

Inverting Prediction Models in Micro Production for Process Design

Phil Gralla^{1,*}, Iwona Piotrowska-Kurczewski¹, Daniel Rippel², Michael Lütjen², and Peter Maaß^{1,3}

¹University of Bremen, Bremen, Center for Industrial Mathematics (ZeTeM), Postfach 330440, 28334 Bremen, Germany

²BIBA – Bremer Institut für Produktion und Logistik GmbH at the University of Bremen, Hochschulring 20, 28359 Bremen, Germany

³University of Bremen, Bremen, MAPEX Center for Material and Processes, Bremen, Germany

Abstract. Databased prediction models are used to estimate a possible outcome for previously unknown production parameters. These forward models enable to test new production designs and parameters virtually before applying them in the real world. Cause-effect networks are one way to generate such a prediction model. Multiple inputs and stages are being connected to one large prediction model. The functional behaviour and correlation of inputs as well as outputs is obtained through data based learning. In general, these models are non-linear and not invertible, especially for micro cold forming processes. While already being useful in process design, such models have their highest impact if inverted to find process parameters for a given output. Combining methods from the mathematical field of inverse problems as well as machine learning, a generalized inverse can be approximated. This allows finding process parameters for a given output without inverting the model directly but still using inherent information of the forward model. In this work, Tikhonov functionals are used to perform a parameter identification. The classical approach is altered by changing the discrepancy term to incorporate tolerances. Thereby, small deviations of a certain pattern are being neglected and the parameter finding process is being stabilized. In addition, different types of regularization are taken into consideration. Besides theoretical aspects of this method, examples are provided to demonstrate advantages and boundaries of an application for the process design in micro cold forming processes.

Keywords. Predictive Model, Optimization, Process Control

1 Introduction

Modern production processes for micro components depend on many process parameters. Each of them is affecting process quality and cost in its own way. Estimating the outcome for a set of process parameters needs a large amount of experiments, as former knowledge from macro production can only partly be applied due to size effects (see [1]). Prediction models can help reducing necessary experiments by using knowledge from other parameter sets as well as former experiments to estimate the influence of a new parameter set on the process outcome.

In this work, cause-effect networks are used to generate a process model and allow the exhibition of new process strategies. These networks consist of a set of interconnected technological and logistic parameters, e.g. representing forces, times or material properties, which are relevant for the process. Thereby, each parameter contains a prediction model, allowing to calculate its value using values from connected parameters.

Cause-effect networks allow estimating the outcome of a complex process for a new parameter set, however they are generally not invertible. To find a suitable set of parameters for a desired output new strategies need to be developed. In mathematics this is considered an inverse

problem where the forward operator and an output are known but the causing parameter is to be identified. In application these are called ill-posed problems, which are either non-invertible or amplify small perturbations of the output in the reconstructed parameter. Since prediction models for complex processes, such as production processes for micro components, tend to lead to ill-posed problems, this issue needs to be solved via regularization methods. Different approaches to deal with ill-posed problems have been developed over time (see [2,3,4]). In this work Tikhonov type functionals will be used to tackle the ill-posed nature of the operators at hand.

Besides estimating the expected value of a production process, its variance in the output is of great interest as well. To ensure a low rate of non-usable components the variance should stay within a predefined area. Thus, the classic approach for parameter identification from inverse problems needs to be adjusted. In this work Tikhonov functionals are being altered in order in to include variances and to further stabilize the parameter identification.

The structure of this work follows the procedure of the presented method as shown in Fig. 1. First prediction models and how these are obtained is described. In this section the state of the art is summarized focusing on the models which are later used in this work. In the following

* Corresponding author: phil.gralla@uni-bremen.de

section the planning tool μ -ProPIAn, which generates process models through cause-effect networks, is inaugurated. Cause-effect relations are models by the means of prediction models as introduced in section 2. In addition mathematical notations are introduced and mathematical properties stated. It forms a basis for section 4 where the introduced notations and stated properties are needed.

