Comparison of approaches to interpolating climate observations in steep terrains with low-density gauging networks

Response to comments by Reviewers

We are very grateful with the two anonymous reviewers who have provided very valuable feedback to improve the manuscript. We are glad that both of them highlighted that the topic of the manuscript is interesting, valuable and within the scope of HESS. We are also happy that reviewer 1 highlighted the value of the temperature results, and that he or she suggested to analyse CHIRPS data in a similar way as WCA. This new method ended up being a very good alternative to interpolate precipitation.

Overall, the key requirements from reviewers involved: reorganising the content of some sections, better explaining the results obtained by the referenced authors on the interpolation of climate variables in mountain areas, providing more information of the general climate in the case study and giving a better explanation of the GLMM.

We addressed all the comments of the reviewers below.

Reviewer 1

Specific Comments

1. **Section 2.1 - Separate the description of the geographical and climate settings.**

The description of the geography was moved to the beginning of Section 2, while the climate settings were kept in Section 2.1.

a. **Climate setting would deserve a more extensive description**

The description of the climate settings in Section 2.1 was considerably increased, to provide more details of the broader climate phenomena affecting the case study, the sources of inter-annual variability (including ENSO and a brief comment on the Pacific Decadal Oscillation), and temperature fluctuations.

Further references were included. A comment on glaciers was also added, including two references which provide further details. We did not go deeper on the subject of glaciers, as their presence is restricted to the highest elevation areas of some of the sub-catchments in the case study. Furthermore, one of the references highlights that, although challenging to quantify, their role in catchment flows seems to be only relevant during dry years, and only for the very upper sub-catchments (Ohlanders et al., 2013). This means that the overall relevance of glaciers in the case study is not that high, thus, we do not consider pertinent to provide much more detail about them.

b. **Eliminate large map from Figure 1. Enlarge the small one. Clearly define the case study. Change colour of catchment delineation.**

The catchment delineation colour has been changed so it is easy to identify it. The whole figure has been enlarged. None of the figures was eliminated as we think they all are useful to clearly locate the catchment.

c. **Include comments on glaciers in the area.**

See 1a.

2. Section 2.2 – Clearly state the total number of time-series and the maximum time-span covered by the considered time-series.

This information was provided in the Appendix, however, we acknowledge that a better explanation was required in the text. Thus, the third paragraph of Section 2.2 was reworded to better link the text with the information provided in the Appendix. This paragraph is included as follows:

"A total of 42 gauges were used in the project, 18 of them measured precipitation and 24 measured temperature. The 42 gauges covered 41 sites, with one site (site 27) having both temperature and precipitation gauges. The locations of the temperature and precipitation gauges are shown in Figure 1, while further details of the gauges (including the periods with information available and the percentage of missing values) are provided in Table A1 in the Appendix."

In addition, throughout the text, whenever we referenced a gauge, we changed or complemented their names with the number of the site in Figure 1, so it is easier for the reader to locate the gauges in the map without going to the Appendix.

***a.* Merge Figure 2 and 3 in a two panel figure.**

The large number of figures was an issue in the previous version. Each figure that we planned to be a multiplot had to be split to follow the Copernicus Latex format, they were presented as multiple separate figures. Now, image files were merged prior to their inclusion in Latex so that many figures are now merged appropriately, including figure 2 and figure 3.

We have also done a major review on the figure and their labels, and have improved their overall presentation.

3. Section 2.3 – Lines 27-29 include references of studies that have evidenced decreased skill of remote sensed products in the mountain environment.

Taking into account the suggestions from Reviwer 2 to reorganise the introduction, we moved this information to the introduction. The new paragraph includes the references supporting each one of the statements.

"A broader review of the performance of satellite products for estimating precipitation in the Andes and other mountain areas (Nikolopoulos et al., 2013, Thiemig et al., 2012, Dinku et al., 2014), suggests that in these regions, satellite products tend to be good at detecting precipitation (except in very dry areas (Zambrano-Bigiarini et al., 2016, Manz et al., 2016)) and its overall spatial variability, but struggle to accurately predict the magnitudes of the events, particularly during extremely dry (e.g. in the north of Chile (Zambrano-Bigiarini et al.)) or extremely wet regions (e.g. western slopes in the Colombian Andes (Dinku et al., 2010)), and for daily and subdaily resolutions (Dinku et al., 2010, Manz et al., 2016, Thiemig et al., 2012)."

a. **Merge Figure 4 and 5 in a 2 panel figure.**

Figures have been merged. See answer to comment 2a.

b. Specify what is the DEM used for.

The DEM was used to define the elevation at all points in the catchment, as this variable is required for some of the interpolation approaches. The first part of this paragraph in section 2.3 was adjusted to include this as follows:

"The third spatial data set used was a Digital Elevation Model (DEM) based on the Shuttle Radar Topography Mission (SRTM) (Jarvis et al., 2008), with a spatial resolution of 90 m. The DEM was used to define the elevation in the catchment, in order to use this variable in some of the interpolation approaches."

c. **Consider including a plot of MEI index to discuss el nino or la nina events during the period of analysis. Given the short length of the used time series, it could be difficult to have enough ENSO cycles to get a significant correlation between observations and MEI index.**

We considered this and concluded that there would not be too much added value by including this plot. As the reviewer highlights, the period of analysis is relatively short compared to the frequency of the ENSO events, and this may have hindered finding a better correlation between the MEI index and the climate variables. This, however, is clearly stated in section 4, where we describe the correlation analysis between variables and covariates.

4. **Section 3 – It is not clear if in the first paragraph the authors discuss literature or the methods**

We have done relevant changes in Section 3 and the introduction. We moved all the review of literature to the introduction, and in Section 3 kept only the detailed explanation of the added value of the GLMM within the context of this case study, how our method differs from alternatives like Kriging or GLMs and why we chose an approximate Bayesian inference method.

a. Include at the beginning a structured list of the methods, possibly including literature on advantages/drawbacks.

This was included at the beginning of Section 3 in Table 1.

b. Discuss in more detail why the GLMM is potentially good for this application.

We provide a detailed explanation of this at the beginning of Section 3.1.

c. Be more clear on what type of data was used for each method.

This was clearly included in Table 1.

d. Provide details of the resolution of the climate outputs.

The GLMM, IDW, LR and ChA methods were used to generate data at centre-points of 5 km x 5km grids, while WCA used the original 1 km x 1 km WC grids. However, we thought it was more important to specify in the paper how we generated the data for comparison with the validation gauges. This is explained in Section 3.3, where we specify that we generated the estimates at the location of the validation gauges at each round of the leave-one-out cross validation, and then we repeated this for each approach, both for temperature and precipitation.

5. **Section 3.1 – Clearly state in Section 2.2 and abstract that monthly precipitation data was used.**

Done

***a.* If monthly precipitation data was used, why including FAR and POD?**

Although POD and FAR are more commonly used for daily analyses, the large numbers of months without precipitation in the catchment make the calculation of these two categorical statistics valuable. This reason was made explicit in the article with the following paragraph:

"Furthermore, two categorical statistics, the False Alarm Ratio (FAR) and the Probability of Detection (POD) (e.g. as applied in (Zambrano-Bigiarini et al., 2016)), were used to assess to what extent the model is able to predict precipitation occurrence (see Table 2). These categorical statistics are relevant, even at a monthly time-scale, considering that in the case study there are several months without any precipitation, thus accurately simulating its occurrence is not a trivial exercise."

***6.* Section 3.2 – Is WCA based on IDW using both station data and WorldClim maps?**

Yes, WCA is based on using IDW to interpolate the residuals between the WorldClim maps and the station data. We thought this was clear enough; however, we reworded the explanation in Table 1 and in Section 3.2 to further clarify. Two references explaining a similar method were included in the revised version, in case the reader wants to have more details about this procedure. Taking into account comment 10 from Reviewer 1, here we also explained how the same method was applied with CHIRPS data.

"The WCA method attempts to couple the benefits of the spatial variability of the WC maps and those of the temporal resolution of the observations in a simple way. Likewise, ChA attempts to improve the performance of raw CHIRPS by doing a straightforward merging of this product with observations. These approaches are similar to the RIDW in Manz et al. (2016) or the bias adjustment in Dinku et al. (2014), but in this case using WC maps and CHIRPS. First, the residual between observations and WC/CHIRPS is computed at each gauge location at a daily resolution for temperature and at a monthly resolution for precipitation. Then, these residuals are interpolated using Inverse Distance Weighting (IDW) to each point in the catchment, and this interpolated surface is added back to the original WC/CHIRPS values. This procedure is repeated for every time-step."

***a.* Merge figures 6 and 7**

To address the previous comment we reviewed again some papers where similar methods were applied, and realised that none of them included this kind of figures, but only a brief explanation with the steps followed. Taking this into account, figure 6 and 7 were eliminated and the explanation of the method was improved by providing a more specific explanation, and some references to obtain further details about the approach.

7. Section 3.3 – Divide LOOCV and sensitivity tests.

The explanation of the LOOCV and the sensitivity test was divided. Section 3.3 explains the LOOCV while Section 3.4 explain the sensitivity tests. The results were also divided in Section 4. First, a table with the results of the LOOCV is provided, together with its explanation, and then another table with the results of the sensitivity analysis is included.

a. Be more clear why for the GLMM it was required to use the expected value as opposed to the others (GLMM is a stochastic method, the others are not). Explain this in a better way for all methods.

Further details of this were provided in the fourth paragraph of Section 3.3, as follows:

"For all tests, the average Root Mean Squared Error (RMSE) was used to assess the performance of temperature and precipitation predictions, following similar comparisons (Cameletti et al., 2013, Manz et al., 2016, Nerini et al., 2015). Being a stochastic approach, for the GLMM this involved the analysis of the expected values of each variable (y in Equation 3 and y^P in Equation 6).

b. Be more specific of the comparisons of raw WC maps and temperature data, and discuss its small RMSE.

We thought sufficient details had been provided, however, we have improved the description of the comparison of this in Section 3.4.

"The sensitivity test was complemented with the estimation of precipitation and temperature values at all locations using raw WC maps and CHIRPS, in order to understand the accuracy of these data sets when used independently of the observations. This involved comparing the observed values at each time-step with those reported by CHIRPS or the WC maps, which in the latter case meant estimating the climate variables based on the long-term averages in the WC maps."

Also, in Tables 4 and 6, which include the results of the sensitivity tests, we included footnotes to make this clearer.

Furthermore, the discussion of these results was enhanced by rewording/adapting the following paragraphs in Section 5.

"Furthermore, it was found that the quality of results of the GLMM are particularly sensitive to the number and location of gauges measuring temperature. As shown in Table 3, the RMSE for this approach rises sharply when only 8 (3.89 ^{o}C), 5 (3.99 ^{o}C) and 2 (14.44 ^{o}C) gauges are used to estimate its parameters. The performances of IDW and LR also decrease considerably (RMSE of 9.34 ^{o}C and 7.78 ^{o}C respectively, with only two gauges), to the extent that using the raw WC maps for this case study (RMSE of 3.36 ^{o}C) may be preferable to any method other than WCA once the density of gauges becomes low."

"The other precipitation interpolation approaches decrease their performance at a relatively similar rate, when facing a reduction in the number estimation gauges. As shown for the LOOCV (see Figure 7B), this may be because errors at high elevation gauges strongly influence the overall RMSE. When only 4 gauges are included, however, ChA and to a lesser extent WCA show a better RMSE (21 mm and 23.5 mm respectively), although the former has a relatively low POD (88.4 %) and the latter a larger FAR (27.9 %). It was also found that CHIRPS as a standalone product is a useful alternative to the interpolation approaches when 4 or fewer gauges are available, with only marginally worse RMSE value than IDW and better RMSE than LR and GLMM (RMSE=26.2 mm, POD=88.5 % and FAR=28.6 %)."

8. Include a paragraph in the methods section discussing the correlation analysis.

A paragraph has been included at the beginning of Section 3.1 before Table 1, briefly describing the correlation analysis and its purpose. The more in-depth discussion of the results was kept in the first paragraph of the results section (Section 4), as we consider that this is a more suitable location than Section 3. The paragraph included is as follows:

"Before using the covariates mentioned in Table 1 (e.g. WC, elevation, CHIRPS), an analysis of their correlation with the climate variables was done. This included plotting temperature and precipitation observations versus the covariates, and computing Pearson Correlation coefficients."

a. **Merge Fig 8-13 in a 6 panel figure.**

The figures have been merged. See answer to comment 2a.

9. **Section 4.1 – Explain how and why the 5 yrs daily average was calculated, and explain that this was for plotting purposes only in Figs 14-16.**

This aggregation was done for illustration purposes only. Our goal with these figures was to show: what methods over and under-estimate observations, by approximately how much, how this changed as a function of the period of the year, and how this changed as a function of different types of stations. The 5-year series of daily data contained too much variability to visually assess the trends, which was achieved using the averaged series.

For the same reason, to facilitate the visualisation of the main trends, values were also smoothed using the LOESS method. Briefly, the method analyses data nearby a point X (how much data is included is a user defined parameter), and does a simple regression using this data. The value of X is adjusted to the value predicted by this regression.

Although this may eliminate day-to-day fluctuations, the overall trend over several days is shown much more clearly, as the noise is reduced. The LOESS is just one of the several methods that could be used to do this (a simple moving average could have also been used). A reference was provided so the reader can have access to more details (Jacoby, 2000). This information was not provided in the previous version because we did not consider it to be very relevant, taking into account that the method is only used for illustration purposes.

To address this, we have included the following paragraph:

"Figure 5 illustrates the daily temperature averaged over the 5-year period of analysis for sites 18, 27 and 28 (similar results were found for the rest of the gauges). Values were averaged in this way purely to facilitate visualisation of results, as the daily variability over the five years makes it difficult to see what approaches over and under-estimate observations, by approximately how much, and how this changes as a function of the period of the year. The performance metrics were calculated with the non-aggregated data."

We have also included a footnote for the figure to better explain the purpose of using the LOESS method.

"All curves were smoothed using the LOESS method (Jacoby, 2000) with \alpha= 0.045, this is similar to a moving average and is used to facilitate the visualisation of the main trends only"

a. **Merge these three in a three panel figure.**

The figures have been merged. See answer to comment 2a.

10. **Why CHIRPS data was not analysed in the same way as WC? Or be more clear how the CHIRPS data was merged with observations.**

In the revised manuscript, CHIRPS has been used in the same way as the WC maps, in order to generate methods WCA and ChA. Table 1 and Section 3 now explain in much more detail how alternative datasets were merged with observations.

See also the response to comments 6 and 7B.

The new method ChA ended up having a very good performance, and this is mentioned in the Discussion and Conclusions.

a. **Merge figures 17 and 23. 19 and 21. 18 and 22.**

Figures 19-21, and 18 and 22 were merged. See answer to comment 2a. We consider that due to the size and content of figures 17 and 23, it is desirable to keep them separated.

Reviewer 2

Specific Comments

Introduction

How have other authors addressed this topic? There is a strong discourse on this issue and a large number of researchers developing precipitation products as MSWEP, CHIRPS and CR2 have dealt with this problem. Please elaborate on the findings of other authors working with high elevation data. Also how do authors deal with missing information in hydrological modelling, which interpolation methods have worked and which were the results of evaluating different satellite based and combined precipitation data sets in data scarce Andean regions? Although you mention some authors, their findings are not described or compared. Ideally, these should help to justify your objectives.

We have made a major revision of the introduction following this comment. We have complemented the literature review with further references, and we have explained in a clearer way how other authors have used other interpolation approaches and alternative datasets within the Andean mountains, for both temperature and precipitation. Then, we use this information to highlight the gaps in the literature, which end up supporting the scope of our paper.

Furthermore, we have narrowed the scope of the paper so it is clearer from the beginning what the reader can expect from the rest of the paper.

We have also explained in Section 3 in much better way, why we selected the GLMM, CHIRPS and WC data, from other approaches and alternative datasets.

Data

The data (input, validation..) should be presented in the main text. Otherwise the numbers in the map are useless. Also in the map, it would help to enlarge it and use other colours for elevation and delineate a stronger catchment area to make the map understandable even in black and white. Numbers in the map should also be visible in Figures 2 and 3.

We are not sure about what the reviewer means by including the "data (input,validation..)" in the main text. We have tried to follow general practice from similar papers working with similar data, which commonly include:

- *A map of the region being analysed including the location of the gauges.*
- *A list, usually in an appendix, of the stations analysed, providing detailed information of the location, variables measured and availability of observations (this is not included when analyses involve a very large number of stations e.g. > 100).*
- *An overview of the data with some figures that shows seasonality patterns and the range of values of some of the gauges analysed..*

Furthermore, we are not sure how we could differentiate validation stations, as a leave-one-out cross validation method was used, which means that all stations were both used for calibration and validation in different runs of the model. We have included a paragraph in Section 3.3 to better explain the LOOCV, and what this means in terms of the gauges used for validation and as input data.

