This manuscript presents a case study in understanding rainy season dynamics in semi-arid regions. The study is intended to develop and validate a new framework for calibrating rainy season metrics. The paper is generally clear.

Dear Jingwei Zhou, thanks a lot for taking the time to review our research and for your thoughtful comments allowing us to significantly improve our manuscript. Please find our point by point responses below. Note that all line number references are referring to the cleaned revised manuscript and not the version with track changes. For your orientation, we use **bold** letters to indicate new or revised text and *italic* letters to indicate citations.

Comment #1.1

Although CMIP5 data seem to be applicable in your research, I will still suggest you using the newer CMIP6 data, which are accessible in many data portals provided by organizations such as ESGF Copernicus. CMIP5 data has been more than ten years since developed, while CMIP6 incorporate numerous updates and incorporate more features, let alone it may address the issues you do mentioned in your text (lines 345-350). In addition, the implications or significance for this novel method in evaluating rainy season and to be applied in future research could be discussed further. A brief comment on the potential transferability of this framework to other semi-arid regions would be valuable as well. Overall, I will give this paper the suggestion of moderate revisions

Thanks for your suggestions.

Regarding CMIP5 vs. CMIP6: Please note that we do not directly use CMIP5 data but rather use a high resolution statistically downscaled CMIP5 data product which was specifically created for the area by Potter et al. (2023). This dataset is the only existing high-resolution, gridded future dataset which combines a large ensemble of future climate models for the Rio Santa basin. While we agree that a dataset based on CMIP6 would be interesting, it however simply does not exist at this point. Given that the main focus of this paper is to compare, validate and improve upon metrics to determine the onset and end of the rainy season, we believe including the creation of a new future climate dataset would obscure the main purpose of this manuscript.

Using statistically-downscaled CMIP5 models based on a bias-corrected regional climate model has advantages over downscaling directly to station data, as the regional climate model is spatially and temporally consistent, without the potential biases which come from using station data (for example there being fewer stations at higher elevations due to inaccessibility). However even using statistical rather than dynamical downscaling methods, creating large-ensembles of high-resolution gridded datasets is computationally expensive. As such, it is not feasible to redo previously-published work for this study.

To this end, we already stated in the preprint that the climate data we use in our analysis is unique and comes from a previously published study:

Introduction, l.105f.:

"*Specifically, we calculate the rainy season metrics based on convection-permitting, bias-corrected Weather Research and Forecasting (WRF) precipitation data and statistically downscaled CMIP5 projections (Potter et al., 2023) [...]"*

We revised the corresponding part in the data section of the manuscript to emphasize our rationale for choosing these data:

Methods/Data, l.145:

A key component is the WRF bias-corrected regional climate model data published by Potter et al. (2023), which provides consistent precipitation estimates at 4 km grid spacing from 1981 to 2018. In addition, Potter et al. (2023) produced statistically downscaled projections based on a 30-member CMIP5 ensemble from 2019 to 2100 using quantile delta mapping for both the RCP4.5 and RCP8.5 scenarios. These data preserve CMIP5 model trends while adjusting precipitation magnitude and the number of wet days based on the bias-corrected WRF data, and are available in the same 4 km grid spacing from 2019 to 2100. The two RCP scenarios allow us to assess multiple trajectories of future changes in the rainy season in the Rio Santa basin and provide a large dataset for metric sensitivity analysis. We do not evaluate metrics for raw, coarse-scale CMIP data in this study, as at their native resolution, they are known to inadequately represent orographic processes and interannual variability (e.g. Gutierrez et al., 2024).

