Gaussian Process for Flight Delay Prediction: Learning a Stochastic Process

Aakarshan Khanal *, Rajnish Bhusal †, Kamesh Subbarao‡, Animesh Chakravarthy§
The University of Texas at Arlington, Arlington, TX 76019, USA

Wendy A. Okolo¶
NASA Ames Research Center, Moffett Field, CA 94035, USA

This paper presents a machine-learning approach to predict flight delays. Whereas neural networks are extensively studied for predictive capabilities, they involve non-intuitive design and extensive analysis, particularly in training and optimization processes. Instead, the proposed framework employs Gaussian Processes as a supervised learning technique for flight delay prediction. This data-driven approach trains the model using prior information, specifically the mean and covariance tied to existing data. The proposed Gaussian Process Regression (GPR) model employs the day of flight as a pivotal feature for delay forecasting. We analyze flights from various routes and gauge the accuracy of the presented learning technique by comparing the predicted delays with the actual ones. Given the inherent challenges in precisely forecasting delays, we predict the delays with a 95 % confidence interval. Also, an error propagation analysis in the prediction horizon is carried out to determine the optimal time frame for prediction. The proposed method for flight delay prediction is important as airlines can strategize flight operations and issue timely advisories.

I. Introduction

In recent years, the volume of aircraft flying within global airspace has increased notably. The significant impact of this substantial increase is most evident within the United States national airspace. In addition, flight delays are more frequent, potentially affecting the reputation and business of airlines [1]. The most common types of delays include carrier delays, weather delays, security delays, and airspace-related delays. Carrier delays are attributable to the airline's management, including delays resulting from maintenance, flight crews, aircraft refueling, etc. Similarly, weather delays arise from extreme weather conditions, airspace delays emanate from congested national airspace, air traffic controller delays, and low visibility impacting aerodrome operation, while security delays are induced by security-related issues.

Presented at AIAA SciTech Forum, 23-27 January 2023, National Harbor, MD, AIAA 2023-1257
*Ph.D. Student, Department of Mechanical and Aerospace Engineering, axk0129@mavs.uta.edu, AIAA Student Member
†Research Engineer Scientist, Department of Mechanical and Aerospace Engineering, rajnish.bhusal@mavs.uta.edu, AIAA Member
‡Professor, Department of Mechanical and Aerospace Engineering, subbarao@uta.edu, Associate Fellow (AIAA)
§Professor, Department of Mechanical and Aerospace Engineering, animesh.chakravarthy@uta.edu, Associate Fellow (AIAA)
¶Aerospace Research Engineer, Intelligent Systems Division, wendy.a.okolo@nasa.gov, AIAA Member

The United States Department of Transportation (DOT), Bureau of Transportation Statistics (BTS) publishes the monthly Air Travel Consumer Report with details on all flights, including the arrivals and departures, delays, cancellations, and diversions. As per the report of 2022, only 76.72 % of flights in the U.S. arrived on time, with 22.81% of the flights delayed at Dallas/Fort Worth International Airport [2]. Therefore, using a flight delay prediction model, airlines can have prior information on delays and develop operational plans to mitigate the underlying causes within their control.

Flight delay, a stochastic process, is affected by random factors like weather, air traffic, etc. These uncertainties cause an impact on flight delays and are inherent. We can employ a mathematical model to quantify and propagate uncertainty within a stochastic process. A mathematical model describing the flight operation is always approximate, and evaluating such a mathematical model can be practical depending on the ability to quantify its uncertainties [3]. It involves randomly sampling uncertain parameters from a probability distribution and propagating the uncertainty. However, the results can be inaccurate if the assumed probability differs from the actual probability distribution. Furthermore, if numerous uncertain variables govern the system, solving mathematical models with Stochastic Differential Equations (SDEs) can be computationally costly. Alternatively, we can consider flight delay prediction as a machine-learning problem. Machine learning refers to methods or algorithms that can "learn" from available information, enabling them to carry out desired tasks [4]. In supervised learning, for instance, we can predict future delays based on prior data on flight schedules and delays.

In the literature, various machine-learning techniques have been implemented for flight delay prediction. Gui et al. [5], used Long Short-Term Memory (LSTM) neural networks and random forest-based models for predicting flight delays using big data. For a binary classification problem, the random forest-based model had a prediction accuracy of 90.2 % . Kim et al. [6] employed Recurrent Neural Networks (RNNs) to predict air traffic delays. Similarly, Yu et al. [7] studied patterns in flight delays by implementing deep learning to large datasets. Esmaeilzadeh et al. [8] evaluated interdependence between flight delays using a Support Vector Machine (SVM) regression model. However, very few research in the literature typically considers the uncertainty associated with the predictions. Mueller et al. [9] modeled the delay-time probability density function using a statistical approach and showed that departure delay fits better with Poisson distribution and arrival delays with Normal distribution. Similarly, Tu et al. [10] estimated departure delay distribution by implementing a statistical approach, and Wesonga et al. [11] used a logistic model to estimate the probability of aircraft delays at Entebbe International Airport. Perez-Rodriguez et al. [12] implemented an asymmetric Bayesian logit model to determine aircraft delay probability. The study showed a strong correlation between the departure and arrival delay for a particular flight. Using Bayesian RNN-LSTM, Vandal et al. [13] predicted aggregate delays in US airports and also implemented the Monte-Carlo Dropout technique for uncertainty quantification. The analysis of the model showed difficulty in predicting flight delays where the RSME was above 10 minutes. Zoutendijk et al. [14] applied Mixture Density Networks and Random Forest Regression to predict flight delays. However, no prediction results are shown for multiple flight days as a time series. Also, weather information is considered a feature

that, for a future flight day itself, depends on a prediction model. Similarly, Wang et al. [15] proposed machine learning techniques to predict the distribution of delays. But this paper also presented the results for a single flight day. Our work stands out from others in its unique focus on predicting flight delays for multiple days for a specific route and a particular flight, whereas most existing work has concentrated on predicting delays at specific airports. Moreover, as airlines frequently change their flight number, this study focuses on using compact training datasets because data for a particular flight are often limited in size. This methodology can give more precise predictions regarding individual flights and routes.

This research applies machine learning to predict flight delays based on prior flight schedules and delay information. The main objective of this research is to learn the trend and periodicity in training data and use them for future delay prediction. While different machine learning methods provide regression models, they don't have the capability to give the uncertainty of their predictions. Predicting flight delays is a critical task, and accounting for uncertainty in these predictions is equally important for the reliability of prediction. Hence, it is essential to implement a machine learning method that can give a prediction and measure of confidence. Gaussian Process Regression (GPR), a probabilistic machine learning framework, has gained significant importance in recent years due to its flexibility, and effectiveness in addressing stochastic problems. This differs from the predominant utilization of more complex methods like neural networks for similar tasks, while also allowing us to calculate the degree of confidence in our predictions. Gaussian Process (GP) has been used in various applications, including uncertainty quantification in aerodynamics [16], model predictive control of quadrotor [17], forecasting time series data [18], and predicting travel time for a vehicle based on traffic data [19]. Furthermore, Amer et al. [20] integrated GPR with other methods for probabilistic damage detection and structural health monitoring, and Mohanty et al. [21] implemented GP to predict the fatigue life of the material. For predicting flight delays with a machine learning technique, working directly with a large dataset is daunting, and it is important to first gain a proper understanding of the correct approach to tackle the problem. The selection of kernels and identifying important features is a critical aspect of GPR. Thus, we can first apply the technique to a familiar problem, such as a periodically excited time series, to understand how kernels capture the system with trends and periodicity, characteristics that are also being examined in flight delay prediction, ensuring a more accurate application of GPR to the task at hand.

The structure of this paper is outlined as follows. Section II discusses the mathematical foundations of the GP, and Section III gives a concise overview of the GPR. In Section IV, an application of GPR to a time series with periodic excitation is discussed. Section V focuses on the application of GPR for flight delay prediction. The meaningful discussion of results is given in Section VI, followed by the conclusion in Section VII. Finally, in Section VIII, the paper explores the limitations of the proposed method and future work..

II. Mathematical Preliminaries

In this section, we provided the mathematical concepts to understand stochastic and Gaussian Processes.

A. Gaussian Random Variable, Multivariate Distribution, and Gaussian Random Vectors

A random variable is defined as a variable whose value depends on a random phenomenon. Any random variable X is said to be Gaussian or normally distributed if its probability density function, $f_X(x)$, is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \text{ for all } x \in \mathbb{R} \tag{1}$$

where $\mu = \mathbb{E}[X]$ is the mean or expected value and σ^2 is the variance of X.

Multivariate distribution is a distribution of multiple random variables. The correlation among the random variables is generally represented by a symmetric, positive semi-definite matrix known as the covariance matrix. In the covariance matrix, the off-diagonal elements represent the covariance between different random variables while the diagonal element represent the variances of random variables. To that end, let us denote the covariance matrix among X_i, $i = 1, 2 \ldots, n$ random variables with $\mathbf{K} = \mathbf{K}^{\mathrm{T}} \in \mathbb{R}^{n\times n}$. For all $i, j = 1, 2, \ldots, n$, the (i, j) element of $\mathbf{K}$ represents the covariance between X_i and X_j denoted by $\mathrm{cov}[X_i, X_j] = \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right]$ such that

$$\mathbf{K} = \begin{bmatrix} \mathrm{cov}[X_1, X_1] & \mathrm{cov}[X_1, X_2] & \cdots & \mathrm{cov}[X_1, X_n] \\ \mathrm{cov}[X_2, X_1] & \mathrm{cov}[X_2, X_2] & \cdots & \mathrm{cov}[X_2, X_n] \\ \vdots & \vdots & \cdots & \vdots \\ \mathrm{cov}[X_n, X_1] & \mathrm{cov}[X_n, X_2] & \cdots & \mathrm{cov}[X_n, X_n] \end{bmatrix} \tag{2}$$

Now, let us consider a random vector $\mathbf{X} = [X_1, X_2, ..., X_n]^T \in \mathbb{R}^n$ and let the distribution of $\mathbf{X}$ be multivariate Gaussian. Then its probability density function is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n\,|\mathbf{K}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{3}$$

where $\boldsymbol{\mu} = [\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_n]]^{\mathrm{T}}$ is the mean vector and $\mathbf{K}$ as defined in Eq. (2) is the covariance matrix of $\mathbf{X}$. In Eq. (3), $|\cdot|$ denotes the determinant operator. The $\sqrt{|\mathbf{K}|}$ ensures that the total volume under the probability density function is one, thus making it a valid probability density function.

B. Gaussian Processes

Gaussian Processes (GPs) are stochastic processes that can be used to model continuous functions and, in addition, are a generalization of the multivariate normal distribution to infinite dimensions [22]. A function $f : \mathbb{R}^n \to \mathbb{R}$ is a

sample drawn from a GP denoted as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}')) \tag{4}$$

if the probability density function of f at any input $\mathbf{x} \in \mathbb{R}^n$ is jointly Gaussian. In Eq. (4), $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$ is the mean function or the expected value of functions present in GP evaluated at $\mathbf{x}$ and $\mathbf{K}(\mathbf{x}, \mathbf{x}')$ is the covariance function given by

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = \mathbb{E}\left[\left(f(\mathbf{x}) - \mu(\mathbf{x})\right)\left(f(\mathbf{x}') - \mu(\mathbf{x}')\right)\right] \tag{5}$$

The covariance function $\mathbf{K} : \mathbb{R}^n \times \mathbb{R}^n$ maps two vector inputs to a scalar output. It is a measure of the similarity between two input vectors. In the GP, kernel functions $\mathbf{K}(\mathbf{x}, \mathbf{x}')$ are used to specify the covariance function [23] and therefore capture the similarity between inputs $\mathbf{x}$ and $\mathbf{x}'$. To that end, a larger value for $\mathbf{K}(\mathbf{x}, \mathbf{x}')$ corresponds to the high correlation between the GP outputs $f(\mathbf{x})$ and $f(\mathbf{x}')$ and thus, these outputs tend to be similar for both inputs $\mathbf{x}$ and $\mathbf{x}'$ [22]. Further, the covariance matrix $\mathbf{K}$ associated with the covariance function must be positive semi-definite [24]. A symmetric matrix $\mathbf{M}$ is said to be positive semi-definite if all the eigenvalues are non-negative. There are several covariance functions for selection, and we can also develop new covariance functions using a combination of existing covariance functions as required. The choice of covariance function depends on the characteristics of the model, such as how smooth the model is, how scattered are the model data, and how unpredictable the model is [25].

III. Learning Time Series Patterns Using Gaussian Process Regression

It is not always feasible to represent a time series with mathematical models because of challenges in formulating such models. In these scenarios, understanding the underlying trends using machine learning becomes important for forecasting. Given the inherent uncertainty in such data, predicting a point value may not be meaningful. Therefore, it's necessary to implement the techniques that give predictions along with associated uncertainties. In the context of time series analysis, the non-parametric characteristic and adaptability to a diverse range of time series of GPR offer notable advantages. A notable feature of GPR is its capability to provide a predictive distribution rather than just a point prediction, which is important for uncertainty quantification. Numerous works in the literature have studied applying GPR for time series forecasting. Yadav et al. [26] implemented GPR for electric load forecasting and demonstrated the superior performance of GPR as compared to Decision Trees and Artificial Neural Networks. Liu et al. [27] applied GPR to predict the volatility of foreign exchange returns. In this work, both non-coregionalized and coregionalized GP were implemented to reduce predictive uncertainty. The GPR is a suitable approach for the nature of the task at hand, predicting flight delays. This is relevant as we need to quantify uncertainty and recognize trends and periodicity inherent in the dataset – capabilities that GPR is well-equipped to handle.

In linear regression techniques, the main objective is to relate a function $f(x)$ to a specific model (e.g., $f(x) = mx+c$)).

In GPR, the function $f(x)$ is not specific and allows data to define the model [28]. GPR is a non-parametric Bayesian approach where target variables are assumed to be random and extracted from the probability distribution. It is a non-parametric approach because the parameters increase with training data. A non-parametric model though computationally complex, is very flexible for data fitting [29]. Let $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ denotes a vector of predictor variables, $\mathbf{y} = [y_1, y_2, ..., y_n]^T$ denotes a vector of target variables, and $\mathbf{X}$ denotes all the observed data associated with predictor variables. Then, the probability distribution of the latent function f^* at new points X^*, conditioned on the observed data $\mathbf{X}$ is specified by Bayes theorem as follows [30]:

$$p(f^*|\mathbf{X}, \mathbf{y}, \mathbf{X}*) = \frac{p(\mathbf{y}|f)p(f, f^*|\mathbf{X}, \mathbf{X}^*)}{p(\mathbf{y}|\mathbf{X})} \tag{6}$$

where $p(\mathbf{y}|f)$ represents likelihood of the observed output $\mathbf{y}$ given the latent function f, $p(f, f^*|\mathbf{X}, \mathbf{X}^*)$ represents prior probability and $p(\mathbf{y}|\mathbf{X})$ is normalization constant. Marginal probability is computed by integrating over all the possible latent function values f.

