https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

A non-stationary trans-Gaussian model for daily rainfall over complex topography

Lionel Benoit[1], Matthew P. Lucas[2], Denis Allard[1], Keri M. Kodama[3], and Thomas W. Giambelluca[2]

[1]Biostatistics and Spatial Processes (BioSP), INRAE, 84914 Avignon CEDEX 9, France
[2]Water Resources Research Center, University of Hawai‘i at Mānoa, Honolulu, Hawaii, USA
[3]University of Hawaii Sea Grant College Program, University of Hawaii at Mānoa, Honolulu, Hawaii, USA

Correspondence: Lionel Benoit (lionel.benoit@inrae.fr)

Abstract. The orographic effects that influence rainfall fields in mountainous regions depend on elevation and the exposure of the topography to prevailing winds. Transitions between wet and dry areas can occur within a few kilometers, creating strong horizontal gradients of various rainfall statistics such as the frequency of occurrence, the distribution of intensity and the structure of spatial correlation.

Most statistical models of daily rainfall assume spatial stationarity (i.e., the spatial homogeneity of rainfall statistics) and are therefore not well suited for studying the highly non-homogeneous characteristics of orographic rainfall. To overcome this limitation, we design a non-stationary trans-Gaussian geostatistical model for the analysis of daily rainfall fields over complex topography.

The modeling framework presented in this paper infers rainfall statistics from sparse rain gauge observations, simulates realistic rainfall fields after calibration and stochastically interpolates rain gauge observations to create rainfall maps. The performance of the model is assessed with data from the Island of Hawai‘i where extreme spatial gradients in rainfall are observed. The results presented in this paper demonstrate that a non-stationary trans-Gaussian model can skillfully reproduce orographic rainfall statistics as well as their variations in space.

1 Introduction

Air masses flowing across mountainous regions are displaced by the topography and forced into successive uplifts and downdrafts. The vertical displacement of air modulates the condensation within the air column, triggering or enhancing precipitation in the case of uplift and, conversely, inhibiting or attenuating rainfall in the case of downdraft. The impact of topography on precipitation is called the orographic effect (Roe, 2005; Houze, 2012) and results in windward slopes being generally wetter than leeward slopes and the highlands generally wetter than the lowlands (Foresti et al., 2018). Additionally, fluctuations in the direction and intensity of prevailing winds combined with changes of precipitation regimes lead to complex patterns of orographic rainfall with strong spatial gradients between wet and dry areas (Giambelluca et al., 2013; Isotta et al., 2014; Benoit and Sichoix, 2023).

Stochastic rainfall models (SRMs) are probabilistic tools that simulate synthetic rainfall datasets with statistical properties that resemble those from observations. These models first infer the probability distribution of rainfall from observations and

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

then generate synthetic datasets by random sampling of this distribution, a process known as stochastic simulation (Wilks and Wilby, 1999; Ailliot et al., 2015). When applied to mountainous regions, SRMs help improve our understanding of orographic rainfall and its variability (Benoit et al., 2022). SRMs can be used to create a database of design storms for dam or drainage system sizing (Niemi et al., 2016), to investigate the effects of different rainfall attributes on the hydrological system (Paschalis et al., 2014), or to assess statistical properties of extreme rainfall (e.g., the intensity-duration-frequency curve or the tail behavior of the intensity distribution for different time aggregation levels) (Evin et al., 2018).

Hydro-meteorological applications such as precipitation mapping (Foehn et al., 2018) or distributed hydrological modeling (Moraga et al., 2022) require the generation of gridded rainfall products from unevenly distributed and sparse observations that are typically recorded by rain gauges. Geostatistics have long been used for the spatial analysis of rainfall fields (Creutin and Obled, 1982; Goovaerts, 2000; Kyriakidis et al., 2001) because they use Gaussian random fields that can model rainfall at any point of the spatial domain while accounting for spatial dependencies (Chilès and Delfiner, 2012). To attain daily to sub-daily resolutions, trans-Gaussian geostatistical methods that apply a parametric transform function to a Gaussian random field have been developed to account for precipitation intermittency and the non-Gaussian distribution of rainfall intensity (Benoit and Mariethoz, 2017; Benoit et al., 2018a; Papalexiou and Serinaldi, 2020).

In mountainous regions, the performance of trans-Gaussian geostatistical models is constrained by the complexity of modeling the spatial variation of rainfall statistics caused by orographic effects. To better model spatially varying statistics, some parameters of the model have to be made location-dependent, a condition referred to as spatial non-stationarity (Sampson and Guttorp, 1992). For example, Frazier et al. (2016) and Lucas et al. (2022) used a non-stationary mean when mapping monthly rainfall in Hawai‘i, whereas Paciorek and Schervish (2006) and Fuglstad et al. (2015) used a non-stationary covariance when mapping annual precipitation in, respectively, Colorado and the continental United States.

However, making one or more parameters non-stationary greatly increases the overall number of model parameters because a statistical property that is modeled by a single parameter in the stationary case will require several parameters in the non-stationary case. As a result, non-stationary geostatistical models tend to have a large number of parameters and require significantly more training data than their stationary counterparts. To limit model complexity, non-stationary rainfall models often focus on one aspect of non-stationarity, the choice of which is driven by the problem at hand.

In contrast, this paper introduces a fully non-stationary trans-Gaussian geostatistical model for daily rainfall, which assumes that all statistics of the model are non-stationary. This is particularly relevant in mountainous regions where orographic effects make most rainfall statistics non-stationary. The performance of the model is illustrated for the Island of Hawai‘i (State of Hawaii, USA) where orographic effects are very strong (Giambelluca et al., 2013), and for which an extensive and high-quality rain gauge dataset enables the calibration of a highly parametrized model.

The rest of the paper describes the design and assessment of the proposed geostatistical model and is structured as follows: Section 2 describes the trans-Gaussian model being used, explains how to make it non-stationary in space and proposes a method for model calibration. Section 3 applies the model to the simulation of orographic rainfall on the Island of Hawai‘i and the performance of the model is assessed with respect to two applications: the stochastic generation of synthetic rainfall

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

fields and rainfall mapping. Finally, Section 4 discusses the advantages and limitations of the model with respect to orographic rainfall modeling and proposes some lines of inquiry for future research.

2 Non-stationary trans-Gaussian rainfall model

In the following, we model daily rainfall as a stochastic process studied over a spatial domain $\mathcal{D}$: $R(\mathbf{s})$ with $\mathbf{s} \in \mathcal{D} \subset \mathbb{R}^2$. As the focus of the study is the spatial modeling of daily rainfall, it is assumed that consecutive daily rainfall fields are independent of each to reduce the complexity of the model, and the temporal variability of rainfall patterns is modeled separately (and beforehand) following a rain-typing approach (Ailliot et al., 2015; Benoit et al., 2018b; Vaittinada Ayar et al., 2020). We therefore assume that consecutive daily rainfall fields are independent, but that they can be pooled into N_{clust} clusters referred to hereafter as rain types, within which rainfall fields are statistically similar to each other. All the information about the temporal variability of rainfall is encoded into the sequence of rain types, which as been shown to be a reasonable assumption for daily rainfall in tropical islands, and in particular in Hawaiʻi (Benoit et al., 2022).

The spatial modeling of rainfall will therefore be performed independently for each rain type, and for the sake of simplicity we assume a single rain type for the description of the model in Sections 2.1 to 2.5. How to deal with multiple rain types when conditioning daily rainfall maps to monthly totals will be addressed in Section 2.6, and the practical delineation of rain types from an actual dataset will be addressed in Section 3.2.

2.1 Trans-Gaussian geostatistics applied to rainfall modeling

When observed at a single site and at daily resolution, rainfall is featured by a large number of zeros (i.e. dry days) and a distribution of non-zero rainfall intensities that strongly differs from the normal distribution (Benoit et al., 2022). In addition, when observed simultaneously at neighboring locations, rainfall is structured spatially with inter-site dependencies decreasing with the separation distance and vanishing at distances of dozens to hundreds of kilometers depending on the regional climate (Creutin et al., 1988; Guillot and Lebel, 1999). In this study we adopt a trans-Gaussian approach to model the above features of rainfall (Allard and Bourotte, 2015; Benoit et al., 2018a; Papalexiou and Serinaldi, 2020; Vaittinada Ayar et al., 2020). Trans-Gaussian geostatistics assume that a rainfall observation $R(\mathbf{s})$ performed at location $\mathbf{s} \in \mathcal{D}$ originates from a latent and standardized Gaussian random variable $Y(\mathbf{s})$ through a transform function Ψ which combines a truncation at threshold a_0 and a monotonic transformation ψ of the non-censored part of the latent variable. The spatial dependencies observed between a set of N rain observations performed at different locations $R(\mathbf{s}_i), i = 1, ..., N_S$ are modeled by assuming that the associated latent variables $Y(\mathbf{s}_i), i = 1, ..., N$ form a random vector $\mathbf{Y} = [Y(\mathbf{s}_1), Y(\mathbf{s}_2), ..., Y(\mathbf{s}_N)]$ that follows a standardized multivariate Gaussian distribution with covariance function C_Y. Here we assume a Gamma distribution for single-site rain intensity (Wilks and Wilby, 1999; Berrocal et al., 2008), and a Gaussian latent random field with a Matérn covariance function for inter-sites dependencies (Allard and Bourotte, 2015; Porcu et al., 2024):

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

$$\begin{cases} R(\mathbf{s}) = 0 \quad \text{if} \quad Y(\mathbf{s}) \leq a_0 \\ R(\mathbf{s}) = \text{Gamma}^{-1}(\Phi(Y(\mathbf{s})); k, \theta) \quad \text{if} \quad Y(\mathbf{s}) > a_0 \end{cases} \tag{1}$$

With $\quad Y \sim \mathcal{N}(0, C_Y) \quad$ and $\quad C_Y(||\mathbf{h}||; \nu, \rho) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{||\mathbf{h}||}{\rho}\right)^{\nu} \mathcal{K}_\nu \left(\frac{||\mathbf{h}||}{\rho}\right)$

where Gamma$(.; k, \theta)$ is the cumulative distribution function (cdf) of the Gamma distribution with parameters $k > 0$ (shape) and $\theta > 0$ (scale); Φ is the cdf of the standardized normal distribution; $\mathbf{h}$ is a separation vector between two locations; Γ is the usual gamma function; $\mathcal{K}_\nu$ is the modified Bessel function of the second kind; and $\nu > 0$ and $\rho > 0$ are the two parameters of the Matérn covariance function C_Y. The shape parameter ν controls the regularity of the latent process Y at short separation lags, while the range parameter ρ controls the variability of Y between distant locations.

