Technical Report

Microstructure Formation and Data Science

Koutarou HAYASHI* Kazumasa TSUTSUI

Abstract

The patterns were analyzed based on the element concentration mapping data to clarify the microstructure formation of ternary Fe-C-Mn alloys and quaternary Fe-C-Si-Mn alloys using various data science techniques. Microstructural patterns in Fe-C-Mn alloys were analyzed by histograms of each element concentration and scatter plots among element concentrations. In particular, the equilibrium conditions of the two-phase microstructure formation were estimated by comparing the scatter plots with the isothermal cross-sections of phase equilibrium. Principal component analysis of element concentration mapping data in Fe-0.20%C-1.5%Si-1.5%Mn alloys shows that the patterns of the two-phase microstructure formation can be expressed by principal components. Knowledge of computational science and microstructure formation theory was required to understand this pattern.

1. Introduction

When considering the materials science of steel, in addition to chemical compositions, heat treatment, and characteristics, microstructure is also a major focal point of research. In particular, as microstructure has an influence on various characteristics like strength,[1–3] researches on the formation of the microstructure of steel continue to attract attention, and have achieved great academic progress.[4–6] On the other hand, since the microstructure is non-uniform in the steel, the recognition of patterns and the interpretation thereof are difficult. To clarify complicated patterns, data science employing machine learning and so forth are widely utilized in various industries.[7]

As an example of applying data science to materials science, materials informatics is one particular area of focus.[8–11] Data science in microstructural formation research is shown in **Fig. 1**. Conventional materials science has consisted of three elements, experimental science, computer science, and theoretical science. Data science has been newly introduced, and materials science has made further progress.[9] In the research on the microstructure formation also, the combination and fusion of data science with these three experimental sciences such as element and structure analysis,[12–15] computer science represented by the Calculation of PHAse Diagram (CALPHAD),[16] and microstructure formation theory such as that of precipitation and/or phase transformation[17] need to be considered. As stated in Chapter 3 and beyond, the accumulation of the series of analysis of the experimental data of element analysis with descriptive statics, the study on the correlations of data based on thermodynamic equilibrium calculation, and the penetration into the microstructure formation have developed data science in the materials science field.[18, 19] However, materials informatics is at the dawn, and the research on such data science remains scarce.

Experimental science has the role of producing data, and in addition to mechanical properties, the experimental data like element concentration mapping developed by Field Emission Electron Probe MicroAnalysis (FE-EPMA) is also obtainable. In order to extract microstructure formation patterns from the data, various kinds of data analysis become necessary. **Table 1** shows the outline of the

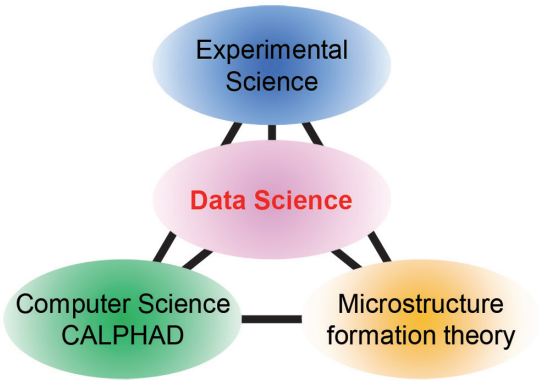

Fig. 1 Data science in microstructural formation research field

* Dr. Eng., Senior Researcher, Research Section-II, Materials Microstructure Characterization Research Dept., Mathematical Science & Technology Research Lab., Advanced Technology Research Laboratories
1-8 Fuso-cho, Amagasaki City, Hyogo Pref. 660-0891

Table 1 Machine learning models in the analysis of materials data

Approach	Algorithm
Supervised learning	Classification
	Regression
Unsupervised learning	Clustering
	Dimensionality reduction

machine learning models utilized for the analysis of material data.[20] The models are divided roughly into two approaches, "supervised learning" and "unsupervised learning". Furthermore, algorithms are developed to achieve various objectives. The first is the classification of the "supervised learning" category which enables the analysis of the microstructure in the steel.[21–25] The second is the "regression" of the "supervised learning" category which enables the prediction of the starting temperature of transformation to martensite or the completion temperature of transformation to austenite.[26, 27] The third is the "clustering" of the "unsupervised learning" category, and the segmentation of microscopic images using the algorithm has been developed.[28] The fourth is the "dimensionality reduction" of the "unsupervised learning" category, and the high dimensional information of microstructure is compressed through principal component analysis, and visualized.[29]

In this research, three test pieces were prepared, each having different alloy chemical compositions and different microstructure. The first test piece is the ternary Fe-0.3%C-0.5%Mn alloy which was heat-treated for a long time, and the microstructure is considered to be in the state of near to equilibrium. The second test piece is the ternary Fe-0.1%C-2.0%Mn alloy which was heat-treated for a short time, and the microstructure is in the state of non-equilibrium. The third test piece is the quaternary Fe-0.2%C-1.5%Si-1.5%Mn alloy in the state of non-equilibrium. Each of the element concentration mapping data was analyzed by using various methods of data science, clarifying the pattern of the respective data, and the microstructure formation was studied based on the patterns. Particularly, with respect to the quaternary Fe-0.2%C-1.5%Si-1.5%Mn alloy, visualization of data with dimensionality reduction was studied. Thus the viewpoint of data science was introduced, and the latent features in the microstructure formation, or patterns, was studied based on the element mapping data.

2. Experimental Method

Test piece materials of high purity were vacuum-smelted, and three test piece alloys were prepared. The chemical compositions of the test piece alloys are shown in **Table 2**. Hot-rolled sheets each with a thickness of 3.5 mm were produced by using a slab heating furnace and hot rolling equipment. After removing scale from the hot-rolled sheets by mechanical grinding, the hot-rolled sheets were cold-rolled to 1.2 mm thick sheets.