"In order to assess the performance of the approaches, one gauge was removed from the group used to interpolate the climate variable, and the set of errors for that gauge were recorded as the difference between the interpolation results for that location and the corresponding observations. After repeating this for all gauges, the concatenated errors are used to calculate the validation metrics. This leave-one-out cross-validation (LOOCV) procedure was applied separately for temperature and precipitation and for each interpolation approach."

The map has been updated following the comments from both reviewers, to make sure that the catchment is easy to identify, terrain elevation is easy to differentiate and the location of the stations is clearer. Figures 2 and 3 were updated as well following comments from reviewers. Also, a CHIRPS figure was provided for the reader to visualise this product and compare it with the WC data, particularly the resolution of both within the area of analysis. We did not think it was worth including the location of gauges in the new Figure 3, as the purpose of this figure is to give an idea of the resolution of the alternative datasets thus it could be redundant to include the gauges.

> It is not well explained why you only used such a short period. There are enough data available to fill gaps (CR2 P dataset, Chirps, MSWEPv2.2, etc.). Temperature of course is difficult but at least different time periods could be compared. The main variable of interest should be precipitation. - Why do you present a spatial distribution of Chirps in May 2009 instead of comparing it with values from observed data?

One of the added values of our project was to include high elevation data for both precipitation and temperature. The dataset we received from the private companies in the area was limited to this period thus we think it was logical to stick to it. We have tried to explain this in much more detail in the following paragraph:

"The period of analysis spans from September 2008 to August 2013 as the data obtained from the high elevation gauges was restricted to these years. Although not long enough to analyse long-term trends, the selected period allows testing of the interpolation approaches over both dry and wet years. Figure 2 provides an overview of the data by showing the monthly average temperature at four representative gauges over the five year period of analysis, and the monthly precipitation at three representative gauges throughout the same period (see Figure 1 for the location of these gauges)."

The temporal infilling using these products could have been inappropriate for creating a dataset for assessing the spatial interpolation methods. We thought that the spatial interpolation methods would be better assessed using observations from gauges only. The use of alternative spatial data sets is useful when used as inputs to the spatial interpolation methods, which he have done, as long as available observations from gauges are used to assess their proficiency. We used datasets such as the ones that the reviewer suggests, e.g. we used CHIRPS and WC maps, however, the scope of the paper was not to use all of the data sets available for the case study but to analyse the interpolation approaches. We have explained in more detail in Section 2.3 why we selected the alternative datasets we used.

We would like to further stress that we have not attempted to claim that the results in the paper are representative of long term trends. We have been cautious highlighting in multiple parts of the paper that our findings are restricted by the limitations of the study, however, this does not mean that they *are not useful. We think that they provide valuable information of the performance of some methods, under a complex climatic region with few observation gauges.*

Methods section 3:

The first paragraphs of this section should be part of the introduction as they deal with the general state of the art. - The advantage of using GLMMs and its exact output in this context is not clear to me. - There should be a conceptual figure explaining the methodology, input data and outputs - You use station data and as Covariates Chirps and ENSO as model input to test different interpolation methods. Then in the results section you correlate station data with Chirps and other data products for the station pixel? This part should be shifted to the data section and justify the method and data input (or not?). - 4.1 difference between input data and validation data not presented.

As explained in the response to comment 4 by reviewer 1, we moved a lot of information on the state of the art of the methods and the datasets from Section 3 to the introduction. We also included a new paragraph in Section 3 better describing the correlation analysis between climate variables and covariates. However, we consider that it is better to keep the outcomes of this analysis in the results section, as they are part of the process to build the GLMM (i.e. defining the covariates to use).

We have also described in detail why we used the GLMM and what were the specific methodological advantages of using it in this case study. We did not include a Figure explaining the GLMM because we provided this information in Table 1, and we think it is now much more clear what are the inputs and outputs of each one of the approaches.

We are not sure what the reviewer means by differences between input data and validation data. By using a leave-on-out cross validation (for example as applied in Manz et al. (2016)), we believe we go a step ahead of using one part of the data for estimation purposes, and the rest for validation. We run each method several times, and in each of them we remove one station at a time, to validate the results of that specific run. We repeat this process for all stations, which means that all stations were used for estimation purposes, but at the same time each of them was used once for validation purposes. The overall output is the average results of all validation stations (i.e. all stations, but only when they were used for validation).

We think that the revised manuscript explains this in much more detail.

Results:

In light of the above described missing information regarding the data input, validation data and output variables, it is difficult to understand the results and their interpretation. Overall presentation structure and language are still very poor. There are too many figures with little information content. Please focus on the main findings and try to present them in fewer self-explanatory figures.

In the revised manuscript we have explained in detail the Leave-one-out cross validation, the inputs/outputs of each approach and the reasons for using the GLMM. We have also improved the structure of the manuscript taking into account the comments from reviewers, particularly the introduction, data and methods section.

The large number of figures was an issue in the previous version and we acknowledge this decreased the presentation quality of that version. However, as explained in the answer to the comment 2A of reviewer 1, we have solved this issue by merging lots of figures in multi-plots.

We have also narrowed the scope so there is consistency throughout the paper on the aims, results and key added value of the paper.

References

CAMELETTI, M., LINDGREN, F., SIMPSON, D. & RUE, H. 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis,* 97**,** 109-131.

DINKU, T., HAILEMARIAM, K., MAIDMENT, R., TARNAVSKY, E. & CONNOR, S. 2014. Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia. *International Journal of Climatology,* 34**,** 2489-2504.

DINKU, T., RUIZ, F., CONNOR, S. & CECCATO, P. 2010. Validation and Intercomparison of Satellite Rainfall Estimates over Colombia. *Journal of Applied Meteorology and Climatology,* 49.

JACOBY, W. G. 2000. Loess:: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies,* 19**,** 577-613.

JARVIS, A., REUTER, H. I., NELSON, A. & GUEVARA, E. 2008. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database.

MANZ, B., BUYTAERT, W., ZULKAFLI, Z., LAVADO, W., WILLEMS, B., ROBLES, L. A. & RODRIGUEZ-SANCHEZ, J. P. 2016. High-resolution satellite-gauge merged precipitation climatologies of the Tropical Andes. *Journal of Geophysical Research: Atmospheres,* 121**,** 1190-1207.

NERINI, D., ZULKAFLI, Z., WANG, L.-P., ONOF, C., BUYTAERT, W., LAVADO, W. & GUYOT, J.-L. 2015. A comparative analysis of TRMM-rain gauge data merging techniques at the daily time scale for distributed rainfall-runoff modelling applications. *Journal of Hydrometeorology*.

NIKOLOPOULOS, E. I., ANAGNOSTOU, E. N. & BORGA, M. 2013. Using high-resolution satellite rainfall products to simulate a major flash flood event in northern Italy. *Journal of Hydrometeorology,* 14**,** 171-185 %@ 1525-755X.

OHLANDERS, N., RODRIGUEZ, M. & MC PHEE TORRES, J. 2013. Stable water isotope variation in a Central Andean watershed dominated by glacier and snowmelt. *Hydrology and Earth System Sciences,* 9.

THIEMIG, V., ROJAS, R., ZAMBRANO-BIGIARINI, M., LEVIZZANI, V. & DE ROO, A. 2012. Validation of satellite-based precipitation products over sparsely gauged African river basins. *Journal of Hydrometeorology,* 13**,** 1760-1783.

ZAMBRANO-BIGIARINI, M., NAUDITT, A., BIRKEL, C., VERBIST, K. & RIBBE, L. 2016. Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. *Hydrology and Earth System Sciences*.

Comparison of approaches to interpolating climate observations in steep terrains with low-density gauging networks

Juan Ossa-Moreno[1], Greg Keir[1], Neil McIntyre[1], Michela Cameletti[2], and Diego Rivera[3]

[1]Centre for Water in the Minerals Industry, Sustainable Minerals Institute, The University of Queensland, Australia
[2]Department of Management, Economics and Quantitative Methods, Universita degli Studi di Bergamo, Bergamo, Italy
[3]School of Agricultural Engineering, Water Research Centre for Agriculture and Mining (WARCAM), Universidad de Concepcion, Concepcion, Chile

Correspondence: Juan Ossa-Moreno (j.ossamoreno@uq.edu.au)

Abstract. The accuracy of hydrological assessments in mountain regions is often hindered by the low density of gauges, coupled with complex spatial variations in climate. Increasingly, spatial data sets (i.e. satellite and ~~gridded~~ other products) and new computational tools are ~~used~~ merged with ground observations to address this problem~~, by assisting with the spatial interpolation of ground observations~~. This paper presents a comparison of approaches of different complexity to spatially interpolate ~~precipitation and~~ monthly precipitation and daily temperature time-series in the upper Aconcagua catchment in central Chile. A Generalised Linear Mixed Model whose parameters are estimated through approximate Bayesian inference is compared with ~~three~~ simpler alternatives: Inverse Distance Weighting, Lapse Rates and ~~a method based on WorldClim~~ two methods that analyse the residuals between observations and WorldClim or CHIRPS data. The assessment is based on a leave-one-out cross validation, with the Root Mean Squared Error being the primary performance criterion for both climate variables, while Probability of Detection and False Alarm Ratio are also used for precipitation. Results show that for spatial interpolation of ~~the expected values of~~ temperature and precipitation, the ~~WorldClim approach~~ approaches based on the WorldClim or CHIRPS residuals may be recommended as being the more accurate, easy to apply and relatively ~~more~~ robust to tested reductions in the number of estimation gauges~~, particularly for temperature~~. The Generalised Linear Mixed Model has comparable performance when all gauges were included and is better for estimating occurrence of precipitation, but is more sensitive to the reduction in the number of gauges used for estimation, which is a constraint in sparsely monitored catchments.

1 Introduction

Climate variables such as temperature and precipitation are key inputs for hydrological modelling and water resources management~~, and generally,~~ . Generally, spatial interpolation of point observations is ~~desirable to support detailed analyses~~a necessary part of developing the climate inputs of models. Many interpolation approaches perform well for gentle terrains, however, their accuracy and precision decreases in mountain areas ~~(Wu and Li, 2013; Frei, 2014; Buytaert et al., 2006; Falvey and Garreaud, 2007)~~ (Wu and Li, 2013; Frei, 2014; Buytaert et al., 2006). As highlighted by Dorninger et al. (2008), challenges include observation errors, anisotropic climate patterns and sensitivity of results to density and location of observations. Strongly non-linear relations between temperature and altitude may be related to physiographic features ~~(Stahl et al., 2006; Diodato, 2005)~~

(Stahl et al., 2006), to cold-air trapped in enclosing hill ranges (Frei, 2014), and also to the presence of glaciers ~~(Ragettli et al., 2014; Peterse~~ (Ragettli et al., 2014). For precipitation, non-linearity can be related to physiographic features (Daly et al., 2008), to the interaction between topography and rain-storms ~~(Falvey and Garreaud, 2007; Garreaud, 2013; Viale and Garreaud, 2015; Diodato, 2005)~~ (Falvey and Garreaud, 2007; Garreaud, 2013) and to summertime convective precipitation events (Viale and Garreaud, 2014).

~~These effects can be incorporated into spatial interpolation through deterministic approaches (Frei, 2014; Masson and Frei, 2014; Thornt~~ ~~, inclusion of physiographic factors (Daly et al., 2002, 2008), geostatistics (Wu and Li, 2013; Goovaerts, 2000), and other stochastic models (Aalto et al., 2013; Kenabatho et al., 2012). The quality of the outputs of these approaches often depends on the reliability and accuracy of climate gauges, which in many catchments are sparsely situated and/or of short record length. As a consequence, there is an increasing interest on alternative sources of data beyond point observations such as satellite and other gridded products (Dinku et al., 2014; Manz et al., 2016; Zambrano-Bigiarini et al., 2016; Dinku et al., 2010; Hobouchian et al., 2017; Demaria et a~~ ~~.~~

The Andes Cordillera in South America is an example of a steep terrain with ~~sparse ground data and~~ complex weather conditions. This mountain range is an important source of natural resources, including water for agriculture, mining and other industries. The stream-flows in the region are highly variable in both time and space ~~(Pellicciotti et al., 2007; Mernild et al., 2017, 2016; Montecir~~ (Pellicciotti et al., 2007; Mernild et al., 2017; Montecinos and Aceituno, 2003; Viale and Garreaud, 2014), therefore under such circumstances, quality of spatial climate data is a key issue when modelling water resources ~~(Zambrano-Bigiarini et al., 2016; Mernild et al.~~ .

~~This paper compares four approaches for interpolating temperature and precipitation in the upper section of the Aconcagua River (in the Chilen Andes). Despite being quite sparsely gauged compared to some mountain ranges globally (~~(Zambrano-Bigiarini et al., 2 . This challenge is further complicated by the lack of gauges (i.e. when compared to mountain regions in Europe or North America), particularly at high elevation points. As a consequence, several hydrological and water resources models in some regions of the Andes, such as central Chile, have applied deterministic interpolation approaches such as Lapse Rates (LR) (Ragettli and Pellicciotti, 2012; Ragettli et al., 2014; Vicuña et al., 2011; Stehr et al., 2008; Correa-Ibanez et al., 2017) to define climate inputs. Although easy to apply, LR in hydrological applications is usually a linear or logarithmic regression using elevation as the only covariate (Ragettli and Pellicciotti, 2012), and hence does not aim to maintain the spatial correlation between observations or to fully explore the spatial dynamics of the climate variables. Therefore, there is an increasing interest in the use of improved interpolation approaches together with alternative sources of data, beyond point observations, such as satellite and other gridded products (Manz et al., 2016; Zambrano-Bigiarini et al., 2016; Dinku et al., 2010; Hobouchian et al., 2017; Dema .

In the Andes, Álvarez-Villa et al. (2011) tested four stochastic interpolation approaches in Colombia and found that Kriging with External Drift (using long term averages of the Tropical Rainfall Measuring Mission - TRMM as the drift term) had the best performance, with RMSEs between 519 and 866 mm, however this analysis was restricted to annual precipitation estimates. In Castro et al. (2014) the authors developed a deterministic method that separated the analysis of occurrence and magnitude of events, and that took into account the influence of topography (i.e. slope orientation and wind direction) to interpolate daily precipitation values in a catchment in central Chile, and found that this method outperformed inverse distance weighting (IDW) and other simple methods. This analysis was restricted to gauges below 1000 masl thus conclusions may not be valid for higher elevation points. This is a common limitation in the south Andes where there are few gauges above this elevation.

In Manz et al. (2016) the authors analysed a database of 735 gauges in Bolivia, Peru, Colombia and Ecuador (including 455 gauges above 1000 masl in the tropical Andes) and merged them with the Tropical Rainfall Measuring Mission Precipitation Radar product (TRMM 2A25). The authors used deterministic (including IDW of residuals between monthly precipitation observations and satellite estimates) and Kriging methods (including KED using mean monthly TRMM 2A25 values as the external drift term). It was found that for this case study, KED had the best performance amongst the Kriging methods, that the overall performance of Kriging methods was similar to the interpolation of residuals to estimate monthly precipitation values, and that this interpolation of residuals was less sensitive to low gauge densities. In that study performance was assessed using leave-one-out cross validation of the gauges, using metrics such as RMSE, and runoff ratios.

A broader review of the performance of satellite products for estimating precipitation in the Andes and other mountain areas (Nikolopoulos et al., 2013; Thiemig et al., 2012; Dinku et al., 2014), suggests that in these regions, satellite products tend to be good at detecting precipitation (except in very dry areas (Zambrano-Bigiarini et al., 2016; Manz et al., 2016)) and its overall spatial variability, but struggle to accurately predict the magnitudes of the events, particularly during extremely dry (e.g. ~~Swiss Alps), in the context of~~ in the ~~central and southern Andes , this catchment has an unusually high number of gauges in high elevation points. Particularly,~~ north of Chile (Zambrano-Bigiarini et al., 2016)) or extremely wet regions (e.g. western slopes in the Colombian Andes (Dinku et al., 2010)), and for daily and subdaily resolutions (Dinku et al., 2010; Manz et al., 2016; Thiemig et al., 20 .

