Bias-Correction of Satellite Rainfall Estimates Through the Use of Metamodels using Gaussian Process and Bayesian Regression. A Case Study for the Imperial Basin (Chile)

Margarita Alejandra Rebolledo Coy[1], Oscar Manuel Baez Villanueva[2], Thomas Bartz-Beielstein[1], Lars Ribbe[2]

[1]SPOTseven Lab, TH Köln
Steinmüllerallee 1, 51643, Gummersbach
E-Mail: margarita.rebolledo@th-koeln.de,
thomas.bartz-beielstein@th-koeln.de

[2] Institute for Technology and resources management in the Tropics and Subtropics (ITT), TH Köln
Betzdorfer Straße 2, 50679 Kön
E-Mail: oscarmbaez89@hotmail.com, lars.ribbe@th-koeln.de

1 Problem Description

Precipitation is a key parameter in the water cycle and plays an important role in both weather and climate. Nevertheless, an accurate assessment of the Spatio-temporal variability of rainfall is severely limited [1]. The traditional approach to determine the spatial distribution of rainfall is to use a dense network of rain gauge stations over the study area [2, 3]. However, in developing countries, hard to access areas and even over the ocean this rain gauge network can be sparse or non-existent. The use of rainfall estimation algorithms on satellite imagery to generate Satellite Rainfall Estimates (SRE) has helped to partially solve this problem, as satellites provide imagery at high temporal and spatial resolution. The SREs are algorithm-based precipitation datasets. These algorithms use information collected by infrared (IR) sensors aboard geostationary (GEO) satellites and passive or active microwave sensors aboard low-earth orbit (LEO) satellites. It is worth to mention that the SREs are estimates of the rainfall constructed with indirect measurements. Therefore, SREs may be used incorrectly if a validation process is not carried out [4].

Several SRE have been recently developed such as the Multi-source Weighted Ensemble Precipitation (MSWEP) [5], the Climate Prediction Center Morphing technique product (CMORPH) [6], the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Climate Data Record (PERSIANN-CDR) [8], the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis [9], and the Climate Hazards group Infrared Precipitation with Stations dataset (CHIRPS) [10] among others.

Zambrano-Bigiarini et al. [1], evaluated the temporal and spatial performance of the above mentioned SREs over Chile in a monthly scale. The authors concluded that despite continuous improvement on most SRE products, different types of discrepancies between SREs and ground observations might be reduced by using local observations to calibrate the satellite estimates. Following these results we propose a bias correction for the SREs using regression methods to correct the satellite estimate error for the Imperial river basin in Chile. The study area has a rain gauge network of 13 station which record the real rainfall value. We model the SRE against the ground station measurements using Gaussian process and Bayesian regression methods. The regression will be carried out using yearly and seasonal time frames to find out which applies better for the error correction. The obtained models will be analysed to discover which method better fits the area of study and the error characteristics will be evaluate to find possible patterns. Additionally it is of our interest to investigate how the two different regression methods perform in this task.

1.1 Study Area

The Imperial river basin is a chilean catchment located in the ninth region named Araucanía, one of the fifteen regions of Chile. The Imperial basin has an area of 12.763 Km^2 with 540599 inhabitants and a river length of about 230 kilometres. Fig. 1 shows the location of the basin together with the location of the 13 rainfall gauge stations it contains.

According to the General Water Direction of Chile (DGA), this basin presents two climate types. A tempered warm rainy climate with a mediterranean influence in the center and low sectors. On the other hand, the highest areas present a tempered cold rainy climate with mediterranean influence. Both climates record a mean on annual precipitation of about 1245 mm and 1850 mm respectively.