The cold-rolled sheets of the three alloys were heat-treated under the conditions described below, and the test pieces were prepared. The cross sections of the test-pieces were buffed with alumina and corroded with Nital. The microstructure of the corroded cross section was observed with a scanning electron microscope. FE-EPMA was used to calibrate the elemental distribution of the test piece. A standard test piece was used to determine the C, Si, and Mn concentrations. The acceleration voltage, irradiation current, probe diameter, and the measurement intervals were: 15 kV, 1.0×10^{-7} A, 0.1 μm, and 0.2 μm, respectively.

For data analysis of the material, the data analysis and AI development platform NS-DIG™ was used.[30]

Table 2 Chemical composition of the alloy studied

(mass%)

Alloy	C	Si	Mn
Fe-0.3%C-0.5%Mn	0.29	< 0.01	0.5
Fe-0.1%C-2.0%Mn	0.10	< 0.01	2.0
Fe-0.2%C-1.5%Si-1.5%Mn	0.20	1.5	1.5

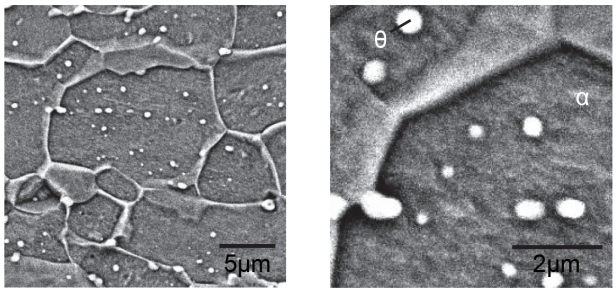

Fig. 2 SEM images of the cross-sectional microstructure for Fe-0.3%C-0.5%Mn alloy annealed at 650°C for 72 h

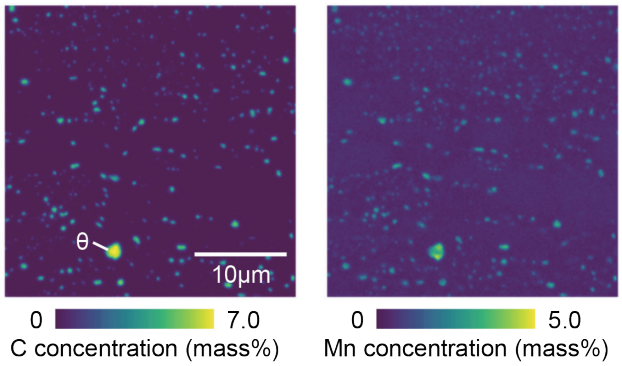

Fig. 3 C and Mn mappings for Fe-0.3%C-0.5%Mn alloy annealed at 650°C for 72 h

3. Microstructure of Ternary Alloy Heat-treated for Long Time

To decrease the strength of steels, in some cases, long time heat treatment under A_{c1} is provided.[31] According to thermodynamic equilibrium calculation, the formation of two-phase structure of ferrite (α) and cementite (θ) is predicted under such heat treatment.[32] This chapter analyzes the microstructure formation due to long-time heat treatment below A_{c1} using the descriptive statistics of element concentration mapping data.

Figure 2 shows the cross-sectional microstructure of the test piece of the Fe-0.3%C-0.5%Mn alloy annealed at 650°C for 72 h. Spherical θ is dispersed in the parent phase α.

Figure 3 shows the C and Mn concentration mappings of the test piece by FE-EPMA analysis. The phases of dispersion of regions of high C concentration and Mn concentration are observed. The regions are of θ, and the surrounding area is of the parent phase α.

The aforementioned element concentration mapping data consist of 75 625 points of the quantitative variables of the Si concentration and Mn concentration. **Figure 4** shows the histograms of the C con-

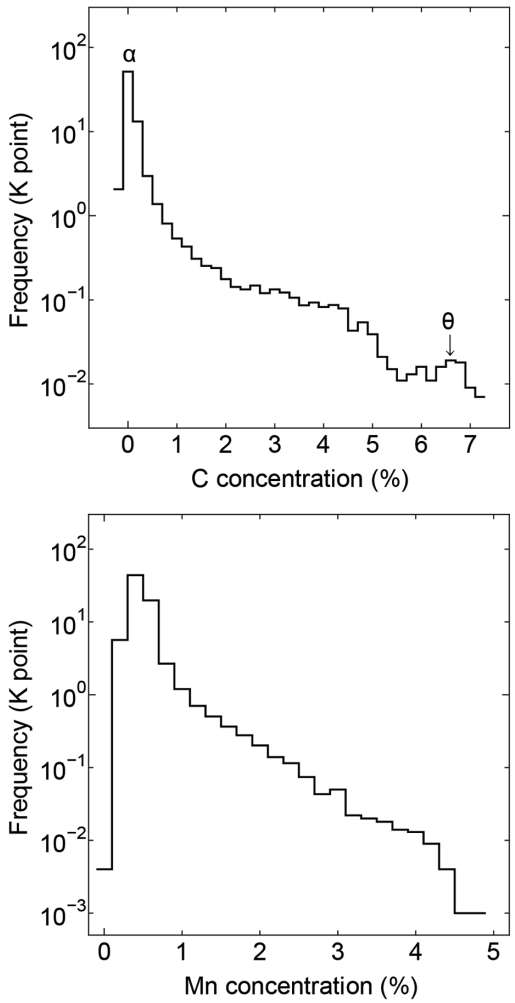

Fig. 4 Histograms of C and Mn concentrations for Fe-0.3%C-0.5%Mn alloy annealed at 650°C for 72 h

centration and Mn concentration of the test piece of the Fe-0.3%C-0.5%Mn alloy annealed at 650°C for 72 h. The vertical axis of the figure shows the logarithm values of frequency, and the horizontal axis shows the C concentration and Mn concentration. Peaks on the parent phase α appear at the lower sides of the C concentration and Mn concentration. As shown by ↓, a small peak of θ is also recognized.