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|f)p(f|\mathbf{X})\, df \tag{7}$$

In GPR, for Gaussian likelihood, the predictive distribution is also Gaussian. This is because any finite subsets from the GP are jointly Gaussian. For instance, let us consider a new dataset $\mathbf{X}^*$. Then the posterior predictive distribution $f(\mathbf{X}^*)$ in GP is Gaussian with

$$\begin{aligned} \mu_* &= {\mathbf{K}_*}^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{X} \\ \Sigma_* &= K(\mathbf{X}^*, \mathbf{X}^*) - {\mathbf{K}_*}^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{K}_* \end{aligned} \tag{8}$$

where μ_* is the predictive mean, Σ_* is the predictive variance, the vector $\mathbf{K}_*$ gives covariance between training data $\mathbf{X}$ and new data $\mathbf{X}^*$ and $\mathbf{K}$ is the covariance matrix of $\mathbf{X}$ with itself.

The accuracy of the GP depends on selecting the covariance function and its hyperparameter because incorrect selection can give poor results. So, hyperparameters of kernels are optimized using the maximum log marginal likelihood [24]. The log marginal likelihood is written as

$$\log p(\mathbf{y}|f, \boldsymbol{\phi}) = -\frac{1}{2}\mathbf{X}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{X} - \frac{1}{2}\log|\mathbf{K}| - \frac{d}{2}\log 2\pi \tag{9}$$

where $\mathbf{X} \in R^n$ is given data , $\boldsymbol{\phi}$ are hyperparameters and $\mathbf{K}$ is the covariance matrix. The terms in log marginal likelihood relation penalize different features [31]. The term $-\frac{1}{2}\mathbf{X}^T\mathbf{K}^{-1}\mathbf{X}$ penalizes how well the hyperparameter values currently

used fit the data. If the data fits well this term should have a small value. Similarly, the term $-\frac{1}{2}\log|\mathbf{K}|$ is the penalty for complexity and the large value of the determinant of covariance matrix represents a large volume in function space. The final term - $\frac{d}{2}\log 2\pi$ is a constant for normalization. The log marginal likelihood can be maximized by taking the partial derivative of the log marginal likelihood relation with respect to $\boldsymbol{\phi}$.

A. Kernels or Covariance Functions

Kernels are functions that describe the similarity between two points in input space. This research implements several distinct kernels, which are outlined as follows.

1. Squared Exponential (SE) Kernel

One of the most common covariance functions is the Squared Exponential, also known as the Radial Basis Function (RBF) kernel, which is given by

$$\mathbf{K}(\mathbf{x},\mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}')^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\mathbf{x}-\mathbf{x}')\right), \text{ for } \mathbf{x},\mathbf{x}' \in \mathbb{R}^n \tag{10}$$

where σ^2 is the signal variance of the function and $\boldsymbol{\Lambda} = \mathrm{diag}([l_1^2, l_2^2, \ldots, l_n^2])$ is a diagonal matrix of squared characteristic length-scales l_i, $i = 1, 2, \ldots, n$. The RBF kernels are useful for capturing non-linear patterns and trends.

2. Matern32 Kernel

The Matern32 kernel is represented as

$$\mathbf{K}(\mathbf{x},\mathbf{x}') = \sigma^2\left(1+\sqrt{3}\|\mathbf{x}-\mathbf{x}'\|_{\mathbf{l}}\right)\exp\left(-\sqrt{3}\|\mathbf{x}-\mathbf{x}'\|_{\mathbf{l}}\right), \text{ for } \mathbf{x},\mathbf{x}' \in \mathbb{R}^n \tag{11}$$

where σ^2 is the signal variance of the function, $\|\mathbf{x}-\mathbf{x}'\|_{\mathbf{l}} = \sqrt{\sum_{i=1}^{n}\left(\frac{x_i - x_i'}{l_i}\right)^2}$ is the distance metric, and $\mathbf{l}$ is length-scale vector. The Matern32 kernel is not infinitely differentiable and has the ability to represent varying levels of smoothness.

3. Exponential Sine Squared (ESS) Kernel

The Exponential Sine Squared (ESS) kernel, also called the periodic kernel, defines periodicity in prior data. The Exponential Sine Squared kernel is given by

$$\mathbf{K}(\mathbf{x},\mathbf{x}') = \exp\left(\left(-2\sin^2\left(\pi(\mathbf{x}-\mathbf{x}')^{\mathrm{T}}\mathbf{p}^{-1}(\mathbf{x}-\mathbf{x}')\right)\right)\boldsymbol{\Lambda}^{-1}\right), \text{ for } \mathbf{x},\mathbf{x}' \in \mathbb{R}^n \tag{12}$$

where σ^2 is the signal variance of the function, $\boldsymbol{\Lambda} = \mathrm{diag}([l_1^2, l_2^2, \ldots, l_n^2])$ is a diagonal matrix of squared characteristic length-scales l_i, $i = 1, 2, \ldots n$, and $\mathbf{p}$ represents the periodicity. The periodic kernel is used for learning time series data

with periodic patterns.

4. Periodic Exponential Kernel

The Periodic Exponential kernel merges the characteristics of the exponential (Laplacian) kernel with a periodic component. This kernel identifies functions that show periodic behavior, yet their amplitude diminishes or shows other changes as one moves farther from a reference point. The Periodic Exponential kernel is represented as

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = \exp\left(-\left(\mathbf{x}-\mathbf{x}'\right)^{\mathrm{T}} \mathbf{\Lambda}^{-1} \left(\mathbf{x}-\mathbf{x}'\right)\right) \cos\left(2\pi \left(\mathbf{x}-\mathbf{x}'\right)^{\mathrm{T}} \mathbf{p}^{-1} \left(\mathbf{x}-\mathbf{x}'\right)\right), \text{ for } \mathbf{x}, \mathbf{x}' \in \mathbb{R}^n \tag{13}$$

where $\mathbf{\Lambda} = \mathrm{diag}([l_1, l_2, \ldots, l_n])$ is a diagonal matrix of length-scales l_i, $i = 1, 2, \ldots n$, and $\mathbf{p}$ represents the periodicity.

5. Cosine Kernel

In Gaussian Processes, the cosine kernel is implemented to identify cyclic trends within the data and is given by

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = \cos\left(2\pi(\mathbf{x}-\mathbf{x}')^{\mathrm{T}}\mathbf{w}\right) \tag{14}$$

where $\boldsymbol{w}$ is a frequency vector in $\mathbb{R}^n$.

6. Combined Kernels

There are two ways of combining kernels to create a new kernel: Multiplication and Addition. Multiplying kernels is the same as an AND operation because the new covariance function will have high covariance only if the covariance of both functions is high. Similarly, adding kernels is equivalent to the OR operation because the new covariance function will have a high covariance if one of them has a high covariance.

B. Regression Metrics

Regression metrics are used to evaluate the performance of the regression model. It measures how well the actual and predicted values match in regression. There are various metrics used for regression based on the application. In this section, we discuss different regression metrics used in the flight delay prediction application.

1. Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) calculates the average percentage error between actual and predicted values. A more accurate model has a lower MAPE, while a higher MAPE suggests a larger error between actual and predicted values. However, it can be largely affected by outliers in prediction. The MAPE is calculated using the given formula.

$$\textbf{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{15}$$

where y_i is the actual value, and $\hat{y}_i$ is the predicted value.

2. Mean Absolute Error

Mean Absolute Error (MAE) measures the average absolute difference between the actual and predicted values. A lower MAE value indicates that the model fits well with the data. The MAE is calculated as follows.

$$\textbf{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{16}$$

where y_i is the actual value, and $\hat{y}_i$ is the predicted value.

3. Mean Squared Error

Mean Squared Error (MSE) measures the average of the squared difference between the actual and predicted values. If a model fits well with the data, the value of MSE is small. The given formula calculates the MSE.

$$\textbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{17}$$

where y_i is the actual value, and $\hat{y}_i$ is the predicted value.

4. Root Mean Squared Error

The Root Mean Square Error (RMSE) provides the square root of the average of the squared differences between the predicted values and the actual observed values. The formula for RMSE is written as

$$\textbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{18}$$

where y_i is the actual value, and $\hat{y}_i$ is the predicted value.

IV. Application of Gaussian Process Regression to Analyze Time Series with Periodic Excitation

In this section, we employed the GPR to predict a time series with periodic excitation. This is relevant for determining the appropriate kernel applicable to data having trend and periodicity, characteristics that are also being studied in flight delay prediction.

A. Methodology

For the training dataset, numerical solutions based on a mathematical model are employed [32]. The numerical solutions for a specific time period are used as training data for the GPR model. Furthermore, a seasonal decomposition technique, using the *statsmodel* Python library, is applied to decompose the time series data into three distinct components: trend, seasonal, and residual. The choice of the kernel is determined by a visual examination of these three components. The framework for prediction using GPR is shown in Fig. 1.

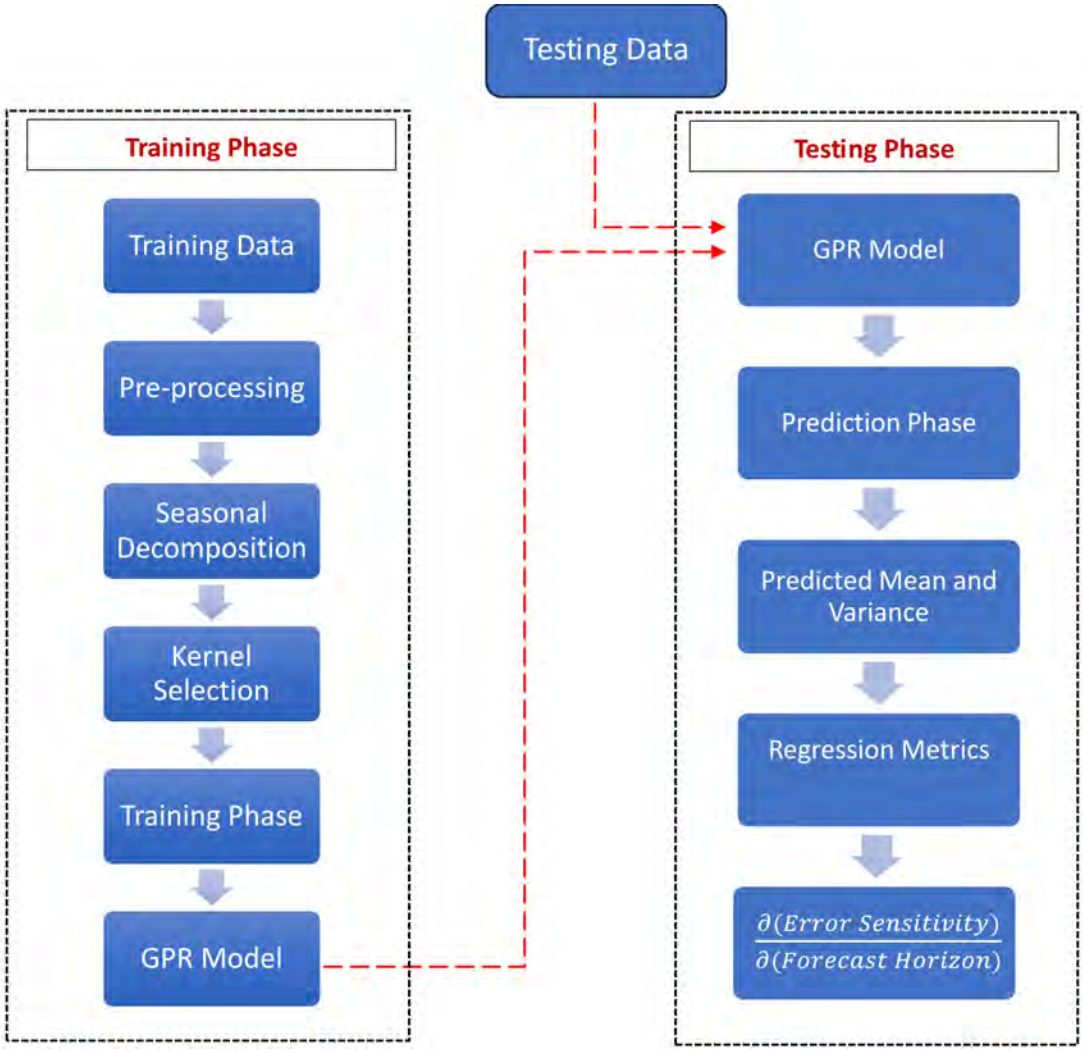


Fig. 1 Framework for Gaussian Process Regression-based prediction.

B. Predicting Time Series Outcomes with Initial Condition Uncertainty

First, we utilized seasonal decomposition to determine the suitable kernel type for the model. As shown in Fig. 2, there exists a clear trend in data with significant periodic patterns, which is obvious for a harmonic system. To achieve that, the potential kernels should display a trend, periodic cycle, or both. In Table 1, we present the performance

assessment of various kernels, including Linear, RBF, Periodic, Cosine, Matern32, (RBF + Periodic), and (RBF + Periodic Exponential). The maximum and minimum MAEs are highlighted in red and green boxes respectively. The training dataset comprised the initial ten seconds of the numerical solution. From the performance assessment results, it was observed that both RBF and (RBF + Periodic Exponential) kernels are favorable choices because of minimal MAE. The Matern32 kernel appeared to overfit the training data, with all the regression metrics having negligible values. Thus, we excluded Matern32 from subsequent analysis. We further evaluated the potential kernels for different testing horizons as shown in Table 2. The regression metrics suggest that the (RBF + Periodic Exponential) kernel has good performance over extended testing horizons. Although the MAE is lowest for a testing horizon of 5 seconds, the predicted mean appears to converge towards the mean of the dataset. This behavior, often observed during extrapolation with GPR, cannot be interpreted as a reliable prediction. Therefore, a prediction horizon of 3 seconds appeared optimal for the given case as shown in Fig. 3.