2.2 Making the trans-Gaussian model non-stationary

The trans-Gaussian model of equation 1 is made non-stationary in space by allowing the parameters of the marginal distribution and those of the dependence structure to vary through space. The parameters of the transform function Ψ (i.e., a_0, k and θ) are defined for each observation location separately, and interpolated at ungauged locations. Similarly the parameters of the covariance function C_Y (i.e., ρ and ν) are first defined independently for a set of climate zones (Luo et al., 2024) within which the covariance is assumed locally stationary, and then interpolated between the climate zones. Since the resulting non-stationary covariance must be a valid covariance function to enable geostatistical simulations (Lantuéjoul, 2002), we adopt the model of Paciorek and Schervish (2006) to derive a parametric form for a valid non-stationary and anisotropic (i.e., ρ depends on the direction of $\mathbf{s}_j - \mathbf{s}_i$) covariance function of Matérn type:

$$C_Y(\mathbf{s}_i, \mathbf{s}_j) = \frac{2^{1-(\nu_i+\nu_j)/2}}{\sqrt{\Gamma(\nu_i)\Gamma(\nu_j)}} |\mathbf{\Sigma}_i|^{\frac{1}{4}} |\mathbf{\Sigma}_j|^{\frac{1}{4}} \left|\frac{\mathbf{\Sigma}_i + \mathbf{\Sigma}_j}{2}\right|^{-\frac{1}{2}} \left(\sqrt{Q_{ij}}\right)^{\frac{\nu_i+\nu_j}{2}} \mathcal{K}_{\frac{\nu_i+\nu_j}{2}} \left(\sqrt{Q_{ij}}\right) \tag{2}$$

with $\quad \mathbf{\Sigma}_i = \mathbf{V}_i \times \mathbf{\Lambda}_i \times {\mathbf{V}_i}^T; \quad \mathbf{V}_i = \begin{bmatrix} \frac{\gamma_{1,i}}{\sqrt{{\gamma_{1,i}}^2 + {\gamma_{2,i}}^2}} & -\frac{\gamma_{2,i}}{\sqrt{{\gamma_{1,i}}^2 + {\gamma_{2,i}}^2}} \\ \frac{\gamma_{2,i}}{\sqrt{{\gamma_{1,i}}^2 + {\gamma_{2,i}}^2}} & \frac{\gamma_{1,i}}{\sqrt{{\gamma_{1,i}}^2 + {\gamma_{2,i}}^2}} \end{bmatrix}; \quad \mathbf{\Lambda_i} = \begin{bmatrix} {\gamma_{1,i}}^2 + {\gamma_{2,i}}^2 & 0 \\ 0 & \rho_{2,i} \end{bmatrix}$

and with $\quad Q_{ij} = (\mathbf{s}_i - \mathbf{s}_j)^T \left(\frac{\mathbf{\Sigma}_i + \mathbf{\Sigma}_j}{2}\right)^{-1} (\mathbf{s}_i - \mathbf{s}_j)$

where $\mathbf{s}_i$ and $\mathbf{s}_j$ are two observation locations, $\gamma_{1,i}$ and $\gamma_{2,i}$ are the two components of the vector defining the major anisotropy axis at location $\mathbf{s}_i$, and $\rho_{2,i}$ is the range parameter along the minor anisotropy axis at the same location.

2.3 Model calibration

The parameters of the model are first estimated at the local scale assuming local stationarity, and then interpolated at the regional scale to create the non-stationary model. The parameter estimation process starts with the estimation of the parameters

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

of the transform function at each rain gauge separately. The truncation threshold $a_0(\mathbf{s})$ is obtained from the proportion $d(\mathbf{s})$ of dry measurements observed at location $\mathbf{s}$ by simply inverting the cdf of a standardized normal distribution $a_0(\mathbf{s}) = \Phi^{-1}(d(\mathbf{s}))$, and then the parameters of the marginal distribution of non-zero daily intensities (i.e. k and θ) are estimated by marginal likelihood maximization. The log-likelihood of the parameters k and θ given a set of rain observations $R(\mathbf{s}_i) > 0$, $i = 1, ..., N_s$ assumed independent and identically distributed following a Gamma distribution is:

$$l(k,\theta) = \sum_{i=1}^{N_s} \left(\ln \left(\frac{1}{\Gamma(k)\theta^k} R(\mathbf{s}_i)^{k-1} e^{-\frac{R(\mathbf{s}_i)}{\theta}} \right) \right) \tag{3}$$

The parameters of the covariance function C_Y of the latent field are subsequently estimated for each climate zone separately using pairwise likelihood maximization. Given a pair of rain observations $R(\mathbf{s}_i)$ and $R(\mathbf{s}_j)$ the log-likelihood of the parameters ν, γ_1, γ_2 and ρ_2 can be written as (Allard and Bourotte, 2015):

$$l(\nu,\gamma_1,\gamma_2,\rho_2) = \ln(\Phi_2(a_0,a_0;C_Y)) \quad \text{if} \quad R(\mathbf{s}_i) = R(\mathbf{s}_j) = 0 \tag{4}$$

$$l(\nu,\gamma_1,\gamma_2,\rho_2) = \ln(\Phi(a_0) - \Phi_2(a_0,a_0;C_Y)) + \ln(\phi(Y(\mathbf{s}_j))) \quad \text{if} \quad R(\mathbf{s}_i) = 0 \quad \text{and} \quad R(\mathbf{s}_j) > 0$$

$$l(\nu,\gamma_1,\gamma_2,\rho_2) = \ln(\phi_2(Y(\mathbf{s}_i),Y(\mathbf{s}_j))) \quad \text{if} \quad R(\mathbf{s}_i) > 0 \quad \text{and} \quad R(\mathbf{s}_j) > 0$$

where Φ_2 and ϕ_2 are respectively the cdf and the probability density function (pdf) of the bivariate standard normal distribution, ϕ is the pdf of the univariate standard normal distribution, and $Y(\mathbf{s}_i)$ is the latent counterpart of the rainfall observation $R(\mathbf{s}_i)$.

Once the parameters of the model are known for each rain gauge and climate zone, they can be propagated to the entire domain of interest by spatial interpolation. Here we use ordinary Kriging with a nugget term and a Matérn covariance function to interpolate the parameters of the transform function Ψ (Bennett et al., 2018), and a squared inverse distance interpolation for the parameters of the covariance function because in the targeted application the small number of climate zones does not allow for more complex interpolation techniques such as Kriging (Babak and Deutsch, 2009).

2.4 Simulating synthetic rainfall fields

Synthetic rainfall fields reproducing the statistics of the training dataset can be generated by unconditional simulations performed at a set of target locations $\mathbf{u}_i \in \mathcal{D}$, $i = 1, ..., N_T$, often located on a regular grid to comply with the requirement of impact models. To this end, an ensemble of simulations of the latent field $\mathbf{Y}_{us}$ is first generated by unconditional geostatistical simulations (Lantuéjoul, 2002). A simple way to draw such simulations is to multiply the lower triangular matrix resulting from the Cholesky decomposition of the covariance matrix C_Y evaluated at the target locations by an independent and identically distributed standardized normal vector (Davis, 1987). However, standard algorithms for the Cholesky decomposition of the covariance matrix have a complexity of $\mathcal{O}({N_T}^3)$ point operations and a memory footprint of $\mathcal{O}({N_T}^2)$ (Abdulah et al., 2019), which becomes prohibitive when more than a few thousands of simulation points are intended. Hence, when more target points are requested, the above simulation method must be replaced by a more scalable approach, such as a spectral simulation

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

method (Emery and Arroyo, 2018). After the simulation of a sample of the latent field $\mathbf{Y}_{us}$, a synthetic rainfall field $\mathbf{R}_{us}$ is obtained by applying the local transform function at every target location $R_{us}(\mathbf{u}) = \Psi_{\mathbf{u}}(Y_{us}(\mathbf{u}))$.

2.5 Rainfall mapping

Rainfall maps and associated uncertainty estimations can be obtained by the stochastic interpolation of the rain gauge observations. To this end, the synthetic rainfall fields $\mathbf{R}_{us}$ described above are modified by a process called conditioning that forces them to reproduce rain gauge observations. An efficient way to perform conditioning in the framework of Geostatistics is conditioning by Kriging (Chilès and Delfiner, 2012). Since conditioning by Kriging requires a target variable with a multivariate Gaussian distribution, it should be performed on the latent field $\mathbf{Y}$. In turn, conditioning the synthetic rainfall fields $\mathbf{R}_{us}$ therefore requires the computation of the latent counterparts of the observations $Y(\mathbf{s}) = \Psi_{\mathbf{s}}^{-1}(R(\mathbf{s}))$, which are referred to as pseudo-observations. For the observation locations $\mathbf{s}_w$ where some rainfall is recorded (i.e., $R(\mathbf{s}_w) > 0$), the pseudo-observation is obtained by simply inverting the parametric transform function ψ. In contrast, at observation locations $\mathbf{s}_d$ where no rainfall is recorded (i.e., $R(\mathbf{s}_d) = 0$), the pseudo-observation $Y(\mathbf{s}_d)$ is not explicitly defined because the latent field Y is censored. To circumvent this problem we propose to use a Gibbs sampler to simulate the censored latent values $Y(\mathbf{s}_d)$ conditional to the uncensored latent values $Y(\mathbf{s}_w)$ under the constraint that $Y(\mathbf{s}_d) < a_0(\mathbf{s}_d)$ (Lantuéjoul, 2002; Benoit et al., 2018a). The Gibbs sampler algorithm adapted to the simulation of the censored pseudo-observations of rainfall is detailed in Appendix A. Once the pseudo-observations $Y(\mathbf{s})$ are known at all observation locations, conditioning by Kriging is performed by adding to the original unconditional simulations $\mathbf{Y}_{us}$ the result of the simple Kriging of the residuals between the unconditional simulations of the latent field $Y_{us}(\mathbf{s})$ and the pseudo-observations $Y(\mathbf{s})$:

$$Y_{cs}(\mathbf{u}) = Y_{us}(\mathbf{u}) + \sum_{i=1}^{N_s} \lambda_i^{SK}(\mathbf{u})\left(Y(\mathbf{s}_i) - Y_{us}(\mathbf{s}_i)\right) \tag{5}$$

with $\lambda_i^{SK}(\mathbf{u})$ being the solution of the linear Simple Kriging system of equations (Chilès and Delfiner, 2012):
$\sum_{j=1}^{N_s} \lambda_j^{SK}(\mathbf{u}) C_Y(\mathbf{s}_i, \mathbf{s}_j) = C_Y(\mathbf{u}, \mathbf{s}_i), \forall i = 1..N_s$

Finally the conditional rainfall field simulations are retrieved by applying the local transform function at each target location: $R_{cs}(\mathbf{u}) = \Psi_{\mathbf{u}}(Y_{cs}(\mathbf{u}))$, and the uncertainty of the stochastic interpolation is assessed by evaluating the dispersion of a large ensemble of conditional simulations.

2.6 Conditioning daily rainfall maps to monthly totals

The uncertainty in daily rainfall maps derived from the stochastic interpolation of rain gauge observations (i.e., R_{cs}) increases with the distance to the rain gauges, and the variance of the interpolation uncertainty tends to the variance of the rainfall signal itself at grid points far from any rain gauge. This leads to a high uncertainty of rainfall interpolation in sparsely gauged areas and at the edges of the interpolation domain. To reduce the uncertainty of daily rainfall maps in data-poor areas, it is possible to generate daily rainfall maps conditioned not only to daily rain gauge observations, but also to monthly rainfall totals derived for

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

instance from monthly rainfall maps incorporating additional observations recorded by rain gauges operating at the monthly resolution, as well as information about long-term rainfall patterns (Giambelluca et al., 2013; Frazier et al., 2016; Lucas et al., 2022).

