3.2.1 Taylor diagram based overall model performance

Taylor diagram highlights both model selection uncertainty and fitting performance. Figure 8 shows the Taylor diagrams for three cases, where the best-fit models according to Z-test show different performances, including: i) all of the top three distributions are performing similarly, ii) there is a relatively bad performing distribution in the top three, and iii) the best-fit (Rank-1) distribution is not the best in terms of the Taylor diagram. Considering these cases allows us to evaluate the BMA performance under different plausible scenarios of distributions combination, and examine whether there will be cases where the BMA would be less desirable compared to the best-fit distribution.

The results in all of the three cases suggest that BMA provides better overall performance for the majority of the datasets - 70% of the total, compared to the best-fit distributions (Figure 9). As Figure 9 shows for the cases where BMA is not better than the best-fit distribution, the difference between them is not big. Figure 8-a shows that all distributions are performing almost equally well, and the BMA does not show much difference from the pooled distributions. In this case, in the absence of an outright best or worst model, making a consensus prediction using BMA is reasonable than relying on a single distribution. Figure 8-b shows that the performances of BMA and best-fit distribution are similar and close to the observed data. The result of the BMA is not affected by the inclusion of the lower ranked distribution(s). The prior that embodies the ranks of the distributions and the likelihood of Rank-1 distribution prevents BMA from drifting to the low performing distributions by providing more weight to Rank-1 distribution. Figure8-c shows that the Rank-1 distribution and BMA performances are different. In this case BMA mostly prefers the alternative distributions than the Rank-1 distribution according to the metrics considered under Taylor diagram. Here, the likelihood in BMA has more effect than the prior in combining the distributions. Partly, the difference is because of the use of Z-test for selection of distribution, which uses higher order moments - skewedness and kurtosis of the data, while the Taylor diagram is based on second order moment. There are 305 instances where Rank-1 is not better than Rank-2 or Rank-3 when Taylor diagram is used as a performance measure. Under this condition, in 95% of the cases BMA improves the performance of Rank-1 distribution in terms of Taylor diagram. Thus, in this case the preferred solution is to consider the result of the averaged consensus.

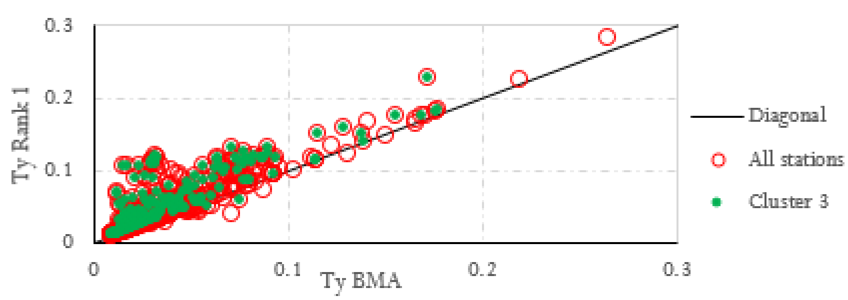

Figure 9. Summarized Taylor diagram result of the BMA and Rank-1 distribution. Cluster 3 indicates cases where the Rank-1 model is not better than Rank-2 in terms of Taylor diagram.

The visual interpretation represented under Taylor diagram implicitly evaluates the distances from the alternative distributions to the observed data. The shorter the distance, the preferred the alternative. This distance has three components, measuring: a) how close is the distribution correlation coefficient to one, b) how close is the standard deviation to the observed standard deviation, and c) how small is the centered root mean square error. In order to quantitatively represent the Taylor diagram, we have summarized these three components as:

$$Ty = \left(1 - r^2\right) + \left|\sigma_{obs} - \sigma_{dist}\right| + CRMS \tag{11}$$

where Ty is the summarized metrics of Taylor diagram and r^2 is the correlation coefficient.

Figure 9 presents the summarized Taylor diagram metrics (Equation 11) comparing Rank-1 and BMA for all the datasets considered. The figure shows that the performances of the BMA and Rank-1 distribution are similar, i.e. most of the points are aligned on the diagonal indicating BMA tends to favor Rank-1 distribution. Overall, the BMA has improved the performance of Rank-1 as there are more points above the diagonal line. In addition, the points that are above the diagonal are scattered relatively farther from the diagonal than those few points below the diagonal, indicating the BMA does not severely degraded the performance of the Rank-1.

