Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 180 (2021) 545–560

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Industry 4.0 and Smart Manufacturing

Large scale predictability analysis of process variables from injection molding machines

Shailesh Tripathi[a,*], Christian Mittermayr[b], David Muhr[a], Herbert Jodlbauer[a]

[a] *Production and Operations Management, University of Applied Sciences Upper Austria, Austria*
[b] *Greiner Packaging International GmbH, Greinerstrasse 70 4550 Kremsmuenster*

Abstract

This paper analyzes the process variable data from injection molding processes to identify the key process variables, which can be predicted by other process variables, which highlights the interdependence among different process variables in various production scenarios. The available data from injection molding machines provide information for the run-time, setup parameters of machines, and measurements of different process variables through sensors. For predictive modeling, we employed different linear regression models with the recursive backward feature selection and SVM regression models using a radial kernel to predict nonlinear process variables. We also applied the linear and SVM regression models for outlier data in the process variables assuming that the upper bound outliers represent the perturbed state of process variables during production. These perturbations are affected by material type, machine type (age and performance), regime changes, or other external effects and subsequently affect the predictability of process variables and production output. Such cases are different compared to the normal or controlled range of process variables. Our analysis shows that the predictability varies for different material types due to the interdependence of the associated process variables. We further highlight that various process variables exhibit nonlinear relationships and cannot be predicted using linear models. We additionally look for the interdependence of process variables used previously by three studies as input features to predict product quality.

Keywords: Injection molding, Process Parameter Setting, Manufacturing, Model Selection, Regression Analysis, Machine Learning, Industry 4.0;

1. Introduction

Injection molding is an important process to produce various types of plastic products. During this process, a melted polymer is injected into a mold cavity, packed under pressure, and cooled until it has sufficiently solidified. This process is performed using an injection molding machine with an appropriate mold. During the process, the material,

* Corresponding author.
E-mail address: shailesh.tripathi@fh-steyr.at

1877-0509
10.1016/j.procs.2021.01.274

mold design, and processing parameters of the injection molding machines interact with each other and determine the quality of the plastic product [4]. As there is a large variety of processing parameters, the complexity of the process requires careful attention to maintain the quality characteristics. If the necessary quality characteristics cannot be achieved, the parts are discarded as scrap. There are different types of quality problems, such as shrinkage, warpage, color and burn marks, surface texture quality, shape distortion, and other aesthetic defects [12]. In real-world industrial production, the product quality varies between production lots for a variety of reasons. These include differing process parameter settings, which can be machine-, product-, and material-related and need to be controlled and monitored throughout the production process.

Various process parameters that depend on the quality of the product are difficult to optimize or are neglected due to the complexity of the production process or the unavailability of data. The complexity of an injection molding process can be described in terms of the interdependence between several process variables. Thus, the quality of the production output is determined from various sets of interdependent process parameters in different cases [8, 3, 18]. These interrelated process parameters are used to assess the product quality and monitor production processes to control the quality of different products using various machines. The lack of efficiently optimized process parameters for the various production scenarios can be due to neglecting their interdependence. The dependence indicates that the process parameters need to be adjusted in accordance with other interdependent process variables. The relationships between process variables can be estimated using various statistical and machine learning models that utilize the process variable data from different types of production scenarios. However, the interdependence of process variables in the production of different product or machine types varies for the different production scenarios, i.e., the dependencies change between process variables if their settings are different for various product types. To analyze the interdependence between process variables in detail, we must consider different production scenarios with a variety of product types (shapes and material type), parameter settings, machine type, and several other varying factors that are sensitive to the process control and production output. The underlying concept for efficient process control and product quality is to not only optimize the key known parameters but to find case-specific dependencies between various process parameters and utilize that information to optimize the process control and product quality.

We analyze data from real-world injection molding processes, which consists of 90 process variables to produce various products of 276 material types in under 15 different groups. In our analysis, we utilize regression models to find the set of interdependent process variables for different production scenarios. We applied the recursive feature elimination (RFE) approach for each regression model with 20-fold cross-validation to select the best subset of input process variables. We further estimated the average R^2, and the average root mean square error (RMSE) with the best subset of dependent process variables. The estimated R^2 and average RMSE are utilized to rank the response process variable predictability.

This approach is divided into three types of models. The first approach is a simple linear model to estimate the predictability of each process variable when linearly predictable. The second approach considers outliers of the response process variable and the corresponding input values of other process variables as input data for the linear model. The underlying assumption when investigating process variables in outliers is that machine parameters are often kept relatively stable, indicating they do not contain much information (variation). By considering only outliers (upper bound data that show higher variations), we assume that these outliers represent interesting events [1]. Therefore, the predictability and interdependence of process variables differ compared to cases when process variables are under controlled conditions. These outliers exhibit variations that can be product- and material-specific, which can be ignored in various production processes where the interdependent process parameters are required to be optimized per the product-shape and material-type to maintain product quality. The third approach estimates a nonlinear dependency among the process variables by utilizing the SVM regression model, which is applied for feature selection using RFE. We give preference to features selected using SVM regression only when it shows a significant improvement in predicting the training and testing errors than the corresponding linear model which shows a lower R^2 estimate of the predictor process variable.

The primary objective of this paper is to find the predictability of process variables (dependent) as predictor variables to determine if they exhibit a stable performance in production processes for various products of different material types and product groups. The secondary objective is to identify the interdependence of process variables in different production scenarios, i.e., to find the process variables (predictors) that are consistently key features when estimating the response process variables. The tertiary objective is to find the interdependent process variables in the previous study [23], which are identified as key features to estimate the scrap rate for similar production cases. Thus, this paper contributes to the application of machine learning methods to identify important interdependent process variables that play important roles in production quality.

The structure of this paper is as follows. In Section 2, we provide a brief overview of related works. In Section 3, we describe the details about the methods for data preprocessing and correlation analyses. In Section 4, we present details about the regression modeling, variable selection, and outlier detection. In Section 5, we evaluate the proposed methods applied to production data and compare the results. Additionally, we compare the predictability of process variables used as input variables in other studies [17, 14, 23] to classify product quality and predict scrap rates. Finally, in Section 6, we provide concluding remarks about the results of our analysis.

2. Literature Review

Several studies have addressed the quality optimization of injection molding processes. Many studies related to quality optimization are based on Taguchi experimentation with fewer key process variables responsible for product quality [21, 20, 24, 4, 2, 16, 5, 15, 11]. Several computational approaches have been considered to optimize product quality, which use gradient-based approaches, evolutionary algorithms, and mixed approaches [28, 29, 15, 27, 5]. Reviews of the frameworks to optimize injection molding methods are described in [12, 6, 19, 9].

[17] used SVM-based approaches by tuning different hyperparameters for error classification. The required data for the analysis was generated with an experimental set up on a *Demag* injection molding machine with *Hostacom DM2 T06* polymer and the *DN502* mold to predict product quality. The classification model used *cycle time*, *dosage time*, *injection time*, *cushion*, *peak melt temperature*, and *ram velocity* as input variables. The response variables described as product defects were classified into 6 classes as Streaks, Strains, Burn marks, Edges, Unfilled parts, and Warped parts. [14] performed feature learning and process monitoring by utilizing a deep learning approach of a convolution-deconvolution autoencoder to predict product quality. The data were experimentally generated on a *JSW J110ADC-180H* electric injection molding machine by tuning different process conditions. The input process variables, *screw displacement*, *injection pressure*, and *cavity pressure* were supplied in the form of a 4D input Tensor as $\chi(B \times V \times T \times C)$, where V is the number of variables, B is the batch size of a product quality class, C is the number of feature channels, and T is the set of time instances. [23] utilized various statistical features for 65 process variables using beta-regression and SVM regression models to predict the scrap rate in a production process. The data were collected from the real-time production of products with different shapes, sizes, and material types as produced by 33 different machines in the company.

