Dear Referee:

Thanks very much for your great support and constructive suggestions with regard to our manuscript. These comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have made our best efforts to improve our paper very carefully following your comments and suggestions. Our point by point response to the comments are given below. We hope the revised manuscript will be acceptable to your requirements. If still there are concerns, we will be happy to take care once we hear from you.

Major improvements:

Comment 1: GSDNN is critical in enhancing the GSARNN model's fitting accuracy. This highlights the significance of obtaining the correct spatial correlation between sample points by taking into account both the dataset's local and global qualities. I'm simply wondering if using GSDNN with traditional approaches will result in better interpolation results.

[Response]: Thanks for your comment. The GSDNN unit plays an important role in GSARNN model, taking variation characteristics of geographical elements in different directions into account. However, since there is no recognized true value of generalized spatial distance for training process, applying GSDNN unit to traditional methods will make the calculation process impossible to carry on. Therefore, the GSDNN unit can only be embedded in the NN method and participates in its overall training and calculation process. In other words, the generalized spatial distance is determined by the spatial characteristics of the elements to be interpolated, owning a specific connotation based on specific context of spatial elements. Some additional explanations have been added in the last paragraph of Section 2.3.1.

"Note that since there is no recognized true value of generalized spatial distance for training process, the GSDNN unit can only be embedded in the neural network-based method and participates in its overall training and calculation process. In other words, the generalized spatial distance is determined by the spatial characteristics of the elements to be interpolated, owning a specific connotation based on specific

context of spatial elements."

Comment 2: Section 2.1.2 should go over the kriging approach in greater detail. Please explain how the weight coefficient $\boldsymbol{\lambda_i}$ is calculated.

[Response]: Thanks for your helpful advice. We have properly supplemented the interpolation calculation process of the Kriging method, including how the weight coefficient λ_i is calculated. However, since the focus of this paper is on the GSARNN model, the derivation processes of some formulas in Kriging method have been omitted to make the length of Section 2.1.2 not too long.

Kriging can be expressed as:

$$z^*(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i) \tag{2}$$

where $z^*(x_0)$ *is the predicted value, and* λ_i *and* $z(x_i)$ *are, respectively, the weight coefficient and observed value of point* i.

Kriging methods involve the calculation of the weight coefficient λ_i*, for which the key is to satisfy the unbiasedness and optimality. Unbiasedness means that* $z^*(x_0)$ *is the unbiased estimate of* $z(x_i)$*, that is:*

$$E[z^*(x_0) - z(x)] = 0 \tag{3}$$

which can derive the following constraints on λ_i*:*

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{4}$$

Optimality means that $z^*(x_0)$ *is the optimal estimate of* $z(x_i)$*, and the variance between the predicted value of the unsampled points and the estimated value of the observed points is the smallest, that is:*

$$\min_{\lambda_i} Var(z^*(x_0) - z(x)) \tag{5}$$

Define the cost function and try to figure out a set of weights λ_i *that satisfy unbiasedness and minimize the cost function. Finally, the following equation set can be derived:*

$$\begin{cases} \sum_{i=1}^{n} r_{ij}\lambda_i = r_{j0}, j = 1,2, \dots, n \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases} \tag{6}$$

where r_{ij} is the semi-variogram between point i and point j, which can be expressed as:

$$r_{ij} = \sigma^2 - Cov(z_i, z_j) = \frac{1}{2}E[(z_i - z_j)^2] \tag{7}$$

where σ^2 is the variance of $z(x)$, which is a constant in OK. r_{ij} can be simply determined by z_i and z_j. Kriging assumes that there is a functional relationship between r_{ij} and d_{ij} (the distance between point i and point j) and $\frac{n(n-1)}{2}$ (r, d) pairs can be generated by taking any two sampled points from the dataset. We can use linear, power, gaussian, spherical or exponential model to fit the relationship between r_{ij} and d_{ij}. Using the fitted function, we can calculate r_{j0} through d_{j0}. Thereby, the optimal weight set λ_i in Formula 6 can be solved.

Comment 3: The phrase "weight matrix" appears for the first time on Page 5, Line 6. Please provide some context.

[Response]: Thanks for your instructive advice. The logic of space weight description in the original manuscript is not coherent enough, which will make some readers confused. The overall presentation of this part (Section 2.2) has been properly reorganized to make it easier to understand.

