

The use of global gridded datasets in a hydrological model for a scarce-data Andean watershed

E. Muñoz¹, M. Billib², J. L. Arumí³ & D. Rivera³

¹*Department of Civil Engineering,*

Universidad Católica de la Santísima Concepción, Concepción, Chile

²*University of Hannover, Germany*

³*University of Concepción, Chile*

Abstract

Global gridded climatological (GGC) datasets including precipitation, temperature, pressure, and circulation indices, among others are becoming more and more precise, accessible and common in climate and hydrological research. In this study, we evaluate these datasets as an alternative input to supply the lack of measured climatological data in a Chilean Andean watershed in order to develop a monthly water balance model. A conceptual model was carried out for the simulation of stream-flows in the mountainous area of the Polcura River basin in south-central Chile. Based on 18 years of simulation and four model performance assessments, we concluded that from GGC datasets it is possible to reproduce observed flows and the snow-rain regime with “good” performance, being an adequate alternative to supply the lack of measured data, but taking the following considerations i) the GGC rainfall datasets in the Andean area are undervalued; this effect is transferred to the simulated flows, and must be fixed through an amplification coefficient (a calibration parameter), ii) due to the nature of the GGC datasets and to the orographic effect produced by the Andean mountains, the rainfall amounts are damped during the rainy season which result in a sub-estimation of the simulated flows during the winter, and limits the model scope to applications where peak flows are dispensable, and iii) the simulated snow-melt minimum and mean flows are close to the observed flows, which suggest that the GGC datasets are an adequate alternative for estimating the evapotranspiration and snow-melting amounts.

Keywords: scarce-data, water balance models, hydrology, water resources.



1 Introduction

Human demands on the world's available freshwater supplies continue to grow as the global population increases. In the endeavour to manage water demand to meet human needs, the needs of freshwater species and ecosystems have largely been neglected (Richter et al. [1]). Actually there exist many tools to support the water resources planning and management. Such tools as hydrological models, have been widely accepted by the hydrological community, helping in researches like the understanding of hydrological processes (Fenicia et al. [2]), water resources availability (Shuol et al., [3]), climate change impact (Alcamo et al. [4], Jiang et al. [5]), and land-use change assessments (Legesse et al. [6], Lin et al. [7]). However the problem is that the hydrological models application and use is still limited by the data availability, and therefore, the scope of investigations on scarce-data environments are limited. This problem is widely known in mountainous areas. Due to the difficulties in accessibility and high cost of operation and maintenance of meteorological stations in the mountains, it is common to not dispose of sufficient meteorological data to perform adequately hydrological studies.

The Andean mountains are the less major mountainous areas investigated, mainly due to the lack of in-situ observations with sufficient density and temporal resolution to resolve the mesoscale phenomena (Gerreaud [8]). As an option for perform hydrological researches in scarce-data environments is the use of globally available meteorological datasets as inputs for hydrological models.

Global gridded climatological (GGC) datasets are becoming with the time and computational capabilities development more common, detailed (better resolution), and precise. Different GGC datasets have been estimated from remotely-sensed observations, point measurements and general circulation models, being the most reliable and less biased those estimated from point measurements (Barrett and Martin [9], Pitcher et al. [10]). However the confidence, scope and limitations under hydrological studies, and moreover under high spatial variability environments (mountainous areas) is still not yet well defined.

This paper aims to evaluate the scope of a hydrological model performed using GGC datasets. As a study case we choose the Chilean Andean watershed of the Polcura River (latitude 37°10'S). The model scope is evaluated through a comparison between observed and simulated flows and using four model performance assessments. Furthermore, comparative advantages and disadvantages of GGC datasets are valuated through a comparison with a model built using rain gauge datasets.

2 Case study

The Polcura river watershed (Figure 1) is located in the temperate zone of south-central Chile. It covers an area of 914(km²) between the 700 and 3090 (masl), and is characterised with high slopes ($\approx 26^\circ$ average).



