A Convex Two-Dimensional Variable Selection Method for the Root-Cause Diagnostics of Product Defects

Chengyu Zhou and Xiaolei Fang[a]

[a]*Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, 915 Partners Way, Raleigh, 27695, NC, USA*

Abstract

Many multistage manufacturing processes consist of multiple identical stages. The root cause diagnostics of the product quality defects of these processes often involves the simultaneous identification of crucial stages and process variables that are related to product anomalies. In the literature, this is typically achieved by using penalized matrix regression that regresses the index of product defect against a matrix whose rows and columns respectively represent the stages and process variables. However, most existing models have some limitations that compromise their applicability and/or performance. For example, some models have an assumption on the rank of the coefficient, which often cannot be satisfied; some others formulate a non-convex optimization criterion that easily results in a local optimum. Also, most models only provide diagnostics results with group-wise (i.e., stage- and variable-wise) sparsity. To address these challenges, this article proposes a novel convex two-dimensional variable selection method that can inspire both group-wise and element-wise sparsity. This is accomplished by proposing a new generalized matrix regression model and simultaneously penalizing the rows, columns, and elements of the regression coefficient matrix using an ℓ_2, ℓ_2, and ℓ_1 norm, respectively. Simulated and real-world data are used to validate the effectiveness of the proposed method.

Keywords: Fault/Defect Diagnostics, Quality Control, Penalized Matrix Regression, Group Lasso, Sparsity, Generalized Linear Model

1. Introduction

Many Multistage Manufacturing Processes (MMPs) consist of identical stages, units, stations, or operations. For example, Figure 1 demonstrates

the schematic layout of a hot strip manufacturing process that is widely used in the steel-making industry. The primary function of the process is to reheat the semi-finished steel slabs to the rolling temperature using furnaces, roll them thinner and longer through a series of roughing and finishing rolling mill stands, and finally coil up the lengthened steel sheet. The finishing mill of the hot strip manufacturing process comprises seven identical stages (denoted as "F1" to "F7" in Figure 1), which have the most significant influence on the final thickness and quality of the product (i.e., steel strip) [9]. Another example of an MMP process with identical stages is the Additive Manufacturing (AM, also known as 3D printing) that fabricates products layer by layer [5], in which each layer can be seen as one stage. One of the characteristics of these MMPs is that all the stages have the same process variables (including process control parameters and sensing data from condition monitoring sensors) since they are identical. For example, each of the seven stages of the finishing mill in Figure 1 has the following process variables: *target speed of rollers*, *the measured speed of rollers*, *looper value*, *target force on both sides of the rollers*, *the measured force on the work side of rollers*, *measured force on the transfer side of rollers*, *roller gap*, *looper height*, and *temperature*, etc [7]. The values of these process variables usually vary from one product to another since they are either directly adjusted or passively affected by a closed-loop feedback control algorithm that tries to monitor and control the process in real-time for product quality guarantee. The inappropriate values of these process variables may result in defects in product quality. For example, Figure 2 illustrates products without and with quality defects from a hot strip manufacturing process. The root cause diagnostics of product quality defects focuses on identifying the crucial process variables (and their stage locations) whose inappropriate values are responsible for the defects of quality. The diagnostic results can be used to guide the revision of the feedback control algorithm to reduce the probability of or even avoid fabricating defective products in the future.

A typical technique for the root cause diagnostics of product defects is the LASSO-based regularization models [13], which have also been widely used in a variety of reliability applications such as risk assessment [1], failure time prediction [4, 10], resilience analysis [12], and uncertainty quantification [11]. This is achieved by regressing the binary quality indices of products (the quality index is 0 if a product is defective and 1 otherwise) against the process variables pooled from all the stages via logistic regression and penalizes the regression coefficients using an ℓ_1 norm to inspire sparsity. Any process

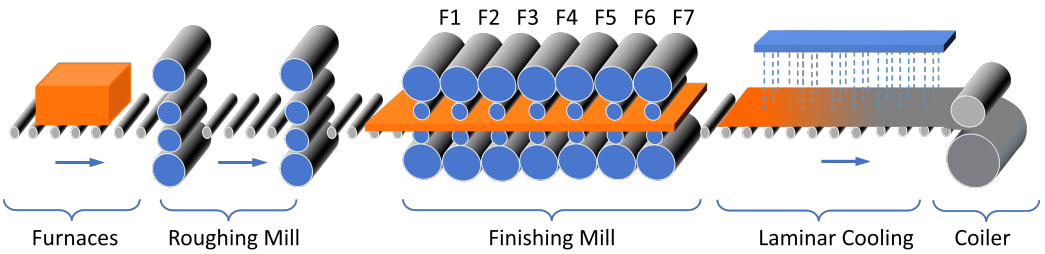
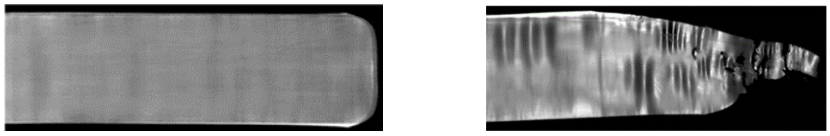


Figure 1: The schematic layout of the hot strip mill.



