Multi-Objective Knowledge-Based Strategy for Process Parameter Optimization in Micro-Fluidic Chip Production

Alexandru-Ciprian Zăvoianu*, Edwin Lughofer*, Robert Pollak*,
Pauline Meyer-Heye†, Christian Eitzinger†, and Thomas Radauer‡

*Department of Knowledge-based Mathematical Systems, Johannes Kepler University Linz, Austria
†Profactor GmbH, Steyr-Gleink, Austria
‡Stratec Consumables, Anif, Austria

***Abstract*—We present an effective optimization strategy for industrial batch processes that is centered around two computational intelligence methods: linear and non-linear predictive mappings (surrogate models) for quality control (QC) indicators and state-of-the-art multi-objective evolutionary algorithms (MOEAs). The proposed construction methodology of the linear and neural network-based mappings integrates implicit expert-based knowledge with a new data-driven sample selection strategy that hybridizes several design of experiments paradigms. Using a case study concerning the production of micro-fluidic chips and 26 QC indicators, we demonstrate how incorporating modeling decisions like cross-validation stability analyses and objective clustering into our optimization strategy enables the discovery of well-performing surrogate models that can guide MOEAs towards high-quality Pareto non-dominated solutions.**

***Index Terms*—process parameters, design of experiments, surrogate modeling, evolutionary multi-objective optimization.**

I. Introduction and Motivation

Optimized production processes and predictive maintenance [1] are two of the most important topics in modern industrial facilities and thus a key issue in several call-objectives within Horizon 2020 work programmes, especially those under the umbrella of the European FoF (Factories of the Future) Research Association (http://www.effra.eu/). The basic strategy is to take action when items or parts of the process show (in advance) certain behaviors that may result (at a later stage) in a machine failure, degraded performance or a downtrend in product quality [2]. Therefore, a core component in production optimization and predictive maintenance systems is the usage of techniques from the fields of forecasting [3] and prognostics [4] to early recognize non-optimal (machining) parameter settings or dynamically arising problems during production cycles [5] and, ideally, trigger an appropriate mechanism to automatically balance out such unpleasant occurrences [6].

Throughout this work, we use the term "*process parameters*" to refer to controllable (machine) settings that are kept fixed over a longer time frame during a production process. These parameters are usually adjusted manually (from time to time) by operators / domain experts according to changing external factors like: new product variants, raw material quality, supply chain disruptions, and others. Using the right settings is of the utmost importance in nearly every industrial production environment in order to maximize product quality and prevent severe quality deterioration that can lead up to production waste. This work is mainly motivated by the fact that, often, the operators themselves do not know the truly ideal settings of the process parameters – i.e., those settings that output products with the best quality control (QC) indicators – as humans tend to set and adjust the parameter values using experience and rules of thumb. Moreover, there is also a strong bias factor since settings that are conventionally used tend to be accepted / adopted by the operators more easily.

Our main goal is to propose and demonstrate the benefits of a generic strategy that combines expert insight and a *novel hybrid design of experiments strategy* (DoE) with surrogate-based search space exploration techniques in order to efficiently and economically discover those parameter settings that result in QC indicator values that are as close as possible to the ideal ones. In [7], a model-based strategy has been suggested for driving a dynamic process to a higher quality but the approach relied on consecutive measurements within a dynamic batch process while our presently proposed strategy aims to provide suggestions on a process and machining parameter setting level fully in advance, before starting the main production cycle. While the approach in [7] does not make use of them, classic DoE strategies have long been used in various process optimization applications [8] [9]. Nevertheless, advanced DoE strategies (like our new hybrid proposal) that are designed to guide the data acquisition process into an ideal direction while considering multiple criteria remain quite scarce in the filed of process optimization.

II. Problem Formalization

Let us assume that a given production process can be controlled using a total number of p process parameters $\{x_1, \cdots, x_p\}$, $x_i \in [l_i, u_i], \cdots, x_p \in [l_p, u_p]$ and that one wishes to assess this process via q QC indicators. Since the process parameters are expected to directly impact each member from the set of QC indicators, we can define the indicators

as $\{f_1(\vec{x}), \cdots, f_q(\vec{x})\}$, where $\vec{x} = \{x_1, \cdots x_p\} \in [l_1, u_1] \times \cdots \times [l_p, u_p]$ denotes the parameter combination (which is in essence a multi-dimensional point).

Theoretically, one aims to find the optimal parameter combination (not., $\vec{x^*} = \{x_1^*, \cdots x_p^*\}$) that simultaneously yields the ideal value for each QC indicator. Assuming a minimization formulation for each of these ideal QC values, $\vec{x^*}$ has the property that it is a solution to:

$$f_i(\vec{x}) = min_i, \quad 1 \leq i \leq q \qquad (1)$$
$$\text{subject to } x_j \in [l_j, u_j], \quad 1 \leq j \leq p$$

where the optimal min_i values and variation intervals of process parameters (i.e., $[l_j, u_j]$) are defined by domain experts.

Nevertheless, in many practical cases there are usually intrinsic conflicts between the various QC indicators meaning that Equation (1) defines a nontrivial multi-objective optimization problem (MOOP) for which there is no single solution (i.e., $\nexists \vec{x^*}$) as it is impossible to synchronously achieve all the ideal indicator values. In the case of nontrivial MOOPs, the solution comes in the form of a Pareto optimal set (*PS*) with the property that $\forall \vec{x}, \vec{y} \in PS$ there is at least one objective $k, 1 \leq k \leq q$ such that $f_k(\vec{x}) < f_k(\vec{y})$ and there is at least one objective $l, 1 \leq l \leq q$ such that $f_l(\vec{y}) < f_l(\vec{x})$. In other words no point in *PS* is better than another point in *PS* with regard to all the considered objectives. The projection of *PS* in objective space is called the Pareto Front (*PF*) and it illustrates the trade-offs between the achievable optimal values of individual objectives. Since for most MOOPs, the *PS* is infinite and/or unknown, one generally settles for discovering a Pareto non-dominated set (*PN*) that provides a (very) good approximation of the *PS* using a fixed number of points.