In the next section, section 4, the forward model is used for parameter identification. A short overview about parameter identification and its challenges is provided before the applied method is introduced. This method uses classic approaches and alters them to further stabilize the parameter identification as well as incorporate variations in the process outcome. Section 5 provides two examples to demonstrate its functionality as well as difference to existing methods.

The overall objective of this work is to derive a reliable method for process parameter identification for micro components while keeping the necessary amount of experiments to a minimum.

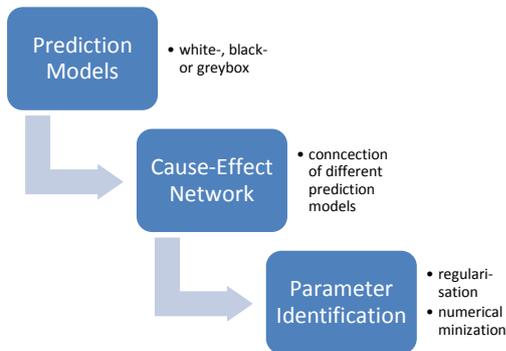


Fig. 1. Workflow of building prediction models and cause-effect networks for performing parameter identification.

2 Prediction Models

Prediction models take an input parameter and map this parameter on a related output. They are the connection of input to output. In this work, prediction models are restricted to production steps and chains. The input parameters are used material and process parameters and the output is the state of manufactured good.

Different approaches to form a suitable model exist and can be classified into two main groups, data based models, also referred to as black box models, and models based on physical laws, also known as white box models. A combination of both approaches is called a grey box model. While each approach has its own advantages and disadvantages, the following section will focus on black box models, however the results and methods of the following sections can be applied on white and grey box models as well, as long as the stated necessary mathematical properties are fulfilled.

Black box models are based on experimental data and are thus often statistical regression models. Other approaches include but are not restricted to linear- or spline interpolation as well as convolution. For this work local regression models are used to obtain a prediction

model, also known as LOWESS (see [10]). Instead of fitting one regression model for the whole data, for each point a separate model is fitted that takes only the nearest neighbours into account.

Rather than giving all points in the neighbourhood equal weight, weights that die off smoothly with distance from the target point can be used. One way to achieve this is by using Epanechnikov quadratic kernels as defined in (1) and (2).

$$K_\lambda(x_0, x) = D\left(\frac{|x-x_0|}{\lambda}\right) \quad (1)$$

$$D(t) = \begin{cases} \frac{3}{4}(1-t^2) & \text{if } |t| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Locally weighted polynomial regression solves a separate weighted least square problem at each target point x_0 by minimizing function (3). The maximum degree of the polynomial is denoted d .

$$\min_{\alpha(x_0), \beta_j(x_0), j=1, \dots, d} \sum_{i=1}^N K_\lambda(x_0, x_i) \cdot J \quad (3)$$

$$J = [y_i - \alpha(x_0) - \sum_{j=1}^d \beta_j(x_0)x_i^j]^2$$

An example of a regression using the above method is given in figure 2. The polynomial degree is set to 2 and $\lambda = 0.2$. Data was generated using a sinus function with added normal noise. The applied method will produce a differentiable function as a regression. For graphical representation of the regression, it is evaluated at a set number of points and linear interpolation is used to draw the missing lines in the graph to achieve the desired resolution, which results in a graph that seems to be only continuous instead of differentiable. For the following parameter identification in section 5, each function evaluation will be done by fitting the desired local quadratic regression. It is therefore necessary to save all data points instead of only one regression function.

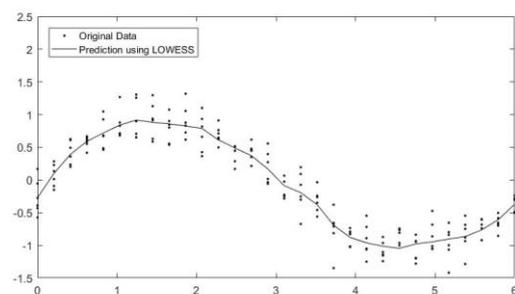


Fig. 2. Example of locally weighted regression to estimate the expected value of a sinus function with additive normal random noise.

