Author Reply to RC1 of essd-2024-137

We are truly grateful to Referee #1 for reviewing our manuscript and offering constructive feedback. These insightful comments played a key role in improving the quality of our work. We sincerely appreciate your patience and valuable thinking. As suggested by the comments, we revised our manuscript and we hope the following responses may adequately address the questions raised in the review.

Q1: The English needs improvement, e.g., page 2, line 19 "due to the limited data density", what's the meaning of data density here? The terminology should be precise, e.g., page 2, line 31, kinetic downscaling should be dynamical downscaling.

A1: "Data density" here means the distributions of in-situ stations (such as gauge), which mainly contain observation value. The distribution of stations cannot be unlimited dense, contributing to hardly producing a high-resolution dataset, especially before the 2000s. Following is the revised description in the manuscript:

> ***Lines 36-44 in the revised version:***
>
> *"However, the distribution of in-situ stations is too sparse to produce a high-quality reanalysis dataset, especially for decades ago."*

Thanks a lot for pointing out the usage of *"dynamical downscaling"*, we have revised our manuscript and further polished the paper writing to avoid the same problems. The incorrect usage of *"kinetic downscaling"* in the whole manuscript has been corrected to *"dynamical downscaling"* in ***lines 36, 41, and 43*** of the revised manuscript.

Q2: Page 2, line 32, "While the existing downscaling methods could produce high-resolution results, the results are unsatisfactory and unable to reconstruct detail and texture information." Please give the support and references for this statement about the limitations of existing methods.

A2: Thanks for the insightful suggestions, as suggested by the reviewer, more details about the related works are introduced in the manuscript:

> ***Lines 36-44 in the revised version:***
>
> *"Dynamical downscaling methods are usually based on Regional Climate Models (RCMs) under the guidance of the initial fields produced by Global Climate Models (GCMs). Although the resolution of RCMs is higher than GCMs, the comprehensible ability to understand the real world is not enough. It leads to a considerable bias (Teutschbein and Seibert, 2012) due to the difficulty of establishing simulation equations that cannot meet the needs of various related tasks. From another opinion, the computational cost of RCMs is huge, and it is an obstacle to producing a wider range of results (Giorgi and Gutowski Jr, 2015; Di Luca et al., 2015). Compared with dynamical downscaling, statistical downscaling maps the relationship between high-resolution and low-resolution from historical data to produce results. The computational cost and bias of statistical downscaling are lower than dynamical downscaling methods."*
>
> ***Lines 53-57 in the revised version:***
>
> *"Super-resolution tasks are similar to geographic downscaling tasks, using SR deep learning*

methods to downscale the geographic data can effectively avoid the problems encountered by former downscaling methods, such as high biases, regional sensitivity, high computational cost, etc. Deep learning methods use deep layers to bridge the relationship from low-resolution to high-resolution data and have robustness against the sensitivity. The computational cost is very low once the model has been trained, during the using step, and the deep learning method can nest a wide range easily."

Q3: The literature review in the introduction should be more comprehensive. The current state of research has not been adequately presented. The authors are suggested to highlight the limitations of existing methods rather than merely listing several studies. For example, the authors list several works using the Transformer architecture without presenting their relationship to the work in this paper.

A3: The valuable comments are meaningful to improve our work, we will enhance the article following the suggestions. Existing advanced deep learning methods of meteorological downscaling mostly used attention architecture (Shen et al., 2023; Liu et al., 2023; Zhong et al., 2023;). However, now existing deep learning downscaling methods only focus on one or two meteorological variables, while different variables have correlations and deep learning could handle multiple variables simultaneously. After that, it's worth noticing that there are no models that can cover a long-term and wide range of historical multiple variables due to the limitation of the training dataset. Our work merged geopotential into the transformer block to enhance the geographic information for multiple meteorological variables not only saving the computational resources but also improving the performance of the model. And we produce a long-term historical dataset to fill the gap in high-resolution historical meteorological datasets. For more introduction of the related work and the limitations, we revised the manuscripts following the suggestions from the reviewer, the revised version is as follows:

