Chained Linear Model of Coregionalization Gaussian Process for Hydrological Forecasting

J. D. Pastrana-Cortés, J. Gil-González, D. Cárdenas-Peña

Universidad Tecnológica de Pereira

August 28, 2025

The Importance of Hydrological Forecasting

Understanding hydrological processes has become increasingly critical in natural resource management. **Suitable forecasting** allows the anticipation capacity of extreme hydrological events such as water shortages.

Challenges Non-linearities, high stochasticity, and complex water resource patterns.

Problem Setting

Consider a vector collecting sequential data across P variables at time instant t, as $\boldsymbol{v}_t \subseteq \mathbb{R}^P$. The flattening of T sequential observations yields an input vector $\boldsymbol{x} \in \mathcal{X}$,

$$\boldsymbol{x} = [\boldsymbol{v}_{t-1}^\top, \boldsymbol{v}_{t-2}^\top, \cdots, \boldsymbol{v}_{t-T}^\top]^\top,$$

$\mathcal{X} \subseteq \mathbb{R}^L$, and $L = PT$. The forecasting task aims to predict the next H sequential values, yielding the output target $\boldsymbol{y} \in \mathcal{Y}$

$$\boldsymbol{y} = [\boldsymbol{v}_t^\top, \boldsymbol{v}_{t+1}^\top, \cdots, \boldsymbol{v}_{t+H-1}^\top]^\top,$$

$\mathcal{Y} \subseteq \mathbb{R}^D$, and $D = PH$. Gathering N input-output i.i.d. observation pairs produces the training dataset as $\mathcal{D} = \{\boldsymbol{x}_n, \boldsymbol{y}_n\}_{n=1}^N = \{\boldsymbol{X}, \boldsymbol{Y}\}$.

The Chained Model

For dataset $\mathcal{D}$, the likelihood function assumes the distribution over $\boldsymbol{Y}$ as the product of D conditionally independent distributions, one by each output as follows:

$$p(\boldsymbol{Y} \mid \boldsymbol{\theta}(\boldsymbol{X})) = \prod_{n=1}^{N} \prod_{d=1}^{D} p\left(y_{n,d} \mid \boldsymbol{\theta}_d(\boldsymbol{x}_n)\right),$$

where $\boldsymbol{\theta}(\boldsymbol{X}) = \{\boldsymbol{\theta}_d(\boldsymbol{x}_n)\}_{n=1,d=1}^{N,D}$, with $\boldsymbol{\theta}_d(\boldsymbol{x}) \subseteq \mathbb{R}^{J_d}$ as a vector containing J_d parameters for d-th output distribution with elements $\theta_{d,j}(\boldsymbol{x})$. **Each likelihood parameter is governed by a Gaussian Process (GP)** $f_{d,j}(\boldsymbol{x})$ via a link function transformation $h_{d,j}(\cdot)$ as $\theta_{d,j}(\boldsymbol{x}) = h_{d,j}(f_{d,j}(\boldsymbol{x}))$.

The LMC Model

Consider a set of Q zero-mean independent GPs $\{g_q\}_{q=1}^{Q}$ with kernel function $k_q(\boldsymbol{x}, \boldsymbol{x}')$ that will be linearly weighed via $a_{(d,j),q} \in \mathbb{R}$ coefficients to generate $f_{d,j}$ as

$$f_{d,j}(\boldsymbol{x}) = \sum_{q=1}^{Q} a_{(d,j),q} g_q(\boldsymbol{x}),$$

proposing a **cross-covariance function for the latent variables $f_{d,j}$** as follows

$$k_{\boldsymbol{f}_{d,j},\boldsymbol{f}_{d',j'}}(\boldsymbol{x}, \boldsymbol{x}') = \operatorname{cov}\{f_{d,j}(\boldsymbol{x}), f_{d',j'}(\boldsymbol{x}')\} = \sum_{q=1}^{Q} a_{(d,j),q} a_{(d',j'),q} k_q(\boldsymbol{x}, \boldsymbol{x}').$$

We call the above as Linear Model of Coregionalization (LMC).

Variational Optimization

We introduce induced vectors $\mathbf{u} \in \mathbb{R}^{MQ}$ with elements $g_q(\mathbf{z}_m)$ at $M \ll N$ inducing points $\{\mathbf{z}_m\}_{m=1}^{M}$ to reduce the complexity $\mathcal{O}(N^3)$ **to** $\mathcal{O}(NM^2)$, and approximating the joint posterior

