# 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 other products) and new computational tools are merged with ground observations to address this problem. This paper presents a comparison of approaches of different complexity to spatially interpolate 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 simpler alternatives: Inverse Distance Weighting, Lapse Rates and 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 temperature and precipitation, the approaches based on the WorldClim or CHIRPS residuals may be recommended as being the more accurate, easy to apply and relatively robust to tested reductions in the number of estimation gauges. 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. Generally, spatial interpolation of point observations is 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). 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), to cold-air trapped in enclosing hill ranges (Frei, 2014), and also to the presence of glaciers (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) and to summertime convective precipitation events (Viale and Garreaud, 2014).

The Andes Cordillera in South America is an example of a steep terrain with 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; 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., 2017). 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; Demaria et al., 2013).

In the Andes, Álvarez-Villa et al. (2011) tested four stochastic interpolation approaches in Colombia and found that Kriging with External Drift (KED - 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 that 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. that 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 that 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. in the 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., 2012).

5    In a comprehensive analysis of precipitation estimates from satellite products in Chile, Zambrano-Bigiarini et al. (2016) found that the satellite product PGFv3 (Sheffield et al., 2006) exhibited the best overall performance for the country, followed by CHIRPS (Funk et al., 2015), TMPA 3B42V7 (Huffman et al., 2007) and MSWEPv1.1 (Beck et al., 2017). The authors mention that the superior performance of PGFv3 is likely due to the 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

10   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 (Hsu et al., 1997)) 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

15   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.

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

20   precipitation magnitudes 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

25   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

30   support hydrological and water resources models in this region.

The aim of this paper is to compare five precipitation and four temperature interpolation approaches in the Upper Aconcagua River in central Chile, a mountainous catchment with steep and complex topography. The paper builds on the literature by: (1) 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) comparing the interpolation approaches, focusing on the relative performance

35   of the simple and complex ones. (3) Testing the sensitivity of the methodologies to the number of available gauges.

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 along the border of Chile and Argentina, where most of the subcatchments run south to north (Blanco and Juncal) or north to south (Colorado). Once these streams connect and create the Aconcagua River, it flows east to west towards the Pacific Ocean. Topography fluctuates from coastal areas to peaks of approximately 5900 m above sea level. The catchment has an area of 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 Aconcagua catchment is Mediterranean, close to semi-arid conditions (Ohlanders et al., 2013). Annual average precipitation is approximately 350 mm, however, most of this is concentrated during the austral winter (frontal rain-storms 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 orographic effects, with the windward slope of the Andes having much more precipitation during winter, compared to the leeward side (Viale and Garreaud, 2015; Viale and Nuñez, 2011).There are few wind gauges in the area, which hinders understanding of the small-scale preferential wind directions and their influence on precipitation. However, it has been reported that at a macro scale the prevalent direction is west to east (Castro et al., 2014; Garreaud et al., 2009).

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).

There is relevant inter-annual variability in the climate variables related to El Niño and La Niña phases (ENSO) (Garreaud et al., 2009). La Niña is an anomalous cooling of the southeastern Pacific leading to dry conditions in Central Chile when the Pacific Anticyclone strengthens, while wet conditions occur during El Niño (Montecinos and Aceituno, 2003; Pellicciotti et al., 2007). Inter-decadal variability related to the Pacific Decadal Oscillation may also affect the case study, although the causes and impacts of these low-frequency fluctuations are less understood than those of ENSO (Garreaud et al., 2009).

