

Article

Not peer-reviewed version

Uncertainty-Aware and Non-Negative Hydrological Forecasting Using Gamma-Likelihood Chained Gaussian Processes for Sustainability-Oriented Water Management

[Yesika Alexandra Bastidas-Pantoja](#)*, [Julián David Pastrana-Cortés](#), [Julián Gil-González](#), [David Augusto Cárdenas-Peña](#), Jhoniers Gilberto Guerrero-Erazo

Posted Date: 3 April 2026

doi: 10.20944/preprints202604.0184.v1

Keywords: sustainable water management; hydrological forecasting; Gaussian processes; uncertainty quantification; hydroclimatic variability



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Uncertainty-Aware and Non-Negative Hydrological Forecasting Using Gamma-Likelihood Chained Gaussian Processes for Sustainability-Oriented Water Management

Yesika Alexandra Bastidas-Pantoja ^{1,*}, Julián David Pastrana-Cortés ¹, Julián Gil-González ¹
David Cárdenas-Peña ¹ and Jhoniers Gilberto Guerrero-Erazo ²

¹ Automatics Research Group, Universidad Tecnológica de Pereira (UTP), Pereira 660003, Colombia
² Water and Sanitation Research Group, Universidad Tecnológica de Pereira, Pereira, Colombia
* Correspondence: yesika.bastidas@utp.edu.co

Abstract

Suitable forecasting of streamflow plays a fundamental role in the sustainable management of water resources. This is especially important in regions with high hydroclimatic variability, such as Colombia. Reliable predictions of water availability are essential to support decision-making on water allocation, risk mitigation, and the long-term resilience of supply systems, including urban aqueduct services. However, traditional deterministic models often require extensive parameterization. They may lack flexibility. Data-driven approaches such as Long Short-Term Memory networks typically fail to provide meaningful uncertainty quantification. In this work, we introduce the Chained Correlated Gaussian Process. This is a scalable probabilistic framework for jointly forecasting multiple hydrological time series. The proposed model represents likelihood parameters through latent Gaussian Processes. This enables the use of non-Gaussian likelihoods to enforce physical constraints on the outputs. In particular, a Gamma likelihood is integrated into a Linear Model of Coregionalization. This approach captures interdependencies among multiple time series while ensuring non-negativity. To achieve computational scalability, we employ a variational inference framework with shared inducing points. This reduces computational complexity from cubic to linear with respect to the number of observations. Additionally, a hybrid optimization strategy is adopted. It combines Adam and Natural Gradient methods to improve numerical stability and avoid poor local optima. The proposed methodology is validated using twelve years of daily observations from 23 reservoirs in Colombia. These reservoirs serve as proxies for water availability dynamics within interconnected resource systems. Experimental results demonstrate that the ChdGP framework consistently outperforms baseline models across multiple forecasting horizons. The Gamma-likelihood configuration shows superior performance. It provides physically consistent predictions, improves uncertainty quantification, and effectively captures asymmetric behaviors and extreme hydrological events. Performance evaluation based on the Negative Log Predictive Density metric confirms the robustness of the proposed approach under non-Gaussian conditions. The results highlight the potential of the proposed framework as a reliable tool for probabilistic forecasting and risk-aware decision-making in sustainable water resource management. It offers a flexible, scalable methodology for modeling complex hydrological systems under uncertainty.

Keywords: sustainable water management; hydrological forecasting; Gaussian processes; uncertainty quantification; hydroclimatic variability

1. Introduction

Water constitutes the most fundamental natural resource as it sustains life, ecosystems, and human societies. As global population growth and development continue, the demand for water has increased

significantly, drawing international attention to the sustainable management of hydrological resources [1]. The excessive exploitation of these resources has led to multiple challenges that hinder economic development and increase environmental concerns, such as declining river flows. Consequently, advancing understanding of hydrological processes is now a key focus in natural resource research, as it enhances the ability to anticipate extreme events such as droughts and periods of intense rainfall [2]. Moreover, changes in climate patterns have increased the temporal and spatial variability of water resources, making their behavior increasingly complex [3].

In the Colombian context, a country characterized by high hydrological endowment, the real challenge is no longer the availability of water, but its sustainable and efficient management under conditions of pronounced hydroclimatic variability [4]. Recent studies have highlighted that the increasing frequency of extreme events—such as prolonged droughts and intense precipitation—combined with anthropogenic pressures, significantly affects the reliability of municipal water supply systems and ecosystem services [5,6]. Consequently, there is a growing need to transition from traditional deterministic planning approaches toward data-driven and probabilistic frameworks that explicitly incorporate uncertainty and variability in hydrological processes.

Initial approaches to forecasting water resources rely on physically based deterministic models that describe natural laws governing hydrological time series. Despite their theoretical foundation, these models pose major challenges, as obtaining reliable results demands substantial effort and detailed knowledge of numerous system parameters, often leading to restricted practical applicability [7,8]. An alternative strategy employs statistical and machine learning techniques that are entirely data-driven, allowing the capture of complex patterns directly from observations [9]. Within this scope, autoregressive (AR) models have been extensively applied; however, they are frequently unable to represent the stochastic and nonlinear behavior characteristic of water resources [10].

To address these limitations, Artificial Neural Networks (ANNs) have been widely adopted in hydrological studies due to their ability to model nonlinear relationships. Nevertheless, ANNs present several limitations, including overfitting, sensitivity to local minima, and issues related to vanishing and exploding gradients during training [11]. Long Short-Term Memory (LSTM) networks are among the most widely recognized neural architectures for time series modeling, as they are specifically developed to process sequential data. In particular, LSTMs effectively address the vanishing gradient problem inherent in standard recurrent architectures [12–14]. Recent research has demonstrated the effectiveness of LSTM-based recurrent neural network for monthly low-flow prediction at the Basantapur station in India [13] and an alternative methodology was presented in [12], where a wavelet transform was first used to filter noise and remove interfering components from hydrological time series. Despite their flexibility, a key limitation of these neural approaches is their inability to provide information about predictive uncertainty [15]. In practical applications, understanding the level of confidence associated with a forecast is essential for informed decision-making and risk management [16].

Gaussian Processes (GPs) offer a robust stochastic modeling alternative by providing full predictive distributions rather than single-point estimates [17]. By linking observations to multivariate normal distributions, GPs effectively capture the random nature and intrinsic non-linearity of hydrological processes [18,19]. The work reported in [18] assessed the effectiveness of several approaches for daily water resource time series forecasting, including Gaussian Process models. The results underscored the capability of GPs to account for data noise and represent nonlinear temporal dynamics. In watershed studies [20] introduced a new method for quantifying uncertainty, employing Gaussian Processes within a Bayesian sensitivity analysis framework. The main goal of this analysis was to detect and localize the factors that most strongly contribute to model uncertainty. Furthermore, hybrid models integrating LSTMs with GPs have been proposed to preserve nonparametric probabilistic characteristics. Within the Colombian context, the study presented in [21] applied a SARIMA model to forecast monthly precipitation. The findings are particularly significant, as they illustrate how

such predictions can support a wide range of hydrological analyses and contribute to effective water resource management.

While Gaussian Processes offer strong probabilistic foundations, many real-world hydrological applications require the simultaneous forecasting of multiple correlated time series, a setting commonly referred to as multi-task or multi-output learning. This problem can be addressed by extending GP frameworks to model multiple outputs jointly, thereby explicitly capturing inter-task dependencies. A widely adopted approach relies on collections of latent GPs combined through coregionalization matrices, commonly known as separable kernel sum (SoS) models [22]. Although these models are highly interpretable, their application in hydrology is often hindered by rapid parameter growth as the number of tasks increases, leading to overparameterization and significant challenges in model estimation and validation [23]. Even with mitigation strategies such as variational inference [24] or low-rank approximations, computational complexity typically scales cubically with the number of tasks, limiting practical applicability in highly expressive settings [25]. Moreover, existing multi-output GP models, including Linear Models of Coregionalization (LMCGP) [25], may assign predictive mass to negative values, violating the inherent non-negativity constraint of streamflow data [26].