3. Methodology

This section provides a brief overview of the methods applied for preprocessing, outlier detection, and regression modeling. First, we describe the available data for our analysis and the major preprocessing steps. After splitting the data into different production lots (segments), we extract relevant process variables for the subsequent regression models. The objective is to train different predictive models for the process variables to obtain information about the various setup parameters and process variables that have the highest impacts on each other. That is, we want to optimize the predictability of all process variables for different product types as processed on different machines. This predictability analysis allows tuning various combinations of interdependent process parameters on different machines with various product and material types. Additionally, we learn the interdependence of the process variables, which are used primarily to predict product quality.

3.1. Data Collection

We collect raw data from 31 injection molding machines. These datasets are recorded in different files during a given production process. Additionally, we use the data from the enterprise resource planning (ERP) system. The relevant data for each production process are recorded in a production file, process data files, and the export file from the ERP system, which have the following information:

Production file: provides timestamps, cycle counter, tool-name, raw-material information, cavities, and set cycle time.
Process data: provides timestamps, cycle counter, set cycle time, and 90 different process variables such as actual cycle time, temperature, pressure, volume, position, and rotational speed.
ERP data: provides the order number, material number, number of produced parts, and number of scrap parts.

The production and process data files contain data from a period when multiple product types are produced. The process data do not contain product-type information. We use the production and ERP data to split the process data into different segments based on the product type and order number. The segments based on product type and order number are described as process segments. Each process segment contains information regarding approximately 90 different process variables for the production process of a product type between the start and the end times of the production order. The segment data file is a multivariate file where the columns are the process variables and the rows are their respective measures at different times. Relevant meta-information about each production order is collected in a separate master data file, which includes the start and end times of production, order number, machine number, raw material number, and total production output in the units. We additionally receive product category information for different materials-types. The frequency of different product groups is shown in Table 1.

Product Category	PPO	PPS	MIX	TPU	TPE	PET	PBT	POM	PS	PP	PC	PE	ABS	PA
frequency	1	2	3	3	7	8	19	27	41	72	86	86	132	183

Table 1: frequency of material types in different product groups.

3.2. Feature Extraction

We assume that there are 31 production machines described as $m = m_1, m_2, \ldots m_n$, with X being a set of r process variables as $X(m_i) = \{X_1, X_2, \ldots X_r\}$. Each $X(m_i, p_j)$ records values during the production-process of product p_j at machine m_i between the time points t_1 and t_n, which is defined as $X(m_i, p_j, r_k) = (x_{t_1}, x_{t_2}, \ldots, x_{t_n})^T$. We extract $|r| = 90$ process variables for each machine m_i during the production of product type p_j.

3.3. Data Filtering

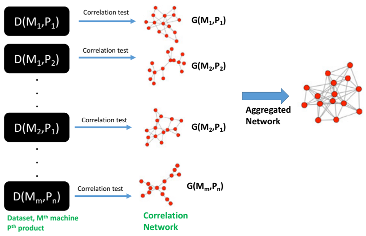
Our analysis filters out process segments that run longer than 120 hours and shorter than 20 minutes. This information is extracted using the master data file. Additionally, we filter out the process variables that show no variability and have a large proportion of missing values for different machines and material types (product). Thus, for 31 machines and 276 material type, we have 686 unique data sets with process variables $2 < p \leq 90$. The list of process variables is given in Table 9 of the Appendix.

3.4. Correlation test and network construction

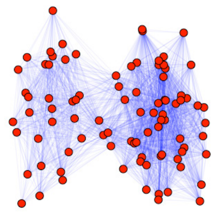
The first step of the analysis is to test the following null hypothesis by performing correlation tests between each pair of process variables of the dataset for machine m_a and product p_k. Let $X_i = X(m_a, p_k, r_i)$ and $X_j = X(m_a, p_k, r_j)$:
$H_0 : \rho(X_i, X_j) = 0$
$H_1 : \rho(X_i, X_j) \neq 0$
Therefore, each dataset tests the $\frac{p(p-1)}{2}$ hypotheses simultaneously. To control the False discovery rate (FDR), we apply multiple testing correction (MTC)[10] and define a graph based on the adjacency matrix using the significance of the correlation between each pair after applying MTC as follows:

$$G_{m_a,p_k}(i, j) = \begin{cases} 1 \text{ if } p(i, j) < \alpha \\ 0 \text{ otherwise} \end{cases}$$

We construct the graph for each data set and construct an aggregated graph, $G = \{G_1 \cup G_2 \ldots G_{|m|,|p|}\}$. A schematic representation of the final graph construction is shown in Figure 1a.



(a) The schematic representation of aggregated network of process variables.



(b) The constructed correlation network of process variables. The network contains $|V| = 90$ vertices and $|E| = 2371$ edges.

Fig. 1: Correlation network construction

4. Regression Model for process variables and feature Selection

To develop models that predict the process variables, we first use the constructed network $G = (V, E)$. Where, V are the set of process variables and E is the edge set representing significant correlation between process variables. The V_i are considered as response variables, and their first order neighbors $nbr(G, V_i)$ are predictor variables. Therefore, one can select process variables that are only correlated with each other. We apply linear and SVM regression models with a radial kernel to predict process variables that are represented as a node in graph G. To select the most important set of input features for the response process variable in each dataset, we apply the recursive feature elimination algorithm for each model. For outlier detection, we apply the algorithm for each response variable of every dataset. The details of these methods are provided in the following sections.

Algorithm 1 Steps for predictability analysis

$D = D_1, D_2, \ldots, D_n$ are $|D|$ datasets.
Let G be the inferred correlation network.
Let, $y = f(X)$ be the regression model to optimize the fitting parameters and predict the process variable y.
$O(x)$ is the outlier function that returns the outlier index vector.
y is the response process variable.
X is a set of predictor variables.
R is the results array.
for i=1 to $|D|$ **do**
  for t=1 to $|V|$ **do**
    $y := t$
    $X = nbr(G, t, 1)$ ▷ nbr selects all the first order neighbors of the input vertex
    Find the best fit solution for $y = f(X)$
    Find the best fit solution for $y(O(y)) = f(X(O(y))$
    R:= Calculate useful statistic, R^2, RMSE, and RMSE on test data feature weights.
  end for
end for

4.1. Recursive feature elimination

The recursive feature elimination algorithm is an iterative process to rank the input features based upon their performance with the model. In each step of the iteration, the lowest ranking feature is eliminated. A brief description is provided in Algorithm 2. The recursive feature elimination is employed to select the best feature subset for each response process variable.

Algorithm 2 Recursive feature elimination algorithm

$F = F_1, F_2, \ldots, F_p$ is the feature vector of size $|F|$.
Let D be the dataset.
Let $R(F, D)$ be the ranking function that returns the lowest ranking feature.
▹ The ranking function assigns a score to a feature based on various ranking methods, such as the accuracy of the model in cross-validation.
n is the subset size.
Let F_n be an empty array.
for t=1 to $|F|$ **do**
    F_{tmp} = R(F,D)
    $F_n(t) = F_{tmp}$
    $F := F - F_{tmp}$
end for
Update F_n by selecting the top n feature.
Repeat the above steps for all the feature subsets for $n = 1, 2, \ldots |F|$.
Compare the performance of each subset F_n on the testing data.
Select the best performing subset.