In the present framework, conditioning daily rainfall maps to monthly totals is made complicated by the fact that the daily rainfall fields can originate from different rain types within a given month. This implies resorting to different statistical models for the spatial modeling of the daily rainfall fields, and makes it difficult to use direct conditioning approaches. We therefore propose to use a Markov Chain Monte Carlo (MCMC) simulation approach to deal with the complex statistical setup imposed by the coexistence of multiple rain types (Bárdossy and Pegram, 2016), and resort to the Metropolis within Gibbs algorithm detailed in Appendix B to condition the daily rainfall simulations to monthly totals. The main limitation of this algorithm is a low computational efficiency, which is typical of MCMC techniques (Gilks et al., 1996). In consequence, the Metropolis within Gibbs algorithm is only applied to a restricted set of locations, which are selected in such a way that the density of conditioning locations becomes relatively homogeneous in space. This leads to the following processing steps, which will be illustrated throughout Section 3:

1. Identify and delineate rain types within which daily rainfall statistics and spatial patterns are similar (described in Sect. 3.2).
2. Infer the parameters of the stochastic rainfall model for each rain type independently (described in Section 3.2).
3. Generate a large ensemble of unconditional rainfall field simulations, which reproduce the statistics of daily rainfall observations (described in Section 3.3).
4. Condition these synthetic rainfall fields to daily rainfall observations to generate daily rainfall maps and assess the uncertainty of rainfall interpolation (described in Section 3.4).
5. Identify areas where rainfall maps are highly uncertain, simulate daily rainfall conditioned to monthly totals at these locations, and use these simulations at *virtual stations* along with daily rain gauge observations to generate improved daily rainfall maps (described in Section 3.5).

3 Model assessment and application

3.1 Orographic precipitation on the Island of Hawai‘i

The non-stationary trans-Gaussian rainfall model is assessed on the Island of Hawai‘i (state of Hawaii, USA, lon=155.5°W, lat=19.5°N) where orographic effects are very strong. The following case study resorts to daily observations from 2000 to 2019 recorded by a network of 79 rain gauges distributed across the island. We use quality-controlled and gap-filled data from (Longman et al., 2018), which are now publicly available on the Hawai‘i Climate Data Portal (HCDP, (Longman et al., 2024), https://www.hawaii.edu/climate-data-portal/).

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

Due to the location of Hawai‘i in the middle of the subtropical Pacific (Fig 1.a), the climate of the island is shaped by prevailing easterly trade winds which blow 85%-95% of the time in summer and 50%-80% of the time in winter (Longman et al., 2021). When they reach the Island of Hawai‘i, trade winds produce the orographic lifting of moist oceanic air masses, which triggers shallow convection leading to frequent and abundant rain showers in the east facing slopes and generally dry conditions on the leeward west side of the island (Fig 1.b) (Giambelluca et al., 2013). An exception to these generally dry weather conditions is a distinct rainfall maximum in the Kona area (West from Mauna Loa, at around 5 km from the coast) which is the result of mechanically diverted tradewinds being directed upslope, enhanced in the afternoon by strong solar heating of the west-facing slope (Huang and Chen, 2019). In Hawai‘i weather conditions are often affected by the presence of the trade wind inversion (TWI, average inversion base height around 2200 m) that caps the moist oceanic air and creates very dry conditions at high altitudes located in the center of the island (Fig 1.b) (Longman et al., 2015). This predominant pattern of precipitation is drastically modified when atmospheric disturbances originating from the mid-latitudes (e.g., cold fronts or Kona lows) eliminate the TWI, produce southerly, southwesterly, or westerly winds at the regional scale, and allow deep convection to develop. Under atmospheric disturbance conditions, widespread rainfall occurs throughout the island, and the west-facing slopes as well as the high altitude locations, which are usually dry, receive a significant part of their annual precipitation (Longman et al., 2021).

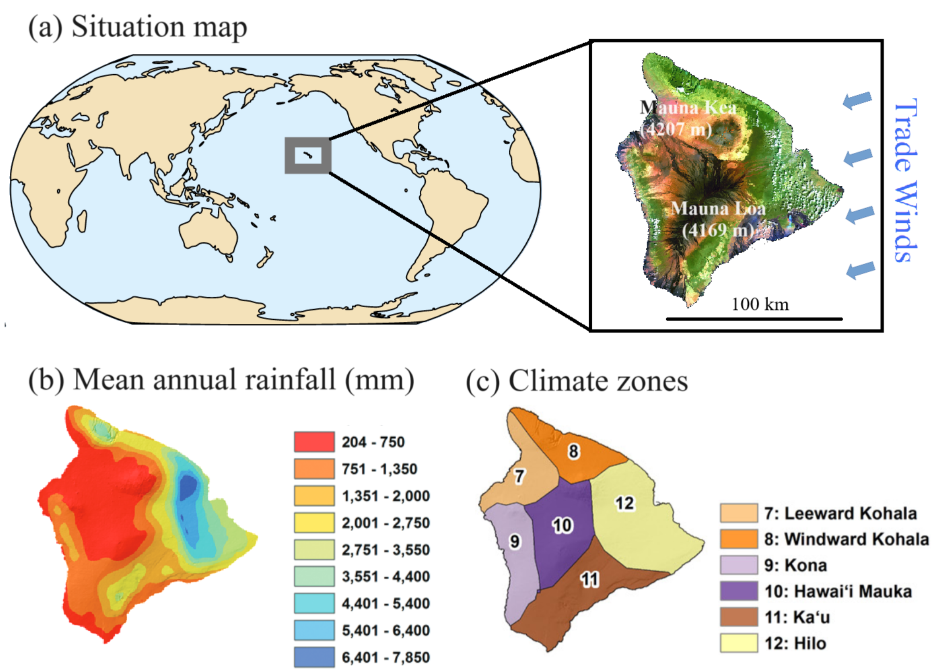

Figure 1. Geographical and climatological context of the Island of Hawai‘i. (a) Situation map and main topo-climatic features (adapted from https://en.wikipedia.org/wiki/Hawaii_(island)), (b) mean annual rainfall (adapted from Giambelluca et al. (2013)), and (c) climate zones (adapted from Luo et al. (2024); the numbering over the Island of Hawai‘i starts at 7 because the original zonation covers the whole Hawai‘i archipelago).

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

3.2 Model settings

To acknowledge the temporal variability of the rainfall patterns described above, we split the 20-year dataset into subsets within which the statistics of rainfall are as homogeneous as possible (Benoit et al., 2018b; Vaittinada Ayar et al., 2020). To this end, the main patterns of daily precipitation over Hawai‘i are delineated by unsupervised clustering of days based on rain gauge observations. We follow the approach of Benoit et al. (2022) and use the following indicators to characterize daily rainfall fields at the scale of the island: (1) the proportion of rain gauges recording a wet day, (2-3) the shape and skewness parameters of the distribution of daily rainfall intensity recorded across the island, and (4-6) the first three components of the Karhunen–Loève expansion of the (preliminarily standardized) rain gauge observations. Based on these indicators, the clustering is performed using a Gaussian Mixture Model (GMM; see e.g., (Fraley and Raftery, 2002)), which approximates the joint pdf of the indicators as a weighted sum of multivariate normal distributions. In the present case the six indicators characterizing rainfall intensity (1-3) and rainfall spatial distribution (4-6) are assumed to be only slightly correlated to each other and the covariance matrices used in the GMM are therefore assumed to be diagonal. The number of clusters, hereafter referred to as rain types, is set to six as a compromise between taking into account the diversity of rainfall patterns and keeping the analysis parsimonious.

The rainfall data from the six rain types are processed separately, and one rainfall model is set up for each cluster. Hence, the observation dataset is split into rain types prior to model calibration, and one set of model parameters is estimated for each rain type separately in order to account for the temporal variability of daily rainfall patterns. Figure 2 displays for each rain type the seasonality of occurrence, the spatial patterns of mean daily rainfall observed by the rain gauge network, as well as the estimated model parameters. In addition, examples of daily rainfall fields corresponding to each rain type are shown in Supplementary Material. Rain typing results show strong contrasts in seasonality and spatial distribution of rainfall between rain types, which translates into contrasting maps of model parameters. It is interesting to notice that the stronger spatial dependencies (denoted by bigger ellipses of anisotropy in figure 2, right column) occur in the North-East of the island, in the wet windward areas. In these areas the spatial dependencies tend to be significantly stronger in the direction parallel to the topography than across it, which is in line with strong rainfall gradients being correlated with gradients of topography. One can also notice the prevalence of rain type 5 during summer months, which brings us to link this rain type with trade wind conditions. This view is reinforced by the observed pattern of daily rainfall, which combines all distinctive marks of trade winds induced orographic rainfall, namely moderately wet windward (East facing) slopes, very dry inland areas because of the presence of the TWI, and some thermally driven rain showers on the Kona coast on the west side of the island (Huang and Chen, 2019).