In the section "Sensitivity analysis" we stated that *"[...] we used four Expert Team on Climate Change Detection and Indices (ETCCDI) climate indices (Zhang et al., 2011) based on the WRF and CMIP5 data by Potter et al., (2023)."* We acknowledge that this might read as if we created some data here - but in fact these data were first published in the previous work by Potter et al., (2023). To avoid this confusion, we changed this sentence to, l.239:

"[...] we used four Expert Team on Climate Change Detection and Indices (ETCCDI) climate indices (Zhang et al., 2011) based on the WRF and **statistically downscaled** *CMIP5 data* ***created by*** *Potter et al., (2023)."*

Finally, to ensure that there are no misunderstandings, we edited a sentence of the Introduction, specifically stating that we make use of results and data of previous studies, l.104:

"[...] we employ a multi-faceted approach capitalizing on previous studies: We combine several precipitation datasets with remote sensing data on temporal vegetation development."

We want to state that we take your comment seriously. Therefore, we have added a comparison to a study (published only after we submitted our original manuscript) by De la Cruz et al. (2025). The authors used the Liebmann metric to determine future changes to rainy season timings over a network of weather stations in Peru. Both that study, which used a statistical downscaling approach of CMIP6 data onto in-situ precipitation observations, and this study based on statistically down-scaled CMIP5 data using a convection-permitting model, find no statistically significant future changes of the onset or end of the rainy season using the Liebmann method. This suggests that conclusions on future changes of rainy season timings are likely to be robust with respect to the input climate dataset (CMIP5 and CMIP6), at least for the Liebmann metric. We revised our introduction now referring to their work:

Introduction, l.86:

"Potential shifts in the timing of the rainy season in the region, despite their profound implications for both societal and ecological systems, have only recently been assessed. Notably, De la Cruz et al. (2025) used an objective metric to derive rainfall sums and rainy season onset and end for a Peru-wide network of meteorological stations based on statistically downscaled CMIP6 projections to derive future changes. They found an increase in future annual precipitation and similar to other studies show that past rainy season dynamics in the broader Andean region reveal high inter-annual variability in rainy season onset, with generally non-significant or weak longer-term trends (Garcia et al., 2007; Giráldez et al., 2020; Gurgiser et al., 2016; Sedlmeier et al.,2023)."

The same authors suggest in their discussion that *"GCMs showed more favourable results for accumulated rainfall, but had limitations in simulating the onset and cessation, with correlations ranging from 0.2 to 0.6. Studies in other regions have demonstrated more optimal results for models with higher resolution."* pointing towards RCMs being the way forward rather than GCM ensembles. We added to the manuscript that we find similar results for the Liebmann method for the future period and make the argument that the data by Potter et al. (2023), based on CMIP5, are thus not compromising our results.

Results & Discussion, Past & Future Trends, l.398:

"The projections by Potter et al. (2023) we use are based on statistical downscaling of CMIP5 models. At the continental scale, many CMIP5 models were previously reported to poorly represent the South American Monsoon System (SAMS) (Bombardi and Carvalho, 2008), a challenge that is particularly pronounced in the topographically complex Andes. We compare our results to those of De la Cruz et al. (2025), who performed statistical downscaling based on meteorological stations in Peru using CMIP6 data and analyzed changes through the LM metric. De la Cruz et al. (2025) also project an increase in total precipitation, consistent with the findings of Potter et al. (2023), whose data informed this study. De la Cruz et al. (2025) also find no significant future changes in rainfall seasonality using the LM metric for the domain in which the Rio Santa basin is located. Furthermore, they highlight that GCMs have limited skill in simulating the interannual variability of rainy season onset and end, noting that many CMIP6 simulations still struggle to adequately represent the SAMS (see also Olmo et al., 2022). This suggests that the results from downscaled CMIP6 models and the downscaled CMIP5 models used in this study are consistent, at least based on the LM metric."