This type of regression is very flexible as it does not require one function that fits all data but instead uses a type of smoothing operator. It is therefore suited for complex processes with none or insufficient theoretical models. It is however computational intensive and does not provide an easily comparable regression function, which depends on few parameters, as do classic least

square methods (for further detail refer to [11]). Due to the complexity of micro production processes, LOWESS with Epanechnikov quadratic kernels and local polynomial degree of 2 will be used as a default option in cause-effect networks generated by the μ -ProPIAn software, which will be described in more detail later.

So far the prediction model is estimating the expected value of the process outcome. For its application in process design additional information of variance is needed as well. For data based prediction models this can be achieved by learning a second model on the sample variance of the given data, i.e. the output of trainings data is $var_j = VAR(y_{j,i})$, where $y_{j,i}$ is the i -th measurements of the input x_i . In this work, LOWESS will be used to find a suitable approximation for the variance over the whole function domain and to keep consistency in the applied methods.

3 Cause-Effect Models

The introduced prediction models estimate the expected value and variance of one process by combining multiple processes into a process chain to form a new process. Cause-effect networks can be used to combine prediction models of single process into a larger process chain and estimate the expected value of the whole process chain. This section introduces one methodology to generate such cause-effect networks and outlines some mathematical basics of such networks.

3.1 Micro – Process Planning and Analysis (μ -ProPIAn)

This article focuses on cause-effect networks generated as part of the methodology “Micro – Process Planning and Analysis” (μ -ProPIAn) (c.f. [5] for further details on the methodology). μ -ProPIAn covers all design phases from the process and material flow planning to the configuration and evaluation of the processes and process chain models. The methodology itself consists of a modeling notation, a procedure model as well as a set of methods and tools for the evaluation of the corresponding models.

The modeling notation consists of three views, representing different levels of detail. The first view focuses on top level process chains. This view’s notation closely follows the classic notation of process chains, as described in [6]. In contrast to the classical approach, process elements as well as operations are connected using process interfaces. These interfaces, which additionally include logistic parameters. Extending the classical approach, operations act as interfaces to the second view: the material flow view.

The material flow view further details operations by assigning those material flow objects that are used to conduct the operation (e.g. machines/devices, work pieces, tools, operating supplies or workers). This enables the modelling of specific production scenarios with specified resources and therefore allows an evaluation of the models regarding logistic aspects. Therefore, μ -ProPIAn offers the option to conduct material flow

simulations based on the specified production system and the modelled process chains.

The third view focuses on the configuration of the processes and process chains using cause-effect networks. Each network consists of a set of parameters and a set of cause-effect relationships, forming a directed graph. The set of parameters consists of all technical and logistic characteristics that are relevant to describe the object’s influence on the production process. In case of work pieces, these are e.g. material properties, costs per piece or geometrical characteristics. As for production processes, these parameters include production speeds, forces or other characteristics that can be set, calculated or measured. From a modelling perspective, the cause-effect networks are modelled hierarchically. Each material flow object (work pieces, machines, tools, workers, etc.) holds its own cause-effect network, or at least a set of describing parameters. When combining these single elements to operations, process elements or process chains, higher-level cause-effect networks are created by describing additional relationships between the parameters of the networks or by connecting them to previously specified process interfaces (c.f. figure 3).

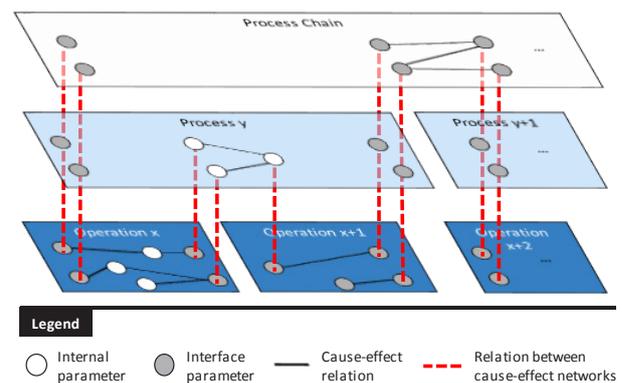


Fig. 3. Structure of Cause-Effect Networks.