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

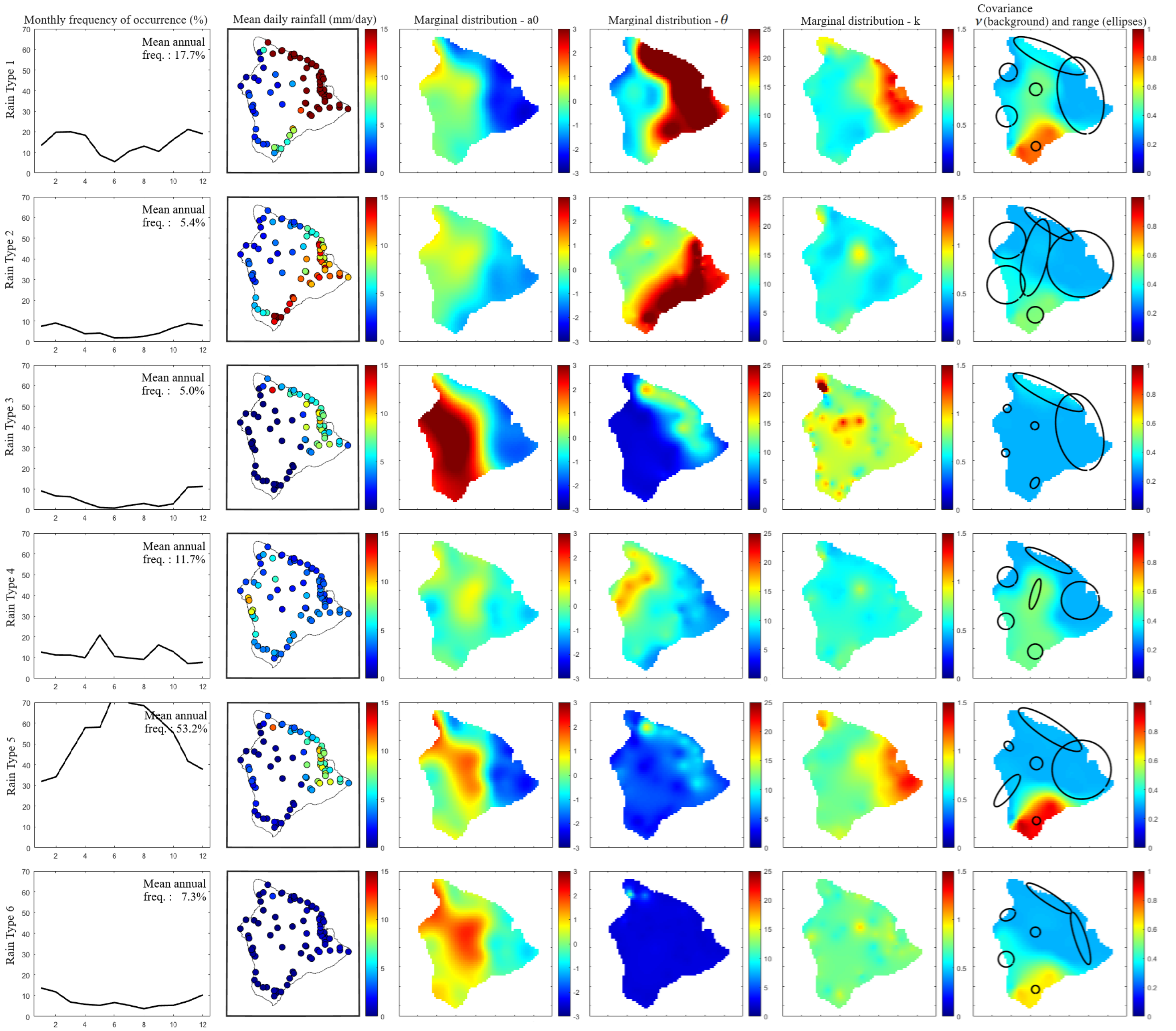

Figure 2. Daily rainfall patterns of the six rain types of the Island of Hawai‘i and associated model parameters. Different rows denote different rain types. The first column shows the seasonality of rain type occurrence, column 2 displays the mean daily rainfall, columns 3-5 display maps of the estimated marginal parameters, and column 6 displays maps of the estimated covariance parameters, namely the shape parameter ν of the covariance function (background color) and ellipses of anisotropy (overprint) that denote the value of the range parameter ρ for different directions and for each climate zone.

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

3.3 Stochastic generation of orographic rainfall

We first assess the ability of our model to capture and reproduce the statistical signature of rainfall at gauge locations, which corresponds to setting-up and deploying a multi-site stochastic rainfall generator. Synthetic rainfall is generated by unconditional simulation (cf. Sect. 2.4) performed on an irregular grid whose nodes are the locations of the rain gauges. For comparison purposes the simulation is repeated the same number of times as the number of days within each rain type, and the 7305 daily simulations are ordered and concatenated to create a 20-year time series with the same rain type timeline as the observations. The statistical distribution of the simulated daily rainfall is subsequently mapped and compared to the distribution of observations. The following statistics are used to assess the distribution of rainfall at each site: the mean daily rainfall, the frequency of rainfall occurrence, the median of daily rainfall during wet days (with wet days defined at the rain gauge level, i.e., the day d is considered as wet at location $\mathbf{s}$ if $R(\mathbf{s},d) > 0$), and the quantile 95% of daily rainfall during wet days. In addition, the whole distributions of simulations and observations are compared at each location using the 1-Wasserstein distance (Panaretos and Zemel, 2019):

$$W_1(\mathbf{R}_{obs},\mathbf{R}_{sim}) = \sum_{i=1}^{n} |\mathbf{R}_{obs(i)} - \mathbf{R}_{sim(i)}| \tag{6}$$

where $\mathbf{R}_{obs(i)}$ (resp. $\mathbf{R}_{sim(i)}$) is the element of rank i of the vector $\mathbf{R}_{obs}$ (resp. $\mathbf{R}_{sim}$).

The performance of the non-stationary trans-Gaussian model is benchmarked against the stochastic rainfall model of Benoit et al. (2022) (hereafter refereed to as BSNLG2022) which has been designed specifically for tropical islands with complex topography. Here the BSNLG2022 model is calibrated using the same training dataset and the same rain type time series as the ones used to calibrate non-stationary trans-Gaussian model (cf. Sect. 3.1).

Figure 3 compares the pointwise statistical signature of rainfall observations for the period 2000-2019 with the one of the 20-year synthetic rainfall simulations performed by the non-stationary trans-Gaussian model and by the BSNLG2022 benchmark model. Results show that both models reproduce very well the four diagnostic statistics, and that the non-stationary trans-Gaussian model slightly outperforms BSNLG2022 in terms of 1-Wasserstein distance with observations, in particular in places where the gradients of rainfall statistics are the strongest (i.e., the northern and eastern endpoints of the island). The only area where the proposed model is less effective than BSNLG2022 is on the East side of the island where the density of rain gauges is the highest. In this area one can notice that our model does not perfectly captures the strong variability of rain statistics between very nearby locations, and tends to slightly over-smooth rainfall statistics in space. This can be explained by the inclusion of a nugget term in the Kriging system used to interpolate the parameters of the transform function Ψ (cf. Sect. 2.3), which is necessary to filter local errors during the interpolation but leads to parameter maps that do not exactly honor point parameter estimates, and therefore results in rainfall simulations that do not perfectly reproduce rainfall statistics at observation locations. Apart for this minor over-smoothing of rain statistics in densely gauged areas, our model is able to simulate synthetic multi-site daily rainfall events that closely mimic observations.

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

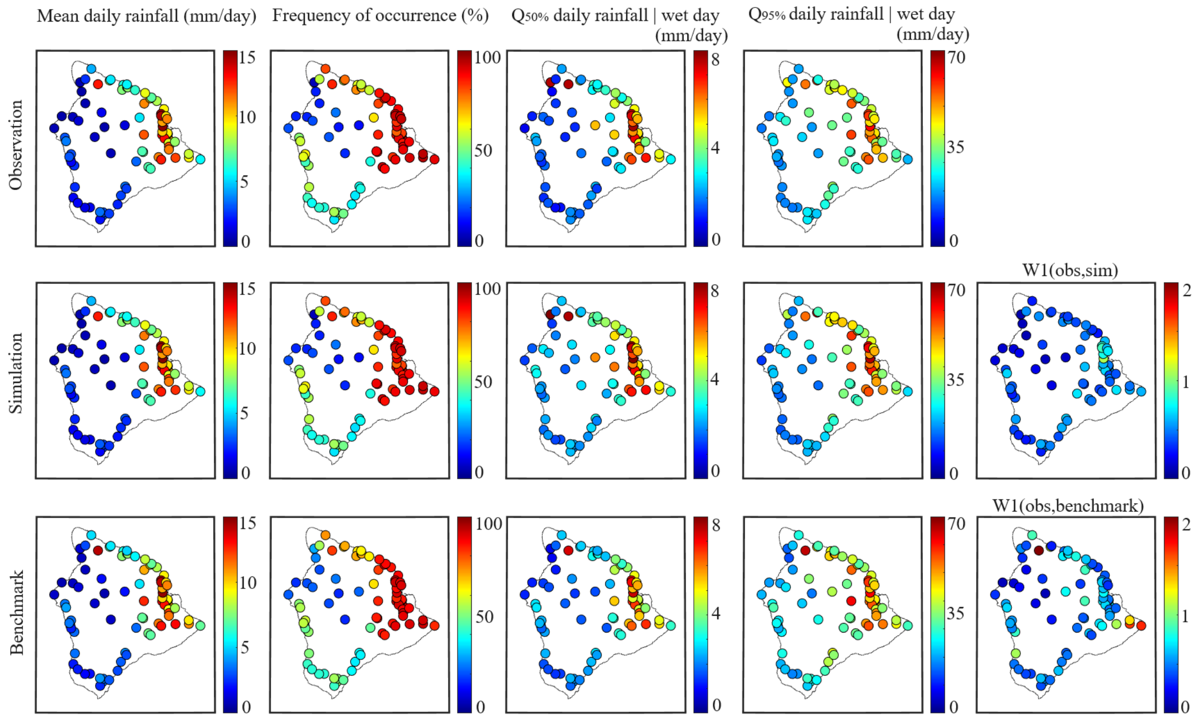

Figure 3. Statistical signature of rainfall derived from observations (top row), simulations performed by the non-stationary trans-Gaussian model (middle row), and simulations performed by the BSNLG2022 benchmark model (bottom row). The first column displays the mean daily rainfall, column 2 displays the frequency of rainfall occurrence, columns 3 and 4 display the quantiles 50% and 95% of daily rainfall during wet days, and column 5 displays the 1-Wasserstein distance between observations and simulations.

We next assess the ability of the model to simulate realistic spatial patterns of rainfall occurrence and intensity. To explore rainfall patterns we map the similarity of rainfall time series simulated at three specific locations with the time series simulated in the rest of the target area. For this experiment the non-stationary trans-Gaussian model is set-up to simulate rainfall on a 2 km-resolution regular grid in order to explore its skills in simulating rainfall fields that are continuous in space, while the BSNLG2022 benchmark model remains multi-site rather than continuous because it is by construction unable to interpolate rainfall between observation locations. The similarity of time series of rainfall occurrence is quantified by the Jaccard Index of Rainfall Occurrence (JI-RO) (Jaccard, 1901), while the similarity of time series of rainfall intensity is quantified by the Pearson Correlation Coefficient of Rainfall Intensity (PCC-RI, evaluated solely on wet days):

$$\mathrm{JI-RO}(\mathbf{s}_i,\mathbf{s}_j) = \frac{M_{11}}{M_{01}+M_{10}+M_{11}} \text{ with :} \tag{7}$$

$$M_{11} = \sum_{t=1}^{N_T} \mathbb{1}\left[R(\mathbf{s}_i,t)>0\right].\mathbb{1}\left[R(\mathbf{s}_j,t)>0\right]$$

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

$$M_{01} = \sum_{t=1}^{N_T} \mathbb{1}\left[R(\mathbf{s}_i,t)=0\right].\mathbb{1}\left[R(\mathbf{s}_j,t)>0\right]$$

$$M_{10} = \sum_{t=1}^{N_T} \mathbb{1}\left[R(\mathbf{s}_i,t)>0\right].\mathbb{1}\left[R(\mathbf{s}_j,t)=0\right]$$

$$\mathrm{PCC-RI}(\mathbf{s}_i,\mathbf{s}_j) = \frac{\sum_{t_w=1}^{N_w}\left(R(\mathbf{s}_i,t_w)-\bar{R}(\mathbf{s}_i)\right)\left(R(\mathbf{s}_j,t_w)-\bar{R}(\mathbf{s}_j)\right)}{\sqrt{\sum_{t_w=1}^{N_w}\left(R(\mathbf{s}_i,t_w)-\bar{R}(\mathbf{s}_i)\right)^2}\sqrt{\sum_{t_w=1}^{N_w}\left(R(\mathbf{s}_j,t_w)-\bar{R}(\mathbf{s}_j)\right)^2}} \tag{8}$$

with $\bar{R}(\mathbf{s}_i) = \frac{1}{N_w}\sum_{t_w=1}^{N_w} R(\mathbf{s}_i,t_w)$ and $\bar{R}(\mathbf{s}_j) = \frac{1}{N_w}\sum_{t_w=1}^{N_w} R(\mathbf{s}_j,t_w)$

where N_T is the number of days in the dataset (here $N_T = 7305$), N_w is the number of wet days, and $\mathbb{1}[.]$ is the indicator function that takes the value 1 if the event in brackets is true and 0 otherwise.