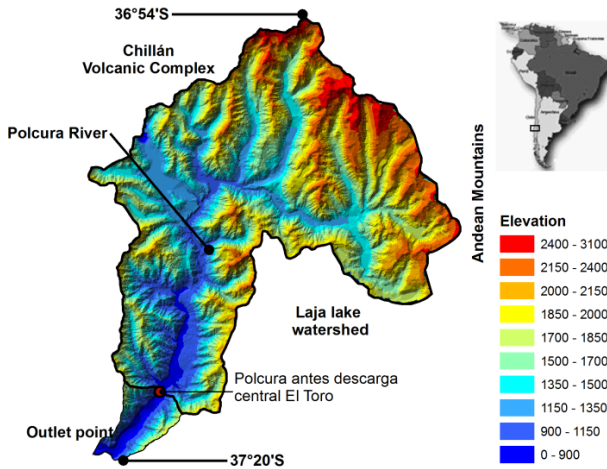


Figure 1: Polcura watershed location, limits and geomorphological characteristics.

The average annual precipitation is 2300 (mm) with a snowfall period during winter and a snow-melting period during the spring. The average monthly temperature is 9 (°C), ranging from 2.5 (°C) in winter to 16.5 (°C) in summer. The watershed is limited by the Andes to the east, the Chillán Volcano to the south and the Laja Lake to the north. The surface soil layers are composed of sedimentary and plutonic rocks overlying volcanic and volcano-clastic deposits (Thiele et al. [11]).

Historically, the Polcura River and surrounding watersheds have been affected by anthropogenic alterations seeking to use the water resources for socio-economic purposes. In this type of watershed, rich in forest and water resources, there exists special economic interest, being essential in the national economical and energy plans.

The Polcura River has a snow-rain regime where both the interaction between natural and industrial components controls the flow regime. On the one hand, there are hydrological processes which produce the stream flows; and on the other hand, there exist industrial activities which change those flows. For example, at the upper part of the river is the canal “Alto Polcura” (a transfer canal) which transfers water from the Polcura watershed to the Laja Lake for hydroelectric industrial purposes. Additionally this activity produces in the lower section of the river (around 9 km from the outlet), an artificial discharge, thus changing completely the hydrologic regime.

Due to the location of the watershed, its mountainous nature and wide elevation range (see Figure 1), the watershed presents high spatial and temporal variability with respect to meteorological and hydrological characteristics, where the orographic effect at the west slopes of the Andes produces an increasing in the rainfall amounts (Gerreaud [8], Vicuña et al. [12]). In addition, the cycle of temperature, the seasonality, and the El Niño phenomena produces temporal

variability on the rainfall patterns (Diaz et al. [13], Escobar and Aceituno [14], Montecinos and Aceituno [15]) and therefore, on the hydrological processes and the stream flows produced.

3 Methodology

The GGC datasets evaluation was achieved by comparing simulated and observed flows using four model performance assessments for a model built with the mentioned datasets as input. To evaluate and identify comparative advantages or disadvantages, a comparison with a model built using rain gauge datasets was also performed.

The watershed was modelled as aggregated. Therefore, calibration, validation, simulation and analysis was performed based on the fluviometric station “Polcura antes de descarga central El Toro” (872 km²), which is the last station (downstream) on the river.

The following presents a brief description of the water balance model used, the inputs collection, its characteristics and processing, and the calibration, validation and simulation procedure.

3.1 Water balance model

The snow-rain and semi-distributed monthly water balance model presented in Muñoz [16] was used. This model simulates the pluvial and snow-melting processes separately and includes the external alterations by adding or subtracting flows.

The pluvial component was modelled through a lumped water balance model that considers the watershed as a double storage system: the sub-superficial (SS) and the underground storage (US). The SS represents the water stored into the unsaturated soil layer as soil moisture. The US is the water which covers the saturated soil layer. The model needs two inputs, the rainfall (PM) and the potential evapotranspiration (PET). The model output is the total runoff (ETOT) at the watershed outlet and includes both the subterranean (ES) and direct runoff (EI) whose amounts are calculated through six parameters of calibration, plus two for the inputs modification (useful in case of non-representative PM and PET data).