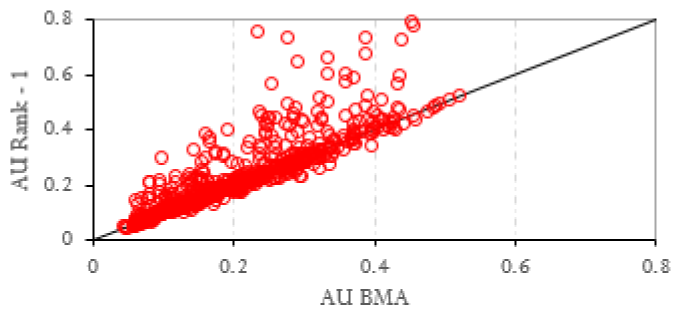

Figure 10. BMA vs. Rank -1 model comparison based on Anderson – Darling performance measure.

3.2.2 Result based on Anderson Darling (AU)

The lower the value of AU the preferred the distribution in representing extreme events. Figure 10 shows that the performance of BMA is mostly dictated by Rank-1 distributions, as most of the data points lie along the diagonal. However, the figure also shows that BMA has performed better than Rank-1 as there are higher number of data points above the diagonal.

3.2.3 Verification of BMA predictive capacity

The bootstrap tests are performed for predicting the five most extreme events. The result of the bootstrap test suggests that BMA has an overall similar predictive performance with the Rank-1 distributions (Figure 11 a-c). The predictive accuracy (Figure 11b) is calculated as the root mean squared difference between the median and the observed five extremes while the predictive uncertainty (Figure 11c) is calculated by a total sum of the difference between the 75% and 25% intervals. In addition to Rank-1 and BMA, we have analyzed the predictive performance of one un-pooled distribution (Beta-K) to illustrate the effectives of ranking based on Z-test.

As shown in Figure 11a, the line fitted to the five extremes fall within the 25% and 75% bands and is close to the median values for the BMA and Rank-1 distribution. On the contrary, the fit for the un-pooled model lies far from the median and is not enclosed by the 25% and 75% bands. The accuracy (Figure 11b) and uncertainty (Figure 11c) summary for the nine stations also showed that the un-pooled model (Beta–K) has a poor accuracy compared to both Rank-1 and BMA when it is not part of the top three pooled models. Beta-K was not ranked as the top model in 6 out of 9 cases, and in those cases, its predictive accuracy is lower than BMA and Rank-1 distribution. This highlights the usefulness of Z-test results for ranking the distributions and its use as a prior in the Bayesian augmentation of the L-moment approach. In terms of uncertainty, BMA has an identical level of uncertainty to Rank-1 and Beta-K distributions (Figure 11c). This is because BMA based uncertainty has as its lower limit the uncertainty band corresponding to the least uncertain distribution among the three distributions and mostly follows rank-1 distribution.