Figure 4 compares the spatial structure of rainfall observations with those of the synthetic rainfall datasets simulated by the tested model on one hand, and of the BSNLG2022 benchmark model on the other hand. Results show that when focusing on the ability of the models to reproduce the spatial patterns of rainfall occurrence evaluated through JI-RO (Fig. 4a-b), both models perform equally well and almost perfectly simulate these patterns. Focusing next on the spatial patterns of rainfall intensity during rainy days evaluated through PCC-RI (Fig. 4c-d), one can notice that the BSNLG2022 benchmark model almost perfectly simulates them while our model presents imperfections, with some long range correlations of moderate intensity being underestimated. This is apparent for the South-East of the island in Fig. 4c-top row, and also detectable for the North-East of the island in Fig. 4c-top row and for the East of the island in Fig. 4c-bottom row. This imperfect modeling of long range correlations can be explained by the fact that the trans-Gaussian model relies on a covariance function that is strictly decreasing with separation distance (cf. Eq. 2), and therefore cannot plateau as would be necessary in some areas to properly match the observations. In contrast, the short to medium range correlations are well simulated by the trans-Gaussian model, including the patterns of non-stationarity and anisotropy. One should in particular notice the ability of the model to simulate sinuous contours of iso-correlation (i.e., not circular nor elliptic), which is made possible by the use of a non-stationary and anisotropic covariance function to model spatial dependencies within the latent field.

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

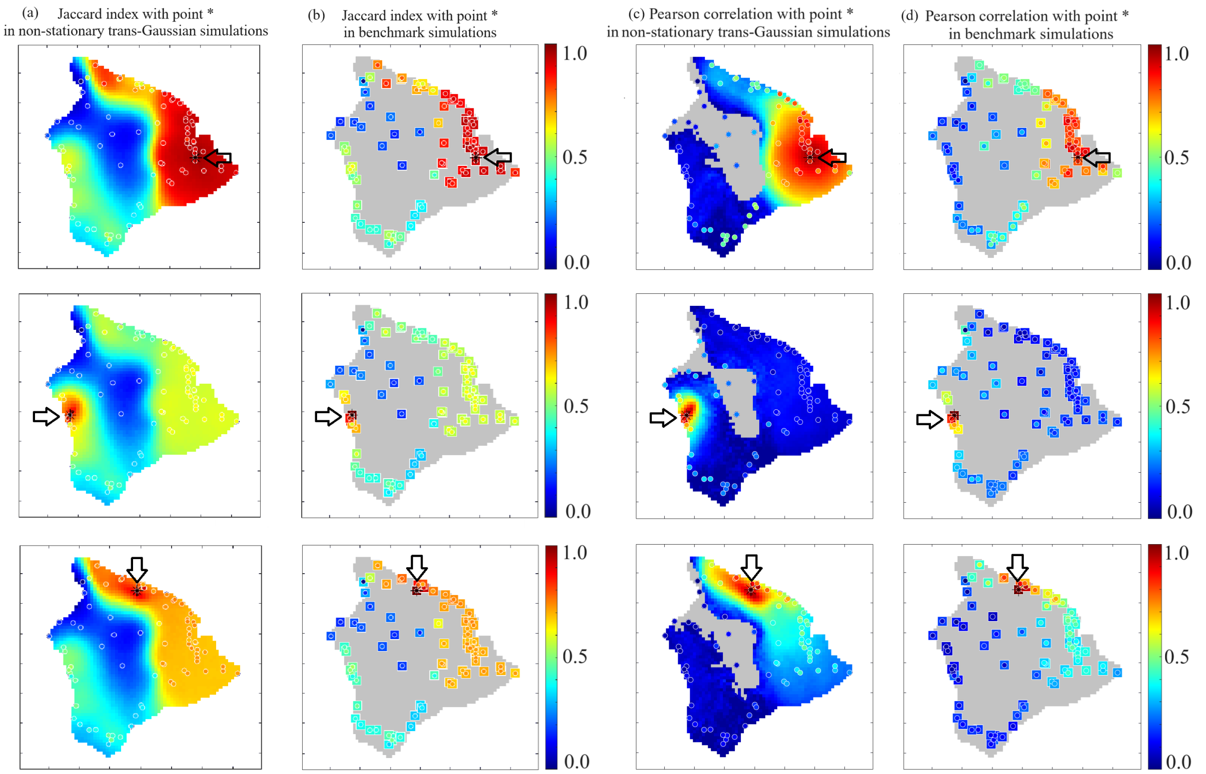

Figure 4. Spatial patterns of rainfall generated by the non-stationary trans-Gaussian model (a, c) and by the BSNLG2022 benchmark model (b, d). The two left columns (a, b) display the Jaccard Index of rainfall occurrence (JI-RO) with respect to three locations denoted by a star (*), and the two right columns display the Pearson Correlation Coefficient of rain intensity evaluated solely on wet days (PCC-RI) with respect to the same locations. In (c) the grey color denotes areas where less than 5% of the days are wet simultaneously with the target location, and where the estimation of PCC-RI is therefore deemed unrealiable. In all subplots the foreground circles denote observation data, the continuous background color (in a, c) denotes simulations from non-stationary trans-Gaussian model, and the background squares (in b, d) denote simulations from the BSNLG2022 benchmark model.

In summary, Figure 3 shows that the non-stationary trans-Gaussian model is able to simulate rainfall fields with point statistics reproducing almost perfectly the marginal distribution of the observations. In addition, Figure 4 shows that our model can simulate rainfall with realistic patterns of rainfall intensity, and is in particular able to reproduce the sinuous contours of iso-correlation emerging from the Hawai'i observation dataset. The trans-Gaussian model is nevertheless outperformed by the BSNLG2022 benchmark model for the simulation of long range and moderate intensity correlations of rainfall intensity, most probably because BSNLG2022 relies on empirical copulas instead of a parametric covariance function to model the spatial dependencies. In return, the use of a parametric covariance function is the keystone for the design of a stochastic rainfall generator able to simulate continuous rainfall fields, and therefore to go beyond multi-site simulation, which is a requirement to generate rainfall maps from sparse rain gauge observations as detailed hereafter.

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

3.4 Mapping daily rainfall over complex topography

Building on the ability of our model to simulate realistic rainfall fields between observation locations, we next assess its ability to predict rainfall at ungauged locations, which corresponds to rainfall mapping. This is evaluated through a cross-validation procedure, in which we iteratively select a target rain gauge, remove it from the training dataset (as well as the data from all gauges belonging to a 5km-radius circular area centered on this target gauge in order to mitigate the effect of near co-located gauges), re-estimate model parameters for the new training dataset, predict rainfall at the target location by conditional simulation (cf. Sect. 2.5), and finally compare the predicted and observed rainfall values. The quality of the rainfall prediction is evaluated with respect to the ability of the model to estimate rainfall occurrence as well as rainfall intensity during wet days. The prediction of rainfall occurrence is evaluated at each gauge location through the following metrics: the mean bias of the predicted rainfall occurrence (MB-RO), the mean absolute error of rainfall occurrence (MAE-RO), and the Brier score of rainfall occurrence (BS-RO) that measures the accuracy of the probabilistic prediction (Brier, 1950):

$$\mathrm{MB-RO}(\mathbf{s}) = \frac{1}{N_T}\sum_{t=1}^{N_T}\left(\mathbb{1}\left[\hat{R}(\mathbf{s},t)>0\right] - \mathbb{1}\left[R_{obs}(\mathbf{s},t)>0\right]\right) \tag{9}$$

$$\mathrm{MAE-RO}(\mathbf{s}) = \frac{1}{N_T}\sum_{t=1}^{N_T}\left|\mathbb{1}\left[\hat{R}(\mathbf{s},t)>0\right] - \mathbb{1}\left[R_{obs}(\mathbf{s},t)>0\right]\right| \tag{10}$$

$$\mathrm{BS-RO}(\mathbf{s}) = \frac{1}{N_T}\sum_{t=1}^{N_T}\left(\frac{1}{N_{sim}}\left(\sum_{sim=1}^{N_{sim}}\mathbb{1}\left[R_{sim}(\mathbf{s},t)>0\right]\right) - \mathbb{1}\left[R_{obs}(\mathbf{s},t)>0\right]\right)^2 \tag{11}$$

where $\hat{R}(\mathbf{s},t)$ is the prediction of rainfall at location $\mathbf{s}$ and day t (here the median of 100 conditional simulations is selected as rainfall prediction for our model), and N_{sim} is the number of conditional simulations (here $N_{sim} = 100$). To complement the evaluation of rainfall occurrence prediction, we assess the prediction of rainfall intensity conditionally to the occurrence of rainfall (in the following, the occurrence of rainfall is always determined from observations). The prediction of rainfall intensity during wet days is evaluated through the following metrics, evaluated solely on wet days: the mean bias of the predicted intensity (MB-RI), the mean absolute error of rainfall intensity (MAE-RI), and the mean continuous ranked probability score of rainfall intensity (MCRPS-RI, (Gneiting and Raftery, 2007)) that measures the accuracy of the probabilistic prediction and corresponds to the integral of the Brier score associated with all possible rainfall intensity thresholds greater than zero:

$$\mathrm{MB-RI}(\mathbf{s}) = \frac{1}{N_w}\sum_{t_w=1}^{N_w}\left(\hat{R}(\mathbf{s},t_w) - R_{obs}(\mathbf{s},t_w)\right) \tag{12}$$

$$\mathrm{MAE-RI}(\mathbf{s}) = \frac{1}{N_w}\sum_{t_w=1}^{N_w}\left|\hat{R}(\mathbf{s},t_w) - R_{obs}(\mathbf{s},t_w)\right| \tag{13}$$

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

$$\mathrm{MCRPS-RI}(\mathbf{s}) = \frac{1}{N_w}\sum_{t_w=1}^{N_w}\int_{r=0}^{+\infty}\left(\mathrm{F}_{R_{sim}(\mathbf{s},t_w)}(r) - \mathbb{1}[r \geq R_{obs}(\mathbf{s},t_w)]\right)^2 dr \tag{14}$$

where N_w is the number of wet days observed at the target location (i.e., days such that $R_{obs}(\mathbf{s},t_w) > 0$), and $\mathrm{F}_{R_{sim}(\mathbf{s},t_w)}$ is the cdf of the rainfall forecast derived from the 100 conditional simulations $R_{sim}(\mathbf{s},t_w)$.

The performance of our model is benchmarked against a Climatically Aided Interpolation (CAI) framework initially developed to map monthly rainfall in Hawai'i (Lucas et al., 2022), which has been adapted to the present case of daily rainfall mapping by processing the data from each rain type separately. In a nutshell, a Climatological Rainfall Pattern (CRP) is first derived by Ordinary Kriging (using a Matérn covariance function with Nugget) of the mean daily rain gauge observations for the rain type of interest. Afterwards, relative rainfall anomalies are computed for each day and rain gauge location: $RFA(\mathbf{s},t) = log\left(\frac{R_{obs}(\mathbf{s},t)}{\mathrm{CRP}(\mathbf{s},t)} + 1\right)$. These relative rainfall anomalies are subsequently interpolated at ungauged locations by Ordinary Kriging (using a Matérn covariance function without Nugget), and rainfall fields are finally retrieved by back-transformation of the anomalies and addition of the CRP.