The design of cause-effect networks is divided into two steps: qualitative modeling and quantification. The qualitative model of the corresponding network is created by collecting all relevant parameters and denoting their influences among each other. At this point the knowledge of a process expert can facilitate the creation of networks by point out the most relevant parameters and indicating the relationships.

The second step concerns the quantification of the cause-effect networks. The objective is to enable the propagation of different parametrizations throughout the network. In case of simple or well-known relations, μ -ProPIAn allows to input mathematical formulas directly (white-box models). However, in the area of micro manufacturing different parameters can have a more significant impact than in the macro domain, resulting in the inclusion of parameters that can be neglected in macro manufacturing. In addition, size-effects may induce a different behavior than usually observed. Therefore, it is often impossible to describe all parameters and cause-effect relations comprehensively directly.

As a result, μ -ProPIAn offers the capabilities to quantify cause-effect relations from experiment- or production data by applying methods from the areas of data mining and statistics. Therefore, the qualitative cause-effect network is subdivided into a set of sub-problems. Each of these consists of one single dependent parameter and its independent parameters. Thus objective of the quantification is the determination of a function or model to estimate the value of the dependent variable based on the values of the independent variables. To achieve this objective, μ -ProPIAn offers a set of statistical regression methods (e.g. linear or polynomial least-square regressions) or learning methods (e.g. artificial neural networks, support vector machines, regression trees or local regression methods). While statistical regression methods usually calculate a mathematical function to describe the expected value of the referring dependent variable, learning methods usually provide a model for a mean estimator. While these estimators can predict the value of the dependent variable based on the values of the independent variables, their models usually provide less insight to mechanics behind the cause-effect relation and are thus treated as black-box models. Their advantage shows in their ability to learn arbitrary relations without prior knowledge of the relation's shape (no prototype of the function has to be provided) and regardless of the problems dimensionality. In practice, the application of locally weighted linear regression (LOWESS) models has yielded promising results in estimating complex relations (e.g. in [7]).

In addition to the mean estimator, μ -ProPIAn characterizes the (local) variance for each set of training data as described in [8]. Using these variances, μ -ProPIAn applies an interpolation to estimate the local variance in addition to the expected mean value for each parameter, provided a given parametrization for the cause-effect network.

3.2 Mathematical Properties

In order to perform a parameter identification on a full prediction model, including cause-effect networks, certain mathematical properties need to be fulfilled. To understand how properties of single process steps are carried over into the full model, a proper mathematical formulation is needed.

In section 2, different ways of generating data based prediction models are introduced and the LOWESS method is explained in more detail. Each prediction model is a mathematical operator, denoted by F , that maps an input, denoted by x , on an output, denoted by y . In most cases, x, y will be vectors of scalars, but are not restricted to it, e.g. one input could be a time-dependent heat source.

For a full cause-effect network, multiple prediction models, denoted by F_j , for different production steps, are linked together. To apply a second process F_i on a process F_j the output space of F_j must match with the domain of F_i . This may not be the case if, for example, new process parameters are introduced for process F_i . To analyse the network, the first operator F_j is extended by mapping the new process parameter of F_i onto itself, i.e.

an identity operator for missing inputs is applied. Thus the new model F for the linked process is as defined in (4).

$$F = F_i(F_j(x)) \quad (4)$$

Parallel processes that do not influence each other can be linked the same way. The order does not matter as long as the input and output space are adjusted accordingly via applying identity operators.

As a result, cause-effect networks can be expressed as one operator, which inherits its mathematical properties from each prediction model within. If, for example, the network consists of a continuous operator applied on a differentiable operator, the whole network will be differentiable.

