https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions
Open Access

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], 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]Competence Center High-Performance Computing, Fraunhofer ITWM, Germany
[4]Eberswalde University for Sustainable Development (HNEE), Germany

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, we evaluate the ~~CCLM~~ performance over the historical period ~~using a simulation driven by ERAInterim reanalysis~~. CCLM's added value, compared to its driving GCM, is significant over CA mountainous areas, which are at higher risk of extreme precipitation events. Furthermore, we downscale the CCLM for 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~~ dynamically downscaled ~~CCLM ones.~~ We show that ~~applied~~ CNN could emulate the GCM-CCLM model chain over large CA areas. However, this specific emulator has shortcomings when applied to a new GCM-CCLM model chain. Our downscaling data and the pre-trained CNN model could be used by scientific communities interested in downscaling CMIP6 models and searching for a trade-off between the dynamical and statistical methods.

Summary of Comments on gmd-2023-227.pdf

Page: 1

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:04

Overall, you wrote a good abstract. It generally reads well and gives a good overview, but it needs some clarifications and minor changes.

Number: 2 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:04

"as reference"

Number: 3 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:04

Number: 4 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:04

Number: 5 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:04

CCLM driven by ERAInterim reanalysis

Number: 6 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:04

Needs to be quantified; how much-added value in mm/day? Or another kind of value.

Number: 7 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:04

This makes it sound like you downscaled the RCM, which is incorrect from what I understood. You looked at CCLM predictions for future projections.

Number: 8 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:04

Unclear: is "the single model high-resolution dynamical downscaling" another name for the CNN? If yes, it's just very long and confusing. I suggest you remove this and write instead: "Finally, we train a convolutional neural network [...]".

Number: 9 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:04

a

Number: 10 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:04

Number: 11 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:04

its

Number: 12 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:04

Number: 13 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:04

"the" or "the applied"

Number: 14 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:04

RCM (CCLM)

Number: 15 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:04

I find this very vague and not completely true. Yes, you've created new downscaled data, which is great and can certainly be used. But you've shown that your CNN doesn't perform well when applied to a new GCM, so unfortunately, your trained CNN might not be very useful for another GCM (which is normal and to be expected). I suggest you separate this into two sentences: one about your new RCM products over CA, which are great, and one that says something like your CNN architecture could be used to downscale other GCMs (not the trained CNN).

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions
Open Access

1 Introduction

It is very well acknowledged that the global mean temperature is increasing due to anthropogenic greenhouse gas emissions (Allan et al., 2021). The most critical challenge for society is to assess and predict the future impact of this warming on the human health, natural ecosystems, and economy for different regions of the World. Studies of vulnerability, impacts and adaptation at the regional scale require reliable high-resolution climate projections (Maraun et al., 2015), which are based on dynamical downscaling through 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).

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 one of the most vulnerable regions to climate change impacts. CA's water resources depend on water from glaciers and rivers that are shrinking due to rising temperatures and decreasing precipitation (Reyer et al., 2017). Food security is at severe risk in CA with reduction of crop yields due to climate change (Allan et al., 2021). Extreme events like floods and landslides are happening more frequently and intensively in the region leading to severe damage to infrastructures, livelihoods and crops, subsequently causing population displacement and migration (Reyer et al., 2017).

Given the above-mentioned concerns, ~~the~~ impact modelling is still hindered in CA, based on the lack of high-resolution climate projections but also on the elevated level of uncertainty 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 CMIP6 project. In some cases, for properly reproducing 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 the orography of the region, the large-scale atmospheric circulation, the sea surface temperature anomalies in Indian Oceans 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 the resolution of a driving Global Circulation Model (GCM) and produce a robust and physically consistent regional state of the climate. This is often considered a critical point for preferring the use of RCMs to 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, RCMs are computationally demanding, especially at the very high resolutions useful for impact studies. At the same time, they suffer from 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

Page: 2

Number: 1 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

These two sentences feel out of place. It's the first (and also never mentioned outside of the introduction) time you use convective precipitation, but it needs to be clarified if that's what the model will actually emulate. Furthermore, after these two sentences, you continue with the pros and cons of statistical and dynamical downscaling, which is an entirely different subject. I suggest you remove or place these two sentences elsewhere.

Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Acronym used before being defined.

Number: 4 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This paragraph lists the pros and cons of stat and dyn downscaling but feels scattered as they are mixed. You're doing:
- pros dynamical downscaling
- what's statistical downscaling
- cons statistical downscaling
- cons dynamical downscaling

which doesn't make a lot of sense.
I suggest you do something like this (switch it if you want, but keep the methods together):
- pros dynamical downscaling
- cons dyn downscaling
- pros statistical downscaling
- cons statistical downscaling

Also, you write that RCMs are computationally demanding without giving any quantitative backup (how much computational time are we talking about here?)

Number: 5 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This abbreviation should have been defined before and not here, as it's already been used earlier in the text. Define it the first time GCM is used.

Number: 6 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

not clear if you mean comparability between stat and dynamical downscaling or within different RCMs.

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions
Open Access
EGU

climate projections using RCMs, over the years, members from different international institutions have joined forces ~~into~~ 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 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 focuses on highly industrialized countries (Allan et al., 2021), and ~~fewer~~ institutes run RCM simulations over CA (refer to https://esgf.llnl.gov/). 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 simulation (except this study) driven by the CMIP6 model simulations has been planned so far for CA (see https://wcrp-cordex.github.io/simulation-status/CMIP6_downscaling_plans.html, last visited on 14.08.2023). The 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

Number: 1 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
"in"

Number: 2 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
,

Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
This sentence was plagiarised from the cordex website: https://rcmes.jpl.nasa.gov/content/cordex

Number: 4 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 5 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 6 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
unnecessary acronym, used only once, remove it.

Number: 7 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
This sentence also feels out of place. No connection.

Number: 8 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
"few"

Number: 9 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 10 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
This feels out of place in this paragraph. It repeats the pros of RCMs mentioned in lines 44-46. I suggest you move them to the beginning of the next paragraph, as it follows that text (keep one idea per paragraph).

Number: 11 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
This was also already mentioned in line 50 - 51. Repeated information.

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions

value in simulating mesoscale convective precipitation, coastal rainfall, and extreme rainfall events (Giorgi and Gutowski Jr, 2015; Russo et al., 2020, 2019; Feser et al., 2011).

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 features from spatial data and capture non-linear mappings 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 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 network architecture (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 ~~feeds~~ 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 unconstrained and constrained CNN approaches.

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 on the 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 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.

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Unnecessary acronym, used only twice.

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 3 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

receives

Number: 4 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This makes it sound as if this paragraph proposes a new setup than the previous paragraph. But from my understanding, this hybrid framework is the same CNN framework as the one discussed in the last paragraph. Unclear.

Number: 5 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

high? there's an adjective missing here.

Number: 6 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

It's not really a starting point because it has already been done for temperature: https://link.springer.com/article/10.1007/s00382-022-06343-9 and mass balance: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022MS003593. I think this should be mentioned (and possibly cited). It looks like it's also been done for precipitation (https://hess.copernicus.org/preprints/hess-2023-55/, https://link.springer.com/article/10.1007/s12145-023-00970-4, etc.). However, your starting point might be on CA so if that's the case, then you should mention it that way.

Number: 7 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

No, that's not true; it's very difficult to generalize a model to new testing data, even if it's the same GCM/RCM chain. It would need to be tested, but it's a big claim to say that it would work on other ssps and other time-frames and cannot be stipulated as such.

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

2 Data and methods

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) RCM. COSMO-CLM 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, 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 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, 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). 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). 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

We create an emulator of CCLM using CNN. We use the output of the COSMO-CLM Version 6.0 RCM, which is driven by the MPI-ESM1-2-HR GCM under four different scenarios: historical, SSP126, SSP370 and SSP585. Historical is based on the data of greenhouse gas levels, land use, and other climate forcings from 1950 to 2014 that were observed. SSP126

Page: 5

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Overall this section needs restructuring and rewriting. There are inconsistencies and quite a few pieces of information need to be included or clarified. You might understand your setup very well, but it's hard to follow exactly what you've done for a new reader and this creates a problem of transparency.

Number: 2 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

this reference feels at the wrong place, should it not be after "numerical weather prediction model"?

Number: 4 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 5 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 6 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 7 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Unclear if setting up the simulations according to CORDEX is the optimal setup provided by Russo et al. or if you're talking about two different things.

Number: 8 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

here you should say which SSPs.

Number: 9 Author: anonymous Subject: Pencil Date: 20.12.2023, 11:28:28

Number: 10 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

The CNN section particularly needs considerable rewriting and more information about your choice of the framework (i.e., what's the perfect and imperfect framework and why did you choose the imperfect over perfect one), the selection of training/validation/testing data (and why), and the different models (HCL, No-CL, SCL) you created (see annotated pdf for more details). Things from the appendix need to be in the CNN section, and your different CNN setups (NoCL, etc.) must be clearly defined there. I think it's also missing a figure for the architecture, and readers would benefit from seeing it in your manuscript instead of having to look it up in another paper.

Number: 11 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This whole paragraph is too long and contains too many different ideas. Cut it at least into two.

Number: 12 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

If CCLM and COSMO-CLM refer to the same thing, they should not be used interchangeably, or it will confuse the reader. Choose one and apply it to the whole manuscript.

Number: 13 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This should already go above in the RCM section when different SSPs are mentioned. Or you should write in the RCM section "different SSPS (see Section 2.1.2)".

Number: 14 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

While the historical period is mentioned, we're missing the future time frame of the SSPs, that was last mentioned in the abstract and is needed here.

