Author response to 'Downscaling precipitation over High Mountain Asia using Multi-Fidelity Gaussian Processes: Improved estimates from ERA5'

Kenza Tazi, Andrew Orr, Javier Hernandez-González, Scott Hosking, Richard E. Turner

August 2, 2024

We thank both reviewers for their feedback on the manuscript. We discuss their comments (italicised) below.

Response to Reviewer 1

We thank Reviewer 1 for their thorough and positive feedback (e.g., "*Overall, this is one of the best scientific papers I have ever read, especially considering its highly technical nature*"). We appreciate their encouraging comments on the manuscript's presentation, including the clarity and conciseness of the writing, the precision of the mathematical formulae and the efficacy of the figures (e.g., "*All aspects of the paper appear to be very carefully prepared, including citation of relevant literature, readability of the text by a broad audience, presentation of methods and results, and technical precision in presentation of the mathematics and statistics*"). We are also grateful they recognised our efforts towards contextualising this research within existing literature and giving a balanced discussion about the advantages and disadvantages of the proposed method.

Response to Reviewer 2

We thank Reviewer 2 for their insightful comments and suggestions. We address the reviewer concerns point by point below.

Comment 1

"*First, ECMWF also provides high-resolution reanalysis precipitation data (ERA5 Land, hourly, 0.1 degree, 9 km), which is not considered in the manuscript. How does the generated MFGP precipitation estimates compare with ERA5 Land precipitation data?*"

ERA5-Land is a reanalysis dataset that provides a consistent view of the evolution of land variables at an enhanced spatial resolution of 0.1° by 0.1° (approx. 9km) compared to ERA5's resolution of 0.25° by 0.25° (approx. 31km). It is produced by running a land surface model to regenerate some of the land components of ERA5 climate reanalysis. For atmospheric forcing, it uses ERA5 atmospheric variables such as air temperature and precipitation at a 0.1° resolution by linearly interpolating the driving variables to the ERA5-Land grid. Although other forcing variables are corrected, this is not the case for precipitation. For further details please see Muñoz-Sabater et al. (2021) and ECMWF (2024). Precipitation characteristics from ERA5-Land are therefore very similar to ERA5 (Gomis-Cebolla et al., 2023; Xu et al., 2022; Xin et al., 2022). They also should theoretically perform worse than the linear regression models presented in our paper (Table 1-3) which also include elevation as a predictor.

Moreover, we would like to stress that we have carefully chosen four gridded precipitation datasets for the High Mountain Asia region, based on existing literature, to evaluate our model against. These are:

- APHRODITE, which is a gridded rain-gauge interpolated dataset for Asia considered the gold-standard for precipitation in High Mountain Asia,
- CRU TS, which is a global gridded rain-gauge interpolated dataset.
- A bias-corrected high-resolution regional climate model simulation, which used the Weather Research and Forecast (WRF) model at a spatial resolution of 5 km, with precipitation output corrected using local rain-gauge data for the region investigated in this manuscript.
- TRMM, which is a satellite-based precipitation dataset, designed to improve our understanding of precipitation in the current climate.

To address the reviewer's concern, we make the connection between ERA5-Land and our linear regression models clearer in a new subsection of the discussion:

"We note that the model presented in this paper is similar to the interpolation scheme used for precipitation in ERA5-Land (Muñoz-Sabater et al., 2021). ERA5-Land is a reanalysis dataset that provides a consistent view of the evolution of land variables at an enhanced spatial resolution of 9 km. This is produced by running a land surface model to regenerate some of the land components of ERA5 climate reanalysis. For atmospheric forcing, it uses ERA5 atmospheric variables including precipitation which are linearly interpolated to the ERA5-Land grid. The linear interpolation model also includes elevation as a predictor which should allow it to perform better than ERA5-Land especially over mountainous regions."

– Section 6.3.3

Comment 2

"*Second, the authors only consider a very simple machine learning model, i.e., linear regression, and complex deep learning models that require a lot of training data, including Convolutional Conditional Neural Processes (ConvCNP) and Convolutional Gaussian Neural Processes (ConvGNP). They neglect simple machine learning methods that do not need many data, such as random forest and support vector machines.*"

The main goal of the paper was to show that the precipitation uncertainty could be narrowed by combining datasets from multiple sources. This information allows hydrologists to quantify the probabilities of extreme events and policymakers to make better decisions with limited resources as highlighted in Section 1. However, we appreciate that the case for the performance of Gaussian processes could be better contextualised by including these models.

