Enhanced Network Bandwidth Prediction with Multi-Output Gaussian Process Regression

Shude Chen[1], Takayuki Nakachi[2]
Graduate School of Engineering and Science, University of the Ryukyus, Okinawa, Japan[1]
Information Technology Center, University of the Ryukyus, Okinawa, Japan[2]

***Abstract*—Modern network environments, especially in domains like 5G and IoT, exhibit highly dynamic and nonlinear traffic behaviors, posing significant challenges for accurate time series analysis and predictive modeling. Traditional approaches, including stochastic ARIMA and deep learning-based LSTM, frequently encounter difficulties in capturing rapid signal variations and inter-channel dependencies, often due to data sparsity or excessive computational cost. To address these issues, this paper proposes a Multi-Output Gaussian Process (MOGP) framework augmented with a novel signal processing strategy, where additional signals are generated by summing adjacent elements over multiple window sizes. Such multi-scale enrichment effectively leverages cross-channel correlations, enabling the MOGP model to discover complex temporal patterns in multi-channel data. Experimental results on real-world network traces highlight that the proposed method achieves consistently lower RMSE compared to conventional single-output or deep learning methods, thereby underscoring its value for robust bandwidth estimation. Our findings suggest that integrating MOGP with multi-scale augmentation holds promise for a wide range of predictive analytics applications, including resource allocation in 5G networks and traffic monitoring in IoT systems.**

***Keywords*—Network traffic prediction; Multi-Output Gaussian Process (MOGP); signal processing; time series analysis; predictive modeling; multi-channel data; IoT traffic monitoring; 5G networks**

I. Introduction

The proliferation of diverse network applications has led to increasingly congested environments, necessitating efficient resource allocation to maintain service quality and operational stability [1], [2]. With the rapid growth in both the volume and complexity of network traffic, the non-stationary and dynamic nature of traffic patterns has become a critical focus in telecommunications research [3], [4]. Accurate and timely network traffic prediction is essential to mitigate delays, prevent data loss, and optimize resource utilization in overloaded networks [5].

Recent advancements in traffic prediction technologies have enabled dynamic resource allocation through accurate traffic forecasting, typically categorized into short-term, medium-term, and long-term forecasts [6]. Short-term forecasting focuses on real-time predictions for the next few seconds or minutes, essential for time-sensitive applications. In contrast, long-term forecasting relies on historical trends over extended periods to guide strategic planning. Medium-term forecasting, which bridges the gap between these two, is particularly challenging due to the inherent variability in network traffic spanning minutes to hours.

Traditional prediction models, such as the Auto-Regressive Integrated Moving Average (ARIMA) [7], excel in capturing linear trends but struggle with the nonlinear and dynamic nature of modern network traffic [8], [9]. Deep learning models, such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks [6], offer superior nonlinear forecasting capabilities but often face challenges with overfitting, high computational demands, and large data requirements.

Gaussian Processes (GP) [10], [11] have emerged as a powerful tool for modeling nonlinear data and providing uncertainty quantification, making them particularly effective for network traffic forecasting with limited data [12], [13]. However, traditional GP methods limit their ability to adapt to dynamic and multi-scale traffic variations. This results in prediction errors over extended horizons, complicating medium-term forecasting tasks.

To address these challenges, Building on the strengths of existing GP-based multi-slot-ahead prediction models [12]-[15], this study introduces a Multi-Output Gaussian Process (MOGP) framework, incorporating an innovative signal augmentation strategy. The proposed approach introduces two key innovations: 1) the utilization of inter-channel correlations to improve prediction accuracy and reduce error propagation over extended periods, and 2) a signal augmentation strategy that generates additional input features by summing adjacent data points with varying window sizes. By enriching the input dataset with these newly generated signals, the model captures complex temporal dependencies and inter-channel interactions more effectively.

The main contributions of this paper are:

- Proposed a network bandwidth prediction framework based on Multi-Output Gaussian Process (MOGP), combined with a signal enhancement strategy.
- Enriched the input feature space and improved prediction accuracy by generating adjacent element sums with varying window sizes.
- Validated the effectiveness of the method through experiments on real multi-channel network traffic datasets, significantly reducing prediction errors.

The remainder of this paper is organized as follows: Section II reviews related methodologies and their limitations. Section III details the proposed forecasting approach, including the signal augmentation strategy and the MOGP framework. Section IV presents the dataset and experimental setup. Section

V discusses simulation results and evaluates the model's performance. Finally, Section VI concludes the paper and outlines future research directions.

II. Related Work

The evolution of network bandwidth prediction techniques has progressed from classical statistical approaches to sophisticated machine learning methods.

A. Methods Based on Probabilistic Processes

Early models like ARIMA have been extensively utilized for time series forecasting in network traffic. These models, though effective for linear trends, struggle with the complex and non-stationary nature of network data [15]. Enhancements like ARIMA/GARCH have been proposed to handle long-term dependencies, but these methods still face challenges in tracking rapidly changing network traffic characteristics [5]. These methods typically focus on single time series data, capturing patterns and periodicity over time, but have limitations in modeling long-term dependencies.

B. Machine Learning Methods

Recent advancements have seen the application of neural network architectures, such as LSTMs, which are adept at capturing nonlinearities and temporal dependencies in traffic data, providing significant improvements over traditional models [16]. Despite their effectiveness, deep learning models require substantial amounts of training data and computational resources, which can be impractical in dynamic network environments. Additionally, methods such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) have been employed to capture temporal and spatial dependencies in network traffic [17]. However, these machine learning methods are often limited by their high computational cost and susceptibility to overfitting in small datasets.

C. Traditional Gaussian Processes

Gaussian Processes (GP) have emerged as a powerful tool for network traffic prediction, given their ability to model uncertainty and non-linear dynamics of network traffic [8]. Particularly, Multi-Output Gaussian Processes (MOGP) have shown great potential in leveraging the correlations among multiple network channels to enhance prediction accuracy. GP models offer the advantage of providing uncertainty measures along with predictions, crucial for robust network management [18]. Incorporating convolutional structures in GPs, researchers have managed to capture spatiotemporal characteristics of traffic data more effectively, enhancing predictive performance [18]. These models have been particularly useful in scenarios involving large-scale data from multiple network sensors [19].

D. Recent Advances and Our Contribution

Recent studies, such as the work by Wang et al., explore a GP-based online learning framework for multi-slot-ahead network traffic forecasts [13]. They emphasize the evolving nature of network traffic and propose a dynamic kernel design using Spectrum Mixture (SM) functions and Process Convolution (PConv) to capture complex traffic patterns over different time scales. This approach addresses both tracking capability and prediction horizon challenges, demonstrating superior performance through simulations [13].

Emerging methodologies such as Graph Gaussian Processes (Graph-GP) [4] and Multi-channel Transformer-based models (MCformer) [20] further enhance the ability to model spatiotemporal correlations and interdependencies in network traffic. Graph-GP adapts well to scenarios with missing data and non-stationary traffic, while MCformer utilizes attention mechanisms to effectively capture multi-channel dependencies, proving highly effective in dynamic environments.However, these methods require extensive labeled datasets and suffer from high computational complexity.

E. Gap in Literature

- Existing deep learning models such as RNN and LSTM suffer from overfitting and large data requirements.
- Conventional GP-based methods do not efficiently utilize inter-channel correlations, limiting their predictive accuracy.
- Recent studies on multi-output regression models for network forecasting have not explored adjacent signal augmentation as a feature engineering technique.
- Our work extends these methodologies by introducing a signal augmentation strategy within an MOGP framework, improving multi-step forecasting accuracy.

