The following supplement accompanies the article

Stochastic generation of daily precipitation amounts: review and evaluation of different models

Jie Chen[1,2,*], François P. Brissette[2]

[1]State Key Laboratory of Soil Erosion and Dryland farming on Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Yangling, Shaanxi, 712100, China

[2]Department of Construction Engineering, École de technologie supérieure, Université du Québec, 1100, rue Notre-Dame Ouest, Montreal, Quebec H3C 1K3, Canada

*Email: jie.chen.1@ens.etsmtl.ca

Climate Research 59: 189–206 (2014)

Supplement 1. Additional data

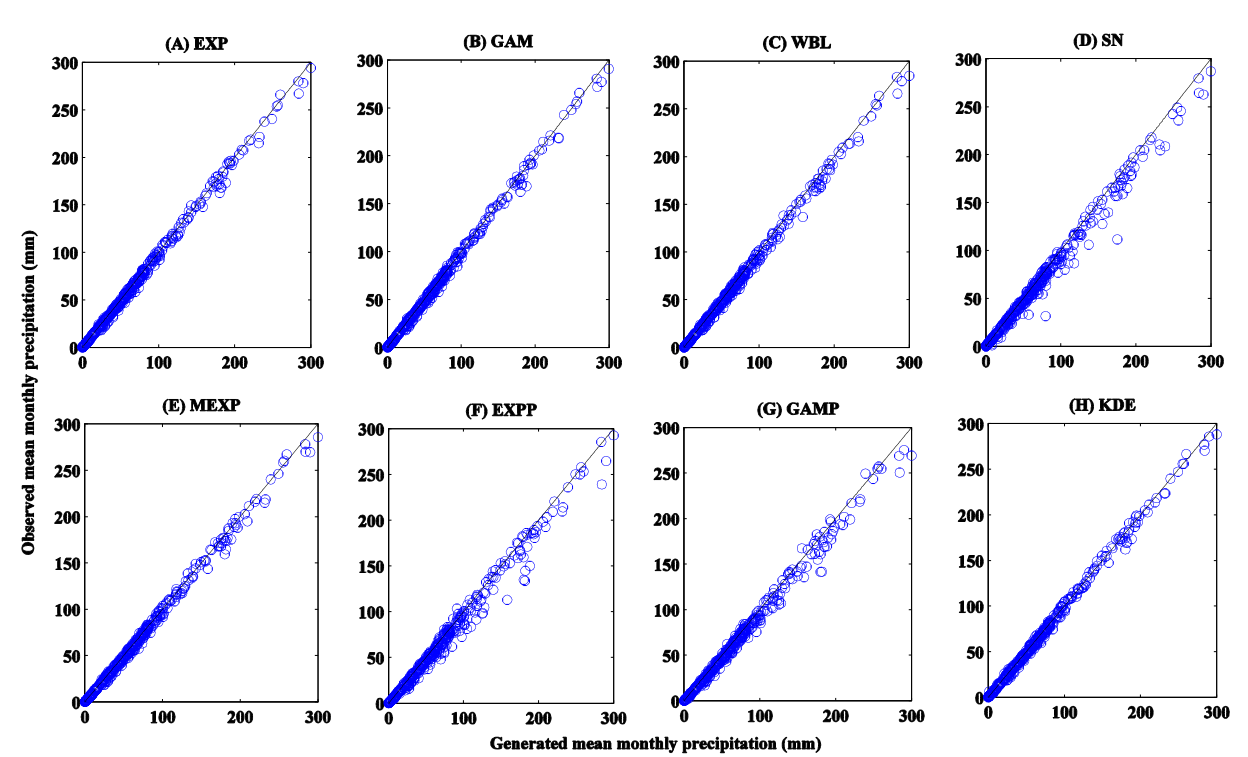

Fig. S1.Observed vs. 8 model-generated mean monthly precipitations for all station-month combinations

