Dear Editor and Reviewers,

Many thanks for your valuable comments and suggestions, which were very useful for improving presentation of results and paper readability. In the following, we will answer all the comments raised by the reviewers in detail. The reviewers' comments are in bold, citations in Italic and our answers are in regular font.

1 Reviewer 2:

1.1 Major comments

The study presents results from two downscaling methods for GCM precip itation. One is dynamical downscaling with the CCLM RCM and the second is a statistical emulator for CCLM based on CNNs. The downscaling is applied to simulations with the MPI−ESM1−2_HR GCM. Both approaches are in principle useful and the results are informative.

Thanks a lot for the positive feedback. We have to mention that we also conducted a simulation where we downscaled EC−Earth3−Veg GCM using CCLM and the trained CNN (see Sec. 4.1 Applying CNN to a different GCM and `https://zenodo.org/records/10417111`). We will add more details of the simulations in a new version of the manuscript.

However, the method evaluation is very limited as it is only based on MAE, while many other evaluation measures could have been used. I suggest to add at least an analysis of the bias in the mean and include a justification for the specific choice of evaluation measures. Moreover, the analysis of MAE of GCM simulations and GCM-driven RCM simulations relative to observations is fundamentally wrong, because the random internal variability in the simulations and in the observations is not synchronised.

We are aware that he simulations are in the so-called "free" mode and do not include any kind of data assimilation and do not "see" the observations. However, we conduct averaging of 30 years on each day, i.e. we have 30 first of January for example and the resulted pattern is not only a random pattern of a single day. However, as suggested by you, we added the new figure 5 in the new version of the manuscript, showing biases of the climatological values. With such figures we again confirm the added values of the RCM downscaling conducted in our study for mean states. The overall added value pattern is similar in both cases, i.e. using MAE or climatological bias. This might be embedded in the 30-years averaging in the MAE calculation, which cancels out small random variablities, which are induced by the dynamical downscaling.

The choice of using MAE for the comparisons are motivated by the fact that a mean of absolute errors over 30 years already cancels out the majority of randomness within the dataset. And we are ultimately interested in the reduction of the mean error via downscaling. Other reason for selection of MAE lays in the distribution of Precipitation. According to the study of Hodson (2022), MAE is more suitable for precipitation than other metrics.

For the rest of the manuscript, we will show the MAE, because we are comparing the emulated pattern against the RCM-pattern and there is no comparison to observations.

The manuscript is also not well written, and important conceptual basics and technical details are not clearly explained. It needs substantial rewriting before it is suited for publication.

We have implemented your critics in the new version of the manuscript and modified the text a lot to meet your expectations.

1.2 Specific comments

The introduction contains many good points, but there is some repetition and it should be better structured. The purpose of the paper should be made clear in the abstract and at the beginning of the introduction. What are the research questions?

We modified the introduction accordingly with a focus on the research questions listed as bullet points in the introduction.

- Research Question 1: How effectively can CMIP6 models be downscaled for the CORDEX Central Asia region to enhance precipitation simulations? (Kendon et al., 2014; Demory et al., 2020; Hess et al., 2022)
- Research Question 2: Can convolutional neural networks (CNNs) effectively downscale GCM outputs, and how do they perform when applied to GCMs they were not initially trained on? (Sun and Lan, 2021; Rasp and Lerch, 2018)

Is this mainly a methodological study or is the main purpose to provide high-resolution scenarios to inform impact and adaptation studies? The first part of the introduction suggests that the CCLM predictions are the main point, and only after line 110 it is said that a ML emulator will be tested, and that three research topics are addressed. The second research topic (line 118) is unclear. What

is meant with the 'dynamical downscaling signal for heavy precipitation'? Signal of what? The third topic is training a CNN-based emulator for CCLM, but why is evaluation not mentioned? It would also be good to discuss whether there are already published findings on the added value of RCMs over Central Asia, for instance from CORDEX.

We have modified the introduction a lot to make those points clear. We added a paragraph listing the few downscaling efforts done in Central Asian domain.

Line 11: 'we downscale CCLM' is not correctly phrased

We have changed it to "Additionally, we employ the CCLM to refine future climate projections."

Lines 16-18: The setup of the CNN training and evaluation, and the applications are not clearly explained in the abstract. Of course, the CNN emulator is model-specific, as it is designed to emulate a specific RCM.

We reformulated this part in the new version of the manuscript.

Line 25: Maraun and Widmann (2018) 'Statistical downscaling and bias correction for climate research', Cambridge University Press, is a standard reference for statistical downscaling and should also be cited.

That is true. We added this source in the new version.

Lines 101-102: If ML is used for postprocessing large-scale data separately for each time step, as is the case for the emulator presented, there is no iterative use of the output. It is therefore unclear how this comment relates to the study.

We have removed this sentence and explained the conservation process of the DL architecture we have used in this study. For example the following part is added to the new introduction:

"CNNs are adept at maintaining spatial coherence in the output, which is critical in downscaling where preserving the geographical patterns of climate variables (like precipitation) is crucial (Kurth et al., 2018). Constrained CNNs integrate physical constraints or laws directly into the training process. The constraining is done by changing the loss function or the network's architecture to enforce compliance with physical laws (i.e., conservation of mass, energy, or momentum). Unconstrained CNNs operate without explicitly incorporating physical laws or constraints into the network's architecture or loss functions. They focus solely on learning from the input data to the output predictions based on the data-driven patterns they detect. This study explores unconstrained and constrained CNN approaches to understand their effectiveness in downscaling and how they perform when applied to GCMs on which they were not initially trained."

Lines 153-159: It is not clear what is meant with 'high/low challenges to adaptation'.

we added the following paragraph to the new introduction: "Challenges to adaptation refer to the degree of difficulty that societies might face in adjusting to the environmental, economic, and social impacts of climate change. Specifically, this term refers to a society's fundamental susceptibility and the accessibility and efficacy of technologies and approaches designed to lessen the impacts of climate change. The adaptation challenges are minimal in the SSP126 scenario, which envisions a sustainable future. This implies that, under this scenario, global cooperation and sustainable practices lead to advancements in technology and governance that significantly reduce vulnerability to climate change impacts. Additionally, societal structures are resilient, and resources are managed to minimise environmental stresses and maximise human well-being."

Section 2.1.1: There should be a reference to Fig. 1a to specify the CCLM domain, and in the figure caption 'study region' should be replaced with 'CCLM simulation domain'.

We added this into the new version.

Section 2.1.2: The setup for the CNN training is not fully clear. What are the simulation periods for the scenario runs? Are the input and output variables daily precipitation?

We implemented this data in the new version of the manuscript.

Line 161-162: If input and output are precipitation, what is the meaning of energy and mass conservation in this context? Is 'energy' and 'mass' in this context the same or are these different quantities? It turns out later that the details are given in appendix A and that for the hard constraint the meaning of 'mass and energy conservation' is that the precipitation over a GCM gridcell is conserved in the high-resolution precipitation. However, the main part needs to be self-contained and written such that it is not confusing. Therefore, a short explanation of the main aspects of the constraints and a reference to the appendix should be given here.

We agree with this point and therefore moved this part to the main part of the manuscript. We explain each layer in the main part of the paper with references to the new figure 2. Description of the constraint layers are now presented in section 2.1.3.

Appendix A is very unsystematic and unclear. Specific problems are listed in the next five points. Line 417: The simplest way to ensure mass conservation would be to scale all small-scale values within a given large-scale gridcell with the ratio of the large-scale value and the sum of the small-

scale values. Why is this not done and what is the reason for the specific choice using the exponential dependency of the scaling factors on the small-scale values?

Thanks a lot for this point, this might need further explanation. The way of scaling that you are describing is also one possibility described in Harder et al., 2023. The work showed though that using the softmax constraints layer gives better results. The exponential both ensures positive predictions and leads to more variance between subpixels in the super-resolved prediction. The multiplicative rescaling you are describing e.g. struggles when the sum of the small-scale values gets close to zero.

Line 418: Why is the MAE loss function mentioned in the context of the hard constraint, which in the way it is formulated does not depend on the loss function for the CNN? It is said already in line 192 that MAE is the loss function for the CNN.

That's correct, the hard constraints layers dont depend on the loss function. There is a formulation of the soft constraining, which includes an addional term in the loss function. We have removed that part in the new version of the manuscript.

Lines 419-424: Is this the loss function for the whole CNN, or is it only relevant for the constraint layer? If it is the former, this contradicts the statement in line 192. If it is the latter, the use of two loss functions needs to be explained. How are the y_i calculated from the values in the previous layer? Why is there an explicit version for calculating y_i for the hard constraint (eqn. A2) but not for the soft constraint?

The loss function is always for the whole CNN. If there are hard constraints are included there is no change in the loss function at all. Next to hard constraints there also exist soft constraints, which is a way of enforcing constraints through an additional term in the loss function. Soft constraints and hard constraints are fundamentally different as hard constraints build on reajusting the outputs during training and inference at the end of the neural network through a non-trainable layer. This is an explicit function that is applied. Soft constraints are not enforced explicitly but through the penalizing term mentioned above.

Lines 425-426: It is not clear why a loss function is mentioned if there is no constraint layer. It is already said in line 192 that MAE is the loss function used for the whole CNN.

The loss function is mentioned because it is used for the soft constraints and changed for this.

Lines 428-429: This sentence says that MAE is an evaluation criterion for the different settings. This is confusing. Are the 'loss function' and the 'evaluation criterion' used differently? If so, it needs

to be explained how, or the statement on the evaluation criterion needs to be moved to the 'metrics' section.

The MAE is both used as a loss function and an evaluation metric. A loss function is used during the training to optimize the neural networks parameters, while an evaluation metric is calculated on the validation or test data set to evaluate the model on an independent dataset. Those are two different use cases, but both can use an MAE. This information is added into the new version for clarifying.

Lines 162-171: Although GCM errors affect the output, the emulator is a statistical model, which should be not very sensitive to the states used for fitting, otherwise there is the usual problem of stability of statistical relationships in statistical downscaling. The phrasing 'introduce biases in the downscaling process' is misleading, as the 'downscaling process' is the CCLM model or the CNN emulator, not the output. The discussion in this part is conceptually unclear as in conflates models and outputs. The fact that biases and errors of the CCLM and CNN output are partly caused by propagation of GCM biases and errors needs to be taken into account in the evaluation.

We changed the whole paragraph into a new one :

"In the context of deep learning for climate modelling, the 'perfect model' approach involves starting with high-resolution data, which is considered accurate or nearly perfect, and intentionally degrading it to a lower resolution. The aim is to simulate a scenario where the 'truth' (the original high-resolution data) is known, and then to recover this high-resolution from the artificially degraded data using deep learning techniques. This approach is a crucial part of training, as it teaches the model the desired mapping from low to high resolution, enabling the model to effectively learn how to upscale or enhance resolution while minimizing the loss of critical information. It's a controlled experiment that helps refine the model's capabilities.

The "imperfect model" approach, on the other hand, acknowledges that both the low-resolution (GCM output) and the high-resolution (RCM output) datasets have their inherent errors and limitations. In this scenario, we do not have a single source of truth but rather two separate sets of data:

- Low-resolution data: This may capture global or large-scale phenomena but miss regional details (Xu et al., 2021; Chokkavarapu and Mandla, 2019).

- High-resolution data: This provides detailed regional information but may still have errors or not perfectly reflect reality due to limitations in data collection, model configuration, or computational constraints (Muttaqien et al., 2021).

In this setup, the challenge for deep learning is to learn a mapping between these two independently imperfect data sets. With using the CNN we try to train a model that can predict high-resolution details from low-resolution inputs as accurately as possible despite the absence of a perfect ground truth. This involves understanding and modeling the uncertainties and biases inherent in both datasets."

Lines 189-190: The statement that the unconstrained CNN works best is based on the evaluation, and the performance ranking depends in principle on the evaluation measures. This is the method section and the reader does not know yet what the evaluation measures are. The statement on the best architecture should be moved to the result section and it should be made clear that the ranking of methods can depend on the evaluation measures.

We agree and have moved this part to the results and explain it in a right order now.

Lines 196-206: How different are the CHIRPS, APHODITE and CPCC data on the coarser grids for which they are all available?

Based on the point raised by reviewer 1, we have removed this lines from the new version of the manuscript.

Lines 210-215: It would be good to also do the evaluation on the coarse grid, or at least comment why this is not done. See also Volosciuk et al. (HESS 2017) 'A combined statistical bias correction and stochastic downscaling method for precipitation' for a discussion on the distinction between biases and differences in statistical properties on different spatial scales.

We did not initially include an evaluation on the coarse grid primarily due to our specific focus on the impacts at a higher spatial resolution, which is more relevant for local climate adaptation strategies, like the work of Russo et al. (2020). Nevertheless, to enhance the robustness of our findings and provide a comprehensive analysis, we will include a discussion on the implications of not evaluating at coarser resolutions in the revised manuscript. This discussion will reference the findings of Volosciuk et al. (2017) to provide a theoretical framework for understanding the scale-dependent nature of biases and differences in statistical properties.

We added the following paragraph to the discussions : "Our study evaluated the downscaling skill primarily using higher resolution observations, which are critical for capturing localized climate phenomena relevant to regional adaptation strategies. However, as Volosciuk et al. (2017) noted, examining downscaling outputs at coarser resolutions can be equally informative. Their work emphasizes that downscaling methods can introduce or fail to correct biases that differ significantly across spatial scales. By evaluating on a coarser grid, it is possible to distinguish between the inherent biases of the model and those introduced by the downscaling process. This distinction is crucial for understanding the limitations and strengths of downscaling methods in representing climatic variables across different scales."

Lines 216-223: The evaluation is very limited because it only addresses temporal variability, and only with one specific measure. Other measures for the agreement of simulated and observed temporal variability could also be used such as correlations of the timeseries or Brier Score for threshold exceedances. Differences in distributions, including the bias in the mean, potentially also in quantiles should also be analysed. The various aspects of evaluation are discussed for instance in Maraun et al. (Earth's Future 2015) 'VALUE: A framework to validate downscaling approaches for climate change studies'.

Indeed we show correlations as numbers in the paranthesis within new Figure 9, new Figure 10 c and d, whenever we do not shuffle the datasets. We also added the simple bias maps in new figure 5. The current model set-up has already been evaluated in a previous study i.e. Russo et al. (2020). Here, we apply that setting for creating the CMIP6-based CORDEX-CA.

Moreover, the terminology is not correct. A bias is a systematic difference between a statistical variable calculated from two datasets. It often refers to variables that characterise distributions (such as the mean, variance, or quantiles), but can also be used for variables that characterise temporal variability (such as autocorrelation or spectra) or spatial variability (such as correlation lengths, for instance Widmann et al. (IJC 2019) 'Validation of spatial variability in downscaling results from the VALUE perfect predictor experiment'. It is not common practice to use the term bias to characterise the agreement of individual time steps, and therefore MAE should not be called bias.

This issue was raised also by reviewer 1 and we agree with it. We now show both MAE and the simple bias maps. Please see answer to reviewer 1.

Section 3.1.1: If I understand correctly the MAE is calculated based on pairs of daily simulated and observed values. If so, this approach is fundamentally wrong, because the precipitation series are realisations of random internal variability, which are different in the observations and in the GCM simulations or GCM-driven RCM simulations. This is different for reanalyses and reanalysis-driven RCM simulations because of the data assimilation in reanalyses. MAE is based on pairs of values for a given time and a measure for how similar the timeseries are. It makes no sense to calculate MAE

for non-synchronised timeseries, because there is no justification for the paring of values. In this situation the MAE is only affected by the difference in variance and provides no meaningful measure of agreement of the specific temporal behaviour. This section and Fig. 3 should therefore be deleted.

We completely agree on this point and removed the subsection and its figures.

Lines 259-261: The GCM biases that affect the climate change signal in RCM simulations are not the MAE for short-term variability but systematic biases for instance in the large-scale mean circulation for the present climate, which may be linked to unrealistic large-scale climate change. The links between biases and climate change signals are complex, and should be discussed more carefully.

As mentioned above, this section is removed from the new version of the manuscript.

Line 322: Project reference is missing

We corrected that.

Lines 351-353: This argument is missing the main point. CCLM is driven at the lateral boundaries by the GCM values for the state variables of CCLM (temperature, pressure, wind speed etc.). Precipitation is not used for driving the RCM. The CNN input is the GCM precipitation, which has different biases in the two GCM, and therefore the mapping from the MPI-GCM-precipitation to the CCLM precipitation cannot be successfully transferred to EC-Earth.

We added this point in the new version.

Discussion and conclusions: This section needs to be rewritten after the issues listed above have been addressed.

We have rewritten those sections.

Line 516: Reference for the Harder et al. 2022 preprint should be updated to the peer-reviewed version Harder et al. 2023.

Done.

On behalf of all authors,

Bijan Fallah

References

Chokkavarapu, N. and Mandla, V. R.: Comparative study of GCMs, RCMs, downscaling and hydrological models: a review toward future climate change impact estimation, SN Applied Sciences, 1, 1698, 2019.

Demory, M.-E., Berthou, S., Fernández, J., Sørland, S. L., Brogli, R., Roberts, M. J., Beyerle, U., Seddon, J., Haarsma, R., Schär, C., et al.: European daily precipitation according to EURO-CORDEX regional climate models (RCMs) and high-resolution global climate models (GCMs) from the High-Resolution Model Intercomparison Project (HighResMIP), Geoscientific Model Development, 13, 5485–5506, 2020.

