Gaussian Process Regression for Predicting Precipitation with Oscillations and Reducing Uncertainty Over Time

Ziqi Wei[1], Mofan Zhang[2], Sarah Fletcher[2]

[1]Weiyang College, Tsinghua University, [2]Department of Civil and Environmental Engineering, Stanford University

Fletcher Lab
WATER RESOURCES PLANNING FOR AN UNCERTAIN FUTURE

Gaussian Process Regression (GPR)

Concept

- **Gaussian Process:** Model each data point as part of a joint Gaussian distribution.
  $f(x) \sim \mathcal{GP}(\mu(x), k(x,x'))$ where $\mu(x)$ is the mean function; $k(x,x')$ is the kernel function.
- **Bayesian Method:** Update the probability of a hypothesis with new evidence.
  $$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)\cdot p(\theta)}{p(\mathbf{X})}$$
  where $p(\theta|\mathbf{X})$ is the posterior probability; $p(\mathbf{X}|\theta)$ is the likelihood; $p(\theta)$ is the prior probability; $p(\mathbf{X})$ is the marginal likelihood(evidence).

Kernel Function

Express the prior knowledge of the relationship between data points, adjusting with hyperparameters. It's a function of:

- $(\boldsymbol{x}-\boldsymbol{x}')$**:** stationary, $\Delta\boldsymbol{x}$-dominant;
- $(\boldsymbol{x}\cdot\boldsymbol{x}')$**:** non-stationary, non-linear change.

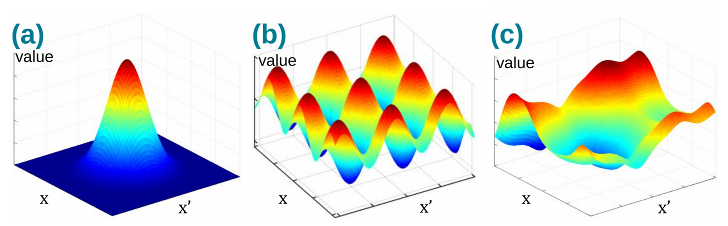


Figure 1: Kernels of **(a)** squared exponential($\boldsymbol{SE}$) **(b)** Periodic($\boldsymbol{PE}$) **(c)** polynomial($\boldsymbol{Poly}$). **(a)(b)** are stationary; **(c)** is non-stationary.

Prediction

Given a set of training data $\mathbf{X}$ with corresponding outputs $\mathbf{y}$, the posterior distribution for a new test point x_* is given by:

$$\mathbf{f}_*|\mathbf{X},\mathbf{y},x_* \sim \mathcal{N}(\mu_*, \mathbf{\Sigma}_*)$$

where
$$\begin{cases}\mu_* = k(x_*,\mathbf{X})[k(\mathbf{X},\mathbf{X})+\sigma_n^2\mathbf{I}]^{-1}\mathbf{y}\\ \mathbf{\Sigma}_* = k(x_*,x_*) - k(x_*,\mathbf{X})[k(\mathbf{X},\mathbf{X})+\sigma_n^2\mathbf{I}]^{-1}k(\mathbf{X},x_*)\end{cases}$$

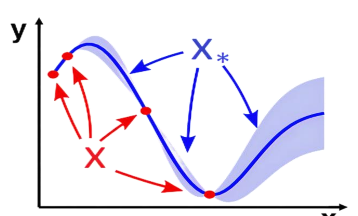


Figure 2: GPR prediction.

