
Article

Filling Gaps in Daily Precipitation Series Using Regression and Machine Learning in Inter-Andean Watersheds

Marcelo Portuguez-Mauryta ^{1,2,3,*}, José Luis Arumi ^{2,4}, Octavio Lagos ^{2,4}, Alejandra Stehr ⁵
and Nestor Montalvo Arquiñigo ³

¹ Doctoral Program in Water Resources and Energy for Agriculture, Universidad de Concepcion, Av. Vicente Mendez 595, Chillan 3812120, Chile

² CRHIAM Water Research Center, Universidad de Concepcion, Victoria 1295, Concepcion 4070386 , Chile; jarumi@udec.cl (J.L.A.); octaviolagos@udec.cl (O.L.)

³ Water Resources Department, College of Agricultural Engineering, Universidad Nacional Agraria La Molina, Av. La Molina s/n, Lima 15024, Peru; nmontalvo@lamolina.edu.pe

⁴ Water Resources Department, College of Agriculture Engineering, Universidad de Concepción, Av. Vicente Mendez 595, Chillan 3812120, Chile

⁵ Centro de Ciencias Ambientales EULA-Chile, Departamento de Sistemas Acuáticos, Facultad de Ciencias Ambientales, Universidad de Concepción, Concepcion 4070386, Chile; astehr@udec.cl

* Correspondence: mportuguez@lamolina.edu.pe; Tel.: +51-1-949-377-610

Supplementary material

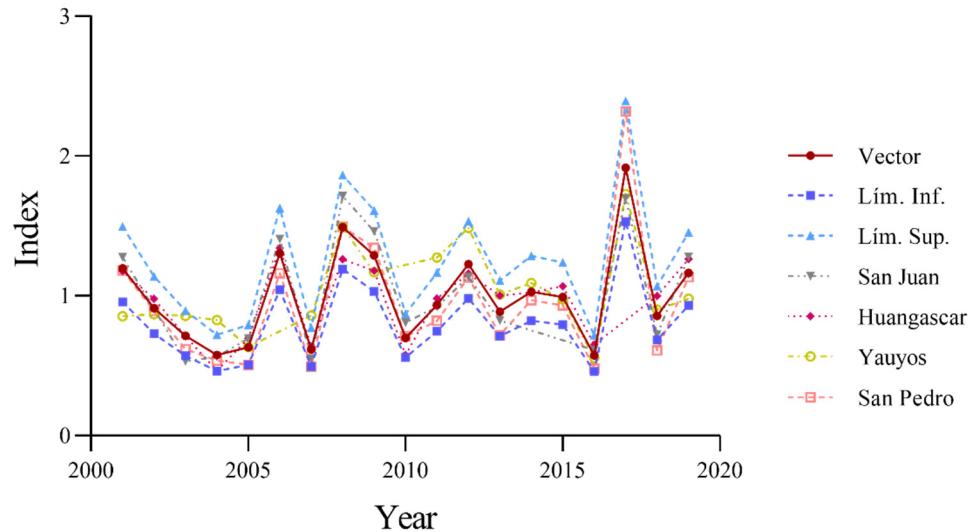


Figure S1. Annual indices of the regional vector and stations in Region 2.

