



# A general regional frequency analysis framework for quantifying local-scale climate effects: A case study of ENSO effects on Southeast Queensland rainfall

X. Sun, M. Thyer, Benjamin Renard, M. Lang

## ► To cite this version:

X. Sun, M. Thyer, Benjamin Renard, M. Lang. A general regional frequency analysis framework for quantifying local-scale climate effects: A case study of ENSO effects on Southeast Queensland rainfall. *Journal of Hydrology*, 2014, 512, p. 53 - p. 68. hal-00997139

**HAL Id: hal-00997139**

**<https://hal.science/hal-00997139>**

Submitted on 27 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**A general regional frequency analysis framework for quantifying local-  
scale climate effects: A case study of ENSO effects on southeast  
Queensland rainfall**

Xun SUN<sup>1</sup>, Mark THYER<sup>2</sup>, Benjamin RENARD<sup>1</sup>, Michel LANG<sup>1</sup>

Correspondence to: Mr. X. Sun, Zhongzhou Rd 68 long 6 hao, 200080, Shanghai, China

Email: [sj133@hotmail.com](mailto:sj133@hotmail.com) Phone: +33 472208928

---

<sup>1</sup> Irstea, UR HHLY, Hydrology-Hydraulics, 5 rue de La Doua, Villeurbanne, F-69626, France

<sup>2</sup> University of Adelaide, School of Civil, Environmental & Mining Engineering, SA5005, Australia

8 *Abstract*

9 There is increasing evidence that the distribution of hydrometeorological variables such as average  
10 or extreme rainfall/runoff is modulated by modes of climate variability in many regions of the  
11 world. This paper presents a general spatio-temporal regional frequency analysis framework that  
12 allows quantifying the effect of climate variability on the distribution of at-site hydrometeorological  
13 variables. Climate effects are described through the parameters of a pre-specified distribution, by  
14 using regression models linking parameter values with time-varying covariates, such as climate  
15 indices. For the regional model copulas are used to incorporate spatial dependency. A Bayesian  
16 framework is used for inference and prediction, which enables quantification of parameter and  
17 predictive uncertainties. A regional approach enables better identification of climate effects which  
18 can be subject to high uncertainty using only at-site (local) analysis. Lastly, model comparison tools  
19 enable considering competing statistical hypotheses on the nature of climate effects and selecting  
20 the most relevant one.

21  
22 This modelling framework is applied to two case studies assessing the effect of El Niño Southern  
23 Oscillation (ENSO) on summer rainfall in Southeast Queensland. The first case study focuses on  
24 summer rainfall totals while the second analysis focuses on extremes using summer daily rainfall  
25 maxima. The Southern Oscillation Index (SOI), a measure of ENSO, is considered as a time-  
26 varying covariate. In order to account for different effects during La Niña and El Niño episodes, an  
27 asymmetric piecewise-linear regression is used to analyse the rainfall data using both local and  
28 regional models. During La Niña episodes, SOI has a significant effect on both summer rainfall  
29 totals and maxima. Conversely, during El Niño episodes, the SOI has little effect on rainfall. It is  
30 found that, during a strong La Niña, the most likely 1 in 100 year summer maximum daily rainfall  
31 for different sites estimated with the local asymmetric model can be 5% to 33% higher than the

estimates from a local symmetric linear model and 20% to 50% higher than the estimates from a stationary model, albeit with significant uncertainty. Results from regional and local models are also compared: the former shows a great advantage in terms of uncertainty reduction and allows a better quantification of the ENSO effect on summer rainfall totals and maxima.

**Keywords:** regional frequency analysis, spatio-temporal extremes, climate-informed model, ENSO, precipitation, Australia

## 1 INTRODUCTION

Extreme precipitations and their consequences (floods) are one of the most threatening natural disasters for human beings. In engineering design, Frequency Analysis (FA) techniques are an integral part of risk assessment and mitigation. FA uses statistical models to estimate the probability of extreme hydrometeorological events which provides information for designing hydraulic structures. However, standard FA methods commonly rely on the assumption of ‘identical distribution’ (Brockwell and Davis, 2006): the distribution of observations does not vary with time. As will be reviewed subsequently, there is now a substantial body of evidence that large-scale modes of climate variability (e.g. El-Niño Southern Oscillation (ENSO); North Atlantic oscillation (NAO); Indian Ocean Dipole (IOD); Pacific Decadal Oscillation (PDO); etc.) exert a significant influence on rainfall in various regions worldwide (e.g. Gershunov and Cayan (2003); Haylock et al. (2006); Henley et al. (2011); Kamruzzaman et al. (2013); Schreck and Semazzi (2004); Willems (2013a); (2013b)). Furthermore, climate change is likely to have an influence on hydrology, thus further challenging the assumption of stationarity (Milly et al., 2008). Therefore, FA techniques need to move beyond this assumption. In order to provide a more accurate risk assessment, it is important to understand and predict the effect of climate variability/change on the severity and

55 frequency of hydrometeorological events (especially extremes). This paper provides an important  
56 step towards this goal, by developing a rigorous regional frequency analysis (RFA) framework for  
57 incorporating the effects of climate variability on hydrometeorological events.

58

59 Climate variability influences hydrology worldwide. ENSO is one of the prominent modes of  
60 climate variability and has global impact on hydrometeorological variables (Hoerling et al., 1997).  
61 For example, during winter (summer) season in northern (southern) hemisphere, during El Niño  
62 phase, positive anomalies were found in Southwest U.S (Castello and Shelton, 2004; Cayan et al.,  
63 1999; Gershunov and Barnett, 1998; Meehl et al., 2007), Southern South America (Grimm and  
64 Tedeschi, 2009) and Southern China (Wu et al., 2003); while during La Niña phase, positive  
65 anomalies were found in Northwest U.S (Castello and Shelton, 2004; Cayan et al., 1999; Gershunov  
66 and Barnett, 1998; Meehl et al., 2007), South Africa (Kruger, 1999; Vanheerden et al., 1988) and  
67 Southeast Queensland, Australia (Cai et al., 2010).

68

69 In order to overcome the assumption of identical distribution and enable the inclusion of climate  
70 information, innovative FA methods have been recently developed. At the local scale (i.e. for a  
71 single site), Renard et al. (2006b) built a non-stationary FA model by estimating time-varying  
72 parameters from a pre-specified distribution. Micevski et al. (2006) used the Inter-decadal Pacific  
73 Oscillation (IPO) to characterize the flood hazard. Ouarda and El-Adlouni (2011) discussed non-  
74 stationary FA models within the Bayesian approach. More generally, Khaliq et al. (2006) reviewed  
75 non-stationary local FA methods. While local FA methods enabling the inclusion of climate  
76 information or non-stationarity are becoming common, such at-site models remain limited by two  
77 important drawbacks:

78 1) Local analysis cannot be applied to ungauged sites.

79 2) Uncertainty in parameter estimates (and hence predictive estimates) tends to be very large  
80 due to the limited number of observations at one location. In addition, if climate information is  
81 included and more complex models are proposed, these observations may not be enough to identify  
82 the parameters (Thyer et al., 2006).

83  
84 This motivates the development of regional frequency analysis (RFA) models that use information  
85 from multiple sites to overcome these shortcomings. Many RFA methodologies have been  
86 developed over the years. Cooley et al. (2007) and Ghosh and Mallick (2011) used Bayesian spatial  
87 models for extreme precipitation, but still under the identical distribution assumption. Cunderlik and  
88 Burn (2003) described a second-order non-stationary approach to pooled flood FA, and Hanel et al.  
89 (2009) introduced a non-stationary index-flood model for extreme precipitation. Recently, several  
90 authors (Aryal et al., 2009; Lima and Lall, 2010; Sang and Gelfand, 2009) started investigating  
91 spatio-temporal models using hierarchical approaches. In the same vein, Gregersen et al. (2013)  
92 also used Poisson regression models to describe the frequency of extreme rainfall in both space and  
93 time. A common difficulty for all these approaches is the treatment of the spatial dependency  
94 existing between data.

95  
96 The main contribution of this paper is the construction of a rigorous spatio-temporal framework that  
97 enables the quantification of the effect of climate variability on the severity/frequency of  
98 hydrometeorological events. This framework builds on the previous works referenced in this  
99 introduction section: it brings together several components separately developed in previous studies  
100 (in particular spatio-temporal regression models, copula-based modeling of spatial dependence,  
101 Bayesian inference, model comparison tools) to derive a general and flexible modeling platform.  
102 This paper has two main objectives:

103 1) Model Development, Inference and Comparison: the construction of the model, using  
104 regressions with spatial and temporal covariates to describe the spatio-temporal variability of the  
105 parameters, is described. Inference accounts for spatial dependence between data and uses a  
106 Bayesian framework, thereby enabling a direct quantification of estimation uncertainty and  
107 predictive uncertainty. In addition, within this general framework, different climate-informed  
108 regression models can be compared (for instance, linear vs. non-linear regression). This helps  
109 identifying the most suitable regression to link climate variability and spatio-temporal  
110 hydrometeorological variability.

111 2) Model Application: Two case studies illustrate the application of the framework to quantify  
112 the ENSO effect on the summer total and extreme rainfall in Southeast Queensland (SEQ),  
113 Australia. The flexibility of the framework enables several competing statistical hypotheses to be  
114 rigorously compared, e.g. to assess whether the effect of ENSO on summer maximum daily rainfall  
115 is symmetric or asymmetric.

116  
117 Similar studies on the summer total and maximum rainfall over SEQ were described by Cai et al.  
118 (2010) and King et al. (2013). In particular, these authors highlighted the existence of an  
119 asymmetric relationship between ENSO and precipitation in Eastern Australia. With reference to  
120 Cai et al. (2010) a simplified physical interpretation for this asymmetric relationship, is as follows:  
121 During La Nina events the warm pool of Pacific Ocean sea surface temperature anomaly's (SSTA)  
122 moves westward, closer to the Eastern Australia coast, producing higher rainfall in SEQ. The  
123 stronger the La Nina event, the further the warm pool moves west and the higher the rainfall in  
124 SEQ. During El Nino events, the warm pool of Pacific SSTA's moves east, producing reduced  
125 rainfall in SEQ. Once this warm pool moves sufficiently east away from the coast of Australia, the  
126 intensity of the El Nino event is irrelevant. Thus while an El Nino event produces reduced rainfall,  
127 the intensity of the El Nino event does not impact on rainfall in SEQ.

128 In this paper, we do not attempt to discuss or discover new physical understanding for the reasons  
129 behind this relationship. Instead, we focus on developing a framework that can provide local multi-  
130 site predictions of this effect of ENSO on rainfall and rainfall extremes, which is of prime  
131 importance for engineering design and operations. More precisely, we aim at addressing the  
132 following shortcomings of previous studies:

133 1) The aforementioned studies are based on precipitation totals or maxima spatially averaged  
134 over a large region (South-East Queensland or Eastern Australia). From an engineering perspective  
135 this type of analysis is inadequate. Practical applications require estimates of the ENSO effects on  
136 rainfall at individual sites, rather than spatially averaged regions. This framework aims to provide  
137 predictions of multi-site local scale climate effects.

138 2) The use of simple statistical approaches such as least-square linear regression may suffice to  
139 estimate an overall ENSO effect over a given region, but is not sufficient for engineering  
140 applications. Indeed, such applications require predicting the full distribution of the target  
141 precipitation variable conditionally on the ENSO state, which in turn allows estimation of extreme  
142 rainfall conditioned on ENSO.

143

144 As such, this paper aims to complement and build on previous studies that have demonstrate ENSO  
145 effects using spatially averaged data, by providing multi-site local predictions of hydro-over spatial  
146 regions suitable for practical engineering applications.

147

148 The paper is organized as follows. Section 2 describes the framework for building general climate-  
149 informed models at both local and regional levels. Section 3 and Section 4 present the case study  
150 with ENSO effect on summer rainfall in Southeast Queensland. Different regression models are  
151 compared in this section as well. Further improvements are discussed in Section 5, while the  
152 conclusion in Section 6 summarizes the main findings of this study.



## 153 2 GENERAL SPATIO-TEMPORAL REGIONAL FREQUENCY ANALYSIS 154 FRAMEWORK

155 The general frequency analysis framework aims to take advantage of spatial and temporal  
156 information (e.g. climate information in ENSO) to enhance the predictions of the  
157 hydrometeorological variable of interest. In this framework, data are assumed to follow a  
158 distribution, whose parameters are linked to temporal or/and spatial covariates using regression  
159 models. In the first part, an at-site (local) model that uses temporal covariates is described. In the  
160 second part, the framework is generalized to a regional scale. Since data at different sites are used,  
161 both spatial and temporal covariates are involved. The spatial dependency of the data is considered  
162 with elliptical copulas. The last part of this section describes criteria for model selection. The  
163 RFortran software library (Thyer et al., 2011), was used as part of this framework to enable fast and  
164 efficient debugging, diagnosis and analysis of results.

### 165 2.1 At-site (local) model with temporal covariates

#### 166 2.1.1 Parent distribution

167 The basis of the at-site (local) model is to assume that data follows a time-varying distribution,  
168 conditioned on temporally varying covariates, such as climate information. More precisely, let  
169  $Y(t)$  be the observation at time  $t$  and  $\mathbf{Y} = (Y(t_1), Y(t_2), \dots, Y(t_n))$  be the collection of observations of  
170 a site at  $n$  time steps. A local model is defined as:

$$171 \quad Y(t) \sim D(\boldsymbol{\beta}(t)) \quad (1)$$

172 where  $D$  is the assumed distribution of  $Y$  and  $\boldsymbol{\beta}(t) = (\beta_1(t), \beta_2(t), \dots, \beta_m(t))$  is the collection of  
173 all  $m$  distribution parameters at time  $t$  ( $m=2$  for a Gaussian distribution,  $m=3$  for a generalized  
174 extreme value distribution (*GEV*), etc.).

## 175 2.1.2 Regression models with temporal covariates

176 The parameters  $\beta$  directly characterize the distribution, such as its location, scale and shape. These  
177 parameters may depend on different covariates, like time, pressure and some climate indices. Thus a  
178 regression function is defined for each distribution parameter as follows:

$$179 \quad \beta_i(t) = h_i(\mathbf{x}(t); \theta_i) \quad i = \{1, 2, \dots, m\} \quad (2)$$

180 where  $h_i$  is the regression function of  $\beta_i(t)$ ,  $\mathbf{x}(t)$  is the collection of temporal covariates and  $\theta_i$  is  
181 the collection of all parameters used in the regression function  $h_i$ .

182 To avoid confusion with the  $D$ -parameters  $\beta(t)$ , we call  $\theta = (\theta_1, \dots, \theta_m)$ , the parameters we are  
183 going to estimate, the regression parameters. Figure 1 illustrates the construction of the local model.

## 184 2.1.3 Parameter estimation

185 With a pre-specified distribution  $D$  and the regression functions  $h$ , regression parameters  $\theta$  are  
186 estimated in a Bayesian framework. The posterior probability distribution function (pdf) of the  
187 regression parameters is computed as follows:

$$188 \quad f(\theta | Y) \propto f(Y | \theta) f(\theta) \quad (3)$$

189 where  $f(\theta)$  is the prior pdf of regression parameters and  $f(Y|\theta)$  is the likelihood function:

$$190 \quad f(Y | \theta) = \prod_t f(Y(t) | \beta_1(t), \beta_2(t), \dots, \beta_m(t))$$

$$191 \quad = \prod_t f(Y(t) | h_1(\mathbf{x}_t, \theta_1), h_2(\mathbf{x}_t, \theta_2), \dots, h_m(\mathbf{x}_t, \theta_m)) \quad (4)$$

192 In equation (4), a time independence assumption is applied:  $\forall t_1 \neq t_2$ ,  $Y(t_1)$  is independent of  
193  $Y(t_2)$  conditionally on  $\beta(t_1), \beta(t_2)$ .

194 The posterior pdf in Eq(3) is estimated using a Markov chain Monte Carlo (MCMC) sampler  
195 described in Renard et al. (2006a).

## 196    **2.2    Regional model with spatio-temporal covariates**

### 197    **2.2.1    Parent distribution**

198    Let  $Y(s, t)$  be the observation at site  $s$  and time  $t$  and  $\mathbf{Y} = (Y(s_j, t_k), j = \{1, 2, \dots, p\}, k = \{1, 2, \dots, n\})$   
199    be the collection of the observed data at all  $p$  observation sites for  $n$  time steps. Similar to the local  
200    model, a common distribution  $D$  is assumed for all sites, but with parameters varying in both space  
201    and time:

$$202 \qquad Y(s, t) \sim D(\boldsymbol{\beta}(s, t)) \qquad (5)$$

203    where  $\boldsymbol{\beta}(s, t) = (\beta_i(s, t), i = \{1, 2, \dots, m\})$  is the collection of all distribution parameters.  $m$  is the  
204    number of distribution parameters of  $D$ .  $\beta_i(s, t)$  is the  $i^{th}$  distribution parameter at time  $t$  and site  $s$ .

