

# Toward a stochastic precipitation generator conditioned on ENSO phase for eastern Australia

**Duc-Anh An-Vo<sup>a,b</sup>, Roger Stone<sup>b</sup>, Duc Ngo-Cong<sup>a</sup> and Torben Marcussen<sup>b</sup>**

<sup>a</sup> *University of Southern Queensland, Computational Engineering and Science Research Centre, Toowoomba, QLD 4350, Australia*

<sup>b</sup> *University of Southern Queensland, International Centre for Applied Climate Sciences, Toowoomba, QLD 4350, Australia*

*Email: [duc-anh.an-vo@usq.edu.au](mailto:duc-anh.an-vo@usq.edu.au)*

**Abstract:** Stochastic generation of the required daily precipitation data offers an attractive alternative to the use of observed historical records. Stochastic precipitation generators are typically built on the statistical structure of historical data and thus can produce synthetic daily rainfall series with statistical characteristics similar to those of observed series. Parameters of precipitation generator have been typically estimated using all historical daily data for a given period. This approach, however, fails to capture signals in the precipitation process associated with an El Niño-Southern Oscillation (ENSO) phenomenon. ENSO signals have long been known to influence the precipitation in eastern Australia with high rainfall in a cold (La Niña) phase and low rainfall in a hot (El Niño) phase. Here, models for daily rainfall occurrence and intensity conditioned on each ENSO phase were developed to acknowledge ENSO signals in the precipitation process of eastern Australia. The developed models can be used to construct a stochastic precipitation generator for eastern Australia.

We parameterised first-order two-state Markov chains for occurrence process and gamma distributions for intensity process in each month, using recorded data of all historical years (primary models) or recorded data for years of each ENSO phase (conditional models). The Akaike information criterion (AIC) was used to select the “best” occurrence and intensity models among a range of parameterisation schemes for 3 typical locations in eastern Australia, an important agricultural region with a clear ENSO precipitation signal in July – December. Relative performance of the conditional models compared to the primary models was demonstrated by graphic diagnostics of lengths of dry spells for occurrence process and daily precipitation amounts for intensity process.

AIC values of conditional precipitation models (occurrence, intensity, or both) were significantly smaller than those of primary models in all of 18 location-month combinations, indicating superior performance of the conditional models. Graphic diagnostics showed that conditional occurrence models successfully captured differences in the number and persistence of dry days (dry spell) among ENSO phases. Similarly, conditional intensity models noticeably improved the agreements between theoretical and empirical distributions of daily rainfall amounts. Precipitation generators based on the conditional precipitation models can be linked to other process models (e.g. crop model) to derive realistic assessments of the likely consequences of ENSO-related variability of agricultural production in eastern Australia. Conditional stochastic precipitation generators, therefore, can be useful tools to translate ENSO forecasts into likely regional impacts on sectors of interest.

**Keywords:** *Precipitation models, Markov chain, weather generators, ENSO, climate variability, Australia*

## 1. INTRODUCTION

Many interactions between crops and weather are nonlinear, and simple statistical approaches usually fail to assess the effect of climate variability on agricultural production risks. Process-level crop simulation models provide alternative tools to assess climate impacts on agriculture and to evaluate the outcomes of a wide range of management decision alternatives (An-Vo et al., 2015a&b; Mushtaq et al., 2017). These models typically require inclusion of daily weather data. Although observed historical weather sequences can be used, they fundamentally represent only one realisation of the weather process (Richardson, 1981). A thorough risk assessment should explore the range of outcomes of the process of interest by using equally likely weather sequences, with the same stochastic structure as the observed series. Stochastic precipitation generators offer a solution to this problem by providing alternative daily precipitation sequences required by a thorough risk assessment. It is also noteworthy that though weather generators create weather data having the same stochastic structures as a historically observed base period we can now employ climate change projection in a weather generator to simulate future climate scenarios (e.g. Fatichi et al., 2011) for climate change impact studies.