In a comprehensive analysis of precipitation estimates from satellite products in Chile, Zambrano-Bigiarini et al. (2016) found that the satellite product PGFv3 exhibited the best overall performance for the country, followed by CHIRPS, TMPA 3B42V7 and MSWEPv1.1. The authors mention that the superior performance of PGFv3 is likely due to the ~~recent installation of several gauges by private companies operating in the area. The methodologies used include a Generalised Linear Mixed Model (GLMM - a spatio-temporal model) (Faraway, 2016), whose parameters were estimated using approximate Bayesian inference (Rue et al., 2009). This approach is relatively common in the statistics literature but is rarer in the hydrology realm.~~ bias-correction of this product, which uses several gauges from Chile. The authors also found that for most products, the performance in central Chile was superior to that in the north of the country (the driest region), that better results were achieved during the wet season and that errors were lower in areas below 1000 masl. In a similar analysis using three satellite products with long historical data records (CHIRPS, TMPA and PERSIANN-CDR) to estimate precipitation and monitor droughts in Chile, Zambrano et al. (2017) found that there were no major differences in the performances of the three products except in the southern most part of the country where PERSIANN-CDR highly underestimated values. The authors also confirmed that errors are lower during the wet season and in relatively humid parts of the country. In these two papers there was no interpolation or merging of satellite products and gauge data, but the authors recommended site-specific analyses before using satellite products in hydrological models. Furthermore, the authors also mentioned the limitations due to the lack of observations at higher elevation points.

~~Also applied are more commonly used (in hydrological and water resources modelling) deterministic approaches for interpolation, including Lapse Rates (LR) (Ragettli et al., 2014)) and Inverse Distance Weighting (IDW) . Finally, it was also useda method based on merging gauged data with WorldClim climate maps (Hijmans et al., 2005))~~In Alvarez-Garreton et al. (2018) authors describe CR2MET (DGA, 2017), a gridded product for Chile, which includes precipitation and temperature. This dataset was developed based on logistic (for precipitation occurrence) and linear (for precipitation maagnitudes and temperature) regressions using covariates such as topography, slope, ERA-Interim reanalysis variables (Balsamo et al., 2015) and in the case of temperature, MODIS satellite data were also used. Estimates of both variables on a 5 km grid were generated, however, performance metrics, particularly at high elevation gauges, were not reported. There are few other analysis of temperature interpolation in the Andes, compared to other regions (Frei, 2014; Wu and Li, 2013). However, there are global gridded datasets such as WorldClim (Hijmans et al., 2005), which are based on regressions using observations from around the world (further details of this product are given in Section 2.3).

This review highlights that there is still a lack of knowledge of how to interpolate point observations at high elevations in the sparsely gauged sub-tropical Andes, and how this process can be supported on a catchment-specific basis by using alternative sources of data. Furthermore, it is not clear what approaches are more suitable for merging different datasets under these conditions (e.g. deterministic or stochastic), particularly when compared to simple alternatives such as LR often used to support hydrological and water resources models in this region.

The aim of ~~the~~ this paper is to compare ~~the performance of the approaches in~~ five precipitation and four temperature interpolation approaches in the Upper Aconcagua River in central Chile, a mountainous catchment with steep and complex topography. ~~It specifically expects to test~~ The paper builds on the literature by: (1) ~~the applicability of a GLMM whose parameters are estimated through approximate Bayesian Inference in a hydrological context,~~ including a unique dataset of precipitation and temperature stations above 2000 masl from private companies in the area, which has not been used in similar analyses before; (2) ~~compare the quality of outcomes of the four approachesand their sensitivity to the reduction of available gauges and~~ The paper compares the approaches, focusing on the relative performance of the simple and complex ones. (3) Finally, the sensitivity of the ~~added value of including alternative data sources. A full rationale for the selection of the GLMM and the approximate Bayesian Inference method is provided later in the paper~~methodologies to the number of available gauges is tested.

~~The paper includes a description of the case study, the methods and the alternative sources of data. This is followed by a description of results, a discussion of the latter and the conclusions~~It is not in the scope of this paper to compare several stochastic interpolation methods such as in Nerini et al. (2015) or Álvarez-Villa et al. (2011); rather the paper selects one stochastic methodology (see section 3) as representative of a complex, computationally expensive approach, for comparison with simple deterministic alternatives.

2 Case study and input data

The Aconcagua River is an important source of water in Central Chile (Pellicciotti et al., 2007). The source is located in the Andean mountains ~~in the border between~~ near the border of Chile and Argentina, and the river flows west towards the Pacific Ocean. Topography fluctuates from coastal areas to peaks of ~~around~~ approximately 5900 m above sea level. The catchment has an area of ~~around 7500km^2~~approximately 7500 km^2; however, the upper section, which is the subject of this research, is only around a third of this and includes the Andean mountains and a portion of the central valley (see Figure 1).

2.1 Climate Settings

Climate within the ~~basin~~ Aconcagua catchment is Mediterranean, close to semi-arid conditions (Ohlanders et al., 2013). ~~Yearly~~ Annual average precipitation is approximately 350 mm, however, most of ~~it~~ this is concentrated during the austral winter (~~i.e.~~ frontal rainstorms during June, July and August), when the South Pacific Anticyclone retreats from the region (Falvey and Garreaud, 2007; Montecinos and Aceituno, 2003). This is complemented by occasional convective storms (Garreaud et al., 2009; Viale and Garreaud, 2014). Furthermore, precipitation is also highly influenced by the orographic effects on the windward slope of the Andes (Viale and Garreaud, 2015). The occurrence of solid or liquid precipitation is determined by the location of the zero isotherm during winter, however, above 3000 masl, low temperatures prevail and precipitation is mostly snowfall. This thermal regime allows a relevant presence of snowpack and glaciers (e.g Juncal Norte) (Janke et al., 2017; Ohlanders et al., 2013).

Streamflow peaks at the beginning of the austral summer, although it remains high between late spring and summer (Pellicciotti et al., 2007) (i.e. the dry season). This means that during this period almost all runoff comes from snowmelt and glacier melt, although the contribution from the latter seems to be ~~particularly~~ relevant during very dry years only (Ohlanders et al., 2013).

~~Water resources management in some sub-catchments of the Aconcagua have received attention from researchers (Ragettli and Pellicciott~~ ~~, who highlighted the importance of properly modelling snow accumulation and melt, which in turn requires accurate estimates of precipitation and temperature. This is particularly important for analysing potential impacts of changes in climate conditions on economic activities (Pellicciotti et al., 2014; Vicuña et al., 2011) .~~

2.2 Precipitation and temperature gauges

Observations of daily average temperature and precipitation in the catchment were sourced from the Chilean General Water Directorate (DGA) and the Chilean Meteorological Directorate (DMC), through the Chilean Centre for Climate and Resilience Research (CR2) databases. Most of these gauges are located in lowlands, whereas the mountain areas are sparsely monitored with the only available gauges ~~from operational mine sites~~ sourced from mine projects in the area. Amongst these ~~gauges~~high-elevation gauges operated by mining companies, there are two that record liquid and solid precipitation (~~Lagunitas and Los Bronces~~sites 27 and 17, see the Appendix for more details). The latter were transformed to snow water equivalents (SWE) before being analysed here.

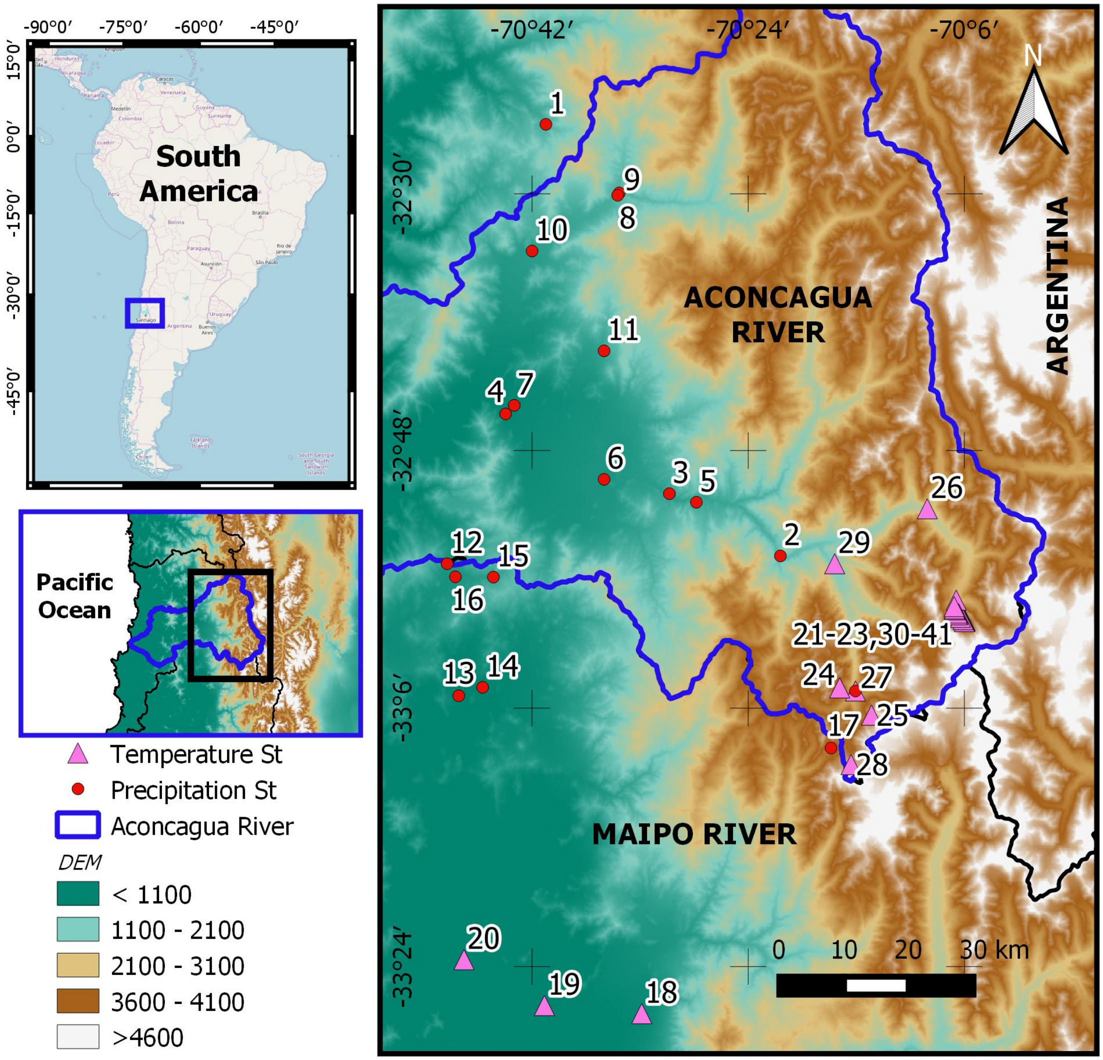

Figure 1. Temperature and precipitation gauges in the catchment with available data during the period of analysis. ~~The numbers correspond to~~ Further details of the ~~numeration~~ gauges are provided in the Appendix.

This data ~~was~~ were complemented with information from Universidad de Chile (Ohlanders et al., 2013) (available for some months only) and with measurements done by ETH-Zurich in the 2008-2009 summer season (sites 21-23 and 30-41 in Figure 1) (Ragettli and Pellicciotti, 2012; Pellicciotti et al., 2010). The latter was available during a very short period, but the measurements were done ~~in an area different than the mine sites and~~ nearby a major glacier ~~, thus~~ and in a different sub-catchment from the one where the private companies installed their gauges. Thus, they provide valuable information to test the interpolation approaches.

A total of ~~41~~ 42 gauges were used in the project, ~~17~~ 18 of them measured precipitation ~~only, 23 measured temperature only and 1 measured both variables. The location~~ and 24 measured temperature. The 42 gauges covered 41 sites, with one site (site 27) having both temperature and precipitation gauges. The locations of the temperature and precipitation gauges ~~is~~ are shown in Figure 1, while further details of the gauges (including the periods with information available and the percentage of missing values) are provided in Table A1 in the Appendix.

The period of analysis spans from September 2008 to August 2013 ~~because this was the period with more data available, as some gauges started or stopped recording measurements during these years(Figure 2 show some of the data)~~as the data obtained from the high elevation gauges was restricted to these years. Although not long enough to analyse long-term trends, the selected period allows testing of the interpolation approaches over both dry and wet years~~(the average yearly precipitation of the gauges analysed during this period was 217 mm~~. Figure 2 provides an overview of the data by showing the monthly average temperature at four representative gauges over the five year period of analysis, and the monthly precipitation at three representative gauges throughout the same period (see Figure 1 for the location of these gauges).

Quality control of climate data was done by analysing double mass plots and Pearson correlation values with patron gauges (e.g. long-term gauges previously used by academic and government sources (Jacquin and Soto-Sandoval, 2013; Ragettli et al., 2014; Correa-Ibanez et al., 2017)). ~~Beyond some issues with precipitation measurements in Hornitos and Saladillo (it was decided not to include these gauges in the precipitation Analysis), no~~ This led to the exclusion of precipitation measurements at sites 26 and 29 (the temperature measurements at these sites did not show any anomaly). No further issues with data quality were noted.

2.3 Spatially distributed data sets

To complement the point observations, the Climate Hazards Group ~~InfraRed~~ Infrared Precipitation with Station data (CHIRPS) satellite product (Funk et al., 2015) was used. ~~Including remotely sensed data to analyse climate variables is increasingly popular amongst researchers, and several examples exist for precipitation in the Andes (Dinku et al., 2010; Zambrano-Bigiarini et al., 2016; and beyond (Nikolopoulos et al., 2013; Thiemig et al., 2012; Dinku et al., 2014). Based on these experiences in mountain regions, it could be said that generally, satellite products tend to be good at detecting precipitation and its overall spatial variability, but struggle to accurately predict the magnitudes of the events, particularly heavy rainfall events, and for daily and subdaily resolutions (Dinku et al., 2010; Manz et al., 2016; Thiemig et al., 2012). This is usually a consequence of orographic effects and convective precipitation events.~~

~~Despite this , merging satellite data is one of the most promising options for spatial interpolation of precipitation, and catchment-specific studies are needed to develop this potential (Zambrano-Bigiarini et al., 2016). CHIRPS was chosen for this case study because its resolution (daily values with~~ Although there is a wide range of products available, this selection was done taking into account the good performance of this product in Chile, as reported by Zambrano-Bigiarini et al. (2016), and its spatial resolution (0.05^o pixels)~~was ideal~~. Most other products (e.g. TMPA 3B42v7, MSWEP and PGFv3) are relatively coarse for the size of the ~~case study, and because recent studies showed good performance in Chile (Zambrano-Bigiarini et al., 2016)~~ catchment (0.25^o pixels). A sample image ~~of CHIRPS~~illustrating CHIRPS' resolution compared to the size of the case study is

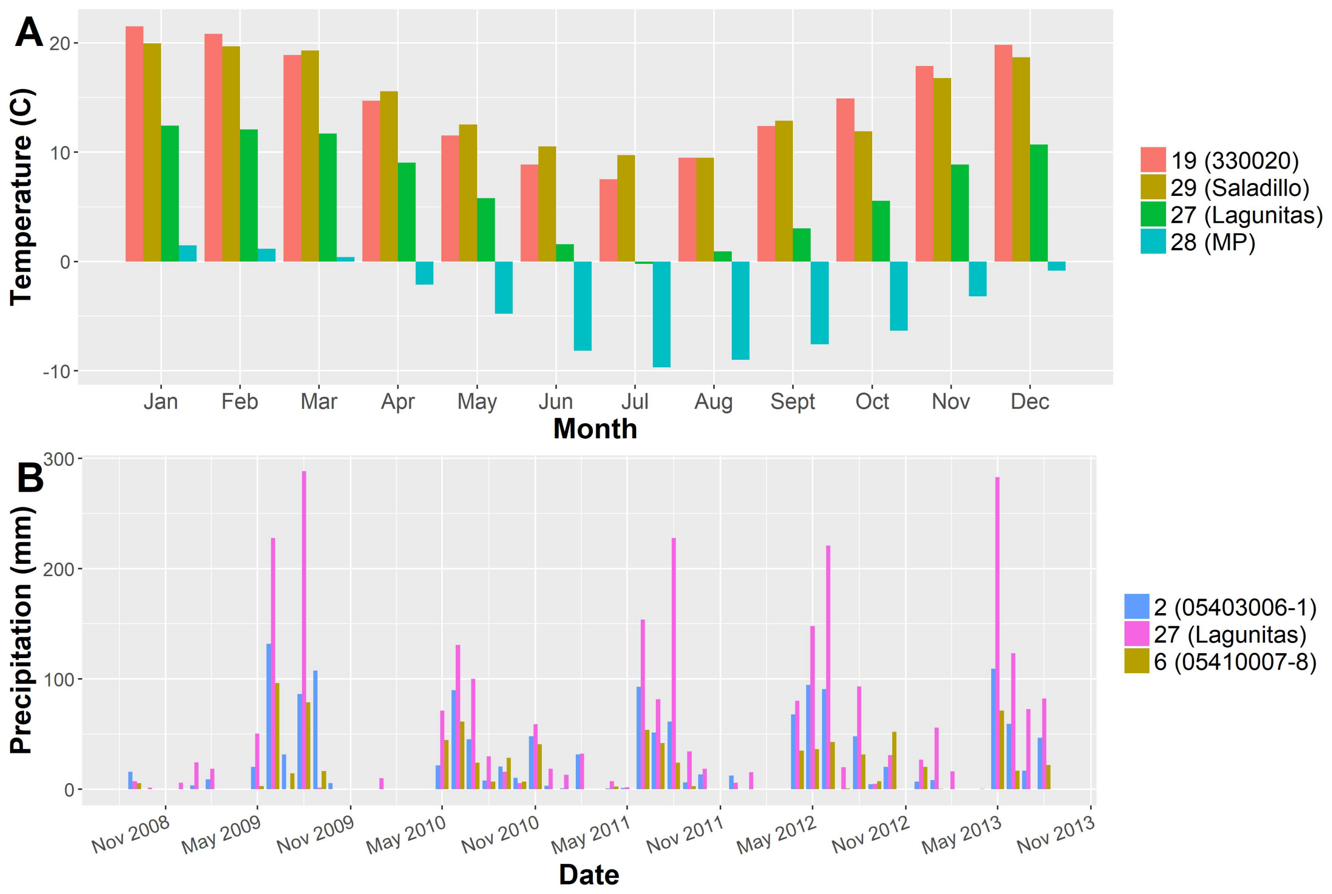

Figure 2. (A) Monthly ~~average~~ temperature ~~aggregated~~ averaged over the period of analysis, 09/2008 - 08/2013, for four of the gauges in the catchment (~~above~~B) ~~and monthly~~ Monthly precipitation ~~between 09/2008 and 08/2013~~ in the period of analysis for three of the gauges in the catchment~~(below)~~. The numbers in the legend correspond to those in Figure 1, while the texts in parenthesis are the names of the gauges. 330020 (527 masl), Saladillo (1580 masl), Lagunitas (2765.5 masl), MP (4250 masl), 05410007-8 (820 masl) and 05403006-1 (1290 masl).~~The location of the gauges can be checked in Figure 1 and the Appendix.~~

presented in Figure 3. CHIRPS does not include estimates of temperature and therefore was only used to support interpolation of precipitation.