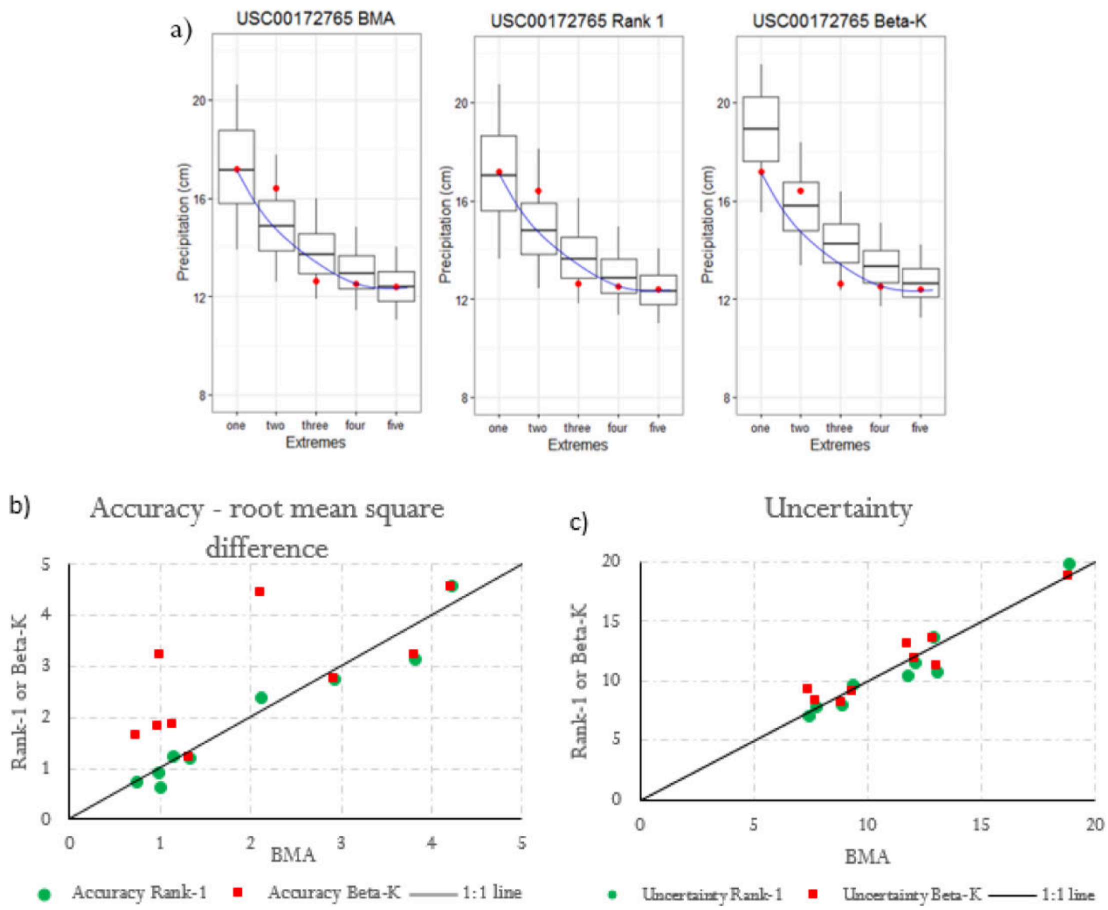

Figure 11. Bootstrap based accuracy and confidence intervals for the 9 stations with record lengths longer than 80 years. a) Whisker plot showing accuracy and uncertainty, b) comparison of accuracies of the distributions in terms of root mean square error, and c) comparison of uncertainties computed based on the 25% and 75% quantiles.

Based on the different performance measures used in this study, BMA has proved to be effective in characterizing extreme events compared to most of the best-fit distributions ranked by Z-test or any model that has a lower rank (un-pooled) which can be recommended. The improvement is attributed to both the likelihood, and the use of prior and ranking under Z-test.

3.3 Implications to IDF

Intensity (or Depth) duration frequency (IDF) curves, are standard tools used for representing the interplay among extreme event magnitudes, return periods and the durations of interest.

Figure 12 shows the implication of uncertainty on DDF curves that leads to overlap of the DDF curves. It indicates that the 95% upper limit for the 50 year return period curve is intersecting within the 100 year return period 95% credible interval. This could have implications on design specifications that need due consideration. Presenting both parameter and model uncertainty confidence bands developed along with the IDF or DDF curves allows a transparent decision making that acknowledges uncertainty and its sources. This benefits an informed design in applied hydrology complementing the safety factor concept to uncertainty quantification which is not transparent enough to quantify and associate the uncertainty sources.

