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 1:

## 1.1 General comments:

**I appreciate that the authors have made some effort to add two new figures, which are important new information, and their research is a lot more detailed and feels reproducible. I still think that the quality of their research is also good. However, the quality of the manuscript still does not feel ready for publishing. I started reviewing until page 7 (see attached reviewed manuscript) but did not continue because I saw that some of my previous comments on the first manuscript were not incorporated. There are still acronyms that have not been introduced, information (or even the same sentences) repeated several times throughout a section, and many grammatical mistakes. I recommend that they have a native English speaker go through the manuscript (or use tools like Grammarly) before sending it again for review. I ask the authors to apply my recommendations to the whole manuscript (not only my comments till page 7) and re-read their paper thoroughly to ensure everything is in its right place.**

We appreciate your effort in reviewing the revised manuscript. Following your recommendations, we have used Grammarly to correct the grammatical mistakes and have thoroughly reviewed the entire manuscript to eliminate any information repetition. We hope that the new version now meets the GMD standards. The numerous changes made are visible in the tracked changes version of the manuscript. We added the report and the score achieved from Grammarly at the end of this answer. As you can notice, we reached the score of 99 and the Plagiarism index of 2 %, which is mainly related to the open discussion version of our paper in GMD.

You mentioned where the abbreviation SSP was first introduced in the main text, but this abbreviation was already spelled out in the abstract.

Regarding the comment on $y_i$, we had two $y$s : $\tilde{y}_i$ and $y_i$.

# 2 Reviewer 2:

## 2.1 Minor comments

**Most of my review comments have been addressed. However, the statement in the response letter on the use of MAE is inconsistent with the manuscript. There are also a few minor points that should be corrected or clarified before publication, which are listed below. Use of MAE I had commented on the fact that free-running GCM simulations and observation are completely different realisations of random internal variability, and that any comparison that uses pairs for a given time, including calculating the MAE between timeseries from GCM-driven RCMs or emulators and from observations makes no sense. In response to this, the authors have removed the original section 3.1.1. and Figure 3. This is however not made clear in section 2.1. of the response letter, which says '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.' This reads as if the averaging over 30-years had been done in the original version, which is not the case, and no wrong analysis had been conducted. It only becomes clear in section 2.2. of the response letter that the original section 3.1.1 of the manuscript has been removed. It is unclear to me why the authors discussed their response to this issue in such an unclear way. More importantly, the statement in the response letter means that there should now be an evaluation of the representation of the annual cycle instead of an evaluation of the temporal variability in the timeseries. However, it seems that this is not the case. The explanation of the MAE calculation (lines 228- 232) does only mention timesteps, which indicates the use of the simulated and observed timeseries, and the averaging over 30 years for each day of the year is not mentioned. The MAE results that are shown are based on setups where the pairing is justified (comparing ERA- Interim-driven CCLM simulations with observations, or comparing GCM-driven CCLM and CNN simulations) and there is no indication that any of these results are based on the annual cycle rather than on the full timeseries. Please clarify the situation and explain the use of MAE in the manuscript such that there is no room for misinterpretations of what has been done.**

We agree fully with this point that the response was not fully clear and apologies for the unclear answer. The

above-mentioned sentences were also not fully explaining the desired calculation. As you correctly hinted, now we use MAE just for pairs of simulations when the usage could be justified. The $t$ in equation 5 will remain the daily time-step and $T$= 365 days $\times$ 30 years, which shall be enough data for shaping the statistics. We will add a clear sentence to clarify this.

**Further points**

- **The terminology for the mapping from low resolution to high resolution randomly switches between 'upscaling' and 'downscaling'. A consistent terminology should be used, preferably using the standard term 'downscaling'.**

- **The reason for randomly shuffling the data (line 160) should be given, at it should be clarified that the input and output datasets are shuffled in the same way in order to retain the original pairing.**

- **The reason for choosing the SoftMax constraining (equation 2) rather than a simple, linear scaling is explained in the response letter, but not in the paper. The explanation should be added to the paper.**

Thanks a lot for the clarifying comment. We changed the up-scaling to downscaling to keep the form homogeneous throughout the manuscript.

We added the following paragraph for clarifying the shuffling of the data: "We randomly shuffled the dataset at the pair level, ensuring that each input (GCM) and its corresponding output (RCM) were shuffled together. We preserve the intrinsic relationships between coarse-resolution and fine-resolution data across diverse forcing scenarios. Maintaining the original pairing ensured that the model accurately captures the dependencies between inputs and outputs, which is crucial for effective downscaling. The rationale for shuffling the dataset was to mitigate any temporal or spatial autocorrelation that could bias the training process. This randomization helps the model generalize better by exposing it to diverse conditions, leading to more robust and reliable performance. Preserving input-output pairs during shuffling ensures that the fine-resolution targets remain correctly aligned with their corresponding coarse-resolution features, retaining the integrity of the data relationships throughout the training process."

We added the following information to the new version of the manuscript :

"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. However, (Harder et al., 2022)

showed 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 struggles when the sum of the small-scale values gets close to zero."

On behalf of all authors,

Bijan Fallah

# References

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.

# Untitled

by Bijan Fallah

## General metrics

| 44,608 | 6,109 | 303 | 24 min 26 sec | 46 min 59 sec |
|---|---|---|---|---|
| characters | words | sentences | reading time | speaking time |

## Score


This text scores better than 99% of all texts checked by Grammarly

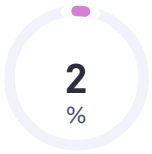
## Writing Issues

**23** Issues left

Critical

**23** Advanced

## Plagiarism




**11** sources

2% of your text matches 11 sources on the web or in archives of academic publications

## Writing Issues

| | | |
|---|---|---|
| 22 | **Clarity** | |
| 14 | Passive voice misuse |  |
| 1 | Intricate text |  |
| 5 | Wordy sentences |  |
| 2 | Unclear sentences |  |
| 1 | **Delivery** | |
| 1 | Inappropriate colloquialisms |  |

## Unique Words

Measures vocabulary diversity by calculating the percentage of words used only once in your document

**22%**

unique words

## Rare Words

Measures depth of vocabulary by identifying words that are not among the 5,000 most common English words.