### 205    **2.2.2    Regression models with spatio-temporal covariates**

206    Similar to the local model, regressions are used to describe spatio-temporal variations in the  
207    parameters  $\beta_i(s, t)$ . However, three different kinds of covariates are considered in the regional  
208    model:

- 209    •    Temporal covariates  $\mathbf{x}(t)$  : e.g. time, SOI (Southern Oscillation Index), NAO (North  
210    Atlantic oscillation)...
- 211    •    Spatial covariates  $\boldsymbol{\omega}(s)$  : e.g. altitude, coordinates...
- 212    •    Spatio-temporal covariates  $\mathbf{z}(s, t)$  : e.g. temperature...

213    Temporal covariates only change over time (but are common to all sites), and spatial covariates  
214    only change over sites (but do not change in time). Spatio-temporal covariates change over both  
215    these dimensions.

216 The regional regression is then established in two steps: specification and spatialization. The first  
217 step establishes site-specific regressions with temporal and spatio-temporal covariates. The second  
218 step establishes a spatial model with the spatial covariates (see Figure 2):

219 1) **Specification step:** specify the time model using at-site regressions for a distribution  
220 parameter  $\beta_i(s, t)$  :

221 For a given site  $s$ :

$$222 \quad \beta_i(s, t) = h_i(\mathbf{x}(t), \mathbf{z}(s, t); \boldsymbol{\theta}(s)) \quad (6)$$

223 where  $h_i$  is the regression function,  $\mathbf{x}$  and  $\mathbf{z}$  are temporal and spatio-temporal covariates, and  $\boldsymbol{\theta}(s)$   
224 are the regression parameters.

225

226 2) **Spatialization step:** regression parameters  $\boldsymbol{\theta}(s)$  are split into two groups:

227  $\boldsymbol{\theta}(s) = \{\boldsymbol{\theta}_{loc}^{(s)}; \boldsymbol{\lambda}(s)\}$ .  $\boldsymbol{\theta}_{loc}^{(s)}$  is the collection of purely local parameters, whose value remains specific to  
228 each site  $s$ .  $\boldsymbol{\lambda}(s)$  represents all the parameters waiting to be spatialized. For each of its  
229 component  $\lambda(s)$ , we apply a spatial regression function. This spatial regression is time-invariant:  
230 neither spatial regression parameters nor covariates change over time. Hence, at this step, only  
231 regional parameters and spatial covariates are used. Thus a spatial regression function  $g$  is  
232 introduced:

$$233 \quad \lambda(s) = g(\boldsymbol{\omega}(s); \boldsymbol{\theta}_{reg}) \quad (7)$$

234 where  $\boldsymbol{\omega}(s)$  is a vector of spatial covariates and  $\boldsymbol{\theta}_{reg}$  is a vector of regional parameters (identical  
235 for all sites).

236 This two-step mechanism is very general and corresponds to standard regionalization reasoning. As  
237 an illustration, consider the following “trend analysis” situation: the mean of some  
238 hydrometeorological variable is assumed to be a linear function of time (step 1, specification).

239 Then, the slope of this linear trend line may be allowed to vary across sites according to elevation  
240 (step 2, spatialization).

### 241 **2.2.3 Spatial dependence**

242 In a region, the observations from different stations are in general not completely independent. The  
243 dependence is expected to increase with decreasing distance.

244 There exist several ways to model this dependence. In this paper, we opt for the use of copulas.  
245 Max-stable processes are an interesting alternative, especially in the context of extremes (Davison  
246 et al., 2012; Padoan et al., 2010). An illustration in hydroclimatic context was described by Westra  
247 and Sisson (2011). However, they are not considered in the proposed framework for the following  
248 reasons:

249 1) Max-stable processes are only suitable for extreme data, but the framework we propose is  
250 not restricted to extreme value distributions and leaves the choice of the marginal distribution open  
251 (see Eq (5)): in this respect, using max-stable processes would result in an important loss of  
252 generality.

253 2) Estimation of max-stable processes is challenging due to the difficulty of computing the  
254 whole likelihood. Pragmatic solutions based on the use of “composite likelihoods” have been  
255 proposed within a maximum-likelihood estimation context (see (Padoan et al., 2010) for further  
256 discussion), including a quantification of estimation uncertainties. However, we chose to use a  
257 Bayesian inference framework, within which max-stable processes estimation remains a work in  
258 progress (see e.g. Reich and Shaby (2012)).

259  
260 Copulas are used to build a joint distribution from a set of marginal distributions (Sklar, 1959). For  
261 a  $p$ -dimensional multivariate random variable  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$  with marginal cumulative  
262 distribution functions (cdf)  $F_1, F_2, \dots, F_p$ , a copula is a function  $C$  :

$$C : [0,1]^p \rightarrow [0,1] \quad (8)$$

$$(F_1(y_1), F_2(y_2), \dots, F_p(y_p)) \mapsto F(y_1, y_2, \dots, y_p)$$

where  $F$  is the joint cdf of the random variable  $Y$ .

Sklar (1959) showed the existence of such a function and pointed out that if the marginal distributions are continuous, then the copula  $C$  is unique. Applications of copulas in an hydrometeorological context have been described by e.g. Favre et al. (2004), Bardossy and Li (2008), Bardossy and Pegram (2009), Renard and Lang (2007), AghaKouchak et al. (2010) and Haslauer et al. (2010). Due to their convenience in highly dimensional setups (typically the case with spatial datasets) (Renard, 2011), elliptical copulas are favored in this paper. The elliptical copulas are linked to elliptical distributions (Genest et al., 2007). The two most commonly used are the Gaussian copula and the Student copula. Renard and Lang (2007) showed that some multivariate datasets could be correctly described by a Gaussian copula, which allows to account for spatially dependent data. It helps improving the estimation of parameter uncertainties, which are always under-estimated when incorrectly assuming spatially independent data. In practice, these two copulas are very convenient since the modeling of spatial dependence is related to the properties of multivariate Gaussian and Student distributions (respectively asymptotically independent and dependent), which are already well known (Fang et al., 2002; Genest et al., 2007). In particular, both copulas are parameterized by a symmetric matrix  $\Sigma$  representing pairwise dependence between sites. The appendix section provides all needed formula for these two particular copulas.

In this study, the dependence matrix  $\Sigma$  is a function of the inter-site distance: for any  $s_i \neq s_j$ ,

$$\Sigma(s_i, s_j) = \psi(\|s_i, s_j\|, \boldsymbol{\eta}) \quad (9)$$

where  $\|\cdot\|$  is the distance function and  $\psi$  is the correlation function whose variables are the distance and the dependence parameters  $\eta$ .

#### 2.2.4 Parameter estimation

Similar to equation(3), the posterior pdf of the regression parameters is given as follows:

$$f(\theta_{loc}, \theta_{reg}, \eta | Y) \propto f(Y | \theta_{loc}, \theta_{reg}, \eta) f(\theta_{loc}, \theta_{reg}, \eta) \quad (10)$$

where  $f(\theta_{loc}, \theta_{reg}, \eta)$  is the prior joint pdf. If the priors are assumed to independent (as in both case studies below) this simplifies to  $f(\theta_{loc}, \theta_{reg}, \eta) = f(\theta_{loc})f(\theta_{reg})f(\eta)$ . The posterior pdf of  $\theta_{loc}, \theta_{reg}, \eta$  is estimated by a MCMC sampler (Renard et al., 2006b). More detailed formulas for the likelihood function  $f(Y | \theta_{loc}, \theta_{reg}, \eta)$  are given in the Appendix.

### 2.3 Model comparison tools

The general framework allows analyzing the effect of different covariates on hydrometeorological data by using the regression models. Thus, a comparison tool is introduced to judge the performance of these models.

The Akaike Information Criterion (AIC) (Akaike, 1974), its modified version AICc (Burnham and Anderson, 2002) and Bayesian Information Criterion (BIC) (Schwarz, 1978) are three criteria based on parameter point-estimates (e.g. maximum likelihood). However, these criteria ignore parameter uncertainties. In the context of short records of extreme rainfall, parameter uncertainty is significant.

Bayesian Model Selection (BMS) techniques (Kass and Raftery, 1995) and Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) are two approaches based on posterior distribution of parameters, which include the parameter uncertainties. Further discussion and interpretation of the

BMS tools were described by Frost (2004). However, the use of BMS tools often requires using informative priors which are not always available in the context of climate-informed rainfall modeling. Therefore, the DIC criterion is used as a general criterion, because it accounts for the effect of prior information when available but remains usable with non-informative or improper priors (provided the posterior is well-posed).

For one given parameter vector  $\theta$ , the deviance is defined as follow:

$$D(\theta) = -2 \log(f(y | \theta)) \quad (11)$$

The DIC criterion is then computed by:

$$DIC = \bar{D} + p_D \quad (12)$$

where  $\bar{D} = E^\theta[D(\theta)]$  is the expectation of the deviance (with respect to the posterior distribution)

and  $p_D = \bar{D} - D(\bar{\theta})$  is the model complexity penalty. Models with small DIC values are preferred.

### 3 CASE STUDY 1: QUANTIFYING THE EFFECT OF ENSO ON SUMMER RAINFALL TOTALS USING LOCAL MODELS

Two case studies are considered in this paper. In both cases, we focus on the summer rainfall over Southeast Queensland (SEQ), Australia (Figure 3). This area was chosen because Cai et al. (2010) found that there is an asymmetric impact of ENSO on the summer rainfall in SEQ: La Niña episodes correspond to marked positive rainfall anomalies in SEQ, and the anomalies are direct function of the strength of the La Niña, while El Niño episodes do not appear to have any noticeable effects on rainfall. Cai et al. (2010) focused on spatially averaged rainfall over a large region, while in this paper, we will investigate if this effect is evident in the individual rainfall sites. This first case study uses a local model to verify and quantify such asymmetric effect on summer rainfall totals over SEQ.



### 3.1 Data and covariates

Rainfall data are provided by the Australian Bureau of Meteorology (BOM). High quality summer (Dec, Jan, Feb) totals (Lavery et al., 1997) are available over 16 observation sites until 2011, with the record starting year among these sites ranging from 1870 to 1913 with most having a record longer than one hundred years. An assessment of autocorrelation was performed by computing the autocorrelation functions of at-site data: 9 of the 10 sites had lag-one autocorrelation that was not statistically significant. Overall, the autocorrelation in the data is quite low. Thus data can be considered to be temporally independent. Figure 3 shows the location of the rain gauges.

The Southern Oscillation Index (SOI) is an index computed from the Mean Sea Level Pressure difference between Tahiti and Darwin. SOI is considered as a measure of ENSO. A positive (resp. negative) value of SOI corresponds to the La Niña (resp. El Niño) episode. The SOI data (1877-2011) were obtained from BOM (<http://www.bom.gov.au/climate/current/soi2.shtml>). The summer average SOI is used in this case study.

Alternative indices that quantify the variability of ENSO were considered as covariates. In preliminary analyses for the period 1950-2011 (not shown here), two indices were compared as potential covariates: SOI (1877-2011) and Niño 3.4 (1950-2011). Similar results were found on each site for both indices. The SOI index was therefore preferred as a covariate in this study due to the availability of a longer series.

### 3.2 Local model for the summer rainfall totals

The previous study by Cai et al. (2010) suggested separating the effect of La Niña (positive SOI) and El Niño (negative SOI) episodes on the summer rainfall in SEQ. Thus the specific

implementation of Eq (1) and (2) for this is to use a lognormal model for the summer total rainfall,  
as follows:

$$Y(t) \sim \log N(\mu(t), \sigma(t)) \quad (13)$$

where the mean  $\mu(t)$  is asymmetric with respect to the positive and negative phases of the SOI,  
while the standard deviation is assumed to be constant:

$$\mu(t) = \begin{cases} \mu_0 + \mu_1^- * SOI(t); SOI(t) < 0 \\ \mu_0 + \mu_1^+ * SOI(t); SOI(t) > 0 \end{cases} \quad (14)$$

$$\sigma(t) = \sigma_0 \quad (15)$$

where  $\mu_0, \mu_1^-, \mu_1^+$  and  $\sigma_0$  are the regression parameters.

The asymmetric regression model applied to  $\mu(t)$  in Eq (14) could also be applied to  $\sigma(t)$  in Eq (15).  
However, preliminary analyses (not shown) suggested no noticeable effect of the SOI on this  
parameter.

In this study, independent flat priors are used for the regression parameters.

### 3.3 Results

The goodness-of-fit of the at-site marginal distribution of the model to the observed data from  
individual sites is evaluated graphically by using a probability-probability plot (pp-plot). The idea  
behind the pp-plot is that if  $(Y_i)_{i=1,n}$  are random variables with cdf  $(F_i)_{i=1,n}$ , then  $(F_i(Y_i))_{i=1,n}$  are  
identically distributed according to a uniform distribution Unif[0,1]. Therefore the sorted values of  
 $F_1(y_1), F_2(y_2), \dots, F_n(y_n)$  are plotted against empirical frequencies  $\left(\frac{i}{n+1}\right)_{i=1,n}$ . If the fit is good, this  
plot should be close to the diagonal. More explanations and usage of pp-plot in a non-identically-  
distributed context can be found in Coles (2001) and Renard et al. (2013). In this study, the pp-plot

of each site is close to the diagonal (not shown), which indicates that the lognormal distribution is compatible with the observations.

The effect of El Niño (negative SOI) and La Niña (positive SOI) on the summer rainfall is characterized by  $\mu_1^-$  and  $\mu_1^+$  respectively. If such effect is significant, the posterior pdf of  $\mu_1^-, \mu_1^+$  should be significantly different from zero. Figure 4 indicates that most sites are significantly influenced by La Niña, whereas El Niño influence is not detected.

To further illustrate the effect of La Niña and El Niño, the p-value of 0 is calculated for the regression parameters  $\mu_1^-$  and  $\mu_1^+$ . This p-value is equal to  $Prob[\mu \leq 0 | Y]$ , which refers to the probability of the posterior distribution of  $\mu_1^-$  or  $\mu_1^+$  being smaller than 0. Figure 5 illustrates the p-value of all 16 sites on a map. During El Niño episodes, the majority of sites show little effect. However, during the La Niña episodes, the significance is quite clear.

Considering the effect of ENSO on summer rainfall, we concentrate on the slope of rainfall quantiles with respect to the SOI value. Figure 6 indicates that during the La Niña episodes, each unit of SOI value increases the summer rainfall by almost 5mm for the 0.5-quantile and by 10mm for the 0.99-quantile (1 in 100 year rainfall). However, during the El Niño episodes, no clear trend is found.

### 3.4 Summary

The analysis of summer totals shows a clear effect of La Niña but no strong effect of El Niño, thereby confirming the results of Cai et al. (2010). This effect can be detected even using a local model. In the remainder of this case study, we assess whether a similar relationship can be detected for extreme summer daily rainfall.

## **4 CASE STUDY 2: QUANTIFYING THE EFFECT OF ENSO ON SUMMER MAXIMUM DAILY RAINFALLS USING LOCAL AND REGIONAL MODELS**

The second case study focuses on the summer maximum daily rainfall over SEQ. King et al. (2013) used a linear regression analysis between spatially averaged annual maxima (of 5-day totals) and SOI, and found that the asymmetry in ENSO-rainfall teleconnection over SEQ also exists in the extreme rainfall, or at least in spatially-averaged extremes, which are quite different from local extremes recorded at rain gauges. In this study, we focus on investigating if the asymmetric effect of ENSO (that was evident in the summer total rainfalls) is also found in the observed summer maximum daily rainfalls, as well as the intensity of the effect. For this case study, the analysis is extended to include both local and regional models. In the case of extreme rainfall, there is considerably more uncertainty in the parameter estimates (cf. summer rainfall totals) – this uncertainty may mask the effect of ENSO. The use of regional model to reduce parameter uncertainty and better identify the effect of ENSO is highlighted. Furthermore, comparison of different models is undertaken to answer questions, such as: “Is the effect of ENSO on summer maximum daily rainfall symmetric or asymmetric?”

### **4.1 Data and covariates**

Among the 16 high quality sites (Figure 3), daily rainfall data is available in 10 sites. The record starting years among these sites are ranging from 1880 to 1906. Summer maximum daily rainfall is extracted from the daily data of these 10 rain gauges. The same covariate (SOI) as in Section 3.1 is used in this section.