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

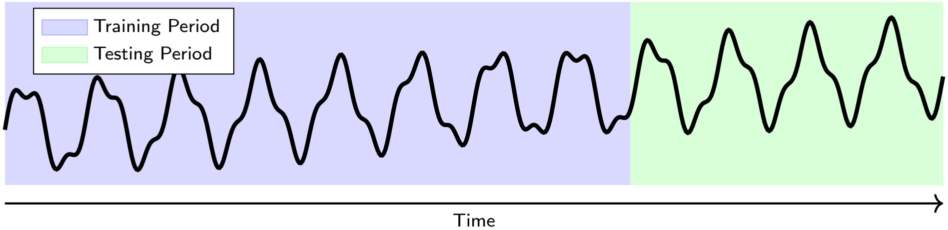
with $q(\mathbf{u}_q) = \mathcal{N}(\mathbf{u}_q \mid \boldsymbol{\mu}_q, \mathbf{S}_q)$, $\boldsymbol{\mu}_q \in \mathbb{R}^M$, $\boldsymbol{S}_q \in \mathbb{R}^{M\times M}$, and $\boldsymbol{f}_{d,j} = [f_{d,j}(\boldsymbol{x}_1), \cdots, f_{d,j}(\boldsymbol{x}_N)]^\top \in \mathbb{R}^N$. We tune the model by optimizing the loss function $\mathcal{L}$

$$\mathcal{L} = \sum_{n=1}^{N} \sum_{d=1}^{D} \mathbb{E}_{q(\mathbf{f}_{d,n})}\{\log p(y_{d,n} \mid f_{d,1}(\mathbf{x}_n), \cdots, f_{d,J_d}(\mathbf{x}_n))\} - \sum_{q=1}^{Q} \mathrm{KL}\{q(\mathbf{u}_q) \parallel p(\mathbf{u}_q)\},$$

using Natural Gradient for $\mathbf{u}_q$ and $\boldsymbol{S}_q$, and Adam for the remaining parameters [1].

Dataset

We use daily time series of streamflow contributions recorded from 23 Colombian reservoirs. The dataset contains 4,442 i.i.d. samples, with $N = 3,554$ used for training and $N_* = 888$ for testing.



The testing dataset $\mathcal{D}_* = \{\boldsymbol{x}_{*n}, \boldsymbol{y}_{*n}\}_{n=1}^{N_*}$, where $\boldsymbol{y}_{*n} = [y_{*n,1}, \cdots, y_{*n,D}]^\top$, collects all tasks observations at every test input $y_{*n,d}$.

Model Setup

The forecasting methodology constructs the covariance function using the squared exponential kernel, facilitating smooth data mapping.

$$k_q(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^\top \boldsymbol{\Phi}_q^{-2}(\boldsymbol{x} - \boldsymbol{x}')\right)$$

$\boldsymbol{\Phi}_q = \mathrm{diag}\{\Delta_{ql}\}_{l=1}^{L} \in \mathbb{R}^{++L \times L}$ contains the length-scales. Regarding the likelihood,

$$p(\boldsymbol{Y} \mid \theta(\boldsymbol{X})) = \prod_{n=1}^{N}\prod_{d=1}^{D} \mathcal{N}\left(y_{nd} \mid h_{d,1}(f_{d,1}(\mathbf{x}_n)), h_{d,2}(f_{d,2}(\mathbf{x}_n))\right),$$

$h_{d,1}(\cdot)$ is the identity function. The baseline model is the standard LMC GP with $h_{d,2}(\cdot)$ as a constant, while the proposed model is the Chained LMC GP with $h_{d,2}(\cdot) = \ln(\exp(\cdot) + 1)$.

Performance Metrics

Given the predictive mean $\hat{\mu}_{n,d}$ and variance $\hat{\sigma}^2_{n,d}$, the Mean Squared Error (MSE) takes the following form

$$\text{MSE} = \frac{1}{N_* D} \sum_{n=1}^{N_*} \sum_{d=1}^{D} \left(y_{*n,d} - \hat{\mu}_{n,d}\right)^2 .$$

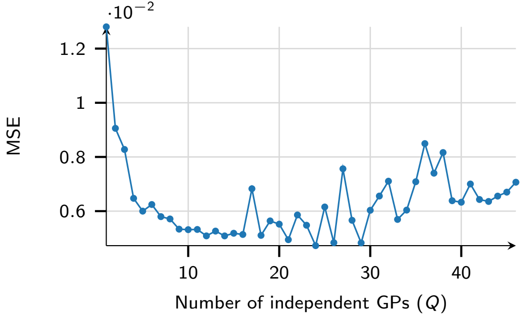
The Mean Standardized Log Loss (MSLL) **assesses probabilistic prediction quality** [2]

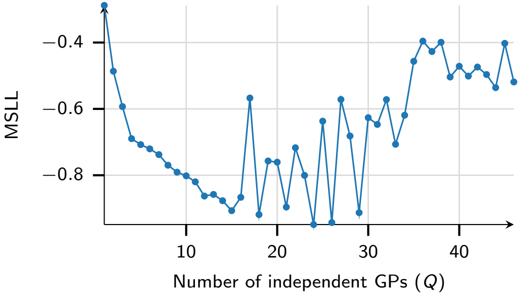
$$\text{MSLL} = \frac{1}{2N_* D} \sum_{n=1}^{N_*} \sum_{d=1}^{D} \frac{\left(y_{*n,d} - \hat{\mu}_{n,d}\right)^2}{\hat{\sigma}^2_{n,d}} - \frac{\left(y_{*n,d} - \mu_d\right)^2}{\sigma^2_d} + \ln\left(\frac{\hat{\sigma}^2_{n,d}}{\sigma^2_d}\right)$$

where $\mu_d = \frac{1}{N_*} \sum_{n=1}^{N_*} y_{*n,d}$ and $\sigma^2_d = \frac{1}{N_* - 1} \sum_{n=1}^{N_*} (\mu_d - y_{*n,d})^2$.

