |
| $\bar{\epsilon}_{\dagger}$ | Smallest achievable mean relative deviation |
| $\check{\epsilon}$ | Upper limit of mean relative deviation |
| ⁹⁰⁰ μ | Homogeneous step size |
| d | Dimensionality of the input vector \mathbf{x} |
| n | Dimensionality of the total space |
| t | Computation time |

IMRed — Vectors

| | |
|--|--|
| ⁹⁰⁵ \mathbf{b}_{ϵ} | Vector for defining fundamental quadric of edge ϵ in its generalized form |
| \mathbf{b} | Vector for defining fundamental quadric in its generalized form |
| \mathbf{e} | Orthonormal basis |
| \mathbf{p} | Point in n -dimensional space |

| | | |
|-----|---------------------------|--|
| 910 | \mathbf{z}^* | Ideal new position/point in n -dimensional space |
| | \mathbf{z} | New position/point in n -dimensional space |
| | $\bar{\mathbf{v}}$ | New vertex resulting from edge contraction |
| | IMRed — Matrices | |
| | \mathbf{A}_ε | Matrix for defining fundamental quadric of edge ε in its generalized |
| 915 | | form |
| | \mathbf{A} | Matrix for defining fundamental quadric in its generalized form |
| | \mathbf{I} | Identity matrix |
| | IMRed — Indices | |
| | σ | Simplex |
| 920 | ε | Edge of a simplex σ consisting of two vertices |
| | IMRed — Sets | |
| | E | Set of all edges |
| | E^- | Reduced set of all edges |
| | E^B | Set of boundary edges |
| 925 | E^C | Set of crossing edges |
| | E^H | Set of hull edges |
| | E^I | Set of inner edges |
| | E^M | Set of mixed edges |
| | E^O | Set of outer edges |
| 930 | $S_{\mathbf{v}}$ | Set of simplices to whom the vertex \mathbf{v} is assigned |
| | V^B | Set of boundary vertices |
| | V^H | Set of hull vertices |
| | V^I | Set of inner vertices |
| | V^O | Set of outer vertices |
| 935 | V_ε | Set of vertices which are assigned to the edge ε |
| | IMRed — Parameters | |
| | ω_σ | The d -dimensional content of the simplex σ |
| | c | Parameter for defining fundamental quadric in its generalized form |

| | |
|---|--|
| c_ε | Parameter for defining fundamental quadric of edge ε in its generalized form |
| 940 $Q_{\mathbf{p}}(\mathbf{z})$ | Fundamental quadric $Q_{\mathbf{p}}$ at the point \mathbf{p} reflecting the geometric error caused by moving the point \mathbf{p} to the position \mathbf{z} |
| $Q_{\mathbf{v}}$ | Fundamental quadric of one vertex \mathbf{v} of a simplex σ |
| Q_σ | Fundamental quadric of a simplex σ |
| 945 Q_ε | Fundamental quadric of edge ε |
| IMRef — Vectors | |
| $\mathbf{x}_{\mathbf{v}}$ | Vector of input values of any vertex \mathbf{v} |
| $\mathbf{y}_{\mathbf{v},\lambda}^\triangleleft$ | Vector of approximated output values of any vertex \mathbf{v} in the respective element λ |
| 950 IMRef — Indices | |
| δ | Dimension |
| λ | Linear element |
| σ | Simplex |
| $\tilde{\lambda}$ | Cutting element |
| 955 IMRef — Sets | |
| \tilde{L} | Set of cutting elements $\tilde{\lambda}$ |
| L | Set of linear elements λ |
| IMRef — Parameters & Variables | |
| α_λ | Constant fitting parameter for linear element λ |
| 960 $\beta_{\delta,\lambda}$ | Fitting parameter for linear element λ specific for dimension δ |
| $\check{\epsilon}_{\mathbf{v},\lambda}$ | Modified relative deviation at vertex \mathbf{v} regarding linear element λ |
| $\chi_{\mathbf{v},\tilde{\lambda}}^+$ | Positive slack variable at vertex \mathbf{v} being assigned to respective cutting element $\tilde{\lambda}$ |
| $\chi_{\mathbf{v},\tilde{\lambda}}^-$ | Negative slack variable at vertex \mathbf{v} being assigned to respective cutting element $\tilde{\lambda}$ |
| 965 $\epsilon_{\mathbf{v},\lambda}$ | Relative deviation at vertex \mathbf{v} regarding linear element λ |
| Ω | Big-M parameter controlling modified relative deviation $\check{\epsilon}_{\mathbf{v},\lambda}$ |
| Ψ | Big-M parameter controlling the slack variables $\chi_{\mathbf{v},\tilde{\lambda}}^+$ and $\chi_{\mathbf{v},\tilde{\lambda}}^-$ |

| | |
|-------|----------------------------------|
| y | Binary variable |
| 970 l | Number of linear elements in L |

