

Stochastic daily rainfall generation on tropical islands with complex topography

Lionel Benoit^{1,2,5}, Lydie Sichoix², Alison D. Nugent³, Matthew P. Lucas⁴, Thomas W. Giambelluca¹

¹Water Resources Research Center, University of Hawai‘i at Mānoa, 96822 Honolulu, Hawai‘i, USA

²GePaSud Laboratory, University of French Polynesia, 98702 Faa’a - Tahiti, French Polynesia

³Department of Atmospheric Sciences, School of Ocean and Earth Science and Technology, University of Hawai‘i at Mānoa, 96822 Honolulu, Hawai‘i, USA

⁴Department of Geography, University of Hawai‘i at Mānoa, 96822 Honolulu, Hawai‘i, USA

⁵Now at Biostatistics and Spatial Processes (BioSP), INRAE, 84914 Avignon Cedex 9, France

Correspondence to: Lionel Benoit (benoitlionel2@gmail.com)

Abstract. Stochastic rainfall generators are probabilistic models of rainfall space-time behavior. During parameterization and calibration, they allow the identification and quantification of the main modes of rainfall variability. Hence, stochastic rainfall models can be regarded as probabilistic conceptual models of rainfall dynamics.

As with most conceptual models in Earth Sciences, the performance of stochastic rainfall models strongly relies on their adequacy in representing the rain process at hand. On tropical islands with high elevation topography, orographic rain enhancement challenges most existing stochastic models because it creates localized precipitations with strong spatial gradients, which break down the stationarity of rain statistics.

To allow for stochastic rainfall modeling on tropical islands, despite non-stationarity of rain statistics, we propose a new stochastic daily multi-site rainfall generator specifically for areas with significant orographic effects. Our model relies on a preliminary classification of daily rain patterns into rain types based on rainfall space and intensity statistics, and sheds new light on rainfall variability at the island scale. Within each rain type, the distribution of rainfall through the island is modeled by combining a non-parametric resampling of past analogs of a latent field describing the spatial distribution of rainfall, and a parametric Gamma transform function describing rain intensity.

When applied to the stochastic simulation of rainfall on the islands of O‘ahu (Hawai‘i, United States of America) and Tahiti (French Polynesia) in the tropical Pacific, the proposed model demonstrates good skills in jointly simulating site-specific and island-scale rain statistics. Hence, it provides a new tool for stochastic impact studies in tropical islands, in particular for watershed water resources management.

1 Introduction

Stochastic rainfall generators are probabilistic tools aiming at simulating synthetic rainfall that mimic as closely as possible the statistical signature of rain observations [Richardson, 1981] [Wilks and Wilby, 1999] [Ailliot et al., 2015]. More specifically, stochastic rainfall modeling consists of statistical learning (i.e., inference) of the joint space-time probability

32 density function (pdf) of rainfall at all sites and times of interest, and sampling this pdf to efficiently generate synthetic rainfall.
33 The ability of stochastic rainfall generators to emulate long and realistic rainfall sequences makes them an appropriate tool for
34 the simulation of design storms [Niemi *et al.*, 2016]. Simulated rainfall can then be used as inputs for impact models assessing
35 the effects of rainfall on different environmental processes including hydrology [Paschalis *et al.*, 2014], water resources
36 [Cappelaere *et al.*, 2020], geomorphology [Peleg *et al.*, 2020], and agronomy [Mavromatis and Hansen, 2001]. The
37 probabilistic approach followed by stochastic rainfall generators enables a comprehensive study of rainfall variability and, in
38 turn, the assessment of uncertainty propagation along the whole modeling chain [Gabellani *et al.*, 2007]. This makes stochastic
39 rainfall generation a key tool for management of rain-induced risk, in particular, for flood [Caseri *et al.*, 2016] and drought
40 risks [Supit *et al.*, 2012]. In addition, the focus of stochastic rainfall models on the statistical signature of rainfall creates new
41 ways to characterize rainfall space-time behavior [Marra and Morin, 2018], and assess the impact of rainfall variability on the
42 hydrosphere [Morin *et al.*, 2019]. Finally, when conditioned to climate model outputs, stochastic rainfall generation can be
43 used for the downscaling of future precipitation projections [Maraun *et al.*, 2010], resulting in local-scale and high-resolution
44 scenarios of the possible evolution of rainfall in the context of climate change [Jha *et al.*, 2014] [Volosciuk *et al.*, 2017].

45 To enable fast and computationally efficient simulations, and thereby allow for investigation of rainfall variability
46 and associated uncertainty through the simulation of large ensembles, stochastic rainfall generators adopt an empirical
47 approach that bypasses the detailed physical modeling of rain generation processes [Bauer *et al.*, 2015]. To avoid the pitfall of
48 physically unrealistic simulations, stochastic rainfall models embed a significant part of our conceptual knowledge about
49 rainfall behavior in their parameterization (i.e., they implement statistical relationships that reflect as closely as possible the
50 physical processes at work). However, rainfall properties [Krajewski *et al.*, 2003] and, in turn, the performance of stochastic
51 rainfall generators [Breinl *et al.*, 2017] [Vu *et al.*, 2018] strongly depend on the climate of the area of interest. Hence, different
52 models have been proposed for different climates with each model focusing on a specific aspect of rainfall, for instance: rainfall
53 seasonality in monsoonal climates [Greene *et al.*, 2011]; rainfall spatial-temporal correlation in temperate climates [Paschalis
54 *et al.*, 2013]; or rainfall occurrence and extreme intensities in arid regions [Wilcox *et al.*, 2021].

55 On high tropical islands, or islands with high elevations and significant topography, rainfall is strongly location
56 dependent due to complex interactions between atmospheric circulation and island topography, which trigger different
57 mechanisms of orographic rain enhancement [Foresti and Pozdnoukhov, 2012] [Houze, 2012]. At monthly to annual scales,
58 the effect of orographic lifting of relatively steady trade winds generates well defined rain patterns [Lyons, 1982]. In these
59 patterns, highlands are usually wetter than lowlands, windward slopes wetter than leeward sides, and in case of successive
60 mountain ridges the first to be reached by the wet air masses is the wettest [Giambelluca *et al.*, 2013] [Laurent *et al.*, 2019].
61 To this first order quasi-static picture is added the important variability of daily rainfall patterns associated with processes
62 ranging from synoptic-scale disturbances [Hopuare *et al.*, 2018] [Longman *et al.*, 2021] to large-scale atmospheric circulations
63 [Hopuare *et al.*, 2015] [Frazier *et al.*, 2018] [Brown *et al.*, 2020]. Large deviations from the long term rainfall patterns are
64 thus observed, and usually dry leeward slopes can become the wettest part of the island, for example during Kona storms
65 (seasonal cyclones) in the Hawai‘i archipelago [Caruso and Businger, 2006].

66 Orographic effects lead to non-stationary (i.e., non-homogeneous) rain statistics in both space and time, which
67 challenges most existing stochastic rainfall models [Nerini et al., 2017]. In the context of high tropical islands, the first
68 difficulty arises from the long term patterns of orographic rain enhancement that create non-stationarities in space. Rainfall
69 generators must therefore account not only for the correlation between locations [Leblois and Creutin, 2013] [Paschalis et al.,
70 2013], but also for the location of the rain within the island. When orographic effects are directly related to topography,
71 Generalized Linear Models (GLMs) have been leveraged to account for spatial rain patterns by linking model parameters to
72 local topographic information such as altitude or slope aspect [Ambrosino et al., 2014] [Chandler, 2020]. In tropical islands
73 however, the complexity of the rain-topography relationships (e.g., successive mountain ridges becoming drier at similar
74 altitude) hinder the direct regression of model parameters on topographic data. An interesting option to overcome the non-
75 uniqueness of topography-rainfall relationships is to interpolate the parameters of the rainfall model in space in order to let
76 data inform non-stationarity [Kleiber et al., 2012] [Bennett et al., 2018]. This however increases model complexity, and
77 therefore requires training datasets with high spatial resolution, which are not yet available in most target areas [Benoit et al.,
78 2021]. Finally, the difficulty of accounting for orographic effects in the context of high tropical islands is further increased by
79 the temporal variability of rainfall patterns (e.g., wet leeward slopes during Kona storms), which calls for the development of
80 models able to identify and capture the main modes of variability of orographic effects over time.

81 A rainfall generator compatible with marine tropical climates and complex rain-topography interactions encountered
82 in tropical islands is, to the best of our knowledge, still missing in the toolbox of stochastic rainfall generators. To fill this gap,
83 the present paper proposes a new rainfall model dedicated to high tropical islands with significant and complex topography,
84 which aims to account for both the long-term quasi-static patterns of rain accumulation and the day-to-day fluctuations of the
85 rainfall spatial distribution. Specifically, the goal is to develop a daily resolution multi-site stochastic rainfall generator able to
86 simulate: (1) site specific rain occurrence, persistence, intensity and seasonality; (2) spatial patterns of daily rain accumulation;
87 and (3) areal rain statistics at the island scale.

88 To achieve these objectives, the remainder of the article is structured as follows. Section 2 briefly reviews the main
89 features of tropical island rainfall and describes our stochastic rainfall model. Section 3 illustrates the performance of the model
90 for the island of O‘ahu (Hawai‘i, USA) in the tropical Pacific, and a similar test study is repeated in supplementary material
91 for the island of Tahiti (French Polynesia) to demonstrate the versatility of the model. Finally, section 4 discusses how the
92 focus on orographic rain enhancement has influenced the design of the model and provides concluding remarks.

93 **2 Data and methods**

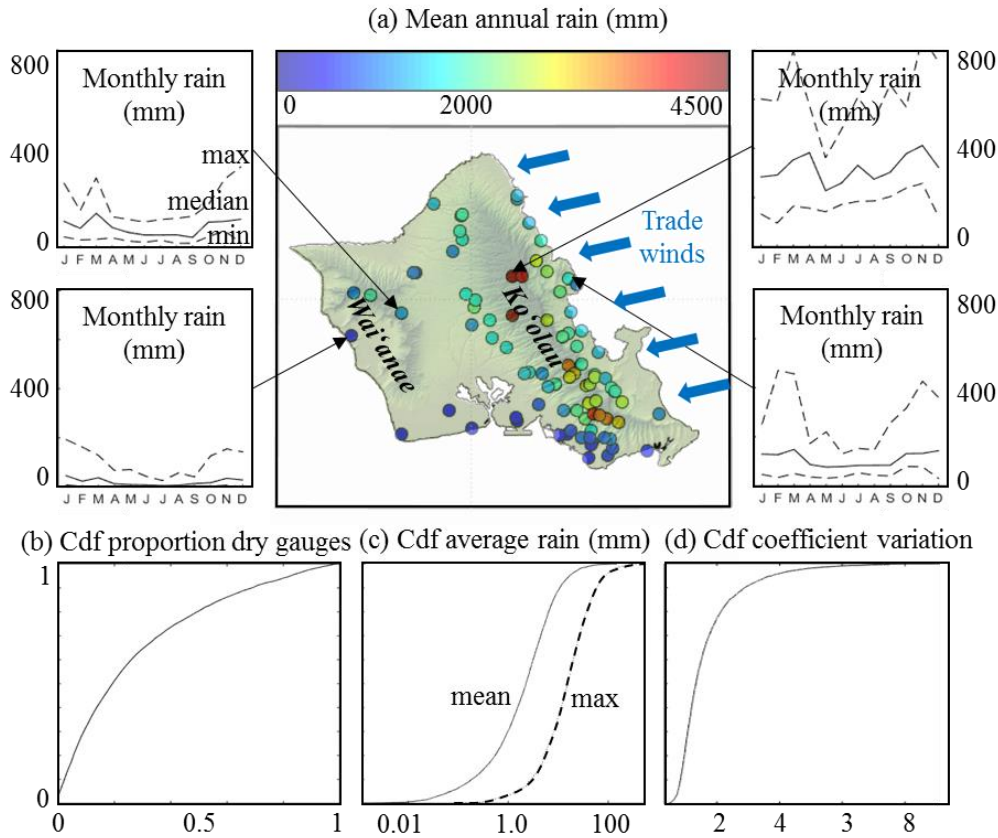
94 **2.1 Rainfall features of interest**

95 Because stochastic rainfall models are data-driven, their structure depends on the rain features one wants to reproduce
96 in simulations. Hence, the identification of the main features of daily rainfall in high tropical islands is a prerequisite for the
97 design of the present model. For illustration purposes, we focus throughout the main text on the island of O‘ahu, Hawai‘i (lon

98 = 158°W, lat = 21.5°N, area = 1545 km², max altitude = 1220 m). The available rain gauge observation dataset consists of
99 daily records from a network of 86 rain gauges spread over the island (Fig. 1a), and covers a 20-year period 1991–2011. It
100 corresponds to a compilation of quality controlled and gap-filled daily observations [Longman *et al.*, 2018]. Gap-filling was
101 performed using the normal ratio method [Paulhus and Kohler, 1952], and only stations with less than 5% gap-filled data are
102 kept for this study in order to minimize the impact of gap-filling on our results. To contextualize the observed rain patterns,
103 several meteorological covariates (e.g., pressure, temperature, humidity and wind) are investigated at the island scale. We use
104 the ERA5 reanalysis dataset [Hersbach *et al.*, 2018] at 12:00 PM HST to inform these covariates and average the values of the
105 12 grid cells (pixel size = 0.25° x 0.25°) encompassing the island of O‘ahu.

