Elliptic Confidence on Correspondence Analysis Stainless Steels Corrosion Problem

Yuli Sri Afrianti*, Udjianna S. Pasaribu, Husaini Ardy, and Karunia Eka Lestari, *Member*, *IAENG*

***Abstract*—Stainless steel is the primary material of the lower-side superheater (LSSH) convection pipe support for petroleum or natural gas. Stainless steel corrosion is a severe challenge in equipment maintenance and operation. Corrosion may cause structural defects, leaks, and even failures in superheater convection pipes that could disrupt production and lead to reduced product quality. The Sigma phase occurrence is an indicator of the existence of stainless-steel corrosion. This corrosion occurs due to the exposure of steel to extremely high temperatures over a long period. The experiment was conducted on ASTM A297 after 24 years of operation. The testing sample was acquired from an Indonesian petrochemical company as part of the LSSH convection pipe support. Specimens of the material were heated at a high temperature over 500 °C. The corrosion process was observed based on the corrosion location and color. To analyze and understand this corrosion problem, correspondence analysis was performed. One statistical technique for examining the relationship between two sets of categorical variables is correspondence analysis. In the context of stainless-steel corrosion, the first variable is corrosion location categorized into three locations: A, B, and C, while the second variable is corrosion color divided into light brown (LBr), dark brown(DBr), bluish blue(BBr), light blue(LBl) and dark blue(DBl). Furthermore, elliptic confidence regions were used to estimate the extent to which points in a two-dimensional space represent different corrosion levels in line with the observed variables. The location and corrosion color of stainless steel are found to be statistically significantly correlated in the results; location A tends to be blue, location B tends to be brown, and location C is uncertain because it depends on whether it is light blue or dark brown. It suggests distribution patterns and identifies areas that are more susceptible to corrosion. These findings assist engineers in making decisions regarding the preservation and maintenance of stainless steel to reduce the corrosion rate.**

***Keywords*—correspondence analysis, elliptic confidence, sigma phase, stainless-steel corrosion**

I. INTRODUCTION

STAINLESS Steel, abbreviated as SS, widely used in a variety of industries, including manufacturing, construction, transportation, medical equipment, and food processing. SS is also the primary material of the lower-side superheater (LSSH) convection pipe support for petroleum or natural gas. SS corrosion is a severe problem that compromises the sustainability of the infrastructure and industries that rely on it. In the LSSH pipe support case, SS corrosion may cause structural defects, decreased product quality, and even pipe leaks that potentially pollute the environment or destroy natural ecosystems. Therefore, the feasibility analysis of SS as the primary material for the pipe support must be continuously researched and analyzed, especially regarding its corrosion resistance.The homogeneously distributed elements in the SS material include iron, chromium, nickel, manganese, silicon, and carbon. On the surface of stainless steel (SS), chromium (Cr) occurs naturally as a self-healing protective oxide layer that prevents corrosion [1]. When heated for an extended period of time, Cr alloy forms the Fe-Cr intermetallic phase in SS. The appearance of intermetallic precipitates reduces the material's ability to withstand corrosion. In ferritic and austenitic stainless steel, Fe-Cr compounds (σ-phase) form between 560 and 980°C [2]. The amount of ferrite increases whenever the heat treatment temperature increases, which may transform into intermetallic, such as σ-phase [3]. Beyond reducing the corrosion resistance of SS, this phase can decrease its mechanical properties [4]-[5]. A limited amount of research has looked at the sigma phase's morphology and how it affects material properties [6], as shown in [7]–[8]. The effect of sigma phase precipitation kinetics on these properties has been the main focus of previous research.

Conversely, the sigma phase morphology was described by Fonseca et al. [9]. The phase transformation gives rise to a color that characterizes the microstructural the sigma phase morphology. Beginning with the appearance of LBr, DBr, BBr, and LBl to DBl colors, this phase develops gradually over time. Initially, the material is filled with brown delta ferrite; however, after being exposed to high temperatures for an extended period of time, blue σ-phase forms. This phase emerges from the edge, moves slowly toward the center, and finally turns from brown (δ-ferrite) to blue. The transformation of δ-ferrite to σ-phase was described in [10]. The sigma σ -phase is associated with blue or dark blue. Distinct colors are associated with the corrosion risk level at the observed location. It implies that even though SS has high corrosion resistance, it often still occurs in practice.

As a result, it's critical to comprehend the variables that affect corrosion's likelihood. The investigation of the location of corrosion provides a better understanding of the factors affecting corrosion. It allows for identifying locations most susceptible to corrosion, for instance, center, surface, or edge locations. Additionally, studying corrosion color is valuable as it can provide insight into the corrosion level that is being witnessed. LBr, DBr, BBr, and LBl to DBl are the possible corrosion colors; these colors represent the corrosion levels

Manuscript received September 6, 2023; revised May 25, 2024.

*Y. S. Afrianti is a doctoral student at the Mathematics and Natural Sciences Faculty, Bandung Institute of Technology, Bandung 40132, Indonesia. (corresponding author to provide email: yuli.afrianti@itb.ac.id)

U. S. Pasaribu is a professor in the Statistics Research Division, Mathematics and Natural Sciences Faculty, Bandung Institute of Technology, Bandung 40132, Indonesia. (email: udjianna@itb.ac.id)

H. Ardy is a professor in the Mechanical Engineering at the Faculty of Mechanical and Aerospace Engineering Bandung Institute of Technology, Bandung 40132, Indonesia. (email: husaini@itb.ac.id)

K. E. Lestari is a Research Scholar at the Mathematics Education Department, Universitas Singaperbangsa Karawang, Indonesia. (email: karunia@fkip.unsika.ac.id)

from lowest to highest. For this reason, this research focuses on exploring the pattern or behavior of SS corrosion based on the location and color.

In the context of SS corrosion problems, correspondence analysis (CA) allows us to understand the association between corrosion location and color since both are categorical variables. This method visually examines the associations between two categorical variables by depicting their correspondence plots [11]. When discussing the underlying theory and applications of CA, the common emphasis is quantifying the associations between variables and then portraying them visually. It implies that a lot of researchers are interested in finding out the principal coordinates. Unfortunately, inferential aspects involving confidence regions have received less attention from researchers [12]. However, plenty of mathematical aspects can be explored from this problem to develop science and technology in various findings, such as theoretical results, formulations, proof of concept, technological applications, or new policies that may be employed for science development. One of the applications includes SS corrosion problems. The elliptical confidence region explains the inferential aspects that ensure the accuracy of the position of each category; thus, it minimizes subjectivity when calculating the number of each color at a location, which depends on determining the grid radius.

