

Stochastic Modeling of Multiple Streamflow Time Series in Colombian Based on Gaussian Processes

Author: Julián David Pastrana-Cortés

Director: Álvaro Angel Orozco-Gutiérrez

Co-director: David Augusto Cardenas-Peña

Automatic Research Group
September 26, 2024



Motivation

Understanding the implications of time series associated with hydrological variables, such as flow rates or reservoir levels, is essential for hydroelectric generation and the planning of other generation systems in Colombia.



(a) Irrigation



(b) Flood control



(c) Hydropower generation

Challenges

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

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 droughts and heavy rainfall.



(a) Drought Condition



(b) Full Dam

Problem Statement and Research Question

	AR	NN	RNN	LSTM	GP
Interpretability [7, 8]	✓	✗	✗	✗	✓
Nonlinearity [3]	✗	✓	✓	✓	✓
Stochasticity	✓	✗	✗	✗	✓
Long-term [4, 5, 6]	✗	✗	✗	✓	✓
Scalability [9, 10, 11]	✓	✓	✓	✓	✗
Output Constraint	✗	✓	✓	✓	✗

Research Question

How to develop a joint probabilistic prediction model for multiple hydrological series associated with electricity generation, that describes the randomness of the forecast, is scalable, utilizes task correlations to improve performance, and incorporates output constraints to ensure feasible predictions?

Objectives

General Objective

Develop a stochastic forecasting model for making multiple simultaneous predictions of hydrological time series. This model will take advantage of cross-correlations among the outputs to improve performance, while maintaining scalability and output constraints.

Specific Objectives

- Develop a model that allows the forecasting of hydrological time series, properly quantifying the **uncertainty** associated with each value within the prediction horizons.
- Design a **multi-output** forecasting methodology that captures and models cross-correlations between hydrological time series, to improve forecast accuracy within forecast horizons.
- Develop a multi-output prediction methodology that handles **data constraints** across reservoirs while maintaining high forecasting performance as measured by probabilistic metrics.

Problem Setting

We model time series using observed vectors. At each time step n , the vector $\mathbf{v}_n \in \mathbb{R}^D$ represents hydrological resources across D outputs. The input vector \mathbf{x}_n for the model is constructed from time n back to $n - T + 1$:

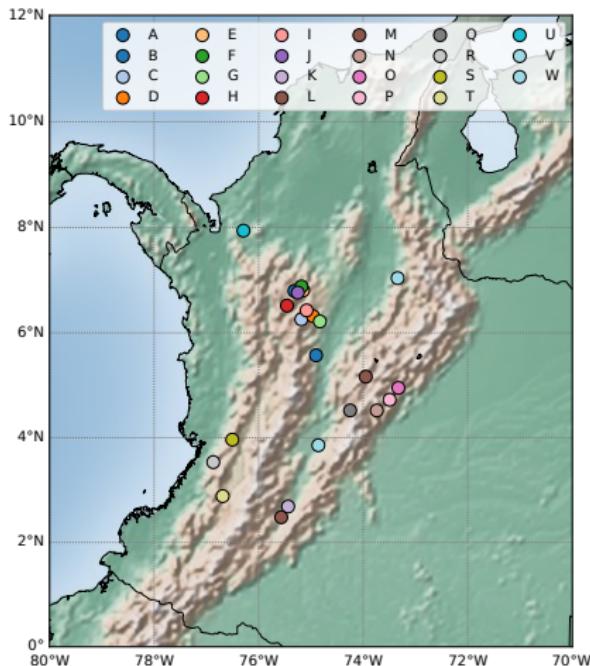
$$\mathbf{x}_n = \begin{bmatrix} \mathbf{v}_n \\ \mathbf{v}_{n-1} \\ \vdots \\ \mathbf{v}_{n-T+1} \end{bmatrix} \in \mathcal{X},$$

where T is the model order, and $\mathcal{X} \subset \mathbb{R}^{DT}$ represents the input space. The target output vector \mathbf{y}_n is given as:

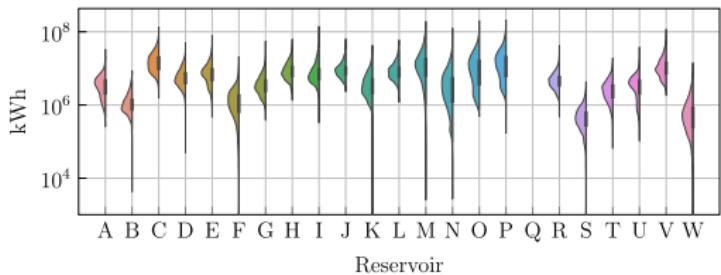
$$\mathbf{y}_n = \mathbf{v}_{n+H} \in \mathbb{R}^D,$$

where H is the prediction horizon. We build a dataset $\mathcal{D} = \{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^N = \{\mathbf{X}, \mathbf{y}\}$, comprising N input-output pairs.

Reservoir Locations and Dataset Overview



The hydrological forecasting task utilizes daily streamflow data from $D = 23$ Colombian reservoirs from January 1, 2010, to February 28, 2022.



Although volumetric measurements are recorded, they are reported in kilowatt-hours (kWh) by the hydroelectric power plants.

Methodology

Performance Metrics

- Mean Squared Error (MSE)
- Mean Standardized Log Loss (MSLL)
- Continuous Ranked Probability Score (CRPS)
- Negative Log Predictive Density (NLPD)

Gaussian Process Models

- ① Start with a Sparse Variational GPs for scalable and stochastic regression.
- ② Extend to Multi-Output GPs, capturing dependencies across multiple outputs.
- ③ Introduce Chained Correlated GPs to handle non-Gaussian likelihoods.

Gaussian Process Regression: Bayesian Non-Parametric Model

Gaussian Process (GP) Framework

In a GP framework, the function $f(\cdot)$ maps inputs x_n to outputs y_n . Adding i.i.d. Gaussian noise ϵ , the model becomes:

$$y_n = f(x_n) + \epsilon$$

For test inputs X_* , the joint distribution of training outputs y and test outputs f_* is:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_y & \mathbf{K}_* \\ \mathbf{K}_*^\top & \mathbf{K}_{**} \end{bmatrix}\right)$$

The posterior distribution for test points is:

$$f_* | X_*, \mathcal{D} \sim \mathcal{N}(\mathbf{K}_*^\top \mathbf{K}_y^{-1} y, \mathbf{K}_{**} - \mathbf{K}_*^\top \mathbf{K}_y^{-1} \mathbf{K}_*)$$

- $\mathbf{K}_y = \mathbf{K} + \Sigma_\epsilon$, where $\mathbf{K} \in \mathbb{R}^{ND \times ND}$ is the covariance matrix for the train set and Σ_ϵ contains task-wise noise.
- $\mathbf{K}_{**} \in \mathbb{R}^{N_* D \times N_* D}$ is the covariance matrix for the test set.
- $\mathbf{K}_* \in \mathbb{R}^{ND \times N_* D}$ represents the cross-covariance matrix between the training and test points.