Hyperparameter Tuning

A grid search on the baseline model fixed $T = 1$ and $M = 64$. We varied $Q \in \{1, \cdots, 46\}$.

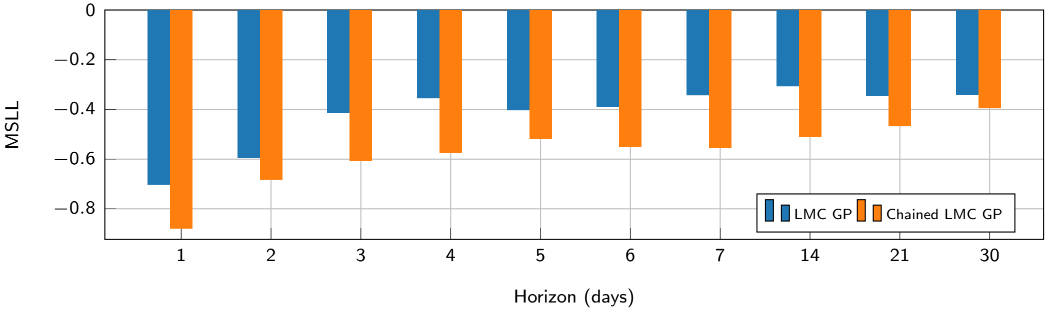




The MSE and MSLL decrease until $Q = 15$. Beyond this threshold, the trend becomes unstable, likely due to the increased model complexity.

Performance Analysis

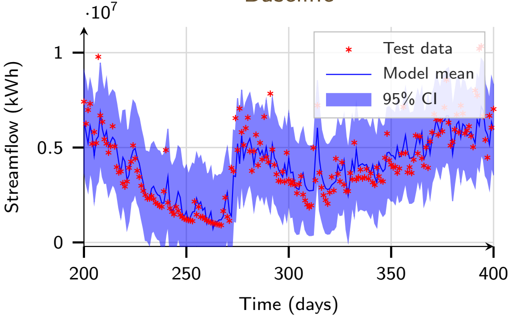
The proposed model outperforms the baseline across all evaluated scenarios



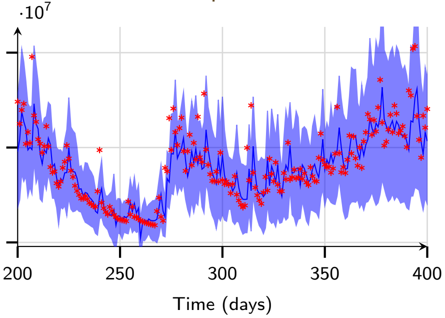
This improvement is due to the latent GP component in the proposed model, which dynamically explains the noise in the data.

One-day-ahead forecasting

Baseline



Proposed



The confidence interval remains practically constant over time, even when the data changes slowly.

The model explains the peaks with higher uncertainty while maintaining a narrow confidence interval when the data changes slowly.

Conclusions

- The Chained LMC GP incorporates **latent Gaussian processes to model likelihood parameters**, enabling more flexible and adaptive uncertainty quantification.
- We evaluated the approach on **daily streamflow time series from 23 Colombian reservoirs**, framing each reservoir as a related task in a multi-output forecasting scenario.
- Across all forecasting horizons, the Chained LMC GP achieved **significant improvements over the standard LMC GP**.
- Its main advantage is its ability to **model data noise with a GP**, which better explains peaks and smooth trends."

Future Work

In future work, we plan to extend this research in three ways.

- Validate the Chained LMC GP with **non-Gaussian likelihood functions** for modeling amplitude-constrained time series.
- Enhance the LMC GP by introducing **non-instantaneous mixing**, similar to a process convolution, which will generalize the Chained model to better capture temporal dependencies.
- Incorporate a **deep learning approach**, such as deep Gaussian Processes, to capture more complex patterns and improve prediction accuracy.

Acknowledgment

- Thanks to the *Doctorate in Engineering*, program from the Universidad Tecnológica de Pereira, for their support.
- This work was carried out under grants of the project 1110-890-82356, funded by MINCIENCIAS.
- Acknowledgment to the Automatics Research Group from the Universidad Tecnológica de Pereira for their invaluable support in the development of the research.

Hugh Salimbeni, Stefanos Eleftheriadis, and James Hensman.
Natural gradients in practice: Non-conjugate variational inference in gaussian process models.
In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 689–697. PMLR, 09–11 Apr 2018.

Carl Edward Rasmussen and Christopher K. I. Williams.
Gaussian processes for machine learning.
Adaptive computation and machine learning. MIT Press, 2006.