Figure 5 shows the scatter plot of the C concentration and Mn concentration of the test piece of the Fe-0.3%C-0.5%Mn alloy annealed at 650°C for 72 h. According to the figure, the C concentration and Mn concentration maintain a positive correlation with each other. The correlation coefficient of the two is 0.84. In the figure, the blue broken line is the tie line of α and θ which runs through the Fe-0.3%C-0.5%Mn alloy.[32] The lower left plots and the upper right plots represent the C concentration and Mn concentration of α and θ, respectively. The tie line inclination is positive, and data are distributed along the tie line. The distribution like this shows that, under annealing at 650°C for 72 h, the microstructure is near to the state of equilibrium of α and θ. Thus, microstructure formation reaching an equilibrium state has been clarified by mapping data of the C concentration and Mn concentration.

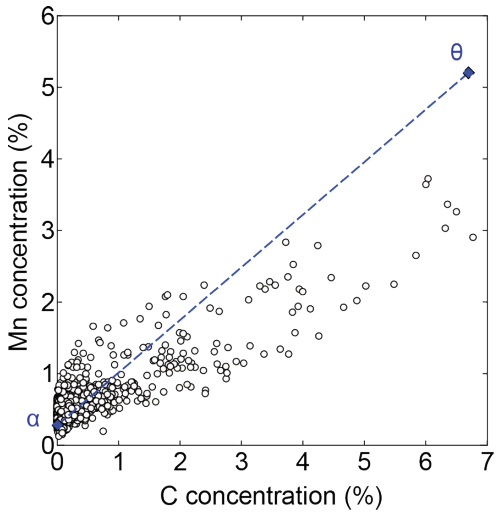

Fig. 5 Scatter plot of C and Mn concentrations for Fe-0.3%C-0.5%Mn alloy annealed at 650°C for 72 h

4. Microstructure of Ternary Alloy after Short Time Heat Treatment

In order to increase the strength of steels, sometimes heat treatment is provided in the austenite (γ) single-phase region. During its heating process, γ is formed from α and θ structures, and the two-phase structure of α and γ is formed. Furthermore, when the heat treatment time is short, the structure is considered to be in a state of non-equilibrium. In this chapter, the microstructure formation in short time heat treatment above A_{c1} is analyzed with the descriptive statistics of element concentration mapping and image processing.

Figure 6 shows the C and Mn mappings for the Fe-0.1%C-2.0%Mn alloy annealed at 750°C for 1 000 s. According to the figure, a two-phase structure consisting of regions α of low C concentration and low Mn concentration, and γ having high concentrations are clearly recognizable. γ becomes martensite (M) by cooling after being isothermally retained. Herein, in addition to the analysis of the C concentration and Mn concentration mappings as explained in the previous chapter, the information of the microstructure of α and γ is also utilized.

Table 3 shows the mean values and the standard deviations of the C concentration and Mn concentration in the mapping data. The mean value and the standard deviation of the C concentration are nearly equal. In the meantime, the standard deviation of the Mn concentration is smaller than its mean value. This is because the diffusion coefficient of C is higher than that of Mn,[33] and therefore the distribution of C between α and γ precedes the distribution of Mn.

Figure 7 shows the scatter plots of the C concentration and Mn concentration for the Fe-0.1%C-2.0%Mn alloy annealed at 750°C for 1 000 s. According to the figure, a positive correlation between the C concentration and Mn concentration is found. The correlation coefficient between the two is 0.54. In addition, bimodal distribution each responding to α and γ is formed. By separating the distribution, information concerning the equilibrium condition and so on between α and γ is obtained. Then, the bimodal distribution was separated according to the following method.

As the element concentration mapping data is of tensor form consisting of the information of space and concentration, analysis in the same way as that of image data is possible.[34] Then, the Mn concentration mapping having high contrast was transformed to a gray-

scale image. Specifically, a grayscale image with 256 gradations between 0 and 255 tone steps has been prepared. With this transformation, the binarization processing based on the characteristics of brightness has been made possible.[35, 36]

Figure 8 shows the gray scale image and its binary image. The gray scale image on the left side was processed with the adaptive binarization method. The purple regions and the yellow regions on the right binary image become α and γ, respectively. From the binary image, the area fraction of microstructure of α and γ are obtained, which are 38% and 62%, respectively. In the meantime, according to thermodynamic equilibrium calculation, the histodifferentiation rate of α is 59% and the rate of γ is 41%,[32] which are different from the values obtained from the binary image. This is considered to be attributed to the non-equilibrium state of the test piece, causing delay of Mn in distribution between α and γ.

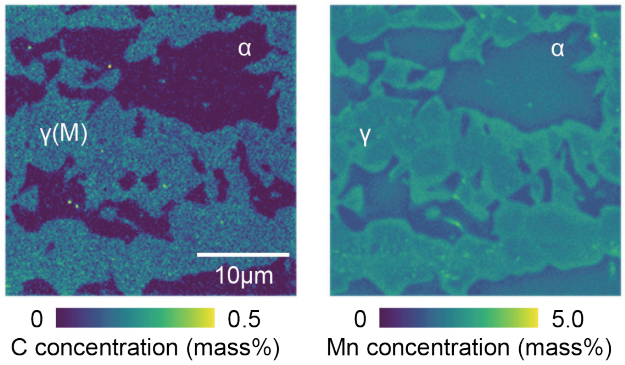

Fig. 6 C and Mn mappings for Fe-0.1%C-2.0%Mn alloy annealed at 750°C for 1 000 s

Table 3 The mean value and standard deviation of C and Mn concentrations for Fe-0.1%C-2.0%Mn alloy annealed at 750°C for 1 000 s