In light of the above, the goal of the present work can be refined to introducing a strategy for discovering *PN*s that provide process operators with an accurate image of the trade-offs between the best achievable QC values.

III. Our Approach

In the past two decades, multi-objective evolutionary algorithms (MOEAs) have emerged as one of the most successful computational intelligence techniques for solving challenging MOOPs [10]. Their popularity is motivated by robust performance on problems stemming from various fields [11] and by the ability to produce accurate Pareto non-dominated sets after single runs. The main drawback of MOEAs lies in the prerequisite to evaluate a large number of individuals / solution candidates (i.e., parameter settings in our case) during the optimization run. This is especially problematic for QC indicators of batch processes as they are determined after the termination of a production cycle (that can last from a few hours to several days). This means that there is a significant time delay between setting the process parameters and observing their impact on product quality. This makes an on-line evaluation of individuals generated during optimization runs unrealistic. Furthermore, no analytical simulation models are available, thus the actual mathematical formulae that define $f_i(\vec{x}), 1 \leq i \leq q$ are unknown and difficult to model explicitly. Moreover, the reported values of the QC indicators can be noisy as they are susceptible to measurement and (to a lesser extent) human error.

In order to alleviate the aforementioned problems, one possible strategy [12] [13] is to construct regression models (i.e., predictive mappings) between the process parameters (inputs) and the QC indicator values (outputs) and to use these regression models as surrogate quality assessment mappings that steer the evaluation-intensive MOEAs towards promising regions of the search space. Nevertheless, the particularities of many production environments (micro-fluidic chips included) impose serious challenges regarding (i) the availability of training data for building reliable surrogate mappings with good generalization capability and (ii) the inherent complexity of the multi-objective optimization task. Therefor, our knowledge-based strategy for improving the performance of production processes is divided in two major parts:

1 In Section IV we propose several modeling decisions that can alleviate the insufficient data problem and help to deliver reliable surrogate models. These include a novel *hybrid design of experiments* approach which is able to operate in a *fully unsupervised manner* by combining optimality criteria based on the Fisher information matrix with space filling design aspects.

2 In Section V, we describe a new effective surrogate-based optimization strategy for simultaneously improving many QC parameters beyond the levels achieved using expert based settings. The focus is on ideas for meaningful objective reduction and robust MOEAs that incorporate proven multi-objective optimization principles.

IV. Construction of Predictive Mappings

Taking into account the significant delay between changes in process parameter settings and the (QC-measured) effect onto the production process, typically only a very limited number of (training) samples can be obtained within a reasonable time frame. In our case study scenario, the delay can last from several hours up to one day. Therefore, it is of utmost importance to carefully select process parameter combinations that are (i) known to have an essential effect on the quality of the chips and (ii) expected to induce a high generalization capacity of the predictive mappings, especially by decreasing their uncertainty / parameter instability. Thus, we propose a knowledge-based construction strategy for QC predictive mappings that aims to combine:

1) operator / domain expert knowledge;
2) purely data-driven insights based on a flexible design of experiments (DoE) methodology;
3) linear and non-linear modeling techniques;

A. Expert Knowledge Initialization and Hybrid Design of Experiments (DoE)

Expert knowledge is mainly integrated in the mappings construction process during the initial data collection / generation stage. First, based on expert input, a so-called Cause-Effect (CE) diagram can be established in order to elicit the process

parameters that are most likely to influence QC indicators. Such a diagram, concerning our case study from Section VI-A, is presented in Figure 1 and the relevant process parameters (X1 to X11) are highlighted in blue and explicitly mentioned in the right lower corner.

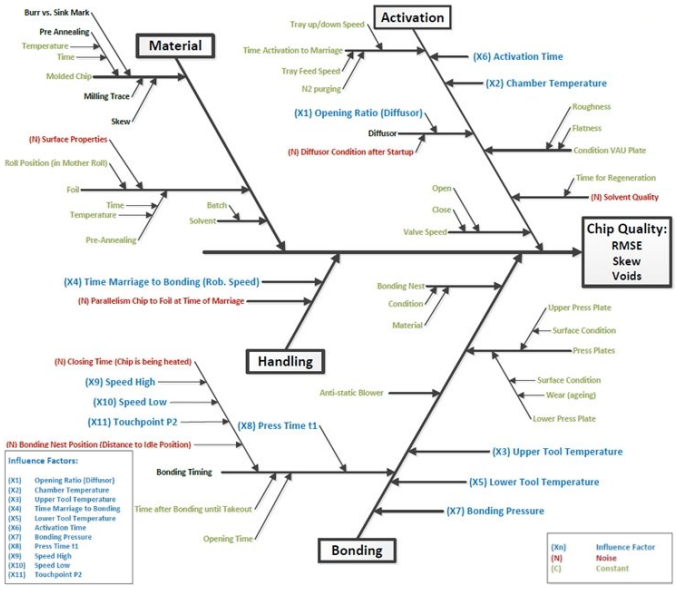

Fig. 1. A cause-effect diagram as constructed for the micro-fluidic chip production process at the bondingliner machine.