Our proposed method extends these traditional GP approaches by using a multi-channel framework. By generating new signals from the original signals through summing adjacent signals [7], we capture the underlying characteristics of the data more effectively. Initially, the two channels with the strongest correlations are selected as the original data. Subsequently, new signals are created by using adjacent sums with varying window sizes for each channel signal. This approach enhances the prediction accuracy by exploiting inter-channel correlations more effectively, resulting in a robust framework for multi-channel network traffic forecasting.

III. Proposed Method

This section introduces the proposed Multi-Output Gaussian Process (MOGP) model for network bandwidth prediction. The method systematically integrates correlation analysis, signal transformation, and a Gaussian Process framework to improve prediction accuracy. The workflow, illustrated in Fig. 1, involves selecting highly correlated input signals, generating new signals with varying temporal scales, and training the MOGP model for predictive tasks.

A. Flowchart Overview of the Prediction Process

The overall workflow is illustrated in Fig. 1. The process begins with correlation analysis to identify a subset of highly correlated signals. These signals are then enriched by generating new signals through adjacent sums with varying window sizes. The original and generated signals are subsequently input into the MOGP model, which leverages their temporal and inter-channel dependencies for bandwidth prediction.

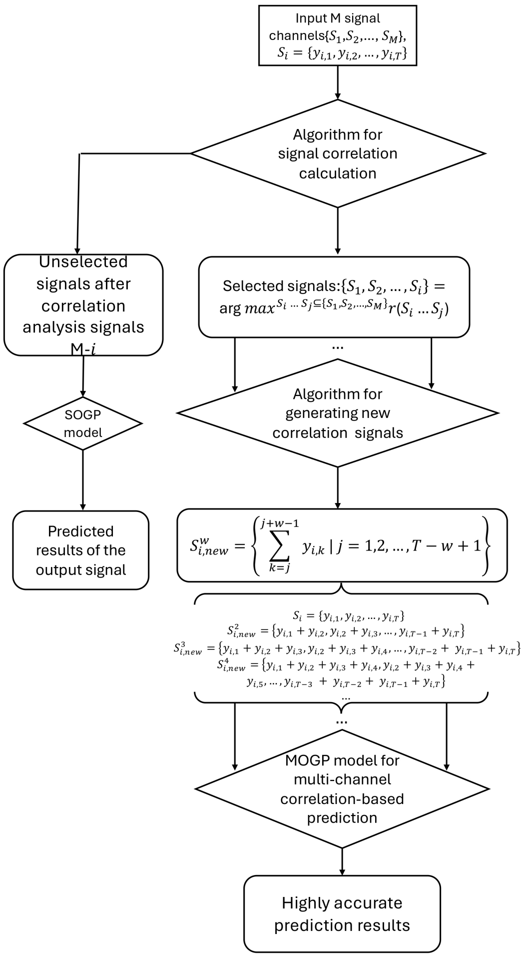

Fig. 1. Flowchart of the proposed prediction method.

B. Input Signal Definition and Correlation Analysis

Time Index and Signals. We define the discrete time axis as

$$t = 1,\, 2,\, \ldots,\, T. \tag{1}$$

For each time t, we have M signals $\{\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_M\}$, where

$$\mathbf{S}_i = \{y_{i,1},\, y_{i,2}, \ldots, y_{i,T}\}, \quad i = 1, \ldots, M, \tag{2}$$

and $y_{i,t}$ is the value of the i-th signal at time t.

Let the output signals be denoted as

$$\{\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_M\}, \tag{3}$$

where each $\boldsymbol{S}_i = \{y_{i,1}, y_{i,2}, \ldots, y_{i,T}\}$ represents a sequence of data points over time.

1) Correlation Metric: The Pearson correlation coefficient is used to quantify the linear relationship between two signals $\mathbf{S}_k$ and $\mathbf{S}_l$[21], [22]:

$$r(\mathbf{S}_k, \mathbf{S}_l) = \frac{\sum_{t=1}^{T}(y_{k,t} - \bar{y}_k)(y_{l,t} - \bar{y}_l)}{\sqrt{\sum_{t=1}^{T}(y_{k,t} - \bar{y}_k)^2}\sqrt{\sum_{t=1}^{T}(y_{l,t} - \bar{y}_l)^2}}, \tag{4}$$

where $\bar{y}_k$ and $\bar{y}_l$ are the mean values of $\mathbf{S}_k$ and $\mathbf{S}_l$, respectively.

C. Selection of Highly Correlated Signals

Identifying highly correlated signals is crucial for reducing model complexity while maintaining predictive accuracy.

To identify the most relevant signals, the following selection criterion is applied. Let L be the maximum number of signals we wish to select. We choose:

$$\{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_L\} = \{\mathbf{S}_k \mid r(\mathbf{S}_k, \mathbf{S}_l) > \text{threshold}, \; \forall \mathbf{S}_k, \mathbf{S}_l \in \{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_M\}\}, \tag{5}$$

where,

- $r(\mathbf{S}_k, \mathbf{S}_l)$: Correlation coefficient between signals $\mathbf{S}_k$ and $\mathbf{S}_l$,
- threshold: Minimum correlation value required for a signal to be included. This yields L original signals with the highest inter-correlations,
- $\{\mathbf{S}_k\}$: Subset of signals that meet the correlation criterion.

D. Generation of New Signals

To enrich the input features, new signals are generated by summing adjacent data points in the selected signals [12]. For a signal $\mathbf{S}_i$ of length T (i.e., $\mathbf{S}_i = \{y_{i,1}, y_{i,2}, \ldots, y_{i,T}\}$), and a chosen window size w, the new signal $\mathbf{S}_{i,\text{aug}}^{(w)}$ is computed as:

$$\mathbf{S}_{i,\text{aug}}^{(w)} = \Big\{ \sum_{k=j}^{j+w-1} y_{i,k} \;\Big|\; j = 1, 2, \ldots,\, T - w + 1 \Big\}. \tag{6}$$

where,

- $y_{i,k}$: The k-th data point of the original signal $\mathbf{S}_i$,
- k: Index of a data point within the summation window, ranging from j to $j + w - 1$,
- j: Starting index of the summation window in $\mathbf{S}_i$,
- w: Window size, i.e., the number of consecutive data points to be summed,
- T: Total length (in data points) of the original signal $\mathbf{S}_i$,
- $\mathbf{S}_{i,\text{aug}}^{(w)}$: The newly generated signal of length $T - w + 1$.

For example,

- For $w = 2$ (adjacent two-term sums):

$$\mathbf{S}_{i,\text{aug}}^{(2)} = \{\, y_{i,1} + y_{i,2},\; y_{i,2} + y_{i,3},\; \ldots,\; y_{i,T-1} + y_{i,T} \}. \tag{7}$$

- For $w = 3$ (adjacent three-term sums):

$$\mathbf{S}_{i,\text{aug}}^{(3)} = \Big\{ y_{i,1} + y_{i,2} + y_{i,3}, \quad y_{i,2} + y_{i,3} + y_{i,4}, \; \ldots, \quad y_{i,T-2} + y_{i,T-1} + y_{i,T} \Big\}. \tag{8}$$

Signal Length Note: Hence, if the original signal $\mathbf{S}_i$ has length T, the generated signal $\mathbf{S}_{i,\text{aug}}^{(w)}$ has length $T - w + 1$. By varying w, we can capture different scales of temporal dependencies, thereby enhancing the feature set for the Multi-Output Gaussian Process (MOGP) model.