Unconditional models of precipitation generators, however, produced distributions of monthly total precipitation in which moderate amounts were over-represented and more extreme amounts were under-represented; i.e. dispersion was underestimated (Katz & Parlange, 1998). Inspired by the conditional precipitation generator developed by Grondona et al. (2000) for a case study in south-eastern South America, the objective of this study is to investigate parameterisation of stochastic weather generators conditioned on ENSO phase for eastern Australia where precipitation regimes are influenced by ENSO events (Nguyen-Huy et al., 2017). As an initial work on this topic for eastern Australia, we modelled the precipitation process by using a common parametric approach, i.e. first order Markov chain for the occurrence process and gamma distribution for the intensity process. Such model requires only a relatively small number of parameters, while they still account for the most important statistical properties of precipitation series. Interested readers can refer to e.g. Verdin et al. (2015) for comprehensive conditional stochastic precipitation generators and e.g. Bennett et al. (2016) for parsimonious daily rainfall field models. Here, model parameters are estimated separately for warm and cold ENSO events, and neutral years. The Akaike information criterion (AIC; Akaike, 1974) was used to select the “best” occurrence and intensity models. To assess whether the primary model should be rejected in favour of a conditional model, we assessed statistical significance of the difference in log-likelihoods by a  $\chi^2$  statistic and its associated  $p$  value. In addition, various graphical diagnostics were employed for occurrence and intensity processes to assess whether the conditional models offer an improvement over simpler alternatives. The hypothesis is that models conditioned on ENSO should capture differences in the precipitation process among ENSO phases, thus providing a more realistic assessment of agricultural risks associated with ENSO-related climate variability.

## 2. METHODOLOGY

The generation of precipitation is based on models for two processes: (i) the occurrence process (i.e., the sequence of “dry” or “wet” days) model; and (ii) the intensity process (i.e., the sequence of precipitation amounts on wet days) model. A wet day is defined as one on which a measurable amount of precipitation (in this study)  $\geq 0.1$  mm occurs.

### 2.1 Occurrence process

In a first-order Markov chain process, the probability of a wet (dry) state on day  $t$  depends only on the precipitation state on day  $t-1$ . Here  $J_t$  is defined as the state indicator random variable for day  $t$ :  $J_t = 0$  if day  $t$  is dry;  $J_t = 1$  for a wet day. This model is characterized by four transition probabilities:

$$P_{ij} = \Pr\{J_t = j | J_{t-1} = i\}, \quad i, j = 0, 1. \quad (1)$$

As transition probabilities  $P_{00} + P_{01} = P_{10} + P_{11} = 1$ , only two parameters are needed to define the first-order, two-state Markov process.

An alternative formulation of this model is in terms of parameters  $d$ , the first-order autocorrelation coefficient, also referred to as the “persistence” parameter, and  $\pi$ , the probability of a wet day (Katz & Parlange, 1993). These parameters are related to the transition probabilities and to one another as (Grondona et al., 2000)

$$d \equiv \text{corr}(J_{t-1}, J_t) = P_{11} - P_{01} \quad (2)$$

$$\pi = \Pr\{J_t = 1\} = \frac{P_{01}}{P_{01} + P_{10}} = \frac{P_{01}}{1 - d} \quad (3)$$

## 2.2 Intensity process

Proposed statistical distributions to model the intensity process i.e. the distribution of rainfall amounts on wet days are typically the log-normal, cubic root normal, mixed exponential, kappa, gamma, and Weibull distributions. This work adopts the frequently used gamma distribution, given by the expression

$$f(y_t | J_t = 1) = \frac{(y_t / \beta)^{\alpha-1} \exp(-y_t / \beta)}{\beta \Gamma(\alpha)} \quad (4)$$

For  $y_t > 0$ ,  $\alpha > 0$  and  $\beta > 0$ , where  $y_t$  denotes the rainfall amount on wet day  $t$ . This distribution is characterised by  $\alpha$ , the “shape” parameter, and  $\beta$ , the “scale” parameter. The mean and variance of the daily amount of rain (on wet days) are

$$\mu = E(Y_t | J_t = 1) = \alpha\beta \quad (5)$$

$$\sigma^2 = \text{var}(Y_t | J_t = 1) = \alpha\beta^2 \quad (6)$$

To obtain parsimonious models, common assumptions are that daily rainfall is independent of the occurrence process and is also independent of rainfall amounts in previous days. Katz and Parlange (1998) explored model extensions that do not require these assumptions.