To address the reviewer's concern we implement random forests and support vector regression for the validation experiments and now include the performance of ConvCNPs. These results are shown below in Tables A-C and are presented in the manuscript under Appendix E with the linear regression model and the models' implementation details. A summary of the results and the rationale for choosing these models are included in the discussion:

"Overall, linear interpolation performs significantly worse over both Europe and the Beas and Sutlej basins than the MFGP, and even its probabilistic counterpart, the GP fit to ERA5. This can be attributed to the GP's generation of non-linear functions that better capture ERA5's physics and data assimilation.

We then contrast the MFGP to random forest and support vector regression. Both random forests (Ho, 1995) and support vector regression (Drucker et al., 1996) have been used extensively to downscale precipitation, including over High Mountain Asia (Sun et al., 2022; Xiang et al., 2024; Ahmed et al., 2020; Yan et al., 2022; Ning et al., 2016; Mei et al., 2020). Both methods work well with small datasets, are non-linear, and, for support vector regression, are kernel-based like GPs. .

The random forest and support vector regression models perform similarly to the MFGPs in terms of RMSE/R^2 for the 'data-rich' Europe experiment. However the MFGP performs consistently better for these metrics and is less sensitive to the reduction of data when moving to the 'data-sparse' setup. Over Europe, the random forests are however better at representing extreme values across all the cross validation folds. Over the Beas and Sutlej basins, the MFGP dominates offering more better and more consistent results with the exception of the 5$^{\text{th}}$ percentile RMSE. The relatively poor performance for the low percentiles values is due to the GP and MFGP models reverting to the observation mean in locations far from the training distribution where they are uncertain rather than confidently predicting lower values like the non-probabilistic models.

Lastly, ConvCNPs are also implemented for the validation experiments. The ConvCNP model is one member of the neural process model family that has shown state-of-the-art performance in spatiotemporal downscaling tasks (Vaughan et al., 2022; Gordon et al., 2019; Andersson et al., 2023). Neural processes offer similar advantages to the MFGP in terms of being able to quantify the probability of extreme events, generalise to multiple locations, predict at arbitrary locations, and overcome gridding biases. The results show that these models overfit these relatively small datasets performing worse than linear regression, in particular, for the Beas and Sutlej experiment. This is not surprising as neural networks generally require a large number of datapoints to be trained adequately. As these models can be used for transfer learning, future work could investigate the using data from other mountainous regions to inform predictions in data-sparse High Mountain Asia. In summary, the MFGPs are best suited to downscaling in the sparse and out-of-distribution settings presented in this paper."

– Section 6.3.3

Model	Training features	RMSE [mm/day]	RMSE5 [mm/day]	RMSE95 [mm/day]	R^2	MLL
Linear reg.	ERA5	1.72±0.46	1.75±0.18	5.21±1.55	0.04±0.06	-
RF	ERA5 + gauges	1.12±0.44	**0.45±0.19**	**2.62±0.93**	0.61±0.09	-
SVR_{RBF}	ERA5 + gauges	1.14±0.46	0.53±0.33	3.03±1.48	0.60±0.12	-
MFGP	ERA5 + gauges	**1.06±0.42**	0.51±0.20	2.72±1.54	**0.65±0.09**	**0.89±0.20**
ConvCNP	ERA5 + gauges	2.16±0.76	2.29±0.93	4.25±1.60	-0.49±0.48	2.40±0.91

Table A: Comparison of model performance metrics trained on ERA5 data for the 'data-rich' setup over Europe. We include a linear interpolation model, a random forest (RF), a support vector regression (SVR) model with a smooth Radial Basis Function (RBF) kernel, a ConvCNP and the MFGP model. The metrics include the average RMSE, the 5th percentile RMSE (RMSE5), the 95th percentile RMSE (RMSE95), the R^2 score, and the MLL. The errors represent the standard deviation across the validation folds.