(a) A product without quality defects

(b) A product with quality defects

Figure 2: Products with and without quality defects from a hot rolling process [2]

variables with nonzero coefficients are considered to be responsible for the product defects. However, one of the key limitations of the LASSO is that it cannot provide a structured solution. Specifically, as illustrated in Figure 3, a specific process variable might be identified as crucial in some stages and non-crucial in other stages, so it is challenging to determine if this process variable is important or not. Similarly, in a specific stage, some process variables are selected but some others are not, and thus the significance of the stage cannot be decided either. Such an unstructured solution does not provide much useful information to guide the revision of the feedback control algorithm since it usually selects many variables in many stages and does not show the overall importance of each variable and stage. Therefore, the control algorithm experts prefer a more structured solution that identifies a few important stages and process variables such as the ones shown in Figure 4, in which stages 2, 4, and 6, and variables 2, 3, and 5 are selected.

To provide a structured result for diagnostics, one straightforward method is to conduct a two-step variable selection. The first step identifies the crucial stages, which can be achieved by applying the group LASSO [14] on the logistic regression model mentioned earlier (i.e., penalizing the regression

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
Variable 1	x		x			x	
Variable 2		x		x			x
Variable 3			x		x	x	
Variable 4	x					x	
Variable 5		x		x	x		x

Figure 3: Unstructured selection results ("x" represents the variable/stage is selected).

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
Variable 1							
Variable 2		x		x		x	
Variable 3		x		x		x	
Variable 4							
Variable 5		x		x		x	

Figure 4: Structured selection results ("x" represents the variable/stage is selected).

coefficients corresponding to the variables in each column in Figure 3 as a group using an ℓ_2 norm). Next, the non-crucial stages are excluded from the logistic regression model, and the group LASSO is used again to select the crucial process variables (i.e., penalizing the regression coefficients corresponding to the variables in each row in Figure 3 as a group using an ℓ_2 norm after removing the non-crucial stages). Of course, the order of the two steps can be switched–that is–we may identify the crucial process variables first and then the important stages. However, one limitation of the two-step selection method is that its selection result is not necessarily reliable since when identifying the crucial stages, the unimportant process variables will compromise the selection accuracy, and vice versa.

Other than the two-step variable selection method, several statistical learning methods have been developed in the literature that can possibly be used to achieve the goal of providing a structured result for diagnostics. Article [15] proposed a structured LASSO method that models a normally distributed quality index as a bilinear product of the process variable matrix. The sparsity is inspired by penalizing the two unknown vectors in the bilinear product using an ℓ_1 norm. Although the structured LASSO can provide a structured variable selection result, it has two limitations. First, it assumes that the quality index is from a normal distribution, which is often not valid

in real-world applications such as the hot strip mill example in Figure 1, in which the quality index is binary. Second, it can be proved that the structured LASSO method is equivalent to a penalized matrix regression where the rank of the coefficient matrix is one (see Proposition 1 in [7] for details). The rank assumption significantly restricts the generalizability of the structured LASSO method. This limitation was later addressed by [16], which developed a trace regression model, which is a penalized matrix regression model that simultaneously penalizes the rows and columns of the regression coefficient matrix using an ℓ_2 norm. However, this model still assumes that the quality index follows a normal distribution, which limits its applicability.

The latest effort in achieving the goal of providing a structured diagnostic result while addressing the aforementioned challenges is the two-dimensional variable selection method developed by [7]. The method decomposes the unknown regression coefficients matrix as a product of two matrices and simultaneously penalizes the rows of the first matrix and the columns of the second matrix using an ℓ_2 norm. By doing so, they address the rank assumption imposed on the coefficient matrix in the structured LASSO [15]. In addition, the two-dimensional variable selection method assumes that the quality index is from the exponential family, which significantly extends the application generalizability of the model since the exponential family consists of many specific distributions such as normal, Bernoulli, binomial, Poisson, gamma, etc. Although numerical studies have demonstrated the effectiveness of the two-dimensional variable selection method, it has two limitations. First, it was formulated as a non-convex optimization criterion, which implies the solution (i.e., the estimated regression coefficients) are usually not the optimal ones since they can easily get trapped in a local optimum. Second, it only provides solutions with group-wise sparsity, which is also a limitation of the models developed by [15] and [16]. Specifically, since it uses the ℓ_2 norm to inspire sparsity, it only identifies the crucial rows and/or columns of the process variable matrix but cannot tell the importance of the elements in each identified row/column [6]. In other words, in the final selection results, if a stage is identified as important, all the process variables in that stage will be identified as crucial. For example, Figure 5 summarizes the selected stages and process variables in Figure 4, where stages 2, 4, and 6 are identified as important, and process variables 2, 3, and 5 are selected. It can be seen that there is no element-wise sparsity since all the elements in the structured selection result are selected. In reality, control experts prefer a structured selection result with element-wise sparsity. This is because

	Stage 2	Stage 4	Stage 6
Variable 2	x	x	x
Variable 3	x	x	x
Variable 5	x	x	x

Figure 5: Structured selection results without element-wise sparsity ("x" represents the variable/stage is selected).