E. Visualization of Signals and Generated Signals

Fig. 2 and 3 illustrate the transformation from two original signals, $\mathbf{S}_1$ and $\mathbf{S}_2$, into their augmented versions using window sizes $w \in \{2, 3, 4\}$.

a) Original Signals: Fig. 2 shows two original signals, $\mathbf{S}_1$ and $\mathbf{S}_2$, used as the starting input for the prediction process. The periodic nature and daily patterns of these signals are evident, highlighting their suitability for modeling temporal dependencies.

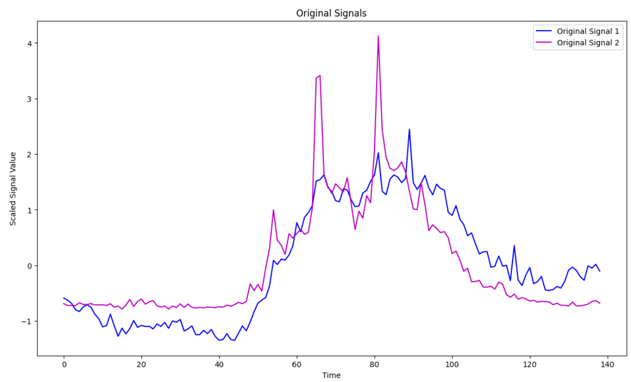

Fig. 2. Original signals ($\mathbf{S}_1$ in blue and $\mathbf{S}_2$ in magenta) with scaled values over time.

b) Augmented Signals via Adjacent Sums: Given a signal $\mathbf{S}_i = \{y_{i,1}, y_{i,2}, \dots, y_{i,T}\}$, we define its w-term adjacent-sum (augmented signal) as

$$\mathbf{S}_{i,\text{aug}}^{(w)} = \{\underbrace{y_{i,1} + y_{i,2} + \cdots + y_{i,w}}_{w \text{ terms}}, \underbrace{y_{i,2} + \cdots + y_{i,w+1}}_{w \text{ terms}}, \dots\}. \tag{9}$$

As w increases, these sums capture progressively larger local trends. Fig. 3 shows newly generated signals (for $w = 2, 3, 4$) from the original $\mathbf{S}_1$ and $\mathbf{S}_2$, demonstrating how the adjacent-sum approach broadens temporal coverage.

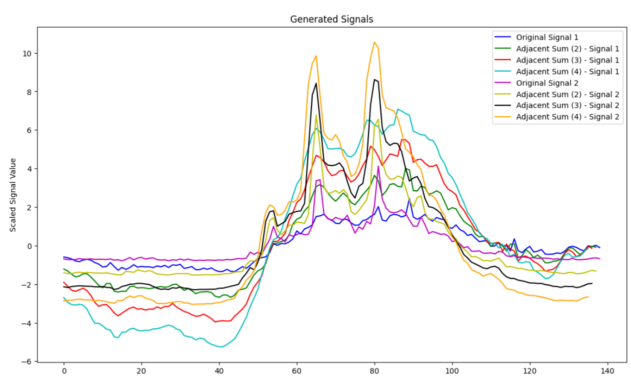

Fig. 3. Newly generated signals from $\mathbf{S}_1$ and $\mathbf{S}_2$, illustrating different window sizes w.

c) Defining MOGP$(L, \tilde{N})$: Suppose we select L original signals based on correlation analysis, and for each original signal, we generate $\tilde{N}$ adjacent-sum signals (e.g. multiple choices of w). We then denote our model by

$$\text{MOGP}(L, \tilde{N}), \tag{10}$$

meaning:

- L is the number of *original* (highly correlated) signals chosen,
- $\tilde{N}$ is the total number of *augmented* signals generated from those L originals,
- Altogether, there are $N = L + \tilde{N}$ channels in the resulting multi-output GP.

For instance,

$$\text{MOGP}(2, 6) \tag{11}$$

indicates two original signals and six augmented signals, making $N = 8$ jointly modeled channels.

d) Example: Generating Up to Eight Channels: If $\mathbf{S}_p$ and $\mathbf{S}_q$ are the selected original signals, then for $w = 2, 3, 4$ each signal spawns three augmented versions:

$$\mathbf{S}_{p,\text{aug}}^{(2)}, \quad \mathbf{S}_{p,\text{aug}}^{(3)}, \quad \mathbf{S}_{p,\text{aug}}^{(4)}, \quad \mathbf{S}_{q,\text{aug}}^{(2)}, \quad \mathbf{S}_{q,\text{aug}}^{(3)}, \quad \mathbf{S}_{q,\text{aug}}^{(4)}. \tag{12}$$

Thus, we obtain up to eight channels in total:

$$\underbrace{\mathbf{S}_p, \ \mathbf{S}_q}_{\text{original signals}}, \ \underbrace{\mathbf{S}_{p,\text{aug}}^{(2)}, \ \mathbf{S}_{q,\text{aug}}^{(2)}}_{\text{2-term}}, \ \underbrace{\mathbf{S}_{p,\text{aug}}^{(3)}, \ \mathbf{S}_{q,\text{aug}}^{(3)}}_{\text{3-term}}, \ \underbrace{\mathbf{S}_{p,\text{aug}}^{(4)}, \ \mathbf{S}_{q,\text{aug}}^{(4)}}_{\text{4-term}}. \tag{13}$$

Each augmented channel reflects a different local summation scale, thereby enriching the feature set that the Multi-Output Gaussian Process (MOGP) exploits for multi-step prediction. Hence, the MOGP$(L, \tilde{N})$ approach systematically combines correlated original signals with their adjacency-sum signals, allowing finer control over both short- and long-term dependencies.

F. Multi-Output Gaussian Process (MOGP) Prediction Framework

After identifying highly correlated signals and generating augmented signals (Sections III-D–III-E), we obtain N total channels (*original* + *augmented*). These N channels are collectively modeled using a *Multi-Output Gaussian Process* (MOGP), which exploits *both* temporal correlations *and* inter-channel correlations to enhance predictive accuracy.

1) Input/Output Setup and Kernel:

a) Inputs $\mathbf{X}$*:* We consider T time indices or feature vectors $x_1, x_2, \dots, x_T \in \mathcal{X}$, collected in $\mathbf{X} = \{x_1, \dots, x_T\}$. (Each x_t can be a scalar time step or a multi-dimensional feature.)

b) Outputs $\mathbf{Y}$*:* All original and augmented signals together form $N = L + \tilde{N}$ channels:

$$\mathbf{Y} = \{\mathbf{S}_1, \dots, \mathbf{S}_L, \ \mathbf{S}_{1,\text{aug}}^{(w)}, \dots, \mathbf{S}_{L,\text{aug}}^{(w)}\}. \tag{14}$$

We collect these N channels (each of length T) into a *multi-channel output matrix* $\mathbf{Y} \in R^{N \times T}$. In a block sense, we may write

$$\mathbf{Y} = [\mathbf{S}_1, \ \mathbf{S}_2, \dots, \mathbf{S}_N], \tag{15}$$

where each $\mathbf{S}_i \in R^T$. Equivalently, to index individual data values.

c) MOGP Kernel: A Gaussian Process (GP) is defined by a mean function (often taken as zero) and a covariance (kernel) function capturing similarities among inputs. For time-series forecasting, we frequently adopt an RBF+noise kernel [23], [24]:

$$k(x,x') \;=\; \theta_1 \exp\Big(-\frac{\|x-x'\|^2}{2\,\theta_2}\Big) \;+\; \theta_3\,\delta(x,x'), \tag{16}$$

where $\theta_1, \theta_2, \theta_3$ are hyperparameters, and $\delta(x,x')$ is the Kronecker delta for noise. In a *multi-output* setting, we extend this kernel to model both *within-channel* and *cross-channel* covariances, often via a block-structured approach (e.g. $K_{\text{MOGP}} = B \otimes K_{\text{RBF}}$, where B is an $N\times N$ matrix of inter-channel correlations).