Regarding your comments on further discussion of applying our framework in future research and potential transferability, there are several important aspects to consider:

1. The widely adapted objective methods which showcase implausible sensitivities should be used with caution. It seems that the general opinion in the community is that these metrics are robust to changes as they do not depend on specific parameters. However, in our sensitivity analysis we show that this is in fact not the case. We now added a sentence further emphasizing this argument into the conclusions, l.428: *"[...] while more objective and flexible metrics have comparably low skill regarding this task. These objective metrics seem to exhibit implausible sensitivities that can potentially render them uninformative or even misleading under certain conditions of rainy season change.* "**We therefore recommend that the usage of such methods should be at least critically reviewed on a case-by-case basis to ensure that no false conclusions are drawn or misleading practical recommendations are made."**
2. We show that our calibration approach is effective and transferable and should be considered by future research. Among other sections in the manuscript where we mention that the approach can be readily adapted, most prominently we revised the abstract following a comment of Reviewer #2 (Comment #2.2), where we now state, l.20: **"Our findings emphasize the need for careful calibration of metrics across diverse climate scenarios and different locations to ensure their reliability for agricultural planning, policymaking, and climate adaptation strategies"**
3. We show that our bucket approach is outperforming other metrics and can be further developed or adapted depending on the specific question. We already dedicated a paragraph of the conclusions to it, which we now revised to further acknowledge this argument, l.444: **"Motivated by limitations in existing metrics, we designed a novel bucket metric, which outperforms other metrics for both the onset and end of the rainy season, shows physically consistent sensitivities and corrects for the vegetation – precipitation lag. The high skill and flexibility of the bucket metric allows for a wide range of applications in the context of hydroclimate in semi-arid areas. Additionally, it can likely be extended e.g. by making evapotranspiration dependent on energy- and/or water availability or by altering parameters over time to simulate changes, while still remaining simplistic and efficient. The bucket metric is to our knowledge also the first attempt to take legacy effects of water availability into account; particularly relevant in regions such as the Rio Santa basin where large inter-annual precipitation anomalies, for example related to ENSO, are common. Future attempts in addressing questions regarding the rainy season across semi-arid regions can readily use or adapt the bucket metric to suit a wide range of requirements"**
4. Finally, while not specifically mentioned in the text, we would like to point out that all the code used for the analysis and the preprocessed data is available in a public repository (https://github.com/lohae/RainySeasonMetrics) and preserved at Zenodo (https://doi.org/10.5281/zenodo.13952139), allowing other researchers to build upon it.
5. Together with answering a comment by reviewer #2 (comment #2.8) we now state in our objective statement more specifically the issue of transferability, l.XY: "**The principal**

objective of this study is to showcase a novel framework for characterizing the rainy season, emphasizing the importance of employing a calibration strategy for inferred rainy season onsets and ends. In addition, we test the sensitivity of rainy season metrics to plausible changes in rainfall intensity and frequency, as might occur due to global warming. By capturing shifts in seasonal rainfall dynamics, our approach provides a foundation for identifying and understanding hydrological changes that may inform future adaptation strategies. The proposed framework is designed to improve our understanding of variations in water availability within semi-arid regions, offering insights that extend beyond the Rio Santa basin and can be applied to similar climates. Regarding the Rio Santa basin, we aim to provide insights into past and future changes. We achieve this by: [...]"

Specific Comments:

Comment #1.2

Line 17: I suggest change it to "only a slight delay in rainy season end, but no consistent trends in rainy season onset," to stress the later parts with the bad results

Thank you. Following your suggestion while still appropriately integrating the latter part of the sentence ("inter-annual variability and ensemble spread being the dominant factors"), we have rephrased the sentence as follows, see l.18:

"Statistically downscaled CMIP5 ensemble projections for the future period suggest only a slight delay in the rainy season end, with no consistent trends in onset timing. Instead, inter-annual variability and ensemble spread remain the dominant influences."

Comment #1.3

Lines 40-45: The review of existing metrics could be more comprehensive, particularly regarding their applications in similar geographical contexts. Some results from the former study can be put here.