	Stage 2	Stage 4	Stage 6
Variable 2		x	x
Variable 3	x		
Variable 5	x		x

Figure 6: Structured selection results with element-wise sparsity ("x" represents the variable/stage is selected).

they are interested in knowing not only the overall importance of each stage but also the process variables with the highest priorities in each important stage. Figure 6 shows an example of the structured results with element-wise sparsity, in which, for example, stage 6 is chosen as an important stage. In the meantime, only variables 2 and 5 are selected in stage 6, and variable 3 is excluded. We would like to point out that although the result in Figure 6 looks similar to that of Figure 3, their statistical meanings are essentially different. The result in Figure 3 is from a LASSO-based method, and thus it cannot tell the group-wise (i.e., stage-wise and process variable-wise) importance. For example, the process variable 2 in stage 4 is selected in Figure 3, which only implies that the element itself (i.e., the variable 2 in stage 4) is important, and one can neither conclude that overall stage 4 is important nor say that overall variable 2 is important. Unlike Figure 3, the results in Figure 6 provide two meanings: group-wise and element-wise importance. For example, it first implies that overall stage 6 is important. Second, it suggests that in stage 6, the two important elements are variables 2 and 5.

To address the limitations of the aforementioned models and provide a structured diagnostics result with element-wise sparsity, this article develops a convex two-dimensional variable selection method. The method is based on a generalized matrix regression that regresses the quality index from the

exponential family against a predictor matrix, in which the rows represent the process variables and columns represent stages. Two types of norms are introduced for regularization. The rows and columns of the regression coefficient matrix are simultaneously penalized using an ℓ_2 norm to inspire group-wise sparsity. In the meantime, all the elements of the coefficient matrix are penalized using an ℓ_1 norm to inspire element-wise sparsity. The proposed method has the following advantages compared with existing methods developed by [7, 16, 15]. First, it can be applied to a variety of applications since it assumes that the quality index is from the exponential family, which consists of various specific distributions. Second, it is formulated as a convex optimization criterion, which converges to the global optimum and is computationally more efficient. Third, it yields a solution with both group-wise and element-wise sparsity, which provides more insights to guide the revision of the feedback control algorithm.

The rest of the article is organized as follows. In Section 2, we introduce the details of the convex 2D variable selection method. Sections 3 and 4 validate the effectiveness of the proposed method by comparing its performance with some benchmarks using simulated data and a real-world dataset from a hot strip steel mill. Finally, Section 5 concludes.

2. The convex 2D variable selection method

This article proposes a convex 2D variable selection method for the quality defects diagnostics of multistage manufacturing processes with identical stages. The proposed method is built based on generalized matrix regression, the response variable and predictor of which are the quality index and process variable matrix (see Figure 3 for an example), respectively. The quality index is assumed to be from the exponential family, which consists of many specific distributions such as normal, binomial, Poisson, exponential, gamma, and inverse Gaussian, etc. It yields diagnostics results with both group-wise and element-wise sparsity. The group-wise sparsity is achieved by simultaneously penalizing the rows and columns of the unknown regression coefficient matrix using an ℓ_2 norm, which helps select a few crucial stages and process variables that significantly affect the quality of products. The element-wise sparsity is accomplished by penalizing the entries of the coefficient matrix using an ℓ_1 norm. It identifies a few crucial process variables in each selected stage, which suggests that only a subset but not all the process variables in a crucial stage are important. Similarly, it also selects a few crucial stages

for each selected process variable, which implies even if a process variable is identified as overall important, it does not affect all but some of the stages.

We assume that there exists a historical dataset for model training (aka. parameter estimation). The dataset consists of the quality index and process variables of n products from the same multistage manufacturing process. We denote the quality index and process variable matrix of product i as $y_i \in \mathbb{R}$ and $\boldsymbol{X}_i \in \mathbb{R}^{s\times t}$, respectively, where s is the number of process variables, t is the number of stages. We assume that y_i follows a distribution from the exponential family. As a result, its probability mass or density function can be expressed as follows [8]:

$$\mathbb{P}(y_i|\theta_i,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\}, \tag{1}$$

where θ_i is location parameter and $\phi > 0$ is scale parameter. $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are known functions determined by the specific distribution in the exponential family. For example, if $y_i \sim \boldsymbol{N}(\mu_i,\sigma^2)$, the parameters are $\theta_i = \mu_i$, $\phi = \sigma^2$ and the functions are $a(\phi) = \phi$, $b(\theta_i) = \theta_i^2/2$ and $c(y_i,\phi) = -\frac{1}{2}(y_i^2/\phi + \log(2\pi\phi))$; if y_i follows a Binomial distribution, i.e., $y_i \sim \boldsymbol{B}(m,p_i)$, then $\theta_i = \log(\frac{p_i}{1-p_i})$, $\phi = 1$, $a(\phi) = 1$, $b(\theta_i) = m\log(1+e^{\theta_i})$, $c(y_i,\phi) = \log\binom{m}{y_i}$; if y_i follows a Poisson distribution, i.e., $y_i \sim \boldsymbol{Poisson}(\lambda_i)$, then $\theta_i = \log(\lambda_i)$, $b(\theta_i) = e^{\theta_i}, a(\phi) = 1$, and $c(y_i,\phi) = -\log(y_i!)$.