Overview

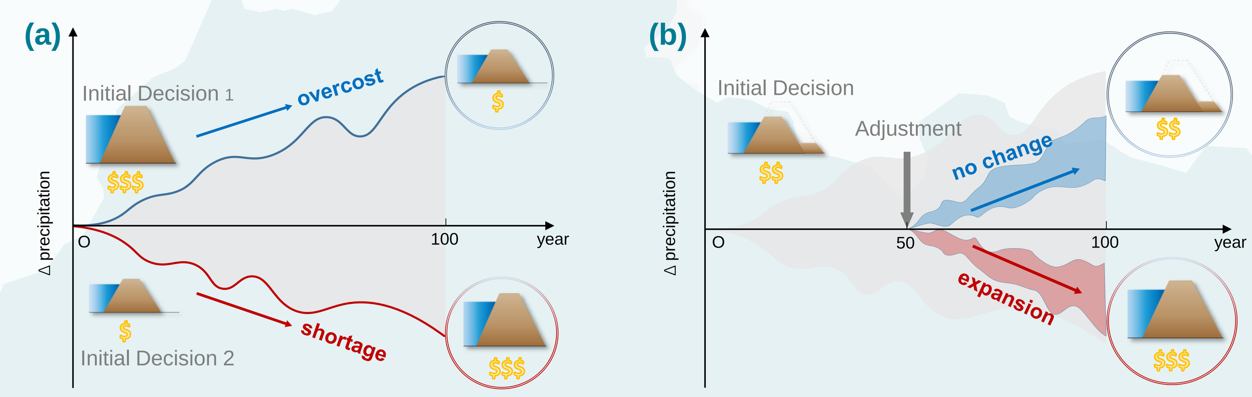


Diagram 1: Motivation – Water infrastructure with **(a)** static design and **(b)** flexible planning.

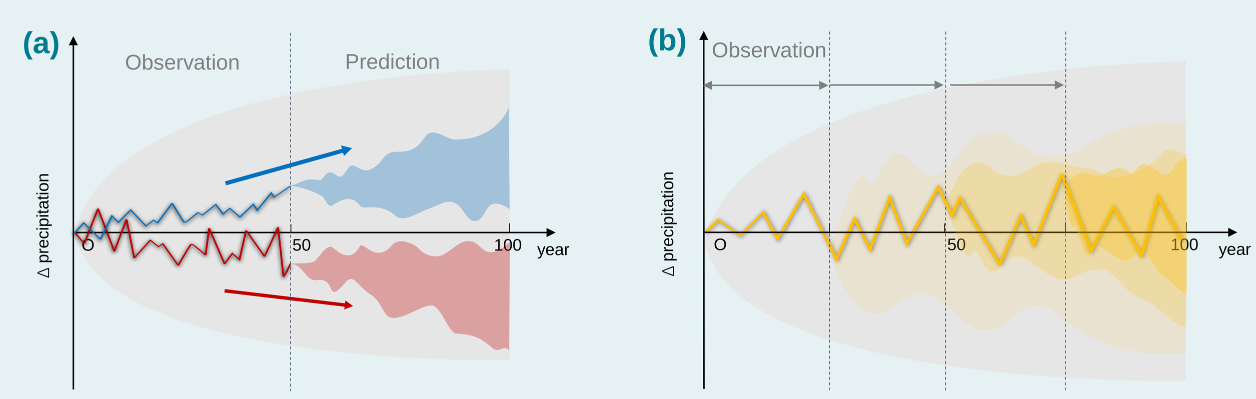


Diagram 2: Goal – Update the prediction **(a)** pattern and **(b)** uncertainty with more observations.

Water infrastructure planning should consider precipitation uncertainty in the future. And dynamic planning can be efficient when we have more observations to update the pattern and uncertainty.

Previous researches apply Gaussian Process Regression to precipitation with long-term trends. However, oscillations are not taken into consideration.

The research goal is: Gaussian Process Regression for **updating the prediction** and **reducing the uncertainty with more observations** when there are **oscillations** in precipitation.

Awash Basin, Ethiopia

Framework

Data Pre-process

Time Range	Frequency	Source	Scenario	Volume	Training	Test	Pre-process
1850 - 2099	Annual	CMIP6[1]	SSP2-4.5[2]	20 Simulations	16	4	Scalization[3]

Table 1: Precipitation data overview. [1] Coupled Model Intercomparison Project Phase 6 for global climate models. [2] Future with intermediate GHG. [3] Every simulation is scaled by its historical mean and standard deviation before 1950, aiming to better see the reduction of uncertainty while keeping the trend and oscillations.

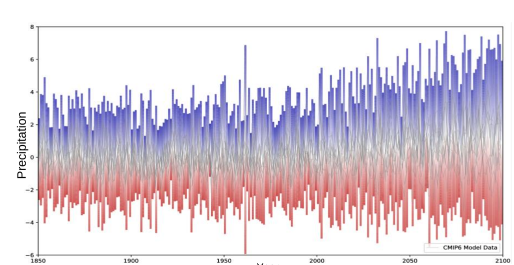


Figure 3: All scaled CMIP6 simulation.