Generally, CA guides to seek whether there is a pattern of association between corrosion location and color in SS. Moreover, elliptic confidence areas examine how significantly each category contributes to the association. Thus, this study's two main goals are to: (1) develop the mathematical foundation for elliptical confidence regions in correspondence analysis; and (2) investigate the applicability of the mathematical results to a case study of SS corrosion data from Petrochemical's LSSH pipe supports in Indonesia. The results of this study provide new insight to engineers for further improvement and corrosion prevention. Hence, it could support the effective and efficient utilization of LSSH materials in Petrochemical companies, especially in Indonesia.

II. PRELIMINARY THEORY

This section introduces the basic concepts and notation of correspondence analysis (CA) or, rather specifically, simple correspondence analysis (SCA). SCA is a powerful statistical tool for the contingency tables analysis [12]. SCA, also known as reducing dimensions methods, naturally represents the low-dimensional plot's association structure between two categorical random variables [13]. Based on the dependence between the row, column, or both categories, the output generates a correspondence plot that offers details on the interaction between two categorical variables [14].

A. Basic Notation on Correspondence Analysis.

The notes that follow are condensed from a variety of literary sources [12]–[13], [15]–[16]. X and Y are two categorical random variables. Let's say that X and Y are made up of the S and T categories, respectively. The X and Y are used to select and categorize a random sample of n people or objects. These cross-classification variables are shown in an $S\ x\ T$ contingency table. In notation of matrix, represented as a cross-tabulation matrix or data matrix, $N = (n_{st})$. The elements of the matrix are denoted as n_{st} such that $n = \sum_{s=1}^{S}\sum_{t=1}^{T} n_{st}$. Rows and columns marginal frequencies, respectively, are presented in diagonal matrices D_S and D_T, which are both non-singular, such that $D_S = \text{diag(r)}$ and $D_T = \text{diag(c)}$, where $r = (n_1, n_2, \dots, n_S)^{\perp}$ dan $c = (n_1, n_2, \dots, n_T)^{\perp}$are referred to row vectors and marginal column frequencies, respectively.

B. Association Measure

In the SCA literature, two categorical variables are measured for association using Pearson's chi-squared statistic. It is calculated using the expected number of people or things that will fall into the ith row and jth column categories, such that $E[n_{st} = (n_{s.}n_{t.})n^{-1}]$. The independence hypothesis is stated as follows in order to determine whether two categorical variables are related:

$$H_0\colon n_{st} = (n_{s.}n_{t.})n^{-1},\ \ vs.\ \ H_1\colon n_{st} \neq (n_{s.}n_{t.})n^{-1}, \quad (1)$$

for all $s = 1,2, \dots, S$ and $t = 1,2, \dots, T$.

The statistical value χ^2_{stat}under the assumption of H_0 is true calculated as:

$$\begin{aligned}
\chi^2 &= \sum_{s=1}^{S}\sum_{t=1}^{T}((n_{st} - E[n_{st}])^2)(E[n_{st}])^{-1} \\
&= \sum_{s=1}^{S}\sum_{t=1}^{T}\left(\left(n_{st} - (n_{s.}n_{t.})(n^{-1})\right)^2\right)\left((n_{s.}n_{t.})(n^{-1})\right)^{-1} \\
&= \sum_{s=1}^{S}\sum_{t=1}^{T}\Big(n^{-2}\big(n_{st} \\
&\qquad - n_{s.}(n_{t.})(n^{-1})\big)^2\Big)\left(\frac{1}{n}\left(n_{s.}(n_{t.})(n^{-1})\right)\right)^{-1} \\
&= n^{-1}\sum_{s=1}^{S}\sum_{t=1}^{T}\Big(\big(n_{st} \\
&\quad - n_{s\cdot}\,(n_{t.})(n^{-1})\big)^2\Big)\Big(\big(n_{s\cdot}\,(n_{t.})(n^{-1})\big)\Big)^{-1} \\
&= n^{-1}\,\chi^2_{stat},
\end{aligned}$$

since

$$\begin{aligned}
\chi^2_{stat} &= \textstyle\sum_{s=1}^{S}\sum_{t=1}^{T}\Big(\big(n_{st} - \\
&n_{s\cdot}\,(n_{t.})(n^{-1})\big)^2\Big)\Big(\big(n_{s\cdot}\,(n_{t.})(n^{-1})\big)\Big)^{-1}, \\
&\text{or } \chi^2_{stat} = n\chi^2
\end{aligned} \quad (2)$$

The chi-squared distribution of this statistic has a degree of freedom of $v = (S-1)(T-1)$. [17] used a standard residual matrix to represent the relationship between row and column in the contingency table:

$$(s_{st}), \text{ where } s_{st} = \left(n_{st} - (n_{s.}n_{t.})(n^{-1})\right)(n_{s\cdot}n_{\cdot t})^{-\frac{1}{2}}. \quad (3)$$

The element s_{st} represents the dependence of $(n^{-1})(n_{s\cdot}\,n_{t\cdot}) \neq n_{st}$, in matrix operation, it is written as

$$\mathbf{S} = \mathbf{D}_S^{-1/2}(\mathbf{N} - n^{-1}\mathbf{r}\mathbf{c}^{\mathrm{T}})\mathbf{D}_T^{-1/2} \quad (4)$$

C. Coordinate Systems and Confidence Regions

The associations between the row and the column categories are visualized by considering the set of singular vectors, $\{\mathbf{a}_{im}\}$ and $\{\mathbf{b}_{im}\}$ for $s = 1, 2, \cdots, S$ and $t = 1, 2, \cdots, T$. Singular value decomposition (SVD) of the residual standard matrix yielded these vectors, which were then called standard coordinates because they were shown as coordinates for the s th row and tth column categories on the m-dimensional plot. The row and column standard coordinate matrix **A** and **B**, is established by

$$\mathbf{A} = \mathbf{U}\,\mathbf{D}_{\sqrt{\lambda}} \text{ and } \mathbf{B} = \mathbf{V}\,\mathbf{D}_{\sqrt{\lambda}}, \tag{5}$$

where $\mathbf{D}_{\sqrt{\lambda}}$, **A**, and **B** are the S's diagonal matrix singular value, which is the matrix in which each column represents the left and right singular vectors of S, respectively, and is derived from the SVD of S [12], [18]. An asymmetric plot results from the standard coordinate. In an asymmetric plot, the separation between a row- and a column-coordinate indicates their association. The row is positioned exactly at the barycenter of the column [19].

