



# A Combination of Geostatistical Methods and Principal Components Analysis for Detection of Mineralized Zones in Seafloor Hydrothermal Systems

Vitor Ribeiro de Sá,<sup>1</sup> Katsuaki Koike ,<sup>1,9</sup> Tada-nori Goto,<sup>2</sup> Tatsuo Nozaki,<sup>3,4,5,6</sup> Yutaro Takaya,<sup>3,6,7</sup> and Toru Yamasaki<sup>8</sup>

Received 30 November 2019; accepted 16 May 2020

Seafloor massive sulfide (SMS) deposits are typical of submarine mineral resources and generally rich in base metals (Cu, Pb, Zn); however, their distribution, configuration, and formation mechanism, especially sub-seafloor mineralization, remain poorly understood because of scant drilling and geophysical data. To address this problem, this study aims to identify and characterize mineralized zones in seafloor hydrothermal areas using limited metal content data from sparse drilling sites. We use principal component analysis to decrease the dimensionality of the content data. High metal content zones are delineated using principal component values by three geostatistical methods: (1) spatial estimation using ordinary kriging; (2) turning bands simulations (TBSIM); and (3) sequential Gaussian simulations. We selected an active seafloor vent area at 1570 m below sea level in the Okinawa Trough, southwest Japan, as a case study. Results from the three methods show two types of high metal content zones: One is around a sulfide mound, and the other is layered in association with lateral flow of hydrothermal fluids from the bottom of the mound. TBSIM is the most effective under scarce data conditions because the model yields the smallest cross-validation error, decreases the smoothing effect, and corresponds well to a conceptual deposit model that shows a stockwork below the sulfide mound. The results contribute to better understanding the formation mechanism of SMS deposits as well as constraining submarine metal reserves and mining.

**KEY WORDS:** Metal content, Principal components analysis, Ordinary kriging, Turning bands simulation, Sequential Gaussian simulation, Okinawa Trough.

<sup>1</sup>Department of Urban Management, Graduate School of Engineering, Kyoto University, Kyoto 615-8540, Japan.

<sup>2</sup>Graduate School of Life Science, University of Hyogo, Himeji, Hyogo 671-2280, Japan.

<sup>3</sup>Submarine Resources Research Center, Research Institute for Marine Resources Utilization, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Kanagawa 237-0061, Japan.

<sup>4</sup>Frontier Research Center for Energy and Resources (FR CER), School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan.

<sup>5</sup>Department of Planetology, Graduate School of Science, Kobe University, Kobe, Hyogo 657-8501, Japan.

<sup>6</sup>Ocean Resources Research Center for Next Generation, Chiba Institute of Technology, Narashino, Chiba 275-0016, Japan.

<sup>7</sup>Faculty of Science and Engineering, Waseda University, Tokyo 169-8555, Japan.

<sup>8</sup>Research Institute of Geology and Geoinformation, Geological Survey of Japan (AIST), Tsukuba, Ibaraki 305-8560, Japan.

<sup>9</sup>To whom correspondence should be addressed; e-mail: koike.katsuaki.5x@kyoto-u.ac.jp

## INTRODUCTION

Spatial modeling of geologic phenomena is challenging and involves the evaluation of uncertainties innate to any estimation or simulation method (e.g., Bowen 2010; Koike et al. 2015; Battalgazy and Madani 2019). Uncertainty is often represented by a variable degree of a probability distribution (Pyrzcz and Deutsch 2014), which originates from sparsity of sampled data and/or heterogeneity of geologic properties. Such spatial uncertainty can be modeled and quantified by geostatistical methods (Delfiner and Chilès 2012) using estimation and stochastic conditional simulations for unsampled locations. The uncertainty measures include estimate error variance (Calder and Cressie 2009) and the variability of multiple equiprobable simulated realizations (Pyrzcz and Deutsch 2014).

Ordinary kriging (OK) is an optimal linear unbiased estimator that has been widely used in earth science (e.g., Isaaks and Srivastava 1989; Shahbeik et al. 2014; Ilyas et al. 2016; Pugliese et al. 2016). However, OK also brings inevitable problems common to any other estimators, particularly a smoothing effect that makes OK estimations much smoother than actual heterogeneities. Simulations are then required to reproduce the heterogeneity by considering stochastic properties. The most popular simulation method is the sequential Gaussian simulation (SGSIM) based on a Gaussian random field (Deutsch and Journel 1998; Chilès and Lantuéjoul 2005; Emery 2007). SGSIM is a straightforward method to obtain sequential values at each new simulation point from a conditional distribution assigned to the data and previously simulated values (Delfiner and Chilès 2012). Another typical simulation method is the turning bands simulation (TBSIM) that also relies on a Gaussian random field (Matheron 1973; Ren 2005; Emery and Lantuéjoul 2006). TBSIM generates  $\mathbb{R}^n$  simulations from multiple independent, unconditional  $\mathbb{R}^1$  simulations along lines that can be rotated in  $\mathbb{R}^3$  space. This unconditional simulation is then corrected to be conditional by kriging so that the simulated values are equal to sample values at the data locations. However, SGSIM is not useful for reproducing short-scale continuities (Lantuéjoul 1994), which causes numerical instability if the data covariance is constant over a short distance, and artifact discontinuities tend to appear in TBSIM results (Lantuéjoul 1994; Olea 1999) when the number of lines is few

(Gneiting 1999; Emery and Lantuéjoul 2006; Eze et al. 2018).

Although many studies have compared geostatistical methods (e.g., Iskandar et al. 2012; Paravarzar et al. 2015; Lu et al. 2016), there are only a few that specifically address situations under severely limited access to data. This situation is common in submarine resource exploration because of the technical difficulty and high costs associated with deep drilling of the seafloor. However, submarine resources have attracted considerable attention with the increasing demand for mineral resources. Seafloor massive sulfide (SMS) deposits are a seafloor mineral resource that are generally rich in base metals (Cu, Pb, Zn), but their distribution, configuration, and formation mechanism, especially sub-seafloor mineralization, remain poorly understood because considerably less survey data are available from drilling and geophysical prospecting compared with on-land areas.