4.2. SVM Regression

The support vector regression is based on the SVM concept [7, 25]. In SVM regression, the input data are mapped to the high dimensional feature space using nonlinear mapping, which is described as: Let $f : X \to \mathbf{R}$

$$f(x) =< w, \phi(x) > +b \tag{1}$$

where $\phi(x)$ is a high dimensional feature space. The kernel trick allows using a kernel function to calculate the inner product in the feature space, which is described as:

$$\hat{f}(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x) + b \tag{2}$$

where $\alpha = (K + \lambda I)^{-1} y$. In our regression model, we use the *radial kernel function*, which is described as:

$$k(x, x^{'}) = exp(\frac{\|x - x^{'}\|}{2\sigma^2}) \tag{3}$$

For feature selection in SVM regression, we apply the RFE method to find the most important features that predict the outcome with a higher accuracy. We apply backward feature elimination, which rejects the weakest predictor variable. In the analysis, the hyper-parameters $\sigma = \{.01, .05, .1 \ldots, 1\}$ and box constraints $C = \{0.01, .1, .5, 1, 2, 4, 8, 16\}$ are tuned for each iteration. We then perform feature elimination on the best model after tuning the hyper-parameters.

4.3. Outlier detection

Outliers are a common phenomenon when collecting variable data samples from various sensors in production processes. Outliers can represent anomalies or deviations from the normal range of the expected outcome, which represent interesting events. We separately modeled outliers of the response variables for the corresponding input variables. We employed univariate outlier detection using the method based on box plots for a skewed distribution. The boundaries of outliers with a skewed distribution are described as follows [26, 13]:

$$[Q1 - k * SIQRL; Q3 + k * SIQRU].$$

where $SIQRL = Q2 - Q1$ and $SIQRU = Q3 - Q2$ are the lower and upper semi-interquartile range, respectively. We select the points that are above the upper outlier fence.

4.4. Selecting key predictors

The key predictors for process variables as shown in Tables 3, 4, and 5 are selected as:

– Let the RFE algorithm select a process variable X_p as a key feature to predict y in the model $y = f(X_1, X_2, \dots X_r)$, for dataset D_i, where $i = \{1, 2, \dots, n\}$, and a total of $|D|$ datasets. The predictor variable X_p is weighted as:

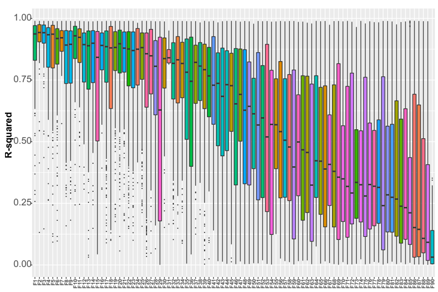
$$w_{y,X_p} = \frac{\#(X_p \text{ is selected as a key feature by RFE})}{|D|}$$

– Select X_p as a key feature to predict y if $w_{y,X_p} > \alpha$. Here, we select $\alpha = .5$.

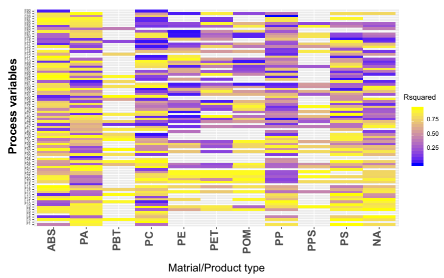
5. Results

5.1. Regression model

By applying Algorithm 1, we rank the process variables based upon their average R^2 values. In Figure 2a, we show R^2 distribution to predict different process variables. The x-axis displays the process variables from the best selecting models, and the y-axis shows the distribution of the R^2 values from the different models built on the various datasets for the different material types. The R^2 results indicate that some process variables consistently show a better performance with the training data and is without substantial variation. Some process variables show good predictability in some cases but vary in other production processes. Half of the process variables show poor predictability in their R^2 measure and vary significantly. Thus, they do not show linear dependency or stability on the other process variables. In such cases, these show low variance or nonlinear dependencies.



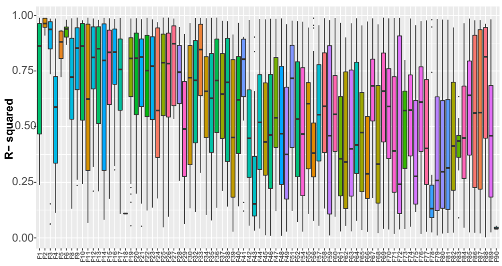
(a) The distribution of R-squared measure of regression models built upon different dataset for the predictability of different process variables (x-axis). The process variables are ordered as $F1, F2, \dots F90$ and mapped from the table in Appendix section.



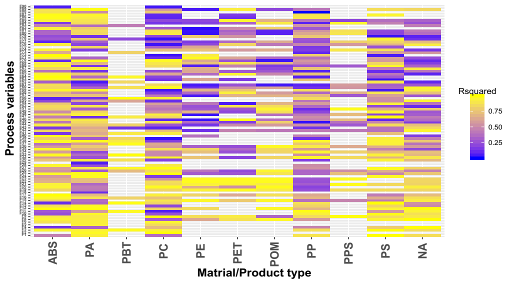
(b) The Average of R-squared measure of regression models built upon different dataset for the predictability of different process variables (x-axis). We grouped datasets based on producttype/material type. The process variables are ordered as $F1, F2, \dots F90$ and mapped from the table in Appendix section.

Fig. 2: Average R^2 of process variables as response variable in Regression model in different production processes.

Figure 2b shows the average R^2 measure of the models when predicting response process variables for different product groups. This analysis suggests that different predictive models for process variables consistently perform better or are product-dependent. The figure shows that the models that have the best predictability of the process variables demonstrate a higher average R^2, but this is not true in all cases. For example, in the product group *PP*, several best-predicted process variable models show comparatively lower average R^2 values, while only 13 process variables show good accuracy in their R^2 measure, as shown in Table 2. However, in the *PPO*, *PBT*, and *TPU* groups, the R^2 is higher in more of the models utilized to predict the process variables. This analysis is useful to classify the



(a) The distribution of R-squared measure of regression models built upon different dataset for the predictability of different process variables (x-axis). The process variables are ordered as $F1, F2, \ldots F90$ and mapped from the table in Appendix section.



(b) The Average of R-squared measure of regression models built upon different dataset for the predictability of different process variables (x-axis). We grouped datasets based on producttype/material type. The process variables are ordered as $F1, F2, \ldots F90$ and mapped from the table in Appendix section.

Fig. 3: Average R^2 of process variables as response variable in Regression model considering outliers of response variables in different production processes.

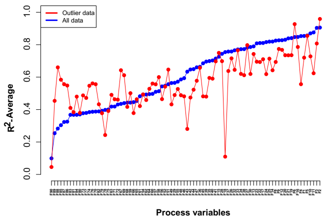


Fig. 4: The comparison of R^2 of linear models for predicting process variables with outlier data and the complete data. The x-axis show the process variables. The spearman rank correlation between R^2 values of models predicting process variables in outlier cases and non-outlier cases is 0.7305637.

process variables based on the product-related features and for tuning per their dependencies on each other for the different product groups.

Next, we show the R^2 distribution in Figure 3a for the process variable models in the case of outliers. The distribution has a high R^2 variation as explained by the input process variables. The average R^2 values for the specific product group cases are shown in Figure 3b. The process variable predictabilities are different from the high predictability with complete datasets for specific material types. The other result is the improved predictability of the process variables, which do not exhibit a high predictable behavior with the complete data. Some of these process variables are *F84, F85, . . . F89*. The improved linear predictability of these process variables means that they are not stable, have a high variance, and show interdependence with the other variables. Therefore, these should be considered in the production process and monitored when deviating from normal values. However, in the case of outliers, we cannot describe what "special events" they represent. For example, these may describe out of control processes or represent high variances in the process variables but with a controlled process or some data aberrations. Such variations are related to product quality issues. We conclude that the higher variance in the case of outliers for these variables and their corresponding predictor variables should be analyzed with the product quality data for a rational interpretation.

