Combining machine learning quantile regression and Gaussian random fields: a general framework for modeling and simulating space-time processes

Said Obakrim, Denis Allard, Lionel Benoit, Gregoire Mariethoz

Combining machine learning quantile regression and Gaussian random fields: a general framework for modeling and simulating space-time processes

Said Obakrim[1], Denis Allard[2], Lionel Benoit[3], and Grégoire Mariéthoz[4]

[1,4]Institute of Earth Surface Dynamics, University of Lausanne, 1015 Lausanne, Switzerland

[2,3]INRAE, BioSP, 84914 Avignon, France

January 5, 2026

Abstract

Modeling and simulating space-time random fields while accounting for possibly misaligned covariates is crucial for many applications requiring uncertainty quantification and risk assessment. To achieve this, we propose a flexible framework that couples machine-learning quantile regression with space-time Gaussian random fields. In this framework, the target variable is modeled as a combination of a latent Gaussian random field and transformed marginals obtained by Machine learning quantile regression conditionally on a set of covariates. We illustrate the approach on a synthetic experiment and on a case study that considers daily maximum temperature over north-east Switzerland conditional on seasonal cycle and large-scale geopotential height.

Keywords: K-nearest neighbors, quantile regression forest, quantile neural network, Gaussian random field, covariance function

1 Introduction

Modeling and simulating space-time processes conditional on covariates is important for representing natural processes in fields such as climate science (Terzago et al., 2018) or

hydrology (Reddy et al., 2023), where decisions depend on uncertainty quantification. In these contexts, modeling only the conditional expectation of the variable of interest is not sufficient (Kneib et al., 2023) and full predictive distributions are required to capture the uncertainty at each site and time. Constructing such distributions involves modeling not only the local marginals but also the dependence across space and time.

A substantial body of work has addressed the modeling of processes characterized by significant spatial or spatio–temporal dependences using geostatistical models built on Gaussian random fields (GRFs) (Seeger, 2004). Within this paradigm, dependence is encoded via covariance functions, ranging from separable to flexible nonseparable families; see, for example, the nonseparable constructions of Gneiting (2002) and Allard et al. (2022) as well as recent reviews (Chen et al., 2021; Porcu et al., 2021). These models offer parsimony and well understood tools for estimation (Sun et al., 2011) and simulation (Allard et al., 2025). However, when the target variables do not have Gaussian marginals, which is the case for many processes, specifications with Gaussian marginals may not capture the full conditional distributional structure.

In such situations, latent GRF approaches with transformation mappings (also known as trans-Gaussian models) extend Gaussian dependence to data with non-Gaussian marginals and have been used, for instance, in multivariate weather generation (Obakrim et al., 2025) and rainfall simulation (Benoit et al., 2021). Yet, the above latent constructions are not explicitly conditional on covariates, which can limit their ability to represent covariate-driven marginal distributions.

There is a growing interest in modeling the conditional distributions of a response conditionally on covariates. A common approach is to choose a parametric distribution for the response and let its parameters depend on covariates through regression, as done for example in Haruna et al. (2025). This approach can be effective, but it still relies on a chosen parametric family for the conditional distribution, which may limit flexibility in some applications. Quantile regression, introduced by Koenker and Bassett Jr (1978), offers a robust, distribution free alternative to mean regression by directly modeling conditional quantiles. In the machine learning era, quantile methods have evolved, enabling representation of flexible, nonlinear, and high-dimensional relationships (Patidar et al., 2023). Prominent approaches include neighborhood based empirical estimators (K-nearest neighbors, KNN; Bhattacharya and Gangopadhyay, 1990), tree-ensemble estimators (Quantile Regression Forests; Meinshausen and Ridgeway, 2006), and neural network based methods (Quantile Regression Neural Network, QRNN; Padilla et al., 2022; Cannon, 2011).

These approaches have been designed for univariate marginal problems but are not directly adapted to space–time problems.

This motivates hybrid approaches that decouple marginal behavior from dependence: retain a GRF to model the space–time correlation structure, while modeling site–wise conditional marginals with machine learning. Examples include the combination of INLA–SPDE with a latent Gaussian field (Lindgren and Rue, 2015; Lindgren et al., 2022) and Random Forest for the mean (Figueira et al., 2025), and regression residual kriging where Random Forest captures covariate effects and kriging accounts for spatial dependence (Guo et al., 2015). These hybrid methods typically model the effects of covariates through the conditional mean (with a Gaussian residual field for dependence), rather than learning the full conditional distribution of the response.

In this study, we present a general framework that couples machine learning quantile regression for the estimation of site–wise conditional marginals with a latent space–time GRF for the modeling of the dependence structure. The marginals can be modeled using a range of possible methods (here we choose KNN, QRF, and QRNN as a representative sample of methods), while the dependence is governed by a parametric space–time covariance. Our contributions are threefold: (i) develop a flexible, covariate-driven marginal modeling via different types of quantile learners; (ii) achieve interpretable and parsimonious control of the space–time dependence using covariance functions; and (iii) provide a fast algorithm for conditional simulation, enabled by recent advances in GRF simulation (Allard et al., 2025). The resulting framework enables full flexibility in modeling the marginal distribution, but the dependence structure remains that of a Gaussian copula. This constraints the range of dependence structures that can be captured by the model but has the advantage to keep inference and simulation tractable while still allowing rich, covariate-driven marginals. We illustrate the approach on a synthetic experiment and a climate case study.

The remainder of the paper is structured as follows. Section 2 presents the modeling framework, including the estimation procedure and simulation algorithm. In section 3, we present a synthetic experiment to assess the performance and flexibility of the framework and in section 4 we illustrate the approach through an application to maximum daily temperature simulation over northwest Switzerland, conditional on large-scale geopotential height and seasonal cycle. Finally, Section 5 concludes the study and outlines directions for future work.

2 Methodology

2.1 Notations

Let $\mathcal{S} \subset \mathbb{R}^d$ denote the spatial domain of interest and $\mathcal{T} \subset \mathbb{R}$ the time index set. Spatial locations are $\mathbf{s} \in \mathcal{S}$ and times are $t \in \mathcal{T}$. For two locations $\mathbf{s}_1, \mathbf{s}_2$ and times t_1, t_2, define the spatial lag $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$ and the temporal lag $u = t_1 - t_2$, with their respective Euclidean distances $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$ and $|u| = |t_1 - t_2|$.

Random quantities are written in uppercase (e.g., Y) and their observed values in lowercase (e.g., y). A space–time random field is denoted by $Y(\mathbf{s}, t)$, with observed values $y(\mathbf{s}_i, t_j)$. Covariates are regarded as deterministic and written in bold lowercase, $\mathbf{x}_\mathbf{s}(t) \in \mathbb{R}^p$. A subscript or superscript $\mathbf{s}$ on an object (e.g., $\mathbf{x}_\mathbf{s}(t)$ or $Q^\mathbf{s}_\tau$) indicates site-specific indexing, meaning that each site has its own vector or function but without implying that the object is itself a spatial random field.

For a given site $\mathbf{s}$ and covariate value $\mathbf{x}$, let $F^\mathbf{s}_{Y|\mathbf{x}}(\cdot \mid \mathbf{x})$ be the conditional cumulative distribution function (CDF) of $Y(\mathbf{s}, t)$ given $\mathbf{x}$. The (left-continuous) conditional τ-quantile at site $\mathbf{s}$ is

$$Q^\mathbf{s}_\tau(\mathbf{x}) = \inf\big\{y \in \mathbb{R} : \ F^\mathbf{s}_{Y|\mathbf{x}}(y \mid \mathbf{x}) \geq \tau\big\}, \qquad \tau \in (0, 1), \tag{1}$$

and

$$Q^\mathbf{s}_0(\mathbf{x}) = \inf\{y : \ F^\mathbf{s}_{Y|\mathbf{x}}(y \mid \mathbf{x}) > 0\}, \qquad Q^\mathbf{s}_1(\mathbf{x}) = \sup\{y : \ F^\mathbf{s}_{Y|\mathbf{x}}(y \mid \mathbf{x}) < 1\}.$$

We write Φ for the standard normal CDF, and use hats to denote estimates (e.g., $\hat{Q}$, $\hat{\theta}$).

2.2 General framework

Let $Y(\mathbf{s}, t) \in \mathbb{R}$ be the target space–time random field on $\mathcal{S} \times \mathcal{T}$. For each space-time coordinate $(\mathbf{s}, t)$, let $\mathbf{x}_\mathbf{s}(t) \in \mathbb{R}^p$ denote the (deterministic) covariate vector available at site $\mathbf{s}$ and time t.

Define a latent zero-mean and unit variance Gaussian random field $Z(\mathbf{s}, t)$ and the associated rank field

$$\tau(\mathbf{s}, t) \ = \ \Phi\big(Z(\mathbf{s}, t)\big) \in (0, 1),$$

where Φ is the standard normal CDF. We model the conditional distribution of $Y(\mathbf{s}, t)$ via site-specific conditional quantiles:

$$Y(\mathbf{s}, t) \ = \ Q^\mathbf{s}_{\tau(\mathbf{s},t)}\big(\mathbf{x}_\mathbf{s}(t)\big), \qquad (\mathbf{s}, t) \in \mathcal{S} \times \mathcal{T}. \tag{2}$$

As an illustration, in the special case where $\tau(\mathbf{s},t) \equiv 0.5$ for all $(\mathbf{s},t)$ (equivalently, $Z \equiv 0$), then equation (2) reduces to a median regression, $Y(\mathbf{s},t) = Q^{\mathbf{s}}_{0.5}\big(\mathbf{x}_{\mathbf{s}}(t)\big)$, and no space–time dependence is modeled.