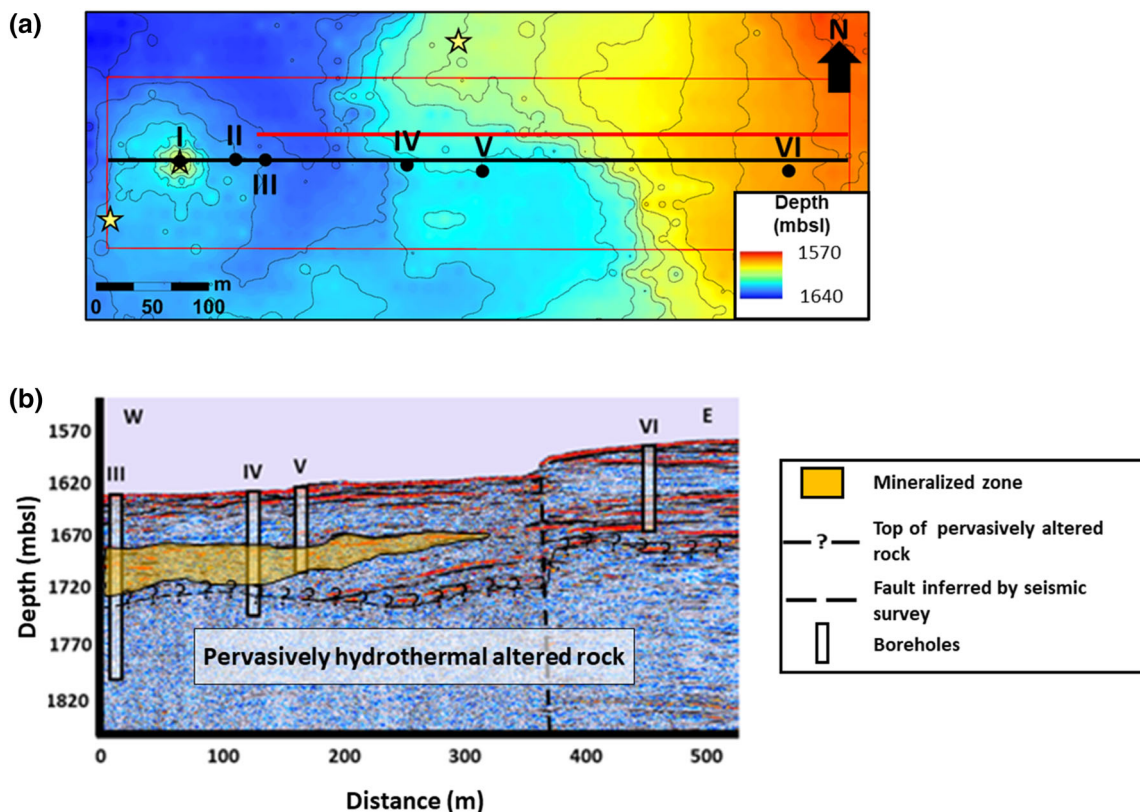
Framed by the background presented above, this study aims to clarify the metal content distribution and locate mineralized zones using a limited amount of metal content data from a few drilling sites. We use principal component analysis (PCA) and three geostatistical methods (OK, SGSIM, and TBSIM) to evaluate the best method for the sparse data scenario. To address global necessity, the method selection must be useful for reserve assessment and deposit modeling. An active seafloor vent area in the Okinawa Trough, southwest Japan, was selected as a case study. The deposits in the subduction-related back-arc setting (Pirajno 2009) are regarded as a modern analog of Kuroko-type volcanogenic massive sulfide (VMS) deposits on land (Halbach et al. 1989; Ishibashi et al. 2015).

## MATERIALS AND METHODS

### Geologic Setting

The study area is a part of a caldera floor in the middle Okinawa Trough (Fig. 1) formed by resurgent rhyolite domes and covered by up to 30-m-thick unconsolidated sediments including hemipelagic, silty Holocene clays, sulfide-bearing layers, tuff breccias, and pumice (Glasby and Notsu 2003; Ishibashi et al. 2015; Nozaki et al. 2018). Several normal faults are present in this area along the rift axis of the Okinawa Trough and trend E–W or ENE–WSW

## A Combination of Geostatistical Methods and Principal



**Figure 1.** (a) Seafloor topography of the study area (Hakurei Site, Izena Hole in the middle Okinawa Trough) overlain with locations of six drill sites (I–VI). The target region for geostatistical modeling is shown in the red rectangle. (b) A seismic profile along the red line in (a) with interpreted lithotypes, mineralized zone, and fault distribution by Asakawa and Lee (2018). The black line in (a) is the location of the cross section of the geostatistical results presented in Figure 6. The three stars in (a) represent the active chimney locations.

(Kato et al. 1989; Kato 1990; Halbach et al. 1993). The faults act as pathways for hydrothermal fluid flow (Pirajno 2009).

An area covering 700-m E–W × 130-m N–S at a depth of 1570 m below sea level (mbsl) (Fig. 1a) was chosen for geostatistical modeling, in which six drill sites (I–VI) are distributed nearly E–W. This area is termed the Hakurei Site, Izena Hole (Ishibashi et al. 2015; Totsuka et al. 2019). The westernmost drill hole (I) is located at a massive sulfide mound with a complex chimney structure. Based on the drill core observations, excluding the westmost sulfide mound, the top layer consists of poorly sorted primary and reworked underwater debris flow sediments with variable grain sizes mixed with volcanoclastic and hemipelagic sediments. The deeper portions consist of hydrothermal altered clay and altered volcanoclastic rocks (Nozaki et al. 2018). The basement presumably consists of intra-caldera ignimbrite with

a dacitic–rhyolitic composition with pervasive hydrothermal alteration (Nozaki et al. 2018; Yamasaki 2018).

The geologic structure underneath the seafloor was imaged from a seismic profile by Asakawa and Lee (2018) that covered a part of the study area from drill sites III to VI (Fig. 1b). The profile outlines two major structures: a stratabound layer with a high-velocity anomaly and fault development between drill sites V and VI from the basement toward the seafloor. The layered structure is interpreted as a concentration zone of polymetallic sulfides and hydrothermal alteration (Asakawa and Lee 2018; Nozaki et al. 2018). Such zones are characterized by high permeability and sulfide precipitation in the diffusive flow and cooling of hydrothermal fluids (Rona et al. 1993; Nozaki et al. 2018). The inferred fault may have been generated during the caldera formation (Halbach et al. 1993; Yamasaki 2018).

## Sample Data

Element contents in the vertical drill cores (ppm or wt%) were measured by inductively coupled plasma quadrupole mass spectrometry (ICP-QMS) following the HF–HNO<sub>3</sub>–HClO<sub>4</sub> acid digestion method (Takaya et al. 2018). In total, 448 samples with six elements (Zn, Pb, Cu, Ba, Ag, and Cd) were selected for geostatistical modeling. These are major elements of the dominant constituent sulfide/sulfate minerals of SMS deposits in the Okinawa Trough (Halbach et al. 1989, 1993; Ishibashi et al. 2015; Nozaki et al. 2016). The sample sizes collected onboard were typically 10 to 20 cm<sup>3</sup>. Approximately 50 mg of powder of each sample was used for the ICP-QMS analyses.

## Analytical Flow of Spatial Modeling

Mineralized zones were identified in the study area using 448 geochemical sample data points from the drill cores, implemented by four steps: pre-processing, PCA, normal score transformation, and spatial modeling by geostatistical estimation and simulations, as shown in the flowchart in Figure 2. The data at each sample point are multivariate and compositional. Because the content magnitude and variance differ substantially between the elements, we perform pre-processing to avoid generating spurious correlations between two elements by decreasing the bias of the content distribution. The centered log-ratio (clr) transformation by Aitchison (1986) was selected for this purpose because its suitability has been demonstrated in several previous case studies (e.g., Aitchison 2002; Pawlowsky-Glahn and Olea 2004; Pawlowsky-Glahn et al. 2011). In this method, the content data of a certain element,  $x_i$ , is divided by the geometric mean ( $g_m$ ) and then log-transformed as:

$$\text{clr}(x_i) = \ln \frac{x_i}{g_m} \quad \text{for } g_m = \left( \prod_{j=1}^m x_j \right)^{1/m} = \sqrt[m]{x_1 \dots x_m} \quad (1)$$

where  $m$  is the number of data points for the target element. The content data after the clr transformation are expressed by adding \*, e.g., Zn\* for the original Zn data.