2) MOGP Posterior for Single-Step and Multi-Step Forecasts:

a) MOGP Prior: We place a joint GP prior over all N latent functions $\{f_i(\cdot)\}$ corresponding to the N channels:

$$\begin{bmatrix} f_1(\mathbf{X}) \\ f_2(\mathbf{X}) \\ \vdots \\ f_N(\mathbf{X}) \end{bmatrix} \;\sim\; \mathcal{GP}\Big(\mathbf{0},\, K_{\text{MOGP}}(\mathbf{X},\mathbf{X})\Big). \tag{17}$$

Observations $\mathbf{Y}$ are then linked to these $f_i(\mathbf{X})$ values (e.g. via additive Gaussian noise).

b) Single-Step Prediction: To predict the next time point $t+1$ for each channel i, we collect all available data up to t. Let $\mathbf{X} = \{x_1, \dots, x_t\}$ and $\mathbf{Y}_{\text{train}}$ be the corresponding outputs. Under the MOGP prior, we form the joint Gaussian $\{\mathbf{Y}_{\text{train}},\, f(x_{t+1})\}$ and condition on $\mathbf{Y}_{\text{train}}$. The resulting posterior mean for channel i is[13]:

$$\hat{y}_i(t+1) \;=\; \mu_i\big(x_{t+1}\big), \quad i = 1, \dots, L, \tag{18}$$

where $\mu_i(\cdot)$ is the i-th component of the MOGP posterior.

c) Multi-Step (K-Step) Prediction: To forecast K future points $\{t+1, \dots, t+K\}$ for each channel, we consider two main strategies.

Iterative (Recursive) Approach: We repeatedly use predicted values as though they are observed for the next step:

$$\begin{aligned} &\hat{y}_i(t+1) = \mu_i\big(x_{t+1}\big), \\ &\hat{y}_i(t+2) = \mu_i\big(x_{t+2} \mid \hat{y}_i(t+1)\big), \\ &\hat{y}_i(t+3) = \mu_i\big(x_{t+3} \mid \hat{y}_i(t+1),\, \hat{y}_i(t+2)\big), \\ &\quad \dots, i = 1, \dots, L, \end{aligned} \tag{19}$$

This simple approach can accumulate errors but allows easy updates for adjacency-sum channels. Inspired by an Eq. (21)-style logic, we can define partial sums or increments. For instance, if

$$\widetilde{y}_i\big(t+k\big) \;=\; f_{i+1}\big(\mathbf{X}^*\big) \quad \text{for } k = 0, \tag{20}$$

and

$$\widetilde{y}_i(t+k) \;=\; f_{i+2}(\mathbf{X}^*) \;-\; f_{i+1}(\mathbf{X}^*) \quad \text{for } k = 1, \dots, K) \tag{21}$$

then each new adjacency-sum is obtained by subtracting the previously predicted partial sum. Meanwhile, the final predicted channel values are:

$$\hat{y}_i(t+k) \;=\; \mu_i\big(x_{t+k}\big), \tag{22}$$

ensuring consistency among all signals as they roll forward in time.

Direct Joint Approach: Instead of predicting each future time point separately, we form a single joint Gaussian over $\{x_{t+1}, \dots, \mathbf{x}_{t+K}\}$ for all N channels[13], [25]:

$$\begin{bmatrix} \mathbf{Y}_{\text{train}} \\ \mathbf{F}_* \end{bmatrix} \;\sim\; \mathcal{N}\Big(\mathbf{0},\, \begin{bmatrix} K_{\text{MOGP}}(\mathbf{X},\mathbf{X}) & K_{\text{MOGP}}(\mathbf{X},\mathbf{X}_*) \\ K_{\text{MOGP}}(\mathbf{X}_*,\mathbf{X}) & K_{\text{MOGP}}(\mathbf{X}_*,\mathbf{X}_*) \end{bmatrix}\Big). \tag{23}$$

where $\mathbf{X}_* = \{x_{t+1}, \dots, x_{t+K}\}$ and $\mathbf{F}_* \in R^{N\times K}$ denotes the unknown function values $\{f_i(x_{t+k})\}$. Conditioning on $\mathbf{Y}_{\text{train}}$ produces a posterior with mean $\hat{\mathbf{F}}_*$. Its (i,m)-th entry gives $\hat{y}_i(t+m)$. This one-shot method often yields better long-horizon accuracy but incurs higher computational cost due to the larger covariance block.

Overall, both approaches yield $\{\hat{y}_i(t+k) \mid i = 1, \dots, N;\, k = 1, \dots, K\}$. The iterative method is simpler yet can accumulate errors; the direct method jointly captures cross-step correlations but is more expensive. In practice, adjacency-sum signals can still be integrated into either approach by updating or subtracting previously predicted sums as new time steps unfold.

3) Recap and Advantages: By modeling all channels (original + augmented) within a unified MOGP, we exploit *inter-channel* correlations and *different temporal scales* simultaneously. This often reduces uncertainty in multi-step forecasting, compared to treating each channel as an independent single-output GP. Hence, our MOGP framework is well-suited to network traffic scenarios, where multiple correlated signals (e.g. raw measurements and adjacent-sum signals) provide complementary information for bandwidth prediction.

G. MOGP-Based Framework for Multi-Step Traffic Prediction

Fig. 4 presents a schematic of our Multi-Output Gaussian Process (MOGP) framework for multi-step bandwidth prediction. Building on the earlier steps of selecting highly correlated signals and generating adjacency-sum augmentations, this framework highlights three major components:

- Adjacency-Sum Channels (green blocks). From each original signal S_i (or S_j), we construct multiple augmented signals $S^{(w)}_{i,\text{aug}}$ by summing adjacent data points with different window sizes w. These channels capture various temporal scales (e.g. short-term bursts for $w = 2$ or medium-range trends for $w = 3, 4, \dots$).
- Block-Structured MOGP Inference (blue boxes). All channels (original + augmented) are modeled jointly using a multi-output Gaussian Process. The block-structured kernel encodes both temporal dependencies (within each channel) and cross-channel correlations (across original and augmented signals). From this MOGP, we obtain a predictive mean $\mu(\mathbf{X}^*)$ and covariance $\Sigma(\mathbf{X}^*)$ for the forecast horizon K.

- Partial-Sum or Difference-Based Recovery (bottom). In multi-step prediction, each adjacency-sum channel (e.g. $S^{(2)}_{i,\text{aug}}$) can be used to recover the single-step prediction $\hat{y}_i(t+k)$ by subtracting already-predicted values. For instance,

$$\begin{aligned}\hat{y}_i(t+2) &= \hat{S}^{(2)}_{i,\text{aug}}(t+1) - \hat{y}_i(t+1),\\ \hat{y}_i(t+3) &= \hat{S}^{(3)}_{i,\text{aug}}(t+2) - \hat{S}^{(2)}_{i,\text{aug}}(t+2),\\ &\dots\end{aligned} \tag{24}$$

  thus ensuring consistency among predicted sums and the underlying pointwise forecasts.