For invertibility of the quantile mapping, the map $\tau \mapsto Q^{\mathbf{s}}_{\tau}(\mathbf{x})$ must be nondecreasing and left-continuous on $(0,1]$ for each fixed $(\mathbf{s},\mathbf{x})$, with limits $Q^{\mathbf{s}}_{0}(\mathbf{x}) = \inf\{y:\ F^{\mathbf{s}}_{Y|\mathbf{x}}(y \mid \mathbf{x}) > 0\}$ and $Q^{\mathbf{s}}_{1}(\mathbf{x}) = \sup\{y:\ F^{\mathbf{s}}_{Y|\mathbf{x}}(y \mid \mathbf{x}) < 1\}$. Under these conditions, the corresponding rank for an observation $y(\mathbf{s},t)$ is given by the generalized inverse

$$\tau(\mathbf{s},t) = \sup\big\{\tau \in [0,1]:\ Q^{\mathbf{s}}_{\tau}(\mathbf{x}_{\mathbf{s}}(t)) \le y(\mathbf{s},t)\big\}.$$

The model in Equation (2) involves two key components: the conditional distribution of $Y(\mathbf{s},t)$ given the covariates $\mathbf{x}_{\mathbf{s}}(t)$ and the latent space-time GRF $Z(\mathbf{s},t)$. The conditional distribution is described by the quantile regression function $Q^{\mathbf{s}}_{\tau(\mathbf{s},t)}(\cdot)$. Regarding the latent space-time GRF, given that $Z(\mathbf{s},t)$ has zero mean and unit variance, it is fully characterized by its second-order moment that can be modeled using a space-time covariance function (here in fact a correlation function since the variance is one). The next two sections detail the main components of the proposed modeling framework.

2.3 Learning marginal conditional distributions

To demonstrate the flexibility of the framework, we illustrate $Q^{\mathbf{s}}_{\tau}(\cdot)$ with three representative machine learning quantile estimators: K-nearest neighbors (KNN), Quantile Regression Forests (QRF), and a neural quantile model (QRNN). Note that our objective here is the methodology for coupling machine learning-based conditional quantiles with a latent GRF, and not to perform a comparative benchmark between quantile regression methods.

For a fixed spatial location $\mathbf{s}$, let $\{(\mathbf{x}_{\mathbf{s}}(t_j),\, y(\mathbf{s},t_j))\}_{j=1}^{n_t}$ be the training pairs used for training the machine learning quantile regression, with n_t the number of training time steps. A unifying view for KNN and QRF is the weighted empirical quantile defined for a new covariate vector $\mathbf{x}$ as

$$Q^{\mathbf{s}}_{\tau}(\mathbf{x}) = \inf\Big\{y \in \mathbb{R} : \sum_{j=1}^{n_t} w_{\mathbf{s},t_j}(\mathbf{x})\, \mathbf{1}\{y(\mathbf{s},t_j) \le y\} \ge \tau\Big\}, \qquad \tau \in (0,1), \tag{3}$$

where $w_{\mathbf{s},t_j}(\mathbf{x}) \ge 0$ and $\sum_{j=1}^{n_t} w_{\mathbf{s},t_j}(\mathbf{x}) = 1$. In KNN, if $V_{\mathbf{s}}(\mathbf{x};K)$ denotes the index set of the K nearest neighbors of $\mathbf{x}$ among $\{\mathbf{x}_{\mathbf{s}}(t_j)\}_{j=1}^{n_t}$ under a chosen distance over $\mathbb{R}^p$, and if $d_{\mathbf{s},t_j}(\mathbf{x}) \ge 0$ is a similarity (e.g., an inverse-distance or kernel weight) that is zero outside

$V_{\mathbf{s}}(\mathbf{x}; K)$, then

$$w_{\mathbf{s},t_j}^{\mathrm{KNN}}(\mathbf{x}) = \frac{d_{\mathbf{s},t_j}(\mathbf{x})\,\mathbf{1}\{\mathbf{x}_{\mathbf{s}}(t_j) \in V_{\mathbf{s}}(\mathbf{x}; K)\}}{\sum_{\ell=1}^{n_t} d_{\mathbf{s},t_\ell}(\mathbf{x})\,\mathbf{1}\{\mathbf{x}_{\mathbf{s}}(t_\ell) \in V_{\mathbf{s}}(\mathbf{x}; K)\}}. \tag{4}$$

In QRF, with trees $\{T_b\}_{b=1}^{B}$ and leaf $L_b(\mathbf{x})$ reached by $\mathbf{x}$ in tree b, the weights average leaf memberships (Meinshausen and Ridgeway, 2006):

$$w_{\mathbf{s},t_j}^{\mathrm{QRF}}(\mathbf{x}) = \frac{1}{B}\sum_{b=1}^{B} \frac{\mathbf{1}\{\mathbf{x}_{\mathbf{s}}(t_j) \in L_b(\mathbf{x})\}}{|L_b(\mathbf{x})|}. \tag{5}$$

where $|L_b(\mathbf{x})|$ is the number of samples that fall into the leaf $L_b(\mathbf{x})$. It is easy to check that for both systems, the sum of the weights is equal to 1.

Quantile regression neural network parametrizes the conditional quantile function with a neural network. Specifically, we write $Q_\tau^{\mathbf{s}}(\mathbf{x}; \boldsymbol{\vartheta}) := f_{\boldsymbol{\vartheta}}(\mathbf{x}, \tau)$, where $f_{\boldsymbol{\vartheta}}$ is a feed-forward (fully connected) neural network. The parameters $\boldsymbol{\vartheta}$ are learned by minimizing the pinball loss over a grid $\mathcal{Q} \subset (0, 1)$:

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} \sum_{j=1}^{n_t} \sum_{\tau \in \mathcal{Q}} \rho_\tau\Big(y(\mathbf{s}, t_j) - Q_\tau^{\mathbf{s}}\big(\mathbf{x}_{\mathbf{s}}(t_j); \boldsymbol{\vartheta}\big)\Big)\,, \tag{6}$$

with $\rho_\tau(u) = u\big(\tau - \mathbf{1}\{u < 0\}\big)$.

The hyperparameters for all marginal learners (KNN, QRF, QRNN) are selected by cross-validation using the pinball loss. For example, the hyperparameters of KNN are the number of neighbors and the bandwidth used for the kernel. Let $\mathcal{Q} \subset (0, 1)$ denote the grid of quantile levels and $\mathcal{F}$ be a set of time-blocked folds. For a candidate hyperparameter vector η, the cross-validation risk is

$$\mathcal{R}_{\mathrm{CV}}(\eta) = \frac{1}{|\mathcal{F}|}\sum_{f \in \mathcal{F}} \frac{1}{|\mathcal{Q}|}\sum_{\tau \in \mathcal{Q}} \frac{1}{|I_f|}\sum_{(\mathbf{s},t) \in I_f} \rho_\tau\Big(y(\mathbf{s}, t) - \hat{Q}_\tau^{\mathbf{s},\eta}\big(\mathbf{x}_{\mathbf{s}}(t)\big)\Big),$$

where I_f indexes validation times in fold f (aggregated over the sites used for cross-validation) and $\hat{Q}_\tau^{\mathbf{s},\eta}$ denotes the fitted quantile function under η. The selected hyperparameters satisfy

$$\hat{\eta} \in \arg\min_{\eta}\ \mathcal{R}_{\mathrm{CV}}(\eta).$$

To reduce computational burden, cross-validation can be performed on a random subset of locations, and the resulting $\hat{\eta}$ is then used for all sites.

2.4 Modeling and estimating the space-time Gaussian field

The latent Gaussian space-time random field $Z(\mathbf{s}, t)$ has zero mean and unit variance so it is fully characterized by its second-order moment i.e, the covariance function C given by

$$\mathrm{Cov}\big(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)\big) = C(\mathbf{s}_1, \mathbf{s}_2; t_1, t_2), \tag{7}$$

for all $(\mathbf{s}_1, t_1) \in \mathcal{S} \times \mathcal{T}$ and $(\mathbf{s}_2, t_2) \in \mathcal{S} \times \mathcal{T}$. The covariance function C is said to be stationary in space and time if and only if it depends only on the spatial lag $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$ and the temporal lag $u = t_1 - t_2$, i.e. with a slight abuse of notation,

$$C(\mathbf{s}_1, \mathbf{s}_2; t_1, t_2) = C(\mathbf{h}, u)$$

For simplicity we focus in the following on stationary and isotropic covariance functions; nonetheless, nonstationary constructions (e.g., spatially and/or temporally varying parameters, see for example Allard et al. (2025)) can be incorporated within the same framework. Numerous stationary space-time covariance functions have been proposed in the literature, encompassing both separable and non-separable models, see Porcu et al. (2021) and Chen et al. (2021) for recent reviews. In this work, we adopt the following Gneiting non-separable covariance function:

$$C_{\mathcal{G}}(\mathbf{h}, u) = \frac{1}{\big(a|u|^{2\alpha} + 1\big)^{\delta}} C_{\mathcal{M}}\left(\mathbf{h}; \frac{r}{\big(a|u|^{2\alpha} + 1\big)^{b/2}}, \nu\right), \tag{8}$$

where $a > 0$, $0 < \alpha \leq 1$, and $0 \leq b \leq 1$, with $\delta > bd/2$, are parameters controlling the temporal and spatial dependence. The function $C_{\mathcal{M}}(\cdot; r, \nu)$ denotes the Matérn covariance function with scale parameter $r > 0$ and smoothness parameter $\nu > 0$:

$$C_{\mathcal{M}}(\mathbf{h}; r, \nu) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(r\|\mathbf{h}\|\right)^{\nu} K_{\nu}\big(r\|\mathbf{h}\|\big), \tag{9}$$

where K_{ν} is the modified Bessel function of the second kind of order ν.