This paper presents the Chained Correlated Gaussian Process (ChdGP) as a scalable, probabilistic model for the joint forecasting of multiple hydrological time series in Colombia. Unlike traditional GP frameworks, the proposed approach represents likelihood parameters through latent processes, enabling the use of non-Gaussian likelihoods to enforce non-negativity in predictions. Additionally, we employ a hybrid optimization strategy that combines Adam and Natural Gradient (NG) methods to enhance stability and avoid poor local optima in variational inference. The proposed model is validated using data from 23 strategic reservoirs, demonstrating that ChdGP effectively exploits inter-task correlations and heteroscedastic noise in the input space, significantly enhancing forecasting accuracy and uncertainty quantification across multiple horizons.

The remainder of this paper is organized as follows. Section 2 presents the mathematical framework, including the problem formulation, the multi-output Gaussian Process models, and the variational inference approach. Section 3 describes the dataset and experimental setup; and discusses the results, highlighting the implications of probabilistic modeling for hydrological forecasting and sustainable water resource management. Finally, Section 4 concludes the paper by summarizing the main findings, proposing directions for future research.

2. Mathematical Framework

2.1. Problem Setting

Let $\mathbf{v}_n \in \mathbb{R}^D$ denote the vector of hydrological observations of the D outputs at time instant n . The proposed model uses the complete sequence of these vectors from time $n - T + 1$ to time n as input to forecast the resource vector at a future time $n + H$. In this formulation, T corresponds to the model order, while H indicates the forecasting horizon. Accordingly, the input vector $\mathbf{x}_n$ is defined as:

$$\mathbf{x}_n = \begin{bmatrix} \mathbf{v}_n^\top \\ \mathbf{v}_{n-1}^\top \\ \vdots \\ \mathbf{v}_{n-T+1}^\top \end{bmatrix} \in \mathbb{R}^{TD}$$

and the target output vector $\mathbf{y}_n$ as follows:

$$\mathbf{y}_n = \mathbf{v}_{n+H}.$$

This formulation allows the model to effectively exploit past hydrological observations to generate reliable forecasts of future values. Based on this approach, we construct a dataset $\mathcal{D} = \{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^N = \{\mathbf{X}, \mathbf{y}\}$ consisting of N input–output pairs, where $\mathbf{X} \subset \mathcal{X} \in \mathbb{R}^{TD}$ represents the dimensional input space. The target is defined as a vector-valued function, with $\mathbf{y}_n \in \mathbb{R}^D$ containing the observations of

all outputs (tasks) corresponding to the same input $\mathbf{x}_n$. Thus, $\mathbf{y}_n = \{y_{d,n}\}_{d=1}^D$, and the complete set of outputs is expressed as $\mathbf{y} = [\mathbf{y}_1^\top, \mathbf{y}_2^\top, \dots, \mathbf{y}_D^\top]^\top \in \mathbb{R}^{DN}$, where each task-specific output vector d is given by $\mathbf{y}_d = [y_{d,1}, y_{d,2}, \dots, y_{d,n}]^\top \in \mathbb{R}^N$.

2.2. LMCGP for Chained GPs Regression Framework

When the output variable is subject to constraints, such as being restricted to non-negative values, standard probabilistic models like Gaussian Processes may encounter limitations. Deterministic neural networks can naturally enforce such restrictions by applying activation functions like the Rectified Linear Unit at the output layer to ensure non-negative predictions. However, Gaussian Processes require appropriate transformations of random variables in order to properly define their probability distributions. The authors in [26] proposed a framework for handling output constraints by selecting an appropriate distribution likelihood family, with each parameter modeled by a single-output Gaussian Process. Additionally, to extend this formulation to multiple outputs, the method introduced in [27] provides a way to incorporate inter-task correlations while simultaneously managing constraints for each task.

Accordingly, we consider the outputs $\mathbf{y}_d$ to be conditionally independent given a set of parameters $\boldsymbol{\theta}_d = [\theta_{d,1}, \theta_{d,2}, \dots, \theta_{d,J_d}]^\top \in \mathbb{R}^{J_d}$, where J_d represents the number of parameters that characterize the likelihood distribution of the d -th output. In the same manner, the complete parameter vector is defined as $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \dots, \boldsymbol{\theta}_D^\top]^\top \in \mathbb{R}^J$, with $J = \sum_{d=1}^D J_d$.

Considering a Gaussian likelihood, we can factorize it over the training set instances. The likelihood of observing all output realizations $\mathbf{y}$ given $\boldsymbol{\theta}$ can be expressed as a product of individual likelihoods:

$$p(\mathbf{y} | \boldsymbol{\theta}) = \prod_{d=1}^D p(\mathbf{y}_d | \boldsymbol{\theta}_d) \quad (1)$$

Each j -th parameter of the d -th likelihood distribution $\theta_{d,j}$ is a deterministic transformation of a latent Gaussian Process prior realization $f_{d,j}$, given by $\theta_{d,j} = g_{d,j}(f_{d,j})$. For further detail, we define:

- $\mathbf{f}_{d,j} = [f_{d,j,1}, f_{d,j,2}, \dots, f_{d,j,N}]^\top \in \mathbb{R}^N$ as the vector of latent function evaluations for the j -th parameter of likelihood associated to the d -th output, where $f_{d,j,n} = f_{d,j}(\mathbf{x}_n)$.
- $\mathbf{f}_d = [\mathbf{f}_{d,1}^\top, \mathbf{f}_{d,2}^\top, \dots, \mathbf{f}_{d,J_d}^\top]^\top \in \mathbb{R}^{J_d N}$ as the stacked vector of all latent function evaluations of likelihood associated to the d -th output
- $\mathbf{f} = [\mathbf{f}_1^\top, \mathbf{f}_2^\top, \dots, \mathbf{f}_D^\top]^\top \in \mathbb{R}^{JN}$ as the grand vector of all evaluations across all outputs.

The conditionally independent likelihood is then formulated as follows:

$$p(\mathbf{y} | \boldsymbol{\theta}) = p(\mathbf{y} | \mathbf{f}) = \prod_{d=1}^D p(\mathbf{y}_d | \mathbf{f}_d) = \prod_{d=1}^D \prod_{n=1}^N p(y_{d,n} | f_{d,1,n}, f_{d,2,n}, \dots, f_{d,J_d,n}) \quad (2)$$

This formulation introduces J latent parameter functions $f_{d,j}(\mathbf{x})$, each governed by a GP prior, therefore:

$$\mathbf{f}_d(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_d(\mathbf{x}, \mathbf{x}')) \quad (3)$$

with $\mathbf{f} : \mathbf{x} \rightarrow \mathbb{R}^{Jd}$ representing the vector-valued function responsible for mapping the input space. If all latent functions are assumed independent, the joint prior over $\mathbf{f}$ evaluated at the training inputs $\mathbf{X}$ follows:

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_D \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{K}_D \end{bmatrix} \right) \quad (4)$$

Here, $\mathbf{K}_d \in \mathbb{R}^{J_d N \times J_d N}$ denotes the covariance matrix associated with the parameters that characterize the likelihood distribution of the d -th output. Under the assumption of independence, the grand covariance matrix of the multi-output model, $\mathbf{K} \in \mathbb{R}^{JN \times JN}$, present a diagonal by block structure. This formulation is referred to as the Independent Gaussian Process (IGP) model and is equivalent to employing J separate single-output Gaussian Processes. Although this approach can be effective for multi-output prediction, the main motivation for jointly learning multiple outputs is to exploit the dependencies among them, allowing each output to inform and enhance the predictions of the others. Such dependencies can be incorporated by introducing cross-covariance terms to replace the zero entries in the covariance matrix, as suggested in earlier works [25,28], where proposed the Linear Model of Coregionalization Gaussian Process (LMCGP).