We also compare the average R^2 values for the models of different process variables in normal and outlier cases, as shown in Figure 4. Overall, the predictive performances of the process variables are worse than for the complete data. However, some cases show higher predictive performances. The rank of the process variables does not change significantly when measuring the Spearman rank correlation, $\rho = 0.730$. The $F18 : switch - overvolume$ significantly loses its predictability compared to other process variables, which means that setting the other dependent parameters cannot describe its variance in outlier cases. The models of the process variables that have significant differences in their R^2 values in outlier and non-outlier cases should be investigated in detail. If a process variable value goes beyond the specified limit, the other dependent parameters cannot be estimated correctly; hence, some processes deviate in this case.

Material Type	$R^2 \in (0.00304, 0.249]$	$R^2 \in (0.249, 0.494]$	$R^2 \in (0.494, 0.739]$	$R^2 \in (0.739, 0.985]$
PP	0.03	0.34	0.39	0.23
PE	0.07	0.30	0.32	0.32
PC	0.01	0.33	0.33	0.34
TPE	0.07	0.30	0.28	0.34
ABS	0.01	0.30	0.34	0.35
PPS	0.41	0.19	0.05	0.35
PA	0.00	0.34	0.29	0.37
PET	0.15	0.35	0.13	0.37
MIX	0.11	0.31	0.20	0.38
POM	0.07	0.16	0.34	0.42
PS	0.03	0.28	0.26	0.43
TPU	0.16	0.10	0.22	0.52
PBT	0.05	0.23	0.19	0.53
PPO	0.05	0.09	0.23	0.63

Table 2: proportion of predective process variables which have R^2 with in different range, $R^2 \in (0.739, 0.985]$ is the best cases where higher R^2 shows a better fit for the model

In the next analysis, we show the top predictive process variables and their respective predictor variables as estimated using the RFE algorithm with complete data and data with outliers only, as shown in Tables 3 and 4. The tables show some of the key process variables, such as *Plasticizing volume*, *Plasticizing time*, *switchover pressure*, *Injection pressure*, *shot volume*, and *cushion parameters*, which play important roles in the final quality of the products and illustrate a high predictability. Table 3 shows that only 20 process variables have $R^2 > 0.75$. The key predictor variables in this model vary from 2 to 17. The RMSEs for the testing data (in some cases in Tables 3 and 4) are lower than the training data RMSE. The RMSE of the training data is an average RMSE, which is calculated using a 20-fold cross-validation, with a set of fixed-size features and leads to an improved accuracy. However, the test RMSE is calculated with only the best fit model and only a single data set. Therefore, in some cases, there is a higher RMSE in the training cases than the testing data. The key process variables that play an important role in maintaining the product quality and the predictor variables (predictor process variables to which the response process variable shows dependence) should be utilized to optimize the process parameters and maintain product quality.

Response variable	Rsquared	RMSE	RMSE (test)	Key predictors
Plasticizing volume	0.91	2179.18	1.99	F6, F19, F21, F22, F23, F46, F48
Specific pressure at switch over	0.90	38.55	58.69	F4, F11, F13, F14, F16, F18, F29, F37, F41, F44, F47, F64
Shot volume end - corrected	0.90	2.44	2.20	F1, F3, F4, F56
Plasticizing time	0.88	0.47	0.52	F1, F3, F4, F10, F13, F14, F18, F33, F43
Material cushion end holding pressure	0.87	0.91	0.63	F1, F4
Plasticizing stroke	0.86	93.80	6.12	F22, F80
Material cushion after holding pressure	0.85	3.03	0.57	F10, F13
Shot volume	0.85	0.98	1.09	F10, F13
Specific injection pressure peak value	0.85	43.04	2306.53	F1, F3, F4, F10, F11, F13, F14, F18, F37, F43, F47
Decompression	0.85	81.50	3.40	F19, F21, F22, F23, F80, F81
Specific injection pressure peak value	0.84	412.64	380.51	F8, F9, F17, F22, F23, F46, F48, F80, F81, F86, F87
Material cushion after holding pressure	0.83	9.19	6.90	F19, F21, F22, F23, F46, F48
Shot volume	0.83	25.43	9.08	F19, F21, F22, F23, F46, F48, F80, F81
Material cushion after holding pressure	0.83	25.03	301.31	F19, F21, F22, F23, F46, F48, F80, F81
Specific holding pressure peak value	0.82	29.41	36.12	F3, F4, F11, F13, F16, F18
Material cushion actual value	0.82	0.24	0.43	F1, F3, F4, F10, F14, F41
Hydraulic pressure switchover value	0.81	293.40	92.00	F8, F9, F17, F19, F21, F46, F48, F80, F81, F86, F87
Specific pressure at switch over	0.81	4574.73	431.43	F8, F9, F17, F19, F21, F46, F48, F80, F81, F86, F87
Material cushion smallest value	0.79	0.72	1.80	F1, F4, F11
Torque mean value current cycle	0.79	21.69	31.00	F1, F4, F10, F13, F26, F27, F33, F38, F39, F55
Cycle time until start button	0.78	1.84	2.60	F1, F4, F26, F28

Table 3: Top 20 Process variables which are exhibiting best predictive behaviors with a Linear regression model.

5.2. SVM Regression Model

In the first step of the analysis, we train the model using SVM regression with Algorithm 1. We further tune the model with twenty-fold cross-validation using various hyper-parameter combinations. The results of the best performing models for different process variables are shown in Table 5, where the features are ordered based on the training error. The SVM model estimates the predictability of different process variables to provide an improved

Response variable	Rsquared	RMSE	RMSE (test)	Key predictors
Plasticizing volume	0.96	9.31	1.17	F80, F81, F23, F22
Decompression	0.93	18.59	5.18	F23, F22, F80, F81, F21, F48, F19, F35, F46
Plasticizing stroke	0.85	0.89	0.35	F22, F80, F81, F48, F23, F35, F46, F17, F87, F86
Specific pressure at switch over	0.81	40.26	1225.30	F16, F4, F44
Cycle time until start button	0.80	1.09	1.55	F26
Specific Injection pressure peak value	0.79	7549.36	1280.71	F3, F4, F1, F18
shot volume 1	0.77	1922.27	1.93	F22, F21, F19, F23, F81, F80, F48, F46
Material cushion after holding pressure	0.77	33.12	0.68	F21, F19, F80, F81, F22, F23
Metering performance	0.77	0.46	0.47	F27, F13, F11
Deviation of cycle time	0.75	8954.70	112.99	F81, F80, F21, F22, F19, F23, F46, F48, F9
Material cushion smallest value	0.74	0.34	0.22	F4, F29, F27
Specific injection pressure peak value	0.74	12594.50	40.56	F9, F81, F80, F46, F48, F8, F23, F22, F86, F87
Integral monitoring 1 micrograph	0.73	1001482.91	32404.83	F9, F22, F8, F80, F48, F81, F23, F19, F21, F46, F17, F36, F32, F26, F86, F25, F87, F35, F34
Spec. Injection pressure peak value	0.73	8175.29	4223.10	F9, F81, F8, F46, F48, F80, F22, F23, F17, F86, F87, F35
Material cushion end holding pressure	0.73	0.29	0.28	F4, F29, F13
Material cushion after holding pressure	0.72	0.27	0.55	F14, F13, F29
cycle time until end of demolding	0.72	774.05	577.17	F81, F21, F80, F22, F19, F23
Injection force 1	0.71	6.33	41.48	F21, F19, F23, F22, F81
material cushion actual value	0.71	6.54	1.09	F22, F19, F23, F21, F81, F80, F46, F48
Standstill time prior to cycle start	0.70	16.32	134992.00	F48, F21, F23, F46, F22, F19

Table 4: Top 20 Process variables which are show best predictive behaviours with Linear regression models trained with outlier data.