After the identification phase, the first few process parameter settings (i.e., data samples) to be tested are the ones recommended by the production process operators. Apart from this, expert (domain) knowledge is also used to restrict the domain of each parameter and to filter invalid parameter combinations. Computer-aided design selection strategies such as the Taguchi L12 method or Full Factorial [14] may support the experts when choosing appropriate combinations based on these limitations. This initial expert-based step is a prerequisite for a reliable design of experiments procedure.

The second step of our proposed data collection stage is based on a domain-independent DOE-based strategy and aims to obtain samples that are well distributed in parameter space in order to reduce the uncertainty of predictive mappings. Because of the severe restriction of having only a couple of expert-based data samples, we applied various optimal design criteria, such as A-optimality, D-optimality and E-optimality [15], under the (plausible [16]) assumption of having linear dependencies between process parameters and QC values (in the expert-based samples). We combined the optimal design criteria with pure space filling techniques based on Latin hypercube (LH) sampling and min-max optimization [17].

In particular, $M+2^p$ samples are to be drawn from the parameter space, with p the dimensionality of the space and $M >> 2^p$. M samples are generated via LH sampling and the remaining 2^p are the corner-points of the parameter space as we aim to reduce the likelihood of extrapolation as much as possible. In each odd iteration of sample selection, each of the remaining $M+2^p-|S|$ generated samples – with S and $|S|$ the set and number of samples selected so far – is evaluated based on how much it improves one of the following criteria:

- *A-optimality* (variant 1): ability to minimize the trace of the inverse of the Fisher information matrix [18].
- *D-optimality* (variant 2): ability to maximize the determinant of the Fisher information matrix.
- *E-optimality* (variant 3): ability to maximize the minimum eigenvalue of the Fisher information matrix.

As we apply linear model-based DoE, the Fisher information matrix is equivalent to the Hessian matrix $X_{ext}^T X_{ext}$, with X_{ext} containing the initial expert-based samples plus all the samples selected so far over the DoE iterations and the new sample to be checked for improvement of the optimality criteria. In each even sample selection iteration, we compute the minimal distances between each of the the remaining $M+2^p-|S|$ generated samples and the already selected samples (including the initial expert-based ones). The sample with the maximal minimal distance is selected as it embeds the highest novelty content.

In light of the mixed selection strategy, we consider that our design of experiments technique (shown in Algorithm 1) displays a strong *hybrid* flavor. It is also noteworthy that Algorithm 1 describes a *completely unsupervised* strategy that only relies on the input data matrix and does not require associated output measurements (i.e., QC values). This can have a great advantage whenever sample outputs are costly or time-intensive to obtain.

B. Predictive Mapping Models

Like in most data-driven modeling tasks, a basic preprocessing step should be performed in order to at least detect and remove those process parameters that are expected to have a minor influence on the QC indicators. This helps to reduce the input dimensionality and thus to increase the generalization capabilities of the models. In our case, process parameters can be seen as factors, which, based on expert knowledge input, can be divided into low, medium and high importance groups. By applying a multi-way ANOVA analysis [19] on the data collected through the proposed hybrid DoE approach, one can detect significant influences (on the QC values) of various factors and groups of factors. Factors that don't appear to have any influence can be deleted.

Given the limited number (i.e., low tens) of data samples one is likely to have gathered after the two-stage data collection phase and their high dimensionality, linear regression models between process parameter settings (inputs) and QC values (targets) are likely to deliver the best (generalization) performance [16]. Thus, training linear models is an intuitive first step. Non-linear regression modeling paradigms like artificial neural networks (ANNs), support vector machines or genetic programming should also be tested – particularly for those QC indicators that are problematic for the linear models.

In order to avoid overfitting (especially for the non-linear models), we always apply an n-fold cross-validation strategy for assessing model performance (i.e., R^2) during training.

Algorithm 1 Hybrid DoE for Sample Selection

function HYBRIDDOE(X, M, N, $optCrit$)
  $S \leftarrow \emptyset$, $G \leftarrow$ GENERATELHSAMPLES($size = M$)
  $G \leftarrow G \cup$ GENERATECORNERSAMPLES($size = 2^p$)
  for $j = 1$ to N **do**
    for $i = 1$ to $M + 2^p$ **such that** $\vec{x}_i \leftarrow G_i \wedge \vec{x}_i \notin S$ **do**
      $A, D, E, Dst \leftarrow \emptyset$
      $X_{ext} \leftarrow [X; S; \vec{x}_i]$
      if ODD(j) **then** ▷ iteration no. is odd
        $A_i \leftarrow \langle key =$TRACE$([X_{ext}^T X_{ext}]),\ val = \vec{x}_i \rangle$
        $D_i \leftarrow \langle key =DET([X_{ext}^T X_{ext}]),\ val = \vec{x}_i \rangle$
        $E_i \leftarrow \langle key =$MINEIG$([X_{ext}^T X_{ext}]),\ val = \vec{x}_i \rangle$
      else ▷ iteration no. is even
        $L_{\vec{x}_i} \leftarrow$ EUCLIDEANDISTANCES($\vec{x}_i$, $[X; S]$)
        $Dst_i \leftarrow \langle key =$SELECTMIN$(L_{\vec{x}_i}),\ val = \vec{x}_i \rangle$
      end if
    end for
    if ODD(j) **then** ▷ iteration no. is odd
      if $optCrit$="A-optimality" **then**
        $\vec{x}_{sel} \leftarrow$ GETVALFORMINKEY($set = A$)
      end if
      if $optCrit$="D-optimality" **then**
        $\vec{x}_{sel} \leftarrow$ GETVALFORMAXKEY($set = D$)
      end if
      if $optCrit$="E-optimality" **then**
        $\vec{x}_{sel} \leftarrow$ GETVALFORMAXKEY($set = E$)
      end if
    else ▷ iteration no. is even
      $\vec{x}_{sel} \leftarrow$ GETVALFORMAXKEY($set = Dst$)
    end if
    $S_j \leftarrow \vec{x}_{sel}$
  end for
  return S
end function