4 Parameter Identification

Prediction models are useful to find outputs for new parameters. Thus reducing the need of experiments to estimate outputs of new parameters. Finding parameters for a given output is called parameter identification. In an ideal case, this can be done by inverting the model to directly calculate parameters to a given output. In practice, this does not always work due to several reasons. The most common ones are that the model is not invertible, and that, due to model approximations or variations in measurements, the given or observed output has small deviations from the predicted output if both are based on the same parameters. These small deviations can lead to large or even unbounded errors in parameter identification. These two cases are examples of ill-posed problems (see definition and examples in [2, 3]). One class of ill-posed problems is the inversion of integral operators. As shown in earlier sections, black-box models can be represented as integral operators and thus lead to ill-posed problems for the parameter identification. To overcome the ill-posed nature of the problem at hand regularization is used. Various methods of regularization exist (for examples see [3, 4, 9]) and depend on a priori knowledge of the parameter and data. In this work Tikhonov functionals are used to stabilize the parameter identification progress.

Tikhonov functionals consist of two terms, a discrepancy and a regularization term. By minimizing both terms, weighted with a coefficient α on the regularization term, a suitable approximation of the true parameter, i.e. the parameter that causes the given or observed data, is found. Let X be the space in which the prediction model F maps onto and all data resides. The discrepancy term then describes the distance of given or observed data y^δ to the predicted data $F(x)$ to the parameter x in the space X . The regularization term $R(x)$ uses some a priori information and maps the parameter u onto the positive real numbers. A common choice for $R(x)$ is (5), where x_0 is often set to 0. For prediction models $x_0 \neq 0$ but instead a density point of known data is applicable as well to avoid extrapolation in the prediction.

$$R(x) = \|x - x_0\|_q^q, \quad 1 \leq q \leq 2 \quad (5)$$

The prediction models in this work do not only give a point estimation of the outcome but a variance as well. To fully utilize this additional information, the given data should not be a single point but a closed set of feasible data points. This set will be denoted as tolerance area of the point v^δ . To compare two sets while maintaining good numerical properties, the distance measurement given in (6) is used. This distance is differentiable and measures the overlapping area, i.e. it is only equal to zero if S_1 is a subset of S_2 . This is important to ensure that the determined parameter is within the tolerance area if such a parameter exists.

$$d(S_1, S_2) = \int_{S_1} \min_{y \in S_2} \|x - y\|_2^2 dx \quad (6)$$

In the following, $S_{F(y)}$ denotes the tolerance set of $F(y)$ obtained in the same way as the tolerance set of y^δ is obtained, which will be denoted by S_δ . For quality management the three sigma area is a common choice to define a process quality if an additive normal error is assumed. This area can be used as a tolerance set as it is closed and generated using expected value and variance, which are both estimated by the cause-effect network.

To perform the parameter identification the functional (7) is minimized. In this functional S_δ is the given target and F the cause-effect network operator. The minimiser of (7) is then the identified parameter.

$$T(x) = d(S_{F(x)}, S_\delta) + \alpha R(x) \quad (7)$$

The minimisation of (7) is done numerically using a standard non-linear solver, such as gradient decent with adaptive step size, inner point algorithm or sequentially quadratic problem solver. Depending on the properties of F the solver may need to be adapted to be able to minimize the functional. In all cases only a local minimum is identified which may not be global minimum, as all algorithms are of iterative nature and depend on the starting value. Different strategies, such as hill climbing and trust region, can be used to soften the dependency on the starting value but yield higher numerical cost.

5 Numerical Results

In this section two numerical examples are outlined. The first one is a theoretical example to showcase the influence of varying variances on parameter identification. The second example is a toy example of a production process with two process parameters and one output quality parameter. A toy example is used to illustrate the whole process of the presented method and still be useful to provide a quick understanding of the workflow of the process. In addition, the true functional dependency of input and output are known which allows to evaluate the results.

In the example a logarithmic grow of the output $y \in \mathbb{R}$ is observed as the input $x \in \mathbb{R}$ is increasing. The variance in y is increasing as the distance between x and 1 is increasing. The prediction function output including estimated variance is shown in figure 4.