106 Figure 1 displays the main features of rainfall over the island of O‘ahu. It shows the strong impact of trade wind
107 induced orographic rain enhancement on the spatial distribution of annual precipitation (Fig. 1a), with windward (northeast)
108 sides significantly wetter than leeward (southwest) ones, and highlands generally wetter than lowlands. Note some important
109 details of this annual rainfall pattern, for example that the rain maximum is observed leeward of the main crest of the Ko‘olau
110 range, and that the Ko‘olau mountains are significantly wetter than the Wai‘anae range despite higher elevation. In addition to
111 prevailing orographic rainfall triggered by the interactions of trade winds with island topography (east-northeasterly trade
112 winds blow more than 280 days per year over the Hawaiian archipelago [Longman *et al.*, 2015]), the island of O‘ahu also
113 experiences spatially widespread rain events, mostly triggered by regional atmospheric disturbances such as cold fronts
114 originating from mid-latitudes and Kona storms (seasonal cyclones) [Longman *et al.*, 2021]. These atmospheric disturbances
115 mostly occur during (boreal) winter, which corresponds to the local rainy season (spanning from October–March). They
116 represent the main source of precipitation for dry leeward locations and are responsible for the enhanced seasonality of rain
117 accumulation in these areas (Fig. 1a).

118 The diversity of rain generation mechanisms (e.g., orographic lifting, cold fronts, or Kona lows) coupled with the
119 steep island topography of volcanic origin result in a complex distribution of rainfall in space and time, which produces highly
120 variable island-scale rain statistics (i.e., statistics summarizing rain behavior throughout the island for a given day). Figure 1
121 b–d shows that at the scale of the island of O‘ahu, daily rainfall is strongly intermittent in space (only 3% of the days record
122 rain at all gauge locations, and half of the time at least 20% of the gauges measure no rain, Fig. 1b), highly skewed (island-
123 scale rain accumulation average < 2.25mm/day 50% of the time, but island-scale maximum accumulation
124 > 15mm/day 50% of the time and reaches 500 mm/day, Fig 1c), and strongly variable in space (coefficient of variation > 1.3
125 50% of the time, and > 2.9 10% of the time).



126

127

Figure 1: Main features of rainfall observed over the island of O'ahu. (a) Mean annual rainfall (central panel) and seasonality of rain accumulation for four specific rain gauges (outer panels). (b) Cumulative distribution function (cdf) of the proportion of gauges measuring no rain for a given day. (c) Cdf of the mean and maximum daily rain accumulation computed over the whole observation network (abscissa is in log-scale). (d) Cdf of the coefficient of variation (i.e., standard deviation/mean) of daily rain accumulation throughout the rain gauge network.

128

129

130

131

132

2.2 Model description

133

2.2.1 Model overview

134

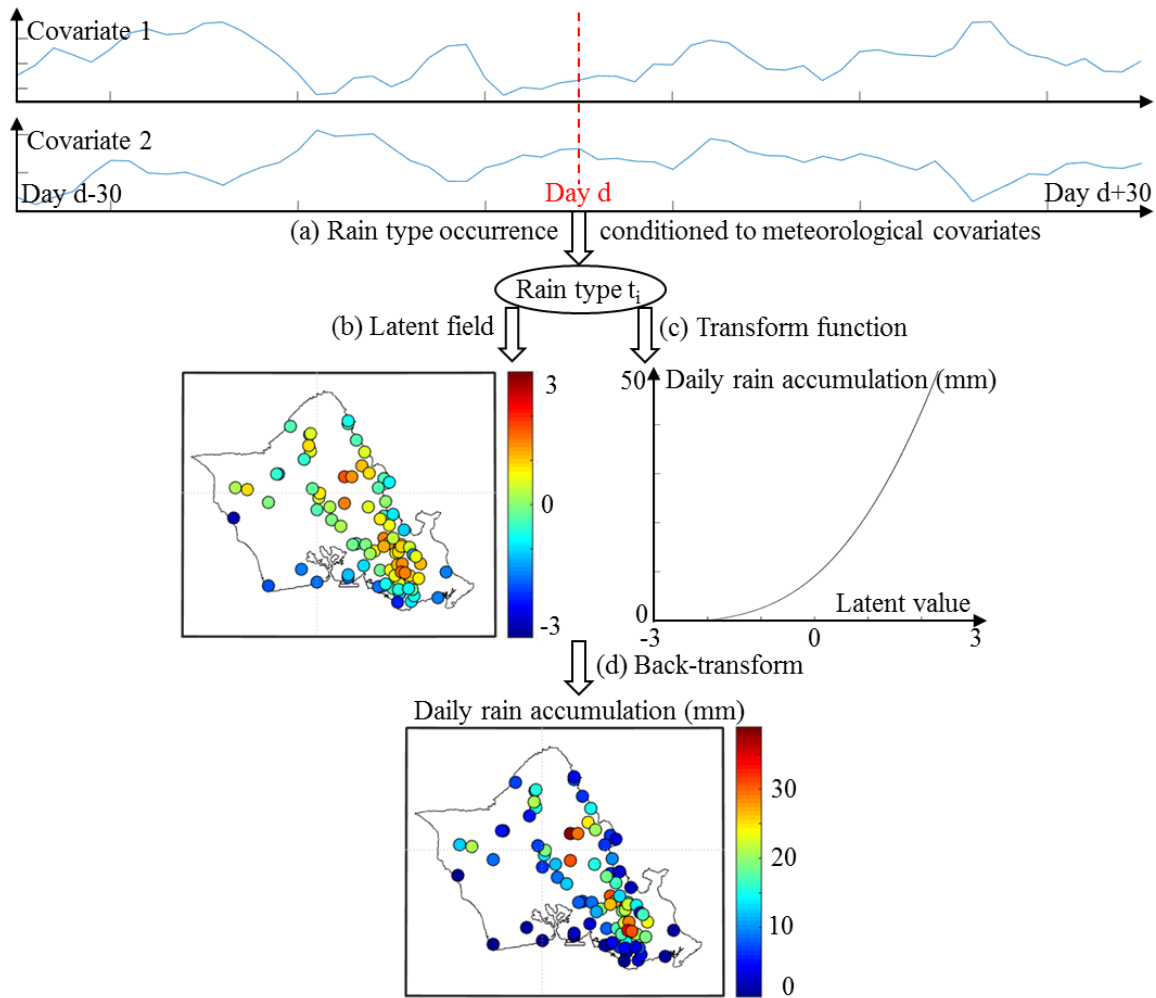
To model the statistical features of daily rainfall in tropical islands while accounting for the main mechanisms of orographic rain enhancement, in particular the variability of rainfall-topography relationships under the influence of changing atmospheric conditions, the proposed model splits rainfall behavior into three components: temporal variability, intensity (i.e., marginal distribution), and spatial distribution. Figure 2 summarizes the structure of the model, which is briefly introduced in the later part of this subsection, and will be discussed in detail in sections 2.2.2 to 2.2.4.

135

136

137

138



139

140 **Figure 2: Overview of the structure of the stochastic rainfall model.** (a) Meteorological covariates (here Geopotential at 950hPa and
 141 Temperature difference between 950hPa and 700hPa) driving the occurrence of rain types, which summarize daily rain statistics. (b) Latent
 142 field modeling of the spatial distribution of rainfall across the island. (c) Transform function linking latent values with actual rain
 143 accumulations. (d) Back-transform combining (b) and (c) to obtain daily rain simulations.

144

145 The temporal variability of rain statistics and its relationships with the state of the atmosphere are modeled following
 146 a rain typing approach (Fig. 2a) [Ailliot *et al.*, 2015] [Benoit *et al.*, 2018b]. In this framework, days with similar rain statistics
 147 are first pooled together in a finite number of rain types, which represent summaries of island-scale daily rain statistics. Next,
 148 rain type occurrence is modeled by a Markov chain describing how rain types transition to each other [Bárdossy and Plate,
 149 1991] [Wilby, 1994]. To preserve climatological consistency and model the influence of atmospheric circulation on orographic
 150 rain enhancement, rain type occurrence is conditioned to local meteorological covariates [Benoit *et al.*, 2020] making the
 151 Markov chain non-homogeneous [Hughes and Guttorp, 1999] [Vrac *et al.*, 2007]. It is important to note here that we do not

152 use the widespread Hidden Markov Model (HMM) approach [Ailliot *et al.*, 2009] [Greene *et al.*, 2011] to introduce rain types
 153 in our stochastic rainfall model, but rather resort to a two-steps approach in which rain types are first determined by a direct
 154 classification of rain gauge observations, and subsequently a statistical rainfall model is built for each rain type. This choice
 155 has been motivated by the use of non-parametric methods to model the distribution of rainfall conditional to rain types, which
 156 prevents the formulation of the likelihood of the full statistical model and in turn the use of a HMM to model rain types.

157 Conditional to each rain type, the distribution of rain across the island is modeled following a meta-Gaussian approach
 158 (also referred to as trans-Gaussian or transformed Gaussian) [Allard and Bourotte, 2015] [Baxevani and Lennartsson, 2015]
 159 [Papalexiou and Serinaldi, 2020]. In this framework, a latent field with standardized Gaussian marginal distribution is non-
 160 linearly transformed to match the marginal distribution of daily rainfall across the island, and the spatial dependencies of the
 161 latent field are used to encode the spatial distribution of rainfall (Fig. 2b-c). The latent field is often assumed to follow a
 162 multivariate-Gaussian distribution, which allows for a parsimonious modeling of the spatial dependencies using geostatistics
 163 [Lantuéjoul, 2002], and in turn the spatial interpolation of rain gauge observations [Benoit *et al.*, 2018a]. In this study however,
 164 a careful investigation of inter-gauges spatial dependencies (Supplementary material 1) shows the inadequacy of the
 165 multivariate-Gaussian distribution to model the latent field. Instead, a non-parametric resampling of past analogs
 166 [Gangopadhyay and Clark, 2005] [Yiou, 2014] of the daily latent field is used to model spatial dependencies. Aside from the
 167 choice of the marginal distribution of the latent field (Gaussian vs. uniform), this model is comparable to the use of empirical
 168 spatial copulas [Bárdossy and Pegram, 2009] coupled with a parametric marginal distribution of non-zero rain accumulation.
 169 This approach has the advantage of faithfully reproducing the complex spatial distribution of rainfall due to orographic effects,
 170 but at the expense of the ability to spatial interpolation. The present model should therefore be regarded as a multi-site
 171 stochastic rainfall generator (and not a rain field generator), and the spatial interpolation of the simulated multi-site rainfall is
 172 kept out of the scope of this paper.

173 174 2.2.2 Meta-Gaussian representation of island-scale daily rainfall

175 Rain intensity and spatial distribution are modeled jointly following a meta-Gaussian approach (Fig. 2b-c). For a
 176 given day, the observed rain accumulations $R_{i=1 \dots N_T}$ across a network of N_T gauges are linked to their latent counterparts Z_i
 177 (which follow a standardized Gaussian marginal distribution, i.e., $Z \sim \mathcal{N}(0,1)$) through a non-linear transform function ψ . This
 178 transformation is performed by first assuming that non-zero rain accumulations observed throughout the island in a given day
 179 follow a Gamma distribution:

$$180 \quad Z_i = \psi(R_i) = \Phi^{-1} \left(\frac{N_d}{N_T} + \frac{N_w}{N_T} \times \text{Gamma}(R_i; k, \theta) \right) \text{ if } R_i > 0 \quad (1)$$

181 where N_d , N_w are the number of dry and wet gauges, Φ^{-1} is the inverse cumulative distribution function (cdf) of the univariate
 182 standardized Gaussian distribution, and $\text{Gamma}(R_i; k, \theta)$ is the cdf of the Gamma distribution with shape parameter $k > 0$
 183 and scale parameter $\theta > 0$.