Figure 5.a-d displays the results of the cross-validation focusing on the spatial prediction of rainfall occurrence. It shows that our model performs very well except for a few stations located in sparsely gauged areas and at the transition between a domain where rainfall is very frequent (North-East coast and mountain slopes) and another domain with very rare precipitation (summits in the center of the island and North-West coast). In that context, the withdrawal of the target rain gauge from the training and conditioning datasets leads to a lack of information about the exact location of the gradient of rainfall occurrence, and in the absence of such information the model generates a smooth transition which misrepresents the actual gradient. This artifact is absent from more densely gauged areas (e.g., the West and North-East coasts), and one can therefore infer that the density of the rain gauge network is crucial to get accurate rain occurrence maps in areas with strong rainfall gradients. When comparing the above results with the CAI interpolation, one can notice that our model outperforms the benchmark both in terms of bias (Fig. 5.b) and magnitude of the prediction errors (Fig. 5.c). This can be explained by the tendency of CAI interpolation to over-predict very light rain in place of dry conditions because it does not explicitly account for the zero-inflated nature of daily rainfall data. In contrast, the explicit modeling of this feature in the trans-Gaussian framework allows our model to map rainfall occurrence more effectively.

Figure 5.e-h investigates the interpolation of rainfall intensity conditionally to the occurrence of rainfall. One can notice that in contrast with rainfall occurrence, rainfall intensity varies substantially between nearby locations, which is particularly visible in the densely gauged area in the East of the island (Fig 5.e). This feature is not perfectly captured by our model nor by the CAI interpolation, and both models smooth-out the local variability of rainfall intensity. This is because all data from the gauges located in a 5km radius around the target gauge are discarded from the cross-validation, which prevents the models from learning the local behavior of rainfall and leads to biases of opposite signs for neighbouring gauges (Fig 5.f). The second artefact in rainfall intensity mapping is the imperfect depiction of rainfall intensity gradients in sparsely gauged areas, which is caused by the same reasons as discussed for rainfall occurrence. When comparing the skills of the two rainfall mapping approaches, one can notice that the benchmark CAI interpolation marginally outperforms the non-stationary trans-Gaussian

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

model when averaging the MB-RI and MAE-RI metrics over all gauges, but that the patterns and overall magnitude of the interpolation errors are very similar for both models.

In summary, figure 5 shows that the non-stationary trans-Gaussian model outperforms CAI interpolation for the spatial prediction of rainfall occurrence, and that both models perform similarly for the prediction of rainfall intensity during wet days. However, a significant advantage of the non-stationary trans-Gaussian model is its ability to provide probabilistic estimates of rainfall by means of an ensemble of conditional simulations, which CAI cannot do. The better scores obtained for the probabilistic forecasts (BS-RO in Fig. 5.d, and MCRPS-RI in Fig. 5.h) than for their deterministic counterparts (MAE-RO in Fig. 5.c, and MAE-RI in Fig. 5.g) show the additional information brought by the ensemble approach.

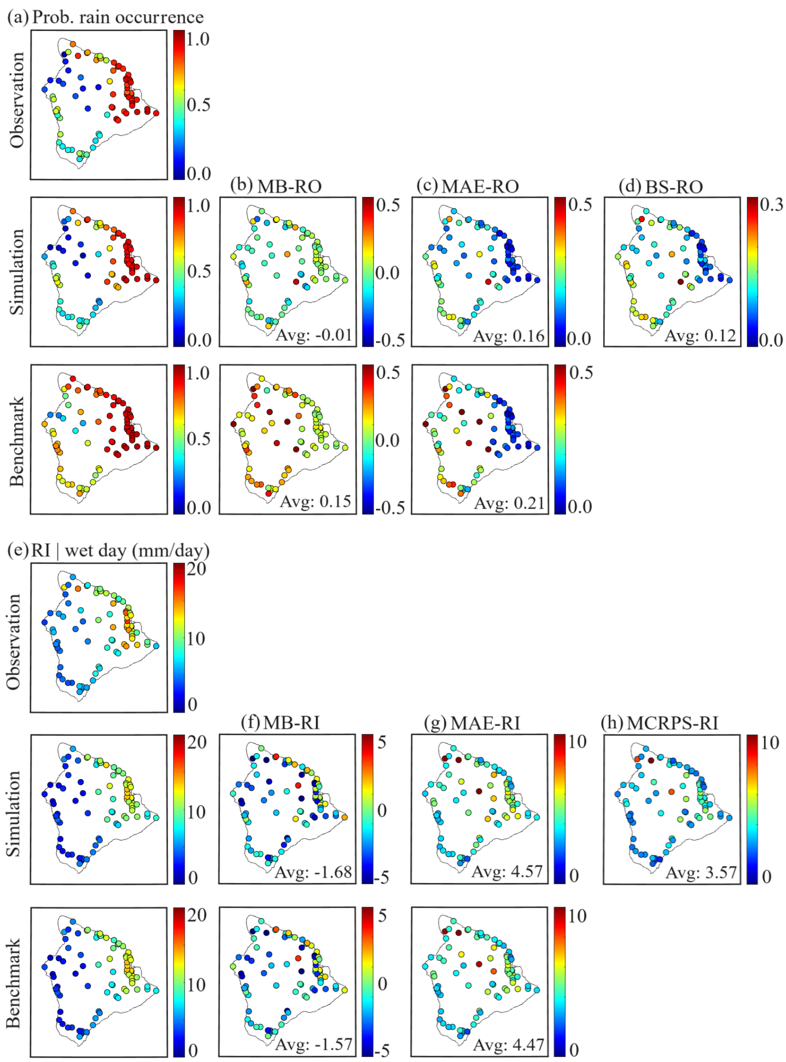

Figure 5. Cross-validation assessment of rainfall mapping performance. (a-d) Evaluation of rain occurrence prediction. (e-h) Evaluation of the prediction of daily rainfall intensity conditionally to the occurrence of rainfall.

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

3.5 Conditioning daily rainfall maps to monthly totals

Section 3.4 showed that, like every interpolation method, the proposed trans-Gaussian approach has decreasing skills in rainfall mapping when the density of the rain gauge network decreases. This is often unavoidable in mountainous regions because the rough topography makes the setup and maintenance of rain gauge networks challenging. To improve the representation of rainfall gradients in poorly gauged areas, we therefore propose conditioning daily rainfall maps not only to daily rain gauge observations as demonstrated in Section 3.4, but also to monthly totals derived from monthly rainfall maps that incorporate more information about rainfall patterns than is provided by the daily rain gauge network. Conditioning daily maps to monthly totals is obtained by adding a Metropolis within Gibbs step in the processing framework (see Sect. 2.6), and here we apply this step at 200 virtual stations spread throughout the Island of Hawai'i (white crosses in Fig. 6-bottom left) to compensate for the large gaps existing in the rain gauge network.

The impact of conditioning daily rainfall maps to monthly totals is illustrated for the months of January, April, July and October 2018, which have been selected to cover a wide range of weather conditions and sample all rain types observed in Hawai'i. Figure 6 compares the monthly rainfall accumulation derived from the reference HCDP monthly rainfall map (Fig. 6-top row) with the monthly accumulation obtained by summing the 30 or 31 daily rainfall maps derived from daily rain gauge observations only (Fig. 6-middle row), and the monthly accumulation obtained by summing the daily rainfall maps conditioned to daily rain gauge observations as well as monthly totals at 200 virtual stations (Fig. 6-bottom row). The daily rainfall fields associated to Fig. 6-middle row and Fig. 6-bottom row are displayed in Supplementary Material.

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

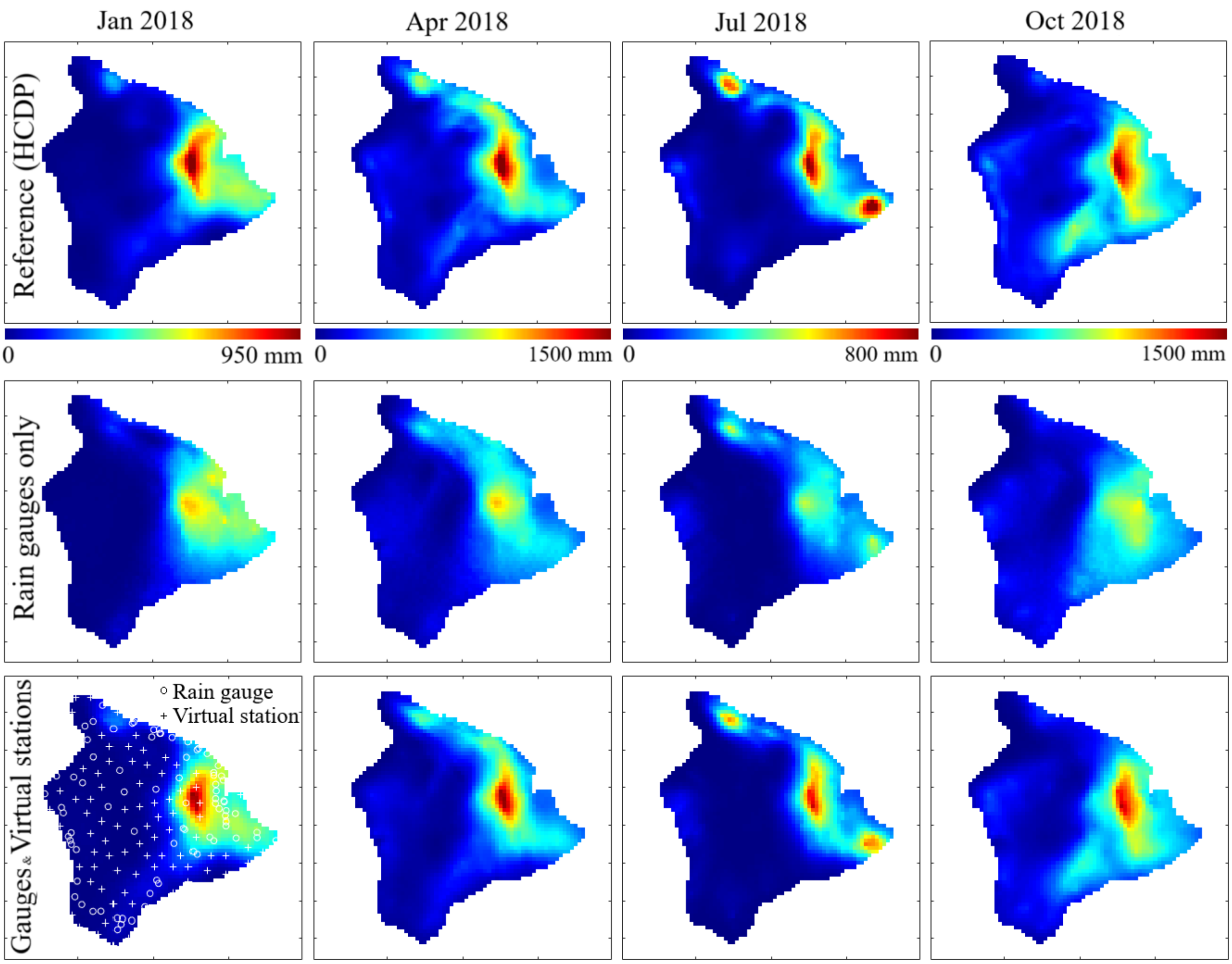

Figure 6. Maps of monthly rainfall for the months of January, April, July and October 2018 derived from the HCDP monthly historical rainfall dataset (top row), the sum of 30 or 31 daily rainfall maps obtained by the stochastic interpolation of rain gauge observations (middle row), and the sum of daily rainfall maps obtained by the stochastic interpolation of rain gauge observations constrained to honor monthly totals at 200 virtual stations (bottom row). In the bottom plot of January 2018 the dots denote the locations of the 79 rain gauges and the crosses denote the locations of the 200 virtual stations.