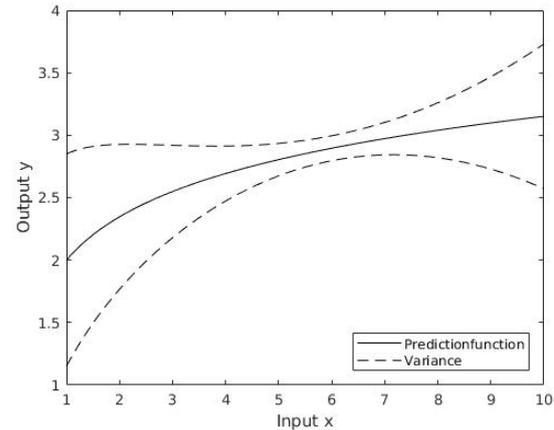


Fig. 4. Example of influence of varying variance in the output on the parameter.

The prediction function can be inverted directly to find a parameter for a desired output. In this example the desired output is $y = 3.07$ with a tolerance area of $[2.83, 0.24]$. This derived from the expected value plus minus allowed variance of 0.24. An input of $x = 8.5$ produces the desired expected value of $y = 3.07$, the variance is 0.2875 and does exceed the given boundary. To find a better parameter set considering the tolerance area a parameter identification using the method introduced in this work is performed. The identified parameter is $x = 6.74$ which produces the output $y = 2.954$ with a variance of 0.1164 and therefore a tolerance area of $[2.8376, 3.0705]$ which lies within the desired boundaries. This example shows how performing a parameter identification on the expected value does not provided the best parameter for a desired outcome with a set acceptable variance.

For the second example a two-step process is considered. The first process F_1 maps $x_1 \in \mathbb{R}$ onto $y_1 \in \mathbb{R}$. The second process F_2 maps $(x_2, x_3) \in \mathbb{R}^2$ onto $y_2 \in \mathbb{R}$. For the cause-effect network process F_2 is applied on the outcome of process F_1 where $y_1 = x_2$ is set. The expanded operator as well as the whole process model is given in (8).

$$\begin{aligned} F_1(x_1, x_3) &= (y_1, x_3)^T \\ F_2(x_2, x_3) &= y_2 \\ F(x_1, x_3) &= F_2(F_1(x_1, x_3)) \end{aligned} \quad (8)$$

The true model without noise is given in (9). This will be used to compare the numerical results based on the prediction models with the true solution. The noise is an additive normal noise on each prediction model with mean 0 and a standard deviation of 0.2. The domain of x_1 is restricted to $[0, 4]$, the domain of x_2 is restricted to $[0, 6.25]$, and the domain of x_3 is restricted to $[0, 1]$.

$$\begin{aligned} \tilde{F}_1(x_1) &= (x_1 - 1.5)^2 \\ \tilde{F}_2(x_2, x_3) &= 5 \cdot \sin(3 \cdot x_3) + \log(x_2^3 + 2) \end{aligned} \quad (9)$$

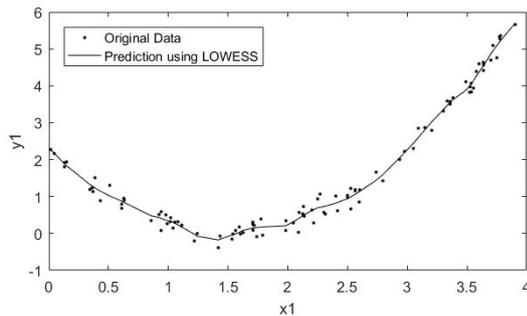


Fig. 5. Data and regression for the first process.

The first model is shown in figure 5, the second model is shown in figure 6. The output of the whole network matches figure 6 as the transformation on x_1 is not visible in such a graph.

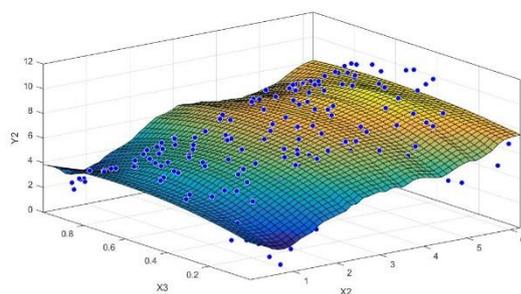


Fig. 6. Data and regression for the second process.