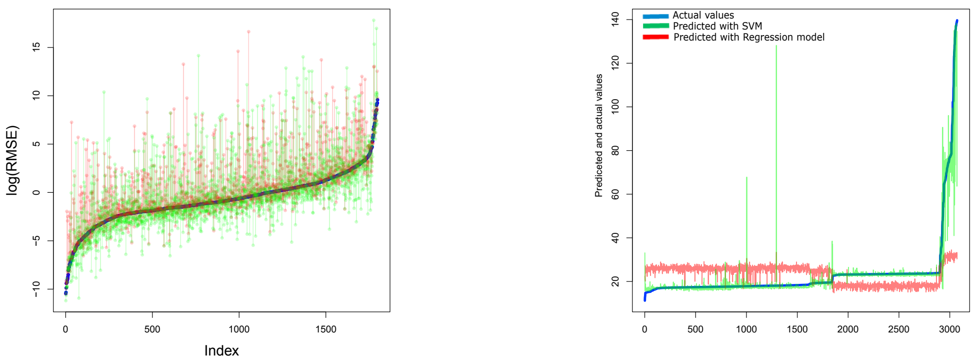
accuracy for the $RMSE$ (training and testing). A higher accuracy for various process variable predictions indicates a nonlinear relationship between the predictive and predictor process variables. In Table 5, some of the important key process variables, such as injection-, speed-, and molding-related process variables, show high accuracies when measured as the average RMSE in both training and testing sets. This also highlights that there is the least overfitting on the training data and, therefore, are fit to the nonlinear relationships. Another reason could be the least variation in the data. The top process variables that are predictable using the linear model are not in Table 5 except for the *mass cushion minimum value*. The key predictor variables differ in the SVM and linear models.

Response variable	$RMSE$	$RMSE(test)$	Key predictors
Injection force	0.01	0.02	F26, F27, F56, F73
Mold protection time	0.01	0.01	F26
Speed peak value	0.01	0.01	F10, F27
Speed peak value 1	0.01	0.01	F17
RPM Speed peak value	0.01	0.01	F22
Cycle time ejector	0.02	0.02	F26, F60, F69
Injection time	0.04	0.05	F26, F33
Injection time	0.04	0.04	F22, F26, F34
Cycle time holding pressure	0.08	0.13	F26, F27, F77
Injection force	0.09	0.10	F8, F17, F26, F32, F35, F36, F51, F61, F80, F81
Cycle time until end demolding	0.10	0.14	F26
Mold pause time	0.11	0.15	F60
Cycle time holding pressure	0.13	0.16	F26
Switchover volume actual value	0.13	0.17	F77
Plasticizing volume	0.18	0.26	F11, F14, F55
Material cushion end holding pressure	0.18	0.24	F26, F38, F62
switch over position actual value	0.24	0.25	F8, F22, F23, F26, F51
Material cushion smallest value	0.24	0.23	F14, F27, F60
Material cushion after holding pressure	0.25	0.29	F26, F37, F41
temperature zone 3	0.25	0.24	F17, F26, F74

Table 5: Top 20 Process variables which are exhibiting best predictive behaviours with SVM regression.

5.3. Comparison of Models

The three types of models identify sets of process variables that are predictable with various accuracies based on other predictor variables. The SVM and linear models are not directly comparable in terms of R^2. Therefore, we compare three types of models for process variables with each other using the average RMSE, as shown in Figure 5a. The RMSE values vary for linear models w.r.t. SVM regression models (blue). A large number of RMSE values for the linear models are higher than those estimated from SVM regression. However, a low RMSE in many SVM regression models can also be from overfitting for various prediction models. Therefore, feature selection using SVM regression is useful only when the results are consistent with the training and testing data. We show one example in Figure 5b for the test data of the linear and SVM regression models for the *Actual pressure value pump* process variable. The RMSE for the training and testing data in this case for the linear model are 16.08332 and 15.44713 while the R^2 measure is 0.065. Comparing these results for the SVM regression with RFE shows a high accuracy with the training and testing results, which are 5.35841 are 5.292136, respectively. Here, our main objective is to select key-dependent features using linear and SVM regressions. We give preference to the feature selection using SVM regression when there is a low R^2 value for the corresponding linear model and a significant improvement in the RMSE for the training and testing cases from the SVM regression model. In the example shown in Figure 5b, the $R^2 < .5$ and the RMSE is significantly higher in the training and testing cases than the SVM regression model.



(a) The RMSE comparison of SVM regression(blue), Linear regression (red) and Linear Regression (green)

(b) Example of *Actual pressure value pump* process variable with SVM regression and linear regression model

Fig. 5: RMSE comparison and example of predicted values of test data in SVM regression and linear regression model

We also show the top 10 models of the process variables for each product category, which are predictable using other input process variables. We rank the linear models based on the R^2 and SVM regression using the average training and testing errors. The results are shown in Table 6. The predictability of the process variables for each material type and different methods indicate 79 different process variables, which are predictable in different cases. Some of the process variables in the linear model are accurately predictable for at least 5 product categories. Some of these are general parameters and are always tuned to control production quality and remain stable for processes under control.

Many process variables in the linear model, which are highly predictable, are material specific. One interesting result is that the process variables modeled with linear regression form 1 to 47 show high predictability for outlier and non-outlier data and, in some cases of SVM models. The other process variables are showing predictability only in the SVM model or models with the outliers data. Which means that these process variables are nonlinearly related (for SVM) as they show lowe RMSE for training and testing for multiple data set. In the outlier cases, 61 to 79 are some of the *Temperature, hydraulic pressure, cycle time, and switching over* process variables are predictable for some material types. The predictability of process variables in the outlier case is a subject of the investigation that if outliers' values indicate the sensitivity to some production error or out of control process. Moreover, what stage of production we observe outliers. Additionally, are these outliers the primary indications of production process error or are they just observed or recorded in later stages of production failures and are not important.

5.4. Comparison with related studies

We compare the predictability of the process variables used to predict product quality and scrap rate in previous studies. We consider [23], which found statistical features of 12 process variables and those from [17, 14, 19]. Some process variables are common in these studies, giving a total of 17 variables. The average R^2 and mean $RMSE$ for the training and testing data of different models are shown in Table 7. Table 8 illustrates the process variables that are modeled using linear and SVM regression along with the corresponding highest weighted input process variables. Therefore, the response process variables show interdependence with these process variables. The results in Table 7 show that the input variables describe 80% of the variance of the top 10 process variables. The temperature zone-related process variables and cycle time process variables are the least predictable using the linear models. The