For clarity of exposition we suppose that observations are available at all spatio-temporal locations, but more general settings can be considered. Let $\boldsymbol{y} = \big[\,[y(\mathbf{s}_i, t_j)]_{i=1}^{n_s}\big]_{j=1}^{n_t}$ and $\boldsymbol{x} = \big[\,[\mathbf{x}_{\mathbf{s}_i}(t_j)]_{i=1}^{n_s}\big]_{j=1}^{n_t}$ denote the observed responses and covariates, respectively. Suppose we have site-wise estimates of the conditional quantile functions $\{\hat{Q}_{\tau}^{\mathbf{s}}(\cdot) : \tau \in (0,1)\}$ (e.g., obtained via KNN, QRF, or QRNN). For each $(\mathbf{s}_i, t_j)$, we obtain a pseudo–percentile by inverting the estimated conditional quantile function at the observed covariate:

$$\hat{\tau}(\mathbf{s}_i, t_j) \;=\; \inf\big\{\tau \in (0,1) :\; \hat{Q}_{\tau}^{\mathbf{s}_i}\big(\mathbf{x}_{\mathbf{s}_i}(t_j)\big) \;\geq\; y(\mathbf{s}_i, t_j)\big\}. \tag{10}$$

The pseudo-observations of the latent GRF are then obtained via

$$\hat{z}(\mathbf{s}_i, t_j) \;=\; \Phi^{-1}\big(\hat{\tau}(\mathbf{s}_i, t_j)\big), \qquad i = 1, \ldots, n_s,\; j = 1, \ldots, n_t. \tag{11}$$

The Gneiting model can then be fitted to $\hat{\boldsymbol{z}} = \big[[\hat{z}(\mathbf{s}_i, t_j)]_{i=1}^{n_s}\big]_{j=1}^{n_t}$ using maximum likelihood. If the number of the data points is high, which is the case in most space-time applications, composite likelihood (Lindsay, 1988) or Vecchia approximation (James and Guinness, 2024) can be used to limit the computational cost of likelihood computation.

2.5 Simulation algorithm

Given a set of parameters, the simulation is done in two steps. First, a space-time Gaussian random field is simulated according to the covariance model. Second, the latent field is mapped to the target variable using the learned site–wise conditional quantile functions, i.e.,

$$Y(\mathbf{s}, t) \;=\; Q^{\mathbf{s}}_{\Phi(Z(\mathbf{s},t))}\big(\mathbf{x}_{\mathbf{s}}(t)\big), \qquad (\mathbf{s}, t) \in \mathcal{S} \times \mathcal{T}.$$

For simulating the GRF, the classical approach is the Cholesky decomposition of the covariance matrix. However, this method is computationally expensive and unfeasible for large space-time domains. To reduce the computational cost of the Cholesky method, one can simulate the time sequentially using an autoregressive assumption, but this method is still costly. An appealing alternative is Gaussian mixtures simulation (Allard et al., 2025), which achieves the simulation of the whole GRF simultaneously with a computational cost scaling linearly with the size of the domain. The simulation algorithm for the Gneiting model used in this study is given in Allard et al. (2020) and in (Allard et al., 2025). The simulation algorithm adapted to the conditional model is presented in Algorithm 1.

3 Synthetic experiment

3.1 Setup

The following synthetic case study aims to illustrate our approach and to assess the estimation of the covariance parameters in a controlled setting. We consider the spatial domain $\mathcal{S} \subset [0, 1]^2$ on which the target $Y(\mathbf{s}, t)$ is defined. $\mathcal{S}$ is a rectangular lattice with $n_s = 15 \times 15 = 225$ locations. The temporal domain is $\mathcal{T} = \{1, \ldots, n_t\}$ with $n_t = 6000$ regular time steps, which we consider as daily time steps for illustration purpose.

Algorithm 1 Simulation algorithm

Require: Spatial grid $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_{n_s}\}$, time grid $\mathcal{T} = \{t_1, \ldots, t_{n_t}\}$

Require: Covariates $\mathbf{x}_{\mathbf{s}}(t) \in \mathbb{R}^p$ for $(\mathbf{s}, t) \in \mathcal{S} \times \mathcal{T}$

Require: Site-wise conditional quantiles $Q^{\mathbf{s}}_{\tau}(\cdot)$ for $\tau \in [0, 1]$

Require: Latent GRF covariance parameters θ_Z

Ensure: Simulated field $Y(\mathbf{s}, t)$ on $\mathcal{S} \times \mathcal{T}$

1: Sample a latent field $Z \sim \mathcal{GP}\big(0,\, C_{\mathcal{G}}^{\theta_Z}(\cdot)\big)$ on $\mathcal{S} \times \mathcal{T}$

2: **for all** $(\mathbf{s}, t) \in \mathcal{S} \times \mathcal{T}$ **do**

3: $\quad \tau(\mathbf{s}, t) = \Phi\big(Z(\mathbf{s}, t)\big)$

4: $\quad Y(\mathbf{s}, t) = Q^{\mathbf{s}}_{\tau(\mathbf{s},t)}\big(\mathbf{x}_{\mathbf{s}}(t)\big)$

5: **end for**

6: **return** $Y(\mathbf{s}, t)$

We generate the target as

$$Y(\mathbf{s}, t) \;=\; \exp\Big\{\mu(\mathbf{s}, t) \;+\; \sigma(\mathbf{s}, t)\, Z(\mathbf{s}, t)\Big\}, \qquad \mathbf{s} \in \mathcal{S},\; t \in \mathcal{T}, \tag{12}$$

with spatially varying mean and scale

$$\mu(\mathbf{s}, t) \;=\; \beta_0 \;+\; \tanh\big(\,\boldsymbol{\beta}(\mathbf{s})^{\top}\mathbf{x}_{\mathbf{s}}(t)\,\big), \qquad \sigma(\mathbf{s}, t) \;=\; \exp\Big\{\gamma_0 \;+\; \sin\big(\,\boldsymbol{\gamma}(\mathbf{s})^{\top}\mathbf{x}_{\mathbf{s}}(t)\,\big)\Big\}.$$

Here $Z(\mathbf{s}, t)$ is Gaussian random field, β_0 and γ_0 are intercepts and

$$\mathbf{x}_{\mathbf{s}}(t) = (x_{\mathbf{s},1}(t), \ldots, x_{\mathbf{s},p}(t))^{\top} \in \mathbb{R}^p$$

is the covariate values at time t. The coefficient vectors $\boldsymbol{\beta}(\mathbf{s}), \boldsymbol{\gamma}(\mathbf{s}) \in \mathbb{R}^p$ are time-invariant and defined on the response grid $\mathcal{S}$. This construction yields log-normal margins for $Y(\mathbf{s}, t)$ and a Gaussian space–time dependence inherited from Z.

For the synthetic setup, we simulate once the covariates $\{\mathbf{x}_{\mathbf{s}}(t)\}$ and the spatial coefficient fields $\boldsymbol{\beta}(\mathbf{s}), \boldsymbol{\gamma}(\mathbf{s})$ from Gaussian models to define $\mu(\mathbf{s}, t)$ and $\sigma(\mathbf{s}, t)$. Afterwards these covariates and coefficients are treated as fixed, and only the latent GRF $Z(\mathbf{s}, t)$ is considered as random (and thus resampled for each synthetic dataset).

Parameterization:

- **Covariates:** At each site, the $p = 10$ covariates are preliminarily simulated according to stationary Gaussian processes, and then they are considered as fixed, i.e., they

vary along time in a deterministic way. The p covariates are independent site-wise. The following time covariance function is used:

$$\mathrm{Cov}\big(x_{\mathbf{s},k}(t),\, x_{\mathbf{s},k}(t+u)\big) \;=\; \frac{1}{\big(1 + a_x\,|u|^{2\alpha_x}\big)^{\delta_x}}, \qquad k = 1, \ldots, p,$$

where $a_x = 0.5$, $\alpha_x = 0.5$, and $\delta = 0.2$. Here, $\mathrm{Cov}(x_{\mathbf{s},k}, x_{\mathbf{s},k'}) = 0$ for $k \neq k'$ and $\mathrm{Cov}(x_{\mathbf{s},k}, x_{\mathbf{s}',k}) = 0$ for $\mathbf{s} \neq \mathbf{s}'$.