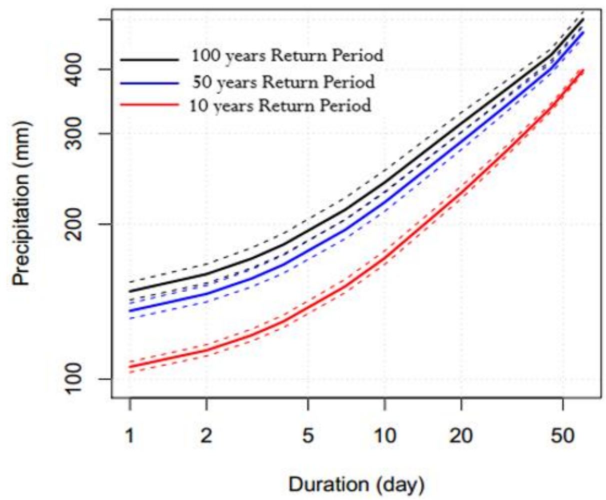

Figure 12. Implication of uncertainty analysis on Depth Duration Frequency (DDF) curves at station USC00279940.

4. SUMMARY AND CONCLUSION

Although any single distribution can be a good fit, this study has rather shown that at any site it is not always clear that a single distribution could be an outright best choice. Instead a Bayesian augmented consensus prediction is found out to be a better alternative as diagnosed by multiple efficiency measures that probe the overall performance, extreme events, and predictive capacity based on a wide range of data. The standard L-moment based scheme in RFA has an obvious practicality for its ease of application, less sensitive to outliers and sampling variability. The core deficits of the method are, however, it can still be influenced by outliers and is reliant on a single statistic particularly in model selection. In addition the scheme has no direct approach to address parameter uncertainty. Therefore, to better enable decision making, avoiding overdesign or risk of failure, it is essential to build on the success of L-moment approach through Bayesian statistics aided by multiple performance measures.

In the Bayesian augmented L-moment approach, any distribution can be ranked based on Z-test, and pooled together. As a result, the limitation of any single model can be improved by the additional skill drawn from the alternative pooled distributions. Moreover, the use of L-moments in ranking and as a prior for the Bayesian approach preserve the benefit of L-moment approach to RFA. In parameter estimation, the Bayesian framework has delivered not just the optimal parameter estimate, but also directly the associated parameter posterior distribution.

REFERENCES

Ahmad, M. I., C. D. Sinclair, and B. D. Spurr. 1988. "Assessment of Flood Frequency Models Using Empirical Distribution Function Statistics." *Water Resources Research* 24(8):1323–28.

Bobée, B., G. Cavadias, F. Ashkar, J. Bernier, and P. Rasmussen. 1993. "Towards a Systematic Approach to Comparing Distributions Used in Flood Frequency Analysis." *Journal of Hydrology* 142(1–4):121–36.

Cohn, T. A., W. L. Lane, and W. G. Baier. 1997. "An Algorithm for Computing Moments-Based Flood Quantile Estimates When Historical Flood Information Is Available." *Water Resources Research* 33(9):2089–96.

Dietterich, Thomas G. 2000. "Ensemble Methods in Machine Learning." 1–15.

Fowler, H. J. and C. G. Kilsby. 2003. "A Regional Frequency Analysis of United Kingdom Extreme Rainfall from 1961 to 2000." *International Journal of Climatology* 23(11): 13–34.

Gaál, Ladislav, Ján Szolgay, Silvia Kohnová, Kamila Hlavčová, and Alberto Viglione. 2010. "Inclusion of Historical Information in Flood Frequency Analysis Using a Bayesian MCMC Technique: A Case Study for the Power Dam Orlík, Czech Republic." *Contributions to Geophysics and Geodesy* 40(2):121–47.

Gaume, E. et al. 2010. "Bayesian MCMC Approach to Regional Flood Frequency Analyses Involving Extraordinary Flood Events at Ungauged Sites." *Journal of Hydrology* 394(1–2):101–17.

Gelman, Andrew and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7(4):457–72.

Hall, M. J., H. F. P. van den Boogaard, R. C. Fernando, and A. E. Mynett. 2004. "The Construction of Confidence Intervals for Frequency Analysis Using Resampling Techniques." *Hydrology and Earth System Sciences* 8(2):235–46.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky. 1999. "Bayesian Model Averaging: A Tutorial (with Comments by M. Clyde, David Draper and E. I. George, and a Rejoinder by the Authors." *Statistical Science* 14(4):382–417.