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 relevant during very dry years only (Ohlanders et al., 2013).
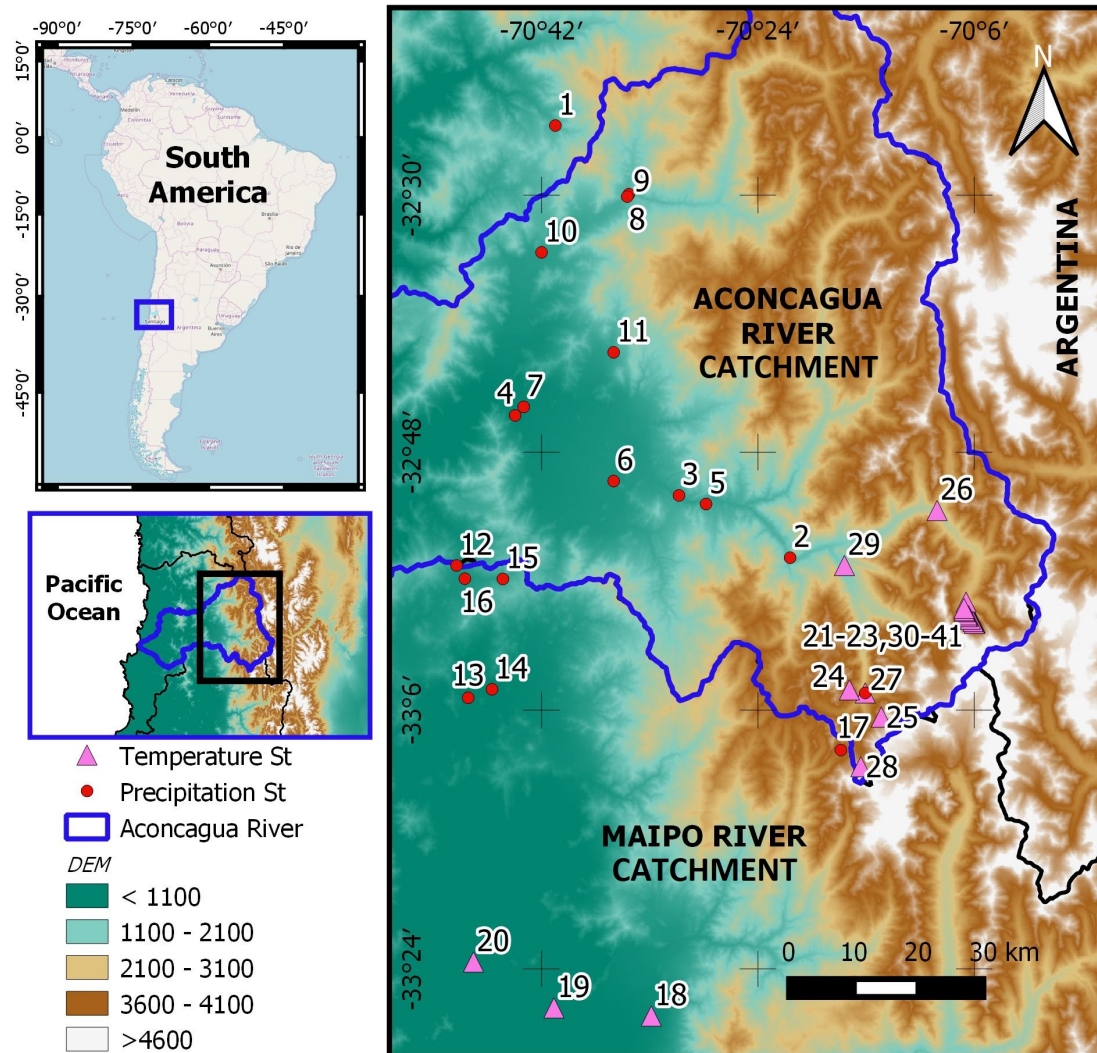
**Figure 1.** Temperature and precipitation gauges in the catchment with available data during the period of analysis. Further details of the gauges are provided in the Appendix.

## 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 sourced from mine projects in the area. Amongst these high-elevation gauges operated by mining companies, there are two that record liquid and solid precipitation (sites 27 and 17, see the Appendix for more details). The latter were transformed to snow water equivalents (SWE) before being analysed here.

These data 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 nearby a major glacier 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 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.

The period of analysis spans from September 2008 to August 2013 as most of the data obtained from the high elevation gauges were 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. In addition, Figure 3 shows monthly average precipitation and monthly observations at three representative gauges throughout the same period (see Figure 1 for the location of these gauges).

Many of the precipitation and temperature data from government institutions have previously been published (Castro et al., 2014; Jacquin and Soto-Sandoval, 2013; Vicuña et al., 2011; Ragettli et al., 2014; Zambrano-Bigiarini et al., 2016), and it has been reported that the precision of the DGA precipitation gauges is 0.1 mm (Castro et al., 2014). In addition, the characteristics of the gauges used by Pellicciotti et al. (2008) were reviewed and the gauges precision was deemed sufficient for the purposes of this project. In order to complement this, data quality was analysed using double mass plots, and plots of the relationship between elevation and the climate variables. This led to the exclusion of precipitation measurements at site 26. Furthermore, all precipitation observations at site 29 were also excluded following the suggestion from one of the mining companies, which stated that the gauge at this site had technical issues when measuring solid precipitation. The temperature measurements at both 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 Precipitation with Station data (CHIRPS) satellite product (Funk et al., 2015) was used. 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). Most other products (e.g. TMPA 3B42v7, MSWEP and PGFv3) are relatively coarse for the size of the catchment ($0.25^o$ pixels). CHIRPS does not include estimates of temperature and therefore was only used to support interpolation of precipitation.