- **Coefficients on $\mathcal{S}$:** Similarly, the coefficients $\boldsymbol{\beta}(\mathbf{s}) = \big(\beta_1(\mathbf{s}), \ldots, \beta_p(\mathbf{s})\big)^T$ and $\boldsymbol{\gamma}(\mathbf{s}) = \big(\gamma_1(\mathbf{s}), \ldots, \gamma_p(\mathbf{s})\big)^T$ are spatially varying coefficients. As done with the covariates, they are simulated using

$$\beta_j \sim \mathcal{GP}\big(0,\, C_{\mathcal{M}}^{\theta_\beta}(\cdot)\big),\; j = 1, \ldots, p$$

$$\gamma_j \sim \mathcal{GP}\big(0,\, C_{\mathcal{M}}^{\theta_\gamma}(\cdot)\big),\; j = 1, \ldots, p$$

with

$$\theta_\beta = (\sigma_\beta^2, \kappa_\beta, \nu_\beta) = (1 \times 10^{-1},\, 2.23,\, 1.6),$$

$$\theta_\gamma = (\sigma_\gamma^2, \kappa_\gamma, \nu_\gamma) = (1 \times 10^{-2},\, 2.88,\, 1.5).$$

Intercepts are fixed at $\beta_0 = 1.0,\ \gamma_0 = -1.0$.

- **Latent GRF.** $Z(\mathbf{s}, t) \sim \mathcal{GP}\big(0,\, C_{\mathcal{G}}^{\theta_Z}\big)$ on $\mathcal{S} \times \mathcal{T}$, where $C_{\mathcal{G}}^{\theta_Z}$ is the Gneiting covariance in (8) with

$$\theta_Z = (\kappa, \nu, a, \alpha, \delta, b) = (3.4,\ 0.7,\ 1.2,\ 0.3,\ 0.9,\ 0.63).$$

All GRFs are simulated by the Gaussian mixture method (Allard et al., 2025). We generate $n_{\text{sims}} = 20$ independent synthetic datasets. Note that the covariates and coefficients are considered fixed and hence simulated once for all 20 synthetic datasets. The last $n_{\text{test}} = 200$ days are held out as a test set for evaluation while the first $n_t - n_{\text{test}}$ epochs form the training period. The hyperparameters of each quantile regression method are selected by cross-validation on the training period using the methodology explained in Section 2. For example, for KNN we use an isotropic Gaussian kernel and choose the bandwidth by cross-validation.

To explore a range of realistic situations, we consider the three scenarios described in Table 1. In the *well-specified* scenario, the model inputs are only the covariates that generated the target, namely $\mathbf{x}(t)$. In the *noisy* scenario, the predictor set is blurred with noisy covariates $\mathbf{x}'_{\mathbf{s}}(t)$ that were not used to generate $Y(\cdot)$ and are simulated using

Scenario	Predictors given to the learner	Variable selection
Well-specified	$\mathbf{x}(t) \in \mathbb{R}^{10}$ (true covariates)	No
Noisy	$\big(\mathbf{x}(t),\, \mathbf{x}'(t)\big) \in \mathbb{R}^{20}$	No
Noisy+VS	$\big(\mathbf{x}(t),\, \mathbf{x}'(t)\big) \in \mathbb{R}^{20}$	RF-based VS

Table 1: Summary of the three experimental scenarios.

the same parametrization as $\mathbf{x}_{\mathbf{s}}(t)$. We add $p' = 10$ such noisy predictors, so the noisy scenario has $p + p' = 20$ covariates. Finally, in the *noisy+VS* scenario we use the same augmented predictors $(\mathbf{x}(t), \mathbf{x}'(t))$ but precede the quantile regression with a variable-selection (VS) step based on Random Forest importance (Genuer et al., 2010). Specifically, we randomly select $n_{\text{sites}}^{\text{VS}} = 50$ response grid locations, fit RF regressions between $Y(\mathbf{s}, t)$ and covariates $(\mathbf{x}(t), \mathbf{x}^{\text{N}}(t))$ on the training period at each selected location, aggregate variable importances across locations (here the median is used for aggregation), rank predictors, and retain the smallest set whose cumulative importance reaches the threshold 0.8. The selected covariates are then used for quantile-regression models on all locations.

In this synthetic experiment we keep the Matérn smoothness parameter ν fixed at its true value. This choice is motivated by identifiability: different pairs (κ, ν) can induce very similar spatial correlation structures, so the likelihood surface is often flat in that direction and many combinations of (κ, ν) are plausible. Fixing ν allows us to focus on the recovery of the remaining covariance parameters under controlled conditions. In contrast, ν will be estimated from the data when analyzing the climate application in Section 4.

3.2 Results

We first generate a single synthetic realization of the target field over the full space-time domain from the data-generating model; this *reference dataset* is treated as the observed data and is used to fit the models and to define the training and test periods. In addition, for validation purposes, we simulate 50 further realizations from the same data-generating process only over the test period; we refer to these as the *oracle ensemble*. We then fit the model using KNN, QRF, and QRNN on the training part of the reference dataset.

We first examine the quantile regression performance. Figure 1 displays the true q-01, median (i.e. q-50), and q-99 at 18–12–2015 (a date in the test period), together with the corresponding estimates from KNN, QRF, and QRNN. All three methods recover the true

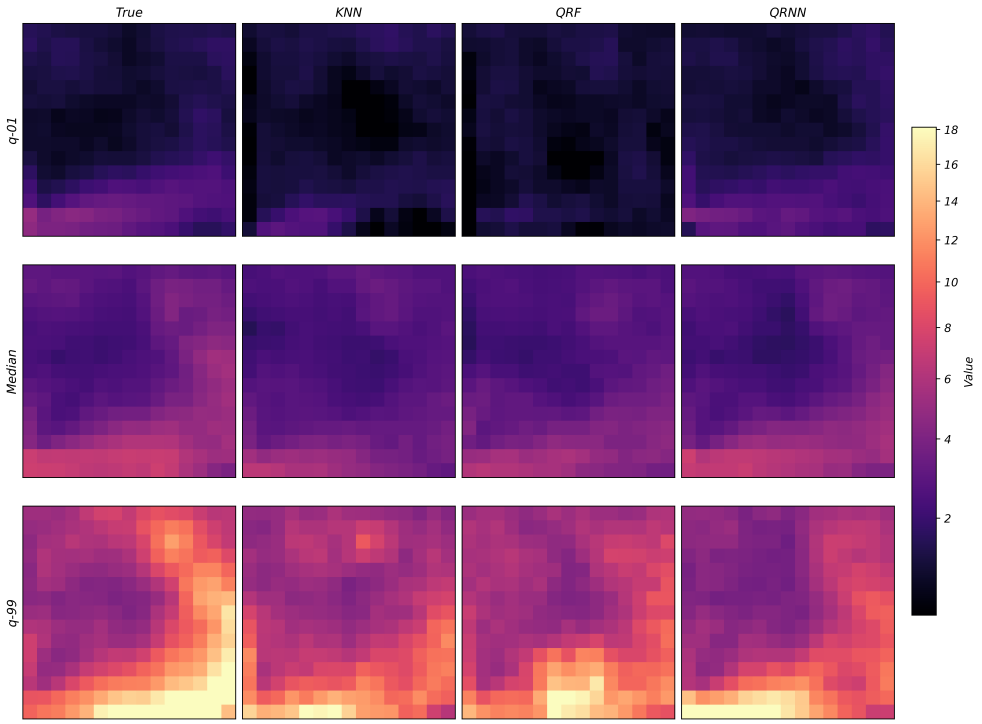

Figure 1: Quantiles 1% (top row), 50% (middle row) and 99% (bottom row) of the target process Y on 18-12-2015. From left to right: reference and estimations with KNN, QRF and QRNN.

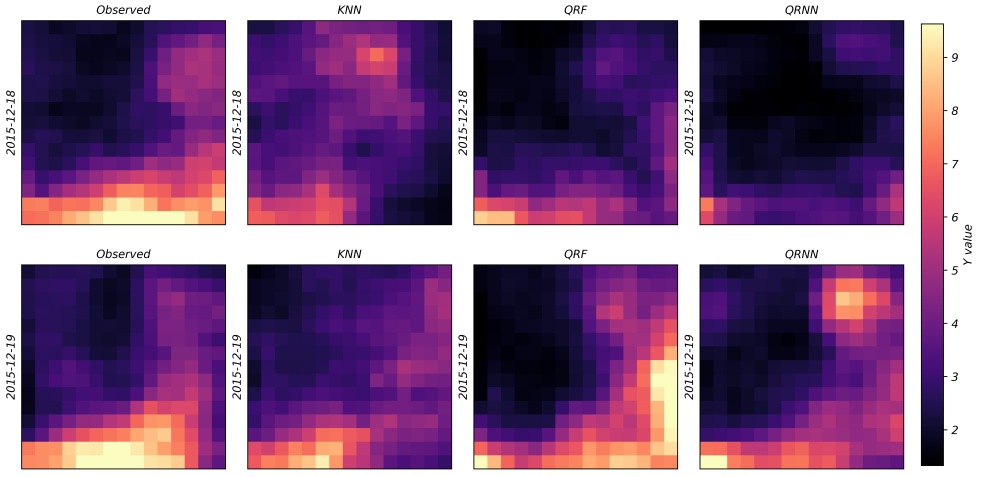

Figure 2: Two consecutive time steps in the test period of the synthetic target field and the corresponding simulations from models fitted with KNN, QRF, and QRNN.