## 4.2 Models for summer rainfall maximum

### 4.2.1 Local model with temporal covariates

Annual/seasonal maxima are often modelled with a generalized extreme value (*GEV*) distribution. The relevant theory was introduced by Fisher and Tippett (1928). In this study, the specific implementation of Eq (1) is a *GEV* model for the summer maximum daily rainfall (Coles et al., 2003; Katz et al., 2002):

$$Y(t) \sim \text{GEV}(\mu(t), \sigma(t), \xi(t)) \quad (16)$$

To consider the ENSO effect on the location ( $\mu(t)$ ) and scale ( $\sigma(t)$ ) of the *GEV* distribution, these parameters are assumed to be dependent on SOI, while the shape parameter is assumed to be constant. This is because the shape parameter  $\xi$  is difficult to estimate at a local scale (Coles, 2001, p106) even in the stationary context.

To determine whether the asymmetric effect of ENSO found in the summer rainfall totals is also observed in the summer maximum daily rainfall, two different regression models are considered. The first one is a symmetric linear model and the other one is an asymmetric piecewise-linear model. To distinguish these two models, the asymmetric model uses the same symbols as in equations (14) and (15), and the symmetric model parameters are denoted with a tilde.

Model 1 (Symmetric linear model)

$$\tilde{\mu}(t) = \tilde{\mu}_0 + \tilde{\mu}_1 * \text{SOI}(t) \quad (17)$$

$$\tilde{\sigma}(t) = \tilde{\sigma}_0 + \tilde{\sigma}_1 * \text{SOI}(t) \quad (18)$$

$$\tilde{\xi}(t) = \tilde{\xi}_0 \quad (19)$$

Model 2 (Asymmetric piecewise-linear model)

$$\mu(t) = \begin{cases} \mu_0 + \mu_1^- * \text{SOI}(t); & \text{SOI}(t) < 0 \\ \mu_0 + \mu_1^+ * \text{SOI}(t); & \text{SOI}(t) > 0 \end{cases} \quad (20)$$

$$\sigma(t) = \begin{cases} \sigma_0 + \sigma_1^- * SOI(t); SOI(t) < 0 \\ \sigma_0 + \sigma_1^+ * SOI(t); SOI(t) > 0 \end{cases} \quad (21)$$

$$\xi(t) = \xi_0 \quad (22)$$

where  $\theta_{M_1} = \{\tilde{\mu}_0, \tilde{\mu}_1, \tilde{\sigma}_0, \tilde{\sigma}_1, \tilde{\xi}_0\}$ ,  $\theta_{M_2} = \{\mu_0, \mu_1^-, \mu_1^+, \sigma_0, \sigma_1^-, \sigma_1^+, \xi_0\}$  are the regression parameters of Model 1 and Model 2. Independent flat priors are used in this study as well.

#### 4.2.2 Regional models

In order to better identify the parameters quantifying the effect of ENSO, regional models are applied in this case study. Following the two-step construction introduced in Section 2.2, in the first step, the time-varying structure at each site is prescribed using the same regression functions as in the equations (17)-(19) and (20)-(22). In the second step, two sets of parameters are spatialized. The first set comprises the ENSO effect parameters (e.g. for the asymmetric model,  $\mu_1^-, \mu_1^+, \sigma_1^-$  and  $\sigma_1^+$ ). Indeed, climate indices, like ENSO, are expected to have similar effects on all the observation sites within a region. We also assume a regional shape parameter  $\xi_0$ . Thus we assume these parameters are the same over the region. Conversely, all other parameters are assumed purely local. The regionalized equations for the asymmetric model thus become:

$$\mu(s, t) = \begin{cases} \mu_{loc_0}^{(s)} + \mu_{reg_1}^- * SOI(t); SOI(t) < 0 \\ \mu_{loc_0}^{(s)} + \mu_{reg_1}^+ * SOI(t); SOI(t) > 0 \end{cases} \quad (23)$$

$$\sigma(s, t) = \begin{cases} \sigma_{loc_0}^{(s)} + \sigma_{reg_1}^- * SOI(t); SOI(t) < 0 \\ \sigma_{loc_0}^{(s)} + \sigma_{reg_1}^+ * SOI(t); SOI(t) > 0 \end{cases} \quad (24)$$

$$\xi(s, t) = \xi_{reg_0} \quad (25)$$

where

$\theta_{loc}^{(s)} = (\mu_{loc}^{(s)}, \sigma_{loc}^{(s)}) = ((\mu_{loc_0}^{(s_1)}, \mu_{loc_0}^{(s_2)}, \dots, \mu_{loc_0}^{(s_p)}), (\sigma_{loc_0}^{(s_1)}, \sigma_{loc_0}^{(s_2)}, \dots, \sigma_{loc_0}^{(s_p)}))$  are local regression parameters.

$\theta_{reg} = (\mu_{reg_1}^-, \mu_{reg_1}^+, \sigma_{reg_1}^-, \sigma_{reg_1}^+, \xi_{reg_0})$  are the regional regression parameters.

464

465 For all the models considered therein, a Gaussian copula is used to describe the spatial dependence.

466 The distance-dependence relationship is characterized by the following function:

$$467 \quad \Sigma(s_i, s_j) = \eta_1 * \exp(-\eta_2 * \|s_i, s_j\|) \quad (26)$$

468 where  $\eta_1$  and  $\eta_2$  are the dependence parameters.

### 469 **4.3 Assessing statistical hypotheses of ENSO effect on summer maximum daily rainfalls**

470 In this case study, we consider three competing statistical hypotheses in terms of the relationship  
471 between ENSO and summer maximum rainfall which lead to different regression models used in  
472 the framework, as follows:

473 1) There is no ENSO influence on the maximum rainfall, leading to a time-invariant model.  
474 This hypothesis will be used as baseline, to compare the predictions of summer maximum rainfall  
475 from the other two hypotheses which include an ENSO effect.

476 2) The ENSO influence on rainfall is symmetric linear with respect to the SOI values. The  
477 physical interpretation of this hypothesis is that the strength of the El Nino event and the strength of  
478 the La Nina event (as measured by the SOI) have a “symmetric” impact on maximum rainfall.  
479 Assuming a symmetric linear relationship between rainfall and ENSO indices is a common  
480 approach used in several previous studies (e.g. Chiew et al. (1998); Nicholls and Kariko (1993);  
481 Pui et al. (2012); Risbey et al. (2009)). Thus a symmetric linear regression model is trialed (Eq. (17)  
482 and (18)).

483 3) The third statistical hypothesis is that ENSO has an asymmetric effect during two different  
484 phases (El Niño and La Niña). The physical interpretations of this hypothesis are that the impact of  
485 the strength of the La Nina is different from the impact of the strength of an El Nino event.  
486 This was motivated by the results of Cai et al. (2010) (who provide evidence of the physical  
487 mechanisms, outlined in the introduction) and King et al. (2013), and the results from the analysis

of the multi-site summer rainfall totals. Therefore an asymmetric linear regression model is trialed (Eq. (23) and (24)).

Furthermore, an important research question we are interested in is whether the multi-site information from regional analysis provides improved identification of the effect of ENSO. Hence, we are interested in comparing local and regional versions of the same models.

Table 1 gives an overview of the different local and regional models. To simplify the notation, we use “L” for local and “R” for regional. The name of the models is denoted by their regression functions on the location and scale parameter ( $\langle \text{location regression function} \rangle_{\langle \text{scale regression function} \rangle}$ ). *Stat* is for the identical (stationary) function. *Sym* is for the symmetric function. *Asy1* is for the asymmetric function as in equations (23) and (24). *Asy2* is another asymmetric function in which the slope during negative SOI episode is fixed at 0 since the El Niño effect is not significant for summer total rainfall as shown in Section 3.4.

## 4.4 Results

### 4.4.1 Goodness-of-fit

Similar to the first case study, the goodness of fit of the at-site marginal distribution of the model to the observed data from individual sites is evaluated using a probability-probability plot described in Section 3.3. Figure 7 shows the results for the local (*LAsy1\_LAsy1*) and regional (*RAsy1\_RAsy1*) asymmetric models for all ten sites. The lines are all close to the diagonal, which indicates that both *GEV* local and regional asymmetric models have good fit with the observation data. The spatial dependence is also acceptably captured with the dependence-distance relationship assumed in eq.



(26) (Figure 8). However, this relationship is certainly perfectible: there is a non-negligible scatter around the red curve, suggesting that inter-site distance only explains a part of spatial dependence.

#### 4.4.2 Identifying the effect of ENSO on summer maximum daily rainfall: None, Symmetric or Asymmetric? Local analysis

The symmetric model (*LSym\_LSym*) does not separate El Niño and La Niña episodes. The p-value shown in the Figure 9 (a) and (b) indicates that 6 out of 10 sites detect a significant ENSO effect on location or scale parameter or both. The asymmetric model (*LAsyI\_LAsyI*) separates the effect of El Niño and La Niña episodes. Similar to the result of the summer total rainfall, the El Niño effect is found neither on the location nor on the scale parameter (not shown) for almost all sites. However, the La Niña effect is detected on either location or scale parameter or both (Figure 9 (c) (d)). The significance on the scale parameter indicates that La Niña also increases the variability of the summer maximum rainfall over the majority of sites. With both models, summer maximum rainfall is found to be affected by ENSO, at least during the La Niña episode.

Compared with the asymmetric model, the symmetric model has two main differences. One is the value of the slope (with respect to SOI) and the other is the significance of the effect. An overview of all ten sites (Figure 10) indicates that 8 out of 10 sites have significant positive slope for the 1 in 100 year rainfall for the asymmetric model during the La Niña episode, and values of the slope are ranging from 4 to 10 mm/unit SOI. In comparison for the symmetric linear model, only half of the sites show a significant effect, and values of the slope are much lower ranging from 1 to 5 mm/unit SOI. From the asymmetric model, a significant effect is found during the La Niña episodes, but not during El Niño episodes, which explains why the analysis based on the symmetric model leads to less significant results.

535 This result is consistent with the physical mechanism postulated by Cai et al. (2010), whereby the  
536 strength of the La Nina event has an effect on summer maximum rainfall, while the strength of a El  
537 Nino event does not.

#### 538 **4.4.3 ENSO-conditional predictions for summer maximum extreme rainfall: Local analysis**

539 Figure 11 illustrates the relationship between the 1 in 100 year rainfall (0.99-quantile) and the SOI  
540 index for site 16. The large slope of the asymmetric model (red) indicates that, for the positive SOI,  
541 each incremental unit increase in the SOI value will increase the 1 in 100 year rainfall by nearly  
542 5mm, whereas the negative SOI does not have a statistically significant effect. Figure 11 also  
543 illustrates that these estimations are affected by very large uncertainties. During a strong La Niña,  
544 the asymmetric model estimates the posterior median of the 1 in 100 year rainfall is almost 25%  
545 higher than the symmetric model and 45% higher than a stationary model (Figure 11). Over all sites  
546 (not shown here), these two values can be up to 33 % and 50% respectively.

547  
548 Although the asymmetric model detects a significant ENSO effect during La Niña, the ENSO-  
549 conditional predictions are affected by large uncertainties. This is due to the difficulty of precisely  
550 identifying the parameters with a local analysis. The regional analysis aims to reduce parameter  
551 uncertainties, hence better quantify the effect of El Niño and La Niña.

#### 552 **4.4.4 Does regional analysis improve the identification of the effect of ENSO on summer** 553 **maximum daily rainfall?**

554 Figure 12 gives the distributions of the La Niña effect parameters on the *GEV* location parameter in  
555 local (*LAsyI\_LAsyI*) and regional (*RAsyI\_RAsyI*) models. There is a significant reduction of the  
556 distribution width for the regional model. Figure 13 illustrates that, for the asymmetric model  
557 *RAsyI\_RAsyI*,  $\mu_{reg_1}^+$  and  $\sigma_{reg_1}^+$  (associated to La Niña) are found significantly larger than 0, whereas

558  $\mu_{reg_1}^-$  and  $\sigma_{reg_1}^-$  (associated to El Niño) are not. This regional analysis gives a more robust conclusion  
559 that the strength of the La Niña event has a significant influence on the summer maximum daily  
560 rainfall, whereas El Niño has not. Furthermore, the reduction of uncertainty on the La Niña effect  
561 parameters ( $\mu_{reg_1}^+, \sigma_{reg_1}^+$ ) and the shape parameter ( $\xi_{reg_0}$ ) provides an important improvement to  
562 decrease the uncertainty on high quantiles. In Figure 14, the 1 in 100 year rainfall of the local and  
563 regional asymmetric models are compared. During a strong El Niño, the uncertainty of the regional  
564 model (measured by the interval width) is reduced by 50% compared with the local model, and  
565 during strong La Niña, this reduction is up to 60%. This clearly shows the benefit of regional  
566 analysis in better identifying the asymmetric effect of ENSO on extreme rainfall.

#### 567 **4.4.5 Model comparison for summer rainfall maxima**

568 In this section, we use the DIC criterion to compare the following three pairs of models. A better  
569 model is denoted by a smaller DIC.

##### 570 i) Local vs. regional modelling

571 Figure 15 illustrates the DIC values for the models in Table 1. The DIC values of at-site models  
572 ( $L\_Stat\_Stat$ ,  $Lsym\_Lsym$ ,  $LAsy1\_LAsy1$ ) are much larger than the regional models ( $R\_Stat\_Stat$ ,  
573  $RSym\_RSym$ ,  $RAsy1\_RAsy1$ ,  $RAsy2\_RAsy2$ ,  $RAsy1\_Stat$ ,  $RAsy2\_Stat$ ). Compared with the regional  
574 models, the local models have many more parameters, which lead to a large penalty on the model  
575 complexity. Thus regional models are preferred (according to the DIC criterion) to at-site models.

##### 576 ii) Stationary model vs. climate-informed model

577 According to i), we assess this point with regional models. Among the six regional models  
578 ( $R\_Stat\_Stat$ ,  $RSym\_RSym$ ,  $RAsy1\_RAsy1$ ,  $RAsy2\_RAsy2$ ,  $RAsy1\_Stat$ ,  $RAsy2\_Stat$ ), the DIC value  
579 of the stationary model ( $R\_Stat\_Stat$ ) is the largest (Figure 15). Thus, this result shows once again  
580 that ENSO influences the summer maximum rainfall over SEQ and justifies the use of a climate-  
581 informed model.

### 582   iii)   Symmetric vs. asymmetric effect of ENSO

583   The comparison between the symmetric and asymmetric models is established with the regional  
584   models listed in Table 1. Table 2 summarizes the DIC difference between the regional models in the  
585   list and the preferred model ( $RA_{sy2\_RA_{sy2}}$ ) with the smallest DIC. This preferred model has  
586   asymmetric regressions on both location and scale parameters. The difference between  
587    $RA_{sy2\_RA_{sy2}}$  and the remaining regional models suggests that the models with asymmetric ENSO  
588   effect are preferred. In particular, model  $R_{Stat\_Stat}$  (no ENSO effect) is strongly discredited  
589   according to the DIC. Lastly, these results also suggest that modelling a trend on the scale  
590   parameter is preferable, since models  $RA_{sy1\_Stat}$  and  $RA_{sy2\_Stat}$  have a lesser performance than  
591   the reference model  $RA_{sy1\_RA_{sy1}}$  and  $RA_{sy2\_RA_{sy2}}$ .

### 592   4.4.6   Summary

593   We use both at-site and regional models to analyze ENSO effects on the summer rainfall maximum  
594   over SEQ. The link between ENSO and summer maximum daily rainfall is strong during La Niña  
595   phase and weak during El Niño phase, confirming the results of Cai et al. (2010) and King et al.  
596   (2013). We demonstrate that using regional model helps to reduce the uncertainty and provides  
597   more robust results. With the DIC criterion, competing models are compared. It is found that the  
598   asymmetric regression on both location and scale parameters is the preferred representation of  
599   ENSO effect on summer maximum daily rainfall.

## 600   5   DISCUSSION

601   This section discusses key assumptions and current limitations of the modeling framework, and  
602   their consequences on the SEQ case study. It also proposes avenues for future improvements.

## 603 **5.1 Assumption of homogeneous regions**

604 An assumption of the regional model is that all data should be subject to similar climate effects.  
605 This raises the question of defining such climatically homogenous regions. Ouarda et al. (2001)  
606 described some approaches to determine homogeneous hydrologic regions. Some Southeast  
607 Australian basins have also been classified into homogeneous regions by Bates et al. (1998). The  
608 SEQ is a relatively small area, thus SEQ is assumed to be inside a same climatic homogenous  
609 region. However, when studying larger areas, the classification of different homogeneous regions  
610 will play an important role.

