Author Reply to RC1 of essd-2024-137

We are pleased to get the comment from the reviewer. Concerning the queries, here are our replies.

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:

(1) “Data density” here means the distributions of in-situ station, which mainly contains station observations. The distribution of stations cannot be unlimited dense, contributing to hardly producing a high-resolution dataset, especially in decades ago. Following is the revised description in manuscript: “…However, the distribution of in-situ stations is too sparse to produce a high-quality reanalysis dataset, especially for decades ago…”

(2) 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.

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:

As suggested by the reviewer, more details are introduced in the manuscript:

“…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 produce 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…Using 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 between low-resolution and high-resolution data. As a result, robustness against the sensitivity can be achieved with the increasing amount of training data. Once the model has been trained, the computational cost is at a low level 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:

(1) Following the suggestions from the reviewer, we presented more details of the literature review, the revised version is as follows:

"…On the one hand, Liang et al. (2021) proposed SwinIR and achieved impressive results in the SR task which can be considered as the benchmark for the SR (Super-Resolution) task. The core algorithm of SwinIR is to use no overlap windows in order to split the input feature for calculating the attention relationship inner each window, then shift the windows by the step of half-width of the windows and calculate the relationship again. Meanwhile, 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… 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 (temperature at 2 meters) was considered in this work and the network was built on CLDAS, which cannot cover long-term years. On the other hand, 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 (temperature at 2 meters, the same as SNCA-CLDASSD) results. The data range of TGAN used began in 2018 and TGAN also cannot be used in the historical situation. What's more, Zhong et al. (2023) proposed a transformer-based learning method Uformer, which used topography data to achieve high-resolution meteorological variables in inner Mongolia province, China. Although topography data can help rebuild the high-resolution, adding into the input low-resolution directly will lose the characters of topography. All of the above, existing advanced deep learning methods of meteorological downscaling mostly employ attention architecture (Transformer is one of the special attention architectures) …"

(2) Transformer architecture has been widely used and considered as the most advanced unit in deep learning methods in recent years due to its excellent performance. 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. Lastly, it's worth noticing that there are no models that can cover a long-term and wide range of historical multiple variables. Our work is under the basis of the transformer architecture, and we merge geopotential into the transformer block to enhance the geographic information for multiple meteorological variables not only to save the computational resources but also to improve the performance of the model. And we produce a long-term historical dataset to fill the gap in high-resolution historical meteorological datasets.

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:

Nowadays, existing methods usually utilize W10m and 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 land surface dataset, and chose the shared variables in both datasets including temperature at 2m, pressure at the surface, wind speed at 10m, and precipitation. Then we evaluated the precipitation of CLDAS and ERA5. Due to the low correlation of precipitation in these two datasets, we discarded the precipitation and used the rest three variables. Noted that the wind speed provided in ERA5 is divided into two components of u and v at 10m and in CLDAS is the synthesized wind speed. We calculated the synthesized wind speed using the data provided by ERA5 and perform the following steps.

Q5: Page 6, line 98, “GeoAB is repeated 18 times”, What considerations are there for setting it to 18?
A5:
On the one hand, 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. On the other hand, the computational expense of deep networks is huge and the training step will last several weeks. We lastly choose 18 as our repeating times, and it serves as the result of balancing the computational resources and the depth of the network. If we choose a deeper network, our GPU (4 RTX 6000 Ada GPU 48G) can’t afford the computational expense and may cause exploding or vanishing gradient problems.

Q6: Page 6, line 105, “GUP memory limitation” should be “GPU memory limitation”?
A6:
Yes, here is a typo, it should be “GPU memory limitation”. We will revise this mistake in the next version and check 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:
As far as we are awarded, our work is the first attempt to establish the mapping relationship from ERA5 to CLDAS to make a long-term 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, which was proposed in 2021, was a novel and SOTA (state of the art) super-resolution model based on transformer block, which can be considered as the benchmark as well as the representative method of super-resolution missions.

In order to migrate the methods to downscale the long-term meteorological dataset from ERA5 to CLDAS, we have to redesign each existing method. These methods may take several weeks to months to retrain so as to be compatible with the task.
In the future, we will continue to make larger comparisons with more existing representative deep learning methods from different fields and tasks, and delve into how these deep learning methods can dig geographic information to help various geographic tasks.

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:
We have downloaded the CLDAS dataset from 2020 to 2023. The validation dataset needs to cover one

full cycle, for meteorological task is one year. The amount of training data is directly related to the performance of the network. Under guaranteeing data spanning at least a whole year for validation, we used as much as we can to train the model. Thus, we used the entire data from 2023 to validate the proposed method and used data from 2020 to 2022 to train our own work. Although CLDAS contains the data since 2008, it's hard to obtain previous data after 2017, thus we cannot access enough long-range data to validate our model. What's more, the increase in the amount of data can also greatly expand and prolong the network's training expenditure and time. In future work, we will try to use other wider range of datasets to validate geography-related tasks.

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.