The WorldClim (WC) Version 1 maps (Hijmans et al., 2005) were a further source of data. 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.
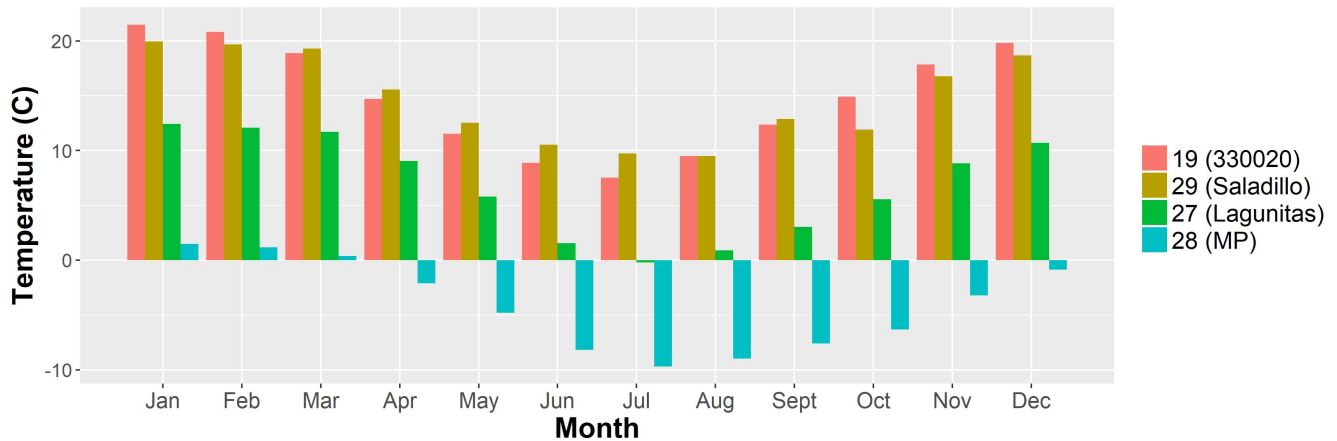
**Figure 2.** Monthly temperature averaged over the period of analysis, 09/2008 - 08/2013, for four of the gauges in the catchment. The numbers in the legend correspond to those in Figure 1, while the text in parenthesis are the names of the gauges. 330020 (527 masl), Saladillo (1580 masl), Lagunitas (2765.5 masl), MP (4250 masl).

WC data provide a historical average for each one of the 12 calendar months (one map for every month) and originate 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 the WC data warn about its potential inaccuracies in mountainous areas (Hijmans et al., 2005).

5 Therefore, the WC data were used only to complement point observations or as a benchmark for testing other interpolation approaches.

Although different in essence, both WC and CHIRPS can be used as a complement to point observations to construct daily or monthly interpolated fields. None of the selected gauged data were used as inputs 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.

10 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, an ENSO (El Niño-Southern Oscillation) index was included to analyse the inter-annual variability of precipitation in the catchment (Wolter and Timlin, 2011).

15 **3    Interpolation of Climate Data**

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

---

[1]The name of the gauges used to calibrate CHIRPS can be checked here.
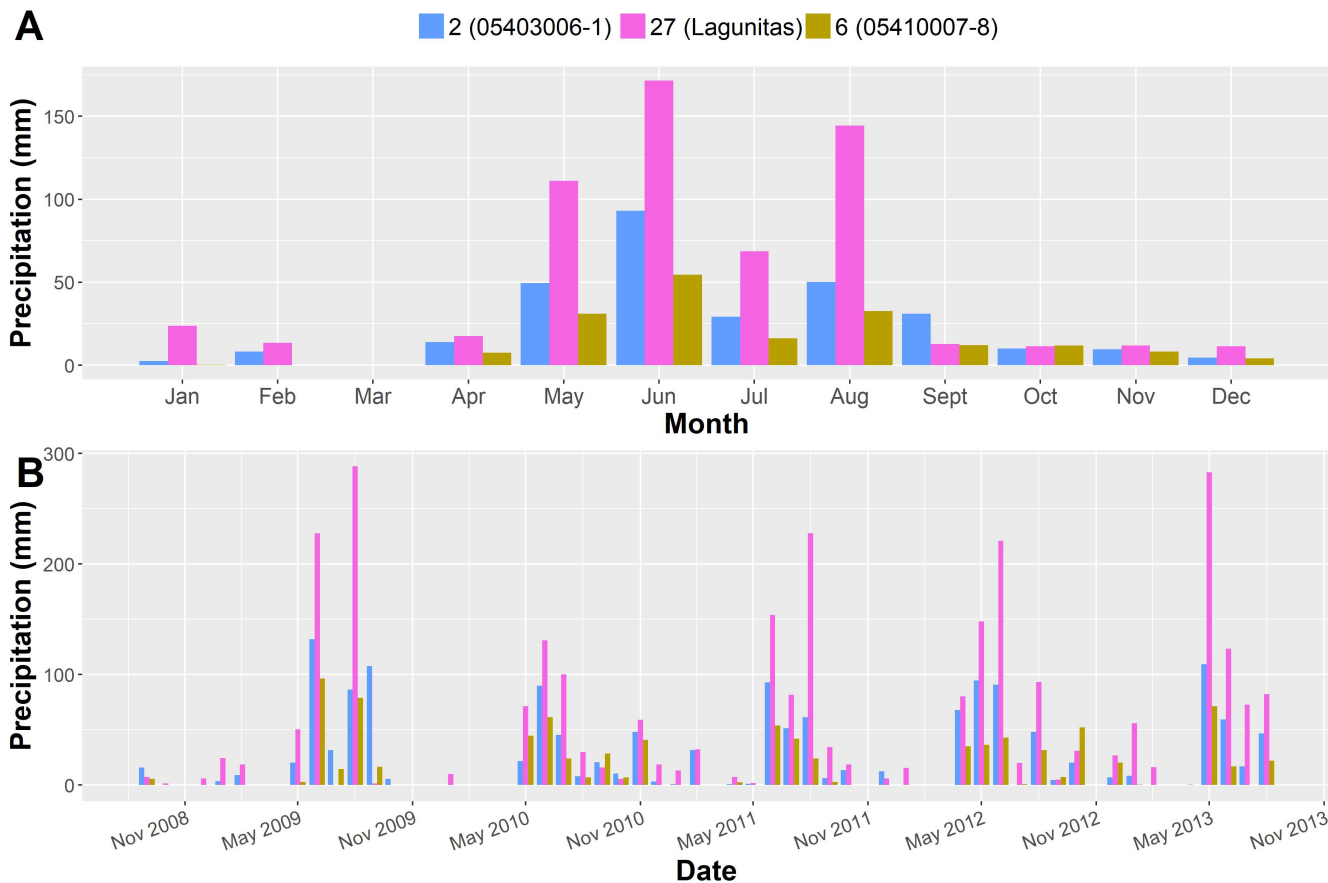
**Figure 3.** (A) Monthly average precipitation and (B) Monthly precipitation, over the period of analysis (09/2008 - 08/2013) for three of the gauges in the catchment. The numbers in the legend correspond to those in Figure 1, while the texts in parenthesis are the names of the gauges. Lagunitas (2765.5 masl), 05410007-8 (820 masl) and 05403006-1 (1290 masl).