Hess, P., Drüke, M., Petri, S., Strnad, F. M., and Boers, N.: Physically constrained generative adversarial networks for improving precipitation fields from Earth system models, Nature Machine Intelligence, 4, 828–839, 2022.

Hodson, T. O.: Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not, Geoscientific Model Development, 15, 5481–5487, 2022.

Kendon, E., Roberts, N., Fowler, H., Roberts, M., Chan, S., and Senior, C.: Heavier summer downpours with climate change revealed by weather forecast resolution model. Nat. Climate Change, 4, 570–576, 2014.

Muttaqien, F. H., Rahadianti, L., and Latifah, A. L.: Downscaling for Climate Data in Indonesia Using Image-to-Image Translation Approach, in: 2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 1–8, IEEE, 2021.

Rasp, S. and Lerch, S.: Neural networks for postprocessing ensemble weather forecasts, Monthly Weather Review, 146, 3885–3900, 2018.

Russo, E., Sørland, S. L., Kirchner, I., Schaap, M., Raible, C. C., and Cubasch, U.: Exploring the parameter space of the COSMO-CLM v5. 0 regional climate model for the Central Asia CORDEX domain, Geoscientific Model Development, 13, 5779–5797, 2020.

Sun, L. and Lan, Y.: Statistical downscaling of daily temperature and precipitation over China using deep learning neural models: Localization and comparison with other methods, International Journal of Climatology, 41, 1128–1147, 2021.

Volosciuk, C., Maraun, D., Vrac, M., and Widmann, M.: A combined statistical bias correction and stochastic downscaling method for precipitation, Hydrology and Earth System Sciences, 21, 1693–1719, 2017.

Xu, Z., Han, Y., Tam, C.-Y., Yang, Z.-L., and Fu, C.: Bias-corrected CMIP6 global dataset for dynamical downscaling of the historical and future climate (1979–2100), Scientific Data, 8, 293, 2021.

Climate Model Downscaling in Central Asia: A Dynamical and a Neural Network Approach

Bijan Fallah[1], Christoph Menz[1], Emmanuele Russo[2], Paula Harder[3], Peter Hoffmann[1], Iulii Didovets[1], Masoud Rostami[1,5], and Fred F. Hattermann[1,4]

[1]Potsdam Institute for Climate Impact Research (PIK), P.O. Box 601203, 14412 Potsdam, Germany

[2]ETH Zürich, Department of Environmental Systems Science, Universitätstrasse 16, 8092 Zürich, Switzerland

[3]Mila Quebec AI Institute, Montreal, Canada

[4]Eberswalde University for Sustainable Development (HNEE), Germany

[5]Laboratoire de Météorologie Dynamique (LMD), Sorbonne University (SU), Ecole Normale Supérieure (ENS), Paris, France

Correspondence: Bijan Fallah (fallah@pik-potsdam.de)

Abstract. To estimate future climate change impacts, usually high-resolution climate projections are necessary. Statistical and dynamical downscaling or a hybrid of both methods are mostly used to produce input datasets for impact modelers. In this study, we use the regional climate model (RCM) COSMO-CLM (CCLM) version 6.0 to identify the added value of dynamically downscaling a general circulation model (GCM) from the sixth phase of the Coupled Model Inter-comparison Project (CMIP6) and its climate change projections' signal over Central Asia (CA). We use the MPI-ESM1-2-HR (at 1° spatial resolution) to drive the CCLM (at 0.22° horizontal resolution) for the historical period of 1985-2014 and the projection period of 2019-2100 under three different shared socioeconomic pathways (SSPs): SSP1-2.6, SSP3-7.0 and SSP5-8.5 scenarios. Using the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) gridded observation dataset as reference, we evaluate the ~~CCLM performance over the historical period using a simulation~~ performance of CCLM driven by ERAInterim reanalysis over the historical period. CCLM's added value, compared to its driving GCM, is significant (more than 5mm/day) over CA mountainous areas, which are at higher risk of extreme precipitation events. ~~Furthermore, we downscale the CCLM for~~ Additionally, we employ the CCLM to refine future climate projections. We present high-resolution maps of heavy precipitation changes based on CCLM and compare them with CMIP6 GCMs ensemble. Our analysis shows a significant increase in heavy precipitation intensity and frequency over CA areas that are already at risk of extreme climatic events in the present day. Finally, ~~applying our single model high-resolution dynamical downscaling,~~ we train a convolutional neural network (CNN) to map ~~the low-resolution GCM simulations to the~~ a GCM simulation to its dynamically downscaled CCLM~~ones~~. We show that ~~applied~~

the CNN could emulate the GCM-CCLM model chain over large CA areas. ~~However, this specific emulator has shortcomings~~ This emulator has added values when applied to a new GCM-CCLM model chain. ~~Our downscaling data and the pre-trained CNN model could be used by scientific~~ Scientific communities interested in downscaling CMIP6 models ~~and searching for a trade-off between the~~ could use our downscaling data. The CNN architecture can be applied as an alternative to dynamical and statistical methods.

1 Introduction

~~It is very well acknowledged that the~~ The increasing global mean temperature ~~is increasing~~ due to anthropogenic greenhouse gas emissions ~~(Allan et al., 2021). The most critical~~ presents a significant challenge for society~~is to assess and predict the future impact of this warming on the~~ , requiring the assessment and prediction of future impacts on human health, natural ecosystems, and ~~economy for~~ economies across different regions of the ~~World. Studies of~~ world (Allan et al., 2021). Regional studies on vulnerability, impacts~~and adaptation at the regional scale require~~ , and adaptation necessitate reliable high-resolution climate projections~~(Maraun et al., 2015)~~, which are ~~based on dynamical downscaling through RCMs(Rummukainen, 2010; Feser et al., 2011)~~ typically achieved through dynamical downscaling via Regional Climate Models (RCMs) (Rummukainen, 2010; Feser et al., 2011) , statistical techniques ~~(Fowler et al., 2007) or hybrid approaches using both (Meredith et al., 2018; Laflamme et al., 2016)~~ (Maraun and Widmann, 2018; Fowler et al., 2007), or a hybrid of both approaches (Maraun et al., 2015; Meredith et al., 2018; Laflamme e .

~~Countries should develop an adaptation and mitigation strategy to cope with potential future risks of climate change. Usually, climate projections are used as the basis for decision-making in spending financial resources on infrastructure, society, and environments (Maraun et al., 2015).~~ Central Asia (CA)~~is assumed to be~~ , recognized as one of the most vulnerable regions to climate change impacts~~. CA's water resources depend on water~~ , is heavily dependent on water resources from glaciers and rivers that are shrinking due to rising temperatures and decreasing precipitation ~~(Reyer et al., 2017). Food securityis at severe risk in CA with reduction of crop yields due to climate change (Allan et al., 2021). Extreme~~ (Reyer et al., 2017; Fallah et al., 2023; Didovets . The area faces significant challenges to food security, characterized by declining crop yields and an increased occurrence of severe and frequent extreme weather events like floods and landslides~~are happening more frequently and intensively in the region leading to severe damage to infrastructures, livelihoodsand crops, subsequently causing~~ . These conditions damage infrastructure, livelihoods, and agriculture, resulting in population displacement and migration ~~(Reyer et al., 2017)~~(Allan et al., 2021; Reyer .

~~Given the above-mentioned~~ Despite these critical concerns, the ~~impact modelling is still hindered in CA, based on the lack~~ development of high-resolution climate projections ~~but also on the elevated level of uncertainty~~ in CA is impeded by the significant uncertainties inherent in the existing high-resolution observational and reanalysis datasets ~~. Motivated by these challenges, in this manuscript we produce a dynamically downscaled state of the projected climate over CA from a single GCM of the~~ (Fallah et al., 2016a). Dynamical downscaling of CMIP6 ~~project. In some cases, for properly reproducing~~ models for the CA region is vital for accurately simulating extreme convective precipitation events~~and local topographical~~

~~effects, downscaling is essential for representing local dynamics (Kendon et al., 2014; Demory et al., 2020). Various factors, such as~~ , which are influenced by the orography of the region, ~~the~~ large-scale atmospheric circulation, ~~the~~ and sea surface temperature anomalies in ~~Indian Oceans~~ the Indian Ocean and the Pacific ~~, and the soil moisture feedback influence convective precipitation events in central Asia (Xu et al., 2022). The main goal of dynamical downscaling is to improve~~ (Kendon et al., 2014; Demory et al., 2020; Xu et al., 2022). Dynamical downscaling enhances the resolution of a driving ~~Global Circulation Model (GCM) and produce a robust~~GCM and produces a robust, physically consistent regional state of the climate. ~~This is often considered a critical point for preferring the use of RCMs to~~ High-resolution atmospheric models have been shown to have better skills over complex topographies in estimating variables like precipitation than in situ observations, satellite-derived and radar datasets (Lundquist et al., 2019). Many studies confirm that RCMs can better represent small-scale atmospheric features, especially for precipitation over complex topographies (Ban et al., 2015; Wang et al., 2013; Frei et al., 2003) . This method is often preferred over statistical downscaling approaches~~that rely on the assumption that statistical relationships found for the present also hold true for the future . Dynamical downscaling reproduce a wide range of local physical processes, especially important for the representation of precipitation (Hess et al., 2022). Traditional statistical downscaling approaches are based on model output statistics and try to improve the spatial resolutions based on statistical relationships and not dynamical processes (Hess et al., 2022; Lange, 2019). The resulting statistically-downscaled data usually lacks physical consistency and might be too smooth (Lange, 2019; Fallah et al., 2023). On the other hand~~, which assume that present statistical relationships will hold in the future (Hess et al., 2022). However, RCMs are computationally demanding ~~, especially at the very high resolutions useful for impact studies. At the same time, they suffer from~~ and inherit a 'cascade ~~' of uncertainties that must be taken into account prior to the performance of climate projections . In order to improve the models inter-comparability and to provide a robust, validated benchmark for the performance of high-resolution climate projections using RCMs , over the years members from different~~ of uncertainty', meaning that the uncertainties in the models will expand from one step or chain to another, highly affecting RCM outcomes and must be considered prior to performing climate projections (Mitchell and Hulme, 1999; Sørlar . Despite these considerations, the added value of RCMs concerning their driving GCM is constantly debated in the community and is highly dependent on the driving GCM (Jacob et al., 2012; Lenz et al., 2017; Fotso-Nguemo et al., 2017; Di Luca et al., 2012, 2015) . An RCM is tuned to perform over the target local region. However, a GCM is tuned to represent energy and water balance globally (Sørland et al., 2018).

Various international institutions have ~~joined forces into~~ collaborated within the Coordinated Regional Climate Downscaling Experiment (CORDEX) ~~-~~

~~CORDEX is a program sponsored by the World Climate Research Program (WCRP) aimed at developing an improved framework for generating regional-scale climate projections for impact assessment and adaptation studies worldwide within~~ to address these issues and improve the models' inter-comparability. CORDEX aims to create a better framework for producing climate projections at a regional scale that is suitable for impact evaluation and adaptation planning globally, aligned with the timeline of the Intergovernmental Panel on Climate Change Sixth Assessment Report (Kikstra et al., 2022)~~timeline and beyond. CORDEX aims to produce regional climate projections and to evaluate their performance through different experiments. The usage of CORDEX-like simulations must be adapted to the needs of the impact modelling. CORDEX data are often affected~~

~~by diverse sources of uncertainty: systematic biases in the driving GCM and RCM itself, uncertainty in scenarios, the model internal variability, model-specific response to driving GCM's boundary forcing and a small population of RCM simulations. We might underestimate/overestimate the uncertainty if the sample is too small (Hewitson et al., 2014).~~

~~Unfortunately, most of the research conducted in/for the CORDEX initiative~~ . However, most CORDEX research focuses on highly industrialized countries ~~(Allan et al., 2021), and fewer institutes run RCM simulations over CA (refer to). Sadly, the developing countries (CA included) are the ones who will suffer the most from the consequences of global warming (Naddaf, 2022). In particular, only two CORDEX model simulations are available to date for CA, driven by the fifth phase of the coupled model intercomparison project (CMIP5) GCMs (Taylor et al., 2012). On the other hand, no~~ (Allan et al., 2021; Taylor et al., 2012) . No simulation (except this study) driven by the CMIP6 model simulations has been planned so far for ~~CA~~ CORDEX-CA (see https://wcrp-cordex.github.io/simulation-status/CMIP6_downscaling_plans.html, last visited on ~~14.08.2023). One motivation to conduct dynamical downscaling, especially over areas with complex topography, as in CA, is that high-resolution atmospheric models have been shown to have better skills in estimating variables like precipitation than in situ observations, satellite-derived and radar datasets (Lundquist et al., 2019). Many studies confirm that RCMs can better represent small-scale atmospheric features, especially for precipitation over complex topographies (Ban et al., 2015; Frei et al., 2003).~~

~~Despite these considerations, the added value of RCMs concerning their driving GCM is constantly debated in the community and is highly dependent on the driving GCM (Jacob et al., 2012; Lenz et al., 2017; Fotso-Nguemo et al., 2017; Di Luca et al., 2012, 2015) . An RCM is tuned to perform over the target local region. However, a GCM is tuned to represent energy and water balance globally (Sørland et al., 2018). Additionally, there is a debate in the community on whether the GCM-RCM chain might suffer from a "cascade of uncertainty", meaning that the uncertainties in the models will expand from one step or chain to another (Mitchell and Hulme, 1999; Sørland et al., 2018), highly affecting RCM outcomes. A significant advantage of the high-resolution RCMs is the use of high-resolution surface forcings like the topography, land use and land cover, soil type, and coastlines (Hong and Kanamitsu, 2014).~~

~~Here, we focus on the added value of the dynamical downscaling for precipitation. Precipitation is one of the most critical variables in vulnerability, impacts and adaptation studies (Jacob et al., 2012). Mountain precipitation is especially vital for studying floods and water availability in the field of hydrology (Smith et al., 2010). Extreme daily precipitation is one of the primary triggers of landslide events in CA , especially in Tajikistan and Kyrgyzstan (Wang et al., 2021). On the other hand, precipitation simulation is challenging for any climate model (Russo et al., 2019). RCMs have been shown to potentially add value in simulating mesoscale convective precipitation, coastal rainfall, and extreme rainfall events (Giorgi and Gutowski Jr, 2015; Russo et~~ ~~.~~ 17.04.2024). Sadly, developing countries, including CA, bear the brunt of global warming's consequences, with only a limited number of CORDEX model simulations available for this region (Naddaf, 2022). The dynamical downscaling in CA can provide detailed insights into regional climate phenomena often not captured by coarser-resolution global models (Russo et al., 2019). Climate projections might be sensible to different parameter settings, emphasizing the need for careful calibration and validation of regional models. Dynamical downscaling's added value lies in its ability to tailor climate projections more closely to regional specifics, thereby improving the utility of climate data for regional climate change impact assessments

(Russo et al., 2020). Despite some systematic biases, dynamical downscaling consistently provides high-quality datasets that accurately describe the climatology of all climate variables in CA (Qiu et al., 2022).

~~In recent years, machine learning (ML) approaches like convolutional neural network (CNNs) have emerged as a promising statistical downscaling tool due to their ability to learn~~ Beyond dynamical methods, recent developments in machine learning, including CNNs as the most popular choice, offer promising and potentially transformative avenues for statistical downscaling (Harder et al., 2023; Rampal et al., 2024). CNNs have proven effective in numerous earth science disciplines besides downscaling, such as classification (Gardoll and Boucher, 2022), segmentation (Galea et al., 2024) and prediction (Watson-Parris et al., 2022) thanks to their capacity to extract features from spatial data and ~~capture~~ identify non-linear ~~mappings~~ relationships between inputs and outputs~~(Sun and Lan, 2021). Unlike the point-wise approaches, they apply an image-to-image translation which might reduce the spatial intermittency problems of post-processing methods (Rasp and Lerch, 2018). CNNs have been successfully applied to various tasks in computer vision, natural language processing, and image super-resolution. In climate science, CNNs have been used for statistical downscaling of temperature and precipitation over different regions and time scales, using distinct types of predictors and predictands (Baño-Medina et al., 2021; Serifi et al., 2021; Yang et al., 2023; Sun and Lan, 2021; Hess et al., 2022). Super-resolution (SR) in ML tries to increase the resolution of images or videos and preserve their content and details. The task is challenging because SR involves recovering high-frequency information lost or degraded in low-resolution images or videos (Dong et al., 2015). ML can generate high-resolution data that looks realistic and has good accuracy in prediction. However, when ML is applied to a physical system like the Earth's atmosphere, it may face a significant challenge: the predicted output values may need to obey physical laws such as energy, momentum, and mass conservation. These violations of constraints can be harmful - causing errors that may accumulate as climate models iteratively run on their own output (Harder et al., 2022). If there exists a physical relationship between low-resolution~~ .