The snow-melting model calculates the snowfall (P_{snow}) based on the rainfall above the 0 (°C) isotherm. P_{snow} is stored in the snow storage system (SN), from where the melting calculations are achieved based on the concept of the degree-day method (see Rango and Martinec [17]). Using the mentioned method, the potential melting (PSP) is estimated and then based on the snow stored; the real melting (PS) is calculated. Latter, PS is distributed into the pluvial model through the parameter of calibration F .

Every parameter of calibration has a conceptual physical meaning, integrating the spatial and temporal variability. Table 1 presents a brief description of the model parameters and its influence on the model.



Table 1: Description of the model parameters and its adjustment factors for the pluvial and snow-melting model.

Parameter	Description	Influence
PLUVIAL	C_{\max} - Maximum runoff coefficient when the sub-surface layer is saturated.	- EI
	P_{Lim} - Limit of rainfall over which exists PPD.	- PPD
	D - Percentage of rainfall over P_{Lim} transformed in PPD.	- PPD
	H_{\max} - Maximum capacity of the soil layer to retain water.	- C_{\max} and ER
	PORC - Fraction of H_{\max} that defines the soil water content restricting the evaporation processes.	- H_{crit} and ER
	C_k - Subterraneous runoff coefficient.	- ES
	A - Adjust the precipitation data	- PM
SNOW-MELTING	B - Adjust the potential evapotranspiration data	- PET and ER
	M - Fraction of snow-melt over a base temperature (T_b) where starts the melting.	- PSP, PS
	T_b - Value of temperature over which starts the melting (usually 0 °C).	- PSP, PS
	DM - Minimum rate of melting when $T_m < T_b$.	- PSP, PS
	F - Fraction of the real snow-melt which goes to EI.	- EI
	FgT - Factor which modifies the thermic gradient (should be 1 if the thermic gradient is measured)	- Pnival

The external alterations model simulates inflows and/or outflows to/from the watershed by adding or subtracting flows as follow:

$$Q_{\text{out}}(t) = \text{ETOT}(t) + Q_{\text{contributions}}(t) - Q_{\text{extractions}}(t) \quad (1)$$

where the watershed outflow (Q_{out}) at the time t , is the watershed runoff (ETOT) plus the contributions ($Q_{\text{contributions}}$) less the extractions ($Q_{\text{extractions}}$) during the same period.

3.2 Input datasets

For the above described model application, it is necessary to dispose of rainfall, temperature and potential evapotranspiration data series.

Globally available data from the University of Delaware (Willmott and Matsuura [18]) and local data from point measurements were collected (hereafter UD and DGA data respectively).

The DGA database has only rainfall data while the UD database has rainfall and temperature. Based on this, two meteorological datasets were built, one with rainfall from rain gauges and one with GGC precipitations. In both cases, the potential evapotranspiration was estimated using the Thornthwaite method (Thornthwaite [19], Thornthwaite and Mather [20]) and the UD temperature dataset. Its spatial distribution over the watershed was calculated using the Thiessen polygons method.

3.3 Calibration, validation and simulation

Simulated and observed flows were compared at the “Polcura antes de descarga central El Toro” station. This station has 26 years (1980-2005) of flow records.



Furthermore, the calibration was performed for the period between Jan.1980 and Dec.1987 using the Monte Carlo method, the Nash Sutcliffe efficiency (NSE) (Nash and Sutcliffe [21]) as objective function, and selecting the set of parameter which maximize the NSE and also has physical sense. After the calibration a simulation and the model validation (Jan.1988–Dec-2005) was performed.

The analysis and discussion were based on the simulation results, and the following model performance assessment functions: i) the NSE, ii) the percent of bias (PBIAS) (described on Gupta et al. [22]), iii) the Kling-Gupta efficiency (Gupta et al. [23]), and iv) the relative root mean squared error (RRMSE).

