Author's response #2

Repeatable high-resolution statistical downscaling through deep learning

Dánnell Quesada-Chacón
02/08/22

1. Replies to reviewer #1

1.1. General comments

This paper evaluates a suite of related Deep Learning models for downscaling precipitation data over the Ore Mountain region in Germany. In addition to evaluating U-Net and U-Net++ architectures with a number of different options, the authors also explore the use of containerization to enable repeatability of the experiment, and provide a Singularity container with the associated code and environment.

This paper is well-written and of interest to the readers of GMD, and the topic is highly relevant and of value to the scientific community. I have only minor revisions to suggest.

It's a good paper, and I enjoyed reading it.

Good work!

R/ We would like to sincerely thank the reviewer for the time dedicated reviewing the manuscript and the encouraging words. We hope to have satisfactorily amended the concrete comments.

1.2. Suggested revisions:

1.2.1. The manuscript has too many novel abbreviations. I recommend replacing all of the following with their full expansions, both in the text and in figures: 1. CC - climate change 2. TF - transfer function 3. OM - Ore Mountains 4. EOM - Eastern Ore Mountains? (This abbreviation is never defined) 5. SN - Saxony 6. DD - Dresden

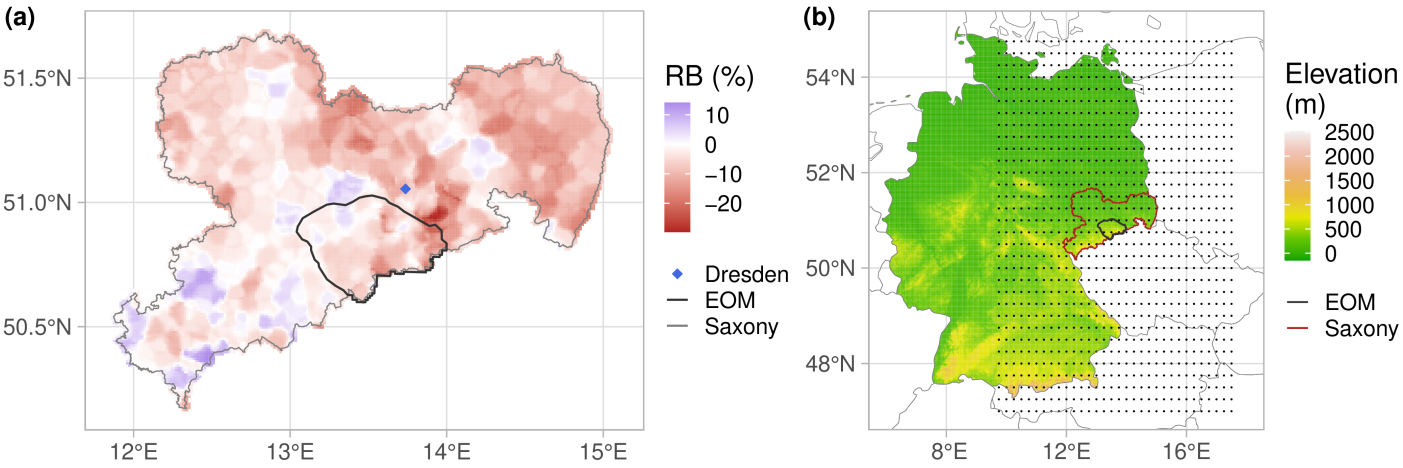
R/ We expanded Ore Mountains, Saxony and Dresden. We kept the others. Eastern Ore Mountains is defined towards the end of the introduction.

C/ Please see the *marked-up* pdf file.

1.2.2. I recommend changing the following abbreviations: 1. IB - international borders - these are obvious from context and can be omitted from the legend. 2. m.a.s.l - meters above sea level - I would call this simply "elevation (m)"

R/ Both of these suggestions were included accordingly.

C/ The following Figure replaced the original Figure 1 in the manuscript:



1.2.3. Consider also expanding: SD - statistical downscaling, DL - deep learning. Although these abbreviations are not unknown, I think the text would read more clearly without them.

R/ Since these two are used quite often throughout the manuscript we decided to keep them.