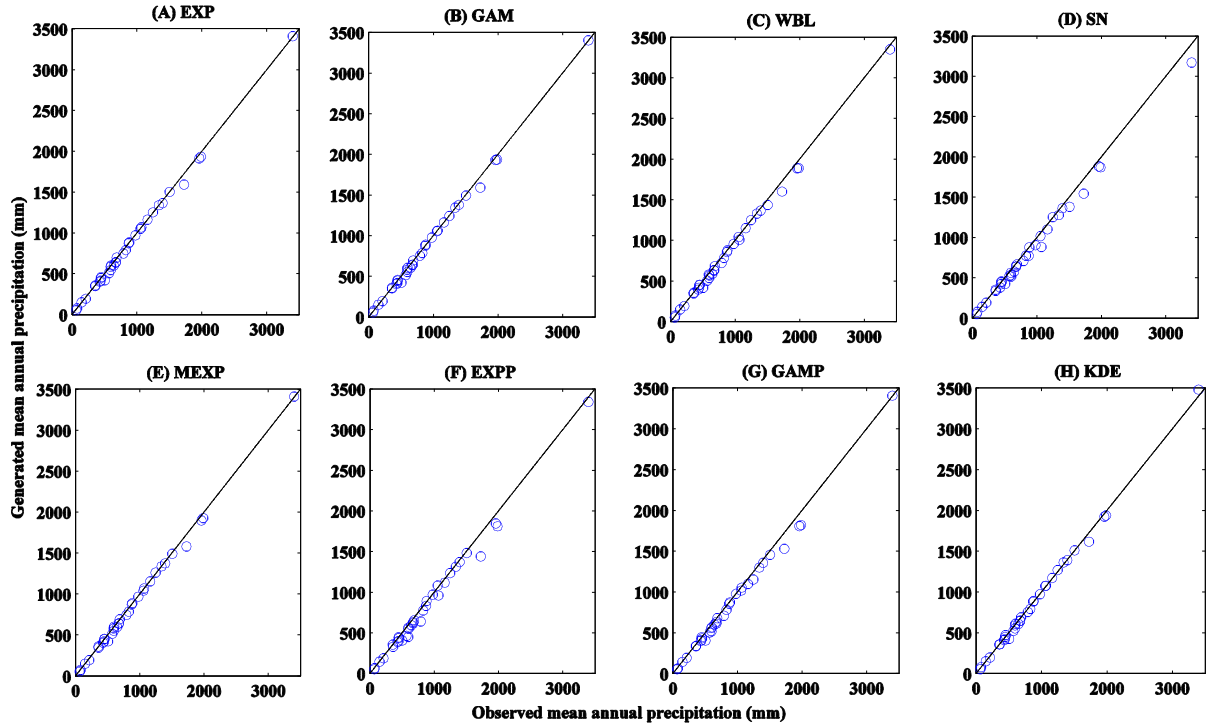

Fig. S2. Observed vs. 8 model-generated mean annual precipitations for all 35 stations

