

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.

Comment 1: The authors only list the parameters of GSARNN model in detail in the manuscript. The configurations of traditional methods, such as the p value of IDW and the variation function adopted in Kriging, should also be mentioned in the comparison experiments.

[Response]: Thanks for your very helpful advice. The power parameter of IDW method is 4 in two cases. In Kriging method, we adopt the gaussian model to fit the functional relationship between the semi-variogram and the spatial distance, which turns out to be the optimal variation function model among linear, gaussian, spherical and exponential models. Some explanations have been added in Section 3.1.2, Paragraph 2.

“Besides, the power parameter of IDW method is 4, and in Kriging method, we adopt the gaussian model to fit the functional relationship between the semi-variogram and the spatial distance, which turns out to be the optimal variation function model among linear, gaussian, spherical and exponential models.”

Comment 2: The results of case 2 turn out that the neural network-based models generate smoother spatial patterns than traditional methods. I wonder if that is worth discussing.

[Response]: Thanks for your very helpful advice. As you mentioned, the interpolation results of neural network-based models exhibit smoother spatial patterns with less noise than those of traditional methods. This indicates that neural network-based

models can greatly reduce the influence of local extreme points on the points to be interpolated and acquire quite reasonable distributions of the geospatial elements through the non-linear fitting ability of neural networks. Some discussion has been added in Section 4 Discussion, Paragraph 2.

“In contrast, neural network-based models generate smoother interpolation surface than traditional methods. This indicates that neural network-based models can greatly reduce the influence of local extreme points on points to be interpolated and acquire quite reasonable spatial patterns of geospatial elements exploiting the non-linear fitting ability of neural networks.”

Comment 3: I think the point of how long it takes to run the model deserves more discussion in the manuscript. The authors briefly mention this as a limitation in the conclusion section, but some basic statistics on how long it takes would be a helpful addition.

[Response]: Thanks for your very insightful advice. As you mentioned, the model complexity of GSARNN is considerably higher than traditional methods. Nonetheless, compared with multifarious models in the fields of neural networks and deep learning, the structure of GSARNN with a few hidden layers is relatively lightweight, so its training and calculation efficiency can be quite high. The GSARNN model usually converges to the optimal state within 15-20 minutes in our cases since it can take advantage of mighty parallel computing capabilities of GPU units and distributed computing structures to accelerate the training process. Although the efficiency of Kriging method is better than GSARNN model, under the same condition, it still takes about 10 minutes to fit the functional relationship between the semi-variogram and the distance using “pykrige”. However, as the number of sampled points increases, the number of input neurons and output neurons of the GSARNN will also increase, resulting in the expansion of network parameters and the extension of training time inevitably. How to maintain a stable and acceptable training time given different sample data volumes is an important problem to be tackled in further researches.

Some discussions have been added in the end of Section 3.1.3 (an additional

paragraph) and Section 5 Conclusion, Paragraph 4.

“In addition, compared with multifarious models in the fields of deep learning, the structure of GSARNN is relatively lightweight, so its training and calculation efficiency can be quite high. Taking advantage of mighty parallel computing capabilities of GPU units and distributed computing structures to accelerate the training process, the GSARNN model usually converges to the optimal state within 15-20 minutes in our cases. Although the efficiency of Kriging method is better than GSARNN model, under the same condition, it still takes about 10 minutes to fit the functional relationship between the semi-variogram and the distance.”

“In addition, as the number of sampled points increases, the number of input neurons and output neurons of the GSARNN will also increase, resulting in the expansion of network parameters and the extension of training time inevitably. Therefore, how to maintain a stable and acceptable training time given different sample data volumes is an important problem to be tackled in further researches.”

Comment 4: Please express the information of GSDNN unit in Figure 2. Maybe Figure 1 is redundant and it can be merged into Figure 2.

[Response]: Thanks for your very instructive advice. The information of GSDNN unit is shown with GSARNN model structure in the same figure (Figure 1 in the revised manuscript) now. Due to the deletion of the original Figure 1, the numbers of subsequent figures and their related text have also been revised accordingly.

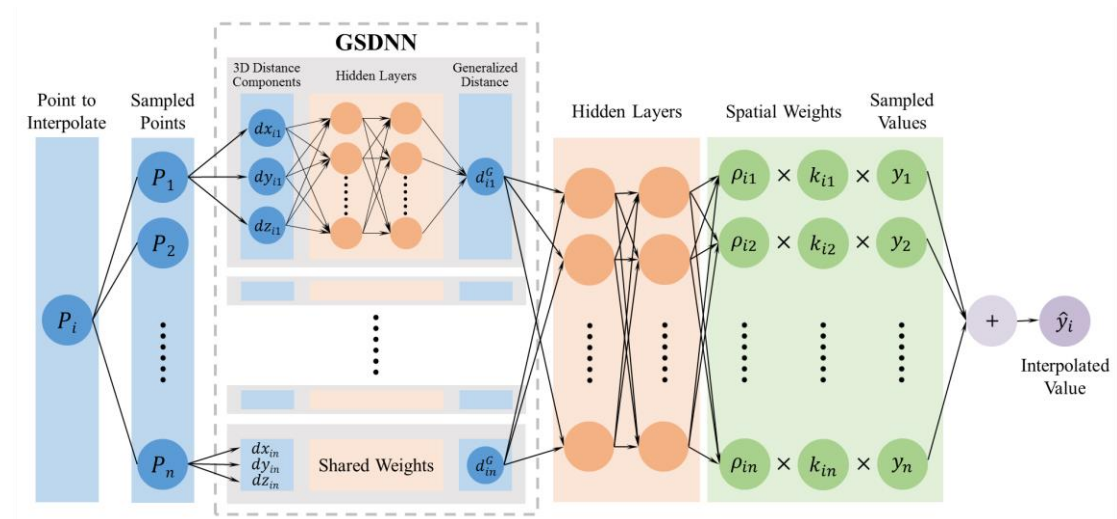


Figure 1. The GSARNN model structure.