It should be noted that there is a weight w_{ii} in the vector $\boldsymbol{w}_i$ which represents the spatial weight of point i to itself. To avoid overfitting, this weight should be set to 0:

$$w_{ij} = \begin{cases} f(d_{i1}^s, d_{i2}^s, \cdots, d_{in}^s)_j, & i \neq j \\ 0, & i = j \end{cases} \tag{9}$$

The spatial weights of all sampled point pairs can be expressed by an $n * n$ weight matrix $\boldsymbol{W}$. According to Formula 9, the weights on the diagonal of $\boldsymbol{W}$ should be reset to 0. Therefore, $\boldsymbol{W}$ can be defined as:

$$\boldsymbol{W} = \boldsymbol{\rho} * \boldsymbol{K} \tag{10}$$

where $\boldsymbol{\rho}$ *is the spatial weight component, and* $\boldsymbol{K}$ *is the standard weight component, which ensures that the neural network weight is independent of the point itself in the training process.* k_{ij} *in* $\boldsymbol{K}$ *can be expressed as:*

$$k_{ij} = \begin{cases} 1, & i \neq j \\ 0, & i = j \end{cases} \tag{11}$$

Comment 4: You say in section 2.3.3 that you utilize variable learning rate for network training and explain how it changes during the training process. You should also consider the benefits of this customized learning rate.

[Response]: Thanks for your very helpful advice. The learning rate starts from α_{start} and increases gradually at the rate of k_1 until α_{max}. A relatively small initial learning rate can prevent excessive fluctuation and convergence obstacle and the following increment of leaning rate can avoid the convergence rate at the early stage of the training process being too low. The maximum learning rate is maintained for n epochs. At this stage, the model can stably learn the spatial characteristics of the elements. The learning rate then gradually decreases exponentially at the rate of k_2, ensuring that the model can sufficiently converge near the optimal position. The description of the customized learning rate benefits has been added in Section 2.3.3, Paragraph 3, which makes this strategy more reasonable.

"where α_{start} *is the initial learning rate, which increases gradually at the rate of* k_1 *until* α_{max}*. A relatively small initial learning rate can prevent excessive fluctuation and convergence obstacle and the following increment of leaning rate can avoid the convergence rate at the early stage of the training process being too low. The maximum learning rate is maintained for* n *epochs. At this stage, the model can stably learn the spatial characteristics of the elements. The learning rate then gradually decreases exponentially at the rate of* k_2*, ensuring that the model can sufficiently converge near the optimal position. The change of the learning rate throughout the training process is shown in Fig. 3."*

Comment 5: The difference between all of these interpolation solutions is difficult to notice in Figure 11. I recommend graphing the difference between the interpolated and real values (interpolation error) and modifying the color scheme accordingly.

[Response]: Thanks for your very instructive advice. According to your recommendation, we graph the cross-validation interpolation results as well as the interpolation errors of the four methods as Figure 10 (Figure 11 in the original manuscript). The difference of the interpolation performances between the four methods can be more easily distinguished now. The caption and paragraph corresponding to this figure have also been appropriately revised.

"Figure 10 shows three-dimensional diagrams of the cross-validation interpolation results generated by the four methods and their corresponding interpolation errors. In interpolation error diagrams, red represents overestimation and blue represents underestimation. Taking the real dataset in Figure 9 as a reference, the four models restore the data features in most areas, which is consistent with the statistical indicator results, with small differences in some details. The IDW method evidently underestimates the temperature at shallow depths, which may be because its interpolation mechanism can produce large errors at the edge points of a given space. In the OK results, the coexistence of underestimation and overestimation around the sea surface is observed, indicating that the OK method also has some limitations in edge-area interpolation. The SARNN and GSARNN models slightly overestimate the temperature of bottom area. The error of GSARNN is generally smaller than that of SARNN. Further quantitative analysis is needed to elucidate more details of the interpolation experiment results."