Element	Mean value	Standard deviation
C	0.108	0.111
Mn	2.05	0.396

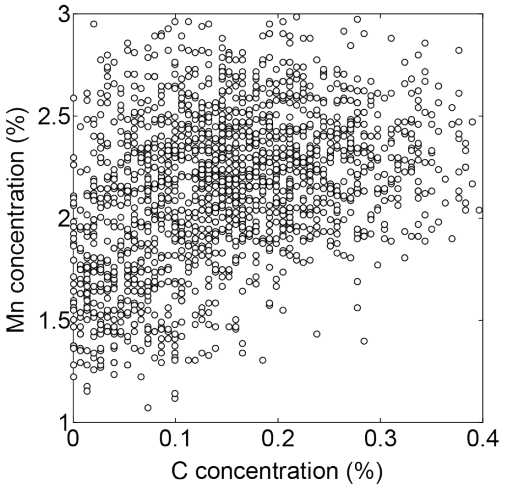

Fig. 7 Scatter plot of C and Mn concentrations for Fe-0.1%C-2.0%Mn alloy annealed at 750°C for 1 000 s

From the above analysis, information of element concentration and microstructure was obtained. By incorporating the two, microstructure formation like that of the two-phase equilibrium of α and γ can be analyzed. With the binary image, the bimodal distribution of the element mapping data was separated into α and γ distributions. As **Fig. 9** shows, such distributions were laid over the isothermal cross-section for the ternary Fe-C-Mn at 750°C.

Purple color plots and yellow color plots denote α and γ, respectively. According to the figure, although the distributions of α and γ overlap with each other, the α distribution is situated at the lower C concentration side and lower Mn concentration side, and the γ distribution is situated at the higher C concentration side and higher Mn concentration side. Furthermore, the right side edge of the γ distribution falls in line with the γ side of the α–γ phase boundary line. This suggests that the formation of the two-phase structure of α and γ follows local equilibrium. In the figure, the blue broken line is the tie line of α and γ going through the Fe-0.1%C-2.0%Mn alloy.[32] The left bottom point and the right top point of the tie line show the C concentration and Mn concentration of α and γ, respectively. As aforementioned, when the microstructure reaches the state of equilibrium, data are distributed along the alloy tie line. However, many of the data in α and γ are distributed away from the tie line. Distribution like this means that the microstructure is in a state of non-equi-

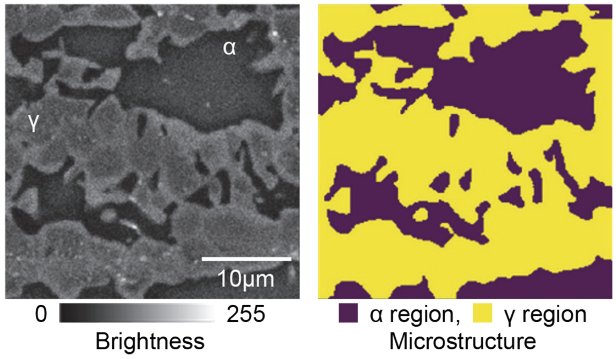

Fig. 8 Grayscale image with transformed Mn mapping and its binary image

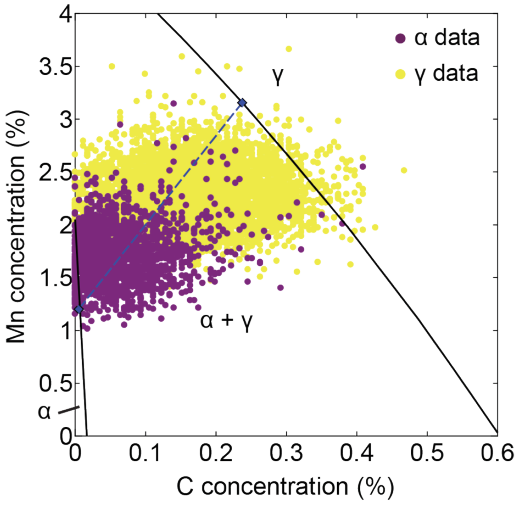

Fig. 9 Overlay of the isothermal cross-section for the Fe-C-Mn ternary system and element mapping data of α and γ at 750°C

librium. Thus, by integrating the information of the C concentration, Mn concentration, and further, the information of microstructure of α and γ, the formation process of non-equilibrium microstructure has been clarified.

5. Microstructure of Quaternary Alloy

For certain high tensile strength steel, microstructure is controlled by heat treatment like austempering after being heated at a high temperature.[2] By heat treatment like this, microstructure of the non-equilibrium state is formed. Furthermore, in order to enhance the characteristics of steels, various alloy elements are added. Among them, Mn and Si are vital alloying elements for high tensile strength steel, and quaternary alloys or of higher dimensions are used.[37] As shown until now, in order to understand the microstructure formation of the ternary Fe-C-Mn alloy, both CALPHAD and data science analyses have been effective, including the use of isothermal cross sections. However, in the isothermal cross-section, the information of the solute elements is limited to two kinds, and the analysis of microstructure formation of the quaternary Fe-C-Si-Mn alloy becomes difficult. With the dimensionality reduction described in Chapter 1, the quaternary Fe-C-Si-Mn alloy data can be expressed on scatter plots. In this chapter, information of element concentration is compressed with the principal component analysis of the element concentration mapping data, and with the analysis of the potential features of the data, the relationship between its features and the microstructure formation is studied.