## 611 **5.2 Spatial dependence modelling**

612 The reason for using simple copulas, like Gaussian and Student copulas, is that they are applicable  
613 to any marginal distribution, which is convenient in the context of the general framework proposed  
614 in this paper. Moreover the parameterization by a dependence matrix enables using geostatistical-  
615 like models (pairwise dependence is a function of distance). Results in Figure 8 suggest that such a  
616 simple relationship, while perfectible, adequately captures spatial correlations. However, different  
617 copulas have different asymptotic behaviour: asymptotically dependent (e.g. Student copula) and  
618 asymptotically independent (e.g. Gaussian copula). The extrapolation of copula is risky because the  
619 asymptotic dependence properties exert a strong leverage on joint probability of exceedance, but the  
620 limited sample size is not enough to identify such asymptotic properties. Therefore, to make an  
621 informed decision between asymptotic dependent and independent copulas, more physical  
622 knowledge on the spatial extent of rainfall or meteorological events is required. Moreover, future  
623 work could also investigate alternatives to copula, in particular max-stables processes that may be  
624 more relevant for extreme data.

### 625 **5.3 Spatial regression modelling**

626 Inside a homogenous region, the distribution of the rainfall may depend on the spatial information  
627 at each site. For example the ENSO effect could vary with elevation or distance to sea. However, in  
628 the case study, we simply assume the same ENSO effect and shape parameter for all sites. Note  
629 however that this does not imply that the rainfall distribution is the same at all sites: since location  
630 and scale parameters remain site-specific, the distribution will differ from site to site, which allows  
631 accounting for e.g. orographic effects.

632 Spatial effects could be investigated in the case study in several aspects. First, some parameters are  
633 purely local, which prevents transferring quantile estimates to ungauged sites. This could be  
634 improved by spatializing these parameters using a spatial regression. Moreover, a more flexible  
635 model could be considered by allowing spatial variations in purely regional parameters (ENSO  
636 effects and shape parameter). This was not attempted in this case study because identifying such  
637 spatial effects is difficult with only ten sites: we therefore favoured the identification of ENSO  
638 effects. However, future case studies based on a spatially denser dataset will investigate in more  
639 depth the construction of such spatial models.

### 640 **5.4 Practical Implications: Utilising predictions of extreme rainfall distributions from the** 641 **climate-informed framework**

642 One of the advantages of using a fully probabilistic model for extremes (as opposed to a simple  
643 linear regression between SOI and spatially averaged rainfall, as undertaken in Cai et al. (2010) and  
644 King et al. (2013)) is that it enables the prediction of the occurrence probability of extreme rainfall  
645 conditioned on climate variability indices. Figure 10, Figure 11 and Figure 14 all provide prediction  
646 of the 1 in 100 year rainfall conditional on values of SOI. For the preferred regional model  
647 (*RAsyI*\_*RAsyI*), during strong La Niña phases (with high SOI), the 1 in 100 summer maximum

648 daily rainfall is about 33% higher than the corresponding estimate obtained with the stationary  
649 model for this particular site (e.g. Figure 14).

650

651 From an operational perspective, the knowledge that the 1 in 100 year quantile is 33% higher during  
652 a strong La Niña could provide useful information for planners, engineers, water resource  
653 managers, emergency response organizations, in order to design operational/response strategies to  
654 mitigate the potential impact due to the increased risk of extreme rainfall. Consider the recent  
655 example of the summer of 2010-2011, when there was a strong La Niña (SOI = 27.1, in December)  
656 and a series of floods hit Queensland, which impacted on more than 70 towns and 200,000 people.  
657 The damage bill was over 5 billion \$AUD (page 4, Operation Queenslander: the State Community,  
658 Economic and Environmental Recovery and Reconstruction Plan, 2011-2013). One of the major  
659 impact was a major flood in the city of Brisbane (a major Australian city with a population of 2.15  
660 million), caused by the release of water from the major Wivenhoe dam upstream of Brisbane (see  
661 Chapter 16, Queensland Floods Commission of Inquiry). Armed with this knowledge of the effect  
662 of ENSO on extreme rainfall, planners/engineers/water resource managers, would be able to  
663 undertake better planning of emergency response, and potentially improve reservoir operating rules  
664 to better control floods, and reduce the impact of extreme rainfall during strong La Niña's. On the  
665 other hand of the hydrologic spectrum, climate-informed frameworks have the same importance for  
666 predicting extreme droughts (Henley et al., 2011).

667

668 From a design perspective, the unconditional marginal distribution of extreme rainfall would be  
669 needed (e.g. for designing a dam or other hydraulic structures). Evaluating the marginal  
670 probabilities involves integrating out the SOI. This requires determining the distribution of SOI.  
671 Historical information could be used to inform this distribution, or alternatively predictions of the  
672 future variation in SOI from climate changes models could also be used. This climate-informed

framework which provides a quantitative link between climate variability and rainfall provides far more useful information than that derived from a stationary model. The comparison of the extreme rainfall risk from a stationary model to the ones obtained by integrating out SOI in a climate-informed model is an important question that will be investigated in future work.

## 5.5 Stationarity or non-Stationarity?

The distinction between the conditional and unconditional distributions introduced in section 5.4 is important with respect to the concept of stationarity. Indeed, a SOI-conditional model yields a distribution that varies in time simply because SOI values vary in time. Yet this does not necessarily imply non-stationarity: indeed, the unconditional distribution (after integrating out SOI) might not depend on time if the SOI values are themselves realizations from a stationary distribution. Conversely, if a temporal trend affects SOI, this trend will propagate to the unconditional distribution of rainfall, yielding a non-stationary unconditional distribution.

Historically, there has been changes in the frequency of ENSO events, and under climate change, the changes to the frequency of ENSO events is unknown (Giese and Ray, 2011; Ray and Giese, 2012). It is therefore difficult to decide whether the models used in this paper are stationary or not. This is why we favoured the expressions “climate-informed models” or “conditional models” over the expression “non-stationary models”.

## 5.6 Perspective on hierarchical models

In the regional model, we proposed two kinds of parameters: local and regional parameters. Local parameters are different for each site, which offers a good flexibility. Regional parameters are common for all sites, yielding reduced uncertainties (Figure 12 and 13). However, this distinction may be too “rigid”. Some parameters may be different at each site, but still have some spatial



consistency. A possible improvement is to use hierarchical models to enable constrained variations of parameters in space. Wikle et al. (1998) described a general hierarchical Bayesian framework in a non-stationary context. Lima and Lall (2009) (2010) used hierarchical models to describe the daily rainfall occurrence and extreme runoff. Renard (2011) and Renard et al. (2013) proposed a general hierarchical approach to regional frequency analysis. The next step of this work could therefore be to generalize the model proposed in this paper to a hierarchical setup.

## 6 CONCLUSIONS

In this paper, we describe a general spatio-temporal regional frequency analysis framework, geared towards detecting and quantifying the effect of climate variability on hydrometeorological variables. This is undertaken by using temporal regression models where the parameters of the probability distribution of hydrometeorological events are a function of climate drivers (e.g. ENSO). A flexible framework is adopted, which allows testing different temporal regression functions to describe the effect of climate variability. This flexibility provides a convenient way to compare these models and to select the most relevant relationship between climate indices and hydrometeorological data. For the regional analysis, spatial dependency is incorporated using copulas and a Bayesian approach is used for inference to enable uncertainties to be easily quantified. The use of a Bayesian regional framework provides the opportunity to assess the value of regional information in better identifying the effect of climate variability on hydrometeorological extremes.

The first case study with the dataset of summer rainfall totals in Southeast Queensland shows that La Niña exerts a significant influence in the region for summer rainfall totals, while the effect of El Niño is not significant.

720

721 In the second case study of summer daily rainfall maxima, the flexible framework enables  
722 comparing numerous models to incorporate the effect of ENSO on extreme rainfall over SEQ.  
723 Stationary, symmetric and asymmetric models in both local and regional setups are compared using  
724 a model selection criterion (in this case, the DIC). Overall, the use of regional models yielded better  
725 identification of the effect of ENSO on extreme rainfall over SEQ compared with using only local  
726 models, for which there was too much uncertainty to enable a clear identification. A variety of  
727 regional models, with different representations of the effect of ENSO (linear symmetric versus  
728 asymmetric) were also compared. Asymmetric models are found to be the best among them. More  
729 precisely, it is found that an asymmetric model, distinguishing between ENSO effect on location  
730 and scale parameters during the positive and negative phases of the SOI, is the most suitable in this  
731 case. These results corroborate the findings of other recent studies (Cai et al. (2010) and (King et  
732 al., 2013)).

733

734 From a practical perspective, it was found that during a strong La Niña the mostly likely 1 in 100  
735 year rainfall for different sites can be 20% to 50% higher than estimates using a stationary model  
736 which ignores the influence of ENSO, albeit with significant uncertainty. This information has the  
737 potential to be used by engineers/planners to provide better informed flood response strategies.

738

739 The framework developed in this paper is general, and in the future can be extended in several  
740 ways. Firstly, spatial effects can be included using spatial regression models with spatial covariates  
741 (e.g. elevation), while hierarchical approaches will also be developed to provide more flexibility for  
742 modeling the effect of spatio-temporal covariates on hydrometeorological variables.

## 7 ACKNOWLEDGMENTS

Xun Sun is supported by a grant from the Région Rhône-Alpes (Explora'doc). We also wish to acknowledge the additional research funding provided by Electricité de France (EDF), Irstea DRI and the University of Adelaide. We wish to acknowledge Prof. Dmitri Kavetski for providing the FORTRAN library DMSL. We also wish to thank all three anonymous reviewers and the associate editor for their useful comments.

## 8 APPENDIX

For a fixed time  $t_k$ , the joint pdf of  $(Y(s_j, t_k))_{j=1,p}$  is given by the Gaussian copula formula (see e.g. Renard (2011)). This yields:

$$\begin{aligned}
 & f(y(s_1, t_k), y(s_2, t_k), \dots, y(s_p, t_k) | (\beta_i(s_j, t_k), i = \{1, 2, \dots, m\}, j = \{1, 2, \dots, p\}), \boldsymbol{\eta}) \\
 &= \left( \frac{\prod_{j=1}^p f_j(y(s_j, t_k) | (\beta_i(s_j, t_k), i = \{1, 2, \dots, m\}))}{\prod_{j=1}^p \phi(u_{j,k})} \right) \Phi_{\Sigma}(u_{1,k}, u_{2,k}, \dots, u_{p,k} | \boldsymbol{\eta}) \\
 &= \left( \frac{\prod_{j=1}^p f_j(y(s_j, t_k) | \boldsymbol{\theta}_{loc}^{(j)}, \boldsymbol{\theta}_{reg})}{\prod_{j=1}^p \phi(u_{j,k})} \right) \Phi_{\Sigma}(u_{1,k}, u_{2,k}, \dots, u_{p,k} | \boldsymbol{\eta}) \quad (27)
 \end{aligned}$$

where

$f_j(y(s_j, t_k) | (\beta_i(s_j, t_k), i = \{1, 2, \dots, m\}))$  is the marginal pdf for site  $s_j$  time  $t_k$ ;

$u_{j,k} = \mathcal{Y}^{-1}(F_j(y(s_j, t_k)))$ ;

$\phi(u)$  is the standard Gaussian pdf or Student pdf with  $\nu$  degree of freedom (the latter being made implicit in the notation);

760  $\Phi(u_{1,k}, u_{2,k}, \dots, u_{p,k})$  is the multivariate Gaussian pdf (with mean=0, correlation matrix  $\Sigma$ ) or  
761 multivariate Student pdf (with mean=0, correlation matrix  $\Sigma$  and degree of freedom  $\nu$ , the latter  
762 being made implicit in the notation)  
763  
764 The derivation of the full likelihood uses a time independence assumption:  $\forall s, \forall t_1 \neq t_2, Y(s, t_1)$  is  
765 independent of  $Y(s, t_2)$  conditionally on  $\beta(s, t_1), \beta(s, t_2)$ . Therefore, the likelihood  
766 function  $f(Y | \theta_{loc}, \theta_{reg}, \eta)$  for all time and all sites is the product of equation(27) applied at all  $n$  time  
767 steps.

## 768 9 REFERENCE

- 769 AghaKouchak, A., Bardossy, A., Habib, E., 2010. Copula-based uncertainty modelling: application to multisensor  
770 precipitation estimates. *Hydrological Processes*, 24(15): 2111-2124.  
771 Akaike, H., 1974. New look at statistical-model identification. *Ieee Transactions on Automatic Control*, AC19(6): 716-  
772 723.  
773 Aryal, S.K. et al., 2009. Characterizing and modeling temporal and spatial trends in rainfall extremes. *Journal of*  
774 *Hydrometeorology*, 10(1): 241-253.  
775 Bardossy, A., Li, J., 2008. Geostatistical interpolation using copulas. *Water Resources Research*, 44(7): W07412.  
776 Bardossy, A., Pegram, G.G.S., 2009. Copula based multisite model for daily precipitation simulation. *Hydrology and*  
777 *Earth System Sciences*, 13(12): 2299-2314.  
778 Bates, B.C., Rahman, A., Mein, R.G., Weinmann, P.W., 1998. Climatic and physical factors that influence the  
779 homogeneity of regional floods in southeastern Australia. *Water Resources Research*, 34(12): 3369-3381.  
780 Brockwell, P.J., Davis, R.A., 2006. *Time series: theory and methods*. Springer.  
781 Burnham, K.P., Anderson, D.R., 2002. *Model selection and multimodel inference : a practical information-theoretic*  
782 *approach*. Springer.  
783 Cai, W., van Rensch, P., Cowan, T., Sullivan, A., 2010. Asymmetry in ENSO teleconnection with regional rainfall, its  
784 multidecadal variability, and impact. *Journal of Climate*, 23(18): 4944-4955.  
785 Castello, A.F., Shelton, M.L., 2004. Winter precipitation on the US Pacific Coast and El Nino Southern oscillation  
786 events. *International Journal of Climatology*, 24(4): 481-497.  
787 Cayan, D.R., Redmond, K.T., Riddle, L.G., 1999. ENSO and hydrologic extremes in the western United States. *Journal*  
788 *of Climate*, 12(9): 2881-2893.  
789 Chiew, F.H., Piechota, T.C., Dracup, J.A., McMahon, T.A., 1998. El Nino/Southern Oscillation and Australian rainfall,  
790 streamflow and drought: Links and potential for forecasting. *Journal of Hydrology*, 204(1): 138-149.  
791 Coles, S., 2001. *An introduction to statistical modeling of extreme values*. Springer.  
792 Coles, S., Pericchi, L.R., Sisson, S., 2003. A fully probabilistic approach to extreme rainfall modeling. *Journal of*  
793 *Hydrology*, 273(1-4): 35 - 50.  
794 Cooley, D., Nychka, D., Naveau, P., 2007. Bayesian spatial modeling of extreme precipitation return levels. *Journal of*  
795 *The American Statistical Association*, 102(479): 824-840.  
796 Cunderlik, J.M., Burn, D.H., 2003. Non-stationary pooled flood frequency analysis. *Journal of Hydrology*, 276(1-4):  
797 210-223.  
798 Davison, A.C., Padoan, S.A., Ribatet, M., 2012. Statistical modeling of spatial extremes. *Statistical Science*, 27(2): 161-  
799 186.  
800 Fang, H.-B., Fang, K.-T., Kotz, S., 2002. The meta-elliptical distributions with given marginals. *Journal of Multivariate*  
801 *Analysis*, 82(1): 1-16.