Lines 23-30 in the revised version:

"He et al. (2020) produced a meteorological dataset with a spatial resolution of 0.1° from 1979 in China. In this paper, the China Meteorological Forcing Dataset was proposed by fusing remote sensing products, reanalysis datasets, and in-situ station data. The most significant contribution of this work was using a larger number of stations to raise the quality of the dataset. A long-term gridded daily meteorological dataset for northwestern North America was proposed by Werner et al. (2019). The authors try to produce a dataset for training statistical downscaling schemes in Canada. The same in Italian, Bonanno et al. (2019) proposed the high-resolution meteorological dataset named MERIDA. MERIDA was produced by dynamical downscaling from the fifth-generation reanalysis dataset for the global climate and weather (ERA5) using WRF."

Lines 48-71 in the revised version:

"Liang et al. (2021) proposed SwinIR and achieved impressive results in the SR task and be considered the benchmark for the SR task. The core algorithm of SwinIR uses no overlap windows to split the input feature to calculate the attention relationship inner each window and shift the windows by the step of the half-width of the windows. Song and Zhong (2022) proposed a novel network to harvest long-range information from global instead of inner the window. The experimental results on SR benchmarks (Bevilacqua et al., 2012; Martin et al., 2001; Huang et al., 2015; Matsui et al., 2017) show this strategy can achieve better results. Super-resolution tasks are similar to geographic downscaling tasks, using SR deep learning methods to downscale the

geographic data can effectively avoid the problems encountered by former downscaling methods, such as high biases, regional sensitivity, high computational cost, etc. Deep learning methods use deep layers to bridge the relationship from low-resolution to high-resolution data and have robustness against the sensitivity. The computational cost is very low once the model has been trained, during the using step, and the deep learning method can nest a wide range easily. Shen et al. (2023) proposed a near-surface air temperature downscaling network SNCA-CLDASSD. In this model, Shen, et al. used two attention blocks to downscale the input data called Cross-Attention based on Light-CLDASSD. However, only near-surface air temperature is considered in this work and the network was built on CLDAS, which cannot cover long-term years. Liu et al. (2023) used the terrain to guide the deep learning network for the downscaling task called terrain-guided attention network (TGAN) in Southwest China. TGAN used the digital elevation model (DEM) to build high-resolution temperature (at 2 meters) results. The range of TGAN used begins in 2018 and TGAN cannot be used in the historical situation. Zhong et al. (2023) proposed a transformer-based learning method Uformer, which directly adds topography data, to achieve high-resolution meteorological variables in inner Mongolia province, China. Although topography data can help rebuild the high-resolution, directly adding into the input low-resolution will lose the characters of topography. All of the above, existing advanced deep learning methods of meteorological downscaling mostly used attention architecture (Transformer is one of the special attention architectures). However, now existing methods focus on one or two meteorological variables, while different variables have correlations and deep learning could handle multiple variables simultaneously. This way, not only saves the computational resources but also improves the performance of the model. Last and most important, there are no models that can cover a long-term and wide range of historical multiple variables."

Q4: The authors are suggested to present the motivations of using the variables of temperature at 2m, pressure at the surface, and wind speed at 10m for GeoAN in this work.

A4: Thanks a lot for the suggestions. Nowadays, existing methods usually utilize W10m or T2m as the downscaling variables (Shen et al., 2023; Liu et al., 2023; Zhong et al., 2023;). We compared the variables between CLDAS and ERA5, and chose the shared variables in both datasets including temperature at 2m, pressure at the surface, wind speed at 10m, and precipitation. We actually used the precipitation in the network first, but there are several reasons we have to abandon this variable even if the precipitation is very valuable for use.

Point 1: The evaluation of precipitation between CLDAS and ERA5 showed the low correlation is the main reason we discarded the precipitation and used the rest three variables. In future work, we will try to solve the problems of downscaling the precipitation.

Point 2: The precipitation is a rapidly changing variable comparing the pressure or temperature. This led to deep learning or other statistical methods very hard to harvest the rules from the training data. Thus, the results will be not very good because of the irregular situation.