CNNs can recognize and encode spatial hierarchies in data (Zhu et al., 2017), making them exceptionally suitable for geospatial data, which is fundamental in climate modelling. Unlike traditional statistical methods that often require manual selection and careful engineering of features, CNN automatically learns the most predictive features directly from the data (Reichstein et al., 2019). CNNs can model complex non-linear relationships between input data and ~~high-resolution datasets via some equations, one could enforce physical constraints between the datasets. This could be achieved by adding a constraint layer at the end of a neural networkarchitecture (Harder et al., 2022). Therefore, we could guarantee that we employ physical constraints (like mass and energy conservation) in the prediction.However, in the GCM-RCM chain, unlike many statistical methods that try to re-distribute the precipitation amount from a coarse grid box to nested finer ones (Lange, 2019), precipitation might not follow the mass conservation. The RCM has its internal variability and lends information from a GCM only at its boundaries. In an unconstrained set-up, a CNN might be able to learn the hidden physical mappings between the RCM and its driving GCM. Therefore, we will explore both the~~ outputs, often present in climate data due to intricate interactions in weather systems. CNNs are generally more straightforward and efficient for tasks that aim to predict or classify based on patterns distributed across the spatial domain, such as temperature or precipitation patterns in climate models (Racah et al., 2017). CNNs are adept at maintaining spatial coherence in the output, which is critical in downscaling where preserving the geographical patterns of climate variables (like precipitation) is crucial (Kurth et al., 2018). Constrained CNNs integrate physical constraints or laws directly into the training process. The constraining is done by changing the loss function or the network's architecture to enforce compliance with physical laws (i.e., conservation of mass, energy, or momentum). Unconstrained CNNs operate without explicitly incorporating physical laws or constraints into the network's architecture or loss functions. They focus solely on learning from the input data to the output predictions based on the data-driven patterns they detect. This study explores unconstrained and constrained CNN approaches ~~,~~ to understand their effectiveness in downscaling and how they perform when applied to GCMs on which they were not initially trained.

The research questions guiding this study are:

– Research Question 1: How effectively can CMIP6 models be downscaled for the CORDEX Central Asia region to enhance precipitation simulations?

– Research Question 2: Can convolutional neural networks (CNNs) effectively downscale GCM outputs, and how do they perform when applied to GCMs they were not initially trained on?

~~Our final goal is to explore a hybrid framework using dynamical downscaling and deep learning to enhance the spatial resolution of GCM-like climate datasets.~~ The ~~tested methodology could be easily and rapidly applied to new climate datasets. Since the dynamical downscaling approaches have high computational costs and require hardware capacities (thousands of central processing units) , scientists, especially impact modelers, must find trade-offs between the dynamically constraint and statistical downscaling methods. Therefore, our study would be a good starting point to test the idea of training the CNN onthe dynamical chain of a single GCM-RCM to find physical relationships between the coarse state of a GCM and the finer state of an RCM. By finding an emulator for a specific GCM-RCM chain, we could apply it to different time periods and forcings, but for the same GCM. Therefore, the~~ manuscript will focus on three main topics: 1-added value of CCLM for the representation of precipitation over CA, 2-dynamical downscaling signal of CCLM for heavy precipitation and 3-training a CCLM emulator using a CNN. We present data and methods in section 2. The results of dynamical and hybrid downscaling are introduced in section 3 and 4, respectively. Finally, we discuss the results and draw conclusions in section 5.

2 Data and methods

The schematic shown in figure 1) depicts the methodology used in this study. In the following we will explain it in more details.

2.1 Employed Models and Experimental Setups

2.1.1 RCM

In our study, we conduct a series of simulations with the COnsortium for Small scale Modelling in CLimate Mode (~~COSMO-CLM~~CCLM) RCM. ~~COSMO-CLM~~ CCLM is a regional climate model developed by the German Weather Service (DWD) and the German Climate Computing Center (Deutsches Klimarechenzentrum, DKRZ) ~~in Germany (Rockel and Geyer, 2008)~~ from the COSMO numerical weather prediction model (Rockel and Geyer, 2008), widely used for short-term weather forecasting. The original core of COSMO-CLM or CCLM, was called Local MOdel (LM), developed by DWD for weather forecasting. The adopted LM version for climate purposes formed the ~~COSMO-CLM (Böhm et al., 2003). COSMO-CLM~~ CCLM (Böhm et al., 2003) . CCLM is designed to simulate the regional climate at high spatial resolution, allowing researchers to study various aspects of the climate system, such as temperature, precipitation, and extreme events. CCLM has been utilized in numerous studies to evaluate the impact of climate change on various regions, including Europe (Russo et al., 2021), Africa (Panitz et al., 2014; Dosio and Panitz, 2016), and Asia (Jacob et al., 2014; Kotlarski et al., 2014; Wang et al., 2013). It has also been used for climate projection studies and to assess the effectiveness of climate adaptation and mitigation strategies. The model has been thoroughly evaluated and validated ~~(Russo et al., 2019; Kjellström et al., 2011)~~(Fallah et al., 2016b; Russo et al., 2019; Kjellström et al., 2011) . Its ability to produce realistic simulations of the current climate and its variability has made it one of the most widely used regional climate models in the scientific community (Sørland et al., 2021).

For our experiments, we ~~have~~ used a similar model set-up as the "optimal" set-up provided in the study of Russo et al. (2019). ~~We set up our simulations in accordance with CORDEX.~~ The CORDEX protocol requires a set of simulations that can be divided into two main groups. The first one, referred to as the evaluation run, consists of a single model experiment performed over the period 1979-2014, using ERAInterim at a spatial resolution of T255 ($\sim 0.7°$) as the driving data. In the second stream (projection), the models must run with boundary conditions from GCMs of the CMIP6 project for the period 1950-2100 under different SSPs (here, we have chosen a single GCM: MPI-ESM1-2-HR and SSP126, SSP370 and SSP585 scenarios). SSPs are baseline scenarios describing the future development pathways depending on population, technology and economic growth, urbanization, investment in healthcare and education, land use and energy (Riahi et al., 2017).

We have chosen the two available CORDEX-CA evaluation simulations from other models, driven by ERAInterim at 0.22° horizontal resolution, for comparison/evaluation of our RCM simulations, which are driven by ERAInterim for the evaluation period. The two simulations are 1) ERAInterim-RMIB-UGent-ALARO-0 (Giot et al., 2016) and 2) ERAInterim-GERICS-REMO2015 (Jacob and Podzun, 1997; Fotso-Nguemo et al., 2017).

2.1.2 ~~CNN~~CNNs

We create an emulator of CCLM using a CNN. We use the output of the ~~COSMO-CLM~~ CCLM Version 6.0 RCM, which is driven by the MPI-ESM1-2-HR GCM under four different scenarios ~~: historical, SSP126, SSP370 and SSP585~~(for 2019-2100). Historical is based on the data of greenhouse gas levels, land use, and other climate forcings from 1850 to 2014 that were observed. SSP126 ~~(Shared Socioeconomic Pathway 1 - RCP2.6)~~ represents a "green" future where global resources are protected, human well-being is improved, and income gaps are narrowed. This scenario has low challenges to adaptation and low greenhouse gas emissions. Challenges to adaptation refer to the degree of difficulty that societies might face in adjusting to the environmental, economic, and social impacts of climate change. Specifically, this term refers to a society's fundamental susceptibility and the accessibility and efficacy of technologies and approaches designed to lessen the impacts of climate change. The adaptation challenges are minimal in the SSP126 scenario, which envisions a sustainable future. This implies that, under this scenario, global cooperation and sustainable practices lead to advancements in technology and governance that significantly reduce vulnerability to climate change impacts. Additionally, societal structures are resilient, and resources are managed to minimise environmental stresses and maximise human well-being. SSP370 ~~(Shared Socioeconomic Pathway 3 - RCP7)~~ depicts a regional rivalry future where nationalism and regional conflicts prevail, global issues are ignored, and inequality is increasing. This scenario has high challenges to adaptation and high greenhouse gas emissions. SSP585 ~~(Shared Socioeconomic Pathway 5 - RCP8.5)~~ portrays a fossil-fueled development future where global markets are connected, technological progress is fast, but environmental policies are weak. This scenario has low challenges to adaptation and very high greenhouse gas emissions. As an additional dataset, we merge the ERA-Interim reanalysis and CCLM simulation driven by it (ERAInterim-CCLM) to our previous ~~simulations~~data pool of GCM and RCM (see Fig. 1). We then train our CNN model based on the architecture proposed by ~~Harder et al. (2022)~~Harder et al. (2023), which can incorporate physical constraints to ensure mass conservation and energy balance. We evaluate our model in the CA domain. ~~Using the GCM as low-resolution data may introduce biases and errors in the downscaling process because the GCM may not capture the regional features and variability of the climate system accurately (Xu et al., 2021; Chokkavarapu and Mandla, 2019). RCM itself is prone to different biases. Therefore, we have both an imperfect input and imperfect output. Upscaling the RCM (the so-called perfect model experiment) may reduce these biases and errors because the RCM can better represent the regional climate characteristics and feedbacks (Muttaqien et al., 2021). However, we are interested in the so-called~~ We have to note that we use not the whole GCM domain as input for the CNN but only the domain covering the CA (Fig. 3).

In the context of deep learning for climate modelling, the 'perfect model' approach involves starting with high-resolution data, which is considered accurate or nearly perfect, and intentionally degrading it to a lower resolution. The aim is to simulate a scenario where the 'truth' (the original high-resolution data) is known, and then to recover this high-resolution from the artificially degraded data using deep learning techniques. This approach is a crucial part of training, as it teaches the model the desired mapping from low to high resolution, enabling the model to effectively learn how to upscale or enhance resolution while minimizing the loss of critical information. It's a controlled experiment that helps refine the model's capabilities.

The "imperfect model" ~~set-up (Stengel et al., 2020; Leinonen et al., 2020), where the dynamical mapping from GCM to RCM is of higher interest.~~ approach, on the other hand, acknowledges that both the low-resolution (GCM output) and the high-resolution (RCM output) datasets have their inherent errors and limitations. In this scenario, we do not have a single source of truth but rather two separate sets of data:

- Low-resolution data: This may capture global or large-scale phenomena but miss regional details (Xu et al., 2021; Chokkavarapu and .

- High-resolution data: This provides detailed regional information but may still have errors or not perfectly reflect reality due to limitations in data collection, model configuration, or computational constraints (Muttaqien et al., 2021).

In this setup, the challenge for deep learning is to learn a mapping between these two independently imperfect data sets. With using the CNN we try to train a model that can predict high-resolution details from low-resolution inputs as accurately as possible despite the absence of a perfect ground truth. This involves understanding and modeling the uncertainties and biases inherent in both datasets.

Many regions of CA receive low precipitation throughout the year and the spatio-temporal variability of precipitation is large. One needs a large dataset of GCM output and the corresponding RCM with various precipitation patterns for training a CNN to find an RCM emulator that captures the mapping from GCM to RCM.

~~We~~ First, the daily datasets are shuffled randomly. We then have used a total number of 68141 (60%), 22714 (20%) and 22714 (20%) RCM simulation days for training, ~~testing and evaluation~~validation and testing, respectively. The low-resolution (GCM) and high-resolution (RCM) datasets ~~(GCM)~~ have 30×60 and 120×240 grid points over latitudes and longitudes, respectively. Therefore, the downscaling factor (N) is 4 in this case. For a complete explanation of the CNN architecture, we refer to the work of ~~Harder et al. (2022)~~ Harder et al. (2023) and the corresponding ~~GitHub repository at~~ Zenodo repository at https://zenodo.org/records/8150694 (last visited on 21st of June 2023). ~~Here, we briefly explain the architecture of the CNN used~~

Figure 2 shows the schematic of the standard CNN (without constraint layers) architecture used for two times up-sampling in this study~~:~~

. We briefly explain the steps shown in the schematic:

- ~~The input layer is a low-resolution (LR)image of size 30 × 60 with only one channel, i. e. , precipitation value in mm/day.~~ Conv (Convolutional Layer): Initially, these layers help in extracting various levels of features from the low-resolution images, such as edges, textures, and other relevant image details.

- ~~The first layer is a convolutional layer with 64 filters of size 3 × 3 × 1 and stride 1. The output is a feature map of size 30 × 60 × 64.~~ ReLU (Rectified Linear Activation Unit): This non-linear activation function is a key player in our model's learning process. It introduces non-linearity, outputting the input directly if it's positive; otherwise, it outputs zero. This intriguing function helps the network learn complex patterns efficiently.

- ~~The second layer is a sub-pixel convolutional layer with 256 filters of size 3 × 3 × 64 and stride 1. The output is a feature map of size 60 × 120 × 64.~~

- ~~The third layer is another sub-pixel convolutional layer with 256 filters of size 3 × 3 × 64 and stride 1. The output is a feature map of size 120 × 240 × 64.~~

- ~~The fourth layer is a convolutional layer with 1 filter of size 3 × 3 × 64 and stride 1. The output is high-resolution (HR) image of size 120 × 240 × 1.~~ TransConv (Transposed Convolutional Layer): This layer is crucial for the task of upscaling. It increases the spatial dimensions of the feature maps, performing a sort of learned interpolation. This reassures us about the model's ability to add details to the upscaled images based on the features extracted and processed in the earlier layers.

- ~~The fifth layer is an optional renormalization layer that applies a linear transformation to the HR image toensure that the total mass or energy is conserved between the LR and HR images.~~ ResBlock (Residual Block): They allow the model to learn corrections (or residuals) to the primary interpolated outputs, refining the details and adding high-frequency information that enhances the perceptual quality of the upscaled images. Adding original input features (from earlier layers) to the output of several convolutional layers ensures that no critical information is lost during processing.

~~For this work, we find the unconstrained CNN(NoCL) performing the best, most likely due to the significant mismatch between low-resolution and~~

2.1.3 Constraint layers

We test the CNN with three different constraining methods in the last CNN layer: 1- soft constraining (SCL), 2- hard constraining (HCL) and 3- without constraining (NoCL). For a detailed information on the settings used we refer to the work of Harder et al. (2023) . In the following, we explain briefly the three different constraining methodologies. The set-up of constraining is as following: consider a factor N for downscaling in all linear directions and let $n := N^2$ and y_i, $i = 1, ..., n$ be the high-resolution ~~samples. A description of the constraint layers can be found in the appendix, see 2.1.3~~patch values that correspond to low-resolution pixel x. The mass conservation law has the following form:

$$\frac{1}{n}\sum_{i=1}^{n} y_i = x. \tag{1}$$

Hard constraining: it uses the SoftMax constraining, which is a constraining for quantities like water content. It enforces the output to be non-negative. For constraining the predicted quantities, we use a SoftMax operator on the intermediate outputs of the neural networks before the constraining layer ($\tilde{y}_i$) and multiply it with the corresponding input pixel value x:

$$y_i = \exp(\tilde{y}_j) \cdot \frac{x}{\frac{1}{n}\sum_{i=1}^{n} \exp(\tilde{y}_i)}. \tag{2}$$

y_i is the final output after applying the constraints. We have used the mean absolute error (MAE) as the loss function.

Soft constraining: This is done by adding a regularization term to the loss function. The MAE loss then extended with an additional constraint violation (CV) loss term to:

$$\text{Loss} = (1-\alpha) \cdot MAE + \alpha \cdot \text{CV}, \tag{3}$$

where CV is the constraint violation, which is the mean-squared error over all constraint violations between an input pixel x and the super-pixel (high-resolution grid-cell) y_i:

$$\text{CV} = \text{MSE}(\frac{1}{n}\sum_{i=1}^{n} y_i, x) \tag{4}$$

We use the $\alpha = 0.99$ in this study.

Without constraining: In this setup we remove the constraining layer after the last convolutional layer in the CNN.

The constraint layers are applied at the end of the CNN architecture, and all satisfy the criteria that the resulting high-resolution patch conserves the values in low-resolution pixels. The performance of the different settings is assessed through the MAE.

We use the ~~MAE~~ mean absolute error (MAE) as the loss function. We use 160 epochs, with a batch size of 64 and a learning rate of 0.001 for training with HCL and NoCL; and 0.00001 for SCL. Training takes 15 hours on an NVIDIA Corporation Graphics Ampere 104 [GeForce Ray Tracing Texel eXtreme (RTX) 3060 Ti Lite Hash Rate] graphics processing unit (GPU). We use the same model set-up as in Harder et al. (2023), and the computational cost of the CNN is very high, therefore, we did not conduct any cross-validation in this study.

We must note that the MAE can be used as both a loss function and an evaluation metric. A loss function is used during training to optimize the neural network parameters, while an evaluation metric is calculated on the validation or test data set to evaluate the model on an independent dataset. Those are two different use cases, but both can use an MAE.

2.2 Evaluation ~~Data~~and testing

According to Ciarlo et al. (2021), the choice of observational data significantly influences the added value calculation of an RCM, as well as the extreme events detection. To reduce these issues, they recommended to use observations with a resolution comparable to the one of the model. Therefore, for assessing the added value of ~~COSMO-CLM~~ CCLM with respect to the driving ~~model~~GCM, we use the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) as our gridded observation. CHIRPS has a resolution of 0.05° and covers the area between 50°S-50°N. CHIRPS is based on satellite information and station data, and, in contrast to reanalysis data, it is independent of climate model simulations. Therefore, CHIRPS could be an excellent alternative to similar but not identical coarse datasets like Global Precipitation Climatology Centre (GPCC) (Becker et al., 2013) for data-sparse regions with convective rainfall (Funk et al., 2015). ~~The APHRODITE's (Asian Precipitation-Highly-Resolved Observational Data Integration Towards Evaluation) dataset might be another alternative of an evaluation dataset. However, the merged domain version which could be used for our study, covering the period 1950-2007, is available only at 0.25° and 0.5° horizontal resolutions (Yatagai et al., 2007).~~

For ~~evaluating~~ testing the CNN methods, instead of ~~using~~ CHIRPS, we use the corresponding CCLM simulation (20% of the data, as mentioned above) as our target~~and~~. We calculate the metrics on ~~CNN and~~ the CNN and interpolated GCM outputs with respect to CCLM output.