184 In many instances, gauges measuring no rain (i.e., $R_i=0$) represent a significant part of the network, which creates a
 185 concentration of zero values in rain accumulation distribution, and prevents a correct Gaussian transform using the function of
 186 Eq. (1). To circumvent this problem, the latent values corresponding to dry gauges are assigned based on the distance of the
 187 dry gauges to the closest wet gauge, such as the marginal distribution of the latent values matches the left portion of a
 188 standardized normal distribution. The idea behind this application-specific solution to deal with the spatial intermittence of
 189 rainfall is that a location far from any wet gauge should remain dry even after combining the associated latent field with slightly
 190 different parameters of the transform function (Eq. 1-2) during the simulation step (Sect. 2.3.3). In case of a gauge measuring
 191 no-rain ($R_i = 0$), the corresponding latent value is given by:

$$192 \quad Z_i = \psi(R_i) = \Phi^{-1} \left(\left(1 - \frac{Dw_i}{\max_{j=1:N_d}(Dw_j)} \right) \times \frac{N_d}{N_T} \right) \text{ if } R_i = 0 \quad (2)$$

193 where Dw_i is the distance of the gauge i observing no rain to the closest gauge measuring non-zero rain. This transformation
 194 has the advantage of creating spatial patterns of censored latent values (i.e., corresponding to dry gauges) that are coherent
 195 with the ones of non-censored latent values (i.e., corresponding to wet gauges), and create smooth transitions between wet and
 196 dry domains.

197 Once latent values (Z_i) are derived from rain observations (R_i), the spatial distribution of rain across the island is
 198 defined by the spatial distribution of the latent field [Bárdossy and Pegram, 2009], i.e., the joint cdf of Z_i . As mentioned in
 199 section 2.1, the spatial distribution of daily rainfall in high tropical islands is complex and strongly non-stationary due to
 200 orographic effects, which prevents the use of a simple parametric form (such as the multivariate Gaussian distribution used in
 201 most meta-Gaussian models of precipitation [Benoit et al., 2018a] [Papalexiou and Serinaldi, 2020]) for the spatial
 202 dependencies. Hence, in the present case, the non-parametric resampling of past latent fields is used to simulate the spatial
 203 distribution of rainfall [Rüschendorf, 2009].

204 2.2.3 Rain typing

205 Based on the above meta-Gaussian representation of daily rain fields, days with similar rain statistics are pooled into
 206 rain types (Fig. 2a) using a non-supervised clustering applied on the six-dimensional feature-space defined by the following:

- 207 - The three parameters of the transform function (ψ) (i.e., $p_0 = \frac{N_d}{N_T}$, k , θ), which inform the marginal distribution of daily
 208 rainfall over the island.
- 209 - The first three components of the Karhunen-Loève expansion [Huang et al., 2001] of the latent field Z
 210 (PC_1, PC_2, PC_3), which inform the spatial distribution of rainfall across the island.

211 Based on this feature-space $\mathbf{Y} = (p_0, k, \theta, PC_1, PC_2, PC_3)^T$, the clustering is performed using a Gaussian Mixture
 212 Model (GMM, [Fraley and Raftery, 2002]) which approximates the pdf of \mathbf{Y} as a weighted sum of multivariate Normal
 213 distributions:

$$214 \quad p_{\mathbf{Y}}(\mathbf{Y} = \mathbf{y}) = \sum_{l=1:N_c} b_l \times \mathcal{N}(\mathbf{y} | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) \quad (3)$$

215 where p_Y is the joint pdf of the random vector \mathbf{Y} , N_C is number of components in the GMM, b_l is a weight assigned to the 1th
 216 component, and $\boldsymbol{\mu}_l$ and $\boldsymbol{\Sigma}_l$ are the mean vector and covariance matrix of the multivariate normal distribution of the 1th
 217 component. Here, the parameters embedded in the vector \mathbf{Y} are assumed to be only slightly correlated and the covariance
 218 matrices ($\boldsymbol{\Sigma}_l$) are therefore assumed to be diagonal. The number of components of the GMM (N_C) is selected by minimization
 219 of the Bayesian Information Criterion (BIC [Schwartz, 1978]) estimated for different numbers of components in order to select
 220 a parsimonious classification (i.e., with as few rain types as possible) while properly fitting the pdf of \mathbf{Y} (i.e., p_Y). Once the
 221 pdf p_Y is known, the probability that an observed vector \mathbf{y}_{obs} belongs to the 1th component C_l is given by:

$$222 \quad p(\mathbf{y}_{obs} \in C_l) = \frac{b_l \times \mathcal{N}(\mathbf{y}_{obs} | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)}{\sum_{k=1}^{N_C} b_k \times \mathcal{N}(\mathbf{y}_{obs} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}. \quad (4)$$

223 And the classification is obtained by assigning each day (d_i) with a rain type (RT) that corresponds to the most probable
 224 mixture component:

$$225 \quad RT(d_i) = \max_{l \in 1..N_C} (p(\mathbf{y}_i \in C_l)). \quad (5)$$

226 2.2.4 Rain type occurrence

227 Once rain types have been defined based on rainfall statistical properties, their occurrence is conditioned to the vector
 228 \mathbf{MC}_d of meteorological covariates observed at day d (Fig. 2a), and rain type occurrence is modeled by a non-homogeneous
 229 Markov Chain of order 1 [Hughes and Guttorp, 1999] [Vrac et al., 2007]:

$$230 \quad p(RT_d = j | RT_{d-1} = i, \mathbf{MC}_d) \propto \gamma_{ij} \exp\left(-\frac{1}{2}(\mathbf{MC}_d - \boldsymbol{\mu}_{ij})\boldsymbol{\Sigma}_{ij}^{-1}(\mathbf{MC}_d - \boldsymbol{\mu}_{ij})^T\right) \quad (6)$$

231 Where RT_d is the state of the Markov chain (i.e., the rain type) at day d , $p(RT_d = j | RT_{d-1} = i, \mathbf{MC}_d)$ is the probability to
 232 transition from rain type i to rain type j , $\boldsymbol{\Sigma}_{ij}$ and $\boldsymbol{\mu}_{ij}$ are the covariance matrix and the mean vector of the meteorological
 233 covariates when the transition from type i to type j occurs, and γ_{ij} is the baseline (i.e., long term average) probability of
 234 transition from type i to type j . This model allows the transition probability $p(RT_d = j | RT_{d-1} = i, \mathbf{MC}_d)$ to vary proportionally
 235 to the conditional density of \mathbf{MC}_d given the transition, and conditions the occurrence of rain types to the state of the atmosphere
 236 characterized by the covariates.

237 2.3 Model implementation

238 2.3.1 Selection of meteorological covariates

239 The set of meteorological covariates used for the conditioning of the non-homogeneous Markov Chain must be chosen
 240 so that: (i) the covariates are only weakly correlated to each other, which ensures model parsimony (i.e., minimal redundancy
 241 between covariates); and (ii) the temporal variations of the covariates are correlated with variations in rain type occurrence
 242 (Supplementary Material 2), which indirectly informs the seasonality and interannual variability of rainfall patterns. Note that
 243 in the present framework, the conditioning to covariates (i.e., the non-homogeneous part of the Markov chain) is used to inform
 244 the low frequency fluctuations of rain type occurrence (seasonal to interannual time scales), with higher frequencies (weekly

245 to daily time scales) being informed by the baseline transition probabilities (γ_{ij}). Hence, meteorological covariates are
246 aggregated at the monthly scale prior to use for the conditioning of the non-homogeneous Markov chain. The monthly-
247 aggregated covariates inform monthly anomalies in atmospheric conditions and, in turn, the likelihood of rain types to occur
248 during a given month.

249 In the present case, we selected the meteorological covariates according to our initial knowledge about rain generation
250 mechanisms in high tropical islands, and their links with the state of the atmosphere [Elison Timm *et al.*, 2014] [Réchou *et al.*,
251 2019] [Sanfilippo, 2020]; this led to the following five covariates.

- 252 1) Geopotential height at 700 hPa ($\text{m}^2.\text{s}^{-2}$). This covariate is correlated with the presence of synoptic-scale weather systems
253 at the vicinity of the island and identifies regional atmospheric disturbances.
- 254 2) Temperature difference between 950 hPa and 700 hPa (K). This covariate is correlated with the lower atmospheric
255 instability and identifies days prone to shallow convection.
- 256 3) Specific humidity at 700 hPa ($\text{kg}.\text{kg}^{-1}$). This covariate informs the presence of humidity above the height of the trade wind
257 inversion and is negatively correlated with the strength of the inversion and positively correlated with the potential for
258 deep convection and cold rain.
- 259 4) Meridional and 5) longitudinal humidity fluxes at 950 hPa (i.e., specific humidity multiplied by the u (east-west) or v
260 (north-south) components of the wind field, $\text{m}.\text{s}^{-1}.\text{kg}.\text{kg}^{-1}$). These covariates provide the amount of moisture crossing over
261 the mountain barrier available for precipitation and are a proxy for orographic precipitation.

262 **2.3.2 Model calibration**

263 The model is calibrated from a training dataset made of N days of rain accumulation recorded by a network of N_T rain
264 gauges (Fig. 1a). Data must be available for all stations and all days of the calibration period, and a preliminary gap-filling
265 step is required in case of incomplete data [Longman *et al.*, 2018] [Oriani *et al.*, 2020]. Once a complete training dataset is
266 available, the first step of model calibration consists of inferring the parameters of the transform function (ψ) for each day of
267 the training period. This is performed by calculating the proportion of dry gauges and then estimating the parameters of the
268 gamma distribution of the wet gauges using a maximum likelihood approach. Once the three parameters of ψ are known, this
269 function can be inverted to derive the latent values at each gauge location.

270 After calibration of the transform function and derivation of the latent values for each day of the calibration dataset,
271 days with similar rain statistics are pooled together by rain typing. The first three principal components of the latent field are
272 preliminarily derived from the Karhunen-Loève transform of all latent values. Next, the parameters of the GMM are inferred
273 using an expectation-minimization approach [Fraleigh and Raftery, 2002]. Finally, rain typing (i.e., clustering) is performed by
274 assigning to each day the type that corresponds to the most probable component of the GMM.

275 After rain typing, the time series of observed rain types is analyzed in relation to observations of the meteorological
276 covariates to calibrate the non-homogeneous Markov chain. The baseline transition matrix (γ_{ij}) is first estimated by counting

277 the transitions between each pair of rain types occurring during the calibration period and normalizing the result by the total
278 number of transitions. Next, the parameters of the mean vector ($\boldsymbol{\mu}_{ij}$) and the covariance matrix ($\boldsymbol{\Sigma}_{ij}$) used to make the Markov
279 chain non-homogeneous are estimated by the method of moments applied to covariates observations.

280 Conditional to each rain type, the joint distribution of the parameters of ψ is inferred by multivariate kernel density
281 estimation using a trivariate Gaussian kernel. The bandwidth of the kernel is selected following the Scott's rule [Scott, 1979]
282 [Scott, 2010], i.e., in the present case:

$$\sqrt{\mathbf{H}_{ii}} = N^{-\frac{1}{7}} \times \sigma_i \quad (7)$$

284 where \mathbf{H} is the bandwidth matrix of the kernel, N the number of days in the calibration dataset, and σ_i the standard deviation
285 of the i^{th} parameter (here $i=1..3$).

286 Finally, because the latent field is simulated using an analog approach (cf next sub-section for details), this part of the
287 model does not require formal inference.

288 **2.3.3 Stochastic rainfall generation**

289 After model calibration, stochastic rainfall generation is performed following the steps summarized in Fig. 2. Starting
290 from a time series of meteorological covariates, rain types are first simulated using the non-homogeneous Markov chain
291 described in Eq. (6). Next, conditional to this simulated rain type time series, the parameters of the transform function are
292 sampled from their joint distribution defined by Eq. (7). Subsequently, one realization of the latent field is simulated using an
293 analog approach [Gangopadhyay and Clark, 2005] [Mezghani and Hingray, 2009] [Yiou, 2014], i.e., by randomly picking the
294 latent field of a day belonging to the same rain type as the day to simulate from the calibration dataset. Finally, the simulated
295 rain field is obtained by back-transformation of the simulated latent field (Eq. 1–2) using the simulated parameters of the
296 transform function.