Thank you. Following your comment #1.1 regarding CMIP5/CMIP6 and the fact that the paper by De la Cruz et al., (2024) was published after we submitted our manuscript, we updated that part, now to more thoroughly discuss results of usage of metrics in the broader region. Please refer to the edits in our answer to your comment #1.1

Comment #1.4

Line 123: maybe you can mention why you choose these two scenarios

Thank you. We are limited to these two scenarios due to the available data published by Potter et al. (2023), see also our response to your comment #1.1 which led to rewriting this part. Specifically related to this comment, we now state in l.150: "**The two RCP scenarios allow us**

to assess contrasting trajectories of future changes in the rainy season in the Rio Santa basin and allow us to have a large data basis for metric sensitivity analysis."

Comment #1.5

Lines 254-269: I think most of these can be moved into methods, just leave some brief descriptions here in the results

Thanks for your suggestion. We have made a new subsection in the methods section where we moved this part. Note that we initially decided to put this method-heavy part in the results section as most of these methods are rather standard repertoire (sums, linear regression…) but thought it would be good if readers did not have to jump back to the methods section here. We however believe both versions to work and decided to follow your suggestion. Please refer to the new section 2.5 for the part we moved into the methods section. The original paragraph in Section 3.2 is now shortened to, at l.308:

To assess the sensitivity of rainy season metrics (RSO/RSE) to hydro-climatological changes, we correlated them with full hydrological year and sub-seasonal (SON, DJF, MAM, JJA) rainfall sums, as well as four ETCCDI climate indices as explained in Section 2.4. The results of these regressions are summarized in Figure 5, with detailed plots provided in Figures A4 and A5.

Comment #1.6

Line 320: "Feb."

Now l.205 and changed to: **"February"**, for consistency throughout the manuscript we did that as well in l. 161.

Comment #1.7

Lines 345-367: Most of these can be moved into conclusions part. They are discussions from my perspective

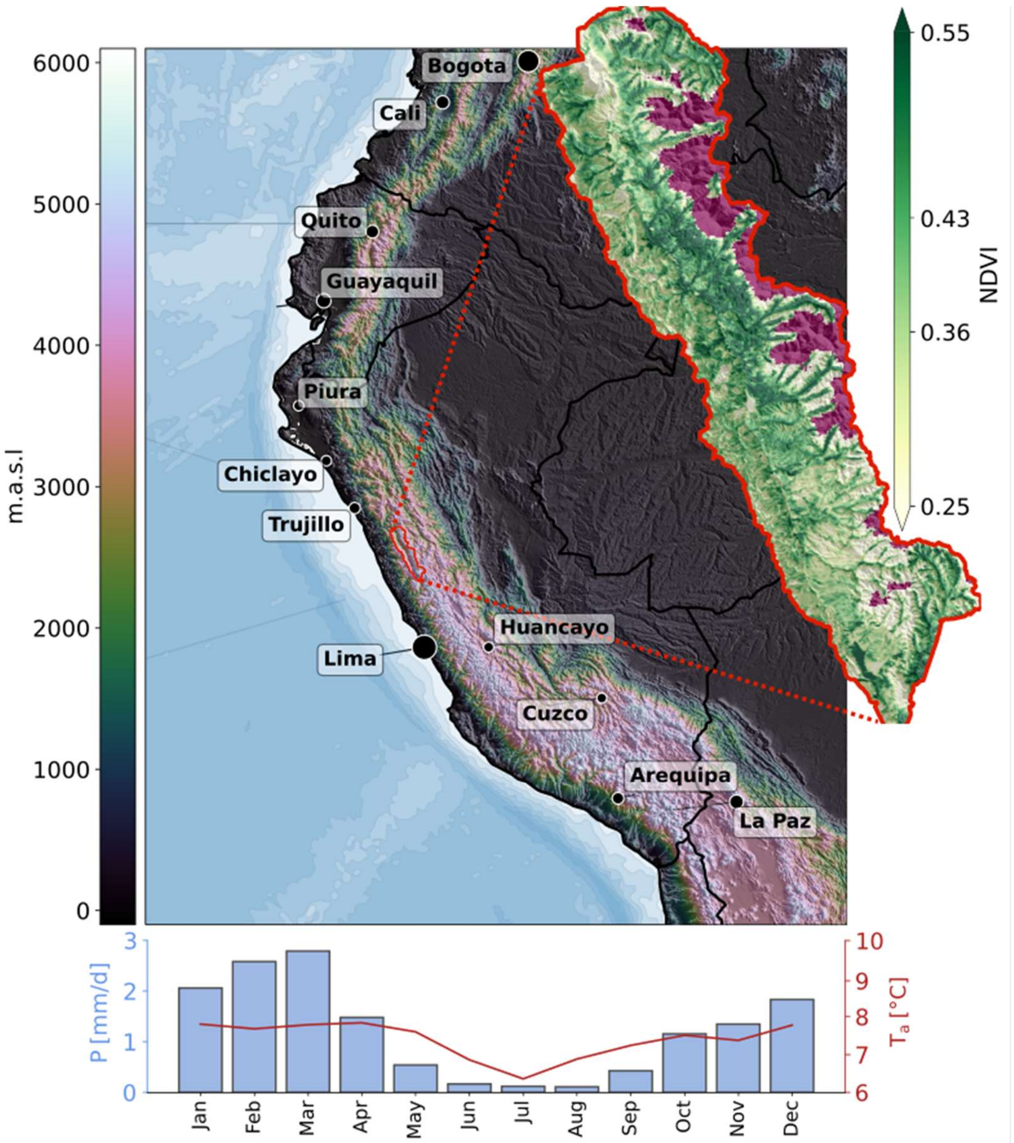
Thanks for your suggestion. Note that we generally combined Results and Discussion here (i.e. Section 3 Results & Discussion), therefore the 3 subsections (~ 3.1 Comparison, 3.2 Metric sensitivity and 3.3 Past/Future Trends) all contain parts which are discussion. We decided to do that as we realized when reviewing an earlier version of the manuscript among the co-authors that the paper is otherwise harder to follow. The conclusions' part on the other hand is strictly summarizing our main findings and giving a broader outlook. We hope you understand that we would therefore prefer not to move this quite specific part related to limitations into the conclusions part.