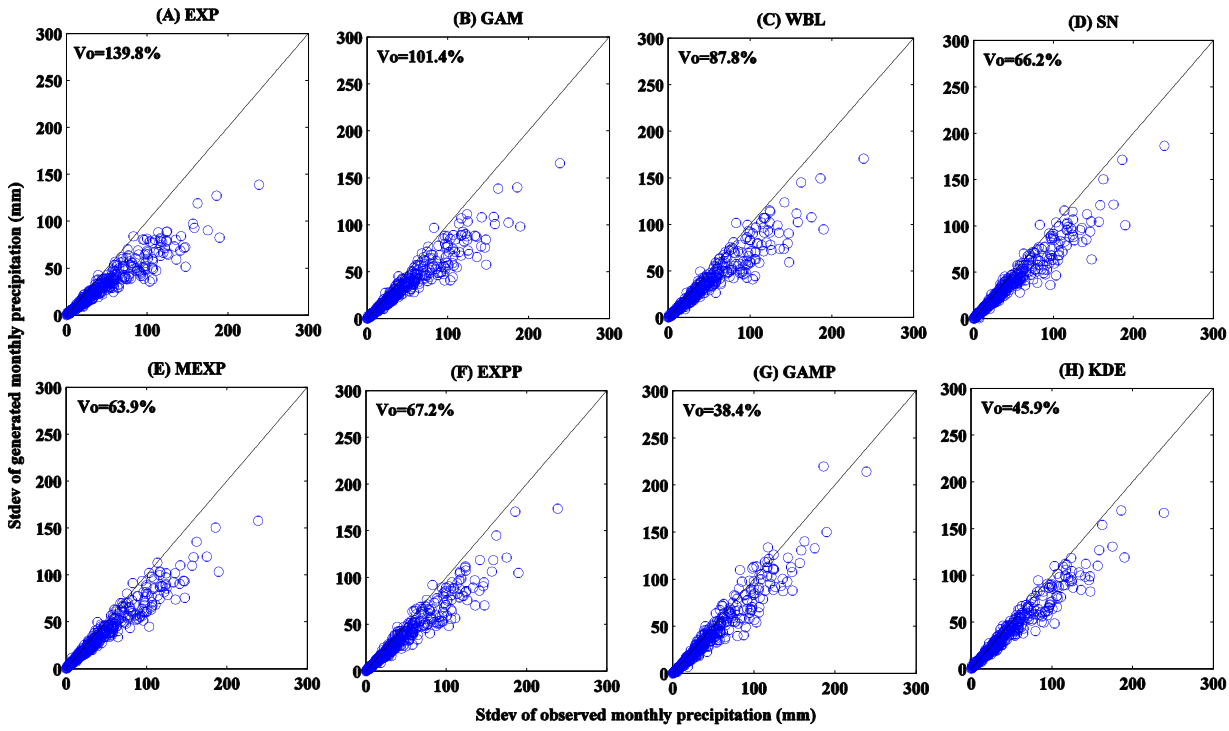

Fig. S3. Standard deviation (Stdev) of observed vs. generated monthly precipitation from 8 models for all station-month combinations. Vo: variance overdispersion

