# Chained Gaussian Processes for Hydrological Forecasting

Julián David Pastrana-Cortés

*Director: Álvaro Angel Orozco-Gutiérrez*

*Co-director: David Augusto Cardenas-Peña*

Automatic Research Group

August 24, 2024

# Motivation

frame

# Objetive

Develop a predictive methodology for thermoelectric plant dispatch based on historical data. This aims to estimate the natural gas demand for each plant within the planning horizon, leveraging hydrological forecasting for enhanced accuracy.

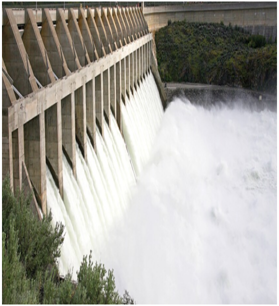
Modeling thermolectic time series is hard. Instead, model hydrological time series forecasting via hydrothermal dispatch, inferring thermoelectric dispatch.

# Motivation

We want a non-linear model to not only forecast water resources but also provide a confidence interval that measures the certainty of that prediction.

(a) Irrigation

(b) Flood control



(c) Hydropower generation

**Difficulties**: non-linearities, high stochasticity, and complex water resource patterns.

# Dataset Construction

Consider $\mathbf{v}_n \in \mathbb{R}^{D+}$ as a value useful volume time series observed at time $n$ for all $D$ reservoirs. The dataset comprises $N$ pairs of input/output observations $\{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^{N} = \{\mathbf{X}, \mathbf{Y}\}$, defined as:

$$\mathbf{y}_n = \mathbf{v}_n, \quad \mathbf{x}_n = \mathbf{v}_{n-1}$$

Each $\mathbf{v}_n$ represents a realization of a random vector $\boldsymbol{y}$. Considering $\boldsymbol{y}(\boldsymbol{x})$ for some input space $\boldsymbol{x} \in \chi$, we model a stochastic process.

# Gaussian Process Framework

A Gaussian Process (GP) models a collection of random variables, any finite number of which follow a joint Gaussian distribution:

$$\boldsymbol{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, \boldsymbol{k}(\mathbf{x}, \mathbf{x}'))$$

The covariance function $\boldsymbol{k}$ computes the correlation between points in the input space.

# Likelihood

We model outcomes $\mathbf{y}(\mathbf{x})$ using parameters $\boldsymbol{\theta}_d(\mathbf{x})$ for each outcome $y_d(\mathbf{x})$. $\boldsymbol{\theta}(\mathbf{x})$ combines all parameters across outcomes. The likelihood, or chance of observing our data $\mathbf{y}(\mathbf{x})$, is calculated by multiplying the likelihoods of each outcome, as shown:

$$p(\mathbf{y}(\mathbf{x}) \mid \boldsymbol{\theta}(\mathbf{x})) = \prod_{d=1}^{D} p(y_d(\mathbf{x}) \mid \boldsymbol{\theta}_d(\mathbf{x}))$$

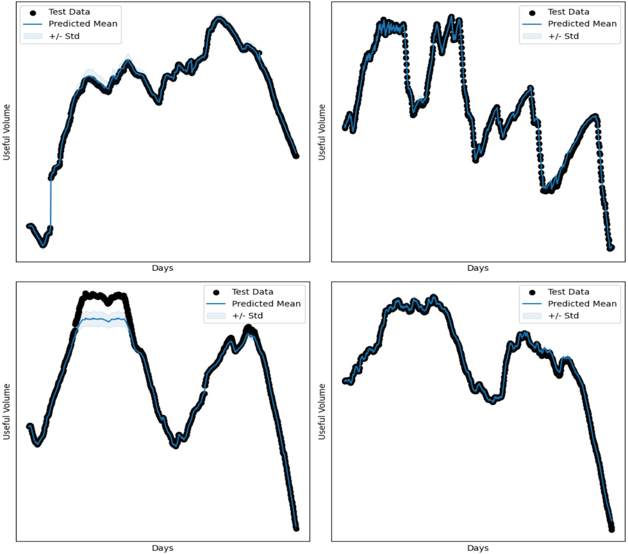
# Chained Gaussian Processes

In each likelihood distribution, parameters $\theta_{d,j}(\mathbf{x})$ are derived from Gaussian Processes (GP), specifically, $\theta_{d,j}(\mathbf{x}) = g_{d,j}(f_{d,j}(\mathbf{x}))$. This means we transform GP outputs to obtain our parameters. Combining all transformed GP outputs, $\mathbf{f}(\mathbf{x})$, allows us to model the likelihood of our observations $\mathbf{y}(\mathbf{x})$ as:

$$p(\mathbf{y}(\mathbf{x}) \mid \mathbf{f}(\mathbf{x})) = \prod_{d=1}^{D} p(y_d(\mathbf{x}) \mid \mathbf{f}_d(\mathbf{x}))$$

Here, each $y_d(\mathbf{x})$ is assumed to follow a LogNormal distribution, ensuring positivity in our predictions.

# Forecasting



# Thank You for Watching!

Thank you for your attention!

Questions or Comments?

Feel free to reach out for further discussion.