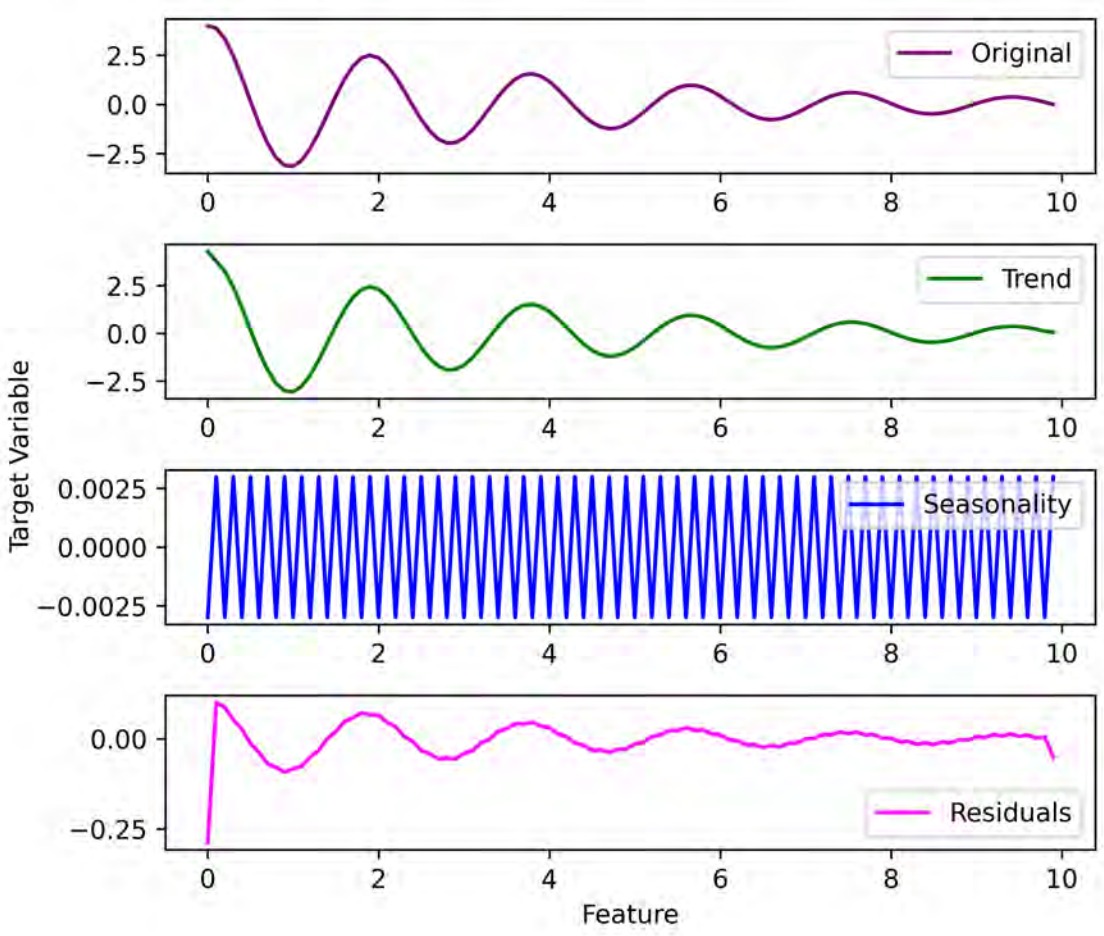


Fig. 2 Seasonal decomposition of training data subject to initial condition uncertainties.

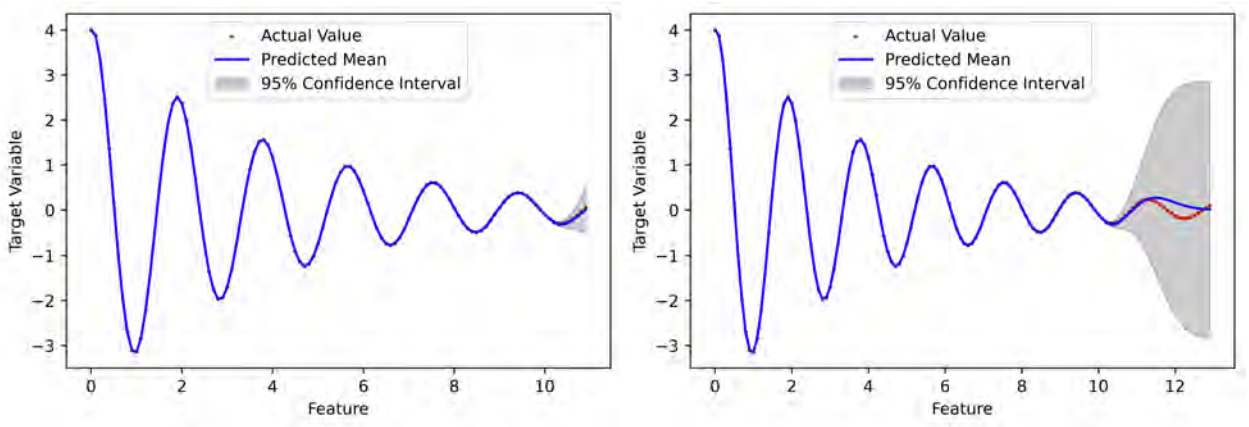
From the obtained results, it can be inferred that GPR can capture both the trend and periodicity inherent to time series data, but only up to a specific testing horizon. Hence, utilizing the proposed GPR framework, we can predict the

Table 1 Performance evaluation of various kernels on periodically excited time series using a 10-second training dataset and 1-second testing period.

Kernels	Training Data				Testing Data			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
Linear	0.9693	1.7455	1.3212	99.99	0.1847	0.0429	0.2071	100
RBF	0.0001	2.304e-08	0.00015	0.0213	0.0059	5.72e-05	0.0075	7.47
Periodic	0.9693	1.7455	1.3212	99.99	0.1847	0.0429	0.2071	100
Cosine	0.5585	0.5633	0.7505	132.55	0.7573	0.713	0.8444	426.03
Matern32	1.23e-08	1.86e-15	4.32e-08	3.557e-06	0.2084	0.0883	0.2971	305.694
RBF + Periodic	0.0001	2.314e-08	0.00015	0.02138	0.02596	0.00157	0.03966	41.07166
RBF + Per. Exponential	0.0001	2.325e-08	0.00015	0.0331	0.0181	5.9e-06	2.437e-03	24.9274

Table 2 Performance comparison of RBF and (RBF + Periodic) kernels on periodically excited time series using a 10-second training dataset on different testing horizons.

Testing Horizon (seconds)	RBF				RBF + Periodic Exponential			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
1	0.0059	5.72e-05	0.0075	7.47	0.0181	5.9e-06	2.437e-03	24.9274
2	0.108	0.0441	0.21	290.44	0.062	0.01	0.1	151.71
3	0.5675	0.815	0.902	1179.79	0.1	0.02	0.144	152.976
5	0.6561	0.8932	0.9451	2919.22	0.093	0.016	0.1263	135.2470



(a) **(b)**

Fig. 3 Prediction using (RBF + Periodic Exponential) kernel for prediction horizon of a) 1 second and b) 3 seconds.

time series outcomes that exhibit trends and periodic patterns.

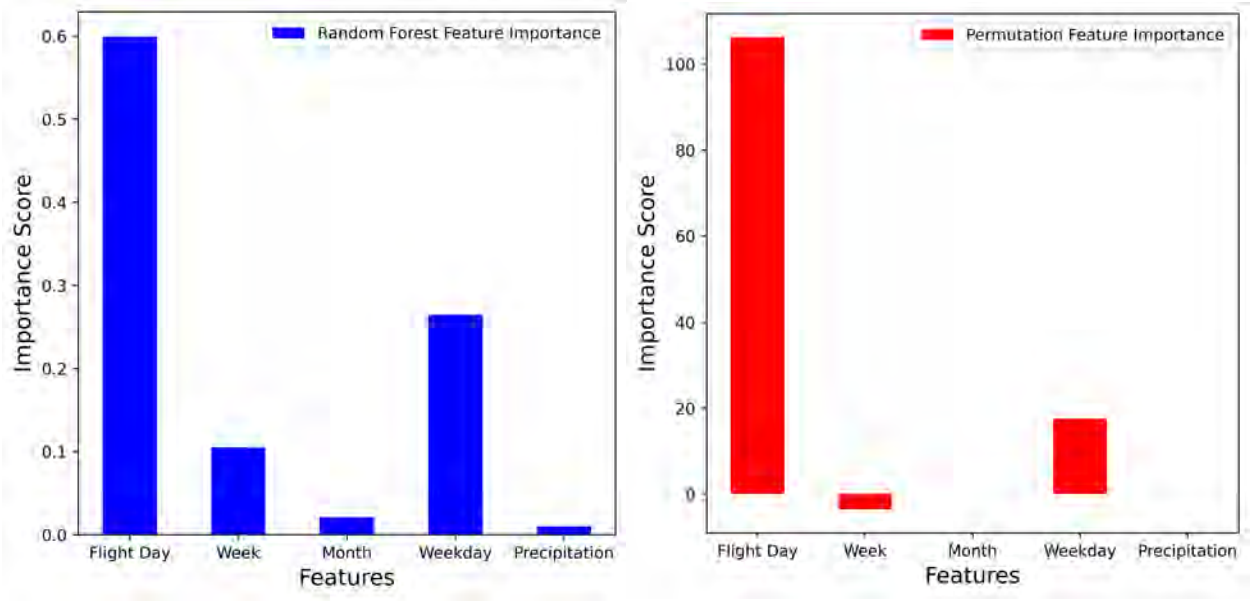
V. Application to Flight Delay Prediction

In this section, we employed GPR to predict arrival/departure delays of a particular flight on a specific flight route. The flight data source for this study is the database maintained by the United States Department of Transportation, Bureau of Transportation Statistics. This database contains information on all commercial flights operating within the

national airspace. The available flight information includes airline flight number, departure airport, arrival airport, scheduled time of departure, the actual time of departure, scheduled time of arrival, the actual time of arrival, and flight time. The flights taken into account are from either before the COVID pandemic or after it, as flight scheduling was greatly disrupted during the COVID phase.

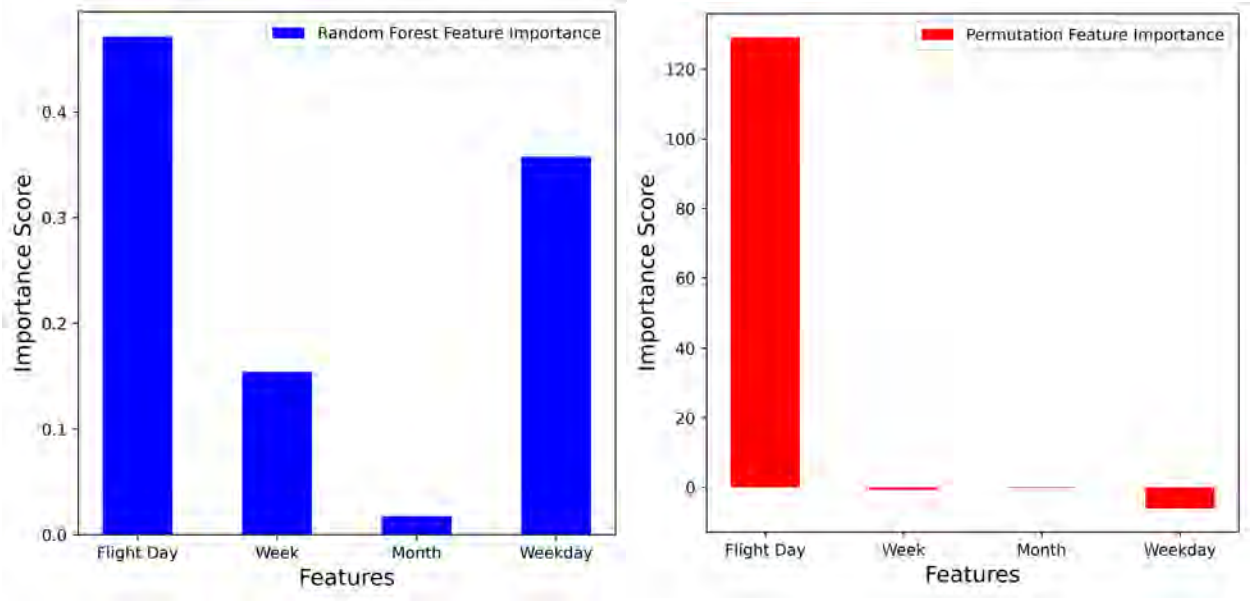
A. Input Feature Selection

For input feature selection, we implemented the Random Forest feature importance, and Permutation feature importance method to calculate the feature importance score. We considered three different flights, AA2754, AA1331 (LAX-DFW), and AA1331 (DFW-LAX), to determine which features have more importance in flight delays. Importance scores calculated from both techniques are compared for feature selection. For each of the flights, flight day has a significantly large importance score compared to other features. For flight AA1331, the random forest feature importance score for weekday was approximately 0.35. However, the importance score with permutation feature importance was significantly less compared to flight day. Since both methods showed higher importance to flight day in each flight, we selected flight day as the only feature for predicting arrival and departure delays. Figure 4, 5, and 6 shows feature importance scores for various features studied on various flights.



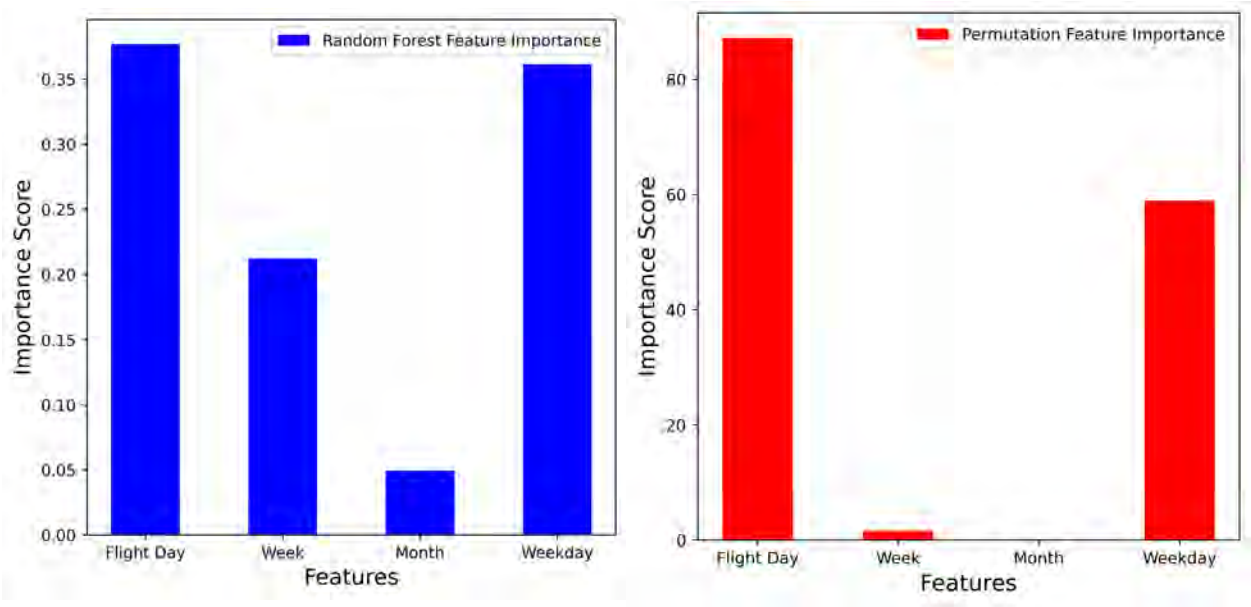
(a) **(b)**

Fig. 4 AA2754 feature importance Score using (a) Random Forest feature importance and (b) Permutation feature importance method.



(a) **(b)**

Fig. 5 AA1331 (DFW-LAX) feature importance Score using (a) Random Forest feature importance and (b) Permutation feature importance method.



(a) **(b)**

Fig. 6 AA1331 (LAX-DFW) feature importance Score using (a) Random Forest feature importance and (b) Permutation feature importance method.

B. Prediction Framework

For this application, we are utilizing the same framework discussed in Section IV. In regression, training is less effective when the data is scattered. Therefore, we pre-processed the data before feeding it into the training model. Based

on our observations, the outliers are infrequent and typically arise from unforeseen factors. These outliers will appear recurrently in predictive outcomes if we incorporate them into the training data, which doesn't align with real-world flight delay patterns. As a result, in this study, we have set an acceptable range of ± 60 minutes for filtering out outliers. Initially, outliers were removed from the dataset, and after that, flights diverted from their intended destination airport were also excluded because of prolonged flight times. Moreover, data normalization was carried out case-by-case depending on the model's suitability.The main focus of this application was to capture seasonal patterns, and removing outliers shouldn't affect the prediction results. Due to non-stationarity, the kernels selected for seasonal patterns can fail to capture peak delays. However, implementing a non-stationary kernel is out of the scope of this application. Readers interested in the non-stationary kernel for GPR can refer to the literature [33].