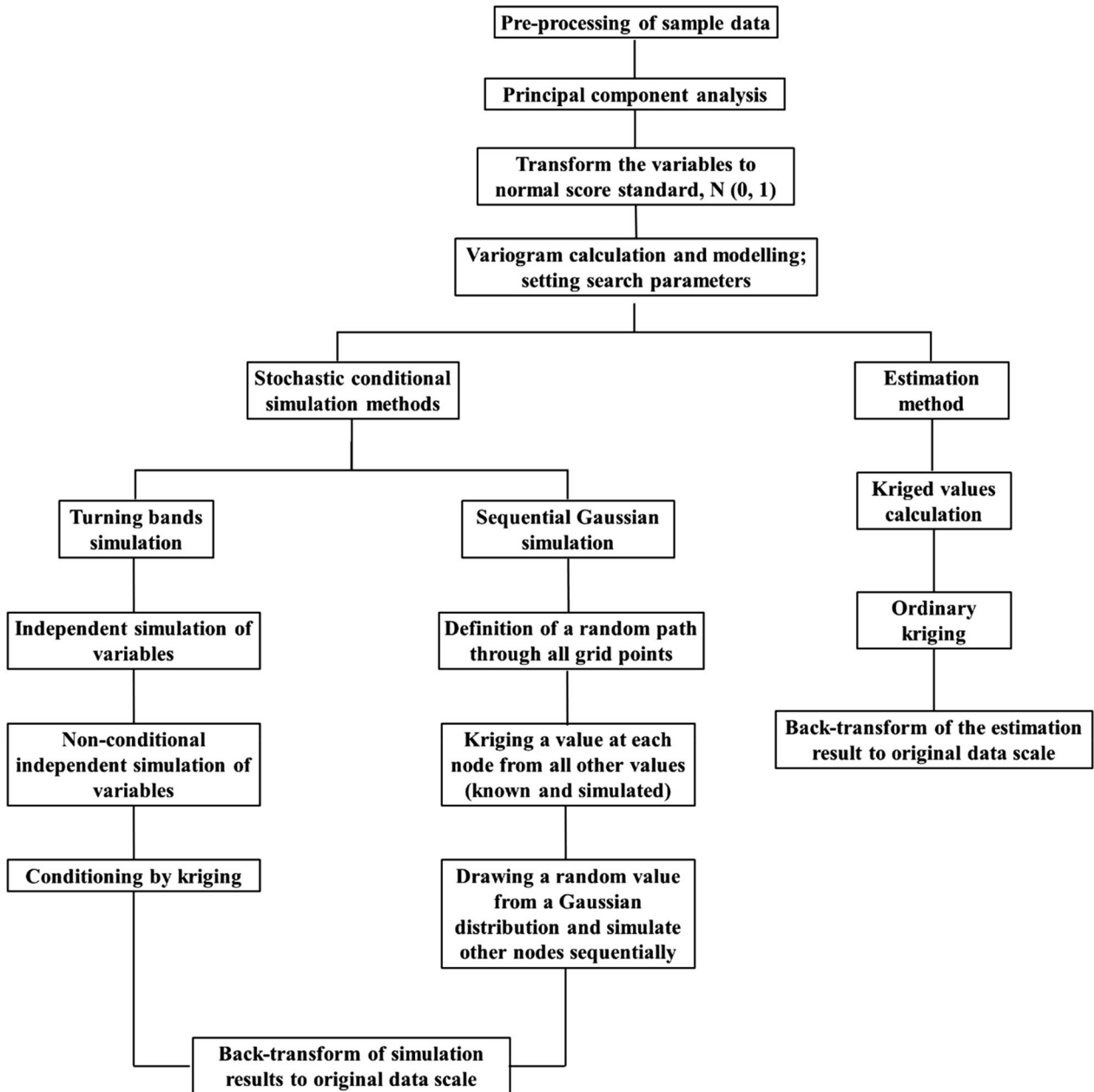
Because correlations generally exist between element contents in metal deposits, PCA was applied to decrease the dimensionality of the content data by linearly combining the correlated elements to yield lower-dimensional variables and principal components (PCs). The PCs were used for subsequent geostatistical analyses. PCA simplifies the geostatistical calculations by changing multivariate to univariate and facilitates the specification of mineralized zones. Each PC was then transformed into a normal score that follows a standard normal distribution with a mean of 0 and variance of 1. Because a dataset following a normal distribution is suitable to geostatistical analyses, this normal transformation is indispensable for the case that the data distribution is biased and far from a normal distribution.

For the actual geostatistical steps, the variography and principles, equations, and calculation procedure of OK, SGSIM, and TBSIM have been described in detail in several references (e.g., Isaaks and Srivastava 1989; Armstrong 1998; Deutsch 1998; Delfiner and Chilès 2012; Pyrcz and Deutsch 2014). TBSIM uses 1000 turning bands to avoid possible artifacts, as proposed by Emery (2008). By averaging 100 realizations for both SGSIM and TBSIM, e-type models of these simulation methods were produced. The estimated and simulated results obtained from the three methods were linearly back-transformed into the original data scale.

## Grid Setting

For the geostatistical modeling, the study area was gridded by voxels of a unit size of 10 m along the  $X$ -axis (E–W), 10 m along  $Y$ -axis (N–S), and 0.4 m along the depth direction,  $Z$ -axis. These sizes were determined by considering the average intervals of neighboring drill sites in the horizontal direction and neighboring sample points along the drill site in the vertical direction. A small vertical size was chosen to reveal small content changes. The bottom location of the longest borehole III was used to set the bottom boundary of the modeling domain. The estimation and simulation were point based, i.e., the calculations were implemented at the grid points.

## A Combination of Geostatistical Methods and Principal



**Figure 2.** Flowchart for clarification of mineralized zones using geochemical sample data from drill cores. The procedure consists of pre-processing, PCA, normal score transformation, and spatial modeling by geostatistical estimation and simulation.

## RESULTS

### Descriptive Statistics of the Content Data and PCs

The basic statistical features of the six selected elements are indicated by descriptive statistics (Table 1) and a correlation matrix between two ele-

ments (Table 2). Based on the median values, Zn, Pb, and Ba are the main enriched elements in the study area. Linear correlation coefficients ( $R_s$ ) in the correlation matrix reveal that four metals: Zn\*, Pb\*, Cu\*, and Cd\*, are correlated with one another with the strongest correlation between Zn\* and Pb\* ( $R = 0.94$ ). Ba\* is not strongly correlated with the

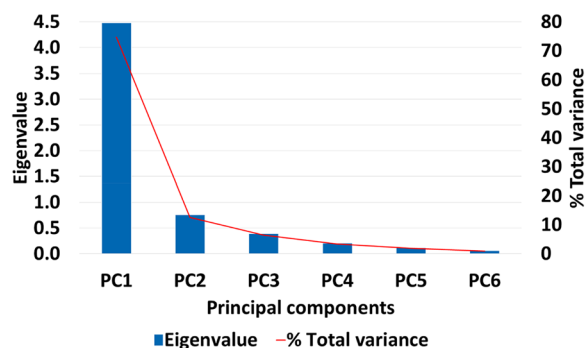


**Table 1.** Descriptive statistics of content data of six elements after the centered log-ratio transformation

Statistical parameters	Zn*	Pb*	Cu*	Ba*	Ag*	Cd*
Number of data	448	448	448	448	448	448
Mean	7.34	6.06	5.43	7.44	0.54	1.61
Variance	9.38	9.24	10.18	3.91	24.87	15.99
Minimum	2.91	1.16	− 22.91	1.63	− 24.05	− 22.49
Lower quartile	5.05	3.79	3.48	6.15	− 1.77	− 1.11
Median	6.11	4.93	4.16	6.97	0.36	0.42
Upper quartile	8.83	8.04	7.76	8.55	3.19	3.94
Maximum	17.24	16.73	14.94	13.31	10.95	12.17

**Table 2.** Correlation matrix of content data after the centered log-ratio transformation with correlation coefficients between two elements

	Zn*	Pb*	Cu*	Ba*	Ag*	Cd*
Zn*	1.00	−	−	−	−	−
Pb*	0.94	1.00	−	−	−	−
Cu*	0.85	0.84	1.00	−	−	−
Ba*	0.59	0.59	0.42	1.00	−	−
Ag*	0.65	0.74	0.64	0.32	1.00	−
Cd*	0.84	0.86	0.77	0.41	0.75	1.00

**Figure 3.** Eigenvalues of six principle components, PC1 to PC6 (red line), and percentage of each eigenvalue for the total variance (blue bars).

other five elements, and Ag\* has moderate correlations with the above four metals with  $R = 0.64$  to  $0.75$ . The existence of those moderate and high correlations among the target elements demonstrates the effectiveness of using PCA, as noted by Swan and Sandilands (1995).