- Favre, A.C., El Adlouni, S., Perreault, L., Thiémonge, N., Bobee, B., 2004. Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40(1): W01101.
- Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(02): 180-190.
- Frost, A.J., 2004. Spatio-temporal hidden markov models for incorporating interannual variability in rainfall, PhD thesis, University of Newcastle.
- Genest, C., Favre, A.C., Beliveau, J., Jacques, C., 2007. Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research*, 43(9): W09401.
- Gershunov, A., Barnett, T.P., 1998. ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results. *J Climate*, 11(7): 1575-1586.
- Gershunov, A., Cayan, D.R., 2003. Heavy daily precipitation frequency over the contiguous United States: Sources of climatic variability and seasonal predictability. *J Climate*, 16(16): 2752-2765.
- Ghosh, S., Mallick, B.K., 2011. A hierarchical Bayesian spatio-temporal model for extreme precipitation events. *Environmetrics*, 22(2): 192-204.
- Giese, B.S., Ray, S., 2011. El Niño variability in simple ocean data assimilation (SODA), 1871-2008. *Journal of Geophysical Research C: Oceans*, 116(2).
- Gregersen, I.B., Madsen, H., Rosbjerg, D., Arnbjerg-Nielsen, K., 2013. A spatial and non-stationary model for the frequency of extreme rainfall events. *Water Resources Research*: doi:10.1029/2012WR012570, in press.
- Grimm, A.M., Tedeschi, R.G., 2009. ENSO and Extreme Rainfall Events in South America. *Journal of Climate*, 22(7): 1589-1609.
- Hanel, M., Buishand, T.A., Ferro, C.A.T., 2009. A nonstationary index flood model for precipitation extremes in transient regional climate model simulations. *Journal of Geophysical Research-Atmospheres*, 114.
- Haslauer, C.P., Li, J., Bardossy, A., 2010. Application of Copulas in Geostatistics. *Geoenviron Vii - Geostatistics For Environmental Applications*, 16: 395-404.
- Haylock, M. et al., 2006. Trends in total and extreme South American rainfall in 1960-2000 and links with sea surface temperature. *J Climate*, 19(8): 1490-1512.
- Henley, B.J., Thyer, M.A., Kuczera, G., Franks, S.W., 2011. Climate-informed stochastic hydrological modeling: Incorporating decadal-scale variability using paleo data. *Water Resources Research*, 47(11): W11509.
- Hoerling, M.P., Kumar, A., Zhong, M., 1997. El Nino, La Nina, and the nonlinearity of their teleconnections. *Journal of Climate*, 10(8): 1769-1786.
- Kamruzzaman, M., Beecham, S., Metcalfe, A., 2013. Climatic influences on rainfall and runoff variability in the southeast region of the Murray - Darling Basin. *International Journal of Climatology*, 33(2): 291-311.
- Kass, R.E., Raftery, A.E., 1995. Bayes Factors. *Journal of the American Statistical Association*, 90(430): 773-795.
- Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8-12): 1287 - 1304.
- Khaliq, M.N., Ouarda, T.B.M.J., Ondo, J.C., Gachon, P., Bobee, B., 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, 329(3-4): 534-552.
- King, A.D., Alexander, L.V., Donat, M.G., 2013. Asymmetry in the response of eastern Australia extreme rainfall to low-frequency Pacific variability. *Geophys Res Lett*: 1-6.
- Kruger, A.C., 1999. The influence of the decadal-scale variability of summer rainfall on the impact of El Nino and La Nina events in South Africa. *International Journal of Climatology*, 19(1): 59-68.
- Lavery, B., Joung, G., Nicholls, N., 1997. An extended high-quality historical rainfall dataset for Australia. *Aust Meteorol Mag*, 46(1): 27-38.
- Lima, C.H.R., Lall, U., 2009. Hierarchical Bayesian modeling of multisite daily rainfall occurrence: Rainy season onset, peak, and end. *Water Resources Research*, 45: W07422.
- Lima, C.H.R., Lall, U., 2010. Spatial scaling in a changing climate: A hierarchical bayesian model for non-stationary multi-site annual maximum and monthly streamflow. *Journal of Hydrology*, 383(3-4): 307-318.
- Meehl, G.A., Tebaldi, C., Teng, H., Peterson, T.C., 2007. Current and future US weather extremes and El Nino. *Geophys Res Lett*, 34(20).
- Micevski, T., Franks, S.W., Kuczera, G., 2006. Multidecadal variability in coastal eastern Australian flood data. *Journal of Hydrology*, 327(1-2): 219-225.
- Milly, P.C.D. et al., 2008. Climate change - Stationarity is dead: Whither water management? *Science*, 319(5863): 573-574.
- Nicholls, N., Kariko, A., 1993. East Australian rainfall events: Interannual variations, trends, and relationships with the Southern Oscillation. *Journal of Climate*, 6(6): 1141-1152.
- Ouarda, T.B.M.J., El-Adlouni, S., 2011. Bayesian Nonstationary Frequency Analysis of Hydrological Variables. *J Am Water Resour As*, 47(3): 496-505.

- Ouarda, T.B.M.J., Girard, C., Cavadias, G.S., Bobée, B., 2001. Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology*, 254(1): 157-173.
- Padoan, S.A., Ribatet, M., Sisson, S.A., 2010. Likelihood-based inference for Max-stable processes. *Journal of the American Statistical Association*, 105(489): 263-277.
- Pui, A., Sharma, A., Santoso, A., Westra, S., 2012. Impact of the El Niño-Southern Oscillation, Indian Ocean Dipole, and Southern Annular Mode on Daily to Subdaily Rainfall Characteristics in East Australia. *Mon Weather Rev*, 140(5): 1665-1682.
- Queensland Floods Commission of Inquiry, 2012 report (<http://www.floodcommission.qld.gov.au/publications/final-report>).
- Queensland Government, Operation Queenslander: The State Community, economic and environmental recovery and reconstruction plan 2011–2013 (<http://www.qldreconstruction.org.au/publications-guides/reconstruction-plans/state-plan>).
- Ray, S., Giese, B.S., 2012. Historical changes in El Nino and La Nina characteristics in an ocean reanalysis. *J Geophys Res-Oceans*, 117.
- Reich, B.J., Shaby, B.A., 2012. A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, 6(4): 1430-1451.
- Renard, B., 2011. A Bayesian hierarchical approach to regional frequency analysis RID G-1524-2011. *Water Resources Research*, 47: W11513.
- Renard, B., Garreta, V., Lang, M., 2006a. An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. *Water Resources Research*, 42(12): W12422.
- Renard, B., Lang, M., 2007. Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, 30(4): 897-912.
- Renard, B., Lang, M., Bois, P., 2006b. Statistical analysis of extreme events in a non-stationary context via a Bayesian framework: case study with peak-over-threshold data. *Stochastic Environmental Research and Risk Assessment*, 21(2): 97-112.
- Renard, B., Sun, X., Lang, M., 2013. Bayesian methods for non-stationary extreme value analysis. In: AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., Sorooshian, S. (Eds.), *Extremes in a changing climate: Detection, analysis and uncertainty*. Water Science and Technology Library. Springer Netherlands, pp. 39-95.
- Risbey, J.S., Pook, M.J., McIntosh, P.C., Wheeler, M.C., Hendon, H.H., 2009. On the remote drivers of rainfall variability in Australia. *Mon Weather Rev*, 137(10): 3233-3253.
- Sang, H.Y., Gelfand, A.E., 2009. Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16(3): 407-426.
- Schreck, C.J., Semazzi, F.H., 2004. Variability of the recent climate of eastern Africa. *International Journal of Climatology*, 24(6): 681-701.
- Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics*, 6(2): pp. 461-464.
- Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 8: 229-231.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.R., van der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of The Royal Statistical Society Series B-Statistical Methodology*, 64: 583-616.
- Thyer, M., Frost, A.J., Kuczera, G., 2006. Parameter estimation and model identification for stochastic models of annual hydrological data: Is the observed record long enough? *Journal of Hydrology*, 330(1-2): 313-328.
- Thyer, M., Leonard, M., Kavetski, D., Need, S., Renard, B., 2011. The open source RFortran library for accessing R from Fortran, with applications in environmental modelling. *Environ Modell Softw*, 26(2): 219-234.
- Vanheerden, J., Terblanche, D.E., Schulze, G.C., 1988. The Southern Oscillation and South-African Summer Rainfall. *J Climatol*, 8(6): 577-597.
- Westra, S., Sisson, S.A., 2011. Detection of non-stationarity in precipitation extremes using a max-stable process model. *Journal of Hydrology*, 406(1-2): 119-128.
- Wikle, C.K., Berliner, L.M., Cressie, N., 1998. Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5: 117-154.
- Willems, P., 2013a. Adjustment of extreme rainfall statistics accounting for multidecadal climate oscillations. *J Hydrol.*
- Willems, P., 2013b. Multidecadal oscillatory behaviour of rainfall extremes in Europe. *Climatic Change*: 1-14.
- Wu, R.G., Hu, Z.Z., Kirtman, B.P., 2003. Evolution of ENSO-related rainfall anomalies in East Asia. *Journal of Climate*, 16(22): 3742-3758.

914 **List of captions**

915 **Tables**

916 Table 1: Candidate models

917 Table 2: DIC difference between the regional models listed on the table *and*  $RA_{sy2\_}RA_{sy2}$  model

918 **Figures**

919 Figure 1: Schematic of the Local Model

920 Figure 2: Schematic of the Regional Model

921 Figure 3: Locations of the rain gauges. Summer rainfall totals are available in all 16 gauges. The  
922 blue dots are the gauges in which daily rainfall data are available, which are used to compute the  
923 summer daily maxima.

924 Figure 4 Boxplot of the posterior distribution of (a)  $\mu_1^-$  (El Niño) and (b)  $\mu_1^+$  (La Niña) for each site  
925 for the summer rainfall totals

926 Figure 5: P-value of zero of (a)  $\mu_1^-$  (El Niño) and (b)  $\mu_1^+$  (La Niña) for each site for the summer  
927 rainfall totals. A p-value smaller than 10% (blue dots) indicates that the parameter is significantly  
928 larger than 0.

929 Figure 6: Quantiles of summer total rainfall with respect to SOI value for site 16. The blue, red and  
930 green lines are respectively the 0.05, 0.5 and 0.99 quantiles with 90% credibility intervals (grey  
931 shaded areas). Black dots are the observations with respect to the SOI value of each year.

932 Figure 7: Probability-Probability plot of summer maximum daily rainfall with (a) local model  
933  $LA_{sy1\_}LA_{sy1}$  and (b) regional model  $RA_{sy1\_}RA_{sy1}$ . Each colour presents one site.

934 Figure 9: summer maximum daily rainfall. P-value of zero of (a)  $\tilde{\mu}_{loc_1}^{(s)}$  and (b)  $\tilde{\sigma}_{loc_1}^{(s)}$  of each site for the  
935 symmetric model  $LSym\_LSym$ , and p-value of zero of (c)  $\mu_{loc_1}^{+(s)}$  and (d)  $\sigma_{loc_1}^{+(s)}$  of each site (during La  
936 Niña episode) for the asymmetric model  $LA_{sy1\_}LA_{sy1}$ . A p-value smaller than 10% (blue dots)  
937 indicates that the parameter is significantly larger than 0.

938 Figure 10: P-value of zero for the slope of 1 in 100 year summer maximum daily rainfall with (a)  
939 the symmetric model  $LSym\_LSym$  and (b) the asymmetric model  $LAsyI\_LAsyI$  during the La Niña  
940 episode.

941 Figure 11: 1 in 100 summer maximum daily rainfall at site 16. The blue line is based on the  
942 stationary model ( $L\_Stat\_Stat$ ). The green and red lines are respectively based on the symmetric  
943 ( $LSym\_LSym$ ) and asymmetric ( $LAsyI\_LAsyI$ ) models. The solid lines are median and areas inside  
944 the dashed line are 90% credibility intervals of each model. Black dots are the observations with  
945 respect to the SOI value of each year.

946 Figure 12: Boxplot of the posterior distribution of location parameter  $\mu_1^+$  ( $\mu_{loc_1}^+$  in local model  
947  $LAsyI\_LAsyI$  of each site and  $\mu_{reg_1}^+$  in regional model  $RAsyI\_RAsyI$ ).

948 Figure 13: Boxplot of the posterior distribution of the regional parameters of model  $RAsyI\_RAsyI$   
949 for the summer maximum daily rainfall

950 Figure 14: 1 in 100 year summer maximum daily rainfall with local ( $L\_Stat\_Stat$  &  $LAsyI\_LAsyI$ )  
951 and regional ( $RAsyI\_RAsyI$ ) models at site 16. The blue line is based on the stationary model  
952 ( $L\_Stat\_Stat$ ). The red and green lines are respectively based on the local ( $LAsyI\_LAsyI$ ) and  
953 regional ( $RAsyI\_RAsyI$ ) models. The solid lines are median and areas inside the dashed line are  
954 90% credibility intervals of each model. Black dots are the observations with respect to the SOI  
955 value of each year.

956 Figure 15: DIC value for the models in Table 1 for the summer maximum daily rainfall.  
957  $L\_Stat\_Stat$ ,  $LSym\_LSym$  and  $LAsyI\_LAsyI$  are local models.  $R\_Stat\_Stat$ ,  $RSym\_RSym$ ,  
958  $RAsyI\_RAsyI$ ,  $RAsy2\_RAsy2$ ,  $RAsyI\_Stat$  and  $RAsy2\_Stat$  are regional models.



959 Table 1: Candidate models

	Regression functions			Regression
Models	Regression functions for $\mu(s,t)$	Regression functions for $\sigma(s,t)$	for $\xi(s,t)$	
Local models				
L_Stat_Stat	$\widehat{\mu}_{loc}^{(s)}$	$\widehat{\sigma}_{loc}^{(s)}$	$\widehat{\xi}_{loc_0}^{(s)}$	
LSym_LSym	$\tilde{\mu}_{loc_0}^{(s)} + \tilde{\mu}_{loc_1}^{(s)} * SOI(t)$	$\tilde{\sigma}_{loc_0}^{(s)} + \tilde{\sigma}_{loc_1}^{(s)} * SOI(t)$	$\tilde{\xi}_{loc_0}^{(s)}$	
LAsy1_LAsy1	$\begin{cases} \mu_{loc_0}^{(s)} + \mu_{loc_1}^{+(s)} * SOI(t); SOI(t) < 0 \\ \mu_{loc_0}^{(s)} + \mu_{loc_1}^{-(s)} * SOI(t); SOI(t) > 0 \end{cases}$	$\begin{cases} \sigma_{loc_0}^{(s)} + \sigma_{loc_1}^{-(s)} * SOI(t); SOI(t) < 0 \\ \sigma_{loc_0}^{(s)} + \sigma_{loc_1}^{+(s)} * SOI(t); SOI(t) > 0 \end{cases}$	$\xi_{loc_0}^{(s)}$	
Regional models				
R_Stat_Stat	$\widehat{\mu}_{loc}^{(s)}$	$\widehat{\sigma}_{loc}^{(s)}$	$\widehat{\xi}_{reg}$	
RSym_RSym	$\tilde{\mu}_{loc}^{(s)} + \tilde{\mu}_{reg} * SOI(t)$	$\tilde{\sigma}_{loc}^{(s)} + \tilde{\sigma}_{reg} * SOI(t)$	$\tilde{\xi}_{reg}$	
RAsy1_RAsy1	$\begin{cases} \mu_{loc_0}^{(s)} + \mu_{reg_1}^{-(s)} * SOI(t); SOI(t) < 0 \\ \mu_{loc_0}^{(s)} + \mu_{reg_1}^{+(s)} * SOI(t); SOI(t) > 0 \end{cases}$	$\begin{cases} \sigma_{loc_0}^{(s)} + \sigma_{reg_1}^{-(s)} * SOI(t); SOI(t) < 0 \\ \sigma_{loc_0}^{(s)} + \sigma_{reg_1}^{+(s)} * SOI(t); SOI(t) > 0 \end{cases}$	$\xi_{reg}$	
RAsy2_RAsy2	$\begin{cases} \mu_{loc_0}^{(s)}; SOI(t) < 0 \\ \mu_{loc_0}^{(s)} + \mu_{reg_1}^{+(s)} * SOI(t); SOI(t) > 0 \end{cases}$	$\begin{cases} \sigma_{loc_0}^{(s)}; SOI(t) < 0 \\ \sigma_{loc_0}^{(s)} + \sigma_{reg_1}^{+(s)} * SOI(t); SOI(t) > 0 \end{cases}$	$\xi_{reg}$	
RAsy1_Stat	$\begin{cases} \mu_{loc_0}^{(s)} + \mu_{reg_1}^{-(s)} * SOI(t); SOI(t) < 0 \\ \mu_{loc_0}^{(s)} + \mu_{reg_1}^{+(s)} * SOI(t); SOI(t) > 0 \end{cases}$	$\sigma_{loc_0}^{(s)}$	$\xi_{reg}$	
RAsy2_Stat	$\begin{cases} \mu_{loc_0}^{(s)}; SOI(t) < 0 \\ \mu_{loc_0}^{(s)} + \mu_{reg_1}^{+(s)} * SOI(t); SOI(t) > 0 \end{cases}$	$\sigma_{loc_0}^{(s)}$	$\xi_{reg}$	