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

(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 . 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. We then train our CNN model based on the architecture proposed by Harder et al. (2022), 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 "imperfect model" set-up (Stengel et al., 2020; Leinonen et al., 2020), where the dynamical mapping from GCM to RCM is of higher interest. 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 have used a total number of 68141 (60%), 22714 (20%) and 22714 (20%) RCM simulation days for training, testing and evaluation, 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 (F) is 4 in this case. For a complete explanation of the CNN architecture, we refer to the work of Harder et al. (2022) and the corresponding GitHub repository at https://github.com/RolnickLab/constrained-downscaling (last visited on 21st of June 2023). Here, we briefly explain the architecture of the CNN used in this study:

- The input layer is a low-resolution (LR) image of size 30×60 with only one channel, i.e., precipitation value in mm/day.
- The first layer is a convolutional layer with 64 filters of size $3 \times 3 \times 1$ and stride 1. The output is a feature map of size $30 \times 60 \times 64$.
- The second layer is a sub-pixel convolutional layer with 256 filters of size $3 \times 3 \times 64$ and stride 1. The output is a feature map of size $60 \times 120 \times 64$.
- The third layer is another sub-pixel convolutional layer with 256 filters of size $3 \times 3 \times 64$ and stride 1. The output is a feature map of size $120 \times 240 \times 64$.

Page: 6

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

this feels very repetitive and weighs the text down. Maybe you could keep "SSP3-RCP7" after having written it in full once

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

I'm guessing you mean that you added the CCLM driven by ERAInterim and not ERAInterm in itself, which is what this sentence alludes too. Reformulate that so that it's clear that you added an RCM and not a GCM.

Number: 4 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This should be a new paragraph.

Number: 5 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This makes it seem as though your model was trained over the whole GCM/RCM domain and evaluated only over CA, which seems unrealistic. The input and target domain should be specified. I would strongly encourage you to add a figure like in these papers: https://link.springer.com/article/10.1007/s00382-022-06343-9 or https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022MS003593. And if that is what Figure 1 shows, then it should be referred to here already.

Number: 6 Author: anonymous Subject: Pencil Date: 20.12.2023, 11:28:28

Number: 7 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

If the reader has not read about this perfect and imperfect model framework before, they will not understand what you mean as your text is right now. You need to explain both frameworks better and why you chose to take the imperfect model framework.

Number: 8 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This should be cited from Doury et al 2021 who proposed this perfect model framework for downscaling a GCM with machine learning: https://link.springer.com/article/10.1007/s00382-022-06343-9

Number: 9 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Again, this feels out of place in this paragraph and has nothing to do with what was just said before. Consider adding this to the beginning of the next paragraph.

Number: 10 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

How were these days selected? Manually, randomly? Also, do they equally cover the three SSPs? That's essential information for the reader because whether you train and test the model on random data is very different from whether it is contiguous in time and will influence the performance a lot. Also, I would strongly advise you to use the words "training," "validation," and "testing" in that order instead of "training, testing, and evaluation" (also because later in the text, you interchange evaluation and testing; agree on one). In machine learning, we commonly use "testing" when testing the model on unseen data, and not "evaluation", which might confuse a reader. So you have 60% for training, 20% for validation (happens during training), and 20% for "testing" the trained model.

Number: 11 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 12 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

unnecessary

Number: 13 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

I would strongly advise you to add a Figure of the emulator architecture because it's very hard to follow the description without looking at it.

Number: 14 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

low-resolution was used before, without the acronym and this acronym is only once more. Either define it before and use it, or remove it.

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions

– The fourth layer is a convolutional layer with 1 filter of size $3 \times 3 \times 64$ and stride 1. The output is high-resolution (HR) image of size $120 \times 240 \times 1$.

– The fifth layer is an optional renormalization layer that applies a linear transformation to the HR image to ensure that the total mass or energy is conserved between the LR and HR images.

For this work, we find the unconstrained CNN (NoCL) performing the best, most likely due to the significant mismatch between low-resolution and high-resolution samples. A description of the constraint layers can be found in the appendix, see A.

We use the 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 MCL 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).

2.2 Evaluation Data

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 with respect to the driving model, 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~~ the CNN methods, instead of using CHIRPS, we use the corresponding CCLM simulation as our target and calculate the metrics on CNN and GCM outputs with respect to CCLM.

2.3 Metrics

In a first step, the selected GCM, RCM and observational data is 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 and errors that depend on the choice of interpolation method, the spatial distribution of the data points and the resolution ratio.

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
same comment as for LR.

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
call it the constraint layer or add it to this sentence otherwise it's confusing with the rest of the text.

Number: 3 Author: anonymous Subject: Pencil Date: 20.12.2023, 11:28:28

Number: 4 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
This should not be here but in the discussion or results.

Number: 5 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
This should be in the bullet point of that said layer above and not here.

Number: 6 Author: anonymous Subject: Sticky Note Date: 20.12.2023, 11:28:28
Here you need to insert this from your appendix, so that readers understand what the acronyms of the different models mean: "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)."

Number: 7 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
How were these hyperparameters selected? Are you doing any cross-validation?

Number: 8 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
These acronyms are only defined in the appendix, the reader has no idea which model you're talking about.

Number: 9 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
GCM

Number: 10 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
Not sure this is necessary information if you do not use it. I suggest you remove it.

Number: 11 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
"testing"

Number: 12 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
testing

Number: 13 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
(20% as mentioned above)

Number: 14 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
"interpolated GCM outputs"? Otherwise you couldn't compare them to the RCM.

Number: 15 Author: anonymous Subject: Pencil Date: 20.12.2023, 11:28:28

Number: 16 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
evaluation

Number: 17 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
This section should be combined with Evaluation data into one section called "Evaluation" or "Testing," however you want to call it. Then, the first paragraph does not feel out of place, and you can just add the evaluation metrics, too.

Number: 18 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
Evaluation

Number: 19 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
This paragraph is out of place in metrics. Maybe you could put it just above and start the paragraph with "For this" instead of "In a first step". Also, you finish this paragraph with how interpolation is a bad idea, so the reader asks, why do it then? I think you meant to say that the simple interpolation of the GCM on the RCM grid is not precise enough, and that's why we need dynamical/stat downscaling (and so you're using simple interpolation as a baseline?). But then you need to say it; otherwise, one wonders why you're doing it and why this paragraph is there at all.

Comments from page 7 continued on next page

– The fourth layer is a convolutional layer with 1 filter of size $3 \times 3 \times 64$ and stride 1. The output is high-resolution (HR) image of size $120 \times 240 \times 1$.

– The fifth layer is an optional renormalization layer that applies a linear transformation to the HR image to ensure that the total mass or energy is conserved between the LR and HR images.

For this work, we find the unconstrained CNN (NoCL) performing the best, most likely due to the significant mismatch between low-resolution and high-resolution samples. A description of the constraint layers can be found in the appendix, see A.

We use the 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).

2.2 Evaluation Data

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 with respect to the driving model, 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~~ the CNN methods, instead of using CHIRPS, we use the corresponding CCLM simulation as our target and calculate the metrics on CNN and GCM outputs with respect to CCLM.

2.3 Metrics

In a first step, the selected GCM, RCM and observational data is 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 and errors that depend on the choice of interpolation method, the spatial distribution of the data points and the resolution ratio.

Number: 20 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 21 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 22 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

(

Number: 23 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

)

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions

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

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

where N 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.,~~

$$\mathrm{AV} = \mathrm{MAE}_{GCM} - \mathrm{MAE}_{CCLM} \tag{2}$$

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

3 Results

Figure 1.a shows the topography of the CORDEX-CA simulation domain. Figure 1.b presents the annual climatology (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 1.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. 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 model in time and space, ~~we show~~ the maps of yearly, winter (DJF), and summer (JJA) mean ~~biases~~ of precipitation ~~between~~ ERAInterim and CCLM, 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.~~ Figures 2.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~~ ($\mathrm{MAE}_{\mathrm{ERAInterim,CHIRPS}} - \mathrm{MAE}_{\mathrm{CCLM,CHIRPS}}$) ~~or the added values~~ are shown in Figures 2.d-f. CCLM ~~bias~~ is high~~er~~ during the Asian summer monsoon, over the South and Southeast of the domain. During winter, the ~~bias~~ is generally lower. CCLM presents a ~~bias~~ reduction for prominent locations within the domain and an increase of ~~bias~~ 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 2.g-l respectively. The CHIRPS dataset is again used as the observational dataset O ~~to calculate MAE and AV according to equations 1 and 2~~. The ~~AV~~ is the most pronounced over areas with complex topography, for all three considered RCMS (Figs.2.d-l). Areas

Page: 8

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

I would use "emulated precipitation" instead of simulations, as you commonly talk about emulators in your study, and otherwise, one might not be sure what you refer to with simulations.

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

this makes it seem as though N is equal to T, then just use the one or the other.

Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

See here you define a new N, so you should definitely remove the N from section 2.1.2 for consistency.

Number: 4 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

:

Number: 5 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

I don't think bias is the right word here since it is a statistical result that measures a systemic distortion and is generally used in a phrase as "bias towards...". You should use big or small MAE or "difference". Otherwise, it's hard to interpret your results.

Number: 6 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Is this the MAE of the interpolated GCM on RCM grid compared to the reference dataset ? If yes, this should be made more clear.

Number: 7 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

If it's annual climatology of daily precipitation, do you mean "mean daily precipitation averaged over all years"? If yes, add this. If not, explain it in more detail.

Number: 8 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

I think this is a grammar mistake. It's already very sparse, right? Not that it "could be sparser". Or do you mean "could be denser"?

Number: 9 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 10 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Figure 2 shows

Number: 11 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

CCLM

Number: 12 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

MAE

Number: 13 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

by

Number: 14 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

daily? annual?

Number: 15 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

interpolated?

Number: 16 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

driven by ERAInterim

Number: 17 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

this repeats the sentence above once my suggestions have been incorporated.