The connection between the quality index and the process variable matrix is established by using the link function. It maps the expectation of product i's quality index (denoted as μ_i) against its process variable matrix $\boldsymbol{X}_i$.

$$g(\mu_i) = \beta + \left\langle \boldsymbol{B}, \boldsymbol{X}_i \right\rangle, \tag{2}$$

where $\mu_i = \mathbb{E}(y_i|\boldsymbol{X}_i)$ is the expectation of quality index. $g(\cdot)$ is the known link function that depends on the specific distribution that y_i follows. For example, if y_i follows a Normal distribution, then $g(\mu_i) = \mu_i$, and the model is a normal matrix regression model $\mathbb{E}(y_i|\boldsymbol{X}_i) = \beta + \left\langle \boldsymbol{B}, \boldsymbol{X}_i \right\rangle$; if y_i follows a binomial distribution, one of the possible link functions is the logit function, $g(\mu_i) = \log(\frac{\mu_i}{1-\mu_i})$, and the model ends up with $\log(\frac{\mu_i}{1-\mu_i}) = \beta + \left\langle \boldsymbol{B}, \boldsymbol{X}_i \right\rangle$; if y_i follows a Poisson distribution, then $g(\mu_i) = \log(\mu_i)$, and the model is $\log(\mu_i) = \beta + \left\langle \boldsymbol{B}, \boldsymbol{X}_i \right\rangle$. β is the intercept, and $\boldsymbol{B} \in \mathbb{R}^{s\times t}$ is the unknown regression coefficient matrix. $\langle\cdot,\cdot\rangle$ is the matrix inner product operator.

The coefficient matrix $\boldsymbol{B}$ in Equation (2) can be estimated by using the

maximum likelihood estimation (MLE) method, which maximizes the following log-likelihood function:

$$\ell\big(\boldsymbol{B}, \beta\big) = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi), \tag{3}$$

where $\theta_i = \beta + \big\langle \boldsymbol{B}, \boldsymbol{X}_i \big\rangle$. Let $\mathcal{L}(\boldsymbol{B}, \beta)$ be the negative log-likelihood function, i.e., $\mathcal{L}(\boldsymbol{B}, \beta) = -\ell(\boldsymbol{B}, \beta)$. MLE works by solving the following optimization criterion:

$$\min_{\boldsymbol{B},\beta} \ \mathcal{L}(\boldsymbol{B}, \beta). \tag{4}$$

To achieve both group-wise (i.e., row- and column-wise) and element-wise selection, we simultaneously penalize each row, each column, and each element of matrix $\boldsymbol{B}$ using an ℓ_2, ℓ_2, and ℓ_1 norm, respectively. Mathematically, we solve the following optimization criterion:

$$\min_{\boldsymbol{B},\beta} \ \mathcal{L}(\boldsymbol{B}, \beta) + w\mathcal{R}(\boldsymbol{B}), \tag{5}$$

where $w > 0$ is the tuning parameter that adjusts the weights of the regularization term, which is defined as follows:

$$\mathcal{R}(\boldsymbol{B}) = \lambda \left(\sum_{i=1}^{s} \sqrt{t} \|\boldsymbol{b}_{i,:}\|_2 + \sum_{j=1}^{t} \sqrt{s} \|\boldsymbol{b}_{:,j}\|_2 \right) + (1-\lambda) \sum_{i=1}^{s} \sum_{j=1}^{t} \|b_{i,j}\|_1 \tag{6}$$

where $\lambda \in [0, 1]$ is a value that controls the weights of the group-wise and element-wise regularization terms; $\|\cdot\|_1$ is the ℓ_1 norm that inspires element-wise sparsity; $\|\cdot\|_2$ is the ℓ_2 norm, which inspires group-wise sparisty; s and t are respectively the number of row and column of $\boldsymbol{B}$; $\boldsymbol{b}_{i,:}$ is the ith row of $\boldsymbol{B}$; $\boldsymbol{b}_{:,j}$ is the jth column of $\boldsymbol{B}$, and $b_{i,j}$ is the (i,j)th element of $\boldsymbol{B}$. $\sqrt{s}$ and $\sqrt{t}$ account for the varying sizes of the penalty term.

The tuning parameters w and λ can be determined using a model selection criterion such as cross-validation (CV), Akaike Information Criterion (AIC), or Bayesian Information Criterion (BIC), etc. The optimization criterion (5) is convex, and thus it can be easily solved using existing optimization packages such as the CVX [3].

3. Simulation Study

In this section, we validate the performance of our proposed method using simulated datasets with different correlation-structures, multiple sample sizes, and different noise levels.

3.1. Data Generation

We first generate the process variable matrix $\boldsymbol{X}_i \in \mathbb{R}^{s\times t}$ by considering two types of entry correlation structures: (i) no correlation (i.e., independent and identically distributed, IID) and (ii) row-wise correlation. For the first correlation structure, all the entries of $\boldsymbol{X}_i$ are generated from an IID standard normal distribution. For the second correlation structure, we set the correlation between the jth row and kth row of matrix $\boldsymbol{X}_i$ as $0.5^{|i-j|}$ to mimic the correlation among process variables, where $i = 1, \ldots, s, j = 1, \ldots, s$, and $s = 10, t = 10$.