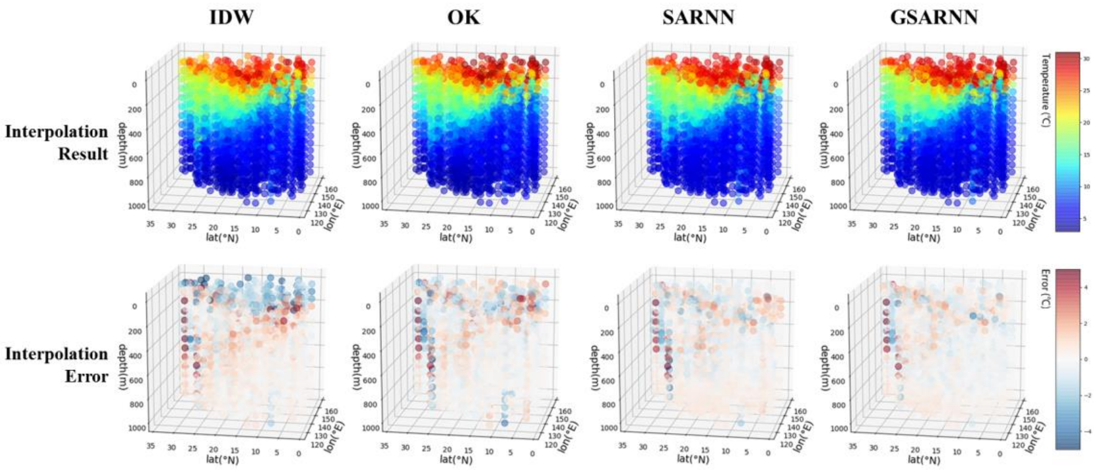


Figure 10. Three-dimensional diagrams of the cross-validation interpolation results and interpolation errors.

Comment 6: It is unclear what the x-axis signifies in Figures 8 and 12. Please include some text and figure descriptions.

[Response]: Thanks for your helpful advice. The x-axis in Figures 8 and Figure 12 represents the identifier number of each sampled point after they are sorted in ascending order of real value. Without an explanation, readers may feel confused. Some descriptions have been added in Figure 8 and Figure 12.