The correlations cause particularly large eigenvalue for the first PC (PC1) in which 74.7% of the total variance is included (Fig. 3). PC2's variance is much smaller (12.6%), and the sum of PC3 to PC6

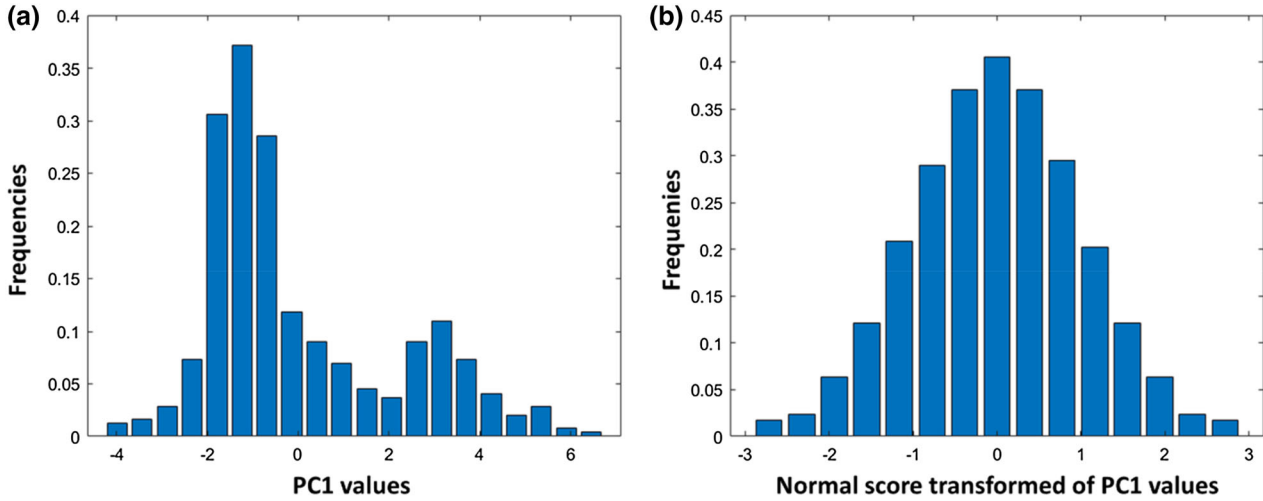
variances is 12.7%. This means that the higher PC1 value corresponds to a higher sum of the six element contents and high-PC1 zones can be indicative of sulfide mineralization zones. Consequently, a set of only the PC1 values was used for the subsequent geostatistical analyses. A histogram of the PC1 values shows two peaks in the low and high values (Fig. 4a). The lower peak in the high value suggests the formation of mineralized zones. This bimodality is also observed in the main base metals: Zn\*, Pb\*, and Cu\*. The histogram in Figure 4b verifies the correct transformation of PC1 into a standard normal distribution.

### Validation of Semi-variogram Model and Data Search Size

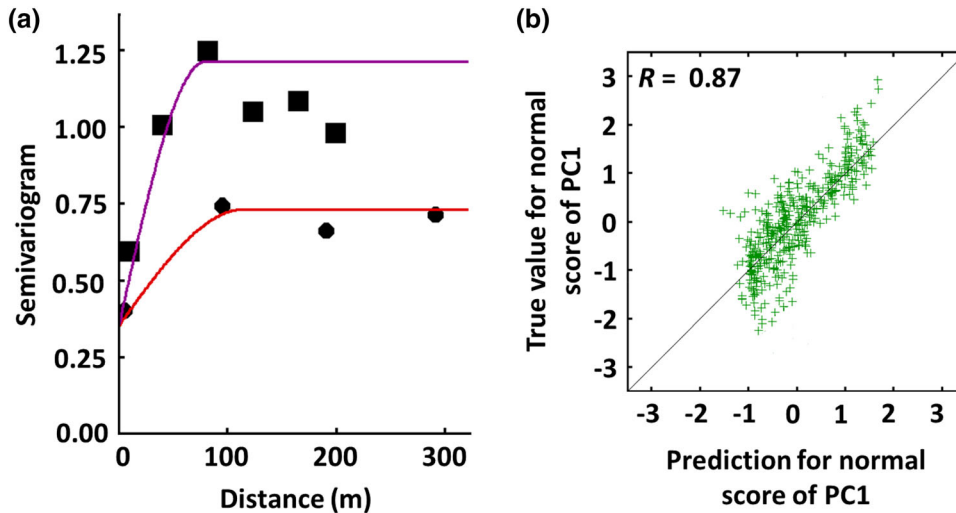
Because the six drill sites are distributed along an E–W line, it was impossible to detect anisotropic behavior of the semi-variogram in the horizontal direction. Accordingly, two experimental semi-variograms for the omnidirectional horizontal and vertical directions were produced and the spherical model was fitted to both the directions (Fig. 5a). The resulting semi-variogram models show geometrical anisotropy and derive ranges of 115 m and 79 m along the horizontal and vertical directions, respectively. Considering the ranges, the neighborhood search area for the three geostatistical methods was set to be an ellipse shape with sizes of 150 m in the horizontal direction and 100 m in the vertical direction. This search area was set to encompass at least the closest four data, following Isaaks and Srivastava (1989).

To check the suitability of the semi-variogram models and size of the neighborhood search, cross-validation was implemented using OK. The results