297 **2.4 Model assessment**

298 The ability of the model to identify climatologically relevant rain types is assessed qualitatively by applying rain
299 typing to the full study dataset of section 2.1 and scrutinizing the emergent spatial-temporal rainfall patterns for each type. The
300 resulting classification is subsequently interpreted in terms of rain generation processes by confronting rain types with co-
301 occurring meteorological covariates. However, in doing so, one should keep in mind that the rain typing procedure is fully
302 statistical and that the rain type description is based on emerging statistical patterns, not on physical modeling (e.g., using a
303 numerical weather model to reproduce the observed patterns).

304 When discussing rain types and their link to rain generation processes, special attention is paid to:

- 305 (1) The emergence of spatial patterns in relation with orographic effects;
- 306 (2) The seasonality of rain type occurrence in relation with the regional annual rain cycle;

307 (3) The relationship of rain types with the state of the atmosphere quantified by the set of climate covariates described in
308 section 2.4 and used here at a daily resolution (i.e., not aggregated at the monthly scale as is the case for the conditioning
309 of the non-homogeneous Markov chain).

310 After the assessment of the climatological realism of rain types, the ability of the model to stochastically generate
311 rainfall is assessed using a leave-one-year-out cross-validation procedure. Data from one year are iteratively removed from the
312 study dataset of section 2.1 and the stochastic model is calibrated using the remaining data (i.e., 19 years of data are used for
313 model calibration). The model is fully recalibrated, which includes rain typing, inference of the transform function, and creation
314 of a training dataset of latent fields. After model calibration, daily rainfall is simulated for each day of the target year, i.e., the
315 year excluded from the calibration dataset. Fifty simulations are generated to assess the uncertainty associated with stochastic
316 rainfall generation. The same procedure is repeated for each year of the study dataset, which leads to a 20-year long validation
317 set made of 50 simulations for each gauge of the O‘ahu rain-monitoring network.

318 Simulated rainfall is compared to observations following a multi-criteria approach. First, simulation results are
319 evaluated qualitatively by visual inspection of rainfall time series for the four target stations of Fig 1.a. Next, a quantitative
320 assessment is performed using the following evaluation statistics:

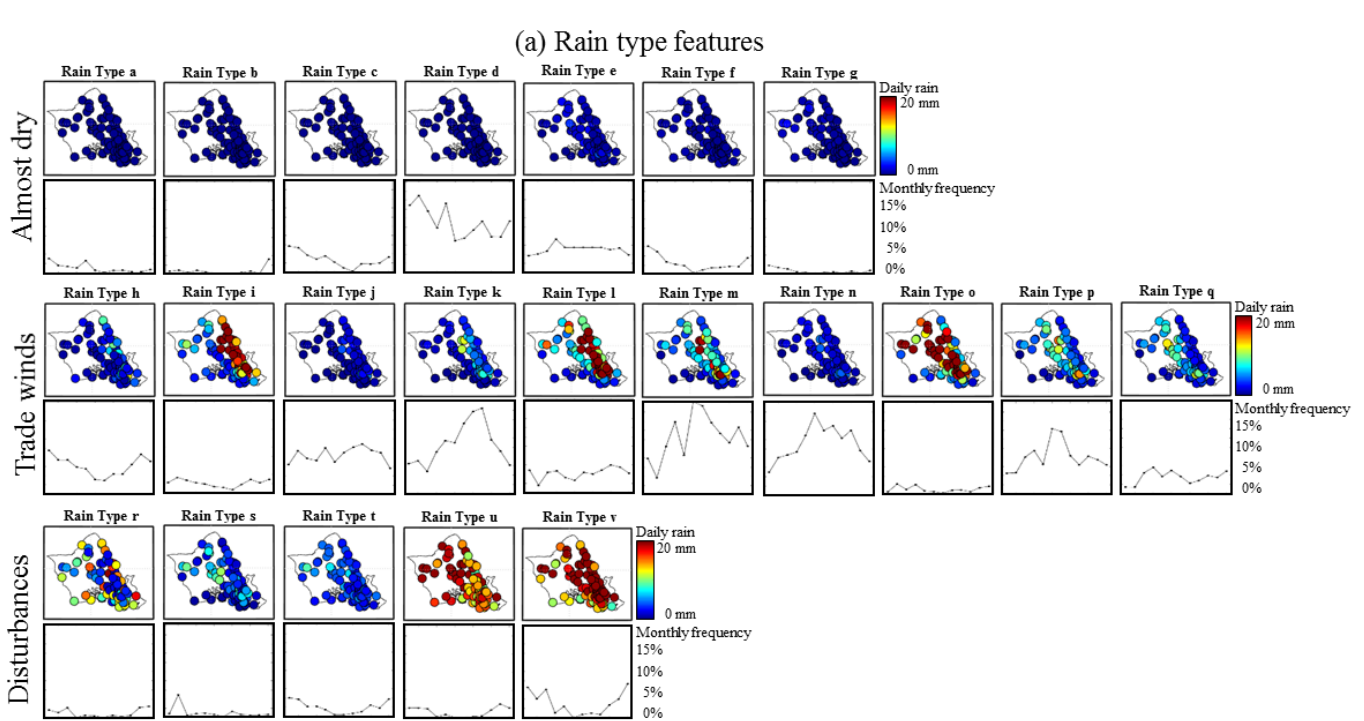
- 321 (1) Site-specific rain statistics. The following statistics are considered: quantiles 10%, 50% and 90% of monthly rain
322 accumulation to assess seasonality; annual rain accumulation to assess interannual variability; quantile-quantile (q-q) plot
323 of the percentiles of daily rain accumulation to assess the probability distribution of daily rainfall; and q-q plot of the
324 percentiles of wet-spell duration to assess rain persistence.
- 325 (2) Spatial patterns of rain distribution across the island. The following statistics are mapped to investigate the spatial
326 distribution of rainfall: quantiles 10%, 30%, 50%, 70% and 90% of daily rain to assess how the probability distribution
327 of rainfall varies in space.
- 328 (3) Areal rain statistics. Q-q plots of the percentiles of (i) the proportion of dry rain gauges, (ii) mean and (iii) max of daily
329 rain, and (iv) the coefficient of variation of rain accumulation across the island to assess island-scale statistics.

330 **3 Results**

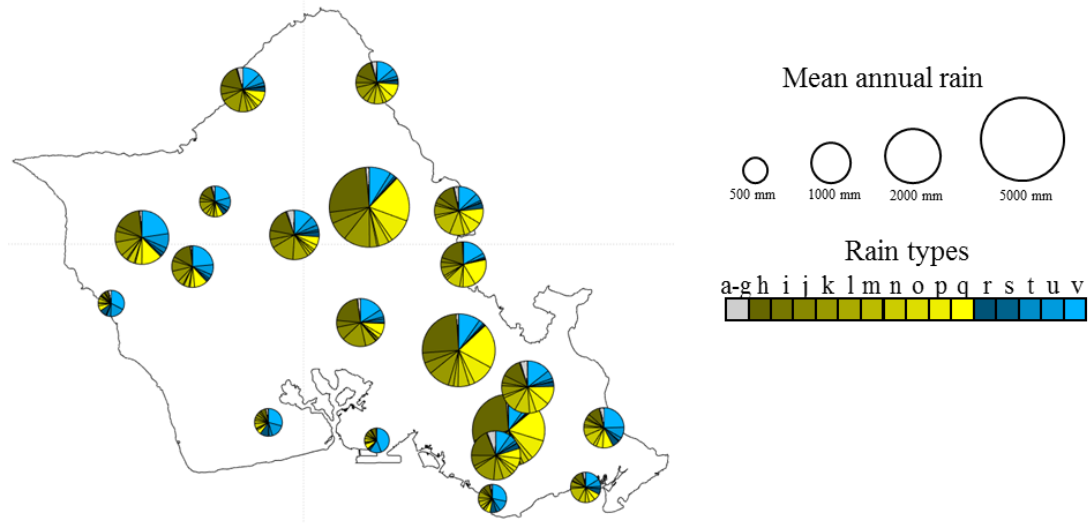
331 **3.1 Rain types in O‘ahu**

332 Figure 3 displays the 22 rain types identified for O‘ahu Island during the period 1991–2011. Although this number
333 may seem high compared to the number of rain types inferred for mid-latitude continental climates (usually less than 10 types,
334 see e.g. [Vrac *et al.*, 2007] [Benoit *et al.*, 2018b]), we believe that it reflects the tremendous variability of rainfall observed in
335 the Hawai‘i archipelago [Giambelluca *et al.*, 2013], which led Hawaiians to use more than one hundred different words to
336 describe rainfall [Akana and Gonzalez, 2015]. In addition to the large number of rain types required to account for rainfall
337 variability in tropical islands, one key attribute of the resulting classification is that although no information is given to the

338 classifier about geographical coordinates, time of occurrence, or meteorological covariates, the identified rain types display
 339 well-defined patterns of spatial rain distribution (Fig. 3a), seasonality of occurrence (Fig. 3a), and correlation with the regional
 340 state of the atmosphere (Supplementary Material 2).



(b) Rain type contribution to annual rain



341
 342 **Figure 3: Rain types identified for the island of O'ahu.** (a) Spatial distribution of daily rain and frequency of occurrence of each rain type.
 343 (b) Contribution of each rain type to the annual rain accumulation for a selection of 20 gauges spread throughout the island.
 344

345 To better identify the main modes of rainfall variability over O‘ahu, rain types are pooled into three hyperclasses (H1-
346 3) that can be linked to the three main rain generation processes in the area (Fig. 3):

- 347 • (H1) Almost dry days (Fig. 3, rain types a–g). During these days, most rain gauges report no rain, and no gauge
348 reports more than 5 mm/day on average. These types of weather conditions are associated with a stable atmosphere
349 and a low moisture flux (Fig. SM 2.1).
- 350 • (H2) Trade wind days (Fig. 3, rain types h–q). This category displays well-defined spatial patterns of rain
351 accumulation caused by orographic lifting, and are associated with a stable atmosphere, a well-defined trade wind
352 inversion, and an important influx of moisture below the inversion layer under the influence of east-northeasterly
353 trade winds (Fig. SM 2.1). When scrutinizing inter-type variability within this category, note that the location of the
354 rain maximum shifts westward with increasing moisture flux, likely due to stronger trade winds causing an overshoot
355 of orographic rain enhancement whereby rain forms over the mountains but falls further downwind on the leeward
356 side [Daly *et al.*, 2017]. In addition, for similar wind conditions and, therefore, spatial patterns (compare for instance
357 types j, k and l), rain intensity is correlated to the instability of the atmosphere (Fig. SM 2.1).
- 358 • (H3) Regional atmospheric disturbance days (Fig. 3, rain types r–v). These types display either unstructured (types r–
359 t) or relatively homogeneous (types u–v) spatial patterns of rain accumulation and are associated with low pressure,
360 unstable atmosphere, and absent (or weak) trade wind inversion. This allows high moisture content at high altitude
361 (Fig. SM 2.1). These rain types mostly occur during winter, i.e., the local rainy season. When scrutinizing inter-type
362 variability within this category, note that rain intensity increases with atmospheric instability and the presence of
363 humidity at high altitude, and that the spatial patterns tend to become more structured when the low-level moisture
364 influx increases (probably due to stronger and more uniform winds).