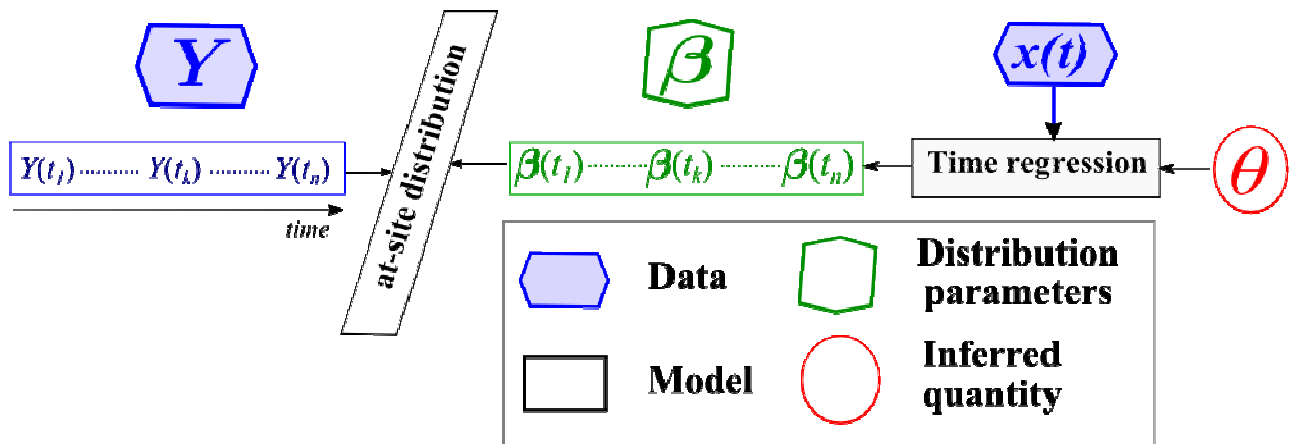
960

961

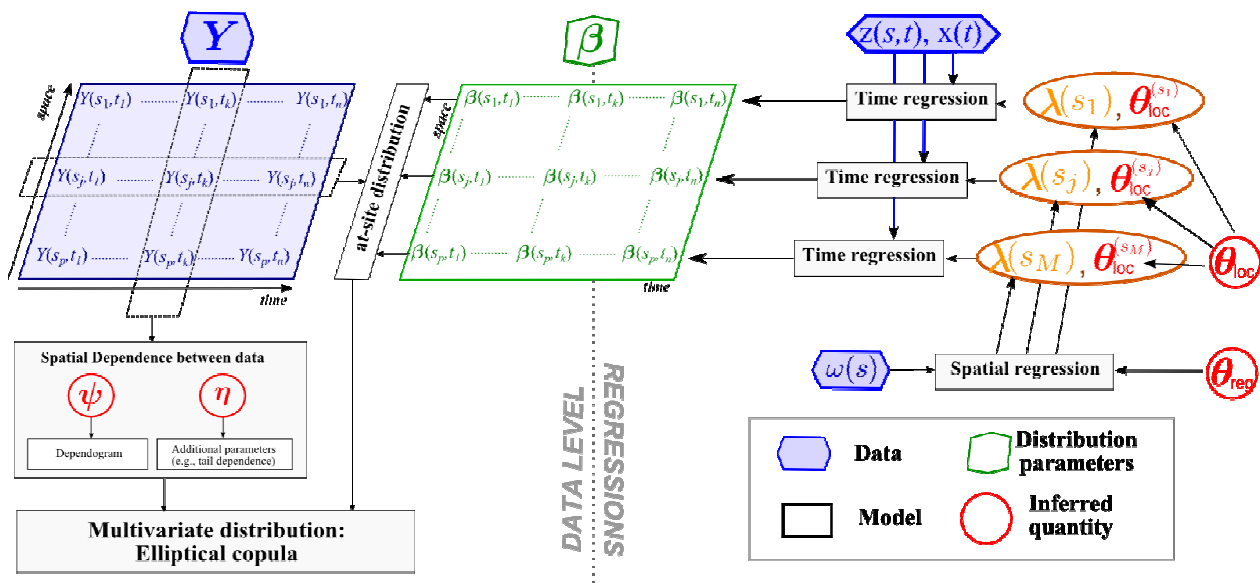
962 Table 2: DIC difference between the regional models listed on the table and *RAsy2\_RAasy2* model

R_Stat_	RSym_	RAasy1_	RAasy2_	RAasy1_	RAasy2_
Stat	RSym	RAasy1	RAasy2	Stat	Stat
11.2	4.4	3.2	0	7.9	6.1

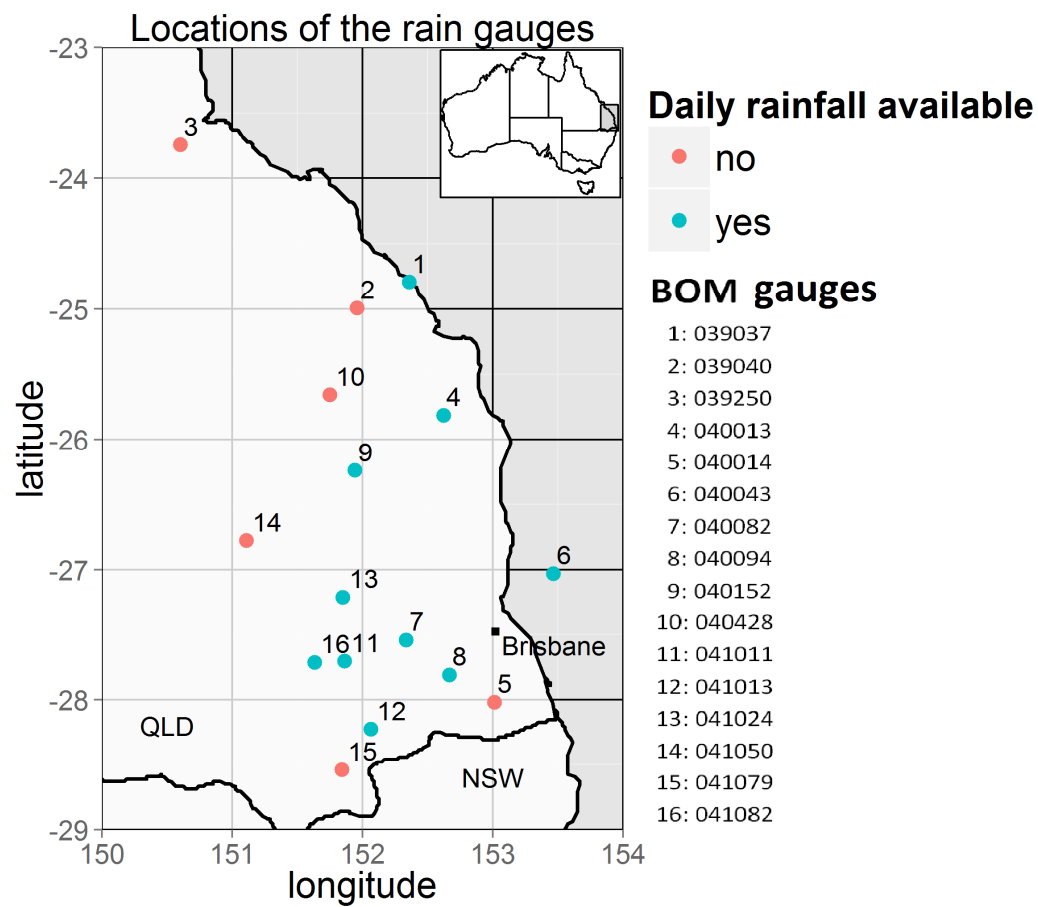
963



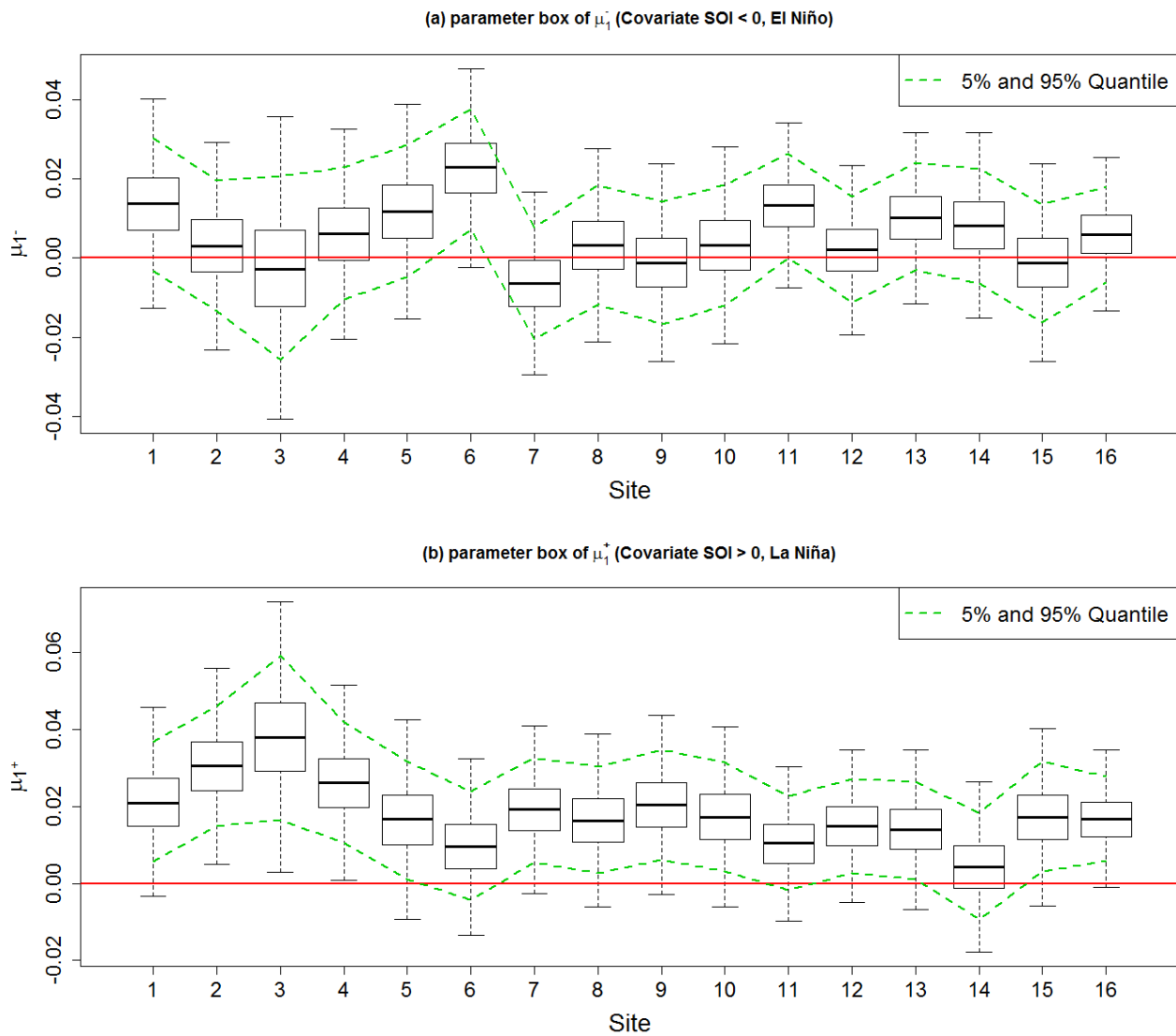
964 **Figure 1: Schematic of the Local Model**



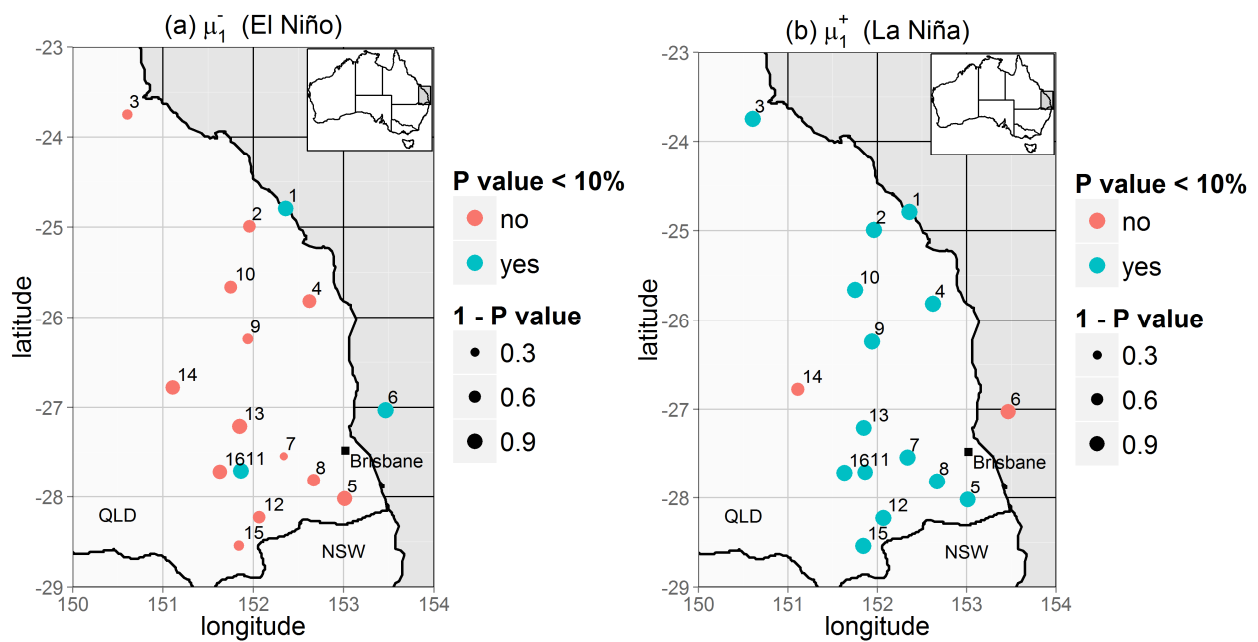
965 **Figure 2: Schematic of the Regional Model**



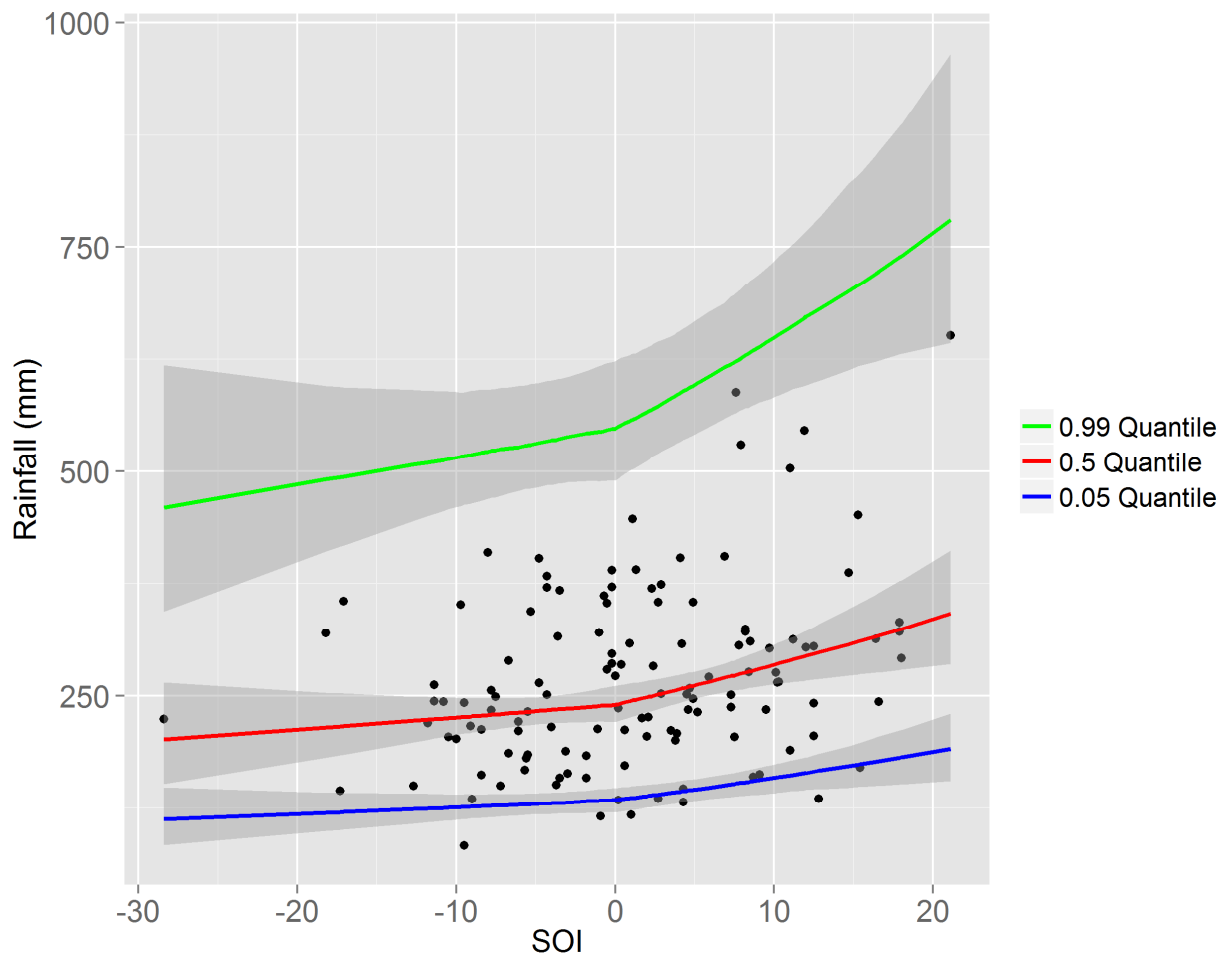
966 **Figure 3: Locations of the rain gauges. Summer rainfall totals are available in all 16 gauges.**  
967 **The blue dots are the gauges in which daily rainfall data are available, which are used to**  
968 **compute the summer daily maxima.**