**42%**

rare words

## Word Length

Measures average word length

**5.5**

characters per word

## Sentence Length

Measures average sentence length

**20.2**

words per sentence

# Untitled

\begin{abstract}

High-resolution climate projections are essential for estimating future climate change impacts. Statistical and dynamical downscaling methods, or a hybrid of both, are commonly employed to generate input datasets for impact modelling. In this study, we utilise the regional climate model (RCM) COSMO-CLM (CCLM) version 6.0 to assess the added value of dynamically downscaling a general circulation model (GCM) from the sixth phase of the Coupled Model Intercomparison Project (CMIP6) and its climate change projections over Central Asia (CA). The MPI-ESM1-2-HR GCM (at 1° spatial resolution) drives the CCLM (at 0.22° horizontal resolution) for the historical period of 1985-2014 and the projection period of 2019-2100 under three 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 a reference, we evaluate the performance of CCLM driven by ERA-Interim reanalysis over the historical period. The added value of CCLM, compared to its driving GCM, is significant (more than 5 mm/day) over mountainous areas in CA, which are at higher risk of extreme precipitation events. Additionally, we employ CCLM to refine future climate projections. We present high-resolution maps of heavy precipitation changes based on CCLM and compare them with the CMIP6 GCM ensemble. Our analysis indicates a significant increase in the intensity and frequency of heavy precipitation events over CA areas already at risk of extreme climatic events by

the end of the century. Finally, we train a convolutional neural network (CNN) to map a GCM simulation to its dynamically downscaled CCLM counterpart. The CNN successfully emulates the GCM-CCLM model chain over large CA areas, demonstrating added value when applied to a new GCM-CCLM model chain. The scientific community interested in downscaling CMIP6 models could use our downscaling data, and the CNN architecture offers an alternative to traditional dynamical and statistical methods.
\end{abstract}

\introduction %% \introduction[modified heading if necessary]

The increasing global mean temperature due to anthropogenic greenhouse gas emissions presents a significant challenge for society, requiring the assessment and prediction of future impacts on human health, natural ecosystems, and economies across different regions of the world \citep{allan2021ipcc}. Researchers conducting regional studies on vulnerability, impacts, and adaptation typically achieve reliable high-resolution climate projections through dynamical downscaling via RCMs \citep{rummukainen2010state, feser2011regional}, statistical techniques \citep{maraun2018statistical, fowler2007linking}, or a hybrid of both approaches \citep{maraun2015value, meredith2018classification, laflamme2016statistical}.

CA, recognised as one of the most vulnerable regions to climate change impacts, heavily depends on water resources from glaciers and rivers that are shrinking due to rising temperatures and decreasing precipitation \citep{reyer2017climate, fallah2023anthropogenic, didovets2024attribution,

fallah2024exploring}. The area faces significant challenges to food security, characterised by declining crop yields and an increased occurrence of severe and frequent extreme weather events like floods and landslides. These conditions damage infrastructure, livelihoods, and agriculture, resulting in population displacement and migration \citep{allan2021ipcc, reyer2017climate}.

Significant uncertainties inherent in the existing detailed observational and reanalysis datasets impede the development of high-resolution climate projections in CA \citep{fallah2016emergence}. One option to complement these datasets is to use dynamical downscaling with RCMs. CMIP6 provides a framework for coordinated climate model experiments, enhancing our understanding of past, present, and future climate changes. Dynamical downscaling of CMIP6 models for the CA region is vital for accurately simulating extreme convective precipitation events, which are influenced by the orography of the region \citep{lundquist2019our, ban2015heavy wang2013regional,frei2003daily,russo2019sensitivity}, large-scale atmospheric circulation, and sea surface temperature anomalies in the Indian Ocean and the Pacific \citep{kendon2014heavier, demory2020european, xu2022central}. This method enhances the resolution of a driving GCM and produces a robust, physically consistent regional state of the climate. Despite some systematic biases, dynamical downscaling consistently provides high-quality datasets that accurately describe the climatology of all climate variables in CA \citep{qiu2022hcpd}.

Various international institutions have collaborated within the Coordinated Regional Climate Downscaling Experiment (CORDEX) to address these issues

and improve the inter-comparability of RCMs. CORDEX aims to create a robust framework for producing climate projections at a regional scale that is suitable for impact evaluation and adaptation planning globally. This effort aligns with the timeline of the Intergovernmental Panel on Climate Change's Sixth Assessment Report \citep{kikstra2022ipcc}. However, most CORDEX research focuses on highly industrialised countries \citep{allan2021ipcc, taylor2012overview}. Developing regions, including CA, bear the brunt of global warming's consequences, yet they have access to only a limited number of CORDEX model simulations \citep{naddaf2022climate}. As of the latest update, no simulation driven by CMIP6 has been planned for CORDEX-CA (see \url{https://wcrp-cordex.github.io/simulation-status/CMIP6_downscaling_plans.html}, last visited on 17.04.2024).

Beyond dynamical methods, recent developments in machine learning, including CNNs, offer promising avenues for statistical downscaling \citep{harder2023hard, rampal2024enhancing}. CNNs have proven effective in numerous earth science disciplines besides downscaling, such as classification \citep{gardoll2022classification}, segmentation \citep{galea2024deep}, and prediction \citep{watson2022climatebench} thanks to their capacity to extract features from spatial data and identify nonlinear relationships between inputs and outputs. CNNs can recognise and encode spatial hierarchies in data \citep{zhu2017deep}, making them exceptionally suitable for analysing geospatial data, a critical component in climate modelling. Unlike traditional statistical methods that often require manual selection and careful engineering of features, CNNs automatically learn the most predictive features directly from the data \citep{reichstein2019deep}. They are generally more straightforward and efficient than traditional

statistical downscaling methods for tasks aiming to predict or classify patterns distributed across spatial domains, such as temperature or precipitation patterns in climate models \citep{racah2017extremeweather}. 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 \citep{kurth2018exascale}.