## 2.3 Parameterisation

To deal with seasonal variation in precipitation we estimated model parameters using daily rainfall data of each month for a considered period (i.e. July – December in this study). This approach thus implicitly assumes stationary, not only within a month but also between years. Low-frequency or inter-annual components such as ENSO circulation or long-term climate trends, however, violate this assumption. As a result, inter-annual changes in the distribution of precipitation should be expected. To acknowledge ENSO-related changes we estimated model parameters conditioned on ENSO phases.

Fitting all model parameters to data of each ENSO phase (full conditional models) would imply that ENSO influences all parameters of the rainfall model, and there is no common structure among phases. An unconditional model (hereafter, “primary”) with a single set of fitting parameters for data of all ENSO phases, however, implicitly indicates no ENSO signal on rainfall distribution. A balance between the two extremes is to assume some common model parameters for all ENSO phases (partial conditional models).

## 2.4 Model selection

The Akaike information criterion (AIC; Akaike, 1974) was used to select the “best” occurrence and intensity models. Briefly, the AIC is the likelihood ratio statistic penalized by twice the degrees of freedom (Wilks, 2011) which allows a balance between model improvement and model complexity, i.e. parsimony of model development. The best model will be the one minimising the AIC.

In each month, four alternative models were explored for both the occurrence and intensity processes: one primary or unconditional model (two parameters per process); two partial conditional models (four parameters per process); and a full conditional model (six parameters per process). For each process, the model considered best among the four alternative models was the one for which the AIC was minimised. AIC values for the four alternative models were calculated for each station, month, and process. We compared the AIC values among the four models to select the one with lowest AIC value as the best. To assess whether the primary model should be rejected in favour of a conditional model, a statistical significance of the difference in log-likelihoods can be assessed knowing that the null distribution of

$$\Lambda = 2(L_c - L) \quad (7)$$

is  $\chi^2$  with  $\nu = k_c - k$  degrees of freedom; where  $k_c$  is the number of parameters in conditional models and  $k$  is the number of parameters in primary models;  $L_c$  and  $L$  are log-likelihoods of the conditional models and primary models, respectively. We thus can have

$$\Lambda = 2 \left( \frac{2 * k_c - AIC_c}{2} - \frac{2 * k - AIC}{2} \right) = (AIC - 2 * k) - (AIC_c - 2 * k_c) \tag{8}$$

where  $AIC_c$  and  $AIC$  denote values of Akaike information criterion, respectively, for conditional and primary models.

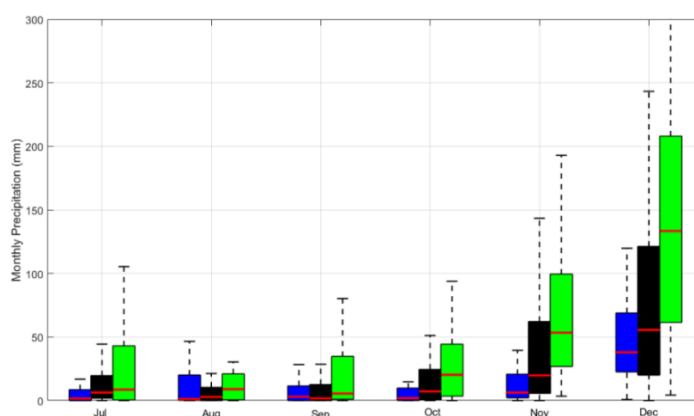
## 2.5 Data

Daily precipitation data was obtained from (<http://www.longpaddock.qld.gov.au/silo>) for 3 stations at Tully, Ayr in Queensland and Bendigo in Victoria for the 1889 – 2015 period. The stations analysed are in areas of agricultural importance. Among the three stations, two stations (Tully and Ayr) are in summer rainfall dominant zone while Bendigo is in winter rainfall dominant zone. The chosen three stations thus cover the two seasonal rainfall patterns of eastern Australia. Moreover, Tully and Ayr are sugarcane growing regions and sugarcane is one of the most important crops in Queensland; Bendigo is a wheat growing region and wheat is a dominant crop in Victoria grain industry. Though Tully and Ayr are both in north-eastern Queensland and in summer rainfall dominant zone Tully has significantly higher annual rainfall than that of Ayr. The sugarcane farming in Tully thus is rain-fed while fully irrigation is required for a viable sugarcane farming in Ayr.