Variational Inference for Scalability

The prediction performance is influenced by the selected parameter set θ and the matrix Σ_ϵ . These parameters are determined by maximizing the marginal log-likelihood:

$$\{\theta_{\text{opt}}, \Sigma_{\epsilon \text{opt}}\} = \arg \max_{\theta, \Sigma_\epsilon} -\frac{1}{2}\mathbf{y}^\top \mathbf{K}_y^{-1} \mathbf{y} - \frac{1}{2} \ln |\mathbf{K}_y| - \frac{ND}{2} \ln 2\pi,$$

with complexity $\mathcal{O}(N^3D^3)$ due to the need to invert the matrix \mathbf{K}_y . For scalability, we introduce $M \ll N$ inducing points Z with inducing variables $q(\mathbf{u}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{S})$, providing the following ELBO:

$$\mathcal{L} = \sum_{d=1}^D \sum_{n=1}^N \mathbb{E}_{q(f_d(\mathbf{x}_n))} \{ \ln p(y_{dn} \mid f_d(\mathbf{x}_n)) \} - \sum_{d=1}^D \text{KL}\{q(\mathbf{u}_d) \parallel p(\mathbf{u}_d)\},$$

where $f_d(\mathbf{x}_n)$ represents the d -th latent function value at input \mathbf{x}_n , and y_{dn} is the corresponding observed value. This reduces the complexity to $\mathcal{O}(NM^2D^3)$.

Model Setup

The GP covariance is factorized into two kernels: $k_{\mathcal{X}}$ for input correlations and k_D for task correlations:

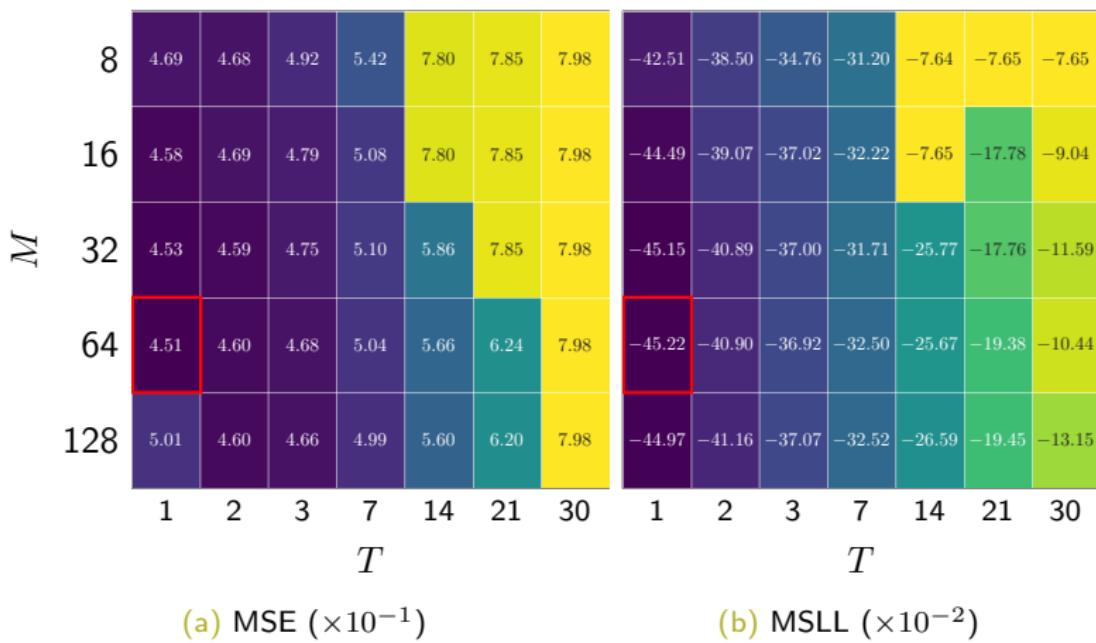
$$k((\mathbf{x}, d), (\mathbf{x}', d')) = k_{\mathcal{X}}(\mathbf{x}, \mathbf{x}' | \Theta_d) k_D(d, d' | \sigma_d),$$

with:

$$k_{\mathcal{X}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^\top \Theta_d^{-2}(\mathbf{x} - \mathbf{x}')\right),$$
$$k_D(d, d') = \sigma_d^2 \delta_{d,d'},$$

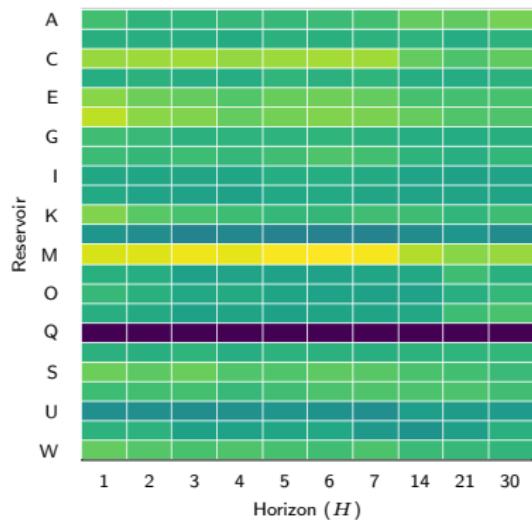
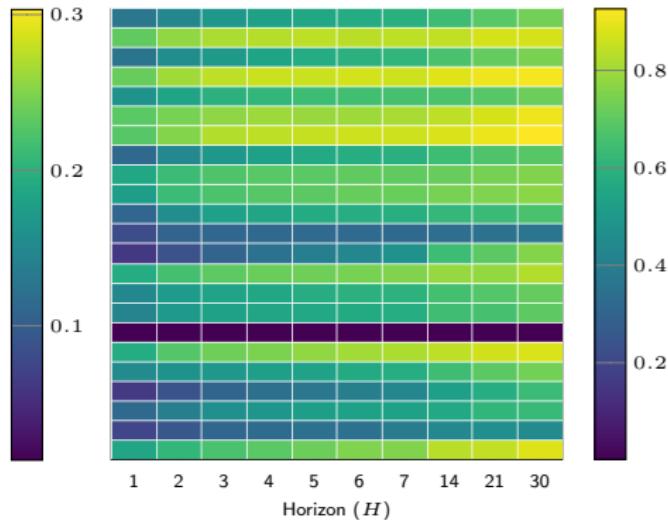
where $\delta_{d,d'}$ is the Kronecker delta, $\Theta_d = \text{diag}\{\Delta_{dl}\}_{l=1}^{DT}$ is the lengthscale matrix, and σ_d^2 is the output scale. This reduces complexity to $\mathcal{O}(NM^2D)$ by avoiding explicit task correlations.