Next, we generate the regression coefficient matrix $\boldsymbol{B} \in \mathbb{R}^{10\times 10}$, each entry of which follows a uniform distribution $\mathcal{U}[-1, 1]$. To mimic the group-wise (i.e., row-wise and column-wise) sparsity, we let the 1st, 3rd, 5th, 7th, 9th rows and the 2nd, 4th, 6th, 8th and 10th columns of $\boldsymbol{B}$ be zeros. After setting the zero rows and columns, there are 25 nonzero elements left in the 100 entries of $\boldsymbol{B}$. To further mimic the element-wise sparsity, we randomly choose 12 of the 25 entries and set them to be zeros.

Third, we generate the product quality index y_i's using the following logistic regression model

$$\log(\tfrac{y_i}{1-y_i}) = \beta + \langle \boldsymbol{B}, \boldsymbol{X}_i + \boldsymbol{E}_i \rangle,$$

where $\beta = 0$; $\boldsymbol{E}_i$ is an IID noise/disturbance matrix from normal distribution $\mathcal{N}(0, \sigma^2)$. In this study, we will test three levels of noise: No Noise ($\sigma = 0$), Low Noise ($\sigma = 0.05$), and High Noise ($\sigma = 0.1$).

Finally, the generated data $\{\boldsymbol{X}_i, y_i\}_{i=1}^n$ are used to validate the performance of our model. We will evaluate the performance of our proposed method under two sample sizes: $n = 100$ and $n = 200$. The turning parameter w and weight λ are selected using 10-fold cross-validation.

3.2. Benchmark and Evaluation Criteria

The performance of our model will be compared with the 2D variable selection method developed by [7], which is known to be a state-of-the-art

Table 1: Average selection accuracy and precision for $\boldsymbol{X}_i$ with IID entries (%)

Noise	Method	Group/Element	Sample Size	TP	TN	FP	FN	Accuracy (SD)
No Noise	Proposed Method	Row+Column	100	99.60	90.80	9.20	0.40	95.20(5.31)
			200	100.00	98.00	2.00	0.00	99.00(2.13)
		Element	100	77.46	95.92	4.08	22.54	86.32(4.59)
			200	94.31	98.50	0.50	5.69	96.42(3.10)
	Benchmark	Row+Column	100	90.30	87.40	12.60	9.70	88.85(6.28)
			200	92.40	88.90	11.10	7.60	90.65(6.17)
		Element	100	81.40	28.60	71.40	18.60	55.00(3.02)
			200	82.70	29.48	70.52	17.30	56.10(2.98)
Low Noise	Proposed Method	Row+Column	100	99.80	88.50	11.50	0.20	94.15(5.46)
			200	100.00	97.50	2.50	0.00	98.75(2.29)
		Element	100	81.46	91.50	8.50	18.54	86.28(5.09)
			200	92.08	97.25	2.75	7.92	94.56(3.62)
	Benchmark	Row+Column	100	88.10	85.80	14.20	11.90	86.95(6.43)
			200	88.20	88.20	11.80	11.80	88.20(6.72)
		Element	100	80.54	28.00	72.00	19.46	54.27(3.07)
			200	81.54	28.08	71.92	18.46	54.81(3.18)
High Noise	Proposed Method	Row+Column	100	99.60	86.00	14.00	0.40	92.80(6.25)
			200	100.00	96.80	3.20	0.00	98.40(3.01)
		Element	100	81.38	91.08	8.92	18.62	86.04(5.07)
			200	92.08	96.42	3.58	7.92	94.16(3.57)
	Benchmark	Row+Column	100	87.30	85.30	14.70	12.70	86.30(6.22)
			200	87.70	86.40	13.60	12.30	87.05(6.67)
		Element	100	79.46	28.33	71.67	20.54	53.90(2.89)
			200	80.92	27.75	22.25	19.08	54.34(2.51)

method. As mentioned earlier, it has two limitations compared with the method proposed in this paper. First, it is not convex, which implies its optimization criterion may converge to a local optimum, and thus its variable selection accuracy might be compromised. Second, it provides a solution with only group-wise sparsity but no element-wise sparsity.

The performance of the proposed method and the benchmark is evaluated using two criteria: group-wise (i.e., row-wise and column-wise) selection accuracy and element-wise selection accuracy. The group-wise selection accuracy is defined as follows:

$$\text{Group-Wise Accuracy} = \frac{\text{TP} + \text{TN}}{\text{No. of rows} + \text{No. of columns}}, \tag{7}$$

where "TP" represents "True Positive", which is the number of crucial rows

and columns that are selected correctly. "TN" represents "True Negative", which is the number of non-crucial rows and columns that are removed correctly. "No. of rows" and "No. of columns" are the number of rows and columns in the process variable matrix, respectively.

The element-wise selection accuracy is computed using the equation below:

$$\text{Element-Wise Accuracy} = \frac{\text{TP} + \text{TN}}{\text{No. of elements}}, \tag{8}$$

where "TP" is the number of crucial entries that are selected correctly. "TN" is the number of non-crucial entries that are removed correctly. The denominator "No. of elements" is the total number of entries in the process variable matrix.