The ENSO signal on rainfall in Australia is most marked in the northern and eastern regions and during winter and spring. This is the period, thus, in which potential improvements associated with ENSO-conditional parameterisation are most likely to be detected. For this reason, and for the sake of brevity, subsequent analyses focus on the period from July to December. An ENSO phase (warm, neutral, or cold) is first assigned to each analysed month (July – December) in the historical precipitation records. There are several alternative definitions of ENSO events (Trenberth, 1997). Here, events were categorized according to values of the Southern Oscillation Index (SOI) between June and March. An original classification proposed by Dr Rob Allan has been modified to account for late-forming El Niño or La Niña events. Threshold values of the SOI have been adjusted such that the frequency of El Niño or La Niña years from 1950-51 to 2009-10 is similar to that obtained by the “WMO RA IV Consensus Index and Definitions of El Niño and La Niña” (<http://www.noaanews.noaa.gov/stories2005/s2428.html>). An “El Niño year” is indicated if the six-month average value of the SOI, ending in any month between November and March, was below a threshold value of negative 6. A “La Niña year” is indicated if the six-month average value of the SOI, ending in any month between November and March, was above a threshold value of positive 6. Lists of El Niño and La Niña events are given in Table 1.

**Table 1.** El Niño and La Niña events between 1889 and 2016, as defined by the SOI index. Note that here an “ENSO year” encompasses the period between June of the listed year and May of the following year. For example, the El Niño event of 1982 includes the period between June 1982 and May 1983. Years not listed are considered as Neutral.

ENSO phase	Years
El Niño	1896 1900 1902 1904 1905 1911 1913 1914 1918 1919 1923 1925 1939 1940 1941 1946 1951 1953 1957 1958 1963 1965 1969 1972 1977 1982 1986 1987 1991 1992 1993 1994 1997 2002 2004 2006 2009 2014 2015
La Niña	1889 1892 1893 1897 1903 1906 1909 1910 1916 1917 1921 1922 1924 1928 1929 1938 1942 1947 1949 1950 1954 1955 1956 1961 1964 1970 1971 1973 1974 1975 1988 1996 1998 1999 2000 2007 2008 2010 2011



**Figure 1.** Box plots of monthly precipitation (July – December) by ENSO phase for Ayr, Queensland. The red line in each box indicates the median. Blue boxes correspond to the El Niño years, black boxes the Neutral years and green boxes the La Niña years. The box bounds the central half of the data, thus its width corresponds to the interquartile range (IQR). The whiskers extend to the most extreme data points being not outliers. Points are considered as outliers if they are larger than  $Q3+1.5*(Q3-Q1)$  or smaller than  $Q1-1.5*(Q3-Q1)$ , where  $Q1$  and  $Q3$  are the 25th and 75th percentiles, respectively.

### 3. CONDITIONAL MODELS OUTPERFORM THE PRIMARY MODELS

#### 3.1 ENSO signal on precipitation

ENSO effects on the precipitation process are typically represented by boxplots of monthly rainfall totals by ENSO phase for Ayr (Figure 1). The central tendency of monthly precipitation at Ayr has a clearly increasing trend from lowest values in the warm phase, to moderate values in the neutral phase, and largest values in the cold phase for the considered period excepting September in which median precipitation of the neutral phase is smaller than that of the warm phase (Figure 1). Rainfall distributions are narrow during warm events while much wider distributions can be seen in cold events.