Grid search average values for tuning T and M



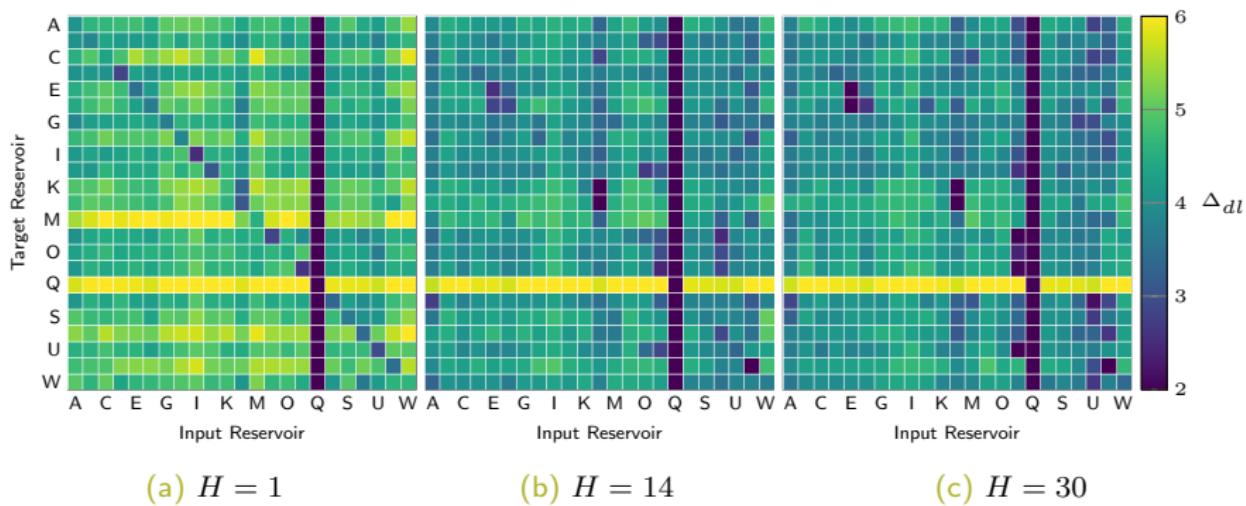
The error increases with larger values of T . Smaller values of M are more sensitive to initialization, while larger values of M tend to overfit. The optimal parameters are $M = 64$ and $T = 1$.

Reservoir-Wise Output Scales and Noise Variance

(a) Output scales σ_d^2 (b) Noise variances Σ_ϵ

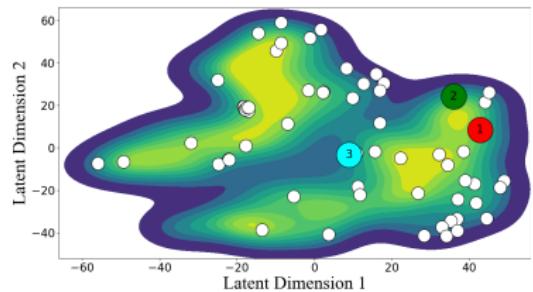
Longer horizons typically exhibit smaller output scales and higher noise variances, indicating weaker correlations and more complex tasks.

Reservoir-Wise Lengthscales

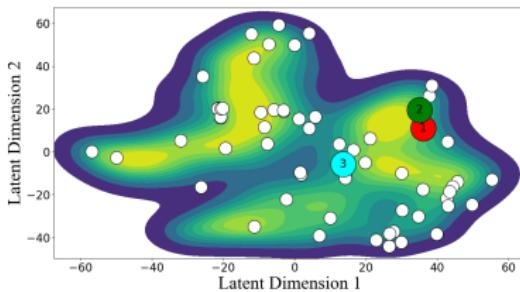


As the horizon increases, main diagonal lengthscales lose relevance, while off-diagonal ones gain importance.

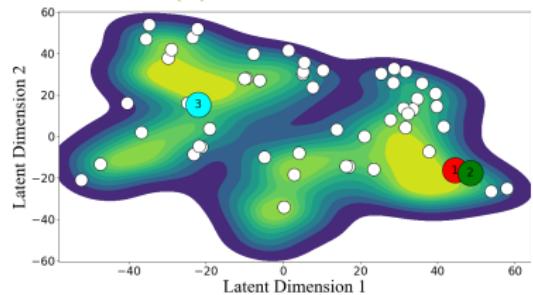
SVGP t-SNE Latent Mapping Space



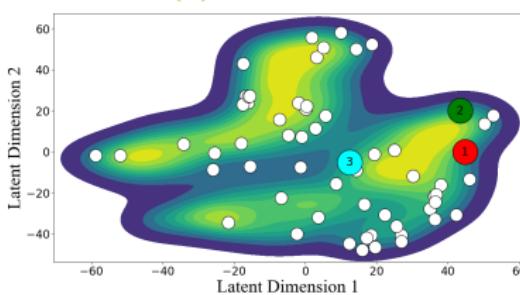
(a) Reservoir E



(b) Reservoir I



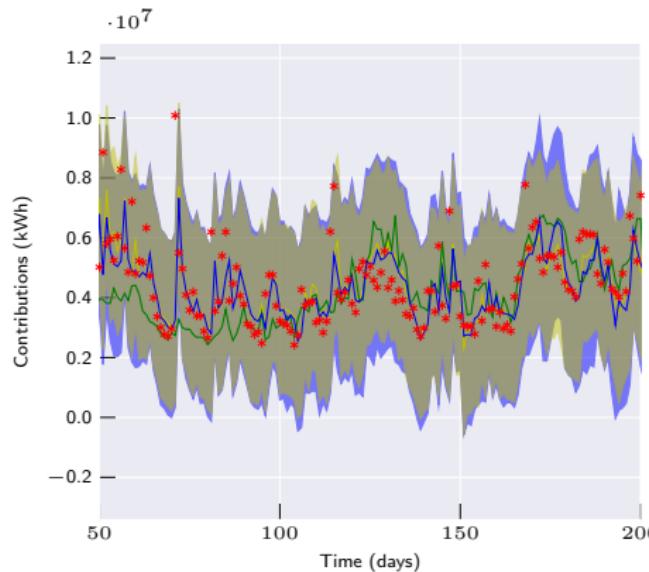
(c) Reservoir L



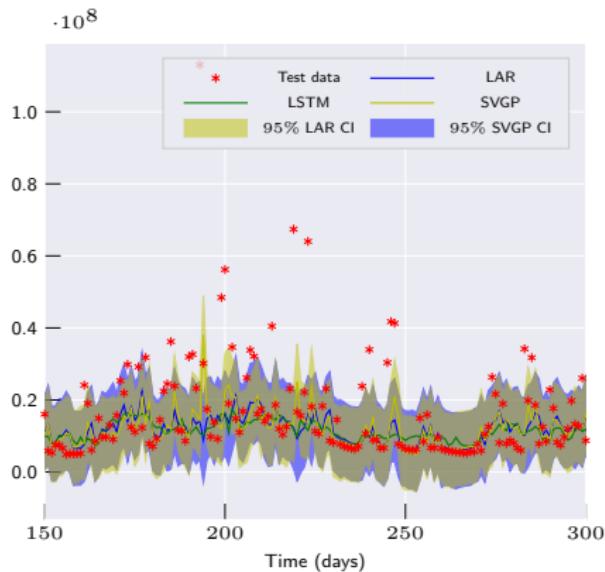
(d) Reservoir U

The shared inducing points allow for the capturing of task-wise, and global information about the streamflow dynamics.