In certain cases, the position (profile) of the row- and column-categories must be taken into account in order to assess the strength of an existing association. Principal coordinates for the row and column, F and G, reflect it and are ascertained by:

$$\mathbf{F} = \sqrt{n}\,\mathbf{D}_S^{-\mathbf{1/2}}\mathbf{A} \text{ and } \mathbf{G} = \sqrt{n}\,\mathbf{D}_T^{-\mathbf{1/2}}\mathbf{B}. \tag{6}$$

The resulting plot of the principal coordinate is called a symmetric plot. In this plot, the row-to-row and column-to-column distances are approximate χ^2 −distance between the respective profiles [13]. Thus, categories whose frequency is rarely plotted far from the origin, and vice versa [15]-[16]. The correlation between the variables is significantly influenced by the principal coordinates that are far from the origin.[12].

On the other hand, the interpretation of both symmetric and asymmetric plots depends on how close a coordinate point is to the positions of other coordinates and how it originated. Equation (2) indicates that, if H_0 is correct, there is no difference between the estimated and observed frequencies. Coordinate point configuration positions for each category are shown on the contingency table, which is centered at the origin. Coordinate points close to the origin, therefore, suggest that the category that the coordinate point is meant to represent is not included in the association structure between the variables [12].

To be regarded as statistically significant when determining the structure of associations between variables, a coordinate's distance from the origin must be measured. By building the coordinate point's confidence regions, its measure can be found [20]. At the beginning of the development of SCA, [21] constructed the $100(1-\alpha)\%$ circle regions of confidence for each category of variables in a configuration of coordinate points on a plot. Unfortunately, the circle region is too weak in representing the association structure between variables since the eigenvalues as scaling factors of the correspondence plot have a descending order [18]. It implies that an eigenvalue larger than the principal inertia of the second principal axis always represents the principal inertia of the first principal axis. These regions do not account for the fact that the axes of the correspondence plot have different weights, nor do they consider any information covered in higher dimensions [12].

For the ith row category and jth column category in a two-dimensional correspondence plot, Beh [11] defined a semi-major length of 100 (1-α)% in order to address the problems.

$$x_i = \sqrt{\lambda_1 \frac{n\,\chi_\alpha^2}{n_{s\cdot}\,\chi_{stat}^2}} \text{ and } x_j = \sqrt{\lambda_1 \frac{n\,\chi_\alpha^2}{n_{\cdot t}\,\chi_{stat}^2}}, \tag{7}$$

and a semi-minor length of

$$y_i = \sqrt{\lambda_2 \frac{n\,\chi_\alpha^2}{n_{s\cdot}\,\chi_{stat}^2}} \text{ and } y_j = \sqrt{\lambda_2 \frac{n\,\chi_\alpha^2}{n_{\cdot t}\,\chi_{stat}^2}}. \tag{8}$$

where λ_1 and λ_2 are the first two eigenvalues of $\mathbf{S}\mathbf{S}^T$ or $\mathbf{S}^T\mathbf{S}$. Here, with two degrees of freedom, the chi-squared distribution's (1-α) percentile is represented by χ_α^2, as two-dimensional displays are typically the preferred ones.

Ginanjar et al. [15] proposed a simplification of CA (SoCA) to estimate the principal coordinates for $3 \times T$ contingency table, where $T = 3, 4, \cdots$. These coordinates are more precise than those from Equation (6) since those are calculated directly from the data matrix. The findings of [15] and [11] in Equations (7) and (8) motivated us to determine a formula for constructing the confidence regions analytically.

III. MATHEMATICAL RESULTS

A. Elliptical Confidence Regions Construction

Let **S** be an $S \times T$ standard residual matrix in Equation (3). If $S < T$, then the size of $\mathbf{S}\mathbf{S}^{\perp} < \mathbf{S}^{\perp}\mathbf{S}$, computing eigenvalues of $\mathbf{S}\mathbf{S}^{\perp}$ will be simplified [14], [16]. The element of $\dot{\mathbf{S}} = \mathbf{S}\mathbf{S}^{\perp}$ as $\dot{\mathbf{S}} = (\dot{s}_{su})$, where

$$\dot{s}_{su} = \left((n_{s\cdot}n_{u\cdot})^{-\frac{1}{2}}\right)\left(\sum_{t=1}^{T}(n_{st}n_{ut})(n_{\cdot t})^{-1} - (n_{s\cdot}n_{u\cdot})(n)^{-1}\right), \tag{9}$$

for $s, u = 1, 2, \cdots, S$, described by Ginanjar et al. [22].

Lemma 3.1 states that eigenvalues corresponding to $\dot{\mathbf{S}}$can be obtained directly from the elements of $\dot{\mathbf{S}}$ in accordance with [14] and [16].

Lemma 3.1. Given $\boldsymbol{N}$ is the $3 \times T$ cross-tabulation matrix for $T = 3, 4, \cdots$. Let $\dot{\mathbf{S}} = \mathbf{S}\mathbf{S}^{\perp}$ as in Equation (9). Then, the eigenvalues that correspond to $\dot{\mathbf{S}}$ are:

$$\lambda_1 = \frac{\left[tr\left(\dot{\mathbf{S}}\right) + \sqrt{\left(tr\left(\dot{\mathbf{S}}\right)\right)^2 + 4\left[\left(\sum_{\substack{s=1\\ s\neq t}}^{2}\sum_{t=2}^{3}\dot{s}_{st}{}^2\right) - \left(\sum_{\substack{s=1\\ s\neq t}}^{2}\sum_{t=2}^{3}\dot{s}_{ss}\dot{s}_{tt}\right)\right]}\right]}{2},$$

$$\lambda_2 = tr\left(\dot{\mathbf{S}}\right) - \lambda_1, \text{and } \lambda_3 = 0.$$