The AIC values for primary models (considered as a benchmark) and the best models are presented in Table 2. The  $\chi^2$  statistic and associated  $p$  value are also listed for each station, month, and process. Remarkably, conditional precipitation models (occurrence, intensity, or both) were identified as an improvement over primary models in all of the 18 location-month combinations analysed. The  $\chi^2$  statistic shows small  $p$  values for all of the 18 location-month combinations with 13 combinations (72 %) having  $p$  values less than  $10^{-3}$ . These results suggest a very significant ENSO signature on precipitation at the 3 locations tested which also give a clear implication for eastern Australia, though results at more locations are needed for a regionally consistent evidence. When compared with ENSO signal identified in analysed sites in southeastern South America (Grondona et al., 2000), more significant ENSO signal was identified in the present selected sites which has implication for eastern Australia.

**Table 2.** AIC values for primary and “best” occurrence and intensity models. “Cond.” stands for partial conditional models and “Full” for full conditional models. For instance, “cond.  $\pi$ ” refers to a partial conditional intensity model with probability of a wet day  $\pi$  being conditional on the ENSO phase and a common persistence parameter  $d$ .  $\chi^2$  statistic and associated  $p$  values are also shown to provide a guide on whether primary models should be rejected in favour of conditional models.

Month	Occurrence					Intensity				
	AIC primary	AIC best	Best model	$\chi^2$	$p$	AIC primary	AIC best	Best model	$\chi^2$	$p$
<b>Tully</b>										
Jul	4048	—	Primary	—	—	10263	10257	Full	13.76	0.008
Aug	3952	3945	Full	14.77	0.005	9372	9318	Full	61.62	$<10^{-3}$
Sep	3673	3667	Cond. $\pi$	9.73	0.008	7903	7897	Full	13.68	0.008
Oct	4013	3998	Cond. $\pi$	18.93	$<10^{-3}$	8254	8122	Full	139.84	$<10^{-3}$
Nov	4273	4253	Cond. $\pi$	23.95	$<10^{-3}$	10156	10153	Cond. $\beta$	7.43	0.024
Dec	4654	4622	Cond. $\pi$	35.73	$<10^{-3}$	13353	13193	Full	167.7	$<10^{-3}$
<b>Ayr</b>										
Jul	2285	2274	Full	18.96	$<10^{-3}$	2285	2193	Full	100.05	$<10^{-3}$
Aug	1937	1930	Cond. $\pi$	11.28	0.004	1793	—	Primary	—	—
Sep	1875	1871	Full	11.97	0.018	2032	1938	Full	101.99	$<10^{-3}$
Oct	2330	2305	Cond. $\pi$	28.69	$<10^{-3}$	2390	2364	Full	34.01	$<10^{-3}$
Nov	3175	3147	Cond. $\pi$	31.94	$<10^{-3}$	4198	4159	Full	46.47	$<10^{-3}$
Dec	3956	3932	Full	31.5	$<10^{-3}$	7265	7220	Full	52.93	$<10^{-3}$
<b>Bendigo</b>										
Jul	5085	—	Primary	—	—	9444	9438	Full	14.21	0.007
Aug	5090	5087	Cond. $\pi$	6.96	0.031	9286	9253	Full	40.99	$<10^{-3}$
Sep	4835	4832	Cond. $\pi$	6.85	0.033	8125	8087	Full	46.63	$<10^{-3}$
Oct	4822	4794	Full	35.76	$<10^{-3}$	7134	7078	Full	63.14	$<10^{-3}$
Nov	4201	—	Primary	—	—	5306	5305	Cond. $\alpha$	5.02	0.081
Dec	3885	3885	Cond. $\pi$	4.29	$<10^{-3}$	4605	—	Primary	—	—