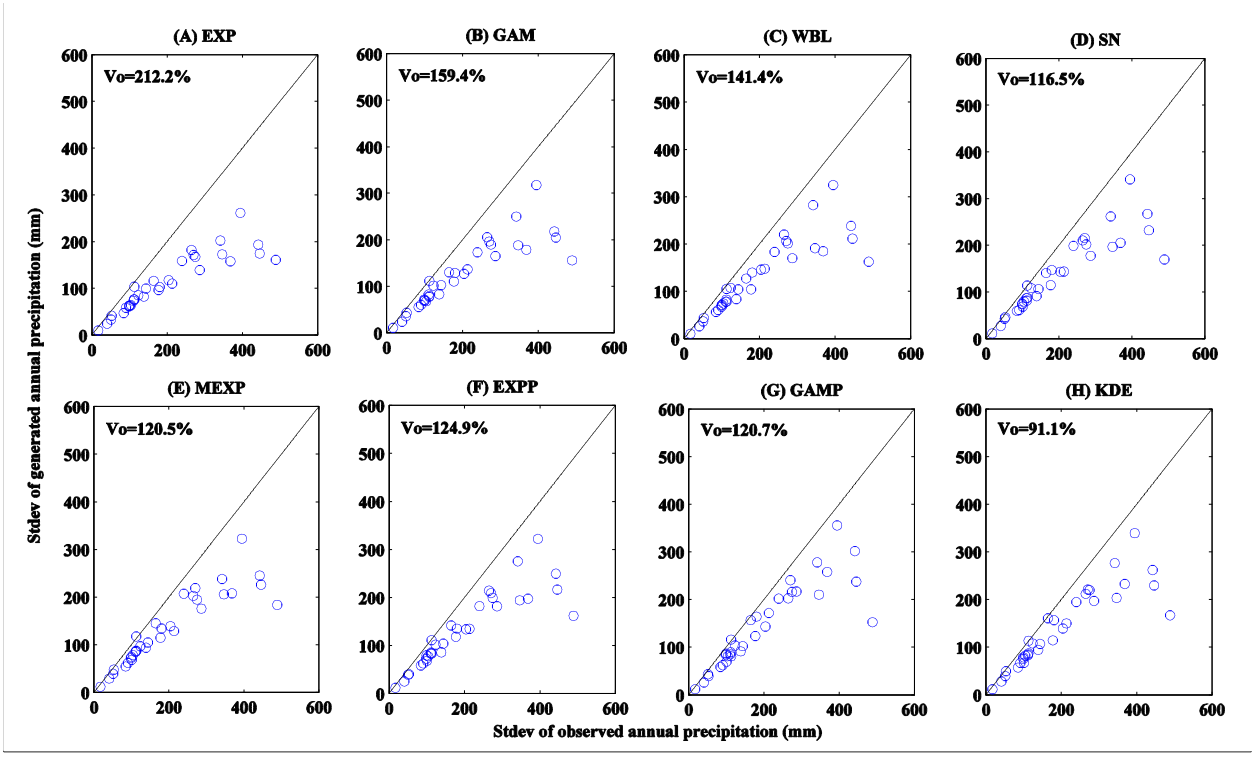

Fig. S4. Standard deviation (Stdev) of observed vs. generated annual precipitation from 8 models for all 35 stations. Vo: variance overdispersion

Table S1. Literature review of the generation of daily precipitation amount

ID	Source	Applied distribution	Study area	Conclusion
1	Bardossy & Plate (1992)	A transformed conditional multivariate autoregressive AR(I) model, with parameter conditioning on the atmospheric circulation pattern	44 stations in the river Ruhr catchment, Germany	Suitable for simulating the local precipitation under both stationary and nonstationary climate conditions.
2	Chapman (1997)	Kappa, Weibull, gamma, mixed exponential and skewed normal distributions, and a Srikanthan-McMahon model	22 islands in Western Pacific	The mixed exponential and skewed normal distributions successfully produced the rainfall amounts. The Srikanthan-McMahon model performed better than other models for 18 out of 22 stations, due to the larger number of parameters.
3	Chen et al. (2009)	Skewed normal distribution used by CLIGEN	12 stations in the Loess Plateau of China	The skewed normal distribution reproduced the mean of daily, monthly and annual precipitation reasonably well. However, the standard deviation and extremes were somewhat less well reproduced.
4	Chen et al. (2012)	Exponential, gamma, skewed normal, and mixed exponential distributions used by WeaGETS	2 stations in Canada	The mixed exponential and skewed normal distributions were consistently better than the exponential and gamma distributions at generating daily precipitation amounts. The mixed exponential distribution performed better than all 3 other distributions at representing the extreme precipitation events.
5	Cowpertwait (2004)	Neyman-Scott Rectangular Pulses model	Auckland, New Zealand (Hourly data)	The Neyman-Scott Rectangular Pulses model reasonably reproduced the mean of the observed precipitation amount and was also able to represent extreme values.
6	Yusof et al. (2007)	Exponential and mixed exponential distributions, and Neyman-Scott Rectangular Pulses model	A station in Wilayah Persekutuan (Sub-daily timescale)	Mixed exponential distribution performed better than other models at reproducing most statistics of the observed precipitation.
7	Furrer & Katz (2007)	Gamma generalized linear modeling (GLM) with a logarithmic link function	Daily precipitation at Pergamino in the Pampas region of central-eastern Argentina	The extreme precipitation and inter-annual variability of precipitation were underestimated.
8	Furrer & Katz (2008)	Gamma distribution, stretched exponential model and hybrid gamma and Pareto distribution	Fort Collins in the USA and Pergamino in Argentina	None of these was able to fully represent the extreme precipitation. However, the hybrid gamma and Pareto distribution is superior to the other models.
9	Hanson & Vogel (2008)	Kappa, gamma and skewed normal distributions	237 stations across 49 USA states	The Kappa and skew normal distribution are superior to the gamma distribution at simulating the daily precipitation.