and alternative datasets. The first of these two methods uses IDW to interpolate the residuals between WC maps and gauged values (precipitation and temperature), 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.

    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.

**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. | 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; Ragettli et al., 2014; Vicuña et al., 2011; Stehr et al., 2008) |
| 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-Bigiarini et al., 2016; Manz et al., 2016) |
| 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 coordinates 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., 2009) |
| 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 coordinates 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 Cameletti, 2015) |

## 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). 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 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 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). However, 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 (Rue et al., 2009).

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 $\mu$ 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), $\mu_i$ can be connected to a structured additive predictor $\eta_i$ through a link function $g(\ )$ as shown below (Rue et al., 2009):

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

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), $\epsilon(s_i,t) \sim N(0, \sigma_\epsilon^2)$ is the random error component, the $f^{(j)}(u_{j(s_i,t)})$ are functions of covariates $u$ and the $\beta s$ 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 $Y(.,.)$ 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), $\beta$s are the coefficients of the covariates, $\epsilon$ is the measurement/observation error component, both serially and spatially uncorrelated ($\epsilon(s_i, t) \sim N(0, \sigma_\epsilon^2)$), and $\xi$ 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 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 variable $\Phi(s_i, t)$ is spatially correlated but independent of other time-steps. The model is described as follows:

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

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

Both Eq. 7 and Eq. 8 share the same covariate coefficients $\beta^P$s, but the latter has an extra parameter ($\beta^{P'}$) connecting the random field in both equations.

It is acknowledged that 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 particulate matter), and this represents a subject for future research. However, taking into account the scope of the paper, it was desired to work with existing GLMMs in the literature (or close adaptations) that have been analysed with the INLA-SPDE approach.

## 3.2 Deterministic approaches

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 or logarithmic regressions to model the relation between temperature or precipitation and elevation. The regressions could be extended to include all the covariates used in 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 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.

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.

## 3.3 Comparison of interpolation approaches

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.

For temperature there was a total of 24 gauges available, thus, the LOOCV analysed 24 combinations of 23 gauges. For precipitation there were 18 gauges available, thus the LOOCV involved analysing 18 combinations of 17 gauges.

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).

This was complemented with an analysis of the distribution of the residuals. 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.

**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 R-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

### 4.1 Preliminary analysis of correlations between covariates and climate variables

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

Figure 4C and Figure 4D show that the correlation between CHIRPS and daily precipitation observations is weak, but considerably improves when both are aggregated to monthly values ($\rho = 0.81$). The $\rho$ for monthly precipitation and WC values is lower but significant ($\rho = 0.62$), while monthly correlation with elevation is above 0.6 for most months. ENSO shows a weak correlation with precipitation $\rho = 0.12$ (not shown in Figure 4 for brevity purposes), and similar results are obtained with time-delayed (1 - 4 months) ENSO index (i.e. as done in Francou et al. (2004)). However, a monthly analysis shows that for several months the correlation is close to $\rho = 0.5$, therefore, it was decided to keep ENSO as a covariate for the precipitation GLMM. These correlations may be stronger in longer-term analyses that cover several Niño-Niña cycles, which last approximately 2-5 years each (Wolter and Timlin, 2011; Garreaud et al., 2017; Montecinos and Aceituno, 2003).

### 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. It was found that the GLMM and WCA have the best performance, while LR and particularly IDW have larger RMSE values.

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

| Approach | RMSE ($^{o}C$) |
| --- | --- |
| GLMM | 1.20 |
| WCA | 1.22 |
| LR | 1.53 |
| IDW | 2.72 |

Table 4 shows the results of the sensitivity analysis. As expected, 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).

