Replies to reviews

"Benchmarking convection-permitting climate simulations for hydrometeorological characterizations: A comparative study of WRF-SAAG and observation-based products in Chile"

Sofía Segovia, Pablo A. Mendoza, Miguel Lagos-Zúñiga, Lucía Scaff, and Andreas Prein

We thank the three reviewers for their time, revision, and suggestions to our paper. We provide responses to each individual point below, and how we have addressed the comments of the reviewers. For clarity, comments are given in black italics, and our responses are given in plain blue text.

Reviewer #1

General Comment:

The manuscript offers an evaluation of daily total precipitation, maximum daily temperature, and minimum daily temperature variables simulated by the WRF-SAAG convective-permitting numerical model for hydrological applications. It compares these outputs with meteorological station measurements and simulations from the CR2MET and RF-MET products in the conterminous Chile. Furthermore, these variables are used as input data in a conceptual hydrological model (HBV-like) in order to assess the performance of the simulated daily streamflow against observed series in pristine catchments with low glacier cover.
The work addresses a fundamental topic in mountain regions, which is the use of numerical models to supplement the lack of meteorological measurements, especially for precipitation. However, the article has certain structural deficiencies that require manuscript ***major revisions****:*

1. The title and the body of the text do not specify which hydrological applications are being adressed to (e.g., flood risk, rain-on-snow events, water supply projections, among others - see a non-exhaustive list in Table 1 of Dominguez et al., 2024). The associated temporal scale and hydrological processes are also not defined.

We agree that the original submission lacked the specification of target hydrological applications. In the revised version, we have specified that the proposed framework is applied for simulating diverse hydrological signatures oriented to mean flow and extremes, including floods and droughts, through the analysis of hydrological signatures across continental Chile (see L126-128, L322-325 and Table 4 in the main text). Additionally, we incorporated the evaluation of key model outputs, including actual evapotranspiration (ET), soil moisture, and snow water equivalent (SWE).

Table 14: Hydrologic signatures.

Signature name	Signature description	Units
Q/P	Runoff ratio.	-
Q_{mean}	Mean daily discharge.	mm/d
Q_{JJA}	Mean daily discharge of winter days (JJA)	mm/d
Q_{DJF}	Mean daily discharge of summer days (DJF)	mm/d
Q_{95}	95% flow quantile (high flow)	mm/d

Q_5	5% flow quantile (low flow)	mm/d
$Q_{50\%}$	Day of year when 50\% of the flow volume is reached.	-
low_q_freq	Average frequency of low-flow days (< 0.2 times the mean daily flow).	d/yr
$high_q_freq$	Average frequency of high-flow days (> 9 times the median daily flow).	d/yr
low_q_dur	Average duration of low-flow events (number of consecutive days < 0.2 times the mean daily flow).	d
$high_q_dur$	Average duration of high-flow events (number of consecutive days > 9 times the median daily flow).	d
BFI	Ratio of mean daily baseflow to mean daily discharge.	-
$slope_fdc$	Slope of the flow duration curve (FDC) between log-transformed 33rd and 66th percentiles.	-
$stream_elas$	Sensitivity of streamflow to changes in precipitation at the annual scale.	-

Additionally, we have clarified that the evaluation of precipitation, air temperature and simulated discharge is conducted at the daily time scale, which is consistent with the temporal resolution of the streamflow records used, and serve as a basis for the dominant processes analyzed.

Finally, we have modified the title to reflect the study domain, and have replaced “hydrological applications” by “hydrometeorological characterizations”, which englobes both the assessment of climate model simulations, as well as the evaluation of hydrological model outputs:

“Benchmarking convection-permitting climate simulations for hydrometeorological characterizations: A comparative study of WRF-SAAG and observation-based products in Chile”

2. In line with the previous point, if by hydrological application the authors mean "hydrological models", thus: Which models are they? at what scales? representing which physical processes?

We have modified the text to clarify the following points: (i) only one hydrological model structure (TUW) is used, (ii) WRF-SAAG and hydrological model simulations are evaluated at the daily time scale, and (iii) the hydrological processes represented in the TUW model. We specified this in section 4.2 Implications for hydrological modeling (L304-309):

“To assess the potential of WRF-SAAG precipitation and temperature data for hydrological modeling applications, and to perform a comparative evaluation against CR2MET and RF-MEP, we configured and calibrated the TUW model (Parajka et al., 2007) in the 44 case study basins at a daily time scale. The TUW model is a conceptual, bucket-style rainfall–runoff model based on the structure of the HBV model (Bergström, 1976; Lindström et al., 1997). It simulates the catchment-scale water balance through snow, soil moisture, response, and routing modules and requires the specification of 15 parameters (Table 3).”

3. The scientific advancement is not made clear. This is also reflected in the poorly developed Conclusions section. What new facets does this work offer? What are the hydrological novelties? Where does this leave us?

This work is motivated by the recent publication of the hourly precipitation and temperature simulations for South America (Dominguez et al., 2024), which provide new meteorological input data for application-relevant research across the continent. One of the areas of interest for the WRF-SAAG community is the evaluation of convection-permitting model output for the characterization of extreme hydrometeorological events through process-based hydrological modeling. This study contributes to this objective by (1) presenting a an assessment of WRF-SAAG daily outputs along a hydroclimatically diverse Andean subdomain, which show comparable performance for replicating hydrometeorological extremes when contrasted against two widely used regional observation-based products – CR2MET (Boisier et al., 2023) and RF-MEP (Baez-Villanueva et al., 2020); and (2) mapping differences between WRF-SAAF, CR2MET and RF-MEP across different hydroclimates. Additionally, this paper expands on previous assessments of WRF precipitation outputs using in-situ observations (e.g., Ikeda et al., 2010; Mendoza et al., 2015) and offline hydrological modeling applications (e.g., Mendoza et al., 2016) conducted in other mountainous regions of the world. The results presented in this paper have shed light on the potential of WRF-SAAG for characterizing diverse hydrological signatures, oriented to average conditions, extreme events| and future avenues of research. Importantly, the work presented here has laid the foundation for using kilometer scale model data to study future hydrological changes in the Andes and interpret them based on the performance of such models under current climate conditions. This is essential for increasing resilience to climate change in Andean countries.

To emphasize this message, we edited the final paragraph of the Introduction (L115-127):
"Despite the WRF-SAAG simulations were not designed to reproduce every singular event precisely, their potential in representing historical weather variability at the daily time scale as well as the spatial and temporal characteristics of key hydroclimatic variables remains unknown. Therefore, the general objective of this study is to evaluate the ability of the WRF-SAAG simulations to represent daily precipitation and maximum and minimum temperatures over continental Chile, and to assess their potential as a meteorological forcing dataset for hydrological modeling, by comparing them with two observation-based meteorological products (CR2MET and RF-MEP). To achieve this, four specific objectives are addressed: (i) to identify the regions and climatic conditions where the main differences among the three meteorological datasets are observed; (ii) to evaluate the performance of daily meteorological time series from WRF-SAAG and observation-based meteorological products for precipitation, maximum temperature, and minimum temperature, against meteorological station observations; (iii) to assess the ability of the precipitation datasets to reproduce daily precipitation events of different magnitudes at meteorological stations; and (iv) to analyze the ability of the meteorological datasets to simulate runoff, reproduce hydrological signatures associated with mean flow and extreme runoff events and simulate other processes of the water cycle in a suite of catchments located across Chile, using a conceptual hydrological model."

4. There is an aporia (a logical contradiction) in the methodology that the authors themselves present. If, according to the manuscript under review, the WRF-SAAG model was not designed to simulate singular events but rather hydroclimatic features in South America (lines 111, 214, and 402), why does the article evaluate the performance of the simulations against daily events of precipitation, maximum temperature, and minimum temperature?

While the WRF-SAAG simulations were not designed as a reanalysis dataset (i.e., they do not include data assimilation or spectral nudging), they are nonetheless capable of reproducing historic weather patterns, particularly when these events are influenced by large-scale atmospheric forcing. Therefore, evaluating WRF-SAAG at the daily scale allows us to assess its general skill in representing historical weather variability and the spatial and temporal characteristics of key hydroclimatic variables, even though the system was not designed to reproduce every singular event precisely. However, given the spatial scale of the WRF simulations, they explicitly resolve convective precipitation, even when it is not fully measured, especially in high-elevation zones.

We clarified this point in the description of the dataset, Section 3.2 WRF-SAAG Simulations (L176-183):
"Since WRF-SAAG simulations are forced with ERA5 reanalysis, they are expected to reproduce large-scale precipitation events and explicitly resolve mesoscale processes such as convective precipitation. However, because the simulations were not designed as a reanalysis dataset (i.e., they do not include data assimilation or spectral nudging), they are not intended to reproduce every individual event with high precision.
WRF simulations are often used in hydrological studies including hydrological modeling (e.g., Wagner et al., 2016) and case study analysis (e.g., Naabil et al., 2017) among others. However, the WRF-SAAG simulations have not been exhaustively evaluated in South America, and their skill as meteorological input forcing for hydrological applications (e.g., Xie et al., 2025) remains unexplored."

5. Following the same line of reasoning, according to Dominguez et al. (2024), the WRF-SAAG runs use a reanalysis product (ERA5) as their initial and boundary conditions, which by definition represents the best snapshot of weather conditions at a specific place and time (Kalnay et al., 1996). Furthermore, Dominguez et al. (2024 - see Fig. 4) present an evaluation of singular events (peak precipitation hour), comparing simulations (Nov. 2018 to Mar. 2019) with station measurements, GPM-IMERG, and ERA5.

We acknowledge that WRF-SAAG uses ERA5 reanalysis data as its initial and boundary conditions, which constrains the simulation to realistic large-scale atmospheric states. While Dominguez et al. (2024) demonstrated the model's capacity to reproduce specific weather events for a particular period, our study has a broader objective: to evaluate WRF-SAAG's overall skill in representing daily hydroclimatic variability across Chile. Because the model is forced by ERA5, it can reproduce many large-scale weather events, although smaller-scale or convective processes may be displaced in space and/or time (please see our response to the previous comment).