References

- Arce-Medina, E., & Paz-Paredes, J. I. (2009). Artificial neural network modeling techniques applied to the hydrodesulfurization process. *Mathematical and Computer Modelling*, 49, 207 – 214. URL: <http://www.sciencedirect.com/science/article/pii/S0895717708001398>.
 975 doi:<https://doi.org/10.1016/j.mcm.2008.05.010>.
- Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM TRANSACTIONS ON MATHEMATICAL SOFTWARE*, 22, 469–483.
- 980 Bellman, R. (2015). *Adaptive Control Processes: A Guided Tour*. Princeton Legacy Library. Princeton University Press. URL: <https://books.google.de/books?id=iwbWCgAAQBAJ>.
- Bhosekar, A., & Ierapetritou, M. (2018). Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers & Chemical Engineering*, 108, 250 – 267. URL: <http://www.sciencedirect.com/science/article/pii/S0098135417303228>.
 985 doi:<https://doi.org/10.1016/j.compchemeng.2017.09.017>.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. URL: <https://doi.org/10.1023/A:1010933404324>. doi:10.1023/A:1010933404324.
- 990 Burnak, B., Katz, J., Diangelakis, N. A., & Pistikopoulos, E. N. (2018). Simultaneous process scheduling and control: A multiparametric programming-based approach. *Industrial & Engineering Chemistry Research*, 57, 3963–3976. URL: <https://doi.org/10.1021/acs.iecr.7b04457>. doi:10.1021/acs.iecr.7b04457. arXiv:<https://doi.org/10.1021/acs.iecr.7b04457>.

- 995 Dyn, N., Levin, D., & Rippa, S. (1986). Numerical procedures for surface fitting of scattered data by radial functions. *SIAM Journal on Scientific and Statistical Computing*, 7. doi:10.1137/0907043.
- Fang, H., & Horstemeyer, M. (2006). Global response approximation with radial basis functions. *Engineering Optimization*, 38, 407–424. doi:10.1080/1000 03052150500422294.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *Ann. Statist.*, 19, 1–67. URL: <https://doi.org/10.1214/aos/1176347963>. doi:10.1214/aos/1176347963.
- Garland, M., & Heckbert, P. S. (1997). Surface simplification using quadric error metrics. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '97* (pp. 209–216). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co. URL: <http://dx.doi.org/10.1145/258734.258849>. doi:10.1145/258734.258849.
- 1010 Garland, M., & Heckbert, P. S. (1998). Simplifying surfaces with color and texture using quadric error metrics. In *Proceedings of the Conference on Visualization '98 VIS '98* (pp. 263–269). Los Alamitos, CA, USA: IEEE Computer Society Press. URL: <http://dl.acm.org/citation.cfm?id=288216.288280>.
- Garland, M., & Zhou, Y. (2005). Quadric-based simplification in any dimension. *ACM Trans. Graph.*, 24, 209–239. URL: <http://doi.acm.org/10.1145/1061347.1061350>. doi:10.1145/1061347.1061350.
- 1015 Hart, W. E., Laird, C. D., Watson, J.-P., Woodruff, D. L., Hachebeil, G. A., Nicholson, B. L., & Siirola, J. D. (2017). *Pyomo-optimization modeling in python* volume 67. (2nd ed.). Springer Science & Business Media.
- 1020 Hart, W. E., Watson, J.-P., & Woodruff, D. L. (2011). Pyomo: modeling and solving mathematical programs in python. *Mathematical Programming Computation*, 3, 219–260.

- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA: Springer New York Inc.
- Ierapetritou, M. G., Wu, D., Vin, J., Sweeney, P., & Chigirinskiy, M. (2002). Cost minimization in an energy-intensive plant using mathematical programming approaches. *Industrial & Engineering Chemistry Research*, *41*, 5262–5277. URL: <http://dx.doi.org/10.1021/ie0111012b>. doi:10.1021/ie0111012b. arXiv:<http://dx.doi.org/10.1021/ie0111012b>.
- Karwan, M. H., & Kebblis, M. F. (2007). Operations planning with real time pricing of a primary input. *Computers & Operations Research*, *34*, 848 – 867. URL: <http://www.sciencedirect.com/science/article/pii/S0305054805001619>. doi:<http://doi.org/10.1016/j.cor.2005.05.014>. Logistics of Health Care Management Part Special Issue: Logistics of Health Care Management.
- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, *52*, 119–139. URL: <http://dx.doi.org/10.2307/3006914>. doi:10.2307/3006914.
- Mitra, S., Grossmann, I. E., Pinto, J. M., & Arora, N. (2012). Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. *Computers & Chemical Engineering*, *38*, 171–184. URL: <http://dx.doi.org/10.1016/j.compchemeng.2011.09.019>. doi:10.1016/j.compchemeng.2011.09.019.
- Mitra, S., Pinto, J. M., & Grossmann, I. E. (2014). Optimal multi-scale capacity planning for power-intensive continuous processes under time-sensitive electricity prices and demand uncertainty. part i: Modeling. *Computers & Chemical Engineering*, *65*, 89 – 101. URL: <http://www.sciencedirect.com/science/article/pii/S0098135414000283>. doi:<http://dx.doi.org/10.1016/j.compchemeng.2014.01.016>.