Furthermore, since the low sample count limits the number of cross-validation folds (i.e., $3 \leq n \leq 5$), we also perform a stability assessment of the most promising non-linear regression models. The idea is to compare if, for a given QC indicator, the non-linear model is able to generally deliver a superior modeling performance (vs. its linear counterpart) over several different cross-validation partitions. If the non-linear model significantly outperforms (e.g., when using statistical hypothesis tests on the residuals), one should use it as the predictive mapping for the QC indicator in question. Otherwise, the simpler linear regression model should be preferred.

V. Process Optimization

When considering processes (i.e., real-life MOOPs) that have more than 5-6 QC indicators to be optimized (i.e., objectives), one is said to deal with a many-objective optimization problem. Specialized algorithms that deal with such types of problems have been recently proposed (e.g., NSGA-III adaptations [20]) but, given the serious inherent difficulty of trying to solve (surrogate-based) many-objective optimization problems, a natural first step is to reduce the number of objectives to be simultaneously optimized.

A. Reducing the Number of Objectives

There are numerous ways of reducing the number of objectives in a many-objective optimization problem but most of these basic objective reduction strategies can be broadly classified as *explicit* or *implicit*.

Explicit objective reduction strategies are based on discussing the relative importance of each objective with the decision maker (DM) – i.e., process operator in our case. Apart from the obvious approach of simply removing the very low priority objectives from the problem formulation, one can also opt for the replacement of these objectives with a newly defined *synthetic objective* that aggregates them. The synthetic objective can be obtained using any of classical objective reduction techniques [21] that require no articulation of preference on behalf of the DM (e.g., Tschebyscheff min-max and global-criterion) or an a priori articulation of preference (e.g., weighted sum, lexicographic ordering, goal programming).

Implicit objective reduction is data-driven and can be achieved by clustering objectives that are cross-correlated. Depending on the strength of intra-cluster cross-correlation, one could choose to reduce the entire cluster of objectives to:

- one of its members that shall act as a "cluster representative" (very strong intra-cluster cross-correlation);
- a new synthetic objective that aggregates all the members of the cluster (mild intra-cluster cross-correlation).

Depending on the complexity of the objective reduction tasks, implicit and explicit strategies can be combined in order to achieve a suitable and useful MOOP formulation.

B. Multi-Objective Optimization Algorithms

The Non-dominated Sorting Genetic Algorithm II (NSGA-II) [22] has become in the 15 years since its proposal one of the default go-to (metaheuristic) multi-objective solvers. Its main feature is a highly elitist evolutionary model underpinned by a two-tier selection for survival operator that features a primary non-dominated sorting criterion and a secondary (tie-breaking) objective-space crowding measure. The non-dominated sorting operator, although simple (when compared to newer MOEAs), is highly robust and enables NSGA-II to discover high-quality *PN*s in many application domains [23]. NSGA-II has also popularized two genetic operators in the multi-objective optimization field: simulated binary crossover (SBX) and polynomial mutation (PM) [24].

DECMO2 [25] is a coevolutionary MOEA that was designed to deliver fast average convergence and well-spaced *PN*s on a wide class of problems. The key feature of DECMO2 is that it tries to combine (and dynamically pivot between) three multi-objective search space exploration paradigms:

- P - one of the two equally sized sub-populations evolved in DECMO2 uses a SPEA2 [26] evolutionary model centered around the *environmental selection* operator which implements a two-tier selection for survival strategy that

is very similar to the one of NSGA-II. Population P is also evolved using the SBX and PM operators.

- Sub-population Q adopts the GDE3 [27] search behavior that aims to benefit from the very good performance of differential evolution operators (e.g., *DE/rand/1/bin*) [28] on continuous optimization problems.
- The third multi-objective optimization paradigm is incorporated in DECMO2 via an archive A of well-spaced elite solutions that are maintained according to a (weighted Tschebyscheff) decomposition-based strategy similar to the one popularized by MOEA/D-DE [29]. Although A largely acts as a passive sub-population, from time to time (especially if the other search paradigms underperform), a few individuals are evolved directly from A using differential evolution.

DECMO2 actively rewards the currently best performing strategy by allowing the sub-population that implements it to generate a total of $m = \frac{2}{9}|P|$ more individuals than usual. A schematic overview of the search strategy proposed by the coevolutionary solver is presented in Figure 2.