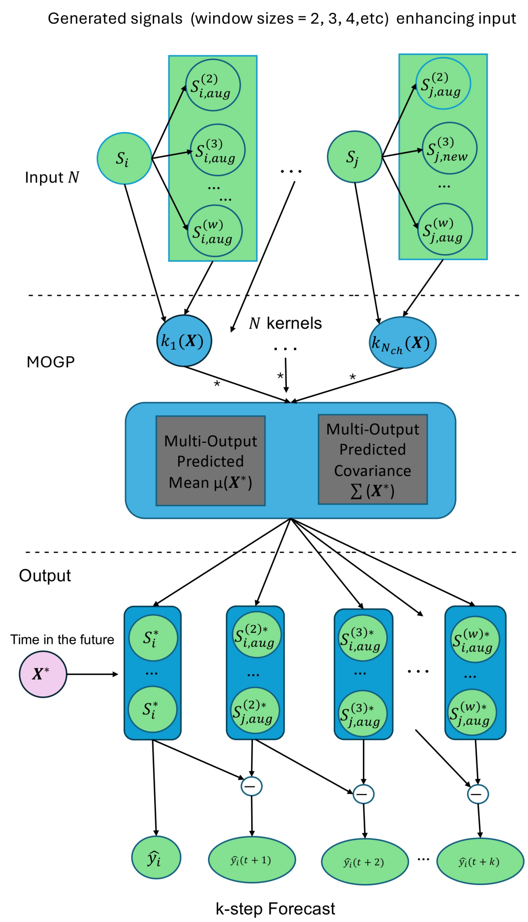

Fig. 4. The MOGP-based framework for multi-step forecasting. After generating adjacency-sum augmentations (green), a multi-output Gaussian Process (blue) provides joint predictions, which are then combined or subtracted to yield final outputs $\hat{y}_i(t+k)$.

a) Multi-Step Prediction: As discussed above, once the MOGP posterior is obtained, there are two main ways to forecast K time steps ahead:

- Iterative Approach: predict $\hat{y}_i(t+1)$, feed it back into the adjacency-sum channels, then predict $\hat{y}_i(t+2)$, and so forth. $\hat{y}_i(t+3)$ can be predicted in a similar manner until $\hat{y}_i(t+k)$.
- Direct (Block) Approach: form a single joint Gaussian over $\{\mathbf{x}_{t+1}, \dots, \mathbf{x}_{t+K}\}$ and solve for all $\hat{y}_i(t+k)$ in one shot, often achieving better accuracy (at a higher computational cost).

b) Advantages at Different Scales: By integrating adjacency-sum channels with an MOGP, the framework:

- *Captures multiple timescales* via small vs. larger summation windows,
- *Leverages cross-channel correlations*, reducing prediction uncertainty,
- *Yields robust multi-step forecasts*, as each newly predicted value can be consistently used to update the adjacency sums in subsequent steps.

In essence, this joint modeling of raw and augmented signals enables more accurate and stable bandwidth predictions across various forecasting horizons.

IV. Introduction to Experimental Model

This section provides background on the WIDE Project, describes the MAWI dataset used for our experiments, and presents an overview of traffic patterns and inter-signal correlations. The goal is to illustrate the data sources and motivations that underpin the MOGP-based forecasting framework described in Section III.

A. WIDE Project

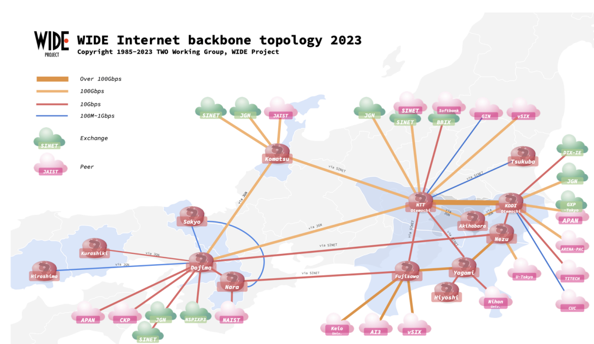

Fig. 5. 2023 WIDE backbone topology.

The Widely Integrated Distributed Environment (WIDE Project[1]) is a Japanese initiative focused on advancing Internet infrastructure. It involves a broad consortium of academic and research institutions that collaborate on designing, standardizing, and operating critical Internet technologies. Over the past decades, the WIDE Project has significantly contributed to IPv6 deployment, mobile Internet, and security innovations, and it continues to foster global partnerships among engineers and researchers.

As of 2023, the WIDE backbone (Fig. 5) consists of a Layer 2 and Layer 3 network interconnecting various Japanese and international sites, including San Francisco and Bangkok. This infrastructure provides a rich environment for collecting and analyzing real-world network traffic, which is crucial for exploring predictive modeling techniques such as Gaussian Process (GP) forecasting.

[1] WIDE Project: http://www.wide.ad.jp/.

1) MAWI Dataset Overview: The MAWI[2] (Measurement and Analysis on the WIDE Internet) group, a key part of the WIDE Project, collects and publicly shares traffic data from global ISP links. In particular, this study uses MAWI's traffic records from 2024/5/17, sampled at 10-minute intervals. As illustrated in Fig. 6, the dataset contains seven major aggregated flow signals and one time-stamp column, totaling eight columns in all. Each row represents a 10-minute observation, leading to 144 potential data points daily, though our final curated dataset has 143 valid entries after cleaning.

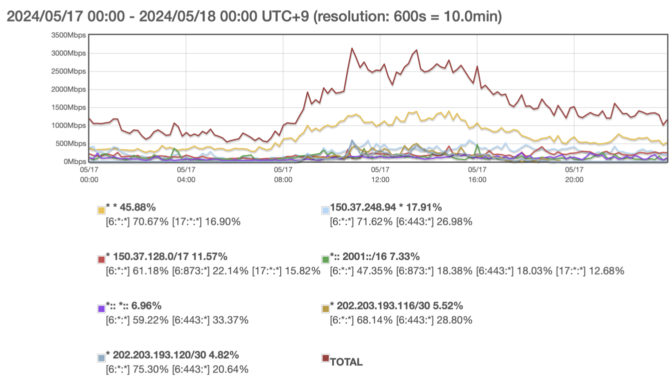

Fig. 6. Traffic data from 2024/5/17, courtesy of the MAWI dataset.

a) Data Preprocessing: We remove rows with missing values and normalize each signal to zero mean and unit variance. The time column is re-indexed to facilitate chronological analysis (e.g. minute-based or 10-minute-based indexing). This ensures that all signals are on a comparable scale before feeding them into the Multi-Output Gaussian Process (MOGP) model.

2) Traffic Pattern Analysis: A preliminary inspection of the MAWI traffic from 2024/5/17 reveals distinct diurnal patterns. Communication volumes begin to rise around 8 AM, peak during midday, and gradually decline overnight. Fig. 6 (from Section III)demonstrates these daily fluctuations. Such periodic behaviors indicate temporal correlations within each signal.

Furthermore, external factors like time zone synchronization across regions can induce correlated traffic surges among different signals. Leveraging these temporal and inter-signal dependencies is central to improving forecast accuracy in network traffic models.

3) Correlation Analysis: To quantify inter-signal relationships, we perform pairwise correlation among the seven flow signals. The resulting correlation matrix (Fig. 7) highlights various strengths of linear association. Signals that exhibit high pairwise correlation are prioritized for the MOGP input, as multi-output GP can exploit these cross-channel correlations more effectively.