Kernel Function Types for Oscillations

Based on experiments on Southern Oscillation Index(SOI, 3-7 year oscillations) and synthetic data(10-30 year oscillations), create two kernel function types:

① $\sum_i \boldsymbol{SE}\cdot\boldsymbol{PE}_{i,p}$

② $\sum_i \boldsymbol{Poly}(\boldsymbol{PE}_{i,p})$

$\boldsymbol{PE}_{i,p}$ means $\boldsymbol{i}^{th}$ periodic kernel with a $\boldsymbol{p}$-year period. Use different periods to simulate different scales' fluctuation. Use ①$\boldsymbol{SE}$ to control $\boldsymbol{PE}$'s magnitude or ②$\boldsymbol{Poly}$ to capture $\boldsymbol{PE}$'s change over time, imitating uncertain oscillations.

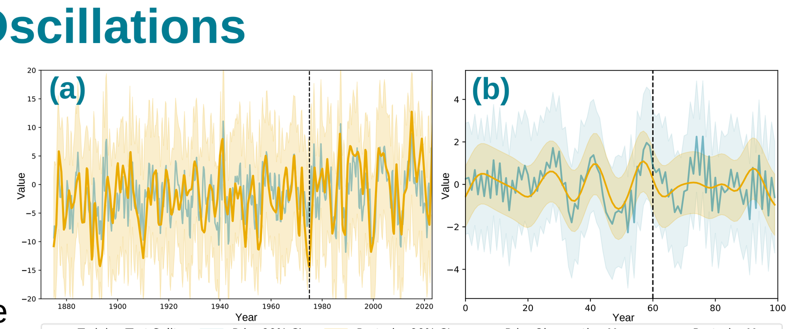
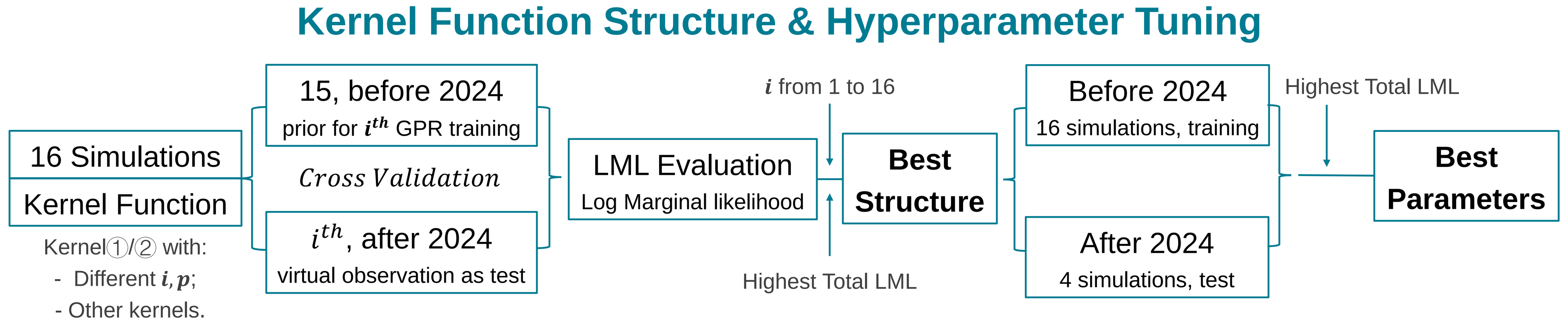


Figure 4: SOI **(a)** and synthetic data **(b)** experiments.

Kernel Function Structure & Hyperparameter Tuning



Characteristics

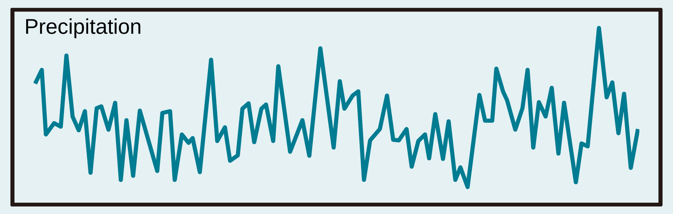


Diagram 3: The Awash Basin's precipitation Diagram.