Researchers classify CNNs into two categories based on their last layer: 1) constrained and 2) unconstrained. Constrained CNNs integrate physical laws directly into the training process, such as mass, energy, or momentum conservation. This integration is achieved by modifying the loss function or the network's architecture to ensure compliance with these laws. In contrast, unconstrained CNNs do not explicitly incorporate physical laws or constraints. Instead, they rely solely on learning from the input data, generating output predictions based on the patterns detected in the data.

This study explores unconstrained and constrained CNN approaches to understand their effectiveness in downscaling and their performance when applied to GCMs not initially used for training.

The research questions guiding this study are:
\begin{itemize}
\item \textbf{Research Question 1:} How effectively can CMIP6 models be downscaled to enhance precipitation simulations for the CORDEX Central Asia region?
\item \textbf{Research Question 2:} Can CNNs effectively downscale GCM outputs, and how do they perform when applied to GCMs that did not initially

train them?
\end{itemize}

This article focuses on two main topics: 1) the added value of CCLM for representing precipitation over Central Asia, and 2) training a CCLM emulator using a CNN. We present data and methods in Section 2. Sections 3 and 4 introduce the results of dynamical and hybrid downscaling, respectively. Finally, we discuss the results and conclude in Section 5.

\section{Data and Methods}

The methodology employed in this study is illustrated in Figure \ref{fig:0}. The following sections provide a detailed explanation of this methodology.

\subsection{Employed Models and Experimental Setups}
\subsubsection{Regional Climate Model (RCM)}
In this study, we conduct simulations using the CCLM regional climate model. Developed by the German Weather Service (DWD) and the German Climate Computing Center (Deutsches Klimarechenzentrum, DKRZ), CCLM originates from the COSMO numerical weather prediction model \citep{rockel2008performance}, which is widely utilised for short-term weather forecasting. Explicitly designed for regional climate simulation, CCLM enables researchers to investigate various aspects of the climate system, including temperature, precipitation, and extreme events. It has been extensively used to assess the impact of climate change across different regions such as Europe \citep{russo2021long}, Africa \citep{panitz2014cosmo,dosio2016climate}, and Asia \citep{jacob2014euro,kotlarski2014regional,wang2013regional}.

Additionally, CCLM has been employed in climate projection studies to evaluate climate adaptation and mitigation strategies. The model has undergone thorough evaluation and validation \citep{fallah2016westerly,russo2019sensitivity,kjellstrom2011century}, and its ability to generate realistic simulations of present climate conditions and variability has established it as one of the most widely used regional climate models in the scientific community \citep{sorland2021cosmo}.

For our experiments, we utilised a model setup similar to the "optimal" configuration described by \cite{russo2019sensitivity}. In their study, \cite{russo2019sensitivity} optimised the CCLM regional climate model for CA by adjusting albedo based on forest fraction ratios and soil conductivity to account for the soil's liquid water and ice proportions. These modifications significantly improved the model's climatological performance and the distribution of incoming radiation, leading to more accurate climate representations for the region. According to the CORDEX protocol, simulations are divided into two primary phases. The first phase, the evaluation run, involves a single model experiment over the period 1979-2014, using ERAInterim reanalysis data at a spatial resolution of T255 ($\sim$ 0.7\textdegree). The second phase, the projection run, utilises boundary conditions from GCMs of the CMIP6 project for the period 1950-2100 under various SSPs. For this study, we selected the MPI-ESM1-2-HR GCM and considered SSP1-2.6, SSP3-7.0, and SSP5-8.5 scenarios. SSPs represent baseline scenarios that describe future pathways based on population growth, technological advancement, economic development, urbanisation, and investments in healthcare, education, land use, and energy \citep{riahi2017shared}.

Historical data for this study are based on greenhouse gas levels, land use, and other climate forcings observed from 1850 to 2014. The Shared Socioeconomic Pathway (SSP) scenarios used in the projections are as follows:
\begin{itemize}
\item \textbf{SSP1-2.6} represents a "green" future, characterised by global efforts to protect resources, improve human well-being, and narrow income gaps. This scenario assumes low challenges to adaptation and low greenhouse gas emissions. Adaptation challenges in this context refer to the difficulties societies might face in adjusting to the impacts of climate change, including their susceptibility and the availability and effectiveness of mitigation technologies and strategies. Under SSP1-2.6, global cooperation and sustainable practices lead to advancements in technology and governance, significantly reducing vulnerability to climate change impacts. Societal structures are resilient, and resources are managed to minimise environmental stresses while maximising human well-being.
\item \textbf{SSP3-7.0} depicts a future characterised by regional rivalry, where nationalism and regional conflicts dominate, global issues are neglected, and inequality increases. This scenario involves high challenges to adaptation and high greenhouse gas emissions.
\item \textbf{SSP5-8.5} represents a future of fossil-fueled development with globally connected markets, rapid technological progress, and weak environmental policies. This scenario has low challenges to adaptation but results in very high greenhouse gas emissions.
\end{itemize}

For comparison and evaluation of our RCM simulations, we have selected two CORDEX-CA evaluation simulations from other models driven by ERAInterim at

a 0.22° horizontal resolution: 1) \textbf{ERAInterim-RMIB-UGent-ALARO-0} \citep{giot2016validation} and 2) \textbf{ERAInterim-GERICS-REMO2015} \citep{jacob1997sensitivity,fotso2017added}.