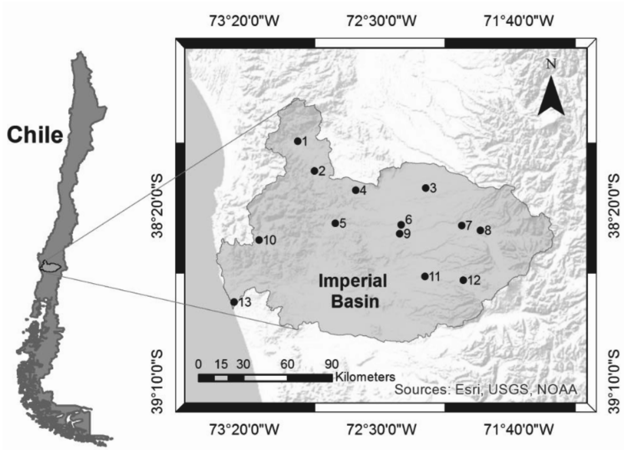


Figure 1: Imperial river basin located in Chile. The image show the area of study and the location of the 13 gauge stations used for this work.

The Imperial basin presents a pluvial hydrological regime. In other words, the water resource comes mainly for the rainfall. This characteristic is related to the low snow accumulation associated with the relative small altitude in the Andean mountain range on its latitude [11].

1.2 Data Description

The selected rainfall dataset covers the period from January 2003 up to December 2015 on a daily basis. The data is observed across 13 rain gauge stations dispersed across the area of study as seen is Fig. 1. This result in a total of 4748 data points for each station. The collected information is organised in 13 data frames, one for each station. All data frames contain the following data:

- The date on which the measurement was taken.
- The precipitation value in millimetres (mm) measured by the rain gauge (observed values).
- The SRE precipitation value in mm yearly recorded for the specific station area (SRE_annual).
- The SRE precipitation value in mm seasonally recorded for the specific station area (SRE_seasonal).

Table 1: Summary of the available rain gauge and SRE information for station 7.

	observed	SRE_annual	SRE_seasonal
1	Min. : 0.000	Min. : 0.000	Min. : 0.000
2	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000
3	Median : 0.000	Median : 0.000	Median : 0.000
4	Mean : 2.489	Mean : 3.081	Mean : 2.933
5	3rd Qu.: 0.000	3rd Qu.: 3.082	3rd Qu.: 2.034
6	Max. :99.000	Max. :77.818	Max. :108.237
		NA's :4	NA's :3

The rain gauge measurements are used as the real precipitation value and the models will use it as its dependent variable. The SRE information is available in two formats annual and seasonal. The seasons are a quarterly subdivision of the year corresponding to December - February, March - May, June - August, and September - November. For each station the SRE which showed more closeness to the observed rainfall values on that season was selected. This means that a year of seasonal data consists of a mixture of several SREs. On the other hand, the SRE_annual data covers the whole year using only one satellite measurement, in this case all the annual data comes from the SRE PERSIANN-CDR. It is good to mention again that these satellites not necessarily represent an accurate reading of the studied area and the error in the estimates is still considerable.
Table 1 shows the summary of the data for station 7 as enumerated by the map in Fig 1. It can be seen that the majority of the data points are close to zero.

The SREs data is obtained from the public data set of the center for hydrometeorology and remote sensing (CHRS). The data from the ground stations are taken from the Chile center of climate science and resilience (CR). Included in the data there were four missing values all of which happened on the same days across all stations. We assumed this was due to a non-recurrent error on the SRE and omitted the missing data points from the analysis.

2 Gaussian Process

Gaussian process regression, also know as kriging, is an interpolation method that works by placing a prior distribution on an unknown function. This function is evaluated at the input data points $X = \{x_i\}_{i=1}^n$ and response observations $Y = \{y_i\}_{i=1}^n$ are collected. These observations are used to update the prior distribution into a posterior distribution by means of bayesian inference. In this sense, kriging assumes the observed points come from some multivariate Gaussian distribution and are related between them by the kernel function. The kernel function used in our study is given by

$$K(x, x') = exp(-\sum(\gamma_i |x_i - \hat{x}_i|^{p_i})) \tag{1}$$

The obtained posterior distribution distribution of y is given by a Gaussian process:

$$y|X, y \sim N(m(x), s^2(x)) \tag{2}$$

Where $m(x)$ is the maximum a posteriori probability at x and $s^2(x)$ is the estimations mean squared error.