Results in Figure 6 show that conditioning daily rainfall maps to monthly totals enables the emergence of a monthly spatial maximum at an ungauged location that is located on the East side of Mauna Loa for the four months of interest. In addition, the spatial gradients of rainfall occurring in the North and South of this wet area are sharper when the daily maps are conditioned to monthly totals, leading to more realistic rainfall patterns in the South-East and East feet of Mauna Loa. At higher altitudes, the wet-dry gradient occurring at the East of Mauna Kea and Mauna Loa summits is more sinuous when adding the monthly conditioning, which again improves the realism of the monthly rainfall patterns. It is worth mentioning that the improved

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

gradients and patterns of monthly rainfall obtained by conditioning daily maps to monthly totals propagate to the daily rainfall maps displayed in Supplementary Material, and that this improvement does not introduce any spurious smoothing nor a drizzle effect (i.e., the under-estimation of dry areas coupled with an over-estimation of very low rainfall intensities) in the daily rainfall maps. Overall, conditioning daily rainfall maps to monthly totals where the density of the rain gauge network is low allows for the generation of daily rainfall maps with improved spatial patterns, in particular where orographic effects are the strongest.

4 Conclusions

In this study we designed a fully non-stationary and trans-Gaussian model that is able to accommodate the steep spatial gradients of daily rainfall observed in mountainous regions. The use of a non-stationary transform function enables the simulation of realistic patterns of daily rainfall occurrence and intensity. In addition, the use of an anisotropic and non-stationary latent field paves the way for the simulation of rainfall fields with realistic spatial dependencies. When applied to data from a network of 79 rain gauges on the Island of Hawai‘i, the proposed model is able to infer the spatial distribution of orographic rainfall and simulate gridded rainfall products that skillfully reproduce observed rainfall statistics. Finally, daily rainfall maps are improved in poorly gauged areas by conditioning rainfall interpolation not only to daily rain gauge observations, but also to monthly totals through a Metropolis within Gibbs approach.

When designing the model, we chose to follow a data-driven approach using only observations to inform the non-stationarity of rainfall. Hence, the marginal distribution of rainfall is first estimated at each rain gauge location separately and subsequently interpolated at ungauged locations. Similarly, the covariance function of the latent field is first assumed to be stationary within relatively small climate zones during the inference step, and a non-stationary covariance function is subsequently obtained at the scale of the entire target area by the convolution of these locally stationary covariance kernels.

The choice of a data-driven non-stationary model has the advantage of removing the need for covariates such as elevation or wind to inform the patterns of rainfall non-stationarity. This is of particular interest over complex topography and for high-resolution rainfall modeling because in this context the link between rainfall and topo-climatic covariates tends to be weak and complex (Giambelluca et al., 2013; Benoit et al., 2024) and therefore challenging to account for in statistical rainfall models. However, data-driven non-stationary models require a lot of observations for calibration, which restricts the use of our model to densely gauged areas and calls for an increased density of rain gauge networks to improve model performance. Such high-density rain gauge network is currently being implemented in the state of Hawaii by the extension of the HCDP weather station network through the construction of Hawai‘i Mesonet (Longman et al., 2024).

Over the past few decades, automatic rain gauges and weather stations have started to record rainfall data at sub-daily time steps and large observation datasets at a 10-min (and sometimes higher) temporal resolution have become common. Generating high-temporal-resolution rainfall maps at the scale of an island or a mountain range using such rain gauge observations could be addressed using trans-Gaussian stochastic rainfall models. The main difficulty we envision in this endeavor is the modeling of the temporal dependencies that emerge in sub-daily resolution rainfall fields in addition to the spatial dependencies addressed

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

in the present study that focused on daily resolution. Accounting for both spatial and temporal dependencies in the presence of orographic effects will require the development of non-stationary space-time covariance functions to model the latent field and robust estimation methods to infer model parameters from rain gauge observations.

In the end, rainfall maps derived from rain gauge observations are expected to complement radar rainfall estimates in mountainous regions where the complex topography challenges this technology with beam blocking and ground clutter, and to substitute the lack of radar products in non-equipped areas. Increasing the accuracy and resolution of rainfall maps is expected to improve hydrological modeling and flood forecasting in small catchments that are widespread in mountainous regions where orographic effects prevail.

Code and data availability. The source code used for this study is freely available on the following repository:
https://github.com/LionelBenoit/StochasticRainfallModel_Orography
The daily and monthly resolution rainfall datasets used to illustrate the study are publicly available on the Hawai‘i Climate Data Portal: https://www.hawaii.edu/climate-data-portal/

Appendix A: Gibbs sampler adapted to the simulation of censored rainfall pseudo-observations

Inputs:

- A vector $\mathbf{R}$ of daily rainfall observations sorted so that its first N_d elements correspond to a dry observation (i.e., $\mathbf{R}(i) = 0, \forall i = 1,...,N_d$), and the other elements correspond to a non-zero observation (i.e., $\mathbf{R}(j) > 0, \forall j = N_d+1,...,N_s$).
- The transform function $\Psi_\mathbf{s}$ at each observation location.
- The covariance function C_Y between all pairs of observation locations.
- The number of iterations of the Gibbs sampler N_{iter}.

Instructions:

1. **For** $d = 1$ **to** N_d

   Initialize $\mathbf{Y}(d) = a_0(\mathbf{s}_d) - 0.1$

   end

2. **For** $w = N_d + 1$ **to** N_s

   Compute $\mathbf{Y}(w) = \psi_{\mathbf{s_w}}^{-1}(\mathbf{R}(w))$

   end

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

3. Pre-compute the covariance matrix $\mathbf{\Sigma}$ between all observation locations using the covariance function C_Y

4. **For** $iter = 1$ **to** N_{iter}

    For $d = 1$ **to** N_d

    (a) Draw $\alpha \sim \mathcal{N}(\mathbf{\Sigma}^t_{-d,d}\mathbf{\Sigma}^{-1}_{-d,-d}\mathbf{Y}_{-d}, 1 - \mathbf{\Sigma}^t_{-d,d}\mathbf{\Sigma}^{-1}_{-d,-d}\mathbf{\Sigma}_{-d,d})$ where $\mathbf{\Sigma}_{-d,-d}$ is the matrix $\mathbf{\Sigma}$ deprived of its d^{th} row and column, $\mathbf{Y}_{-d}$ is the vector $\mathbf{Y}$ deprived of its d^{th} element and $\mathbf{\Sigma}_{-d,d}$ is the d^{th} column of the matrix $\mathbf{\Sigma}$ deprived of its d^{th} element

    (b) **While** $\alpha > a_0(\mathbf{s}_d)$

        Go to 4.a

        end

    (c) Set $\mathbf{Y}(d) = \alpha$

    end

   end

Output: The vector of pseudo-observations $\mathbf{Y}$

Appendix B: Metropolis within Gibbs sampler to simulate an ensemble of meta-Gaussian random fields with prescribed sum

Inputs:

– A set of N_f mutually independent random fields Y_f characterized by their covariance: $Y_f \sim \mathcal{N}(0, C_{Y_f})$. A transform function Ψ_f is associated to each latent field Y_f to link it with the actual rainfall.

– A vector $\mathbf{R}$ of daily rainfall observations (length N_s).

– A vector $\mathbf{R}_{sum}$ of monthly rainfall totals (length N_v) embodying the monthly rainfall data at the virtual stations $\mathbf{u}_j$, associated with an uncertainty vector $\mathbf{U}_{sum}$ (specified as one standard deviation).

– The number of iterations N_{iter} of the Gibbs sampler.

Instructions:

for $f = 1$ **to** N_f

1. Create a vector $\mathbf{Y}_{f,obs}$ (length N_s) so that: $\forall i = 1, ..., N_s : \mathbf{Y}_{f,obs}(i) = {\Psi_{f,\mathbf{s}_i}}^{-1}(R(\mathbf{s}_i))$. If $R(\mathbf{s}_i) = 0$ the corresponding latent value is undefined and has to be simulated using Algorithm 1.

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

2. Create a vector $\mathbf{Y}_{f,sim}$ (length N_v) of latent values at the virtual station locations $\mathbf{u}_j$ using the unconditional simulation method described in section 2.4 combined with conditioning Kriging described in section 2.5.
3. Set $\mathbf{Y}_f = [\mathbf{Y}_{f,sim}, \mathbf{Y}_{f,obs}]$ and compute its covariance matrix $\mathbf{\Sigma}_f$.

end

for $iter = 1$ **to** N_{iter}

for $f = 1$ **to** N_f

for $k = 1$ **to** N_v

1. [*Gibbs sampling step*] Draw $\alpha \sim \mathcal{N}({\mathbf{\Sigma}_f}_{-k,k}^T {\mathbf{\Sigma}_f}_{-k,-k}^{-1} \mathbf{Y}_{f_{-k}}, 1 - {\mathbf{\Sigma}_f}_{-k,k}^T {\mathbf{\Sigma}_f}_{-k,-k}^{-1} \mathbf{\Sigma}_{f_{-k,k}})$ where $\mathbf{\Sigma}_{f_{-k,k}}$, $\mathbf{\Sigma}_{f_{-k,-k}}$ and $\mathbf{Y}_{f_{-k}}$ are defined in a similar way as in Appendix 1.
2. Set $\mathbf{Z}_{k,0}$ = $\mathbf{Z}_{k,1}$ = $[\mathbf{Y}_1(k), ..., \mathbf{Y}_{N_f}(k)]$ and assign $\mathbf{Z}_{k,1}(f) = \alpha$
3. Compute the likelihoods:
$l_0 = \frac{1}{\sqrt{2\pi}\mathbf{U}_{sum}(k)}.\exp\left(-\frac{1}{2.\mathbf{U}_{sum}(k)^2}\left(\left(\sum_{f=1}^{N_f}\Psi_f(\mathbf{Z}_{k,0}(f))\right)-\mathbf{R}_{sum}(k)\right)^2\right)$
$l_1 = \frac{1}{\sqrt{2\pi}\mathbf{U}_{sum}(k)}.\exp\left(-\frac{1}{2.\mathbf{U}_{sum}(k)^2}\left(\left(\sum_{f=1}^{N_f}\Psi_f(\mathbf{Z}_{k,1}(f))\right)-\mathbf{R}_{sum}(k)\right)^2\right)$
4. Draw $\beta \sim U_{[0,1]}$
5. [*Metropolis acceptance rule*] **if** $\beta > 1 - \frac{l1}{l0}$

Set $\mathbf{Y}_f(k) = \alpha$

end

end

end

Set $\mathbf{Y}_{f,iter}$=$\mathbf{Y}_f$

end

Output: A set of N_{iter} random vectors of latent field pseudo-observations at the virtual station locations.