2.3 ~~Metrics~~

~~In a first step, the~~ The selected GCM, RCM and observational data ~~is~~ are interpolated onto the RCM grid using the distance-weighted average method. Interpolation of the coarser grid to a higher resolution ~~one~~ might create unrealistic values. This issue was discussed in the work of Ciarlo et al. (2021). Usually, the interpolation does not account for the physical processes and constraints that govern the original data, the statistical properties (like mean, variance and skewness) are not preserved, and it introduces ~~artifacts~~ artefacts and errors that depend on the choice of interpolation method, the spatial distribution of the data points and the resolution ratio. Therefore, dynamical/statistical downscaling is used to increase the resolution of the climate data, and we use simple interpolation as a baseline in our study.

Since precipitation does not follow a normal distribution, following Hodson (2022), we use the ~~mean absolute error (MAE)~~ MAE to explore the bias of the ~~simulations~~ emulated and dynamically downscaled precipitation (F) against observations(O):

$$\mathrm{MAE} = \cancel{\frac{1}{N}}\frac{1}{T}\sum_{t=1}^{T}|F_t - O_t| \tag{5}$$

where ~~N~~ T is equal to the number of time steps~~(T)~~. We quantify the added value (AV) as the ability of the downscaling approach to decrease the MAE of the driving GCM when calculated against the reference dataset (CHIRPS or target CCLM simulation), i.e.~~,~~:

$$\mathrm{AV} = \mathrm{MAE}_{GCM} - \mathrm{MAE}_{CCLM} \tag{6}$$

where MAE_{GCM} and MAE_{CCLM} are the ~~biases of~~ differences of interpolated GCM and RCM with respect to the reference dataset.

As an additional metric we also use the climatological bias, i.e., the difference between the model and observations:

$$\mathrm{BIAS} = \mathrm{PR}_{MODEL} - \mathrm{PR}_{OBS} \tag{7}$$

3 Results

Figure 3.a shows the topography of the CORDEX-CA simulation domain. Figure 1.b presents the ~~annual climatology~~ mean daily precipitation averaged over all years (mm/day) ~~of daily precipitation~~ as derived from CHIRPS data for the period 1985-2014. The regions with the highest values of precipitation are the mountainous areas of CA. Additionally, also the Asian summer monsoon region north of India and along the Himalayas in the southeastern part of the domain present pronounced precipitation values. Figure 3.c shows the distribution of the WorldClim weather stations (Fick and Hijmans, 2017) over CA, representing a proxy for the density of the station data used in the CHIRPS dataset. Over East China, especially over the Tibetan Plateau, the observation data distribution ~~could be sparser~~is sparse. The data-model comparison is ~~to be~~ considered unreliable over this region (Randall et al., 2007; Cui et al., 2021; Yan et al., 2020; Russo et al., 2019).

3.1 Added value of CCLM driven by ERAInterim

To characterize the overall performance of the CCLM model in time and space, ~~we~~ Figures 4 and 5 show the maps of ~~yearly~~annual, winter (DJF), and summer (JJA) MAE and mean biases of precipitation between interpolated ERAInterim and CCLM driven by ERAInterim, calculated over the period 1985-2014 with respect to CHIRPS ~~. We calculate the MAE of daily precipitation for 1985-2014 from ERAInterim and CCLM driven by ERAInterim~~(Eq. 5 and Eq. 6). Figures 4.a-c show the MAE of ERAInterim with respect to CHIRPS for annual, winter and summer averages. The ~~differences in MAEs between ERAInterim and CCLM ($MAE_{ERAInterim,CHIRPS} - MAE_{CCLM,CHIRPS}$) or the added values~~ added value of the CCLM RCM compared to the interpolated ERAInterim GCM are shown in Figures 4.d-f. CCLM~~bias is higher~~ 's MAE is high during the Asian summer monsoon, over the South and Southeast of the domain (regions in magenta). During winter, the ~~bias~~ MAE is generally lower. CCLM presents a ~~bias reduction for prominent locations within the domain~~ MAE reduction for mountainous areas of Afghanistan, Kyrgizstan and Tajikistan and an increase of ~~bias~~ MAE near the boundaries: South of the domain throughout the year, South and Southeast during the summer.

Added values of GERICS-REMO2015 and RMIB-UGent-ALARO-0 driven by ERAInterim are shown in Figure 4.g-l respectively. The CHIRPS dataset is again used as the observational dataset O~~to calculate MAE and AV according to equations 5 and 6. The AV~~ . The added value of RCM is the most pronounced over areas with complex topography and especially during summer, for all three ~~considered RCMS~~ RCMS considered (Figs.4.d-l). Areas where the ~~downscaling reduces the bias of~~ RCM has smaller MAE than the reanalysis with respect to observations are located over Tajikistan, Kyrgyzstan, North of Afghanistan and part of the Himalayas. Mountain areas of Tajikistan and Kyrgyzstan are the main source of water for the former Soviet Union countries. However, precipitation during the colder seasons might be of more importance for water availability. The annual AV patterns still show positive values over those areas (Figure 4.d,g and j). Considering the whole domain, all three RCMs sensibly reduce the large and local-scale bias of ERAInterim against CHIRPS~~(Figure 4)~~, especially for complex topographies. The nested RCMs show similar values of MAE near their lateral boundaries, with respect to their driving model (Figure 4, ~~panels~~ *a*,*b*,*c*). Therefore, negative AV quantities might originate from the boundary effect, especially near the east and south-eastern boundaries, where the monsoonal precipitation is dominant. GERICS-REMO2015 shows pronounced negative added values for annual and winter above Tibet.

3.1.1 ~~Added value of CCLM driven by MPI-ESM1-2-HR~~

~~We showed that COSMO-CLM can reduce the bias of its driving reanalysis for daily precipitation, especially over areas with a complex topography like Tajikistan and Kyrgyzstan. In particular, our model simulation shows similar skills as in the previously published CORDEX-CA simulations. Here, we calculate the added value of the CCLM simulations driven by MPI-ESM1-2-HR for 1985-2014. It can be seen in figure **??**.a that the MPI-ESM1-2-HR shows less bias than the ERAInterim over Tajikistan and Kyrgyzstan. According to Déqué et al. (2007),~~ As an additional check, we also show the bias in the ~~GCM bias is one of the most important sources of uncertainty in the RCM's regional climate projection, and the smaller $MAE_{MPI\text{-}ESM1\text{-}2\text{-}HR}$ compared to $MAE_{ERAInterim}$ over Tajikistan and Kyrgyzstan might increase the skill of the final regional projections (under the assumption that the model bias remains conserved under other radiative forcings). The added value of CCLM driven by MPI-ESM1-2-HR shows smaller values over those areas compared to the simulation driven by ERAInterim, especially for summer season (Figure **??**.f). Our analysis of the two driving datasets (ERAInterim and MPI-ESM1-2-HR) tends to confirm the findings of the Sørland et al. (2018), at least for daily precipitation, that the biases of the GCM-RCM chain are not additive~~

~~and not independent. For example, in all regions with high values of yearly precipitation, where GCM has a slight bias, the RCM does not present higher biases or vice versa. The large-scale patterns in the parent GCM are usually a limiting factor for the dynamical downscaling following the "garbage in, garbage out" problem (Rummukainen, 2010). AV in an RCM is achieved by the improved representation of surface processes, which usually are present over areas with complex topography (Torma et al., 2015) or over coastal areas with strong land–sea differences (Feser et al., 2011)~~climatologies of models in figures 5. Once again the biases are pronounced on the right bottom corner of the domain during the JJA and south Tibetan Plateau throughout the year.

3.1.1 Extreme precipitation patterns in CCLM and CMIP6 GCMs

~~We explore climate change signals in the high-resolution output, given~~ Given that the CCLM simulation has shown some added value for precipitation over mountainous areas of CA, we explore climate change signals in its high-resolution output. The resulting high-resolution maps might have biases inherited from the GCM-RCM selection. We assume that many model biases remain conserved among the different time slices and, therefore, ~~could~~ can be removed when calculating the changes between the historical (1985-2014) and future periods (2070-2099).

We present the resulting climate change trends in CCLM and the CMIP6 GCMs ensemble statistics (ensemble mean and standard deviation). We analyzed 31, 33 and 38 models for SSP126, SSP370 and SSP585 scenarios with a total number of simulations of 158, 185 and 242, respectively (see Supplementary materials for the list of models used in this study). To give the same weight to individual models, we first calculate the statistics over all the members of each model and then build the final statistics. We have chosen the yearly 99^{th} percentile of daily precipitation (PR99 hereafter), which considers the three days of the year with the highest precipitation. We also chose the number of very heavy precipitation days in the period (ECA_RX20mm) as a different index, one of several precipitation-related indices used to monitor and analyze climate variability and change. For example, this index is often used in climate research to assess the impacts of very heavy precipitation events on water resources, agriculture, and natural ecosystems (Klok and Klein Tank, 2008). Figure 6 presents the changes in averaged PR99 at the end of the century (2070-2099) with respect to the historical period (1985-2014) for CCLM (a,d and g) and CMIP6 GCMs (b,e and h) under different scenarios. The downscaling signals indicate that those characteristics depend on the scenario and time period. The large-scale patterns remain the same among all three selected scenarios with intensification when the anthropogenic influence increases. The standard deviation of the models' ensemble is shown in Figures 6.c,f and i. According to our analysis, the Himalayas, especially Nepal, North India, and Bhutan, have the highest uncertainty among the GCMs and in all scenarios. Except for this area and the eastern boundary of the domain, the standard deviation remains under 3 mm/day. Under the pessimistic SSP585 and the regional rivalry SSP370 scenarios, areas with more than 9 mm/day increase in PR99 for CCLM over Northwest India, North Pakistan, North Iran, Southwest of Iran exist and South and Southeast of Black Sea. A reduction pattern is detected East of the Mediterranean Sea in Jordan, Syria, and South of Turkey. Similar patterns are also observed in the CMIP6 ensemble mean. However, due to the averaging, the GCMs' ensemble mean patterns are around $\pm$ 5 mm/day over those areas. Under the SSP126 scenario, which agrees with the 2°C target, the increasing patterns of more than $\pm$ 9 mm/day for CCLM and $\pm$ 5 mm/day for GCMs disappeared. In CA, areas of increased PR99 over Kyrgyzstan, Tajikistan, North of Pakistan and Southwest Iran are regions with a considerable risk of rainfall-triggered events like landslides (Wang et al., 2021; Kirschbaum et al., 2010) and floods (for example, Pakistan floods of 2010 and 2022).

Figures 7.1,d and g show the ECA_RX20mm values for CCLM for the three scenarios at the end of the century. The patterns are like those shown in Figure 6, indicating that the number (frequency) of very heavy precipitation days also increases with an enhanced anthropogenic influence, particularly over the Tibetan Plateau. From Figures 7.b,e and h, we conclude that the CMIP6 GCM ensemble also presents a very similar behavior to CCLM. The ensemble standard deviations, however, increase over Tajikistan and Kyrgyzstan for ECA_RX20mm values (Figures 7.c,f and i). The increased frequency and intensity of extreme precipitation over elevated areas of CA due to anthropogenic forcing is alerting (Fallah et al., 2023). The presented CCLM simulation contributes to study the sensitivity of dynamical downscaling to different levels of anthropogenic forcing at the local scale. This information might be of interest for the scientific community working on the impact of climate change in CA.

4 CCLM emulator using a ~~Convolutional Neural Network~~CNN

We have shown that the dynamical downscaling added value to explore the local effects of climate change during the historical period, especially over areas with enhanced topographical forcings. Here, we create an emulator of ~~COSMO-CLM~~ CCLM for precipitation over CA. ~~We demonstrate that the unconstrained CNN model could reconstruct high-resolution features from a coarse GCM, which are like the target COSMO-CLM simulations.~~ As explained previously, a CNN could be trained on our GCM-RCM chain and be applied as a fast and computationally cheap downscaling method. However, the skill of such a model must be explored and verified.

~~One major source of error in training a CNN is usually the problem of over-fitting. However, in our case, we have an overly complex climate system (i.e. COSMO-CLM) with highly complex precipitation fields as input, and a low-complex CNN on the model side. Therefore, our problem is of an under-fitting nature.~~ Here we want to demonstrate that the emulator ~~has significantly more skill~~ is better at downscaling than a simple interpolation, especially for areas receiving extreme precipitation values. More specifically, our goal is to show that the ~~COSMO-CLM~~ CCLM emulator can produce ~~COSMO-CLM-like~~ CCLM-like patterns when fed by the parent GCM.

For the CNN approach, we focus on the CA domain ~~introduced as a domain covering~~ covering only the former Soviet Union countries (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan) and not the CORDEX-CA domain previously shown in Figure 3. This domain is the region of interest in the Green Central Asia project ~~?~~https://www.greencentralasia.org/en, which is financed by the German Foreign office. Figure 8.a shows the MAE ~~from~~ of the interpolated MPI-ESM1-2-HR ~~with respect to the COSMO-CLM~~ using the CCLM driven by it from the test dataset ~~, i.e., MAE(MPI-ESM1-2-HR, MPI-ESM1-2-HR-CCLM)~~ as the "true" precipitation. As can be seen, ~~COSMO-CLM~~ CCLM produces different precipitation values compared to its driving GCM, especially over regions with complex topography. ~~This has been noticed in the added value and downscaling signal maps of COSMO-CLM. To explore a potential skill in~~ Here, we assume that the CCLM is the ground truth and check if the CNN can produce it using the GCM as input data. To evaluate the performance of the emulator, we show the maps of ~~MAE~~

~~reduction, i.e., $\mathrm{MAE}_{\mathrm{GCM,CCLM}} - \mathrm{MAE}_{\mathrm{CNN,CCLM}}$ in figures~~ added value in Figures 8.b-d. Comparison of MAE reduction maps shows that the unconstrained CNN produces significant skills over elevated regions of CA and the constrained runs do not present considrable patterns of changes. For example, there are areas of negative and positive added values remarkably close together over elevated areas of CA created by HCL and SCL emulators. NoCL, in contrast, shows systematic positive values over large parts of the domain. ~~The fingerprint of the GCM is detectable~~ There are several artifacts in the MAE reduction maps of constrained models, especially over North of India, which represent the GCM grid shape. We produce the boxplots of daily precipitation over the ~~newly-considered domain~~ CA domain covering the former soviet union to explore the improvement in the distributions (Figure 9). The correlation coefficients between the time-series of average precipitation over the domain with respect to CCLM are also presented in Figure 9 (values in the parentheses). For the daily averages, NoCL presents the best performance (highest correlation coefficient). However, the values of outliers are ~~less~~ smaller than the ones from CCLM and all other model simulations. The distribution is more condensed around the median (smallest interquartile range). The distribution of ~~all constrained models~~ both constrained models (HCL, SCL) is like the interpolated GCM one. This was expected, since the constraining conserves the mass of high-resolution grid-boxes within the corresponding low-resolution grid-box (Equation 1).

4.1 Applying the CNN to a different GCM

~~Here, we~~ We evaluate the emulator's generalization ability, i.e. the ability to create reliable predictions ~~of~~ on a new data set. We conduct ~~here~~ a new 15-year dynamical simulation with ~~COSMO-CLM~~ CCLM driven by the EC-Earth3-Veg (Döscher et al., 2022) GCM under ssp370 from 2019 to 2033. We use this data as input to our ~~COSMO-CLM~~ CCLM emulator, which was previously trained ~~on the MPI-ESM1-2-HR and its COSMO-CLM run~~to emulate CCLM using MPI-ESMI-2 HR as input GCM. We now use the emulator to reconstruct the local features of ~~COSMO-CLM~~ CCLM driven by EC-Earth3-Veg. Figure 10.a presents the MAE of the interpolated EC-Earth3-Veg with respect to the ~~dynamically downscaled simulation using COSMO-CLM~~dynamical downscaling with CCLM, i.e., the ~~COSMO-CLM~~ CCLM simulation driven by EC-Earth3-Veg. The MAE pattern of EC-Earth3-Veg is remarkably like the one from MPI-ESM1-2-HR (Figure 8.a). However, the ~~COSMO-CLM~~ CCLM emulator based on the NoCL CNN model does not show positive error reduction everywhere in the domain (Figure 10.b). We chose the NoCL CNN because it showed the best performance among the constrained ones. Training the CNN on the MPI-ESM1-2-HR/CCLM might have ignored learning processes which overcome considerable biases in the driving GCM. The ~~COSMO-CLM~~ CCLM emulator tries to find relations between the MPI-ESM1-2-HR and ~~COSMO-CLM~~CCLM, which might be specific to these two models and there is no guarantee that those relationships also apply to the new EC-Earth3-Veg and ~~COSMO-CLM~~ CCLM driven by EC-Earth3-Veg. This new GCM-RCM chain ~~contains new sets of models and~~ is extremely sensitive to the characteristics of the EC-Earth3-Veg model because, as we showed previously, the RCM state follows the state of its driving GCM. We note that CCLM is driven at the lateral boundaries by the GCM values for the state variables of CCLM (temperature, pressure, wind speed etc.). Precipitation is not used for driving the RCM. The CNN input is the GCM precipitation, which has different biases in the two GCM, and therefore the mapping from the MPI-ESM1-2-HR-precipitation to the CCLM precipitation cannot be successfully transferred to EC-Earth3-Veg.