Proof. Consider a cross-tabulation matrix **N** of size $3 \times T$. The matrix $\dot{\mathbf{S}}$ has a size of 3×T. As a result of $\dot{\mathbf{S}}$ being a real symmetric matrix, the eigenvalues are non-negative.
Suppose that:

$$b = -\sum_{s=1}^{3}\dot{s}_{ss} = -tr\left(\dot{\mathbf{S}}\right),$$

$$c = \sum_{\substack{s=1\\ s\neq t}}^{2}\sum_{t=2}^{3}\dot{s}_{ss}\dot{s}_{tt} - \sum_{\substack{s=1\\ s\neq t}}^{2}\sum_{t=2}^{3}\dot{s}_{st}^2, \text{ and}$$

$$d = \sum_{\substack{s=1 \\ s\neq t\neq u}}^{3}\sum_{t=1}^{2}\sum_{u=2}^{3} \dot{s}_{ss}\dot{s}_{su}^2 - \dot{s}_{11}\dot{s}_{22}\dot{s}_{33} - 2\dot{s}_{12}\dot{s}_{13}\dot{s}_{23}$$

The cubic characteristic equation must be solved in order to determine the eigenvalues of $\dot{\mathbf{S}}$, $\lambda^3 + b\lambda^2 + c\lambda + d = 0$. Since $\lambda_3 = 0$ is the eigenvalue of $\dot{\mathbf{S}}$ [14], we have a quadratic equation $\lambda^2 + b\lambda + c = 0$. Let $p = \sum_{\substack{s=1 \\ s\neq t}}^{2}\sum_{t=2}^{3} \dot{s}_{ss}\dot{s}_{tt}$ and $q = \sum_{\substack{s=1 \\ s\neq t}}^{2}\sum_{t=2}^{3} \dot{s}_{st}^2$. The roots of the equation are $\lambda_1 = \frac{-b+\sqrt{(-b)^2-4(p-q)}}{2}$ and $\lambda_2 = -b - \lambda_1$. Plugging the values of b, p, and q into the roots completes the proof ■

Lemma 3.1 is a valuable tool for constructing elliptical confidence regions. By applying it to Equations (7) and (8), for the sth row-and tth column-category, we have the length of semi-major as follows:

$$x_s = \left(\left(\frac{\lambda_1 n\, \chi_\alpha^2}{n_{s\cdot}\, \chi_{stat}^2}\right)^{\frac{1}{2}}\right) \text{ and } x_t = \left(\frac{\lambda_1 n\, \chi_\alpha^2}{n_{\cdot t}\, \chi_{stat}^2}\right)^{\frac{1}{2}}. \quad (10)$$

and the length of the semi-minor as follows:

$$y_s = \left(tr\left(\dot{\mathbf{S}}\right) - \lambda_1\right)\sqrt{\frac{n\, \chi_\alpha^2}{n_{s\cdot}\, \chi_{stat}^2}} \text{ and}$$
$$y_t = \left(tr\left(\dot{\mathbf{S}}\right) - \lambda_1\right)\sqrt{\frac{n\, \chi_\alpha^2}{n_{\cdot t}\, \chi_{stat}^2}}. \quad (11)$$

The final pair of equations demonstrates that the first eigenvalue and the components of the $\dot{\mathbf{S}}$ matrix can be used to construct the elliptical confidence regions. This formula helps simplify the computational process since calculating all the eigenvalues is unnecessary.

B. The Eccentricity of Elliptical Regions

The ratio of distances from the center to the semi-major axis is called the ellipse's eccentricity. It is the ovalness of an ellipse, whose value lies in the interval [0,1].
The eccentricity measured the deviation of the curve. From the SCA point of view, It shows how much the first principal inertia weighs in relation to the second principal inertia. Weir [23] determines the elliptical region's eccentricity for the sth row category and tth column category by

$$E_s = \sqrt{1 - \frac{\lambda_2}{\lambda_1}} \text{ and } E_j = \sqrt{1 - \frac{\lambda_2}{\lambda_1}}. \quad (12)$$

By applying Lemma 3.1 to Equation (12), we have a new formula for the elliptical region eccentricity as follows:

$$E_s = \sqrt{\frac{2\lambda_1 - tr(\dot{\mathbf{S}})}{\lambda_1}} \text{ and } E_t = \sqrt{\frac{2\lambda_1 - tr(\dot{\mathbf{S}})}{\lambda_1}}. \quad (13)$$

C. p_{value} Approximation for Confidence Ellipses

The p_{value} is used as the statistic for significance tests in hypothesis testing. It represents the likelihood of getting test results that are at least as extreme as those found if the null hypothesis were to be true. In a SCA analysis, we compute the pvalue of elliptical regions to determine the statistical significance of a row or column category to the association structure between the variables.

The p_{value} reflects the proximity of a row (or column) principal coordinates from the origin in dimensions higher than the Dth explained by [20]. The first D dimensions may visualize the association between variables for practical reasons. It implies we may ignore the higher dimensions, losing no actual information about this structure. By considering [12], the approximation p_{value} of confidence ellipses for the sth row category and tth column category derive to:

$$pvalue_{sD} \approx P\left\{\chi^2 > \chi_{stat}^2 \frac{n_{s\cdot}}{n}\sum_{m=1}^{D}\left(\frac{f_{sm}}{\sqrt{\lambda_m}}\right)^2\right\} \text{ and}$$
$$pvalue_{tD} \approx P\left\{\chi^2 > \chi_{stat}^2 \frac{n_{\cdot t}}{n}\sum_{m=1}^{D}\left(\frac{g_{tm}}{\sqrt{\lambda_m}}\right)^2\right\}. \quad (14)$$

where λ_m is the mth eigenvalue of $\dot{\mathbf{S}}$. Here, f_{sm} and g_{tm} are the elements of matrices $\mathbf{F}$ and $\mathbf{G}$ referred to as the sth row and tth column principal coordinates in the m-dimensional plot. According to equation (14) it is possible to consider the principal inertia values, which show how each axis is relative weighted in a D-dimensional correspondence plot [12].

IV. PRACTICAL RESULTS

A case study on SS data is provided in this section. The variables in this study are corrosion location and color. The location of corrosion is a categorical variable that describes the position where corrosion occurs on SS. There are three locations observed, namely the center (A), surface (B), and edge (C), as illustrated in Fig. 1. These locations were chosen since they represent distinct positions, which may provide insight into the corrosion behavior based on the location factor. There are three locations: A (center), B (near surface), and C (center but near the edge).