One-day-ahead Models Forecasting



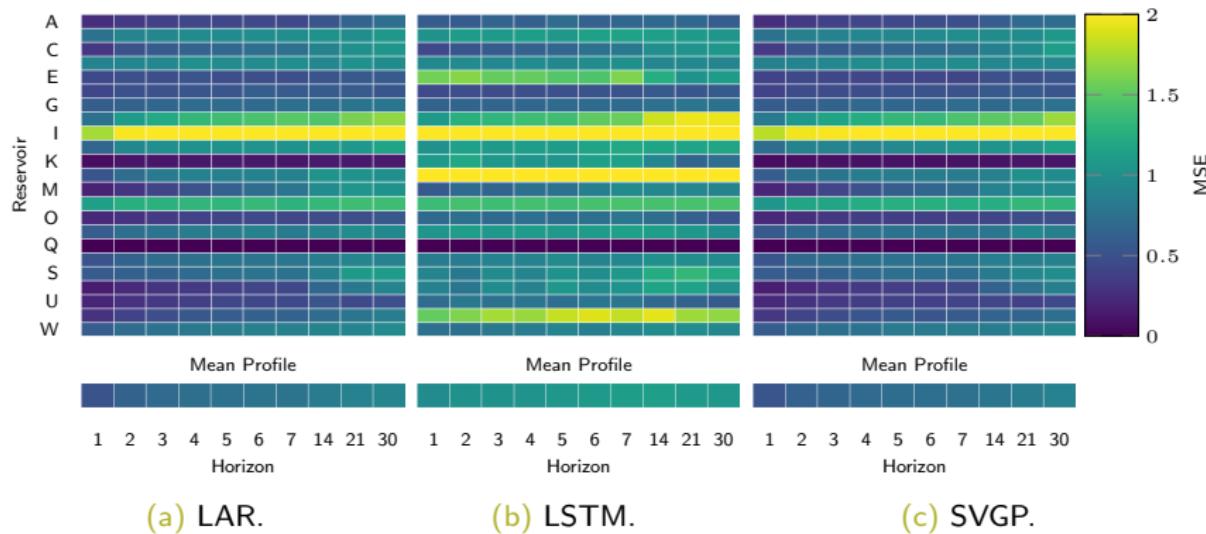
(a) Reservoir A



(b) Reservoir I

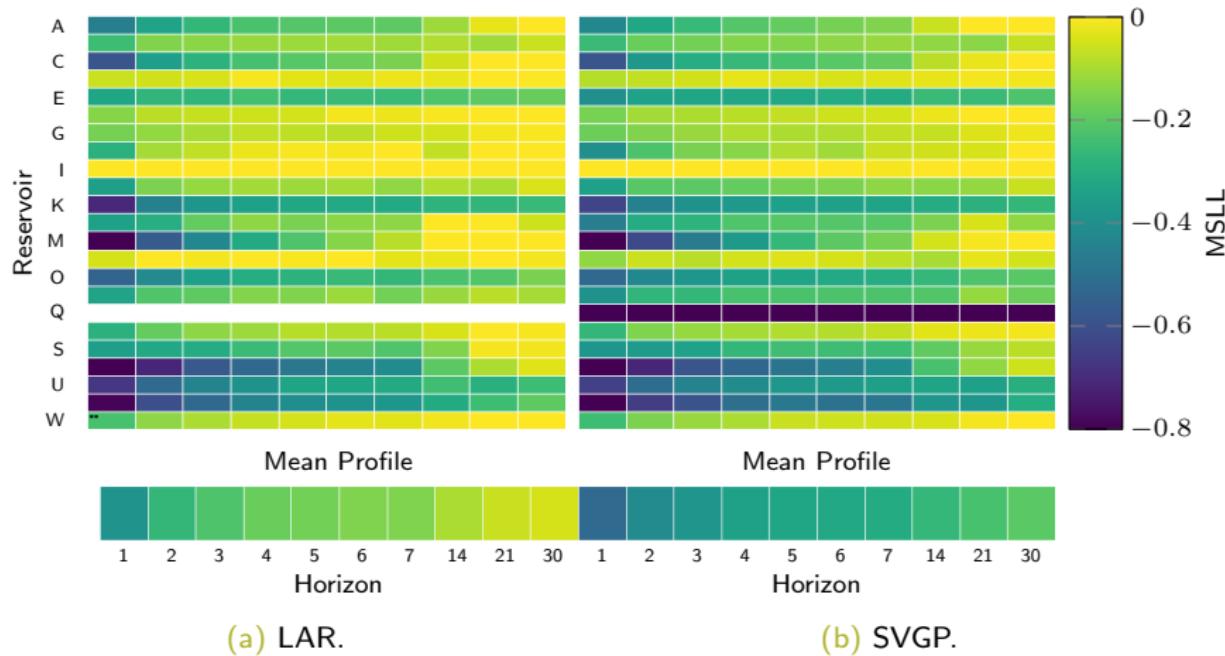
The SVGP close follows to time-series data and captures its stochastic nature through the predictive distribution.

MSE achieved by LAR, LSTM, and SVGP



Longer horizons yield larger errors. The LAR and SVGP models significantly outperform the LSTM models across all scenarios.

MSLL achieved by LAR, and SVGP



The SVGP models exhibits lower error and a slower error increase with the horizon compared to the LAR models.