10	Hundecha et al. (2009)	Transformed truncated normal distribution	122 stations within the Unstrut catchment in central eastern Germany	The model reproduced a number of daily precipitation statistics fairly well at both individual stations and at the catchment scale.
11	Ison et al. (1971)	Gamma distribution	3 stations (Garden City, Manhattan and Columbus) in Kansas, USA	Goodness-of-fit tests supported the proposed probability model for describing variation in observed data.
12	Jamaludin & Jemain (2007)	Exponential, gamma and Weibull distributions and their mixed distributions	18 rain gauge stations in Peninsular Malaysia	The compound distributions were better than the single distribution; the mixed Weibull was the most appropriate model for the majority of sites in Peninsular Malaysia. However, results were varied depending on the geographical, topographical and climatic conditions of the studied sites.
13	Johnson et al. (1996)	Skewed normal distribution used by CLIGEN and the mixed exponential distribution used by USCLIMATE	6 locations across the USA	Both models adequately reproduced the characteristics of daily precipitation, with the exception of extreme daily amounts. Both models significantly underestimated daily precipitation amounts for several locations.
14	Kilsby et al. (2007)	Neyman-Scott Rectangular Pulses model	UK	A daily weather generator used for climate change impact studies.
15	Li et al. (2012)	Kappa, exponential, gamma, mixed exponential, dynamic mixture of gamma and generalized Pareto, hybrid gamma and generalized Pareto and hybrid exponential and generalized Pareto distribution	49 stations across Texas, USA	Commonly used single distributions cannot realistically represent extreme rainfall events, but compound distributions displayed the better performance. However, the fitting of these compound distributions is subject to several drawbacks, such as functional complexity, numerical instability, data sensitivity and supervised learning. The hybrid exponential and Pareto distribution is relatively simple and was reliable in modeling the entire range of precipitation distribution.
16	Li et al. (2005)	Generalized Pareto distribution	5 weather stations in southwest and western Australia	The Generalized Pareto distribution adequately reproduced the upper tail of the daily precipitation distribution.
17	Liu et al. (2011)	Exponential, gamma, mixed-exponential and lognormal distributions	10 stations in the Yishu River watershed, China	The lognormal distribution overestimated the statistics of monthly precipitation. The mixed exponential distribution also underestimated the monthly mean precipitation, while the exponential and gamma distributions produced the monthly precipitation well. All exponential, gamma and mixed exponential distributions underestimated the monthly maximum daily precipitation.

18	Makhnin & Mcallister (2009)	Truncated and power-transformed normal distribution	22 rain gauges in southwestern Colorado, 27 sites in Minnesota and 25 sites in the northern part of New York State	This method unified the precipitation occurrence and amount process. It has the advantage of overcoming the spatial intermittence problem and producing the lag-1 cross covariances. However, the parameters' estimation is computationally intensive.
19	Mckague et al. (2005)	Weibull distribution used by ClimGen	21 stations in Southern Ontario, Canada	ClimGen underestimated the daily precipitation in the winter, while overestimating in the summer. Moreover, ClimGen underestimated both the higher and lower extremes for most sites, with the difference being <5%.
20	Mehrotra & Sharma (2007)	Nonparametric kernel density estimation	30 rain gauge stations around Sydney in eastern Australia	The nonparametric kernel density estimation reproduced the main rainfall statistics fairly accurately at the daily as well as at longer timescales.
21	Nguyen & Mayabi (1991)	Exponential, gamma, Weibull and mixed exponential distributions	9 rain gauge stations in the Montreal region	The mixed exponential distribution was the best model, followed by the Weibull and gamma distributions. The exponential distribution showed the worst performance.
22	Nicks & Gander (1994)	Skewed normal distribution used by CLIGEN	CLIGEN was widely used in North America, Australia and China	The skewed normal distribution reproduced the mean daily, monthly, and annual precipitation reasonably well. The observed standard deviation and extremes are slightly underestimated.
23	Qian et al. (2004)	The semi-empirical (using 10 intervals from observed daily precipitation amount to simulate daily precipitation) used by LARS-WG and a semi-empirical distribution estimated from logarithm-transformed precipitation used by AAFC-WG	9 stations across Canada	Both weather generators not only reasonably reproduced the daily precipitation-associated statistics but also satisfactorily reproduced the extreme values of daily precipitation.
24	Rajagopalan & Lall (1999)	K-nearest-neighbor (K-NN) resampling Algorithm	Salt Lake City, Utah, USA	The precipitation amount properties were well preserved. The k-NN approach is better than CLIGEN at preserving cross-dependence and the frequency structure.
25	Rajagopalan et al. (1997)	Nonparametric kernel density estimation	2 synthetic data sets	The nonparametric method worked nearly as well as the parametric method, where the latter was appropriate. The former performed better, where the latter was inappropriate.
26	Richardson (1981)	Exponential distribution used by WGEN	WGEN was widely used in North America and Europe	
27	Richardson & Wright (1984)	Gamma distribution used by WGEN	WGEN was widely used in North America and Europe	