Number: 18 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

(Eq 1)

Number: 19 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

you've defined the added value in metrics, use it here :)

Number: 20 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

interpolated

Comments from page 8 continued on next page

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

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

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

where N 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.,~~

$$\mathrm{AV} = \mathrm{MAE}_{GCM} - \mathrm{MAE}_{CCLM} \tag{2}$$

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

3 Results

Figure 1.a shows the topography of the CORDEX-CA simulation domain. Figure 1.b presents the annual climatology (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 1.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. 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 model in time and space, ~~we show~~ the maps of yearly, winter (DJF), and summer (JJA) mean ~~biases~~ of precipitation ~~between~~ ERAInterim and CCLM, 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.~~ Figures 2.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~~ ($\mathrm{MAE}_{\mathrm{ERAInterim,CHIRPS}} - \mathrm{MAE}_{\mathrm{CCLM,CHIRPS}}$) ~~or the added values~~ are shown in Figures 2.d-f. CCLM ~~bias~~ is higher during the Asian summer monsoon, over the South and Southeast of the domain. During winter, the ~~bias~~ is generally lower. CCLM presents a ~~bias~~ reduction for prominent locations within the domain and an increase of ~~bias~~ 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 2.g-l respectively. The CHIRPS dataset is again used as the observational dataset O ~~to calculate MAE and AV according to equations 1 and 2~~. The ~~AV~~ is the most pronounced over areas with complex topography, for all three considered RCMS (Figs.2.d-l). Areas

Number: 21 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
added value of the CCLM RCM compared to the interpolated ERAInterim GCM (Eq 2)

Number: 22 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 23 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 24 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
MAE

Number: 25 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
"higher" would refer as a comparison to something else. Here you just mean "high".

Number: 26 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
(regions in magenta)

Number: 27 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
MAE

Number: 28 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
MAE

Number: 29 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
MAE

Number: 30 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
Could you also add which regions (in green) it seems to fit better than the interpolated GCM? "Prominent locations" is vague and the reader is not sure what it refers to.

Number: 31 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 32 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
either use the acronym or not but not a mix of both otherwise it's confusing. I think that "added value" is small enough not to need an acronym in the text.

Number: 33 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 34 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
of RCMs

Number: 35 Author: anonymous Subject: Pencil Date: 20.12.2023, 11:28:28

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions

Where the downscaling ~~reduces the bias of 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 2.d,g and j). Considering the whole domain, all three RCMs sensibly reduce the large and local-scale bias of ERAInterim against CHIRPS ~~(Figure 2),~~ especially for complex topographies. The nested RCMs show similar values of MAE near their lateral boundaries, with respect to their driving model (Figure ~~2~~, panels *a*,*b*,*c*). Therefore, negative AV quantities might originate from the boundary effect, especially near the east and southeastern boundaries, where the monsoonal precipitation is dominant.

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~~ 3.a that the MPI-ESM1-2-HR shows ~~less bias~~ than the ERAInterim over Tajikistan and Kyrgyzstan. According to Déqué et al. (2007), ~~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 3.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).

3.1.2 Extreme precipitation patterns in CCLM and CMIP6 GCMs

~~We explore climate change signals in the high resolution output, given that the CCLM simulation has shown some added value for precipitation over mountainous areas of CA.~~ 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~~ 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 https://www.pik-potsdam.de/~fallah/papers/model_lists.pdf for the list of

Page: 9

Number: 1 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

(RCM) has a smaller MAE than the

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

But from your figure, this looks strongest in summer, not in annual and winter. This seems to be missing from this sentence.

Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Refer to last comment.

Number: 4 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 5 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

In these figures, adding the spatial means and standard derivation of AV (or MAE of both models) would also be valuable. Because right now, you qualitatively claim that "all three RCMs sensible reduce the MAE compared to the GCM" (I guess by looking by eye at the figures) but without any quantitative backup. Spatial means could (hopefully) show that the overall AV is positive.

Number: 6 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

before you never wrote panels, use the same notations for all references of figures.

Number: 7 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

What about negative AVs for annual and winter GERICS above Tibet? These seem to be strongly pronounced but not explained here. Would that still be considered the boundary?

Number: 8 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

improve local-scale predictions of daily precipitation compared to its driving reanalysis

Number: 9 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 10 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

here as I aforementioned, you should stick to either CCLM or COSMO-CLM everywhere in the text, otherwise it generates confusion. And it should read "CCLM driven by ERAInterim" not CCLM alone, as that is what you have shown.

Number: 11 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

RCM

Number: 12 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

is that a reference to GERICS-REMO2015 and RMIB-UGent-ALARO-0 or something else? It should be specified.

Number: 13 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

s

Number: 14 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

In Figure 3, we show

Number: 15 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Figure

Number: 16 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

smaller MAEs

Number: 17 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

interpolated

Number: 18 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 19 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

However, that is not what happens, so you need to make this more clear.

Number: 20 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

However,

Comments from page 9 continued on next page

where the downscaling ~~reduces the bias of 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 2.d,g and j). Considering the whole domain, all three RCMs sensibly reduce the large and local-scale bias of ERAInterim against CHIRPS ~~(Figure 2),~~ especially for complex topographies. The nested RCMs show similar values of MAE near their lateral boundaries, with respect to their driving model (Figure 2, panels *a*,*b*,*c*). Therefore, negative AV quantities might originate from the boundary effect, especially near the east and southeastern boundaries, where the monsoonal precipitation is dominant.

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~~ 3.a that the MPI-ESM1-2-HR shows ~~less bias~~ than the ERAInterim over Tajikistan and Kyrgyzstan. According to Déqué et al. (2007), ~~the~~ GCM bias is one of the most important sources of uncertainty in the RCM's regional climate projection, and the smaller $MAE_{\text{MPI-ESM1-2-HR}}$ compared to $MAE_{\text{ERAInterim}}$ 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 3.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).

3.1.2 Extreme precipitation patterns in CCLM and CMIP6 GCMs

~~We explore climate change signals in the high resolution output, given that the CCLM simulation has shown some added value for precipitation over mountainous areas of CA.~~ 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~~ 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 https://www.pik-potsdam.de/~fallah/papers/model_lists.pdf for the list of

Number: 21 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
Tajikistan and Kyrgyzstan

Number: 22 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
CCLM

Number: 23 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
New paragraph

Number: 24 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
that

Number: 25 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 26 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
The formulation of this sentence is confusing with the opposition; I would suggest you write it like this: "For example, in all regions with high values of yearly precipitation, where GCM has a small MAE, the RCM also has a small MAE and vice versa."

Number: 27 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
"small MAE" and give examples of regions

Number: 28 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
MAE

Number: 29 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
the

Number: 30 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
Added value

Number: 31 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
I'm not sure what the purpose of this sentence is, and it feels out of place with the rest of the text. Maybe think about which message you want to convey.

Number: 32 Author: anonymous Subject: Sticky Note Date: 20.12.2023, 11:28:28
What I find confusing is that you said that the hypothesis was that if the GCM has a small MAE (which is the case here that MPI is better than ERAInterim), hopefully the RCM might have it too. But that is not what happens in this case right? Seeing how CCLM(ERAInterim) performs better over Tajikistan and Kyrgyzstan than CCLM(MPI) - CCLM(MPI) has a smaller added value than CCLM(ERAInterim). So how do you explain this?

Number: 33 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
Grammar: "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."

Number: 34 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
can

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions
Open Access
EGU

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 4 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 4.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 ± 5 mm/day over those areas. Under the SSP126 scenario, which agrees with the 2°C target, the increasing patterns of more than ± 9 mm/day for CCLM and ± 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 5.1,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 4, indicating that the number (frequency) of very heavy precipitation days also increases with an enhanced anthropogenic influence, particularly over the Tibetan Plateau. From Figures 5.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 5.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

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 1 ~~COSMO-CLM~~ for

Number: 1 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

CCLM

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

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~~ than a simple interpolation, especially for areas receiving extreme precipitation values. More specifically, our goal is to show that the ~~COSMO CLM~~ emulator can produce ~~COSMO CLM~~-like patterns when fed by the parent GCM.

For the CNN approach, we focus on the CA domain ~~introduced as a domain~~ covering the former Soviet Union countries (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan) and not the CORDEX-CA domain previously shown in Figure 1. This domain is the region of interest in the Green Central Asia project, which is financed by the German Foreign office. Figure 6.a shows the MAE from the interpolated MPI-ESM1-2-HR with respect to the ~~COSMO CLM~~ from the test dataset, i.e., MAE(MPI-ESM1-2-HR, MPI-ESM1-2-HR-CCLM). As can be seen, ~~COSMO CLM~~ produces different precipitation values, 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~~ the emulator, we show the maps of ~~MAE reduction~~, i.e., $MAE_{GCM,CCLM} - MAE_{CNN,CCLM}$ in ~~figures~~ 6.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 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 in the MAE reduction maps of constrained models, especially over North of India. We produce the boxplots of daily precipitation over the newly-considered domain to explore the improvement in the distributions (Figure 7). The correlation coefficients between the time-series of average precipitation over the domain with respect to CCLM are presented in Figure 7 (values in the parentheses). For the daily averages NoCL presents the best performance (highest correlation coefficient). However, the values of outliers are ~~less~~ 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 is like the GCM one. This was expected, since the constraining conserves the mass of high-resolution grid-boxes within the corresponding low-resolution grid-box (Equation A1).

4.1 Applying CNN to a different GCM

~~Here,~~ we evaluate the emulator's generalization ability, i.e. the ability to create reliable predictions ~~of~~ a new data set. We conduct ~~here~~ a new 15-year dynamical simulation with ~~COSMO-CLM~~ 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~~ emulator, which was previously trained ~~on the MPI-ESM1-2-HR and its COSMO-CLM run~~. We now use the emulator to reconstruct the local features of COSMO-CLM driven by EC-Earth3-Veg. Figure 8.a presents the MAE of the EC-Earth3-Veg with respect to the ~~dynamically~~

Page: 11

Number: 1 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

can

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This sentence feels inappropriate here; it feels like you're already giving results, but you want to talk about CNN's potential. It should be removed or rewritten.

Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

where?