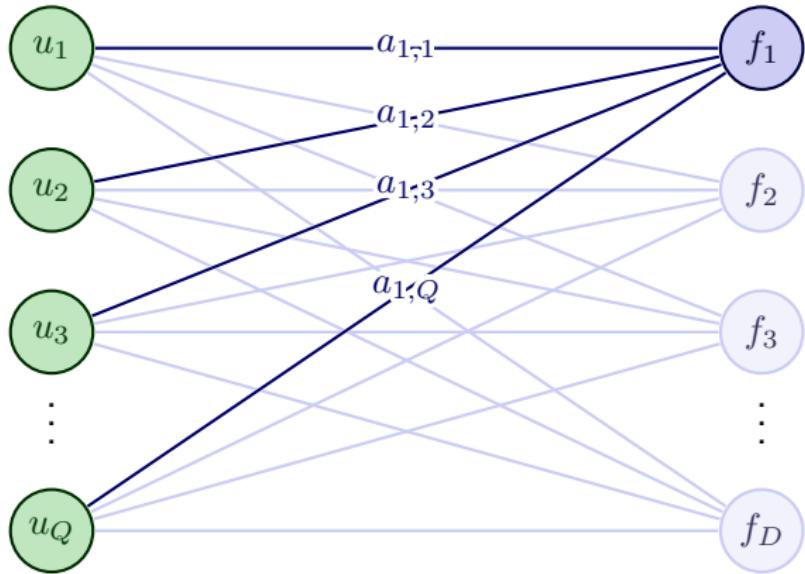
Performance Assessment

H	MSE			MSLL		CRPS	
	LAR	LSTM	SVGP	LAR	SVGP	LAR	SVGP
1	0.51	0.96	0.52 *	-0.39	-0.53	0.34	0.32
2	0.63	1.01	0.61 *	-0.27	-0.42	0.39	0.36
3	0.68	1.02	0.65 *	-0.22	-0.38	0.41	0.38
4	0.72	1.03	0.69 *	-0.18	-0.34	0.42	0.39
5	0.74	1.06	0.71 *	-0.17	-0.33	0.43	0.40
6	0.76	1.07	0.72 *	-0.16	-0.32	0.44	0.40
7	0.76	1.11	0.74 *	-0.15	-0.31	0.44	0.41
14	0.83	1.12	0.79 *	-0.10	-0.27	0.46	0.43
21	0.88	1.08	0.83 *	-0.07	-0.23	0.48	0.45
30	0.91	1.06	0.88 *	-0.05	-0.20	0.49	0.46
Grand Average	0.74	1.05	0.71 *	-0.18	-0.33	0.43	0.40

Bold and asterisk indicate a p -value $p < 1\%$ (LAR vs. SVGP, LSTM vs. SVGP). SVGP outperforms all models, except LAR at $H = 1$, where linear dependence is stronger. As horizon increases, SVGP captures complex input-output relations, significantly outperforming the other models.

Objective 2: Multi-Output Gaussian Processes

Linear Model of Coregionalization GP (LMCGP)



Independent Process (IGP)

$$u_q(\mathbf{x}) \sim \mathcal{GP}(0, k_q(\mathbf{x}, \mathbf{x}'))$$

Latent Process (LMCGP)

$$f_d(\mathbf{x}) = \sum_{q=1}^Q a_{d,q} u_q(\mathbf{x})$$

Variational Inference, ELBO, and Predictive Distribution

We extend variational inference to include the independent set, utilizing the inducing variables u_q derived from independent processes. The ELBO is given by:

$$\mathcal{L} = \sum_{d=1}^D \sum_{n=1}^N \mathbb{E}_{q(f_{dn})} \{ \log p(y_{dn} | f_{dn}) \} - \sum_{q=1}^Q \text{KL}\{q(\mathbf{u}_q) \parallel p(\mathbf{u}_q)\}$$

The posterior over test points X_* , $p(\mathbf{f}_* | \mathbf{y})$, is given by:

$$p(\mathbf{f}_* | \mathbf{y}) \approx q(\mathbf{f}_*) = \int p(\mathbf{f}_* | \mathbf{u}) q(\mathbf{u}) d\mathbf{u}$$

Gaussian noise σ_{Nd}^2 is added to obtain the predictive distribution.

Model Setup

Covariance Function (LMCGP)

The LMCGP model uses a squared exponential kernel:

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

Here, Θ_q is the lengthscale matrix, and $a_{d,q}$ work as outputscales.

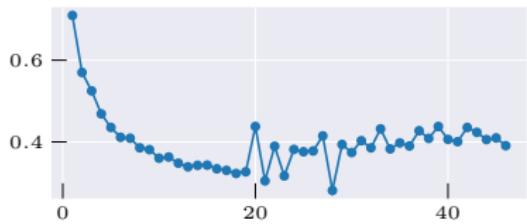
Optimization and Model Variants

Strong dependencies between parameters may cause poor local minima [12]. We address this by combining Natural Gradient (NG) to optimize variational parameters, and Adam for the rest [13].

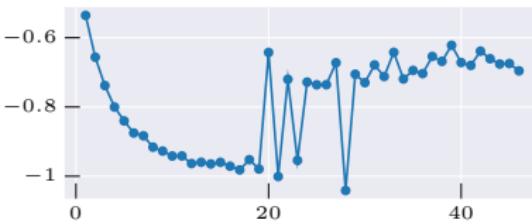
Variants:

- IGP: Independent GP (Adam).
- IGP+: Independent GP (Adam+NG).
- LMCGP: Correlated GP (Adam+NG)