	Response process variable	Linear model	Linear model with outlier	SVM
1	Plasticizing volume	ABS, PA, PBT, PE, PET, PP, PS	PA	
3	Specific pressure at switch over	ABS, PA, PBT, PC, PE, POM, PS	PA, PBT, PC, PP	
13	Shot volume	MIX, PC, PET, PP, PPO, PPS, TPE	ABS, PC, PET, PP, PPS	
22	Specific injection pressure peak value	PA, PBT, PC, PET, PPO, PPS	PE, PET, PPS	
4	Material cushion end holding pressure	ABS, PC, PE, PS, TPU	PA, PP, PS	TPU
5	Material cushion after holding pressure	ABS, PE, POM, PS, TPU	PA, PP	TPU
6	Decompression	ABS, PA, PE, PP, TPE	ABS, PA, PC	
8	Plasticizing time	ABS, PA, PBT, PE, PP	PBT, PP	
10	Specific injection pressure peak value	ABS, PA, PC, PP, PS	PA, PC	
11	material cushion actual value	MIX, PBT, PPO, PPS, TPE		TPE
2	Shot volume end - corrected	ABS, PA, PBT, PE		PE
7	Plasticizing stroke	ABS, PE, PET, PP	PA, PC	
15	Material cushion after holding pressure	MIX, PBT, PPS, TPE	PET, POM, PS	TPE
17	Injection force	MIX, PA, PC, PET	PPS	PBT, PC, PET, POM, PP, PS, TPE
21	Specific injection pressure peak value	PA, PBT, PC, PET	PE, PPS	
9	Material cushion after holding pressure	ABS, MIX, PP	PE, POM, PS	
12	Integral monitoring 1 micrograph	MIX, PC, PET	ABS, PC, POM	
14	material cushion actual value	MIX, PET, PPS		TPE
23	Torque mean value current cycle	PBT, PS, TPU	PBT	
24	Specific holding pressure peak value	PBT, POM, PS	PC, PP, PS	
25	material cushion actual value	PC, POM, PS	PA, PS	
26	Shot volume	PC, PE, POM		
27	Material cushion smallest value	PE, POM, TPU	ABS, PA, PP, PS	
28	Hydraulic pressure switchover value	PET, PPO, PPS	PE, PET	
29	Deviation of cycle time	PET, TPE, TPU	PE, POM	
36	Cycle time until end demolding	PPO, TPE, TPU		
37	Torque mean value current cycle	PPO, PPS, TPE	PE, POM	
38	Plasticizing number	PPO, PPS, TPU	PE, POM	
18	Specific holding pressure peak value	MIX, PPO		
19	Hydr. holding pressure peak value	MIX, PPO	PET	PET
31	Torque peak value current cycle	POM, PS		
33	Plasticizing number	POM, PP	PBT, PP	
39	Specific pressure at switch over	PPO, PPS	PE, PET, PPS	
16	temperature zone 1	MIX		
20	Cycle time acknowledgement take-off station	PA		
30	Clamping force at switch over	POM		
32	Switchover volume actual value	POM		POM
34	Injection force	PP	PBT	ABS, PA, PBT, PC, PE, POM, PP, TPU
35	Injection time	PP	PBT, PC	ABS, PA, PE, PP, TPU
40	temperature zone 8	PPS	PS	
41	Metering performance	PS	ABS, PS	TPU
42	Cycle time automatic	TPE		
43	Standstill time prior to cycle start	TPE	ABS, PC	PS
44	Cycle time	TPE	POM	
45	Cycle time until start button	TPU	PA, PBT, PS	
46	Screw speed	TPU		
47	Speed peak value	TPU		ABS, PA, PBT, PC, PE, PET, POM, PP, PS
48	Speed peak value			ABS, PA, PC, PE, PET, POM, PP
49	Mold protection time			ABS, PA, PBT, PC, PE, POM, PP, TPU
50	Speed peak value			ABS, PA, PBT, PC, PE, POM, TPU
51	Cycle time ejector		PPS	ABS, PA, PBT, PC, PE, PET, PP, PS, TPE
52	Injection time		ABS, POM	ABS, PA, PBT, PC, PE, PET, POM, PP, PS
53	Mold pause time			ABS, PC, PE
54	Cycle time holding pressure			PA, POM, PP, TPU
55	Switchover volume actual value			PA, POM, PP, TPU
56	Cycle time holding pressure			PBT, PET, PS
57	Plasticizing volume			PBT
58	Peak injection speed/ rate			PBT
50	Screw speed			PC
60	switch over position actual value		PP	PET, PS, TPE
61	Switchover volume actual value		PP, PS	PET, PS, TPE
62	Cycle time until end demolding		PPS	PET, TPE
63	temperature zone 11		ABS	PS
64	Cooling time		PET	TPE, TPU
65	temperature zone 12		ABS	
66	temperature zone 7		ABS	
67	Closing force		PBT	
68	Peak injection speed/ rate		PBT, PC	
69	Cycle time		PBT, POM	
70	temperature zone 14		PE	
71	temperature zone 13		PE	
72	temperature zone 2		PET	
73	pump pressure acutla value		PET	
74	temperature zone 3		PET	
75	pump volume actual value		POM	
76	temperature zone 5		PPS	
77	Back pressure peak value		PPS	
78	Specific back pressure peak value		PPS	

Table 6: Different Process Variables (respons), which show high predictability in terms of R^2 (linear model), and RMSE (svm) in different product categories in different predictive models

training and testing errors for some of the process variables using linear models are relatively large for both outlier and non-outlier cases. Such errors can be random aberrations or due to more specific characteristics of products and machines. These values should be examined to determine what production steps are needed to better understand their predictability. The temperature zones, cycle time, and torque-related process variables show consistent average RMSEs in the training and testing cases for the SVM models. This suggests that these process variables are nonlinearly related to other input process variables. Table 7 shows the various input process variables that describe the predictability of the output response process variables. Therefore, the response process variables that play significant roles in maintaining the product quality depend on different process variables. The predictor process variables include *hydraulic pressure*, *injection pressure*, *shot volume*, *switch over position*, *cushion*, *closing force*, *temperature*, and *torque related parameters*. The key input predictor process variables are also common for response process variables, which shows complex interdependence characteristics between the different process variables. To maintain an efficient production process, we should consider tuning all the relevant process parameters. However, many of the process variables can be redundant or non-adjustable and generated after the process ends. Further interpretation of the product and material-dependent process variables for tuning is left to domain experts.

S. No.	process varialbes (response)	Linear model			Linear model (outliers)			SVM	
		mean R^2	mean RMSE	mean RMSE (test)	mean R^2	mean RMSE	mean RMSE (test)	mean RMSE	mean RMSE (test)
1	Decompression	0.85	81.50	3.40	0.93	18.59	5.18	1.43	1.50
2	Material cushion after holding pressure	0.83	25.03	301.31	0.69	14.35	0.75	1.24	1.71
3	Shot volume	0.83	25.43	9.08	0.77	1922.27	1.93	2.00	2.03
4	Plasticizing time	0.88	0.47	0.52	0.62	3.48	0.72	0.39	0.63
5	Material cushion after holding pressure	0.83	9.19	6.90	0.77	33.12	0.68	0.59	0.73
6	material cushion actual value	0.81	$> 10^8$	0.74	0.71	6.54	1.09	0.28	0.34
7	Specific injection pressure peak value	0.84	412.64	380.51	0.74	12594.50	40.56	3.50	3.27
8	Integral monitoring 1 micrograph	0.83	228797.56	199983.04	0.74	1001482.91	32404.83	4900.36	5149.22
9	Hydraulic pressure switchover value	0.81	293.40	92.00	0.69	2708.07	56.23	4.17	4.58
10	Injection force	0.82	$> 10^8$	1.91	0.71	6.33	41.48	0.09	0.10
11	Standstill time prior to cycle start	0.74	16.87	3.23	0.70	16.32	134992.00	1.52	1.79
12	Torque peak value current cycle	0.70	18.63	63.24	0.48	51.86	26.56	8.54	8.50
13	Torque peak value current cycle	0.69	$> 10^8$	2527.24	0.48	233.09	2411.02	11.34	11.97
14	temperature zone 11	0.44	677.51	9.95	0.50	73.66	1.91	4.29	4.24
15	Cycle time	0.50	72.54	31.74	0.56	2838.16	120.46	1.93	2.20
16	temperature zone 7	0.43	25.43	13.08	0.64	11.84	1.36	0.39	0.38
17	temperature zone 6	0.39	41.71	16.58	0.56	283.89	1703.28	0.58	1.33

Table 7: List of process variable which are used by some previous studies to model scrap rate.