## A Combination of Geostatistical Methods and Principal



**Figure 4.** Comparison of two histograms of (a) original and (b) normal score transformed PC1 values.



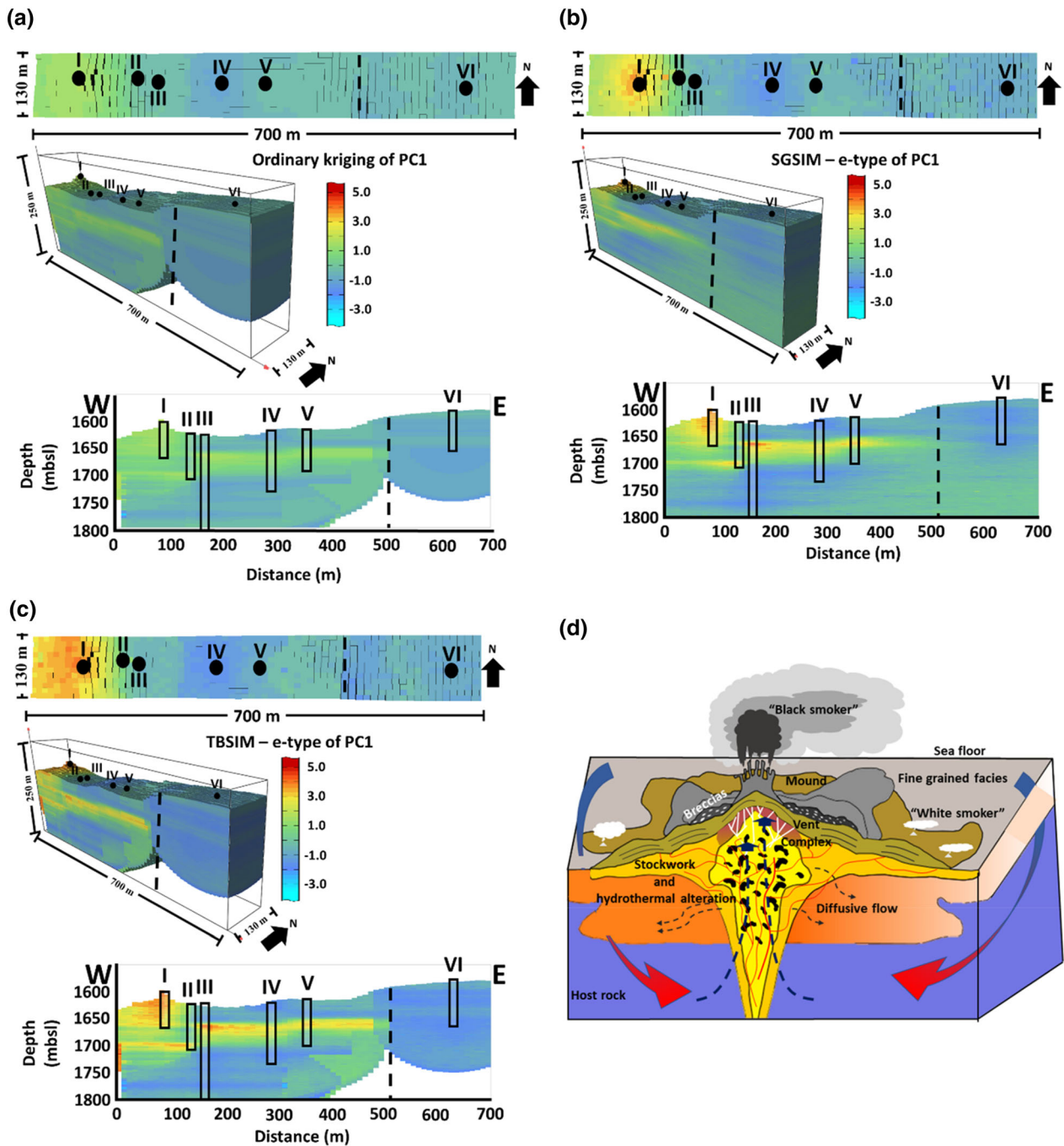
**Figure 5.** (a) Experimental semi-variograms along omnidirectional horizontal (circles) and vertical (squares) directions and their fitting to the spherical model as shown by the curves. (b) Cross-plot between the true and predicted PC1 normal scores by ordinary kriging, showing the accuracy of the kriging calculation.

show a cross-plot between the true and predicted PC1 values at each sample data point, indicating adequate prediction accuracy with  $R = 0.87$  (Fig. 5b), which demonstrates its suitability.

### Comparison of Spatial Models

Three PC1 spatial models generated through the OK and e-type of SGSIM and TBSIM are compared using the same color scale in Figure 6

from three viewpoints: the distribution on the sea-floor, perspective view from the southeast, and vertical E–W cross section along the black line in Figure 1a. The three models are similar and have common features with high-content (i.e., high-PC1) zones that extend around the sulfide mound and likely horizontal stratabound mineralization, consistent with the seismic profile interpretation in Figure 1b. The former shape appears as an upside-down ring and is likely a stockwork feature (Fig. 6d). However, smoothing effects appear in the OK



**Figure 6.** Comparison of spatial models of PC1 values using the same color scale from three views: distribution of the seafloor, perspective view from the southeast, and vertical cross section along the black line in Figure 1a. The models were produced by the geostatistical estimation method (a) OK and two simulation methods, (b) SGSIM, and (c) TBSIM. The average of 100 realizations is used as an e-type model of the two simulations. A schematic model of a seafloor massive sulfide system accompanying the sulfide mound is shown in (d) to illustrate the development of a stockwork under the sulfide mound and fluid flows (arrows) by compiling data from previous studies (Lydon 1988; Herzig and Hannington 1995; Ohmoto 1996; Tornos et al. 2015).



## A Combination of Geostatistical Methods and Principal

model for which the value range is much narrower than those from the SGSIM and TBSIM models. The mineralization zones are difficult to be specified in the OK model. In addition, the high-content zone is narrow and the stockwork feature is not entirely clear in the SGSIM model. One remarkable feature in the TBSIM model is that the stratabound mineralization is disconnected at the inferred fault (dashed line). This suggests that the fault acts as a path of downward recharge flow of seawater from the seafloor or upward discharge flow of hydrothermal fluids.

The simulation result quality is checked by comparing semi-variograms of each location with the semi-variogram model for both the horizontal and vertical directions (Fig. 7), as proposed by Jewbali and Dimitrakopoulo (2011). We use the normal score transformed PC1 data for this comparison. Ideally, the average of 100 semi-variograms should coincide with the semi-variogram model for both the directions, which proves an unbiased simulation result (Eze et al. 2018). Common to both the methods and directions, the variability of semi-variograms increases with separation distance. Owing to the paucity of sampled data particularly in the horizontal direction, the horizontal semi-variograms are largely variable at each realization and their averages are far from the semi-variogram models in both the SGSIM and TBSIM results (Fig. 7a, c). On the contrary, the average semi-variogram of the TBSIM results along the vertical direction approaches the semi-variogram model because of the substantially closer data intervals than in the horizontal direction. The TBSIM results show a more similar trend of the average semi-variogram to the model than the SGSIM results (Fig. 7b, d).

To check the calculation accuracy,  $R$  and the standard error of the three models are compared between the true and predicted PC1 values (Table 3). Although differences are small among the three models, TBSIM has the highest  $R$  and smallest standard error.

## DISCUSSION

Seafloor hydrothermal systems are comprised of a heat source (underlying magma chamber), recharge zone, circulation cell, and discharge zone that emits white and black smoke (Robb 2004). Because the heat source and circulation cells are located at 2–8 km depth underneath the seafloor (Pirajno 2009),