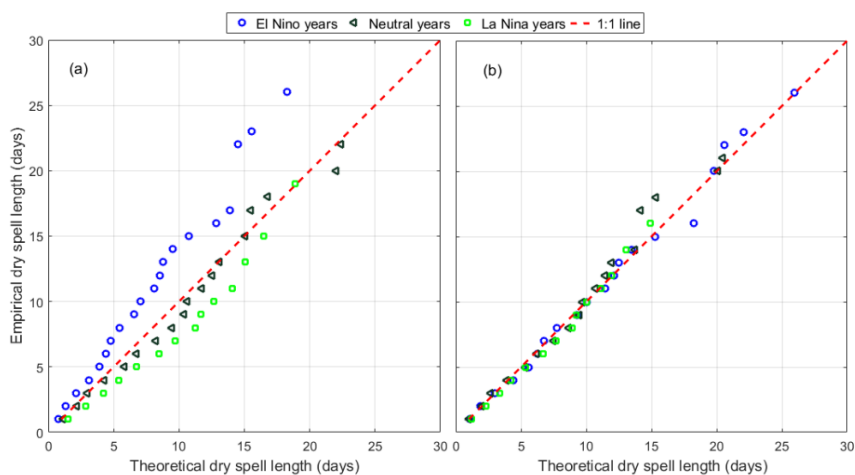
### 3.2 Diagnostics

Two diagnostics are presented to assess the performance of ENSO-conditional precipitation generators from the point of view of practical applications. It means the parsimony of model development is required where potential improvements over simple models should be balanced against increased model complexity. For brevity, only typical results are presented here.

#### *The occurrence process*

We used the distribution of persistently consecutive dry days (dry spell) to diagnose the performance of primary and conditional occurrence models because of its importance for drought related risk assessment which is a major concern for agricultural production in eastern Australia. The probability of observing a dry spell longer than  $k$  days is  $P(K > k) = P_{00}^k$  (Grondona et al., 2000).

The theoretical and empirical distributions of dry spell lengths longer than  $k$  days are compared for each of the alternative theoretical model by using quantile-quantile (Q-Q) plots. Fig. 2a shows the Q-Q plot for the primary occurrence model. For cold (warm) ENSO events, most points fall below (above) the 1:1 line, indicating that the primary occurrence model overestimates (underestimates) the probability of observing dry spells longer than  $k$  days. When an occurrence model with conditional  $\pi$  is used (Fig. 2b), points for all ENSO phases lie much closer to the 1:1 line meaning that the theoretical and empirical distributions of dry spell lengths agree much better. This suggests that the conditional models successfully capture differences in the dry spells distribution among ENSO phases.



**Figure 2.** Q-Q plots of theoretical and observed distributions of dry spell lengths longer than  $k$  days for October in Bendigo. Points are plotted separately for each ENSO phase: (a) simple occurrence model, (b) conditional occurrence model (conditional  $\pi$ ).

#### *The intensity process*

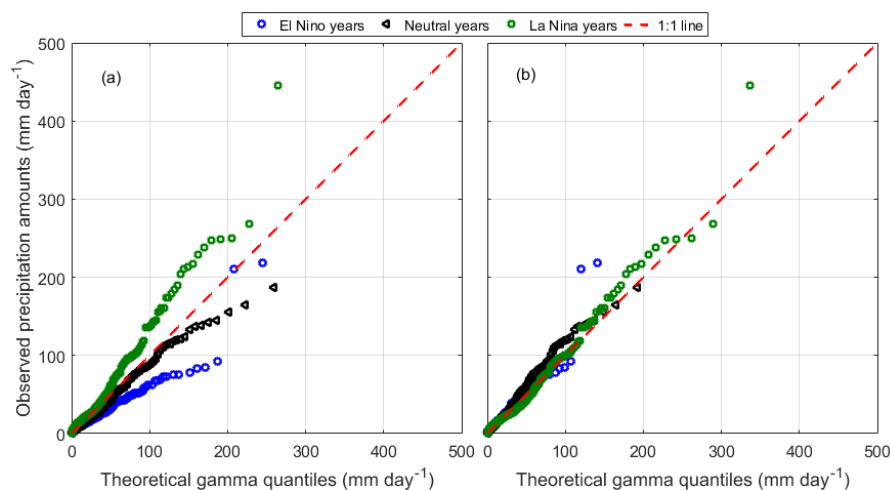
In this case, empirical quantiles of daily precipitation amounts for a given month (computed separately for each ENSO phase) are plotted against theoretical quantiles (derived from gamma distribution) for unconditional parameterisation and ENSO-conditional parameterisation (Fig. 3). In Fig. 3a, theoretical quantiles are computed using a primary model. The points for cold (warm) events fall above (below) the 1:1 line, indicating that the primary model underestimates (overestimates) daily precipitation amounts during cold (warm) events. When theoretical quantiles are computed using a partial conditional model (conditional  $\beta$ , Fig. 3b), points for all ENSO phases locate much closer to the 1:1 line, indicating a much better improvement in the fit of the theoretical model to the empirical data.