1.2.4. Line 107: change "the Fichtelberg with 1215 m.a.s.l and the Kahleberg (905 m.a.s.l.)" to "the Fichtelberg (elevation 1215 m) and the Kahleberg (elevation 905 m)"

R/ Changed accordingly.

C/ Please see the *marked-up* pdf file.

1.2.5. URLs should be listed as part of a reference, not included in the text of the paper. For example, on line 132, you could simply write "The code needed to recalculate these results can be found on GitHub (Quesada-Chacon, 2022), with all the modifications...". Please replace the URLs on lines 113, 132, 198, 212, 214, and 217 (and any others I may have missed) with citations.

R/ The originally used format is found in most recent papers on GMD (e.g. https://doi.org/10.5194/gmd-15-5807-2022, https://doi.org/10.5194/gmd-15-5567-2022, https://doi.org/10.5194/gmd-15-5857-2022). Still, we introduced the suggested changes throughout the manuscript, since it does look tidier. Following other GMD papers, we left the links in the *Code and data availability* section, but not the last access dates. Also, it was not clear how to include all the details of the links in the bibtex format. We had to use the `note` field for the `@misc` entry type to add the *last access* date, instead of `urldate`, since the latter was not showing it at all, which might have unexpected effects.

C/ Please see the *marked-up* pdf file. E.g.

- "Nevertheless, newer versions of TensorFlow have included further deterministic implementations of GPU algorithms (https://github.com/NVIDIA/framework-determinism, accessed on 12 December 2021)" → "Nevertheless, newer versions of TensorFlow have included further deterministic implementations of GPU algorithms (NVIDIA, 2021)"
    - With the Reference entry: NVIDIA: Framework Determinism, GitHub, https://github.com/NVIDIA/framework-determinism, last access: 12 December 2021, 2021.

- “A subset of the ReKIS gridded dataset for the Free State of Saxony was used as predictand, with a spatial resolution of 1 km at a daily temporal resolution, available at https://rekis.hydro.tu-dresden.de (accessed on 20 November 2021)” → “A subset of the ReKIS (2021) gridded dataset for the Free State of Saxony was used as predictand, which has a spatial resolution of 1 km at a daily temporal resolution.”
    - With the Reference entry: ReKIS: Regionales Klimainformationssystem Sachsen, Sachsen-Anhalt, Thüringen, https://rekis.hydro.tu-dresden.de, last access: 11 July 2022, 2021.

1.2.6. In section 3.2, please add a definition of the Bernoulli-Gamma loss function used.

R/ Added accordingly.

C/ The next two equations, the Bernoulli-Gamma loss function and the negative log-likelihood of it, were added to the manuscript:

$$f(y;\rho,\alpha,\beta) = \begin{cases} 1-\rho & \text{if} \quad y=0 \\ \dfrac{\rho\cdot(y/\beta)^{\alpha-1}\cdot e^{-y/\beta}}{\beta\cdot\Gamma(\alpha)} & \text{if} \quad y>0 \end{cases}$$

$$-lnf = \begin{cases} -\ln(1-\rho) & \text{if} \quad y=0 \\ \dfrac{\alpha\cdot\ln\beta+\ln\Gamma(\alpha)+y/\beta}{\ln\rho+(\alpha-1)\cdot\ln y} & \text{if} \quad y>0 \end{cases}$$

1.2.7. Line 165: A sentence or two discussing why batch normalization and spatial dropout were included as options (i.e., what effect they have and when one would want to use them) would be valuable to the reader.

R/ A couple of sentences addressing these options were added.

C/ “The basic “convolutional unit” (ConvUnit) consisted of a convolutional layer (kernel size 3 by 3), with the respective activation function and, optional batch normalization and spatial drop-out of feature channels. Two successive ConvUnits constitute a “convolutional block” hereafter referred to as ConvBlock.” → “The basic “convolutional unit” (ConvUnit) consisted of a convolutional layer (kernel size 3 by 3), with the respective activation function and, optional batch normalization and spatial drop-out. The last two options are used to avoid overfitting of the model. Batch normalization standardizes the layer’s input data, which also improves learning speed, while spatial drop-out (2D version of Keras) randomly ignores entire feature maps during training. Two successive ConvUnits constitute a “convolutional block” hereafter referred to as ConvBlock.”