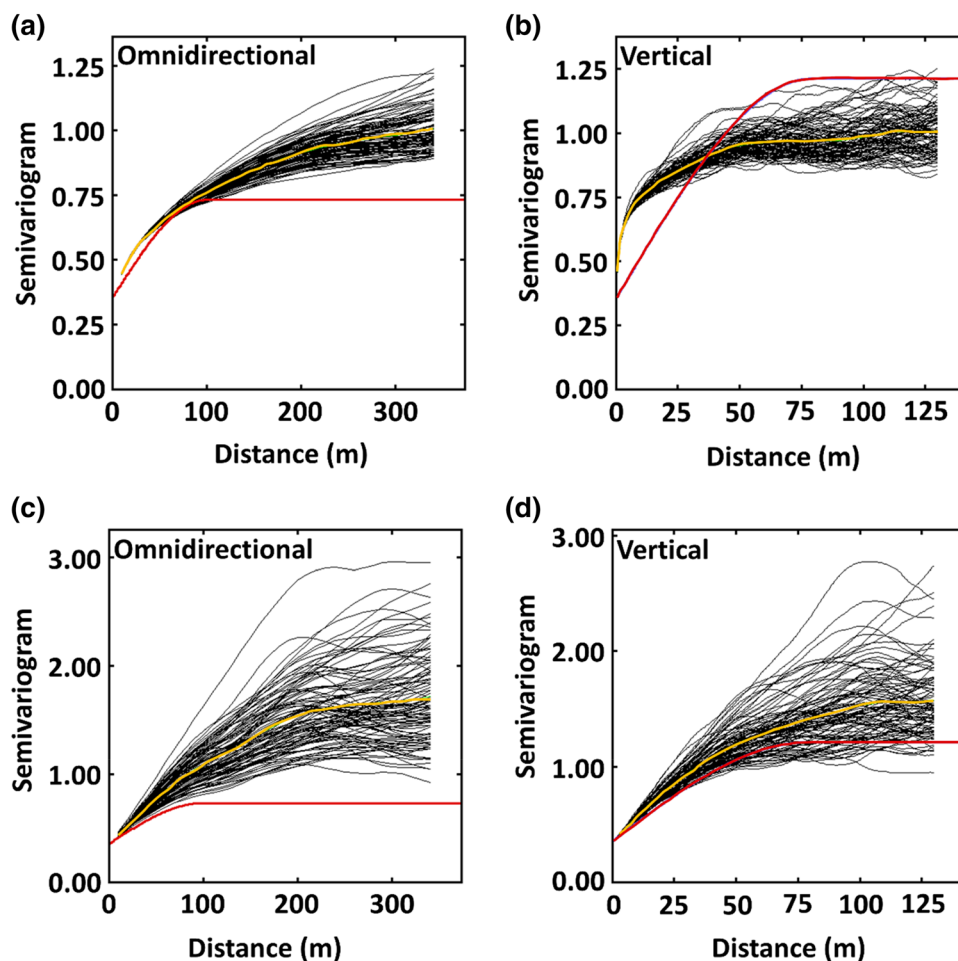
these features do not appear in geostatistical models constructed using the drill site data with a 180-m depth maximum. Instead, the models highlight mineralization features of a stockwork around the sulfide mound and of a stratabound feature on the eastern side of the mound.

The most plausible mechanism of this mineralization is mixing of hydrothermal fluids with cold ambient seawater in pore spaces in permeable strata (Shanks and Thurston 2012). Mineralization at two separated zones with different configurations and massive and stratified shape at several tens of meters below the seafloor was recently confirmed by a resistivity distribution from a deep-tow marine electric sounding at the Iheya North, middle Okinawa Trough (Ishizu et al. 2019). The geostatistical models are therefore geologically appropriate, and the combination of PCA, normal score transformation, and geostatistics for clarifying the mineralization features is effective.

As a comparison of the three geostatistical models, the mineralization shapes at the bottom of the sulfide mound and layer by OK are much smoother than those by SGSIM and TBSIM (Fig. 6a–c). This smoothing effect is revealed quantitatively by a comparison of the value range of the PC1 data,  $-4.26$  (minimum) to  $6.73$  (maximum). The ranges of the resultant three PC1 models are OK:  $-2.95$  to  $1.88$ , SGSIM:  $-2.54$  to  $4.95$ , and TBSIM:  $-3.73$  to  $5.29$ . Accordingly, the TBSIM best follows the variability of the sample values even under sparse data conditions in the horizontal direction, whereas OK induces a strong smoothing effect.

The three models are compared in Figure 8 by selecting PC1 zones  $> 1$  and focusing on the stockwork structure beneath the sulfide mound. The high-PC1 zones  $> 3$  colored by orange and red show the suggested mineralization zones. The stockwork structure does not appear in the OK model (Fig. 8a), and the high-PC1 zones are limited around the mound and near the seafloor and not vertically continuous in the SGSIM model. In contrast, the TBSIM model is best fitted to the stockwork conceptual model (Fig. 6d) and low-resistivity distribution (Ishizu et al. 2019).

The superiority of TBSIM over SGSIM under these conditions can be explained by the difference in neighboring data used for the kriging calculation at a certain voxel. SGSIM is a sequential algorithm that adds simulated values to the sample dataset for subsequent calculations, which reduces the distance



**Fig. 7.** Semi-variograms of realizations along the omnidirectional horizontal and vertical directions by SGSIM ((a) and (b)) and TBSIM ((c) and (d)) shown as black curves. Red and yellow curves represent the spherical models in Figure 5a and semi-variogram averages, respectively.

**Table 3.** Cross-validation results showing correlation coefficient ( $R$ ) and standard error between true and predicted PC1 values by OK and e-types of SGSIM and TBSIM

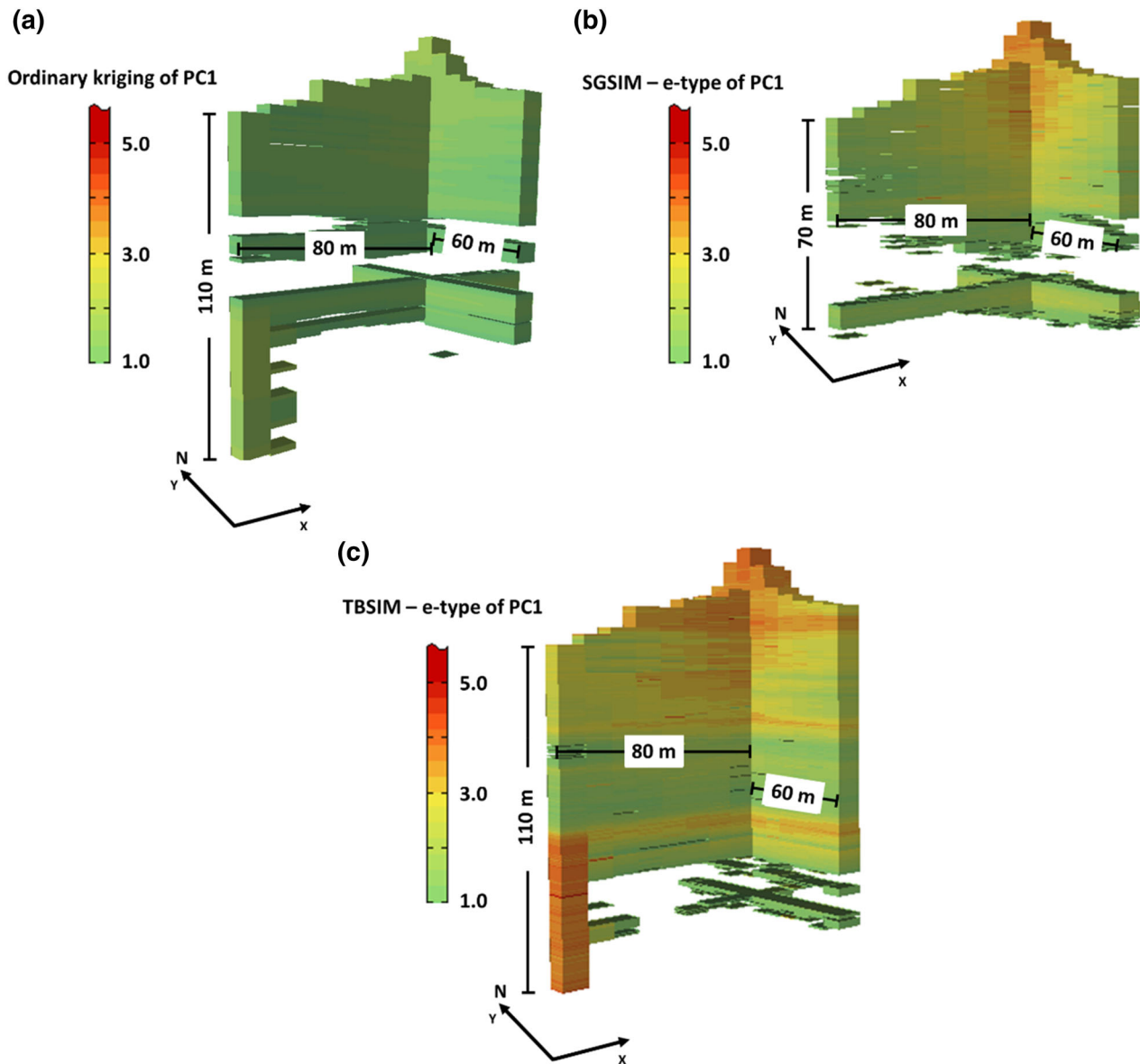
OK		SGSIM		TBSIM	
$R$	Standard error	$R$	Standard error	$R$	Standard error
0.87	1.03	0.88	1.00	0.89	0.99

to nearest neighbor data with and consequently induces a smoothing effect owing to the decrease in kriging variance (Delfiner and Chilès 2012). In contrast, TBSIM can preserve the variability of sample by using the sample data only for the simulation and setting many turning lines (Gneiting 1999; Emery 2008).

