Replies to Reviewer 2

1. General comments

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 his time and hope that we have addressed satisfactorily the specific details under scrutiny and the questions.

2. Abstract

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.

3. Introduction

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.

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

R/ The citation was corrected.

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.

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.

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.

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.

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.

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.

4. Data

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.

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

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

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.

5. Methods

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.

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

R/ Changed accordingly.

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.

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.

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.

6. Results and discussion

6.1. “This could be applied to by …” this sentence needs correction, line 285.

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

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.

7. Conclusion

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.