969 **Figure 4** Boxplot of the posterior distribution of (a)  $\mu_1^-$  (El Niño) and (b)  $\mu_1^+$  (La Niña) for  
970 each site for the summer rainfall totals

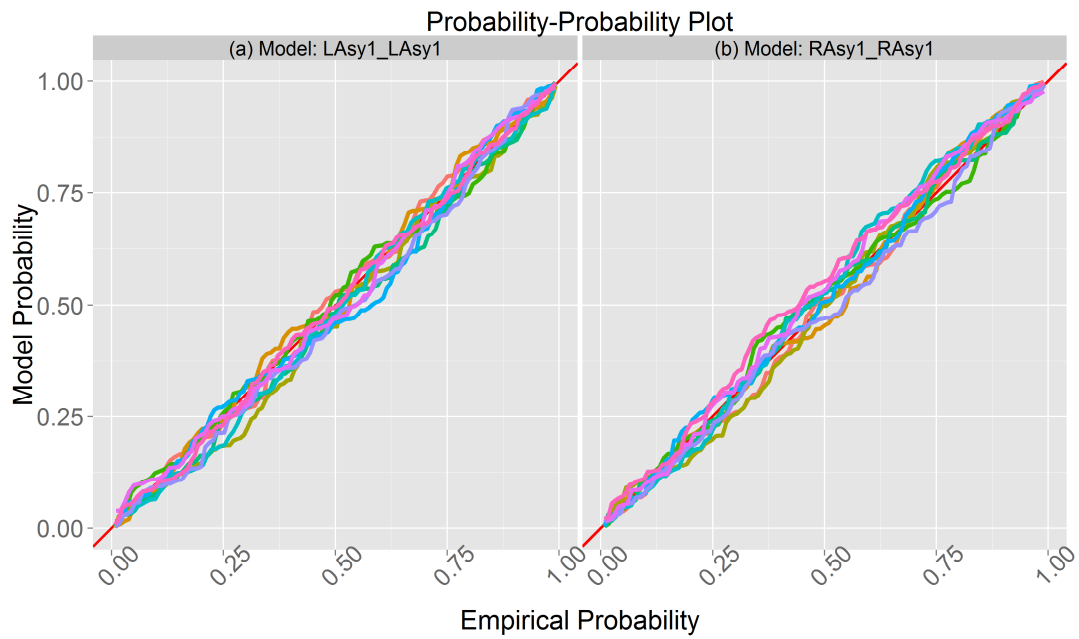


971 **Figure 5: P-value of zero of (a)  $\mu_1^-$  (El Niño) and (b)  $\mu_1^+$  (La Niña) for each site for the**  
 972 **summer rainfall totals. A p-value smaller than 10% (blue dots) indicates that the parameter**  
 973 **is significantly larger than 0.**

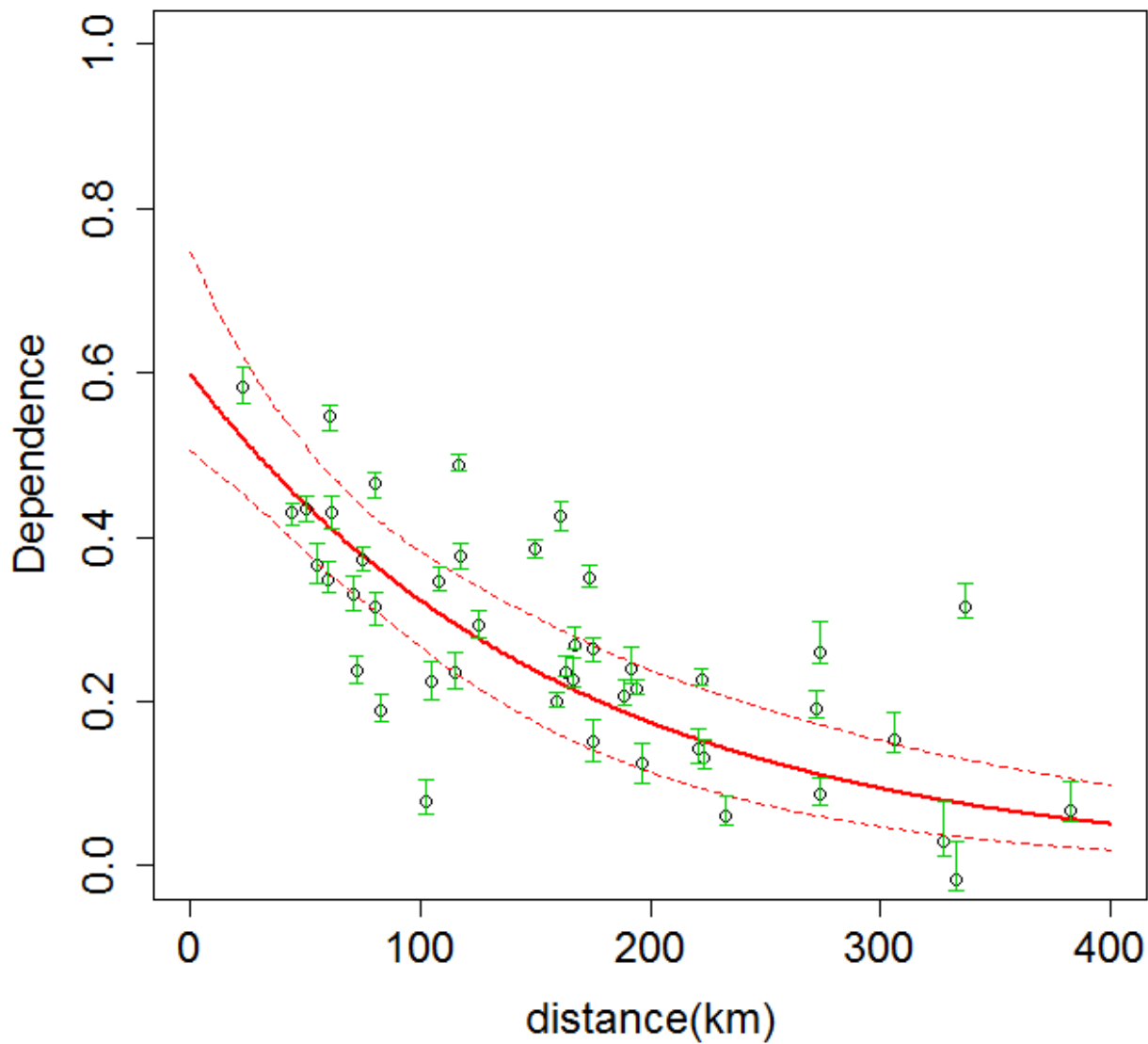


974 **Figure 6: Quantiles of summer total rainfall with respect to SOI value for site 16. The blue,**  
 975 **red and green lines are respectively the 0.05, 0.5 and 0.99 quantiles with 90% credibility**  
 976 **intervals (grey shaded areas). Black dots are the observations with respect to the SOI value of**  
 977 **each year.**

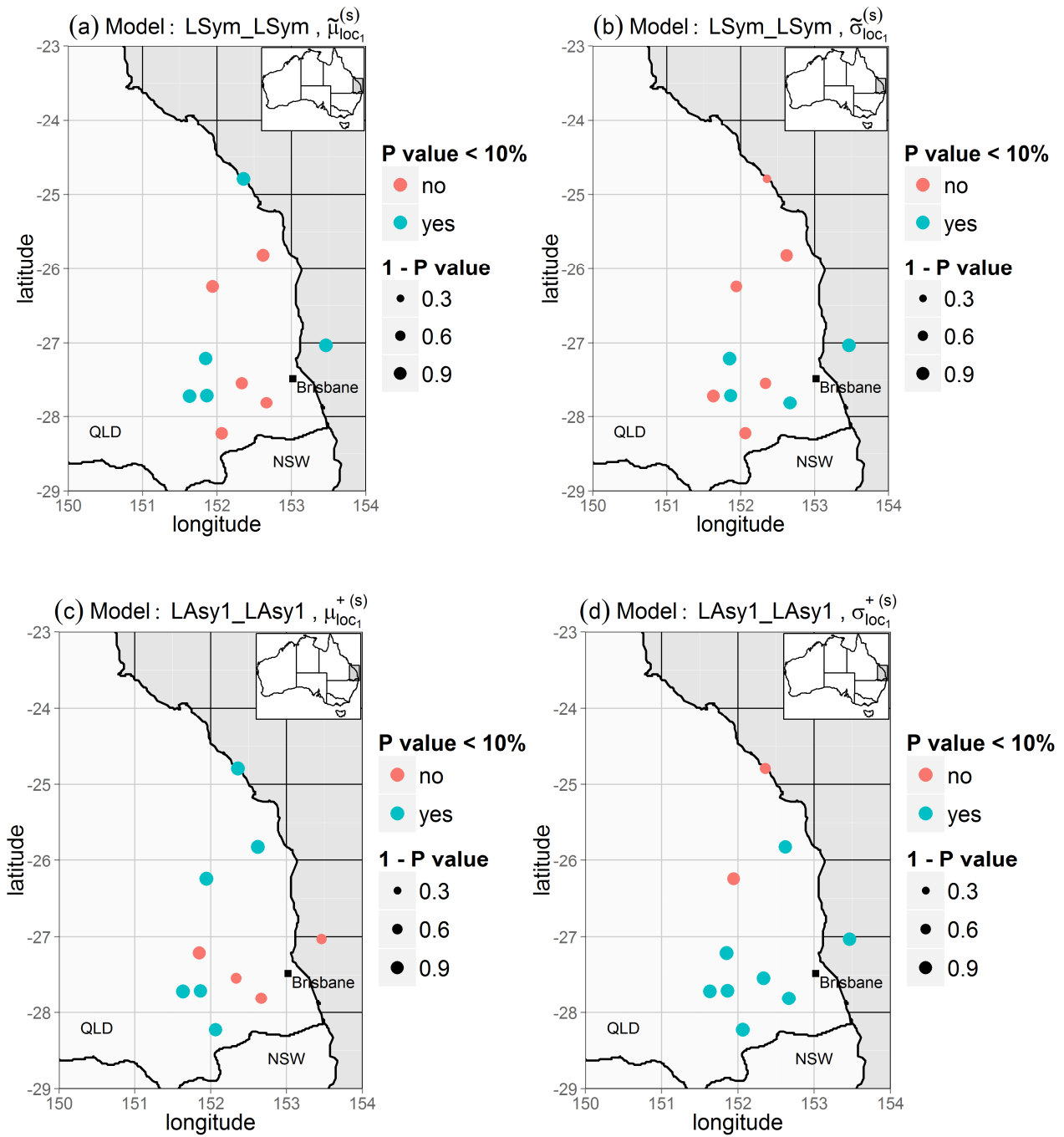




978 **Figure 7: Probability-Probability plot of summer maximum daily rainfall with (a) local model**  
979 ***LAsy1\_LAsy1* and (b) regional model *RAsy1\_RAsy1*. Each colour presents one site.**

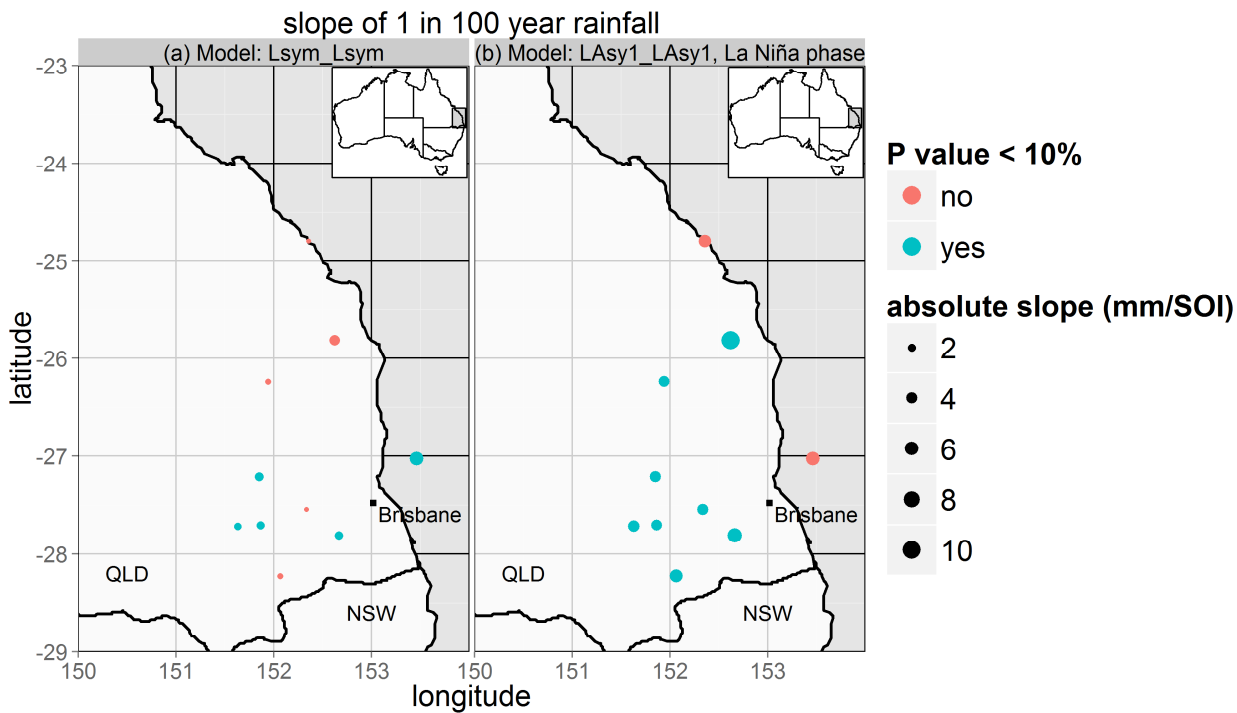


980 **Figure 8: Dependence-distance relationship of the transformed data. In the time-varying**  
981 **context, data  $y$  are transformed into Gaussian quantiles  $u$  according to Eq (27). Red lines are**  
982 **the estimated Dependence-distance with 90% credibility interval. Black circles are empirical**  
983 **correlation between transformed data, and their 90% credibility interval due to the marginal**  
984 **transformation is presented in green.**