28	Safeeq & Fares (2011)	Weibull distribution used by ClimGen	4 tropical watersheds located in Hawai'i USA	The Weibull distribution was only able to reproduce the observed precipitation distribution for 1 of 4 locations. It also underestimated the number of extreme rainfall events.
29	Schoof (2008)	Mixed exponential distribution	29 stations in the USA	The mixed exponential distribution closely reproduced the observed precipitation amount but underestimated the inter-annual variability.
30	Schoof et al. (2005)	Mixed exponential distribution	9 stations in the southeast USA	The mean precipitation amount was adequately reproduced across a range of timescales, but the variability was generally underestimated.
31	Selker & Haith (1990)	Beta-P model, Weibull distribution and 2 other single-parameter precipitation distributions	33 sites in the east of Rocky Mountains, USA	The Weibull model was considerably better than the other models at reproducing precipitation probability distributions.
32	Semenov (2008)	Semi-empirical distribution with 10 intervals used by LARS-WG	20 sites with diverse climates (10 sites in Europe, 5 sites in USA, 1 site in Australia and 1 site in New Zealand)	LARS-WG accurately reproduced the mean yearly maximum daily precipitation at the 10 and 20 yr return periods for all sites.
33	Semenov & Barrow (2002)	Semi-empirical distribution	LARS-WG was widely used in Europe.	
34	Semenov et al. (1998)	Gamma distribution used by WGEN and semi-empirical distributions used by LARS-WG	18 sites in the USA, Europe and Asia	Both WGEN and LARS-WG underestimated the monthly and inter-annual variability. Neither generator performed uniformly well in simulating the variance of daily precipitation.
35	Skiles & Richardson (1998)	Gamma distribution used by WGENAL	8 stations in Alaska, USA	WGENAL was developed based on WGEN, especially for the US state of Alaska. It had a similar performance to that of WGEN.
36	Stöckle et al. (1999)	Weibull distribution used by ClimGen	ClimGen was used in North America	
37	Todorovic & Woolhiser (1975)	Exponential distribution	Austin, Texas, USA	The exponential distribution was able to simulate the small and medium precipitation events but underestimated the heavy precipitation.
38	Van Montfort & Witter (1986)	Generalized Pareto distribution	One station in Netherlands (De Bilt) and 32 Dutch rainfall stations	The generalized Pareto distribution is appropriate to reproduce the series of peaks-over-threshold rainfall depth.
39	Vrac & Naveau (2007)	Hybrid gamma and Pareto distribution	37 stations in Illinois, USA	It is very important to use Extreme Value Theory (EVT) distributions to improve the modeling and downscaling of local extremes.

40	Wallis & Griffiths (1995)	Skewed normal distribution by WXGEN	5 stations in Texas, USA	The weather sequences generated by WXGEN are often unrealistic. *F*- and *t*-tests carried out on monthly rainfall showed 2 of the lowest failure rates (~10%) encountered overall.
41	Wan et al. (2005)	Exponential, gamma, skewed normal and mixed exponential distributions	657 stations across Canada	The mixed exponential distribution is generally superior to the other distributions, especially during warmer months, while the gamma distribution is adequate for the winter.
42	Wilks (1999b)	Gamma, Common-a gamma and mixed exponential distributions	30 locations across the USA	The commonly used gamma distribution is inferior to the mixed exponential distribution at reproducing the daily precipitation for all locations.
43	Wilks (1998)	Gamma and mixed exponential distributions	25 locations in New York State, USA	The mixed exponential distribution not only provided better fits than the gamma distribution but was also better at representing the tendency for smaller amounts at locations near the edges of wet areas. The mixed exponential distribution was also better than the gamma distribution at preserving the inter-annual variability.
44	Wilks & Wilby (1999)	Exponential, gamma and mixed exponential distributions	Ithaca, New York, USA	The exponential distribution underestimated the very small and very large precipitation amounts. The gamma distribution substantially improved the representation of small amounts but only slightly improved the representation of large amounts. The mixed exponential distribution displayed the best performance among all 3 distributions.
45	Woolhiser & Roldán (1982)	Chain-dependent and independent exponential, gamma and mixed exponential distributions	5 USA stations	The independent mixed exponential distribution displayed the best performance, followed by the independent gamma distribution. The chain-dependent gamma distribution performed the worst.
46	Zhang and Garbrecht (2003)	Skewed normal distribution used by CLIGEN	4 stations in Oklahoma, USA	The skewed normal distribution reproduced the means of daily, monthly and annual precipitation reasonably well. However, the seasonal and inter-annual variability and maximum daily precipitation were not as well- simulated.
47	Zheng and Katz (2008)	Mixture of generalized chain-dependent process conditioned on a climate variable	4 rainfall stations in New Zealand	The mixture of generalized chain-dependent processes conditional on the inter-decadal Pacific Oscillation was an effective way to reduce the overdispersion of seasonal rainfall.