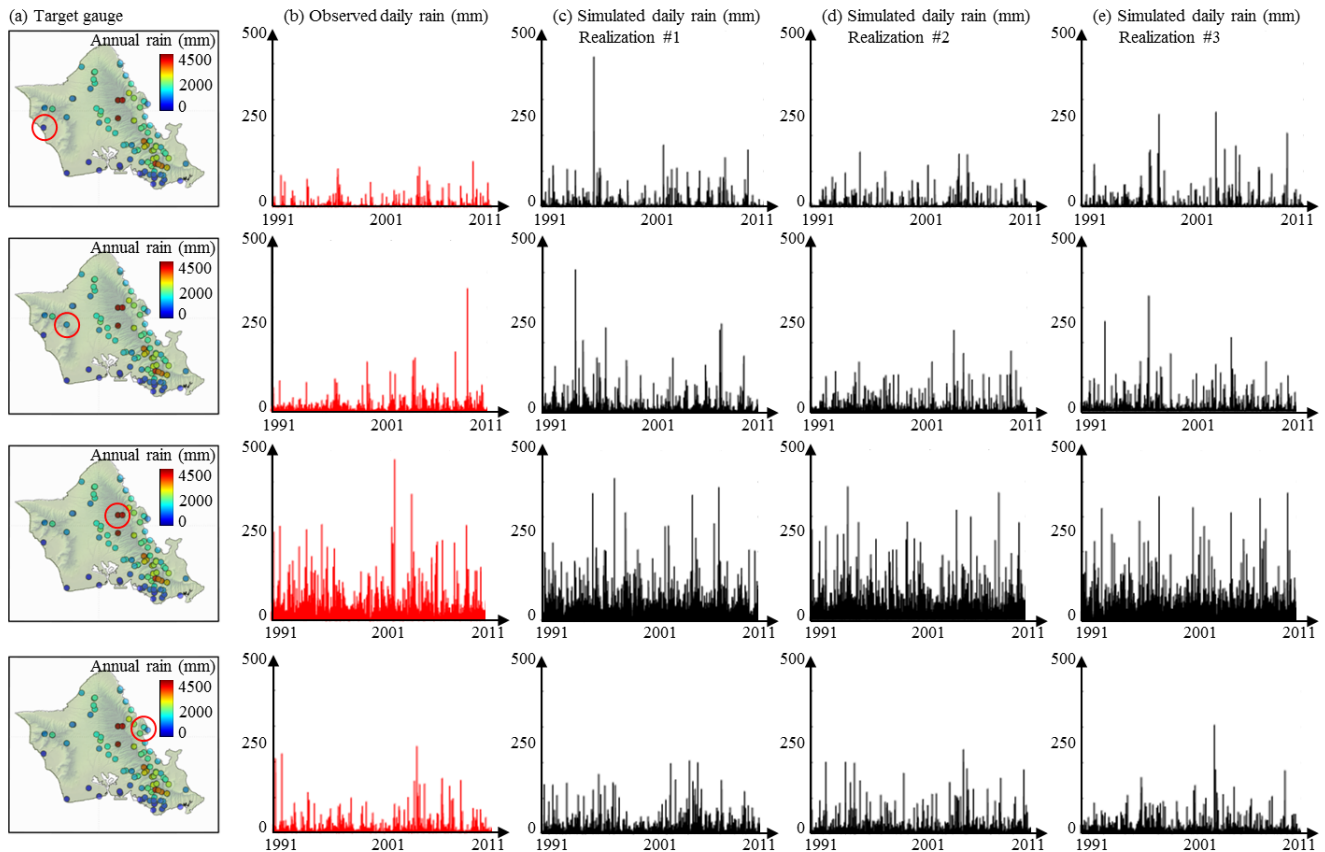
365 Hence, rain typing provides new insights on island-scale rain climatology (Fig. 3b). In particular, this step helps us
366 gain a better understanding of how different atmospheric conditions lead to different rain generation processes that, when
367 interacting with island topography, generate contrasting orographic effects. In the case of the island of O‘ahu, orographic rain
368 enhancement occurring during days influenced by trade winds is the main explanation for the high annual rain accumulations
369 in the Ko‘olau mountains (up to 5000 mm annual rainfall), while widespread rainfall linked to regional atmospheric
370 disturbances is the main source of rain at leeward locations despite their relative temporal scarcity.

371

372 **3.2 Simulation of site-specific rainfall time series**

373 Figure 4 displays stochastic rainfall generation outputs for the four rain gauges of Fig 1.a which experience different
374 rainfall climatologies. Visual inspection of the simulated time series shows that our stochastic rainfall generator is able to
375 simulate synthetic rainfall that is almost indistinguishable from the observed one. The dry-wet ratio as well as the marginal
376 distribution of daily rain accumulation seem properly replicated, except for the dry leeward gauge (first row in Fig. 4) where
377 the 20-year maximum tends to be overestimated (see also Fig. 5.d). In terms of temporal variability, the simulated time series

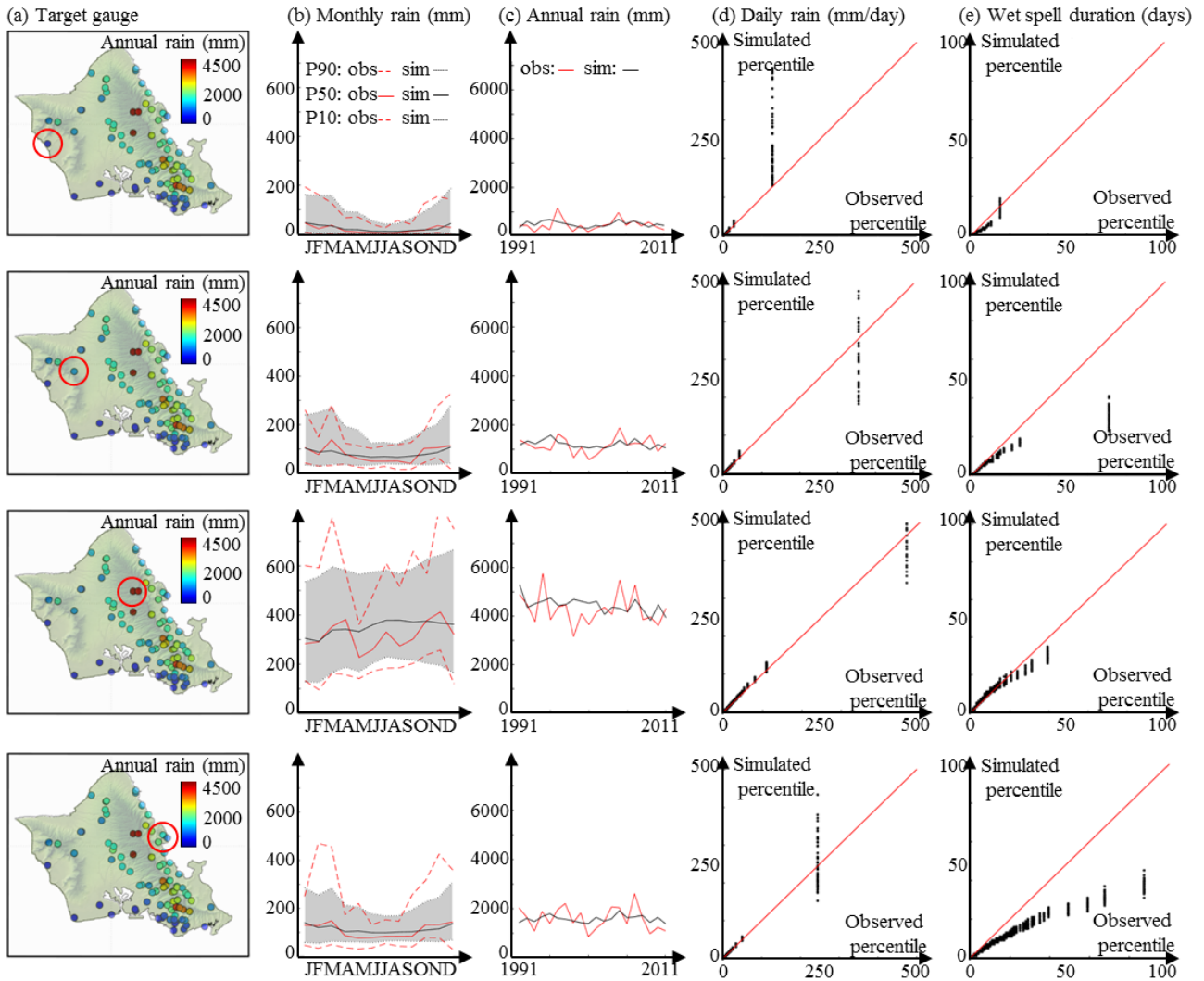
378 properly capture the seasonal cycle visible for the two leeward gauges (rows 1-2 in Fig. 4), as well as the inter-annual variability
379 visible for the two coastal gauges (rows 1 and 4 in Fig. 4). Finally, the rainfall generator captures the relatively steady behavior
380 of the wet gauge located in the Ko‘olau range (row 3 in Fig. 4).
381



382
383 **Figure 4: Ability of the model to simulate realistic rainfall time series on O'ahu.** (a) Target locations. (b) Observed rainfall time series
384 for the period 1991-2011. (c-e) Simulated time series. For readability, only the three first realizations of the 50-member ensemble are
385 displayed.
386

387 3.3 Simulation of site-specific rain statistics

388 To complement the qualitative assessment of site-specific rainfall time series in Figure 4, Figure 5 investigates the
389 statistical features of the results of the cross-validation procedure (50 realizations are drawn) for the same four rain gauges.



390

391

392

393

394

395

396

397

398

399

400

401

Figure 5: Ability of the model to simulate site-specific rain statistics on O'ahu. (a) Target locations. (b) Observed (red) and simulated (black) monthly rain accumulation. Dashed lines denote quantiles 10% and 90%, and solid lines denote the quantile 50%. In the case of simulations (black), for readability, we report in the figure only the median of each quantile (10%, 50%, 90%) instead of 50 simulated quantiles. The grey band between the dashed black lines denotes the interval Q10%-Q90% derived from simulations. (c) Observed (red) and simulated (black) annual rain accumulation. In the case of simulations (black), for readability, we report in the figure only the median of the 50 simulations. (d) Q-q plot of daily rain percentiles. (e) Q-q plot of wet-spell duration percentiles. In (d-e) we display q-q plots between one observation time series (abscissa) and 50 simulations (ordinates). Hence, black dots line up vertically in the q-q plots because for each percentile, 50 simulations are compared to a single observation.

The results in Fig. 5 show that the proposed model correctly simulates rainfall seasonality (Fig. 5b) and interannual variability (Fig. 5c). Note that simulations capture both the stronger seasonality at leeward locations (compared to windward

402 locations) as well as the near absence of seasonality at the wettest gauge located in Ko‘olau Mountains (Fig. 5, third row). The
403 interannual variability of rain accumulation is also properly simulated, in particular, at leeward locations where the impact of
404 winter storms is the highest. These results suggest that the non-homogeneous Markov chain of order 1 conditioned to monthly-
405 aggregated meteorological covariates adequately models the long-term variability of rain accumulation, and that the selected
406 covariates properly capture rain type occurrence in a tropical marine climate.

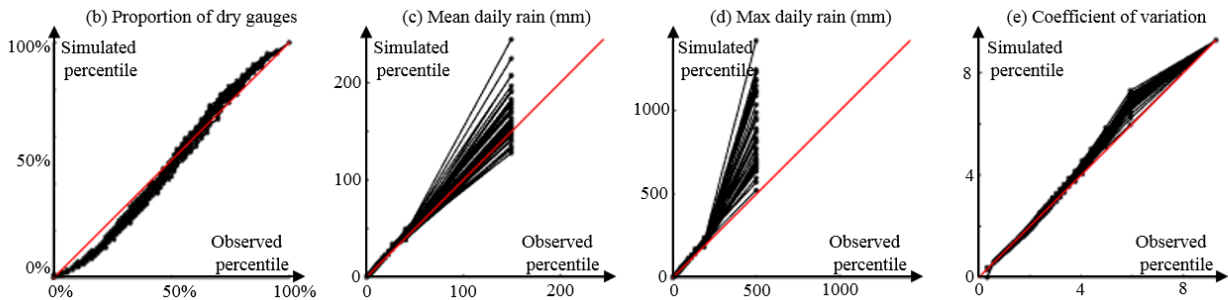
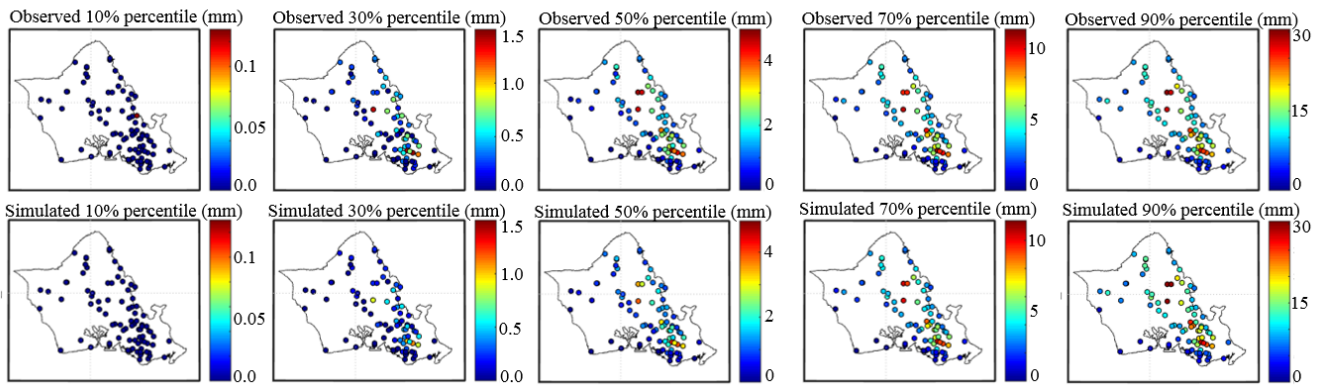
407 However, rain persistence is slightly underestimated at some locations, especially for the high percentiles, i.e., long-
408 lasting wet spells (Fig. 5e). This result exposes limitations in the use of the non-homogeneous Markov chain of order 1 for
409 modeling weekly-scale temporal variability of rainfall. This may be explained by the fact that daily-scale and seasonal-scale
410 rainfall fluctuations are informed, respectively, by the Markov chain of order 1 and conditioning to monthly-aggregated
411 meteorological covariates, but that the weekly-scale is not explicitly included in the model. Although the resulting errors are
412 of low magnitude, they should be kept in mind in case of applications requiring a precise estimation of rain persistence, for
413 example crop models in semi-arid environments that can be found in some leeward areas of the target islands.