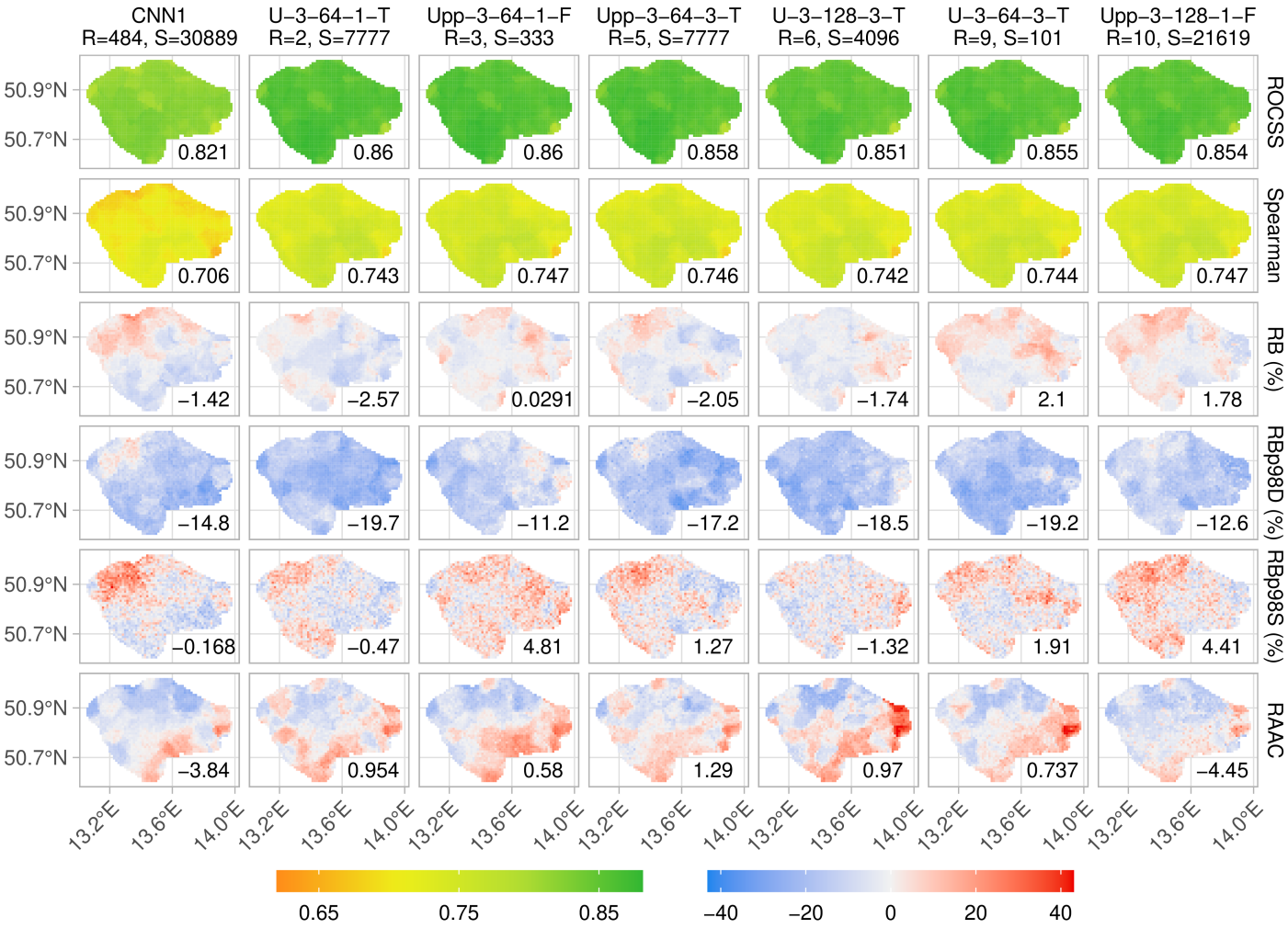
1.2.8. Although many writers use the passive voice in scientific writing, active writing using the first person is much clearer and easier to understand. The authors could improve the paper by switching to active voice throughout; for example, on line 123, replacing "The reanalysis dataset employed as a predictor is ERA5..." with "For the predictor, we used the ERA5 reanalysis..." makes the text much easier to follow. Since this would require extensive editing, I do not expect the authors to make this change, but mention it merely for the sake of future papers.