2.1 Cluster Kriging

One of the major problems of Kriging is its high time and space complexity for large datasets. To help with this issue we divided the data set into smaller not overlapping clusters. For each cluster a Kriging model is built. A global posterior model is obtained by a weighted linear combination of the models as suggested by [13]. The global posterior is defined then as:

$$y|X, y \sim N(\sum_{i=1}^{q} w_i m_i(x), \sum_{i=1}^{q} w_i s_i^2(x)) \tag{3}$$

Where m_i and s_i represent the mean and variance of the model in the $i-th$ cluster. The weights w_i in the equation were calculated by minimizing the variance of the process [13].

$$w_i = s_i^{-2}(x) / \sum_{i=1}^{q} s_i^{-2}(x) \tag{4}$$

3 Bayesian Regression

The essential of the Bayesian methods is their use of probability for quantifying uncertainty in inferences based on statistical analysis [15]. Bayesian regression follow the Bayes' rule to make inferences from the data. Let y represent the observed precipitation values, x the SRE and θ a sequence of unknown parameters. We can infer the posterior distribution in terms of the likelihood and prior knowledge as shown in Eq. 5.

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{5}$$

Given the nature of the data, we choose to use a logarithmic transformation on the observed values to allow the posterior to be sampled from a normal distribution. In these terms the likelihood $p(y|\theta)$ is defined as:

$$log(y+1) \sim N(\mu, \sigma) \tag{6}$$

The prior distributions for the parameters μ and σ are assumed to be normally distributed since not much previous information is known.

4 Experiments

The experiments were executed using SPOT [12] for the Gaussian process regression and rstan [14] for the Bayesian regression.

Two regression models were generated for each method, one using the SRE_annual data and other using the SRE_seasonal data. For each case the first twelve years of data were used as the training set and the last year was used as the validation set. The goodness-of-fit measures used to evaluate the models performance are the root mean square error (RMSE) and the Kling-Gupta efficiency (KGE). KGE is a goodness-of-fit measure obtained through the decomposition of the squared mean error [16]. This decomposition facilitates the analysis of the different components of the model: The correlation (r), the bias (β), and the relative variability in the simulated and observed values (α). All three components have their optima at unity. The KGE range from ∞ to 1 and its value is defined as shown in Eq. 7. The closer to 1 the more accurate the model is

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{7}$$

For the Gaussian process regression the kernel terms γ and p from Eq. 1 were determined using the maximum likelihood method. The terms values are updated in every SPOT iteration. After some tests this approach was deemed to deliver better results than with static γ or p values.
To implement the cluster kriging the SRE_annual data set was divided into 20 clusters using the K-means method. The regression was run 10 times for each station. Each run had a different random seed to proof the stability of the system.
Since the SRE_seasonal data was already divided into subgroups, i.e. the seasons, the cluster kriging was not implemented here. In this case the analysis was also run 10 times for each station with different random seeds each.

The Bayesian regression was implemented with a logarithmic transformation on the response variable with the form:

$$ln(y+1)|x,\theta \sim N(\alpha + \beta * x, \sigma) \tag{8}$$

Uninformative proper priors are used for the parameters α, β and σ since no additional information is previously known.

$$\alpha \sim N(0,5), \qquad \beta \sim N(0,5), \qquad \sigma \sim cauchy(0.001, 10) \tag{9}$$

The sampling of the posterior distribution of y is done using Hamiltonian Monte Carlo sampling. A total of 4 chains were generated for the posterior. After checking that all the chains converged and were accurate representations of the posterior distribution the number of iterations for each chain was set at 10000 including 2000 warmup iterations.