~~Gridded products such as~~ The WorldClim (WC) Version 1 ~~(Hijmans et al., 2005) provide another potentially valuable~~ maps (Hijmans et al., 2005) were a further source of data (see Figure 3). ~~These climate surfaces~~ WorldClim was suitable due to its spatial resolution ($1km$), because it provides both temperature and precipitation values, and as for CHIRPS, because it is available worldwide and so may be used to support interpolation in any case study.

WC data provide a historical average for each one of the 12 calendar months (one map for every month) ~~, with a $1km$ spatial resolution.~~

~~WC data originates from an~~ and originates from a statistical analysis of weather observations worldwide between 1950 and 2000, through an algorithm included in the ANUSPLIN interpolation package (Hutchinson, 2004), using latitude, longitude and elevation as independent variables in a regression. The developers of ~~this~~ the WC data warn about ~~potential inaccuracies of WC~~ its potential inaccuracies in mountainous areas (Hijmans et al., 2005)~~, therefore,~~ . Therefore, the WC data were ~~never used independently but~~ used only to complement ~~point-observations,~~ point observations or as a benchmark for testing other interpolation approaches.

Although different in essence, both WC and CHIRPS can be used ~~as a~~ to complement to point observations to construct daily or monthly interpolated fields. None of the selected gauged data were used as input in the construction of WC or CHIRPS [1], furthermore the 5-year period of analysis here does not overlap with the period used to develop WC.

The third spatial data set used was a Digital Elevation Model (DEM) based on the Shuttle Radar Topography Mission (SRTM) (Jarvis et al., 2008), with a spatial resolution of $90m$. The DEM was used to define the elevation in the catchment, in order to use this variable in some of the interpolation approaches. Finally, although not spatially distributed, a multivariate ENSO (El Niño-Southern Oscillation) index was included to analyse the inter-annual variability of precipitation in the catchment (Wolter and Timlin, 2011).

3 ~~Analysis~~ Interpolation of Climate Data

~~The analysis and interpolation of climate variables in hydrology is done using a wide range of approaches~~A stochastic approach, a Generalised Linear Mixed Model (GLMM), was compared to simpler deterministic approaches: IDW and LR (Pellicciotti et al., 2014; Ragettli et al., 2014), and two methods based on the residuals between observations and alternative datasets. The first of these uses IDW to interpolate the residuals between WC maps and gauged values (precipitation and temperature), ~~including simple methods such as Inverse Distance Weighting (Lu and Wong, 2008; Chen and Liu, 2012) and linear regressions (Ragettli et al., 2014; Ragettli and Pellicciotti, 2012; Meza et al., 2014; Masson and Frei, 2014). Other approaches like non-linear functions and Generalised Linear Models (GLM)(Frei, 2014; Aalto et al., 2013) have also been used, sometimes including parameters that analyse the spatial correlation between observations (i. e. inter-site dependency) (Kenabatho et al., 2012; Kigobe e~~ ~~. The Kriging family of methods, borrowed from the geostatistics literature, has also been widely used to analyse climate~~

[1]The name of the gauges used to calibrate CHIRPS can be checked here.

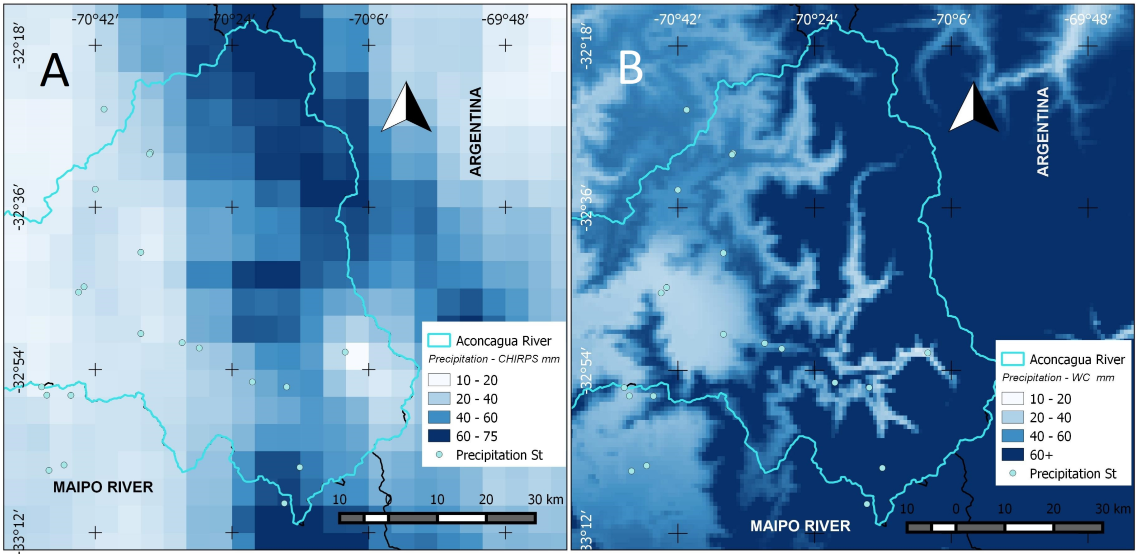

Figure 3. ~~Sample of~~ (A) CHIRPS ~~daily values aggregated~~ precipitation for May 2009 (~~left~~B) ~~and sample of~~ Worldclim precipitation values for May (long-term average)~~(right)~~.

~~variables(Nerini et al., 2015; Álvarez-Villa et al., 2011; Benavides et al., 2007; Yao et al., 2013) (a more detailed review of examples can be seen in Li and Heap (2014); Bivand et al. (2013))~~ the second uses IDW to interpolate the residuals between CHIRPS and precipitation observations. These two methods are from now on called WC Adjustment (WCA) and CHIRPS Adjustment (ChA). A summary of all interpolation approaches including the data required is in Table 1. The following sections describe the methods in more detail and their application.

~~GLMMs allow the analysis of non-normal observations as GLMs do, but the former are an extension of the latter due to their larger flexibility to analyse~~ Before using the covariates mentioned in Table 1 (e.g. WC, elevation, CHIRPS), an analysis of their correlation with the climate variables was done. This included plotting temperature and precipitation observations versus the covariates, and computing Pearson Correlation coefficients.

3.1 Stochastic Approach - GLMM

In addition to including the effects of covariates, GLMMs allow modelling of the spatio-temporal variability of the data (after removing the effect of the covariates) by means of random effects (Faraway, 2016). ~~GLMMs are frequently specified by means of a set of equations connected hierarchically; for this reason they are also known as multilevel or hierarchical models (Kéry and Royle, 2015; Rue et al., 2009). GLMMs can be estimated using Bayesian or likelihood-based approaches; the former being adopted in this paper. In both cases, they avoid~~ For example, the temporal correlation of temperature observations in this case study was analysed through an autoregressive (AR1) term (although further alternatives such as random walks could also be used). Furthermore, spatial correlation of precipitation and temperature was modelled as random

Table 1. Summary of approaches to interpolate climate variables.

Approach	Description	Input Data	Advantages	Disadvantages	References
IDW (Precipitation and Temperature)	Interpolation based on the inverse of the distance between gauges for each time-step independently.	Observations and distances between gauges	Simple and easy to implement approach.	Ignores the effects of elevation on the climate variables and does not include information from alternative datasets.	
LR (Precipitation and Temperature)	Interpolation based on linear (temperature) and logarithmic (precipitation) regressions using elevation as independent variable, for each time-step independently.	Observations and elevation of gauges	Simple and easy to implement approach that takes into account the effects of elevation on the climate variables. ~~However, as in many other fields beyond statistics, Generalised Linear Mixed Models (GLMM - also defined as Generalised Linear Mixed Effects Models) are less frequent.~~	Although alternative datasets could be included as covariates, in similar applications in nearby catchments it is more common to find elevation as the only independent variable.	(Ragettli and Pellicciotti, 2012; R
WCA (Precipitation and Temperature)	Interpolation of residuals between observations and values in WC maps. Each time-step is analysed independently.	WC Maps and observations.	Simple and easy to implement. The effects of spatial location and elevation are included to some extent through the WC values.	WC maps are not a continuous dataset but only a monthly long-term average.	(Hijmans et al., 2005)
ChA (Precipitation)	Interpolation of residuals between observations and CHIRPS. Each time-step is analysed independently.	CHIRPS data and observations.	Simple and easy to implement. Makes direct use of remotely sensed data.	Remotely sensed products tend to underperform when analysing extreme weather conditions.	(Funk et al., 2014; Zambrano-Bi
GLMM (Temperature)	Spatio-temporal model whose parameters are estimated through approximate Bayesian inference. The model includes a first order autoregressive process with spatially correlated innovations for temperature.	Observations, elevation and coordinate of gauges, and WC maps.	Takes into account multiple covariates, and analyses the random component of the climate variable thought a spatio-temporal model.	Computationally expensive compared to the rest of the models.	(Cameletti et al., 2013; Rue et al.
GLMM (Precipitation)	Spatio-temporal model whose parameters are estimated through approximate Bayesian inference. Precipitation is modelled as a spatially correlated variable with monthly dummy variables.	Observations, elevation and coordinate of gauges, CHIRPS, ENSO index and WC maps.	Takes into account multiple covariates including satellite data, reproduces both occurrence and magnitude of precipitation events, and analyses the random component of this climate variable thought a spatial model.	Computationally expensive compared to the rest of the models.	(Rue et al., 2009; Blangiardo and

variables whose covariance matrix is defined by a covariance function (in this case the Matern Function (Minasny and McBratney, 2005)) which depends on the distance between gauges and some spatial parameters (as opposed to intersite dependence functions that do not take into account distance between observations (Yang et al., 2005)).

In addition, inference on GLMMs is performed jointly for all the parameters, without having to split the estimation problem into separate steps (i.e. one for each time-step or doing the covariates regression first and the spatio-temporal analysis second

(Hengl et al., 2003)). This approach differs from Kriging methods, as it avoids using the *method of moments* to define empirical/experimental variograms (Minasny and McBratney, 2005), and the subsequent adjustment of a theoretical variogram through a curve-fitting exercise (Ecker and Gelfand, 1997; Müller, 1999), as sometimes done for Kriging applications in hydrology (Goovaerts, 2000; Nerini et al., 2015). Further details of GLMMs and the different alternatives to model spatio-temporal variables can be found in Faraway (2016); Rue et al. (2009); Lindgren et al. (2011); Cameletti et al. (2013).

The main drawback of ~~analysing~~ using GLMMs with the Bayesian approach, as done here, is the computational requirements of the classical simulation-based methods such as Markov Chain Monte Carlo (MCMC) (Cameletti et al., 2011)~~, and this is perhaps why they are less attractive compared to simpler alternatives in fields like hydrology~~. However, ~~the relatively recent~~ here we use the Integrated Nested Laplace Approximation together with the Stochastic Partial Differential Equation approach (INLA-SPDE) (Rue et al., 2009; Lindgren et al., 2011; Cameletti et al., 2013), which represents a computationally efficient way to do approximate Bayesian inference on GLMMs ~~and other models belonging to the class of latent Gaussian models~~ (Rue et al., 2009).

3.2 ~~The GLMM in the Aconcagua case study~~

~~The~~ In this approach, the climate variables in the case study (temperature and precipitation) are assumed to be realisations (e.g. observations) of a spatio-temporal process (random field) of the form:

$$Y(s,t) \equiv \{y(s,t), :(s,t) \in \mathbb{D} \subseteq \mathbb{R}^2 \times \mathbb{R}\} \tag{1}$$

where s and t denote the spatial location and time. This process has a mean μ and covariance function $Cov(y(s,t), y(s',t')) = \sigma^2 C((s,t),(s',t'))$ (Blangiardo et al., 2013; Cameletti et al., 2013). Assuming that climate observations, $\mathbf{y} = \{y(s_i,t), i = 1,...,N, t = 1,...,T\}$, follow an exponential family probability distribution function (PDF), μ_i can be connected to a structured additive predictor η_i through a link function $g(\ )$ as shown below (Rue et al., 2009):

$$g(\mu(s_i,t)) = \eta(s_i,t) = \alpha + \sum_{j=1}^{n_f} f^{(j)}(u_{j(s_i,t)}) + \sum_{k=1}^{n_\beta} \beta_k z_{k(s_i,t)} + \epsilon(s_i,t) \tag{2}$$

$$\mathbf{x} = (\alpha, \{f^{(j)}(.)\}, \{\beta_k\}, \{\eta(s_i,t)\})$$

$$\epsilon(s_i,t) \sim N(0, \sigma_\epsilon^2)$$

~~where **x**~~ where $\mathbf{x} = (\alpha, \{f^{(j)}(.)\}, \{\beta_k\}, \{\eta(s_i,t)\})$ is the vector including the Gaussian latent processes (i.e. the parameters describing the random field), ~~$\epsilon(s_i,t)$~~ $\epsilon(s_i,t) \sim N(0,\sigma_\epsilon^2)$ is the random error component, the $f^{(j)}(u_{j(s_i,t)})$ are functions of covariates u and the βs are the multipliers of covariates z.

For temperature, the model in this project was defined based on the one described in Cameletti et al. (2013) and Cameletti et al. (2011) for particulate matter, with daily time-steps. This selection was done taking into account that both variables are affected by their values in previous time-steps, but also because both of them have a spatial correlation. The model is described as follows:

$$y(s_i,t) = z(s_i,t)\beta + \xi(s_i,t) + \epsilon(s_i,t) \tag{3}$$

$$\xi(s_i,t) = a\xi(s_i,t-1) + \omega(s_i,t) \tag{4}$$

where $y(s_i,t)$ represents a realisation of the ~~gaussian field (GF)~~ Gaussian field $Y(.,.)$ for site s_i and time t, $z(s_i,t) = (z_1(s_i,t),...,z_p(s_i,t))$ are the covariates (fixed effects), βs are the coefficients of the covariates, ϵ is the measurement/observation error component, both serially and spatially uncorrelated ($\epsilon(s_i,t) \sim N(0,\sigma_\epsilon^2)$), and ξ ~~represents the random component in the model. The latter~~ is defined as a first-order autoregressive (AR) component with spatially correlated innovations $\omega(s_i,t)$ (a is the parameter of the AR1 process). The covariates included latitude, longitude, elevation and WC. Data from WC maps were included in the model as covariates, after extracting the values of the pixels containing the gauges.