Point 3: Precipitation is more like an extreme event, and deep learning is lean to produce an average result. Therefore, the results of the deep learning is hard to restore the value from the low-resolution input.

The related discussion can be found in the revised manuscript:

Lines 96-99 in the revised version:

"There are four meteorological variables, temperature at 2m, pressure at the surface, wind speed

at 10m, and daily total precipitation (TP) considered in GeoAN. Considering it is hard to process the downscale of TP, only three other variables are produced by GeoAN in MDG625."

Lines 217-221 in the revised version:

"The dataset referenced in MDG625 includes three meteorological variables (PRS, T2M, and WS10m), but it lacks precipitation data. The low correlation of total precipitation (TP) between the ERA5 and CLDAS datasets is the primary reason why the GeoAN model struggles to accurately restore the spatial structure of precipitation. Inspired by the work of (Zou et al., 2024), more comprehensible algorithms could be implemented in GeoAN in the future"

Q5: Page 6, line 98, "GeoAB is repeated 18 times", What considerations are there for setting it to 18?

A5: The repeated times is an adjustable hyperparameter. The reasons we set it to 18 are listed in the following discussion.

Point 1: Deep learning methods commonly use repeated blocks to reach a deep network for harvesting deep information, and that's why it is called "deep learning". In this theory, the deeper the network is, the more performance it will have. However, when the network is too deep, the chain rule will cause exploding or vanishing gradient problems. The computational expense of a deeper network is huge and the training step will last several weeks or months. Thus, the repeated times can not be unlimited deep, and it is constricted by the hardware.

Point 2: The data is the soul of the deep learning methods. More training data can produce better performance and emergence. This performance is based on a large model and enough data. If the data is not enough for the large model, it will damage the performance and the general practice is using a smaller model. Thus, the repeated times are not allowed too big.

We last chose 18 as our repeating times, and it served as the result of balancing the computational resources and the depth of the network. If we choose a deeper network, our GPUs (4 RTX 6000 Ada GPU 48G) can't afford the computational expense and may cause exploding or vanishing gradient problems. From another opinion, considering the training data, 18 is a reasonable setup. We add the above reasons of setting the repeated times to 18 in the caption in Fig. 1 in the manuscript.

Caption in Fig. 1 in the revised version:

"The results of the two blocks of head in the diagram have the same channels of 108. GeoAB, which is repeated 18 times constricted by the hardware and the data amount, is the attention block for extracting deep information using geopotential."

Q6: Page 6, line 105, "GUP memory limitation" should be "GPU memory limitation"?

A6: Sorry about this typo, and we thank you for your attention to the detail, it should be *"GPU memory limitation"*. We have revised this mistake in the manuscript and checked the whole paper to avoid the same mistake.

Q7: In addition to UNet and SwinIR, the authors are suggested to compared the results of the proposed method with more existing representative deep learning based downscaling methods.

A7: Your suggestion is insightful for us to improve this work. As to the questions in the discussion, here are our replies. As far as we know, our work is the first attempt to establish the mapping relationship from ERA5 to CLDAS to make a long-term daily meteorological dataset. Thus, there are no existing works that can be compared directly. Downscaling is a similar task to super-resolution in computer vision, so we choose three super-resolution methods migrating to this task to make a comparison:

1) Bilinear is the most classic and commonly used algorithm to improve resolution.

2) U-net has been one of the most used deep learning methods in recent years in various tasks for its strong universality.

3) SwinIR, proposed in 2021, was a novel SOTA (state-of-the-art) super-resolution model based on a transformer block. It can be considered the benchmark and representative deep learning method of SR-related missions.

To migrate the methods to downscale the long-term meteorological dataset from ERA5 to CLDAS, each existing method has to be redesigned. These methods may take several weeks to months to retrain to be compatible with the task. The above discussion can be found in a new section "Limitations" in the revised version:

Line 221 in the revised version:

"We will try to compare a wider range of models to evaluate our performance."