414 The simulations properly reproduce site-specific marginal distributions of daily rain accumulation (Fig. 5d), except
415 for the driest gauge (top row in Fig. 5) where the 20-year maximum tends to be overestimated. The satisfactory simulation of
416 rainfall distribution at several sites suggests that a type-dependent gamma distribution is an adequate model for the non-zero
417 daily rain accumulations across the island. It is noteworthy that all percentiles of the marginal distribution of rain accumulation
418 are properly reproduced in simulations, which suggests that our model is able to simulate the whole spectrum of daily
419 precipitation, from dry days to intense rainfall.

420 **3.4 Simulation of island-scale rain fields**

421 Figure 6 displays the results of the cross-validation procedure focusing on island-scale features. Figure 6a compares
422 observed and simulated spatial patterns for five quantiles of daily rain accumulation across the island of O‘ahu. Results show
423 very good model performance in reproducing the spatial patterns of daily rainfall. This result was expected because the use of
424 empirical copulas combined with rain typing is almost equivalent to resampling the observed spatial patterns conditional to
425 meteorological covariates. However, satisfactory simulation results ensure that the rain-type-based resampling of latent fields
426 is unbiased and that the choice and calibration of the meta-Gaussian model are relevant for the study island.

(a) Spatial patterns of daily rain percentiles



427
428 **Figure 6: Assessment of island-scale statistics simulation in O‘ahu.** (a) Spatial patterns of observed (upper row) and simulated (lower
429 row) percentiles of daily rain accumulation. From left to right: 10%, 30%, 50%, 70% and 90% percentiles. (b–d) Q–q plots of key rain
430 statistics aggregated over the whole rain gauge network: (b) proportion of dry gauges; (c–d) mean and max daily rain; (e) coefficient of
431 variation. Corresponding pdfs are displayed in Supplementary material 3. In (b–e) we display q–q plots between one observation time series
432 (abscissa) and 50 simulations (ordinates). Hence, black dots line up vertically in the q–q plots because for each percentile, 50 simulations are
433 compared to a single observation. To distinguish between simulations, quantiles related to the same simulation are connected by a solid
434 black line.

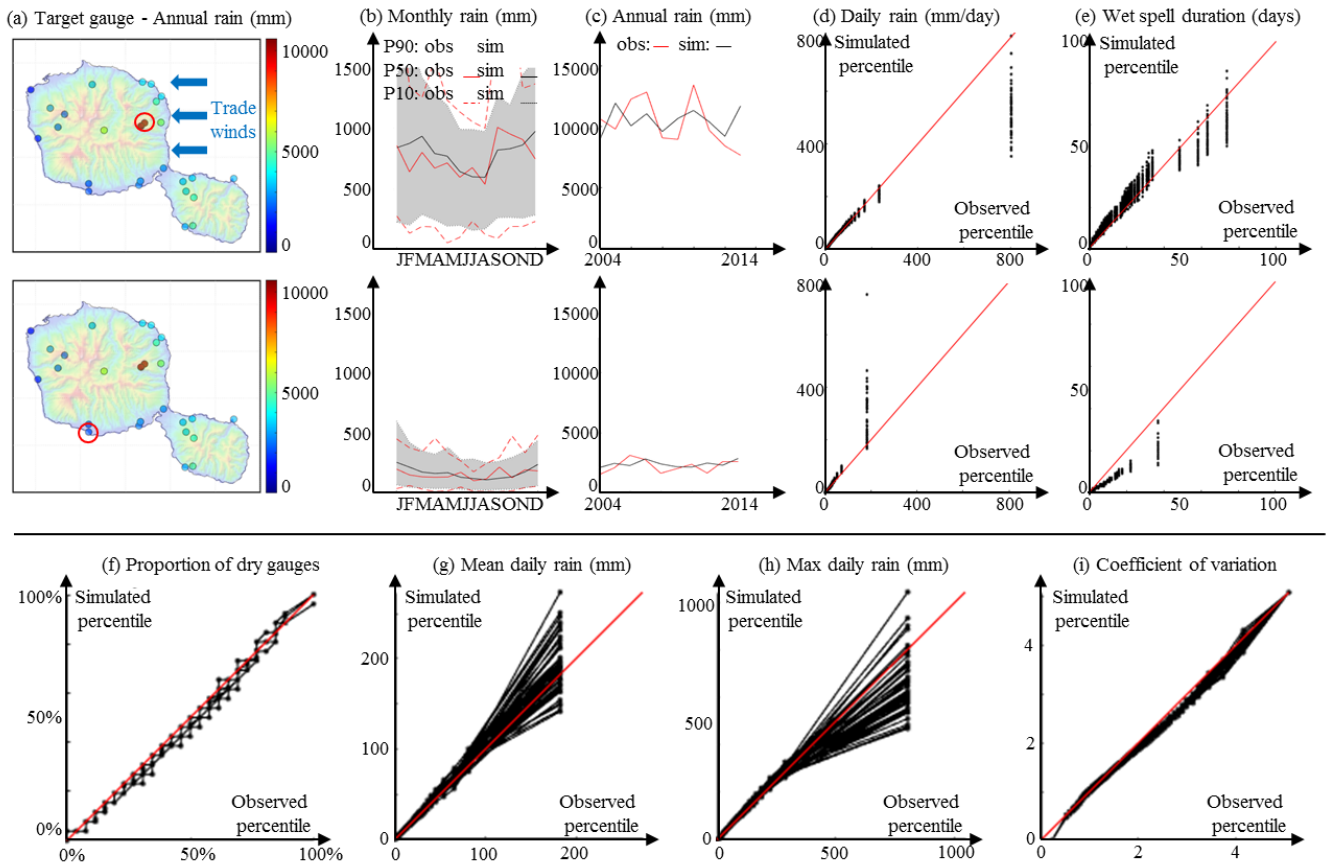
435
436
437 Figure 6b–e assesses the ability of the model to simulate four key rain statistics - the proportion of dry gauges, mean
438 and max of daily rain accumulation, and coefficient of variation of daily rain across the island - aggregated over all rain gauges
439 of the rain monitoring network of O‘ahu. Results show a slight underestimation of the low percentiles of the proportion of dry
440 gauges (Fig. 6b), which corresponds to a slight overestimation of the frequency of proportions of dry rain gauges below 5%
441 (i.e., 4 gauges over 86 in the present setting). That is, our model tends to simulate rain at all 86 gauges when a very small
442 number of gauges (less than 4) actually records no rain, leading to a slight drizzle effect in space. A careful examination of the
443 spatial patterns of daily rain percentiles (Fig 6a) shows that the patterns of spatial intermittency are properly simulated, which
444 suggests that the drizzle effect is randomly distributed amongst locations, thus reducing its potential impact for applications.
445 This level of accuracy in the simulation of the rain fraction shows that a truncated Gaussian latent field is an appropriate model

446 for rain intermittency. In addition, the correct simulation of the spatial patterns of dry locations in Fig. 6a suggests that the
447 distance-based modeling of the censored latent values (Eq. 2) coupled with empirical copulas is a proper model for the spatial
448 distribution of dry locations. Similarly, the good agreement between observed and simulated coefficients of variation (Fig. 6e)
449 coupled with the correct simulation of spatial patterns of non-zero daily rain accumulation in Fig. 6a suggest that the selected
450 meta-Gaussian framework captures the spatial distribution of non-zero rain accumulations.

451 Finally, Fig. 6c-d shows that island-scale daily mean and maximum rain accumulation are properly simulated, despite
452 an overestimation of the last percentile of the maximum, i.e., the 20-year maximum observed over the whole island. This result
453 suggests that the meta-Gaussian framework coupled with the kernel estimation of the transform function parameters performs
454 reasonably well to reproduce the marginal distribution of island-scale rain accumulation. However, the attempt to reproduce
455 both island scale statistics and site-specific marginal distributions (from dry days to heavy rainfall) results in an inaccurate
456 simulation of the island-scale 20-year extreme precipitation. This limitation calls for additional developments before the
457 proposed model can be used for simulating extremes in a spatial context [Opitz *et al.*, 2021].

458 **3.5 Model versatility**

459 To investigate the versatility of our stochastic daily rainfall model, the case study performed in sections 3.1–3.3 for
460 the island of O‘ahu (Hawai‘i, USA) located in the North Pacific was repeated in supplementary material 4 for the island of
461 Tahiti (French Polynesia) located in the South Pacific. Figure 7 provides a brief overview of the results. This additional cross-
462 validation shows that our model also performs very well for Tahiti, despite a wetter (annual rain reaches 10 000 mm in Tahiti)
463 and more seasonal climate than the O‘ahu case study. In addition, the model adapts automatically to different dataset sizes (86
464 rain gauges x 21 years for O‘ahu, 26 gauges x 11 years for Tahiti) and rain climatologies due to the selection of different
465 numbers of rain types. These results suggest that our model may be adapted to most high tropical islands across the globe.



466
 467 **Figure 7: Overview of stochastic daily rainfall simulation for the island of Tahiti, French Polynesia.** (a-e) Site-specific statistics: (a)
 468 Target locations. (b) Observed (red) and simulated (black) monthly rain accumulation. (c) Observed (red) and simulated (black) annual rain
 469 accumulation. (d) Q-q plot of daily rain percentiles. (e) Q-q plot of wet-spell duration percentiles. (f-i) Island-scale statistics: (f) proportion
 470 of dry gauges; (g-h) mean and max daily rain; (i) coefficient of variation.
 471

472 **4 Discussion and conclusion**

473 **4.1 Discussion: stochastic modeling of orographic rainfall patterns**

474 Validation results in section 3 show that the proposed model is able to simulate realistic multi-site rainfall time series,
 475 which accurately reproduce site-specific and island-scale daily rain statistics for two different tropical islands. This has been
 476 made possible by a two-step modeling approach (rain typing and meta-Gaussian representation of island-scale daily rainfall),
 477 which takes into account our conceptual knowledge about orographic rain enhancement in tropical islands.

478 The first component consists of rain types, which summarize island-scale rain statistics. Unlike weather type based
 479 approaches [Ailliot *et al.*, 2015] [Réchou *et al.*, 2019], we define rain types based on rain features only, i.e., no information

480 about meteorological covariates or large-scale circulation are included during the classification step. This leads to a
481 classification centered on rainfall intensity and spatial distribution, which allows us to explore how island-scale rainfall
482 variability is impacted by orographic effects (section 3.1). The links between rain types and local climate are established in a
483 second step by conditioning the non-homogeneous Markov model of rain type occurrence to meteorological covariates. We
484 conceptualize rain types as the main modes of island-scale daily rainfall variability, which is assumed to be primarily
485 influenced by orographic effects caused by interactions between changing atmospheric conditions and fixed island topography.
486 In this context, one interesting contribution of this study is the refinement of the meteorological predictors proposed by
487 [Sanfilippo, 2020] for rain type occurrence in a tropical marine climate, in particular, to distinguish between shallow convection
488 occurring during typical trade wind situations and deeper convection in the vicinity of atmospheric disturbances.