R/ We really appreciate the suggestion, it indeed could make the text much easier to follow. Unfortunately, it is not possible for this specific manuscript but, we will keep it in mind for future works.

1.2.9. Figure 4: I think this figure would be easier to understand if the green-to-orange color bar were reversed, with orange indicating lower values (worse performance) and green indicating higher values (better performance).

R/ The Figure was modified accordingly.

C/ The following Figure replaced the original one in the manuscript:



1.2.10. Line 294: rather than giving Lat & Lon coordinates, I think it would be clearer to point to this location as the southeast corner of the region.

R/ Changed accordingly.

C/ "… with a noticeable decrease in its performance around 50.7°N and 13.9°E …" → "… with a noticeable decrease in its performance in the south-eastern corner of the region …"

1.3. Minor corrections:

Since all of the following corrections were included as suggested by the reviewer, no change comment (***C/***) was added, please see the *marked-up* pdf file.

1.3.1. Line 59: change "1 500" to "1500"

R/ Changed accordingly.

1.3.2. Line 120: change "which leads to a number of 1916 pixels" to "giving a region with 1916 pixels"

R/ Changed accordingly.

1.3.3. Line 123: change "with a spatial resolution" to "which has a spatial resolution"

R/ Changed accordingly.

1.3.4. Line 209: I think you want "interface" here, not "interfere"

R/ Changed accordingly.

1.3.5. Line 225: change "functions" to "function"

R/ Changed accordingly.

1.3.6. Line 227: change "improved significantly" to "significantly improved"

R/ Changed accordingly.

1.3.7. Line 252, 289: change "yet," to "although" (no comma)

R/ Changed accordingly.

1.3.8. Line 323: change "observations" to "failures" (not a correction, but clearer)

R/ Changed accordingly.

1.3.9. Line 334: change "Besides" to "In addition"

R/ Changed accordingly.

1.3.10. Line 380: change "Yet" to "However"

R/ Changed accordingly.

1.3.11. Line 500: "https://doi.org/" has been doubled in this URL

R/ Changed accordingly.

C/ A couple of other references had the same issue, which was ammended.

2. Replies to reviewer #2

2.1. General comments

2.1.1. In this paper, the author employed deep learning models to downscale rainfall at a regional scale (1km) over the Eastern Ore Mountains in Saxony. The author used different deep learning algorithms including the state-of-the-art U-Net and U-Net++ models and compared their performance with CNN1 (benchmark). The aim of this work was not only to downscale precipitation but also to explore the repeatability aspect of the downscaling experiment. The findings in this paper are very interesting. In general terms, this paper falls within the scope of this journal, the figures and tables are well organized, and the results are properly discussed. However, the paper needs extensive revision, especially the introduction section.

R/ We would like to thank the reviewer for time dedicated reviewing the manuscript and hope that we have addressed satisfactorily the specific details under scrutiny and the questions.

2.2. Abstract

2.2.1. The abstract is well written however, some very interesting findings mentioned in the conclusion could also be included in the abstract.

R/ Great suggestion. Some brief details were added, such as relative amount of parameters, the training time and the attainable datasets along their potential applications.

C/ "The results were validated with the VALUE framework. The introduced architectures show a clear performance improvement when compared to previous SD benchmarks. Characteristics of the DL models configurations that promote their suitability for this specific task were identified, tested and argued. Full model repeatability was achieved employing the same physical GPU." → "The results were validated with the robust VALUE framework. The introduced architectures show a clear performance improvement when compared to previous SD benchmarks. The best performing architecture had a small increase in total number of parameters, in contrast with the benchmark, and a training time of less than 6 minutes with one NVIDIA A-100 GPU. Characteristics of the DL models configurations that promote their suitability for this specific task were identified, tested and argued. Full model repeatability was achieved employing the same physical GPU, which is key to build trust in DL applications. The EURO-CORDEX dataset is meant to be coupled with the trained models to generate a high-resolution ensemble, which can serve as input to multi-purpose impact models."