Specific Comment:

Title
It should reflect which hydrological application the authors aim to address, ideally indicating the process and its temporal scale. I recommend incorporating the study area (i.e., continental Chile).

We have modified the title following the reviewer's recommendation:

"Benchmarking convection-permitting climate simulations for hydrometeorological characterizations: A comparative study of WRF-SAAG and observation-based products in Chile"

Abstract

L6. State explicitly that you will evaluate daily maximum and minimum temperature. This should be made clear in the abstract.

We have been explicit in the abstract that the minimum and maximum daily temperature is evaluated (L8-10):

"In particular, we analyze spatial differences among climatological estimates from the three products, and we assess the quality of daily precipitation and daily maximum and minimum temperature simulations using observations from meteorological stations over the period April/2001-March/2018"

L18. If you are going to use a hydrological model, please declare this in the sentence where you reference the methodology used.

The abstract was edited and now specifies that (L10-12):
"We also compare the potential of these climate datasets for hydrological modeling using a conceptual, bucket-style rainfall-runoff model, evaluating the results against streamflow observations, satellite remote sensing (SRS) products, and a snow water equivalent (SWE) reanalysis product."

Introduction

L37. You use the expression "Satellite-based products," which, for example, is not the case for CR2MET or RF-MET. Change 'to "Gridded products."

We appreciate the reviewer's comment. The introduction follows a logical structure that first discusses different sources of meteorological data, and the cited text intentionally refers to satellite-derived datasets to highlight their limitations in complex terrain and high-elevation regions, which is a relevant background for motivating the use of alternative products. To clarify this point, we revised the wording and added examples of satellite-based products (L35-39):

"Although spatial interpolation schemes based on in-situ measurements and topographic descriptors have been the standard approach (Daly et al., 1994, 2008; Clark and Slater, 2006; Isotta et al., 2014; Newman et al., 2015a, 2020), their accuracy heavily depends on the spatial density of stations, which is typically higher in valleys and populated areas. **On the other hand, satellite-based products (e.g., TRMM and CMORPH)** have become an attractive alternative in sparse data regions …"

L62/63. The sentence should conclude with at least one citation.

We have added the reference Prein et al., 2023 and Lundquist et al., 2019 associated with the lines (L63-66):

"As a result, high-resolution atmospheric models can perform similarly (e.g., Prein et al., 2023) or even outperform (e.g., Lundquist et al., 2019) gridded observational products in capturing total precipitation over complex terrain. Additionally, these models offer a physically consistent and

spatially continuous representation of precipitation, making them a viable alternative for hydrological modeling applications."

L72. The authors state that there is "little" information on the performance of the CR2MET and RF-MET products; however, in L78 (in the same paragraph), they provide some numbers and cite an article that has already evaluated their performance. There is a logical contradiction in the writing of this paragraph.

We appreciate this suggestion. With "limited information", we aimed to state that no systematic assessments of CR2MET and RF-MEP daily precipitation have been conducted using ground measurements as the observational reference. We have reworded the text in that section to clarify this point:

"While these observation-based datasets have been widely used for different applications (e.g., Hernandez et al., 2022; Murillo et al., 2022 in the case of CR2MET, and Chen et al., 2022; Al-Saeedi et al., 2024 in the case of RF-MEP), no systematic assessments of CR2MET and RF-MEP daily precipitation have been conducted using ground measurements as the observational reference.".

Further, the study cited (Baez-Villanueva et al., 2021) did not evaluate the products against daily station observations; instead, it presented a comparison of annual precipitation amounts retrieved from different products across macro-regions, with the aim to assess the impact of the choice of forcing dataset on the regionalization of hydrological model parameters.

Study domain

Fig. 1. Where does the precipitation for the catchments used to calculate the runoff coefficient come from? A gridded product? Which one?

The study area figure (Fig. 1) now includes mean annual runoff, instead of the runoff coefficient. Initially, the runoff coefficient was calculated using the CR2MET gridded precipitation product. In the current version of the manuscript, the runoff ratio (together with a set of other hydrological signatures) was calculated using the three datasets as part of the hydrological evaluation.

L140. Same comment.

Corrected.

Methods - General Comments

If both CR2MET and RF-MET were constructed using station measurements, does it make sense to compare their performance at those same locations? I agree that the reported errors can be used as a reference for the performance of WRF-SAAG, but many lines of text are wasted on the analysis of these two products. It would be more fruitful to calculate the difference between the grids (WRF-SAAG vs. CR2MET and RF-MET) to visualize substantial differences. Furthermore, the introduction emphasizes the lack of measurements in high-mountain areas, which further highlights the importance of performing this grid-to-grid comparison; otherwise, the potential of WRF-SAAG remains very limited.

We appreciate the comment and, after discussing this point, we decided to keep the assessment of CR2MET and RF-MEP against station observations because it provides a better understanding of

how gridded data represents observational data, even if that data information was considered in the production of the gridded dataset. Further, we find relevant to the hydrological community, that tries to apply these gridded data to modelling tools, the following points: (1) the results demonstrate that, despite the precipitation products combine reanalysis data and ground observations, they were not designed to match station measurements perfectly; and (2) the reported errors are valuable for the hydrometeorology community since this is the first assessment against in-situ observations at the daily time scale.

Additionally, we appreciate and agree with the reviewer's recommendation regarding a grid-to-grid comparison. Therefore, we now include figures showing differences between products for different temporal aggregations, with the aim of identifying areas with scarce or no available observations. The precipitation and temperature results are shown in Figures 2 and 3 (results from section 5.1 Comparison between gridded products), attached below:

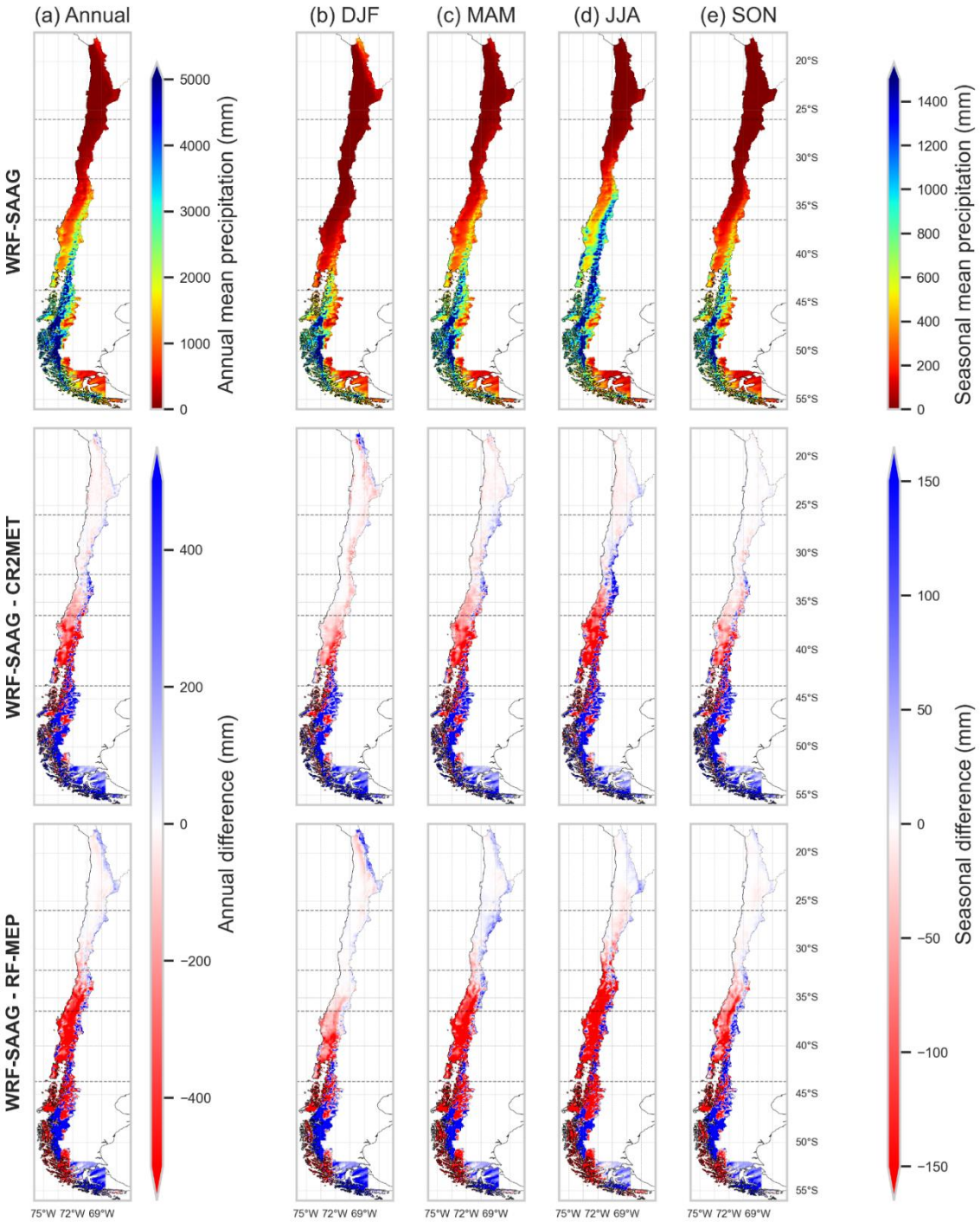


Figure 12: Spatial distribution of mean precipitation simulated by WRF-SAAG and precipitation differences relative to CR2MET and RF- MEP over Chile (april/2001 - march/2018). Columns show (a) annual means and seasonal means for (b) summer, (c) autumn, (d) winter, and (e) spring. The top row displays mean precipitation from WRF-SAAG, while the middle and bottom rows show precipitation differences between CR2MET and WRF-SAAG, and between RF-MEP and WRF-SAAG, respectively. Horizontal dashed lines indicate the boundaries of the five latitudinal macrozones: Far North, Near North, Central Chile, Southern, and Austral.