4 Results and discussion

The Table 2 presents the NSE values for the models built using DGA and UD inputs during the calibration period, and four model performance assessments for the validation and analysis during the simulation period. Figure 2 shows the comparison between observed and simulated flows obtained using DGA and UD rainfall inputs for the calibration and simulations periods.

Table 2: Objective function values for the calibration and simulation periods using DGA/UD inputs.

	NSE	RRMSE	PBIAS	KGE
Calibration	0.68/0.57	-	-	-
Simulation	0.78/0.65	0.49/0.60	10.7/11.5	0.71/0.55

The figure demonstrates that both models (with DGA and UD inputs) reproduce properly the pluvial and snow-melting processes. There are observed peak flows during winter, spring, and even during early summer, reproducing the observed behaviour. Moreover it is observed that the simulated flows are underestimated, which is supported by positive PBIAS values (10.7 and 11.5 for DGA and UD inputs respectively) (Gupta et al. [22]).

The base-flow and mean flows in both models are well reproduced with similar results. The major differences are observed on the peak flows, where the model built with DGA rainfall inputs present better results than the model built with UD rainfall inputs. Using the peak differences (PDIFF) assessment (described in Gupta et al. [24]), it is observed that the peak flow differences are larger for the UD model (53 and 87 m³/s for DGA and UD respectively). Moreover the positive values confirm the peak flow underestimation observed in Figure 2 and the PDIFF amounts suggest that the UD rainfall data has the inability to capture the highest rainfall intensity periods.

The large differences in the UD model (in comparison with the DGA model) are due to that the interpolated data are calculated from at least the 20 rain gauges nearest to the center of the grid, which produces a damping effect in the UD datasets reducing the influence of the Andean mountains over the local climatic patterns.



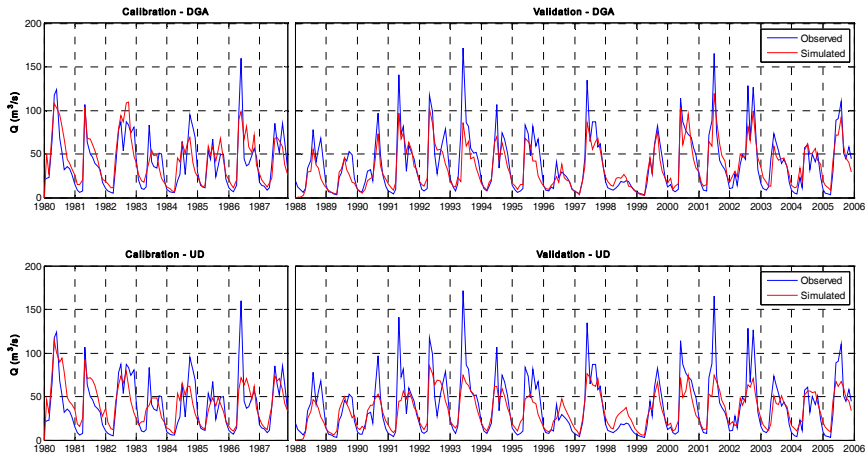


Figure 2: Observed and simulated flows for the calibration and simulation periods using DGA (top) and UD (bottom) inputs.

Table 2 shows that:

- i) The four model performance assessments confirm that the model with DGA inputs reproduces better the stream flows on the Polcura River than with UD inputs.
- ii) Motovilov et al. [25] categorized the NSE as “unsatisfactory” (if $NSE < 0.36$), “satisfactory” (if $0.36 < NSE < 0.75$) and “good” (if $NSE > 0.75$), latter Van Liew et al. [26] defined the PBIAS as “unsatisfactory” (if $PBIAS > |40\%|$), “satisfactory” (if $|20\%| < PBIAS < |40\%|$) and “good” (if $PBIAS < |20\%|$). Based on that, and analysing the simulation period, the NSE and PBIAS suggest that both models have a “good” performance, except the UD model, where the NSE presents a “satisfactory” performance assessment.
- iii) Reviewing the DGA and UD models during the calibration period, it is observed NSE values are worse during the simulation period, suggesting that the NSE is affected by its sensitive to the longitude of the observed data series and to the magnitude of bias (see McCuen et al. [27], Gupta et al. [23]).