Before comparing the models we need to define a baseline against which to compare the models with. Since our goal is to approximate the SRE to the observed rain gauge values as much as possible, a model will be good when its RMSE and KGE values are better than those on the raw data. For example, for the annual model the baseline RMSE will be the RMSE between the station observed and SRE_annual values without any model.

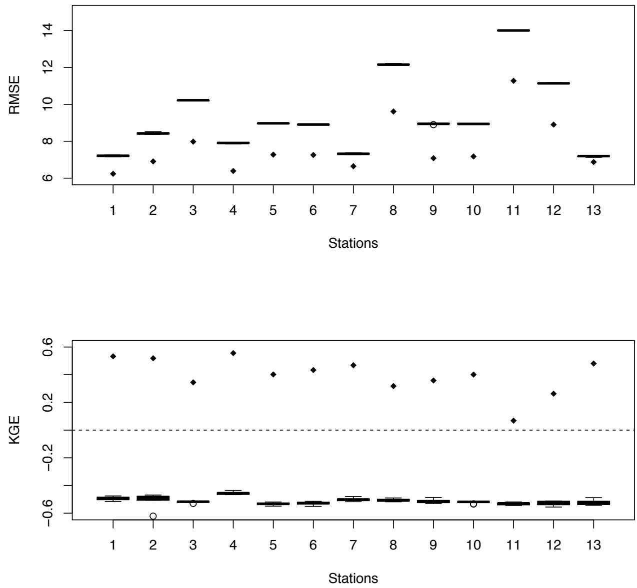


Figure 2: Boxplot of the RMSE and KGE values between the fitted annual Gaussian model response and the real rain gauge values respectively. The values are taken for all runs across all stations. The baseline RMSE and KGE values are depicted as diamonds. The results are stable between runs, yet the model had no improvement with respect to the baseline values.

4.1 Results Gaussian Process

The annual Gaussian process regression was done using the cluster kriging method. The RMSE and KGE fitness measures were computed for all stations in all 10 runs. The results indicated that the model was stable since there is no considerable difference between the different runs results. However, the model did not bring any improvement to the baseline SRE values. Fig. 2 illustrates this. The RMSE between the model fitted values and the real observed values for all runs across all stations is plotted against the baseline RMSE values, depicted in the image as diamonds. Every station shows a worse RMSE as that of the Baseline. The same is the case with the KGE, in all stations the baseline value is better than the model fitted values.

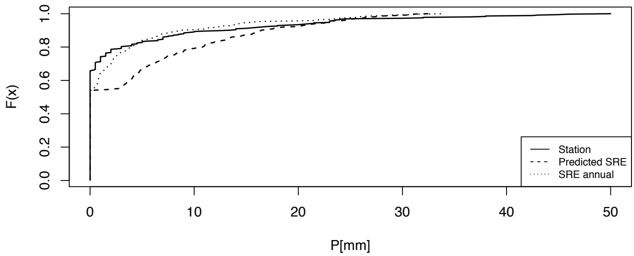
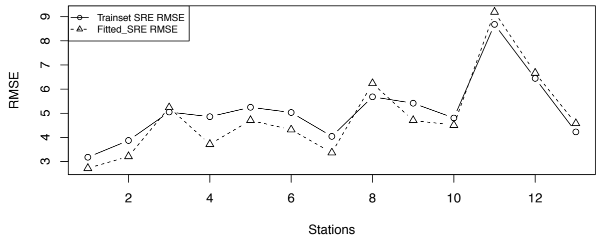


Figure 3: CDF of the validation set observed values of station 6. The Solid line indicate the CDF for the validation set observed values. The dashed line shows the annual Gaussian model predicted values. The dotted line are the validation set SRE_annual values. It can be seen that the model tends to underestimate the precipitation. The model predicted values and the SRE_annual values also fail to approximate the rain gauge maximum value.

