Dear Reviewer,

The comments offered have been immensely helpful. We appreciate your insightful comments on our paper. We have responded to every question, indicating exactly how we addressed each concern.

General Comments:

The paper is presenting a method for spatial interpolation of air temperature data for China from meteorological stations based on machine learning tools. The authors analyze the technique used and present the limitations of the experiment. Three ML models were tested and three interpolation method and the Gaussian Process Regression was chosen based on its better performance. The results compared with existing published datasets. A detailed trend analysis of the predicted dataset is also presented. ML techniques are very promising as they are addressing current challenges in computational research.

Response: Many thanks for the constructive comments on our study.

Specific comments:

Q1: It is not clear to me if all stations contribute equally to the analysis, for example red and blue stations (Figure S2). Is there any weighting technique applied to the training model(s)? If yes, I think it could be mentioned.

Response: In our study, we used the “subset features” option of the Geostatistical Analyst Tools in ArcGIS10.8 to divide the original dataset into 70% training dataset and 30% testing dataset. This tool considers the randomness both in the data and the spatial distribution of the data. There is no weighting technique applied in the training models.

Q3: Besides the trends statistical analysis, are there available error spatial distributions of the predicted temperatures so as to illustrate the confidence level of the analysis results especially in station empty areas?

Response: GPR is a full Bayesian learning algorithm. A process is referred to as a Gaussian process if it is assumed that the joint probability distribution of model outputs is Gaussian (Zhu et al., 2018). Because a GPR model is probabilistic, it is possible to compute the prediction intervals using the trained model. As you suggested, we will provide the spatial confidence level graph for each month (with a significance level of 95%). Here, we displayed the spatial distribution of the width of the predicted intervals with a significance level of 95% (using the upper limit minus the lower limit) for the 12 months in 2010 using the trained models (Figure 1). Figure 2 provides the histogram of the width of the confidence intervals. The calculation was conducted in Matlab 2021b. For more detailed information about the prediction intervals of GPR models, please see https://www.mathworks.com/help/stats/gaussian-process-regression-models.html.