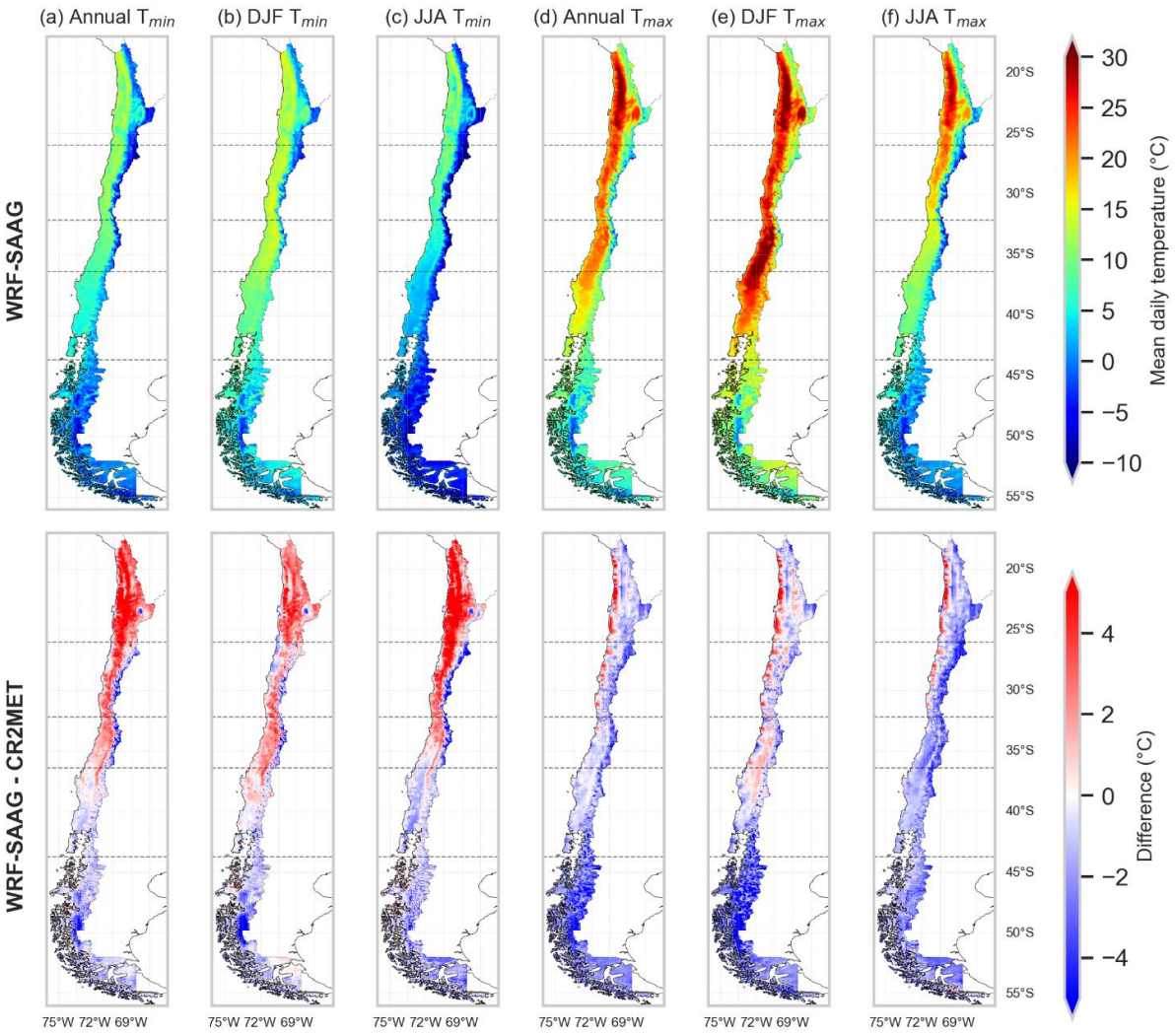


Figure 23: (Top row) Spatial distribution of climatological averages for annual, summer and winter daily (a-c) minimum (Tmin) and (d-f) maximum (Tmax) air temperatures over Chile derived from the WRF-SAAG dataset, and (bottom row) differences relative to CR2MET (april/2001 - march/2018). Horizontal dashed lines indicate the boundaries of the five latitudinal macrozones: Far North, Near North, Central Chile, Southern, and Austral.

The hydrological application referred to by the authors in the title is hydrological modeling. Even so, this is very general and therefore weak. Models are subject to multiple sources of uncertainty, and parameter calibration can, in turn, ***yield correct results for the wrong reasons****, especially when the only facet being evaluated is the catchment streamflow (Beven 2006; Kirchner 2006).*
If the authors decide to incorporate the broad area of hydrological models as their application in a revised version, they should define the working scales, processes, and model types from the beginning *(this must be reflected in the methodology). Using a numerical model for the sole purpose of running it does not reveal new advancements in hydrology. For example, does it make sense to*

apply a temperature-index model, like the TUWmodel, in the Near-North and Far-North macrozones where sublimation can account for more than 70% of the seasonal snowpack (e.g., Ayala et al. 2023)?

We appreciate the reviewer's feedback. In the revised manuscript, we have expanded the hydrological evaluation considerably by (i) comparing a broader set of hydrological signatures to assess the model's capability to replicate different runoff responses, and (ii) including an assessment of evapotranspiration (ET), soil moisture and SWE. We stress, however, that the examination of sources of uncertainty beyond the choice of meteorological forcing dataset is out of the scope of this study, and we have proposed related topics for future work (L657-659):

"Future comparative assessments between WRF-SAAG and other gridded products could incorporate the effects of parameter equifinality on hydrological model simulations (e.g., Muñoz-Castro et al., 2023), examine model structural uncertainty (e.g., Saavedra et al., 2022) …"

Regarding the working scales, processes and model types, we specify them at the beginning of section 4.2 Implications for hydrological modeling (L304-309):
"To assess the potential of WRF-SAAG precipitation and temperature data for hydrological modeling applications, and to perform a comparative evaluation against CR2MET and RF-MEP, we configured and calibrated the TUW model (Parajka et al., 2007) in the 44 case study basins at a daily time scale. The TUW model is a conceptual, bucket-style rainfall–runoff model based on the structure of the HBV model (Bergström, 1976; Lindström et al., 1997). It simulates the catchment-scale water balance through snow, soil moisture, response, and routing modules and requires the specification of 15 parameters (Table 3)."

Additionally, we recognize that the TUW model has several limitations in terms of process representations, especially in arid regions. Hence, we have modified the text in the Discussion section 6.4 Limitations and future work, as follows (L652-656):
"Fourthly, WRF-SAAG simulates all atmospheric and land surface variables with physical consistency, making it a powerful resource for conducting offline numerical experiments with complex process-based hydrological models, which require not only precipitation and temperature but also wind speed, humidity, air pressure, and radiative fluxes. The performance of such models could be thoroughly evaluated in terms of process fidelity, especially in arid areas given the TUW model limitations regarding sublimation processes and baseflow generation mechanisms."

Finally, one of the specific goals of this research was to test the applicability of WRF-SAAG simulations for hydrological applications starting with simple models, as it had not yet been tested in continental Chile. To highlight the specific goals, we explicitly stated them in the introduction (L120-127) to settle the readers' expectations:
"… four specific objectives are addressed: (i) to identify the regions and climatic conditions where the main differences among the three meteorological datasets are observed; (ii) to evaluate the performance of daily meteorological time series from WRF-SAAG and observation-based meteorological products for precipitation, maximum temperature, and minimum temperature, against meteorological station observations; (iii) to assess the ability of the precipitation datasets to reproduce daily precipitation events of different magnitudes at meteorological stations; and (iv) to analyze the ability of the meteorological datasets to simulate runoff, reproduce hydrological signatures associated with mean flow and extreme runoff events and simulate other processes of the water cycle in a suite of catchments located across Chile, using a conceptual hydrological model."

Methods - Regarding the Subsections

Section 4.1 should be entitled "Evaluation of daily precipitation and maximum and minimum temperatures."

We thank the reviewer for this suggestion. The section with this methodology is now titled “4.1.2 Evaluation of daily precipitation and daily maximum and minimum temperatures”.

In the introduction, the authors emphasize that in mountain areas (e.g., Chile), most stations are located at low altitudes and are scarce, which leverages the use of high-resolution dynamic models (e.g., WRF-SAAG) to capture total precipitation patterns along mountain ranges like the Andes. However, in the proposed methodology, they evaluate the performance of WRF-SAAG using station measurements, the majority of which are located below 3000 m a.s.l. and with a low-density network in the Cordillera.

We acknowledge the limitations associated with the sparse and low-altitude distribution of meteorological stations in mountainous regions such as the Andes. However, the evaluation of WRF-SAAG and the other gridded products was conducted using the maximum number of available quality-controlled stations, which represents the best observational information currently accessible for continental Chile.

To complement the station-based evaluation and provide additional insights into the representation of those fields in ungauged regions, the revised manuscript has incorporated a direct comparison of the spatial fields from the three meteorological products (CR2MET, RF-MEP and WRF-SAAG), in Figures 2 and 3 (see previous response). This gridded intercomparison highlights the main differences among products across elevation gradients and complex terrain, particularly in the Andes.

Perhaps it would be more interesting, given that CR2MET and RF-MET are built with station measurements, to conduct an analysis of the differences (quarterly?) between grids so that the reader can visualize latitudinal and altitudinal discrepancies. In which quarter are the differences smaller (larger)? Why?

We agree that a comparison between the gridded products would provide valuable insights, particularly because CR2MET and RF-MEP are derived from station observations while WRF-SAAG originates from a dynamical model. As noted in the previous response, the revised manuscript includes a direct intercomparison of the gridded fields from the three meteorological products to highlight latitudinal and altitudinal differences across Chile. Different temporal aggregations have also been evaluated to identify the main differences among the products.

In Figure 1.d, the authors show the reference runoff coefficient for each catchment. After clarifying the source of the precipitation, what values does this coefficient yield when using precipitation simulated by WRF-SAAG? CR2MET? RF-MET? How do the time series and the climatological value for each catchment compare? Given the hydroclimatic regime of each catchment, what values would be logical to expect? Are the absolute values of total precipitation reasonable?

In the revised version of the manuscript, we update Figure 1 (study area, attached below), and panel d now shows the mean annual runoff. Originally, the figure displayed the runoff ratio using CR2MET

data as a reference. However, after expanding the objectives of the paper, the runoff ratio was recalculated using three available products: WRF-SAAG, CR2MET, and RF-MET, together with a set of additional hydrological signatures.