Figure 10 shows the C, Si, and Mn concentration mappings of the Fe-0.2%C-1.5%Si-1.5%Mn alloy test pieces, each heat-treated under the respective conditions. Figures of the left column show the data of the test piece intercritically annealed at 800°C for 100 s. In the said test piece, regions with low C concentration are like islands, and scattered, and the C concentration within the test piece is inhomogeneous. The regions with low C concentration are of α, and the regions with high C concentration are of M, or γ during intercritical annealing. In addition, the Si concentration and Mn concentration are also heterogeneous. Regions with a high Si concentration and low Mn concentration are of α, and the regions with a low Si concentration and high Mn concentration are of γ during intercritical annealing. In the thermodynamic equilibrium calculation for the Fe-0.2%C-1.5%Si-1.5%Mn alloy, concentrations of C, Si, and Mn of α at 800°C are 0.011%, 1.7%, and 0.90%, respectively. Concentrations of C, Si, and Mn of γ are 0.32%, 1.4%, and 1.9%, respectively. Accordingly, the features of the element distribution at 800°C agreed with the result of the thermodynamic equilibrium calculation.

Figures of the right column show the data of the test piece isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C for 100 s. According to the comparison of this test piece with the one intercritical annealing at 800°C for 100 s, C is unevenly distributed by the isothermal holding at 400°C. This uneven distribution is considered to be attributed to the formation of C-rich γ or the precipitation of θ through transformation of γ to bainite. In addition, according to the Si and Mn concentration mappings, change in phase due to isothermal holding at 400°C is not recognized.

Figure 11 shows the 3D scatter plots of the C, Si, and Mn concentrations of the test piece isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C for 100 s. According to the figure, the extent of scatter of the C concentration is large. However, since the plots are concentrated, it is difficult to understand from the 3D scatter plots the features of data like the relationship between the Si concentration and Mn concentration. Thus, dimensionality reduction becomes necessitated in the analysis of element concentration mappings of quaternary alloys.

The principal component analysis is the method of introducing new variables expressed by the linear combinations of the variables obtained by analyzing the observed data of multi-dimensional variables. The new variables are termed as principal components, and the background of data, or specifically, the causes of the scatter are defined. Furthermore, with focus on the principal components having high contribution ratios, dimensionality reduction of data also becomes possible.

The relationship between the observed variables and the latent variables in the principal component analysis of element concentration mapping is described. The observable variables in the data are C, Si, and Mn concentrations. The principal components are considered as latent variables as they influence the observable variables. Since the scattering of element concentration is produced by the

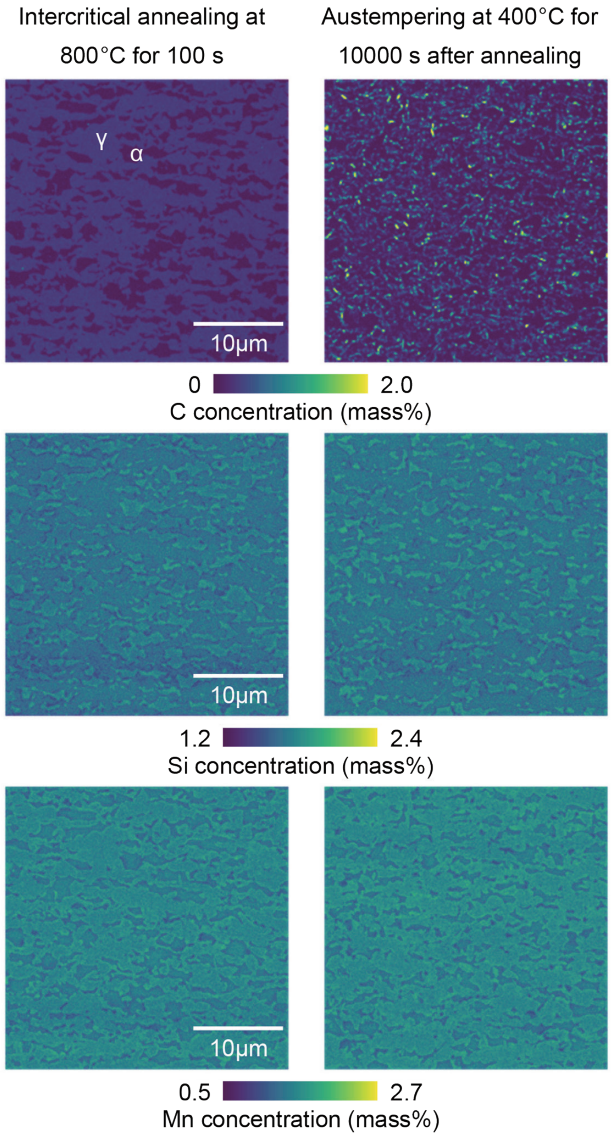

Fig. 10 **C, Si, and Mn mappings for Fe-0.2%C-1.5%Si-1.5%Mn alloys. One is intercritically annealed at 800°C for 100 s. The other is isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C for 100 s.**

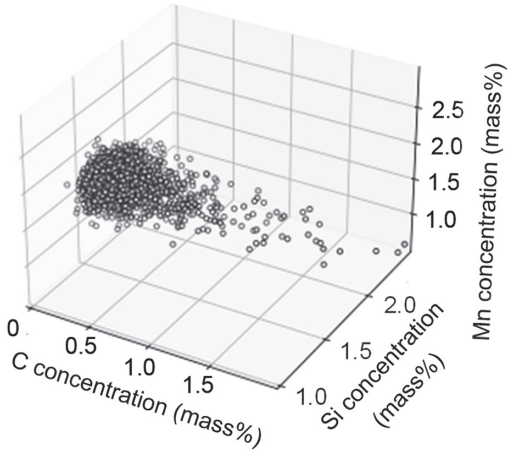

Fig. 11 3D scatter plot of C, Si, and Mn concentrations for Fe-0.2%C-1.5%Si-1.5%Mn alloy isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C for 100 s

structure formation, the principal components of latent variables are considered to represent the microstructure formation. Namely, with the principal component analysis of element distribution data, the factor of structure forming can be estimated.