\subsubsection{CNNs}

In this study, we develop a CNN-based emulator for the CCLM driven by the MPI-ESM1-2-HR GCM. This CNN utilises outputs from the GCM, covering both the historical period from 1985 to 2014 and future scenarios spanning 2019 to 2100, as inputs to model the responses of the CCLM, which serves as the target. Given the low annual precipitation and significant spatio-temporal variability in many regions of CA, a comprehensive dataset that includes various precipitation patterns from both GCMs and RCMs is essential for effectively training the CNN to map from GCM to RCM outputs. To enhance model training, we have augmented our dataset with ERA-Interim reanalysis data and corresponding CCLM simulations driven by it (ERAInterim-CCLM) (see Fig. \ref{fig:0}).

We train our CNN model based on the architecture proposed by \cite{harder2023hard}, which incorporates physical constraints to ensure mass conservation and energy balance. The model architecture features:

\begin{itemize}
\item Conv (Convolutional Layer): These layers help extract various levels of features from low-resolution images, such as edges, textures, and other relevant image details.

\item ReLU (Rectified Linear Activation Unit): This nonlinear activation function introduces non-linearity and returns the input unchanged if it is positive; otherwise, it returns zero. This function enables the network to learn complex patterns efficiently.
\item TransConv (Transposed Convolutional Layer): This layer is crucial for downscaling. It increases the spatial dimensions of the feature maps, performing a sort of learned interpolation. This allows the model to add details to the downscaled images based on the features extracted and processed in the earlier layers.
\item ResBlock (Residual Block): These blocks allow the model to refine the initial lower-resolution predictions, which are downscaled (interpolated outputs) to a higher resolution. They enhance the model's ability to add fine details and textures (high-frequency information), improving the perceptual quality and sharpness of the images at the increased resolution."
\end{itemize}

In the context of deep learning for climate modelling, the "perfect model" approach involves starting with high-resolution data and intentionally upscaling it to a lower resolution. The machine learning model is subsequently trained to reproduce the high-resolution data while receiving this artificial low-resolution input. 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 upscaled data. This approach teaches the model the desired mapping from low to high resolution, enabling the model to effectively learn how to downscale or enhance resolution while minimising the loss of critical information. It is 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:
\begin{itemize}
\item Low-resolution data: may capture global or large-scale phenomena but miss regional details \citep{xu2021bias,chokkavarapu2019comparative}.
\item High-resolution data: 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 \citep{muttaqien2021downscaling}.
\end{itemize}

In this setup, CNN's challenge is learning to map between two independently imperfect datasets. The CNN is trained to predict high-resolution details from low-resolution inputs as accurately as possible despite the absence of perfect ground truth. This process involves understanding and modelling the uncertainties and biases inherent in both datasets.

Prior to training, the dataset was randomly shuffled at the pair level to ensure that each GCM input and its corresponding RCM output remained together, preserving the intrinsic relationships between the coarse and fine-resolution data. This approach prevents temporal or spatial autocorrelation from biasing the training process. It also improves the model's generalisation and performance by exposing it to various conditions. For the dataset distribution, 68,141 days (60$\%$) of RCM simulation data were used for training, 22,714

days (20$\%$) for validation, and 22,714 days (20$\%$) for testing. The low-resolution (GCM) dataset consists of 30 $\times$ 60 grid points, and the high-resolution (RCM) dataset comprises 120 $\times$ 240 grid points over latitudes and longitudes, respectively, resulting in a downscaling factor (N) of 4.

\subsubsection{Constraint layers}\label{constraint_layers}

We test the CNN with three different constraining methods in the last CNN layer \citep{harder2023hard}: 1) soft constraining (SCL), 2) hard constraining (HCL) and 3) without constraining (NoCL). In the following, we explain briefly the three different constraining methodologies. The setup of constraining is as follows: 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 following form:

\begin{equation}
\frac{1}{n} \sum_{i=1}^{n} y_{i} = x .
\label{eq:1}
\end{equation}

\textbf{Hard constraining:} uses the SoftMax, which constrains quantities like water content by enforcing the output to be non-negative. The simplest way to ensure mass conservation would be to scale all small-scale values within a given large-scale grid cell with the ratio of the large-scale value and the sum of the small-scale values. However, \cite{harder2022} demonstrated that employing the SoftMax constraints layer gives better results. The exponential

ensures positive predictions and leads to more variance between subpixels in the super-resolved prediction. The multiplicative rescaling struggles when the sum of the small-scale values gets close to zero. Therefore, the SoftMax operator is used on the intermediate outputs of the CNN before the constraining layer ($\tilde{y}_{i}$) and multiplies it by the corresponding input pixel value $x$:

\begin{equation}
y_{i} = \exp(\tilde{y}_{j}) \cdot \frac{x}{\frac{1}{n} \sum_{i=1}^{n} \exp (\tilde{y}_{i})}.
\label{eq:2}
\end{equation}

$y_{i}$ is the final output after applying the constraints. We have used the mean absolute error (MAE) as the loss function (Eq. \ref{eq:5}).

\textbf{Soft constraining:} This is done by adding a regularisation term to the loss function. The MAE loss is then extended[1] with an additional constraint violation (CV) loss term:
\begin{equation}
\text{Loss} = (1 – \alpha) \cdot MAE + \alpha \cdot \text{CV},
\label{eq:3}
\end{equation}

Where CV is the mean-squared error over all constraint violations between an input pixel x and the super-pixel (high-resolution grid-cell) $y_{i}$:

\begin{equation}
\text{CV} = \text{MSE} (\frac{1}{n} \sum_{i=1}^{n} y_{i}, x)
\label{eq:4}
\end{equation}

We use $\alpha = 0.99$.

\textbf{Without constraining}: In this setup, we remove the constraining layer after the last convolutional layer in the CNN.

We use 160 epochs, using 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 setup as in \cite{harder2023hard}.

It is important to note that the MAE can serve both as a loss function and an evaluation metric. As a loss function, it is used during training to optimise the neural network's parameters. Conversely, when used as an evaluation metric, it is calculated on the validation or test data sets to assess the model's performance using an independent dataset. Despite their different applications, MAE is suitable for both roles.