Comment #1.8

Figure 1: you could move the index plot of Rio Santa Basin with NDVI a little further away from the basic topography map. The Basin looks like it's a big part located within the eastern part of the map

We adjusted the position of the inset map so that the limits of both the x and y axes extend beyond the boundaries of the larger map. This ensures it is clear that the basin does not lie within the larger map. Additionally, we changed the perimeter color to red to draw the viewer's attention to the inset map. We acknowledge that another version like 2 independent panels of similar size might be another solution but as this would heavily increase the figure size, the readability of other features of the map would be reduced. We hope we found a satisfactory compromise with the new version.

New Figure 1:



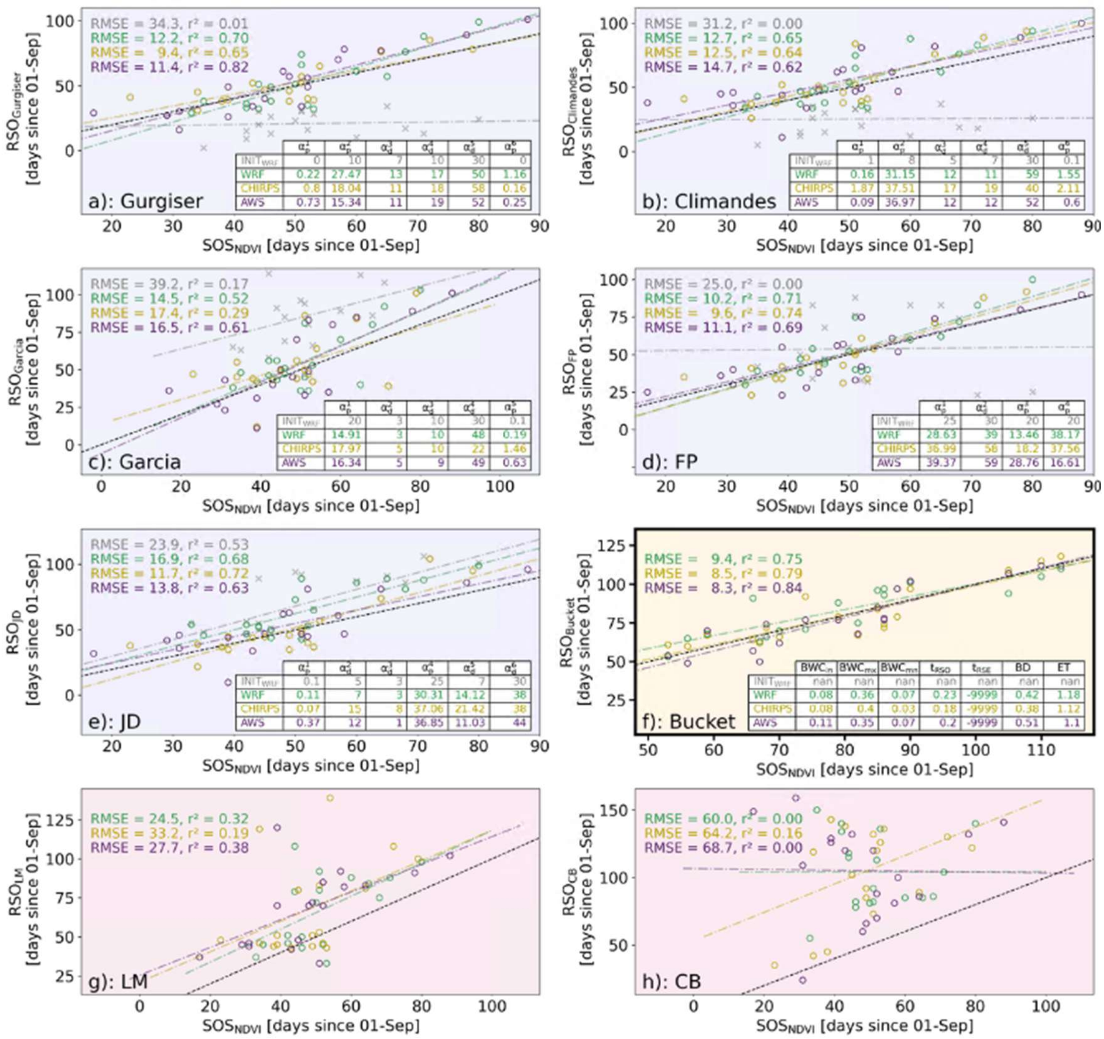
Comment #1.9

Figures 2 and 3: you could add a legend indicating different data sources and also different methodsdon't