After the general comments, if you still wish to incorporate the hydrological model, the following lines should be taken into account,

- *In section 4.2, the hydrological model is poorly presented. TUWmodel is one of the many versions of the original HBV. First, present and cite HBV, then TUWmodel.*

  The manuscript has been revised to first present the original HBV model, including appropriate citations (L304-309):
  "To assess the potential of WRF-SAAG precipitation and temperature data for hydrological modeling applications, and to perform a comparative evaluation against CR2MET and RF-MEP, we configured and calibrated the TUW model (Parajka et al., 2007) in the 44 case study basins at a daily time scale. The TUW model is a conceptual rainfall–runoff model based on the structure of the HBV model (Bergström, 1976; Lindström et al., 1997). It simulates the catchment-scale water balance through snow, soil moisture, response, and routing modules and requires the specification of 15 parameters (Table 3)."

- *What daily temperature value do you use as input data for the model? Minimum and maximum? Daily mean? Up to this section, you have stated that you are evaluating the daily maximum and minimum temperatures.*

  For the hydrological model, the daily mean temperature calculated as the average of the daily minimum and maximum temperatures, is used as input. The manuscript has been revised to explicitly state this in Section 4.2.1 Model configurations (L315-318):
  "The model's required input forcing variables are daily precipitation, temperature, and potential evapotranspiration (PET). Hence, we extracted precipitation time series from the WRF-SAAG, CR2MET, and RF-MEP datasets, and mean temperature time series from WRF-SAAG and CR2MET – computed as the average of daily maximum and minimum temperatures – datasets to compare different combinations of forcing dataset…"

- *Figure 6.c. Except for the river in the Southern macrozone, the model is incapable of simulating the observed annual cycle of the rivers. Is this because WRF-SAAG does not capture the seasonality of precipitation? Could it be that the model does not adequately simulate the dominant physical processes?*

  The reviewer makes a good point. The TUW model may fail to replicate observed annual runoff cycles during the evaluation period (Figure 8) due to (i) poor parameter transferability associated with differences in hydroclimatic conditions between the calibration (which includes wet, dry and normal years) and the evaluation (which includes the central Chile megadrought) periods – which is a well-documented problem in hydrological modeling (e.g., Merz et al., 2011; Fowler et al., 2018); (ii) differences in mean monthly precipitation among climate datasets (e.g., up to 100 mm/month during winter between RF-MEP and the other datasets at Emperador Guillermo) that the hydrological model is not able to compensate for through parameter calibration (e.g., Mendoza et al., 2016); and (iii) structural deficiencies of

the TUW model. We have added these points in the Discussion section (L629-641), which already includes a discussion on the limitations of the TUW model for our study domain.

Final suggestion
After these comments are addressed, I look forward to revisiting the Methods, Results, Discussion, and a richer, more substantive Conclusions section.

We thank the reviewer for the constructive feedback and valuable suggestions. All points raised have been addressed in the revised manuscript, including clarifications in the Methods, Results, and Discussion sections, as well as an expanded and more substantive Conclusions section.

Reviewer #2

This manuscript evaluates precipitation and temperature from WRF-SAAG, CR2MET, and RF-MEP against in-situ stations across continental Chile. The topic is important and relevant. Below are my major and minor comments.

Major Comment:

1. Since WRF-SAAG is not expected to reproduce specific events, I am unclear about the motivation for this evaluation. Is WRF-SAAG typically used as forcing for hydrological models to obtain hydrological simulations? If so, is this recommended practice? If not, the value of evaluating individual events is limited. In that case, comparing climatological characteristics across datasets might be more meaningful.

The motivation for our evaluation is not to assess event-scale accuracy, but rather to examine the skill of the WRF-SAAG dataset in representing hydroclimatic variability at the daily time scale. This assessment constitutes a necessary step to determine whether WRF-SAAG provides sufficiently realistic precipitation and temperature fields to be considered as a potential meteorological forcing for subsequent hydrological modeling experiments. In the revised version of the manuscript, we have made this motivation explicit in the Introduction section (L215-227):

"Despite the WRF-SAAG simulations were not designed to reproduce every singular event precisely, their potential in representing historical weather variability at the daily time scale as well as the spatial and temporal characteristics of key hydroclimatic variables remains unknown. Therefore, the general objective of this study is to evaluate the ability of the WRF-SAAG simulations to represent daily precipitation and maximum and minimum temperatures over continental Chile, and to assess their potential as a meteorological forcing dataset for hydrological modeling, by comparing them with two observation-based meteorological products (CR2MET and RF-MEP). To achieve this, four specific objectives are addressed: (i) to identify the regions and climatic conditions where the main differences among the three meteorological datasets are observed; (ii) to evaluate the performance of daily meteorological time series from WRF-SAAG and observation-based meteorological products for precipitation, maximum temperature, and minimum temperature, against meteorological station observations; (iii) to assess the ability of the precipitation datasets to reproduce daily precipitation events of different magnitudes at meteorological stations; and (iv) to analyze the ability of the meteorological datasets to simulate runoff, reproduce hydrological signatures associated with mean flow and extreme runoff events and simulate other processes of the water cycle in a suite of catchments located across Chile, using a conceptual hydrological model."

Additionally, regarding the application of WRF-SAAG simulations in hydrological modeling, the following text was incorporated into Section 3.2 WRF-SAAG simulations to clarify their intended use and limitations as meteorological forcing (L176-183):
"Since WRF-SAAG simulations are forced with ERA5 reanalysis, they are expected to reproduce large-scale precipitation events and explicitly resolve mesoscale processes such as convective precipitation. However, because the simulations were not designed as a reanalysis dataset (i.e., they do not include data assimilation or spectral nudging), they are not intended to reproduce every individual event with high precision.
WRF simulations are often used in hydrological studies including hydrological modeling (e.g., Wagner et al., 2016) and case study analysis (e.g., Naabil et al., 2017) among others. However, the WRF-SAAG simulations have not been exhaustively evaluated in South America, and their skill as meteorological input forcing for hydrological applications (e.g., Xie et al., 2025) remains unexplored."

2. As this is an evaluation study, the manuscript would benefit from clearer take-home messages. For example: Which dataset should be preferred under certain conditions or in specific regions? Are there areas where none of the datasets are recommended?

We thank the reviewer for the suggestion. We added the following paragraph in Section 7, Summary and Conclusions, which includes the main take-home messages (L678-687):

"The main outcome of this work is that, despite substantial spatial differences in climatological averages and despite not being designed to replicate individual precipitation events in space and time, WRF-SAAG delivered performance comparable to observation-based products at the daily time scale. The hydrological modeling results showed similar streamflow performance metrics and SWE model simulations for the three products, with differences in simulated hydrological signatures, ET and soil moisture that varied with macrozone. For streamflow simulations, the results suggest that in the Southern macrozone all forcing combinations provide reasonable performance. In contrast, in the Austral macrozone, configurations using RF-MEP precipitation as input tend to provide lower performance compared to those forced with WRF-SAAG or CR2MET. In the Far and Near North macrozones, the results do not allow us to conclude that WRF-SAAG-driven simulations are unsuitable; rather, they suggest that testing alternative hydrological model structures may be necessary to better represent hydrological processes in these regions."

3. Figures 3–5: Consider presenting KGE values as boxplots, since the authors mention medians of the KGE frequently. The spatial maps make it difficult to assess the median.

We thank the reviewer for this suggestion. To better visualize both the spatial distribution of KGE_T and its statistical distribution, we added a panel showing boxplots of KGE_T to all figures that evaluate gridded products against meteorological stations. An example is provided below in Figure 5, which presents the evaluation results for daily precipitation amounts during summer:

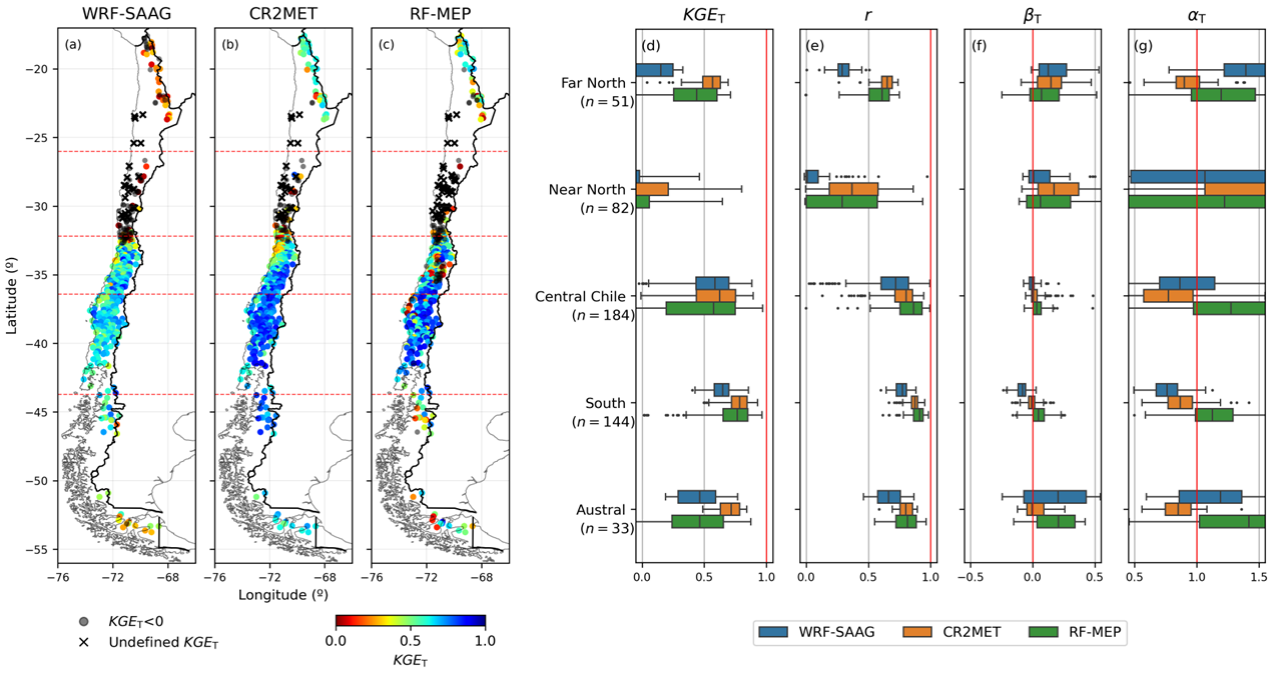


Figure 5: (a)–(c) Spatial distribution of Tang's Kling-Gupta Efficiencies (KGE_T) for summer (DJF) daily precipitation retrieved from WRG- SAAG, CR2MET, and RF-MEP (period April/2001–March/2018) across continental Chile, using ground-based observations as the reference. Red dashed lines indicate the limits of the macrozones analyzed. (d) to (g) boxplots with KGE_T, the Pearson correlation coefficients (r), bias term (β_T), and variability ratios of Tang's Kling-Gupta Efficiency (α_T) obtained with the products examined for each macrozone, with the red line indicating the optimal value. Each boxplot comprises results for all the stations within a specific macrozone. The boxes correspond to the interquartile range (IQR, i.e., 25th and 75th percentiles); the vertical line in each box is the median, and the whiskers extend to the ±1.5 · IQR of the ensemble of stations.