Hosking, J. R. M. and J. R. Wallis. 1993. "Some Statistics Useful in Regional Frequency Analysis." *Water Resources Research* 29(2):271–81.

Hosking, Jonathan R. M. 1990. "L-Moment – Analysis and Estimation of Distributions Using Linear-Combinations of Order-Statistics." *Journal of the Royal Statistical Society. Series B: Methodological* 52(1).

Kidson, R. and K. S. Richards. 2005. "Flood Frequency Analysis: Assumptions and Alternatives." *Progress in Physical Geography* 29(3):392–410.

Liang, Zhongmin, Wenjuan Chang, and Binquan Li. 2011. "Bayesian Flood Frequency Analysis in the Light of Model and Parameter Uncertainties." *Stochastic Environmental Research and Risk Assessment* 26(5):721–30.

Moges, Edom, Yonas Demissie, and Hong-Yi Li. 2016. "Hierarchical Mixture of Experts and Diagnostic Modeling Approach to Reduce Hydrologic Model Structural Uncertainty." *Water Resources Research* n/a-n/a.

Murshed, Md.Sharwar, Sangil Kim, and Jeong-Soo Park. 2011. "Beta-κ Distribution and Its Application to Hydrologic Events." 25(7):897–911.

Öztekin, Tekin. 2007. "Wakeby Distribution for Representing Annual Extreme and Partial Duration Rainfall Series." *Meteorological Applications* 14(4):381–87.

Palmer, T. N., F. J. Doblas-Reyes, R. Hagedorn, and A. Weisheimer. 2005. "Probabilistic Prediction of Climate Using Multi-Model Ensembles: From Basics to Applications."

Philosophical transactions of the Royal Society of London. Series B, Biological sciences 360(1463):1991–98.

Polikar, R. 2006. “Ensemble Based Systems in Decision Making.” *IEEE Circuits and Systems Magazine* 6(3):21–45.

Raftery, Adrian E., Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski. 2005. “Using Bayesian Model Averaging to Calibrate Forecast Ensembles.” *Monthly Weather Review* 133(5):1155–74.

Reis, Dirceu S. and Jery R. Stedinger. 2005. “Bayesian MCMC Flood Frequency Analysis with Historical Information.” *Journal of Hydrology* 313(1–2):97–116.

Sankarasubramanian, A. and K. Srinivasan. 1999. “Investigation and Comparison of Sampling Properties of L-Moments and Conventional Moments.” *Journal of Hydrology* 218(1–2):13–34.

Singh, Vijay P. and W. G. Strupczewski. 2002. “On the Status of Flood Frequency Analysis.” *Hydrological Processes* 16(18):3737–40.

Stedinger, Jery R. and Timothy A. Cohn. 1986. “Flood Frequency Analysis With Historical and Paleoflood Information.” *Water Resources Research* 22(5):785–93.

Taylor, Karl E. 2001. “Summarizing Multiple Aspects of Model Performance in a Single Diagram.” *Journal of Geophysical Research: Atmospheres* 106(D7):7183–92.

Vrugt, J. A. and C. J. F. Ter_Braak. 2011. “DREAM$_{(D)}$: An Adaptive Markov Chain Monte Carlo Simulation Algorithm to Solve Discrete, Noncontinuous, and Combinatorial Posterior Parameter Estimation Problems.” *Hydrology and Earth System Sciences* 15(12):3701–13.

Wallis, J. R., M. G. Schaefer, B. L. Barker, and G. H. Taylor. 2007. “Regional Precipitation-Frequency Analysis and Spatial Mapping for 24-Hour and 2-Hour Durations for Washington State.” *Hydrology and Earth System Sciences* 11(1):415–42.

Wilks, Daniel S. 1993. “Comparison of Three-Parameter Probability Distributions for Representing Annual Extreme and Partial Duration Precipitation Series.” *Water Resources Research* 29(10):3543–49.