Meanwhile, corrosion color is a categorical variable that reflects the corrosion level, graded from LBr, DBr, BBr, LBl to DBl, which is the highest level. The procedure for counting the occurrence of each color in a particular location is accomplished by discretizing the location into a circular area that is determined based on the grid radius. The ASTM A297 material was used in the experiment, which ran for 24 years. This material's low Cr content (about 24%) and 0.38% C indicate that delta ferrite (iv) transformation is the mechanism at work in this experiment. The testing sample was acquired from an Indonesian petrochemical company as part of the LSSH convection pipe support. There are three main steps in the process of obtaining the material image. The specimens were first prepared through mounting, grinding, polishing, and cutting. The microstructure specimens were then subjected to an etching procedure that took place at 25°C for 5 seconds using a modified Murakami reagent. The last step is to image the microstructure of the samples. Fig. 1 depicts the 50% thickness image of the sample, which is the output of all three steps.

The image of every place needs to be divided into grids, and each grid's frequency of each color needs to be computed. In positions A, B, and C, the circular grid was created (refer to Fig. 2). A contingency table displaying the frequency of each color at A, B, and C observation locations is provided, as mentioned in Table 1.

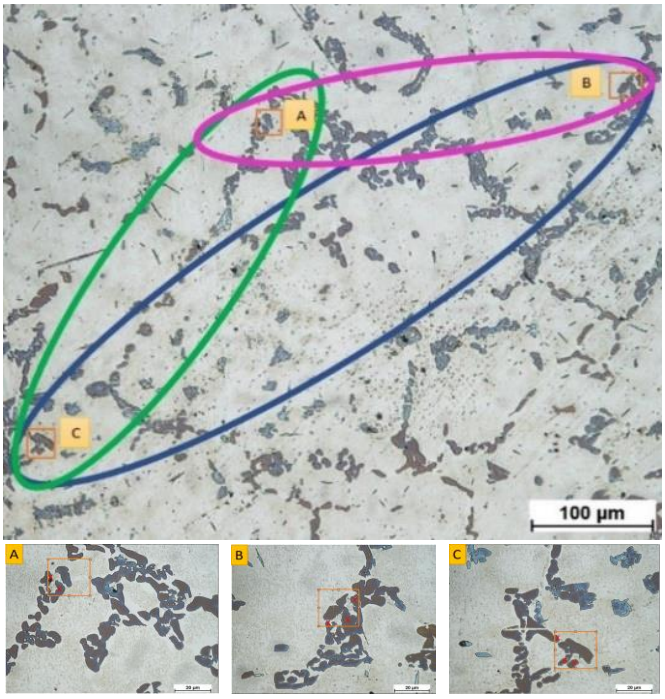
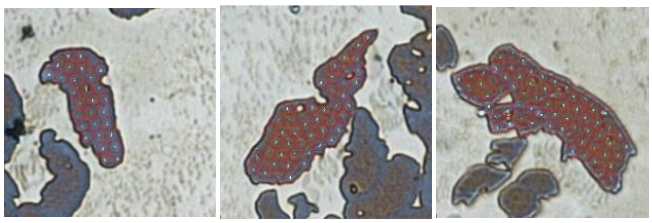


Fig. 1. An image of the specimen at three carefully chosen locations, A, B, and C, at 50% thickness.



Loc A Loc B Loc C

Fig. 2. The outcome of discretizing the circle grid at every location with radius R.

Descriptive statistics can be used to analyze the data in the contingency table and the visualization method as a column vector diagram [24] or SCA. In this case, SCA analysis was performed to verify the independence of the chosen locations. A two-way contingency table's description of categorical data can be visualized and presented in two dimensions using the SCA, a multivariate method [25]–[27]. How the current plots can visualize the association in the low-dimensional plot (dimensional reduction) while absorbing as much information as possible is the main concern in the SCA [14]. The contingency table's rows and columns are represented as coordinate points in low-dimensional space in a correspondence plot [12]. It makes it possible to interpret results in a more direct and useful manner, particularly when relating particular patterns to the characteristic variables [28].

TABLE 1
A TWO-WAY CONTINGENCY TABLE USING LOCATION AND COLOR TO CROSS-CLASSIFIED THE NUMBER OF STAINLESS-STEEL CORROSION

Categorical Variables		Color					
		Light Brown	Dark Brown	Bluish Brown	Light Blue	Dark Blue	Total
Location	A	2	5	10	13	0	30
	B	10	18	13	5	2	48
	C	11	24	22	6	2	65
Total		23	47	45	24	4	143

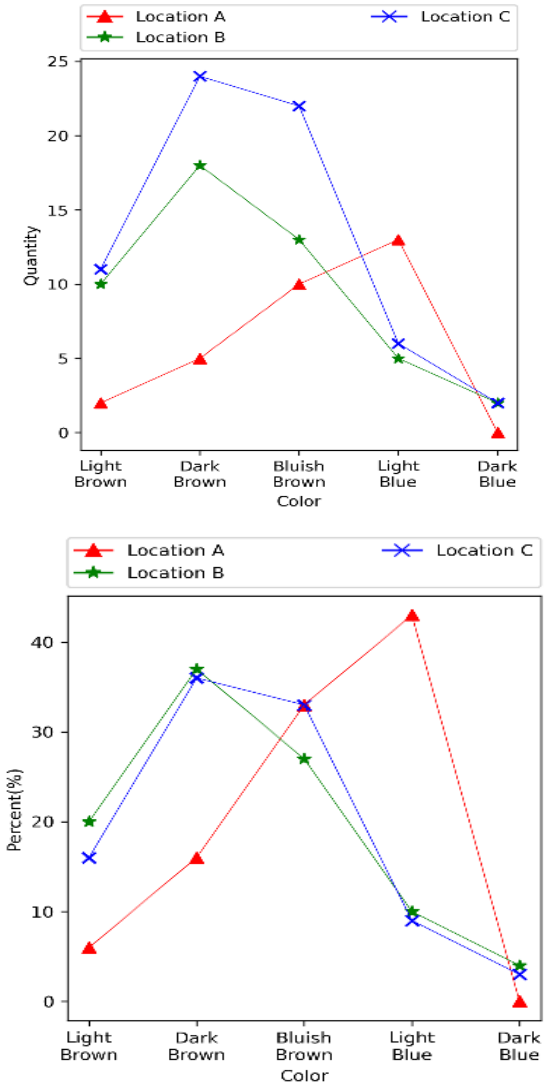


Fig. 3. Plots of frequency of color occurrence (top) and frequency proportion (bottom) for each location.