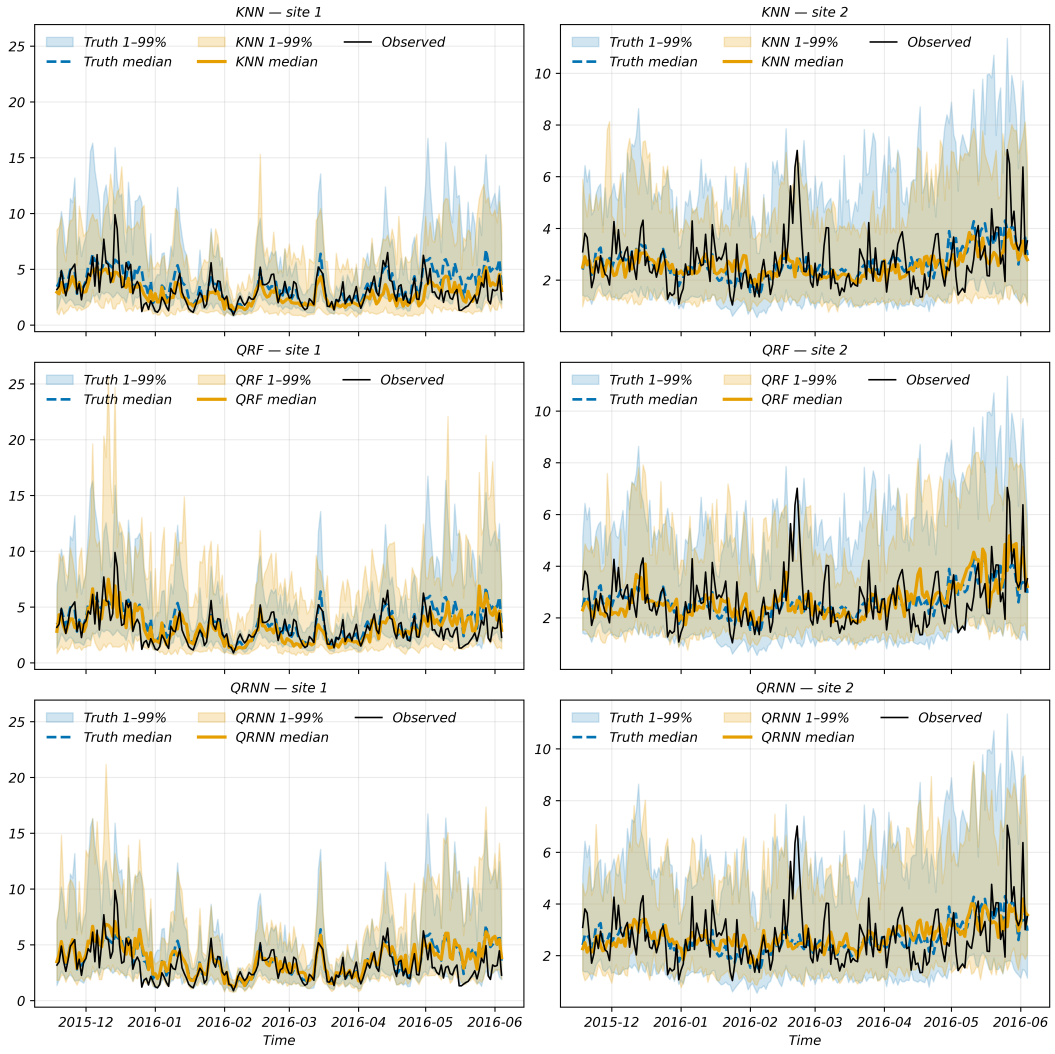

Figure 3: Time series over the test period at two sites, showing the reference dataset, the oracle ensemble 1–99% band and median (computed from oracle ensemble), and the corresponding 1–99% bands and medians from simulations generated by KNN, QRF, and QRNN.

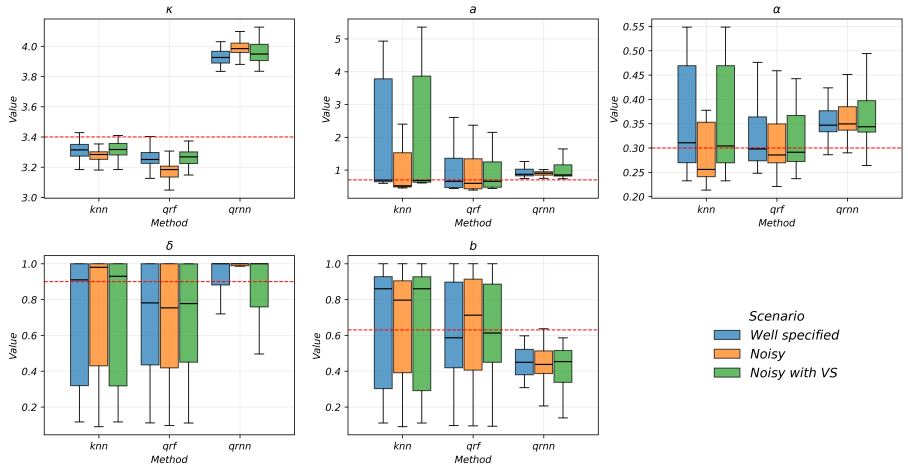

Figure 4: Boxplots of the estimated Gneiting parameters κ, a, α, δ, and b across synthetic replicates in three scenarios (Well-specified, Noisy, Noisy+VS). The red dashed line marks the true value. Methods are KNN, QRF, and QRNN.

quantiles reasonably well, with QRNN providing the closest estimates overall. We then draw 50 simulations from each fitted model over the test period. Figure 2 displays one of these 50 simulations for each method, together with the reference dataset field, at two consecutive time steps in the test period. Visually, all three methods reproduce the main spatial patterns of the reference, including the location and extent of high and low values and the overall smoothness of the field. Note that this figure shows one random realization for each method and therefore are not expected to look the same as the observed fields.

Figure 3 shows the time series during the test period at two selected sites. For each method we plot the median and the $1-99\%$ band derived from the 50 simulations, together with the reference time series and the corresponding oracle band calculated from the oracle ensemble. At both sites, the medians from KNN, QRF, and QRNN closely match the true median (median of the reference) over the test period with QRNN having the best match. The uncertainty bands are also well calibrated, becoming wider when the process is more variable (at higher values) and narrower when it is more stable (at lower values).

Overall, this experiment shows that the proposed framework can recover both the spatial structure and the marginal behavior at each site, as well as the temporal variability, in a realistic way for all three quantile regression methods. In details, KNN tend to produce smoother median and, along with QRF, slightly wider bands. QRNN often leads to more accurate bands that remain close to the oracle envelope.

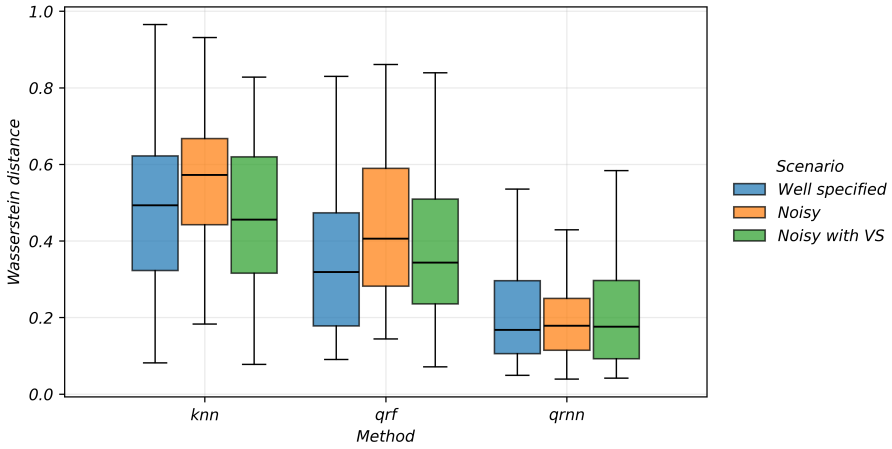

Figure 5: Boxplots of the Wasserstein distance between the oracle and simulated target in the test period with KNN, QRF, and QRNN in three scenarios (Well-specified, Noisy, Noisy+VS).

Since the quantiles have been correctly estimated, as seen in Figures 1 and 3, we are able to derive latent GRF values $\hat{\boldsymbol{z}}$, from which the parameters of the Gneiting covariance function can be estimated. Figure 4 displays boxplots of the estimated parameters of the Gneiting covariance across synthetic replicates for three scenarios with ν being known. Overall, the patterns are consistent across parameters but with notable differences between methods. Regarding variability, it decreases from KNN to QRF to QRNN, with QRNN exhibiting the tightest spreads. Parameterwise, κ is the least variable, whereas the temporal parameters (α, a, δ, b) are more dispersed, especially a and δ. For κ, KNN remains close to the true value in the well-specified case, while QRF shows a slight underestimation and QRNN tends to overestimate the parameter. For a, all methods are broadly unbiased; however, variability is the largest for KNN, smaller for QRF, and the lowest for QRNN. A similar pattern holds for α, with an additional mild positive bias for QRNN. For δ and b, KNN and QRF again exhibit wider spreads, whereas QRNN is comparatively more stable but unbiased. Importantly, the presence of noisy covariates degrades KNN and QRF more strongly, although the variable-selection recenters estimates. In contrast, QRNN is less sensitive to the presence of noise and remains relatively stable across scenarios. Notice that for the spatio-temporal covariance in (9), the parameters a and δ are difficult to estimate jointly. Due to the fact with small temporal lags, the temporal correlation behaves as $1 - a\delta|u|^{2\alpha}$, the product $a\delta$ is more efficiently estimated

than each parameter separately. On Figure 4, higher values of $\hat{\delta}$ correspond in general to lower values of $\hat{a}$, and conversely.