Stochastic Simulation of Daily Precipitation at Multiple Sites: II Performance Evaluation of the Model

Wai Wah Ng, Ph.D.[1]; Umed Panu, Ph.D.[2]; and Peter F. Rasmussen, Ph.D.[3]

[1]Dept. of Earth System Science and Policy, Univ. of North Dakota, Grand Forks, ND 58202, USA. E-mail: ng_waiwah@ymail.com
[2]Dept. of Civil Engineering, Lakehead Univ., Thunder Bay, ON L3V 0B9, Canada. E-mail: Umed.Panu@lakeheadu.ca
[3]Dept. of Civil Engineering, Univ. of Manitoba, Winnipeg, MB R3T 2N2, Canada

ABSTRACT

Weather generation models based on multivariate censored distribution (WG-MCD) and multivariate autoregressive censored process (WG-MACP) have been developed and presented in the paper entitled “Stochastic Simulation of Daily Precipitation at Multiple Sites: I Model Development”. In this paper, the performance of these models is evaluated by comparing discrepancies in attributes (e.g., mean, variance, correlation) obtained from the historical and simulated precipitation. Precipitation records from years 1961–1990 at ten climatic stations located in Manitoba, Canada, are adopted for the performance evaluation of models. Three performance measures (i.e., the coefficient of determination, the coefficient efficiency, and the root mean square error) identify a fairly strong relationship between the historical and simulated precipitation. Proposed models has been found to be suitable for reproducing the statistical characteristics of daily historical precipitations at multiple sites. The spatial and temporal dependencies appear to have been reasonably captured by the covariance and lag-1 covariance of the WG-MACP model. Other descriptors, such as probabilities of wet/dry-day, mean values, and variances, show good statistical agreement with similar statistical characteristics of the historical and simulated data sets. It was found that a better performance of the model could be obtained with use of a smaller number of stations, and with less number of statistical attributes to be preserved.

INTRODUCTION

Records of daily precipitation are probably the most extensively used data in environmental, climatological, hydrological, and water resources studies. For example, conducting a flood risk analysis for a river requires comprehensive historical river flow records, but availability of records may be limited in length. In such cases, rainfall-runoff models are required to provide the river flow information through the input of precipitation data. Precipitation series at times are too short or contain missing records, thus making reliable and meaningful analyses difficult. In such situations, stochastic precipitation generation models can be a great asset in providing alternate precipitation series that are consistent with the observed characteristics of the historical precipitation records.

When multiple sites are considered, a multivariate stochastic model is needed to generate the weather that should recognize the spatial and temporal variation of daily precipitation. Precipitation modeling at multiple sites is a challenging task due to difficulties in modeling the spatial dependence of precipitation amounts that have mixed distributions (i.e., consisting of discrete zero for dry-day and distributed as continuous function for above-zero records for wet-day).

In the accompanying paper entitled "Stochastic Simulation of Daily Precipitation at Multiple Sites: I Model Development", an alternative approach utilizing the concept of a multivariate censored distribution for the estimation of spatial dependence of daily precipitation was proposed. Weather generation models based on multivariate censored distribution and multivariate autoregressive censored process (WG-MCD and WG-MACP) have been developed. In this paper, the adequacy of multivariate censored normal distribution (MCD) is first verified and validated, and the performance of WG-MCD and WG-MACP models are then evaluated.

Since both generation models (i.e., WG-MCD and WG-MACP) share a core probability distribution structure of MCD, the correctness of computer coding and the adequacy of this component is first examined. Secondly, the performance in terms of preserving the statistical characteristics of the historical precipitation of the WG-MCD and WG-MACP models were compared. Further, a sensitivity analysis was conducted to evaluate the capability of models in handling a large number of variables. Therefore, this paper consists of four sections: (1) Validation of the MCD, (2) comparative analysis of the WG- MCD and WG-MACP models, and (3) Sensitivity analysis of different number of variables and their impact on parameter estimation.

STUDY AREA AND RELEVANT DATA SET

For the validation and rest of the section, a historical precipitation records was used. Daily precipitation records from year 1961 to 1990 from 10 stations located in southern Manitoba, Canada, were utilized. Availability of sufficient and reliable daily precipitation data is the reason for the selection of this region for the analysis. Ten stations along with pertinent information are described in Table 1. Stations 1 to 7 are located inside the Red River Basin, station 8 is located in the Assiniboine River Basin, and stations 9 and 10 are located inside the Winnipeg River Basin.