2.3. Introduction

2.3.1. The introduction requires an extensive language revision. The flow of the paragraphs needs to be adjusted.

R/ Several changes were made to the introduction, please see the comparison pdf.

C/ Since several changes were made, please see the *marked-up* pdf file.

2.3.2. The IPCC report citation is wrong, this is the correct citation (IPCC 2021).

R/ The citation was corrected.

C/ (IPCC et al., 2021) → (IPCC, 2021)

2.3.3. "2011-2020, being the warmest on record." And "while 2020 tied with 2016 as the hottest" something is not correct here, the author needs to check which years are the hottest (2020 and 2021) or (2020 and 2016).

R/ The manuscript was written in late 2021 and submitted at the beginning of the present year, thus, the updated information for 2021 was still not available. Now, 2021 is included and a reference was added.

C/ "The last six years (2015-2020) are the warmest on record as well, while 2020 tied with 2016 as the hottest." → "The last seven years (2015-2021) have been the warmest on record as well, while 2020, tied with 2016, was the hottest (WMO, 2022)."

2.3.4. "…and simultaneous numerous heat waves", this sentence should be improved. "On a smaller…in Germany in decades." this sentence is ambiguous and too long. It should be divided into two or three sentences.

R/ This sentence was rewritten and punctuated more clearly.

C/ "On a smaller scale, there have been several signs of the effects that CC can have on extreme events, e.g., the 2020 fires in Siberia, California and Australia, the 2021 summer flash floods events in Western Europe and China, and simultaneous numerous heat waves on the northern hemisphere, after observing the coldest April in Germany in decades." → "On a smaller scale, there have been several signs of the effects that CC can have on extreme events, e.g. the 2020 fires in Siberia, California and Australia; the 2021 summer flash floods events in Western Europe and China; and the 2021 summer heat waves on the northern hemisphere soon after observing the coldest April in Germany for decades."

2.3.5. The same remarque for the second paragraph, the first sentence is too long. The author is advised to rewrite it using short and clear sentences.

R/ Since the first sentence of the second paragraph is, to our criteria, of regular length, we interpreted that the reviewer meant the second one, which we split and rephrased.

C/ "Nevertheless, despite the remarkable improvements of GCMs over recent years, their spatial resolution of up to a few hundred kilometers, depending on the GCM generation, and large regional biases when contrasted to station data (Flato et al., 2013), yields their output unfit to be used directly for regional CC impact studies in fields such as hydrology, agronomy, ecology and risk management (Maraun and Widmann, 2018)." → "Nevertheless, despite the remarkable improvements of GCMs over recent years, their spatial resolution yields their output unfit to be used directly for regional CC impact studies in fields such as hydrology, agronomy, ecology and risk management (Maraun and Widmann, 2018). The resolution of GCMs can be of up to a few hundred kilometers, depending on the GCM generation, which results in large regional biases when contrasted to station data (Flato et al., 2013)."

2.3.6. “which depending on the application can be a…”, this does not sound correct. “which depending on the application can be a…into climate4R (Iturbide et al., 2019).” The author is advised to rewrite this paragraph.

R/ Several sentences between the ones mentioned by the reviewer were modified.

C/ Since several changes were made, please see the *marked-up* pdf file.

2.3.7. The references used in paragraph 4 are old, the author might consider exploring recent papers.

R/ The idea of this paragraph is to convey a chronological order of which methods were applied to statistical downscaling for the first time (roughly). Therefore, we start with rather old references about simple methods and finish with references about deep learning models on gridded data which are quite recent (2018-2021). Thus, we believe that the mentioned references are suitable for the message we want to communicate. However, we added Maraun and Widmann, 2018 to the first sentence of this paragraph to make clear that the concept behind the "transfer functions" has not changed since the originally added reference.

C/ Please see the *marked-up* pdf file.