Numerous machine-learning libraries exist to implement a GPR. In this study, we employed *GPy*, a Python library tailored for classification and regression [34]. *GPy* library has an extensive collection of built-in covariance functions or kernels. As discussed earlier, these kernels can be combined as required to create a new kernel, reflecting the specifics of the problem at hand. When employing priors with zero mean and unit variance, the target variable's values can be standardized by scaling them to a unit variance. In this work, we used the gradient-based L-BFGS-B optimizer to tune hyperparameters. In addition, the Gaussian noise variance was also considered as a hyperparameter to be optimized.

In machine learning problems, it's important to include only those features in the training dataset that significantly impact the target variable. When predicting flight delays, the target variable is the flight's arrival/departure delay. Incorporating irrelevant features adds computational overhead and can degrade the prediction outcomes. For a particular flight, as determined by the random forest feature importance and permutation feature importance scores, the arrival delay is primarily influenced by the day of flight [35]. Therefore, in this research, only flight day was integrated as a feature in the training data. Also, the seasonal decomposition detailed in section IV guided the kernel selection. However, the final decision on the kernels was made strictly based on the regression metrics. The regression metrics MAE, MSE, RMSE, and MAPE are expressed in minutes, minutes squared, minutes, and percentages respectively.

C. Arrival Delay Prediction for American Airlines Flight AA1331 (LAX-DFW)

The American Airlines flight AA1331 was scheduled to fly from Los Angeles International Airport (LAX) to Dallas-Fort Worth International Airport (DFW), in 2022. The training dataset was limited to 180 days, as airlines frequently change their flight numbers for specific routes. From the seasonal decomposition, we observed a trend and periodicity in the dataset as shown in Fig. 7. In addition, the residuals from seasonal decomposition suggested noise within the data. Thus, the potential kernels should show a trend, periodicity, or both.

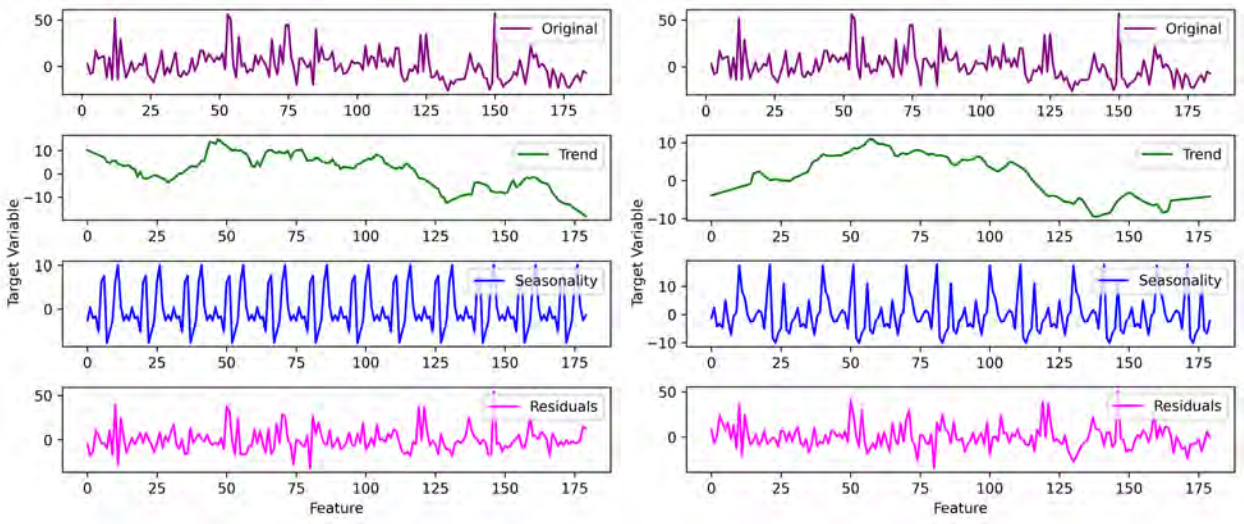


Fig. 7 Seasonal decomposition of training data for AA1331, assuming a periodicity of (a) 15 days and (b) 30 days.

1. Kernel Performance Comparative Analysis

The performance assessment of various kernels with the given training dataset of AA1331 is shown in Table 3. The maximum and minimum MAEs are highlighted in red and green boxes. The results show that the Periodic and RBF kernels give similar outputs. Also, the combined (RBF + Periodic) kernel and the (RBF + Periodic Exponential + Periodic) kernel show identical results. Additionally, the MAE in minutes for the (RBF + Periodic) kernel was low during training, but it significantly rose when applied to the testing data. From this observation, we can infer that complex kernels may not be effective for training datasets of smaller sizes. The delay predictions for flight AA1331, utilizing two distinct kernels, are illustrated in Fig. 8.

Table 3 Performance evaluation of various kernels on a 180-day training dataset (unnormalized) with a 7-day testing period for flight AA1331.

Kernels	Training Data				Testing Data			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
Linear	11.27	238.92	15.45	96.26	18.87	619.72	24.894	88.85
RBF	10.49	208.75	14.45	120.16	17.69	658.72	25.66	81.55
Periodic	10.49	208.78	14.44	120.16	17.69	658.71	25.66	81.55
Cosine	10.67	210.59	14.51	126.25	18.00	647.19	25.44	80.03
Matern32	10.34	204.66	14.30	117.71	17.74	660.04	25.69	81.78
RBF + Periodic	8.91	154.30	12.42	107.27	14.95	482.49	21.96	62.11
RBF + Per. Exponential + Periodic	8.91	154.30	12.42	107.27	14.95	482.49	21.96	62.11

From Fig. 8, it is evident that the 95% confidence interval is wide. This can be attributed to the use of Gaussian noise variance as a hyperparameter that undergoes optimization. Considering the MAE, the predictions are sufficiently

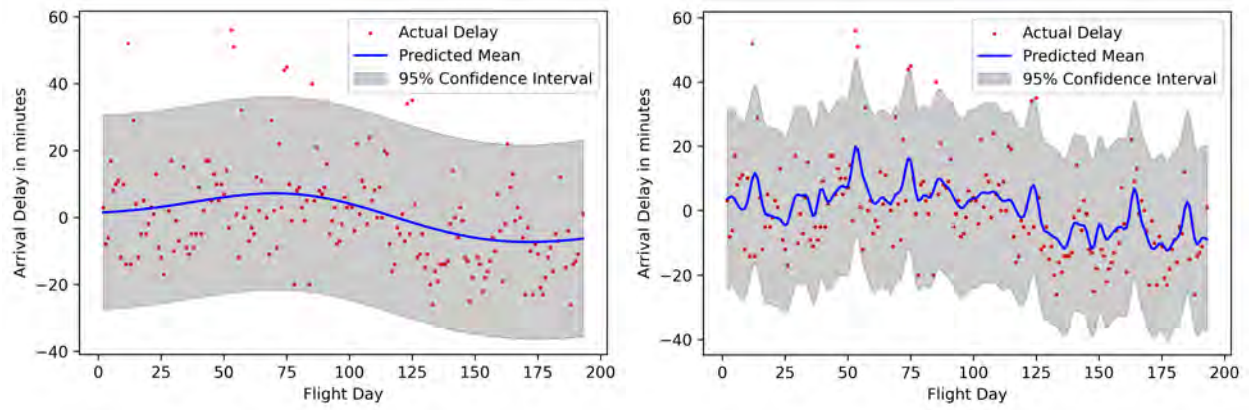


Fig. 8 Arrival delay prediction for AA1331 using 180 days of training data (normalized) and a 10-day testing period using (a) RBF and (b) (RBF + Periodic) kernels.

accurate to provide insights into potential delays or early arrivals. Based on this information, airlines can make informed decisions and plan their operations accordingly.

2. Comparison of GPR with Other ML Techniques

In this section, we compared the results of the GPR model with the Long Short-Term Memory (LSTM) and Support Vector Machine (SVM) regression model. For LSTM, the number of LSTM layers was 200, and the number of iterations was 1000. Since we had a single output model, we set the number of the sequence input layers and fully connected layers as 1. Similarly, we used a polynomial kernel for SVM. For both techniques, a bootstrapping method was used to calculate the confidence interval. The size of the training and testing datasets were kept the same as used in GPR for an unbiased comparison. The computational cost of each method is given in Table 4. The SVM model predicts delays with less computational time and MAE than GPR. However, if we observe the predicted mean for 10 seconds, it is a linear line, and the predicted confidence interval is smaller than the 2-sigma of the actual data. Similarly, the LSTM model also has a confidence interval that is smaller than that of GPR. The predicted mean of LSTM shows some nonlinearity captured, but it is very small as compared to the predicted mean of GPR. In addition, very high computational cost makes the LSTM less effective than GPR. The predicted mean and confidence interval for each methods are shown in Fig. 9.

Table 4 Computational cost comparison of various methods for arrival delay prediction of flight AA1331 (180 days of normalized training data and a 10-day testing period).

Method	Computational Time (in seconds)	MAE
GPR	0.506 s	17.04 min
LSTM	227.6 s	18.62 min
SVM	0.2344 s	16.49 min

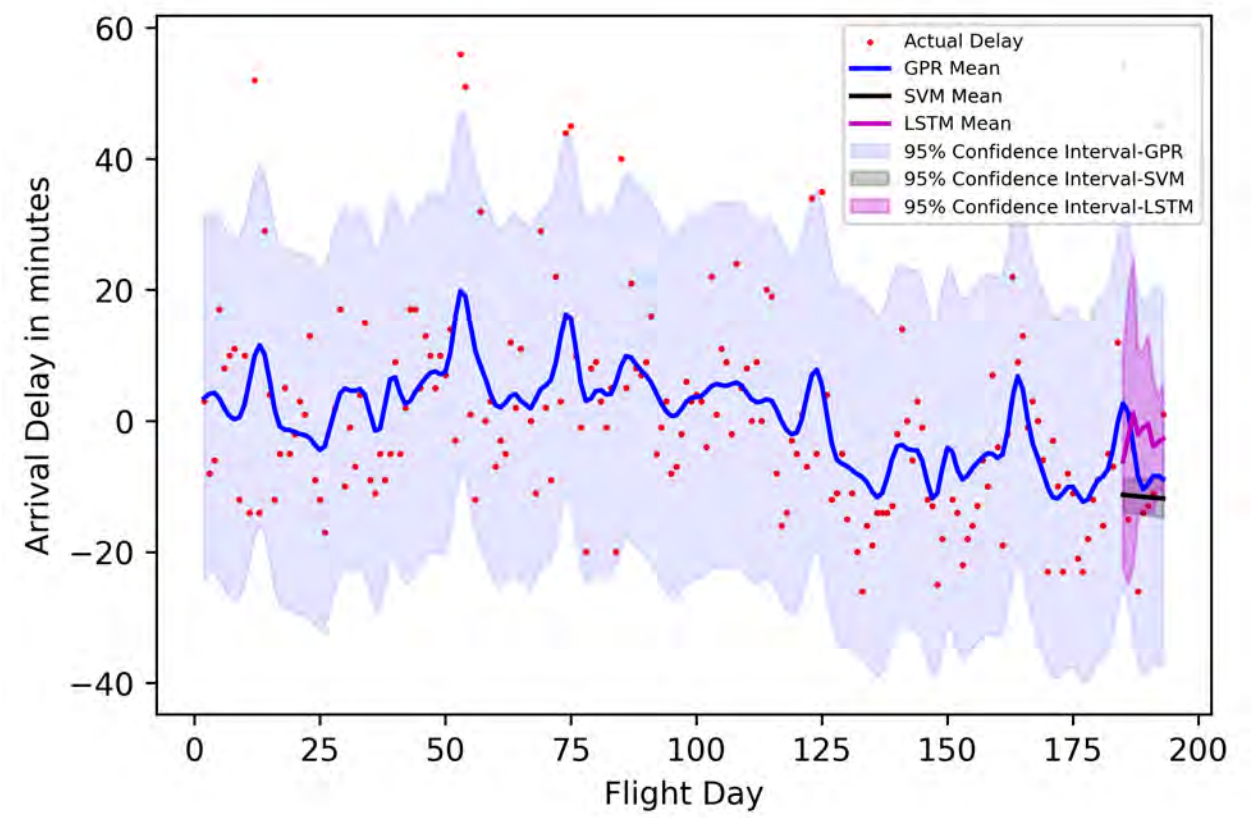


Fig. 9 Arrival delay prediction for flight AA1331 using 180 days of training data (normalized) and a 10-day testing period using different methods.

D. Arrival Delay Prediction for American Airlines Flight (DCA-DFW)

In this section, we focused on predicting arrival delays for an American Airlines flight that was scheduled to fly from Ronald Reagan Washington National Airport (DCA) to Dallas/Fort Worth International Airport (DFW) between 9:00 AM and 10:00 AM from January 2016 through October 2017. In contrast to the previous scenario, instead of targeting a particular flight number, we had chosen flights based on a specific time frame. First, we applied seasonal decomposition to determine the most appropriate kernel types for this model. Figure 10 displays the seasonal decomposition, highlighting a prominent trend feature with some degree of periodicity evident in the arrival delay dataset.