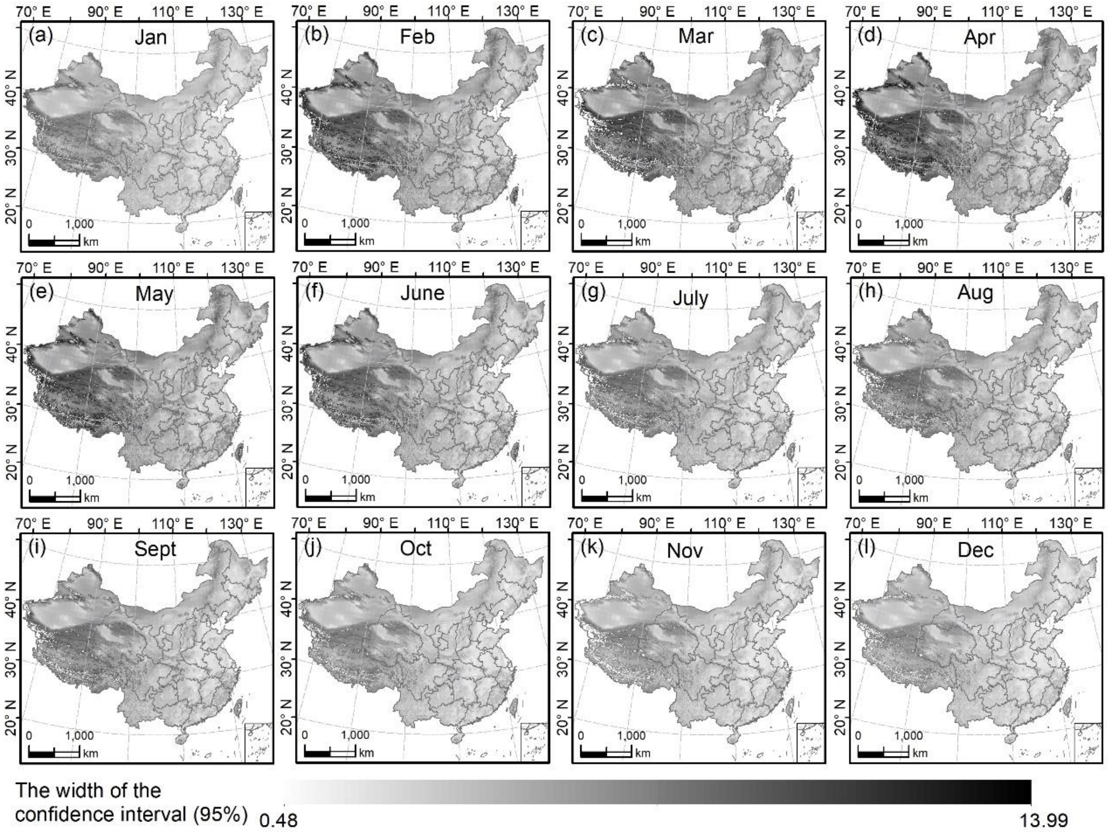

Figure 1 The spatial distribution of the width of the 95% prediction intervals (the upper limit minus the lower limit) for 12 months in 2010

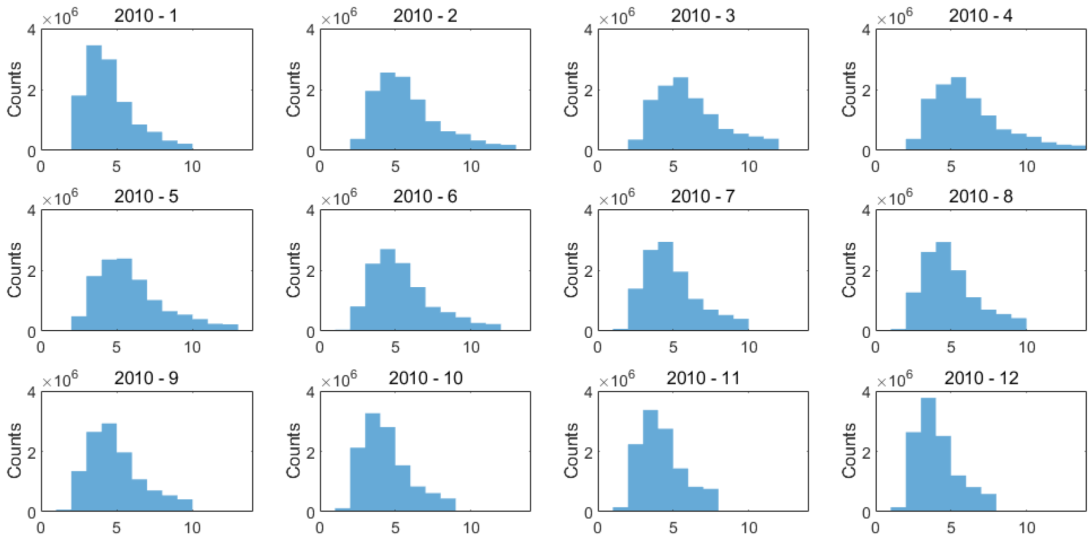

Figure 2 The histogram of the width of the 95% prediction intervals (the upper limit minus the lower limit) for 12 months in 2010

Q4: In addition, as the study of the climatic dynamics is in the epicenter of this work, the use of remote sensing data jointly with the land meteo stations could overcome the data scarcity, improve the results and reveal trends with more accuracy after 2000.

Response: The goal of our study is to generate the long-term time series of high-resolution temperature data. Due to the limitation of the remote sensing data, we did not consider remote sensing data in our study. We discussed the topic in the discussion session.