Author contributions. LB, MPL and TWG designed the study. LB and DA developed the stochastic rainfall model. LB implemented the model and performed the numerical experiments. MPL, KMK and TWG compiled the rainfall dataset. LB wrote the paper with contributions and editing from all co-authors.

Competing interests. The authors declare that they have no conflicts of interest.

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

Acknowledgements. LB and DA acknowledge the financial support of the Chair Geolearning, funded by Andra, BNP Paribas, CCR and the SCOR Foundation for Science. MPL, KMK and TWG acknowledge the financial support of the National Science Foundation ChangeHI EPSCoR Research Infrastructure Improvement Award #OIA-2149133. KMK and TWG acknowledge the financial support of the Cooperative Institute for Research to Operations in Hydrology (CIROH) NA22NWS4320003.

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

References

Abdulah, S., Ltaief, H., Sun, Y., Genton, M. G., and Keyes, D. E. (2019, December). Geostatistical modeling and prediction using mixed precision tile Cholesky factorization. In *2019 IEEE 26th international conference on high performance computing, data, and analytics (HiPC)* (pp. 152-162). IEEE.

Ailliot, P., Allard, D., Monbet, V., and Naveau, P. (2015). Stochastic weather generators: an overview of weather type models. *Journal de la societe francaise de statistique*, 156(1), 101-113.

Allard, D. and Bourotte, M. (2015). Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process, *Stochastic Environmental Research and Risk Assessment* 29: 453-462.

Babak, O., and Deutsch, C. V. (2009). Statistical approach to inverse distance interpolation. *Stochastic Environmental Research and Risk Assessment*, 23, 543-553.

Bárdossy, A., and Pegram, G. G. (2016). Space-time conditional disaggregation of precipitation at high resolution via simulation, *Water Resources Research*, 52(2), 920-937.

Bennett, B., Thyer, M., Leonard, M., Lambert, M., and Bates, B. (2018). A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model. *Journal of Hydrology*, 556, 1123-1138.

Benoit, L., and Mariethoz, G. (2017). Generating synthetic rainfall with geostatistical simulations. *Wiley Interdisciplinary Reviews: Water*, 4(2), e1199.

Benoit, L., Allard, D. and Mariethoz, G. (2018). Stochastic Rainfall Modeling at Sub-kilometer Scale, *Water Resources Research*, 54, 4108-4130.

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

Benoit, L., Sichoix, L., Nugent, A. D., Lucas, M. P., and Giambelluca, T. W. (2022). Stochastic daily rainfall generation on tropical islands with complex topography, *Hydrology and Earth System Sciences*, 26(8): 2113-2129.

Benoit, L. and Sichoix, L. (2023). Sub-daily rainfall patterns in the mountainous regions of the Island of Tahiti: insights from a one-year rain gauge network expansion, *Journal of Hydrology: Regional Studies*, 50, 101559.

Benoit, L., Koch, E., Peleg, N., and Mariethoz, G. (2024). Precipitation-elevation relationship: non-linearity and space-time variability prevail in the Swiss Alps. *Journal of Hydrology X*, 25, 100186.

Berrocal, V. J., Raftery, A. E., and Gneiting, T. (2008). Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *The Annals of applied statistics*, 2(4), 1170-1193.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1), 1-3.

Chilès, J. P., and Delfiner, P. (2012). Geostatistics: modeling spatial uncertainty (Vol. 713). *John Wiley and Sons*.

Creutin, J. D., and Obled, C. (1982). Objective analyses and mapping techniques for rainfall fields: an objective comparison. *Water resources research*, 18(2), 413-431.

Creutin, J. D., Delrieu, G., and Lebel, T. (1988). Rain measurement by raingage-radar combination: a geostatistical approach. *Journal of Atmospheric and oceanic technology*, 5(1), 102-115.

Davis, M. W. (1987). Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Mathematical geology*, 19, 91-98.

Emery, X., and Arroyo, D. (2018). On a continuous spectral algorithm for simulating non-stationary Gaussian random fields, *Stochastic Environmental Research and Risk Assessment*, 32: 905-919.

Evin, G., Favre, A. C., and Hingray, B. (2018). Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrology and Earth System Sciences*, 22(1), 655-672.

Foehn, A., Hernández, J. G., Schaefli, B., and De Cesare, G. (2018). Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments. *Journal of Hydrology*, 563, 1092-1110.

Foresti, L., Sideris, I. V., Panziera, L., Nerini, D., and Germann, U. (2018). A 10-year radar-based analysis of orographic precipitation growth and decay patterns over the Swiss Alpine region. *Quarterly Journal of the Royal Meteorological Society*, 144(716), 2277-2301.

Fraley, C. and Raftery, A. E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation, *Journal of the American Statistical Association*, 97, 611–631.

Frazier, A. G., Giambelluca, T. W., Diaz, H. F., and Needham, H. L. (2016). Comparison of geostatistical approaches to spatially interpolate month-year rainfall for the Hawaiian Islands.

Fuglstad, G. A., Simpson, D., Lindgren, F., and Rue, H. (2015). Does non-stationary spatial data always require non-stationary random fields?. *Spatial Statistics*, 14, 505-531.

Giambelluca, T. W., Chen, Q., Frazier, A. G., Price, J. P., Chen, Y.-L., Chu, P.-S., Eischeid, J. K., and Delparte, D. M. (2013). Online Rainfall Atlas of Hawai‘i, *Bulletin of the American Meteorological Society*, 94, 313–316.

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996). Markov Chain Monte Carlo in practice. Dordrecht, the Netherlands: *Springer Science - Business Media*.

Gneiting, T., and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477), 359-378.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of hydrology*, 228(1-2), 113-129.

Guillot, G., and Lebel, T. (1999). Approximation of Sahelian rainfall fields with meta-Gaussian random functions; Part 2: parameter estimation and comparison to data. *Stochastic Environmental Research and Risk Assessment* 13: 113-130.

Houze Jr, R. A. (2012). Orographic effects on precipitating clouds. *Reviews of Geophysics*, 50(1).

Huang, Y. F., and Chen, Y. L. (2019). Numerical simulations of seasonal variations of rainfall over the Island of Hawaii. *Journal of Applied Meteorology and Climatology*, 58(6), 1219-1232.

Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassegues, P., Rudolf, B., et al. (2014). The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology*, 34(5), 1657–1675.

Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Societe Vaudoise des Sciences Naturelles*, 37, 241-272.

Kyriakidis, P. C., Kim, J., and Miller, N. L. (2001). Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics. *Journal of Applied Meteorology and Climatology*, 40(11), 1855-1877.

Lantuéjoul, C. (2002). Geostatistical Simulation: Models and Algorithms, Springer Verlag, Berlin Heidelberg, 254p.

Longman, R. J., Diaz, H. F., and Giambelluca, T. W. (2015). Sustained increases in lower-tropospheric subsidence over the central tropical North Pacific drive a decline in high-elevation rainfall in Hawaii. Journal of Climate, 28(22), 8743-8759.

https://doi.org/10.5194/egusphere-2025-2181
Preprint. Discussion started: 19 June 2025

Longman, R. J., Giambelluca, T. W., Nullet, M. A., Frazier, A. G., Kodama, K., Crausbay, S. D., Krushelnycky, P. D., Cordell, S., Clark, M. P., Newman A. J., and Arnold, J. R. (2018). Compilation of climate data from heterogeneous networks across the Hawaiian Islands,*Scientific data*, 5(1): 1-17.

Longman, R. J., Timm, O. E., Giambelluca, T. W., and Kaiser, L. (2021). A 20-year analysis of disturbance-driven rainfall on O ‘ahu, Hawai ‘i. *Monthly Weather Review*, 149(6), 1767-1783.

Longman, R. J., Lucas, M. P., Mclean, J., Cleveland, S. B., Kodama, K., Frazier, A. G., Kamelamela, K., Schriber, A., Dodge, M., Jacobs, G., and Giambelluca, T. W. (2024). The Hawai ‘i Climate Data Portal (HCDP). *Bulletin of the American Meteorological Society*, 105(7), E1074–E1083.

Lucas, M. P., Longman, R. J., Giambelluca, T. W., Frazier, A. G., Mclean, J., Cleveland, S. B., Huang, Y-F., and Lee, J. (2022). Optimizing automated kriging to improve spatial interpolation of monthly rainfall over complex terrain, *Journal of Hydrometeorology*, 23(4): 561-572.

Luo, X., Frazier, A. G., Diaz, H. F., Longman, R., and Giambelluca, T. W. (2024). Routine Climate Monitoring in the State of Hawai ‘i: Establishment of State Climate Divisions. *Bulletin of the American Meteorological Society*, 105(6), E1045–E1061.

Moraga, J. S., Peleg, N., Molnar, P., Fatichi, S., and Burlando, P. (2022). Uncertainty in high-resolution hydrological projections: Partitioning the influence of climate models and natural climate variability. *Hydrological Processes*, 36(10), e14695.

Niemi, T. J., Guillaume, J. H., Kokkonen, T., Hoang, T. M., and Seed, A. W. (2016). Role of spatial anisotropy in design storm generation: Experiment and interpretation. Water Resources Research, 52(1), 69-89.

Paciorek, C. J., and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions, *Environmetrics*, 17(5):483-506.

Panaretos, V. M., and Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annual review of statistics and its application*, 6(1), 405-431.

Papalexiou, S.M. and Serinaldi, F. (2020). Random fields simplified: Preserving marginal distributions, correlations, and intermittency, with applications from rainfall to humidity. *Water Resources Research*, 56(2), e2019WR026331.

Paschalis, A., Fatichi, S., Molnar, P., Rimkus, S., and Burlando, P. (2014). On the effects of small scale space–time variability of rainfall on basin flood response. *Journal of Hydrology*, 514, 313-327.

Porcu, E., Bevilacqua, M., Schaback, R., and Oates, C. J. (2024). The Matérn model: A journey through statistics, numerical analysis and machine learning. *Statistical Science*, 39(3), 469-492.

Roe, G. H. (2005). Orographic precipitation. *Annual Review of Earth and Planetary Sciences*, 33(1), 645-671.

Sampson, P. D., and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417), 108-119

Vaittinada Ayar, P., Blanchet, J., Paquet, E., and Penot, D. (2020). Space-time simulation of precipitation based on weather pattern sub-sampling and meta-Gaussian model. *Journal of Hydrology*, 581, 124451.

Wilks, D. S., and Wilby, R. L. (1999). The weather generation game: a review of stochastic weather models. *Progress in physical geography*, 23(3), 329-357.