Performance metrics vs Q



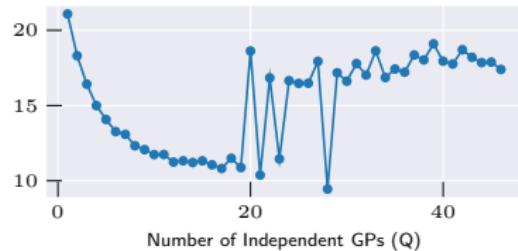
(a) MSE



(b) MSLL



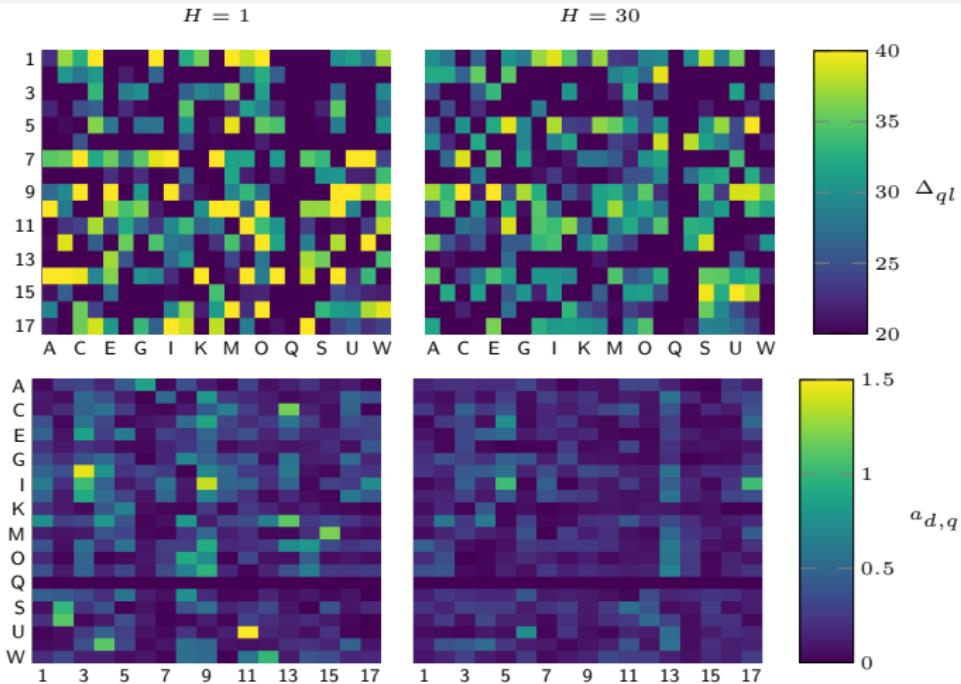
(c) CRPS



(d) NLPD

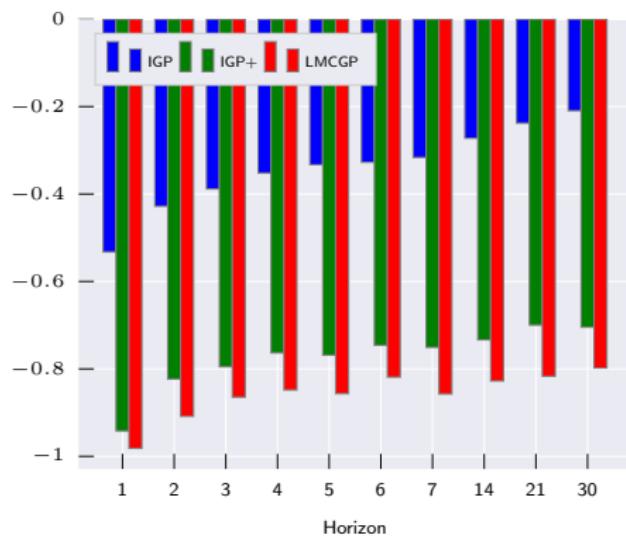
Beyond $Q = 19$, adding more independent GPs becomes counterproductive due to handling a more complex model. We select $Q = 17$ as the proper parameter.

Reservoir-wise Lengthscale and $a_{d,q}$

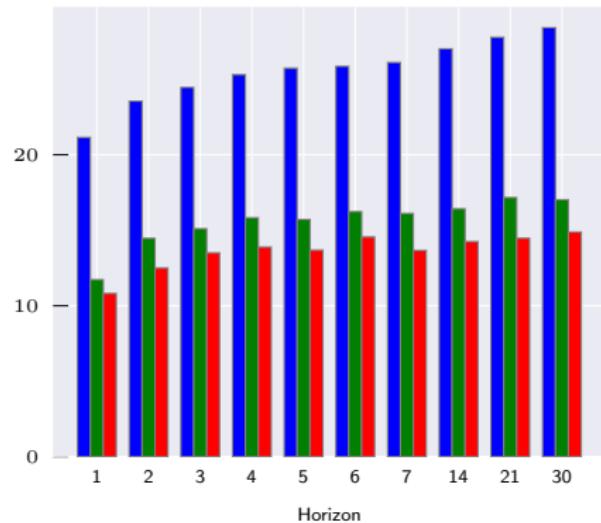


Increase horizon, lengthscale values show less selective. All independent GPs incorporate more features due to the extended time gap. The $a_{d,q}$ coefficients are smaller, indicating weaker individual feature contributions to each output.

Bar plots comparing IGP, IGP+, and LMC GP



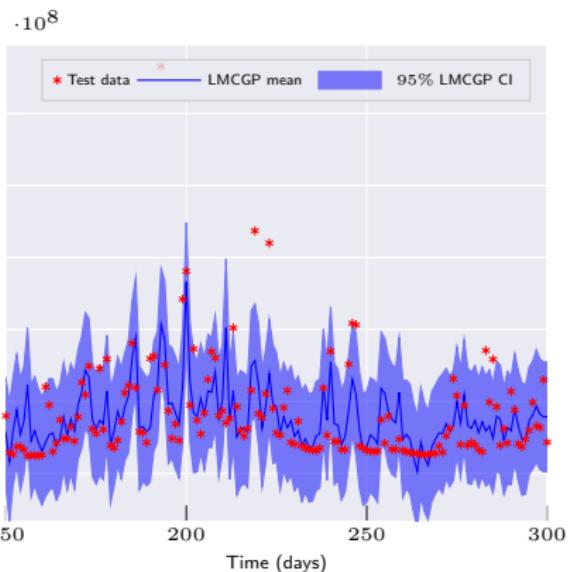
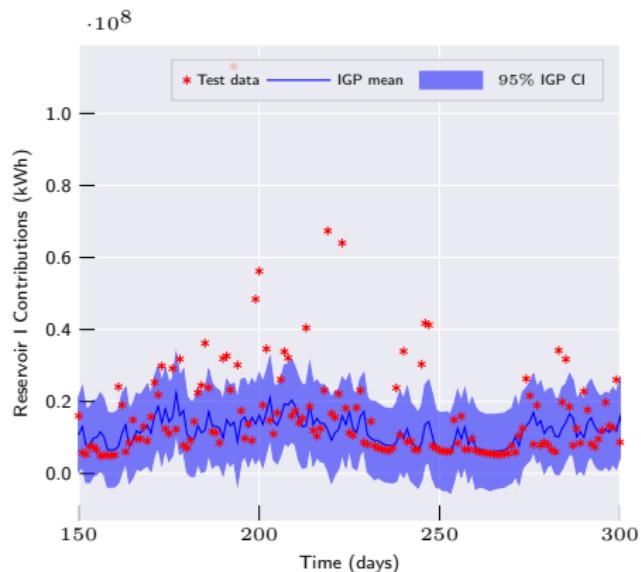
(a) MSLL



(b) NLPD

The Adam+NG optimizer significantly boosts performance, and with LMC GP showing the most improvement, especially for larger horizons.

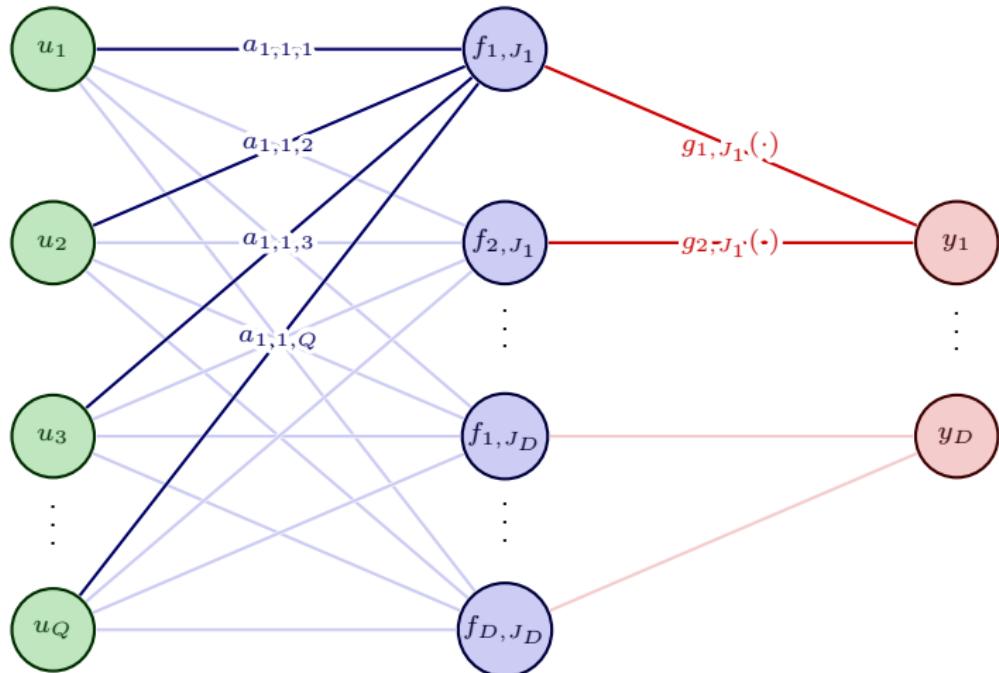
One-day-ahead Models Forecasting



The LMCGP model more accurately captures the peaks in the data, attributable to its ability to model complex behaviors through the incorporation of output correlations.