4. Line 285: When precipitation events > 1 mm/day are considered, CR2MET performs best overall, but for thresholds > 5, > 10, and > 20 mm/day, RF-MEP performs best. Could this pattern be influenced by the precipitation distribution? For example, are events between 1–5 mm/day dominant when precipitation events > 1 mm/day? If so, it might cause RF-MEP's superior performance at higher intensities to be diluted when all events > 1 mm/day are included.

The reviewer makes a good point. We tested his/her hypothesis by computing - for each macrozone and for the different climate datasets (i.e., station observations, WRF-SAAG, CR2MET and RF-MEP) the fraction of events (i.e., combinations of station locations and days) with P > 1 mm/d that lie within the categories 1–5 mm/d, 5-10 mm/d, 10-20 mm/d, and > 20 mm/d (Table R1). The results show that the events between 1–5 mm/d dominate when precipitation events > 1 mm/d, as the reviewer points out. Considering this information, we have modified the text in the first paragraph of section 5.2.1 Ability to simulate precipitation events as follows (L406-412):

"In general, RF-MEP achieves the highest POD values for events >5 mm/d, >10 mm/d and >20 mm/s, particularly in Central Chile ($POD_{RF\text{-}MEP} \geq 0.65$). Nevertheless, in the Near North and Central Chile, RF-MEP's superior performance at higher intensities is diminished when events with P >1 mm/d are

considered in the evaluation. This result may be attributed to precipitation distribution, since the fraction of events (i.e., combinations of station locations and days) between 1–5 mm/d dominate among all events with P > 1 mm/d (not shown). On the other hand, CR2MET achieves the highest POD values (medians ≥ 0.85) across all study zones for events >1 mm/d."

Table R1: Fraction of events with P > 1 mm/d that fall within different categories

Zone	Dataset	1-5 mm	5-10 mm	10-20 mm	> 20 mm
Far North	Station	61.8%	24.4%	11.8%	2.0%
Far North	WRF-SAAG	57.5%	22.2%	15.6%	4.7%
Far North	CR2MET	72.3%	23.0%	4.5%	0.3%
Far North	RF-MEP	63.3%	23.0%	11.6%	2.1%
Near North	Station	45.3%	20.0%	17.2%	17.5%
Near North	WRF-SAAG	37.7%	20.1%	20.4%	21.8%
Near North	CR2MET	57.2%	19.3%	13.1%	10.4%
Near North	RF-MEP	37.1%	18.8%	18.8%	25.3%
Central Chile	Station	35.5%	19.3%	19.8%	25.5%
Central Chile	WRF-SAAG	34.3%	20.1%	21.7%	23.9%
Central Chile	CR2MET	40.1%	21.1%	19.5%	19.2%
Central Chile	RF-MEP	30.9%	16.2%	18.4%	34.5%
South	Station	33.2%	20.9%	22.4%	23.4%
South	WRF-SAAG	38.3%	21.6%	21.9%	18.2%
South	CR2MET	35.8%	22.6%	22.1%	19.5%
South	RF-MEP	34.0%	19.4%	21.2%	25.4%
Austral	Station	55.3%	19.9%	15.1%	9.7%
Austral	WRF-SAAG	57.0%	23.2%	14.2%	5.6%
Austral	CR2MET	61.5%	18.8%	12.5%	7.1%
Austral	RF-MEP	52.3%	18.6%	14.6%	14.6%

4. Line 262: Please specify the formula used to estimate PET and list the variables involved.

PET was calculated using the formulation proposed by Oudin et al. (2005), as implemented in the R package airGR (Coron et al., 2023), considering the latitude of the centroid of each elevation band. In the revised manuscript, the exact equation and all input variables have been included in the Section 4.2.1 Model configurations.

Minor Comments:

6. *Abstract: WRF-SAAG data are available until 2021. Why was the evaluation restricted to 2001–2018?*

The analysis period (2001–2018) was limited by the availability of the RF-MEP dataset, which covers the period 1983–2018 (Baez-Villanueva et al., 2021). We explicitly justify the choice of analysis period at the end of section 4.1 (L251-253):

"Importantly, all the assessments presented in this study were conducted for the period April/2001-March/2018, for which the three forcing datasets have available data (Table 1)."

7. Line 120: Consider adding polygons to the maps to delineate the four geographical units, which would help international readers.

In the current version of the figure, the four geographical units (macrozones) are delineated by horizontal red lines, and the name of each macrozone is indicated on the left side of the panel. We have ensured that this is more clearly described in the figure caption to guide readers' interpretation.

8. The first part of the abstract reads like that WRF-SAAG was evaluated against CR2MET and RF-MEP, but in fact all three datasets were compared against in-situ observations. Please revise for clarity.

We have revised the abstract to clarify that all three datasets—WRF-SAAG, CR2MET, and RF-MEP—were evaluated against in-situ station observations (L6-10):

"Here, we evaluate WRF-SAAG outputs across continental Chile and compare them with two widely used gridded meteorological products — CR2MET and RF-MEP — which combine reanalysis data with in-situ measurements. In particular, we analyze spatial differences among climatological estimates from the three products, and we assess the quality of daily precipitation and daily maximum and minimum temperature simulations using observations from meteorological stations over the period April/2001-March/2018."

9. Line 135: What do "daily observations" refer to? Are they discharge (Q) and precipitation (P)? 1Does this also include temperature?

We refer to daily discharge observations. We have modified the text to clarify this (L147-148):

"The selected catchments fulfill the following criteria: (i) at least 80% coverage of daily discharge observations during the period April/2001 – March/2018…"

10. Line 210: It seems the reference should be "1b (1c)" instead of "1c (1d)."

Thanks for catching this! We have corrected the references in the revised manuscript.

11. Line 255: The DEM data should be introduced in the Data section.

In response to this comment, we have restructured the Data section, which is now titled "Datasets" and includes the following subsections:

3 Datasets
3.1 Ground-based observations
3.2 Climate datasets
3.2.1 WRF-SAAG simulations
3.2.2 CR2MET
3.2.3 RF-MEP
3.3 Digital Elevation Model

3.4 Remote sensing and reanalysis reference datasets
3.4.1 MOD16A2GF evapotranspiration
3.4.2 RemoC3S satellite soil moisture
3.4.3 SWE

12. Equation 8: Please clarify what Q and 1/Q represent in the KGE' formulation.

Q is daily discharge, 1/Q is the reciprocal of daily discharge. We have modified the text as follows to explain this (L336-337):

"where Q is daily discharge, 1/Q is the reciprocal of daily discharge, and KGE′ is the modified Kling-Gupta efficiency (KGE', Kling et al., 2012)."

We will also add the following lines of text in section 4.2.2 Calibration (L315-316):

"Finally, KGE'(Q) (KGE'(1/Q)) denotes the KGE obtained from comparing daily time series of simulated and observed Q (1/Q)".

13. Line 220: When referring to precipitation events > 1 mm/day, do these include events exceeding 5, 10, and 20 mm/day, or are they limited to events between 1–5 mm/day? Please clarify.

In our analysis, we are including all events with a magnitude larger than the specified threshold, without limiting them to the next lower limit. We have clarified this in the revised version to avoid ambiguity (L275-276):

"We used metrics formulated from contingency tables to assess the ability of the datasets to replicate historically observed daily precipitation events exceeding thresholds of 1, 5, 10, and 20 mm/d (i.e., events >1, >5, >10, and >20 mm/d)."

14. Line 297: This section appears to discuss only summer results. Why are "all seasons" mentioned here?

The intention was to highlight that, although Figure 5 focuses on summer (DJF), the Southern macrozone consistently shows high KGE_T median values (≥ 0.65) not only in summer but also across the other seasons.

We have modified the text to clarify this (L421-423):

"Figure 5 shows the spatial distribution of KGE_T (and its components) for summer (DJF) daily precipitation estimates from WRF-SAAG, CR2MET, and RF-MEP. The highest KGE_T values in summer are obtained in the Southern macrozone, where median KGE_T values reach ≥ 0.65, and remain at or above this level throughout all seasons."

Reviewer #3

General Comment:

This study presents a comprehensive evaluation of the high-resolution, long-term WRF-SAAG climate simulation (2000-2021) against station observations and two gridded meteorological products (CR2MET and RF-MEP) over Chile. The subsequent use of WRF-SAAG outputs to drive a hydrological model (TUW) successfully demonstrates the dataset's utility for hydrological applications. The paper highlights the good performance of WRF-SAAG in capturing precipitation and temperature, particularly in complex mountainous terrain where observational records are sparse. This is a valuable contribution to the regional climate modeling and hydrology communities.

The manuscript is well-structured and the analysis is thorough. However, I have several suggestions for improvement that I believe will enhance the clarity, presentation, and overall impact of the paper. My main suggestions focus on making the writing more concise, improving the presentation of results and data, and expanding the discussion to better guide potential users of these datasets.