Figure 5 reports the Wasserstein-1 distance between the distribution of the simulated and the distribution of the observed values. For each scenario and each method, we pool all simulated values over space, time, and compare their empirical distribution (constructed from the 50 realizations) to a reference distribution built in the same way from the oracle ensemble. QRNN is the best, then QRF, and KNN is the least-performing method. This matches the training goal; indeed QRNN minimizes the pinball loss, so it matches the marginal distributions very well. QRNN is also less sensitive to noisy covariates, keeping good marginal fit even when extra noisy predictors are added.

4 Application to simulating temperature fields conditional on Geopotential height

4.1 Study area and data

To illustrate the methodology, we consider daily gridded maximum daily temperature over northwest Switzerland as a target variable. As commonly used in the literature (Pang et al., 2017; Huth, 2002), daily gridded 500 hPa geopotential heights over central Europe are used as covariates (Figure 6). The temperature data is a reanalysis provided by MeteoSwiss (MeteoSwiss, 2016) with 2×2 km spatial resolution, which is available daily from 1961 to present. The covariates derive from the ERA5 reanalysis (Hersbach et al., 2023) which has a spatial resolution of 31×31 km over central Europe.

The target area encompasses northwest Switzerland, a region with a range of geographic features including the relatively flat Swiss plateau as well as parts of the Alps. In consequence, significant local variations in temperature can be expected. Note that here the domain of the covariates is very different from the target domain, both in terms of spatial extent and spatial resolution.

4.2 Model setup

Dates are split into a training period (01-01-1978 to 31-12-2017) and a test period (01-01-2018 to 31-12-2022). The training period serves both for fitting the quantile learners and for estimating the parameters of the latent GRF; these parameters are then held fixed when simulating on the test period.

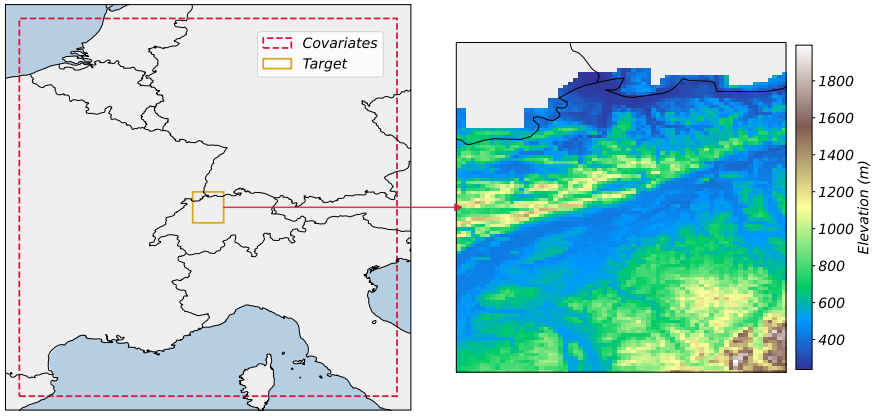

Figure 6: The covariates area and the target area of interest with the corresponding elevation.

For this application we set

$$Y(\mathbf{s}, t) = T_{\max}(\mathbf{s}, t),$$

the daily maximum temperature at site $\mathbf{s}$ and day t. For large-scale predictors, we first apply principal component analysis (PCA) to the geopotential height field over the training period and retain the first m principal components explaining at least 95% of the total variance; in our case this yields $m = 2$ components. Let $\mathrm{PC}_1(t)$ and $\mathrm{PC}_2(t)$ denote the corresponding scores at day t. To represent the annual cycle, we define

$$s_1(t) = \sin\Big(\tfrac{2\pi\,\mathrm{DOY}(t)}{365}\Big), \qquad c_1(t) = \cos\Big(\tfrac{2\pi\,\mathrm{DOY}(t)}{365}\Big),$$

where $\mathrm{DOY}(t)$ is the day of year. All four predictors are then standardized on the training period to have zero mean and unit variance. Denoting these standardized versions by $\widetilde{\mathrm{PC}}_1(t)$, $\widetilde{\mathrm{PC}}_2(t)$, $\tilde{s}_1(t)$, and $\tilde{c}_1(t)$, we define, for every site $\mathbf{s}$,

$$\mathbf{x}_{\mathbf{s}}(t) = \big[\widetilde{\mathrm{PC}}_1(t),\, \widetilde{\mathrm{PC}}_2(t),\, \tilde{s}_1(t),\, \tilde{c}_1(t)\big]^{\top}.$$

Thus all locations share the same large-scale covariates at each time step. This choice reflects the assumption that synoptic conditions (captured by the leading PCs of geopotential height) influence all locations in the study area in a broadly similar way, while site-specific behavior is handled by the site-wise quantile functions and the latent dependence.

4.3 Simulation results

To better understand the role of the covariates, we compute variable importance using QRF at a subset of sites. Concretely, we fit a QRF at 20 randomly chosen locations on

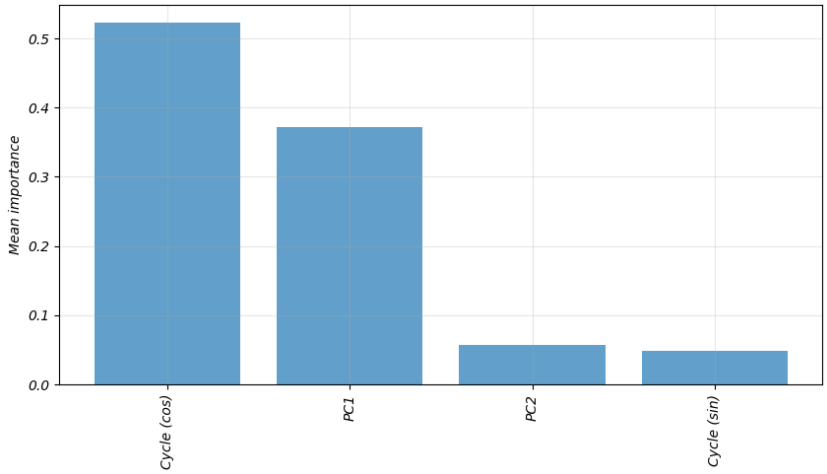

Figure 7: Mean variable importances at randomly selected 20 sites calculated using QRF.

the training period, compute feature importances at each site, and then average them across sites. The resulting aggregated importances are shown in Figure 7. The cosine term of the annual cycle is the most important predictor, followed by the first and second principal components (PC1 and PC2) of geopotential height, while the sine term of the annual cycle is less important but still non-negligible. This indicates that most of the explained variance in temperature in our setup is driven by the seasonal cycle and a small number of large-scale circulation patterns, so that a low-dimensional representation of geopotential height appears sufficient for this application.

We fit the model with the three quantile regression methods KNN, QRF, and QRNN. For each method, hyperparameters are selected by cross-validation on the training period. In particular, for KNN we use an isotropic Gaussian kernel, since all covariates have a unit variance, and choose the bandwidth by cross-validation. After fitting the three models, we generate 50 independent simulations for each method over the test period. Figure 8 shows an example of one instance simulation at 01-08-2018, ensemble mean, and ensemble standard deviation computed from the 50 simulations of each method. The simulations reproduce the main spatial gradients and mesoscale structures seen in the observations, including the cool Alpine band and the warmer Plateau. The median fields closely match the observed temperatures and are similar across methods, whereas the spread (quantified through the ensemble standard deviation) differs between methods: KNN exhibits the largest standard deviation ($\approx 3.6\,^{\circ}\mathrm{C}$), QRF is intermediate ($\approx 2.8\,^{\circ}\mathrm{C}$), and QRNN shows the smallest spread ($\approx 2.4\,^{\circ}\mathrm{C}$). The significantly larger spread produced by KNN can