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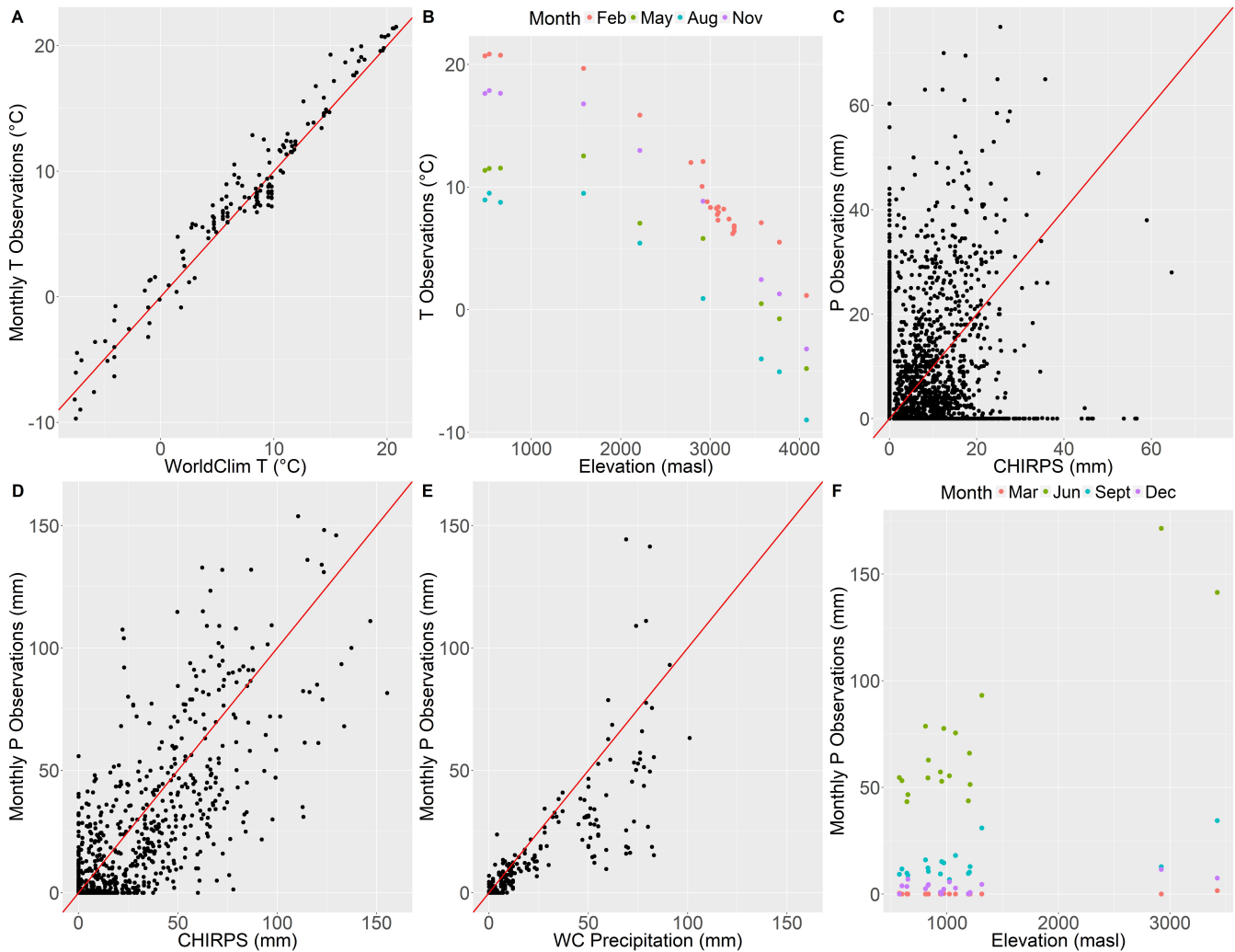
**Figure 4.** (A) WC values versus monthly aggregated (averaged) temperature values. (B) Elevation of gauges versus average temperature in four months. (C) CHIRPS versus precipitation. Daily values for all stations used. (D) Monthly aggregated (sum) CHIRPS versus monthly aggregated (sum) precipitation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

**Table 4.** Sensitivity test of the temperature interpolation approaches.

| Approach | Number of estimation gauges | RMSE ($^oC$) |
|---|---|---|
| GLMM | 8 | 3.89 |
| | 5 | 3.99 |
| | 2 | 14.44 |
| WCA | 8 | 1.77 |
| | 5 | 1.98 |
| | 2 | 2.54 |
| | 0 (Raw WC Maps)* | 3.36 |
| LR | 8 | 2.12 |
| | 5 | 4.14 |
| | 2 | 7.78 |
| IDW | 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).**

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 reproduce the observed temperatures relatively well except for site 28 (the one at the highest elevation - 4250 masl). LR and particularly IDW tend to underestimate temperature at all gauges, except
5   at site 28 where they overestimate it.

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

Figure 6 shows histograms of the validation residuals. It can be seen that the GLMM, WCA and LR 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. Figure 7A shows the relationship between temperature RMSE values and elevation.