Specific Comment:

Abstract

- *The abstract, and the paper in general, could be more concise. Please review for opportunities to shorten sentences and state the main findings more directly.*

  We have revised the abstract and the rest of the manuscript to shorten the text and advocate for conciseness. The abstract now reads as follows:

  Over the last few years, significant progress has been made in the development of convection-permitting climate models (CPCMs), especially in improving precipitation estimates in regions with complex terrain. Recently, the South American Affinity Group (SAAG) developed a novel high-resolution dataset —hereafter referred to as the WRF-SAAG dataset —by dynamically downscaling the ERA5 reanalysis using the Weather Research and Forecasting (WRF) model over South America for the period 2000–2021. Here, we evaluate WRF-SAAG outputs across continental Chile and compare them with two widely used gridded meteorological products — CR2MET and RF-MEP — which combine reanalysis data with in-situ measurements. In particular, we analyze spatial differences among climatological estimates from the three products, and we assess the quality of daily precipitation and daily maximum and minimum temperature simulations using observations from meteorological stations over the period April/2001-March/2018. We also compare the potential of these climate datasets for hydrological modeling using a conceptual, bucket-style rainfall-runoff model, evaluating the results against streamflow observations, satellite remote sensing (SRS) products, and a snow water equivalent (SWE) reanalysis product. We found that, detection accuracy varied within each Chilean macrozone, with worse performance in the Far North (17.5° – 26°S) and Patagonia (43.7°– 56°S) — median Critical Success Index (CSI < 0.49) for events > 5 mm/d— compared to the central region (CSI ≥ 0.44 for events > 5 mm/d). The evaluation of daily precipitation against station observations using Tang's Kling-Gupta efficiency (KGE_T) reveals that all datasets performed better in reproducing precipitation in rainy regions (median $KGE_T \geq 0.65$ in the Southern macrozone), while in arid areas such as the Near North during summer, the median KGET was negative. The CR2MET product consistently provided the best performance for extreme precipitation and temperature, partly because it includes information from the stations used for evaluation. Finally, the application of the TUW hydrological model shows that although WRF-SAAG simulations achieved runoff estimations comparable to the best observation-based gridded products (i.e., CR2MET

or RF-MEP) – with the best results in the southern macrozone –, they provided relatively lower performance in evapotranspiration (ET) and soil moisture estimates, as well as SWE simulations of competitive quality. More broadly, the results presented here show that -– despite some remaining challenges in arid climate regions — kilometer-scale climate models can deliver information of a quality comparable to that of observation-based products for hydrological applications in Chile.

- *The sentence on L9-21 is very long and difficult to parse. Please break this down into two or more sentences for clarity.*

  We completely restructured and summarized the abstract based on the new results included in the paper (please see our previous response).

- *In that same sentence, it is unclear which "precipitation products" are being referred to. Please be specific.*

  We completely restructured and summarized the abstract based on the new results included in the paper (please see previous answer).

Introduction

- *L62: The sentence beginning "As a result, high-resolution atmospheric models…" feels out of place. The preceding text introduces various observational and reanalysis datasets, but there has been no proper introduction to the concept of using high-resolution models as a data source. I suggest moving this sentence to a more logical position, perhaps after L91, where the rationale for using such models is better established.*

  We agree with this reviewer that the alluded text was out of place. Hence, we have moved it to the sixth paragraph of the introduction (L89-97):

  "During the last decade, convection-permitting climate models (CPCMs) have become increasingly popular (Lucas-Picher et al., 2021) because they offer an enhanced representation of precipitation (e.g., Fosser et al., 2020), with the potential to outperform gridded observational products in capturing total precipitation over complex terrain (Lundquist et al., 2019). Additionally, CPCMs do not rely on cumulus parameterizations – detected as an important source of errors in regional climate modeling –, improving land-atmosphere interactions (Prein et al., 2015). CPCMs also offer the opportunity to advance hydrometeorological understanding at kilometer-scale resolution, and have been used for a myriad of purposes, including snowpack analysis (Ikeda et al., 2021), cloud band detection (Zilli et al., 2024), and flood studies (Li et al., 2022) over continental domains (e.g., Liu et al., 2025). In particular, CPCMs offer a physically consistent and spatially continuous representation of precipitation, making them a viable alternative for process-based hydrological modeling applications."

- *L75: The text discusses a "high disagreement among CR2MET, RF-MEP, and ERA5," but the RF-MEP dataset has not been properly introduced at this point. Please ensure all datasets are introduced before they are compared or discussed.*

In the revised manuscript, we have introduced RF-MEP earlier in the manuscript (L69-72):

"For example, Boisier et al. (2018) created the gridded meteorological product CR2MET based on the combination of in-situ observations and ERA5 (Hersbach et al., 2020) reanalysis outputs, whereas Baez-Villanueva et al. (2020) developed the Random Forest based MErging Procedure (RF-MEP) for precipitation estimation, which consists of the combination of observational data, meteorological products (e.g., ERA5 reanalysis) and topographic covariates."

- *L105: This paragraph should more clearly and explicitly state the aims of the study. Currently, it seems the primary goal is to assess WRF-SAAG, but much of the paper also focuses on the inter-comparison of the three gridded products. Clarifying the primary and secondary objectives here would help frame the paper for the reader.*

  The paragraph has been revised to clearly and explicitly state the study objectives, distinguishing the primary aim and the secondary objectives in the revised version of the manuscript (L117-127*):

  "The general objective of this study is to evaluate the ability of the WRF-SAAG simulations to represent daily precipitation and maximum and minimum temperatures over continental Chile, and to assess their potential use as meteorological forcing for simulating hydrological signatures, by comparing them with observation-based meteorological products. To achieve this, four specific objectives are addressed: (i) to evaluate the performance of daily meteorological series from WRF-SAAG and observation-based meteorological products for precipitation, maximum temperature, and minimum temperature, against meteorological station observations; (ii) to assess the ability of the precipitation datasets to reproduce daily precipitation events of different magnitudes at meteorological stations; (iii) to identify the regions and climatic conditions where the main differences among the three meteorological datasets are observed; and (iv) to analyze the ability of the meteorological datasets to reproduce hydrological signatures associated with mean flow and extreme runoff events in catchments across Chile, using a conceptual hydrological model."

Study Domain

- *Figure 1b-c: The colormap used for temperature could be improved. The minimum temperature values around 10°C are close to white, making them difficult to distinguish. Please consider using a different colormap that provides better contrast across the full range of values.*

  In the revised version of the manuscript, we have updated the colormap to one that provides better contrast across the full range of temperature values, ensuring that both low and high temperatures are clearly distinguishable in Figures 1b–c:

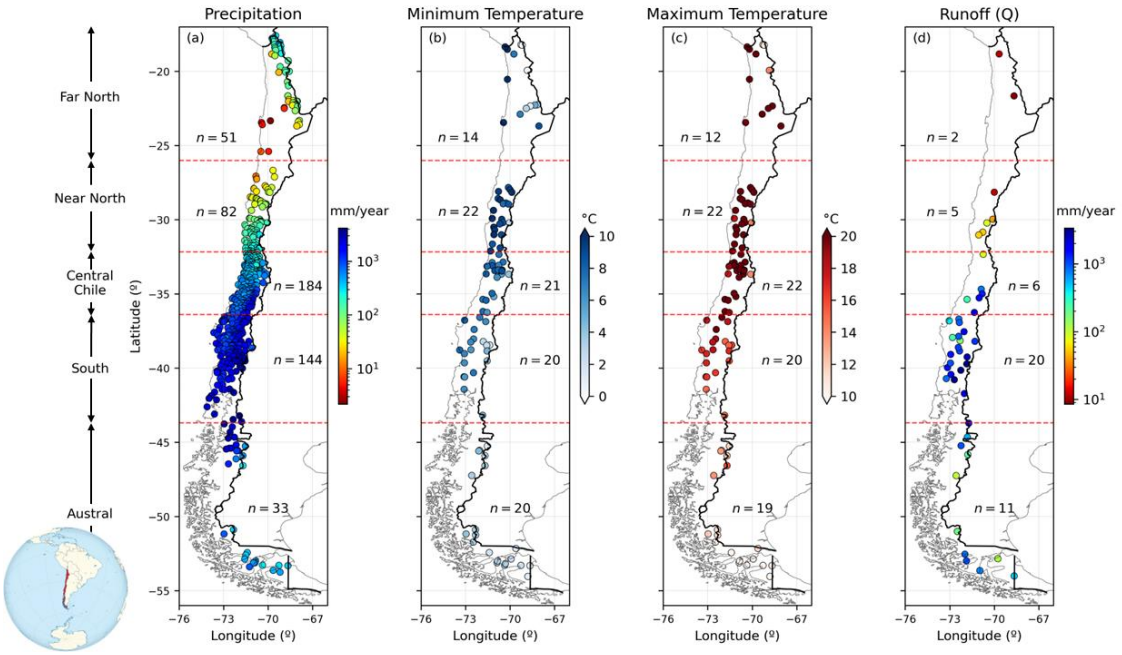


Figure 1: Annual means of (a) precipitation, (b) minimum temperature, (c) maximum temperature, and (d) streamflow measured at stations. Annual means were computed considering only hydrological years with at least 80% data availability. Red dashed lines indicate the boundaries of the study macrozones. All hydroclimatic variables are computed for the period April 2001–March 2018.

Hydrometeorological Datasets

- *This section introduces four different datasets. To improve clarity and provide an easy reference, I strongly recommend summarizing their key attributes (e.g., spatial resolution, temporal coverage, variables, post-processing methods) in a table.*

  In the revised version of the manuscript, we have incorporated a summary table with meteorological data sets in section 3.2 Climate datasets:

Table 2: Meteorological datasets.