Our goal is to represent $\mathbf{f}(\mathbf{x})$ with a multi-output Gaussian process that explicitly captures correlations among the latent parameter functions $f_{d,j}(\mathbf{x})$. To this end, we adopt the Linear Model of Coregionalization Gaussian Process (LMCGP), which encodes the dependence between any pair of latent variables $f_{d,j}(\mathbf{x})$ and $f_{d',j'}(\mathbf{x}')$ through an instantaneous, time-invariant linear combination of a set of independent Gaussian processes $\{u_q(\mathbf{x})\}_{q=1}^Q$. Each of these processes is defined by its own covariance function $k_q(\mathbf{x}, \mathbf{x}')$ and shares the same block diagonal structure specified in Equation (4).

Under this formulation, every latent process $f_{d,j}$ can be expressed as follows:

$$f_{d,j}(\mathbf{x}) = \sum_{q=1}^Q a_{d,j,q} u_q(\mathbf{x}) \quad u_q(\mathbf{x}) \sim \mathcal{GP}(0, k_q(\mathbf{x}, \mathbf{x}')) \quad (5)$$

being k_q the kernel function of the independent process q , governed by the parameters set Φ_q . Due to the independency assumption $\text{cov}\{u_q(\mathbf{x}), u_{q'}(\mathbf{x}')\} = k_q(\mathbf{x}, \mathbf{x}') \delta_{qq'}$. The cross-covariance function of the latent parameter GP $\mathbf{f}$ is as follows:

$$\begin{aligned} k_{(d,j),(d',j')}(\mathbf{x}, \mathbf{x}') &= \text{cov}\{f_{d,j}(\mathbf{x}), f_{d',j'}(\mathbf{x}')\} \\ &= \sum_{q=1}^Q \sum_{q'=1}^Q a_{d,j,q} a_{d',j',q'} \text{cov}\{u_q(\mathbf{x}), u_{q'}(\mathbf{x}')\} \\ &= \sum_{q=1}^Q a_{d,j,q} a_{d',j',q} k_q(\mathbf{x}, \mathbf{x}') \\ &= \sum_{q=1}^Q b_{(d,j),(d',j')}^q k_q(\mathbf{x}, \mathbf{x}') \end{aligned} \quad (6)$$

where $b_{(d,j),(d',j')}^q = a_{d,j,q} a_{d',j',q}$ encodes the interactions among the outputs models by q -th independent process, meanwhile, $k_q(\mathbf{x}, \mathbf{x}')$ manage the interaction among input space view from q -th independent process. This kernel formulation, termed as the sum of separable kernels (SoS kernels), enables the cross-covariance function to operate in an extended input space $\{1, 2, \dots, D\} \times \mathbf{X}$. Instead of a scalar kernel function, we now have a matrix kernel function:

$$\mathbf{k}(\mathbf{x}, \mathbf{x}') = \sum_{q=1}^Q \mathbf{B}_q k_q(\mathbf{x}, \mathbf{x}') \quad (7)$$

with $\mathbf{B}_q \in \mathbb{R}^{D \times D}$ as the coregionalization matrix, with elements $(\mathbf{B}_q)_{(d,j),(d',j')} = b_{(d,j),(d',j')}^q$.

The generative process can be summarized as follows. First, the latent vector $\mathbf{f}$ is sampled from a multivariate normal distribution, $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$, where $\mathbf{K} \in \mathbb{R}^{JN \times JN}$ is the covariance matrix with elements $k_{(d,j),(d',j')}(\mathbf{x}, \mathbf{x}')$, computed in some set of inputs. Once we obtain $\mathbf{f}$, we evaluate the vector of parameters $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \dots, \boldsymbol{\theta}_D^\top]^\top$. Subsequently, the output vector $\mathbf{y} = [\mathbf{y}_1^\top, \mathbf{y}_2^\top, \dots, \mathbf{y}_D^\top]^\top$ is generated, where each component $\mathbf{y}_d$ is obtained by sampling the conditional distribution $p(\mathbf{y}_d | \boldsymbol{\theta}_d)$.

2.3. Variational Inference

The computational complexity of Gaussian Process (GP) models is dominated by the inversion of the covariance matrix $\mathbf{K}$. A common strategy to alleviate this limitation is to introduce a reduced set of $M \ll N$ inducing locations, denoted by $\mathbf{Z}$. Instead of directly working with the latent processes $\mathbf{f}_{d,j}$, the model is reformulated in terms of inducing variables associated with a set of Q independent latent processes $\{u_q\}_{q=1}^Q$. For each process u_q , the corresponding inducing variables are defined as $\mathbf{u}_q \in \mathbb{R}^M$, containing the evaluations of u_q at the inducing locations $\mathbf{Z} \in \mathbb{R}^M$, which are obtained from the input space $\mathbf{X}$. That is, $\mathbf{u}_q = [u_q(\mathbf{z}_1), u_q(\mathbf{z}_2), \dots, u_q(\mathbf{z}_M)]^\top \in \mathbb{R}^M$ becomes the vector of inducing variables for the inducing locations set and $\mathbf{u} = [\mathbf{u}_1^\top, \mathbf{u}_2^\top, \dots, \mathbf{u}_D^\top]^\top \in \mathbb{R}^{QM}$ the grand vector of all inducing variables across all inducing locations.

Since the joint Gaussian prior distribution can be written as $p(\mathbf{f}, \mathbf{u}) = p(\mathbf{f} | \mathbf{u})p(\mathbf{u})$, the latent functions $\mathbf{f}_{d,j}$ are conditionally independent given $\mathbf{u}$. Hence, the conditional distribution is expressed as:

$$p(\mathbf{f} | \mathbf{u}) = \prod_{d=1}^D \prod_{j=1}^{J_d} p(\mathbf{f}_{d,j} | \mathbf{u}) = \prod_{d=1}^D \prod_{j=1}^{J_d} \mathcal{N}(\mathbf{f}_{d,j} | \mathbf{K}_{\mathbf{u},(d,j)}^{-1} \mathbf{K}_{\mathbf{u},\mathbf{u}} \mathbf{u}, K_{(d,j)} - \mathbf{K}_{\mathbf{u},(d,j)}^\top \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1} \mathbf{K}_{\mathbf{u},(d,j)}) \quad (8)$$

Let $K_{(d,j)} \in \mathbb{R}^{N \times N}$ denote the covariance matrix corresponding to the j -th latent parameter function for the d -th output, $\mathbf{f}_{d,j}$. Additionally, $\mathbf{K}_{\mathbf{u},\mathbf{u}} \in \mathbb{R}^{QM \times QM}$ represents the covariance matrix for the inducing variables at inducing locations $\mathbf{u}$. Finally, $\mathbf{K}_{\mathbf{u},(d,j)} \in \mathbb{R}^{QM \times N}$ signifies the covariance matrix between the j -th latent parameter of the d -th output, $\mathbf{f}_{d,j}$, and the inducing variables $\mathbf{u}$.