Knowing these limitations, the CNN model shows added values of more than 1 mm/day over the Alborz Mountains and South of the Caspian Sea in the North of Iran (black rectangular in Figures 10.a and b) and some parts of Tajikistan and Kyrgyzstan. Exploring the field mean of daily precipitation distribution indicates that the CNN's median value and the outliers are lower than both the EC-Earth3-Veg and ~~COSMO-CLM~~ CCLM simulations (Figure 10.c). Only the day-to-day correlation is being improved. ~~The model was~~ As mentioned before, all model were trained on the shuffled dataset and ignored the memory in the time series but here ~~we fed the original (without shuffling) dataset and calculated the correlations~~the trained NoCL model was given unshuffled EC-EARTH3-Veg to make new predictions. The correlation coefficient increases using the NoCL model from 0.815 (EC-Earth3-Veg) to 0.844 (NoCL). Over the black rectangular box in Figure 10.b, the region where the NonCL model reduces the MAE, i.e., the ~~black rectangular box in Figure 10.b, the~~ distribution of precipitation converges to the one from ~~COSMO-CLM~~ CCLM (Figure 10.d) ~~. Only the outliers larger than 20 mm/day are not reconstructed by the NoCL. This region~~ and receives the highest amount of precipitation in Iran and supplies water for a large portion of population in the country, including the capital city Tehran with a population of over 10 million people. Only the outliers larger than 20 mm/day are not reconstructed by the NoCL.

As a new test for generalization, we intentionally did not include a scenario (SSP370) in the training process. This move allowed us to apply the model to a specific simulation and witness its ability to reproduce an unknown forcing. Figure 11 demonstrates the AV of the CNN emulator for SSP370 in comparison to the dynamical downscaling with CCLM, i.e., the CCLM simulation driven by SSP370. The AV pattern is strikingly similar to the one shown in Figure 8.d. We conclude that the CNN can learn patterns it was not trained for, as evidenced by the SSP370 scenario.

5 Discussion and conclusions

Regional climate change impact assessments require high resolution climate projections. The main strategies to produce such datasets are statistical and dynamical downscaling, as well as a hybrid of the two methods. Statistical downscaling (SD) usually has limited capability to consider the dynamic influences of the complex topography. The large-scale domain does not reflect the spatial diversity and variation of the local climate and the topography, which may affect the accuracy of the statistical relationships (Li et al., 2022). For SD applied to precipitation, the observations need to contain detailed information about the precipitation distribution in areas with complex topography (Lundquist et al., 2019). On the other hand, dynamical downscaling requires ~~a massive amount of~~ massive computational time and data storage space. A 30-year CCLM simulation driven by ERAInterim took roughly one week to finish using 216 processors of the HLRE-4 Levante computer at the German Climate Computing Center (DKRZ). Additionally, the added value of RCMs is still debated ~~, since~~ since they are highly dependent on the driving GCMs.

In this study, we contribute to the few dynamical downscaling efforts over the CORDEX-CA domain, a small step towards an RCM ensemble creation for CA. A single RCM simulation can help identify model biases and uncertainties that need to be addressed in future model improvements. It is essential to note that relying solely on a single model run for CMIP6 instead, of an RCM ensemble, may not provide any comprehensive understanding of the potential climate change impacts on a region.

Therefore, it is recommended that researchers conduct multiple simulations with different initial and boundary conditions and different model configurations to account for the uncertainty associated with climate projections.

In a first part of the study we demonstrate the added value of RCMs (here we chose to use the COSMO-CLM/CCLM model) over GCMs for CA in the representation of precipitation. Our ~~COSMO-CLM run shows AV~~ CCLM run shows added value with respect to its driving GCM, comparable to the range of values obtained for other RCMs applied to the CORDEX-CA domain over the evaluation period. It also reproduces extreme precipitation changing patterns like the CMIP6 ensemble mean at the end of the century. Both ~~COSMO-CLM~~ CCLM and CMIP6 ensemble present elevated risk (frequency and intensity) of heavy precipitation events over vulnerable areas of CA due to different anthropogenic influences.

Our study evaluated the downscaling skill primarily using higher resolution observations, which are critical for capturing localized climate phenomena relevant to regional adaptation strategies. However, as Volosciuk et al. (2017) noted, examining downscaling outputs at coarser resolutions can be equally informative. Their work emphasizes that downscaling methods can introduce or fail to correct biases that differ significantly across spatial scales. By evaluating on a coarser grid, it is possible to distinguish between the inherent biases of the model and those introduced by the downscaling process. This distinction is crucial for understanding the limitations and strengths of downscaling methods in representing climatic variables across different scales.

Additionally, acknowledging the computational and memory constraints of ~~an RCM to be run at very~~ running an RCM at high resolution, here we also show that a single GCM-RCM model chain can be used to train a climate emulator based on a CNN model. It can learn some nonlinear and physical relationships between the coarse and fine-resolution datasets~~, based on atmospheric governing equations~~. This can overcome the problem of spatial intermittency seen in some statistical downscaling approaches ~~(Harder et al., 2022)~~(Harder et al., 2023). However, we have also shown that the CNN model has limitations, namely when generalizing, as it did not achieve a robust error-reduction pattern ~~when applied to a different GCM-CCLM chain~~given a different GCM as input. The learning process depends strongly on the GCM/CCLM relationships. More importantly, an RCM is forced to follow its driving GCM and only on local scales can produce extra information. ~~Therefore, we recommend running a GCM-RCM simulation for several years and evaluating the model performance before applying it to a new specific GCM.~~ An application of the presented CCN ~~is~~ could be to apply it for other experiments of the same GCM: One ~~can~~ could use the trained emulator for paleo-climate experiment of the parent GCM to create more than 10,000 years of downscaled simulation. One can also downscale the volcanic forcing experiments using the trained emulator. This will aid the paleo-climate community in conducting proxy-model comparisons at local scales. However, previous studies have shown that the CNN suffers from the same generalisation problem as when applied to a new GCM and such applications must be tested (Jouvet and Cordonnier, 2023).

In an effort to evaluate the model's generalization capabilities, we deliberately excluded the SSP370 scenario from the training dataset. This strategy was implemented to assess whether the model could effectively infer and replicate patterns from untrained scenarios. Remarkably, the model's output for the SSP370 scenario exhibits an AV pattern that closely mirrors the dynamical downscaling results obtained with the CCLM, driven by the same SSP370 scenario. This alignment strongly supports the notion that our CNN emulator is not only capable of learning from its training data but also proficient in generalizing to new, unseen conditions. The similarity in AV patterns between the model output and the CCLM simulation underscores the robustness and adaptability of our model, affirming its potential for broader applicative scenarios in climate modelling.

We note that this work is only a step to demonstrate the potential of such a hybrid approach, and we encourage the community to explore different model structures and parameter combinations for further improvement. For example, our few model set-ups showed that ~~the constrained model~~ using a physically constrained CNN set-up, that applies a linear transformation to the high-resolution image to ensure that the total mass or energy is conserved between the low and high-resolution images, did not successfully downscale the precipitation. The constraints might not be satisfied in the original dataset and therefore the constrained model set-up did not lead to better results. In contrast, with a higher degree of freedom, the unconstrained ~~model run produced more realistic patterns . Alternative~~ CNN produced patterns closer to the target RCM. Alternative machine learning models, such as generative adversarial networks (~~GAN~~GANs), which can generate more high-frequency patterns, might improve the downscaled pattern, and ~~should~~ could be tested in future studies. An additional set-up might be to ~~add~~ provide more information to the CNN by adding characteristics like surface height, vegetation, land-cover, land-use, etc. as new channels within the input layer.

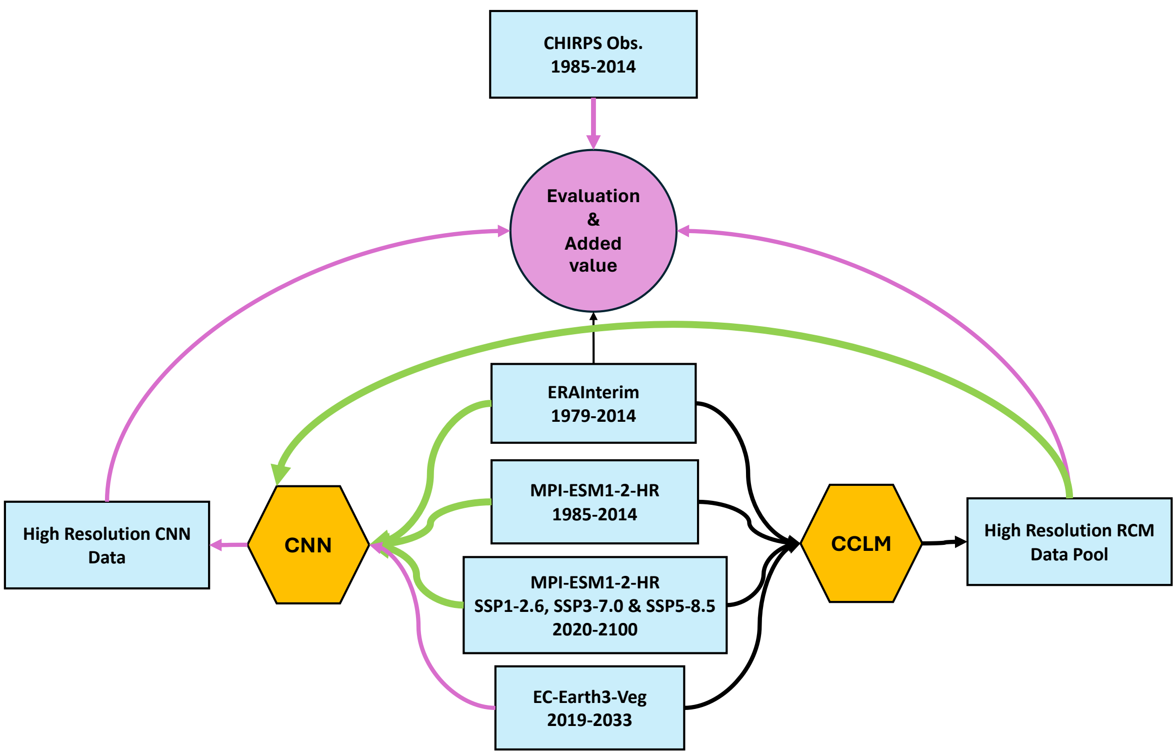


Figure 1. Schematic of the methodology used in this study. Green arrows show the data flow used for training the CNN and magenta for evaluation and calculation of the added values. Datasets are shown by rectangular, downscaling models by hexagonal and evaluation analysis by circle.

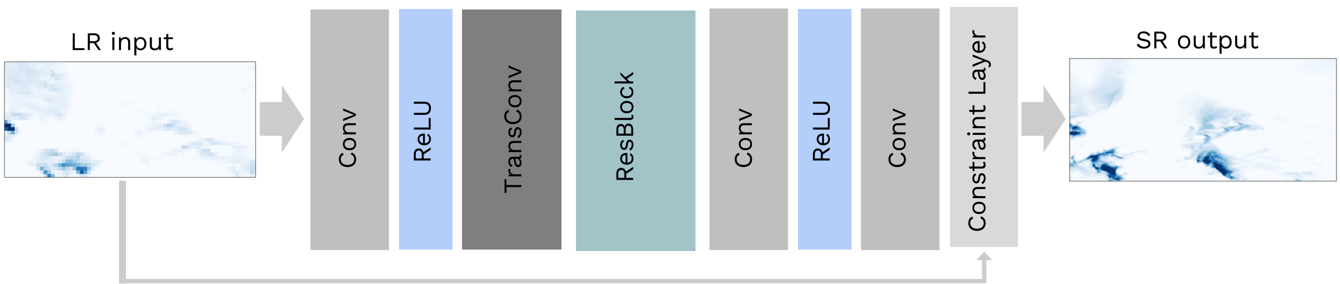


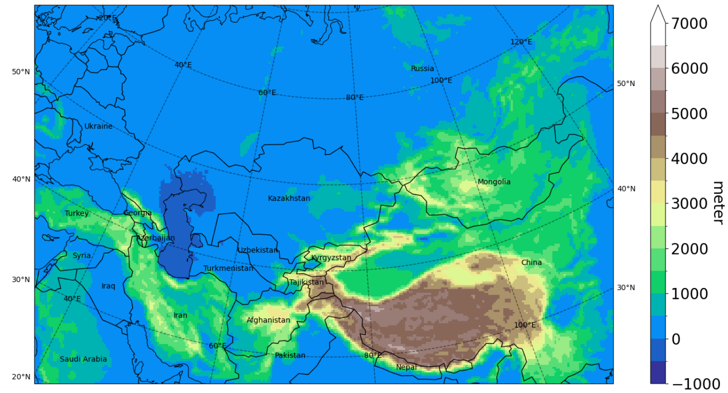
Figure 2. Schematic of the CNN architecture for 2 times upsampling with the constraints layer. The inputs are low-resolution (LR) images of size 30×60 and the output is a super-resolution (SR) image of size 60×120. This figure is modified from (Harder et al., 2023).

Code availability. The code for "Physics-Constrained Deep Learning for Climate Downscaling," is available on Zenodo at the following DOI: https://zenodo.org/uploads/8150694. The input, output, trained models, a snapshot of the code employed in the deep-learning downscaling process, COSMO-CLM model setups for all Regional Climate Model (RCM) simulations conducted, a list of CMIP6 model information used for comparative analysis, and a Jupyter notebook for executing a test case of the "Physics-Constrained Deep Learning for Climate Downscaling" as described in the manuscript are available at Zenodo with the following DOI: https://zenodo.org/records/10417111.

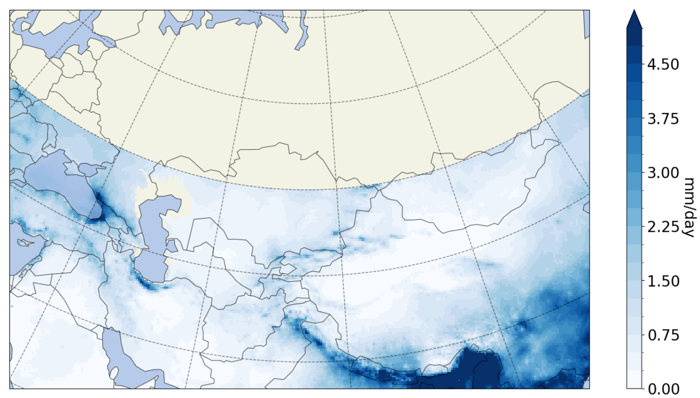
Appendix A: ~~Constraint layers~~

~~We test the CNN with three different constraining methods in the last CNN layer: 1- soft constraining (SCL), 2- hard constraining (HCL) and 3- without constraining (NoCL). For a detailed information on the settings used we refer to the work of Harder et al. (2022). In the following we explain briefly the three different constraining methodologies. The set-up of constraining is as following: Consider a factor N for downscaling in all linear directions and let~~ $n := N^2$ ~~and~~ $y_i, i = 1, ..., n$ ~~be the high-resolution patch values that correspond to low-resolution pixel~~ x~~. The mass conservation law has the form of the following constraint:~~

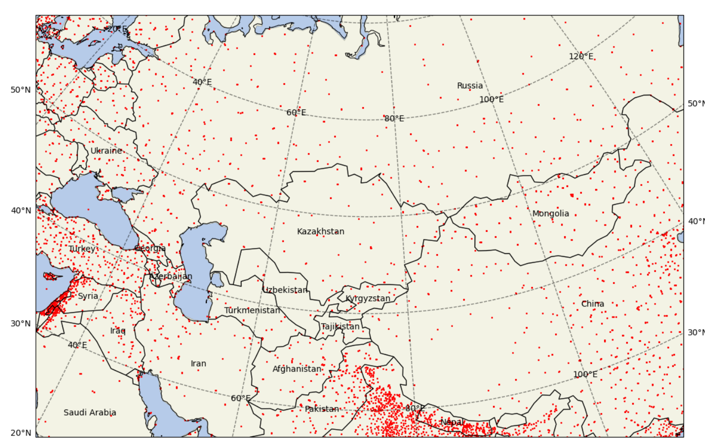
$$\frac{1}{n}\sum_{i=1}^{n} y_i = x.$$



(a)