Data set	Spatial resolution	temporal resolution	Period	Variable	Reference
WRF-SAAG	4 km	hourly	2000-2021	Precipitation; Temperature	Dominguez et al. (2024)
CR2MET (v2.5)	0.5° × 0.5	daily	1960-2021	Precipitation; Minimum and maximum temperature	Boisier et al. (2018)
RF-MEP (v2)	0.5° × 0.5	daily	1983–2018	Precipitation	Baez-Villanueva et al. (2020, 2021)

Results

- *Figure 2a: The y-axis scale (currently showing 0.0-1.0) makes the results difficult to read, as all the data points are clustered at the very top of the plots. Please adjust the y-axis scale to a more appropriate range (e.g., 0.5-1.0) to better visualize the differences.*

  We agree that the original y-axis scale (0.0–1.0) compresses the data points at the top of the plots, making it difficult to distinguish differences. We have added a zoomed view of the first column of Figure 4a, adjusting the y-axis scale (0.6–1.0) to better visualize the variations among data points:

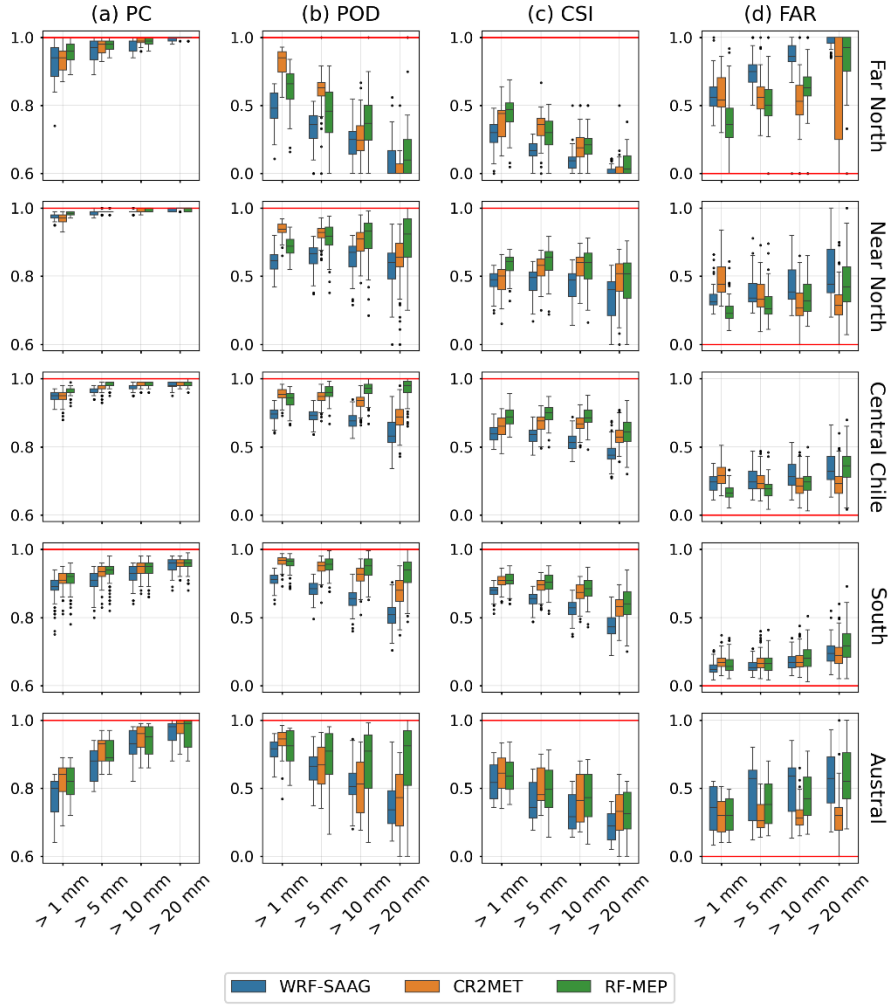


Figure 4: Ability of WRF-SAAG, CR2MET, and RF-MEP to replicate historically observed daily precipitation events with amounts larger or equal than 1, 5, 10, and 20 mm/d in the study macrozones (rows). The metrics displayed in each column are: (a) Percentage Correct (P C), (b) Probability of detection (P OD), (b) Critical Success Index (CSI), and (d) False Alarm Ratio (F AR). Each boxplot comprises results from all the stations within each macrozone. The boxes correspond to the interquartile range (IQR, i.e., 25th and 75th percentiles), the horizontal line in each box is the median, and whiskers extend to the $\pm 1.5 \cdot$ IQR of the ensemble. The red line represents the optimal values.

- *Figure 3: The caption appears to be missing the "WRF-SAAG" label.*

  Thanks for catching this! We have made the correction.

- *Evaluation of Gridded Products: Since CR2MET and RF-MEP both incorporate ground station data using different statistical methods (regression vs. random forest), a brief discussion on the potential sources of uncertainty and discrepancies between these two products would be valuable. Is the disagreement due to the selection of different stations, or the uncertainties inherent in the respective post-processing procedures?*

  The reviewer makes a good point, and in response we have incorporated these points in section 6.1 Precipitation Simulations of the revised manuscript (L591-597):

  "Although both CR2MET and RF-MEP are built upon ERA5 outputs, they have large differences in terms of annual precipitation amounts (Baez-Villanueva et al., 2021), which translate into differences in their ability to replicate daily precipitation events (Figure 4) and the accuracy of daily precipitation amounts (Figures 5 and 6). A detailed examination of discrepancies between CR2MET and RF-MEP is outside the scope of this study; nevertheless, we speculate that such differences can be attributed to (i) the target application for which those products were designed, (ii) the sample of meteorological stations used to develop the products, and (iii) the statistical models and methodological decisions (e.g., training period, predictors, optimization algorithms) adopted for calibration and verification."

- *Beyond Abstract Metrics: Figures 2 and 3 provide a good statistical summary, but the information is quite abstract. To give readers a more intuitive understanding of model performance, please supplement the KGE and contingency table metrics with an evaluation of the raw precipitation and temperature fields. For example, providing maps or summary statistics of the seasonal or annual mean biases (e.g., wet/dry bias, warm/cold bias) would be extremely helpful.*

  We appreciate the reviewer's sentiment. To address this comment, we have elaborated maps of bias for daily precipitation, maximum temperature and minimum temperature, indicating the type of bias (i.e., wet/dry, warm/cold). Since the manuscript's length has increased considerably (from 7 to 13 figures), we decided to include this information in the Supporting information section. As an example, we include only the bias maps for daily precipitation (Figure S.4) in this document.

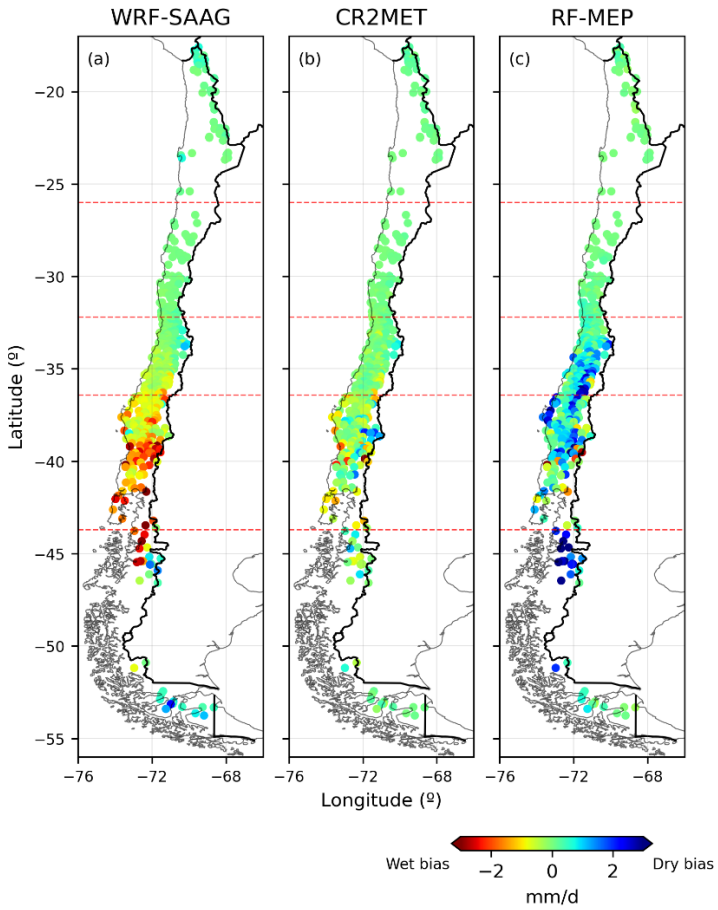


Figure S.4: Spatial distribution of precipitation biases (difference between simulated and observed mean daily precipitation) from WRF-SAAG, CR2MET, and RF-MEP (April 2001–March 2018) across continental Chile, using ground-based observations as the reference. Red dashed lines indicate the boundaries of the analyzed macrozones.

Discussion

- *L452: This paragraph provides a good summary of limitations and future work. To increase the impact of the paper, please also provide some specific insights and examples of how the WRF-SAAG and the two observational datasets could be used in practical application studies (e.g., water resource management, agricultural planning, climate change impact assessments). This would provide valuable guidance to other researchers and stakeholders in the region.*

  In response to this comment and with the aim to improve the organization of the ideas, we have divided the Discussion section into subsections, including one entitled "Limitations and future work". We have added the following text within that subsection (L657-667):

  "Future comparative assessments between WRF-SAAG and other gridded products could incorporate the effects of parameter equifinality on hydrological model simulations (e.g., Muñoz-Castro et al., 2023), examine model structural uncertainty (e.g., Saavedra et al., 2022),

compare parameter regionalization results for streamflow prediction in ungauged basins (e.g., Baez-Villanueva et al., 2021), and conduct drought propagation analyses (e.g., Lema et al., 2025). Finally, the WRF-SAAG dataset could serve as a baseline for future climate change impact assessments aligned with the efforts of the SAAG community. Recenty, Liu et al. (in preparation) conducted a climate perturbation experiment with the WRF model using the Pseudo Global Warming (PGW) approach (Schär et al., 1996; Hara et al., 2008; Kawase et al., 2009), which applies climate perturbations derived from global climate model projections to adjust the reanalysis-based initial and boundary forcings used in the baseline historical regional climate simulations. Since the method assumes that the storm tracks and frequency entering the domain remain the same in both control and future simulations, these could be used to examine changes in hydrologic processes because of shift in thermodynamic conditions."