The upward shift of semi-variograms of all TBSIM realizations along the horizontal direction

(Fig. 7c) is caused by the particular sparsity of the data locations. In particular, large increases of the ranges from 115 m of the semi-variogram model to 300 m of the averaged semi-variogram seem inappropriate. This increase is owing to the layer structure of mineralization having such extent, whose plausibility is supported by the seismic profile in Figure 1d. The mismatch therefore does not neces-

## A Combination of Geostatistical Methods and Principal



**Figure 8.** Comparison of PC1 zones > 1.0 by (a) OK, (b) SGSIM, and (c) TBSIM using the same color scale to highlight the stockwork structure underneath the sulfide mound. These are parts of the 3D models in Figure 6.

sarily signify a defect of TBSIM, as suggested by Emery (2004).

## CONCLUSIONS

The aims of this study were to clarify the mineralization structure in an active seafloor vent area 1570 mbsl in the middle Okinawa Trough using 448 content data from only six drill sites. Under these

particularly sparse data conditions along the horizontal direction, three geostatistical methods (one estimation, OK, and two simulations, SGSIM and TBSIM) were compared by selecting the content data of six elements: Zn, Pb, Cu, Ba, Ag, and Cd, as typical elements of seafloor massive sulfide (SMS) deposits. Because these elements are strongly correlated, PCA was adopted to decrease the data dimensionality and the content information was consequently found to be condensed in the first

principal component (PC1). Through cross-validation between the true and predicted PC1 values by the three methods, TBSIM was identified as the best method for these data conditions by reducing the smoothing effect and reproducing the semi-variogram of the sampled data along the more densely data-distributed, vertical direction.

The most significant result obtained by the e-type of TBSIM was clarification of two mineralization zones with different configurations: a massive shape in the seafloor vicinity and a stratified shape at several tens of meters below the seafloor. These shapes are concordant with the resistivity distribution obtained by a deep-tow marine electric sounding. The massive shape similar to an upside-down ring is likely a stockwork. In addition, the stratabound mineralization feature also appears in the seismic profile. The most plausible mechanism of this mineralization is mixing of hydrothermal fluids with cold ambient seawater in pore spaces in permeable strata. Consequently, PCA and geostatistical simulations contribute to the interpretation and formation mechanism of SMS deposits and reserve assessment.

Our next step is geologic modeling using core description data and geostatistical simulations in combination with the content model to specify mineralized zones and their geologic features in more detail to determine their formation mechanism.

## ACKNOWLEDGMENTS

This work was supported by the Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), “Next-generation technology for ocean resources exploration” (Funding agency: Japan Agency for Marine-Earth Science and Technology, JAMSTEC). We thank captain, crew, and onboard members of the cruise CK16-05 (Exp. 909). Sincere thanks are extended to Glen Nwaila for valuable comments and suggestions that helped improve the clarity of the manuscript.

## REFERENCES

Aitchison, J. (1986). *The statistical analysis of compositional data, monographs on statistics and applied probability*. London: Chapman & Hall.

- Aitchison, J. (2002). Simplicial inference. In M. A. G. Viana & D. S. P. Richards (Eds.), *Algebraic methods in statistics and probability* (Vol. 287), Contemporary Mathematics Series Rhode Island: American Mathematical Society.
- Armstrong, M. (1998). *Basic linear geostatistics*. Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-58727-6>.
- Asakawa, E., Murakami, F., Tara, K., Shutaro, S., Tsukahara, H., & Lee, S. (2018). Multi-stage seismic survey for seafloor massive sulphide (SMS) exploration. In *Conference: OCEANS'18 MTS/IEEE Kobe/Techno-Ocean 2018*, Kobe, Japan.
- Battalagazy, N., & Madani, N. (2019). Categorization of mineral resources based on different geostatistical simulation algorithms: A case study from an iron ore deposit. *Natural Resources Research*, 28(4), 329–1351.
- Bowen, G. J. (2010). Isoscapes: Spatial pattern in isotopic biogeochemistry. *Annual Review of Earth and Planetary Sciences*, 38, 161–187.
- Calder, C. A., & Cressie, N. (2009). Kriging and variogram models. *International Encyclopedia of Human Geography*. <https://doi.org/10.1016/b978-008044910-4.00461-2>.
- Chilès, J. P., & Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty*. New York: Wiley.
- Chilès, J.-P., & Lantuéjoul, C. (2005). Prediction by conditional simulation: Models and algorithms. In M. Bilodeau, F. Meyer, & M. Schmitt (Eds.), *Space, structure and randomness: Contributions in honor of Georges Matheron in the field of geostatistics, random sets and mathematical morphology* (pp. 39–68). New York: Springer. <https://doi.org/10.1007/0-387-29115-6>.
- Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical software library and user's guide*. New York: Oxford University Press.
- Egozcue, J. J., & Pawlowsky-Glahn, V. (2011). Basic concepts and procedures. In V. Pawlowsky-Glahn & A. Buccianti (Eds.), *Compositional data analysis: Theory and applications* (pp. 12–28). Chichester: Wiley.
- Emery, X. (2004). Testing the correctness of the sequential algorithm for simulating Gaussian random fields. *Stochastic Environmental Research and Risk Assessment*, 18(6), 401–413.
- Emery, X. (2007). Conditioning simulations of Gaussian random fields by ordinary kriging. *Mathematical Geology*, 39(6), 607–623.
- Emery, X. (2008). A turning bands program for conditional co-simulation of cross-correlated Gaussian random fields. *Computers & Geosciences*, 34(12), 1850–1862.
- Emery, X., & Lantuéjoul, C. (2006). TBSIM: A computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method. *Computers & Geosciences*, 32(10), 1615–1628.
- Eze, P. N., Madani, N., & Adoko, A. C. (2018). Multivariate mapping of heavy metals spatial contamination in a Cu–Ni exploration field (Botswana) using turning bands co-simulation algorithm. *Natural Resources Research*, 28(1), 109–124.
- Glasby, G. P., & Notsu, K. (2003). Submarine hydrothermal mineralization in the Okinawa Trough, SW of Japan: An overview. *Ore Geology Reviews*, 23(3–4), 299–339.
- Gneiting, T. (1999). The correlation bias for two-dimensional simulations by turning bands. *Mathematical Geology*, 31(2), 195–211.
- Halbach, P., Nakamura, K.-I., Wahsner, M., Lange, J., Sakai, H., Käselitz, L., et al. (1989). Probable modern analogue of Kuroko-type massive sulphide deposits in the Okinawa Trough back-arc basin. *Nature*, 338, 496–499.
- Halbach, P., Pracejus, B., & Marten, A. (1993). Geology and mineralogy of massive sulfide ores from the Central Okinawa Trough, Japan. *Economic Geology*, 88, 2210–2225.