489 The second component of the model consists of a meta-Gaussian representation of island-scale daily rainfall. By
490 explicitly separating rain intensity and spatial distribution, this representation contributed to the performance of the rain typing
491 procedure detailed above and in the identification of rain types with well-defined spatial patterns. When used for stochastic
492 rainfall generation, the adopted meta-Gaussian representation performed well in simulating site-specific rain statistics as well
493 as island-scale spatial patterns of daily rain accumulation. This good performance can be explained by two factors. First, the
494 determination of the censored latent values based on the distance to the closest wet gauge (Eq. 2) generates realistic spatial
495 patterns of dry areas and dry-wet transition [Schleiss *et al.*, 2014]. This contributes to the proper modeling of the spatial
496 intermittency of daily rain fields in tropical islands, which is caused by the drying effect of sinking air masses after crossing
497 mountains. The second innovation of the model is the joint use of empirical copulas and a parametric transform function to
498 model the spatial patterns of non-zero rain. It has the advantage of faithfully preserving the spatial rainfall patterns while
499 generating unobserved values through the kernel density estimation of the transform function parameters distribution. The
500 choice of mimicking the observed spatial rainfall patterns as closely as possible is justified by the complexity of orographic
501 effects and associated rain gradients in tropical islands [Giambelluca *et al.*, 2013] [Laurent *et al.*, 2019] [Benoit *et al.*, 2021].

502 **4.2 Concluding remarks**

503 In this paper we presented a new stochastic daily rainfall generator dedicated to high tropical islands. The combination
504 of (i) rain types, (ii) a non-homogeneous Markov model of rain type occurrence conditioned to meteorological covariates, and
505 (iii) a meta-Gaussian representation of the spatial distribution of daily rainfall allowed us to generate realistic daily rain fields
506 honoring both site-specific and island-scale rain statistics. The performance of the model was carefully tested and illustrated
507 for the islands of O‘ahu (Hawai‘i, USA) and Tahiti (French Polynesia), both located in the tropical Pacific. Cross-validation
508 results prove the ability of the model to capture and simulate the main features of daily rainfall over these two high tropical
509 islands.

510 The main strength of our model is its ability to simulate diverse patterns of multi-site daily rainfall, as well as their
511 linkage with regional atmospheric conditions. It represents a new tool for stochastic investigation and modeling of orographic
512 rain enhancement on tropical islands with complex topography. The main limitations, however, are (i) the inaccurate

513 simulation of extreme rainfalls, which calls for caution when using our model for flood risk assessment, and (ii) the restriction
514 to multi-site simulation, which calls for an additional step of stochastic interpolation of the multi-site patterns when gridded
515 outputs are required.

516 Because of the above strengths and limitations, the main envisioned applications relate to impact studies that require
517 detailed knowledge of daily precipitation in tropical islands, in particular, when the spatial pattern of rainfall plays an important
518 role. This includes watershed water resources management and eco-hydrological studies.

519 Our model can also be used for the stochastic downscaling of future precipitation projections and thereby contribute
520 to the current efforts to better understand, manage, and secure tropical island water resources in a changing climate. An
521 essential future investigation in this direction will be to assess how well GCMs simulate the set of meteorological covariates
522 we selected to drive our stochastic rainfall generator, in particular the vertical temperature gradient used to inform shallow
523 convection.

524
525

526 *Code and data availability.*

527 The implementation of the proposed stochastic rainfall model is open source (MATLAB implementation) and freely available
528 in the following repository (https://github.com/LionelBenoit/StochasticRainfallGenerator_TropicalIslands). The dataset of
529 daily rainfall observations on O‘ahu is open data and freely available on the Hawai‘i Climate Data Portal
530 (<https://www.hawaii.edu/climate-data-portal/data-portal/>). An extract of this dataset is available in MATLAB format as a code
531 demo in the same repository as the source code of the model. The dataset of daily rainfall observations on Tahiti is available
532 upon request from Météo France (contact.polynesie-francaise@meteo.fr) and Groupement d’Etudes et de Gestion du Domaine
533 Public de Polynésie Française (secretariat@equipement.gov.pf).

534 *Author contributions.*

535 LB, LS and TWG designed the experiment. MPL and LS compiled the daily rainfall datasets of O‘ahu and Tahiti respectively.
536 LB and ADN selected the meteorological covariates and designed the non-homogeneous Markov Chain. LB and MPL designed
537 the meta-Gaussian model and the rain typing method. LB implemented the model and performed the numerical experiments.
538 LB wrote the paper with input and corrections from all co-authors.

539 *Acknowledgments.*

540 The work of Lionel Benoit is funded by the Swiss National Science Foundation (SNSF), grant number P2LAP2_191395. The
541 work of Lydie Sichoix is supported by the Government of French Polynesia - Ministère de la Recherche through the project
542 E-CRQUEST, grant number 05832 MED 08/26/2019. The authors are grateful to the Hawai‘i Climate Data Portal for providing
543 the daily rainfall dataset of the island of O‘ahu, and to the French Weather Agency (Direction Interrégionale en Polynésie
544 française - Météo France) and the Polynesian public service named Direction de l'Équipement (Groupement d'Études et de
545 Gestion du Domaine Public de Polynésie Française - GEGDP) for providing the daily rainfall dataset of the island of Tahiti.
546 The authors are grateful to May Izumy from the Publication services of the School of Ocean and Earth Science and Technology,
547 University of Hawai‘i for proofreading this manuscript.

548 **References**

- 549 Ailliot, P., C. Thompson, and P. Thomson (2009), Space–time modelling of precipitation by using a hidden Markov model
550 and censored Gaussian distributions, *Applied Statistics*, 58, 405-426, doi:10.1111/j.1467-9876.2008.00654.x
- 551 Ailliot, P., D. Allard, V. Monbet, and P. Naveau (2015), Stochastic weather generators: an overview of weather type models,
552 *Journal de la Société Française de Statistiques*, 156(1), 101-113, ISSN: 2102-6238
- 553 Akana, C. L., and K. Gonzalez (2015), *Hanau ka Ua Hawaiian Rain Names*, 327 pp., Kamehameha Publishing,
554 ISBN13:9780873362467
- 555 Allard, D., and M. Bourotte (2015), Disaggregating daily precipitations into hourly values with a transformed censored latent
556 Gaussian process, *Stochastic Environmental Research and Risk Assessment*, 29, 453-462, doi:10.1007/s00477-014-
557 0913-4
- 558 Ambrosino, C., R. E. Chandler, and M. C. Todd (2014), Rainfall-derived growing season characteristics for agricultural impact
559 assessments in South Africa, *Theoretical and applied climatology*, 115, 411-426, doi: 10.1007/s00704-013-0896-y
- 560 Bárdossy, A., and E. J. Plate (1991), Modelling daily rainfall using a semi-Markov representation of circulation pattern
561 occurrence, *Journal of Hydrology*, 122, 33-47, doi:10.1016/0022-1694(91)90170-M
- 562 Bárdossy, A., and G. G. S. Pegram (2009), Copula based multisite model for daily precipitation simulation, *Hydrology and*
563 *Earth System Sciences*, 13, 2299-2314, doi:10.5194/hess-13-2299-2009
- 564 Bauer, P., A. Thorpe, and G. Brunet (2015), The quiet revolution of numerical weather prediction, *Nature*, 525, 47-55, doi:
565 10.1038/nature14956
- 566 Bennett, B., M. Thyer, M. Leonard, M. Lambert, and B. C. Bates (2018), A comprehensive and systematic evaluation
567 framework for a parsimonious daily rainfall field model, *Journal of Hydrology*, 556, 1123-1138, doi:
568 10.1016/j.jhydrol.2016.12.043
- 569 Baxevani, A., and J. Lennartsson (2015), A spatiotemporal precipitation generator based on a censored latent Gaussian field,
570 *Water Resources Research*, 51, 4338-4358, doi:10.1002/2014WR016455

571 Benoit, L., D. Allard, and G. Mariethoz (2018a), Stochastic Rainfall Modeling at Sub-kilometer Scale, *Water Resources*
572 *Research*, 54, 4108-4130, doi:10.1029/2018WR022817

573 Benoit, L., M. Vrac, and G. Mariethoz (2018b), Dealing with non-stationarity in sub-daily stochastic rainfall models,
574 *Hydrology and Earth System Sciences*, 22, 5919–5933, doi:10.5194/hess-22-5919-2018

575 Benoit, L., M. Vrac, and G. Mariethoz (2020), Nonstationary stochastic rain type generation: accounting for climate drivers,
576 *Hydrology and Earth System Sciences*, 24, 1-14, doi:10.5194/hess-24-2841-2020

577 Benoit, L., M. Lucas, H. Tseng, Y.-F. Huang, Y.-P. Tsang, A. D. Nugent, T. W. Giambelluca, and G. Mariethoz (2021), High
578 Space-Time Resolution Observation of Extreme Orographic Rain Gradients in a Pacific Island Catchment, *Frontiers*
579 *in Earth Sciences*, 8, 546246, doi: 10.3389/feart.2020.546246

580 Breinl, K., G. Di Baldassarre, M. Girons Lopez, M. Hagenlocher, G. Vico, and A. Rutgersson (2017), Can weather generation
581 capture precipitation patterns across different climates, spatial scales and under data scarcity? *Scientific reports*, 7,
582 5449, doi:10.1038/s41598-017-05822-y

583 Brown, J. R., M. Lengaigne, B. R. Lintner, M. J. Widlansky, K. van der Wiel, C. Dutheil, B. K. Linsley, A. J. Matthew, and J.
584 Renwick (2020), South Pacific Convergence Zone dynamics, variability and impacts in a changing climate, *Nature*
585 *Reviews Earth and Environment* 1, 530-543, doi:10.1038/s43017-020-0078-2

586 Cappelaere, B., et al. (2020), Modeling Land Surface Fluxes from Uncertain Rainfall: A Case Study in the Sahel with Field-
587 Driven Stochastic Rainfields, *Atmosphere*, 11, 465, doi:10.3390/atmos11050465

588 Caruso, S. J., and S. Businger (2006), Subtropical Cyclogenesis over the Central North Pacific, *Weather and forecasting*, 21,
589 193-205, doi:10.1175/WAF914.1

590 Caseri, A., P. Javelle, M. H. Ramos, and E. Leblois (2016), Generating precipitation ensembles for flood alert and risk
591 management, *Journal of Flood Risk Management*, 9, 402-415, doi:10.1111/jfr3.12203

592 Chandler, R. E. (2020), Multisite, multivariate weather generation based on generalized linear models, *Environmental*
593 *Modelling & Software*, 134, 104867, doi:10.1016/j.envsoft.2020.104867

594 Daly, C., M. E. Slater, J. A. Roberti, S. H. Laseter, and L. W. Swift (2017), High-resolution precipitation mapping in a
595 mountainous watershed: ground truth for evaluating uncertainty in a national precipitation dataset, *International*
596 *Journal of Climatology*, 37, 124-137, doi:10.1002/joc.4986

597 Elison Timm, O., T. W. Giambelluca, and H. F. Diaz (2014), Statistical downscaling of rainfall changes in Hawai'i based on
598 the CMIP5 global model projections, *Journal of Geophysical Research: Atmospheres*, 120, 92-112,
599 doi:10.1002/2014JD022059

600 Foresti, L., and A. Pozdnoukhov (2012), Exploration of alpine orographic precipitation patterns with radar image processing
601 and clustering techniques, *Meteorological applications*, 19, 407-419, doi:10.1002/met.272

602 Fraley, C., and A. E. Raftery (2002), Model-Based Clustering, Discriminant Analysis, and Density Estimation, *Journal of the*
603 *American Statistical Association*, 97(458), 611-631, doi:10.1198/016214502760047131

604 Frazier, A. G., O. E. Timm, T. W. Giambelluca, and H. F. Diaz (2018), The influence of ENSO, PDO and PNA on secular
605 rainfall variations in Hawai‘i, *Climate Dynamics*, 51, 2127-2140, doi:10.1007/s00382-017-4003-4

606 Gabellani, S., G. Boni, L. Ferraris, J. Von Hardenberg, and A. Provenzale (2007), Propagation of uncertainty from rainfall to
607 runoff: A case study with a stochastic rainfall generator, *Advances in Water Resources*, 30, 2061-2071,
608 doi:10.1016/j.advwatres.2006.11.015

609 Gangopadhyay, S., and M. Clark (2005), Statistical downscaling using K-nearest neighbors, *Water Resources Research*, 41,
610 W02024, doi:10.1029/2004WR003444

611 Giambelluca, T. W., Q. Chen, A. G. Frazier, J. P. Price, Y.-L. Chen, P.-S. Chu, J. K. Eischeid, and D. M. Delporte (2013),
612 Online Rainfall Atlas of Hawai‘i, *Bulletin of the American Meteorological Society*, 94, 313-316, doi:10.1175/BAMS-
613 D-11-00228.1

614 Greene, A. M., A. W. Robertson, P. Smyth, and S. Triglia (2011), Downscaling projections of Indian monsoon rainfall using
615 a non-homogeneous hidden Markov model, *Quarterly Journal of the Royal Meteorological Society*, 137, 347-359,
616 doi:10.1002/qj.788

617 Hersbach, H., et al. (2018), Operational global reanalysis: progress, future directions and synergies with NWP, *ERA Rep. Ser.*
618 27, 1-63 pp, doi: 10.21957/tkic6g3wm.