### 4. CONCLUSION

In this study, we presented a parametric precipitation model conditional on the ENSO phase to acknowledge ENSO-related effects toward generating daily rainfall field for eastern Australia. Application of the parameterisation methodology to historical data in 3 typical locations revealed that conditional precipitation models outperform the unconditional models and are parsimonious by means of the Akaike information criterion. Graphic diagnostics of dry spell lengths for the occurrence process and daily precipitation amounts for the

intensity process demonstrated that the conditional models successfully capture differences in precipitation regimes among ENSO phases in eastern Australia. Conditional models, therefore, can be used to translate ENSO forecasts into likely regional impacts on sectors of interest such as by generating daily rainfall fields for biophysical crop growth models to assess likely crop production outcomes.

The main limitation of the two-state, first-order Markov approach is the accurate simulation of long runs of dry or wet sequences at some locations while the gamma distribution is limited in representing the extreme rainfall events. Works are ongoing to improve such limitations.



**Figure 3.** Q-Q plots of theoretical and observed distributions of daily precipitation amounts on wet days for December in Tully. Points are plotted separately for each ENSO phase: (a) simple intensity model, (b) conditional intensity model (conditional  $\beta$ ).

## ACKNOWLEDGEMENTS

This research was supported by University of Southern Queensland's strategic research fund.

## REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.
- An-Vo, D.-A., Mushtaq, S., Nguyen-Ky, T., Bundschuh, J., Tran-Cong, T., Maraseni, T., & Reardon-Smith, K. (2015). Nonlinear optimisation using production functions to estimate economic benefit of conjunctive water use for multicrop production. *Water Resources Management*, 29(7), 2153-2170.
- An-Vo, D.-A., Mushtaq, S., & Reardon-Smith, K. (2015). Estimating the value of conjunctive water use at a system-level using nonlinear programming model. *Journal of Economic and Social Policy*, 17(2), 9.
- Bennett, B., Thyer, M., Leonard, M., Lambert, M., & Bates, B. (2016). A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model. *Journal of Hydrology* (in press <https://doi.org/10.1016/j.jhydrol.2016.12.043>).
- Faticchi, S., Ivanov, V. Y., & Caporali, E. (2011). Simulation of future climate scenarios with a weather generator. *Advances in Water Resources*, 34(4), 448-467.
- Grondona, M. O., Podestá, G. P., Bidegain, M., Marino, M., & Hordij, H. (2000). A stochastic precipitation generator conditioned on ENSO phase: a case study in southeastern South America. *Journal of Climate*, 13(16), 2973-2986.
- Katz, R. W., & Parlange, M. B. (1993). Effects of an index of atmospheric circulation on stochastic properties of precipitation. *Water Resources Research*, 29(7), 2335-2344.
- Katz, R. W., & Parlange, M. B. (1998). Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate*, 11(4), 591-601.
- Mushtaq, S., An-Vo, D.-A., Christopher, M., Zheng, B., Chenu, K., Chapman, S. C., . . . Alam, G. M. M. (2017). Economic assessment of wheat breeding options for potential improved levels of post head-emergence frost tolerance. *Field Crops Research*, 213, 75-88. doi:<https://doi.org/10.1016/j.fcr.2017.07.021>
- Nguyen-Huy, T., Deo, R. C., An-Vo, D.-A., Mushtaq, S., & Khan, S. (2017). Copula-statistical precipitation forecasting model in Australia's agro-ecological zones. *Agricultural Water Management*, 191, 153-172.
- Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1), 182-190.
- Trenberth, K. E. (1997). The definition of el nino. *Bulletin of the American Meteorological Society*, 78(12), 2771-2777.
- Verdin, A., Rajagopalan, B., Kleiber, W., Podestá, G., & Bert, F. (2015). A conditional stochastic weather generator for seasonal to multi-decadal simulations. *Journal of Hydrology* (in press <https://doi.org/10.1016/j.jhydrol.2015.12.036>).
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100): Academic press.