15  **4.3 Precipitation results**

Table 5 shows that the performances of all interpolation approaches are relatively similar, in terms of RMSE, although ChA has slightly smaller RMSE values. All probability of detection (POD) indices are above 90%, although WCA and IDW have
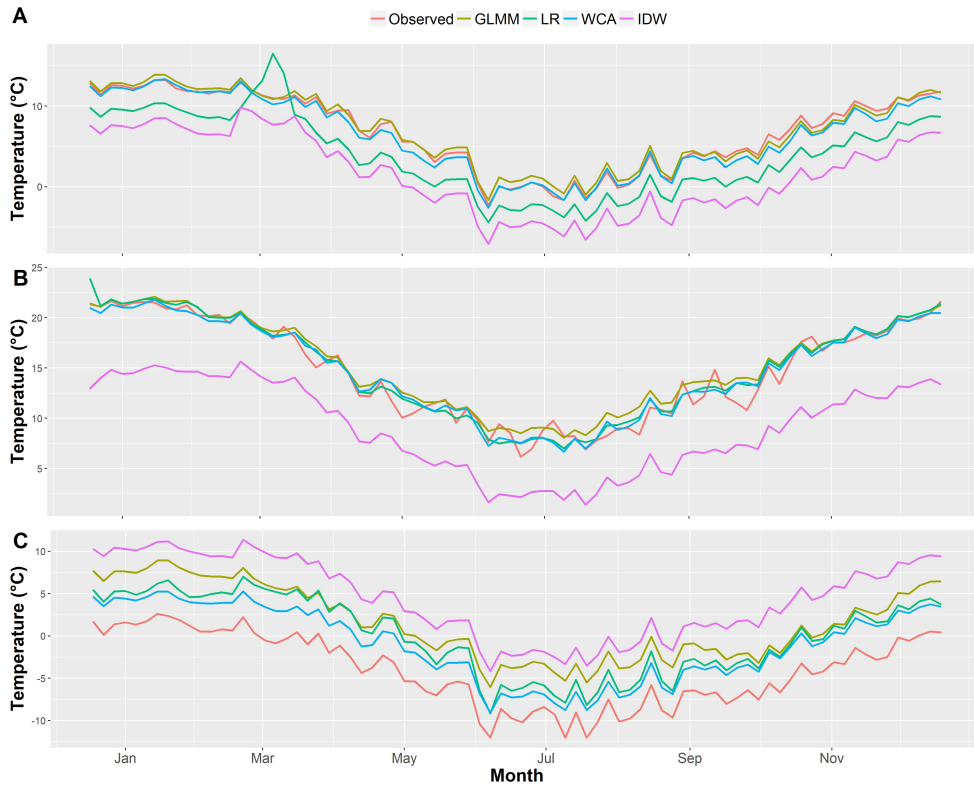
**Figure 5.** Daily temperature averaged over the 5 years of analysis for gauge (A) Site 27 (Lagunitas) (B) Site 18 (330019) (C) Site 28 (MP) (All curves were smoothed using the LOESS method (Jacoby, 2000) with $\alpha = 0.045$, this is similar to a moving average and is used to facilitate the visualisation of the main trends only).

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.** 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    |

Table 6 shows the sensitivity of performances to reductions in the number of estimation gauges. It can be seen that the GLMM is quite sensitive to these changes, and its RMSE performance decreases sharply when moving from 17 to 14 gauges,
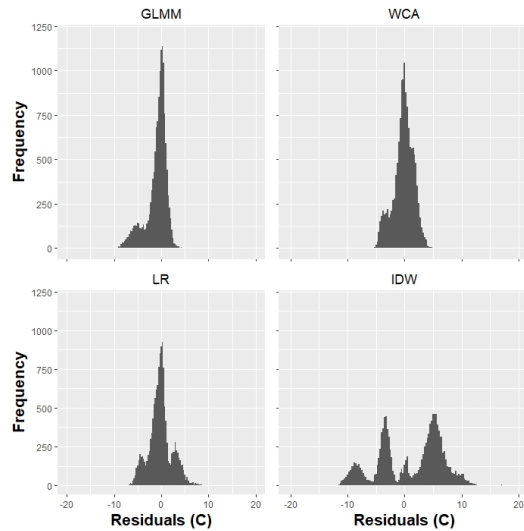
**Figure 6.** Residuals of the temperature LOOCV for each interpolation approach.
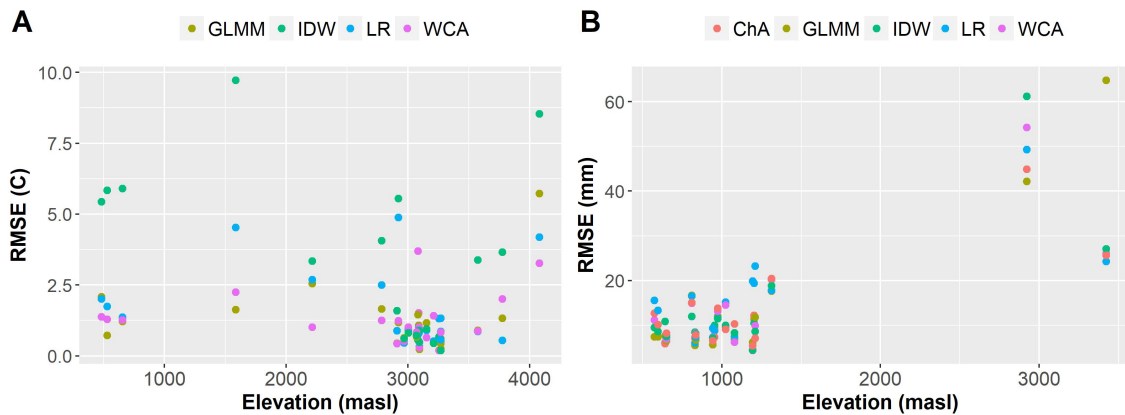
**Figure 7.** (A) Elevation of gauges vs Average temperature RMSE in the LOOCV. (B) Elevation of gauges vs Average precipitation RMSE in the LOOCV.