Model	Training features	RMSE [mm/day]	RMSE5 [mm/day]	RMSE95 [mm/day]	R^2	MLL
Linear reg.	ERA5	1.77±0.46	1.88±0.25	5.19±1.76	-0.02±0.13	-
RF	ERA5 + gauges	1.16±0.39	**0.41±0.20**	**2.92±1.39**	0.57±0.10	-
SVR_{RBF}	ERA5 + gauges	1.53±0.62	0.73±0.23	4.64±1.99	0.29±0.19	-
MFGP	ERA5 + gauges	**1.13±0.47**	0.57±0.23	3.02±1.62	**0.62±0.11**	**0.90±0.20**
ConvCNP	ERA5 + gauges	1.92±0.51	1.77±0.78	4.84±1.70	-0.21±0.34	2.36±1.38

Table B: As Table A for the 'data-sparse' setup over Europe.

Model	Training features	RMSE [mm/day]	RMSE5 [mm/day]	RMSE95 [mm/day]	R^2	MLL
Linear reg.	ERA5	4.21±0.99	2.21±0.45	14.33±4.03	-0.08±0.05	-
RF	ERA5 + gauges	3.05±1.30	**0.52±0.46**	9.87±5.47	0.45±0.23	-
SVR_{RBF}	ERA5 + gauges	3.36±1.66	0.66±0.38	11.05±6.14	0.34±0.33	-
MFGP	ERA5 + gauges	**3.00±0.92**	1.66±0.95	**9.62±3.63**	**0.46±0.11**	**1.79±0.22**
ConvCNP	ERA5 + gauges	4.89±0.93	3.57±0.79	14.16±3.85	-0.51±0.32	3.95±0.88

Table C: As Table B for the upper Beas and Sutlej basins.

Comment 3

"Finally, in the model comparison shown in Table 1-3, GP is only trained on ERA5 data at station locations. Why not use all ERA5 data in the study region to train and test the model? I would expect better model performance even for GPs."

The model has information about ERA5 at all the training and test locations for the validation experiments. These locations fall within the ERA5 grid boxes, so there is theoretically little to no additional information to be gained by including neighbouring grid box values. To check this, we ran an experiment that was trained on all ERA5 data for the Beas and Sutlej basins.

Results from this experiment are shown in Table D (below), and confirm that there is no added benefit in including this data. However, we have clarified this in the revised manuscript by adding additional text which states:

"The experiments were also conducted with all the ERA5 data for the study area (not shown), but showed no significant improvement over using the ERA5 data at the station locations only".

– Section 4.2

We note that we did not rerun these experiments over Europe, as we would have needed to apply methodological approximations to overcome the memory and computational bottlenecks that comes with this larger domain.

Model	RMSE [mm/day]	RMSE5 [mm/day]	RMSE95 [mm/day]	R^2	MLL
$\mathrm{MFGP}_{\mathrm{limited}}$	**3.00±0.92**	1.66±0.95	**9.62±3.63**	**0.46±0.11**	1.79±0.22
$\mathrm{MFGP}_{\mathrm{all}}$	5.16±2.51	**0.84±0.56**	19.48±9.79	0.32±0.27	**1.68±0.34**

Table D: Comparison of MFGP performance using ERA5 for the whole study area (all) and using only ERA5 at the training and test site locations (limited) over Upper Beas and Sutlej Basins. The metrics include the average RMSE, the 5^{th} percentile RMSE (RMSE5), the 95^{th} percentile RMSE (RMSE95), the R^2 score and the mean log loss (MLL).The bolded values highlight the best scores

Comment 4

"*I believe the authors use the Nash-Sutcliffe efficiency (NSE) in the manuscript, rather than R2, which should always be non-negative values.*"

We confirm that we are using the coefficient of determination or R^2 score. This metric, defined and explained in Appendix B, and is given by:

$$\mathrm{R}^2 = 1 - \frac{\mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{tot}}} = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \overline{y})^2}$$

where f_i is the i^{th} predicted value, y_i is the i^{th} observed value, and $\overline{y}$ is the mean of the observations. $\mathrm{SS}_{\mathrm{res}}$ is therefore the sum of the squared residuals and $\mathrm{SS}_{\mathrm{tot}}$ is the total sum of squares. A negative R^2 is possible and would indicate that the model is predicting worse than the precipitation mean. Although negative R^2 scores are unlikely in interpolation settings, they are possible when making predictions outside of the training distribution. To address the reviewer's concerns, we make this interpretation of negative R^2 clearer in the main body of the paper and the appendix:

"Of these methods, the GP with the custom kernel extrapolating only from gauges yields the poorest results with a negative R^2 indicating that the model is predicting worse than the precipitation mean."

– Section 4.2

"An R^2 of 1 indicates a perfect fit whilst a negative R^2 means the model performs worse than the mean. Although negative R^2 scores are unlikely in interpolation settings, they are possible when making predictions outside of the training distribution."