References:

Al-Saeedi, B. A., Baez-Villanueva, O. M., and Ribbe, L.: An optimized representation of precipitation in Jordan: Merging gridded precipitation products and ground-based measurements using machine learning and geostatistical approaches, in: EGU General Assembly Conference Abstracts, p. 11510, 2024.

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Beck, H. E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J. D., and Thinh, N. X.: RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements, Remote Sensing of Environment, 239, 111 606, 2020.

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Mendoza, P. A., McNamara, I., Beck, H. E., Thurner, J., et al. (2021). On the selection of precipitation products for the regionalisation of hydrological model parameters. *Hydrology and Earth System Sciences*, *25*(11), 5805–5837. https://doi.org/10.5194/hess-25-5805-2021

Bergstrom, S.: Development and application of a conceptual runoff model for Scandinavian catchments, 1976

Boisier, J. P., Alvarez-Garretón, C., Cepeda, J., Osses, A., Vásquez, N., and Rondanelli, R.: CR2MET: A high-resolution precipitation and temperature dataset for hydroclimatic research in Chile, in: EGU general assembly conference abstracts, p. 19739, 2018.

Chen, C., He, M., Chen, Q., Zhang, J., Li, Z., Wang, Z., and Duan, Z.: Triple collocation-based error estimation and data fusion of global gridded precipitation products over the Yangtze River basin, Journal of Hydrology, 605, 127 307, 2022.

Clark, M. P. and Slater, A. G.: Probabilistic quantitative precipitation estimation in complex terrain, Journal of Hydrometeorology, 7, 3–22, 2006.

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., Michel, C., Andreassian, V., Bourgin, F., Brigode, P., Moine, N. L., Mathevet, T., Mouelhi, S., Oudin, L., and Pushpalatha, R.: Suite of GR Hydrological Models for Precipitation-Runoff Modelling, 2023.

Daly, C., Neilson, R. P., and Phillips, D. L.: A statistical-topographic model for mapping climatological precipitation over mountainous terrain, Journal of Applied Meteorology and Climatology, 33, 140–158, 1994.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., and Pasteris, P. P.: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, International Journal of Climatology: a Journal of the Royal Meteorological Society, 28, 2031–2064, 2008.

Dominguez, F., Rasmussen, R., Liu, C., Ikeda, K., Prein, A., Varble, A., Arias, P. A., Bacmeister, J., Bettolli, M. L., Callaghan, P., et al.: Advancing South American water and climate science through multidecadal Convection-Permitting Modeling, Bulletin of the American Meteorological Society, 105, E32–E44, 2024.

Fosser, G., Kendon, E. J., Stephenson, D., and Tucker, S.: Convection-permitting models offer promise of more certain extreme rainfall projections, Geophysical Research Letters, 47, e2020GL088 151, 2020.

Fowler, K., Peel, M., Western, A., & Zhang, L. (2018). Improved Rainfall-Runoff Calibration for Drying Climate: Choice of Objective Function. *Water Resources Research*, *54*(5), 3392–3408. https://doi.org/10.1029/2017WR022466

Hara, M., Yoshikane, T., Kawase, H., & Kimura, F. (2008). Estimation of the Impact of Global Warming on Snow Depth in Japan by the Pseudo-Global-Warming Method. *Hydrological Research Letters*, *2*, 61–64. https://doi.org/10.3178/hrl.2.61

Hernandez, D., Mendoza, P. A., Boisier, 820 J. P., and Ricchetti, F.: Hydrologic sensitivities and ENSO variability across hydrological regimes in central Chile (28–41 S), Water Resources Research, 58, e2021WR031 860, 2022.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly journal of the royal meteorological society, 146, 1999–2049, 2020.

Ikeda, K., Rasmussen, R., Liu, C., Gochis, D., Yates, D., Chen, F., Tewari, M., Barlage, M., Dudhia, J., Miller, K., Arsenault, K., Grubišić, V., Thompson, G., & Guttman, E. (2010). Simulation of seasonal snowfall over Colorado. *Atmospheric Research*, *97*(4), 462–477. https://doi.org/https://doi.org/10.1016/j.atmosres.2010.04.010

Ikeda, K., Rasmussen, R., Liu, C., Newman, A., Chen, F., Barlage, M., Gutmann, E., Dudhia, J., Dai, A., Luce, C., et al.: Snowfall and snowpack in theWestern US as captured by convection permitting climate simulations: Current climate and pseudo global warming future climate, Climate Dynamics, 57, 2191–2215, 2021.

Isotta, F. A., Frei, C., Weilguni, V., Percec Tadic, M., Lassegues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S. M., et al.: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data, Int. J. Climatol, 34, 1657–1675, 2014.

Kawase, H., Yoshikane, T., Hara, M., Kimura, F., Yasunari, T., Ailikun, B., et al. (2009). Intermodel variability of future changes in the Baiu rainband estimated by the pseudo global warming downscaling method. *Journal of Geophysical Research*, *114*(D24), D24110. https://doi.org/10.1029/2009JD011803

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal +of hydrology, 424, 264–277, 2012.

Lema, F., Mendoza, P. A., Vásquez, N. A., Mizukami, N., Zambrano-, M., Vargas, X., et al. (2025). Technical note: What does the Standardized Streamflow Index actually reflect? Insights and implications for hydrological drought analysis. *Hydrology and Earth System Sciences*, *29*(8), 1981–2002. https://doi.org/10.5194/hess-29-1981-2025

Li, Z., Gao, S., Chen, M., Gourley, J. J., and Hong, Y.: Spatiotemporal characteristics of US floods: Current status and forecast under a future warmer climate, Earth's Future, 10, e2022EF002 700, 2022.

Lindstrom, G., Johansson, B., Persson, M., Gardelin, M., and Bergstrom, S.: Development and test of the distributed HBV-96 hydrological model, Journal of hydrology, 201, 272–288, 1997.

Liu, C., Ikeda, K., Prein, A., Scaff, L., Dominguez, F., Rasmussen, R., Huang, Y., Dudhia, J., Wang, W., Chen, F., et al.: Convection-permitting climate simulations over South America: Experimentation during different phases of ENSO, Atmospheric Research, 316, 107 936, 2025.

Lucas-Picher, P., Argueso, D., Brisson, E., Tramblay, Y., Berg, P., Lemonsu, A., Kotlarski, S., and Caillaud, C.: Convection-permitting modeling with regional climate models: Latest developments and next steps, Wiley Interdisciplinary Reviews: Climate Change, 12, e731, 2021.

Lundquist, J., Abel, M. R., 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.

Mendoza, P. A., Rajagopalan, B., Clark, M. P., Ikeda, K., & Rasmussen, R. M. (2015). Statistical Postprocessing of High-Resolution Regional Climate Model Output. *Monthly Weather Review*, *143*(5), 1533–1553. https://doi.org/10.1175/MWR-D-14-00159.1

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., & Rajagopalan, B. (2016). How do hydrologic modeling decisions affect the portrayal of climate change impacts? *Hydrological Processes*, *30*(7), 1071–1095. https://doi.org/10.1002/hyp.10684

Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, *47*(2), W02531. https://doi.org/10.1029/2010WR009505

Muñoz-Castro, E., Mendoza, P. A., Vásquez, N., & Vargas, X. (2023). Exploring parameter (dis)agreement due to calibration metric selection in conceptual rainfall-runoff models. *Hydrological Sciences Journal*. https://doi.org/10.1080/02626667.2023.2231434

Murillo, O., Mendoza, P. A., Vasquez, N., Mizukami, N., and Ayala, A.: Impacts of subgrid temperature distribution along elevation bands in snowpack modeling: Insights from a suite of Andean catchments, Water Resources Research, 58, e2022WR032 113, 2022.

Naabil, E., Lamptey, B., Arnault, J., Olufayo, A., and Kunstmann, H.:Water resources management using theWRF-Hydro modelling system: Case-study of the Tono dam in West Africa, Journal of Hydrology: Regional Studies, 12, 196–209, 2017.}

Newman, A. J., Clark, M. P.,Wood, A.W., and Arnold, J. R.: Probabilistic spatial meteorological estimates for Alaska and the Yukon, Journal of Geophysical Research: Atmospheres, 125, e2020JD032 696, 2020.

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., Mizukami, N., Brekke, L., and Arnold, J. R.: Gridded ensemble precipitation and temperature estimates for the contiguous United States, Journal of Hydrometeorology, 16, 2481–2500, 2015a.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andreassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of hydrology, 303, 290–306, 2005.

Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, Hydrological Processes: An International Journal, 21, 435–446, 2007.

Prein, A. F., Ban, N., Ou, T., Tang, J., Sakaguchi, K., Collier, E., Jayanarayanan, S., Li, L., Sobolowski, S., Chen, X., et al.: Towards ensemble-based kilometer-scale climate simulations over the third pole region, Climate Dynamics, 60, 4055–4081, 2023.

Saavedra, D., Mendoza, P. A., Addor, N., Llauca, H., & Vargas, X. (2022). A multi-objective approach to select hydrological models and constrain structural uncertainties for climate impact assessments. *Hydrological Processes*, *36*(1). https://doi.org/10.1002/hyp.14446

Schär, C., Frei, C., Lüthi, D., & Davies, H. C. (1996). Surrogate climate-change scenarios for regional climate models. *Geophysical Research Letters*, *23*(6), 669–672. https://doi.org/10.1029/96GL00265

Wagner, S., Fersch, B., Yuan, F., Yu, Z., & Kunstmann, H. (2016). Fully coupled atmospheric-hydrological modeling at regional and long-term scales: Development, application, and analysis of WRF-HMS. Water Resources Research, 52(4), 3187–3211. https://doi.org/https://doi.org/10.1002/2015WR018185

Xie, K., Li, L., Chen, H., & Xu, C. (2025). Assessing the performance of convection-permitting climate model in reproducing basin-scale hydrological extremes: A western Norway case study. *Journal of Hydrology*, *656*, 132989. https://doi.org/10.1016/j.jhydrol.2025.132989

Zilli, M. T., Lemes, M. R., Hart, N. C., Halladay, K., Kahana, R., Fisch, G., Prein, A., Ikeda, K., and Liu, C.: The added value of using convective-permitting regional climate model simulations to represent cloud band events over South America, Climate Dynamics, 62, 10 543–10 564, 2024.