The model doesn't seem to fit the data well but for completeness and to understand better the annual data behaviour we also analyse the model response on the validation set. As was expected based on the results on the training set ,the predicted model response across stations was steadily performing worse than the baseline SRE_annual. To illustrate the divergence between (1) the predicted values, (2) the observed values and (3) the SRE_annual values, we compare the three measurements cumulative distribution functions (CFD). Fig. 3 shows the CDFs of the three measurements for station 6 as an example. It was observed that the predicted values tend to underestimate the real precipitation before being truncated at a lower maximum value than the observed values.

Apart from the model fitted and predicted values, the model uncertainty was also explored. The standard deviation of the fitted values has a maximum value of $6.326e^{-01}$ while for the predicted values it has a maximum of $4.404e^{-02}$. This uncertainty values are well inside the tolerance limits for rainfall and thus the model uncertainty is not an issue.

The Gaussian process seasonal regression showed a considerable improvement in terms of RMSE and KGE fitness. The stability of the model runs was also improved with almost no identifiable difference between the different runs. For the training set Fig. 4 shows the fitted model



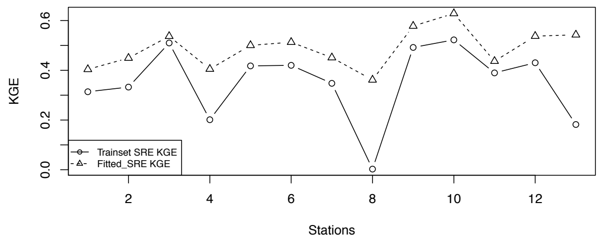


Figure 4: RMSE and KGE for the seasonal Gaussian model fitted values for one season across all stations. The baseline RMSE and KGE are depicted for comparison. For the most cases the model fitted values RMSE improved the baseline measurements and approximated the train set observed values better than the SRE_season data.

values RMSE and KGE for all stations for one season. Again the baseline RMSE and KGE values are illustrated as triangles in the figure. It can be seen that in almost all cases the model fitted values approximated better the station observed values than the SRE_season values did. It muss be mentioned that this behaviour was not stable for all seasons in terms of RMSE. There were some seasons that performed poorly against the baseline RMSE values. However, the KGE values was for all stations and all seasons always better than the baseline.

Since our main use for the fitted model values is on retrospective applications where the stations real values are always available, we will analyse more closely the model results on the training set. The training set observed values CDF for one season of station 2 can be seen on Fig. 5. The behaviour observed in the figure is replicated in the other stations. Here it can be seen that the model fitted values CDF follows the observed values closely despite the difference in RMSE previously noted.

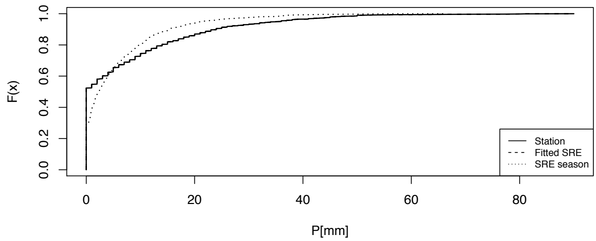


Figure 5: CDF of the training set observed values for station 2 in one season. The Solid line indicates the CDF for the observed values. The dotted line are the values for the validation set satellite measurements. Here the dashed line indicating the seasonal Gaussian model fitted values is not visible due to it overlapping the observed values in spite of the worse RMSE performance.

This performance improvement was also observed on the validation set where the model predicted values were able to predict the whole range of precipitation, contrary to the case observed in Fig. 3.

Evaluated on the training set the standard deviation of the fitted values had a maximum value of $8.444e^{-2}$. However, on the validation set the predicted values had a maximum deviation of up to 1.840. This value is worse than what we expected but still inside the tolerance limits for precipitation values.