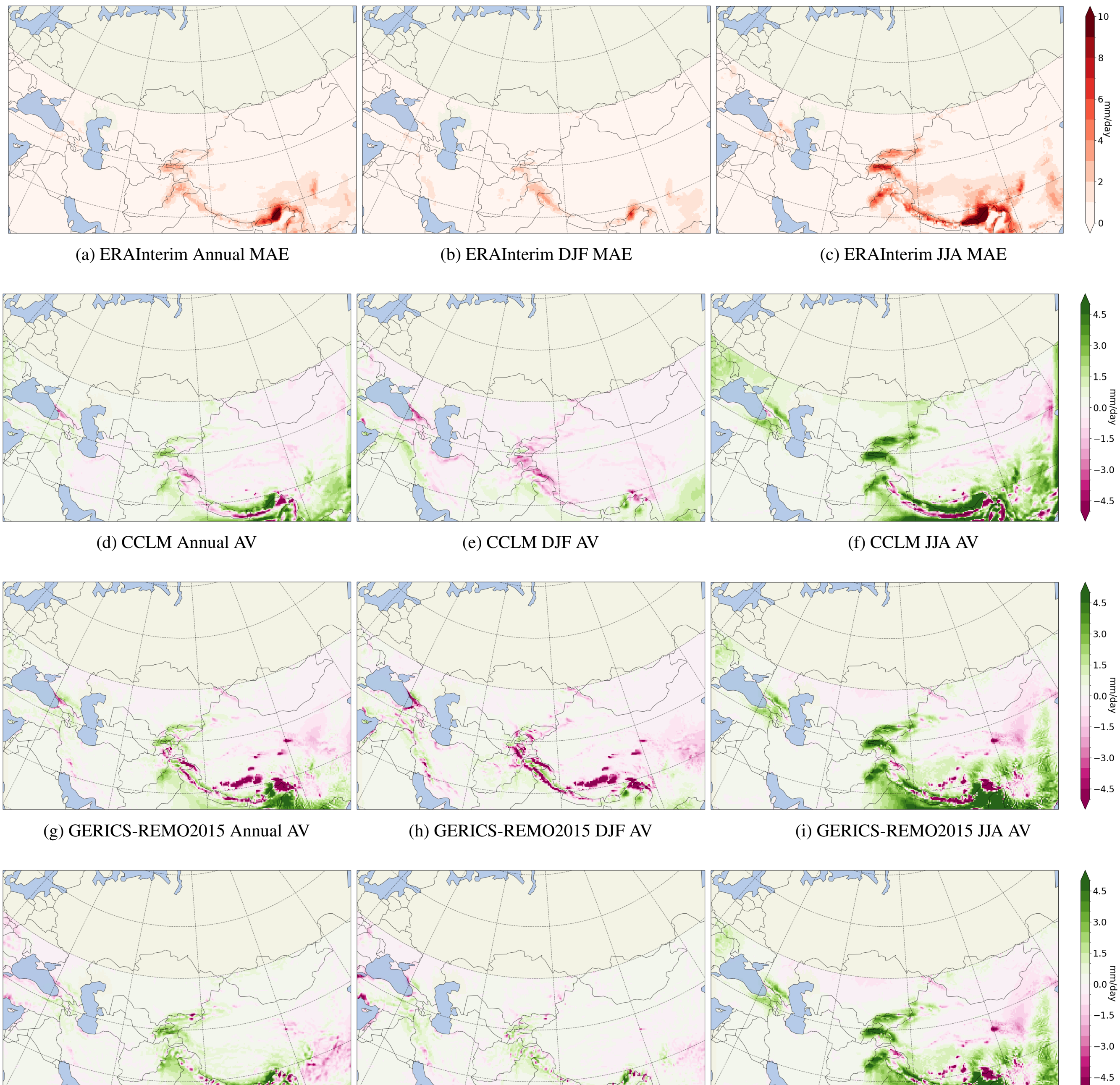


(b)



(c)

Figure 3. a) ~~Study region~~ CCLM simulation domain over Central Asia and the topography (m), (b) CHIRPS climatology for 1985-2014 (average of daily values over all years in mm/day), and (c) WorldClim's weather stations (red dots).



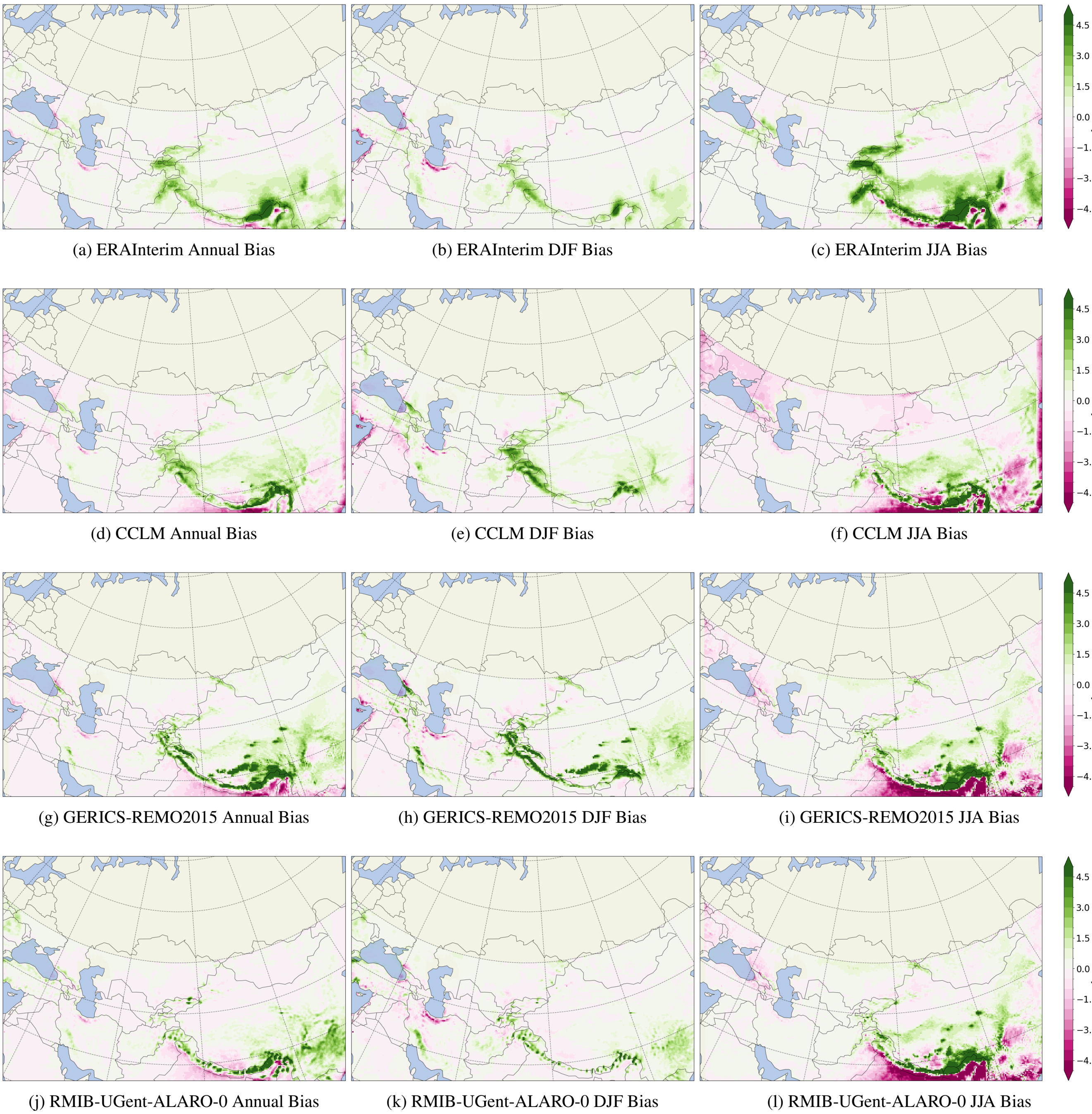
(a) ERAInterim Annual MAE (b) ERAInterim DJF MAE (c) ERAInterim JJA MAE

(d) CCLM Annual AV (e) CCLM DJF AV (f) CCLM JJA AV

(g) GERICS-REMO2015 Annual AV (h) GERICS-REMO2015 DJF AV (i) GERICS-REMO2015 JJA AV

(j) RMIB-UGent-ALARO-0 Annual AV (k) RMIB-UGent-ALARO-0 DJF AV (l) RMIB-UGent-ALARO-0 JJA AV

Figure 4. Mean average error (MAE) of daily precipitation (mm/day) from ERAInterim, as well as, added value (AV) as measured by MAE differences between ERAInterim and RCMs ($MAE_{ERAInterim} - MAE_{RCM}$) in mm/day for annual (a,d,j,i), ~~winter~~ December, January, February (b,e,h,k) and ~~summer~~ June, July, August (c,f,i,l). CHIRPS is used as observation. All daatasets are interpolated to the CCLM grid.



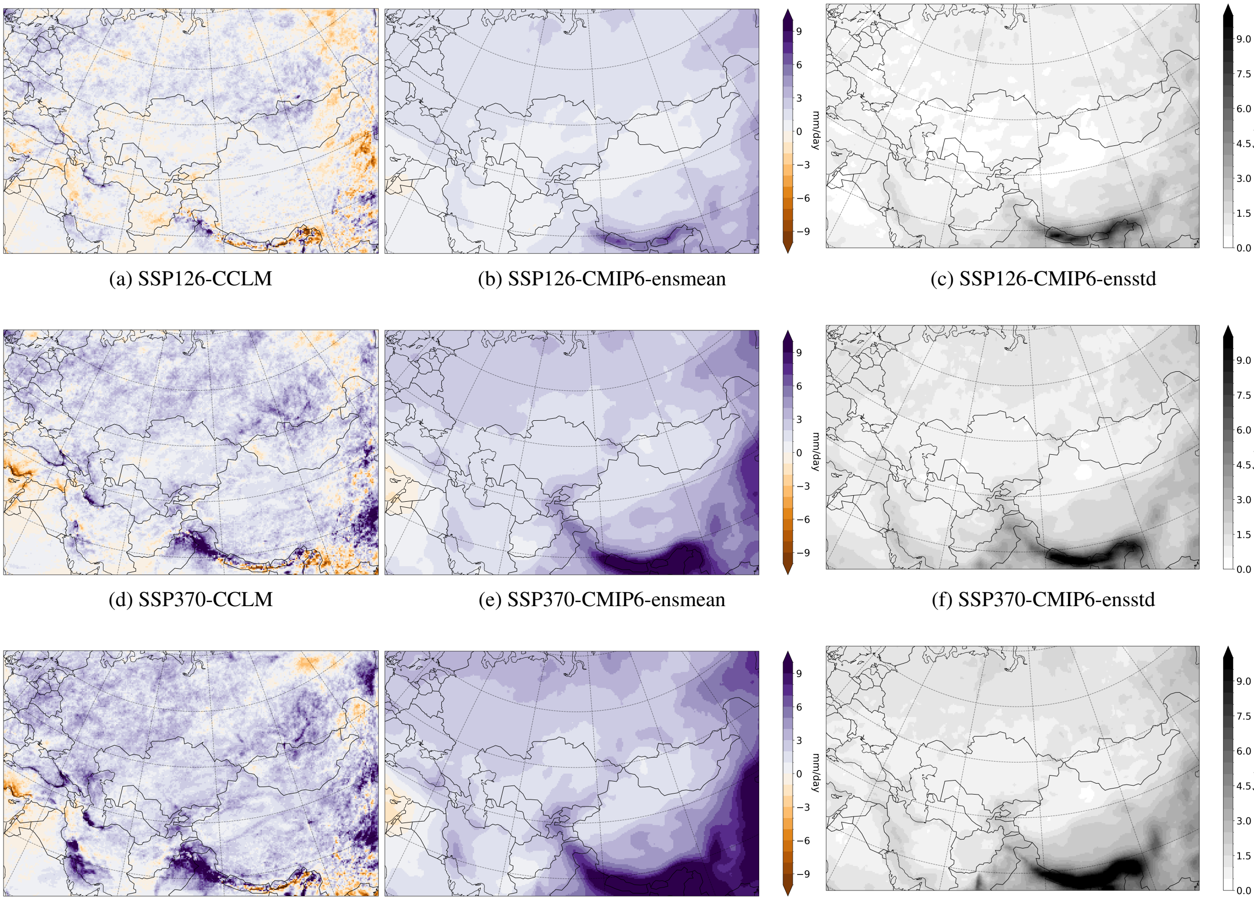
(a) ERAInterim Annual Bias (b) ERAInterim DJF Bias (c) ERAInterim JJA Bias

(d) CCLM Annual Bias (e) CCLM DJF Bias (f) CCLM JJA Bias

(g) GERICS-REMO2015 Annual Bias (h) GERICS-REMO2015 DJF Bias (i) GERICS-REMO2015 JJA Bias

(j) RMIB-UGent-ALARO-0 Annual Bias (k) RMIB-UGent-ALARO-0 DJF Bias (l) RMIB-UGent-ALARO-0 JJA Bias

Figure 5. ~~MAE~~ Bias of ~~daily~~ climatological precipitation (mm/day) from ~~MPI-ESM1-2-HR~~ERAInterim, as well as, ~~added value (AV) as measured by MAE differences between MPI-ESM1-2-HR and~~ ERAInterim-driven RCMs (~~$MAE_{MPI-ESM1-2-HR} - MAE_{RCM}$~~$PR_{ERAInterim-CCLM} - PR_{OBS}$) in mm/day for annual (a~~and~~,d,j,i), ~~winter~~ December, January, February (b~~and~~,e,h,k) and ~~summer~~ June, July, August (c~~and~~,f,i,l). CHIRPS is used as observation.



(a) SSP126-CCLM (b) SSP126-CMIP6-ensmean (c) SSP126-CMIP6-ensstd

(d) SSP370-CCLM (e) SSP370-CMIP6-ensmean (f) SSP370-CMIP6-ensstd

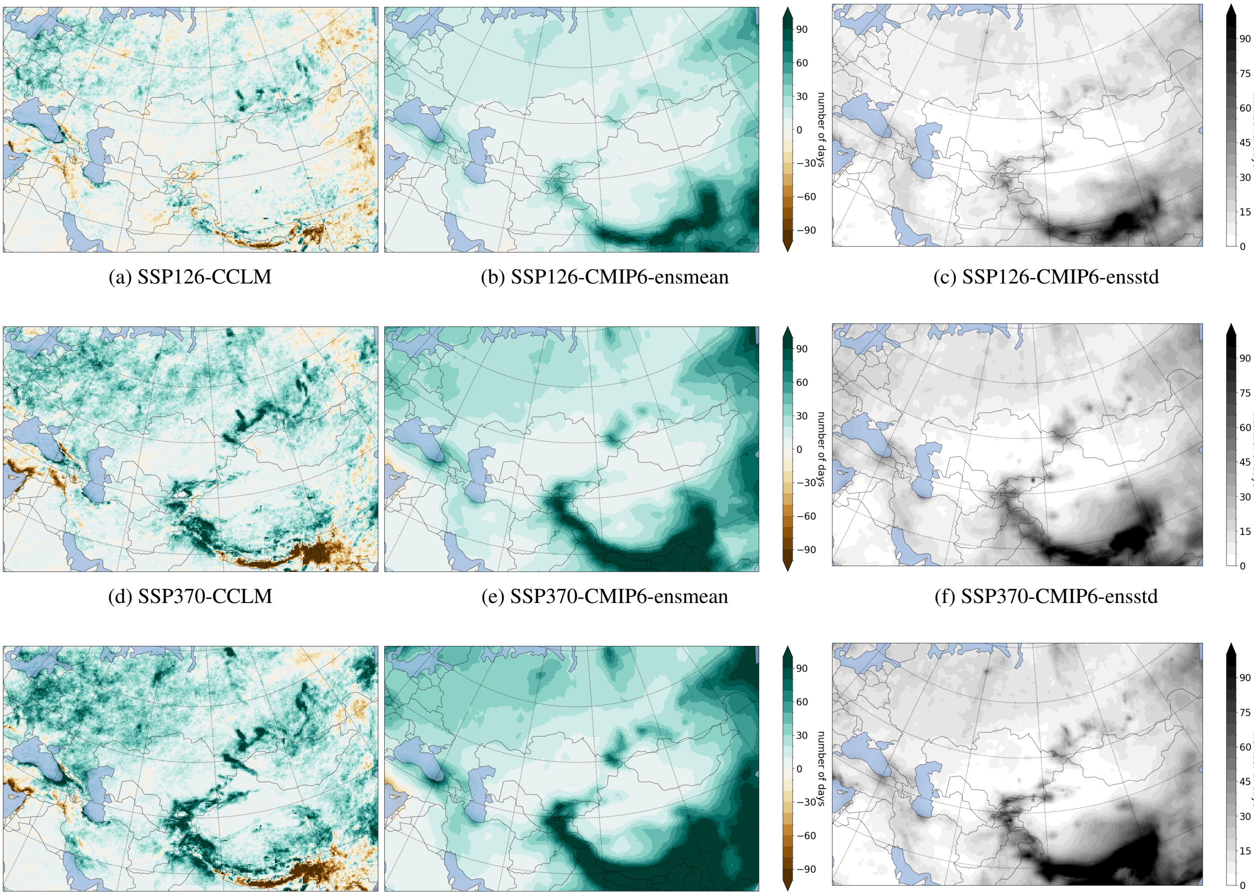
(g) SSP585-CCLM (h) SSP585-CMIP6-ensmean (i) SSP585-CMIP6-ensstd

Figure 6. Changes in averaged yearly 99^{th} percentile (3 days per year) of total precipitation (mm/day) with respect to 1985-2014 references for a,b) SSP126, d,e) SSP370 and g,h) SSP585 at the end of the century (2070-2099) from CCLM and CMIP6 GCMs' ensemble mean. The ensemble's standard deviations are shown in c,f and i.