LITERATURE CITED

Bardossy A, Plate E (1992) Space-time model for daily rainfall using atmospheric circulation patterns. Water Resour Res 28: 1247–1259

Chapman TG (1997) Stochastic models for daily rainfall in the Western Pacific. Math Comput Simul 43: 351-358

Chen J, Brissette FP, Leconte R, Caron A (2012) A versatile weather generator for daily precipitation and temperature. Transactions of the ASABE 55: 895-906

Chen J, Zhang XC, Liu WZ, Li Z (2009) Evaluating and Extending CLIGEN precipitation Generation for the Loess Plateau of China. J Am Water Res Assoc 45: 378-396

Cowpertwait PSP (2004) Mixed rectangular pulses models of rainfall. Hydrol Earth Syst Sci 8: 993-1000

Furrer EM, Katz RW (2007) Generalized linear modeling approach to stochastic weather generators. Clim Res 34: 129-144

Furrer EM, Katz RW (2008) Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resour Res 44: W12439, doi:10.1029/2008WR007316

Hanson LS, Vogel R (2008) The probability distribution of daily rainfall in the United States. World Environmental and Water Resources Congress 2008, p 1-10, doi: 10.1061/40976(316)585

Hundecha Y, Pahlow M, Schumann A (2009) Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes. Water Resour Res 45: W12412

Ison NT, Feyerherm AM, Bark LD (1971) Wet period precipitation and the Gamma distribution. J Appl Meteor 10: 658–665

Jamaludin S, Jemain AA (2007) Fitting daily rainfall amount in Peninsular Malaysia using several types of exponential distributions. J Appl Sci Res 3: 1027-1036

Johnson GL, Hanson CL, Hardegree SP, Ballard EB (1996) Stochastic Weather Simulation: Overview and Analysis of Two Commonly Used Models. J Appl Meteorol 35: 1878-1896

Kilsby CG, Jones PD, Burton A, Ford AC and others (2007) A daily weather generator for use in climate change studies. Environ Model Softw 22: 1705–1719

Li C, Singh VP, Mishra AK (2012) Simulation of the entire range of daily precipitation using a hybrid probability distribution. Water Resour Res 48: W03521, doi:10.1029/2011WR011446.

Li Y, Cai W, Campbell EP (2005) Statistical Modeling of Extreme Rainfall in Southwest Western Australia. J Clim 18: 852-863

Liu Y, Zhang W, Shao Y, Zhang K (2011) A comparison of four precipitation distribution models used in daily stochastic models. Adv Atmos Sci 28: 809-820

Makhnin O, Mcallister DL (2009) Stochastic precipitation generation based on a multivariate autoregression model. Hydrometeorol 10: 1397-1413

Mckague K, Rudra R, Ogilvie J, Ahmed I, Gharabaghi B (2005) Evaluation of weather generator ClimGen for Southern Ontario. Can Water Resour J 30: 315-330

Mehrotra R, Sharma A (2007) A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. J Hydrol (Amst) 335: 180–193