We repeat the evaluation process 100 times. The mean selection accuracy and its standard deviations for the $\boldsymbol{X}_i$'s generated from the first correlation structure (i.e., IID standard normal distribution) are reported in Table 1. Table 2 reports the mean selection accuracy and its standard deviations for the $\boldsymbol{X}_i$'s generated from the second correlation structure (i.e., row-wise correlation). In addition to the "TP" and "TN", we also report the "FP" (i.e., the number of non-crucial rows/columns/elements that are falsely identified as crucial) and "FN" (i.e., the number of crucial rows/columns/elements that are falsely selected as non-cricial) in the two tables.

3.3. Results and Analysis

Table 1 indicates that, when there is no correlation among the entries of the process matrix $\boldsymbol{X}_i$, our proposed convex 2D variable selection method outperforms the benchmark in terms of (group-wise and element-wise) selection accuracy and precision under all different noise levels and sample sizes. For example, when there is no noise added to the quality index, and the sample size is 100, the group-wise selection accuracy (and the standard deviation, SD) of our proposed method is 95.20(5.31), while it is 88.85(6.28) for the benchmark; the element-wise selection accuracy (and SD) of our proposed method is 86.32.00(4.59), while it is 55.00(3.02) for the benchmark. When the noise level is low, and the sample size is 200, the group-wise selection accuracy (and SD) of our proposed method and the baseline model are respectively 98.75(2.29) and 88.20(6.72); the element-wise selection accuracy of the two methods are 94.56(3.62) and 54.81(3.18), respectively. When the noise level is high, and the sample size is 200, the selection accuracy (and SD)

of the proposed convex 2D variable selection method and the benchmark are 98.40(3.01 and 86.30(6.22), respectively; the element-wise selection accuracy (and SD) of the two methods are respectively 94.16(3.57) and 54.34(2.51). We believe this is due to two reasons. First, our proposed method is convex while the benchmark is nonconvex, which implies that our proposed method converges to a global optimum while the baseline model might converge to a local optimum. Second and more importantly, our proposed method is able to achieve both group-wise and element-wise selection, while the benchmark can only inspire group sparsity. Since the true response variables (i.e., the quality indices) are generated using a regression coefficient matrix with both group-wise and element-wise sparsity, the group-wise selection accuracy (and precision) of the benchmark is compromised. This is because the benchmark tends to choose a row/column in which all the elements are nonzero. However, this is usually not the case in the generated data, in which even if a row/column is crucial, some of its entries are set to be zeros. In addition, since the benchmark cannot inspire element-wise sparsity, its element-wise selection accuracy is consistently around 50%, which is close to the accuracy of a random guess.

Table 1 also illustrates that the selection accuracy and precision of both the proposed method and the benchmark decrease with the increase of noise. For example, when the sample size is 100, the group-wise selection accuracy (and SD) of our proposed method are 95.20(5.31), 94.15(5.46), and 92.80(6.25), when there is no noise, low noise, and high noise, respectively; the group-wise selection accuracy (and SD) of the benchmark model are respectively 88.85(6.28), 86.95(6.43), and 86.30(6.22). This is reasonable since higher noise added to the response variable will have more compromise on the accuracy of the parameter estimates, and thus results in a lower selection accuracy. In addition, we can observe from Table 1 that the selection accuracy and precision increase with the increase of sample size, which is also reasonable.

The accuracy and precision data in Table 2 imply that all the conclusions drawn from Table 1 still hold: The proposed convex selection method always outperforms the baseline model in terms of both (group-wise and element-wise) selection accuracy and precision; data with smaller noise result in higher selection accuracy and precision; a smaller sample size will yield worse selection accuracy and precision. In addition, comparing Tables 1 with 2, we notice that both the proposed method and the baseline model have worse selection accuracy and precision when there is a correlation among the rows

of the process variable matrices. For example, when the noise level is high, and the sample size is 200, the group-wise selection accuracy (and SD) of the proposed method is 98.40(3.01) in Table 1, while it is 85.40(8.12) in Table 2. Similarly, the element-wise selection accuracy (and SD) of the proposed method is 94.16(3.57) in Table 1, while it is 82.64(3.91 in Table 2. Similar performance degradation can also be observed on the baseline model in Table 2. This is not something unexpected since it is known that the correlation among the rows/columns of the process variable matrices will compromise the selection accuracy and precision. The correlation between two rows represents that the two process variables contain similar information. If the correlation is high, then the regression model only needs one of them, and the other one is not needed due to the existence of the first one since the information provided by the two variables is overlapped.

4. Case Study

In this section, we validate our proposed method using data from a real-world application. The dataset is from the hot strip mill shown in Figure 1. It consists of the process variables and quality indices of 490 strip steel products from the same rolling mill stand, including 264 good quality products and 226 bad quality products. The product quality is defined based on the percentage of width that is smaller than a predefined width threshold. Specifically, as shown in Figure 7(a), the width of each product was measured at 1500 locations uniformly distributed along with the head of the strip. Any measurement point whose width is smaller than the predefined width threshold is considered as a defective point. In other words, as illustrated in Figure 7(b), the measurement points with a negative width error are considered to be defective. We divide the number of defective points by the total number of measurement points $(1,500)$ to yield the quality index, whose range is $[0,1]$.