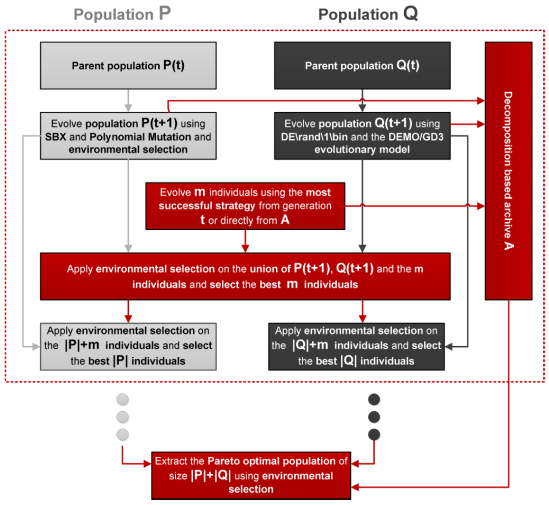

Fig. 2. The DECMO2 evolutionary model

VI. Experimental Setup

A. Application Scenario

We demonstrate the proposed multi-objective process parameter optimization strategy on a batch production process of micro-fluidic chips used for sample preparation in DNA (deoxyribonucleic acid) sequencing. Currently, $q = 26$ QC indicators are supervised at the end of a production cycle (batch) via a diagnostic procedure based on closed loop surface inspection using machine learning classifiers [30]. The a posteriori nature of this inspection process means that faulty chips may be disregarded after having been produced. While this does indeed improve customer satisfaction, it does not reduce the waste and the extra costs associated with the production of low-quality chips because of unlucky parameter settings. Therefore, there is a strong case for optimizing the parameters that control the micro-fluidic chip production process.

The 26 QC indicators that characterize the quality of a given process parameter setting belong to there main groups (*RMSE*, *skew* and *void defects*) and were obtained as follows:

- for the two indicators (i.e., f_1 and f_2) in the *RMSE* group only a single value was computed for the whole production plate by averaging the individual QC values of every micro-chip in the plate;
- for the indicators regarding *skew*, the compliance of each of the 6 micro-chips in a production box was measured at two different time intervals, resulting in a total of 12 QC indicators (f_3 to f_{14});
- the 12 QC indicators (f_{15} to f_{26}) regarding *void defects* were obtained in the same manner as in case of *skew*.

B. Data Characteristics

Based on discussions with the experts, the cause-effect diagram shown in Figure 1 has been established. It lists comprising $p = 11$ essential process parameters that are expected to influence the quality criteria of micro-fluidic chips. Furthermore, experts also defined the relevant domain ranges of each parameter as this information is required by both our hybrid DoE approach from Algorithm 1 and the MOEAs.

During the initial step of the data collection procedure, the experts proposed 12 parameter combinations that are expected to have a positive impact on the quality of the chips. The expert samples were further used as input for the hybrid DoE strategy that aims to improve the generalization capability of predictive mappings by reducing model uncertainty.

In order to detect the best performing setting for our hybrid DoE strategy (i.e., optimality Criterion + value of N) given the 12 expert parameter combinations we performed numerical experiments aimed to measure the evolution of model uncertainty. Figure 3 shows the trend lines of the condition of the parameter covariance matrix for different DoE criteria (y-axis) when selecting more and more samples (x-axis) up to $N = 100$. It is defined by $cond(X) = \frac{max(eig(X^T X))}{min(eig(X^T X))}$, a well-known and widely-applied measure for parameter and model uncertainty in case of regression/mapping problems [19]. A-optimality

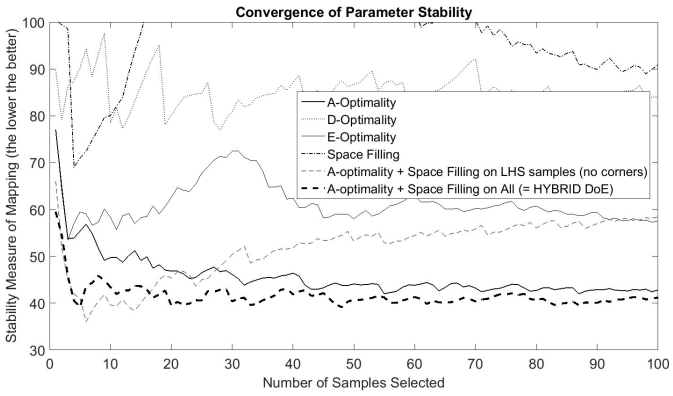

Fig. 3. Trace of parameter stability for various DoE-based sample selection criteria when incrementally (step-wise) adding up to 100 samples.

can significantly outperform the other variants, either when

being used stand-alone (which is state-of-the-art [15]) or in the hybrid combination with a space filling approach. The latter also can significantly outperform the classical SoA A-optimality criterion, as the hybrid combination reduces model uncertainty faster, especially during the first 10-12 samples. The grey dashed line shows the case when only Latin hypercube samples are considered for space filling. Even though this strategy delivers slightly better performance after generating the first 7,8 samples when comparing with hybrid DoE (that also uses corner points), the former is not able to ensure the stability of the expected model certainty further on.

Based on the above observations, 11 additional process parameter combinations have been (i) generated via the hybrid DoE strategy and (ii) tested at the production site in order to elicit QC values. Considering the 12 expert-based suggestions, the total input for our predictive mapping construction phase (from Section VII-A) is a data set with 23 samples.

C. Evaluation Strategy

For each of the 26 QC indicators, apart from linear modeling, we also tried to obtain high quality non-linear predictive mappings based on multi-layer perceptrons (MLPs) with a single hidden layer – i.e., shallow ANNs [31]. In an effort to avoid overfitting the MLPs, we combined a best-training-parameter grid search with a selection process based on 5-fold cross-validation performance. Thus, we varied the number of hidden units between 2 and 28 and we also varied the two control parameters of the back-propagation [32] algorithm:

- the learning rate (between 0.05 and 0.5, step size of 0.05)
- the momentum (between 0 and 0.9, step size of 0.1)

Finally, when considering all the combinations tested during the best-traning-parameter grid searches, the *best-performing MLP* model for each QC target function was selected from 2700 candidates by opting for a strategy aimed to balance predictive accuracy – estimated by imposing a 5-fold cross-validation R^2 performance higher than the average performance of the best 2% of all trained MLP models – and model simplicity (i.e.,a lower number of hidden units) [33].