1. Kernel Performance Comparative Analysis

In this section, we evaluated the performance of various kernels for the given training dataset for DCA-DFW route. The assessment was conducted using regression metrics: MAE, MSE, RMSE, and MAPE, which are expressed in minutes, minutes squared, minutes, and percentages, respectively. From the performance metrics shown in Table 5, the low MAE for the training dataset was achieved when using the RBF, (RBF + Periodic), and (RBF + Periodic Exponential

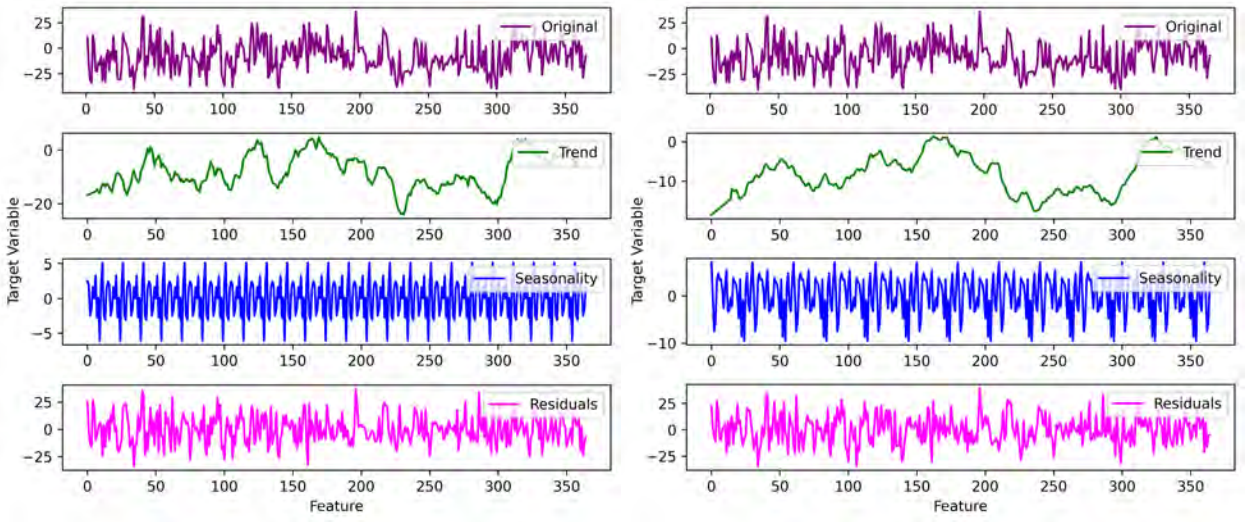


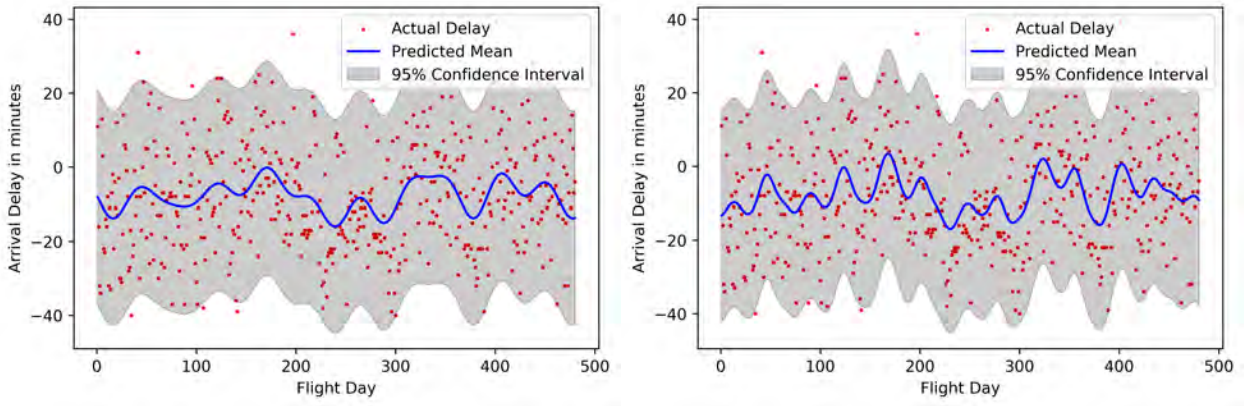
Fig. 10 Seasonal decomposition of training data for DCA-DFW flight route, assuming a periodicity of (a) 15 days and (b) 30 days.

+ Periodic) kernel. Furthermore, when predicting using (RBF + Periodic Exponential + Periodic) kernel for a prediction horizon of 30 days as shown in Fig. 11, the results had an MAE of 8.78 minutes and a MAPE of 114.68%. Therefore, we employed an Exponentiated Weighted Moving Average (EWMA) approach to enhance data smoothness. The primary purpose of employing EWMA was to smooth out abrupt variations in delays, which could otherwise negatively impact model training. When the training data exhibits a pronounced periodic pattern, using a low smoothing factor in EWMA can help maintain that periodicity. The evaluation of various kernels with smoothed data is shown in Table 6. The maximum and minimum MAEs are highlighted in red and green boxes. The evaluation results show that all the kernels perform better on the smoothed dataset, indicated by lower MAE values. Based on the MAE of the training dataset, four potential kernels, namely RBF, (RBF + Periodic), (RBF + Periodic Exponential), and (RBF + Periodic Exponential + Periodic), were selected for further evaluation. The performance of selected kernels on various testing horizons is shown in Table 7. It is apparent that the combination of (RBF + Periodic Exponential + Periodic) gives the best results for all testing horizons, with an MAE of 5 minutes for a testing period of 60 days.

We further evaluated the kernel performance on various training horizons with a fixed testing dataset of 30 days as shown in Table 8. It was observed that the MAE for the training horizons of 100, 200, 400, and 500 days were comparable. However, when trained on a 300-day dataset, the prediction MAE was higher. This suggested that the chosen kernel didn't train the model effectively with a 300-day training dataset. This indicates that model performance doesn't always improve with an increase in the size of the training data. Hence, it can be inferred that increasing the size of the training dataset might degrade the model's performance when the data are contaminated with noise. Another possibility is that the chosen kernel failed to grasp relevant characteristics within this particular span of training data.

Table 5 Performance evaluation of various kernels on a 450-days training dataset (unnormalized) with a 7-day testing period for the DCA-DFW flight.

Kernels	Training Data				Testing Data			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
Linear	13.09	253.84	15.93	118.44	6.72	72.65	8.52	159.40
RBF	6.61	63.22	7.95	46.49	9.00	132.71	11.52	100
Periodic	10.53	163.79	12.79	115.96	7.85	95.29	9.76	106.01
Cosine	12.40	232.11	15.23	128.98	6.02	68.15	8.25	107.98
Matern32	10.59	167.17	12.92	116.64	7.67	111.47	10.55	76.41
RBF + Periodic	6.84	69.42	8.33	75.59	5.71	67.01	8.18	123.36
RBF + Per. Exponential + Periodic	6.91	72.12	8.49	71.89	6.02	68.07	8.25	108.25



(a) (b)

Fig. 11 Arrival delay prediction for the DCA-DFW route using 450 days of training data (normalized) and a 30-day testing period using (a) (RBF + Periodic Exponential) and (b) (RBF + Periodic Exponential + Periodic) kernels.

Table 6 Performance evaluation of various kernels on an EWMA (α = 0.3) smoothed dataset spanning 450 days with a 7-Day testing period for the DCA-DFW flight.

Kernels	Training Data				Testing Data			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
Linear	8.00	94.46	9.72	278.82	5.21	33.50	5.78	134.76
RBF	1.68	4.55	2.13	132.62	8.14	71.38	8.44	155.80
Periodic	6.46	60.59	7.78	434.93	7.049	54.20	7.36	170.59
Cosine	7.07	73.01	8.54	982.75	2.44	7.09	2.66	56.40
Matern32	3.64e-05	2.08e-09	4.56e-05	0.0077	6.26	44.93	6.70	113.51
RBF + Periodic	1.49	3.54	1.88	154.71	4.65	25.80	5.07	85.08
RBF + Per. Exponential	1.68	4.55	2.13	132.77	8.14	71.38	8.44	155.79
RBF + Per. Exponential + Periodic	4.73	34.61	5.88	372.32	2.38	10.31	3.21	39.54

This could be attributed to an irregular delay pattern in this range of data compared to other intervals. Figure 12 shows arrival delay prediction for the DCA-DFW route with 450 days of training and 10-days of testing data using (RBF + Periodic) and (RBF + Periodic Exponential + Periodic) kernels.

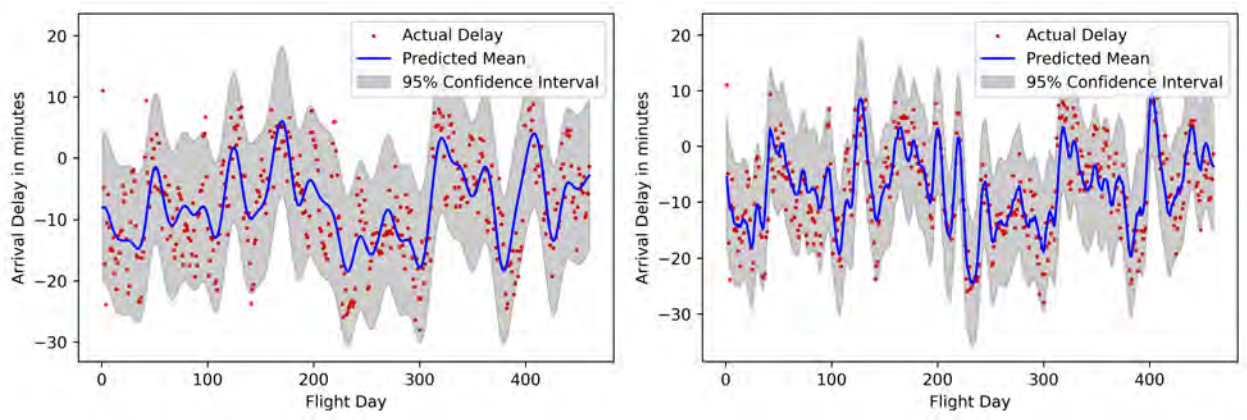
Table 7 Performance evaluation of various kernels on an EWMA (α = 0.3) smoothed dataset spanning 450 days with various testing periods for the DCA-DFW Flight.

Testing Horizon (Days)	RBF				RBF + Periodic			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
10	7.02	57.17	7.56	139.40	4.01	20.03	4.47	81.33
15	8.10	78.45	8.85	126.27	5.05	34.03	5.83	75.13
30	7.99	82.82	9.10	113.13	5.21	36.50	6.04	120.18
60	6.81	64.43	8.02	106.568	6.90	72.618	8.52	294.65

Testing Horizon (Days)	RBF + Per. Exponential				RBF + Per. Exponential + Periodic			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
10	7.02	57.17	7.56	139.39	2.34	8.82	2.97	47.17
15	8.10	78.46	8.85	126.26	4.10	30.01	5.47	52.76
30	7.99	82.82	9.10	113.14	4.56	34.74	5.89	70.00
60	6.81	64.43	8.02	106.60	5.00	37.41	6.11	108.54

Table 8 Performance evaluation of (RBF + Per. Exponential + Periodic) kernel on an EWMA (α = 0.3) smoothed dataset with different training dataset sizes.

Training Horizon (Days)	Training Dataset				Testing Dataset			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
100	3.00	15.25	3.90	2897.32	8.29	85.26	9.23	317.53
200	2.74	12.27	3.50	306.08	8.15	104.70	10.23	164.78
300	3.60	22.08	4.699	677.10	21.67	588.95	24	1102.70
400	4.62	33.59	5.79	238.33	9.06	129.09	11.36	228.97
500	4.58	34.03	5.83	408.42	8.822	111.55	10.56	289.31
600	5.57	48.01	6.92	577.92	13.67	216.60	14.71	825.47



(a) **(b)**

Fig. 12 Arrival delay prediction for the DCA-DFW route using 450 days of training data and 10-day testing using (a) RBF + Periodic and (b) RBF + Periodic Exponential + Periodic kernels.

From the above results, we can conclude that a combined kernel formed by the addition of RBF, Periodic Exponential, and Periodic is the suitable kernel for predicting arrival delays for DCA-DFW flights scheduled to fly between 9:00 AM

to 10:00 AM. This kernel can predict the moving average of arrival delays with an MAE of 4.56 minutes and MAPE of approximately 70 %. Also, the optimal prediction horizon for the (RBF + Periodic exponential + Periodic) kernel is 30 days, given that the MAE is less than 5 minutes and other pertinent metrics are similarly low. For the lowest MAE in the original scale, we can implement weighted transformations or ensemble methods. However, such approaches are beyond the scope of this study as this research is specifically centered on determining whether GPR can effectively capture trends and periodicity in flight delay data.

2. Comparison of GPR with Other ML Techniques

We followed the approach discussed in the previous section to compare with the Long Short-Term Memory (LSTM) and Support Vector Machine (SVM) regression models. In this case, the same parameter values were used as in AA1331 (LAX-DFW). For both techniques, a bootstrapping method was used to calculate the confidence interval, and the number of samples was the same for consistency. The computational cost of each method is given in Table 9. From the result, we can observe that both SVM and LSTM models predict delays with less MAE than GPR. However, if we observe the predicted mean for 30 seconds of testing data, the predicted confidence interval is smaller than the 2-sigma of the actual data. In addition, the LSTM model has a significantly large computational time. Thus, GPR outperforms both LSTM and SVM in the predicted 95% confidence interval. The predicted mean and confidence interval for each method are shown in Fig. 13.

Table 9 Computational cost comparison of various methods for predicting delays of flight for DCA-DFW route (450 days of training data and a 30-day testing period).

Method	Computational Time (in seconds)	MAE
GPR	7.59 s	4.57 min
LSTM	393.48 s	4.18 min
SVM	0.25 s	4.39 min

E. Arrival Delay Prediction for American Airlines Flights on Various Flight Routes

In this section, first, we evaluated the efficacy of various kernels in predicting arrival delays over the course of a year. The flight sector under study includes arrival to Dallas/Fort Worth International Airport (DFW) from ten of the busiest airports in the United States: Hartsfield-Jackson Atlanta International Airport (ATL), Denver International Airport (DEN), O'Hare International Airport (ORD), Los Angeles International Airport (LAX), Harry Reid International Airport (LAS), Miami International Airport (MIA), Charlotte Douglas International Airport (CLT), Seattle-Tacoma International Airport (SEA), Minneapolis-Saint Paul International Airport (MSP), and LaGuardia Airport (LGA) as shown in Fig. 14. All the analyzed flights were scheduled to depart from their respective origin airports with destinations set for DFW, during the time frame spanning from 3:00 PM to 6:00 PM. Subsequently, we studied the effectiveness

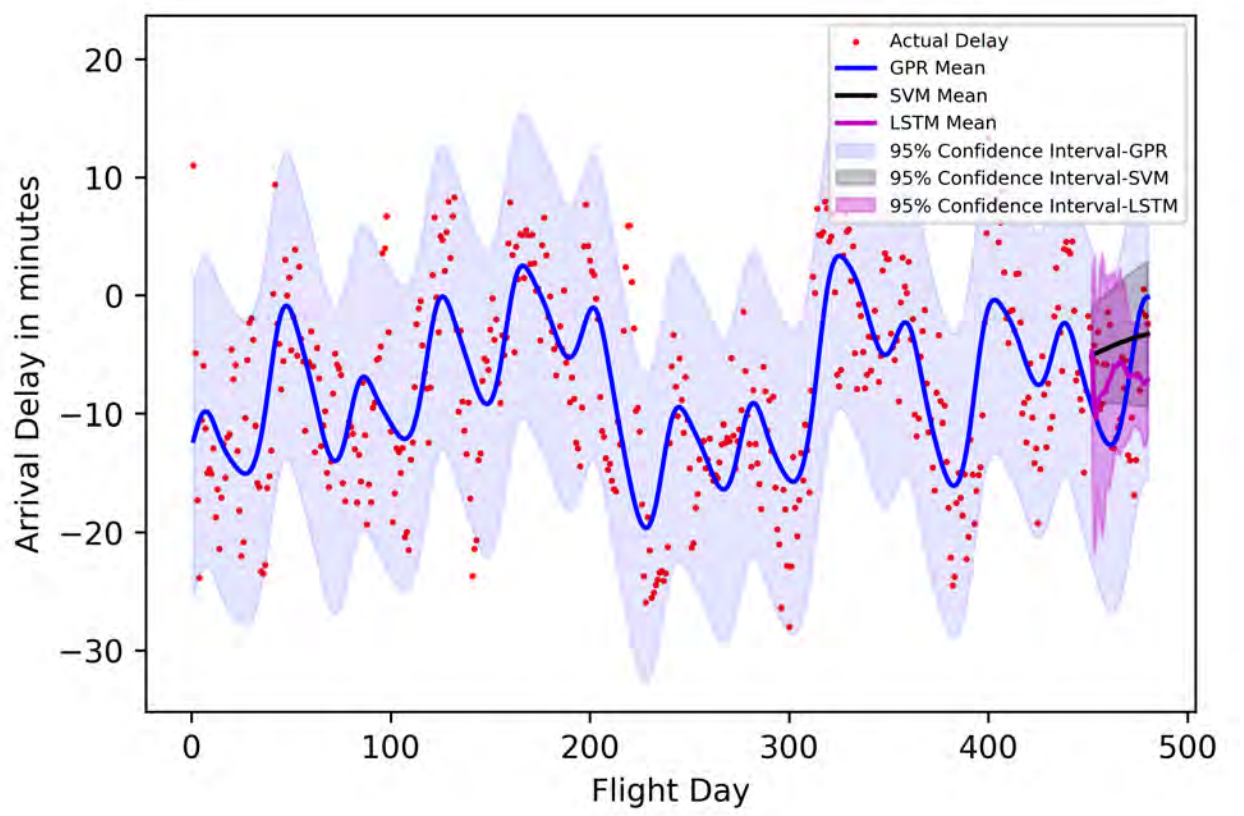


Fig. 13 Arrival delay prediction for DCA-DFW route using 450 days of training data and a 30-day testing period using different methods.

of a specific kernel for predicting arrival delays of different flights. We employed the (RBF + Periodic exponential + Periodic) kernel to forecast arrival delays of American Airlines flights across various sectors. The primary aim of this section was to gauge how well this identical kernel could predict arrival delays for American Airlines flights on different flight routes. We employed data from the entire year of 2016 as our training set and predicted the delays for every month in 2017. Additionally, the training data incorporated flight information up to the month prior to the month being predicted. For instance, to make predictions for March 2017, we utilized data spanning from January 2016 to February 2017 as our training set. For each case, the kernel's initial hyperparameters and the smoothing parameter (α) for the EWMA were kept constant.