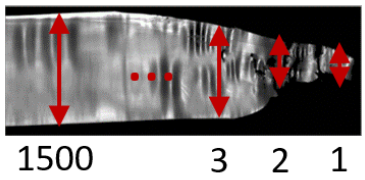
Following the suggestion of the engineers who work in the field, we focus on 9 process variables: *target speed of rollers*, *the measured speed of rollers*, *looper value*, *target force on both sides of the rollers*, *the measured force on the work side of rollers*, *measured force on the transfer side of rollers*, *roller gap*, *looper height*, and *temperature*. At each stage, the data for each process variable is a time series acquired during the production of the product. The time series consists of $1,500$ observations, which are corresponding to the $1,500$ measurement points in Figure 7(a). For each process variable, we

Table 2: Average selection accuracy and precision for $\boldsymbol{X}_i$ with row-wise correlation (%)

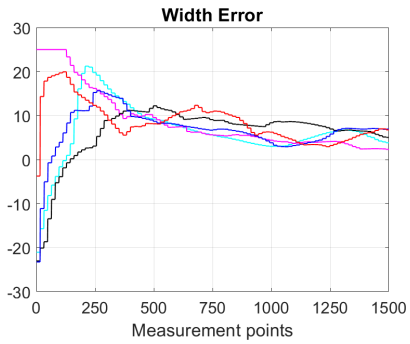
Noise	Method	Group/Element	Sample Size	TP	TN	FP	FN	Accuracy (SD)
No Noise	Proposed Method	Row+Column	100	96.80	76.90	23.10	0.00	86.35(8.34)
			200	98.90	92.80	7.20	1.10	95.85(4.77)
		Element	100	71.31	89.58	10.42	28.69	80.08(4.51)
			200	72.62	93.67	6.33	27.38	82.72(3.89)
	Benchmark	Row+Column	100	97.20	67.70	32.30	2.80	82.45(8.39)
			200	93.40	81.40	18.60	6.60	87.40(6.34)
		Element	100	95.62	6.92	93.08	4.38	53.04(1.94)
			200	89.77	16.92	83.08	10.23	54.80(2.64)
Low Noise	Proposed Method	Row+Column	100	97.10	74.10	25.90	2.90	85.60(8.08)
			200	99.00	92.10	7.90	1.00	95.55(5.27)
		Element	100	71.46	89.21	10.79	28.54	80.06(4.86)
			200	72.23	94.00	6.00	27.77	82.68(3.82)
	Benchmark	Row+Column	100	96.70	66.80	33.20	3.30	81.75(8.66)
			200	93.20	80.30	19.70	6.80	86.75(6.57)
		Element	100	94.77	8.08	91.92	5.23	53.16(1.99)
			200	89.54	17.17	82.83	10.46	54.80(2.76)
High Noise	Proposed Method	Row+Column	100	96.80	74.00	26.00	3.20	85.40(8.12)
			200	98.70	92.40	7.60	1.30	95.52(4.97)
		Element	100	71.20	88.90	11.10	28.80	80.02(4.88)
			200	72.62	93.50	6.50	27.38	82.64(3.91)
	Benchmark	Row+Column	100	94.00	66.20	33.80	6.00	80.10(6.35)
			200	86.70	75.70	24.30	13.30	81.20(8.12)
		Element	100	90.46	14.67	85.33	9.54	52.57(2.57)
			200	78.85	30.00	70.00	21.15	55.40(3.03)

take the average of the $1,500$ observations from each stage and set it as the value of the process variable at that stage. By doing so, we construct a 9×7 process variable matrix $\boldsymbol{X}_i$ for product i, $i = 1, \ldots, 490$. Same as what has been done by [7], we randomly select 400 samples from the dataset to construct a sub dataset. We then apply our method to the sub dataset to identify crucial rows/columns and crucial elements. We repeat this procedure 100 times and then compute the selection accuracy and precision. Any row/columns/elements whose selection rates are higher than 0.5 are considered as important ones. The selection results are reported in Tables 3, 4, and 5.

Table 3 suggests that *Stage 2* and *Stage 4* are identified as crucial stages that are responsible for the quality defects of the hot strip rolling mill. Table 4



(a) Measurement points for width



(b) Difference between desired and measured width of 5 products.

Figure 7: Measurement points of steel strip products and corresponding width error.

Table 3: Selection rates of the stages (%)

Stage	1	2	3	4	5	6	7
Selection Rate	6	85	23	100	4	0	0

indicates that the three crucial process variables are the *Looper Value*, *Roller Gap*, and *Transfer Force*. The method developed by [7] selected *Stages 3, 4, 6* as crucial stages and *Looper Value*, *Roller Gap*, and *Looper Height* as the important process variables. Our proposed method selects less crucial stages than the model in [7], and the same stage identified as crucial by both the two methods is *Stage 4*. In addition, both the methods proposed in this article and [7] identify *Looper Value* and *Roller Gap* as important process variables that might be responsible for the quality defects of the hot strip mill. Both the two process variables are closely related to the performance control of the looper in the hot strip mill, which is known to be the most challenging control component. The process variable *Looper Value* is used to control the tension of the steel strip between two stages. The process variable *Roller Gap* is used to control the thickness of the steel strip, which also significantly affects the real-time looper performance.