We redesign the deep learning networks for similar parameter quantities or GPU costs. We keep the hyperparameters for a fair comparison during the training and testing step. The details of the redesign are listed as follows:

Lines 150-156 in the revised version:

"To ensure a fair comparison, the U-Net architecture was modified for the downscaling task, resulting in a customized version referred to as U-Net Evolution (U-Net Evol.). The original U-Net implementation is available at https://github.com/milesial/Pytorch-UNet, while the SwinIR code can be accessed at https://github.com/JingyunLiang/SwinIR. To maintain consistency, all deep learning models were configured with equivalent parameters or computational complexity, and they were trained for 100 epochs using identical hyperparameters under the same environmental conditions."

In the future, we will continue to make larger comparisons with more existing representative deep learning methods and delve into how these deep learning methods can dig geographic information to help downscale the variables.

Q8: The proposed method has been used to generate a long-term dataset; however, the validation dataset in this paper is quite limited. It is suggested to evaluate the accuracy and reconstruction quality in terms of PSNR and SSIM using a larger dataset.

A8: Thank you so much for the insightful suggestions. More evaluation data could produce a more convincing result. We will try to do a more comprehensive comparison in future work. However, there are some considerations about how many years we should use to evaluate this work we have considered.

Table AC1-1. The data range of different deep learning downscaling methods is used.

	Training dataset	Testing dataset
Zhong et al. (2023)	2020.01 – 2021.05	2021.06 – 2021.09
Liu et al. (2023)	2019.01 – 2019.12	2018
Shen et al. (2023)	90% of 2016, 2017, 2019, 2020	2018

Point 1: Due to many researchers (Zhong et al. (2023); Liu et al. (2023); Shen et al. (2023)), as shown in **Tab. AC1-1**, only using the data in one or two years for downscaling one or two meteorological parameters and one or fewer years for evaluation, we finally decided to use 3 years for training and 1 year to evaluate the model.

Point 2: The amount of the total data is fixed and if we use more data to evaluate the model, less data can be used in training. For a better performance, we use 3 years for training and 1 year for evaluation.

Point 3: The validation dataset needs to cover one full cycle, for the meteorological task is one year. Although there are slight distributing differences between different years, using a whole year of data for validation is enough to cover most situations.

Point 4: The performance of training the deep learning can be obtained from the training log (If the model has converged). The training log is shown in **Fig. AC1-1**. We used the pre-trained model to train the GeoAN and we can find the network convergence is relatively optimistic.