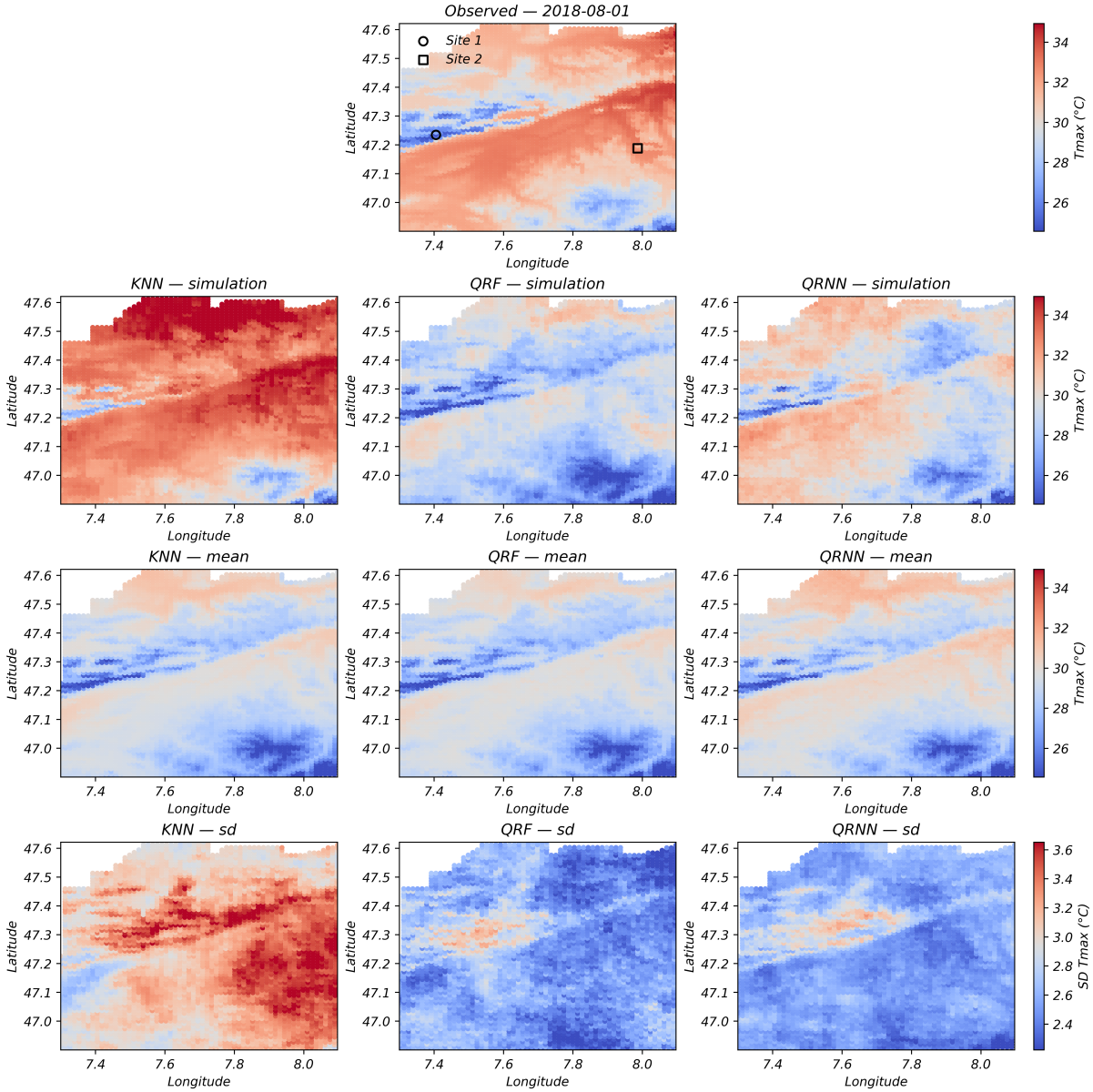

Figure 8: Observed temperature on 01-08-2018 (top row), and for the same date one simulation (second row), ensemble mean (third row), and ensemble standard deviation (bottom row) from the KNN, QRF, and QRNN simulations.

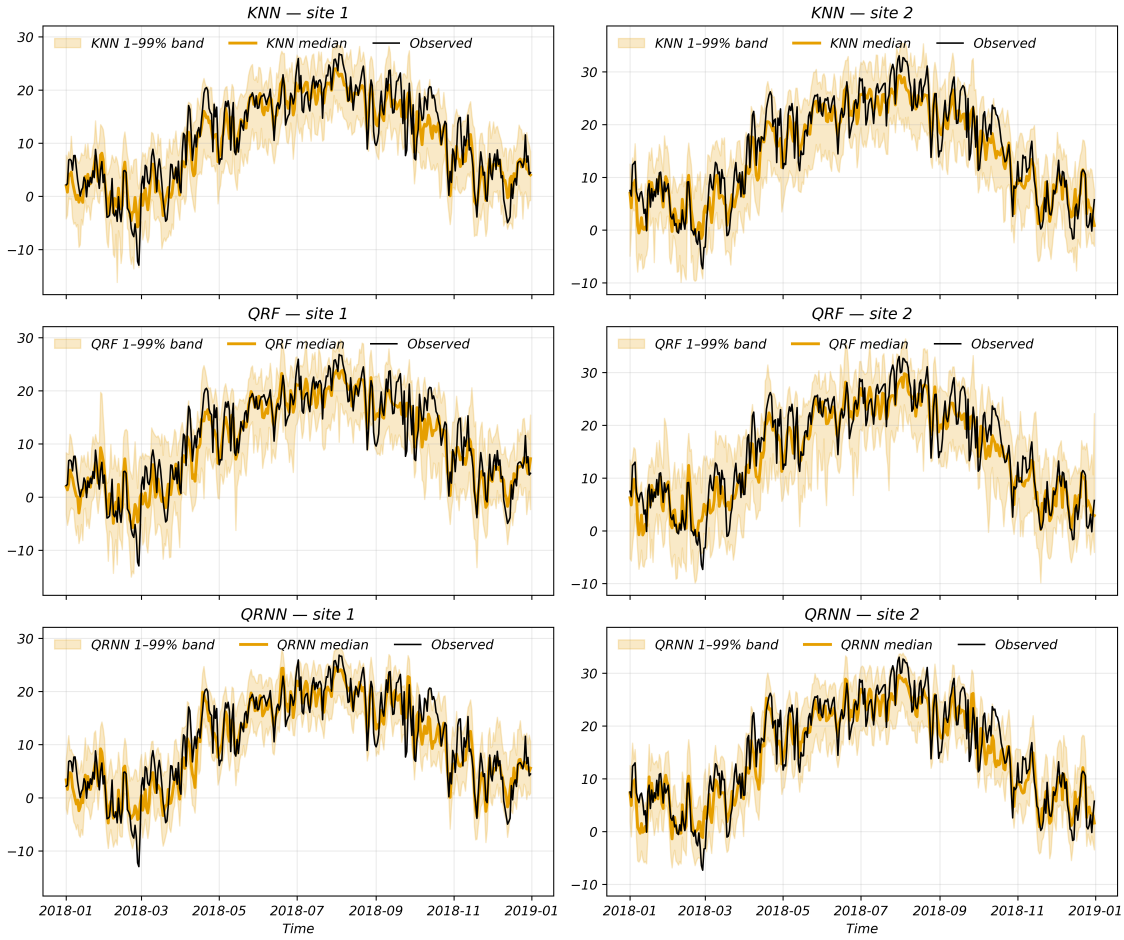

Figure 9: Time series of observed temperature at two locations in 2018 and the median and 1-99% band calculated from 50 independent simulations of KNN, QRF, and QRNN.

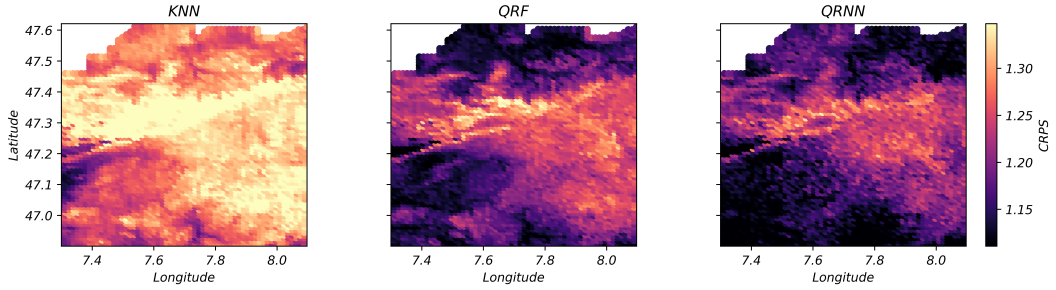

Figure 10: Summer median of continuous ranked probability score (CRPS) in the target area at the test period for each method (KNN, QRF, and QRNN).

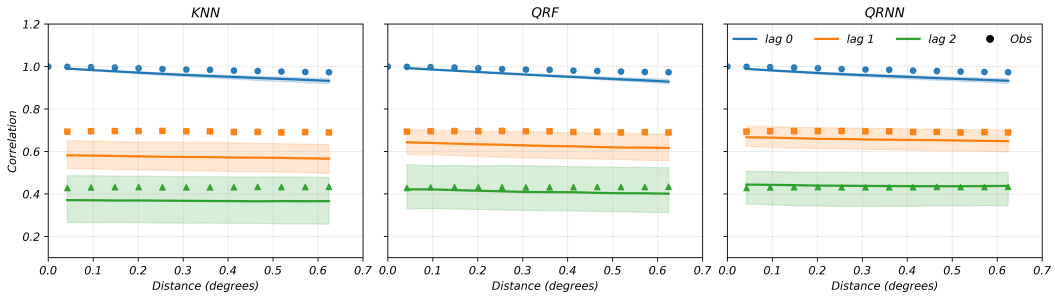

Figure 11: Summer space-time correlations on the test set of observed temperature and the corresponding median and 1–99% bands of the simulation's space-time correlations from KNN, QRF, and QRNN.

be attributed to the nature of the KNN quantile estimator: because KNN relies on local averaging of neighboring values, its estimated conditional quantile functions are smoother and less sharply defined. This smoothing effect leads to a broader conditional distribution, which results in wider predictive spreads.

Consistently, figure 9 shows time series at two sites (whose locations are shown in the top panel of figure 8) with medians that closely match observations and 1–99 % uncertainty bands. QRNN medians are often the closest with tighter bands and KNN's band tends to be wider, as was also seen in Figure 8 in a spatial setting. These qualitative results align with the summer CRPS median maps (Figure 10), where QRNN achieves the lowest scores over most of the domain. All methods have higher CRPS in complex topography (high altitude mountains). Note that, unlike in the synthetic experiment, the Wasserstein distance is not computed here because the true reference distribution of the target variable is unknown in the real dataset.

Figure 11 displays space–time correlations in summer. Correlations decrease with distance and with 1–2-day lags, and all methods capture this pattern. QRNN matches the observed curves best across lags, QRF is next, and KNN is slightly lower.