Nguyen VTV, Mayabi A (1991) Probabilistic analysis of summer daily rainfall for the Montreal region. Can Water Resour J 16: 65–80

Nicks AD, Gander GA (1994) CLIGEN: a weather generator for climate inputs to water resource and other models. Proc 5th Int Conf Comp Agricult. American Society of Agricultural Engineers, St. Joseph, MI, p 3–94

Qian B, Gameda S, Hayhoe H, Jong RD, Bootsma A (2004) Comparison of LARS-WG and AAFC-WG stochastic weather generators for diverse Canadian climates. Clim Res 26: 175–191

Rajagopalan B, Lall U (1999) A k-nearest-neighbor simulator for daily precipitation and other variables. Water Resour Res 35: 3089–3101

Rajagopalan B, Lall U, Tarboton DG (1997) Evaluation of kernel density estimation methods for daily precipitation resampling. Stoch Hydrol Hydraul 11: 523–547

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 17: 182–190

Richardson CW, Wright DA (1984) WGEN: a model for generating daily weather variables. US Department of Agriculture, Agricultural Research Service, ARS-8

Safeeq M, Fares A (2011) Accuracy evaluation of ClimGen weather generatorand daily to hourly disaggregation methodsin tropical conditions. Theor Appl Climatol 106:321-341

Schoof JT (2008) The multivariate spectral weather generator: improvements and an application to diverse climates within the contiguous United States. Agric For Meteorol 148: 517–521

Schoof JT, Arguez A, Brolley J, O'Brien JJ (2005) A new weather generator based on spectral properties of surface air temperatures. Agric For Meteorol 135: 241–251

Selker JS, Haith DA (1990) Development and testing of single- parameter precipitation distributions. Water Resour Res 26: 2733–2740

Semenov MA (2008) Simulation of extreme weather events by a stochastic weather generator. Clim Res 35: 203–212

Semenov MA, Barrow EM (2002) LARS-WG, a stochastic weather generator for use in climate impact studies, user manual. Rothamstead Research, Harpenden

Semenov MA, Brooks RJ, Barrow EM, Richardson CW (1998) Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. Clim Res 10: 95–107

Skiles JW, Richardson CW (1998) A stochastic weather generation model for Alaska. Ecol Model 110: 211–231

Stöckle CO, Campbell GS, Nelson R (1999) ClimGen manual. Washington State University, Pullman, WA

Todorovic P, Woolhiser DA (1975) A stochastic model for n-day precipitation. J Appl Meteorol 14: 17–24

Van Montfort MAJ, Witter JV (1986) The Generalized Pareto distribution applied to rainfall depths. Hydrol Sci J 31: 151–162

Vrac M, Naveau P (2007) Stochastic downscaling of precipitation: from dry events to heavy rainfalls. Water Resour Res 43: W07402, doi: 10. 1029/ 2006 WR 005308

Wallis TWR, Griffiths JF (1995) An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC). Agric For Meteorol 73: 115–133

Wan H, Zhang X, Barrow EM (2005) Stochastic modelling of daily precipitation for Canada. Atmos-ocean 43: 23–32

Wilks DS (1998) Multisite generalization of a daily stochastic precipitation generation model. J Hydrol (Amst) 210: 178–191

Wilks DS (1999b) Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric For Meteorol 93: 153–169

Wilks DS, Wilby RL (1999) The weather generation game: a review of stochastic weather generator. Prog Phys Geogr 23: 329–357

Woolhiser DA, Roldán J (1982) Stochastic daily precipitation models. 2. A comparison of distributions of amounts. Water Resour Res 18: 1461–1468

Yusof F, Mohd. Daud Z, Nguyen, VTV, Yusop Z (2007) Perfomance of mixed exponential and exponential distribution representing rain cell intensity in Neyman-Scott Rectangular Pulse model. Malays J Civ Eng 19: 55-72

Zhang XC, Garbrecht JD (2003) Evaluation of CLIGEN precipitation parameters and their implication on WEPP runoff and soil loss prediction. Trans ASAE 46: 311–320

Zheng X, Katz RW (2008) Mixture model of generalized chain dependent processes and its application to simulation of interannual variability of daily rainfall. J Hydrol (Amst) 349: 191–199