For the principal component analysis, a method of using the correlation coefficient matrix is shown below.[38] The variables of the C, Si, and Mn concentrations are expressed as w_C, w_{Si}, and w_{Mn}, respectively. The unit is mass%. These variables are standardized with the respective mean value and standard deviation. The concentration of the respective solute i so standardized is denoted with s_i.

The first principal component of z_1 is expressed as below.

$$z_1 = a_{11}s_C + a_{12}s_{Si} + a_{13}s_{Mn}$$

where a_{11}, a_{12}, and a_{13} are the coefficients which maximize the scatter of z_1. The combinations of these coefficients are expressed by the inherent vector of the maximum inherent value in the correlation coefficient matrix of w_C, w_{Si}, and w_{Mn}.

Similarly, the k_{th} principal component z_k is expressed as below.

$$z_k = a_{k1}s_C + a_{k2}s_{Si} + a_{k3}s_{Mn}$$

where a_{k1}, a_{k2}, and a_{k3} are expressed by the inherent vector of the k_{th} inherent value in the aforementioned correlation coefficient matrix.

The element scatter plot data of the test piece of the quaternary Fe-0.2%C-1.5%Si-1.5%Mn alloy isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C was provided for principal component analysis. For the analysis of the principal component, scikit-learn of the machine learning library was used.[39]

The contribution ratio of the first principal component, of the second principal component, and of the third principal component was 0.55, 0.27, and 0.18, respectively. Since the accumulated contribution ratio up to the second principal component is 0.82, the first principal component and the second principal component contain about 82% of the information of the element distribution data.

Table 4 shows the coefficient of the respective principal components. In the first principal component, the absolute value of the respective coefficient is similar, and the signs of the coefficients of the C concentration and Mn concentration are positive, while the sign of the coefficient of the Si concentration is negative. According to the aforementioned thermodynamic equilibrium calculation, due to structure formation at an intercritical annealing at 800°C, the C concentration and Mn concentration in γ are high, and the Si concentration becomes lower in γ. The features of the element distribution in

Table 4 The coefficients of the principal components for Fe-0.2%C-1.5%Si-1.5%Mn alloy isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C for 100 s. The principal components contain standardized concentration terms

Primary component	z_1	z_2	z_3
Coefficient of S_C	0.48	0.88	−0.05
Coefficient of S_{Si}	−0.62	0.37	0.69
Coefficient of S_{Mn}	0.63	−0.30	0.72

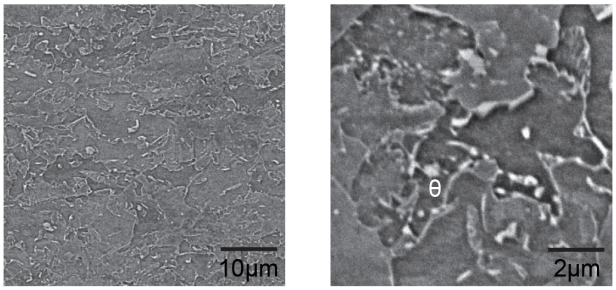

Fig. 12 SEM images of the cross-sectional microstructure for Fe-0.2%C-1.5%Si-1.5%Mn alloy isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C for 100 s

γ correspond to the respective coefficient of the first principal component. Accordingly, the first principal component is considered to represent the structure formation in the two-phase temperature region. In the second principal component, the absolute value of the coefficient of the C concentration is larger than those of other coefficients. If θ precipitates during isothermal-holding at 400°C, due to distribution of C between α and θ, the C concentration scatter becomes larger. Accordingly, the second principal component is considered to represent the structure formation during isothermal holding at 400°C, namely the precipitation of cementite.

Figure 12 shows the cross-sectional microstructure of the test piece of the Fe-0.2%C-1.5%Si-1.5%Mn alloy isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C. According to the left figure, the test piece is of α and bainite microstructure. The figure on the right shows the enlarged photography of the left figure. According to the right figure, the precipitation of θ from the bainitic structure is found. Thus, the aforementioned second principal component is confirmed to relate to the precipitation of cementite.

Secondly, the C, Si, and Mn concentration data were standardized, and the first principal component scores and the second principal component scores were calculated. **Figure 13** shows the scatter plots of the test piece of the Fe-0.2%C-1.5%Si-1.5%Mn alloy isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C for 100 s.

The upper figure shows the scatter plots of the test piece of the Fe-0.2%C-1.5%Si-1.5%Mn alloy intercritically annealed at 800°C for 100 s, and the horizontal axis and the vertical axis show the first principal component scores and the second principal component scores, respectively. According to the figures, regarding the distribution of data, the scores of the first principal component spread in the lateral direction. In the figure, the blue broken line is the tie line of α and γ going through the Fe-0.2%C-1.5%Si-1.5%Mn alloy. The upper left plot and the lower right plot respectively show the plot of the first principal component score and the second principal component score calculated respectively from the equilibrium density of α and γ