It is noteworthy that QRNN has the capacity to learn complex relationships and provides reliable marginal fits. QRF is also capable of modeling complex patterns and has demonstrated effectiveness; however, it generally requires additional storage capacity after the fitting process. Table 2 show the computation times for training and simulation for each method in a 32-core machine. KNN trains substantially faster than the other methods. Despite its heavier training phase, QRNN is the most efficient at simulation,

Table 2: Computation times (minutes) for training and simulation (one simulation on test set).

Method	Training	Simulation
KNN	11.4	1.2
QRNN	411.6	0.48
QRF	535.2	16.8

which is advantageous when many stochastic realizations are required. In contrast, QRF requires more storage and is the slowest to simulate. The selection of a model is contingent upon the application's specific requirements. In applications where optimal marginal modeling is important and the resources for training are available, QRNN is a viable option. Conversely, when a robust and interpretable method is needed, QRF is a suitable choice, but with an increased model size. Despite having generally lower scores and using all standardized covariates with equal weights, KNN has some advantages: it is a more straightforward and fast training algorithm, a benefit that is particularly advantageous in large-scale applications. For scenarios where computational efficiency is more important than accuracy, KNN thus can be an option.

5 Conclusions and perspectives

We introduced a general framework for modeling and simulating spatio-temporal processes. This framework combines machine learning quantile regression for local marginals with a latent Gaussian random field for dependence. The concept is simple: first, learn the conditional distribution at each site from the covariates. Then, use a space-time covariance function to model the space-time dependence. This decoupling makes the model flexible through the choice of quantile regression methods while describing the space–time dependence with only a small set of covariance parameters. An appealing feature of the proposed framework is that the covariates do not need to be co-localized with the target variable. In the synthetic experiment, the covariates were co-localized, but in the application, there were 4 covariates covering the whole region. Any other configuration is possible. In the synthetic experiment, the method performed well. With a light variable-selection step, it remained robust to irrelevant covariates. Moreover, in the case study, the approach accurately reproduced space-time patterns and site-wise uncertainty. Beyond the examples

shown here, the method has many potential applications. It is well suited for stochastic downscaling, conditional stochastic weather generators, and impact studies that require ensemble simulations conditional on large-scale predictors. More generally, it provides a practical way to generate space–time fields that respect both covariate information and a prescribed dependence structure.

Several extensions of the framework are possible. On the dependence side, we have focused on a stationary covariance model, but more complex space–time covariances could be used, for example with anisotropy, nonstationarity, or advection, building on recent developments in Gaussian random field modeling (Allard et al., 2025). In such cases, the main change would occur at the estimation stage, where more advanced inference tools are needed to handle the increased number and complexity of parameters. Furthermore, while the simulations can extend beyond the observed range, this framework is not suited for extreme tails. This can be modeled, though, through quantile regression, using methods adapted for extreme values, for example, as described in Pasche and Engelke (2024). Finally, the framework can also be extended to multivariate space-time target variables using multivariate GRFs with multivariate space-time covariance functions, learning marginals independently for each variable.

Alongside these positive results, some limitations remain. A key one is that, despite the flexibility on the marginal side, the dependence structure is effectively Gaussian-copula based. This implies that some types of dependence, in particular strong tail dependence or asymmetric joint behavior, cannot be captured. Relaxing this Gaussian copula assumption, for example by considering more general latent structures, is an important direction for future work.

Acknowledgments

The authors acknowledge the funding of the Swiss National Science Foundation through project number 200021_204130, entitled "Deep-time synthetic data cubes to enable long-term hydrological modelling". The authors also acknowledge the financial support of the Chair Geolearning, funded by Andra, BNP Paribas, CCR and the SCOR Foundation for Science.

References

Denis Allard, Xavier Emery, Céline Lacaux, and Christian Lantuéjoul. Simulating space-time random fields with nonseparable gneiting-type covariance functions. *Statistics and Computing*, 30(5):1479–1495, 2020.

Denis Allard, Lucia Clarotto, and Xavier Emery. Fully nonseparable gneiting covariance functions for multivariate space–time data. *Spatial Statistics*, 52:100706, 2022.

Denis Allard, Lionel Benoit, and Said Obakrim. Modeling and simulating spatio-temporal, multivariate and nonstationary gaussian random fields: a gaussian mixtures perspective. *hal-05034982*, 2025.

Lionel Benoit, Lydie Sichoix, Alison D Nugent, Matthew P Lucas, and Thomas W Giambelluca. Stochastic daily rainfall generation on tropical islands with complex topography. *Hydrology and Earth System Sciences Discussions*, 2021:1–22, 2021.

Pallab K Bhattacharya and Ashis K Gangopadhyay. Kernel and nearest-neighbor estimation of a conditional quantile. *The Annals of Statistics*, pages 1400–1415, 1990.

Alex J Cannon. Quantile regression neural networks: Implementation in r and application to precipitation downscaling. *Computers & geosciences*, 37(9):1277–1284, 2011.

Wanfang Chen, Marc G Genton, and Ying Sun. Space-time covariance structures and models. *Annual Review of Statistics and Its Application*, 8(1):191–215, 2021.

Mario Figueira, Michela Cameletti, and Luca Patelli. Inla-rf: A hybrid modeling strategy for spatio-temporal environmental data. *arXiv preprint arXiv:2507.18488*, 2025.

Robin Genuer, Jean-Michel Poggi, and Christine Tuleau-Malot. Variable selection using random forests. *Pattern recognition letters*, 31(14):2225–2236, 2010.

Tilmann Gneiting. Nonseparable, stationary covariance functions for space–time data. *Journal of the american statistical association*, 97(458):590–600, 2002.

Peng-Tao Guo, Mao-Fen Li, Wei Luo, Qun-Feng Tang, Zhi-Wei Liu, and Zhao-Mu Lin. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. *Geoderma*, 237:49–59, 2015.

Abubakar Haruna, Juliette Blanchet, and Anne-Catherine Favre. Joint estimation of trend in bulk and extreme daily precipitation in switzerland. *Weather and Climate Extremes*, page 100769, 2025.

Hans Hersbach, Bill Bell, Paul Berrisford, Gionata Biavati, András Horányi, Joaquín Muñoz Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Iryna Rozum, et al. Era5 hourly data on single levels from 1940 to present, copernicus climate change service (c3s) climate data store (cds)[data set], 2023.

Radan Huth. Statistical downscaling of daily temperature in central europe. *Journal of Climate*, 15(13):1731–1742, 2002.

Zachary James and Joseph Guinness. Implementation and analysis of gpu algorithms for vecchia approximation. *Statistics and Computing*, 34(6):207, 2024.

Thomas Kneib, Alexander Silbersdorff, and Benjamin Säfken. Rage against the mean–a review of distributional regression approaches. *Econometrics and Statistics*, 26:99–123, 2023.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

Finn Lindgren and Håvard Rue. Bayesian spatial modelling with r-inla. *Journal of statistical software*, 63:1–25, 2015.

Finn Lindgren, David Bolin, and Håvard Rue. The spde approach for gaussian and non-gaussian fields: 10 years and still running. *Spatial Statistics*, 50:100599, 2022.

Bruce G Lindsay. Composite likelihood methods. In *Statistical Inference from Stochastic Processes: Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference Held August 9-15, 1987, with Support from the National Science Foundation and the Army Research Office*, volume 80, page 221. American Mathematical Soc., 1988.

Nicolai Meinshausen and Greg Ridgeway. Quantile regression forests. *Journal of machine learning research*, 7(6), 2006.

MeteoSwiss. Documentation of meteoswiss grid-data products daily mean, minimum and maximum temperature: Tabsd, tmind, tmaxd. 2016.

Said Obakrim, Lionel Benoit, and Denis Allard. A multivariate and space-time stochastic weather generator using a latent gaussian framework. *Stochastic Environmental Research and Risk Assessment*, pages 1–25, 2025.

Oscar Hernan Madrid Padilla, Wesley Tansey, and Yanzhen Chen. Quantile regression with relu networks: Estimators and minimax rates. *Journal of Machine Learning Research*, 23(247):1–42, 2022.

Bo Pang, Jiajia Yue, Gang Zhao, and Zongxue Xu. Statistical downscaling of temperature with the random forest model. *Advances in Meteorology*, 2017(1):7265178, 2017.

Olivier C Pasche and Sebastian Engelke. Neural networks for extreme quantile regression with an application to forecasting of flood risk. *The Annals of Applied Statistics*, 18(4): 2818–2839, 2024.

Vivek Kumar Patidar, Rajesh Wadhvani, Sanyam Shukla, Muktesh Gupta, and Manasi Gyanchandani. Quantile regression comprehensive in machine learning: a review. In *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pages 1–6. IEEE, 2023.

Emilio Porcu, Reinhard Furrer, and Douglas Nychka. 30 years of space–time covariance functions. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(2):e1512, 2021.

Nagireddy Masthan Reddy, Subbarayan Saravanan, and Devanantham Abijith. Streamflow simulation using conceptual and neural network models in the hemavathi subwatershed, india. *Geosystems and Geoenvironment*, 2(2):100153, 2023.

Matthias Seeger. Gaussian processes for machine learning. *International journal of neural systems*, 14(02):69–106, 2004.

Ying Sun, Bo Li, and Marc G Genton. Geostatistics for large datasets. In *Advances and challenges in space-time modelling of natural events*, pages 55–77. Springer, 2011.

Silvia Terzago, Elisa Palazzi, and Jost von Hardenberg. Stochastic downscaling of precipitation in complex orography: A simple method to reproduce a realistic fine-scale climatology. *Natural Hazards and Earth System Sciences*, 18(11):2825–2840, 2018.