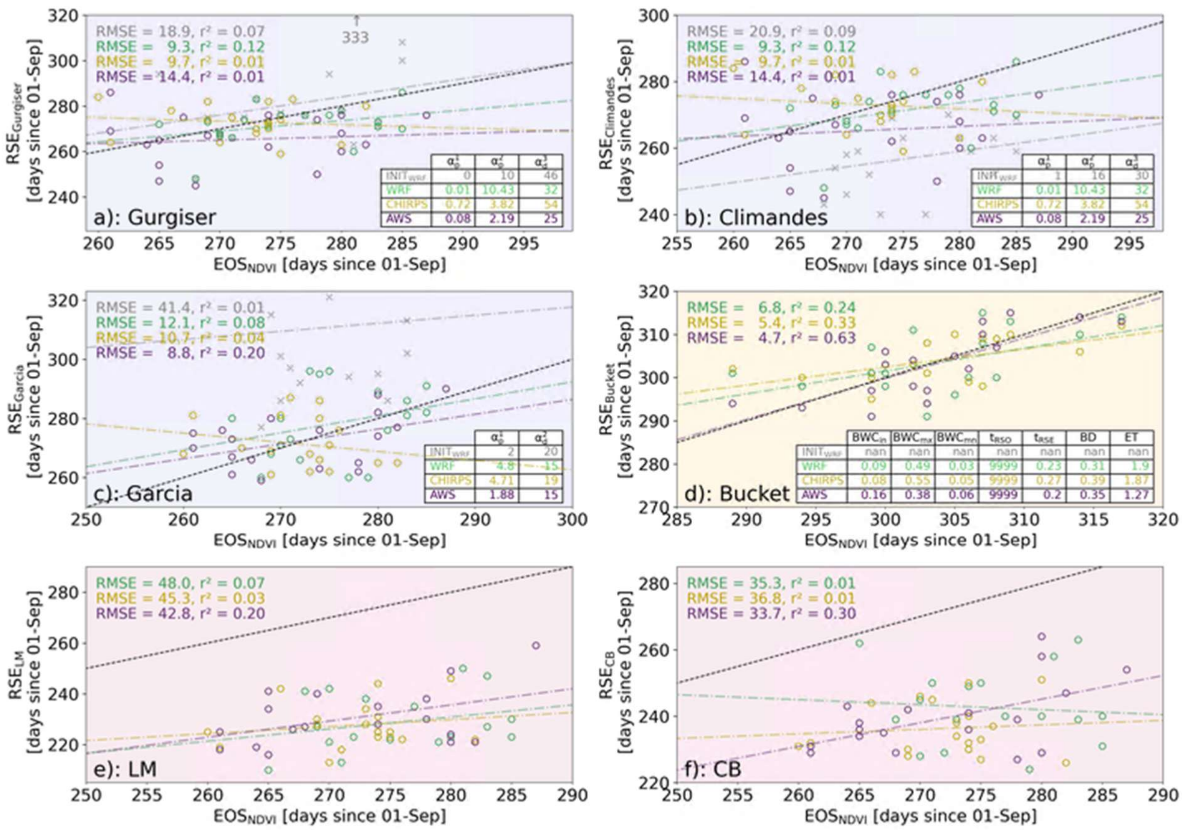
Thanks, following your recommendation we critically revisited the figure but realized that the information is already there. The table does contain the color-coding of the data source. A legend would just repeat this information and make the figure unnecessarily busy.

So far each panel was one different rainy season metric, however indicated only by the relatively small y-label, per category by the background color and indexed by a),b)....h). As we agree that this information was not very easy to find in such a busy figure, we now added the name of each metric to the letters.

New Figure 2:



New Figure 3:

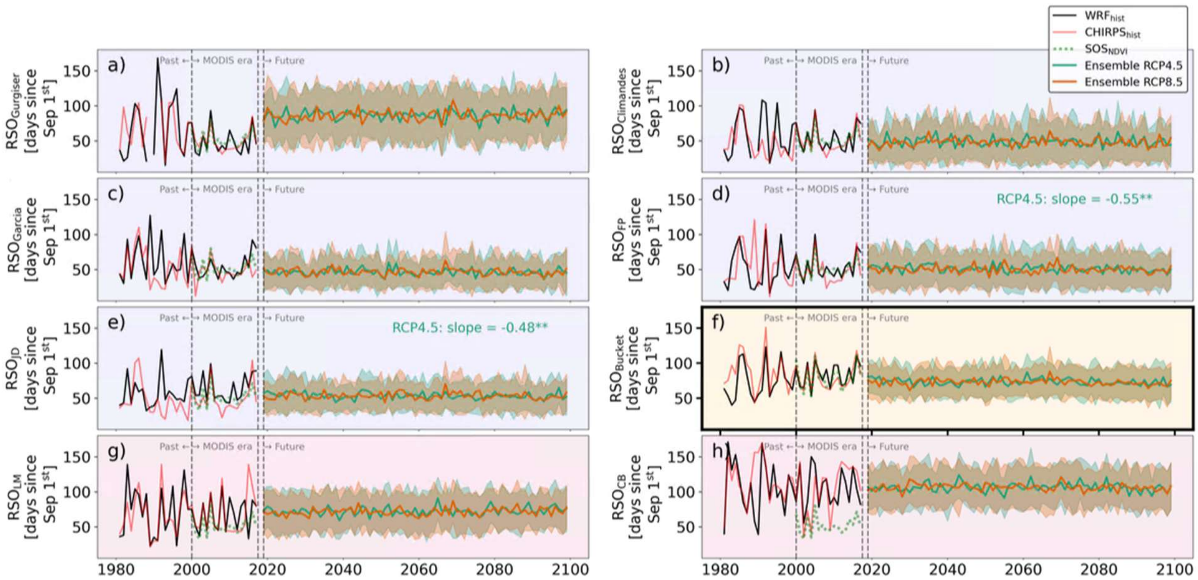

a): Gurgiser
b): Climandes
c): Garcia
d): Bucket
e): LM
f): CB
EOS_NDVI [days since 01-Sep]
RSE [days since 01-Sep]
RMSE = 18.9, r² = 0.07
RMSE = 9.3, r² = 0.12
RMSE = 9.7, r² = 0.01
RMSE = 14.4, r² = 0.01
RMSE = 20.9, r² = 0.09
RMSE = 41.4, r² = 0.01
RMSE = 12.1, r² = 0.08
RMSE = 10.7, r² = 0.04
RMSE = 8.8, r² = 0.20
RMSE = 6.8, r² = 0.24
RMSE = 5.4, r² = 0.33
RMSE = 4.7, r² = 0.63
RMSE = 48.0, r² = 0.07
RMSE = 45.3, r² = 0.03
RMSE = 42.8, r² = 0.20
RMSE = 35.3, r² = 0.01
RMSE = 36.8, r² = 0.01
RMSE = 33.7, r² = 0.30
WRF
CHIRPS
AWS

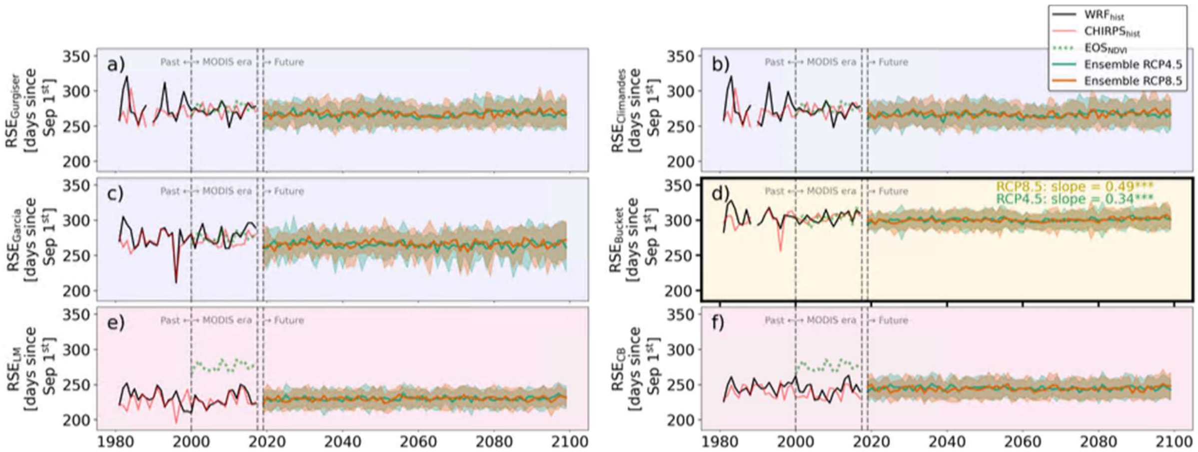
Comment #1.10

Figures 6 and 7: maybe add a legend showing different data sources and different scenarios

Thanks, we agree, the figure was indeed missing a legend. We added a legend to both of the figures according to your suggestions.

New Figures 6 & 7:





Comment #1.11

Some paragraphs have indentation while others , please keep them consistent

Apologies for the oversight, the revised manuscript has now consistent indentation.