The longitude, latitude and elevation are static factors, but we construct the model for each month, respectively, which can reflect the changes of temperature from month to month. The remote sensing data such as NDVI, land-use change and surface temperature are usually not available before 2000 since our data is from 1951 to 2020. Furthermore, the MODIS data are not available for each month from January 2000 to December 2020. As shown in Figure 3, the percentage of the available MODIS images are low in northeast China and southern areas. So we did not use the remote sensing data for generating long-term temperature data in our study.

We will consider using the remote sensing data in future studies to further increase the accuracy.

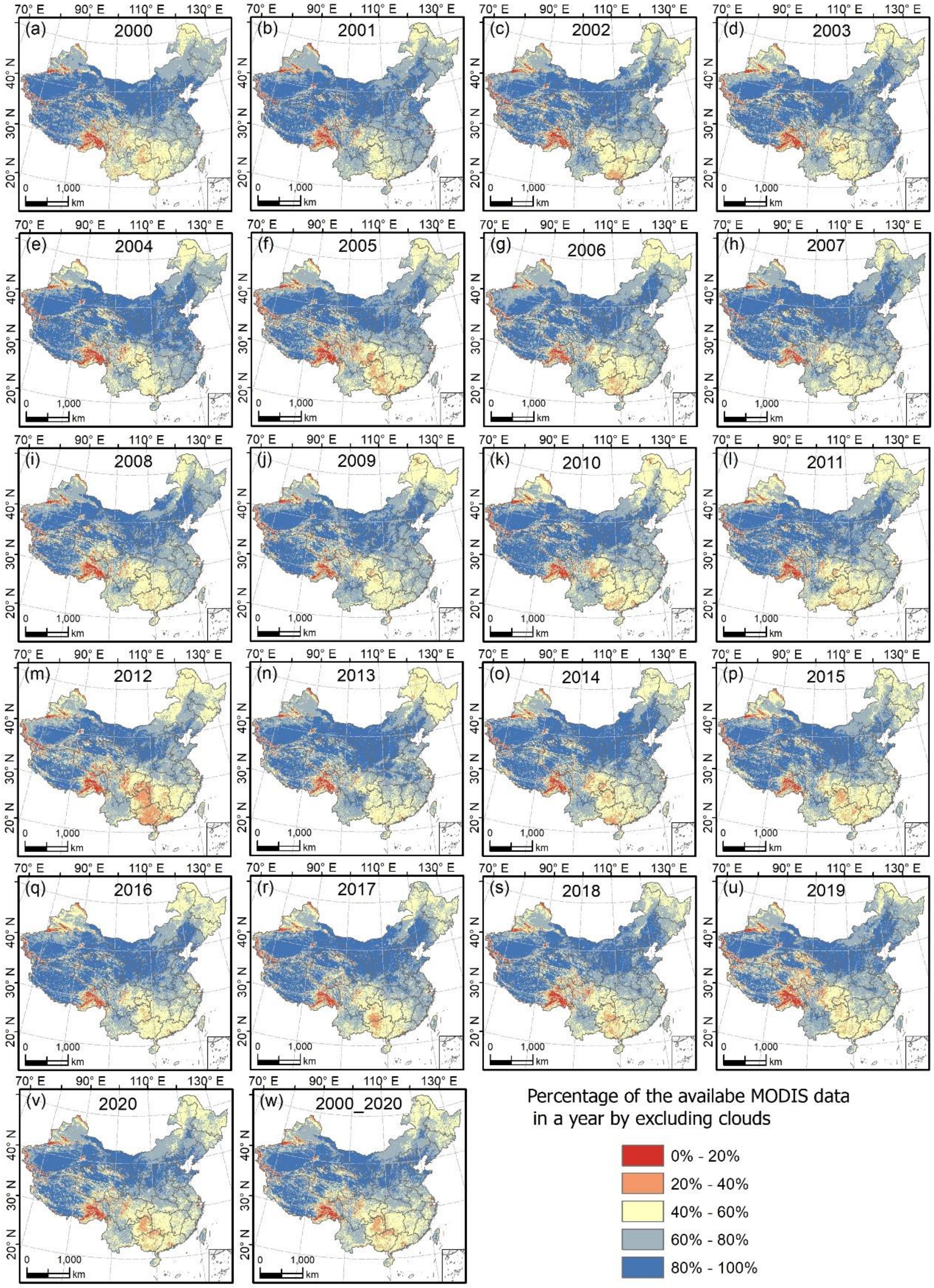

Figure 3 Spatial distribution of the percentage of the available MODIS images in each year (2000 - 2020) by excluding clouds.