4.2 Bayesian regression

The converge of all the chains in the Bayesian regression was controlled using three different criteria: The effective sample size (ESS), Rhat, and the Monte Carlo standard error (MCSE).
ESS gives insight to the autocorrelation within chains. A large value indicates there are enough independent samples in a chain. The Rhat criteria checks that the chains have converged to a common distribution by measuring the variance within and between chains. If the chains are not converging then its value will be greater than one. MCSE measures the uncertainty associated with the Monte Carlo approximation and has zero as its ideal value.

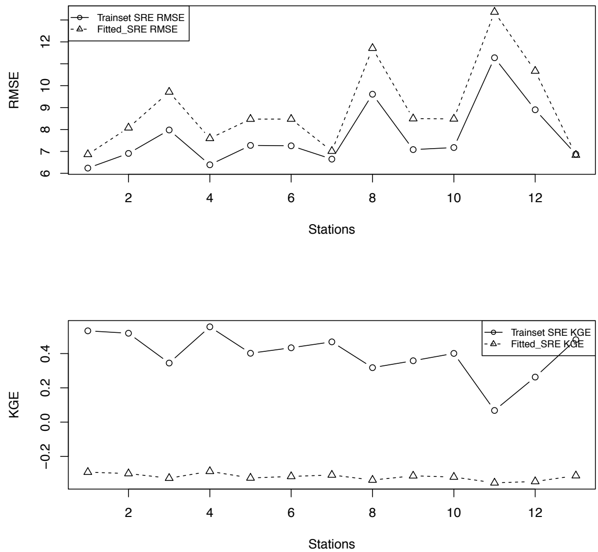


Figure 6: RMSE and KGE measurements of the Bayesian annual regression fitted values. The baseline fitness measurements are also included for comparison. The model performed worse for both fitness measurements. However, it improved in a small scale the KGE Gaussian process annual regression performance.

During the execution of the Hamiltonian Monte Carlo on all models there was no evidence to assume that the chains were not converging or that the Monte Carlo samples were not representative of the posterior distribution.

As with the Gaussian process regression we start by generating and analysing the model using the SRE_annual data. The model follows close in behaviour and improves in a small scale the Gaussian process annual results KGE measures. Fig. 6 shows the RMSE and KGE behaviour of the model against the baseline values for all stations. Here the model RMSE and KGE where computed using the mean of the 4 chains converged results.

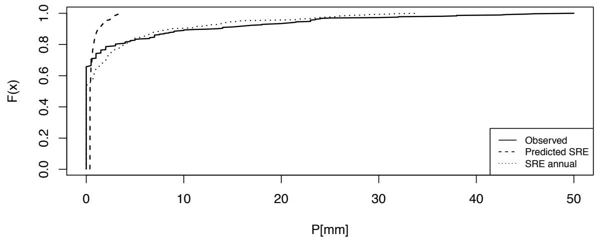


Figure 7: CDF of the validation set observed values for station 6. The solid line indicates the observed values CDF. The dotted line are the SRE_annual values. The dashed line are the annual Bayesian model predicted values. The model fails to predict non-zero values and is truncated on a lower precipitation value. This is due to the predicted posterior having its density centred at zero.

Again the CDF of the model predicted values was examined to find differences between the Bayes and Gaussian process regression on data forecasting. Fig. 7 shows the CDF of the observed, predicted and SRE_annual values on the validation set for station 6. The divergence of the model from the real data is clear. Since the posterior distribution of the model predicted values has most of its density around zero, it fails to predict non-zero values and has a maximum predicted precipitation value of only 4 mm.

Following the same criteria as exposed in the annual analysis, the convergence and representativeness of the chains for the seasonal regression was confirmed.
As was already seen on the Gaussian process regression analysis, the model behaviour across seasons changes. This implies that at some seasons the model fitted RMSE outperforms the baseline RMSE values. However, for all seasons the model fitted KGE was worse than the baseline in all stations, as seen in Fig. 8. This means that overall the model was not doing better that the SRE_season data.

Fig. 9 shows the divergence of the fitted values of the seasonal Bayes regression against the observed values on the training set. In the figure it can again be seen how the model fails to fit precipitation values not close to zero.