S. No.	Predictive process variables which were identified in previous studies		Predictor process variables.
1	F6	Decompression	F19, F21, F81
2	F8	Material cushion after holding pressure	F19, F21, F22, F23, F46, F48, F80, F81
3	F9	Shot volume	F19, F21, F22, F23, F46, F48, F80, F81
4	F11	Plasticizing time	F13, F18, F24
5	F12	Material cushion after holding pressure	F19, F21
6	F17	material cushion actual value	F19, F21, F22, F23, F80, F81
7	F19	Specific injection pressure peak value	F8, F9, F22, F23, F46, F48, F80, F81
8	F20	Integral monitoring 1 micrograph	F8, F9, F17, F19, F21, F22, F23, F46, F48, F80, F81, F86, F87
9	F22	Hydraulic pressure switchover value	F8, F9, F19, F21, F46, F48, F80, F81
10	F25	Injection force	F21, F22, F23
11	F28	Standstill time prior to cycle start	F19, F21, F22
12	F39	Torque peak value current cycle	F1, F4, F30
13	F45	Torque peak value current cycle	F8, F9, F19, F21, F22, F23, F46, F48, F80, F81
14	F59	temperature zone 11	F19, F21, F22
15	F60	Cycle time	F22, F80, F81
16	F67	temperature zone 7	F9, F19, F21, F22, F23, F80, F81
17	F70	temperature zone 6	F19, F21

Table 8: Process variables from previous studies which play a key role in product quality and its corresponding key predictor variables in the current analysis.

6. Conclusion

This paper analyzes process variable data from injection molding processes to identify interdependent process variables. We construct a correlation network of process variables that is far from trivial. First, it is based on several datasets that search for process variables that are correlated during production processes. This allows filtering out other process variables that are not correlated to that particular response process variable. Therefore, there are common input features that can be used to train the model. Additionally, network-based analysis can be further extended by examining nonlinear relationships or conditional independencies through mutual information or partial correlation tests. However, this is beyond the scope of the current analysis. The networks' structural properties are also used to understand and optimize the characteristics of the process parameters during production for different product types and adversarial cases in production. This was evident in our analysis when the modeled process variables showed a varying dependence with each other in both outlier and non-outlier data.

We see that many process variables where the variance is not explained from regression models (low R^2) show higher R^2 values in the case of outliers. Similarly, many response process variables in SVM regression models consistently exhibit lower RMSEs for the training and testing data. These process variables can be non-linearly related to input process variables. Therefore, we select them as nonlinearly dependent features with the response process variables only when the estimated R^2 is below some threshold. For complex production processes that require several process parameters, such as the production time, production type, material type, and product type variations, we need to understand the underlying relationships between different processes parameters. In this analysis, we examined more than 50,000 (linear and SVM) models to assess the predictability of each process variable available under normal and outlier cases. This analysis can be further taken to domain experts to classify the most important process parameters in production processes, who consider various other factors or process variables that influence them. Therefore, more accurate and comprehensive tuning of process parameters ensures efficient production outputs. The predicted key process variables and their dependencies with other process variables (predictors) in different cases can be further utilized to develop functional descriptors or features as process fingerprints. Such process fingerprints are described as specific features to be monitored or controlled for efficient production [22]. The other benefit is to efficiently monitor process control charts. The underlying dependencies, as estimated with machine learning models between different process parameters, allow predicting or optimizing future effects in advance. Thus, the production team in-charge can take necessary actions in advance by readjusting process parameters or other relevant maintenance steps. Additionally, the meta-knowledge and previous understanding of process variables' dependence and predictability can be compared with current results from various products and material-specific production scenarios. Any significant results from this analysis that provide a newer understanding of process variables' dependence and predictability can be used to update the process meta-knowledge.

From this analysis, we can better understand the more general and underlying relationship between process variables and their predictability, which affect the quality of the production process. A regression model with cross-validation (CV) and recursive feature selection for each process variables as a response and the other process variables as predictors provides good predictability in terms of R^2 and a reliable set of features for many process variables. However, the explained variance of these process variables in terms of R^2 varies for different product types and in the case of outliers. The variation in R^2 of the response variables can depend on many other aspects of the production process, which can be various product-specific and machine-specific features. The product- and machine-specific features that are absent in the model can be the reason for the unexplained variance in the R^2 of the predictive models.

In some cases, the process variable remains stable and shows no variability, indicating it is not dependent on other process variables. In the case of outliers, we see that the average explained variance, R^2, of the response process variables is lower in some cases compared to the linear model with a complete data set. A lower predictability means the coordination among process variables is different from under normal situations, which can be due to production-related in-between changes in the process parameters. This may require tuning of process parameters for maintaining processes or it is showing out of control processes. However, we find that the process variables with a lower predictability in the models with complete data sets exhibit a higher R^2 when outliers are present. The higher predictability in these cases means that the process variables and other predictor variables related to machine activity vary and increase significantly compared with the normal range. These results are important for machine-operators and engineers for interpretation. One interesting result is the comparison of top predictive variables from the models where we see many common process variables between the regression models with and without outliers. However, we do not see any significant overlap of the top process variables (predictive variables) between the linear and SVM models. These results show that these process variables have a nonlinear relationship with other predictor process variables, or some process variables remain stable and are not influenced by others.

The process variables identified as important features related to the scrap rate and product quality in previous studies also show a dependence on other process variables, as shown in Table 7. The process variables related to scrap and product quality mainly show dependencies with the *Hydraulic pressure*, *Injection pressure*, *Shot volume*, *Torque*, *Cushion pressure*, and *Screw position* related process variables. These process variables are used for modeling and optimization regarding product quality and process control in injection molding processes.

Acknowledgements

This paper is a part of X-pro project. The project is financed by research subsidies granted by the government of Upper Austria.
The data in this paper was used from the project ADAPT - FFG Number: 862011 funded by the Government of Upper Austria in their programme "Innovative Upper Austria 2020".

References

[1] Charu C. Aggarwal. *An Introduction to Outlier Analysis*, pages 1–34. Springer International Publishing, Cham, 2017.

[2] Mahmoud A. Barghash and Faisal Alkhannan Alkaabneh. Shrinkage and warpage detailed analysis and optimization for the injection molding process using multistage experimental design. *Quality Engineering*, 26(3):319–334, 2014.

[3] M. Berins and Society of the Plastics Industry. *Plastics Engineering Handbook Of The Society Of The Plastics Industry*. Plastics Engineering. Springer US, 1991.

[4] Tao C. Chang and Ernest Faison III. Shrinkage behavior and optimization of injection molded parts studied by the taguchi method. *Polymer Engineering & Science*, 41(5):703–710, 2001.

[5] Wen-Chin Chen, Manh-Hung Nguyen, Wen-Hsin Chiu, Te-Ning Chen, and Pei-Hao Tai. Optimization of the plastic injection molding process using the taguchi method, rsm, and hybrid ga-pso. *The International Journal of Advanced Manufacturing Technology*, 83(9):1873–1886, Apr 2016.

[6] Xuan-Phuong Dang. General frameworks for optimization of plastic injection molding process parameters. *Simulation Modelling Practice and Theory*, 41:15 – 27, 2014.

[7] Harris Drucker, Chris J. C. Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. Support vector regression machines. In *Proceedings of the 9th International Conference on Neural Information Processing Systems*, NIPS'96, pages 155–161, Cambridge, MA, USA, 1996. MIT Press.

[8] S. Ebnesajjad and P.R. Khaladkar. *Fluoropolymer Applications in the Chemical Processing Industries: The Definitive User's Guide and Databook*. Plastics Design Library. Elsevier Science, 2004.