The Pearson correlation coefficient, as defined in Eq. (4), is used to measure the strength of the linear relationship between two signals.

[2]MAWI Working Group Traffic Archive: https://mawi.wide.ad.jp/mawi/.

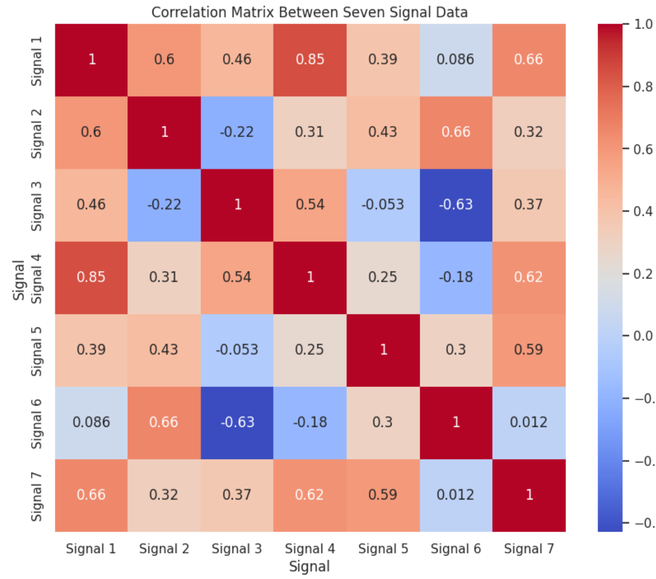

Fig. 7. Correlation matrix between seven MAWI signals, illustrating strong pairs for model input.

B. Data Setting and GP Models

Having introduced the WIDE Project and the MAWI dataset, we now discuss the GP model setup for our experiments. The experimental conditions are as follows:

- Training/Testing Split: We utilize the first T_{train} points (e.g. day 1's data) to train the GP model and reserve the subsequent points (e.g. final few observations) for testing.
- Window Size (Sliding Window): A fixed window size window_size is chosen (e.g. 50). In each step, the model sees $\mathbf{X}_{\text{train}}$ (time indices or features) and $\mathbf{y}_{\text{train}}$ (normalized signal values), then predicts the next point. The window slides forward by one time step iteratively.
- Multiple Channels (N=1,2,8): We compare three main GP setups:
  1) Single-Output GP ($N = 1$): Only one original signal (e.g. `signal1`) is modeled. This captures broad temporal trends but does not exploit cross-signal information.
  2) MOGP(2,0): We pick two highly correlated signals (e.g. $\mathbf{S}_p$ and $\mathbf{S}_q$), as identified by Pearson correlation. Then both signals are jointly modeled in a MOGP setting. This allows cross-channel correlation to inform the prediction, often improving accuracy compared to single-output.
  3) MOGP(2,6): We take two original signals $(\mathbf{S}_p, \mathbf{S}_q)$ plus their six adjacent-sum signals with window sizes $w = 2, 3, 4$. Hence, $N = 2 + 2 \times 3 = 8$ total input channels. By providing multi-scale features (short-term sums, medium-term sums, etc.), the MOGP(2,6) model can capture both short-range fluctua-

tions and daily-scale patterns across the correlated signals.

- Single-Step vs. Multi-Step Forecast:
  - Single-Step (1-step): Predict y_{t+1} given data up to time t as described in Eq. (18).
  - Multi-Step (K-step): Predict $\{y_{t+1}, \ldots, y_{t+K}\}$ either iteratively using Eq. (19) or via a direct block-covariance approach.

a) Summary of Experimental Model: Overall, the data feeding process, combined with the MOGP approach, is illustrated in Fig. 1. We use real traffic from the WIDE Project's MAWI dataset, apply standard preprocessing, select correlated signals, generate additional adjacent-sum signals if needed, and then run sliding-window GP predictions. We will perform performance metrics (RMSE) for single output versus multioutput comparisons, demonstrating how correlation exploitation and multi-scale features can enhance forecast accuracy.

1) Single-output Gaussian Process: In this scenario, only one signal is used as input for the prediction. The model captures the overall trends but struggles to predict sharp fluctuations due to the limited input information. Fig. 8 demonstrates the predicted results along with confidence intervals, showcasing the performance of the single-output GPR model for future-step prediction.

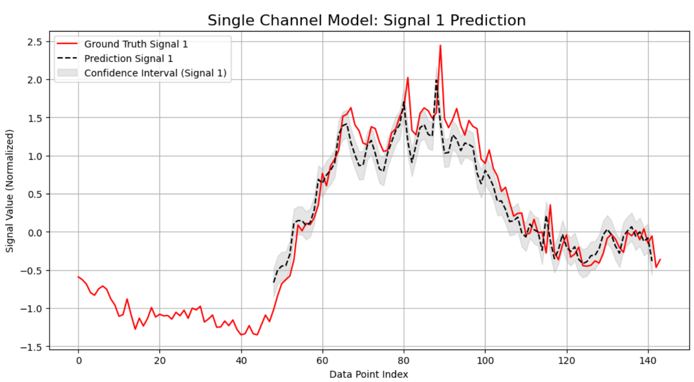

Fig. 8. Gaussian process prediction of signal trends with confidence intervals.

2) MOGP(2,0): Single-Step Prediction: In this setup, two highly correlated signals are used as inputs to the Multi-Output Gaussian Process (MOGP) model (here, MOGP(2,0) indicates two original signals and zero augmentations). This approach leverages inter-channel correlations to enhance the accuracy of predictions. Unlike the single-output model, the MOGP(2,0) model demonstrates improved robustness in capturing signal variations. Fig. 9 illustrates the predictive performance of the MOGP(2,0) model when considering two input signals, highlighting the advantage of incorporating multiple correlated channels.

3) MOGP(2,6): Single-Step Prediction: This model extends the input set by incorporating two original signals and their corresponding generated signals at varying temporal scales ($w = 2, 3, 4$), resulting in a total of eight input signals (2 original + 6 augmented). The inclusion of generated signals captures both short-term fluctuations and long-term patterns, significantly improving the accuracy of prediction. Fig. 10

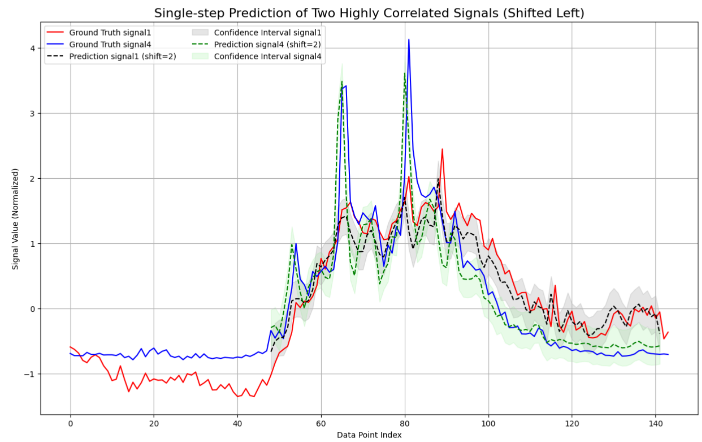

Fig. 9. Prediction results using the MOGP(2,0) model for single-step forecasting.

shows the results of the MOGP(2,6) model, demonstrating its superior predictive capacity compared to the single output and MOGP (2,0) models.