There are 3 main characteristics in the Awash Basin's precipitation:

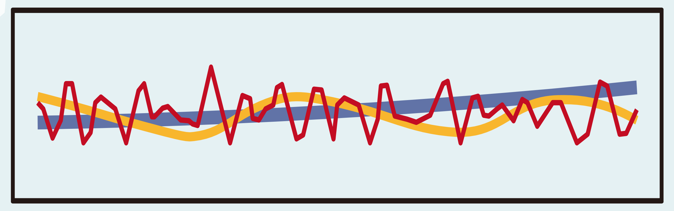


Diagram 4: Sketches for 3 characteristics.

- **3 - 7 year oscillation:** Influenced by El Niño-Southern Oscillation.
- **10 - 30 year oscillation:** Influenced by Pacific Decadal Oscillation.
- **long-term trend:** Influenced by Overall Precipitation.

Note that **oscillations are uncertain**: There is a roughly defined period range for each oscillation, but the exact period length varies unpredictably.

Results

Use 16 CMIP6 simulations as prior knowledge to develop a GPR model structure, and other 4 simulations as virtual observations individually. For one virtual observation, assuming x years as observational data, fine tune hyperparameters to predict precipitation for the next $(250-x)$ year.

Predictions Vary with Different Observations

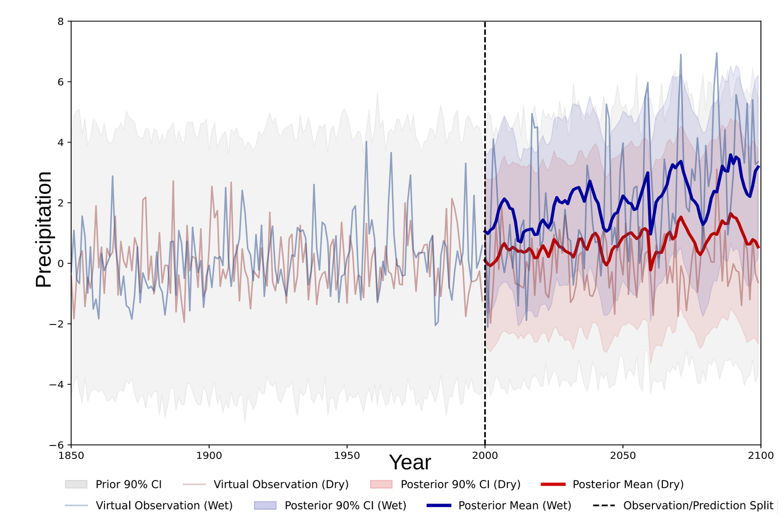


Figure 5: Predictions of two observations by a $\sum\boldsymbol{SE}\cdot\boldsymbol{PE}$-based kernel, with two 3-7 year periods and two 10-30 year periods.

With 150-year observations, the wet simulation tends to be wetter in the future, while the dry simulation maintains lower precipitation. They are **distinguished by predicted mean and uncertainty range**, but **maintain a similar oscillation pattern**, which is determined by prior knowledge.