- Mitra, S., Sun, L., & Grossmann, I. E. (2013). Optimal scheduling of industrial combined heat and power plants under time-sensitive electricity prices. *Energy*, *54*, 194 – 211. URL: <http://www.sciencedirect.com/science/article/pii/S0360544213001448>.
1055 doi:<https://doi.org/10.1016/j.energy.2013.02.030>.
- Oberdieck, R., Diangelakis, N. A., Nascu, I., Papathanasiou, M. M., Sun, M., Avraamidou, S., & Pistikopoulos, E. N. (2016). On multi-parametric programming and its applications in process systems engineering. *Chemical Engineering Research and Design*, *116*, 61 – 82. URL: <http://www.sciencedirect.com/science/article/pii/S0263876216303276>.
1060 doi:<https://doi.org/10.1016/j.cherd.2016.09.034>. Process Systems Engineering - A Celebration in Professor Roger Sargent's 90th Year.
- Obermeier, A., Windmeier, C., Esche, E., & Repke, J.-U. (2019). A discrete-time scheduling model for power-intensive processes taking fatigue of equipment into consideration. *Chemical Engineering Science*, *195*, 904 – 920. URL: <http://www.sciencedirect.com/science/article/pii/S0009250918307553>. doi:<https://doi.org/10.1016/j.ces.2018.10.036>.
1065
- Papadopoulos, H., Vovk, V., & Gammerman, A. (2011). Regression conformal prediction with nearest neighbours. *J. Artif. Intell. Res.*, *40*, 815–840.
1070
- Raman, R., & Grossmann, I. (1994). Modelling and computational techniques for logic based integer programming. *Computers & Chemical Engineering*, *18*, 563 – 578. URL: <http://www.sciencedirect.com/science/article/pii/0098135493E00107>. doi:[http://dx.doi.org/10.1016/0098-1354\(93\)E0010-7](http://dx.doi.org/10.1016/0098-1354(93)E0010-7).
1075
- Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, *48*, W07401. doi:10.1029/2011WR011527.
- Strang, G. (2006). *Linear algebra and its applications*. Belmont, CA: Thomson, Brooks/Cole.
1080

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288. URL: <http://www.jstor.org/stable/2346178>.
- Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solutions of ill-posed problems*.
1085 Washington, D.C.: John Wiley & Sons, New York: V. H. Winston & Sons.
Translated from the Russian, Preface by translation editor Fritz John, Scripta
Series in Mathematics.
- Tsay, C., & Baldea, M. (2019). 110th anniversary: Using data to
bridge the time and length scales of process systems. *Industrial*
1090 *& Engineering Chemistry Research*, 58, 16696–16708. URL: <https://doi.org/10.1021/acs.iecr.9b02282>. doi:10.1021/acs.iecr.9b02282.
arXiv:<https://doi.org/10.1021/acs.iecr.9b02282>.
- Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. New York,
NY, USA: Springer-Verlag New York, Inc.
- 1095 Wang, G. G., & Shan, S. (2007). Review of metamodeling techniques in
support of engineering design optimization. *Journal of Mechanical Design*,
129, 370–380. URL: <http://dx.doi.org/10.1115/1.2429697>. doi:10.1115/
1.2429697.
- Yang, L., Liu, S., Tsoka, S., & Papageorgiou, L. G. (2016). Mat-
1100 hematical programming for piecewise linear regression analysis.
Expert Systems with Applications, 44, 156 – 167. URL: <http://www.sciencedirect.com/science/article/pii/S0957417415005904>.
doi:<https://doi.org/10.1016/j.eswa.2015.08.034>.
- Zhang, Q., Grossmann, I. E., Heuberger, C. F., Sundaramoorthy, A., &
1105 Pinto, J. M. (2015). Air separation with cryogenic energy storage: Op-
timal scheduling considering electric energy and reserve markets. *AIChE*
Journal, 61, 1547–1558. URL: <http://dx.doi.org/10.1002/aic.14730>.
doi:10.1002/aic.14730.

- Zhang, Q., Grossmann, I. E., Sundaramoorthy, A., & Pinto, J. M. (2016a).
1110 Data-driven construction of convex region surrogate models. *Optimization and Engineering*, *17*, 289–332. URL: <http://dx.doi.org/10.1007/s11081-015-9288-8>. doi:10.1007/s11081-015-9288-8.
- Zhang, Q., Sundaramoorthy, A., Grossmann, I. E., & Pinto, J. M. (2016b).
1115 A discrete-time scheduling model for continuous power-intensive process networks with various power contracts. *Computers & Chemical Engineering*, *84*, 382 – 393. URL: <http://www.sciencedirect.com/science/article/pii/S0098135415003142>. doi:<https://doi.org/10.1016/j.compchemeng.2015.09.019>.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the
1120 elastic net. *Journal of the Royal Statistical Society, Series B*, *67*, 301–320.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

Andreas Obermeier: Conceptualization, Methodology, Writing- Original draft, Editing. **Nikolaus Vollmer:** Software, Validation, Investigation. **Christoph Windmeier:** Supervision, Reviewing. **Erik Esche:** Supervision, Reviewing. **Jens-Uwe Repke:** Supervision, Reviewing.

Journal Pre-proof