The posterior joint distribution over the latent variable $\mathbf{f}$ and inducing variable $\mathbf{u}$ takes the following form:

$$p(\mathbf{f}, \mathbf{u} | \mathbf{y}) = p(\mathbf{f} | \mathbf{u})p(\mathbf{u} | \mathbf{y}) \quad (9)$$

One of the main issues of the Chained GP is that exact posterior inference is generally intractable due to the presence of non-Gaussian likelihoods. We use variational inference to compute a lower bound $\mathcal{L}$ for the marginal log-likelihood $\log p(\mathbf{y})$, and for approximating the joint posterior distribution $p(\mathbf{f}, \mathbf{u} | \mathbf{S})$ as follows:

$$p(\mathbf{f}, \mathbf{u} | \mathbf{S}) \approx q(\mathbf{f}, \mathbf{u}) = p(\mathbf{f} | \mathbf{u})q(\mathbf{u}) = \prod_{d=1}^D \prod_{j=1}^{J_d} p(\mathbf{f}_{d,j} | \mathbf{u}) \prod_{q=1}^Q q(\mathbf{u}_q) \quad (10)$$

where $q(\mathbf{u}_q) = \mathcal{N}(\mathbf{u}_q | \boldsymbol{\mu}_q, S_q)$ are variational distributions assumed Gaussian whose parameters $\{\boldsymbol{\mu}_q, S_q\}_{q=1}^Q$ must be optimized. The lower bound $\mathcal{L}$ for $\log p(\mathbf{y})$ is obtained as follows:

$$\begin{aligned} \log p(\mathbf{y}) &= \log \int \int p(\mathbf{y} | \mathbf{f})p(\mathbf{f} | \mathbf{u})p(\mathbf{u})d\mathbf{f}d\mathbf{u} \\ &= \log \int \int q(\mathbf{f}, \mathbf{u}) \frac{p(\mathbf{y} | \mathbf{f})p(\mathbf{f} | \mathbf{u})p(\mathbf{u})}{q(\mathbf{f}, \mathbf{u})} d\mathbf{f}d\mathbf{u} \\ &\geq \int \int q(\mathbf{f}, \mathbf{u}) \log \frac{p(\mathbf{y} | \mathbf{f})p(\mathbf{f} | \mathbf{u})p(\mathbf{u})}{q(\mathbf{f}, \mathbf{u})} d\mathbf{f}d\mathbf{u} = \mathcal{L} \end{aligned} \quad (11)$$

We can further simplify $\mathcal{L}$ to obtain:

$$\begin{aligned}
 \mathcal{L} &= \int \int q(\mathbf{f}, \mathbf{u}) \log \frac{p(\mathbf{y} | \mathbf{f}) p(\mathbf{f} | \mathbf{u}) p(\mathbf{u})}{q(\mathbf{f}, \mathbf{u})} d\mathbf{f} d\mathbf{u} \\
 &= \int \int q(\mathbf{f}, \mathbf{u}) \log p(\mathbf{y} | \mathbf{f}) d\mathbf{f} d\mathbf{u} - \int q(\mathbf{u}) \log \frac{q(\mathbf{u})}{p(\mathbf{u})} d\mathbf{u} \\
 &= \int q(\mathbf{f}) \log p(\mathbf{y} | \mathbf{f}) d\mathbf{f} - \sum_{q=1}^Q \text{KL}\{q(\mathbf{u}_q) \parallel p(\mathbf{u}_q)\} \\
 &= \mathbb{E}_{q(\mathbf{f})} \{\log p(\mathbf{y} | \mathbf{f})\} - \sum_{q=1}^Q \text{KL}\{q(\mathbf{u}_q) \parallel p(\mathbf{u}_q)\} \\
 &= \sum_{d=1}^D \mathbb{E}_{q(\mathbf{f}_{d,1}), \dots, q(\mathbf{f}_{d,J_d})} \{\log p(\mathbf{y}_d | \mathbf{f}_{d,1}, \dots, \mathbf{f}_{d,J_d})\} - \sum_{q=1}^Q \text{KL}\{q(\mathbf{u}_q) \parallel p(\mathbf{u}_q)\} \\
 &= \sum_{d=1}^D \sum_{n=1}^N \mathbb{E}_{q(\mathbf{f}_{d,1,n}), \dots, q(\mathbf{f}_{d,J_d,n})} \{\log p(y_{d,n} | f_{d,1,n}, \dots, f_{d,J_d,n})\} - \sum_{q=1}^Q \text{KL}\{q(\mathbf{u}_q) \parallel p(\mathbf{u}_q)\}
 \end{aligned} \tag{12}$$

where the approximate marginal posterior for $\mathbf{f}_{d,j}$ is $q(\mathbf{f}_{d,j}) = \int p(\mathbf{f}_{d,j} | \mathbf{u}) q(\mathbf{u}) d\mathbf{u}$ which have a close form giving by

$$q(\mathbf{f}_{d,j}) = \mathcal{N}\left(\mathbf{f}_{d,j} | \mathbf{K}_{\mathbf{u},(d,j)}^\top \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1} \boldsymbol{\mu}, K_{(d,j)} - \mathbf{K}_{\mathbf{u},(d,j)}^\top \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1} (\mathbf{S} - \mathbf{K}_{\mathbf{u},\mathbf{u}}) \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1} \mathbf{K}_{\mathbf{u},(d,j)}\right) \tag{13}$$

with $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top, \dots, \boldsymbol{\mu}_Q^\top]^\top$ and $\mathbf{S}$ denoting a block-diagonal matrix with blocks given by S_q , and $\text{KL}\{q(\mathbf{u}_q) \parallel p(\mathbf{u}_q)\}$ representing the Kullback-Leibler divergence between two Gaussian distributions, with a closed form solution. The learnable parameters in our model encompass the variational parameters $\{\mathbf{Z}_q, \boldsymbol{\mu}_q, S_q\}_{q=1}^Q$, and kernel parameters $\{\mathbf{B}_q, \Phi_q\}_{q=1}^Q$. The computational complexity of this model is now $\mathcal{O}(JNQ M^2)$.

Now, consider a test point set $\mathbf{X}_* \in \mathcal{X}$. Assuming a good approximation of the variational posterior $p(\mathbf{u} | \mathbf{y}) \approx q(\mathbf{u})$, the posterior of latent parameter function vector at test points $\mathbf{f}_*$ is

$$q(\mathbf{f}_*) = \int p(\mathbf{f}_* | \mathbf{u}) q(\mathbf{u}) d\mathbf{u}. \tag{14}$$

The predictive distribution for a new observation $\mathbf{y}_*$, given the observed data $\mathbf{y}$, can thus be approximated as:

$$p(\mathbf{y}_* | \mathbf{y}) \approx \int p(\mathbf{y}_* | \mathbf{f}_*) q(\mathbf{f}_*) d\mathbf{f}_*. \tag{15}$$

which integrates over the latent function vector $\mathbf{f}_*$ to account for its uncertainty in the prediction of $\mathbf{y}_*$.

The optimization procedure for ChdGP relies on the estimation of several groups of parameters. These include the kernel hyperparameters, such as characteristic lengthscales and the output scale associated with the exponential quadratic kernel, the likelihood parameters like data noise $\{\sigma_{Nd}^2\}_{d=1}^D$, and the variational distribution parameters $\{\mathbf{Z}_q, \boldsymbol{\mu}_q, S_q\}_{q=1}^Q$. According to [29], there is a strong dependency between the variational parameters and the other parameters, making the model extremely sensitive to any changes in these variables. Additionally, since the ELBO loss function is generally non-convex, stochastic gradient optimization methods often converge to poor local minima.

Recent studies have shown that optimization with Natural Gradient (NG) can effectively address the challenges described above. In particular, the work presented in [30] highlights the benefits of applying NG for variational GPs optimization, demonstrating that it can be implemented with little additional complexity and how to combine it with the Adam optimizer, providing a hybrid optimization framework. Following this approach, we adopt an optimization scheme in which the variational parameters $\{\mathbf{Z}_q, \boldsymbol{\mu}_q, S_q\}_{q=1}^Q$ are updated using Natural Gradient (NG), while the remaining model parameters are optimized with Adam; we denote this strategy as Adam + NG. This hybrid

method enables optimization performance comparable to that obtained when training an exact model using Adam alone.

2.4. Model Setup

The proposed methodology constructs the LMCGP covariance function, as defined in Equation (7), using the widely applied squared exponential kernel, shown in Equation (16). This kernel facilitates smooth data mapping:

$$k_q(\mathbf{x}, \mathbf{x}' | \Phi_q) = \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^\top \Phi_q^{-2}(\mathbf{x} - \mathbf{x}')\right) \quad (16)$$

where the diagonal matrix $\Phi_q = \text{diag}\{\Delta_{q_l}\}_l^{DT} \in \mathbb{R}^{DT \times DT}$ contains the length scale factors $\Delta_{q_l} \in \mathbb{R}^{++}$ from each input dimension. Upon examining Equation (16), one might question the absence of an output scale parameter. However, a detailed look at Equation (7) shows that the elements in the matrix $\mathbf{B}_q$ perform the function of rescaling the exponential term in the k_q kernel.

