Bayesian Inference for Hydrological Modeling: A Case Study of Catchment Flow in the CAMELS Dataset

February 13, 2025

Abstract

This study applies Bayesian inference to estimate the relationship between catchment area and mean streamflow using hydrological data from the CAMELS dataset. A Bayesian linear regression model is constructed to analyze the impact of catchment size on mean discharge, incorporating informative priors derived from empirical data. Our results highlight the significant influence of catchment area on flow rates, providing insights into hydrological scaling relationships.

1 Introduction

Hydrological models play a crucial role in understanding water cycle dynamics. This study applies a Bayesian linear regression approach to estimate the relationship between catchment area and mean discharge, leveraging probabilistic programming tools for uncertainty quantification.

2 Methodology

Data were sourced from the CAMELS dataset, including hydrological, topographic, and geospatial attributes of river basins. We processed the data using Python, incorporating spatial information via GeoPandas and performing unit conversions where necessary. The spatial distribution of all catchments in the full dataset, along with their corresponding catchment area and mean discharge are shown in figure 1. A sample of 100 catchments was selected for analysis.

A Bayesian regression model was specified as follows:

- Response variable: **Log mean discharge ($\mathbf{m^3/s}$)**
- Predictor variable: **Log catchment area ($\mathbf{km^2}$)**
- Priors:

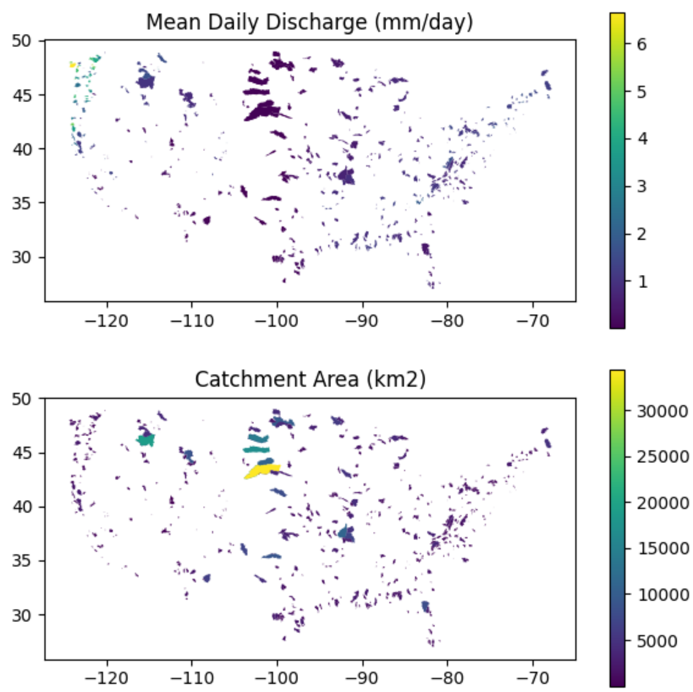

Figure 1: Spatial distribution of mean daily discharge (top) and catchment area (bottom).

- Intercept (α) $\sim$ Normal($\mu = -3.093$, $\sigma = 10$)
- Slope (β) $\sim$ Normal($\mu = 0.749$, $\sigma = 10$)
- Residual standard deviation (σ) $\sim$ HalfNormal($\sigma = 1$)

Sampling was performed using PyMC, with 2000 posterior samples drawn across three chains. Posterior predictive checks were conducted to assess model fit.

3 Results

Posterior estimates for the regression coefficients indicate a significant positive relationship between log-transformed catchment area and log-transformed discharge:

- **Intercept (α)**: -3.462 ± 0.529 (95% HDI: [-4.425, -2.438])
- **Slope (β)**: 0.755 ± 0.085 (95% HDI: [0.588, 0.909])
- **Residual standard deviation (σ)**: 1.151 ± 0.084

These results demonstrate a robust scaling relationship between catchment area and mean discharge, aligning with theoretical expectations in hydrology. The trace plots of the regression coefficients are given in figure 2, and the posterior predictive interval for the regression model with and without log adjustment are given in figure 3.

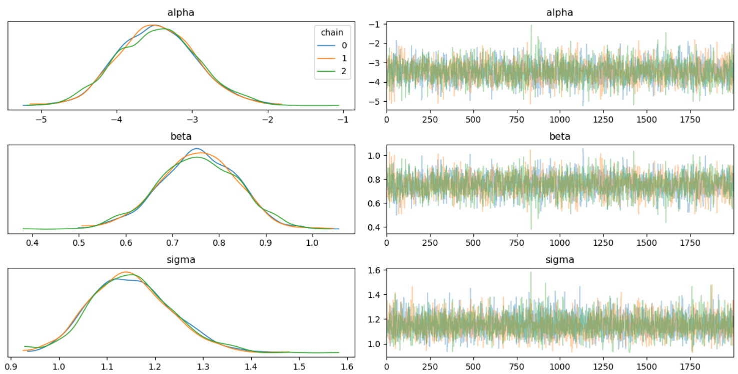

Figure 2: Trace plots for posterior samples of the regression coefficients.

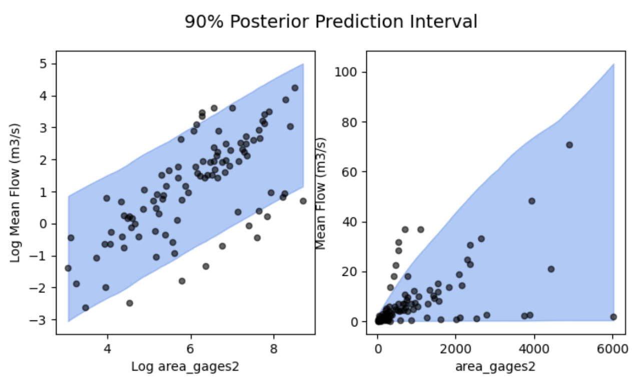

Figure 3: 90% posterior predictive interval for the regression model.

4 Discussion

The analysis reveals a strong dependence of mean streamflow on catchment area, consistent with hydrological scaling laws. The Bayesian framework enables quantification of uncertainty, providing probabilistic estimates for key parameters. The results suggest that catchment area is a dominant predictor of streamflow magnitude, reinforcing established hydrological principles.

5 Conclusion

This study demonstrates the application of Bayesian inference in hydrological modeling, illustrating the utility of probabilistic approaches for parameter estimation. Future work could extend this analysis to incorporate additional catchment characteristics and explore nonlinear relationships.