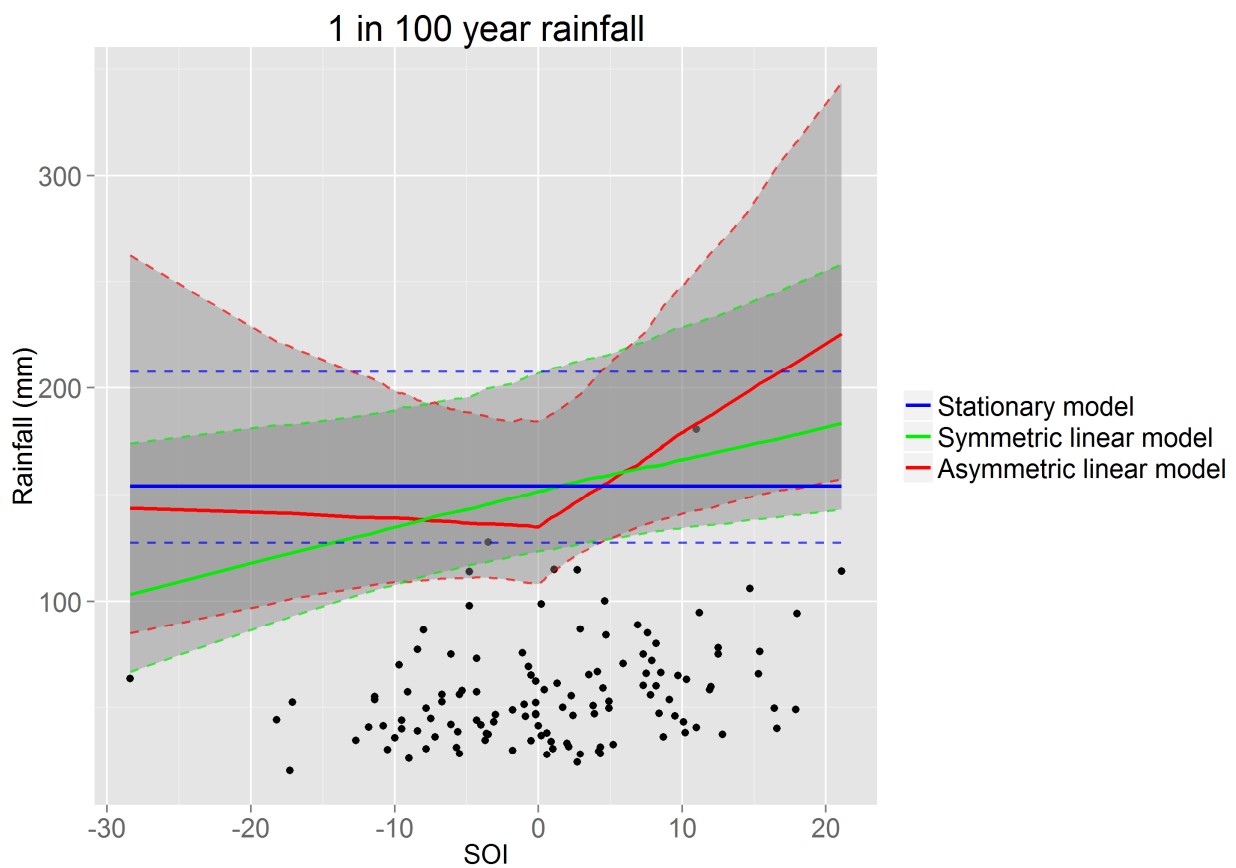


985 **Figure 9: summer maximum daily rainfall. P-value of zero of (a)  $\tilde{\mu}_{loc_1}^{(s)}$  and (b)  $\tilde{\sigma}_{loc_1}^{(s)}$  of each site**  
 986 **for the symmetric model  $LSym\_LSym$ , and p-value of zero of (c)  $\mu_{loc_1}^{+(s)}$  and (d)  $\sigma_{loc_1}^{+(s)}$  of each site**

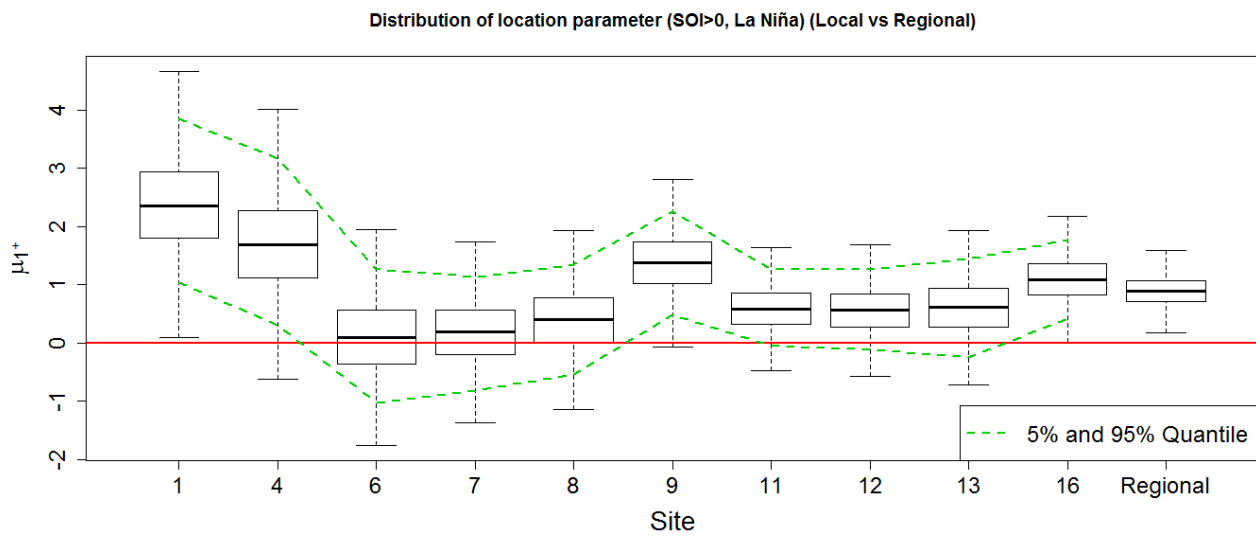
987 (during La Niña episode) for the asymmetric model  $L_{AsyI\_L_{AsyI}}$ . A p-value smaller than  
988 10% (blue dots) indicates that the parameter is significantly larger than 0.



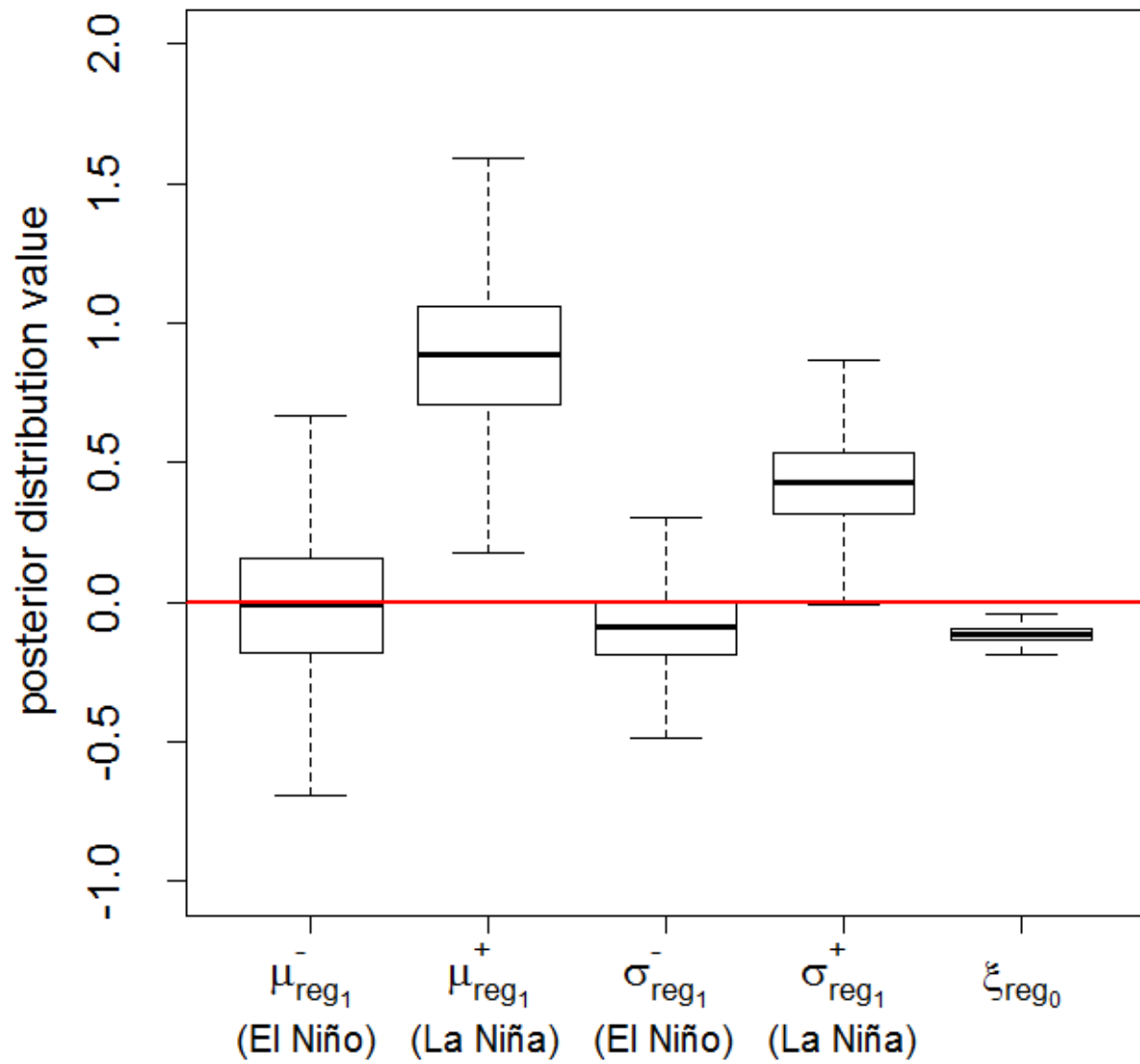
989 **Figure 10: P-value of zero for the slope of 1 in 100 year summer maximum daily rainfall with**  
990 **(a) the symmetric model *LSym\_LSsym* and (b) the asymmetric model *LAsy1\_LAsy1* during the**  
991 **La Niña episode.**



992 **Figure 11: 1 in 100 summer maximum daily rainfall at site 16. The blue line is based on the**  
 993 **stationary model ( $L_{Stat\_Stat}$ ). The green and red lines are respectively based on the**  
 994 **symmetric ( $LSym\_LSym$ ) and asymmetric ( $LAsyI\_LAsyI$ ) models. The solid lines are median**  
 995 **and areas inside the dashed line are 90% credibility intervals of each model. Black dots are**  
 996 **the observations with respect to the SOI value of each year.**

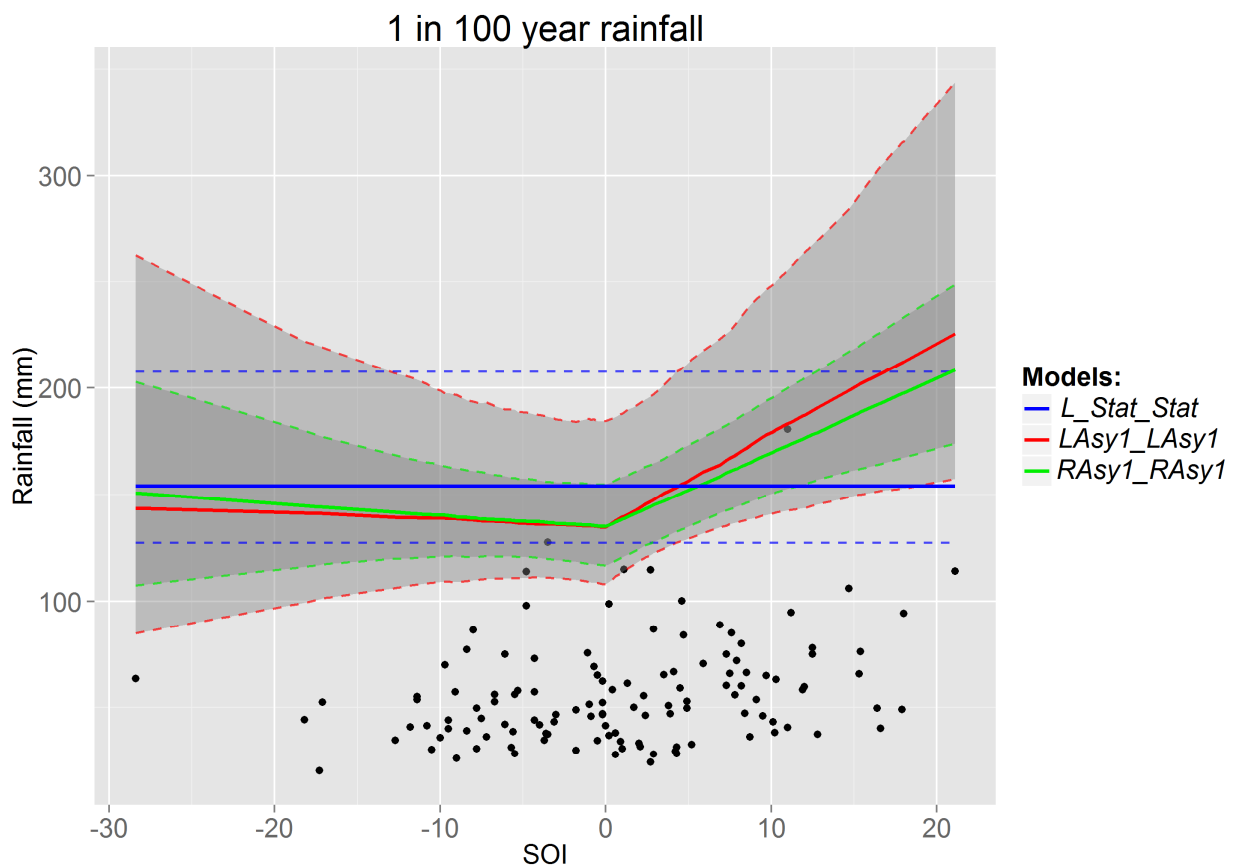


997 **Figure 12: Boxplot of the posterior distribution of location parameter  $\mu_1^+$  ( $\mu_{loc_1}^+$  in local model**  
998  **$L_{AsyI\_L_{AsyI}}$  of each site and  $\mu_{reg_1}^+$  in regional model  $R_{AsyI\_R_{AsyI}}$ ).**

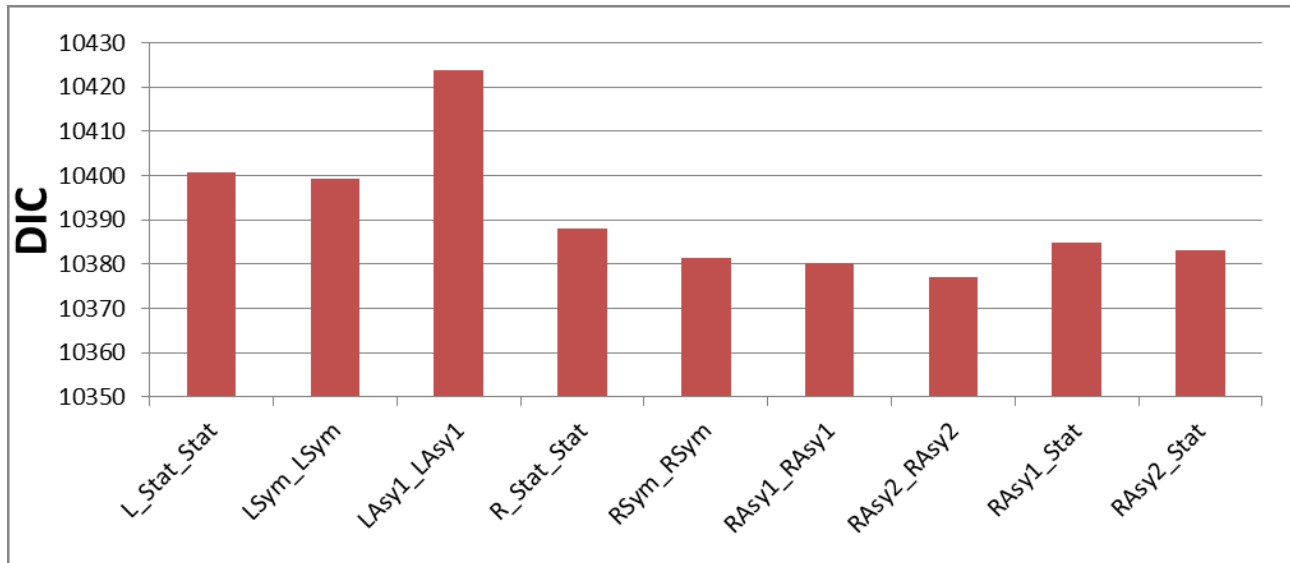


999 **Figure 13: Boxplot of the posterior distribution of the regional parameters of model**  
1000 ***RAsyI\_RAyI* for the summer maximum daily rainfall**





1001 **Figure 14: 1 in 100 year summer maximum daily rainfall with local (*L\_Stat\_Stat* &**  
1002 ***L\_Asy1\_L\_Asy1*) and regional (*R\_Asy1\_R\_Asy1*) models at site 16. The blue line is based on the**  
1003 **stationary model (*L\_Stat\_Stat*). The red and green lines are respectively based on the local**  
1004 **(*L\_Asy1\_L\_Asy1*) and regional (*R\_Asy1\_R\_Asy1*) models. The solid lines are median and areas**  
1005 **inside the dashed line are 90% credibility intervals of each model. Black dots are the**  
1006 **observations with respect to the SOI value of each year.**



1007 **Figure 15: DIC value for the models in Table 1 for the summer maximum daily rainfall.**  
1008 *L\_Stat\_Stat*, *LSym\_LSym* and *LAsy1\_LAsy1* are local models. *R\_Stat\_Stat*, *RSym\_RSym*,  
1009 *RAsy1\_RAsy1*, *RAsy2\_RAsy2*, *RAsy1\_Stat* and *RAsy2\_Stat* are regional models.