\subsection{Evaluation and testing}

According to \cite{ciarlo2021new}, the choice of observational data can significantly influence the perceived added value of an RCM, particularly in detecting extreme events, where poor-quality data might misleadingly suggest improved model performance. They recommend using observations with spatiotemporal resolutions comparable to the model's for enhanced accuracy. In line with this, we use CHIRPS as our gridded observation to assess the added value of the CCLM driven by the GCM. CHIRPS provides a resolution of 0.05°, covers latitudes from 50°S to 50°N, and offers independent observations derived from satellite information and station data. This[2] contrasts with reanalysis data, which depend on climate model simulations \citep{funk2015climate}. We allocate 20$\%$ of the CCLM simulation data as the target to evaluate our CNN emulator instead of using CHIRPS directly. We measure the added value of the CNN by comparing the MAE of both the CNN outputs and the interpolated GCM outputs against the target CCLM output. This comparison assesses whether the CNN outperforms simple interpolation. The selected GCM and observational data are interpolated onto the RCM grid using the distance-weighted average method. \cite{ciarlo2021new} previously noted that such interpolation might create unrealistic values, as it does not account for the physical processes and could introduce artefacts depending on the interpolation method, the spatial distribution of data points, and the resolution ratio. Therefore, we use simple interpolation as a baseline, recognising its limitations in preserving the statistical properties of precipitation, which does not follow a normal distribution. Following \citep{hodson2022root}, we apply the MAE to quantify the biases in emulated and dynamically downscaled precipitation ($F$) against observations ($O$):

\begin{equation}

\text{MAE} = \frac{1}{T}\sum\limits_{t=1}^{T}|F_{t}-O_{t}| \label{eq:5}
\end{equation}

Where $T$ represents the number of time steps over 30 years of daily data. We define added value (AV) as the reduction in MAE achieved by the downscaling relative to the driving GCM:

\begin{equation}
\text{AV} = \text{MAE}_{GCM} - \text{MAE}_{CCLM}
\label{eq:6}
\end{equation}

Where $\text{MAE}_{GCM}$ and $\text{MAE}_{CCLM}$ are the differences between interpolated GCM and RCM with respect to[3] the reference dataset.

As an additional metric, we also use the climatological bias, i.e., the difference between the model and observations:
\begin{equation}
\text{BIAS} = \text{PR}_{MODEL} - \text{PR}_{OBS}
\label{eq:6-1}
\end{equation}

\section{Results}

Figure \ref{fig:1}.a illustrates the topography of the CORDEX-CA simulation domain. Figure 1.b displays the mean daily precipitation, averaged over the years 1985-2014 (mm/day), derived from CHIRPS data. The regions with the

highest precipitation are the mountainous areas of CA, particularly notable in the Asian summer monsoon region north of India and along the Himalayas in the southeastern part of the domain, where precipitation values are pronounced. Figure \ref{fig:1}.c depicts the distribution of WorldClim weather stations \citep{fick2017worldclim} across CA, serving as a proxy for the density of station data used in the CHIRPS dataset. Observational data are sparsely distributed in East China, especially over the Tibetan Plateau. Consequently, data-model comparisons are considered unreliable in this region \citep{randall2007climate,cui2021evaluation,yan2020surface,russo2019sensitivity}.

\subsection{Added value of CCLM driven by ERAInterim}

To characterise the overall performance of the CCLM model across time and space, Figures \ref{fig:2} and \ref{fig:2-1} present maps displaying annual, winter (DJF), and summer (JJA) MAE and mean biases. These biases in precipitation are calculated between the interpolated ERAInterim data and CCLM outputs driven by ERAInterim for the period 1985-2014, in comparison to CHIRPS (see Eq. \ref{eq:5} and Eq. \ref{eq:6}). Figures \ref{fig:2}.a-c illustrate the MAE for ERAInterim for annual, winter, and summer averages. The added value of the CCLM RCM compared to the interpolated ERAInterim is depicted in Figures \ref{fig:2}.d-f. During the Asian summer monsoon, CCLM's MAE is high over the south and southeast of the domain (regions in magenta), whereas it is generally lower during winter. CCLM shows an MAE reduction in the mountainous areas of Afghanistan, Kyrgyzstan, and Tajikistan, as well as an increase near the domain's southern boundaries throughout the year and in the south and southeast during summer.

The AVs of GERICS-REMO2015 and RMIB-UGent-ALARO-0 driven by ERAInterim are presented in Figures \ref{fig:2}.g-l, using CHIRPS as the observational dataset. The added value of RCM is most pronounced in areas with complex topography, especially during summer, across all three RCMs (Figs.\ref{fig:2}.d-l). Areas where the RCM has a smaller MAE than the reanalysis in comparison to observations are found[8] over Tajikistan, Kyrgyzstan, northern Afghanistan, and part of the Himalayas—regions that are crucial water sources for former Soviet Union countries. Nevertheless, precipitation during the colder seasons may be more critical for water availability. The annual AV patterns still show positive values in these regions (Figure \ref{fig:2}.d,g, and j). Across the entire domain, all three RCMs significantly reduce the large and local-scale bias of ERAInterim, especially in complex topographies. The nested RCMs exhibit similar $\text{MAE}$ values near their lateral boundaries, relative to their driving model (Figure \ref{fig:2}, \textit{a},\textit{b},\textit{c}). Thus, negative AV quantities may result from boundary effects, particularly near the eastern and southeastern boundaries where monsoonal precipitation dominates. GERICS-REMO2015 displays pronounced negative added values annually and during winter above Tibet.

Additionally, model climatology biases are displayed[9] in Figures \ref{fig:2-1}. Once again, these biases are noticeable in the lower right corner of the domain during JJA and across the southern Tibetan Plateau throughout the year.

\subsubsection{Extreme precipitation patterns in CCLM and CMIP6 GCMs}

Given that the CCLM simulation has demonstrated added value for precipitation over the mountainous regions of CA, we explore climate change signals in its high-resolution output. These high-resolution maps may inherit biases from the GCM-RCM selection and could vary under different[10] anthropogenic forcings. We assume that many model biases are consistent across different time slices and, therefore, can be removed when calculating changes between the historical period (1985-2014) and future periods (2070-2099).