and even more from 14 to 4 gauges. Its POD and FAR remain similar. The RMSE performance of the other four approaches decreases by a similar rate (3 - 4 mm) when moving to 14 gauges, although LR and ChA have lower POD and FAR. When moving from 14 to 4 gauges ChA shows the smallest increase in RMSE, followed by WCA, while LR has a 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%.

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. The accuracy of CHIRPS gets closer to

that of the interpolation approaches when only 4 gauges are used suggesting its potential value for especially poorly gauged regions; however, still, ChA and WCA perform better with 4 estimation gauges.

**Table 6.** Sensitivity test of the precipitation interpolation approaches.

| Approach | Number of estimation gauges | RMSE (mm) | POD (%) | FAR (%) |
|---|---|---|---|---|
| GLMM | 14 | 32.1 | 91.8 | 7.1 |
| | 4 | 135.8 | 87.8 | 10.6 |
| LR | 14 | 18.9 | 90.6 | 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 | 14 | 17.8 | 97.2 | 22.1 |
| | 4 | 25.4 | 94 | 19.1 |
| ChA | 14 | 16.1 | 90.2 | 13.4 |
| | 4 | 21.0 | 88.4 | 16.4 |
| | 0 (Raw CHIRPS data)** | 26.2 | 88.5 | 28.6 |

**\* Using the monthly long term values provided by WC to approximate monthly precipitation at all sites.**

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

Figure 8 shows the observed and simulated monthly precipitation values for three representative gauges. Figure 8A shows the performance of the low elevation gauge at site 1, which is representative of the performance at the other low elevation gauges. It can be seen that most approaches reproduced observed precipitation at this lowland gauge well compared to the 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 8B shows the performance of all approaches for 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 seen for the lowland gauge (see Figure 8A). Figure 8C illustrates results for 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, although this approach has the best results in site 27.

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 $masl$ are above this threshold, some of them are beyond 40 mm and two 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 high precipitation during the wet season.

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

Regarding the distribution of residuals (see Figure 9), all approaches 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.

## 5    Discussion

5    The LOOCV analysis of air temperature in Section 4.2 shows that for this case study, the GLMM and the WCA have the best performance (i.e. smallest RMSE values - see Table 3). 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).
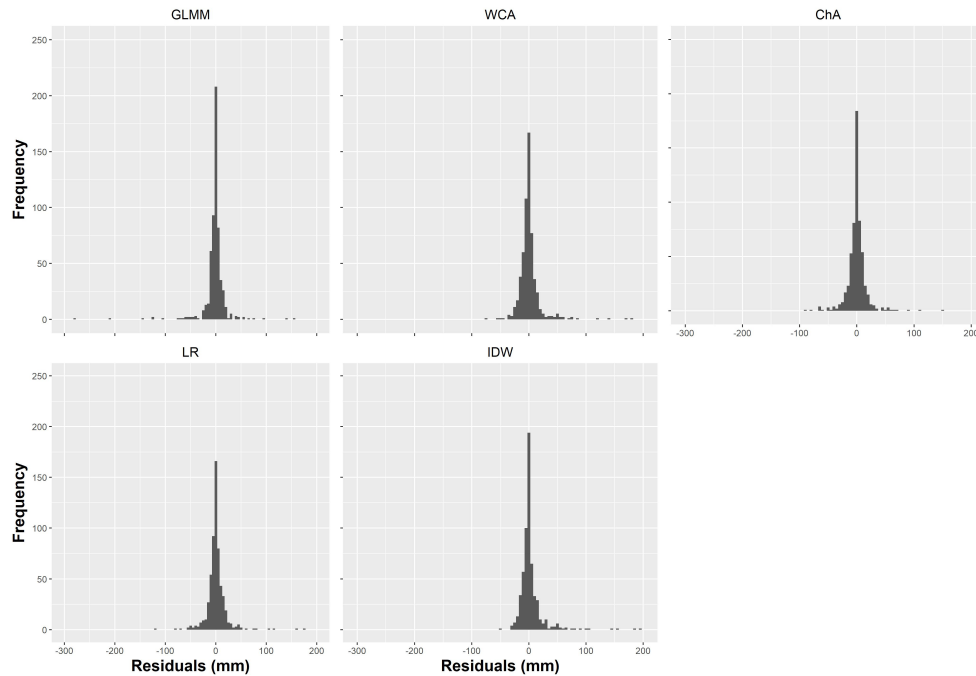
**Figure 9.** Residuals of the precipitation LOOCV for each interpolation approach.