Number: 4 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

"very close to" (and give metric)

Number: 5 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

CCLM

Number: 6 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 7 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

overfitting

Number: 8 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

No, model complexity cannot be quantitatively compared to input complexity. Overfitting/underfitting is measured when comparing the model's performance on training and testing data, which you're not doing right now. So, you cannot talk about overfitting and underfitting here. I suggest you remove this.

Number: 9 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

CCLM

Number: 10 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 11 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

"given as input to"

Number: 12 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 13 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

is better at downscaling

Number: 14 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 15 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

CCLM

Number: 16 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

CCLM

Number: 17 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 18 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

only

Number: 19 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

missing reference

Number: 20 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

CCLM

Number: 21 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

In these four figures, it would again be very interesting to have the spatial mean values of MAE or AV (over the domain) to be able to compare the three models with other than by eye.

Number: 22 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

is there any way this notation could be made more readable? I think it's also somewhere else in the text but very

Comments from page 11 continued on next page

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development Discussions
Open Access

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~~ than a simple interpolation, especially for areas receiving extreme precipitation values. More specifically, our goal is to show that the ~~COSMO-CLM~~ emulator can produce ~~COSMO-CLM~~-like patterns when fed by the parent GCM.

For the CNN approach, we focus on the CA domain ~~introduced as a domain~~ covering the former Soviet Union countries (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan) and not the CORDEX-CA domain previously shown in Figure 1. This domain is the region of interest in the Green Central Asia project ?, which is financed by the German Foreign office. Figure 6.a shows the MAE from the interpolated MPI-ESM1-2-HR with respect to the ~~COSMO-CLM~~ from the 25 ~~test~~ dataset, i.e., MAE(26 MPI-ESM1-2-HR, MPI-ESM1-2-HR-CCLM). As can be seen, 23 ~~COSMO-CLM~~ produces 24 ~~different~~ precipitation values, especially over regions with 28 complex topography. 27 This has been noticed in the added value and down-scaling signal maps of COSMO-CLM. 29 ~~To explore a potential skill in~~ the emulator, we show the maps of 30 ~~MAE reduction~~, i.e., $\mathrm{MAE}_{\mathrm{GCM,CCLM}} - \mathrm{MAE}_{\mathrm{CNN,CCLM}}$ in 31 ~~figures~~ 6.b-d. Comparison of 32 ~~MAE reduction~~ maps shows that the unconstrained CNN 33 34 ~~produces significant skills~~ over elevated regions of CA 35 ~~and~~ the constrained runs 36 ~~do not present considrable patterns of changes~~. For example, there are areas of 37 negative and positive 38 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. 39 The fingerprint of the GCM is detectable in the MAE reduction maps of 40 constrained models, especially over North of India. We produce the boxplots of daily precipitation over 41 ~~the~~ newly-considered domain to explore the improvement in the distributions (Figure 7). The correlation coefficients between the time-series of average precipitation over the domain with respect to CCLM are 42 presented in Figure 7 (values in the parentheses). For the daily averages, 43 CL presents the best performance (highest correlation coefficient). However, the values of outliers are ~~less~~ 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 is like the GCM one. This was expected, since the constraining conserves the mass of high-resolution grid-boxes within the corresponding low-resolution grid-box (Equation A1).

4.1 Applying CNN to a different GCM

~~Here,~~ we evaluate the emulator's generalization ability, i.e. the ability to create reliable predictions ~~of~~ a new data set. We conduct ~~here~~ a new 15-year dynamical simulation with ~~COSMO-CLM~~ 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~~ emulator, which was previously trained ~~on the MPI-ESM1-2-HR and its COSMO-CLM run~~. We now use the emulator to reconstruct the local features of COSMO-CLM driven by EC-Earth3-Veg. Figure 8.a presents the MAE of the EC-Earth3-Veg with respect to the ~~dynamically~~

heavy. For example: CCLM(MPI) or CCLM_{MPI}

Number: 23 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

CCLM

Number: 24 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Different, ok, but are they better? In Figure 6, it would be helpful to add the map of Figure 3d) to show the added value of the RCM compared to the GCM over this new domain. I find the map of MAE(int GCM, RCM) hard to interpret without knowing whether they reproduce the reference data well.

Number: 25 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

here you use test but do you mean the evaluation dataset you mentioned before? Or is that another one. Refer to the section it was explained.

Number: 26 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

interpolated

Number: 27 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Which figures? Too vague.

Number: 28 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

such as? Give geographical indicators

Number: 29 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

To evaluate the performance of

Number: 30 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

added value

Number: 31 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Figures

Number: 32 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

added value

Number: 33 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

(NoCL)

Number: 34 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

that doesn't mean anything. I suggest "has high added value"

Number: 35 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

but

Number: 36 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

this is also vague/unclear, reformulate

Number: 37 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

But this pattern also appears when looking at the added value of the RCM compared to reference data (Fig 3d). It seems like it's just inverted. In Fig3d, where the RCM had an added value compared to the GCM on the reference data, the constrained CNN has no added value compared to the interpolated GCM, and vice versa (green and pink are inverted in the two images). Is that a coincidence?

Number: 38 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

added

Number: 39 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This is also vague, what do you qualitatively mean?

Number: 40 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

the

Number: 41 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

is that a new domain, or the CA domain covering the former soviet union mentioned above. Not clear

Number: 42 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

also

Number: 43 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

,

Comments from page 11 continued on next page

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

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~~ than a simple interpolation, especially for areas receiving extreme precipitation values. More specifically, our goal is to show that the ~~COSMO-CLM~~ emulator can produce ~~COSMO-CLM~~-like patterns when fed by the parent GCM.

For the CNN approach, we focus on the CA domain ~~introduced as a domain~~ covering the former Soviet Union countries (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan) and not the CORDEX-CA domain previously shown in Figure 1. This domain is the region of interest in the Green Central Asia project ?, which is financed by the German Foreign office. Figure 6.a shows the MAE from the interpolated MPI-ESM1-2-HR with respect to the ~~COSMO-CLM~~ from the test dataset, i.e., MAE(MPI-ESM1-2-HR, MPI-ESM1-2-HR-CCLM). As can be seen, ~~COSMO-CLM~~ produces different precipitation values, 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~~ the emulator, we show the maps of ~~MAE reduction~~, i.e., $\mathrm{MAE}_{\mathrm{GCM,CCLM}} - \mathrm{MAE}_{\mathrm{CNN,CCLM}}$ in ~~figures~~ 6.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 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 in the MAE reduction maps of constrained models, especially over North of India. We produce the boxplots of daily precipitation over the newly-considered domain to explore the improvement in the distributions (Figure 7). The correlation coefficients between the time-series of average precipitation over the domain with respect to CCLM are presented in Figure 7 (values in the parentheses). For the daily averages NoCL presents the best performance (highest correlation coefficient). However, the values of outliers are ~~less~~ 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 is like the GCM one. This was expected, since the constraining conserves the mass of high-resolution grid-boxes within the corresponding low-resolution grid-box (Equation A1).

4.1 Applying CNN to a different GCM

~~Here,~~ we evaluate the emulator's generalization ability, i.e. the ability to create reliable predictions ~~of~~ a new data set. We conduct ~~here~~ a new 15-year dynamical simulation with ~~COSMO-CLM~~ 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~~ emulator, which was previously trained ~~on the MPI-ESM1-2-HR and its COSMO-CLM run~~. We now use the emulator to reconstruct the local features of COSMO-CLM driven by EC-Earth3-Veg. Figure 8.a presents the MAE of the EC-Earth3-Veg with respect to the ~~dynamically~~

Number: 44 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
smaller

Number: 45 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
both

Number: 46 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
(HCL, SCL)

Number: 47 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
interpolated

Number: 48 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
the

Number: 49 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 50 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
on

Number: 51 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 52 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
CCLM

Number: 53 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
CCLM

Number: 54 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
"to emulate CCLM using MPI-ESMI-2 HR as input GCM."

Number: 55 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 56 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 57 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 58 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
interpolated

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions

~~downscaled simulation using~~ COSMO-CLM, ~~i.e., the COSMO CLM simulation driven by EC-Earth3-Veg.~~ The MAE pattern of EC-Earth3-Veg is remarkably like the one from MPI-ESM1-2-HR (Figure 6.a). However, the COSMO-CLM emulator based on the NoCL CNN model does not show positive error reduction everywhere in the domain (Figure 8.b). 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 emulator tries to find relations between the MPI-ESM1-2-HR and COSMO-CLM, 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 ~~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. 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 8.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 simulations (Figure 8.c). Only the day-to-day correlation is being improved. The model was 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 correlation coefficient increases using the NoCL model from 0.815 (EC-Earth3-Veg) to 0.844 (NoCL). Over the region where the NonCL model reduces the MAE, ~~i.e.,~~ the black rectangular box in Figure 8.b, the distribution of precipitation converges to the one from COSMO-CLM (Figure 8.d). Only the outliers larger than 20 mm/day are not reconstructed by the NoCL. ~~This region~~ 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.

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 computational time and data storage space. Additionally, the added value of RCMs is still debated, since 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

Page: 12

Number: 1 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

dynamical downscaling with

Number: 2 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

What would actually be interesting here, is not to see the MAE(Earth3-veg, CCLM) but the MAE(Earth3-veg, MPI-ESMI) to see how different the two GCMs are. Otherwise one has to skip several pages to compare with a figure before. I'd suggest you add it or change Fig 8a fort that.

Number: 4 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 5 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 6 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 7 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

you should explain before why you used only the not constrained CNN.

Number: 8 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

this does not mean anything. I'd say remove this sentence completely and just keep the next one.

Number: 9 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

chain

Number: 10 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 11 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 12 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 13 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 14 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 15 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 16 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

it looks like 1 mm/day is the maximum of the colorbar in the figure so how can it be more than that?