We present climate change trends in CCLM and the CMIP6 GCMs ensemble statistics (ensemble mean and standard deviation). We analysed 31, 33, and 38 models for SSP126, SSP370, and SSP585 scenarios, respectively, with a total of[1] 158, 185, and 242 simulations (see Supplementary materials for the list of models used). We calculate statistics over each model's members to ensure equal weighting for individual models before building the final statistics. We have selected the yearly 99th percentile of daily precipitation (PR99), which accounts for the three days with the highest precipitation each year. Additionally, we chose the number of very heavy precipitation days during the period (ECA-RX20mm) as another index, which is commonly used in climate research to assess the impacts of heavy precipitation events on water resources, agriculture, and natural ecosystems \citep{klok2008updated}.

Figure \ref{fig:4} shows the changes in averaged PR99 at the end of the century (2070-2099) compared to the historical period (1985-2014) for CCLM (a,d,g) and CMIP6 GCMs (b,e,h) under different scenarios. The large-scale patterns remain consistent across all three scenarios, intensifying with increased anthropogenic influence. The standard deviation of the models' ensemble is

depicted in Figures \ref{fig:4}.c,f, i. Our analysis indicates that the Himalayas, particularly Nepal, North India, and Bhutan, exhibit the highest uncertainty among the GCMs in all scenarios. Except for this region and the eastern boundary of the domain, the standard deviation remains below 3 mm/day. Under the SSP585 and SSP370 scenarios, regions including Northwest India, North Pakistan, North and Southwest Iran, and the South and Southeast of the Black Sea are projected to experience increases in PR99 values exceeding 9 mm/day. A reduction in PR99 is detected in the eastern Mediterranean, specifically in Jordan, Syria, and southern Turkey. Similar patterns are observed in the CMIP6 ensemble mean, but due to averaging, the ensemble mean patterns are approximately ±5 mm/day over these areas. Under the SSP126 scenario, which is aligned with the 2°C warming target, the previously observed increases in precipitation exceeding ±9 mm/day for CCLM and ±5 mm/day for GCMs are no longer evident. In CA, areas such as Kyrgyzstan, Tajikistan, northern Pakistan, and southwestern Iran are particularly vulnerable to rainfall-induced hazards, including landslides \citep{wang2021atmospheric, kirschbaum2010global} and floods (e.g., the Pakistan floods of 2010 and 2022).

Figures \ref{fig:5}.1, d, and g illustrate the ECA-RX20mm values for CCLM at the end of the century across three scenarios. The observed patterns align with those in Figure \ref{fig:4}, underscoring an increase in the frequency of very heavy precipitation days, particularly marked over the Tibetan Plateau, as anthropogenic influences intensify. Similarly, Figures \ref{fig:5}.b, e, and h reveal that the CMIP6 GCM ensemble mirrors the behaviour observed in CCLM. However, the ensemble standard deviations for ECA-RX20mm values rise over Tajikistan and Kyrgyzstan, as shown in Figures \ref{fig:5}.c, f, and i. The growing frequency and intensity of extreme precipitation events over the elevated

regions of Central Asia, driven by anthropogenic factors, are a cause for concern \citep{fallah2023anthropogenic}. This CCLM simulation enhances our understanding of how dynamical downscaling's sensitivity to different levels of anthropogenic forcing can vary locally.

\section{CCLM emulator using a CNN}

We have demonstrated that dynamical downscaling adds significant value in capturing local climate change effects, particularly over areas influenced by complex topography. In this study, we create a CCLM emulator for precipitation over CA. As previously explained, a CNN trained on our GCM-RCM chain could serve as a fast, cost-effective downscaling method, though its efficacy needs to be rigorously assessed.

We aim to establish that this emulator outperforms simple interpolation, particularly in areas experiencing extreme precipitation. We aim to show that the CCLM emulator can replicate CCLM-like precipitation patterns when driven by the parent GCM.

Focusing on the CA domain, which encompasses the former Soviet Union countries (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan), we exclude the broader CORDEX-CA domain shown in Figure \ref{fig:1}. This domain is the region of interest in the Green Central Asia project \url{https://www.greencentralasia.org/en}, which the German Foreign Office finances. Figure \ref{fig:6}.a illustrates the MAE of the interpolated MPI-ESM1-2-HR, using the CCLM output as the 'true' precipitation. CCLM generates

distinct precipitation patterns, particularly in areas with complex topography. Assuming CCLM as the ground truth, we examine whether the CNN can replicate these outputs using the GCM as input. To assess the emulator's effectiveness, we present added value maps (relative to the parent GCM) in Figures \ref{fig:6}.b-d. A comparison of MAE reduction maps reveals that the unconstrained CNN demonstrates significant skill over elevated regions of CA, whereas constrained runs show less noticeable pattern changes. For instance, the HCL and SCL emulators generate closely mingled negative and positive added values across elevated areas, while NoCL consistently exhibits positive values across the domain. Several artefacts in the MAE reduction maps of constrained models, particularly over northern India, reflect the shape of the GCM grid. We also produce boxplots of daily precipitation for the CA domain to explore distribution improvements (Figure \ref{fig:7}). Correlation coefficients between time-series averages of precipitation[17] across the domain and CCLM are presented in Figure \ref{fig:7} (values in parentheses). Among daily averages, NoCL achieves the best performance (highest correlation coefficient), although it records fewer outliers than CCLM and other model simulations. The distribution is concentrated around the median, exhibiting the narrowest interquartile range. The distribution profiles of both constrained models (HCL, SCL) resemble those of the interpolated GCM, expected since the constraints maintain mass consistency within corresponding grid boxes (Equation \ref{eq:1}).