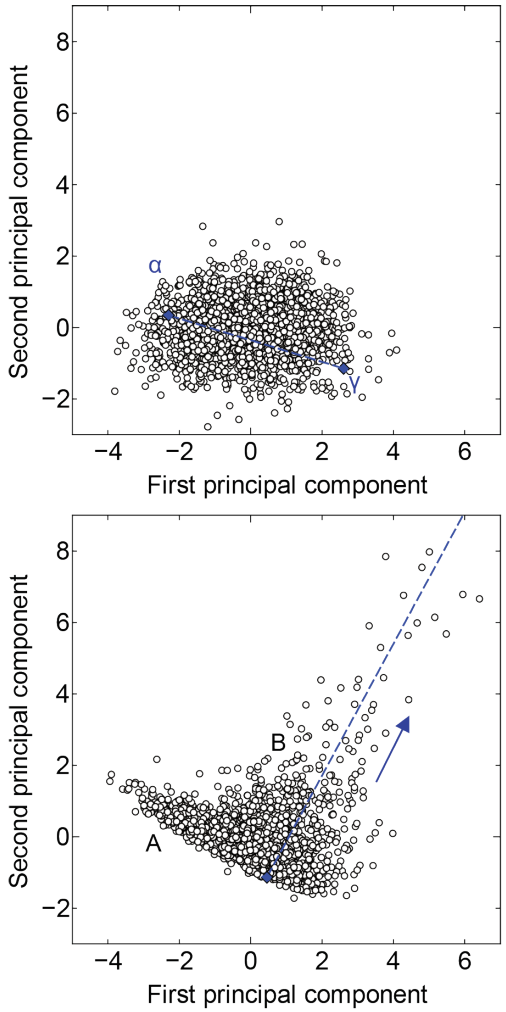

Fig. 13 Scatter plots of the first and second primary component scores for Fe-0.2%C-1.5%Si-1.5%Mn alloys. One is intercritically annealed at 800°C for 100 s. The other is isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C for 100 s

at 800°C. According to the figure, the respective equilibrium concentration lies at the edge of the distribution. Accordingly, this distribution is considered to be formed by the distributions of C, Si, and Mn in the formation of two-phase structure at 800°C.

The lower figure shows the plots of the test piece of the Fe-0.2%C-1.5%Si-1.5%Mn alloy isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C. According to the comparison of the plots and the aforementioned 3D scatter plots, the principal component scores plots is a method which visualizes the features of data for easy understanding. According to the figure, two types of distribution are recognized across the first principal component scores and the second principal component scores. The first distribution is A which stretches from upper left to lower right. The second distribution is B which resides at the upper right above A. The distribution A was produced in the aforementioned two-phase structure formation. The blue broken line is the trajectory of the first principal component scores and the second principal component scores when the C concentration is arbitrarily changed with the Si concentration and Mn concentration fixed at 1.5%. In the direction of the arrow mark (→) in the figure, the C concentration increases. Along this trajectory, the distribution B expands. With the formation of bainite, the C concentration in γ increases, and furthermore, θ becomes precipitated. Therefore, the distribution B represents the data of γ and θ, and is considered to be formed in the decomposition of γ. In this way, with principal component analysis, the multidimensional element distribution information is compressed to two dimensional distribution, and two-phase structure formation of α and γ, and further, the features of the precipitation process of θ become visualized.

6. Conclusion

The element concentration mapping data of the ternary Fe-C-Mn alloy and quaternary Fe-C-Si-Mn alloy were analyzed using various methods of the data science technique, and the patterns of the respective data were clarified, and the relationship between their features and the microstructure formation was studied. The microstructure patterns in the ternary Fe-C-Mn alloy were analyzed based on the histogram of each solute element concentration, and further based on the scatter plots of the C concentration and Mn concentration. When the Fe-0.3%C-0.5%Mn alloy is annealed at 650°C for 72 h, its microstructure consists of two-phase of α and θ, and the element distribution is close to the equilibrium state. In the case that the Fe-0.1%C-2.0%Mn alloy is annealed at 750°C for 1 000 s, its microstructure consisted of two-phase of α and γ. According to the comparison of scatter plots of the C concentration and Mn concentration with an isothermal cross-section of phase equilibrium, it is suggested that the formation of the two-phase microstructure of α and γ follows local equilibrium. According to the principal component analysis of the element mapping data of the quaternary Fe-0.20%C-1.5%Si-1.5%Mn alloy isothermally held at 400°C for 10 000 s after intercritical annealing at 800°C for 100 s, the signs of the coefficients of the first principal component of the C concentration and Mn concentration are positive, while the sign of the Si concentration is negative. The respective coefficient of the first principal component corresponds to the element distribution developed by the isothermal-holding in the two-phase region of α and γ. The absolute value of the coefficient of the second principal component of the C concentration is larger than those of other coefficients, and its features correspond to the C enrichment in γ and the precipitation of θ due to bainitic transformation. This microstructure was confirmed through observation with a scanning electron microscope. Thus, the first principal component represents two-phase microstructure formation, and the second principal component represents bainite structure formation. Furthermore, in the scatter plots of the first principal component scores and the second principal component scores, two types of distribution appeared, each depending on the respective microstructure formation. Thus, by analyzing the element concentration mapping data with various data science techniques, patterns which represent microstructure formation have been clarified.

In this paper, we describe the development of experimental science, including high-resolution analysis by FE-EPMA, as well as the spread of data science, which has deepened the interpretation of microstructure. Incidentally, in the "Introduction to Thermodynamics of Microstructures"[4], a quotation from Saint Exupéry states: "It is only with the heart that one can see rightly. What is essential is invisible to the eye"[40]. Mapping data of high dimensions is dimensionally reduced by principal component analysis, and data generation is visualized on scatter plots. However, the interpretation of principal components requires viewpoints of CALPHAD. Thus, elemental research such as computational science is at the heart of the

material researcher, and this heart gives insight into the microstructure formation, and enables us to understand its essence.

Acknowledgments

We are deeply grateful to Professor Manabu Takahashi of Kyushu University for discussing and teaching the role of data science in materials science, to Professor Hidenori Terasaki of Kumamoto University for teaching data science, and to Ms. Shizuka Otsuki of Nippon Steel Technology Co., Ltd. for her help in carrying out the experiments.