1. Performance Comparison of Various Kernels on Different Flight Routes

In this analysis, we explored various kernel functions to evaluate performance disparities in predicting arrival delays across various flight routes. The comparison of various kernels on different flight routes is shown in Table 10, 11, 12, and 13. For each month, the maximum and minimum MAE are highlighted with red and green boxes. For the month of January, the prediction MAE for arrival delays from the majority of the airports was consistent across all examined

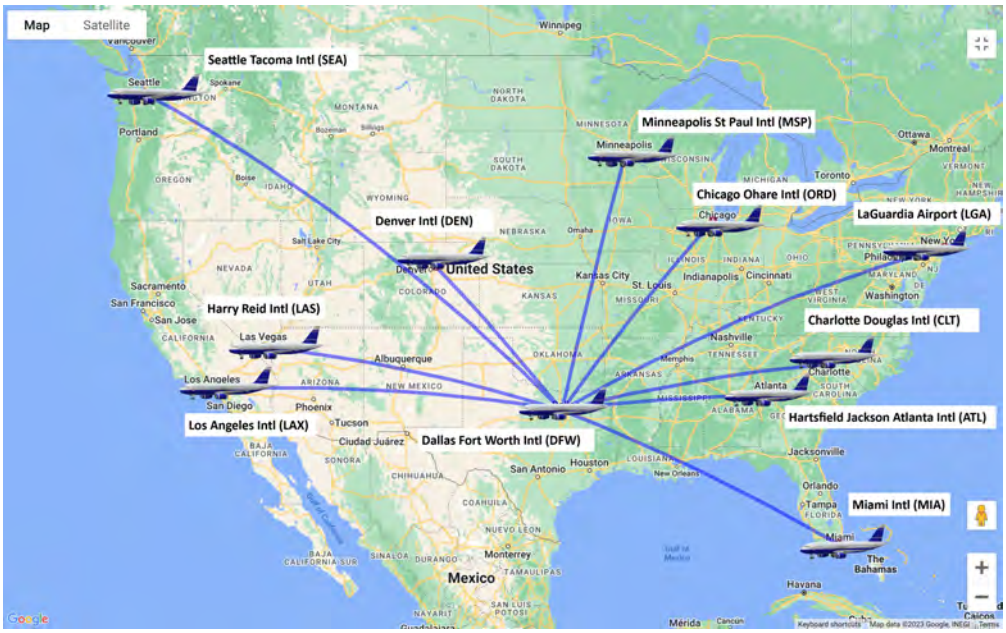


Fig. 14 Map representation of American Airlines flight route for predicting arrival delays.

kernels. Notably, flights originating from LAX had a distinctive behavior, with none of the investigated kernels having satisfactory predictions. In the subsequent monthly analysis, for February, most kernels continued to exhibit subpar performance for flights originating from LAX, with the notable exception of the Cosine kernel. Conversely, the (RBF + Periodic) kernel displayed deficient performance for flights arriving from MSP and ORD airports. In March, the Periodic kernel under-performed for flights associated with LAX and MIA. Similarly, the Cosine kernel had less than optimal outcomes for flights from MSP and ORD. Noteworthy among these observations, the (RBF + Periodic) kernel consistently lagged in its predictive accuracy, registering an MAE exceeding 10 minutes for five distinct flight routes. In April, a decline in performance was observed across all kernels for flights originating from MIA and ORD, with the combined (RBF + Periodic) kernel being the least accurate. Progressing to May, the flight arriving from MIA consistently underperformed with all the investigated kernels. Moreover, the RBF and Periodic kernels individually exhibited large MAE for flights from ORD and SEA, and the combined (RBF + Periodic) kernel underperformed for LGA and MIA routes. Transitioning into June, flights from ORD and SEA remained problematic for most kernels. The Cosine kernel, in particular, displayed suboptimal outcomes for additional routes such as LAX and MIA. Unlike the previous months, the prediction for July was inaccurate primarily for flights from ATL across all kernels, excluding the (RBF + Periodic) kernel. Performance deficits were also noted for LAS and MIA routes. Notably, the (RBF + Periodic) kernel demonstrated higher accuracy for most routes this month, maintaining an MAE below 10 minutes. In August, the predicted MAE for MIA flights was large for all kernels. Furthermore, MSP flights had an MAE exceeding 10 minutes for three kernel configurations. As we moved into September, flights from LGA and ORD proved challenging for all kernels, with the (RBF + Periodic) kernel having a large MAE on half of the examined routes. October had pronounced

deficiencies in five of the six kernels for flights from LGA and MIA. The (RBF + Periodic) kernel had an unusually high MAE of 45 minutes for the LGA-DFW route and an MAE above 10 minutes for five routes. Again, in November, flights from ATL had higher MAE across all kernels. The Cosine kernel was notably the least effective, with three out of ten flight routes yielding an MAE above 10 minutes. Finally, in December, we observed prediction inaccuracy with SEA flights across all kernels. The (RBF + Periodic) kernel's performance, while improved in previous months, still had 2 out of 10 routes with an MAE surpassing 10 minutes.

Table 10 Performance evaluation of various kernels on an EWMA (α = 0.3) smoothed dataset for arrival delay prediction across different flight routes (January-March).

Month	Kernel	Origin Airport and Arrival Delay MAE									
		ATL	CLT	DEN	LAS	LAX	LGA	MIA	MSP	ORD	SEA
January	Linear	4.89	6.63	5.71	5.81	16.47	8.08	7.72	8.59	7.53	6.45
	RBF	4.99	7.15	6.77	5.55	17.06	8.55	6.84	7.64	6.58	6.91
	Periodic	5.95	7.30	7.10	4.91	22.79	5.84	5.49	10.17	6.58	7.37
	Cosine	5.20	7.13	8.06	5.37	13.57	7.80	4.98	8.42	7.19	6.48
	Matern32	4.98	7.15	6.78	5.53	15.77	8.61	6.94	7.42	6.62	6.80
	RBF + Periodic	4.93	7.15	4.69	5.66	42.97	8.55	6.84	7.64	6.57	6.91
February	Linear	7.57	7.35	7.02	6.29	8.72	7.57	6.33	5.52	8.00	4.07
	RBF	7.53	7.32	4.32	6.09	13.30	7.57	6.59	5.62	6.45	5.08
	Periodic	7.64	6.67	3.85	8.00	16.58	7.34	4.81	5.44	9.80	6.90
	Cosine	7.97	9.96	5.52	5.37	4.45	7.86	5.76	8.81	7.07	5.40
	Matern32	7.61	7.17	4.30	6.04	12.72	7.56	6.68	5.69	6.16	4.96
	RBF + Periodic	7.71	7.32	6.36	7.22	13.30	7.57	4.88	14.65	11.06	4.35
March	Linear	4.93	7.37	7.68	4.70	8.20	7.83	8.97	7.87	6.71	5.95
	RBF	4.85	8.29	6.61	4.25	8.69	7.35	8.71	9.20	6.82	6.20
	Periodic	4.78	9.65	6.47	5.18	12.28	8.03	10.37	8.71	6.82	6.44
	Cosine	4.89	9.89	6.76	4.56	8.32	9.79	8.98	13.28	11.12	6.60
	Matern32	4.79	8.26	6.63	4.20	8.65	7.42	8.61	9.18	6.73	6.13
	RBF + Periodic	4.85	10.20	6.61	17.06	13.25	15.38	10.39	9.20	7.30	6.87

From the above observation, we can infer that no specific kernel can perform well in prediction for all the months. A composite kernel, which combines RBF, Periodic Exponential, and Periodic kernel, can be a favorable choice in scenarios where different kernels demonstrate commendable performance on different datasets. A combined kernel, having three different kernels present, can capture different features of the data. Thus, we will further evaluate the (RBF + Periodic Exponential + Periodic) kernel in the next section.

2. Performance Comparison of (RBF + Periodic Exponential + Periodic) Kernel on Different Flight Routes

The training dataset utilized in this study adhered to the aforementioned specifications. The arrival delay prediction error for each flight route employing (RBF + Periodic Exponential + Periodic) kernel is given in Table 14.

The large prediction error correlates with an erratic delay pattern, adding outliers in training data. Such irregular delays can arise from factors like weather conditions or sudden increases in passenger volume. Consequently, the kernel

Table 11 Performance evaluation of various kernels on an EWMA (α = 0.3) smoothed dataset for arrival delay prediction across different flight routes (April-June).

Month	Kernel	Origin Airport and Arrival Delay MAE									
		ATL	CLT	DEN	LAS	LAX	LGA	MIA	MSP	ORD	SEA
April	Linear	7.11	6.28	6.82	5.27	5.45	8.94	11.86	5.24	9.19	7.76
	RBF	6.56	6.52	5.50	5.11	9.57	8.60	12.01	6.41	10.80	8.08
	Periodic	7.17	6.48	6.90	7.04	9.23	11.79	11.35	8.91	12.70	10.07
	Cosine	6.69	6.61	4.21	5.75	4.66	8.56	11.97	5.75	13.36	8.74
	Matern32	6.25	6.54	5.51	5.03	9.35	8.57	12.19	6.41	10.82	7.96
	RBF + Periodic	6.30	17.03	9.02	5.11	44.33	7.59	12.01	7.90	10.80	6.79
May	Linear	7.87	5.70	4.31	5.03	8.05	9.40	10.04	5.68	9.13	9.40
	RBF	7.76	4.93	4.07	4.79	5.70	9.34	10.04	7.57	10.05	10.07
	Periodic	7.08	5.57	4.18	5.03	6.69	9.89	14.18	10.94	13.30	12.85
	Cosine	6.96	4.29	5.16	5.16	9.06	9.32	9.40	6.56	8.01	8.22
	Matern32	7.75	4.99	4.11	4.73	6.11	9.41	10.35	7.68	9.73	9.99
	RBF + Periodic	7.93	4.72	6.39	4.79	5.70	10.03	21.06	6.50	22.36	10.07
June	Linear	6.81	7.68	4.77	3.39	12.17	5.87	9.87	6.50	13.62	7.55
	RBF	6.43	7.06	4.11	3.73	5.46	5.86	9.89	5.66	15.18	10.25
	Periodic	6.60	6.56	3.87	3.19	7.01	7.18	7.69	5.28	13.98	13.48
	Cosine	6.59	5.40	5.04	3.12	10.02	5.38	11.72	5.88	12.15	4.88
	Matern32	6.30	6.98	4.10	3.79	4.71	5.83	9.79	5.66	14.99	10.06
	RBF + Periodic	8.49	18.26	6.46	3.60	9.93	5.86	7.84	8.87	15.59	11.49

Table 12 Performance evaluation of various kernels on an EWMA (α = 0.3) smoothed dataset for arrival delay prediction across different flight routes (July-September).

Month	Kernel	Origin Airport and Arrival Delay MAE									
		ATL	CLT	DEN	LAS	LAX	LGA	MIA	MSP	ORD	SEA
July	Linear	10.76	4.21	3.52	9.77	5.01	8.44	15.86	10.57	6.66	8.24
	RBF	10.81	2.96	3.58	10.01	4.92	8.42	15.17	8.38	4.85	10.59
	Periodic	12.06	3.71	4.32	11.51	7.58	6.18	16.82	9.79	7.86	12.46
	Cosine	12.46	7.40	5.77	8.12	7.43	7.99	15.99	6.97	9.77	6.11
	Matern32	10.80	2.94	3.63	10.08	5.08	8.41	14.95	8.36	4.71	10.46
	RBF + Periodic	9.39	2.96	8.33	9.77	4.92	10.61	7.14	9.34	4.85	9.04
August	Linear	7.29	4.87	3.40	4.91	4.20	5.56	12.14	11.73	4.36	6.26
	RBF	7.36	4.77	3.83	5.09	5.98	5.84	12.65	9.78	3.99	5.34
	Periodic	7.36	3.62	3.43	5.46	6.02	7.35	19.03	10.68	4.68	6.36
	Cosine	7.10	4.06	7.55	4.59	8.91	6.53	10.66	7.65	6.55	8.82
	Matern32	7.37	4.84	3.81	4.95	5.53	5.90	12.07	9.69	3.94	5.36
	RBF + Periodic	10.87	4.59	5.05	5.31	4.75	11.38	23.76	15.84	3.99	6.02
September	Linear	10.30	4.24	5.40	5.54	4.81	12.23	7.54	5.06	13.47	4.82
	RBF	9.34	4.54	4.99	5.78	4.43	12.17	8.73	5.28	12.52	4.67
	Periodic	11.36	4.08	4.66	6.07	6.29	11.02	8.15	3.80	11.60	4.67
	Cosine	11.07	6.32	6.53	5.39	7.60	11.41	6.34	11.00	14.46	6.34
	Matern32	9.24	4.70	4.98	5.77	4.62	12.14	8.37	5.09	12.45	4.65
	RBF + Periodic	7.48	12.05	8.22	6.07	4.43	21.55	10.16	14.77	12.56	4.67

Table 13 Performance evaluation of various kernels on an EWMA (α = 0.3) smoothed dataset for arrival delay prediction across different flight routes (October-December).