\subsection{Applying the CNN to a different GCM}

We evaluate the emulator's generalisation ability, i.e., its capacity to generate reliable predictions on new datasets. We conduct a new 15-year dynamical

simulation using CCLM, driven by the EC-Earth3-Veg \citep{doscher2022ec} GCM under the SSP370 scenario from 2019 to 2033. This data serves as input to our CCLM emulator, which was previously trained to emulate CCLM outputs using MPI-ESMI-2 HR. We now use the emulator to reconstruct the local features of CCLM driven by EC-Earth3-Veg. Figure \ref{fig:8}.a presents the MAE of the interpolated EC-Earth3-Veg with respect to the dynamical downscaling with CCLM. Remarkably, the MAE pattern of EC-Earth3-Veg closely mirrors that of MPI-ESM1-2-HR (Figure \ref{fig:6}.a). However, the NoCL emulator does not uniformly show positive error reduction across the domain (Figure \ref{fig:8}.b). We chose NoCL for its superior performance among the three CNNs. The emulator attempts to establish relationships between MPI-ESM1-2-HR and CCLM, which may be specific to these models and might not necessarily apply to the new EC-Earth3-Veg and CCLM configuration. As demonstrated previously, the RCM state depends on the state of its driving GCM. CCLM is driven at the lateral boundaries by the GCM values for state variables (temperature, pressure, wind speed, etc.) and not by precipitation, which is the CNN's input. The precipitation inputs from the two GCMs carry different biases, complicating the transfer of mapping from MPI-ESM1-2-HR-driven CCLM outputs to those driven by EC-Earth3-Veg.

Despite these challenges, the CNN model demonstrates added values exceeding 1 mm/day in regions such as the Alborz Mountains and the southern Caspian Sea in northern Iran (highlighted in black rectangles in Figures \ref{fig:8}.a and b) and parts of Tajikistan and Kyrgyzstan. Exploration of the daily precipitation distribution field-mean indicates that the CNN's median value and outliers are lower than those of the EC-Earth3-Veg and CCLM simulations (Figure \ref{fig:8}.c). The day-to-day correlation has improved,

although all models were trained on a shuffled dataset that ignored the memory in the time series. The trained NoCL model was provided with unshuffled EC-EARTH3-Veg data for new predictions, increasing the correlation coefficient from 0.815 (EC-Earth3-Veg) to 0.844 (NoCL). Over the highlighted area in Figure \ref{fig:8}.b, where the NoCL model reduces MAE, the distribution of precipitation converges towards that of CCLM, encompassing the region with the highest rainfall in Iran, vital for a large portion of the population, including Tehran. Only the outliers larger than 20 mm/day are not reconstructed by NoCL.

As a further test of generalisation, we intentionally excluded the SSP370 scenario from the training process. This allowed us to apply the model to a specific simulation and assess its ability to handle an unknown forcing. Figure \ref{fig:81} demonstrates the AV of the CNN emulator for SSP370 in comparison to the dynamical downscaling with CCLM, revealing that the AV pattern is strikingly similar to that shown in Figure \ref{fig:6}.d. This confirms that the CNN can learn and reproduce patterns under different forcing scenarios it was not explicitly trained on, as demonstrated by its performance with the SSP370 scenario.

\section{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 often struggles to account for the dynamic influences of complex landscapes, including topography and varying surface parameters such as vegetation, soil types, and water bodies like lakes, which may affect the

accuracy of statistical relationships \citep{li2022spatial}. For statistical downscaling methods applied to precipitation, observations need to contain detailed information about precipitation distribution in areas with complex topography \citep{lundquist2019our}.

Conversely, dynamical downscaling requires massive computational time and data storage. For example, a 30-year CCLM simulation driven by ERAInterim took roughly one week to complete 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, as they are highly dependent on the driving GCMs.

In this study, we contributed to the dynamic downscaling efforts over the CORDEX-CA domain, taking a small step towards creating an RCM ensemble for CA. A single RCM simulation helps identify model biases and uncertainties that need to be addressed in future model improvements. It is essential to note that a single model run for CMIP6, instead of an RCM ensemble, may not provide a comprehensive understanding of potential climate change impacts on a region. Therefore, it is recommended that researchers conduct multiple simulations with different initial and boundary conditions and model configurations to account for the uncertainty associated with climate projections.

In the first part of the study, we demonstrated the added value of RCMs (using the CCLM model) over GCMs for CA in representing precipitation. Our CCLM run showed 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 reproduced extreme precipitation patterns similar to

the CMIP6 ensemble mean projections for the end of the century. Both the CCLM and CMIP6 ensembles indicated an increased risk (in terms of intensity and frequency) of heavy precipitation events in vulnerable regions of CA due to various human activities.

Our study evaluated the downscaling skill using high-resolution observations, a crucial step for accurately capturing localised climate phenomena. This evaluation was essential before further study steps and regional adaptation strategies could be implemented. However, as \cite{volosciuk2017combined} noted, examining downscaling outputs at coarser resolutions can be equally informative. Their work emphasises 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.

We showed that a single GCM-RCM model chain could be used to train a climate emulator based on a CNN model. It learned nonlinear and physical relationships between the coarse and fine-resolution datasets, addressing the issue of spatial intermittency—where data points are unevenly distributed or missing across space—common in some statistical downscaling approaches \citep{harder2023hard}. However, we also demonstrated that the CNN model had limitations when generalising, as it did not achieve a robust error-reduction pattern when given a different GCM as input. The learning process strongly depended on the GCM/CCLM relationships. More importantly, an RCM was

forced to follow its driving GCM and could only produce extra information on a local scale. The presented CNN could be applied to other experiments of the same GCM, such as using the trained emulator for paleo-climate experiments or downscaling volcanic forcing experiments. This would aid the paleo-climate community in conducting proxy-model comparisons at local scales. However, previous studies have shown that the CNN suffered from the same generalisation problem when applied to a new GCM, and such applications must be tested \citep{jouvet2023ice}.