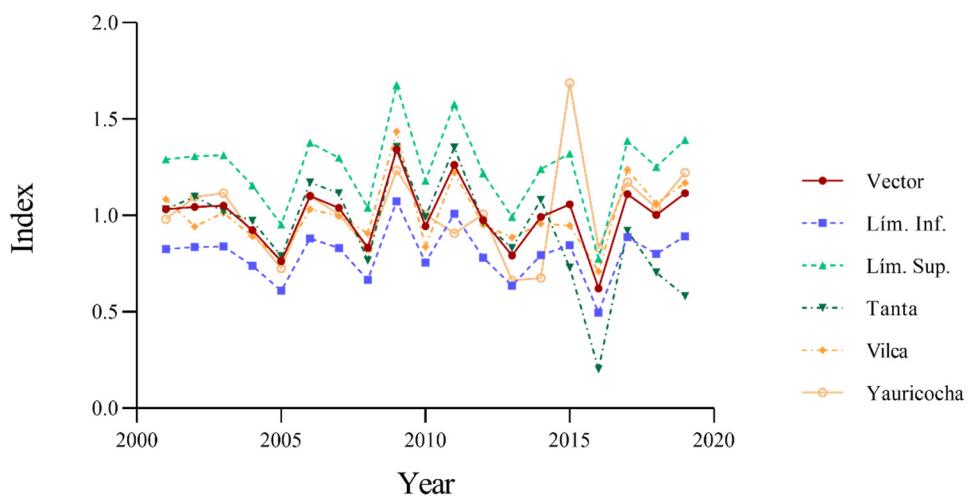
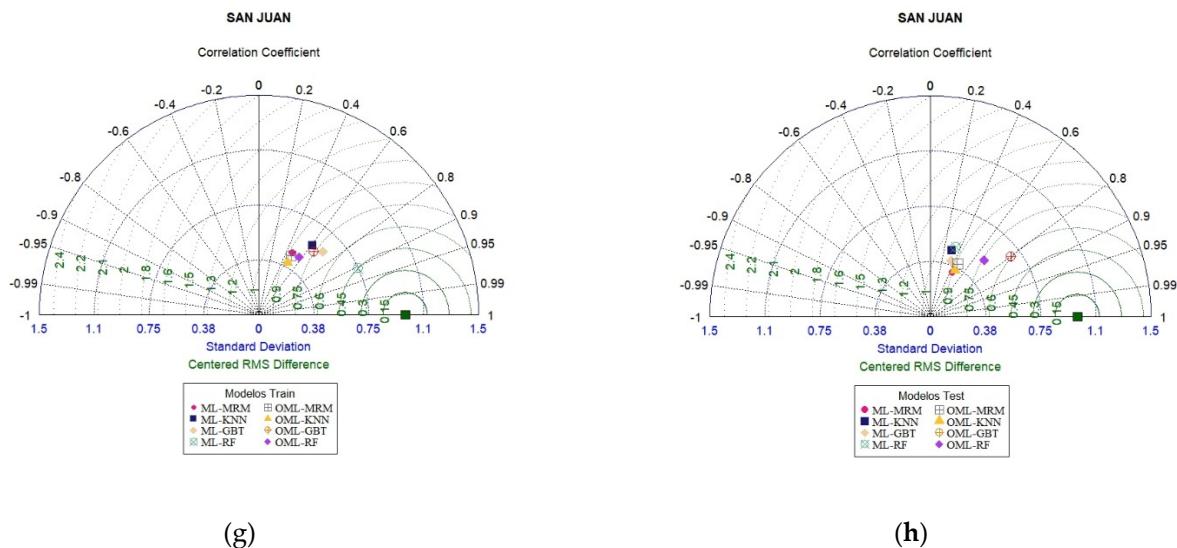


Figure S2. Annual indices of the regional vector and stations in Region 4.

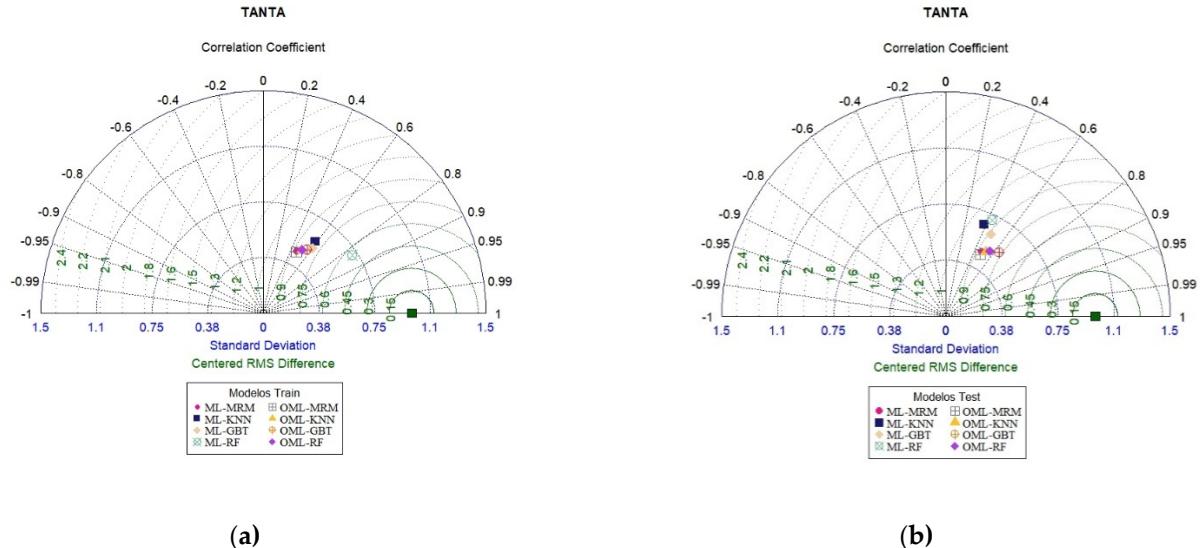




(g)

(h)

Figure S3. Taylor diagrams that show a statistical comparison (normalized standard deviation and correlation coefficient) of observed precipitation and modeled precipitation based on precipitation datasets (training and test) for four stations: (a) San Pedro de Pilas (training), (b) San Pedro de Pilas (test), (c) Huangascar (training), (d) Huangascar (test), (e) Yayos (training), (f) Yayos (test), (g) San Juan de Yanac (training), and (h) San Juan de Yanac (test).