References

1) VISUAL ENGINEERING Book for understanding steel sheet and thick plate. Edited by Nippon Steel Corporation, NIPPON JITSUGYO PUBLISHING, 2009
2) Takahashi, M.: Tetsu-to-Hagané. 100, 82 (2014)
3) Umemoto, M.: Tetsu-to-Hagané. 81, 157 (1995)
4) Nishizawa, T.: Course: Contemporary Metallography Materials Vol. 2 Thermodynamics of Microstructures. The Institute of Metals and Materials, 2005
5) Maki, M.: Microstructure Control of Steel, its Principle and Method. UCHIDA ROKAKUHO PUBLISHING CO., LTD., 2015
6) Enomoto, M.: Phase Transformation of Metals, Introduction to the Science of Materials Microstructure. UCHIDA ROKAKUHO PUBLISHING CO., LTD., 2000
7) Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, 2006
8) Ramakrishna, S., Zhang, T.-Y., Lu, W.-C., Qian, Q., Low, J.S.C., Yune, J.H.R., Tan, D.Z.L., Bressan, S., Sanvito, S., Kalidindi, S.R.: J. Intell. Manuf. 30, 2307 (2019)
9) Iwasaki, Y.: Materials Informatics. Nikkan Kogyo Shimbun, Ltd., 2019
10) Yoshida, R.: Proceedings of the Institute of Statistical Mathematics. 69, 5 (2021)
11) Tsutsui, K., Namba, T., Kihara, K., Hirata, J., Matsuo, S., Ito, K.: Tetsu-to-Hagané. 109, 464 (2023)
12) Liu, Z.-Q., Miyamoto, G., Yang, Z.-G., Furuhara, T.: Acta Mater. 61, 3120 (2013)
13) Yamashita, T., Tanaka, Y., Nagayoshi, M., Ishida, K.: Scientific Reports. 6, 29825 (2016)
14) Amino, T.: Nippon Steel & Sumitomo Metal Technical Report. (118), 3 (2018)
15) Kobayashi, Y., Takahashi, J., Kawakami, K.: Nippon Steel & Sumitomo Metal Technical Report. (118), 22 (2018)
16) Saunders, N., Miodownik, A.P.: CALPHAD (Calculation of Phase Diagram): A Comprehensive Guide. Elsevier Science, 1998
17) Aaronson, H.I., Enomoto, M., Lee, J.K.: Mechanisms of Diffusional Phase Transformations in Metals and Alloys. CRC Press, 2010
18) Tsutsui, K., Moriguchi, K.: Calphad. 74, 102023 (2021)
19) Hayashi, K., Kinoshita, Y., Fujimura, R., Takahashi, M.: Tetsu-to-Hagané. 109, 623 (2023)
20) Chibani, S., Coudert, F.-X.: APL Mater. 8, 080701 (2020)
21) Adachi, Y., Taguchi, M., Hirokawa, S.: Tetsu-to-Hagané. 102, 722 (2016)
22) Taguchi, M., Hirokawa, S., Yasuda, I., Tokuda, K., Adachi, Y.: Tetsu-to-Hagané. 103, 142 (2017)
23) Holm, E.A., Cohn, R., Gao, N., Kitahara, A.R., Matson, T.P., Lei, B., Yarasi, S.R.: Metall. Trans. A, 51A, 5985 (2020)
24) Tsutsui, K., Terasaki, H., Maemura, T., Hayashi, K., Moriguchi, K., Morito, S.: Comput. Mater. Sci. 159, 403 (2019)
25) Tsutsui, K., Terasaki, H., Uto, K., Maemura, T., Hiramatsu, S., Hayashi, K., Moriguchi, K., Morito, S.: Mater. Today Commun. 25, 101514 (2020)
26) Singh, S.B., Bhadeshia, H.K., Mackay, D.J.C., Carey, H., Martin, I.: Ironmaking Steelmaking. 25, 355 (1998)
27) Kim, H., Inoue, J., Okada, M., Nagata, K.: ISIJ Int. 57, 2229 (2017)
28) Kim, H., Inoue, J., Kasuya, T.: Scientifc Reports. 10, 17835 (2020)
29) Kalidindi, S.R.: J. Appl. Phys. 128, 041103 (2020)
30) https://www.nipponsteel.com/news/20190425_200.html
31) New edition Steels and Alloying Elements. Edited by The Iron and Steel Institute of Japan, The Iron and Steel Institute of Japan, 2015
32) Thermo-Calc, https://thermocalc.com/products/thermo-calc
33) Revised edition 4 Metal Data Book. Edited by The Institute of Metals and Materials, Maruzen Publishing Co.,Ltd., 2004
34) F.-Zelaia, P., Rossy, A.M., Campbell, Q., Nycz, A., Ledford, C., Kirka, M.M.: Mater. Charact. 185, 111759 (2022)
35) OpenCV, https://opencv.org/
36) Scikit image, https://scikit-image.org/
37) Ushioda, K., Yoshimura, M., Kaidoh, H., Kimura, K.: Tetsu-to-Hagané. 100, 716 (2014)
38) Nagata, Y., Munechika, M.: Introduction to Multivariable Analysis. SAIENSU-SHA Co., Ltd, 2001
39) scikit-learn, https://scikit-learn.org/
40) Saint Exupéry: The Little Prince. Translated by Naito, A., Iwanami Shoten, Publishers, 1953

Koutarou HAYASHI
Dr. Eng., Senior Researcher
Research Section-II, Materials Microstructure Characterization Research Dept.
Mathematical Science & Technology Research Lab.
Advanced Technology Research Laboratories
1-8 Fuso-cho, Amagasaki City, Hyogo Pref. 660-0891

Kazumasa TSUTSUI
Ph.D., Researcher
Research Section-I, Computational Engineering Research Dept.
Mathematical Science & Technology Research Lab.
Advanced Technology Research Laboratories