In the case of the two MOEAs, we used their respective literature recommended settings to parameterize the genetic operators, a (total) population size of 200 and each optimization run was stopped after evaluating 100,000 individuals.

VII. Results

A. Predictive Mapping Construction

In Table I, we present the comparative performance between linear and best-performing MLP (non-linear) predictive mappings when considering a given 5-fold cross-validation partition. Although, the tabulated data seems to suggest that non-linear modeling brings obvious benefits since it improves results for 19 out of the 26 QC indicators, the scarcity of training data does warrant a more in depth (stability) analysis.

In Figure 4 we plot the average R^2 values obtained by the linear and best-performing MLP regression models for 3 expertly-selected QC indicators – f_1, f_{12} and f_{18} – when considering 25 different random partitions of the training data into 5 cross-validation folds. The plotted results offer a slightly clearer picture, indicating that:

- some QC indicators (e.g., f_1) are generally difficult to model as R^2 values depend more on the cross-validation partition than on the used modeling method;
- some QC indicators (e.g., f_{12} and f_{18}) appear far easier to model but using advanced non-linear methods does not seem to bring a consistent / stable advantage and can even deliver slightly worse results (e.g., f_{12}).

TABLE I
Comparative performance of linear and best-performing MLP predictive mappings. Superior results are highlighted.

QC Indicator	5-fold cross-validation R^2			
	Lin. reg.		MLP	
	μ	σ	μ	σ
f_1	0.581	0.317	**0.795**	0.092
f_2	0.517	0.114	**0.759**	0.177
f_3	0.863	0.080	**0.879**	0.061
f_4	0.882	0.060	**0.913**	0.080
f_5	**0.786**	0.118	0.742	0.336
f_6	**0.922**	0.065	0.874	0.093
f_7	**0.887**	0.061	0.829	0.100
f_8	**0.920**	0.038	0.896	0.089
f_9	0.645	0.062	**0.744**	0.184
f_{10}	0.926	0.043	**0.938**	0.031
f_{11}	**0.895**	0.066	0.841	0.097
f_{12}	**0.947**	0.046	0.923	0.111
f_{13}	**0.870**	0.079	0.827	0.094
f_{14}	0.888	0.059	**0.901**	0.091
f_{15}	0.580	0.284	**0.793**	0.124
f_{16}	0.727	0.247	**0.850**	0.056
f_{17}	0.732	0.115	**0.854**	0.086
f_{18}	0.847	0.152	**0.905**	0.057
f_{19}	0.656	0.227	**0.845**	0.121
f_{20}	0.781	0.108	**0.831**	0.058
f_{21}	0.371	0.633	**0.863**	0.100
f_{22}	0.783	0.157	**0.875**	0.080
f_{23}	0.620	0.368	**0.817**	0.070
f_{24}	0.801	0.210	**0.880**	0.054
f_{25}	0.618	0.415	**0.849**	0.137
f_{26}	0.814	0.145	**0.878**	0.120

B. Process Optimization

Based on the available 23 data samples, we obtained the cross-correlation matrix shown in Figure 5 for the 26 QC indicators that are of interest to the industrial partner.

The cross-correlation results clearly indicate that the original measurement-based classification of the QC indicators into three main groups is relevant as:

- QC indicators f_1 and f_2 are strongly intercorrelated and, from an optimization perspective, can be reunited in a *RMSE cluster*;
- QC indicators f_3 to f_{14} can form the *skew cluster*;
- QC indicators f_{15} to f_{26} can form the *void defects cluster*.

It also noteworthy that there is (i) a small postive correlation between the indicators in the RMSE and skew clusters, (ii) a small negative correlation between the QC indicators in the RMSE cluster and those in the void defects cluster, (iii) no apparent correlation between skew and void defects indicators.

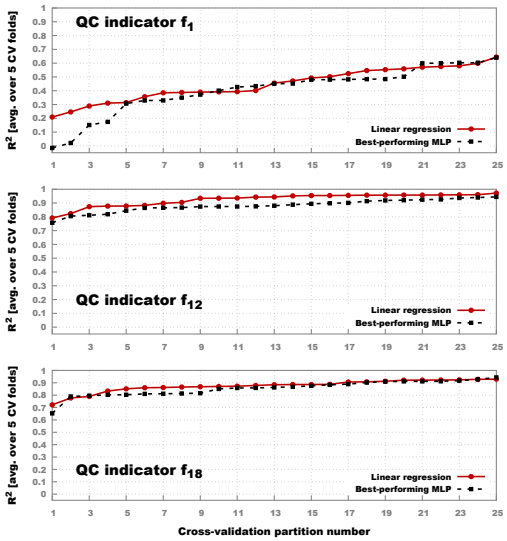

Fig. 4. Comparative performance of linear and best-performing MLP models for three QC indicators over 25 different cross-validation partitions.

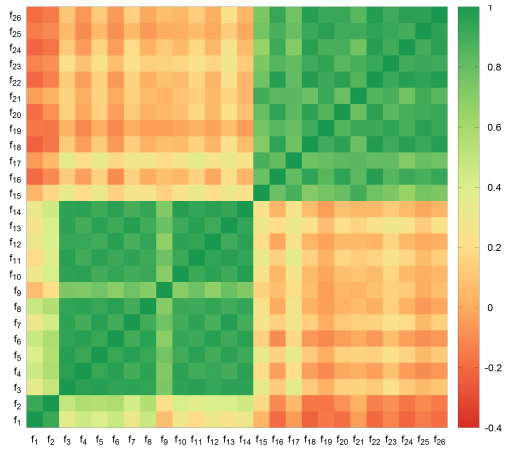

Fig. 5. Cross-correlation between the 26 QC indicators; note the block-type structure with three clusters.