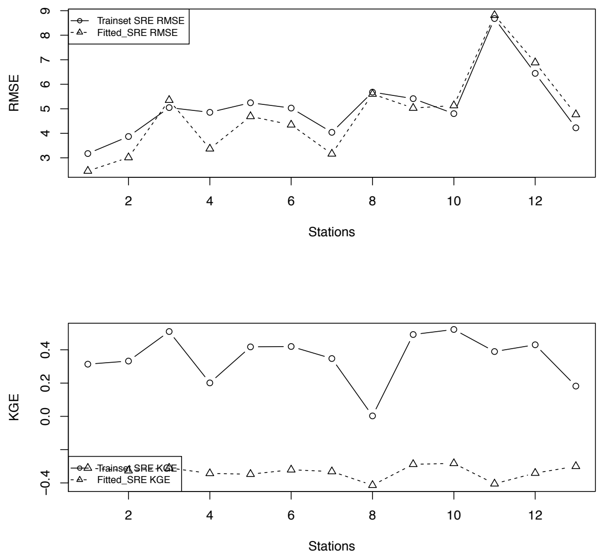


Figure 8: RMSE and KGE of the fitted seasonal bayesian model values for one season across all stations. The baseline RMSE and KGE are also showed for comparison. Even though some stations showed an improved performance on the fitted RMSE values, the KGE performance was consistently worse.

5 Conclusions

The aim of this work was to implement and test a bias correction method for satellite rainfall estimates on the Imperial basin in Chile. This was encouraged by the low accuracy SREs can have under some space and time conditions. We tested two regression methods on two SRE data . The first SRE data class was annual data, this data was collected using only one SRE. The second SRE data class was seasonal data, this data consisted on a combination of SREs. These two data classes were used to test the data aggregation procedures implemented to fit the SRE time scales (annual and seasonal). The two tested regression models were Gaussian process and Bayesian regression. From our analysis it was clear that any bias correction needed to be carried out using seasonal data. Since the combination of two or more SRE improved the rainfall estimates in the varying conditions of the seasons.

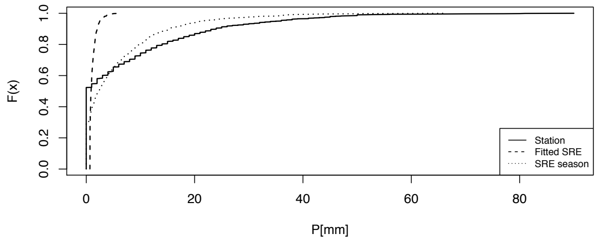


Figure 9: CDF of the validation set observed values of station 6. The Solid line indicates the CDF of the validation set observed values. The dashed line shows the seasonal Bayes model predicted values. The dotted line are the validation set SRE_season values. It is apparent that the model fails to predict precipitation values ranges away from zero.

Overall Gaussian process regression delivered better and more accurate results when working with seasonal data while the Gaussian process annual model showed no improvement at all. There are some factors that could have contributed to the lack of improvement in the annual Gaussian process regression. First, the use of cluster kriging may have introduced an unknown amount of error into the annual model. This error could also have been aggravated by the varying size of the clusters. It is possible that the use of a cluster method that generates fix sized non-overlapping clusters could bring some improvement to the annual analysis. This however needs to come in hand with a better SRE.

Even though the Bayesian regression did not show satisfying results it was interesting to discover that it was considerably faster that the Gaussian regression. Given that the intention of the bias correction model is for it to be used in bigger areas with denser rain gauge networks, the speed in which the model can be generated is important. Taking this into account, it is in our interests to keep exploring the possibility of implementing the bias correction using the Bayesian regression. For this we propose future experiments using hurdle or zero-inflated models that allow to have separate distributions for zero and non-zero values.