[9] Célio Fernandes, António José Pontes, Júlio César Viana, and António Gaspar-Cunha. Modeling and optimization of the injection-molding process: A review. *Advances in Polymer Technology*, 37(2):429–449, 2018.

[10] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

[11] Suchana A. Jahan and Hazim El-Mounayri. Optimal conformal cooling channels in 3d printed dies for plastic injection molding. *Procedia Manufacturing*, 5:888 – 900, 2016. 44th North American Manufacturing Research Conference, NAMRC 44, June 27-July 1, 2016, Blacksburg, Virginia, United States.

[12] Satadru Kashyap and Dilip Datta. Process parameter optimization of plastic injection molding: a review. *International Journal of Plastics Technology*, 19(1):1–18, Jun 2015.

[13] A. C. Kimber. Exploratory data analysis for possibly censored data from skewed distributions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39(1):21–30, 1990.

[14] Ting Mao, Yun Zhang, Yufei Ruan, Huang Gao, Huamin Zhou, and Dequn Li. Feature learning and process monitoring of injection molding using convolution-deconvolution auto encoders. *Computers Chemical Engineering*, 118:77 – 90, 2018.

[15] Erfan Oliaei, Behzad Shiroud Heidari, Seyed Mohammad Davachi, Mozhgan Bahrami, Saeed Davoodi, Iman Hejazi, and Javad Seyfi. Warpage and shrinkage optimization of injection-molded plastic spoon parts for biodegradable polymers using taguchi, anova and artificial neural network methods. *Journal of Materials Science Technology*, 32(8):710 – 720, 2016.

[16] Michael Packianather, Christian Griffiths, and Wan Kadir. Micro injection moulding process parameter tuning. *Procedia CIRP*, 33:400 – 405, 2015. 9th CIRP Conference on Intelligent Computation in Manufacturing Engineering - CIRP ICME '14.

[17] B. Ribeiro. Support vector machines for quality monitoring in a plastic injection molding process. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 35(3):401–410, Aug 2005.

[18] D.V. Rosato and M.G. Rosato. *Injection Molding Handbook*. Springer US, 2012.

[19] Gurjeet Singh and Ajay Verma. A brief review on injection moulding manufacturing process. *Materials Today: Proceedings*, 4(2, Part A):1423 – 1433, 2017. 5th International Conference of Materials Processing and Characterization (ICMPC 2016).

[20] G. Taguchi, E.A. Elsayed, and T.C. Hsiang. *Quality engineering in production systems*. McGraw-Hill series in industrial engineering and management science. McGraw-Hill, 1989.

[21] G. Taguchi, S. Konishi, S. Konishi, and American Supplier Institute. *Orthogonal Arrays and Linear Graphs: Tools for Quality Engineering*. Taguchi methods. American Supplier Institute, 1987.

[22] Guido Tosello. Product/process fingerprint in micro manufacturing. *Micromachines*, 10:340, 05 2019.

[23] Shailesh Tripathi., Sonja Strasser., Christian Mittermayr., Matthias Dehmer., and Herbert Jodlbauer. Approaches to identify relevant process variables in injection moulding using beta regression and svm. In *Proceedings of the 8th International Conference on Data Science, Technology and Applications - Volume 1: DATA,*, pages 233–242. INSTICC, SciTePress, 2019.

[24] Resit Unal and Edwin B. Dean. Taguchi approach to design optimization for quality and cost : An overview. 1991.

[25] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, Berlin, Heidelberg, 1995.

[26] M. Walker, Yinaze Dovoedo, Subha Chakraborti, and Curry Hilton. An improved boxplot for univariate data. *The American Statistician*, pages 1–13, 03 2018.

[27] Fei Yin, Huajie Mao, and Lin Hua. A hybrid of back propagation neural network and genetic algorithm for optimization of injection molding process parameters. *Materials Design*, 32(6):3457 – 3464, 2011.
[28] Fei Yin, Huajie Mao, Lin Hua, Wei Guo, and Maosheng Shu. Back propagation neural network modeling for warpage prediction and optimization of plastic products during injection molding. *Materials Design*, 32(4):1844 – 1850, 2011.
[29] Boštjan Zafošnik, Uroš Božič, and Blaž Florjanič. Modelling of an analytical equation for predicting maximum stress in an injections moulded undercut geometry during ejection. *International Journal of Precision Engineering and Manufacturing*, 16(12):2499–2507, nov 2015.

APPENDIX

S. No.	feature id	Process variable
1	F1	Material cushion after holding pressure
2	F2	Plasticizing volume
3	F3	Specific pressure at switch over
4	F4	Shot volume
5	F5	Plasticizing stroke
6	F6	Decompression
7	F7	Shot volume end - corrected
8	F8	Material cushion after holding pressure
9	F9	Shot volume
10	F10	Material cushion end holding pressure
11	F11	Plasticizing time
12	F12	Material cushion after holding pressure
13	F13	Material cushion smallest value
14	F14	Specific holding pressure peak value
15	F15	Deviation of cycle time
16	F16	Specific injection pressure peak value
17	F17	material cushion actual value
18	F18	Switchover volume actual value
19	F19	Specific injection pressure peak value
20	F20	Integral monitoring 1 micrograph
21	F21	Specific injection pressure peak value
22	F22	Hydraulic pressure switchover value
23	F23	Specific pressure at switch over
24	F24	material cushion actual value
25	F25	Injection force
26	F26	Cycle time until end demolding
27	F27	Cycle time until start button
28	F28	Standstill time prior to cycle start
29	F29	Switchover volume actual value
30	F30	Torque mean value current cycle
31	F31	Cycle time acknowledgement take-off station
32	F32	Plasticizing number
33	F33	Metering performance
34	F34	Torque mean value current cycle
35	F35	material cushion actual value
36	F36	Cycle time
37	F37	Injection force
38	F38	Plasticizing number
39	F39	Torque peak value current cycle
40	F40	Integral monitoring 2 micrograph
41	F41	Injection time
42	F42	Closing force
43	F43	Clamping force at switch over
44	F44	Clamping force peak value
45	F45	Torque peak value current cycle
46	F46	Hydr. holding pressure peak value
47	F47	Peak injection speed/ rate
48	F48	Specific holding pressure peak value
49	F49	temperature zone 1
50	F50	
51	F51	Injection time
52	F52	pump volume actual value
53	F53	Cycle time holding pressure
54	F54	temperature zone 8
55	F55	Plasticizing time
56	F56	Plasticizing volume
57	F57	Screw speed
58	F58	Cycle time automatic
59	F59	temperature zone 11
60	F60	Cycle time
61	F61	Peak injection speed/ rate
62	F62	Speed peak value
63	F63	temperature zone 12
64	F64	Closing force
65	F65	pump pressure acutla value
66	F66	Plasticizing volume
67	F67	temperature zone 7
68	F68	Speed peak value
69	F69	Cooling time
70	F70	temperature zone 6
71	F71	temperature zone 9
72	F72	temperature zone 14
73	F73	Mold protection time
74	F74	temperature zone 5
75	F75	temperature zone 13
76	F76	temperature zone 4
77	F77	Cycle time ejector
78	F78	Specific back pressure peak value
79	F79	temperature zone 10
80	F80	Back pressure peak value
81	F81	Specific back pressure peak value
82	F82	Speed peak value
83	F83	Mold pause time
84	F84	Cycle time until end demolding
85	F85	temperature zone 3
86	F86	switch over position actual value
87	F87	Switchover volume actual value
88	F88	Cycle time holding pressure
89	F89	temperature zone 2
90	F90	Screw speed

Table 9: Process variables mapped with feature ids.