We deliberately excluded the SSP370 scenario from the training dataset to evaluate the model's generalisation capabilities for other scenarios of the same GCM. This strategy allowed us to assess whether the model could effectively infer and replicate patterns from untrained scenarios. Remarkably, the model's output for the SSP370 scenario exhibited an AV pattern mirrored the dynamical downscaling results of the CCLM driven by the same SSP370 scenario. This alignment strongly supported the notion that our CNN emulator could learn from its training data and generalise to new, unseen conditions. The similarity in AV patterns between the model output and the CCLM simulation underscored the robustness and adaptability of our model, affirming its potential for broader applications in climate modelling.

This work was an initial step in demonstrating the potential of such a hybrid approach. We encourage the community to explore different model structures and parameter combinations for further improvement. For example, our initial setups showed that a physically constrained CNN setup that applies a linear transformation to ensure mass or energy conservation between the low and high-resolution images did not successfully downscale precipitation. The

original dataset might not satisfy the constraints, leading to suboptimal results. In contrast, with a higher degree of freedom, the unconstrained CNN produced patterns closer to the target RCM. Future studies could test alternative machine learning models, such as generative adversarial networks (GANs), which can generate more high-frequency patterns and improve the downscaled output. Additionally, incorporating more information into the CNN by adding characteristics like surface height, vegetation, land cover, and land use as new channels within the input layer could enhance model performance.

| | | | |
|---|---|---|---|
| 1. | *is then extended* | Passive voice misuse | Clarity |
| 2. | *This* | Intricate text | Clarity |
| 3. | ~~with respect to~~ → concerning, for, to | Wordy sentences | Clarity |
| 4. | *are pronounced* | Passive voice misuse | Clarity |
| 5. | *are sparsely distributed* | Passive voice misuse | Clarity |
| 6. | *are calculated* | Passive voice misuse | Clarity |
| 7. | *is depicted* | Passive voice misuse | Clarity |
| 8. | *are found* | Passive voice misuse | Clarity |
| 9. | *are displayed* | Passive voice misuse | Clarity |
| 10. | ~~different~~ | Wordy sentences | Clarity |
| 11. | ~~a total of~~ | Wordy sentences | Clarity |
| 12. | *is depicted* | Passive voice misuse | Clarity |
| 13. | *is detected* | Passive voice misuse | Clarity |
| 14. | *Similar patterns are observed in the CMIP6 ensemble mean, but due to averaging, the ensemble mean patterns are approximately ±5 mm/day over these areas.* | Unclear sentences | Clarity |
| 15. | *is aligned* | Passive voice misuse | Clarity |
| 16. | *Figures \ref{fig:5}.1, d, and g illustrate the ECA-RX20mm values for CCLM at the end of the century across three scenarios.* | Unclear sentences | Clarity |
| 17. | precipitation averages | Wordy sentences | Clarity |

| | | | |
|---|---|---|---|
| 18. | *was previously trained* | Passive voice misuse | Clarity |
| 19. | ~~with respect to~~ → concerning, for, to | Wordy sentences | Clarity |
| 20. | *is driven* | Passive voice misuse | Clarity |
| 21. | *etc.* | Inappropriate colloquialisms | Delivery |
| 22. | *were trained* | Passive voice misuse | Clarity |
| 23. | *are not reconstructed* | Passive voice misuse | Clarity |
| 24. | *the added value of dynamically downscaling a general circulation model (GCM) from the sixth phase of the Coupled Model* | GMDD - Climate Model Downscaling in Central Asia: A Dynamical and a Neural Network Approach https://gmd.copernicus.org/preprints/gmd-2023-227/ | Originality |
| 25. | *the CCLM (at 0.22° horizontal resolution) for the historical period of 1985-2014 and the projection period of 2019-2100 under three 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) gri...* | GMDD - Climate Model Downscaling in Central Asia: A Dynamical and a Neural Network Approach https://gmd.copernicus.org/preprints/gmd-2023-227/ | Originality |
| 26. | *future climate projections. We present high-resolution maps of heavy precipitation changes based on CCLM and compare them with* | GMDD - Climate Model Downscaling in Central Asia: A Dynamical and a Neural Network Approach https://gmd.copernicus.org/preprints/gmd-2023-227/ | Originality |
| 27. | *the intensity and frequency of heavy precipitation events over* | Heavy Rainfal Relations Over Chicago And Northeastern Illinois | Originality |
| 28. | *Finally, we train a convolutional neural network (CNN* | 3D Visual Object Detection from Monocular Images | Originality |

| | | | |
|---|---|---|---|
| 29. | *the Intergovernmental Panel on Climate Change's Sixth Assessment Report* | Norway : The climate negotiations Norway's main lines at the negotiation meeting in Bonn (SB 58) | Originality |
| 30. | *Sections 3 and 4 introduce the results of* | Modeling and optimization of the insecticidal effects of Teucrium polium L. essential oil against red flour beetle (Tribolium castaneum Herbst) using response surface methodology | Originality |
| 31. | *x}{\frac{1}{n} \sum_{i=1}^{n* | A Study on Decision-Making Opinion Exploration in Windows-Based Information Security Monitoring Tool Development | Originality |
| 32. | *a batch size of 64 and a learning rate of 0.001* | Development and Verification of Time-Series Deep Learning for Drug-Induced Liver Injury Detection in Patients Taking Angiotensin II Receptor Blockers: A Multicenter Distributed Research Network Approach https://e-hir.org/journal/view.php?number=1170 | Originality |
| 33. | *to assess the impacts of heavy precipitation events on* | Water \| Free Full-Text \| Small Catchment Runoff Sensitivity to Station Density and Spatial Interpolation: Hydrological Modeling of Heavy Rainfall Using a Dense Rain Gauge Network https://www.mdpi.com/2073-4441/13/10/1381?utm_campaign=releaseissue_waterutm_medium=emailutm_source=releaseissueutm_term=doilink33 | Originality |
| 34. | *machine learning models, such as generative adversarial networks (GANs), which can* | Public Perception of AI-Generated NSFW Imagery - cyprusvipcard https://www.cyprusvipcard.com/post/1900 | Originality |