The spatio-temporal model for precipitation was defined based on previous experiences of applications of INLA-SPDE ~~on GLMMs~~ for this variable. This involved dividing the analysis into occurrence (Eq. 5) and magnitude (Eq. 6) components, based on Eq. 8.5 and Eq. 8.6 in Blangiardo and Cameletti (2015). However, it was decided to use monthly time-steps as preliminary results of daily runs were far from satisfactory. In addition, CHIRPS and the ENSO index were included as covariates to complement the ones used for temperature.

$$O(s_i,t) \sim Binomial(\pi(s_i,t),1) \tag{5}$$

$$y^P(s_i,t) \sim Gamma(a(s_i,t),b(s_i,t)) \tag{6}$$

Dummy variables for each calendar month were included as additional covariates, in order to better represent the strong seasonality of precipitation in the case study (Falvey and Garreaud, 2007; Montecinos and Aceituno, 2003). In this way, the random process for this variables $\Phi(s_i,t)$ is spatially correlated but independent of other time-steps. The model is described as follows:

$$logit(\pi(s_i,t)) = z^P(s_i,t)\beta^P + \Phi(s_i,t) + \epsilon^P(s_i,t) \tag{7}$$

$$\log(\mu^P(s_i,t)) = z^P(s_i,t)\beta^P + \epsilon^P(s_i,t) + \beta^{P'}\Phi(s_i,t) \tag{8}$$

~~The link functions connecting the mean of the GF and the predictors are not unitary, as for temperature, but *logit* (occurrence) and *log* (magnitude).~~ Both Eq. 7 and Eq. 8 share the same covariate coefficients β^Ps, but the latter has an extra parameter ($\beta^{P'}$) connecting the random field in both equations.

It is acknowledged that ~~many more models~~ other models (i.e. with different random effects) could be tested with these climate variables ~~(~~after changing covariates, spatio-temporal components, the prior distributions (currently we use the default in the R-INLA package) and correlation functions (e.g. as done in Cameletti et al. (2011) for ~~PM10~~particulate matter), and this represents a subject for future research. However, ~~this project is focused on the comparison of the performance of GLMMs (whose parameters are estimated with INLA-SPDE) with simpler methods often used in hydrology, and on the inclusion of alternative data sets. Therefore,~~ taking into account the scope of the paper, it was desired to work with ~~GLMMs already available~~ existing GLMMs in the literature (or close adaptations) ~~, which have been previously~~ that have been analysed with the INLA-SPDE approach.

3.2 ~~Alternative~~ Deterministic approaches

~~The GLMM was compared to simpler deterministic approaches: IDW, LR (Pellicciotti et al., 2014; Ragettli et al., 2014), and a simple method developed in this project based on WC maps, which from now on is defined as WC Adjusting (WCA).~~ It is assumed that the reader is familiar with IDW and LR. Briefly, the former estimates variables at unsampled locations $y(s_j,t)$ as a function of the inverse of the distance $d(s_j,s_i)$ between s_j and all sampled locations s_i following

$$y(s_j,t) = \frac{\sum_{i=1}^{n} y(s_i,t)\frac{1}{d(s_j,s_i)}}{\sum_{i=1}^{n}\frac{1}{d(s_j,s_i)}} \tag{9}$$

where $y(s_i,t)$ are the values at the n sampled locations. This method does not consider elevation effects. LR, on the other hand, uses linear and logarithmic regressions to model the relation between temperature or precipitation and elevation. The regressions could be extended to include all the covariates of the GLMM, however, the objective here was to apply the methods as they are commonly used to define inputs of hydrological and water resources models in nearby catchments (Ragettli et al., 2014; Vicuña et al., 2011; Meza et al., 2014).

The WCA method attempts to couple the benefits of the spatial variability of the WC maps and those of the temporal resolution of the observations ~~. This method is described as follows:~~

~~For each time-step, each climate observation $y(s_i,t)$ was compared with the corresponding value in the WC maps $WC(s_i,t)$ (i. e. the value of the pixel where the observation is located), to define a residual R. Assuming that in the centre left pixel of map **A** in Figure **??** there is a gauge, and that $WC(s_i,t) = 5.1(^oC)$ and **B** $y(s_i,t) = 7.1(^oC)$~~in a simple way. Likewise, ChA attempts to improve the performance of raw CHIRPS by doing a straightforward merging of this product with observations. These approaches are similar to the RIDW in Manz et al. (2016) or the bias adjustment in Dinku et al. (2014), but in this case using WC maps and CHIRPS. First, the residual ~~can be defined as $R = 2.0(^oC)$.~~ between observations and WC/CHIRPS is computed at each gauge location at a daily resolution for temperature and at a monthly resolution for precipitation. Then, these residuals are interpolated using Inverse Distance Weighting (IDW) to each point in the catchment, and this interpolated surface is added back to the original WC/CHIRPS values. This procedure is repeated for every time-step.

~~This residual R was added to the WC map (map A in Figure **??**) to define an adjusted map (map C in Figure **??**). Steps 1 and 2 were repeated for all gauges available, in order to generate one adjusted map for each gauge at each time-step.~~

~~Step 1 Compares a WC value (A) with the observed value (B) to define the residual, which is added to the original WC map (A) to define an adjusted WC map (C) in step 2. This process is repeated for every gauge available at each time-step.~~

~~All adjusted maps for the same time-step were merged to define the values at all pixels (see Figure **??** for an example of the top left pixel). This was done using inverse distance weighting, between the location of the gauges and the pixel being interpolated (the example in Figure **??** uses 2 gauges). The same process is done for temperature and precipitation.~~

~~(A) shows two pixels of map C in Figure **??**. (B) Shows the observed value of a second gauge in the bottom left corner (28.4 ^oC) and the adjusted value in the upper left corner (11.6 ^oC), after applying Steps 1 and 2. Taking into account that distance from gauge A (red dot) to the upper left pixel is half of that from gauge B (pink dot), Step 3 uses inverse weighting to define the merged value (9.6 ^oC). This is repeated for all pixels first, and then for all time-steps.~~

~~A summary of all interpolation approaches is provided in Table 1.~~ For precipitation, due to the spatial smoothing that is inherent to all approaches, it is common to have very low values of precipitation where none is observed. Therefore, a threshold of 1 mm/month was set ~~,~~ below which all values were deemed to be 0.

~~Summary of approaches to interpolate climate variables. Approach Description Error Model Temporal Correlation IDW Interpolation based on the inverse of the distance No random component The method is run for every time-step independently LR Interpolation based on regressions using elevation as independent variable No random component The method is run for every time-step independently WCA Interpolation based on residuals of observations and values in WC maps No random component The method is run for every time-step independently GLMM Spatio-temporal model whose parameters are estimated through approximate Bayesian inference First order autoregressive process with spatially correlated innovations for temperature and spatially correlated innovations for precipitation AR1 for temperature and monthly dummy variables for precipitation~~

~~The approximate Bayesian inference approach (INLA-SPDE) used to estimate the parameters of the GLMM was run using the INLA package for R (Rue et al., 2013), and this required using the Euramoo and Flashlite High Performance Computers (HPC) system from the Queensland Cyber Infrastructure Foundation (QCIF). All other interpolation approaches were run on a computer with 16 Gb of memory, an i7 processor and 4 cores.~~

3.3 Comparison of interpolation approaches

~~A~~ In order to assess the performance of the approaches, one gauge was removed from the group used to interpolate the climate variable, and the set of errors for that gauge were recorded as the difference between the interpolation results for that location and the corresponding observations. After repeating this for all gauges, the concatenated errors are used to calculate the validation metrics. This leave-one-out cross-validation (LOOCV) ~~(Manz et al., 2016) method was used to assess the performance of the approaches for both~~ procedure was applied separately for temperature and precipitation ~~. Then, the sensitivity of performance to the number of gauges used for estimation was tested~~and for each interpolation approach.

For temperature there ~~were~~ was a total of 24 gauges available, thus, the LOOCV analysed ~~the~~ 24 combinations of 23 gauges~~, leaving one at a time for validation. For the sensitivity analysis it was only used the 9 gauges with relatively large observation periods (i.e. 15 out of the 24 gauges had observations for the 2008-2009 summer season only - see Figure 1 around $-70^o06'$ longitude and $-33^o00'$ latitude). Including the 24 gauges in the sensitivity analysis would have been a problem when using a reduced number of gauges for estimation, as several combinations would have had no data for most of the period of analysis.~~

~~In this way, the performance of all approaches was tested by using all combinations of 8 gauges to estimate results, and using the remaining gauges of each combination (plus the 15 with few data) for validation purposes. The overall metric was the average result of the validation groups. This was then repeated for all combinations of 5 and 2 gauges.~~

~~For precipitation ,~~ . For precipitation there were 18 gauges available~~during most of the period of analysis, however only two of them were located in the mountains. The same procedure was followed for this variable for the LOOCV (~~, thus the LOOCV involved analysing 18 combinations of 17 gauges~~) and for the sensitivity analysis, but this time the latter was done with 14 and 4 gauges~~.

For all tests, the average Root Mean Squared Error (RMSE) ~~of the validation group~~ was used to assess the performance of temperature and precipitation predictions, following similar comparisons (Cameletti et al., 2013; Manz et al., 2016; Nerini et al., 2015). ~~For the GLMM , this involved analysing the~~ Being a stochastic approach, for the GLMM this involved the analysis of the expected values of each variable ~~for each time-step.~~ (y in Equation 3 and y^P in Equation 6).

This was complemented with an analysis of the distribution of the residuals~~of the validation groups of the LOOCV~~. Furthermore, two categorical statistics, the False Alarm Ratio (FAR) and the Probability of Detection (POD) (e.g. as applied in Zambrano-Bigiarini et al. (2016)), were used to assess to what extent the model is able to predict precipitation occurrence (see Table 2)~~(Castro et al., 2015; Tobin and Bennett, 2012).~~ . These categorical statistics are relevant, even at a monthly time-scale, considering that in the case study there are several months without any precipitation, thus accurately simulating its occurrence is not a trivial exercise.

Table 2. Categorical statistics used to assess the capacity of the interpolation approaches to predict the occurrence of precipitation.

Precipitation	Observed	Not Observed
Predicted	A	B
Not Predicted	C	D
POD	$\frac{A}{A+C}$	
FAR	$\frac{B}{A+B}$	

3.4 Sensitivity to the number of estimation gauges

The sensitivity of the performance of the different approaches to the number of estimation gauges was also tested. For temperature, only 9 gauges with relatively long observation periods were used as estimation gauges in this sensitivity analysis. The other 15 gauges were operational for only one summer period, 2008-2009, and the variability in record length they introduced made it difficult to isolate sensitivity to number of estimation gauges. These 15, however, remained as validation gauges.

This allowed 9 combinations of 8 estimation gauges. The 9 validation results were averaged for the purpose of the sensitivity analysis. This was repeated using different numbers of estimation gauges: all possible combinations of 5 and 2 gauges out of the 9. The sensitivity analysis for the precipitation results was done in a similar way, but this time with all combinations of 14 and 4 gauges.

The sensitivity test was complemented with the estimation of precipitation and temperature values at all locations using raw WC maps and CHIRPS, in order to understand the accuracy of these data sets when used independently of the observations. This involved comparing the observed values at each time-step with those reported by CHIRPS or the WC maps, which in the latter case meant estimating the climate variables based on the long-term averages in the WC maps.

Regarding the computational requirements, the approximate Bayesian inference approach (INLA-SPDE), which was run on the INLA package for R (Rue et al., 2013), required using the Euramoo and Flashlite High Performance Computers (HPC) system from the Queensland Cyber Infrastructure Foundation (QCIF). All other interpolation approaches were run on a computer with 16 Gb of memory, an i7 processor and 4 cores.

4 Results

~~Before starting the interpolation of variables, their correlation with the covariates was assessed. As expected a priori,~~

4.1 Preliminary analysis of correlations between covariates and climate variables

Figure 4 shows that monthly temperature values ~~were~~ are inversely correlated to elevation (Pearson Correlation Coefficient $\rho = -0.81$~~, see Figure 4). There was also~~). The figure also shows a strong correlation between WC values and monthly temperatures ($\rho = 0.98$~~, see Figure 4). Daily temperature values showed~~). Likewise, daily temperature values show considerable correlation with elevation ($\rho = -0.77$) and WC ($\rho = 0.93$)~~as well. ENSO showed a~~. In contrast, ENSO has a a low correlation with temperature ($\rho = 0.04$), thus it was decided not to include this covariate in ~~this model~~the GLMM.

~~The~~ Figure 4C and Figure 4D show that the correlation between CHIRPS and daily precipitation observations ~~was weak(see Figure 4)~~is weak, but considerably ~~improved when both were~~ improves when both are aggregated to monthly values ($\rho = 0.81$~~- see Figure 4~~). The ρ for monthly precipitation and WC values ~~was lower~~ is lower but significant ($\rho = 0.62$~~, see Figure 4~~), while monthly correlation with elevation ~~was~~ is above 0.6 for most months~~(see Figure 4). ENSO showed~~. ENSO shows a weak correlation with precipitation ($\rho = 0.12$), however, a monthly analysis ~~showed~~ shows that for several months the correlation ~~was~~ is close to $\rho = 0.5$, therefore, it was decided to keep ENSO as a covariate for the precipitation ~~model~~GLMM. These correlations may be stronger in longer-term analyses that cover several Niño-Niña cycles, which last ~~around~~ approximately 2-5 years each (Wolter and Timlin, 2011; Garreaud et al., 2017; Montecinos and Aceituno, 2003).

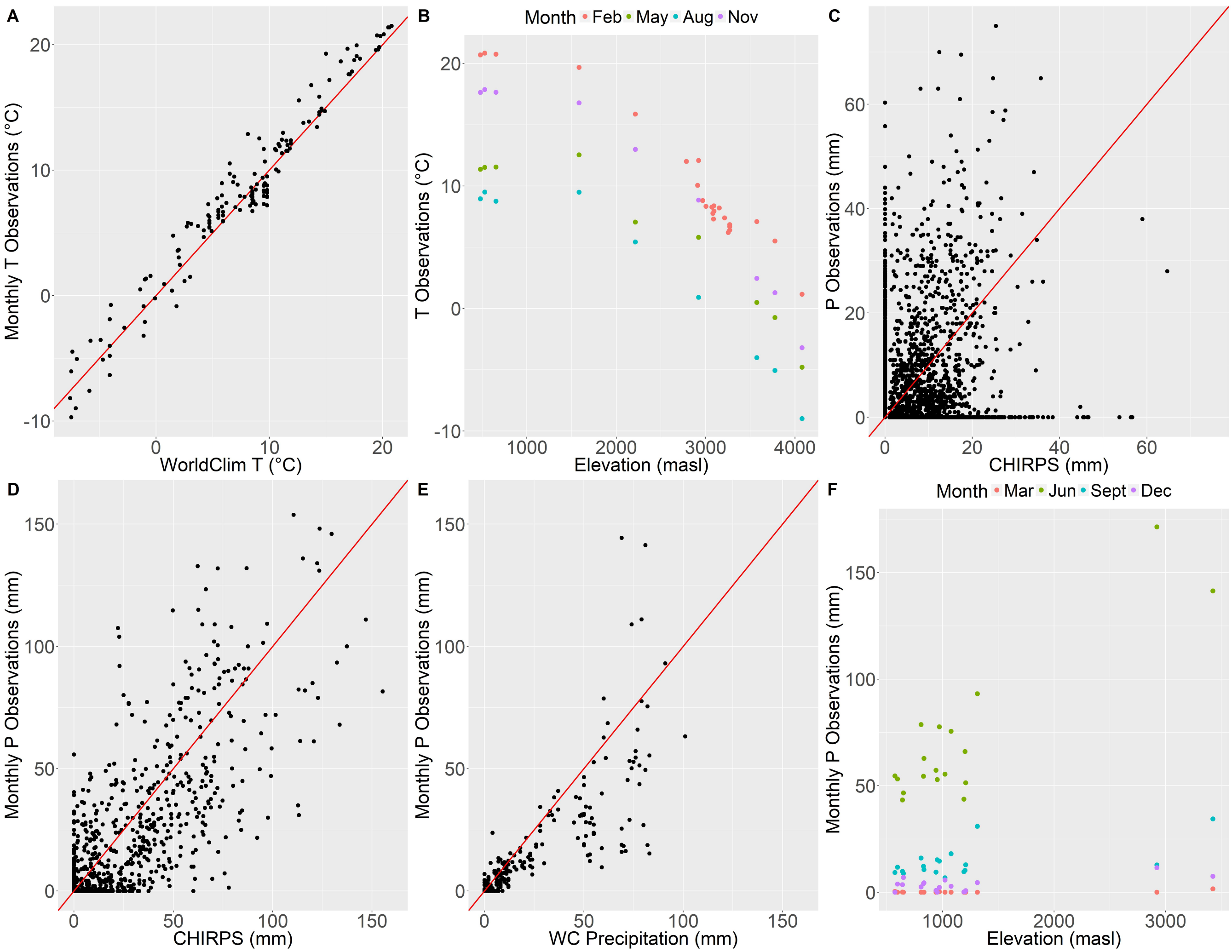

Figure 4. (A) WC values ~~vs.~~ versus monthly aggregated (averaged) temperature values. (B) Elevation of gauges ~~vs~~ versus average temperature in four months. (C) CHIRPS ~~vs~~ versus precipitation. Daily values for all stations used. (D) Monthly aggregated (sum) CHIRPS ~~vs~~ versus monthly aggregated (sum) precipitation values. (E) WC values ~~vs.~~ versus monthly aggregated (sum) precipitation values. (F) Elevation of gauges ~~vs~~ versus average precipitation ~~in~~ for four months. The red lines correspond to the 1:1 line.

