Response to Referee 2

Comments: The original CorrDiff framework relies on the assumption of spatially stationary and homogeneous residual distributions, which is theoretically invalid over a large domain with highly heterogeneous surfaces (e.g. the Tibetan Plateau, deserts, coasts, and mountains in China). Furthermore, the uncertainly estimated from diffusion ensemble variance is empirically motivated and lacks a solid theoretical link to atmospheric intrinsic predictability, initial condition errors, or model error sources.
In this sense, the authors should justify the application of the stationary residual assumption over complex regions, for example by spatially stratified error analysis or regional subdomain modeling.

Response:

We sincerely appreciate the reviewer's thorough and constructive comments, which greatly help improve the quality of this manuscript. Detailed responses are presented below.

First, regarding the assumption of spatially stationary and homogeneous residual distributions. As pointed out, the study area of China features highly heterogeneous underlying surfaces including the Tibetan Plateau, deserts, coasts and mountains, so the residual field naturally violates the strict spatial stationarity and homogeneity assumption. This spatial heterogeneity is well validated by Figures 1 and 2, which exhibit differentiated prediction performances at different vertical levels and geographical subdomains. We have revised the manuscript to explicitly state that this work does not assume a globally consistent residual distribution from a physical perspective.

Technically, the residual correction model used in this work is a conditional generative model rather than an unconditional residual sampler. The diffusion model is conditioned on the interpolated low-resolution fields, the regression-model output, and high-resolution orography. Therefore, although a single set of network parameters is shared over the whole domain, the learned residual distribution is conditional on spatially varying meteorological states and terrain information. This design allows the model to represent spatially dependent residual structures to some extent. In particular, the global residual connection is used to decompose the downscaling problem into a deterministic large-scale component and a stochastic residual correction component, so that the diffusion model mainly learns unresolved local-scale structures rather than the full high-resolution field.

Nevertheless, we recognize that the current conditional structure cannot fully resolve the challenges caused by regional heterogeneity. To quantify this issue, we have added a spatially stratified error analysis in the revised manuscript. The results show that the prediction errors and CorrDiff corrections exhibit clear regional and vertical dependence. For example, stronger corrections are observed over complex-terrain and oceanic regions for near-surface wind, while the correction magnitude and error pattern differ across pressure levels and variables. These results confirm the reviewer's concern that the residual statistics are not uniform over the domain, but they also show that the present conditional CorrDiff model can partially adapt to such heterogeneity through its meteorological and orographic conditioning.

We have supplemented relevant discussions to clarify this inherent limitation. In future work, we will investigate more region-aware CorrDiff variants, including spatially stratified training/evaluation, terrain- or climate-zone-dependent normalization, regional expert modules, and mixture-of-experts or domain-adaptive diffusion heads for different subdomains and variables. On this basis, we will further implement targeted differentiated modeling strategy for different subdomains and meteorological variables at different pressure levels, such as employing distinct neural network modules for diverse regions and variables, to better account for the distinct residual distributions over the Tibetan Plateau, arid regions, coastal areas, and mountainous regions.

Second, with respect to uncertainty estimation based on diffusion ensemble variance. We agree that the variance of the diffusion ensemble should not be interpreted as a rigorous measure of atmospheric intrinsic predictability, initial-condition uncertainty, or model-error covariance. The core goal of this study is to demonstrate the feasibility of the residual corrective diffusion model for high-resolution downscaling and explore whether the ensemble spread from the diffusion process can serve as a practical empirical proxy for predictive uncertainty. This approach is empirically inspired by the successful applications of diffusion ensembles in other fields such as computer vision, where stochastic sampling during inference can produce diverse and physically plausible outputs. We have never intended to claim that the ensemble spread can quantify classical predictability limits or error covariance in a dynamically consistent manner.

In the revised manuscript, we have softened the relevant claim and now describe the diffusion ensemble variance as an empirical proxy for predictive uncertainty. Our analysis shows that grid points with lower CorrDiff ensemble variance generally have lower absolute error, and that variance-based grouping can separate regions with different error levels. Thus, the ensemble spread provides useful diagnostic information about the reliability of the downscaled prediction, but it does not by itself provide a complete dynamical uncertainty decomposition. We have clarified this distinction in the revised manuscript.