2.3.8. The section where the term “reproducibility” term is explained is too long, the author might consider summarising it.

R/ Since: 1. "Reproducibility" is a pivotal subject matter of the manuscript. 2. The literature about these terms can be confusing. and 3. That we use several concepts explained in the mentioned section, such as "repeatability", "similarity", "general reproducibility" and "methods reproducibility". We believe that this section should not be shortened.

C/ A couple of minor changes were introduced, please see the *marked-up* pdf file.

2.4. Data

2.4.1. “The raw station data…to Deutsch (1996) for the amounts. This sentence is not clear.

R/ This sentence was split in two and further details were added.

C/ “The raw station data for precipitation was corrected after Richter (1995) and interpolated using Indicator Kriging (Deutsch and Journel, 1998) for the probabilities and Ordinary Kriging (Wackernagel, 2010) with a negative weight correction and exponential semivariogram model according to Deutsch (1996) for the amounts.” → “The raw station data for precipitation was corrected after Richter (1995) and interpolated using Indicator Kriging (Deutsch and Journel, 1998) for the probabilities of precipitation. Ordinary Kriging (Wackernagel, 2010) with a negative weight correction and exponential semivariogram model according to Deutsch (1996) was employed to estimate the amounts of precipitation.”

2.4.2. “https://github.com/dquesadacr/Rep_SDDL“ this link is not accessible.

R/ The link seems to work properly from our side.

2.4.3. The author indicated that the precipitation dataset was used as a predictand, while several variables were considered from the predictor. In the training phase, shouldn't the author use the same variable from the predictor and the predictand to train the model?

R/ Under the perfect prognosis approach, the transfer functions create the relationship between the predictand (precipitation from ReKIS) and the predictors (the 20 different layers of information). Precipitation (from ERA5) could have been used as a predictor, but since there are large biases between reanalysis datasets and the output from RCMs, such as EURO-CORDEX (which is meant to be used in follow up research), we did not considered prudent to include it. In the case of the Model Output Statistics (MOS) statistical downscaling approach, the predictor is indeed the same variable as the predictand, but is a completely different methodology.

2.5. Methods

2.5.1. In the caption of Figure 1, The author didn't mention which variable is considered to calculate the relative bias. Is it precipitation?

R/ Yes, it is precipitation. The caption was modified.

C/ "**Figure 1.** Location of the study region and the predictor's domain. (a) Relative bias (RB) between training and validation periods for the whole ReKIS domain for Saxony." → "**Figure 1**. Location of the study region and the predictor's domain. (a) Relative bias (RB) of precipitation between training and validation periods for the whole ReKIS domain for Saxony."

2.5.2. Table 1. Replace "d" with "day".

R/ Changed accordingly.

C/ Please see the *marked-up* pdf file.

2.5.3. The links provided are not accessible: https://github.com/dquesadacr/Rep_SDDLhttps://bit.ly/ 215 dl-determinism-slides-v3, https://bit.ly/ 215 dl-determinism-slides-v3

R/ The links seems to work properly from our side. There are some numbers and spaces (215) included in the comment of the reviewer, which are not in the manuscript and could cause the issue.

2.5.4. The focus of this work was on precipitation, however, the author also mentioned that several variables are selected from the predictor (zonal and meridional wind, temperature, geopotential, and specific humidity). How did the author use these variables to downscale precipitation?

R/ Following the perfect prognosis approach by means of deep learning, the 20 different input layers (predictors from ERA5) were used to train the models, that try to simulate the observed values (predictand). Both predictors and predictand need to have a time correspondence, thus, the atmospheric conditions seen, for example, the first of July of 1995, should be properly modelled in the reanalysis output, which is in turn associated by the model with the observed precipitation during the same day. The deep learning architectures, with all their different layers, characteristics and arrangements, are the ones that learn spatial patterns from these different layers of information. The added robustness of the U-like architectures allow the mentioned performance

improvement. The latter is a quite brief and over-simplified explanation of the perfect prognosis approach used also in Baño-Medina et al., 2019 and more in-detail explained in Maraun and Widmann, 2018.