It is important to point out that the parameter A (parameter for adjustment the precipitation data) was defined during the calibration as 1.13 and 1.58 for DGA and UD rainfall inputs respectively. Furthermore, it is known that at the western slope of the Andes, the orographic effect may exert a strong influence over local rainfall patterns (Falvey and Garreaud [28], Garreaud [8]), producing a high spatial variability and increasing the rainfall amounts. The only way to reproduce this effect is through a dense network of rain gauges. In this case, the DGA rainfall data were estimated from two rain gauges, and the UD rainfall data was estimated from at least 20 rain gauges mostly located at the central valley of Chile (where there is no orographic effect), resulting the models inputs undervalued and in the case of UD, also damped.

5 Conclusions

The GGC datasets are a useful alternative for perform hydrological models and research in the Andean mountains; reproducing with “good” performance the snow-rain regime, and being an adequate alternative to supply the lack of measured data in scarce-data environments.

The simulated snow-melt, minimum and mean flows are close to the observed flows, suggesting that the GGC datasets are an adequate alternative in estimating the evapotranspiration and snow-melt amounts in Andean watersheds.

Due to orographic effect and it related high spatial variability on the rainfall patterns, and due to the nature of GGC dataset (spatially averaged), the GGC rainfall data in the Andean area are undervalued and damped (mainly during the rainy season). This effect is transferred to the simulated flows, and must be fixed through an amplification coefficient (a parameter of calibration). Despite this problem, the simulated peak flows tend to be underestimated; being the differences between observed and simulated flows larger than a model built using measured rainfall data as input and limiting the model scopes.

The scope of investigations under high spatial variability environments as the mountainous area of the Chilean Andes, are limited to those studies where the peak flows are dispensable (e.g. agricultural or human consumption purposes) and where the main objective are water resources availability, otherwise in case of winter flows or peak flows, the model results are not reliable.

References

- [1] Richter, B., Matthews, R., Harrison, D. & Wigington, R., Ecologically sustainable water management: managing river flows for ecological integrity. *Ecological Applications*, **13**(1), pp. 206–224, 2003.
- [2] Fenicia, F., McDonnell, J. & Savenije, H., Learning from model improvement: On the contribution of complementary data to process understanding. *Water Resources Research*, **44**(6), pp. 1–13, 2008.
- [3] Schuol, J., Abbaspour, K., Srinivasan, R. & Yang, H., Estimation of freshwater availability in the West African sub-continent using the SWAT hydrologic model. *Journal of Hydrology*, **352**(1-2), pp. 30–49, 2008.
- [4] Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T. & Siebert, S., Development and testing of the Water GAP 2 global model of water use and availability. *Hydrological Sciences Journal*, **48** (3), pp. 317–337, 2003.
- [5] Jiang, T., Chen Y., Xu, C., Chen, X., Chen, X. & Singh, V., Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. *Journal of Hydrology*, **336**(3-4), pp. 316–333, 2007.
- [6] Legesse, D., Vallet-Coulomb, D. & Gasse, F., Hydrological response of a catchment to climate and land use changes in Tropical Africa: case study South Central Ethiopia. *Journal of Hydrology*, **275**(1-2), pp. 67–85, 2003.