619 Hopuare, M., M. Guglielmino, and P. Ortega (2018), Interactions between intraseasonal and diurnal variability of precipitation
620 in the South Central Pacific: The case of a small high island, Tahiti, French Polynesia, *International Journal of*
621 *Climatology*, 39, 670-686, doi:10.1002/joc.5834

622 Hopuare, M., M. Pontaud, J.-P. Céron, P. Ortega, and V. Laurent (2015), Climate change, Pacific climate drivers and observed
623 precipitation variability in Tahiti, French Polynesia, *Climate Research*, 63, 157-170, doi:10.3354/cr01288

624 Houze, R. A. (2012), Orographic effects on precipitating clouds, *Reviews of Geophysics*, 50, RG000365,
625 doi:10.1029/2011RG000365

626 Huang, S. P., S. T. Quek, and K. K. Phoon (2001), Convergence study of the truncated Karhunen–Loeve expansion for
627 simulation of stochastic processes, *International Journal for Numerical Methods in Engineering*, 52, 1029-1043,
628 doi:10.1002/nme.255

629 Hughes, J. P., and P. Guttorp (1999), A non-homogeneous hidden Markov model for precipitation occurrence, *Applied*
630 *Statistics*, 48, 15-30, <http://www.jstor.org/stable/2680815>

631 Jha, S. K., G. Mariethoz, J. Evans, M. F. McCabe, and A. Sharma (2014), A space and time scale-dependent nonlinear
632 geostatistical approach for downscaling daily precipitation and temperature, *Water Resources Research*, 51, 6244-
633 6261, doi: 10.1002/2014WR016729

634 Kleiber, W., R. W. Katz, and B. Rajagopalan (2012), Daily spatiotemporal precipitation simulation using latent and
635 transformed Gaussian processes, *Water Resources Research*, 48, doi:10.1029/2011WR011105

636 Krajewski, W. F., G. Ciach, and E. Habib (2003), An analysis of small-scale rainfall variability in different climatic regimes,
637 *Hydrological Sciences Journal*, 48, 151-162, doi:10.1623/hysj.48.2.151.44694

638 Lantuéjoul, C. (2002), *Geostatistical Simulation: Models and Algorithms*, ISBN: 978-3-662-04808-5

639 Laurent, V., K. Maamaatuaiahutapu, I. Brodien, S. Lombardo, M. Tardy, and P. Varney (2019), *Atlas climatologique de la*
640 *Polynésie française*, 232 pp., Météo France, Délégation Interrégionale de Polynésie Française, ISBN:9782111551916

641 Leblois, E., and J. D. Creutin (2013), Space-time simulation of intermittent rainfall with prescribed advection field: Adaptation
642 of the turning band method, *Water Resources Research*, 49, 3375-3387, doi:10.1002/wrcr.20190

643 Longman, R. J., H. F. Diaz, and T. W. Giambelluca (2015), Sustained Increases in Lower-Tropospheric Subsidence over the
644 Central Tropical North Pacific Drive a Decline in High-Elevation Rainfall in Hawaii, *Journal of climate*, 28, 8743-
645 8759, doi:10.1175/JCLI-D-15-0006.1

646 Longman, R. J., et al. (2018), Compilation of climate data from heterogeneous networks across the Hawaiian Islands, *Scientific*
647 *Data*, 5, 180012, doi:10.1038/sdata.2018.12

648 Longman, R. J., O. Elison Timm, T. W. Giambelluca, and L. Kaiser (2021), A 20-Year Analysis of Disturbance-Driven
649 Rainfall on O‘ahu, Hawai‘i, *Monthly Weather Review*, 6, 1767-1783, doi:10.1175/MWR-D-20-0287.1

650 Lyons, S. W. (1982), Empirical Orthogonal Function Analysis of Hawaiian Rainfall, *Journal of applied meteorology*, 21, 1713-
651 1729, doi:10.1175/1520-0450(1982)021<1713:EOFAOH>2.0.CO;2

652 Maraun, D., et al. (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between
653 dynamical models and the end user, *Reviews of Geophysics*, 48, doi:10.1029/2009RG000314

654 Marra, F., and E. Morin (2018), Autocorrelation structure of convective rainfall in semiarid-arid climate derived from high-
655 resolution X-Band radar estimates, *Atmospheric Research*, 200, 126-138, doi:10.1016/j.atmosres.2017.09.020

656 Mavromatis, T., and J. W. Hansen (2001), Interannual variability characteristics and simulated crop response of four stochastic
657 weather generators, *Agricultural and Forest Meteorology*, 109, 283-296, doi:10.1016/S0168-1923(01)00272-6

658 Mezghani, A., and B. Hingray (2009), A combined downscaling-disaggregation weather generator for stochastic generation
659 of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper
660 Rhone River basin, *Journal of Hydrology*, 377, 245-260, doi:10.1016/j.jhydrol.2009.08.033

661 Morin, E., T. Ryb, I. Gavrieli, and Y. Enzel (2019), Mean, variance, and trends of Levant precipitation over the past 4500
662 years from reconstructed Dead Sea levels and stochastic modeling, *Quaternary Research*, 91, 751-767,
663 doi:10.1017/qua.2018.98

664 Nerini, D., N. Besic, I. V. Sideris, U. Germann, and L. Foresti (2017), A non-stationary stochastic ensemble generator for radar
665 rainfall fields based on the short-space Fourier transform, *Hydrology and Earth System Sciences*, 21, 2777-2797,
666 doi:10.5194/hess-21-2777-2017

667 Niemi, T. J., J. H. A. Guillaume, T. Kokkonen, T. M. T. Hoang, and A. W. Seed (2016), Role of spatial anisotropy in design
668 storm generation: Experiment and interpretation, *Water Resources Research*, 52, 69-89, doi:10.1002/2015WR017521

669 Opitz, T., D. Allard, and G. Mariethoz (2021), Semi-parametric resampling with extremes, *Spatial statistics*, 42, 100445,
670 doi:10.1002/2015WR017521

671 Oriani, F., S. Stisen, M. C. Demirel, and G. Mariethoz (2020), Missing Data Imputation for Multisite Rainfall Networks: A
672 Comparison between Geostatistical Interpolation and Pattern-Based Estimation on Different Terrain Types, *Journal*
673 *of Hydrometeorology*, 21, 2325-2341, doi:10.1175/JHM-D-19-0220.1

674 Papalexiou, S. M., and F. Serinaldi (2020), Random Fields Simplified: Preserving Marginal Distributions, Correlations, and
675 Intermittency, With Applications From Rainfall to Humidity, *Water Resources Research*, 56, e2019WR026331,
676 doi:10.1029/2019WR026331

677 Paschalis, A., P. Molnar, S. Fatichi, and P. Burlando (2013), A stochastic model for high-resolution space-time precipitation
678 simulation, *Water Resources Research*, 49, doi:10.1002/2013WR014437.

679 Paschalis, A., S. Fatichi, P. M. Molnar, P. S., and P. Burlando (2014), On the effects of small scale space–time variability of
680 rainfall on basin flood response, *Journal of Hydrology*, 514, 313-327, doi:10.1016/j.advwatres.2013.11.006

681 Paulhus, J. L., and M. A. Kohler (1952), Interpolation of missing precipitation records, *Monthly Weather Review*, 80, 129-
682 133, doi:10.1175/1520-0493(1952)080<0129:IOMPR>2.0.CO;2

683 Peleg, N., C. Skinner, S. Fatichi, and P. Molnar (2020), Temperature effects on the spatial structure of heavy rainfall modify
684 catchment hydro-morphological response, *Earth Surface Dynamics*, 8, 17-36, doi:10.5194/esurf-8-17-2020

685 Réchou, A., O. Flores, G. Jumaux, V. Dufлот, O. Bousquet, C. Pouppeville, and F. Bonnardot (2019), Spatio-temporal
686 variability of rainfall in a high tropical island: Patterns and large-scale drivers in Réunion Island, *Quarterly Journal*
687 *of the Royal Meteorological Society*, 145, 893-909, doi:10.1002/qj.3485

688 Richardson, C. W. (1981), Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resources*
689 *Research*, 17, 182-190, doi:10.1029/WR017i001p00182

690 Rüschemdorf, L. (2009), On the distributional transform, Sklar’s theorem, and the empirical copula process, *Journal of*
691 *Statistical Planning and Inference*, 139, 3921-3927, doi:10.1016/j.jspi.2009.05.030

692 Sanfilippo, K. M. (2020), Predictor selection and model evaluation for future rainfall projection in Hawai’i, 124 pp, University
693 of Hawai’i at Mānoa, Honolulu, <http://hdl.handle.net/10125/73340>

694 Schleiss, M., S. Chamon, and A. Berne (2014), Nonstationarity in Intermittent Rainfall: The “Dry Drift”, *Journal of*
695 *Hydrometeorology*, 15, 1189-1204, doi:10.1175/JHM-D-13-095.1

696 Schwartz, G. (1978), Estimating the dimension of a model, *The Annals of Statistics*, 6, 461-464, doi: 10.1214/aos/1176344136

697 Scott, D. W. (1979), On optimal and data-based histograms, *Biometrika*, 66(3), 605-610, doi:10.1093/biomet/66.3.605

698 Scott, D. W. (2010), Scott’s rule, *WIREs Computational Statistics*, 2, 497-502, doi:10.1002/wics.103

699 Supit, I., C. A. van Diepen, A. J. W. de Wit, J. Wolf, P. Kabat, B. Baruth, and F. Ludwig (2012), Assessing climate change
700 effects on European crop yields using the Crop Growth Monitoring System and a weather generator, *Agricultural and*
701 *Forest Meteorology*, 164, 96-111, doi:10.1016/j.agrformet.2012.05.005

702 Volosciuk, C., D. Maraun, M. Vrac, and M. Widmann (2017), A combined statistical bias correction and stochastic
703 downscaling method for precipitation, *Hydrology and Earth System Sciences*, 21, 1693-1719, doi:10.5194/hess-21-
704 1693-2017

- 705 Vrac, M., M. Stein, and K. Hayhoe (2007), Statistical downscaling of precipitation through nonhomogeneous stochastic
706 weather typing, *Climate Research*, 34, 169-184, doi:10.3354/cr00696
- 707 Vu, T. M., A. K. Mishra, G. Konapala, and D. Liu (2018), Evaluation of multiple stochastic rainfall generators in diverse
708 climatic regions, *Stochastic Environmental Research and Risk Assessment*, 32, 1337-1353, doi:10.1007/s00477-017-
709 1458-0
- 710 Wilby, R. L. (1994), Stochastic weather type simulation for regional climate change assessment, *Water Resources Research*,
711 30, 3395-3403, doi:10.1029/94WR01840
- 712 Wilcox, C., C. Aly, T. Vischel, G. Panthou, J. Blanchet, G. Quantin, and T. Lebel (2021), Stochastorm: A Stochastic Rainfall
713 Simulator for Convective Storms, *Journal of Hydrometeorology*, 22, 387-404, doi:10.1175/JHM-D-20-0017.1
- 714 Wilks, D. S., and R. L. Wilby (1999), The weather generation game: a review of stochastic weather models, *Progress in*
715 *Physical Geography*, 23, 329-357, doi:10.1177/030913339902300302
- 716 Yiou, P. (2014), AnaWEGE: a weather generator based on analogues of atmospheric circulation, *Geoscientific Model*
717 *Development*, 7, 531-543, doi:10.5194/gmd-7-531-2014