(a)

(b)

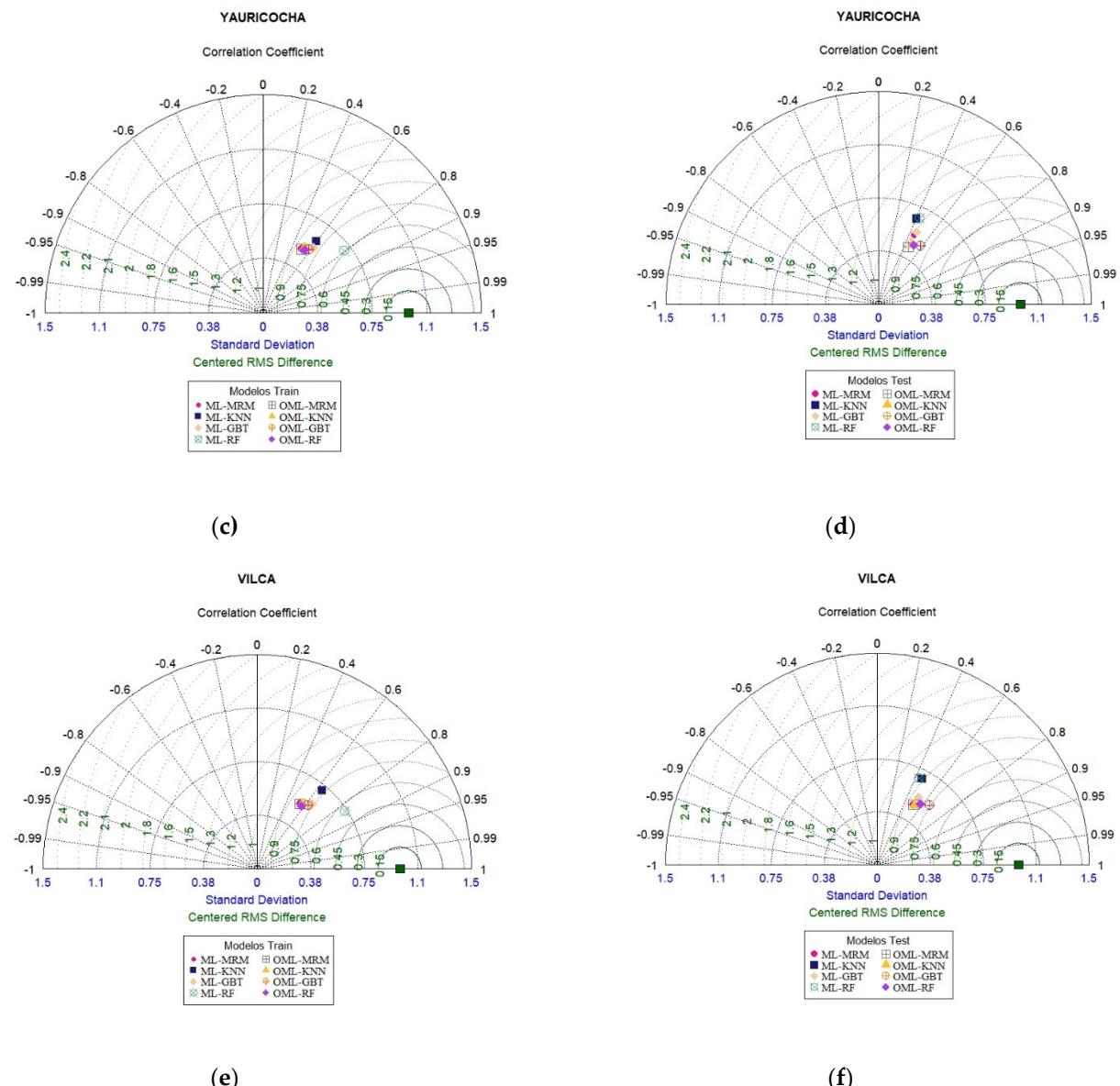


Figure S4. Taylor diagrams that show a statistical comparison (normalized standard deviation and correlation coefficient) of observed precipitation and modeled precipitation based on precipitation datasets (training and test) for four stations: (a) Tanta (training), (b) Tanta (test), (c) Yauricocha (training), (d) Yauricocha (test), (e) Vilca (training), (f) Vilca (test).