- [7] Lin, Y., Hong, N., Wu, P., Wu, C. & Verburg, P., Impacts of land use change scenarios on hydrology and land use patterns in the Wu-Tu watershed in Northern Taiwan. *Landscape and Urban Planning*, **80(1-2)**, pp. 67–85, 2007.
- [8] Gerreaud, R., The Andes climate and weather. *Advances in Geosciences*, **22**, pp. 3–11, 2009.
- [9] Barrett, E. & Martin, D., The use of satellite data in rainfall monitoring. Academic Press, London, 340 pp., 1981.
- [10] Pitcher, E., Malone, R., Ramanathan, V., Blackmon, M., Puri, K., & Bourke, W., January and July simulations with a Spectral General Circulation Model. *Journal of the Atmospheric Sciences*, **40(3)**, pp. 580–604, 1983.
- [11] Thiele, R., Moreno, H., Elgueta, S., Lahsen, A., Rebollo, S. & Petit-Breuilh, M., Evolución geológico-geomorfológica cuaternaria del tramo superior del valle del río Laja. *Revista Geológica de Chile*. **25(2)**, pp. 229–253, 1998.
- [12] Vicuña, S., Garreaud, R. & McPhee, J., Climate Change Impacts on the hydrology of a snowmelt driven basin in semiarid Chile. *Climatic Change* (online version), 2010.
- [13] Diaz, A., Studzinski, C. & Mechoso, C., Relationships between precipitation anomalies in Uruguay and Southern Brazil and sea surface temperature in the Pacific and Atlantic Oceans. *Journal of Climate*, **11(2)**, pp. 251–271, 1998.
- [14] Escobar, F. & Aceituno, P., Influencia del fenómeno ENSO sobre la precipitación nival en el sector andino de Chile central durante el invierno. *Bulletin de l'Institut français d'études andines*, **27(3)**, pp. 753–759, 1998.
- [15] Montecinos, A. & Aceituno, P., Seasonality of the ENSO-related rainfall variability in central Chile and associated circulation anomalies. *Journal of Climate*, **16(2)**, pp. 281–296, 2003.
- [16] Muñoz, E., Desarrollo de un Modelo Hidrológico como Herramienta de Apoyo para la Gestión del Agua. Aplicación a la Cuenca del Río Laja, Chile. Master Thesis, University of Cantabria, Spain, 53 pp., 2010.
- [17] Rango, A. & Martinec, J., Revisiting the degree-day method for snowmelt computations. *Journal of the American Water Resources Association*, **31(4)**, pp. 657–669, 1995.
- [18] Willmott, C., & Matsuura, K., Terrestrial air temperature and precipitation: Monthly and annual time series (1900–2008) Version 1.02. University of Delaware Web Site, <http://climate.geog.udel.edu/climate>.
- [19] Thornthwaite, C., An approach toward a rational classification of climate. *Geographical Review*, **38(1)**, pp. 55–94, 1948.
- [20] Thornthwaite, C., & Mather, J., The water balance. *Publications in Climatology*, **8**, pp. 1–104, 1955.
- [21] Nash, J. & Sutcliffe, J., River flow forecasting through conceptual models part I: A discussion of principles. *Journal of Hydrology*, **10(3)**, pp. 282–290, 1970.



- [22] Gupta, H., Sorooshian, S. & Yapo, P., Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal Hydrologic Engineering*, **4(2)**, pp. 135-143, 18999.
- [23] Gupta, H., Kling, H., Yilmaz, K. & Martinez, G., Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, **377(1-2)**, pp. 80-91, 2009.
- [24] Gupta, H., Sorooshian, S. & Yapo, P., Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resources Research*, **34(4)**, pp. 751-763, 1998.
- [25] .Motovilov Y., Gottschalk, L., Engeland, K. & Rodhe, A., Validation of distributed hydrological model against spatial observations. *Agricultural and Forest Meteorology*, **98-99**, pp. 257-277, 1999.
- [26] Van Liew, M., Arnold, J. & Bosch, D., Problems and potential of autocalibrating a hydrologic model. *Transaction of the American Society of Agricultural Engineering*, **48(3)**, pp. 1025-1040, 2005.
- [27] McCuen, R., Knight, Z. & Cutter, A., Evaluation of the Nash-Sutcliffe efficiency index. *Journal of Hydrologic Engineering*, **11(6)**, pp. 597-602, 2006.
- [28] Falvey, M. & Garreaud, R., Wintertime precipitation episodes in central Chile: Associated meteorological conditions and orographic influences. *Journal of Hydrometeorology*, **8**, pp. 171-193, 2007.