The contingency table shows the frequency of color appearance at each location. For the analysis process, the components in the contingency table can be presented on a graph as in Fig. 3 with red indicating the color frequency for location A, green for location B, and blue for location C. Presenting a correspondence matrix with the percentage of each color frequency for each location relative to the total is the next step. This correspondence matrix can be seen in Table 2. One of the two essential elements of contingency table analysis is the degree of association, either within or between variables [29]. SCA is used in this study to characterize the independence of the locations and colors.The correspondence plots visualize their association, both within or between variables.

The symmetric and asymmetric correspondence plots are presented in Fig. 4. On a symmetric correspondence plot, the association between the location variable and the color variable that appears on the SS surface can be detected separately, whereas, on an asymmetric correspondence plot, the association between these two variables can be seen simultaneously, so that it can help further the analysis process and provide an initial estimate of the location which ones are subject to certain corrosion conditions.

TABLE 2
CORRESPONDENCE MATRIX OF THE NUMBER OF STAINLESS STEEL CORROSION BASED ON LOCATION AND COLOR

Categorical Variables		Color					Total
		Light Brown	Dark Brown	Bluish Brown	Light Blue	Dark Blue	
Location	A	$\frac{2}{143}$	$\frac{5}{143}$	$\frac{10}{143}$	$\frac{13}{143}$	0	$\frac{30}{143}$
	B	$\frac{10}{143}$	$\frac{18}{143}$	$\frac{13}{143}$	$\frac{5}{143}$	$\frac{2}{143}$	$\frac{48}{143}$
	C	$\frac{11}{143}$	$\frac{24}{143}$	$\frac{22}{143}$	$\frac{6}{143}$	$\frac{2}{143}$	$\frac{65}{143}$
Total		$\frac{23}{143}$	$\frac{47}{143}$	$\frac{45}{143}$	$\frac{24}{143}$	$\frac{4}{143}$	$\frac{143}{143} = 1$

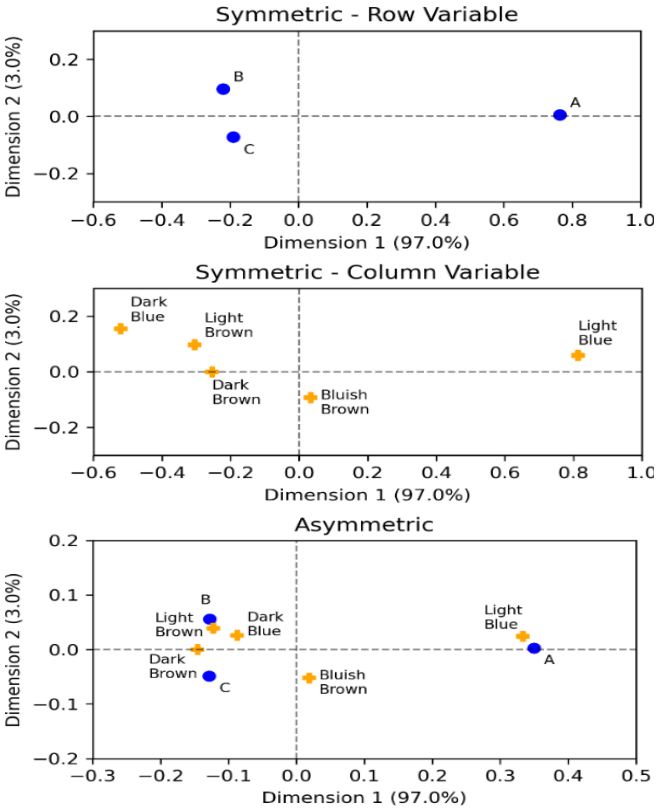


Fig. 4. The symmetric and asymmetric correspondence plots for SS corrosion data in Table 1.

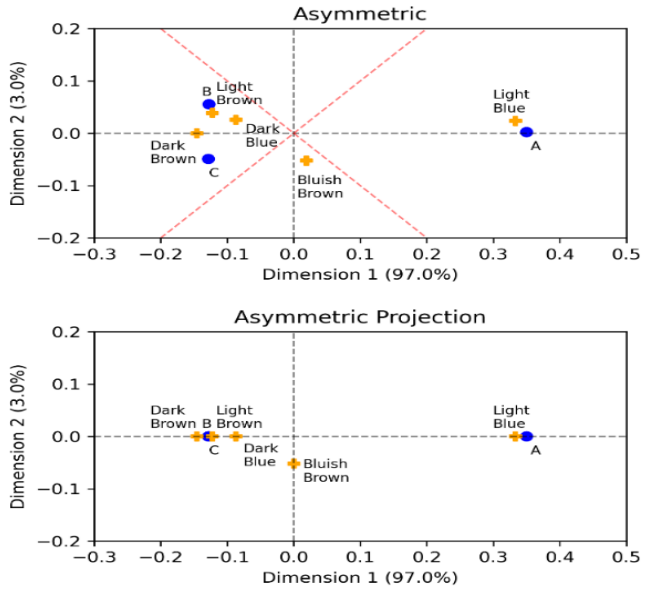


Fig. 5. The projection of asymmetric correspondence plots for SS corrosion data in Table 1.

For the analysis process, a projection is carried out for each coordinate on an asymmetric correspondence plot to the Cartesian axis. The projection results can be seen in Fig. 5. This projection shows that the pairs that are not mutually exclusive are location A with Light Blue, location B with Light Brown, and location C with Dark Brown. By applying Lemma 3.1 to Equations (10), (11), and (14), we get an elliptical confidence region for each row-and column-category, as summarized in Table 3. The confidence region and pvalue can be used to determine how much a category contributes to the structure of the relationship between two categorical variables [30]. Fig. 6 displays the elliptical confidence region for each row category and column category based on the data in Table 3.