Month	Kernel	Origin Airport and Arrival Delay MAE									
		ATL	CLT	DEN	LAS	LAX	LGA	MIA	MSP	ORD	SEA
October	Linear	7.71	7.10	7.19	6.10	6.38	12.88	12.21	5.54	7.88	7.30
	RBF	7.36	6.95	6.18	5.76	4.91	13.34	10.12	5.51	8.29	5.70
	Periodic	7.68	7.37	6.25	5.94	5.77	9.33	7.44	5.28	9.03	5.92
	Cosine	6.47	8.34	4.23	6.50	5.13	16.58	14.89	8.88	8.46	10.14
	Matern32	7.20	6.88	6.04	5.65	4.99	13.56	10.03	5.52	8.46	5.94
	RBF + Periodic	7.33	6.16	4.81	5.23	12.84	45.16	12.10	17.43	11.64	5.17
November	Linear	12.44	7.33	8.42	5.55	5.30	9.26	5.31	6.07	6.97	8.69
	RBF	11.67	7.04	8.29	5.75	6.74	8.52	4.61	5.87	8.93	6.50
	Periodic	13.36	9.52	8.84	5.47	8.68	9.80	5.55	8.79	8.63	5.31
	Cosine	10.28	6.73	11.43	8.43	11.29	7.94	5.55	7.05	8.75	7.00
	Matern32	11.41	6.95	8.25	5.64	6.74	8.36	4.64	5.86	8.87	6.68
	RBF + Periodic	14.34	7.22	8.29	5.61	7.21	9.06	4.67	6.06	9.65	6.50
December	Linear	7.66	6.28	3.72	4.64	5.98	8.63	7.85	8.76	8.61	10.24
	RBF	6.99	6.13	3.62	4.50	6.60	8.41	6.87	8.78	9.02	11.71
	Periodic	6.86	6.09	3.80	4.41	7.83	8.62	8.11	12.89	11.24	10.97
	Cosine	7.73	8.05	4.55	4.89	10.12	8.53	7.94	6.91	7.76	13.49
	Matern32	6.64	5.92	3.60	4.77	6.07	8.47	6.72	9.09	8.63	11.83
	RBF + Periodic	12.05	6.13	5.43	4.73	6.60	8.85	28.80	9.95	8.62	11.69

Table 14 Performance evaluation of (RBF + Per. Exponential + Periodic) kernel on an EWMA (α = 0.3) smoothed dataset for arrival delay prediction across different flight routes.

Month	Origin Airport and Arrival Delay MAE									
	ATL	CLT	DEN	LAS	LAX	LGA	MIA	MSP	ORD	SEA
January	5.63	6.67	4.34	5.58	8.67	10.44	7.68	64.88	10.05	13.94
February	7.95	6.30	5.02	4.81	9.60	7.76	5.00	26.55	9.06	6.37
March	4.93	4.80	6.45	5.60	25.54	9.68	9.56	5.26	13.19	13.29
April	6.42	7.53	10.71	7.75	14.95	24.23	12.35	5.39	5.77	7.17
May	8.93	4.50	7.2	5.15	7.23	12.12	11.51	4.68	8.10	7.95
June	11.19	9.33	4.49	3.84	5.24	25.48	14.28	6.87	15.365	4.91
July	9.04	18.46	6.47	8.98	13.27	7.68	7.22	17.31	13.98	6.91
August	9.38	5.71	4.74	4.62	5.57	9.54	10.91	14.66	8.17	12.38
September	7.33	7.37	5.37	5.46	4.69	9.71	7.07	11.72	11.10	6.15
October	4.81	4.64	3.65	3.82	8.10	24.29	14.02	23.58	6.30	12.77
November	3.46	13.99	9.48	12.13	5.50	13.04	4.32	5.64	7.64	9.30
December	9.23	37.83	4.71	5.23	6.80	10.70	5.91	24.11	7.56	20.61

under study exhibited a large MAE in predicted arrival delay for all flights heading to Dallas/Fort Worth International Airport (DFW) from airports in the northern region during the month of January. Such anomalies in prediction can be predominantly attributed to the marked delays during this month, a consequence of substantial snowfall across the northern territories. In February, a noticeable MAE was present in the prediction of flights departing from MSP. As we

advanced to March, a surge in prediction error became evident for flights from the West Coast, specifically from hubs like LAX and SEA, with the MAE larger than 10 minutes. This trend evolved in April, as predicted arrival delays for flights from airports, including DEN, LAX, LGA, MIA, and ORD, had large MAEs. However, in May, the pronounced MAE was largely limited to the prediction of flights operating from the East Coast to DFW. By June, the flights arriving from ATL, LGA, MIA, and ORD became hotspots for significant MAE figures.

The large MAE shifted in July with a prediction for flights from CLT, LAX, MSP, and ORD, having substantial errors. This pattern of increased MAEs for flights from MSP persisted and became larger in August. In September, both MSP and ORD flights had huge differences in predicted and actual delays. October's delay prediction revealed larger prediction inaccuracies for flights departing LGA, MIA, MSP, and SEA. In November, substantial MAE values were associated with flights from CLT, LAS, and LGA. In addition, the month of December had larger MAEs on flights originating from CLT, MSP, and SEA. Overall, approximately 80% of the predictions maintained an MAE under 10 minutes.

From the above observation, it can be inferred that the proposed kernel, integrating RBF, Periodic Exponential, and Periodic kernel, demonstrates capability in predicting flight delays for American Airlines across diverse routes. However, during months affected by adverse weather conditions or witnessing increased passenger volume, the kernel's performance might be sub-optimal due to the noisy training data. But, with appropriate hyperparameter tuning, its predictive capability can be enhanced for specific flight scenarios.

F. Departure Delay Prediction for American Airlines Flight (DFW-SEA)

In this section, we focused on predicting departure delays for an American Airlines flight that was scheduled to fly from Dallas/Fort Worth International Airport (DFW) to Seattle-Tacoma International Airport (SEA) between the hours of 10:00 AM and 11:00 AM. We have taken data from 2016 to 2018 for training and testing purposes. For the departure dataset, there was a clear trend characteristic accompanied by a subtle periodic pattern as shown in Fig. 15. In the context of arrival delays, the training dataset was more dispersed compared to the departure delays, which in this instance remained largely within a range of ±10 minutes.

1. Kernel Performance Comparative Analysis

In this section, using a dataset of the DFW-SEA route, we assessed the performance of various kernels. From Table 15, we can observe that the Linear, RBF, and Matern32 kernels have similar performance. The maximum and minimum MAEs are highlighted with red and green boxes. In addition, the MAE for the (RBF + Periodic) kernel was 6.88 minutes, and the (RBF + Periodic Exponential + Periodic) kernel was 6.81 minutes. The results for the departure delay predictions are shown in Figure 16. From the figure, we can conclude that the noise present in the data doesn't significantly influence the departure delay prediction, as most of the predicted values lie in the areas of higher data

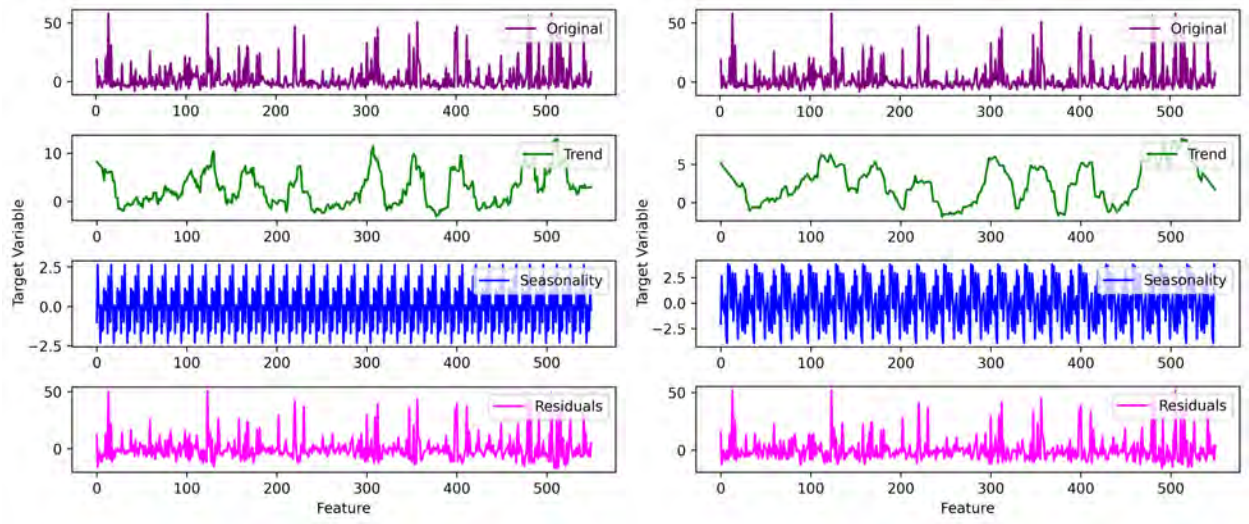


Fig. 15 Seasonal decomposition of training data for the DFW-SEA flight route, assuming a periodicity of (a) 15 days and (b) 30 days.

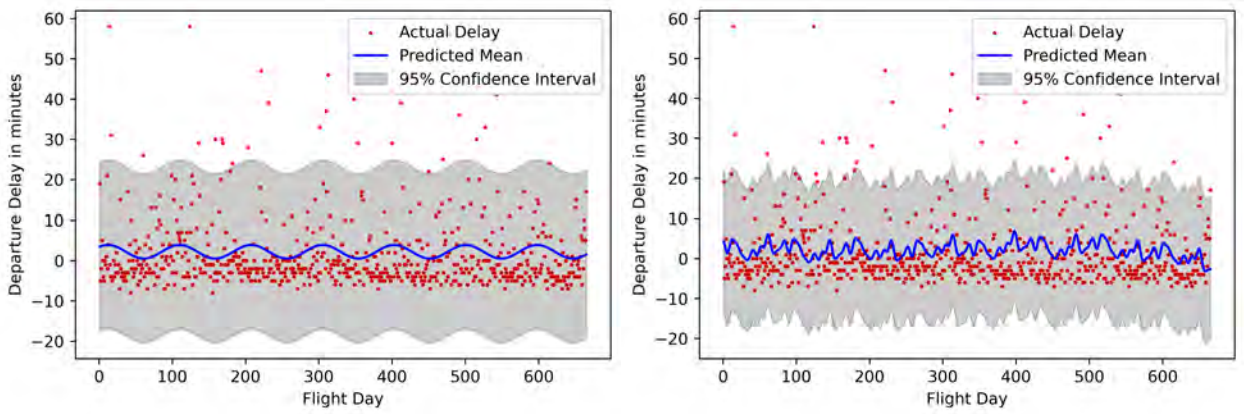
density.

Table 15 Performance evaluation of various kernels on a 550-day training dataset (normalized target variable) with a 7-day testing period for the DFW-SEA flight.

Kernels	Training Data				Testing Data			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
Linear	6.96	115.27	10.73	144.66	5.14	39.28	6.26	174.44
RBF	6.96	115.27	10.73	144.66	5.14	39.28	6.26	174.44
Periodic	6.85	112.05	10.58	143.74	4.78	38.05	6.16	149.64
Cosine	6.93	114.50	10.70	145.18	4.55	37.51	6.12	134.01
Matern32	6.96	115.27	10.73	144.66	5.14	39.28	6.26	174.44
RBF + Periodic	6.88	113.24	10.64	144.41	4.14	37.58	6.13	107.55
RBF + Per. Exponential + Periodic	6.81	109.83	10.48	147.49	3.71	30.71	5.54	132.46

We further investigated the RBF, (RBF + Periodic), and (RBF + Periodic Exponential + Periodic) kernels across different training horizons with a fixed testing horizon of 15 days and the results are presented in Table 16. From the results, we can observe that the training error for all kernels remains consistent across various training horizons. This can be attributed to the delays data being predominantly clustered in the same region, i.e., ±10 minutes. Subsequently, to determine the best testing horizon, we evaluated the RBF, (RBF + Periodic), and (RBF + Periodic Exponential + Periodic) kernels across different testing horizons as presented in Table 17.

From our analysis, it was evident that all three kernels could forecast departure delays up to 60 days in advance with an MAE of less than 7 minutes. The notably extended optimal testing horizon of 60 days can be attributed to the minimal noise present in the actual delays. Thus, we can conclude that GPR is suitable for predicting departure delays.



(a) **(b)**

Fig. 16 Departure delay prediction for the DFW-SEA flight with 650 days of training data (normalized) and a 15-day testing period using (a) (RBF + Periodic) and (b) (RBF + Periodic Exponential + Periodic) kernels.

Table 16 Performance evaluation of different kernels on varied training datasets for the DFW-SEA flight.

Training Horizon (Days)	RBF				RBF + Periodic			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
250	6.62	99.55	9.97	148.29	6.62	99.55	9.97	148.29
350	6.37	95.40	9.67	143.64	6.37	95.40	9.76	143.64
450	6.56	102.15	10.10	143.13	6.45	98.56	9.92	142.45
650	6.96	115.27	10.73	144.66	6.80	113.24	10.64	144.41
750	6.96	112.11	10.58	145.58	6.89	109.93	10.48	144.03

Training Horizon (Days)	RBF + Per. Exponential + Periodic			
	MAE	MSE	RMSE	MAPE
250	6.63	99.74	9.98	149.17
350	6.14	91.64	9.57	142.83
450	6.38	98.86	9.94	142.95
650	6.81	109.83	10.48	147.49
750	6.75	104.24	10.29	151.334

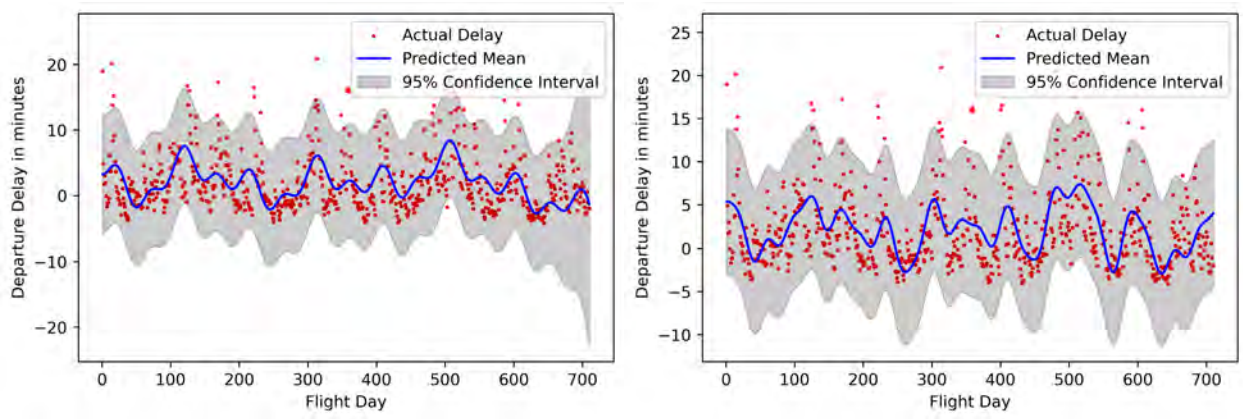
However, further efforts are required to reduce the MAE.

Furthermore, we implemented the EWMA method to smoothen the data and remove outliers. The departure delay prediction using the preprocessed data is shown in Fig. 17. The findings indicate that using EWMA-smoothed data allows for more precise predictions of departure delays compared to the unprocessed data. The (RBF + Periodic) kernel was able to forecast the moving average of departure delays for 60 days with an MAE of 3.18 minutes. According to Table 18, the (RBF + Periodic Exponential + Periodic) kernel yielded an MAE of 1.81 minutes over a testing horizon of 10 days. Remarkably, even with an extended testing horizon of 60 days, the MAE only rose to 3.42 minutes, which indicated a good predictive performance. Hence, we can conclude that the prediction of departure delays is effectively enhanced using the EWMA method.