4.2 Temperature results

Table 3 shows the results of all interpolation approaches in terms of the average RMSE of the validation gauges in the LOOCV (23 gauges). It was found that the GLMM and WCA have the best performance, while LR and particularly IDW have larger RMSE values.

~~In addition, Table 3 also~~

Table 3. Temperature RMSEs in the leave-one-out cross validation for each interpolation approach.

Approach	RMSE (^oC)
GLMM	1.20
WCA	1.22
LR	1.53
IDW	2.72

Table 4 shows the results of the ~~validation groups of all combinations in the~~ sensitivity analysis. As expected~~a priori~~, it can be seen that errors increase when the number of estimation gauges decreases. However, values for WCA increase the least, and its loss of performance is relatively small even when only two estimation gauges are used. On the other hand, the performances of all other approaches, including the GLMM, show a sharp decline, to the point that some of their RMSE values are comparable with the range of observed temperatures (see Figure 2).

Table 4. ~~Average RMSE in~~ Sensitivity test of the ~~leave-one-out cross validation for each~~ temperature interpolation ~~approach~~approaches.

Approach	Number of estimation gauges	RMSE (^oC)
GLMM	~~23 (LOOCV) 1.2~~ 8	3.89
	5	3.99
	2	14.44
WCA	~~23 (LOOCV) 1.22~~ 8	1.77
	5	1.98
	2	2.54
	0 (Raw WC Maps)*	3.36
LR	~~23 (LOOCV) 1.53~~ 8	2.12
	5	4.14
	2	7.78
IDW	~~23 (LOOCV) 2.72~~ 8	4.42
	5	6.15
	2	9.34

*** Using the monthly long term values provided by WC to approximate daily temperature at all sites (i.e. one value applied to all days in the month).**

~~Figures ??, ?? and ?? show~~ Figure 5 illustrates the daily temperature averaged over the 5-year period of analysis for ~~three validation gauges (~~sites 18, 27 and 28 ~~in Figure 1 -~~ (similar results were found for the rest of ~~them). It~~ the gauges). Values were averaged in this way purely to facilitate visualisation of results, as the daily variability over the five years makes it difficult to see what approaches over and under-estimate observations, by approximately how much, and how this changes as a function of the period of the year. The performance metrics were calculated with the non-aggregated data.

In the figure it can be seen that the GLMM and WCA ~~manage to reproduce the values from most of them relatively well , except for the MP gauge~~ reproduce the observed temperatures relatively well except for site 28 (the one at the highest elevation - 4250 masl)~~where larger differences can be seen~~. LR and particularly IDW tend to underestimate ~~values in~~ temperature at all gauges, except ~~for MP, which they overestimate~~ at site 28 where they overestimate it.

~~It is worth mentioning that in Figure ??, the~~ In Figure 5A, the anomalous overestimation of temperature ~~for~~ with the LR method around March ~~, was~~ is because during March 2009 all other ~~gauges in the mine site~~ high elevation gauges stopped measuring, thus the predictions for ~~Lagunitas~~ site 27 were done with the lower elevation data only. This generated large errors for this gauge and this ~~method, highlighting the issues with LR when few~~ approach, which may highlight the limitations of the latter when few estimation gauges are available ~~for estimation purposes~~or when it is required to extrapolate results far beyond the elevation of available gauges. This will be further discussed later in this section.

~~Daily temperature averaged over the 5 years of analysis for Lagunitas (All curves were smoothed using the LOESS method ((Jacoby, 2000)) with $\alpha = 0.045$ to facilitate visualisation).~~

~~Daily temperature averaged over the 5 years of analysis for MP (All curves were smoothed using the LOESS method ((Jacoby, 2000)) with $\alpha = 0.045$ to facilitate visualisation).~~

Figure 6 shows histograms of the validation residuals. It can be seen that the GLMM, WCA and LR ~~have~~ give residuals that are more or less evenly distributed around zero, although those of the GLMM are more peaked. The distribution of IDW residuals is strongly multi-modal indicating consistent over or under-estimation at particular gauges. ~~Furthermore, Figure ??~~ Figure 7A shows the relationship between temperature RMSE values and elevation.

4.3 Precipitation results

Table 5 shows that the performances of all interpolation approaches are ~~similar in the LOOCV~~relatively similar, in terms of RMSE, although ~~WCA and IDW have~~ ChA has slightly smaller RMSE values. All probability of detection (POD) indices are above 90%, although WCA and IDW have values closer to 100%. Differences in false alarm ratios (FAR) are larger, as the GLMM has a ratio of only 7.1%, which is almost half of the one for LR and ChA, and less than a third of that of IDW and WCA.

~~Table 5 also shows how sensitive are the interpolation approaches to the reduction~~

Table 6 shows the sensitivity of performances to reductions in the number of estimation gauges. It can be seen that the GLMM ~~was~~ is quite sensitive to these changes, and its ~~performance decreased sharply with~~ RMSE performance decreases sharply when moving from 17 to 14 ~~and particularly with~~ gauges, and even more from 14 to 4 gauges~~in terms of the RMSE~~. Its POD and FAR ~~remained similarto the values in the LOOCV. The other 3 approaches behaved similarly with~~ remain similar. The

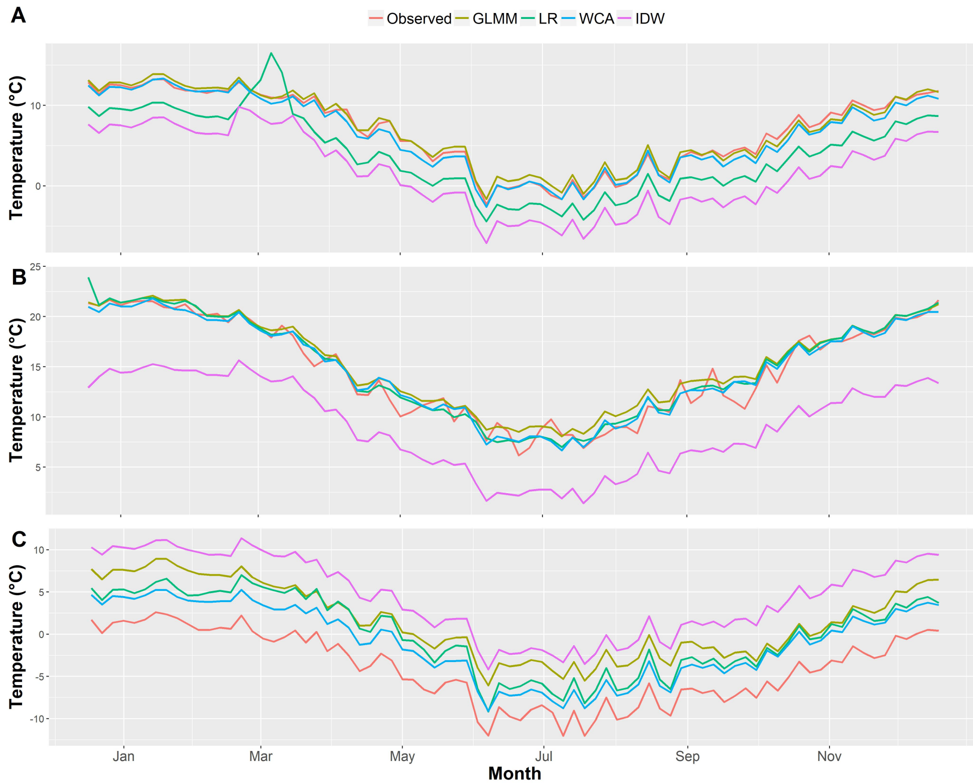

Figure 5. Daily temperature averaged over the 5 years of analysis for gauge (A) Site 27 (Lagunitas) (B) Site 18 (330019) (C) Site 28 (MP) (All curves were smoothed using the LOESS method ~~(~~(Jacoby, 2000) ~~)~~ with $\alpha = 0.045$, this is similar to a moving average and is used to facilitate the visualisation of the main trends only).

Table 5. Precipitation results in the leave-one-out cross validation for each interpolation approach.

Approach	RMSE (mm)	POD (%)	FAR (%)
GLMM	14.2	92.3	7.1
LR	15.5	93.7	12.9
WCA	13.4	97.3	24
IDW	13.5	98	22.7
ChA	12.8	90.1	12.2

RMSE performance of the other four approaches decreases by a similar rate (3 - 4 mm) when moving to 14 gauges, although LR ~~had~~ and ChA have lower POD and FAR. ~~With~~ When moving from 14 to 4 gauges ~~WCA~~ ChA shows the smallest increase in RMSE, ~~although its FAR has the largest increase. On the other hand,~~ followed by WCA, while LR has a ~~larger RMSE but a low FAR again.~~ large increment. PODs and FARs of these four methods remain similar when moving to 4 gauges, except for the LR POD which drops around 6%.

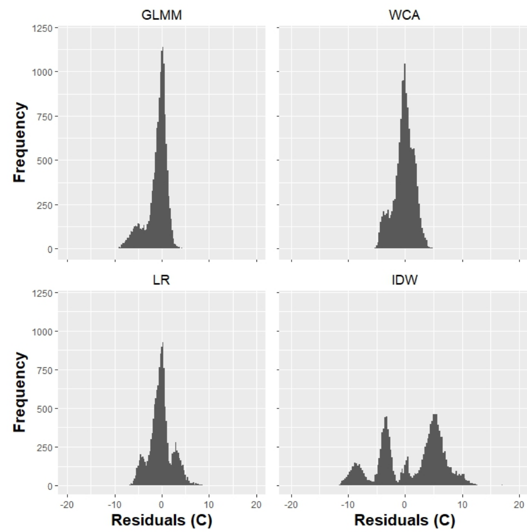

Figure 6. Residuals of the temperature LOOCV for each interpolation approach.

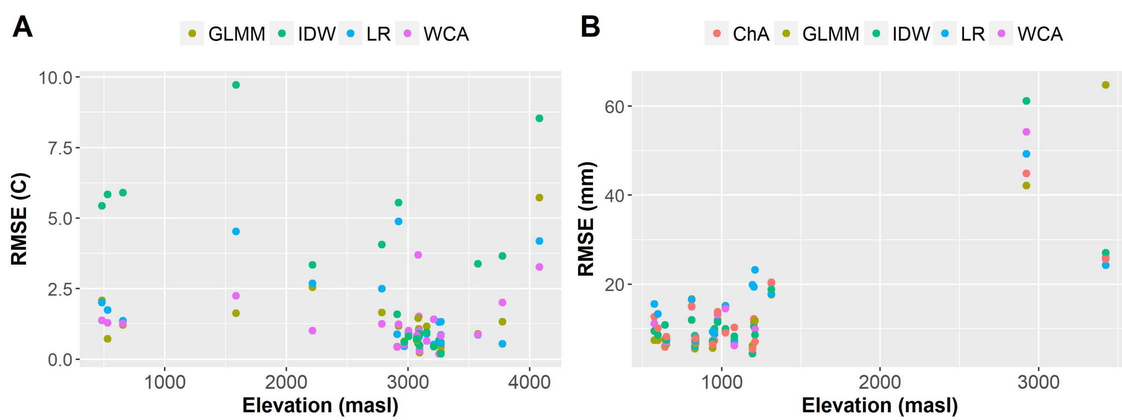

Figure 7. (A) Elevation of gauges vs ~~Temperature~~ Average temperature RMSE ~~for all gauges~~ in the ~~validation groups~~ LOOCV. (B) Elevation of gauges vs Average precipitation RMSE in the LOOCV.

When these values are compared with raw CHIRPS and WC values, it can be seen that the performance of both alternative data sets by themselves is not competitive when there are 17 or 14 gauges available. ~~However, the quality of CHIRPS predictions is similar to results with~~ The accuracy of CHIRPS gets closer to that of the interpolation approaches when only 4 gauges ~~in terms of RMSE, POD and FAR. This suggests that if there were four or less gaugesavailable in the catchment, using CHIRPS would be a useful alternative~~are used suggesting its potential value for especially poorly gauged regions; however, still, ChA and WCA perform better with 4 estimation gauges.

Table 6. ~~Results~~ Sensitivity test of the ~~approaches tested to interpolate~~ precipitation interpolation approaches.

Approach	Number of estimation gauges	RMSE (mm)	POD (%)	FAR (%)
GLMM	~~17 (LOOCV) 14.2 92.3 7.1~~ 14	32.1	91.8	~~7.07~~ 7.1
	4	135.8	87.8	10.6
~~WCA~~ LR	~~17 (LOOCV)~~ 14	~~13.4~~ 18.9	~~97.3~~ 90.6	~~24~~ 15.7
	4	26.4	84.4	11.7
WCA	14	17.4	97.5	25.8
	4	23.5	95.3	27.9
	0 (Raw WC Maps)*	34.1	98.6	40.5
IDW	~~17 (LOOCV) 13.5 98 22.7~~ 14	17.8	97.2	22.1
	4	25.4	94	19.1
~~LR~~ ChA	~~17 (LOOCV) 15.5 93.7 12.9~~ 14	~~18.9~~ 16.1	~~90.6~~ 90.2	~~15.7~~ 13.4
	4	~~26.4~~ 21.0	~~84.4~~ 88.4	~~11.7~~ 16.4
~~CHIRPS~~	0 (Raw CHIRPS data)**	26.2	88.5	28.6

*** Using the monthly long term values provided by WC to approximate daily temperature at all sites (i.e. one value applied to all days in the month).**

**** Using the monthly CHIRPS values at all sites.**

~~Figures ??, ?? and ?? show~~ Figure 8 shows the observed and simulated monthly precipitation values for three representative gauges. Figure ~~??~~ 8A shows the performance of ~~05200007 − 6~~the low elevation gauge at site 1, which is ~~quite similar to that of all gauges in the lowlands~~representative of the performance at the other low elevation gauges. It can be seen that most approaches reproduced observed ~~values relatively well (~~precipitation at this lowland gauge well compared to the ~~gauges in the mountains). Furthermore, it~~ high elevation gauges. It can also be seen that IDW and WCA predicted small amounts of precipitation in several months during the dry season when no precipitation was observed, which causes a larger FAR for both of them (see Table 5).

~~Figure ??~~

Figure 8B shows the performance of all approaches for ~~Lagunitas~~site 27, which is in the mountains at 2765 $masl$. In this plot it can be seen that observed precipitation is larger than in the lowlands, and that all approaches fail to reproduce observations with the level of accuracy ~~shown in Figure ??. Figure ??~~ seen for the lowland gauge (see Figure 8A). Figure 8C illustrates results for ~~Los Bronces gauge~~site 17, the highest of the precipitation gauges (3420 $masl$). Once more, larger errors can be seen compared to the gauges in the lowlands, particularly for the GLMM, ~~which nevertheless had~~ although this approach has the best results ~~for Lagunitas.~~ in site 27.

~~Validation monthly precipitation estimates for gauge 05200007-6.~~

~~Validation monthly precipitation estimates for Lagunitas.~~

~~Validation monthly precipitation estimates for Los Bronces.~~

This behaviour can be better appreciated after plotting the elevation of the gauges versus their average RMSEs (see Figure ~~??~~7B). While RMSE values below 1500 $masl$ are rarely above 20 mm, all the RMSE values of the two gauges above 1500

Figure 8. Validation monthly precipitation estimates for sites (A) 1 (05200007-6) (B) 27 (Lagunitas) (C) 17 (Los Bronces).

$masl$ are above this threshold, some of them are beyond 40 mm and two ~~points~~ are above 60 mm. This suggests that the performance of all approaches is likely to be determined by inaccuracies at high elevation gauges, where frontal systems interact with the topography to create ~~very wet conditions~~ high precipitation during the wet season.

~~Elevation vs Precipitation RMSE for all gauges in the validation groups of the LOOCV.~~

~~Regarding the residuals of all approaches~~ Regarding the distribution of residuals (see Figure 9), ~~as for temperature, the residuals using GLMM are more peaked around zero. Nevertheless,~~ all aproaches show values that are more or less equally distributed around 0. The GLMM residuals are particularly peaked at 0, nevertheless, its greater number of very large residuals gives the GLMM a higher RMSE than ChA, WCA or IDW.