## A Combination of Geostatistical Methods and Principal

- Herzig, P. M., & Hannington, M. D. (1995). Polymetallic massive sulfides at the modern seafloor. A review. *Ore Geology Reviews*, 10, 95–115.
- Ilyas, A., Kashiwaya, K., & Koike, K. (2016). Ni grade distribution in laterite characterized from geostatistics, topography and the paleo-groundwater system in Sorowako, Indonesia. *Journal of Geochemical Exploration*, 165, 174–188.
- Isaaks, E. H., & Srivastava, R. M. (1989). *An introduction applied geostatistics*. New York: Oxford University Press.
- Ishibashi, J.-I., Ikegami, F., Tsuji, T., & Urube, T. (2015). Hydrothermal activity in Okinawa Trough back-arc basin: Geological background and hydrothermal mineralization. In J. I. Ishibashi, K. Okino, & M. Sunamura (Eds.), *Subseafloor biosphere linked to hydrothermal systems: TAIGA concept* (pp. 337–359). Tokyo: Springer. <https://doi.org/10.1007/978-4-431-54865-2>.
- Ishizu, K., Goto, T. N., Ohta, Y., Kasaya, T., Iwamoto, H., Vachiriatienchai, C., et al. (2019). Internal structure of a seafloor massive sulfide deposit by electrical resistivity tomography. *Okinawa Trough, Geophysical Research Letters*, 46, 11025–11035.
- Iskandar, I., Koike, K., & Sendjaja, P. (2012). Identifying groundwater arsenic contamination mechanisms in relation to arsenic concentrations in water and host rocks. *Environmental Earth Sciences*, 65, 2015–2026.
- Jewbali, A., & Dimitrakopoulos, R. (2011). Implementation of conditional simulation by successive residuals. *Computers & Geosciences*, 37(2), 129–142.
- Kato, Y. (1990). Geology and topography on the ridge south of the active hydrothermal deposits of the Izena Hole—the results of diving survey in 1989. *JAMSTEC Deep Sea Research*, 6, 27–31. **(in Japanese with English abstract)**.
- Kato, Y., Nakamura, K., Iwabuchi, Y., Hashimoto, J., & Kaneko, Y. (1989). Geology and topography in the Izena Hole of the middle Okinawa Trough—The results of diving surveys in 1987 and 1988. *JAMSTEC Deep Sea Research*, 5, 163–182. **(in Japanese with English abstract)**.
- Koike, K., Kubo, T., Liu, C., Masoud, A. A., Amano, K., Kurihara, A., et al. (2015). 3D geostatistical modeling of fracture system in a granitic massif to characterize hydraulic properties and fracture distribution. *Tectonophysics*, 660(7), 1–16.
- Lantuéjoul, C. (1994). Non conditional simulation of stationary isotropic multigaussian random functions. In M. Armstrong & P. A. Dowd (Eds.), *Geostatistical simulations* (pp. 147–177). Dordrecht: Springer.
- Lu, L., Kashiwaya, K., & Koike, K. (2016). Geostatistics-based regional characterization of groundwater chemistry in a sedimentary rock area with faulted setting. *Environment Earth Sciences*, 75, 829.
- Lydon, J. W. (1988). Volcanogenic massive sulphide deposits: I. A descriptive model. In: Roberts, R. G., & Sheenan, P. A. (Eds.), *Ore deposit models* (pp. 14–56), Geoscience Canada, Reprint Series 3.
- Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability*, 5(3), 439–468.
- Nozaki, T., Ishibashi, J.-I., Shimada, K., et al. (2016). Rapid growth of mineral deposits at artificial seafloor hydrothermal vents. *Scientific Reports*, 6, 22163.
- Nozaki, T., Takaya, Y., Nagase, T., Yamasaki, T., Ishibashi, J.-I., Kumagai, H., Maeda, L., & CK16-05 Cruise Members (2018). Subseafloor mineralization at the Izena Hole, Okinawa Trough from the aspect of drill cores obtain by the CK16-05 Cruise (Exp. 909). Conference: Goldshmidt2018, Boston, U.S.A.
- Ohmoto, H. H. (1996). Formation of volcanogenic massive sulfide deposits: the Kuroko perspective. *Ore Geology Reviews*, 10, 135–177.
- Olea, R. A. (1999). *Geostatistics for engineers and earth scientists*. Dordrecht: Kluwer.
- Paravarzar, S., Emery, X., & Madani, N. (2015). Comparing sequential Gaussian and turning bands algorithms for cosimulating grades in multi-element deposits. *Comptes Rendus - Geoscience*, 347(2), 84–93.
- Pawlowsky-Glahn, V., & Olea, R. A. (2004). *Geostatistical analysis of compositional data*, International Association for Mathematical Geosciences., Studies in mathematical geosciences New York: Oxford University Press.
- Pirajno, F. (2009). *Hydrothermal processes and mineral systems*. Berlin: Springer. <https://doi.org/10.1007/978-1-4020-8613-7>.
- Pugliese, A., Farmer, W. H., Castellarin, A., Archfield, S. A., & Vogel, R. M. (2016). Regional flow duration curves: Geostatistical techniques versus multivariate regression. *Advances in Water Resources*, 96, 11–22.
- Pyrzcz, M. J., & Deutsch, C. V. (2014). *Geostatistical reservoir modelling*. New York: Oxford University Press.
- Ren, W. (2005). Short note on conditioning turning bands realizations. Centre for Computational Geostatistics, *Report Five*, University of Alberta, 405-1–11.
- Robb, L. (2004). *Introduction to ore-forming processes*. Oxford: Blackwell Publishing.
- Rona, P. A., Hannington, M. D., Raman, C. V., Thompson, G., Tivey, M. K., Humphris, S. E., et al. (1993). Active and relict sea-floor hydrothermal mineralization at the TAG hydrothermal field, Mid Atlantic Ridge. *Economic Geology*, 88, 1989–2018.
- Shahbeik, S., Afzal, P., Moarefvand, P., & Qumarsy, M. (2014). Comparison between ordinary kriging (OK) and inverse distance weighted (IDW) based on estimation error. Case study: Dardevey iron ore deposit, NE Iran. *Arabian Journal of Geosciences*, 7(9), 3693–3704.
- Shanks, W. C. P., III, & Thurston, R. (2012). *Volcanogenic massive sulfide occurrence model*. U.S. Geological Survey Scientific Investigations Report 2010–5070–C.
- Swan, A. R. H., & Sandilands, M. (1995). *Introduction to geological data analysis*. Oxford: Blackwell Science.
- Takaya, Y., Yasukawa, K., Kawasaki, T., et al. (2018). The tremendous potential of deep-sea mud as a source of rare-earth elements. *Scientific Reports*, 8, 5763.
- Tornos, F., Peter, J. M., Allen, R., & Conde, C. (2015). Controls on the siting and style of volcanogenic massive sulphide deposits. *Ore Geology Reviews*, 68, 142–163.
- Totsuka, S., Shimada, K., Nozaki, T., Kimura, J. I., Chang, Q., & Ishibashi, J.-I. (2019). Pb isotope compositions of galena in hydrothermal deposits obtained by drillings from active hydrothermal fields in the middle Okinawa Trough determined by LA-MC-ICP-MS. *Chemical Geology*, 514, 90–104. <https://doi.org/10.1016/j.chemgeo.2019.03.024>.
- Yamasaki, T. (2018). The role of bimodal magmatism in seafloor massive sulfide (SMS) ore-forming systems at the middle Okinawa Trough, Japan. *Ocean Science Journal*, 53, 413–436.