The performance of the models is evaluated using the Negative Log Predictive Density (NLPD), due to its flexibility in accommodating various likelihood functions, which is particularly useful when Gaussian assumptions are not appropriate. The NLPD metric is defined as:

$$\begin{aligned} \text{NLPD} &= -\frac{1}{N_*} \log p(\mathbf{y}_* | \mathbf{y}) \\ &= -\frac{1}{N_*} \sum_{n=1}^{N_*} \log p(\mathbf{y}_{n_*} | \mathbf{y}) \end{aligned} \quad (17)$$

3. Results and Discussions

3.1. Dataset Collection

The hydrological forecasting task used to validate the proposed model is based on time-series data of streamflow contributions from 23 reservoirs in Colombia. This dataset is particularly relevant within the context of sustainable water resource management, as streamflow dynamics reflect the availability, variability, and regulation of water resources that ultimately support multiple uses, including potable water supply, agriculture, and energy production. From a systemic perspective, reservoir inflows constitute a key indicator of water availability within the water–energy nexus, enabling the analysis of how hydroclimatic variability propagates through interconnected resource systems.

The streamflow contributions were recorded on a daily basis from 1 January 2010 to 28 February 2022, yielding a total of 4442 observations. Although these contributions originate from volumetric hydrological processes, they are reported in kilowatt-hours (kWh) by hydroelectric operators, as this unit directly represents the effective energy potential associated with water inflows under operational conditions. This representation does not alter the underlying hydrological information but instead provides a functional transformation that integrates hydraulic head, turbine efficiency, and system constraints. Consequently, modeling streamflow in kWh allows capturing the operational relevance of water availability, facilitating the analysis of how variability in hydrological inputs impacts resource allocation decisions. In the context of this research, this perspective is consistent with a broader sustainability framework, where efficient water use is evaluated not only in volumetric terms but also in terms of its contribution to essential services within interconnected systems. The spatial distribution of the reservoirs is shown in Figure 1, which provides geographical context for the forecasting problem. It is worth noting that the reservoirs are primarily located along mountainous regions to take advantage of streamflow generated by precipitation events.

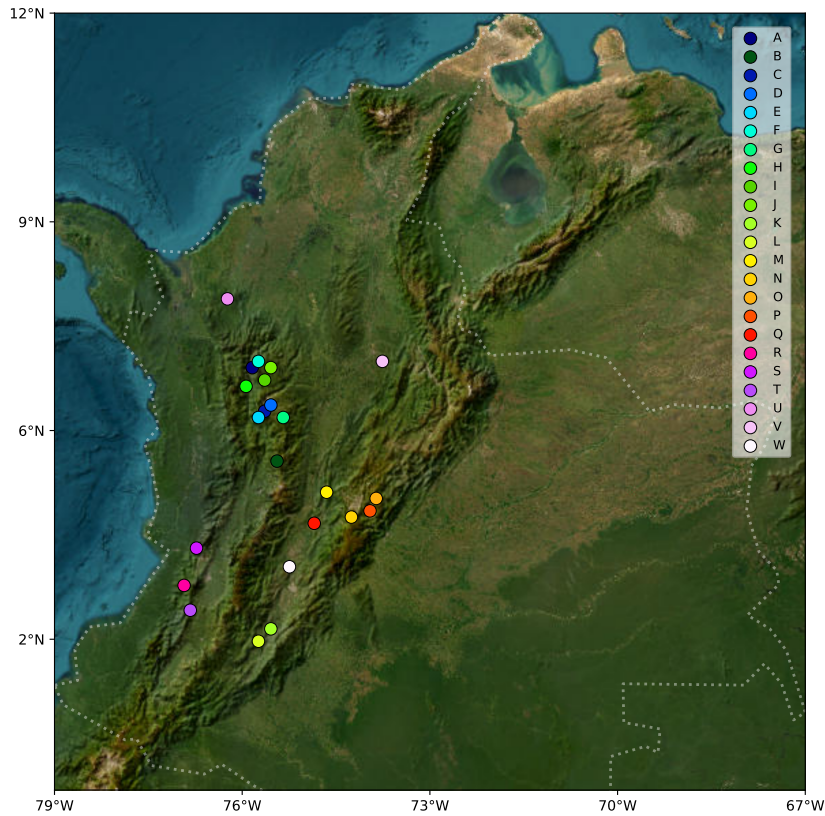


Figure 1. Reservoir locations in Colombia considered in the current study.

Figure 2 illustrates the distribution of streamflow contributions for each reservoir. First, a substantial variation in streamflow levels across reservoirs can be observed, which is mainly attributed to differences in their power generation capacities. Second, the majority of reservoirs exhibit non-Gaussian behavior, including platykurtic distributions (e.g., reservoir C), negative skewness (reservoir A), positive skewness (reservoir B), heavy-tailed behavior (reservoir F), and bimodal patterns (reservoir O). These heterogeneous distributions arise from factors such as irregular rainfall events, extended dry periods characteristic of Colombian climate conditions, and operational strategies aimed at either conserving water or maximizing energy production. In addition, reservoir Q, which consistently reports zero streamflow, provides a useful benchmark for evaluating forecasting performance in trivial scenarios. For model validation, the final two years of observations are used as the testing set, while the preceding ten years constitute the training set, resulting in $N = 3554$ training samples.

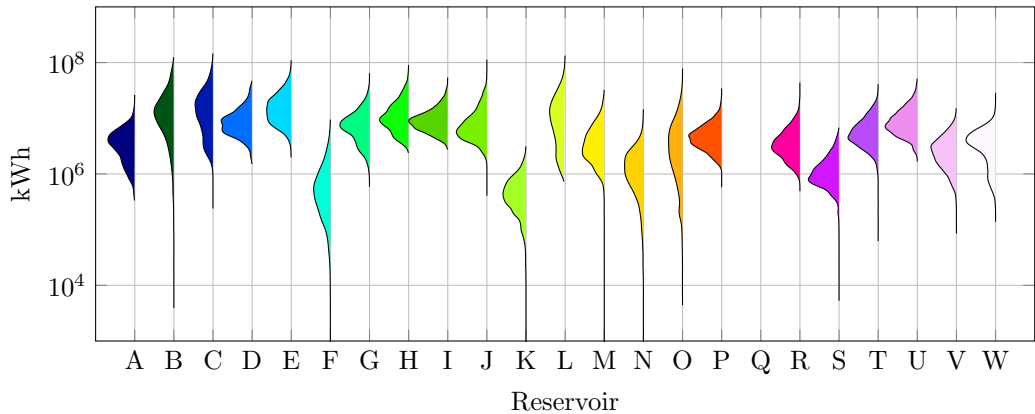


Figure 2. Violin plot depicting the distribution of streamflow time-series for each reservoir within the dataset.

3.2. Model Setup and Hyperparameter Tuning

The model setup takes into account that the time series cannot take negative values, meaning that a zero probability must be assigned to impossible scenarios. Fortunately, the Chained GP framework offers the possibility of using non-Gaussian likelihoods to model the data under the latent process, termed ChdGamma. Therefore, we propose using a Gamma likelihood function, which has positive support, as given by:

$$p(\mathbf{y} | \mathbf{f}) = \prod_{d=1}^D \prod_{n=1}^N \text{Gamma}(y_{d,n} | g_{d,1}(f_{d,1,n}), g_{d,2}(f_{d,2,n})) \quad (18)$$

where the deterministic functions $g_{d,1}(\cdot) = g_{d,2}(\cdot) = \ln(\exp(\cdot) + 1)$ represent the softplus function, ensuring positivity in the modeling of the shape and scale parameters for the Gamma likelihood. This approach allows the model to respect the inherent non-negativity constraints of the data more effectively, thereby providing a more accurate and realistic representation of the underlying processes.