Comment 5: In Figure 13, it would be better to make clear that the values in the left column represent depths below the sea surface.

[Response]: Thanks for your helpful advice. A description (“section depth”) of the values in the left column in Figure 13 (Figure 12 in the revised manuscript) has been added to avoid readers’ confusion.

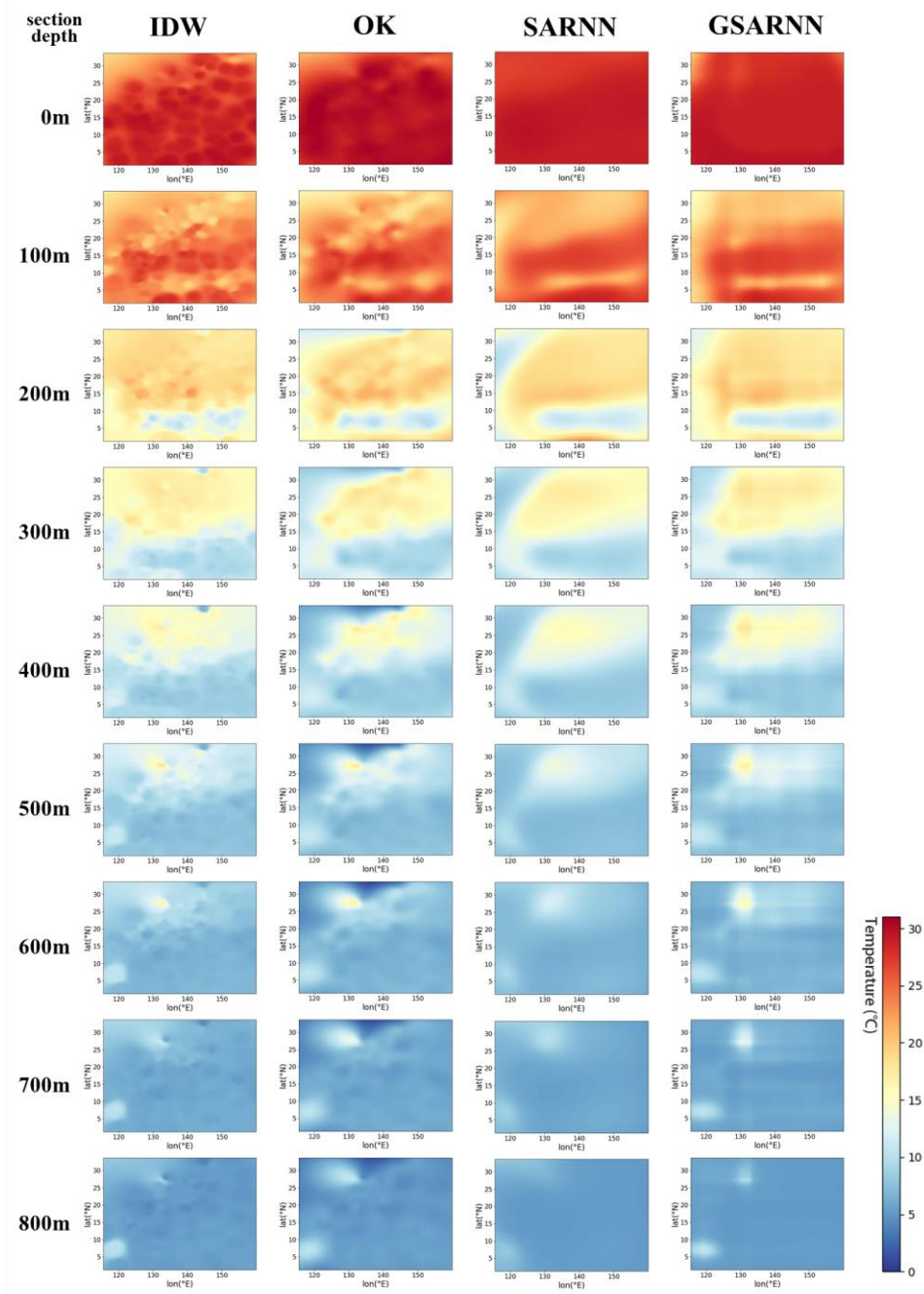


Figure 12. Comparisons of interpolated horizontal sections at 100 m depth intervals

generated by the four methods.

Comment 6: In Table 1 and Table 4, you'd better change 'Hyperparameters' to 'Hyper-parameters'.

[Response]: Thanks for your advice. All the "hyperparameters" in the manuscript have been revised to "hyper-parameters".

Comment 7: When a matrix or a vector is represented by a word or a character, it should be written in bold, such as in Formula 10.

[Response]: Thanks for your very helpful advice. All the matrices and vectors in formulas and paragraphs have been revised to bold. In addition, matrices are uniformly represented by upper case letters, and vectors are represented by lower case letters.