~~**Hard constraining:** It uses the SoftMax constraining, which is a proper constraining for quantities like water content. It enforces the output to be non-negative. For constraining the predicted quantities, we use a SoftMax operator on the intermediate outputs of the neural networks before the constraining layer ($\tilde{y}_i$) and multiply it with the corresponding input pixel value x:~~

$$y_i = \exp(\tilde{y}_j) \cdot \frac{x}{\frac{1}{n}\sum_{i=1}^{n}\exp(\tilde{y}_i)}.$$

~~y_i is the final output after applying the constraints. We have used the mean absolute error (MAE) as the loss function.~~



(a) SSP126-CCLM (b) SSP126-CMIP6-ensmean (c) SSP126-CMIP6-ensstd

(d) SSP370-CCLM (e) SSP370-CMIP6-ensmean (f) SSP370-CMIP6-ensstd

(g) SSP585-CCLM (h) SSP585-CMIP6-ensmean (i) SSP585-CMIP6-ensstd

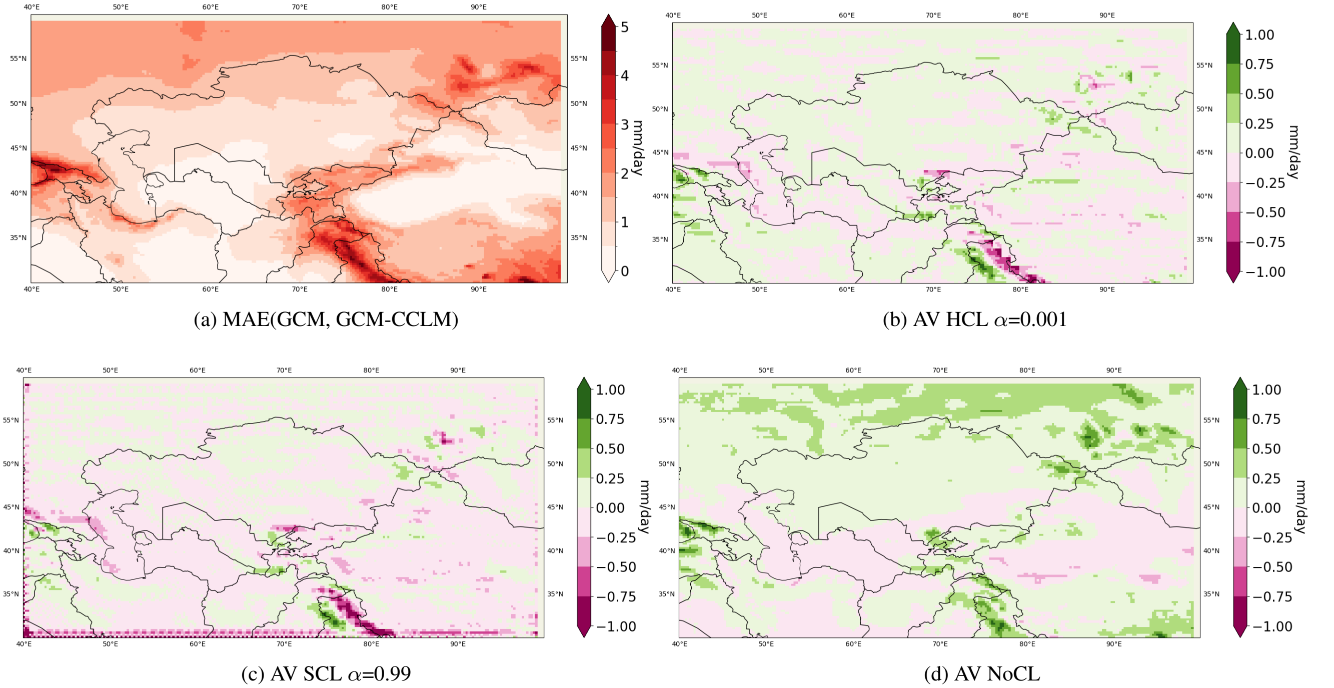
Figure 7. Changes in number of days with precipitation more than ~~20mm~~ 20 mm in the period with respect to 1985-2014 references for a,b) SSP126, d,e) SSP370 and g,h) SSP585 at the end of the century (2070-2099) from CCLM and CMIP6 GCMs' ensemble mean. The ensemble's standard deviations are shown in c,f and i.

~~**Soft constraining:** This is done by adding a regularization term to the loss function. MAE is then changed to the following:~~

$$\mathrm{Loss} = (1-\alpha)\cdot MAE + \alpha\cdot \mathrm{CV},$$

~~where CV is the constraint violation, which is the mean overall constraint violations between an input pixel x and the super-pixel (high-resolution grid-cell)~~ ~~y_i:~~

$$\mathrm{CV} = \mathrm{MSE}(\frac{1}{n}\sum_{i=1}^{n} y_i, x)$$



(a) MAE(GCM, GCM-CCLM) (b) AV HCL α=0.001

(c) AV SCL α=0.99 (d) AV NoCL

Figure 8. a) MAE (MPI-ESM1-2-HR,CCLM). MPI-ESM1-2-HR is remapped bilinearly to the 0.25×0.25 grid. b-d) Added Value (AV) or MAE(MPI-ESM1-2-HR,CCLM) - MAE(CNN,CCLM) for different constraining method.

~~We use the $\alpha = 0.99$ in this study.~~

~~**Without constraining**: In this setup we remove the constraining layer after the last convolutional layer in the CNN. We use the MAE as the loss function.~~

~~The constraint layers are applied at the end of the CNN architecture, and all satisfy the criteria that the resulting high-resolution patch shall conserve the values in low-resolution pixels. The performance of the different settings will be assessed through the MAE.~~

Appendix A: CNN runs

We used the following commands for training the CNN model based on the ~~Harder et al. (2022)~~Harder et al. (2023):

```
# for the run with soft constraining run, with a factor of alpha 0.99 :

$ python main.py --dataset dataset --model cnn --model_id
twc_cnn_soft_constraints_epochs_160_lr_0.00001_alpha_0.99
--constraints soft --loss mass_constraints --alpha 0.99
 --epochs 160 --batch_size 64 --lr 0.00001
```

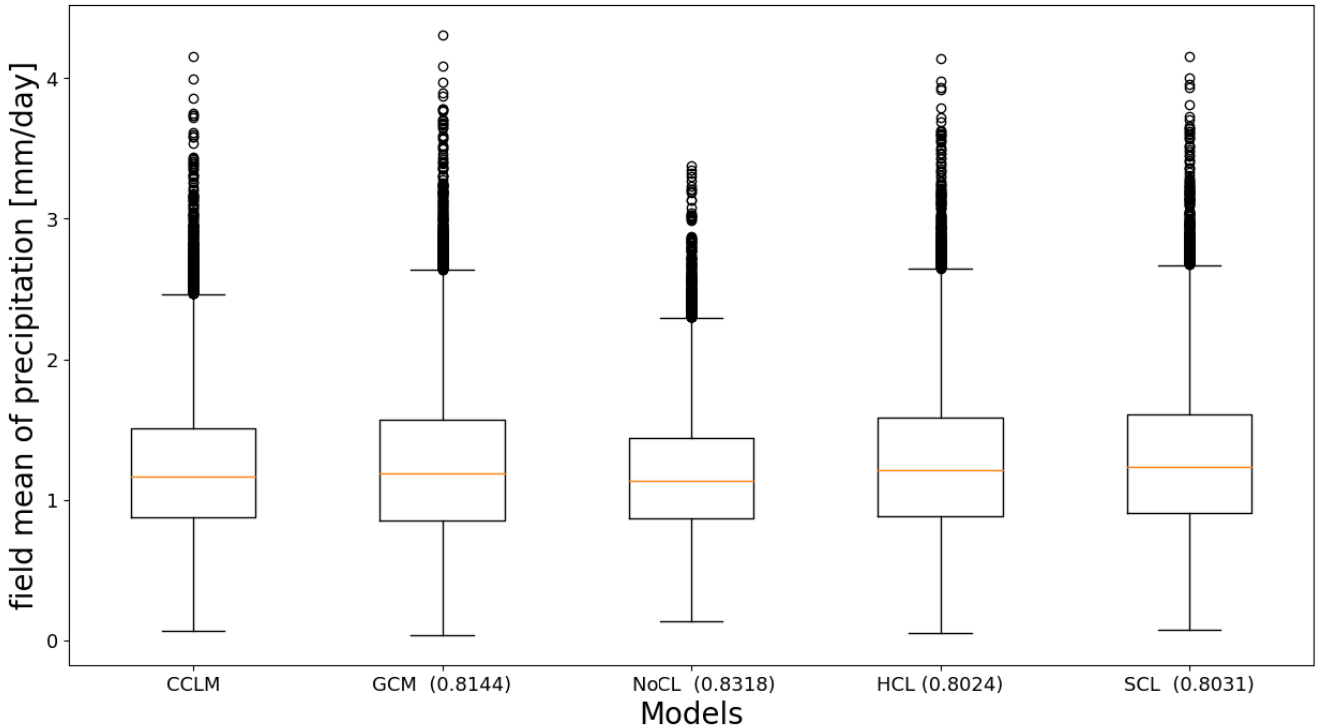
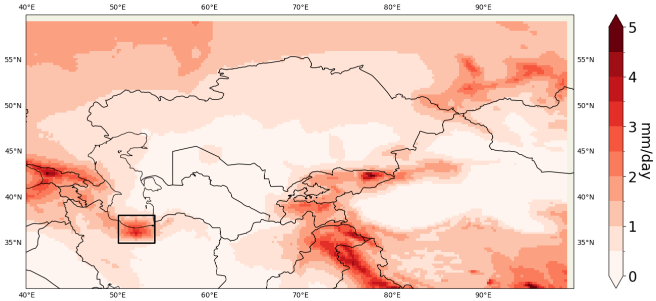


Figure 9. Boxplot of averaged daily precipitation over the Central Asian domain (shown in Figure 7) for different models and test dataset (22714 days or 62.2 years). Numbers in the parenthesis indicate the correlation coefficients between each model and the CCLM simulation.

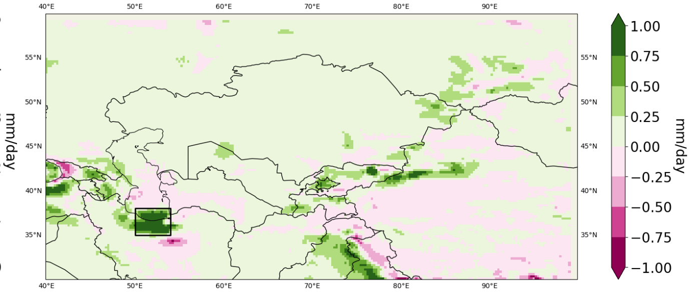
```
# for the run with softmax constraining or hard constraining:
$ python main.py --dataset dataset --model cnn --model_id
twc_cnn_softmaxconstraints_epochs_200_batch_size_64_lr_0.001
--constraints softmax --lr 0.001 --epochs 160 --batch_size 64 --loss mae

# for the standard CNN run without constraining:
$ python main.py --dataset dataset --model cnn --model_id
twc_cnn_noneconstraints_epochs_160_batch_size_64_lr_0.001
--constraints none  --lr 0.001 --epochs 160 --batch_size 64 --loss mae
```

Note that the datasets ~~available at shall be downloaded in a folder called dataset~~ and codes are available at Zenodo (DOI: https://zenodo.org/records/10417111) with comprehensive details utilized in the paper.

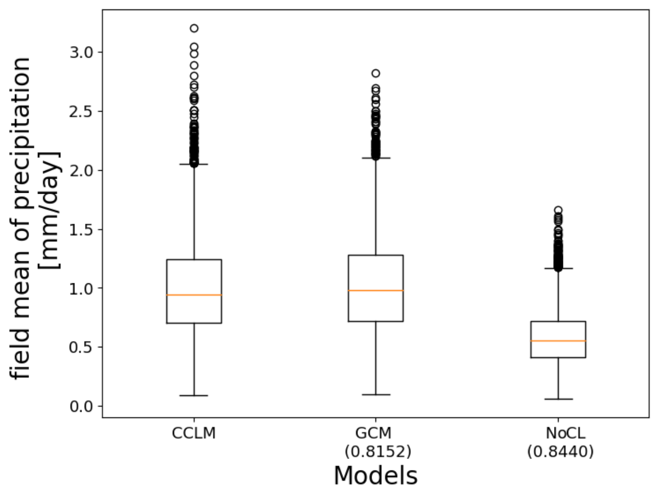
~~The final model run outputs could be find here : .~~



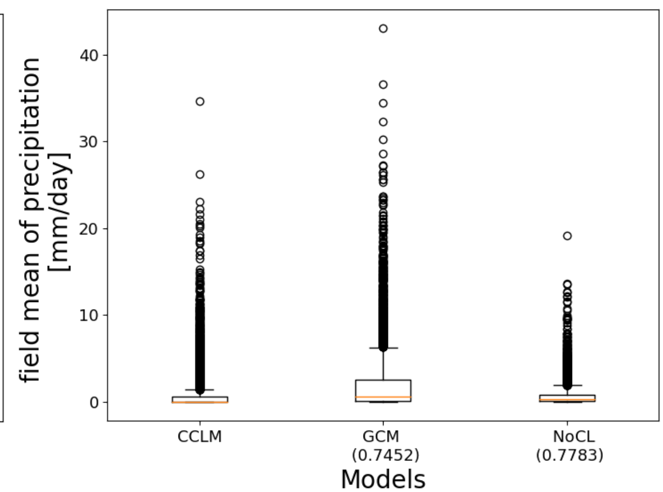
(a) MAE(EC-Earth3-Veg, CCLM)



(b) AV NoCL



(c) CA



(d) Box

Figure 10. a) MAE of GCM (EC-Earth3-Veg) vs CCLM run. GCM is remapped bilinearly to the 0.25×0.25 grid. b) Added value (AV) or MAE reduction (MAE(EC-Earth3-Veg,CCLM) - MAE(CNN,CCLM) for unconstrained method. c) and d) boxplots of averaged daily precipitation over the CA domain and the black box shown in a and b over North of Iran. Numbers in the parenthesis indicate the correlation coefficients of each model with respect to CCLM.

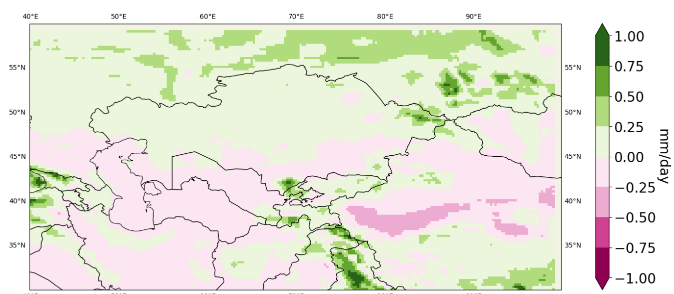


Figure 11. Added value (AV) or MAE reduction (MAE(EC-MPI-ESM1-2HR, CCLM)—MAE(CNN, CCLM)) for an unconstrained method that was not trained but applied to the SSP370 scenario.

Author contributions. BF, conducted the dynamical and statistical downscaling with help of ER and PH, respectively. ER provided the CCLM set-up. PH provided the deep learning model code and set-up. All authors contributed to the analysis of the results and writing the manuscript.

Competing interests. No competing interests.

Acknowledgements. BF thanks the German Climate Computing Center (DKRZ) for its support in using supercomputer data and resources. The German Foreign Office funds BF and ID via the Green Central Asia project (http://greencentralasia.org/en, last access: 4 July 2023). The DKRZ and PIK provided the computational resources. The authors gratefully acknowledge the German Federal Ministry of Education and Research and the Land Brandenburg for supporting this project by providing resources on the high performance computer system at the Potsdam Institute for Climate Impact Research. BF thanks the CCLM-community for providing the model code and the pre-processing code to convert the GCM to CCLM input files.

References

Allan, R. P., Hawkins, E., Bellouin, N., and Collins, B.: IPCC, 2021: summary for Policymakers, IPCC, 2021.

Ban, N., Schmidli, J., and Schär, C.: Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster?, Geophysical Research Letters, 42, 1165–1172, 2015.

Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M.: On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections, Climate Dynamics, 57, 2941–2951, 2021.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., and Ziese, M.: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present, Earth System Science Data, 5, 71–99, 2013.

Böhm, U., Gerstengarbe, F.-W., Hauffe, D., Kücken, M., Österle, H., and Werner, P. C.: Dynamic regional climate modeling and sensitivity experiments for the northeast of Brazil, Global Change and Regional Impacts: Water Availability and Vulnerability of Ecosystems and Society in the Semiarid Northeast of Brazil, pp. 153–170, 2003.

Chokkavarapu, N. and Mandla, V. R.: Comparative study of GCMs, RCMs, downscaling and hydrological models: a review toward future climate change impact estimation, SN Applied Sciences, 1, 1698, 2019.

Ciarlo, J. M., Coppola, E., Fantini, A., Giorgi, F., Gao, X., Tong, Y., Glazer, R. H., Torres Alavez, J. A., Sines, T., Pichelli, E., et al.: A new spatially distributed added value index for regional climate models: the EURO-CORDEX and the CORDEX-CORE highest resolution ensembles, Climate Dynamics, 57, 1403–1424, 2021.