TABLE 3
SUMMARY STATISTIC OF ELLIPTICAL CONFIDENCE REGIONS FOR STAINLESS STEEL CORROSION DATA

Category	F1	F2	Statistic	Semi-mayor	Semi-minor	*p*-value
A	0,7642	0,0048	18,1600	0,7066	0,1327	0,0201
B	-0,2200	0,0956	15,2672	0,5586	0,1049	0,0542
C	-0,1903	-0,0728	12,5352	0,4800	0,0902	0,1289
Light Brown	-0,3051	0,0968	8,5353	0,8070	0,1516	0,3830
Dark Brown	-0,2541	-0,0007	3,1421	0,5645	0,1061	0,9251
Bluish Brown	0,0336	-0,0934	11,5604	0,5769	0,1084	0,1719
Light Blue	0,8136	0,0580	18,8180	0,7900	0,1484	0,0159
Dark Blue	-0,5206	0,1540	3,9065	1,9351	0,3635	0,8655

V. DISCUSSION

Taking into account the two-way contingency table for the SS corrosion data in Table 1, $\chi^2_{stat} = 22.9812$is the observed value of Pearson's chi-squared. With eight degrees of freedom, it exceeds the critical value $\chi^2 = 15.5073$. It suggests that the location and color of SS corrosion are statistically significantly correlated. A correspondence plot depicts the graphical representation of this association. Based on the symmetric plot in Fig. 4, the coordinate point of the row category for location A is further from the origin than the other row categories. It shows that location A is a row category that contributes mostly to the association structure between variables. Considering Fig. 1, location A lies in the center of the specimen; The temperature is higher because there is no heat transfer. Similarly, the column category light blue contributes more than the other column categories since its coordinates are relatively far from the origin. The area has corroded because blue, according to the morphology theory of SS, is the color of the sigma phase.

Furthermore, the association within the row and column categories using the principal coordinates is represented by the symmetric plot. It suggests that it shows the column and row profiles in the same area at the same time. In this case, the only interpretable distance is that which lies inside row or column coordinates. The row coordinates of locations B and C, as well as the column coordinates of light and dark brown, are near to one another, as seen in Fig. 4. It demonstrates how strongly these categories are associated.

Using standard coordinates, we create an asymmetric correspondence plot (Fig. 4) to look for a relationship between row and column coordinates. As can be seen in the figure, there is a relative proximity between A and LBl, B and

LBr, and C and DBr, indicating their dependence. Because it depends on DBr and LBl, these results indicate that A is blue, B is brown, and C is unsure. By location, the results correspond to the physical characteristics of SS. Since Location B is close to the surface, heat transfers there are led outside. Since B is either not yet formed or is in the sigma phase stage, it is brown in this instance. All of the heat in A has been absorbed. A becomes blue due to the formation of several sigma phases.

We create each category's elliptical confidence regions and p_{values} in order to determine which categories statistically contribute to the association between variables. These results are shown in Table 3 and Fig. 6. Here, the eigenvalues of $\dot{\mathbf{S}}$, which monotone decreases ($\lambda_n \geq \lambda_{n+1}$). The weight of the semi-major axis is reflected by the first eigenvalue, λ_1, and the weight of the semi-minor axis by the second eigenvalue, λ_2. Since $\lambda_1 = 0.1552$ and $\lambda_2 = 0.0055$ are relatively different, hence the eccentricity is 0.9822. It shows that the ellipse resembles flattened. In addition, the relative position of each category variable to the origin reflects its contribution to the association. The closer to the origin suggests a smaller contribution. If such a category is omitted, it will not change the association structure.

Based on Fig. 6, a row category of location A and a column category of light blue contribute statistically significantly to the association since its elliptical region does not contain the origin. Their p_{value} can clarify this result in Table 3, which is smaller than the level of significant $\alpha = 0.05$. Lastly, adding up the percentage of variance accounted for by the horizontal and vertical axes suggests that the visualization shows all of the data's variance.

VI. CONCLUSION

The mathematical results provide a formulation for analytically determining the elliptical confidence regions in correspondence analysis, specific to the $3 \times J$ contingency table for $J \geq 3$. These theoretical findings provide insights for academics and researchers for scientific development in the field of correspondence analysis and related methods. Practically, these findings can be applied to solve problems involving two categorical random variables that allow practitioners or companies in decision-making.

The practical findings of a case study involving SS corrosion data led to the conclusion that SS corrosion color and location are statistically significantly correlated. According to this result, location A tends to be blue in extremely hot weather, location B tends to be brown, and location C is unsure because it depends on whether the temperature is dark brown or light blue.

Furthermore, location A and light blue contribute statistically significantly to the association, as their elliptical regions do not contain the origin. This result is confirmed by their respective p-values being smaller than the specified significance level. These practical findings can be recommended to engineers in Petrochemical companies, especially in Indonesia, to develop more effective and efficient preventive and repair strategies for controlling the SS corrosion problems. Some strategies that can be done include (1) choosing SS that has a high Cr composition because the higher the Cr content, the more resistant to the sigma phase; (2) reducing the exposure of SS to high temperatures for an extended period, especially in the center; and (3) conducting regular testing and inspection of SS to detect corrosion.

Despite some advantages, this study has some limitations: (1) the selection of three locations of corrosion (center, surface, edge) may not be generalized to the overall condition of SS; and (2) the frequency of occurrence colors for each location is calculated manually using a grid with one radius size. In future works, improvements will be made by the augmentation of locations and automatic calculation of color frequency by considering and comparing several radius sizes of the grid. Thus, the CA results represent entirely the actual condition of the SS.