The Bayesian regression results also highlighted the differences between RMSE and KGE measures of fitness. Where the RMSE for the Bayesian annual model showed no improvement, the KGE showed an small improvement. The same was the case with the seasonal Bayesian model where the RMSE showed an improvement but KGE steadily showed a worse performance. It is clear that in the latter case the RMSE was overestimating the model skill. This can be due to the choice of baseline with which the models are compared. A more detailed inspection of the KGE components showed an improvement in the bias (β) values for the Bayes regression models, while the correlation and variability did not show any improvement.

References

[1] M. Zambrano-Bigiarini, A. Nauditt, C. Birkel, K. Verbist, L. Ribbe. "Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile". In: *Hydrology and Earth System Sciences*. 21.2. 2017.

[2] C. Berndt, E. Rabiei, U. Haberlandt. " Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios". In: *Journal of Hydrology* 508. 2014.

[3] T.K. Chang, A. Talei, S. Alaghmand, M.P. Ooi. "Choice of rainfall inputs for event-based rainfall-runoff modeling in a catchment with multiple rainfall stations using data-driven techniques". In: *Journal of Hydrology* 545. 2017.

[4] M. Gebremichael, E.N. Anagnostou, M.M. Bitew. "Critical Steps for Continuing Advancement of Satellite Rainfall Applications for Surface Hydrology in the Nile River Basin". In: *jJAWRA Journal of The American Water Resources Assosiation* 46.2. 2010.

[5] H.E. Beck, A.I.J.M. van Dijk, V. Levizzani, J. Schellekens, D.G. Miralles, B. Martens, A. de Roo. "MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data". In: *Hydrology and Earth System Sciences* 21. 2017.

[6] R. Joyce, J.E. Janowiak, P.A. Arkin, P. Xie. "CMORPH: A Method that Produces Global Precipitation Estimates fron Passive Microwave and Infrared Data at High Spatial and Temporal Resolution". In: *Journal of Hydrometeorology* 5.3. 2004.

[7] K.L. Hsu, X. Gao, S. Sorooshian, H.V. Gupta. "Precipitation estimation from remotely sensed information using artificial neural networks". In: *Journal of Applied Meteorology* 36.9. 1997.

[8] H. Ashouri, K.L. Hsu, S. Sorooshian, D.K. Braithwaite, K.R. Knapp, L.D. Cecil, B.R. Nelson, O.P. Prat. "PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies". In: *Bulletin of the American Meteorological Society* 96.1. 2015.

[9] G.J Huffman, D.T. Bolvin, E.J. Nelkin, D.B. Wolff, R.F. Adler, G. Gu, Y. Hong, K.P. Bowman, E.F. Stocker. "The TRMM multi-satellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales". In: *Journal of Hydrometeorology* 8.1. 2007.

[10] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdi, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, J. Michaelsen. "The climate hazards infrared precipitation with stations?a new environmental record for monitoring extremes". In: *Scientific Data 2*. 2015.

[11] N.R. Rivera, F. Encima, A. Muñoz-Pedreros, P. Mejias. "Water Quality in the Cautín and Imperial Rivers, IX Region-Chile". In: *Información Technológica* 15.5. 2004.

[12] T. Bartz-Beielstein, C. Lasarczyk, M. Preuss "Sequential Parameter Optimization". In: *IEEE Congress on evolutionary computation.* 2005.

[13] H. Wang, B. van Stein, M. Emmerich, T. Bäck "Time complexity reduction in efficient global optimization using cluster kriging". In: *Proceedings of the Genetic and Evolutionary Computation Conference. ACM.* 2017

[14] Stan Development Team. "RStan: the interface for Stan in R" Package version 2.16.2 *http://mc-stan.org* 2017

[15] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B Rubin "Bayesian Data Analysis". *Chapman &Hall/CTC Press, Third edition* 2013

[16] H.V. Gupta, H. Kling, K.K. Yilmaz, G.F. Martinez. "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling". In: *Journal of Hydrology* 377. 2009.