Precipitation records were obtained from the Canadian Daily Climate Data 2002 West CD-ROM [3]. Daily precipitation records obtained from the period between 1961 and 1990 is common in precipitation and climate change studies, as it is neither too short, nor too recent to include a strong global change signal [6]. With 30 years of daily precipitation records, a sample size of 30 is available to estimate the parameters for each Julian day at each station.

Table 1. Geographical information on selected stations.

Station	Name	Station Number	Latitude	Longitude	Elevation (m)
1	Deer Wood	5020720	49.4N	98.9W	338
2	Emerson	5020880	49.4N	97.2W	238
3	Morden	5021848	49.1N	98.5W	298
4	Plum Coulee	5022245	49.3N	97.8W	265
5	Steinbach	5022780	49.2N	96.6W	254
6	Altona	5020040	49.6N	97.3W	248
7	Morris 2	5021965	49.6N	97.9W	238
8	Minnedosa	5011760	50.6N	99.0W	521
9	Beausejour	5030160	50.2N	96.8W	251
10	Arborg	5030080	50.8N	96.3W	302

At 10 stations with 30-years records, a total of 109,575 daily precipitation records were available for the analysis. Of these, 29,069 (27%) records are above zero; 75,986 (69%) records are zeros, and 4,520 (4%) records are missing. The number of above-zero records and probabilities of dry-day were used for the estimation of parameters (mean vectors and correlation matrices) of the MCD and its variant forms.

Table 2. Annotation of types of data used for the calculation of attributes.

Notation	Description
MCD(m)	Refers to the multivariate censored normal distribution generated by the MCD, in which attributes (e.g., μ and σ) have been estimated using the historical data as input. The distribution consists of negative and positive values. Such attributes can also be calculated using a sufficient size of simulated data generated by the MCD.
MCD(s)	Refers to the right portion of a truncated point under a multivariate censored normal distribution generated by the MCD, in which s denotes that attributes have been calculated using the simulated data generated by the MCD. The distribution only consists of above-zero values (wet day) that has been back transformed by the inverse of power transformation.
MCD(m,l) or MCD(s, 1)	Equivalent to the MCD(m) or MCD(s) except the MCD does not involve the use of the periodic function.
Notes:	Annotation is not only restricted to MCD, but also applicable to the WG-MCD and WG-MACP models to be discussed in the following section.

VALIDATION OF THE MCD

After the parameters of MCD have been estimated, a string of say 1,000 normally distributed numbers can be generated. The negative values can be discarded and positive values are retained. The positive values can be back transformed (through the inverse of power transformation). The distribution obtained from the above simulation route is designated as MCD(m,1). Table 2 gives the notation conventions for MCD(.). The mean and variance of this simulated data can be estimated from the back transformed positive values.

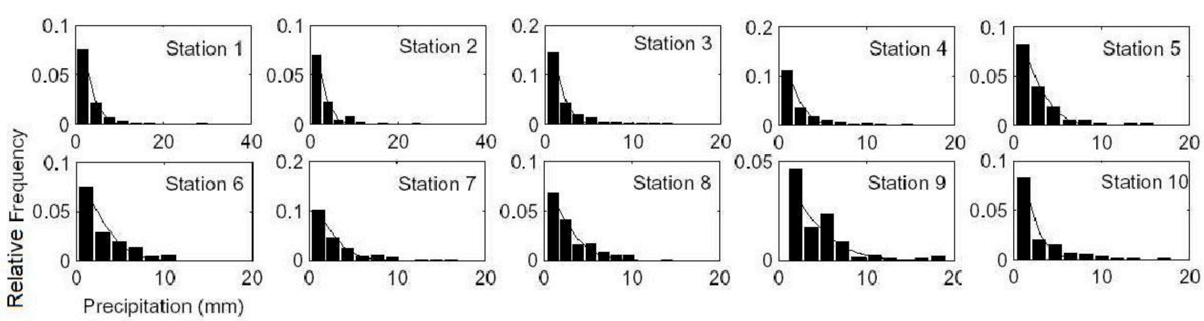

Figure 1: Observed and estimated probability distributions of MCD for the month of January.