Number: 17 Author: anonymous Subject: Pencil Date: 20.12.2023, 11:28:28

Number: 18 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 19 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This also needs context... which model? NoCL trained on MPI-ESMI/CCLM? Also ignoring memory in time-series is not a concept that was mentioned before and it definitively should have been when presenting the training/validation/testing datasets in the methods section.

It's also unclear what you mean by "fed the original (without shuffling)". Do you mean that the trained NoCL model was given unshuffled EC-EARTH3-Veg to make new predictions? If yes, specify and why.

Number: 20 Author: anonymous Subject: Pencil Date: 20.12.2023, 11:28:28

Number: 21 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 22 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

Comments from page 12 continued on next page

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions

~~downscaled simulation using~~ COSMO-CLM, ~~i.e., the COSMO-CLM simulation driven by EC-Earth3-Veg.~~ The MAE pattern of EC-Earth3-Veg is remarkably like the one from MPI-ESM1-2-HR (Figure 6.a). However, the COSMO-CLM emulator based on the NoCL CNN model does not show positive error reduction everywhere in the domain (Figure 8.b). 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 emulator tries to find relations between the MPI-ESM1-2-HR and COSMO-CLM, 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 ~~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. 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 8.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 simulations (Figure 8.c). Only the day-to-day correlation is being improved. The model was 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 correlation coefficient increases using the NoCL model from 0.815 (EC-Earth3-Veg) to 0.844 (NoCL). Over the region where the NonCL model reduces the MAE, ~~i.e.,~~ the black rectangular box in Figure 8.b, the distribution of precipitation converges to the one from COSMO-CLM (Figure 8.d). Only the outliers larger than 20 mm/day are not reconstructed by the NoCL. ~~This region~~ 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.

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 computational time and data storage space. Additionally, the added value of RCMs is still debated, since 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 compared to the interpolated GCM

Number: 23 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

It's actually hard to be able to assess that seeing how squashed the boxplots are because of the many outliers. It needs to have a second panel where we can compare the distributions without outliers i.e., in the range of 0-5 mm/day

Number: 24 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 25 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

The region in the black box

Number: 26 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 27 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 28 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

statistical downscaling

Number: 29 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

I don't think you ever quantified the computational time and storage of your RCM in your paper. So, this is a big claim that needs evidence. Seeing how you ran an RCM, it might be an interesting addition because you say that the CNN took 15 hours, so readers could compare if it's actually faster. Otherwise, this is all very vague.

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development Discussions

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 over GCMs for CA in the representation of precipitation. Our ~~COSMO-CLM~~ run shows ~~AV~~ with respect to its driving GCM, comparable to the range of values obtained for other RCMs applied to 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~~ and CMIP6 ensemble present elevated risk (frequency and intensity) of heavy precipitation events over vulnerable areas of CA due to different anthropogenic influences.

Additionally, acknowledging the computational and memory constraints of a RCM ~~to be run~~ at ~~very~~ 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). However, we have also shown that the CNN model has limitations, as it did not achieve a robust error-reduction pattern when ~~applied to a different GCM-CCLM chain~~. 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~~ to apply it for other experiments of the same GCM: One ~~can~~ 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. 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 set-up 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 models, such as generative adversarial networks (GAN), which can generate more high-frequency patterns, might improve the downscaled pattern, and ~~should~~ be tested in future studies. An additional set-up might be to ~~add~~ more information to CNN by adding characteristics like surface height, vegetation, land-cover, land-use, etc. as new channels within the input layer.

Code and data availability. The main code for the CNN ~~could~~ be accessed via https://github.com/RolnickLab/constrained-downscaling. The training code and the corresponding data-sets used for this paper are addressed in the Jupyter notebook along with a simple test at https://github.com/bijanf/Climate_Model_Downscaling_GMD. RCM simulation output data ~~could~~ be provided upon request.

Number: 1 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
(here we chose to use the COSMO-CLM/CCLM model)

Number: 2 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
CCLM

Number: 3 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
"added value". You cannot use "added value" in the sentence above and then the acronym in the next sentence.

Number: 4 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
the

Number: 5 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
CCLM

Number: 6 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 7 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 8 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
running

Number: 9 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
I would not say this for sure, you don't really know what it learnt.

Number: 10 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
"namely when generalizing,"

Number: 11 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
given a different GCM as input.

Number: 12 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
Is that something you observed or the reason why it's hard to generalize?

Number: 13 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
isn't that what you already did? Because you ran it approximately from 1850 to 2014 right? I would consider that several years.

Number: 14 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
machine learning

Number: 15 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
's

Number: 16 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
could be

Number: 17 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
could

Number: 18 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
You've proven that it works for pretty recent climate data (1850s to now). But that does not mean it will work for climate data from 10'000 years ago. If that data is different from what you have trained the model on, then the CNN will suffer from the same generalisation problem as when applied to a new GCM. For example, models that were trained on paleo climate like Jouvet et al. (https://www.cambridge.org/core/journals/journal-of-glaciology/article/iceflow-model-emulator-based-on-physicsinformed-deep-learning/8C4D103C0F34DA690D9B524DF1461C5C) struggle to generalise to recent climate. You make a strong claim and it would need to be tested before. Otherwise you should just write that it needs to be tested.

Number: 19 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
could

Number: 20 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
same comment as above

Comments from page 13 continued on next page

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

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 over GCMs for CA in the representation of precipitation. Our ~~COSMO-CLM~~ run shows ~~AV~~ with respect to its driving GCM, comparable to the range of values obtained for other RCMs applied to 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~~ and CMIP6 ensemble present elevated risk (frequency and intensity) of heavy precipitation events over vulnerable areas of CA due to different anthropogenic influences.

Additionally, acknowledging the computational and memory constraints of an RCM ~~to be run~~ at ~~very~~ 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). However, we have also shown that the CNN model has limitations, as it did not achieve a robust error-reduction pattern when ~~applied to a different GCM-CCLM chain~~. 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~~ to apply it for other experiments of the same GCM: One ~~can~~ 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. 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 [21] ~~the~~ constrained model set-up 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 [22] ~~model~~ [23] ~~run~~ produced [24] ~~more realistic patterns~~. Alternative [25] models, such as generative adversarial networks (GAN), which can generate more high-frequency patterns, might improve the downscaled pattern, and [26] ~~should~~ be tested in future studies. An additional set-up might be to [27] ~~add~~ more information to [28] CNN by adding characteristics like surface height, vegetation, land-cover, land-use, etc. as new channels within the input layer.

Code and data availability. The main code for the CNN [29] ~~could~~ be accessed via https://github.com/RolnickLab/constrained-downscaling. The training code and the corresponding data-sets used for this paper are addressed in the Jupyter notebook along with a simple test at https://github.com/bijanf/Climate_Model_Downscaling_GMD. RCM simulation output data [30] ~~could~~ be provided upon request.

Number: 21 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
you should repeat in the discussion what this means to constrain the model.

Number: 22 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
CNN

Number: 23 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 24 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
patterns closer to the target RCM.

Number: 25 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
machine learning

Number: 26 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
could

Number: 27 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
provide

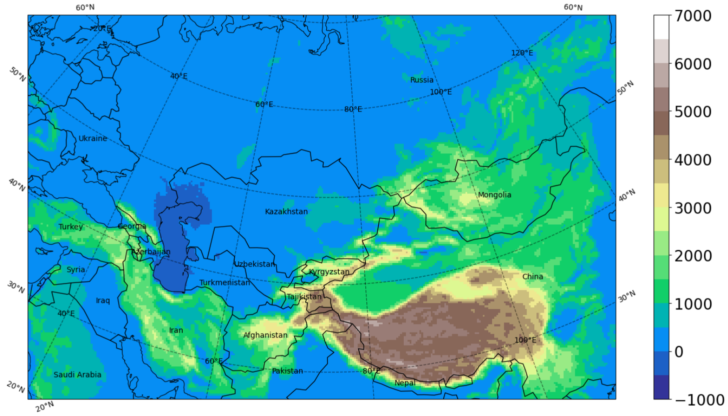
Number: 28 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
the

Number: 29 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
can

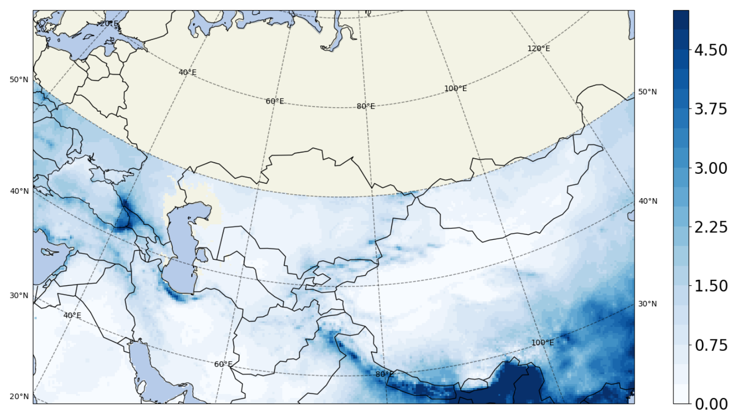
Number: 30 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
can

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

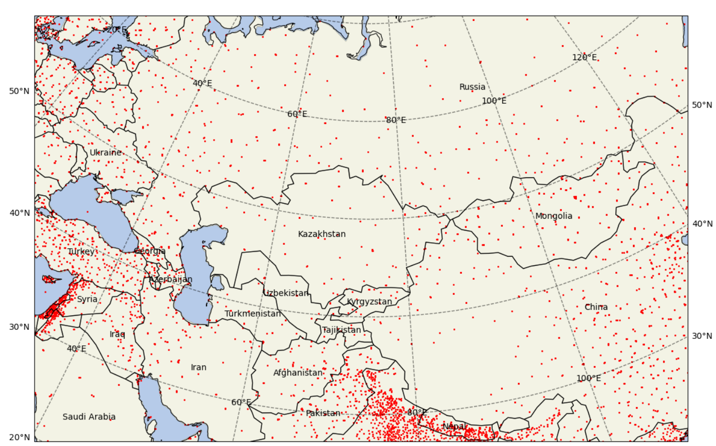
Geoscientific Model Development
Discussions



(a)



(b)



(c)

Figure 1. a) Study region over Central Asia and the topography (m), (b) CHIRPS climatology for 1985-2014 (mm/day), and (c) WorldClim's weather stations (red dots).