2.5.5. It is advised to add another Figure to show the details of the model used (including the resolution of the input and output), the author is referred to check Figure 3 in (Baño-Medina et al., 2019).

R/ We appreciate this suggestion but since we tested a considerable number of different deep learning architectures (including the benchmark from Baño-Medina et al., 2019, CNN1), it is complicated to achieve such a Figure. As an analog to Figure 3 of the aforementioned paper: 1. The left hand side will look the same, 20 different input layers or variables. 2. The upper right hand side will also be the 3 parameters but for the EOM region, which results in precipitation (lower right) at 1 km resolution. 3. The convolutions and dense layers would be substituted by the different U-like architectures.

2.6. Results and discussion

2.6.1. "This could be applied to by …" this sentence needs correction, line 285.

R/ The "to" was indeed incorrect and was removed.

C/ "This could be explained by to the joint effect of the already shown RB in Figure 1a and the extreme events of 2013, employed in the independent validation metrics calculation only." → "This could be explained by the joint effect of the already shown RB in Figure 1a and the extreme events of 2013, included in the independent validation metrics calculation only."

2.6.2. Figure 4. These matrices are calculated on which years, is it the validation period (2010-2015)?

R/ Yes, as mentioned in Section 3.1, the independent validation dataset is the one between 2010 and 2015, which was not employed during training.

2.7. Conclusion

2.7.1. The author mentioned that 5 variables from the predictor (ERA5) were used to downscale precipitation, however, in the conclusion, the author stated that 20 variables were used. Which one is correct?

R/ In the Data section it was indeed mentioned that 5 variables were used, but at 4 different levels each, thus, the 20 different variables mentioned in the conclusions. For example, the temperature related variables or input layers are T@1000, T@850, T@700 and T@500 (all the numbers are in hPa). To avoid further misunderstandings, we added: the word "atmospheric" before "variables" and ", for a grand total of 20 different variables." after the levels in the Data section.

3. Additional relevant changes

3.1. Code and data availability

C/ This section was sustantially modified to incorporate the comments of reviewer #1 regarding the *url* links. Therefore, we removed the *last access* dates, kept the links (as in other GMD papers) and added all the links to the references, with the last access dates. Also, the *GitHub* link to the code was changed to the permanent *Zenodo* one. Since the rendered *README* file with further details on how to modify and run the source code of the project shown in the *GitHub* repository is lost in the *Zenodo* link (the source code of the *README* file is shown), we added another reference for *GitHub* with the pertinent changes in the section. Thus:

- "*Code and data availability.* The processed predictors and predictand used for the development of the models can be found in https://doi.org/10.5281/zenodo.5809553 (Quesada-Chacón, 2021b) (last access: 31 December 2021). The singularity container used for the calculations can be downloaded at https://doi.org/10.5281/zenodo.5809705 (last access: 10 April 2022) (Quesada-Chacón, 2021a). The repository https://github.com/dquesadacr/Rep_SDDL (last access: 16 January 2022) hosts the code employed in this project (Quesada-Chacón, 2022)."

Was changed to:

- "*Code and data availability.* The processed predictors and predictand used for the development of the models can be found in https://doi.org/10.5281/zenodo.5809553 (Quesada-Chacón, 2021b). The singularity container used for the calculations can be downloaded at https://doi.org/10.5281/zenodo.5809705 (Quesada-Chacón, 2021a). The version of the code employed in this paper can be found at https://doi.org/10.5281/zenodo.5856118 (Quesada-Chacón, 2022a). The repository https://github.com/dquesadacr/Rep_SDDL (Quesada-Chacón, 2022b) hosts the rendered description of the software with further details to properly run and modify the source code."

With the following new reference entry: "Quesada-Chacón, D.: Rendered description of the source code of "Repeatable high-resolution statistical downscaling through deep learning", GitHub, https://github.com/dquesadacr/Rep_SDDL, last access: 11 July 2022, 2022b." . We are open to suggestions about how to add this entry, or the title itself, since the title on GitHub is the title of the manuscript and could be confusing for the reader, thus "Rendered description of the source code of..." was added. We wanted to help the reader to access the properly rendered instructions for modifying and running the code, hope this helps.