For a good simulation, the mean and variance of the observed and simulated data should show 1:1 correspondence. The MCD seems to provide satisfactory results as the observed and simulated parameters are in close proximity of each other. Another example is illustrated in

Figure 1 using daily precipitation data for all stations for the month of January. The distribution of actual precipitation is displayed through histograms, whereas the simulated counterparts are shown by the dashed curve generated by the MCD(m,1). The close correspondence between observed histograms and simulated curves is a good evidence of the capability of the MCD to simulate the frequency distribution of the daily precipitation data.

Table 3. Results summary of two-sample Kolmogorov-Smirnov test. Level of Significance (*P*-values)

	Station									
Month	1	2	3	4	5	6	7	8	9	10
Jan	0.37	0.23	0.34	0.18	0.32	0.27	0.37	0.48	0.15	0.17
Feb	0.24	0.32	0.22	0.52	0.53	0.20	0.15	0.37	0.12	0.28
Mar	0.23	0.49	0.36	0.31	0.52	0.30	0.41	0.31	0.25	0.40
Apr	0.35	0.70	0.46	0.58	0.65	0.54	0.64	0.60	0.23	0.47
May	0.19	0.52	0.21	0.40	0.67	0.54	0.50	0.44	0.42	0.52
Jun	0.51	0.60	0.59	0.52	0.55	0.58	0.49	0.44	0.51	0.35
Jul	0.44	0.49	0.57	0.46	0.50	0.40	0.48	0.42	0.34	0.46
Aug	0.42	0.11	0.39	0.23	0.41	0.50	0.21	0.43	0.35	0.53
Sep	0.47	0.60	0.46	0.40	0.22	0.32	0.35	0.55	0.18	0.06
Oct	0.39	0.42	0.36	0.31	0.30	0.39	0.50	0.42	0.28	0.29
Nov	0.23	0.36	0.31	0.35	0.40	0.22	0.31	0.43	0.26	0.30
Dec	0.24	0.42	0.13	0.34	0.44	0.08	0.37	0.34	0.11	0.33

Table 4. Results summary of Mann-Whitney U-test. Level of Significance (*P*-values)

	Station 5									
Month	1	2	3	4	5	6	7	8	9	10
Jan	0.53	0.54	0.58	0.48	0.44	0.50	0.54	0.42	0.54	0.34
Feb	0.13	0.48	0.15	0.51	0.63	0.56	0.13	0.54	0.15	0.27
Mar	0.33	0.49	0.48	0.59	0.59	0.52	0.42	0.39	0.40	0.56
Apr	0.51	0.60	0.56	0.50	0.56	0.47	0.60	0.59	0.15	0.57
May	0.32	0.55	0.40	0.39	0.58	0.58	0.51	0.53	0.34	0.59
Jun	0.55	0.55	0.60	0.55	0.58	0.55	0.57	0.51	0.58	0.44
Jul	0.53	0.46	0.49	0.41	0.54	0.53	0.40	0.51	0.50	0.57
Aug	0.50	0.30	0.55	0.41	0.46	0.47	0.45	0.55	0.46	0.51
Sep	0.58	0.58	0.57	0.33	0.30	0.20	0.24	0.52	0.24	0.21
Oct	0.44	0.57	0.54	0.23	0.42	0.31	0.55	0.33	0.39	0.42
Nov	0.54	0.30	0.45	0.55	0.52	0.55	0.54	0.56	0.34	0.37
Dec	0.25	0.46	0.09	0.56	0.49	0.09	0.51	0.46	0.15	0.57

For the K-S and M-W based null hypothesis tests, when the *P*-value is less than 5%, the null hypothesis is rejected in this case study. The tests were performed on the daily data within a specific month and at a specific site, corresponding to MCD simulations and historical observations. The calculated *P*-values of the K-S and M-W tests are respectively reported in Tables 3 and 4. Each *P*-value reported in the tables is essentially the average of the *P*-values corresponding to the number of days in a specific months at a specific location. Each of these values is calculated based on the empirical distribution functions of observations and simulations