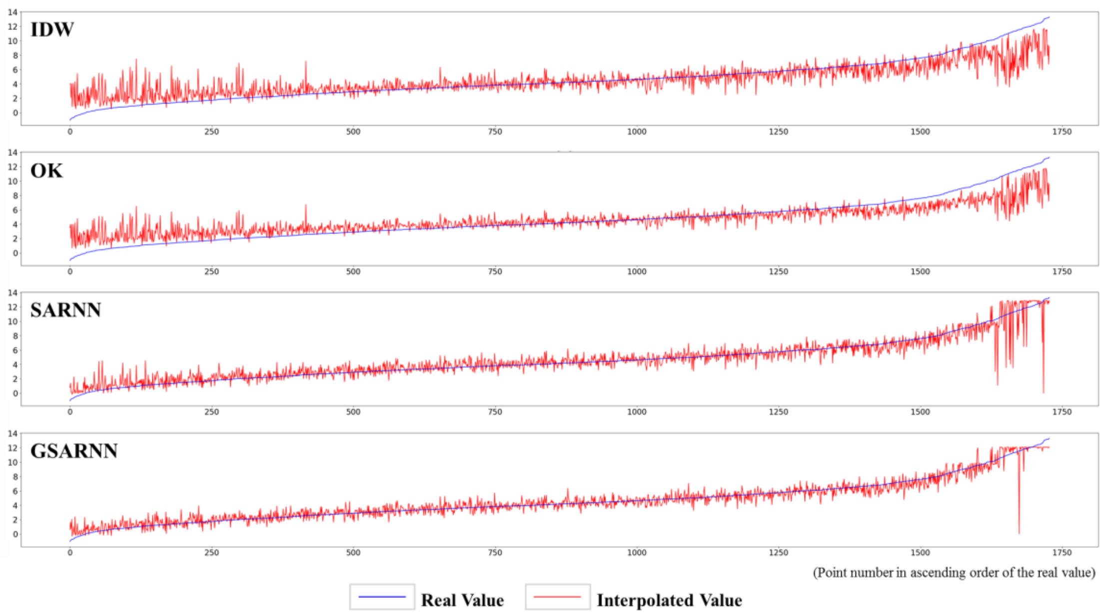
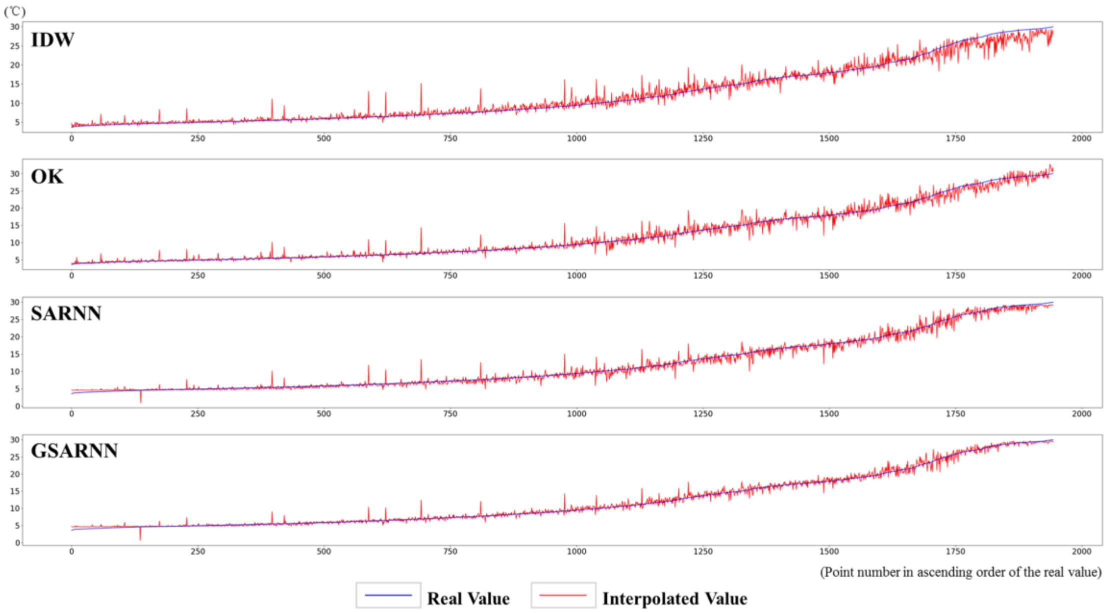
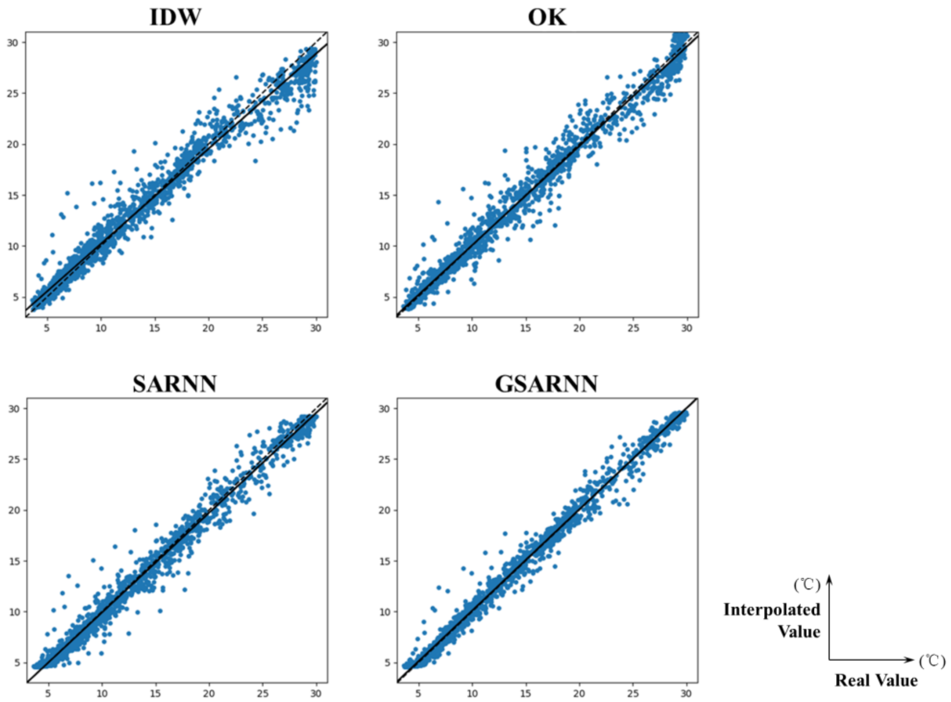


Figure 7. Line charts of real and interpolated values of the four models in Fig. 5.



(a)



(b)

Figure 11. The quantitative analysis results of the four models. (a) line charts in ascending order of real value; (b) scatter diagrams with trend lines.

Minor suggestions:

Comment 1: In section 3.1.1, from Line 19 to Line 22, I suggest not overly emphasizing the benefits of this experiment using the simulated data. This point has been mentioned in earlier paragraphs.

[Response]: Thanks for your helpful advice. Redundant statements have been removed. In order to keep Section 3.1.1 consistent with Section 3.2.1, a sentence is added at the end of Section 3.1.1.

"This case compares the interpolation abilities of the GSARNN model and the other three models in three-dimensional space using the simulated dataset above."