Based on the strong inter-correlation inside the identified objective clusters we defined a surrogate-based MOOP that contains one representative from each cluster with the reasoning that, by simultaneously aiming to minimize f_1, f_12, and f_18 (for example), we are in fact searching for process parameter settings that deliver Pareto optimal solutions (i.e., QC values) related to all 26 RMSE, skew and void defects indicators. When defining this surrogate-based MOOP, we opted for linear predictive mappings and formulations where optimizing an objective meant minimizing its value towards 0.

The 3D Pareto front obtained by DECMO2 for the 3-objective surrogate-based MOOP are presented in Figure 6. Given the fact that objective f_1 has proven harder to model (more unstable and lower R^2 values) in our 5-fold cross-validation experiments, considering the precise numerical values of this indicator in further analyses is not recommended. Therefore, in Figure 7, which contains the final comparative

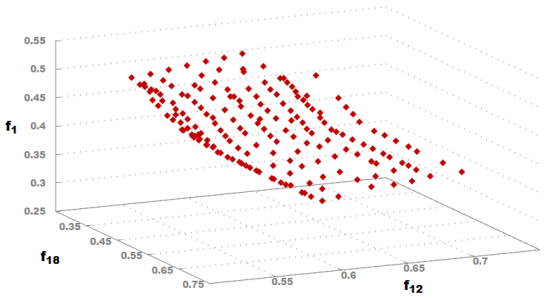

Fig. 6. 3D Pareto Front obtained using DECMO2.

optimization results obtained by NSGA-II and DECMO2, we report performance regarding f_1 by considering 4 broad quality groups. The two Pareto non-dominated sets indicate that:

- the integration of the decomposition-based space exploration paradigm enables DECMO2 to maintain a better spread across the entire PF;
- NSGA-II also performs robustly as it is able to deliver a large number of (albeit more poorly spread) solutions in a key section of the PF where the harder to model f_1 objective is also minimized.

The multi-objective optimization results have received very positive appreciation from domain experts (i.e., process operators) and the industrial partner and several parameter settings discovered using the presented methodology are currently being tested / used on the production site. For example, when comparing to the QC values obtained by one of the best parameter combinations from those obtained from experts and the hybrid DoE strategy, one very promissing DECMO2-discovered parameter setting is estimated to deliver a simultaneous improvement of 9.3% over f_1 and 2.1% over f_{12} while obtaining perfect QC values for f_{18}.

VIII. Conclusions and Future Work

In the present work we propose an efficient strategy for optimizing process parameters with respect to multiple quality control indicators. As the application domain consists of industrial batch production processes that are characterized by very time and cost-intensive evaluations of parameter settings, our strategy is centered on creating linear and non-linear predictive mappings that can be used as surrogate fitness estimators by state-of-the-art multi-objective evolutionary algorithms. Results on a micro-fluidic chip production case study support the proposed methodology and underline that one key part of obtaining useful predictive mappings (and, in turn, very good optimization results) is a two-stage data collection phase that combines expert knowledge with a novel hybrid design of experiments strategy.

We plan to further refine our optimization strategy by using surrogate-based MOEA runs to augment the current hybrid DoE strategy by suggesting a handful of parameter combinations that can simultaneously fill multiple blank spaces in the objective-wise histograms, thus further improving predictive mapping quality and optimization performance.

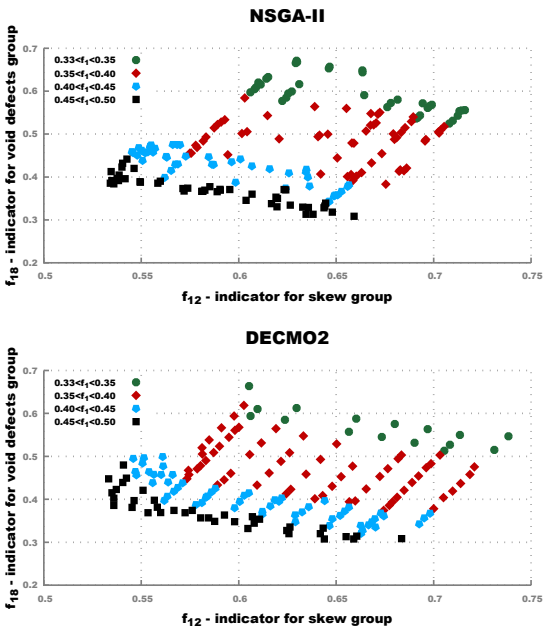

Fig. 7. Comparative f_{12} vs. f_{18} 2D PFs obtained using NSGA-II and DECMO2 (best result out of 5 runs for each algorithm).

ACKNOWLEDGMENT

The authors would like to acknowledge the Austrian Research Funding Association (FFG) within the scope of the 'IKT of The Future' program, project 'Generating process feedback from heterogeneous data sources in quality control (mvControl)' (contract no. 849962).

REFERENCES

[1] J. Levitt, *Complete Guide to Preventive and Predictive Maintenance*. New York: Industrial Press Inc., 2011.

[2] R. Mobley, *An Introduction to Predictive Maintenance — Second Edition*. Woburn, Massachussetts, U.S.A.: Elsevier Science, 2002.

[3] G. Box, G. Jenkins, and G. Reinsel, *Time Series Analysis, Forecasting and Control*. Engelwood Cliffs, New Jersey: Prentice Hall, 1994.

[4] S. Ekwaro-Osire, A. Gonalves, and F. Alemayehu, *Probabilistic Prognostics and Health Management of Energy Systems*. New York: Springer, 2017.