Update Predictions with More Observations

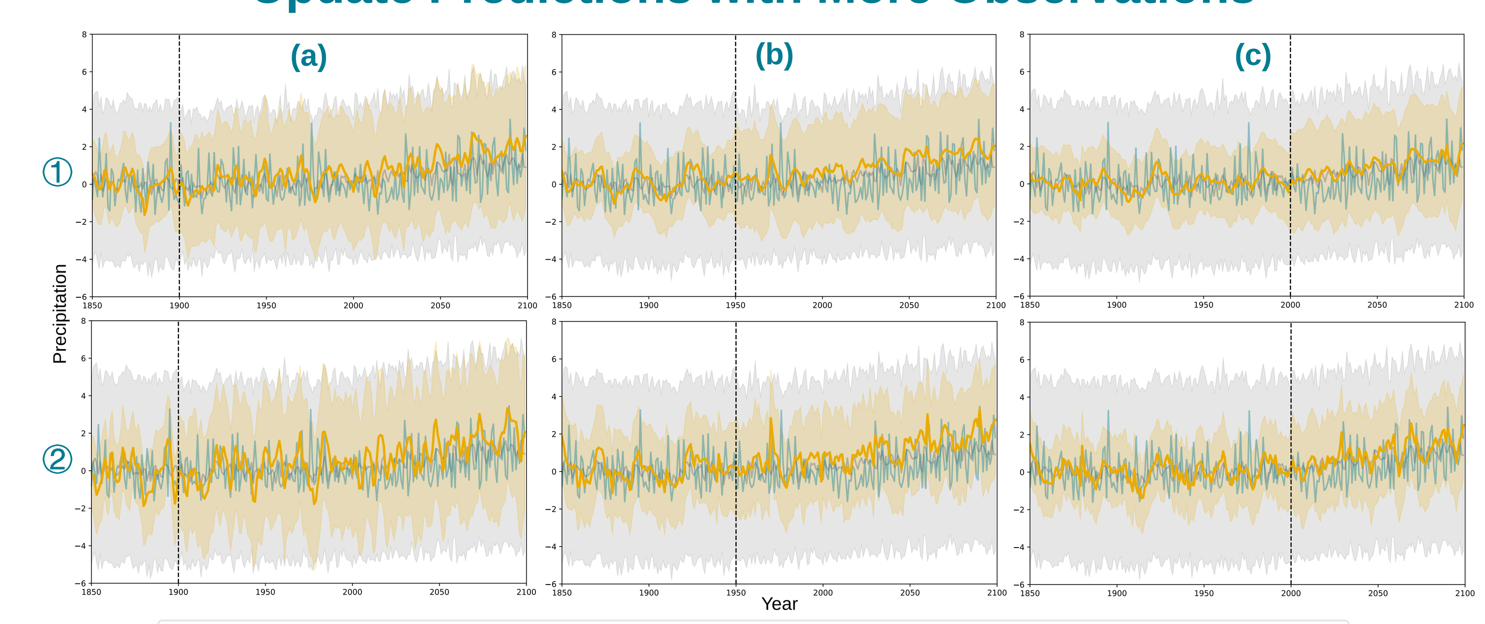


Figure 6: Predictions with **(a)**50 **(b)**100 **(c)**150 years as observations by ①$\sum\boldsymbol{SE}\cdot\boldsymbol{PE}$-based ②$\sum\boldsymbol{Poly}(\boldsymbol{PE})$-based kernel.

With more observations, the posterior uncertainty gradually decreases and the posterior mean increasingly reflects the future oscillation pattern. Compared to the $\sum\boldsymbol{SE}\cdot\boldsymbol{PE}$-based kernel, the $\sum\boldsymbol{Poly}(\boldsymbol{PE})$-based kernel fits the data more closely, but also inevitably introduces more noise.

Takeaways & Future Work

- Gaussian Process Regression can model the Awash Basin's precipitation with the 3-7 year oscillation, 10-30 year oscillation and long-term trend, based on the CMIP6 simulations.
- Two kernel function structures $\sum\boldsymbol{SE}\cdot\boldsymbol{PE}$ and $\sum\boldsymbol{Poly}(\boldsymbol{PE})$ can capture uncertain oscillations.
- The prediction can be updated with observations and the uncertainty can be reduced with more observations.
- Future work includes developing more reasonable evaluation metrics, applying this method to other regions with oscillations in precipitation, and combing the precipitation prediction with dynamic water infrastructure planning under deep uncertainty.

Acknowledgements & Reference

Sincere gratitude to Prof. Sarah Fletcher for continued encouragement and support, to Mofan Zhang for her warm guidance and help, and to all Fletcher's Lab members, without whom this work wouldn't be possible to complete. Thanks to the Stanford School of Engineering and the UGVR program for providing this research opportunity. And many thanks to all the UGVR members for our shared exploration and discussions in this wonderful summer.

[1] *Fletcher, S., Zaniolo, M., Zhang, M., & Lickley, M. (2023). Climate oscillation impacts on water supply augmentation planning. Proceedings of the National Academy of Sciences, 120(e2215681120). https://doi.org/10.1073/pnas.2215681120*
[2] *Wei, P., Vo, T. V., Qu, X., Ong, Y. S., & Ma, Z. (2023). Transfer Kernel Learning for Multi-Source Transfer Gaussian Process Regression. IEEE transactions on pattern analysis and machine intelligence, 45(3), 3862–3876. https://doi.org/10.1109/TPAMI.2022.3184696*
[3] *Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. The MIT Press.*
[4] *Duvenaud, D. (2014). Automatic model construction with Gaussian processes (Ph.D. dissertation). University of Cambridge.*
[5] *Wang, Jie. (2023). An Intuitive Tutorial to Gaussian Process Regression. Computing in Science & Engineering. PP. 1-8. 10.1109/MCSE.2023.3342149.*