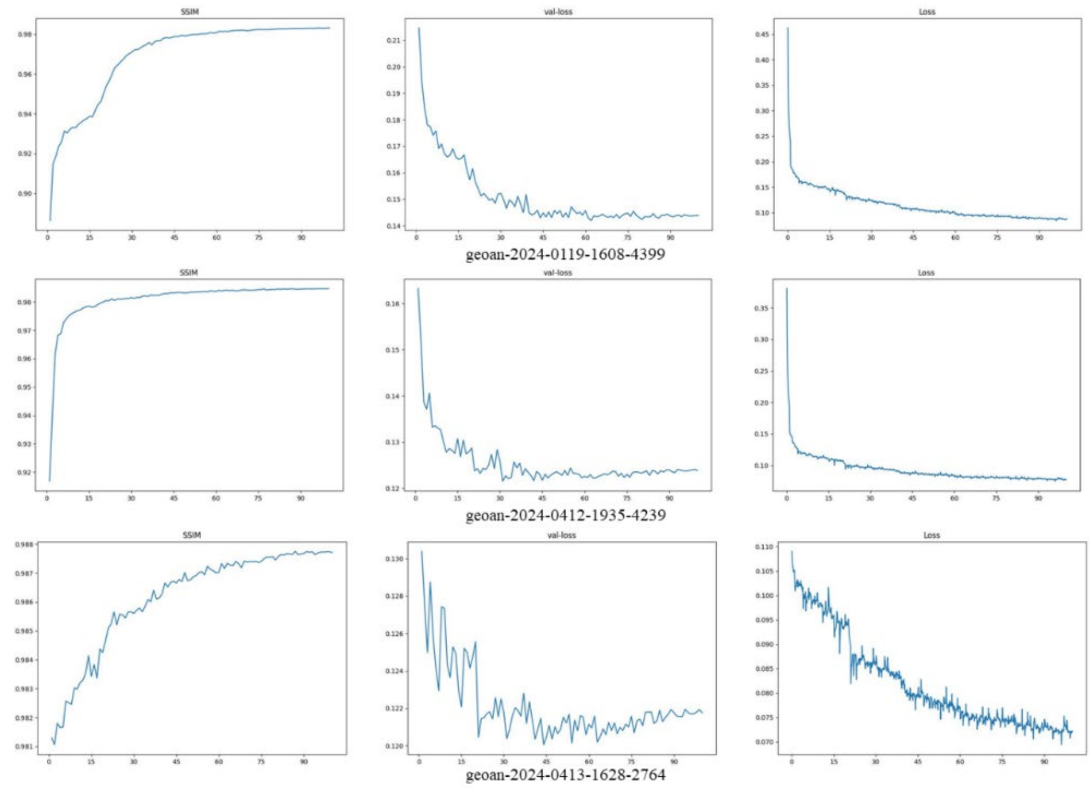

Figure AC1-1. The original training logs of GeoAN. The final GeoAN is trained from the pre-trained model: geoan-2024-0412-1935-4239 (recorded as version-4239), and the version-4239 is trained from version-4399.

A larger dataset and more compared methods are helpful to evaluate GeoAN and MDG625's quality. Limited by the hardware and the availability of data, the larger-scale evaluation has not been conducted yet. In the future, we are going to consider multimodal data (such as rain gauges, satellites, etc.) and use more types of algorithms for analysis to downscale more variables such as precipitation and draw a more comprehensive conclusion.

References:

Bevilacqua, M., Roumy, A., Guillemot, C., and Alberi-Morel, M. L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding, in: The British Machine Vision Conference, pp. 1–10, 2012.

Di Luca, A., de Elía, R., and Laprise, R.: Challenges in the quest for added value of regional climate dynamical downscaling, Current Climate Change Reports, 1, 10–21, 2015.

Giorgi, F. and Gutowski Jr, W. J.: Regional dynamical downscaling and the CORDEX initiative, Annual review of environment and resources, 40, 467–490, 2015.

Huang, J.-B., Singh, A., and Ahuja, N.: Single image super-resolution from transformed self-exemplars, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5197–5206, 2015.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R.: Swinir: Image restoration using swin transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1833–1844, 2021.

Liu, G., Zhang, R., Hang, R., Ge, L., Shi, C., and Liu, Q.: Statistical downscaling of temperature distributions in southwest China by using terrain-guided attention network, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 1678–1690, 2023.

Martin, D., Fowlkes, C., Tal, D., and Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, vol. 2, pp. 416–423, IEEE, 2001.

Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., and Aizawa, K.: Sketch-based manga retrieval using manga109 dataset, Multimedia Tools and Applications, 76, 21 811–21 838, 2017.

Song, Z. and Zhong, B.: A Lightweight Local-Global Attention Network for Single Image Super-Resolution, in: Proceedings of the Asian Conference on Computer Vision, pp. 4395–4410, 2022.

Shen, Z., Shi, C., Shen, R., Tie, R., and Ge, L.: Spatial Downscaling of Near-Surface Air Temperature Based on Deep Learning Cross-Attention Mechanism, Remote Sensing, 15, 5084, 2023.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of hydrology, 456, 12–29, 2012.

Zhong, X., Du, F., Chen, L., Wang, Z., and Li, H.: Investigating transformer-based models for spatial downscaling and correcting biases of near-surface temperature and wind-speed forecasts, Quarterly Journal of the Royal Meteorological Society, 2023.

Zou, W., Hu, G., Wiersma, P., Yin, S., Xiao, Y., Mariethoz, G., and Peleg, N.: Multiple-point geostatistics-based spatial downscaling of heavy rainfall fields, Journal of Hydrology, 632, 130 899, 2024.