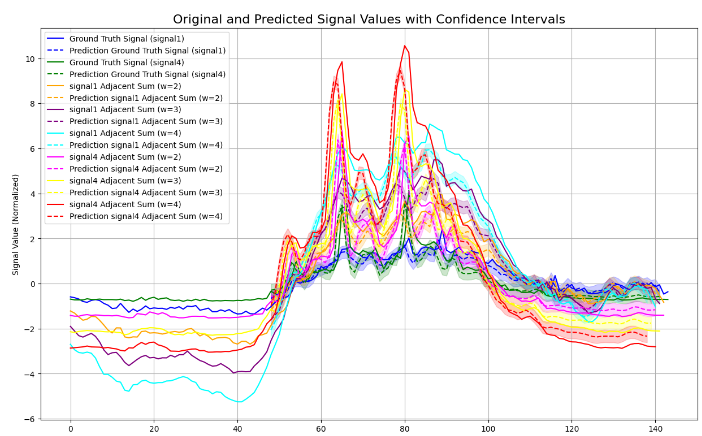

Fig. 10. Enhanced prediction accuracy using the MOGP(2,6) model for single-step forecasting.

4) MOGP(2,0): Multi-Step Prediction: In this experiment, the MOGP(2,0) model is extended to predict three future steps. This approach demonstrates the capability of the model to handle sequential forecasting tasks, utilizing inter-channel correlations effectively. Fig. 11 illustrates the predictive performance of the MOGP(2,0) model for three-step predictions.

5) MOGP(2,6): Multi-Step Prediction: The MOGP (2,6) model is used to predict three future steps, further leveraging the generated signals across varying temporal scales ($w = 2, 3, 4$). By incorporating more input signals, the model captures intricate temporal dependencies, enhancing prediction accuracy. Fig. 12 shows the prediction results for the MOGP(2,6) model.

V. Simulation Results and Discussion

A. Comparison and Summary of Experimental Results

The performance of the proposed models was evaluated using the Root Mean Squared Error (RMSE) metric, defined

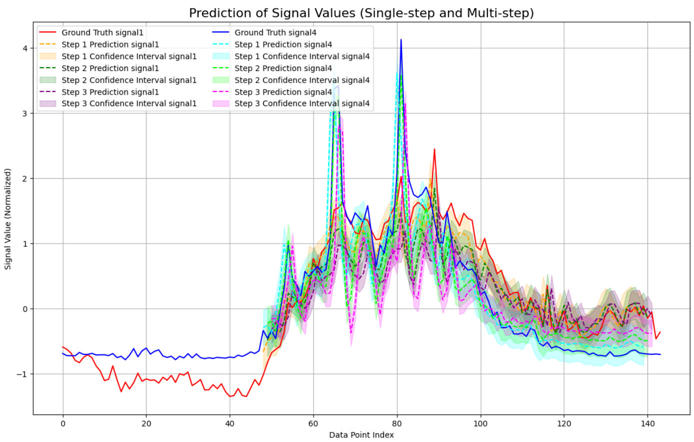

Fig. 11. Prediction results using the MOGP(2,0) model for three-step forecasting.

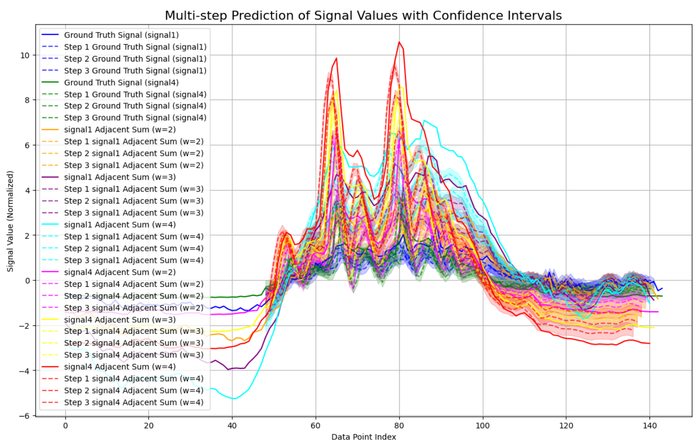

Fig. 12. Enhanced prediction accuracy using the MOGP(2,6) model for three-step forecasting.

as [26]:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(\hat{y} - y_{\text{test}})^2}$$

where $\hat{y}$ represents the predicted values for the unknown input signal x_{test}, and y_{test} denotes the corresponding ground truth values.

Table I compares the RMSE of single-output, MOGP(2,0), and MOGP(2,6) models for predicting network traffic signals.

TABLE I. RMSE Comparison of Gaussian Process Models

Model	RMSE
Single-output Gaussian Process (1-step)	0.1967
MOGP(2,0) (1-step)	0.1552
MOGP(2,6) (1-step)	0.1273
MOGP(2,0) (3-step)	0.1943
MOGP(2,6) (3-step)	0.1607

The results in Table I demonstrate the effectiveness of the MOGP models in leveraging inter-channel correlations for prediction accuracy. The MOGP(2,6) model significantly outperforms the others, achieving the lowest RMSE value. This improvement highlights the importance of integrating generated signals with varying time scales into the prediction framework.

1) Error Analysis: Table II presents the percentage reduction in prediction error achieved by the multi-output models compared to the single-output Gaussian Process model. The error reduction underscores the effectiveness of multi-channel approaches in improving predictive performance.

TABLE II. Percentage Error Reduction by Multi-Output Models

Comparison	Error Reduction (%)
Single-output vs. MOGP(2,0) (1-step)	21.2%
Single-output vs. MOGP(2,6) (1-step)	35.3%
MOGP(2,0) (1-step) vs. MOGP(2,6) (1-step)	17.9%
MOGP(2,0) (3-step) vs. MOGP(2,6) (3-step)	17.3%

The error analysis results indicate that the multi-output models significantly reduce RMSE compared to the single-output model. MOGP(2,6), in particular, achieves substantial error reduction, confirming the advantages of utilizing generated signals to capture both short-term and long-term signal variations.

B. Comparison with Recent Methods

A closely related study by Wang *et al.* [13] has already demonstrated that Gaussian process (GP)–based forecasting can outperform statistical approaches like ARIMA as well as deep learning models such as LSTM, RC-LSTM, and RNN. In particular, their method leverages a multi-output GP framework and multi-scale adjacency-sum augmentation for multi-slot-ahead traffic prediction.

Our proposed approach builds upon and extends the work in [13] by introducing **multi-channel** input signals. Rather than relying on a single channel with adjacency-sum signals, we first select multiple original signals based on their correlations, then generate adjacency-sum augmentations for each. This strategy further exploits cross-channel relationships, allowing the model to capture subtle variations across different but related traffic flows. By doing so, we retain the core strengths of multi-output GP and adjacency-sum augmentation from [13], while enhancing predictive performance through an expanded, correlation-aware input space. Our experimental results confirm that this multi-channel extension yields higher accuracy than single-channel baselines, illustrating the practical benefits of incorporating inter-channel correlations into GP-based network traffic forecasting.

Although direct numerical comparisons are hindered by different datasets and experimental conditions, we attribute our method's strong performance to two main factors:

- *Multi-Output Modeling:* By capturing inter-channel correlations, MOGP leverages shared temporal dynamics across multiple signals—an advantage that traditional single-output methods lack.
- *Multi-Scale Adjacency-Sum Augmentation:* Generating additional signals by summing adjacent data points over varying window sizes enriches the feature space, enabling the model to capture both short-term bursts and longer-range trends more effectively.