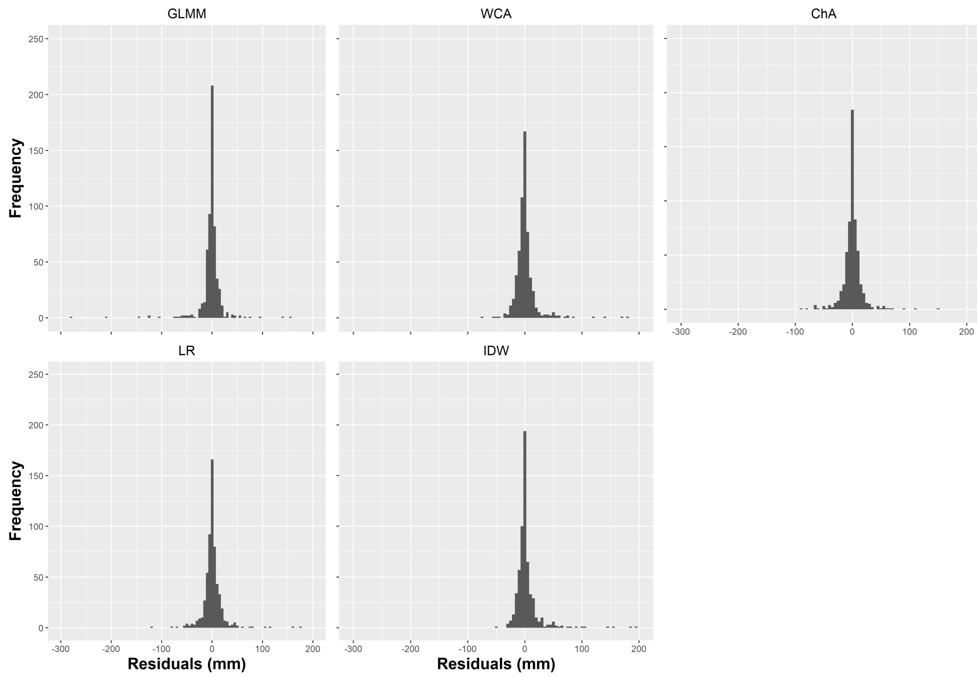

Figure 9. Residuals of the precipitation LOOCV for each interpolation approach.

5 Discussion

The LOOCV analysis of air temperature in Section 4.2 ~~showed~~ shows that for this case study, the GLMM and the WCA have the best performance (i.e. smallest RMSE values - see Table 3)~~, although the magnitudes of LRresults are also~~ . These results, and those of LR, are comparable with those obtained from similar analyses in USA and Canada (Stahl et al., 2006; Wu and Li, 2013). However, compared to the GLMM, WCA has less computational requirements thus is easier to implement (i.e. WCA was run on a desktop computer as described in Section 3.4, while the GLMM was run on 20 HPC cores in parallel).

On the other hand, IDW ~~had the largest~~ has the largest temperature errors and this, together with the skewed and multi-modal nature of its residuals, ~~showed~~ shows the limitations of this approach~~to analyse this variable. Figure ?? and ??~~ . Figure 5C and 7A suggest that IDW residuals can sometimes be related to the high elevation (e.g. ~~MP~~site 28) or isolation (e.g. ~~Saladillo~~site 29) of gauges. ~~The~~ Temperature observations from the 2008-2009 summer season have the best RMSE values for IDW, ~~and~~ but this is likely to be due to the proximity and quantity of gauges in this period. ~~The latter, however, could be a problem for the estimation of values of other gauges as IDW does the interpolation based on distances between gauges only, but does not necessarily analyse the effects of data clustering.~~

In terms of the influence of the elevation of gauges on temperature results, WCA~~and LR showed~~ , LR and the GLMM show similar performance across all elevations, ~~but the GLMM had~~ although the latter has an outstanding error ~~in~~ at the highest gauge (~~MP~~site 28). This may suggest that compared to WCA and LR, this approach is more sensitive to the extrapolation of results beyond the altitude ranges of the estimation gauges.

Furthermore, it was found that the quality of results of the GLMM are particularly sensitive to the number and location of gauges measuring temperature. As shown in Table ~~3~~4, the RMSE for this approach ~~rose considerably~~ rises sharply when only 8 (3.89^oC), 5 (3.99^oC) and 2 ~~gauges were~~ (14.44^oC)gauges are used to estimate its parameters. The ~~performance~~ performances of IDW and LR also ~~decreased~~ decrease considerably (RMSE of 9.34^oC and 7.78^oC respectively, with only two gauges), to the extent that using ~~estimates of long-term monthly average temperatures provided by the WC maps would be a better alternative~~ the raw WC maps for this case study (RMSE of 3.36^oC) may be preferable to any method other than WCA once the density of gauges becomes low.

On the other hand, WCA ~~was~~ is quite resilient to the reduction of estimation gauges~~, and its errors were lower, even with two gauges only (RMSE of~~. Even with two estimation gauges the average RMSE was only 2.54^oC~~)~~. This may be because the raw WC maps have internalised the average effect of elevation, longitude and latitude ~~on this climate variable~~ through the long-term analysis (a worldwide generalisation), which can then be adapted to local conditions by including a small number of gauges. This ~~suggest~~ suggests that WCA is an accurate and easy to use alternative to model air temperature in the case study.

Regarding precipitation, the LOOCV ~~showed~~ shows that all approaches have ~~a similar performance~~ similar performances in terms of RMSE~~($13-15mm$)~~, although the simple merge of CHIRPS and observations (ChA) has a slightly better value ($12.8mm$). However, the GLMM also stands out due to its lower FAR (7.1%), which may be a positive outcome of separating the analysis of precipitation into occurrence and magnitude. This could also be related to the fact that the GLMM analyses the ~~correlation between measurements~~randomness of occurrences and their spatial correlation (see Equation 7), thus limits the possibility of one or few gauges with non-zero precipitation overly influencing the precipitation estimate at all points (e.g. smoothing).

As opposed to this, other alternatives, particularly WCA and IDW, tend to predict precipitation when at least one (IDW) or even when no gauges (WCA - due to the inclusion of long-term averages) record non-zero values. This is ~~evidenced with~~ evident from the prediction of dry-season precipitation events that were never observed (see Figure 8). Preliminary results obtained using a different threshold (0.3 mm) for the detection of precipitation were similar, thus, the preference for GLMM in terms of FAR and POD performance seems not to be sensitive to the selection of this threshold.

When the ~~interpolation approaches were~~ precipitation interpolation approaches are tested with a reduced number of estimation gauges, it ~~was~~ is found that the ~~GLMM failed to maintain relatively good precipitation results, and its precipitation RMSE values rose~~ RMSE values of the GLMM rise drastically (beyond $100mm$ with 4 gauges only). Once more, this suggests that compared to the alternatives, in this case study the GLMM is more sensitive to the number and distribution of estimation gauges. The importance of the latter is highlighted when using only 14 gauges for model estimation but including at least one of the high elevation gauges ~~, Los Broncesor Lagunitas~~at site 17 (Los Bronces) or site 27 (Lagunitas). This gives an RMSE of $19mm$, which is considerably less than the ~~overall~~ average RMSE for the GLMM with 14 gauges(32.1 mm).

~~All other interpolation approaches for precipitation behaved similarly~~ The other precipitation interpolation approaches decrease their performance at a relatively similar rate, when facing a reduction in the number estimation gauges. As shown for the LOOCV (see Figure ~~??~~7B), this may be because errors ~~in~~ at high elevation gauges strongly influence the overall RMSE. When only 4 gauges ~~were~~ are included, however, ~~WCA showed a slightly~~ ChA and to a lesser extent WCA show a better RMSE ($21mm$ and $23.5mm$ ~~) but a~~ respectively), although the former has a relatively low POD (88.4%) and the latter a larger FAR (27.9%). It was also found that CHIRPS ~~,~~ as a standalone product ~~, represents~~ is a useful alternative ~~source of data (i.e. compared to the methods tested in this project, some of which are used in hydrological modelling)~~ to the interpolation approaches when 4 or ~~less gauges were available~~fewer gauges are available, with only marginally worse RMSE value than IDW and better RMSE than LR and GLMM ($RMSE = 26.2mm$, $POD = 88.5\%$ and $FAR = 28.6\%$).

The results in this paper show how ~~a simple approach that~~ simple approaches, which can be easily reproduced elsewhere, may perform at least as well as other more complex or more commonly used approaches, in a catchment with sparse monitoring networks and complex climate dynamics. Based on this evidence and its simplicity, it would be desirable to use WCA to ~~analyse~~ estimate temperature in this case study. For precipitation, ~~WCA is also~~ ChA or WCA may be preferable, unless the modeller ~~was~~ is particularly interested in the occurrence of precipitation in the dry season, in which case the GLMM would be desirable if computational requirements are not an issue ~~, or LR otherwise~~and there is a reasonable coverage of gauges. Analyses of further case studies are required to ~~be able to generalise~~ test the generality of these findings.

~~The fact that 15 temperature gauges in the mountain areas measured during one summer season only, or just started measuring values after 2008, means that although ideal to increase the reliability of results, an analysis of a longer period is still not possible. For precipitation, it would also have been desirable to have good quality gauges between 1300 and 2700 masl, to better understand what happens between the observations in low elevation points and the two high elevation gauges in the mine sites.~~

Beyond the issues with the number and location of gauges to estimate the parameters of the GLMM, this paper shows how approximate Bayesian inference methods can be applied to estimate parameters of these models in a hydrological context. Despite there being high computational requirements with the the R-INLA package, these are lower than those of MCMC, and this facilitates the use of GLMMs. It would now be useful to test if the benefits of GLMMs and Bayesian approaches discussed in this paper and in the non-hydrology literature (Pilz and Spöck, 2008; Ecker and Gelfand, 1997) can equally be achieved by stochastic approaches like Kriging and GLMs that are more common in hydro-climate applications. It would be particularly interesting to analyse how these approaches behave in well and poorly monitored regions, and how this influences hydrological modelling. ~~Furthermore~~

Results in this case study are of course limited by the fact that 15 temperature gauges in the mountain areas measured during one summer season only. For precipitation, it would also ~~be useful to analyse further prior distributions for the Bayesian estimation of the GLMM's parameters. This project used the default priors of the R-INLA package, to facilitate its implementation, but this could be enhanced~~have been desirable to have good quality gauges between 1300 and 2700 masl, to better understand what happens between the low and high elevation gauges.

6 Conclusions

Interpolation of climate variables is a major field of research in hydrology ~~, and its relevance is related to the importance of this data for~~ due to their importance in water resources modelling. ~~The scope of this paper was to compare four interpolation approaches for~~ This paper compared five approaches to interpolating temperature and precipitation gauged data in a catchment with complex and steep terrain, and ~~a low density network of gauges.~~ tested their sensitivity to the reduction of the number of estimation gauges. High elevation gauges, not previously used before for this type of research, were employed to partially test the ability of the approaches to extrapolate to the high Andes.

For temperature, ~~the~~ a Generalised Linear Mixed Model (GLMM) reproduced observations in this case study in the best way (i.e. smallest Root Mean Squared ~~Errors~~ Error - RMSE, in a leave-one-out cross validation - LOOCV), although it was closely followed by a ~~more simple~~ simpler alternative based on merging observations and WorldClim maps (WCA). The latter performed relatively well at all high elevation points. Inverse Distance Weighting (IDW) and Lapse Rates (LR - i.e. a linear regression using only elevation as a covariate) showed a worse performance.

Furthermore, for temperature only WCA demonstrated resilience to the reduction of the number of estimation gauges, ~~and so shows~~ showing good prospects for ~~using this alternative to generate input climate data of hydrological models~~ supporting hydrological modelling in sparsely monitored catchments. The GLMM, IDW and LR, ~~on the other hand,~~ had larger errors to the point that for this case study, ~~it was desirable to use the long-term temperature estimates in the raw WorldClim maps instead, when few gauges were made available~~for temperature interpolation using few gauges, long-term estimates from WorldClim maps gave better RMSE results.

For precipitation, ~~the LOOCV evidenced that~~ no alternative was clearly superior in ~~this case study in terms of RMSE~~the LOOCV, and this may be because errors in high elevation points~~seem to be more determining the overall RMSE , than the quality of calculations of each approach. All~~ , which were large irrespective of the approach, dominated the RMSE values. ChA showed smaller RMSE values in the sensitivity test (although it had lower probabilities of detection - POD), which highlights the desirability of the method in this case study, unless detecting most events was fundamental for the user. The other approaches showed a relatively similar resilience to the reduction of estimation gauges, except for the GLMM, which had a poorer performance with ~~a~~ an RMSE value larger than $100mm$ when only 4 gauges were ~~made available. WCA showed slightly better RMSE than the rest in most cases, which together with its simplicity, highlights the desirability of the method, even if it did not outperform others as much as for temperature.~~ used.

In terms of the added value of alternative datasets, it was found that in this case study the inclusion of CHIRPS and WorldClim ~~maps was as relevant and illustrative as the comparison of interpolation approaches~~was valuable. On the one hand, ~~WorldClim maps allowed developing a very simple but quite~~ using the residuals between WorldClim maps or CHIRPS and climate observations, represented a simple but efficient method that showed good performance and high resilience when working with few gauges. On the other hand, CHIRPS~~was~~ , as a standalone product, demonstrated to be a useful source of precipitation data ~~where few gauges are available. Thus, both represent alternatives to support the development of water resources models in regions with few point observations.~~

~~The paper also illustrated how approximate Bayesian inference methods, particularly the INLA-SPDE method available in the R-INLA package, can be used to estimate the parameters of spatio temporal models in hydrological contexts. Further research could explore other spatio-temporal models and prior distributions, and how the results of this approach compare to those of GLMs or Kriging~~when no or few gauges were available.

Appendix A: Gauges Used

Author contributions. Juan Ossa-Moreno, Neil McIntyre and Greg Keir conceived the paper. Juan Ossa-Moreno drafted the paper, did the literature review and conducted the research. Greg Keir helped writing the code to run the GLMM. Neil McIntyre advised on hydrological processes and interpolation of climate variables. Michela Cameletti advised on the use of INLA-SPDE, and Diego Rivera advised on the Chilean context, on the interpolation of climate variables in mountain regions and on the literature review. All authors reviewed the paper drafts.

Competing interests. There are no competing interests.

Acknowledgements. This research was supported by use of the NeCTAR Research Cloud and by QCIF (http://www.qcif.edu.au). The NeCTAR Research Cloud is a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy. FlashLite was supported by ARC LIEF grant LE140100061, The University of Queensland, Queensland University of Technology, Griffith University, Monash University, University of Technology Sydney, and The Queensland Cyber Infrastructure Foundation. The University of Queensland International Scholarships support the main author. The authors would like to thank the support from James McPhee in Universidad de Chile, Codelco and Anglo-American for providing climate data to the project.

References

Aalto, J., Pirinen, P., Heikkinen, J., and Venäläinen, A.: Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models, Theoretical and Applied Climatology, pp. 1–13, 2013.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Cortes, G., Garreaud, R., and McPhee, J.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies-Chile dataset, Hydrology and Earth System Sciences, 22, 5817–5846, 2018.

Álvarez-Villa, O. D., Vélez, J. I., and Poveda, G.: Improved long-term mean annual rainfall fields for Colombia, International Journal of Climatology, 31, 2194–2212, 2011.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., and Pappenberger, F.: ERA-Interim/Land: a global land surface reanalysis data set, Hydrology and Earth System Sciences, 19, 389–407, 2015.

Benavides, R., Montes, F., Rubio, A., and Osoro, K.: Geostatistical modelling of air temperature in a mountainous region of Northern Spain, Agricultural and Forest Meteorology, 146, 173–188, 2007.

Bivand, R. S., Pebesma, E. J., and Gómez-Rubio, V.: Applied spatial data analysis with R, 2013.

Blangiardo, M. and Cameletti, M.: Spatial and spatio-temporal Bayesian models with R-INLA, John Wiley & Sons, 2015.

Blangiardo, M., Cameletti, M., Baio, G., and Rue, H.: Spatial and spatio-temporal models with R-INLA, Spatial and spatio-temporal epidemiology, 7, 39–55, 2013.

Buytaert, W., Celleri, R., Willems, P., De Bievre, B., and Wyseure, G.: Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes, Journal of hydrology, 329, 413–421, 2006.

Cameletti, M., Ignaccolo, R., and Bande, S.: Comparing spatio-temporal models for particulate matter in Piemonte, Environmetrics, 22, 985–996, 2011.

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H.: Spatio-temporal modeling of particulate matter concentration through the SPDE approach, AStA Advances in Statistical Analysis, 97, 109–131, 2013.