As baseline models, the LMC and Chained GP with Gaussian likelihood (ChdNormal) are adopted. These models consider a Gaussian likelihood function given by:

$$p(\mathbf{y} | \mathbf{f}) = \prod_{d=1}^D \prod_{n=1}^N \mathcal{N}(y_{d,n} | g_{d,1}(f_{d,1,n}), g_{d,2}(f_{d,2,n})) \quad (19)$$

where the identity function $g_{d,1}(\cdot)$ is used to represent the mean parameter, while the softplus function for $g_{d,2}(\cdot) = \ln(\exp(\cdot) + 1)$ ensures the positivity of the variance parameter. The choice of the softplus function is based on empirical findings that it better prevents overflow compared to other functions, such as the exponential function. All models hereafter presented are optimized using the Adam + NG strategy. Prior to training, the time series are normalized to the interval $[0, 1]$ to handle their non-negative nature and improve numerical stability.

We tuned the hyperparameter Q for the chained GP models over a range from 1 to 46, evaluating the performance using the NLPD metric. It is important to note that other metrics mentioned earlier are no longer suitable due to the non-Gaussian assumption. The results of this search are shown in Figure 3 for both the Gaussian and Gamma likelihood models. Unlike the baseline Gaussian model (in Figure 3a), the Gamma likelihood one (in Figure 3b) turns more stable as Q increases, indicating better data learning without overfitting. Based on such observation, we selected $Q = 15$ and $Q = 26$ as the optimal number of independent GPs for the Gaussian and Gamma models, respectively.

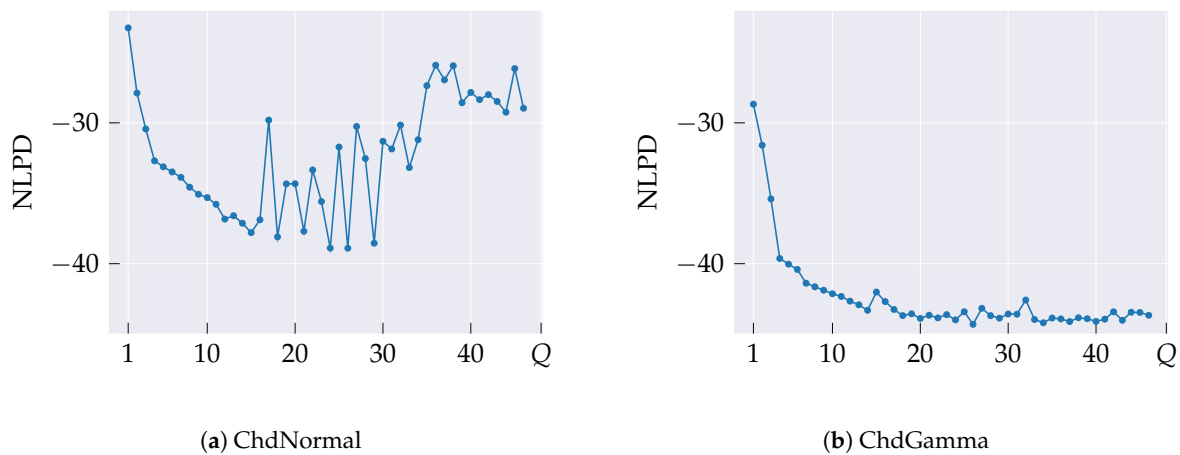


Figure 3. Performance metric for ChdGP Normal and Gamma models as a function of the number of independent GPs. The plot shows the NLPD metric value against the number of independent GPs used in the model

3.3. Performance Analysis

We compare the performance of the LMC, ChdNormal and ChdGamma models after retrained with the newly rescaled data. Figure 4 compares the Negative Log Predictive Density (NLPD) across various prediction horizons for the models. Firstly, the ChdNormal model consistently outperforms the LMC model across all scenarios, which can be attributed to its enhanced ability to model data noise using a Gaussian Process (GP). A particularly notable difference is observed at horizon $H = 7$, likely due to a week periodicity that the ChdNormal model can effectively exploit. In turn, the ChdGamma model consistently exhibits a lower NLPD across all horizons, indicating superior robustness in handling short-term and long-term predictions over LMC and ChdNormal. The Gamma likelihood, employed in the ChdGamma model, is particularly effective for non-negative and asymmetric data, typical in hydrological datasets, and avoids assigning predictive mass to impossible negative values, thereby improving accuracy and confidence.

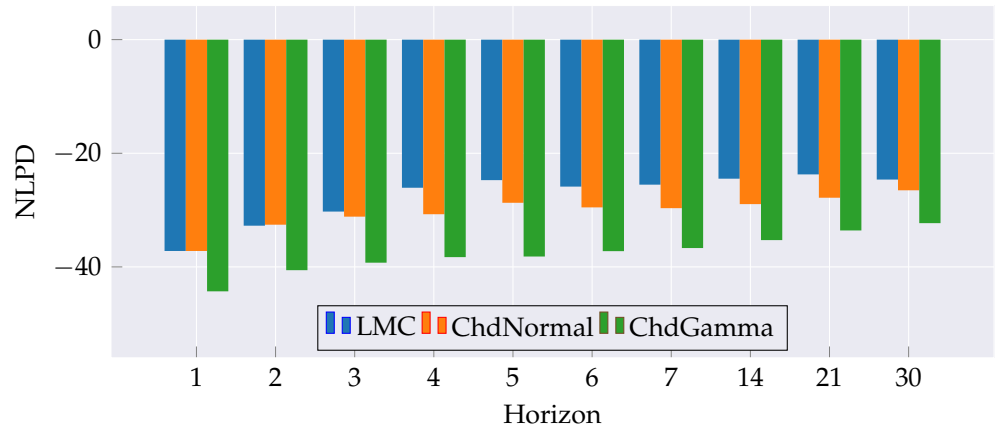


Figure 4. Comparison of NLPD metric across different prediction horizons for the three contrasted models on the testing set.

For a qualitative analysis, Figure 5 presents the forecasting results of both chained models on four selected reservoirs using a one-day-ahead horizon ($H = 1$). The confidence interval corresponds to the probability mass enclosed between the 2.5th and 97.5th percentiles, and the median of the predictive distribution is now also plotted. On the first two selected reservoirs (first two rows) and compared to the traditional Gaussian likelihood, ChdGamma yields narrower confidence intervals, indicating greater confidence in the model’s predictions. Additionally, in these plots, the mean and median predictions are very close, indicating a highly symmetric distribution concerning the mean due to the absence of extreme peaks. On the last two reservoirs, the time series exhibit abrupt peaks. Similar to the ChdNormal forecast, the Gamma likelihood model attempts to account for these peaks with a wider confidence interval. However, unlike the baseline, the ChdGamma setting does not allocate predictive distribution mass to negative values, which is more appropriate for non-negative time series. Consequently, the mean tends to be higher than the median in the presence of peaks, as the predictive distribution favors positive values while recovering symmetry and a narrower confidence interval during periods without peaks.