Q5: While the height of the air Temperature is mentioned for the ERA5 dataset (2m), this is not the case for the GPRChinaTemp1km product or the other datasets mentioned in the analysis.

Response: We will include more detailed descriptions of the data we used. In our study, we use the weather station data to interpolate the gridded temperature datasets. The station data used in the study records the temperature at 2 m height above ground (Liu et al., 2011; Zhang et al., 2010). Thus, the generated GPRChinaTemp1km product also represents the temperature data at 2 m height.

For TerraClimate data, it is produced based on other datasets including WorldClim, CRUTs4.0 and JRA-55 (Abatzoglou et al., 2018, p.1958–2015). The temperatures in WorldClim are at 2 m height (Fick and Hijmans, 2017; Chou et al., 2020). The temperature from CRU Ts and JRA-55 are also at 2 m height (Harris et al., 2020; https://jra.kishou.go.jp/JRA-55/document/JRA-55_handbook_LL125_en.pdf). Therefore, the TerraClimate dataset also represents the 2m temperature.

The height of the temperature data from FLDAS is also 2 m (McNally et al., 2017; https://ldas.gsfc.nasa.gov/fldas/specifications).

We will include more detailed information on the height of the data we used in the revised manuscript.

Q6: Could the experiment be tested in other atmospheric parameters? If so, I think that a few sentences on the perspectives of the specific approach would be beneficial.

Response: Thanks a lot for your suggestion. Our current study is only aiming for temperature. We have not done some experiments on other meteorological variables. This actually is our next work. We are trying to apply the GPR model to other meteorological variables. We will add a few sentences to describe that in the Discussion session.

References:

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C.: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015, 5, 170191, https://doi.org/10.1038/sdata.2017.191, 2018.

Chou, S. C., de Arruda Lyra, A., Gomes, J. L., Rodriguez, D. A., Alves Martins, M., Costa Resende, N., da Silva Tavares, P., Pereira Dereczynski, C., Lopes Pilotto, I., Martins, A. M., Alves de Carvalho, L. F., Lima Onofre, J. L., Major, I., Penhor, M., and Santana, A.: Downscaling projections of climate change in Sao Tome and Principe Islands, Africa, Clim Dyn, 54, 4021–4042, https://doi.org/10.1007/s00382-020-05212-7, 2020.

Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, 37, 4302–4315, https://doi.org/10.1002/joc.5086, 2017.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Sci Data, 7, 109, https://doi.org/10.1038/s41597-020-0453-3, 2020.

Liu, X., Luo, Y., Zhang, D., Zhang, M., and Liu, C.: Recent changes in pan-evaporation dynamics in China, 38, https://doi.org/10.1029/2011GL047929, 2011.

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C. D., and Verdin, J. P.: A land data assimilation system for sub-Saharan Africa food and water security applications, Sci Data, 4, 170012, https://doi.org/10.1038/sdata.2017.12, 2017.

Zhang, X., Kang, S., Zhang, L., and Liu, J.: Spatial variation of climatology monthly crop reference evapotranspiration and sensitivity coefficients in Shiyang river basin of northwest China, Agricultural Water Management, 97, 1506–1516, https://doi.org/10.1016/j.agwat.2010.05.004, 2010.

Zhu, S., Nyarko, E. K., and Hadzima-Nyarko, M.: Modelling daily water temperature from air temperature for the Missouri River, PeerJ, 6, e4894, https://doi.org/10.7717/peerj.4894, 2018.