Table S1. Annual regional vector indices – Region 2.

Station	Nº Years	Standard deviation	Station/vector correlation
San Juan de Yanac	17	0.120	0.952
San Pedro de Pilas	19	0.137	0.976
Huangascar	18	0.096	0.937
Yauyos	17	0.179	0.861

Table S2. Annual regional vector indices – Region 4.

Station	Nº Years	Standard deviation	Station/vector correlation
Tanta	19	0.196	0.727
Vilca	19	0.074	0.903
Yauricocha	19	0.196	0.551

Table S3. K-means clustering (2001-2019 period).

Regions	Stations
Region 1	Langa
	San Lazaro de Escomarca
	Ayaviri
	Huancata
	Huañec
	Huarochiri
Region 2	Caranía
	San Juan de Yanac
	San Pedro de Pilas
	Huangascar
	Yauyos
Region 3	Pacaran
	Socsi
	Cañete
Region 4	Tanta
	Vilca
	Yauricocha

Table S4. Correlation coefficient - Region 2.

San Pedro de Pilas	1			
Huangascar	0.58	1		
Yauyos	0.43	0.40	1	
San Juan de Yanac	0.43	0.36	0.32	1

San Pedro de Pilas	Huangascar	Yauyos	San Juan de Yanac
-----------------------	------------	--------	----------------------

Table S5. Correlation coefficient - Region 4.

Tanta	1			
Yauricocha	0.37	1		
Vilca	0.43	0.45	1	

Tanta	Yauricocha	Vilca
-------	------------	-------

Table S6. Model efficiency according to fit statistics - Region 2.

Stations	Sam-ples	Statis-tics	LRM	MRM	Machine Learning				Optimized Machine Learning			
					MRM	KNN	GBT	RF	MRM	KNN	GBT	RF
San Pedro de Pilas	Train	R ²	0.34	0.42	0.43	0.54	0.71	0.87	0.43	0.43	0.63	0.63
	Train	RMSE	1.97	1.84	1.88	1.69	1.35	0.96	1.88	1.90	1.52	1.54
	Train	NSE	0.34	0.42	0.43	0.54	0.71	0.85	0.43	0.41	0.63	0.62
	Train	PBIAS	0.00	0.00	0.00	10.82	0.00	-1.16	0.00	16.93	0.00	1.02
	Test	R ²			0.39	0.32	0.28	0.31	0.39	0.40	0.64	0.63
	Test	RMSE			1.73	1.86	1.95	1.89	1.71	1.72	1.34	1.37
	Test	NSE			0.37	0.28	0.20	0.25	0.39	0.38	0.63	0.61
	Test	PBIAS			-11.84	0.82	-11.52	-13.62	0.00	16.83	0.00	0.92
Huangascar	Train	R ²	0.34	0.37	0.39	0.52	0.60	0.86	0.39	0.39	0.54	0.54
	Train	RMSE	1.89	1.84	1.85	1.65	1.49	0.93	1.85	1.86	1.61	1.63
	Train	NSE	0.34	0.37	0.39	0.51	0.60	0.84	0.39	0.38	0.54	0.53
	Train	PBIAS	0.00	0.00	0.00	10.41	0.00	-0.97	0.00	15.74	0.00	1.75
	Test	R ²			0.34	0.29	0.34	0.27	0.34	0.37	0.66	0.63
	Test	RMSE			1.80	1.89	1.80	1.95	1.80	1.77	1.31	1.38
	Test	NSE			0.34	0.27	0.34	0.22	0.34	0.36	0.65	0.61
	Test	PBIAS			-1.58	6.13	-3.02	-6.04	0.00	0.67	0.00	1.18
Yauyos	Train	R ²	0.18	0.23	0.25	0.45	0.51	0.75	0.25	0.32	0.45	0.36
	Train	RMSE	2.19	2.12	2.05	1.76	1.67	1.21	2.05	1.95	1.76	1.89
	Train	NSE	0.18	0.23	0.25	0.45	0.50	0.74	0.25	0.32	0.45	0.36
	Train	PBIAS	0.00	0.00	0.00	-4.64	0.00	-0.69	0.00	10.91	0.00	0.18
	Test	R ²			0.18	0.18	0.26	0.21	0.19	0.28	0.46	0.34
	Test	RMSE			2.32	2.38	2.20	2.34	2.31	2.17	1.90	2.08
	Test	NSE			0.18	0.13	0.26	0.17	0.18	0.28	0.45	0.34
	Test	PBIAS			10.58	1.14	9.23	3.36	0.00	13.39	0.00	0.12
San Juan de Yanac	Train	R ²	0.18	0.22	0.23	0.38	0.51	0.82	0.23	0.24	0.43	0.33
	Train	RMSE	1.40	1.37	1.25	1.13	1.01	0.64	1.25	1.25	1.08	1.17
	Train	NSE	0.18	0.22	0.23	0.37	0.50	0.80	0.23	0.23	0.42	0.32
	Train	PBIAS	0.00	0.00	0.00	-9.28	0.00	-2.27	0.00	5.76	0.00	0.46
	Test	R ²			0.20	0.09	0.12	0.12	0.21	0.22	0.64	0.48
	Test	RMSE			1.69	1.81	1.76	1.79	1.67	1.67	1.14	1.39
	Test	NSE			0.19	0.07	0.12	0.09	0.21	0.20	0.63	0.45
	Test	PBIAS			18.62	7.19	18.19	12.24	0.00	15.25	0.00	0.59