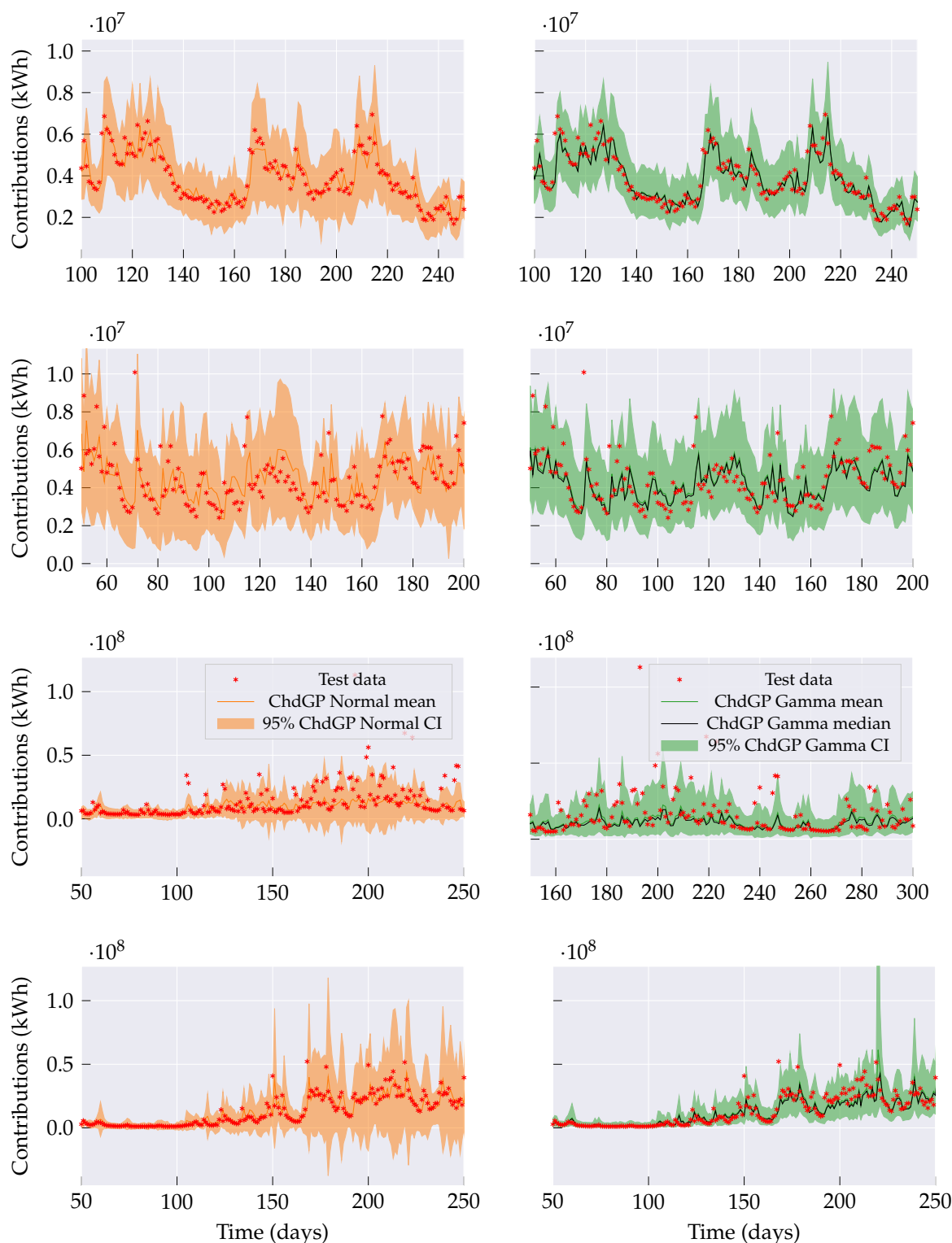


Figure 5. One-day-ahead forecasting on four selected reservoirs (top to bottom) using the ChdNormal (left) and the ChdGamma (right) models ($H = 1$). The shaded areas represent the 95% confidence interval ranging from the 2.5th to the 97.5th percentile of the model's predictions. The mean and median of the predictive distribution are depicted in colored and black continuous lines, respectively.

4. Concluding Remarks and Future Work

This work introduces the Chained Correlated Gaussian Process (ChdGP) model, which models likelihood parameters via GPs. This approach offers greater flexibility and enables the use of non-Gaussian likelihood functions to account for the natural constraints on the outputs, in this particular case, streamflow levels. The results of this study demonstrate that the proposed ChdGP with a Gamma

likelihood constitutes an advancement in the probabilistic modeling of hydrological time series, particularly in contexts characterized by high variability, nonlinearity, and structural constraints, such as non-negativity constraints. Beyond the improvements in predictive accuracy, the main contribution of this work lies in its ability to provide a coherent stochastic representation of water availability dynamics, which is essential for decision-making under uncertainty in water resource systems.

From a methodological perspective, the transition from deterministic, point-based forecasting to fully probabilistic frameworks represents a critical step toward addressing the inherent uncertainty of hydrological processes. The superior performance of the ChdGP model, especially when equipped with the Gamma likelihood, confirms that explicitly incorporating distributional constraints and heteroscedasticity yields more realistic and physically consistent predictions. This is particularly relevant in hydrological systems, where extreme events such as peak inflows and drought periods play a disproportionate role in system reliability and risk management. As shown in the experimental results, the proposed model dynamically adapts its predictive variance, widening confidence intervals during high-variability periods while maintaining sharp estimates under stable conditions, thereby providing meaningful uncertainty quantification. A key insight of this work is the relevance of modeling interdependencies across multiple hydrological series. Thanks to the multi-output structure in the Linear Model of Coregionalization, the proposed framework captures shared and task-specific patterns across reservoirs, improving forecasting performance across all horizons. This finding is particularly important in the Colombian context, where hydrological systems are spatially interconnected and influenced by common climatic drivers. Consequently, joint modeling not only enhances predictive accuracy but also provides a more systemic understanding of water dynamics, which is essential for integrated resource management.

From an application standpoint, the ChdGamma model's ability to generate full predictive distributions opens new opportunities for risk-aware decision-making. For instance, probabilistic forecasts can be directly integrated into optimization models for reservoir operation, early warning systems, and demand management strategies in urban water supply systems. This is particularly relevant in the context of increasing hydroclimatic variability, where traditional deterministic forecasts fail to capture the range of possible scenarios and may lead to suboptimal or vulnerable decisions. Although provided in energy units (kWh), the dataset fundamentally represents transformations of water inflows within the water–energy nexus. Therefore, the proposed modeling approach transcends its immediate application in hydropower forecasting and can be interpreted as a proxy for water availability dynamics. In this sense, the results are directly aligned with the broader objective of promoting sustainable water resource management, as established in the underlying research project. Reliable forecasting of hydrological inputs enables more efficient allocation of water across competing uses, including potable water supply systems, where variability in source availability directly affects service continuity, infrastructure operation, and long-term planning. As highlighted in the manuscript, reservoir inflows encapsulate the combined effects of climatic variability, watershed processes, and operational decisions, making them a valuable indicator of systemic pressures on water resources.

Future research should focus on extending the proposed framework toward integrated water management applications. In particular, coupling probabilistic hydrological forecasts with exogenous variables such as precipitation, water demand, and large-scale climate indices could further enhance predictive performance and interpretability, enabling the identification of consumption patterns and the design of adaptive allocation strategies under uncertainty. Additionally, integrating socio-environmental variables and alternative distributions or hierarchical likelihood structures for categorical data could improve the representation of extreme events and multimodal behaviors observed in some reservoirs, while allowing identification and classification of sustainable consumption profiles. Finally, deploying the proposed methodology within decision-support systems could facilitate its practical adoption by water utilities and policymakers, contributing to the development of resilient and sustainable water management strategies in Colombia and similar contexts.

Author Contributions: Conceptualization, Y.A.B.-P., J.D.P.-C. and D.A.C.-P.; methodology, J.D.P.-C.; validation, Y.A.B.-P. and D.A.C.-P.; formal analysis, J.G.-G.; resources, J.G.G.-E.; data curation, Y.A.B.-P. and J.G.-G.; writing—original draft preparation, J.D.P.-C.; writing—review and editing, D.A.C.-P. and J.G.G.-E.; supervision, J.G.G.-E. All authors have read and agreed to the published version of the manuscript.

Funding: Under grants provided by the research project 111089082356, funded by MINCIENCIAS.