[5] E. Lughofer, R. Pollack, A.-C. Zăvoianu, P. Meyer-Heye, H. Zörrer, C. Eitzinger, J. Heim, and T. Radauer, "Self-adaptive time-series based forecast models for predicting quality criteria in microfluidics chip production," in *Proceedings of the IEEE Cybernetics Conference*, 2017.

[6] E. Permin, F. Bertelsmeier, M. Blum, J. Bützler, S. Haag, S. Kuz, D. Özdemir, S. Stemmler, U. Thombansen, R. Schmitt, C. Brecher, C. Schlick, D. Abel, R. Popraw, P. Loosen, W. Schulz, and G. Schuh, "Self-optimizing production systems," *Procedia CIRP*, vol. 41, pp. 417–422, 2016.

[7] S. Aumi, B. Corbett, and P. Mhaskary, "Model predictive quality control of batch processes," in *2012 American Control Conference*, Fairmont Queen Elizabeth, Montral, Canada, 2012, pp. 5646–5651.

[8] S. Gu, J. Ren, and G. Vancso, "Process optimization and empirical modeling for electrospun polyacrylonitrile (pan) nanofiber precursor of carbon nanofibers," *European polymer journal*, vol. 41, no. 11, pp. 2559–2568, 2005.

[9] K. Chockalingam, N. Jawahar, K. Ramanathan, and P. Banerjee, "Optimization of stereolithography process parameters for part strength using design of experiments," *The International Journal of Advanced Manufacturing Technology*, vol. 29, no. 1, pp. 79–88, 2006.

[10] C. Coello Coello, G. Lamont, and D. Van Veldhuisen, *Evolutionary Algorithms for Solving Multi-Objective Problems*, ser. Genetic and Evolutionary Computation Series. Springer, 2007.

[11] C. A. Coello Coello and G. B. Lamont, *Applications of multi-objective evolutionary algorithms*. World Scientific, 2004.

[12] Y.-S. Hong, H. Lee, and M.-J. Tahk, "Acceleration of the convergence speed of evolutionary algorithms using multi-layer neural networks," *Engineering Optimization*, vol. 35, no. 1, pp. 91–102, 2003.

[13] Y. S. Ong, P. B. Nair, A. J. Keane, and K. W. Wong, "Surrogate-assisted evolutionary optimization frameworks for high-fidelity engineering design problems," in *Knowledge Incorporation in Evolutionary Computation*. Springer, 2005, pp. 307–331.

[14] D. Montgomery, *Design and Analysis of Experiments*. New York Chichester Brisbane Toronto Singapore: John Wiley & Sons, 1991.

[15] G. Franceschini and S. Macchietto, "Model-based design of experiments for parameter precision: State of the art," *Chemical Engineering Science*, vol. 63, no. 19, pp. 4846–4872, 2008.

[16] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction - Second Edition*. New York Berlin Heidelberg: Springer, 2009.

[17] M. McKay, R. Beckman, and W. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics (JSTOR Abstract)*, vol. 21, no. 2, pp. 239–245, 1979.

[18] B. Frieden and R. Gatenby, *Exploratory Data Analysis Using Fisher Information*. New York: Springer Verlag, 2007.

[19] F. Harrel, *Regression Modeling Strategies*. New York, USA: Springer, 2001.

[20] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints." *IEEE Trans. Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, 2014.

[21] K. Miettinen, *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, 1999.

[22] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[23] Y. Yusoff, M. S. Ngadiman, and A. M. Zain, "Overview of nsga-ii for optimizing machining process parameters," *Procedia Engineering*, vol. 15, pp. 3978–3983, 2011.

[24] K. Deb, *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, 2001.

[25] A.-C. Zăvoianu, E. Lughofer, G. Bramerdorfer, W. Amrhein, and E. P. Klement, "DECMO2: a robust hybrid and adaptive multi-objective evolutionary algorithm," *Soft Computing*, vol. 19, no. 12, pp. 3551–3569, 2014.

[26] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization," in *Evolutionary Methods for Design, Optimisation and Control with Application to Industrial Problems (EUROGEN 2001)*. International Center for Numerical Methods in Engineering (CIMNE), 2002, pp. 95–100.

[27] S. Kukkonen and J. Lampinen, "GDE3: The third evolution step of generalized differential evolution," in *IEEE Congress on Evolutionary Computation (CEC 2005)*. IEEE Press, 2005, pp. 443–450.

[28] R. Storn and K. V. Price, "Differential evolution - a simple and effcient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, December 1997.

[29] Q. Zhang, W. Liu, and H. Li, "The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances," School of CS & EE, University of Essex, Tech. Rep., February 2009.

[30] E. Weigl, W. Heidl, E. Lughofer, T. Radauer, and C. Eitzinger, "On improving performance of surface inspection systems by online active learning and flexible classifier updates," *Machine Vision and Applications*, vol. 27, no. 1, pp. 103–127, 2016.

[31] S. Haykin, *Neural Networks: A Comprehensive Foundation (2nd Edition)*. Upper Saddle River, New Jersey: Prentice Hall Inc., 1999.

[32] P. J. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Harvard University, 1974.

[33] A.-C. Zăvoianu, E. Lughofer, G. Bramerdorfer, W. Amrhein, and E. P. Klement, "An effective ensemble-based method for creating on-the-fly surrogate fitness functions for multi-objective evolutionary algorithms," in *Proceedings of SYNASC 2013*. IEEE Computer Society Conference Publishing Services (CPS), 2014, pp. 237–244.