Page: 14

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

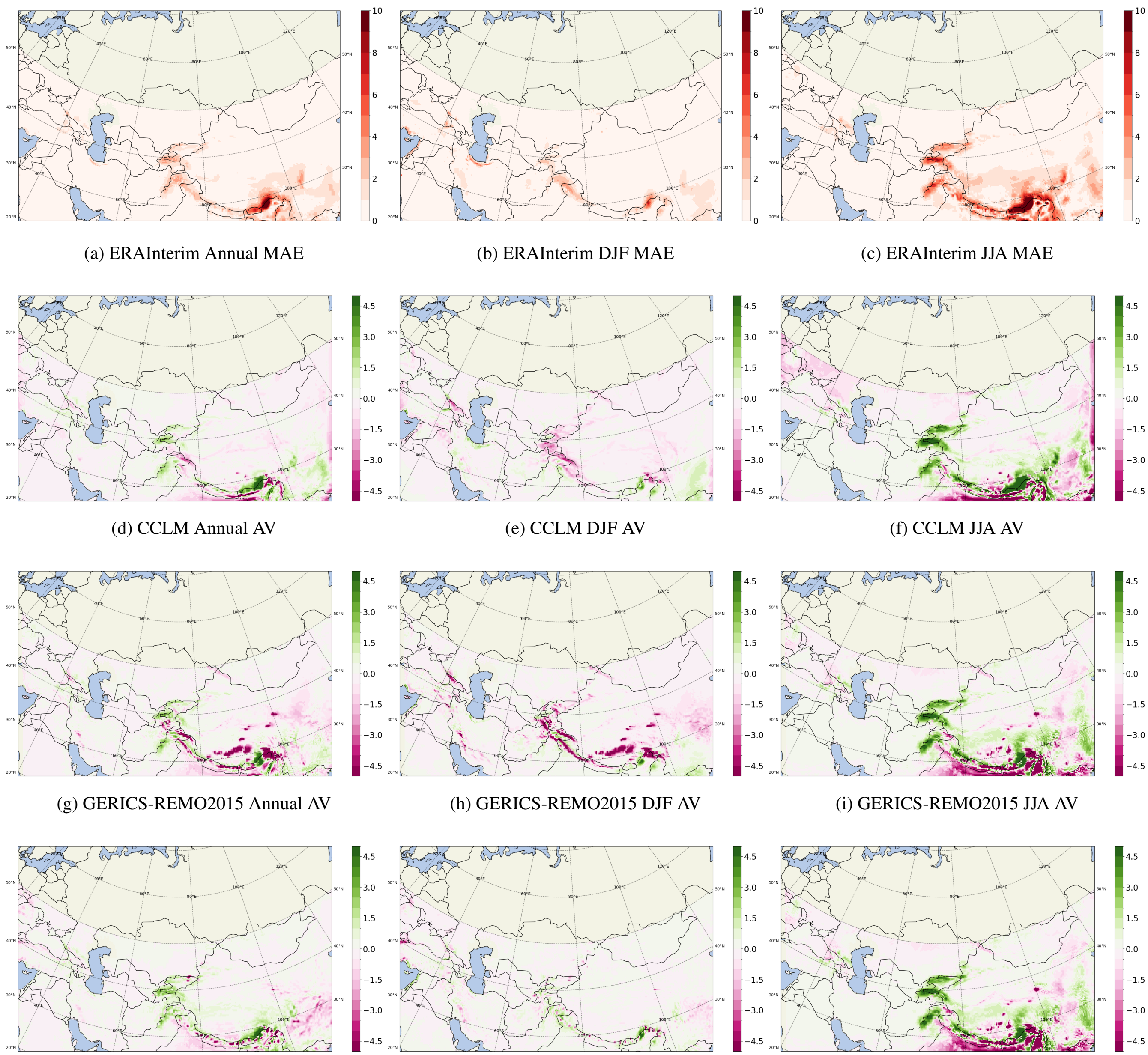
The colorbars of (a) and (b) should have legends.

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

here refer to my comment in the first part of the results, it's not clear if this is the average of daily values over all years, or daily prec averaged per year and then averaged over all years.

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions



(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 2. MAE of daily precipitation (mm/day) from ERAInterim, as well as, added value (AV) as measured by MAE differences between ERAInterim and RCM ($MAE_{ERAInterim} - MAE_{RCM}$) in mm/day for annual (a,d,j,i), winter (b,e,h,k) and summer (c,f,i,l). CHIRPS is used as observation.

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

This figure also needs legends on colorbars (mm/day).

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

MAE needs to be defined in the legend at least once.

Number: 3 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

interpolated on grid [...] (As in Figure 6 and others)

Number: 4 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

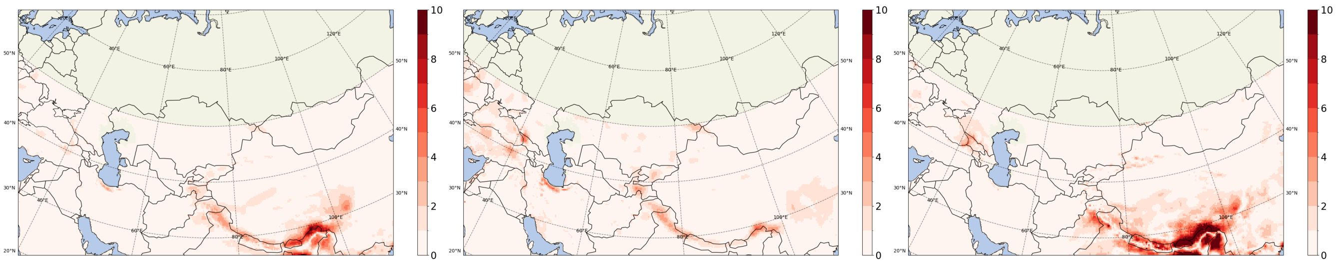
Number: 5 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

December, January, February

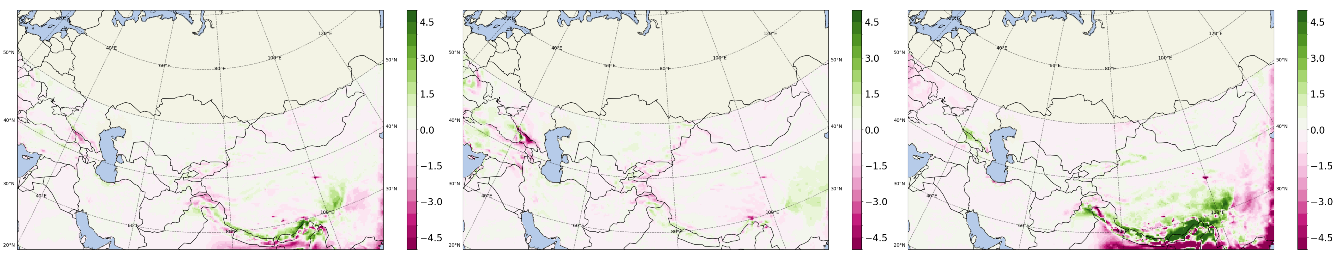
Number: 6 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

June, July, August

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023



(a) MPI-ESM1-2-HR Annual MAE (b) MPI-ESM1-2-HR DJF MAE (c) MPI-ESM1-2-HR JJA MAE



(d) MPI-ESM1-2-HR-CCLM Annual AV (e) MPI-ESM1-2-HR-CCLM DJF AV (f) MPI-ESM1-2-HR-CCLM JJA AV

Figure 3. MAE of daily precipitation (mm/day) from MPI-ESM1-2-HR, as well as, added value (AV) as measured by MAE differences between MPI-ESM1-2-HR and RCMs ($\text{MAE}_{\text{MPI-ESM1-2-HR}} - \text{MAE}_{\text{RCM}}$) in mm/day for annual (a and d), winter (b and e) and summer (c and f). CHIRPS is used as observation.

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~~:

$$\frac{1}{n}\sum_{i=1}^{n} y_i = x. \tag{A1}$$

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:

$$y_i = \exp(\tilde{y}_j) \cdot \frac{x}{\frac{1}{n}\sum_{i=1}^{n} \exp(\tilde{y}_i)}. \tag{A2}$$

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

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
same comment as figure 2 on colorbars (and for all other figures)

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
MAE needs to be defined in the legend at least once.