Data Availability Statement: The data supporting the findings of this study are available on request from the authors. The data are not publicly available due to confidential agreements.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Hojjati-Najafabadi, A.; Mansoorianfar, M.; Liang, T.; Shahin, K.; Karimi-Maleh, H. A review on magnetic sensors for monitoring of hazardous pollutants in water resources. *Science of the Total Environment* **2022**, *824*, 153844. <https://doi.org/10.1016/j.scitotenv.2022.153844>.
- Lavers, D.; Ramos, M.H.; Magnusson, L.; Pechlivanidis, I.; Klein, B.; Prudhomme, C.; Arnal, L.; Crochemore, L.; Van Den Hurk, B.; Weerts, A.; et al. A vision for hydrological prediction. *Atmosphere* **2020**, *11*.
- Yang, D.; Yang, Y.; Xia, J. Hydrological cycle and water resources in a changing world: A review. *Geography and Sustainability* **2021**, *2*, 115–122. <https://doi.org/10.1016/j.geosus.2021.05.003>.
- Morales-Marín, L.A.; Rodríguez, E.A.; Jaramillo, F. Water resources challenges in Colombia: hydrological science solutions for a sustainable future. *Hydrological Sciences Journal* **2026**, *0*, 1–20, [\[https://doi.org/10.1080/02626667.2026.2631143\]](https://doi.org/10.1080/02626667.2026.2631143). <https://doi.org/10.1080/02626667.2026.2631143>.
- Yang, D.; Yang, Y.; Xia, J. Hydrological cycle and water resources in a changing world: A review. *Geography and Sustainability* **2021**, *2*, 115–122. <https://doi.org/10.1016/j.geosus.2021.05.003>.
- Lavers, D.A.; Ramos, M.H.; Magnusson, L.; Pechlivanidis, I.; Klein, B.; Prudhomme, C.; Arnal, L.; Crochemore, L.; Van den Hurk, B.; Weerts, A.H.; et al. A vision for hydrological prediction. *Atmosphere* **2020**, *11*, 237. <https://doi.org/10.3390/atmos11030237>.
- Yaseen, Z.; Allawi, M.; Yousif, A.; Jaafar, O.; Hamzah, F.; El-Shafie, A. Non-tuned machine learning approach for hydrological time series forecasting. *Neural Computing and Applications* **2018**, *30*, 1479–1491. <https://doi.org/10.1007/s00521-016-2763-0>.
- Kashani, M.; Ghorbani, M.; Dinpashoh, Y.; Shahmorad, S. Integration of Volterra model with artificial neural networks for rainfall-runoff simulation. *Journal of Hydrology* **2016**, *540*, 340–354. <https://doi.org/10.1016/j.jhydrol.2016.06.028>.
- Cheng, M.; Fang, F.; Kinouchi, T.; Navon, I.; Pain, C. Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology* **2020**, *590*, 125376. <https://doi.org/10.1016/j.jhydrol.2020.125376>.
- Sit, M.; Demiray, B.; Xiang, Z.; Ewing, G.; Sermet, Y.; Demir, I. A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology* **2020**, *82*, 2635–2670. <https://doi.org/10.2166/wst.2020.369>.
- Wang, X.; Wang, Y.; Yuan, P.; Wang, L.; Cheng, D. An adaptive daily runoff forecast model using VMD-LSTM-PSO hybrid approach. *Hydrological Sciences Journal* **2021**, *66*, 1488–1502. <https://doi.org/10.1080/02626667.2021.1937631>.
- Wang, Z.; Lou, Y. Hydrological time series forecast model based on wavelet de-noising and ARIMA-LSTM. In Proceedings of the IEEE ITNEC, 2019, pp. 1697–1701. <https://doi.org/10.1109/ITNEC.2019.8729441>.
- Sahoo, B.; Jha, R.; Singh, A.; Kumar, D. LSTM recurrent neural network for low-flow hydrological forecasting. *Acta Geophysica* **2019**, *67*, 1471–1481. <https://doi.org/10.1007/s11600-019-00330-1>.
- Li, Y.; Yang, J. Hydrological time series prediction model based on attention-LSTM neural network. In Proceedings of the Proceedings of the 2019 2nd International Conference on Machine Learning and Machine Intelligence (MLMI '19), New York, NY, USA, 2019; pp. 21–25. <https://doi.org/10.1145/3366750.3366756>.
- Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *arXiv* **2017**, [1612.01474].
- Gal, Y.; Ghahramani, Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the ICML, 2016, pp. 1050–1059.
- Quilty, J.; Adamowski, J. A stochastic wavelet-based framework for forecasting hydrological processes. *Environmental Modelling and Software* **2020**, *130*, 104718. <https://doi.org/10.1016/j.envsoft.2020.104718>.

18. Niu, W.; Feng, Z. Evaluating AI methods for streamflow forecasting. *Sustainable Cities and Society* **2021**, *64*, 102562. <https://doi.org/10.1016/j.scs.2020.102562>.
19. Sun, N.; Zhang, S.; Peng, T.; Zhang, N.; Zhou, J.; Zhang, H. Random Forest and Gaussian Process Regression for streamflow forecasting. *Water* **2022**, *14*, 1828. <https://doi.org/10.3390/w14111828>.
20. Yang, J.; Jakeman, A.; Fang, G.; Chen, X. Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian process emulator. *Environmental Modelling and Software* **2018**, *101*, 289–300. <https://doi.org/10.1016/j.envsoft.2017.11.037>.
21. Martínez-Acosta, L.; Medrano-Barboza, J.; López-Ramos, Á.; Remolina López, J.; López-Lambrano, Á. SARIMA approach to generating synthetic monthly rainfall in the Sinú river watershed in Colombia. *Atmosphere* **2020**, *11*, 602. <https://doi.org/10.3390/atmos11060602>.
22. Pastrana-Cortés, J.D.; Gil-Gonzalez, J.; Álvarez Meza, A.M.; Cárdenas-Peña, D.A.; Orozco-Gutiérrez, A.A. Scalable and Interpretable Forecasting of Hydrological Time Series Based on Variational Gaussian Processes. *Water* **2024**, *16*. <https://doi.org/10.3390/w16142006>.
23. Perrin, C.; Michel, C.; Andréassian, V. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology* **2001**, *242*, 275–301. [https://doi.org/10.1016/S0022-1694\(00\)00393-0](https://doi.org/10.1016/S0022-1694(00)00393-0).
24. Hensman, J.; Matthews, A.; Ghahramani, Z. Scalable Variational Gaussian Process Classification. In Proceedings of the Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS), San Diego, CA, USA, 2015; Vol. 38, *Proceedings of Machine Learning Research*, pp. 351–360.
25. Bruinsma, W.; Perin, E.; Tebbutt, W.; Hosking, S.; Solin, A.; Turner, R. Scalable Exact Inference in Multi-Output Gaussian Processes. In Proceedings of the Proceedings of the 37th International Conference on Machine Learning (ICML), 2020, Vol. 119, *Proceedings of Machine Learning Research*, pp. 1190–1201.
26. Saul, A.; Hensman, J.; Vehtari, A.; Lawrence, N. Chained Gaussian Processes. In Proceedings of the Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS), Cadiz, Spain, 2016; Vol. 51, *Proceedings of Machine Learning Research*, pp. 1431–1440.
27. Moreno-Muñoz, P.; Artés, A.; Álvarez, M. Heterogeneous Multi-output Gaussian Process Prediction. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Red Hook, NY, USA, 2018; Vol. 31, pp. 6711–6720.
28. Álvarez, M.; Rosasco, L.; Lawrence, N. *Kernels for Vector-Valued Functions: A Review*; Now Publishers Inc.: New York, NY, USA, 2012.
29. Giraldo, J.J.; Álvarez, M. A fully natural gradient scheme for improving inference of the heterogeneous multi-output Gaussian process model. *IEEE Transactions on Neural Networks and Learning Systems* **2021**, *33*, 6429–6442. <https://doi.org/10.1109/TNNLS.2021.3080238>.
30. Salimbeni, H.; Eleftheriadis, S.; Hensman, J. Natural Gradients in Practice: Non-conjugate Variational Inference in Gaussian Process Models. In Proceedings of the Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS), Playa Blanca, Lanzarote, Spain, 2018; Vol. 84, *Proceedings of Machine Learning Research*, pp. 689–697.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.