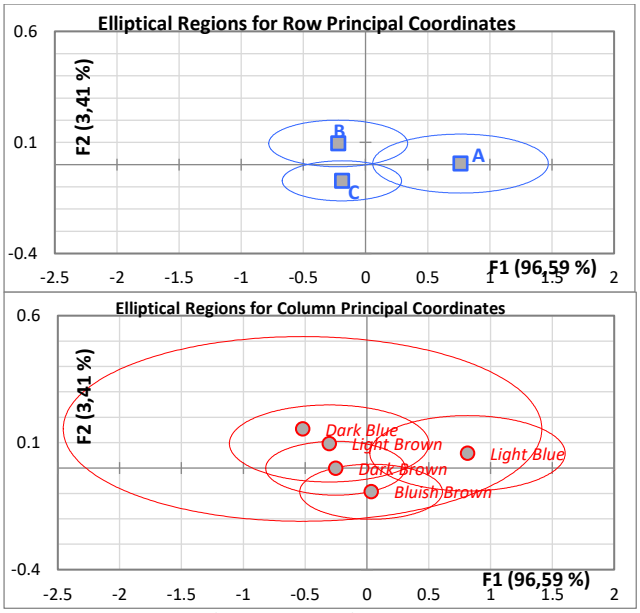


Fig. 6. The elliptical confidence regions for SS data

REFERENCES

[1] P. Moore, "Technical handbook of stainless steels," Atlas steels, pp. 1–45, 2013

[2] J. Hau and A. Seijas, "Sigma Phase Embrittlement of Stainless Steel," in FCC Service, CORROSION NACExpo 61st annual conference & exposition, paper number 06578, 2006

[3] H. Hwang and Y. Park, "Effects of heat treatment on the phase ratio and corrosion resistance of duplex stainless steel," Mater. Trans., vol. 50, no. 6, pp. 1548–1552, Jun. 2009

[4] M. V. Biezma, U. Martin, P. Linhardt, J. Ress, C. Rodriguez, and D. M. Bastidas, "Non-destructive techniques for the detection of sigma phase in duplex stainless steel: A comprehensive," Review Engineering Failure Analysis, vol. 122, 105227, 2021

[5] C. C. Hsieh and W. Wu, "Overview of intermetallic sigma phase precipitation in stainless steels," ISRN Metall, vol. 2012, pp. 1–16, Mar. 2012

[6] V. A. Hosseini, L. Karlsson, S. Wessman, and N. Fuertes, "Effect of sigma phase morphology on the degradation of properties in a super duplex stainless steel," Materials (Basel), vol. 11, no. 6, Jun. 2018

[7] I. J. Marques, A. De Albuquerque Vicente, J. A. S. Tenório, and T. F. De Abreu Santos, "Double kinetics of intermetallic phase precipitation," in UNS S32205 Duplex Stainless Steels Submitted to Isothermal Heat Treatment, Mater. Res., vol. 20, pp. 152–158, Jun. 2017

[8] P. Ferro, A. Fabrizi, and F. Bonollo, "Non-isothermal dissolution modeling of sigma phase in duplex stainless steels," Acta Metall. Sin. (English Lett.), vol. 29, no. 9, pp. 859–868, 2016

[9] G. S. da Fonseca, P. S. N. Mendes, and A. C. M. Silva, "Sigma phase: nucleation and growth," Metals (Basel), vol. 9, no. 1, Jan. 2019

[10] D. M. Aditya, Bachelor thesis, Institut Teknologi Bandung Indonesia, 2020

[11] E. J. Beh and L. D'Ambra, "Non-symmetrical correspondence analysis with concatenation and linear constraints," Aust. New Zeal. J. Stat., vol. 52, no. 1, pp. 27–44, Mar. 2010

[12] E. J. Beh and R. Lombardo, "Correspondence Analysis: Theory, Practice, and New Strategies," pp. 1–560, 2014
[13] M. Greenacre, "Correspondence Analysis in Practice," - 3rd Edition, 2017
[14] I. Ginanjar, U. S. Pasaribu, and A. Barra, "Simplification of correspondence analysis for more precise calculation which one qualitative variable is two categorical data," ARPN Journal, vol. 11, no. 3, 2016
[15] I. Ginanjar, U. S. Pasaribu, and S. W. Indratno, "A measure for objects clustering in principal component analysis biplot: A case study in inter-city buses maintenance cost data," AIP Conference Proceedings, 1827 (1), 020016, 2017
[16] K. E. Lestari, U. S. Pasaribu, S. W. Indratno, and H. Garminia, "Generating roots of cubic polynomials by Cardano's approach on correspondence analysis," Heliyon, vol 6, no. 6, e03998. 2020
[17] M. Greenacre, and J. Blasius, "Multiple Correspondence Analysis and Related Methods", 1st edition, Chapman and Hall/CRC, 2006
[18] K. E. Lestari, M. R. Utami, and M. R. Yudhanegara, "Simple algorithm to construct circular confidence regions in correspondence analysis using R," BAREKENG: Jurnal Ilmu Matematika dan Terapan, vol. 16, no. 1, pp. 065-074. 2022
[19] H. Abdi, M. Béra, "Correspondence Analysis," 2014
[20] E. J. Beh and R. Lombardo, "Correspondence analysis: theory, practice, and new strategies," pp. 1–560, 2015
[21] L. Lebart, L. Morineau, and K. M. Warwick, "Multivariate descriptive analysis: correspondence analysis and related techniques for large matric," John Wiley and Sons, 1984
[22] I. Ginanjar, U. S. Pasaribu, and S. W. Indratno, "Identification the characteristics of Indonesian credit card frauds by trough correspondence analysis (an application of simplification quantitative analysis)," 2nd International Conference on Technology, Informatics, Management, Engineering & Environment, 2014
[23] J. P. Weir, "Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM," in J Strength Cond Res. Feb;19(1):231-40, 2005
[24] Y. S. Afrianti. H. Ardy, U. S. Pasaribu, and F. D. E. Latief, "Identification of inhomogeneous temperature on stainless steel using statistical analysis," Journal of Physics: conference series. ICMSCT 2021, IOP Publishing, 2084, 2021
[25] S. Tufféry, "Data mining and statistics for decision making," John Wiley & Son, 2011
[26] A. C. Rencher, "Methods of multivariate analysis," J. Wiley, 2002
[27] N. Sourial et al., "Correspondence analysis is a useful tool to uncover the relationships among categorical variables," J. Clin. Epidemiol, vol. 63, no. 6, pp. 638–646, 2010
[28] F. Tekaia, "Genome data exploration using correspondence analysis: correspondence analysis and genome data," Bioinform. Biol. Insights, vol. 10, pp. 59–72, 2016
[29] E. J. Beh and R. Lombardo, "An algebraic generalization of some variants of simple correspondence analysis," Metrika, vol. 81, no. 4, pp. 423–443, 2018
[30] Fajriatus Sholihah, Irlandia Ginanjar, and Yuyun Hidayat, "A Hybrid Singly Ordered Correspondence with Correlation Approach to Analyzing the Relationship between Age Groups and Happiness Level in Indonesia," IAENG International Journal of Applied Mathematics, vol. 53, no.3, pp916-923, 2023