Objective 3: Chained Correlated Gaussian Processes

Chained Correlated GP (ChdGP)



Independent Process (IGP)
 $u_q(\mathbf{x}) \sim \mathcal{GP}(0, k_q(\mathbf{x}, \mathbf{x}'))$

Latent Process (LMCGP)

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

Likelihood (Chd GP)

$$\mathbf{y} | \mathbf{f} \sim \prod_{d=1}^D p(\theta_{d,1}, \dots, \theta_{d,J_d})$$

$$\theta_{d,j} = g_{d,j}(f_{d,j})$$

Variational Inference, ELBO and Predictive Distribution

We extent our variational inference, providing the following ELBO:

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

The approximated posterior over test points is given by:

$$p(\mathbf{f}_* | \mathbf{y}) \approx q(\mathbf{f}_*) = \int p(\mathbf{f}_* | \mathbf{u}) q(\mathbf{u}) d\mathbf{u}$$

And the predictive distribution for a new output \mathbf{y}_* :

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

The expectation values can be approximated via Monte Carlo methods.

Model Setup

We again make use of squared exponential kernel to construct the covariance function and Adam + NG framework to train the models.

Gaussian Likelihood: ChdGP Normal

$$p(\mathbf{y} \mid \mathbf{f}) = \prod_{d=1}^D \mathcal{N}(y_d \mid g_{d,1}(f_{d,1}), g_{d,2}(f_{d,2}))$$

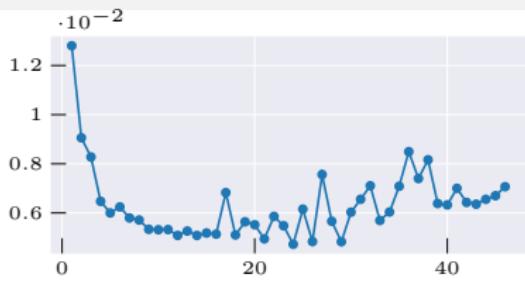
In this formulation, $g_{d,1}(\cdot) = \cdot$, while $g_{d,2}(\cdot) = \ln(\exp(\cdot) + 1)$.

Gamma Likelihood: ChdGP Gamma

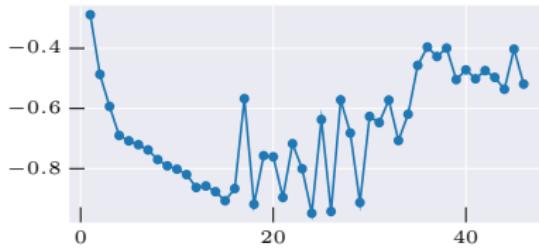
$$p(\mathbf{y} \mid \mathbf{f}) = \prod_{d=1}^D \text{Gamma}(y_d \mid g_{d,1}(f_{d,1}), g_{d,2}(f_{d,2}))$$

In this formulation $g_{d,1}(\cdot) = g_{d,2}(\cdot) = \ln(\exp(\cdot) + 1)$.

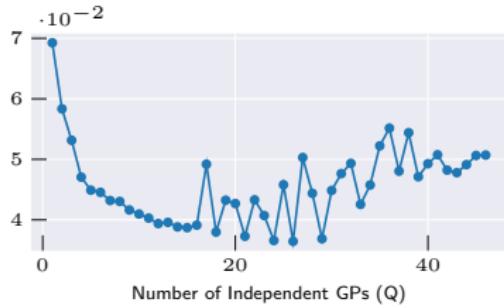
Performance metrics vs Q



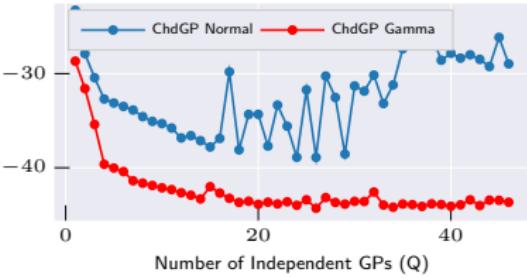
(a) MSE



(b) MSLL



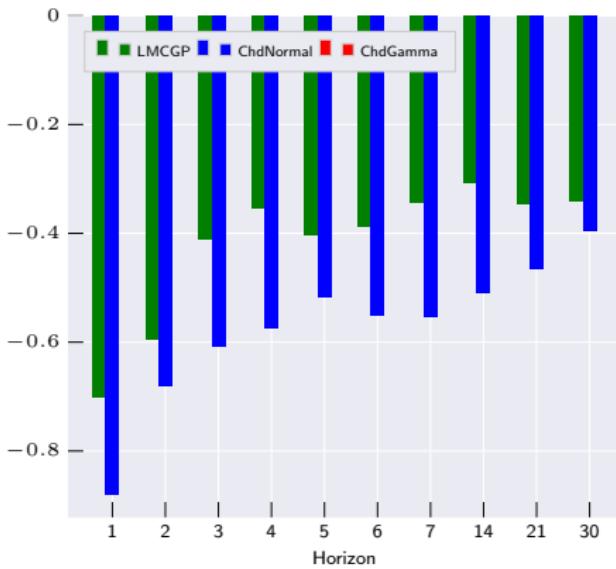
(c) CRPS



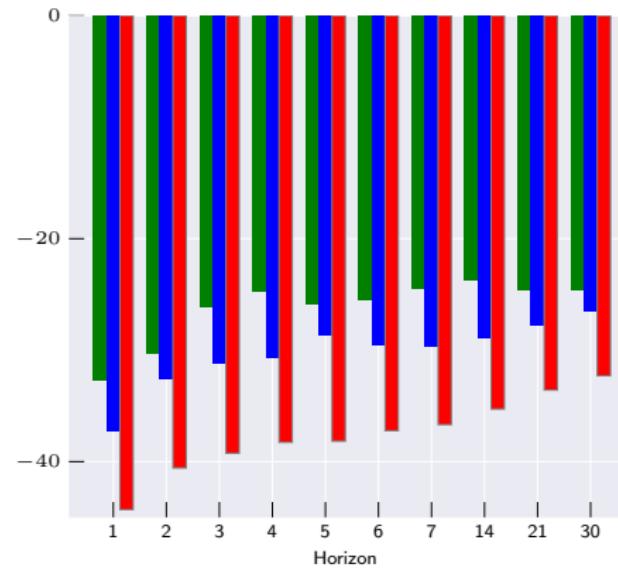
(d) NLPD

The Gamma likelihood provides greater stability by offering a more precise description of the data. We select $Q = 15$ for the Normal likelihood and $Q = 26$ for the Gamma likelihood.