Table S7. Model efficiency according to fit statistics - Region 4.

Stations	Samples	Statis-tics	LRM	MRM	Machine Learning				Optimized Machine Learning			
					MRM	KNN	GBT	RF	MRM	KNN	GBT	RF
Tanta	Train	R ²	0.18	0.22	0.22	0.34	0.36	0.70	0.22	0.26	0.32	0.27
	Train	RMSE	4.84	4.72	4.82	4.46	4.38	3.06	4.82	4.69	4.51	4.67
	Train	NSE	0.18	0.22	0.22	0.33	0.36	0.68	0.22	0.26	0.32	0.27
	Train	PBIAS	0.00	0.00	0.00	-16.18	0.00	-0.50	0.00	1.78	0.00	0.12
	Test	R ²			0.23	0.14	0.23	0.18	0.23	0.27	0.40	0.31
	Test	RMSE			4.40	4.88	4.47	4.74	4.40	4.28	3.90	4.17
	Test	NSE			0.23	0.05	0.21	0.10	0.23	0.27	0.40	0.31
	Test	PBIAS			-1.49	-19.69	-1.99	-3.03	0.00	4.87	0.00	0.14
Yauricocha	Train	R ²	0.21	0.24	0.26	0.35	0.37	0.63	0.26	0.28	0.34	0.29
	Train	RMSE	4.26	4.16	4.20	3.93	3.88	3.01	4.20	4.14	3.97	4.09
	Train	NSE	0.21	0.24	0.26	0.35	0.36	0.62	0.26	0.28	0.34	0.29
	Train	PBIAS	0.00	0.00	0.00	-4.35	0.00	-1.14	0.00	6.98	0.00	-0.02
	Test	R ²			0.20	0.16	0.21	0.18	0.20	0.23	0.33	0.26
	Test	RMSE			4.04	4.29	4.02	4.21	4.00	3.94	3.68	3.87
	Test	NSE			0.19	0.09	0.20	0.12	0.20	0.23	0.33	0.26
	Test	PBIAS			-7.50	-11.81	-6.61	-7.52	0.00	7.95	0.00	0.14
Vilca	Train	R ²	0.21	0.28	0.29	0.40	0.41	0.70	0.29	0.31	0.39	0.33
	Train	RMSE	3.75	3.56	3.52	3.26	3.21	2.34	3.52	3.47	3.29	3.44
	Train	NSE	0.21	0.28	0.29	0.39	0.41	0.69	0.29	0.31	0.38	0.33
	Train	PBIAS	0.00	0.00	0.00	6.39	0.00	-0.51	0.00	5.07	0.00	0.01
	Test	R ²			0.25	0.21	0.27	0.18	0.26	0.27	0.43	0.33
	Test	RMSE			3.67	3.90	3.64	3.98	3.66	3.62	3.23	3.48
	Test	NSE			0.25	0.16	0.27	0.12	0.26	0.27	0.42	0.33
	Test	PBIAS			1.97	7.35	1.59	-2.08	0.00	2.97	0.00	0.23