Comment 2: In Formula 10, since $\boldsymbol{k_{ij}}$ is 1 for the situation $\boldsymbol{i \neq j}$, then should the off-diagonal elements simply be written as $\boldsymbol{\rho_{ij}}$?

[Response]: Thanks for your helpful advice. The issue has been corrected.

$$\boldsymbol{W} = \boldsymbol{\rho} * \boldsymbol{K} = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{bmatrix} * \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 0 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 0 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 0 \end{bmatrix} \quad (13)$$

Comment 3: Multiline formulas, such as Formula 6, Formula 19, Formula 30, should be left aligned. The "int" in Formula 26 and Formula 27 seems to be redundant.

[Response]: Thanks for your helpful advice. Formula 6, Formula 19, Formula 30 have been corrected to be left aligned. The "int" in Formula 26 and Formula 27 is to take the integer part of the operation result, which is the coordinate of the point in the cube.

$$w_{ij} = \begin{cases} f(d_{i1}^s, d_{i2}^s, \cdots, d_{in}^s)_j, & i \neq j \\ 0, & i = j \end{cases} \quad (9)$$

$$\alpha = \begin{cases} \alpha_{start} + k_1 epoch, epoch < epoch_{up} \\ \alpha_{max}, & epoch \in [epoch_{up}, epoch_{down}] \\ k_2^{epoch} \alpha_{max}, & epoch > epoch_{down} \end{cases} \quad (22)$$

$$V_2 = \begin{cases} 1 - \frac{1}{36}(9-(3-x_i)^2)(9-(3-y_i)^2)(9-(3-z_i)^2), dist[(x_i, y_i, z_i), (3,3,3)] \in [1.5,3] \\ 1 + \frac{1}{36}(9-(3-x_i)^2)(9-(3-y_i)^2)(9-(3-z_i)^2), dist[(x_i, y_i, z_i), (3,3,3)] \notin [1.5,3] \end{cases} \quad (33)$$

Comment 4: Page 2, Abstract, Line 3: "which is one of the most important" to "a fundamental".

[Response]: Thanks. The issue has been corrected.

Comment 5: Page 2, Abstract, Line 10: "compared" to "compared with traditional methods".

[Response]: Thanks. The issue has been corrected.

Comment 6: Page 2, Line 17: "continuous data" to "continuous field".

[Response]: Thanks. The issue has been corrected.

Comment 7: Page 2, Line 32: "They" to "These methods".

[Response]: Thanks. The issue has been corrected.

Comment 8: Page 3, Line 7: "Zeng et al." to "In particular, Zeng et al.".

[Response]: Thanks. The issue has been corrected.

Comment 9: Page 3, Line 23: "by combining the GSDNN unit with the SARNN to integrate generalized distances into the spatial interpolation method, we developed" to "by combining the GSDNN unit with the SARNN, we integrated generalized distances into the spatial interpolation method and developed".

[Response]: Thanks. The issue has been corrected.

Comment 10: Page 4, Line 14: "deposit reserves" to "mineral deposit predication".

[Response]: Thanks. The issue has been corrected.

Comment 11: Page 6, Line 6: "speed" to "rate".

[Response]: Thanks. The issue has been corrected.

Comment 12: Page 9, Line 15: "the feature extraction and fitting ability of the GSARNN model are fully and persuasively tested" to "we can fully test the feature extraction and fitting ability of the GSARNN model".

[Response]: Thanks. The issue has been corrected.

Comment 13: Page 9, Line 17: “the most authentic” to “the authentic”.

[Response]: Thanks. The issue has been corrected.

Comment 14: Page 10, Line 4: “sudden change” to “sudden variation”.

[Response]: Thanks. The issue has been corrected.

Comment 15: Page 10, Line 17: “adds” to “imposes”.

[Response]: Thanks. The issue has been corrected.

Comment 16: Page 10, Line 18: “ε” to “The term ε”.

[Response]: Thanks. The issue has been corrected.

Comment 17: Page 11, Line 25: “has” to “achieves”.

[Response]: Thanks. The issue has been corrected. The similar issue in Section 3.2.3 has also been corrected.

Comment 18: Page 16, Line 21: “in low depth area” to “at shallow depths”.

[Response]: Thanks. The issue has been corrected.

Comment 19: Page 17, Line 5: “indicating that there may be individual outliers” to “indicating the presence of potential outliers”.

[Response]: Thanks. The issue has been corrected.