LMCGP vs ChdGP across horizons



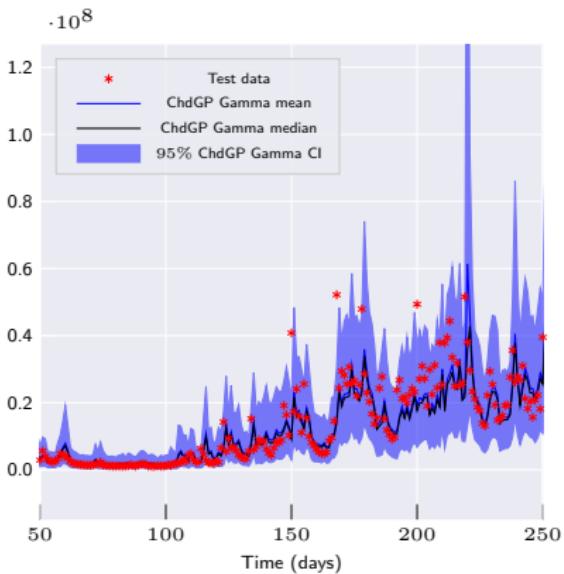
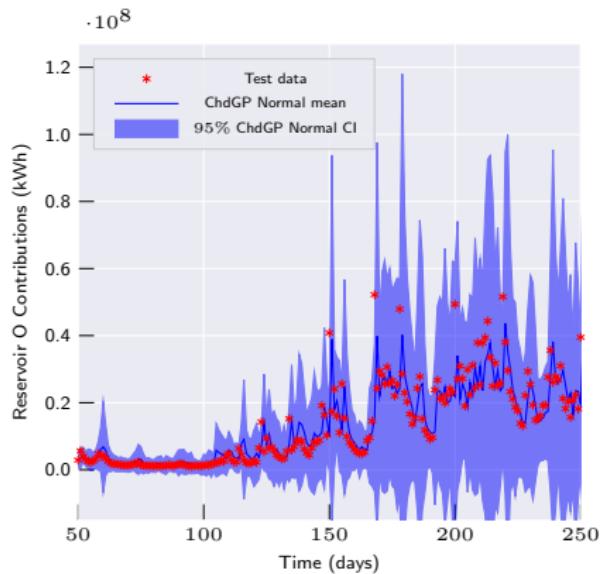
(a) MSLL



(b) NLPD

The ChdGP model consistently outperforms the LMCGP model in all scenarios. Additionally, the Gamma likelihood offers a superior explanation of the data, leading to enhanced performance.

One-day-ahead Models Forecasting



The predictive variance dynamically adapts to the complexity of uncertainty. Additionally, for the ChdGP Gamma setting, the predictive distribution does not accommodate negative values, and the presence of peaks results in a discrepancy between the mean and median.

Grand Conclusions

The Sparse Variational Gaussian Process (SVGP) model demonstrated superior performance over state-of-the-art methods like LSTM and linear autoregression. It effectively captured nonlinear relationships and provided uncertainty estimates, outperforming alternative models across various forecasting horizons.

By linearly combining independent GPs, we developed the Linear Model of Coregionalization GP (LMCGP), enhancing prediction accuracy through shared features and task-specific information. To improve optimization, we introduced a combined Adam + Natural Gradients (NG) framework, resulting in more stable solutions.

Finally, to ensure non-negative streamflow predictions, we introduced the Chained Correlated Gaussian Process (ChdGP) with a Gamma likelihood. This model showed improved stability and accuracy, particularly in capturing periodic patterns and peak occurrences, making it a promising approach for hydrological forecasting.

Generation of Knowledge Products

Article Published (Q1)



Article

Scalable and Interpretable Forecasting of Hydrological Time Series Based on Variational Gaussian Processes

Proceeding Published



Proceeding Paper

Multi-Output Variational Gaussian Process for Daily Forecasting of Hydrological Resources [†]

Registered Software *hydrogpower*



MINISTERIO DEL INTERIOR
DIRECCION NACIONAL DE DERECHO DE AUTOR
UNIDAD ADMINISTRATIVA ESPECIAL
OFICINA DE REGISTRO

CERTIFICADO DE REGISTRO DE SOPORTE LOGICO - SOFTWARE

Libro - Tomo - Partida

13-98-419

Fecha Registro

29-may.-2024

References

- [1] [[1]]Zaher Mundher Yaseen, Mohammed Falah Allawi, Ali A. Yousif, Othman Jaafar, Firdaus Mohamad Hamzah, and Ahmed El-Shafie. Non-tuned machine learning approach for hydrological time series forecasting. *Neural Computing and Applications*, 30(5):1479–1491, 2018.
- [2] [[2]]Mahsa H. Kashani, Mohammad Ali Ghorbani, Yagob Dinipashoh, and Sedaghat Shahmorad. Integration of volterra model with artificial neural networks for rainfall-runoff simulation in forested catchment of northern iran. *Journal of Hydrology*, 540:340–354, 2016.
- [3] [[3]]Muhammed Sit, Bekir Z. Demiray, Zhongrun Xiang, Gregory J. Ewing, Yusuf Sermet, and Ibrahim Demir. A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, 82(12):2635–2670, 08 2020.
- [4] [[4]]Sajjad Abdollahi, Jalil Raeisi, Mohammadreza Khalilianpour, Farshad Ahmadi, and Ozgur Kisi. Daily mean streamflow prediction in perennial and non-perennial rivers using four data driven techniques. *Water Resources Management*, 31(15):4855–4874, 2017.
- [5] [[5]]Jenq-Tzong Shiau and Hui-Ting Hsu. Suitability of ann-based daily streamflow extension models: a case study of gaoping river basin, taiwan. *Water Resources Management*, 30(4):1499–1513, 2016.
- [6] [[6]]Md. Munir Hayet Khan, Nur Shazwani Muhammad, and Ahmed El-Shafie. Wavelet based hybrid ann-arima models for meteorological drought forecasting. *Journal of Hydrology*, 590:125380, 2020.
- [7] [[7]]Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles, 2017.
- [8] [[8]]Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning*, pages 1050–1059, 2016.
- [9] [[9]]John Quilty and Jan Adamowski. A stochastic wavelet-based data-driven framework for forecasting uncertain multiscale hydrological and water resources processes. *Environmental Modelling & Software*, 130:104718, 2020.
- [10] [[10]]Wen jing Niu and Zhong kai Feng. Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. *Sustainable Cities and Society*, 64:102562, 2021.
- [11] [[11]]Wessel P. Bruinsma, Eric Perim, Will Tebbutt, J. Scott Hosking, Arno Solin, and Richard E. Turner. Scalable exact inference in multi-output gaussian processes, 2020.
- [12] [[12]]Juan-José Giraldo and Mauricio A Alvarez. A fully natural gradient scheme for improving inference of the heterogeneous multioutput gaussian process model. *IEEE Transactions on Neural Networks and Learning Systems*, 33(11):6429–6442, 2021.
- [13] [[13]]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.