– Appendix B

Summary of changes

- Make the connection between ERA5-Land and the linear regression model clearer in a new subsection of the discussion (Section 6.3.3).
- Implement random forests and support vector regression for the validation experiments and which we present in Appendix E. The results are discussed alongside the linear regression and ConvCNP performances in Section 6.3.3.
- Clarify the limited benefit of using additional ERA5 data for the validation experiments in Section 4.2.
- Make the interpretation of negative R^2 clearer in the main body of the paper and Appendix B.

References

Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth system science data*, 13(9):4349–4383, 2021.

ECMWF. ERA5-Land: Data documentation. Technical report, ECMWF, 2024. URL `https://confluence.ecmwf.int/display/CKB/ERA5-Land%3A+data+documentation`.

José Gomis-Cebolla, Viera Rattayova, Sergio Salazar-Galán, and Félix Francés. Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020). *Atmospheric Research*, 284:106606, 2023.

Jintao Xu, Ziqiang Ma, Songkun Yan, and Jie Peng. Do ERA5 and ERA5-land precipitation estimates outperform satellite-based precipitation products? A comprehensive comparison between state-of-the-art model-based and satellite-based precipitation products over mainland China. *Journal of Hydrology*, 605: 127353, 2022.

Ying Xin, Yaping Yang, Xiaona Chen, Xiafang Yue, Yangxiaoyue Liu, and Cong Yin. Evaluation of IMERG and ERA5 precipitation products over the mongolian plateau. *Scientific Reports*, 12(1):21776, 2022.

Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

Harris Drucker, Christopher J Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. Support vector regression machines. *Advances in neural information processing systems*, 9, 1996.

He Sun, Tandong Yao, Fengge Su, Zhihua He, Guoqiang Tang, Ning Li, Bowen Zheng, Jingheng Huang, Fanchong Meng, Tinghai Ou, et al. Corrected ERA5 precipitation by machine learning significantly improved flow simulations for the third pole basins. *Journal of Hydrometeorology*, 23(10):1663–1679, 2022.

Yuxuan Xiang, Chen Zeng, Fan Zhang, and Li Wang. Effects of climate change on runoff in a representative Himalayan basin assessed through optimal integration of multi-source precipitation data. *Journal of Hydrology: Regional Studies*, 53:101828, 2024. ISSN 2214-5818.

Kamal Ahmed, Zafar Iqbal, Najeebullah Khan, Balach Rasheed, Nadeem Nawaz, Irfan Malik, and Mohammad Noor. Quantitative assessment of precipitation changes under CMIP5 RCP scenarios over the northern sub-Himalayan region of Pakistan. *Environment, Development and Sustainability*, 22:7831–7845, 2020.

Yilin Yan, Hao Wang, Guoping Li, Jin Xia, Fei Ge, Qiangyu Zeng, Xinyue Ren, and Linyin Tan. Projection of future extreme precipitation in China based on the CMIP6 from a machine learning perspective. *Remote Sensing*, 14(16):4033, 2022.

Chen Ning, Yudan Wang, Zhuotong Nan, Hao Chen, and Canran Liu. Study on correction of daily precipitation data of the Qinghai-Tibetan plateau with machine learning models. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 517–520. IEEE, 2016.

Yiwen Mei, Viviana Maggioni, Paul Houser, Yuan Xue, and Tasnuva Rouf. A nonparametric statistical technique for spatial downscaling of precipitation over High Mountain Asia. *Water Resources Research*, 56(11):e2020WR027472, 2020.

Anna Vaughan, Will Tebbutt, J Scott Hosking, and Richard E Turner. Convolutional conditional neural processes for local climate downscaling. *Geoscientific Model Development*, 15(1):251–268, 2022.

Jonathan Gordon, Wessel P Bruinsma, Andrew YK Foong, James Requeima, Yann Dubois, and Richard E Turner. Convolutional conditional neural processes. *arXiv preprint arXiv:1910.13556*, 2019.

Tom R Andersson, Wessel P Bruinsma, Stratis Markou, James Requeima, Alejandro Coca-Castro, Anna Vaughan, Anna-Louise Ellis, Matthew A Lazzara, Dani Jones, Scott Hosking, et al. Environmental sensor placement with convolutional gaussian neural processes. *Environmental Data Science*, 2:e32, 2023.