Number: 3 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
interpolated on [...] see comment above

Number: 4 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 5 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Number: 6 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
same comment as Figure 2

Number: 7 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
,

Number: 8 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
c

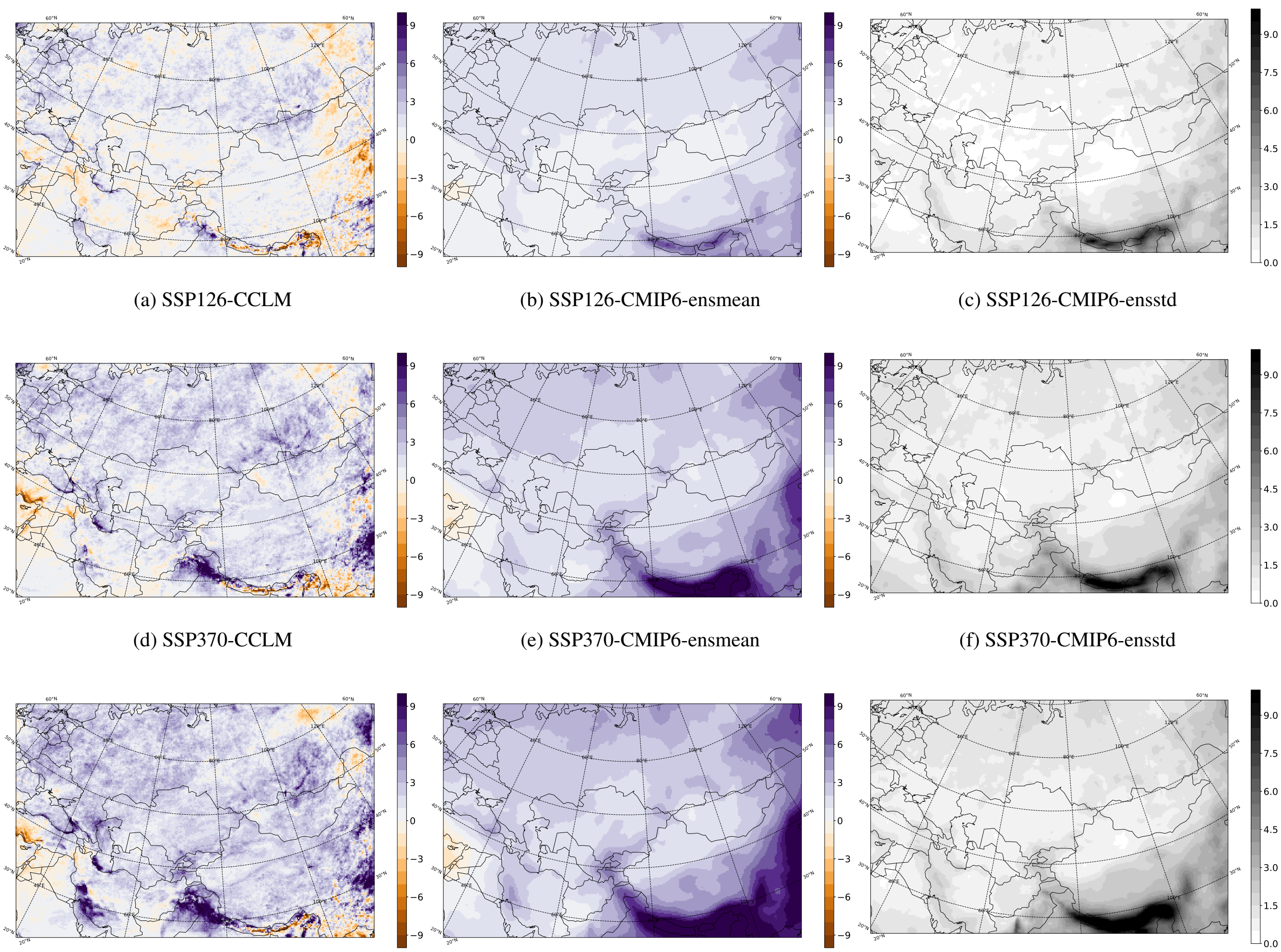
Number: 9 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
the following form

Number: 10 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

Number: 11 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28
not the right word

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Open Access
Discussions



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

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

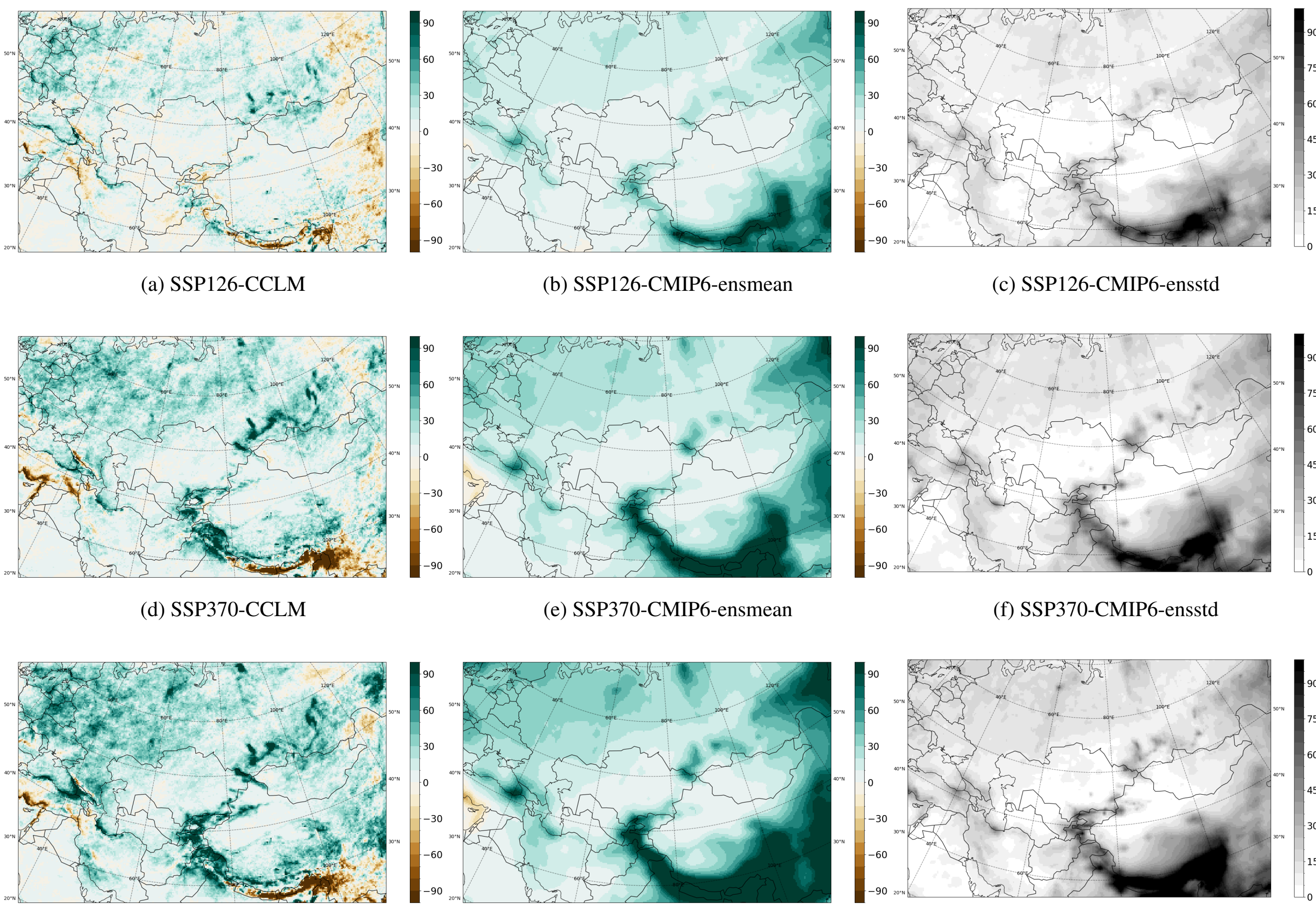
(g) SSP585-CCLM (h) SSP585-CMIP6-ensmean (i) SSP585-CMIP6-ensstd

Figure 4. 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.

This page contains no comments

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Open Access
Discussions



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

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

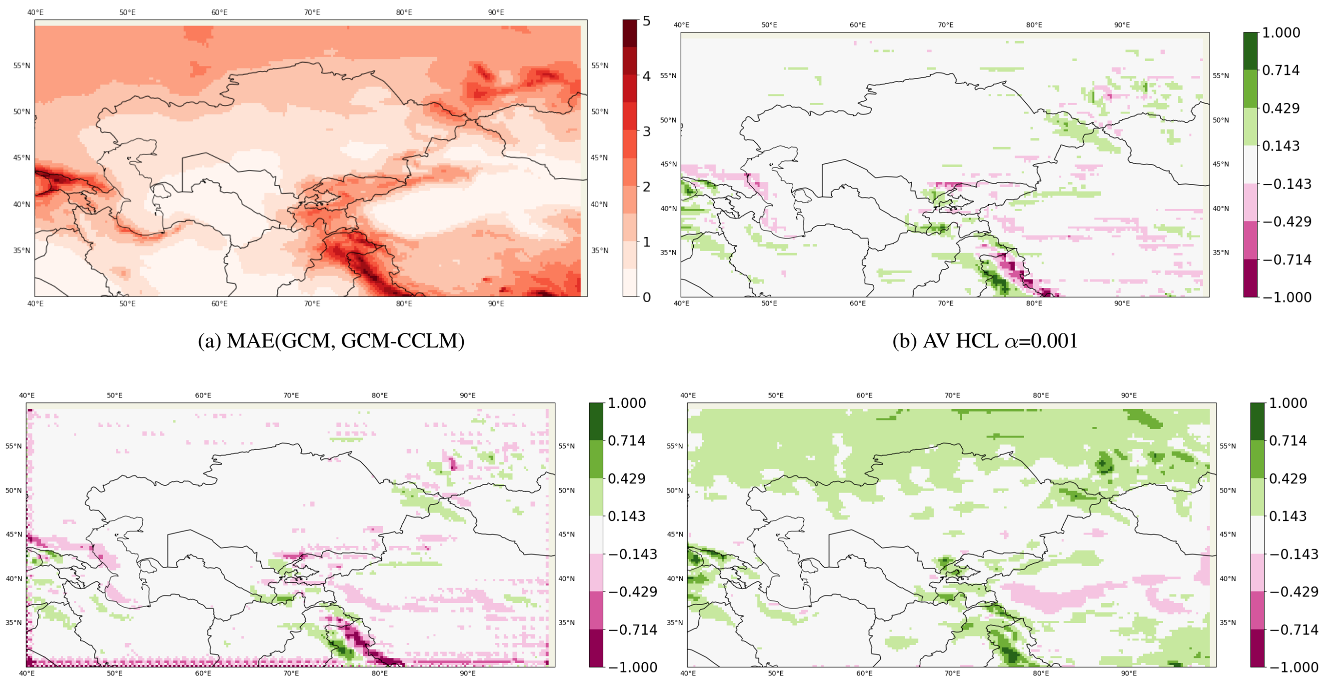
(g) SSP585-CCLM (h) SSP585-CMIP6-ensmean (i) SSP585-CMIP6-ensstd

Figure 5. Changes in number of days with precipitation more than 20mm 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.