Overall, these comparisons underscore the benefits of integrating multi-channel Gaussian Processes with adjacency-sum augmentation, offering robust predictive accuracy and principled uncertainty estimates for complex network traffic scenarios.

C. Assumptions and Limitations

This study makes primary assumptions: The selected subset of MAWI channels is sufficiently representative of typical network traffic. Although these assumptions are practical for many real-world scenarios, they may limit generalization to networks with irregular sampling patterns or strong non-stationary bursts.

Another limitation lies in the computational overhead of handling large-scale datasets. Because our approach generates multiple augmented signals for each original channel, the dimensionality grows along with the dataset size, increasing both memory usage and training time for multi-output Gaussian Processes. In future work, we plan to investigate sparsity-inducing kernels and online GP methods to better accommodate high-volume streaming data.

VI. Conclusions and Future Work

In this study, we proposed a Multi-Output Gaussian Process (MOGP) model for network traffic prediction, introducing a novel approach that integrates original input signals with additional correlated signals generated using adjacent terms. This methodology leverages inter-channel correlations and multi-scale temporal dependencies to improve prediction performance. Experimental results validated through RMSE metrics demonstrated that the MOGP model consistently outperforms single-output Gaussian Processes, showcasing robust performance across both one-step and three-step prediction tasks. The MOGP(2,6) model achieved the best overall results, reducing prediction errors by 35.3% compared to the single-output baseline. These findings confirm that incorporating correlated signals significantly enhances prediction accuracy. Furthermore, the model demonstrated consistent effectiveness in single- and multi-step forecasting scenarios, highlighting its adaptability to various temporal scales and complex datasets.

Future work will focus on exploring alternative kernel functions and hyperparameter optimization strategies to further improve prediction precision and computational efficiency. We also intend to extend our framework to 5G and IoT traffic prediction, and to real-time or large-scale environments, evaluating the model's scalability and adaptability under streaming conditions. Another promising direction is to adapt our MOGP approach for anomaly detection. Such anomalies not only consume additional network resources but also pose significant risks to overall network performance and security [27]. By exploiting the pervasive inter-channel correlations, our framework could identify deviations from typical diurnal or weekly patterns. For example, if two channels $\mathbf{S}_k$ and $\mathbf{S}_l$ with similar usage trends both show high traffic after 8 AM and taper off after 7 PM, a sudden increase in $\mathbf{S}_k$ without a corresponding change in $\mathbf{S}_l$ might signal anomalous behavior. This line of research could play a vital role in proactive network management and early threat detection.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP22K04089.

References

[1] Y. Chen, S. Jain, V. K. Adhikari, Z. L. Zhang, and K. Xu, "A first look at inter-data center traffic characteristics via yahoo! datasets," Proc. of IEEE INFOCOM 2011, pp. 1620-1628, Apr. 2011.

[2] J. Koo, V. B. Mendiratta, M. R. Rahman, and A. Walid, "Deep reinforcement learning for network slicing with heterogeneous resource requirements and time varying traffic dynamics," Proc. of IEEE CNSM 2019, pp. 1-5, Oct. 2019.

[3] A. Baiocchi, Network Traffic Engineering Stochastic Models and Applications, Hoboken, NJ, USA:Wiley, 2020.

[4] Mehrizi, S., & Chatzinotas, S. (2022). Network traffic modeling and prediction using graph Gaussian processes. IEEE Access, 10, 132644–132655.

[5] M. Beshley, et al. "End-to-End QoS 'smart queue' management algorithms and traffic prioritization mechanisms for narrow-band internet of things services in 4G/5G networks." Sensors, vol. 20, no. 8, p. 2324, 2020.

[6] C. Gijon, et al. "Long-term data traffic forecasting for network dimensioning in LTE with short time series." Electronics, vol. 10, no. 10, p. 1151, 2021.

[7] B. Zhou, D. He, Z. Sun, and W. H. Ng, "Network traffic modeling and prediction with ARIMA/GARCH," Proc. of HET-NETs 2005, pp. 1-10, Jul. 2005.

[8] A. Azari, P. Papapetrou, S. Denic and G. Peters, "Cellular traffic prediction and classification: A comparative evaluation of LSTM and ARIMA", Proc. Int. Conf. Discovery Sci., pp. 129-144, 2019.

[9] A. Azzouni and G. Pujolle, "NeuTM: A neural network-based framework for traffic matrix prediction in SDN", Proc. IEEE/IFIP Netw. Oper. Manage. Symp., pp. 1-5, Apr. 2018.

[10] A. Bayati, K.-K. Nguyen and M. Cheriet, "Gaussian process regression ensemble model for network traffic prediction", IEEE Access, vol. 8, pp. 176540-176554, 2020.

[11] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[12] Y. Wang, T. Nakachi, T. Inoue and T. Mano, "Adaptive multi-slot-ahead prediction of network traffic with Gaussian process," 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 2021, pp. 1-6, doi: 10.1109/GLOBECOM46510.2021.9685249.

[13] Y. Wang, T. Nakachi, and W. Wang, "Pattern discovery and multi-slot-ahead forecast of network traffic: A revisiting to Gaussian process," IEEE Transactions on Network and Service Management, 2022.

[14] M. van der Wilk, et al. "A framework for interdomain and multioutput Gaussian processes." arXiv preprint arXiv:2003.01115, 2020.

[15] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control. John Wiley & Sons, 2015.

[16] H. Zhang, G. Wang, J. Liu, H. Hu, and S. Liu, "Network Traffic Prediction Based on Deep Learning." IEEE Access, vol. 6, pp. 23302-23310, 2018.

[17] Z. Zhao, Y. Zhang, Y. Xu, and H. Liang, "Deep learning based network traffic prediction: Methods, datasets and analysis." IEEE Access, vol. 5, pp. 5143-5153, 2017.

[18] H. Liu, J. Cai, and Y. S. Ong, "Remarks on Multi-output Gaussian Process Regression." Knowledge-Based Systems, vol. 144, pp. 102-121, 2018.

[19] E. V. Bonilla, K. M. Chai, and C. K. Williams, "Multi-task Gaussian process prediction." In Advances in Neural Information Processing Systems, pp. 153-160, 2008.

[20] Han, W., Zhu, T., Chen, L., Ning, H., Luo, Y., & Wan, Y. MCformer: Multivariate Time Series Forecasting with Mixed-Channels Transformer. IEEE Internet of Things Journal, 2024.

[21] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation." Anesthesia & Analgesia, vol. 126, no. 5, pp. 1763-1768, 2018.

[22] H. Xu, and Y. Deng, ”Dependent evidence combination based on shearman coefficient and pearson coefficient,” IEEE Access, vol. 6, 11634-11640. 2017.

[23] D. J. MacKay et al., “Introduction to Gaussian processes,” NATO ASI series F computer and systems sciences, vol. 168, pp. 133–166, 1998.

[24] A. G. Wilson, ”Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes”, 2014.

[25] L. Yang, K. Wang, and L. Mihaylova, ”Online sparse multi-output Gaussian process regression and learning,” IEEE Transactions on Signal and Information Processing over Networks, vol.5, no. 2, pp. 258-272, 2018.

[26] T. Chai, and R. R. Draxler, ”Root mean square error (RMSE) or mean absolute error (MAE).” Geoscientific Model Development Discussions, vol. 7, no. 1, pp. 1525-1534, 2014.

[27] Y. Wang, T. Nakachi, ”Network traffic anomaly detection: A revisiting to Gaussian process and sparse representation,” IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. 107, no. 1, pp. 125-133, 2024.