On the other hand, IDW has the largest temperature errors and this, together with the skewed and multi-modal nature of its residuals, shows the limitations of this approach. Figure 5C and 7A suggest that IDW residuals can sometimes be related to the high elevation (e.g. site 28) or isolation (e.g. site 29) of gauges. Temperature observations from the 2008-2009 summer season have the best RMSE values for IDW, but this is likely to be due to the proximity and quantity of gauges in this period.

5      In terms of the influence of the elevation of gauges on temperature results, WCA, LR and the GLMM show similar performance across all elevations, although the latter has an outstanding error at the highest gauge (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

10      of gauges measuring temperature. As shown in Table 4, 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.

On the other hand, WCA is quite resilient to the reduction of estimation gauges. Even with two estimation gauges the average

15      RMSE was only $2.54^{o}C$. This may be because the raw WC maps have internalised the average effect of elevation, longitude and latitude through the long-term analysis (a worldwide generalisation), which can then be adapted to local conditions by including

a small number of gauges. This suggests that WCA is an accurate and easy to use alternative to model air temperature in the case study.

Regarding precipitation, the LOOCV shows that all approaches have similar performances in terms of RMSE, 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 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 (i.e. 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 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 precipitation interpolation approaches are tested with a reduced number of estimation gauges, it is found that the 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 at site 17 (Los Bronces) or site 27 (Lagunitas). This gives an RMSE of $19mm$, which is considerably less than the average RMSE for the GLMM with 14 gauges ($32.1mm$).

The other precipitation interpolation approaches decrease their performance at a similar rate to each other, when facing a reduction in the number of 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 better RMSE values ($21mm$ and $23.5mm$ 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 ($RMSE = 26.2mm$, $POD = 88.5\%$ and $FAR = 28.6\%$).

The results in this paper show how 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 estimate temperature in this case study. For precipitation, ChA or WCA may be preferable, unless the modeller 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 and there is a reasonable coverage of gauges. Analyses of further case studies are required to test the generality of these findings.

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 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.

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 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 due to their importance in water resources modelling. This paper compared five approaches to interpolating temperature and precipitation gauged data in a catchment with complex and steep terrain, and 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, a Generalised Linear Mixed Model (GLMM) reproduced observations in this case study in the best way (i.e. smallest Root Mean Squared Error - RMSE, in a leave-one-out cross validation - LOOCV), although it was closely followed by a 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, showing good prospects for supporting hydrological modelling in sparsely monitored catchments. The GLMM, IDW and LR, had larger errors to the point that for this case study, for temperature interpolation using few gauges, long-term estimates from WorldClim maps gave better RMSE results.

For precipitation, no alternative was clearly superior in the LOOCV, and this may be because errors in high elevation points, 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 an RMSE value larger than $100mm$ when only 4 gauges were used.

In terms of the added value of alternative datasets, it was found that in this case study the inclusion of CHIRPS and World-Clim was valuable. Firstly, 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, as a standalone product, was demonstrated to be a useful source of precipitation data 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. Diego Rivera was supported by CONICYT/FONDAP/15130015 grant.

# References

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.

5 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.

Beck, H. E., Van Dijk, A. I., Levizzani, V., Schellekens, J., Gonzalez Miralles, D., Martens, B., and De Roo, A.: MSWEP: 3-hourly 0.25
10 global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences, 21, 589–615, 2017.

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.

15 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
20 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.

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.

25 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.

30 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.

35 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.

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

Francou, B., Vuille, M., Favier, V., and Cáceres, B.: New evidence for an ENSO impact on low-latitude glaciers: Antizana 15, Andes of Ecuador, 0 28' S, Journal of Geophysical Research: Atmospheres, 109, 2004.

Frei, C.: Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances, International Journal of
10 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.

15 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,
20 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.

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 $https : //webapps.itc.utwente.nl/librarywww/papers_2003/misca/hengl_comparison.pdf$,
25 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.

30 Hsu, K.-l., Gao, X., Sorooshian, S., and Gupta, H. V.: Precipitation estimation from remotely sensed information using artificial neural networks, Journal of Applied Meteorology, 36, 1176–1190, 1997.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., Hong, Y., Stocker, E. F., and Wolff, D. B.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, Journal of Hydrometeorology, 8, 38–55, 2007.

35 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.

5 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.

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.

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

15 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
20 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.

25 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., Rivera, A., Favier, V., Corripio, J., Araos, J., Sicart, J., and Carenzo, M.: A study of the energy balance and melt regime on Juncal Norte Glacier, semi-arid Andes of central Chile, using melt models of different complexity, Hydrological Processes, 22, 3980–3997, 2008.

30 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.

Pilz, J. and Spöck, G.: Why do we need and how should we implement Bayesian kriging methods, Stochastic Environmental Research and
35 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.

5    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.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, Journal of climate, 19, 3088–3111, 2006.

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.

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.

20   Viale, M. and Nuñez, M. N.: Climatology of winter orographic precipitation over the subtropical central Andes and associated synoptic and regional characteristics, Journal of Hydrometeorology, 12, 481–507, 2011.

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.

25   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.

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.

30   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 |