Number: 1 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
[insert space]

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions



(a) MAE(GCM, GCM-CCLM)

(b) AV HCL α=0.001

(c) AV SCL α=0.99

(d) AV NoCL

Figure 6. 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.

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

$$\text{Loss} = (1-\alpha)\cdot MAE + \alpha \cdot \text{CV}, \tag{A3}$$

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:

$$\text{CV} = \text{MSE}(\frac{1}{n}\sum_{i=1}^{n} y_i, x) \tag{A4}$$

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.

Page: 19

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

this is missing in the other figures above when you mention the GCM. You need to add to the other figures that it was interpolated.

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

MAE needs to be defined in the legend at least once.

Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

remove space

Number: 4 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

[insert space here] and also at the other MAEs

Number: 5 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

The

Number: 6 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

conserves

Number: 7 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

is

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

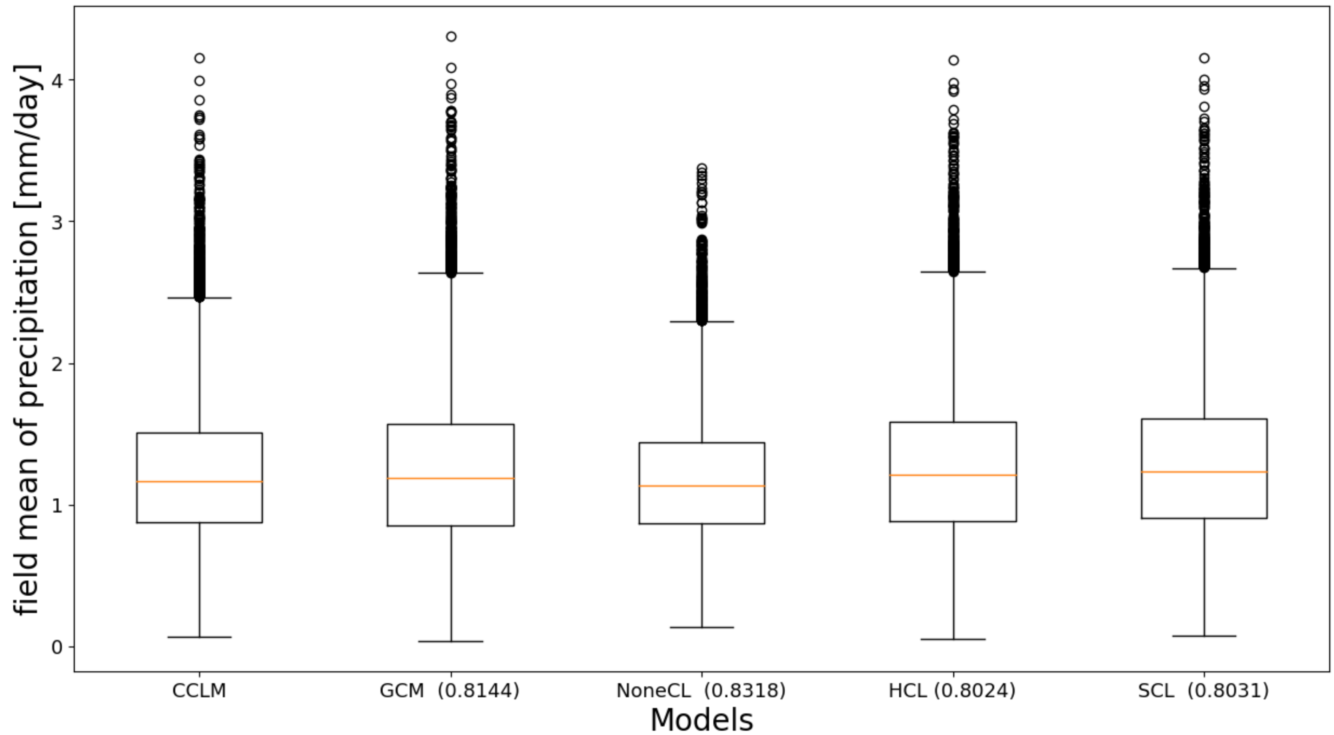


Figure 7. Boxplot of averaged daily precipitation over the domain for different models. Numbers in the parenthesis indicate the correlation coefficients between each model and the CCLM simulation.

Appendix B: CNN runs

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

```
# 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

# 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:
```

Page: 20

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
In this figure, you use NoneCL instead of NoCL as in other places.

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
over how many sampels? How long is the timeseries ?

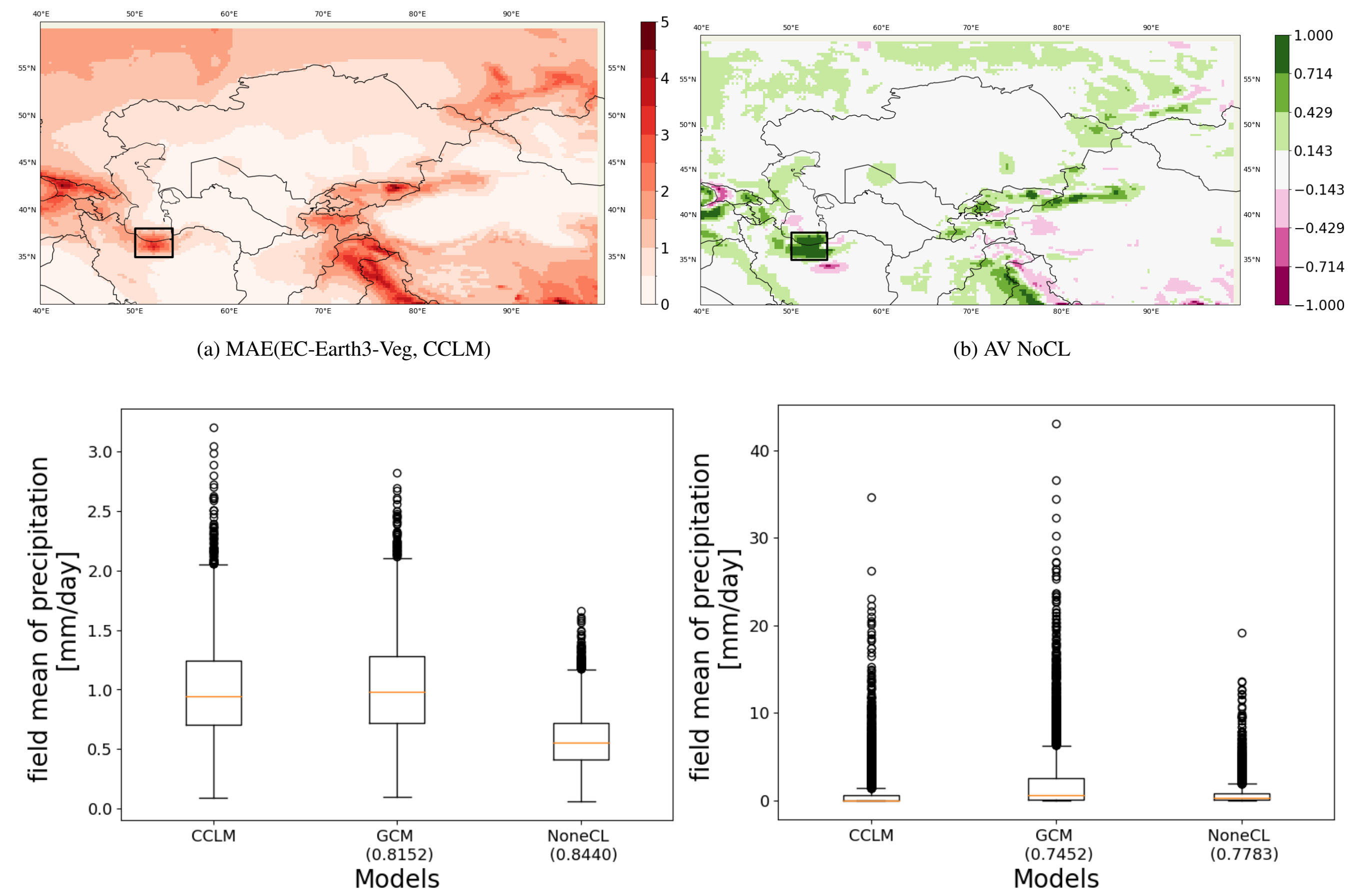
Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
which domain ? Because in your study you have two.

Number: 4 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28
The

Number: 5 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28
I'm not sure this is necessary information seeing how it repeats what you've already told in the method.

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Open Access
Discussions



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

(b) AV NoCL

(c) CA

(d) Box

Figure 8. 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.

Page: 21

Number: 1 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

Same comment as before, you use NoneCL instead of NoCL

Number: 2 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

MAE needs to be defined

Number: 3 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

[insert space]

Number: 4 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

[insert space]

Number: 5 Author: anonymous Subject: Inserted Text Date: 20.12.2023, 11:28:28

The

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development Discussions
Open Access

```
$ 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
The final model run outputs ~~could~~ be find here :

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. BF thanks the CCLM-community for providing the model code and the pre-processing code to convert the GCM to CCLM input files.

Number: 1 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

will

Number: 2 Author: anonymous Subject: Cross-Out Date: 20.12.2023, 11:28:28

can

Number: 3 Author: anonymous Subject: Highlight Date: 20.12.2023, 11:28:28

? do you mean in the folder where you ran the command or did you forget to add the folder? Also there should be no space befroe a :

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

Geoscientific Model Development
Discussions

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.

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.

This page contains no comments

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

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.

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.

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.

This page contains no comments

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.

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., 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.

This page contains no comments

https://doi.org/10.5194/gmd-2023-227
Preprint. Discussion started: 5 December 2023

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.

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.

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.

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.

This page contains no comments

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.

This page contains no comments