Finally, we have summarized two key research directions for follow-up work: (1) Optimize the CorrDiff framework via subdomain- and variable-specific customized modeling to adapt to heterogeneous regions; (2) Establish physical constraints and theoretical connections between diffusion-based uncertainty quantification and atmospheric predictability, initial perturbations and model errors, to improve the physical interpretability of uncertainty estimation.

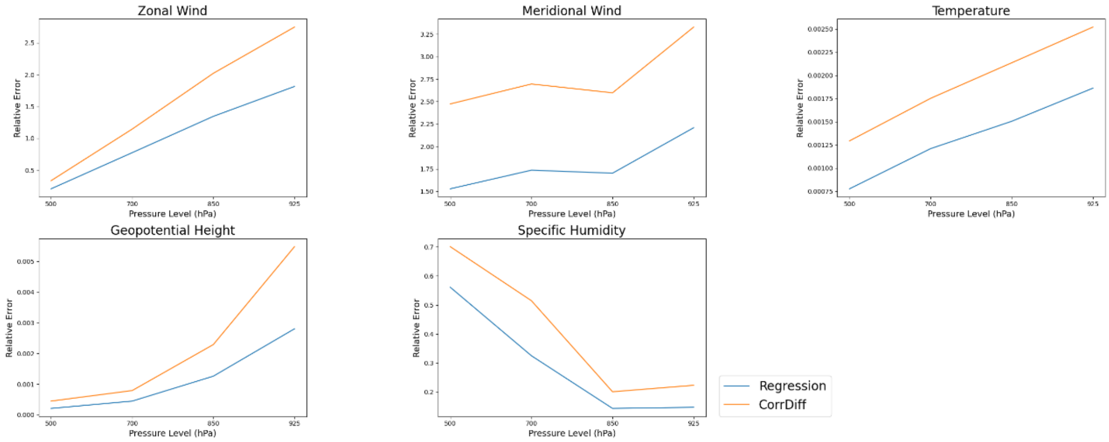

Figure 1. Relative error at different pressure levels, based on data from 00, 06, 12, and 18 UTC on 2023-03-01, 2023-03-02, and 2023-03-03.

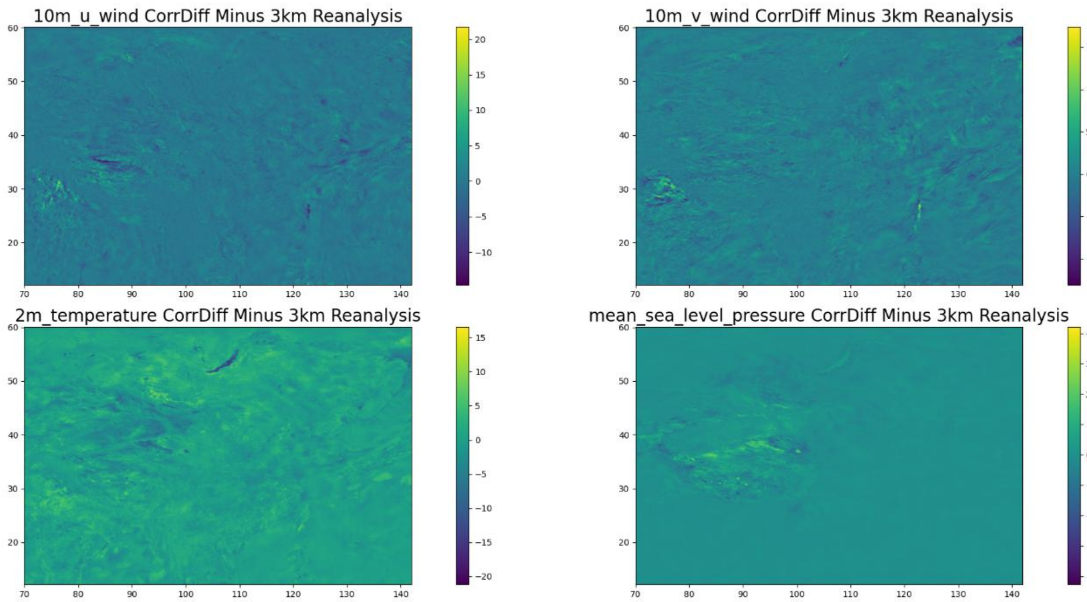

Figure 2. Spatial distribution of error, based on data from 00 UTC on 2023-03-01.