Development of a tool for long-term planning of the Colombian natural gas transportation system

Julián David Pastrana-Cortés David Augusto Cárdenas-Peña

Universidad Tecnológica de Pereira

October 8, 2025

1. Introduction

Motivation

Natural gas has become one of the most important energy sources worldwide. Moreover, long-term planning must also account for external factors such as climatic behavior, planning of new elements, and avoid shortage scenario.

Challenges

High dispatch uncertainty, operational costs, and integration of new injection sources.

General Objective

Develop a tool for the optimal long-term planning of the operation of the Colombian natural gas transportation system. The tool should enable the evaluation of multiple investment scenarios by considering the existing infrastructure, uncertainties in supply and demand, and the impact of diverse climatic conditions, while ensuring the efficient use of thermal and electrical energy.

Specific Objectives

1. Develop a methodology for predicting the dispatch of thermoelectric generation plants based on historical information provided by the grid operator. This methodology should establish the average natural gas demand of each plant for every period considered within the planning horizon.
2. Design a stochastic and/or exact optimization strategy that identifies a set of alternatives and their respective execution periods for the long-term planning of the Colombian natural gas transportation network. The strategy should minimize both operational and investment costs, taking into account the current system infrastructure and the developed prediction models.
3. Implement a long-term planning tool for the Colombian natural gas transportation network based on the developed optimization strategy and predictive models.

2. Objective 1

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

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

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 S_q, and Adam for the remaining parameters [1].

Linear Hydrothermal Dispatch Model

Thermal demand can be predicted from hydrological resources by solving a hydrothermal dispatch problem. In its simplest form, it is modeled as a linear optimization problem:

$$\min_{E_T} \sum_{i=1}^{N_T} \eta_i E_{T,i}$$

$$\text{s.t.} \quad E_{H,j} = V_{j,t} - V_{j,t+1} + A_{j,t}$$

$$\sum_{i=1}^{N_T} E_{T,i} = E_D - \sum_{j=1}^{N_H} E_{H,j}$$

$$E_{T,i}^{\min} \leq E_{T,i} \leq E_{T,i}^{\max}$$

- E_T: Thermal generation
- E_H: Hydropower generation
- E_D: Daily demand
- N_T, N_H: Number of thermal/hydro plants
- η_i: Heat rate of thermal plant i
- V: Reservoir volume
- A: Reservoir inflow
- t: Time index
- $E_T^{\min}, E_T^{\max}$: Thermal limits

3. Objective 2

[1] 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.