Castro, L. M., Gironás, J., and Fernández, B.: Spatial estimation of daily precipitation in regions with complex relief and scarce data using terrain orientation, Journal of Hydrology, 517, 481–492, 2014.

Castro, L. M., Miranda, M., and Fernández, B.: Evaluation of TRMM Multi-satellite precipitation analysis (TMPA) in a mountainous region of the central Andes range with a Mediterranean climate, Hydrology Research, 46, 89–105, 2015.

Chandler, R. E. and Wheater, H. S.: Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland, Water Resources Research, 38, 10–1–10–11, 2002.

Chen, F.-W. and Liu, C.-W.: Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan, Paddy and Water Environment, 10, 209–222, 2012.

Correa-Ibanez, R., Keir, G., and McIntyre, N.: Climate-resilient water supply for a mine in the Chilean Andes, Proceedings of the Institution of Civil Engineers-Water Management, 2017.

Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L., and Pasteris, P.: A knowledge-based approach to the statistical mapping of climate, Climate research, 22, 99–113, 2002.

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, 28, 2031–2064, 2008.

Demaria, E. M. C., Maurer, E. P., Sheffield, J., Bustos, E., Poblete, D., Vicuña, S., and Meza, F.: Using a gridded global dataset to characterize regional hydroclimate in central Chile, Journal of Hydrometeorology, 14, 251–265, 2013.

DGA: Actualizacion del Balance Hidrico Nacional, Report, 2017.

Dinku, T., Ruiz, F., Connor, S., and Ceccato, P.: Validation and Intercomparison of Satellite Rainfall Estimates over Colombia, Journal of Applied Meteorology and Climatology, 49, 2010.

Dinku, T., Hailemariam, K., Maidment, R., Tarnavsky, E., and Connor, S.: Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia, International Journal of Climatology, 34, 2489–2504, 2014.

Diodato, N.: The influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain, International Journal of Climatology, 25, 351–363, 2005.

Dorninger, M., Schneider, S., and Steinacker, R.: On the interpolation of precipitation data over complex terrain, Meteorology and Atmospheric Physics, 101, 175–189, 2008.

Ecker, M. D. and Gelfand, A. E.: Bayesian variogram modeling for an isotropic spatial process, Journal of Agricultural, Biological, and Environmental Statistics, 2, 347–369, 1997.

Falvey, M. and Garreaud, R.: Wintertime Precipitation Episodes in Central Chile: Associated Meteorological Conditions and Orographic Influences, Journal of Hydrometeorology, 8, 171–193, 2007.

Faraway, J. J.: Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models, vol. 124, CRC press, 2016.

Frei, C.: Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances, International Journal of Climatology, 34, 1585–1605, 2014.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., and Hoell, A.: The climate hazards infrared precipitation with stations–a new environmental record for monitoring extremes, Scientific data, 2, 2015.

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C., and Verdin, A. P.: A quasi-global precipitation time series for drought monitoring, US Geological Survey Data Series, 832, 2014.

Garreaud, R.: Warm winter storms in Central Chile, Journal of Hydrometeorology, 14, 1515–1534, 2013.

Garreaud, R., Alvarez-Garreton, C., Barichivich, J., Boisier, J. P., Christie, D., Galleguillos, M., LeQuesne, C., McPhee, J., and Zambrano-Bigiarini, M.: The 2010–2015 mega drought in Central Chile: Impacts on regional hydroclimate and vegetation, Hydrology and Earth System Sciences Discussions, 2017, 1–37, http://www.hydrol-earth-syst-sci-discuss.net/hess-2017-191/, 2017.

Garreaud, R. D., Vuille, M., Compagnucci, R., and Marengo, J.: Present-day south american climate, Palaeogeography, Palaeoclimatology, Palaeoecology, 281, 180–195, 2009.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, Journal of hydrology, 228, 113–129, 2000.

Hasenauer, H., Merganicova, K., Petritsch, R., Pietsch, S. A., and Thornton, P. E.: Validating daily climate interpolations over complex terrain in Austria, Agricultural and Forest Meteorology, 119, 87–107, 2003.

Hengl, T., Heuvelink, G. B., and Stein, A.: Comparison of kriging with external drift and regression kriging, Technical Note. ITC Enschede, Netherlands. Available online at $https://webapps.itc.utwente.nl/librarywww/papers_{2}003/misca/hengl_{c}omparison.pdf$, 2003.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, International journal of climatology, 25, 1965–1978, 2005.

Hobouchian, M. P., Salio, P., Skabar, Y. G., Vila, D., and Garreaud, R.: Assessment of satellite precipitation estimates over the slopes of the subtropical Andes, Atmospheric Research, 190, 43–54, 2017.

Hutchinson, M. F.: Anusplin Version 4.3. Centre for Resource and Environmental Studies, The Australian National University: Canberra, Australia., 2004.

Jacoby, W. G.: Loess:: a nonparametric, graphical tool for depicting relationships between variables, Electoral Studies, 19, 577–613, 2000.

Jacquin, A. P. and Soto-Sandoval, J. C.: Interpolation of monthly precipitation amounts in mountainous catchments with sparse precipitation networks, Chilean journal of agricultural research, 73, 406–413, 2013.

Janke, J. R., Ng, S., and Bellisario, A.: An inventory and estimate of water stored in firn fields, glaciers, debris-covered glaciers, and rock glaciers in the Aconcagua River Basin, Chile, Geomorphology, 296, 142–152, 2017.

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database, http://srtm.csi.cgiar.org, 2008.

Kenabatho, P., McIntyre, N., Chandler, R., and Wheater, H.: Stochastic simulation of rainfall in the semi-arid Limpopo basin, Botswana, International Journal of Climatology, 32, 1113–1127, 2012.

Kigobe, M., McIntyre, N., Wheater, H., and Chandler, R.: Multi-site stochastic modelling of daily rainfall in Uganda, Hydrological sciences journal, 56, 17–33, 2011.

Kéry, M. and Royle, J. A.: Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and Static Models, Academic Press, 2015.

Li, J. and Heap, A. D.: Spatial interpolation methods applied in the environmental sciences: A review, Environmental Modelling & Software, 53, 173–189, 2014.

Lindgren, F., Rue, H., and Lindstrom, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, Journal of the Royal Statistical Society, Series B, 73, 423–98, 2011.

Lu, G. Y. and Wong, D. W.: An adaptive inverse-distance weighting spatial interpolation technique, Computers & Geosciences, 34, 1044–1055, 2008.

Manz, B., Buytaert, W., Zulkafli, Z., Lavado, W., Willems, B., Robles, L. A., and Rodriguez-Sanchez, J.: High-resolution satellite-gauge merged precipitation climatologies of the Tropical Andes, Journal of Geophysical Research: Atmospheres, 121, 1190–1207, 2016.

Masson, D. and Frei, C.: Spatial analysis of precipitation in a high-mountain region: exploring methods with multi-scale topographic predictors and circulation types, Hydrology and Earth System Sciences, 18, 4543–4563, 2014.

Mernild, S. H., Liston, G. E., Hiemstra, C., Beckerman, A. P., Yde, J. C., and McPhee, J.: The Andes Cordillera. Part IV: spatio-temporal freshwater run-off distribution to adjacent seas (1979–2014), International Journal of Climatology, 37, 3175–3196, 2016.

Mernild, S. H., Liston, G. E., Hiemstra, C. A., Malmros, J. K., Yde, J. C., and McPhee, J.: The Andes Cordillera. Part I: snow distribution, properties, and trends (1979–2014), International Journal of Climatology, 37, 1680–1698, 2017.

Meza, F. J., Vicuña, S., Jelinek, M., Bustos, E., and Bonelli, S.: Assessing water demands and coverage sensitivity to climate change in the urban and rural sectors in central Chile, Journal of Water and Climate Change, 5, 192–203, 2014.

Minasny, B. and McBratney, A. B.: The Matérn function as a general model for soil variograms, Geoderma, 128, 192–207, 2005.

Müller, W. G.: Least-squares fitting from the variogram cloud, Statistics & probability letters, 43, 93–98, 1999.

Montecinos, A. and Aceituno, P.: Seasonality of the ENSO-related rainfall variability in central Chile and associated circulation anomalies, Journal of Climate, 16, 281–296, 2003.

Nerini, D., Zulkafli, Z., Wang, L.-P., Onof, C., Buytaert, W., Lavado, W., and Guyot, J.-L.: A comparative analysis of TRMM-rain gauge data merging techniques at the daily time scale for distributed rainfall-runoff modelling applications, Journal of Hydrometeorology, 2015.

Nikolopoulos, E. I., Anagnostou, E. N., and Borga, M.: Using high-resolution satellite rainfall products to simulate a major flash flood event in northern Italy, Journal of Hydrometeorology, 14, 171–185

Ohlanders, N., Rodriguez, M., and Mc Phee Torres, J.: Stable water isotope variation in a Central Andean watershed dominated by glacier and snowmelt, Hydrology and Earth System Sciences, 9, 2013.

Pellicciotti, F., Burlando, P., and Vliet, K. V.: Recent trends in precipitation and streamflow in the Aconcagua River basin, central Chile, IAHS Publications-Series of Proceedings and Reports, 318, 17–38, 2007.

Pellicciotti, F., Helbing, J., Carenzo, M., and Burlando, P.: Changes with elevation in the energy balance of an Andean Glacier, Juncal Norte Glacier, dry Andes of central Chile, in: EGU General Assembly Conference Abstracts, vol. 12, p. 5302, 2010.

Pellicciotti, F., Ragettli, S., Carenzo, M., and McPhee, J.: Changes of glaciers in the Andes of Chile and priorities for future work, Science of The Total Environment, 493, 1197–1210, 2014.

Petersen, L. and Pellicciotti, F.: Spatial and temporal variability of air temperature on a melting glacier: atmospheric controls, extrapolation methods and their effect on melt modeling, Juncal Norte Glacier, Chile, Journal of Geophysical Research: Atmospheres, 116, 2011.

Pilz, J. and Spöck, G.: Why do we need and how should we implement Bayesian kriging methods, Stochastic Environmental Research and Risk Assessment, 22, 621–632, 2008.

Ragettli, S. and Pellicciotti, F.: Calibration of a physically based, spatially distributed hydrological model in a glacierized basin: On the use of knowledge from glaciometeorological processes to constrain model parameters, Water Resources Research, 48, 2012.

Ragettli, S., Cortés, G., McPhee, J., and Pellicciotti, F.: An evaluation of approaches for modelling hydrological processes in high-elevation, glacierized Andean watersheds, Hydrological Processes, 28, 5674–5695, 2014.

Rue, H., Martino, S., and Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, Journal of the royal statistical society: Series b (statistical methodology), 71, 319–392, 2009.

Rue, H., Martino, S., Lindgren, F., Simpson, D., and Riebler, A.: R-INLA: Approximate Bayesian Inference Using Integrated Nested Laplace Approximations, Trondheim, Norway. URL http://www.r-inla.org, 2013.

Stahl, K., Moore, R. D., Floyer, J. A., Asplin, M. G., and McKendry, I. G.: Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density, Agricultural and Forest Meteorology, 139, 224–236, 2006.

Stehr, A., Debels, P., Romero, F., and Alcayaga, H.: Hydrological modelling with SWAT under conditions of limited data availability: evaluation of results from a Chilean case study, Hydrological sciences journal, 53, 588–601, 2008.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V., and De Roo, A.: Validation of satellite-based precipitation products over sparsely gauged African river basins, Journal of Hydrometeorology, 13, 1760–1783, 2012.

Thornton, P. E., Running, S. W., and White, M. A.: Generating surfaces of daily meteorological variables over large regions of complex terrain, Journal of Hydrology, 190, 214–251, 1997.

Thornton, P. E., Thornton, M. M., Mayer, B. W., Wilhelmi, N., Wei, Y., Devarakonda, R., and Cook, R. B.: Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2, 2014.

Tobin, K. J. and Bennett, M. E.: Validation of Satellite Precipitation Adjustment Methodology From Seven Basins in the Continental United States1, Journal of The American Water Resources Association, 48, 221–234, 2012.

Viale, M. and Garreaud, R.: Summer precipitation events over the western slope of the subtropical Andes, Monthly Weather Review, 142, 1074–1092, 2014.

Viale, M. and Garreaud, R.: Orographic effects of the subtropical and extratropical Andes on upwind precipitating clouds, Journal of Geophysical Research: Atmospheres, 120, 4962–4974, 2015.

Vicuña, S., Garreaud, R. D., and McPhee, J.: Climate change impacts on the hydrology of a snowmelt driven basin in semiarid Chile, Climatic Change, 105, 469–488, 2011.

Wolter, K. and Timlin, M. S.: El Nino/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext), International Journal of Climatology, 31, 1074–1087, 2011.

Wu, T. and Li, Y.: Spatial interpolation of temperature in the United States using residual kriging, Applied Geography, 44, 112–120, 2013.

Yang, C., Chandler, R., Isham, V., and Wheater, H.: Spatial-temporal rainfall simulation using generalized linear models, Water Resources Research, 41, 2005.

Yao, X., Fu, B., Lü, Y., Sun, F., Wang, S., and Liu, M.: Comparison of four spatial interpolation methods for estimating soil moisture in a complex terrain catchment, PloS one, 8, e54 660–e54 660, 2013.

Zambrano, F., Wardlow, B., Tadesse, T., Lillo-Saavedra, M., and Lagos, O.: Evaluating satellite-derived long-term historical precipitation datasets for drought monitoring in Chile, Atmospheric Research, 186, 26–42, 2017.

Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., and Ribbe, L.: Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile, Hydrology and Earth System Sciences, 2016.

Table A1. Details of gauges used

	Station	Elevation	Long	Lat	Variable	Dates available	% of Missing Gaps in the 5 year period
1	05200007-6	1202	-70.68	-32.42	P	All period	0
2	05403006-1	1313	-70.36	-32.92	P	All period	1.67
3	05410002-7	954	-70.51	-32.85	P	All period	5
4	05410005-1	642	-70.74	-32.76	P	All period	3.33
5	05410006-K	1078	-70.47	-32.86	P	All period	0
6	05410007-8	830	-70.6	-32.83	P	All period	0
7	05410008-6	650	-70.72	-32.75	P	All period	0
8	05414001-0	1193	-70.58	-32.50	P	All period	23.33
9	05414004-5	1209	-70.57	-32.49	P	All period	5
10	05414005-3	943	-70.7	-32.57	P	All period	0
11	05415004-0	1023	-70.6	-32.68	P	All period	1.67
12	05422002-2	835	-70.82	-32.93	P	All period	1.67
13	05732001-K	575	-70.8	-33.09	P	All period	1.67
14	05732002-8	597	-70.77	-33.08	P	All period	5
15	05733006-6	973	-70.75	-32.95	P	All period	0
16	05733010-4	809	-70.81	-32.95	P	All period	1.67
17	Los Bronces	3423	-70.29	-33.15	P	All period	0
18	330019	654	-70.55	-33.45	T	All period	44.58
19	330020	529	-70.68	-33.45	T	All period	0.16
20	330021	481	-70.79	-33.39	T	All period	1.04
21	AWS1	3088	-70.11	-32.99	T	Summer 08-09	96.11
22	AWS2	2785	-70.11	-32.97	T	Summer 08-09	96.11
23	AWS3	3269	-70.1	-33	T	Summer 08-09	96.66
24	Angela	3573	-70.27	-33.08	T	All period	1.81
25	Barroso	3776	-70.23	-33.11	T	All period	4.05
26	Hornitos	2214	-70.15	-32.87	T	From Sept/12	80.01
27	Lagunitas	2922	-70.25	-33.08	P and T	All period	0
28	MP	4080	-70.26	-33.17	T	All period	2.35
29	Saladillo	1585	-70.28	-32.93	T	From Dec/11	66.32
30	TLog1	3254	-70.1	-33	T	Summer 08-09	96.22
31	TLog10	3004	-70.11	-32.99	T	Summer 08-09	96.22
32	TLog11	2968	-70.11	-32.98	T	Summer 08-09	96.22
33	TLog12	2911	-70.11	-32.98	T	Summer 08-09	96.22
34	TLog2	3269	-70.1	-33	T	Summer 08-09	96.22
35	TLog3	3269	-70.1	-33	T	Summer 08-09	96.22
36	TLog4	3212	-70.1	-33	T	Summer 08-09	96.22
37	TLog5	3153	-70.11	-32.99	T	Summer 08-09	96.22
38	TLog6	3081	-70.11	-32.99	T	Summer 08-09	96.22
39	TLog7	3094	-70.11	-32.99	T	Summer 08-09	96.22
40	TLog8	3092	-70.11	-32.99	T	Summer 08-09	96.22
41	TLog9	3070	-70.11	-32.99	T	Summer 08-09	96.22