In addition to identifying the crucial stages and process variables, the proposed method can also select the important variables in each of the stages identified to be crucial. The results are presented in Table 5, which indicates

that the only important process variable in *Stage 2* is the *Roller Gap*, and the two important process variables in *Stage 4* are the *Looper Value* and *Transfer Force*. This provides more useful information to guide the control experts to revise the control algorithm. Unlike our model, the method proposed in [7] cannot distinguish the importance of the process variables in each crucial stage.

Table 4: Selection rates of the process variables (%)

Process variable	Selection Rate
Target speed	0
Measured speed	0
Looper value	100
Both side target force	10
Work side force	1
Transfer force	93
Roller gap	85
Looper height	0
Temperature	0

Table 5: Element-wise Selection Rate (%)

Method	2	4
Looper value	0	100
Transfer force	6	93
Roller gap	85	0

5. Conclusions

This paper proposed a generalized matrix regression model that can be used for the quality defect diagnostics of multistage manufacturing processes

with multiple identical stages such as hot strip mills in the steel-making industry and 3D printing processes in additive manufacturing. The proposed model is a new generalized matrix regression model with regularization, which is formulated as a convex optimization criterion that consists of a negative log-likelihood function, two ℓ_2 penalization terms, and an ℓ_1 penalization term. By simultaneously using both ℓ_2 and ℓ_1 penalization terms, the proposed method can inspire both group-wise and element-wise sparsity. This implies that, unlike the existing models, the proposed method can provide two levels of information. First, it provides the overall importance of each stage and process variable, which helps identify a few crucial stages and process variables that need to be investigated from the stage-level or process variable-level when revising the control algorithm. Second, it suggests the importance of the process variables in each crucial stage and the stage importance information of each crucial process variable, which provides element-level information that can be used to guide the control algorithm revision as well.

Numerical studies were conducted to validate the effectiveness of the proposed method. The results demonstrated that our proposed method outperformed the benchmark in terms of selection accuracy and precision. We believe this is because the proposed model can achieve both group-wise and element-wise selection, while the benchmark can only inspire group-wise sparsity. A real-world dataset from a hot strip steel mill was also used to evaluate the performance of the proposed method. The results indicated that the proposed method selected less crucial stages and process variables compared with the baseline model. More importantly, the proposed model can provide information regarding the process variables with top priority in each crucial stage, which can provide more information to the control experts to help them revise the feedback control algorithm to reduce the probability of producing defective products in the future.

References

[1] Adland, R., Jia, H., Lode, T., and Skontorp, J. (2021). The value of meteorological data in marine risk assessment. *Reliability Engineering & System Safety*, 209:107480.

[2] Balmashnova, E., Bruurmijn, M., Dissanayake, R., Duits, R., Kampmeijer, L., and van Noorden, T. (2012). Image recognition of shape defects

in hot steel rolling. *Proceedings of the 84th European Study Group Mathematics with Industry (SWI 2012)*, pages 22–38.

[3] Boyd, S., Boyd, S. P., and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

[4] Fang, X., Paynabar, K., and Gebraeel, N. (2017). Multistream sensor fusion-based prognostics model for systems with single failure modes. *Reliability Engineering & System Safety*, 159:322–331.

[5] Gibson, I., Rosen, D. W., Stucker, B., and Khorasani, M. (2021). *Additive manufacturing technologies*, volume 17. Springer.

[6] Hastie, T., Tibshirani, R., and Wainwright, M. (2019). *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC.

[7] Jeong, C. and Fang, X. (2021). Two-dimensional variable selection and its applications in the diagnostics of product quality defects. *IISE Transactions*, pages 1–11.

[8] McCullagh, P. and Nelder, J. (1989). Generalized linear models ii.

[9] Peng, K., Zhang, K., and Li, G. (2013). Quality-related process monitoring based on total kernel pls model and its industrial application. *Mathematical Problems in Engineering*, 2013.

[10] Rifaai, T. M., Abokifa, A. A., and Sela, L. (2022). Integrated approach for pipe failure prediction and condition scoring in water infrastructure systems. *Reliability Engineering & System Safety*, 220:108271.

[11] Schöbi, R. and Sudret, B. (2019). Global sensitivity analysis in the context of imprecise probabilities (p-boxes) using sparse polynomial chaos expansions. *Reliability Engineering & System Safety*, 187:129–141.

[12] Shen, L., Tang, Y., and Tang, L. C. (2021). Understanding key factors affecting power systems resilience. *Reliability Engineering & System Safety*, 212:107621.

[13] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

[14] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

[15] Zhao, J. and Leng, C. (2014). Structured lasso for regression with matrix covariates. *Statistica Sinica*, pages 799–814.

[16] Zhao, J., Niu, L., and Zhan, S. (2017). Trace regression model with simultaneously low rank and row (column) sparse parameter. *Computational Statistics & Data Analysis*, 116:1–18.