The variance in the data was analysed for the whole cause-effect network at once. The variance is stable for the input x_3 and decreases as the distance of x_1 to 1.5 decreases. This matches the expected behaviour of a sensitivity analysis of (9) connected as in (8). As a true parameter [1, 0.2] is set which will result in $y = 3.5241$. The variance is set to be 0.2. The calculated parameter is [0.993, 0.182] which has an estimated expected value of $y = 3.5661$ and a variance of 0.18. It is therefore within the set limit. The identified parameter does match the true parameters within a reasonable span. Without any regularization the identified parameter will be [0.89, 1.62] which has an ideal match in the expected value but does not reconstruct the true parameter so well. It is therefore more influenced by the ill-posed nature of the cause-effect network.

6 Summary and Conclusion

This work presents a new method to invert prediction models in micro production for process design. It alters classic approaches from the mathematical field of inverse problems to account for variations in production and enables a successful parameter identification for a given desired output.

The forward model is generated by cause-effect networks which are generated using the μ -ProPIAn methodology and GUI. As the generated models are usually non-linear and not invertible, special numerical methods have to be applied for a parameter identification.

To choose a suitable method further knowledge about mathematical properties of the model are needed. Therefore, prediction models and how they are derived is discussed in this work in detail. LOWESS as a method for generating data based prediction models is presented in further detail, as it is used as a default option for the μ -ProPIAn method. To demonstrate the challenges and advantages of the presented method two numerical examples are provided.

The presented method allows comprehending complex structures and dependencies in micro production and uses mathematical methods to estimate suitable parameters for a complex process with desired output. It is restricted by the amount of information, theoretical or data based, on each production step available. However, through novel approaches to parameter identification and forward modelling the amount of necessary experiments can be greatly reduced. While its application is not only limited to micro production it was developed and tailored to overcome the challenges faced in micro production.

Due to the possible ill-posed nature of prediction models alternative methods, such as Design of Experiments or Artificial Neural Networks yield a large amount of required experiments and data if applicable. In addition Neural Networks have limitations in solving ill-posed inverse problems as the inverse operator may not exist or is non-differentiable as well as non-injective (see [12]). They can be used in combination with classic approaches but are not suitable to solve the inverse problem on their own.

Overall, this work provides a first proof of a novel concept for parameter identification based on prediction models in micro production for process design. Current work focuses on optimizing numerical efficiency as well as generating new data for further validation.

Acknowledgement

The authors acknowledge the financial support by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the SFB 747 (Collaborative Research Center) "Mikrokaliumformen – Prozesse, Charakterisierung, Optimierung".

P. Gralla is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within in the framework of GRK 2224/1.

References

1. F. Vollertsen, Product Engineering, 2, 2008, 4, S.377-383
2. A. K. Louis, Inverse und Schlecht Gestellte Probleme, Teubner Studienbücher, 2012
3. A. Rieder, Keine Probleme mit Inversen Problemen – Eine Einführung in ihre stabile Lösungen, Vieweg, 2003
4. J. Bangti and P. Maaß, Inverse Problems 28, 2012, 123001
5. D. Rippel, M. Lütjen and B. Scholz-Reiter, Journal of Manufacturing Technology Management 25, 2014

6. B. Denkena and H.K. Tönshoff, Prozessauslegung und –integration in die Prozesskette, in : B. Denkena and H.K. Tönshoff (Eds.), Spanen – Grundlagen, Springer Verlag, Berlin Heidelberg, 2011, pp. 339-362
7. D. Rippel, M. Lütjen and B. Scholz-Reiter, Journal of Manufacturing Technology Management, 25, 2014
8. D. Rippel, M. Lütjen and M. Freoitag, Journal of Manufacturing Systems, 43, 2017, pp. 79-87
9. M. Grasmair, M. Haltmeier and O. Scherzer, Inverse Problems 24, 2008, 055020
10. J. Fox, Nonparametric Regression in R: An Appendix to An R Companion to Applied Regression, 2010, 2nd Edition
11. T. Hastie, The Elements of Statistical Learning – Data Mining, Inference, and Prediction, 2016, Springer Series in Statistics, p. 194-207
12. J. Adler, O. Öktem, Solving ill-posed inverse problems using iterative deep neural networks, arXiv:1704.04058v2, 2017.