Cui, T., Li, C., and Tian, F.: Evaluation of temperature and precipitation simulations in CMIP6 models over the Tibetan Plateau, Earth and Space Science, 8, e2020EA001 620, 2021.

Demory, M.-E., Berthou, S., Fernández, J., Sørland, S. L., Brogli, R., Roberts, M. J., Beyerle, U., Seddon, J., Haarsma, R., Schär, C., et al.: European daily precipitation according to EURO-CORDEX regional climate models (RCMs) and high-resolution global climate models (GCMs) from the High-Resolution Model Intercomparison Project (HighResMIP), Geoscientific Model Development, 13, 5485–5506, 2020.

Déqué, M., Rowell, D., Lüthi, D., Giorgi, F., Christensen, J., Rockel, B., Jacob, D., Kjellström, E., De Castro, M., and van den Hurk, B.: An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections, Climatic Change, 81, 53–70, 2007.

Di Luca, A., de Elía, R., and Laprise, R.: Potential for added value in precipitation simulated by high-resolution nested regional climate models and observations, Climate dynamics, 38, 1229–1247, 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.

Didovets, I., Krysanova, V., Nurbatsina, A., Fallah, B., Krylova, V., Saparova, A., Niyazov, J., Kalashnikova, O., and Hattermann, F. F.: Attribution of current trends in streamflow to climate change for 12 Central Asian catchments, Climatic Change, 177, 16, 2024.

Dong, C., Loy, C. C., He, K., and Tang, X.: Image super-resolution using deep convolutional networks, IEEE transactions on pattern analysis and machine intelligence, 38, 295–307, 2015.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., et al.: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. Geosci. Model Dev. 15, 2973–3020, 2022.

Dosio, A. and Panitz, H.-J.: Climate change projections for CORDEX-Africa with COSMO-CLM regional climate model and differences with the driving global climate models, Climate Dynamics, 46, 1599–1625, 2016.

Fallah, B. and Rostami, M.: Exploring the impact of the recent global warming on extreme weather events in Central Asia using the counterfactual climate data ATTRICI v1. 1, Climatic Change, preprint, https://doi.org/10.21203/rs.3.rs-2106031/v1, 2024.

Fallah, B., Saberi, A. A., and Sodoudi, S.: Emergence of global scaling behaviour in the coupled Earth-atmosphere interaction, Scientific Reports, 6, 34 005, 2016a.

Fallah, B., Sodoudi, S., and Cubasch, U.: Westerly jet stream and past millennium climate change in Arid Central Asia simulated by COSMO-CLM model, Theoretical and Applied Climatology, 124, 1079–1088, 2016b.

Fallah, B., Russo, E., Menz, C., Hoffmann, P., Didovets, I., and Hattermann, F. F.: Anthropogenic influence on extreme temperature and precipitation in Central Asia, Scientific Reports, 13, 6854, 2023.

Feser, F., Rockel, B., von Storch, H., Winterfeldt, J., and Zahn, M.: Regional climate models add value to global model data: a review and selected examples, Bulletin of the American Meteorological Society, 92, 1181–1192, 2011.

Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, International journal of climatology, 37, 4302–4315, 2017.

Fotso-Nguemo, T. C., Vondou, D. A., Pokam, W. M., Djomou, Z. Y., Diallo, I., Haensler, A., Tchotchou, L. A. D., Kamsu-Tamo, P. H., Gaye, A. T., and Tchawoua, C.: On the added value of the regional climate model REMO in the assessment of climate change signal over Central Africa, Climate Dynamics, 49, 3813–3838, 2017.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, International Journal of Climatology: A Journal of the Royal Meteorological Society, 27, 1547–1578, 2007.

Frei, C., Christensen, J. H., Déqué, M., Jacob, D., Jones, R. G., and Vidale, P. L.: Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps, Journal of Geophysical Research: Atmospheres, 108, 2003.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Scientific data, 2, 1–21, 2015.

Galea, D., Ma, H.-Y., Wu, W.-Y., and Kobayashi, D.: Deep Learning Image Segmentation for Atmospheric Rivers, Artificial Intelligence for the Earth Systems, 3, 230 048, 2024.

Gardoll, S. and Boucher, O.: Classification of tropical cyclone containing images using a convolutional neural network: Performance and sensitivity to the learning dataset, Geoscientific Model Development, 15, 7051–7073, 2022.

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

Giot, O., Termonia, P., Degrauwe, D., De Troch, R., Caluwaerts, S., Smet, G., Berckmans, J., Deckmyn, A., De Cruz, L., De Meutter, P., et al.: Validation of the ALARO-0 model within the EURO-CORDEX framework, Geoscientific Model Development, 9, 1143–1152, 2016.

Harder, P., Yang, Q., Ramesh, V., Sattigeri, P., Hernandez-Garcia, A., Watson, C., Szwarcman, D., and Rolnick, D.: Generating physically-consistent high-resolution climate data with hard-constrained neural networks, 2022.

Harder, P., Hernandez-Garcia, A., Ramesh, V., Yang, Q., Sattegeri, P., Szwarcman, D., Watson, C., and Rolnick, D.: Hard-Constrained Deep Learning for Climate Downscaling, Journal of Machine Learning Research, 24, 1–40, 2023.

Hess, P., Drüke, M., Petri, S., Strnad, F. M., and Boers, N.: Physically constrained generative adversarial networks for improving precipitation fields from Earth system models, Nature Machine Intelligence, 4, 828–839, 2022.

Hewitson, B., Daron, J., Crane, R., Zermoglio, M., and Jack, C.: Interrogating empirical-statistical downscaling, Climatic change, 122, 539–554, 2014.

Hodson, T. O.: Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not, Geoscientific Model Development, 15, 5481–5487, 2022.

Hong, S.-Y. and Kanamitsu, M.: Dynamical downscaling: Fundamental issues from an NWP point of view and recommendations, Asia-Pacific Journal of Atmospheric Sciences, 50, 83–104, 2014.

Jacob, D. and Podzun, R.: Sensitivity studies with the regional climate model REMO, Meteorology and atmospheric physics, 63, 119–129, 1997.

Jacob, D., Elizalde, A., Haensler, A., Hagemann, S., Kumar, P., Podzun, R., Rechid, D., Remedio, A. R., Saeed, F., Sieck, K., et al.: Assessing the transferability of the regional climate model REMO to different coordinated regional climate downscaling experiment (CORDEX) regions, Atmosphere, 3, 181–199, 2012.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., et al.: EURO-CORDEX: New high-resolution climate change projections for European impact research, Regional Environmental Change, 14, 563–578, 2014.

Jouvet, G. and Cordonnier, G.: Ice-flow model emulator based on physics-informed deep learning, Journal of Glaciology, pp. 1–15, 2023.

Kendon, E., Roberts, N., Fowler, H., Roberts, M., Chan, S., and Senior, C.: Heavier summer downpours with climate change revealed by weather forecast resolution model. Nat. Climate Change, 4, 570–576, 2014.

Kikstra, J. S., Nicholls, Z. R., Smith, C. J., Lewis, J., Lamboll, R. D., Byers, E., Sandstad, M., Meinshausen, M., Gidden, M. J., Rogelj, J., et al.: The IPCC Sixth Assessment Report WGIII climate assessment of mitigation pathways: from emissions to global temperatures, Geoscientific Model Development, 15, 9075–9109, 2022.

Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., and Lerner-Lam, A.: A global landslide catalog for hazard applications: method, results, and limitations, Natural Hazards, 52, 561–575, 2010.

Kjellström, E., Nikulin, G., Hansson, U., Strandberg, G., Ullerstig, A., Willén, U., and Wyser, K.: 21st century changes in the European climate: uncertainties derived from an ensemble of regional climate model simulations, Tellus A: Dynamic Meteorology and Oceanography, 63, 24–40, 2011.

Klok, E. and Klein Tank, A.: Updated and extended European dataset of daily weather observations, International Journal of Climatology, 28, 2081–2095, 2008.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., Van Meijgaard, E., et al.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geoscientific Model Development, 7, 1297–1333, 2014.

Kurth, T., Treichler, S., Romero, J., Mudigonda, M., Luehr, N., Phillips, E., Mahesh, A., Matheson, M., Deslippe, J., Fatica, M., et al.: Exascale deep learning for climate analytics, in: SC18: International conference for high performance computing, networking, storage and analysis, pp. 649–660, IEEE, 2018.

Laflamme, E. M., Linder, E., and Pan, Y.: Statistical downscaling of regional climate model output to achieve projections of precipitation extremes, Weather and climate extremes, 12, 15–23, 2016.

Lange, S.: Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1. 0), Geoscientific Model Development, 12, 3055–3070, 2019.

Leinonen, J., Nerini, D., and Berne, A.: Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network, IEEE Transactions on Geoscience and Remote Sensing, 59, 7211–7223, 2020.

Lenz, C.-J., Früh, B., and Adalatpanah, F. D.: Is there potential added value in COSMO–CLM forced by ERA reanalysis data?, Climate Dynamics, 49, 4061–4074, 2017.

Li, L., Bisht, G., and Leung, L. R.: Spatial heterogeneity effects on land surface modeling of water and energy partitioning, Geoscientific Model Development, 15, 5489–5510, 2022.

Lundquist, J., Hughes, M., Gutmann, E., and Kapnick, S.: Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks, Bulletin of the American Meteorological Society, 100, 2473–2490, 2019.

Maraun, D. and Widmann, M.: Statistical downscaling and bias correction for climate research, Cambridge University Press, 2018.

Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A.: VALUE: A framework to validate downscaling approaches for climate change studies, Earth's Future, 3, 1–14, 2015.

Meredith, E. P., Rust, H. W., and Ulbrich, U.: A classification algorithm for selective dynamical downscaling of precipitation extremes, Hydrology and Earth System Sciences, 22, 4183–4200, 2018.

Mitchell, T. D. and Hulme, M.: Predicting regional climate change: living with uncertainty, Progress in Physical Geography, 23, 57–78, 1999.

Muttaqien, F. H., Rahadianti, L., and Latifah, A. L.: Downscaling for Climate Data in Indonesia Using Image-to-Image Translation Approach, in: 2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 1–8, IEEE, 2021.

Naddaf, M.: Climate change is costing trillions-and low-income countries are paying the price., Nature, 2022.

Panitz, H.-J., Dosio, A., Büchner, M., Lüthi, D., and Keuler, K.: COSMO-CLM (CCLM) climate simulations over CORDEX-Africa domain: analysis of the ERA-Interim driven simulations at 0.44 and 0.22 resolution, Climate dynamics, 42, 3015–3038, 2014.

Qiu, Y., Feng, J., Yan, Z., and Wang, J.: HCPD-CA: high-resolution climate projection dataset in central Asia, Earth System Science Data, 14, 2195–2208, 2022.

Racah, E., Beckham, C., Maharaj, T., Ebrahimi Kahou, S., Prabhat, M., and Pal, C.: Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events, Advances in neural information processing systems, 30, 2017.

Rampal, N., Hobeichi, S., Gibson, P. B., Baño-Medina, J., Abramowitz, G., Beucler, T., González-Abad, J., Chapman, W., Harder, P., and Gutiérrez, J. M.: Enhancing Regional Climate Downscaling through Advances in Machine Learning, Artificial Intelligence for the Earth Systems, 3, 230 066, 2024.

Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., et al.: Climate models and their evaluation, in: Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR), pp. 589–662, Cambridge University Press, 2007.

Rasp, S. and Lerch, S.: Neural networks for postprocessing ensemble weather forecasts, Monthly Weather Review, 146, 3885–3900, 2018.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat, f.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, 2019.

Reyer, C. P., Otto, I. M., Adams, S., Albrecht, T., Baarsch, F., Cartsburg, M., Coumou, D., Eden, A., Ludi, E., Marcus, R., et al.: Climate change impacts in Central Asia and their implications for development, Regional Environmental Change, 17, 1639–1650, 2017.

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., et al.: The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview, Global environmental change, 42, 153–168, 2017.

Rockel, B. and Geyer, B.: The performance of the regional climate model CLM in different climate regions, as simulated in a transient climate change experiment, Climate Dynamics, 31, 713–728, 2008.

Rummukainen, M.: State-of-the-art with regional climate models, Wiley Interdisciplinary Reviews: Climate Change, 1, 82–96, 2010.

Russo, E., Kirchner, I., Pfahl, S., Schaap, M., and Cubasch, U.: Sensitivity studies with the regional climate model COSMO-CLM 5.0 over the CORDEX Central Asia Domain, Geoscientific Model Development, 12, 5229–5249, 2019.

Russo, E., Sørland, S. L., Kirchner, I., Schaap, M., Raible, C. C., and Cubasch, U.: Exploring the parameter space of the COSMO-CLM v5. 0 regional climate model for the Central Asia CORDEX domain, Geoscientific Model Development, 13, 5779–5797, 2020.

Russo, E., Fallah, B., Ludwig, P., Karremann, M., and Raible, C. C.: The long-standing dilemma of European summer temperatures at the mid-Holocene and other considerations on learning from the past for the future using a regional climate model, Climate of the Past Discussions, 2021, 1–27, 2021.

Serifi, A., Günther, T., and Ban, N.: Spatio-temporal downscaling of climate data using convolutional and error-predicting neural networks, Frontiers in Climate, 3, 656 479, 2021.

Smith, J. A., Baeck, M. L., Villarini, G., and Krajewski, W. F.: The hydrology and hydrometeorology of flooding in the Delaware River Basin, Journal of Hydrometeorology, 11, 841–859, 2010.

Sørland, S. L., Schär, C., Lüthi, D., and Kjellström, E.: Bias patterns and climate change signals in GCM-RCM model chains, Environmental Research Letters, 13, 074 017, 2018.

Sørland, S. L., Brogli, R., Pothapakula, P. K., Russo, E., Van de Walle, J., Ahrens, B., Anders, I., Bucchignani, E., Davin, E. L., Demory, M.-E., et al.: COSMO-CLM regional climate simulations in the Coordinated Regional Climate Downscaling Experiment (CORDEX) framework: a review, Geoscientific Model Development, 14, 5125–5154, 2021.

Stengel, K., Glaws, A., Hettinger, D., and King, R. N.: Adversarial super-resolution of climatological wind and solar data, Proceedings of the National Academy of Sciences, 117, 16 805–16 815, 2020.

Sun, L. and Lan, Y.: Statistical downscaling of daily temperature and precipitation over China using deep learning neural models: Localization and comparison with other methods, International Journal of Climatology, 41, 1128–1147, 2021.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American meteorological Society, 93, 485–498, 2012.

Torma, C., Giorgi, F., and Coppola, E.: Added value of regional climate modeling over areas characterized by complex terrain—Precipitation over the Alps, Journal of Geophysical Research: Atmospheres, 120, 3957–3972, 2015.

Volosciuk, C., Maraun, D., Vrac, M., and Widmann, M.: A combined statistical bias correction and stochastic downscaling method for precipitation, Hydrology and Earth System Sciences, 21, 1693–1719, 2017.

Wang, D., Menz, C., Simon, T., Simmer, C., and Ohlwein, C.: Regional dynamical downscaling with CCLM over East Asia, Meteorology and Atmospheric Physics, 121, 39–53, 2013.

Wang, X., Otto, M., and Scherer, D.: Atmospheric triggering conditions and climatic disposition of landslides in Kyrgyzstan and Tajikistan at the beginning of the 21st century, Natural Hazards and Earth System Sciences, 21, 2125–2144, 2021.

Watson-Parris, D., Rao, Y., Olivié, D., Seland, Ø., Nowack, P., Camps-Valls, G., Stier, P., Bouabid, S., Dewey, M., Fons, E., et al.: ClimateBench v1. 0: A benchmark for data-driven climate projections, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 954, 2022.

Xu, P., Wang, L., and Ming, J.: Central Asian precipitation extremes affected by an intraseasonal planetary wave pattern, Journal of Climate, 35, 2603–2616, 2022.

Xu, Z., Han, Y., Tam, C.-Y., Yang, Z.-L., and Fu, C.: Bias-corrected CMIP6 global dataset for dynamical downscaling of the historical and future climate (1979–2100), Scientific Data, 8, 293, 2021.

Yan, Y., You, Q., Wu, F., Pepin, N., and Kang, S.: Surface mean temperature from the observational stations and multiple reanalyses over the Tibetan Plateau, Climate Dynamics, 55, 2405–2419, 2020.

Yang, D., Liu, S., Hu, Y., Liu, X., Xie, J., and Zhao, L.: Predictor selection for CNN-based statistical downscaling of monthly precipitation, Advances in Atmospheric Sciences, 40, 1117–1131, 2023.

Yatagai, A., Kitoh, A., Kamiguchi, K., Arakawa, O., Kawamoto, H., Tahashima, H., Watanabe, T., Kuboto, J., Taneguchi, M., and Kanae, S.: Asian precipitation–highly resolved observational data integration towards evaluation of the water resources, in: Presentation at PHERPP meeting, WMO, Geneva December, 2007.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., and Fraundorfer, F.: Deep learning in remote sensing: A comprehensive review and list of resources, IEEE geoscience and remote sensing magazine, 5, 8–36, 2017.