Table 17 Evaluation of different kernels on a 650-day training dataset with varied testing intervals for the DFW-SEA flight.

Testing Horizon (Days)	RBF				RBF + Periodic			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
10	5.08	36.45	6.03	144.73	4.50	35.32	5.94	102.79
20	6.54	66.47	8.15	142.85	6.28	67.50	8.21	118.15
30	5.85	51.75	7.19	166.52	5.81	53.74	7.33	155.69
60	6.25	63.32	7.95	144.03	6.58	67.48	8.21	152.88

Testing Horizon (Days)	RBF + Per. Exponential + Periodic			
	MAE	MSE	RMSE	MAPE
10	4.60	37.24	6.10	128.48
20	6.58	88.71	9.41	111.48
30	5.07	63.23	7.95	94.86
60	5.60	75.88	8.71	107.08



(a) **(b)**

Fig. 17 DFW-SEA departure delay prediction with 650 days of training (smoothed and normalized) and 15-day testing data using (a) (RBF + Periodic) and (b) (RBF + Periodic Exponential + Periodic) kernels.

Table 18 Evaluation of different kernels on an EWMA (α = 0.3) smoothed 650-Day training dataset with varied testing intervals for the DFW-SEA flight.

Testing Horizon (Days)	RBF + Periodic				RBF + Per. Exponential + Periodic)			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
10	2.47	9.98	3.16	177.47	1.86	6.75	2.59	105.17
20	3.91	23.05	4.80	260.94	3.05	16.10	4.01	141.20
30	3.70	21.50	4.63	271.06	2.93	14.64	3.82	135.27
60	3.18	18.34	4.28	192.34	3.42	17.03	4.12	220.56

G. Departure Delay Prediction for American Airlines Flights on Various Flight Routes

In this section, we first studied the performance of different kernels for predicting departure delays of American Airlines flights across various sectors. The flight sector under study includes departure from Dallas/Fort Worth

International Airport (DFW) to ten of the busiest airports in the United States: Hartsfield-Jackson Atlanta International Airport (ATL), Denver International Airport (DEN), O'Hare International Airport (ORD), Los Angeles International Airport (LAX), John F. Kennedy International Airport (JFK), Harry Reid International Airport (LAS), Miami International Airport (MIA), Charlotte Douglas International Airport (CLT), Seattle-Tacoma International Airport (SEA), and LaGuardia Airport (LGA) as shown in Fig. 18. All the flights under investigation were scheduled to depart from DFW within the time frame of 3:00 PM to 6:00 PM. Subsequently, we employed (RBF + Periodic Exponential + Periodic) kernel to study its effectiveness across all flight routes. In the experimental setup, the dataset utilized for training adhered to the methodology defined in the preceding section. For each scenario, the initial hyperparameters of the kernel, as well as the smoothing parameter (α) for the EWMA, were kept constant.

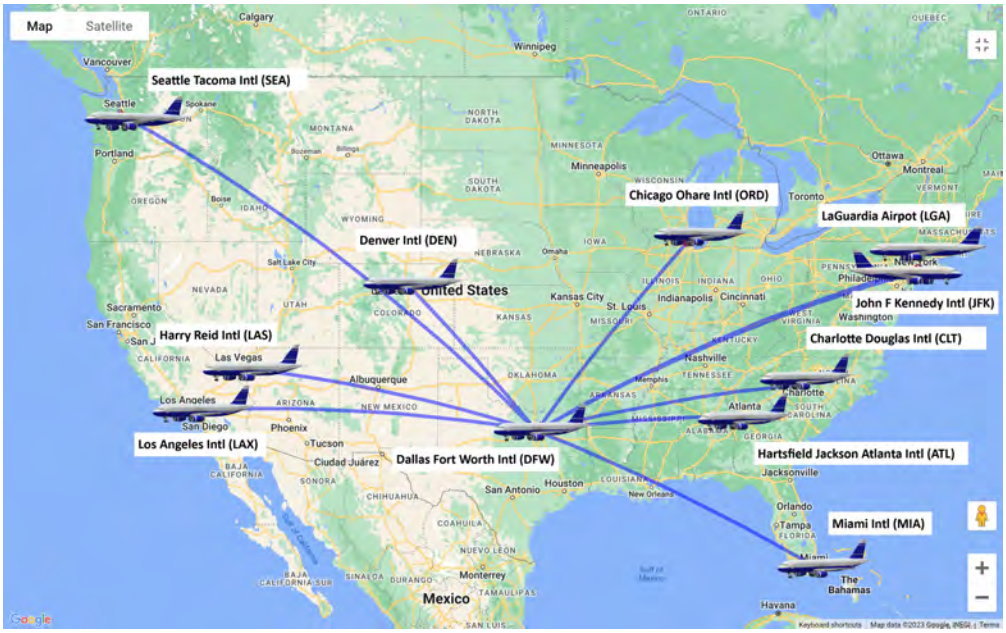


Fig. 18 Map representation of American Airlines flight route for predicting departure delays.

1. Performance Comparison of Various Kernels on Different Flight Routes

In the presented study, an examination of various kernels was undertaken to determine their efficacy in forecasting departure delays over a range of flight routes. The comparative performance of these kernels across different flight routes is tabulated in Table 19, 20, 21, and 22. For each month, the maximum and minimum MAEs are highlighted in red and green boxes. During the months of January and February, all kernels had similar performance across all flight routes, registering an MAE below the 10-minute threshold for departure delay predictions. However, in March, the (RBF + Periodic) kernel began to exhibit deviations, with MAEs of 13.39, 18.72, and 12.82 minutes for the flight to ATL, JFK, and LAS respectively. In the subsequent month of April, it was observed that only the routes to DEN and LGA demonstrated suboptimal performance specifically under the influence of the (RBF + Periodic) kernel. In

the month of May, only the DFW-LAX route breached the 10-minute MAE mark. A large anomaly was observed in June, where flights destined for MIA consistently reported MAEs over 10 minutes for most of the kernel. Concurrently, the LGA route approximated an MAE of 10 minutes with all the kernels. The DFW-ATL flight route had large MAE, specifically when implementing the Linear and Periodic kernels. In July the (RBF + Periodic) kernel recorded an MAE of 13.20 minutes for flights bound for DEN, whereas the Linear kernel reported an MAE of 10.44 minutes on the ORD route. Interestingly, a consistent performance was observed from August through November, with all kernels yielding satisfactory results. However, in December, the (RBF + Periodic) kernel logged an MAE of 11.45 minutes for MIA-bound flights.

Table 19 Performance evaluation of various kernels on an EWMA (α = 0.3) smoothed dataset for departure delay prediction across different flight routes (January-March).

Month	Kernel	Destination Airport and Departure Delay MAE									
		ATL	CLT	DEN	JFK	LAS	LAX	LGA	MIA	ORD	SEA
January	Linear	3.64	3.70	3.74	3.65	4.92	6.71	5.09	4.74	3.99	6.71
	RBF	3.55	3.78	3.92	3.62	4.59	6.51	5.10	4.59	3.96	6.51
	Periodic	3.73	4.17	3.67	3.62	5.11	5.32	5.90	5.48	5.42	5.32
	Cosine	5.50	3.52	4.35	3.81	5.04	6.50	5.04	4.98	3.81	6.50
	Matern32	3.50	3.81	4.15	3.66	4.46	6.58	5.16	4.68	3.95	6.58
	RBF + Periodic	4.50	3.91	3.93	3.74	5.47	5.32	6.12	4.84	3.62	5.32
February	Linear	5.76	4.51	6.89	5.33	4.41	3.73	4.91	5.14	4.73	3.73
	RBF	5.96	4.52	6.39	5.04	4.52	3.85	4.88	5.32	4.55	3.85
	Periodic	7.26	4.87	4.03	7.22	2.84	3.98	4.62	5.05	6.14	3.98
	Cosine	8.67	4.53	7.21	8.14	5.68	3.52	4.65	4.24	3.83	3.52
	Matern32	5.93	4.51	6.07	4.94	4.43	3.74	4.85	5.27	4.52	3.74
	RBF + Periodic	4.37	4.68	4.30	8.93	4.47	4.24	4.62	5.32	4.84	4.24
March	Linear	3.10	4.46	5.70	3.83	5.55	4.82	6.00	3.98	5.57	4.82
	RBF	3.22	4.62	5.37	3.53	5.25	4.63	6.05	4.08	4.91	4.63
	Periodic	4.00	4.86	4.91	3.13	5.25	4.71	5.43	4.29	5.64	4.71
	Cosine	4.98	4.50	5.90	4.07	5.18	4.56	5.51	4.64	5.51	4.56
	Matern32	3.17	4.67	5.38	3.35	5.27	4.63	5.99	4.00	5.01	4.63
	RBF + Periodic	13.39	4.85	5.37	18.72	12.82	4.90	5.72	3.78	4.91	4.90

From the above results, it became evident that different kernels could exhibit superior performance in predicting flight departure delays during different months. Consequently, there arises a need to investigate a composite kernel that amalgamates features from these diverse kernels, to assess its potential effectiveness. To address this objective, the subsequent section of our study is dedicated to evaluating the (RBF + Periodic exponential + Periodic) kernel.

2. Performance Comparison of (RBF + Periodic exponential + Periodic) kernel on Different Flight Routes

The methodology detailed in the preceding section was consistently applied throughout this section as well. The MAE in minutes pertaining to predictions across different flight routes is shown in Table 23.

For the months of January and February, the MAE remained under 7 minutes for the departure delay predictions

Table 20 Performance evaluation of various kernels on an EWMA (α = 0.3) smoothed dataset for departure delay prediction across different flight routes (April-June).

Month	Kernel	Destination Airport and Departure Delay MAE									
		ATL	CLT	DEN	JFK	LAS	LAX	LGA	MIA	ORD	SEA
April	Linear	3.75	7.39	4.30	3.83	4.81	5.11	6.65	4.90	3.04	5.11
	RBF	3.81	7.56	4.52	3.86	4.63	5.14	6.42	4.68	3.32	5.14
	Periodic	3.27	9.96	6.24	4.09	5.90	5.41	8.75	5.12	4.74	5.41
	Cosine	3.78	7.91	5.13	3.83	6.99	4.86	9.14	4.90	2.89	4.86
	Matern32	3.84	7.57	4.51	3.90	4.60	5.12	6.35	4.67	3.33	5.12
	RBF + Periodic	3.81	8.16	10.39	3.77	4.63	5.44	16.15	4.68	3.32	5.44
May	Linear	4.06	3.94	3.99	6.74	4.52	4.32	7.16	4.16	4.11	4.32
	RBF	3.99	4.01	4.72	6.71	4.55	4.45	6.49	3.96	4.09	4.45
	Periodic	3.49	3.99	5.86	7.36	3.67	3.98	6.51	4.84	4.69	3.98
	Cosine	3.91	3.89	4.34	5.85	5.83	5.77	6.74	4.16	4.30	5.77
	Matern32	4.02	3.95	5.19	6.73	4.49	4.59	6.38	3.99	4.14	4.59
	RBF + Periodic	3.99	3.88	6.04	7.28	4.55	12.21	8.81	3.96	4.78	12.21
June	Linear	10.16	8.97	4.73	4.37	6.64	5.93	9.40	11.01	6.14	5.93
	RBF	9.77	9.08	5.06	4.40	5.59	5.86	9.90	11.17	6.31	5.86
	Periodic	11.06	8.70	6.90	5.17	5.24	6.13	10.94	9.37	6.14	6.13
	Cosine	8.74	8.97	4.61	4.43	6.15	5.86	10.47	11.01	5.17	5.86
	Matern32	9.80	9.04	5.12	4.44	5.45	5.78	10.09	11.09	6.36	5.78
	RBF + Periodic	9.69	9.47	4.96	4.40	5.59	9.92	9.90	11.17	6.31	9.92

Table 21 Performance evaluation of various kernels on an EWMA (α = 0.3) smoothed dataset for departure delay prediction across different flight routes (July-September).

Month	Kernel	Destination Airport and Departure Delay MAE									
		ATL	CLT	DEN	JFK	LAS	LAX	LGA	MIA	ORD	SEA
July	Linear	5.56	3.56	3.34	4.92	3.90	5.80	5.79	5.37	10.44	5.80
	RBF	5.35	4.05	3.48	4.87	3.84	5.61	5.71	5.42	9.65	5.61
	Periodic	6.13	4.12	3.74	4.83	4.02	6.02	5.96	8.79	9.88	6.02
	Cosine	4.98	3.77	3.69	6.18	4.28	5.81	5.79	4.92	8.08	5.81
	Matern32	5.36	3.48	3.43	4.99	4.00	5.87	5.74	5.67	9.28	5.87
	RBF + Periodic	5.50	3.52	13.20	4.82	3.84	5.61	6.04	7.60	9.66	5.61
August	Linear	3.76	4.55	8.30	4.07	4.24	8.79	5.34	4.49	8.54	8.79
	RBF	3.73	3.82	8.05	3.97	4.21	8.71	5.39	4.91	7.31	8.71
	Periodic	3.86	3.09	8.61	4.76	4.37	8.44	5.74	5.40	9.53	8.44
	Cosine	4.33	4.97	7.95	4.46	4.18	8.79	5.34	4.41	8.04	8.79
	Matern32	3.70	3.85	8.12	3.90	4.09	8.65	5.51	4.82	6.92	8.65
	RBF + Periodic	3.73	6.19	8.05	4.84	4.12	7.18	5.39	4.91	8.72	7.18
September	Linear	5.64	4.00	3.81	4.72	2.63	5.64	5.70	5.35	5.95	5.64
	RBF	5.51	3.81	4.01	4.47	2.55	5.06	5.63	4.43	5.09	5.06
	Periodic	5.98	4.50	3.61	5.99	3.13	4.33	5.86	4.26	3.74	4.33
	Cosine	5.74	3.72	4.01	3.45	2.98	6.13	5.58	5.75	7.34	6.13
	Matern32	5.60	3.79	3.94	4.42	2.55	4.94	5.56	4.21	4.93	4.94
	RBF + Periodic	5.25	3.74	4.19	3.48	4.73	4.93	5.63	4.20	9.44	4.93

Table 22 Performance evaluation of various kernels on an EWMA (α = 0.3) smoothed dataset for departure delay prediction across different flight routes (October-December).

Month	Kernel	Destination Airport and Departure Delay MAE									
		ATL	CLT	DEN	JFK	LAS	LAX	LGA	MIA	ORD	SEA
October	Linear	4.76	5.18	6.07	4.35	5.28	5.22	5.86	4.52	5.45	5.22
	RBF	4.41	4.50	6.06	4.04	5.13	5.03	5.35	3.96	5.04	5.03
	Periodic	4.91	5.34	3.76	4.09	4.69	4.57	4.49	4.86	5.84	4.57
	Cosine	4.10	4.00	7.57	4.15	5.57	5.45	5.86	4.85	3.89	5.45
	Matern32	4.29	4.54	6.02	3.90	5.14	5.09	5.17	4.05	4.91	5.09
	RBF + Periodic	4.41	4.72	6.06	4.10	5.13	4.64	4.50	4.59	5.23	4.64
November	Linear	2.83	5.93	4.47	3.28	3.75	6.02	7.04	4.75	4.09	6.02
	RBF	3.11	5.57	4.35	2.99	3.72	5.52	7.03	4.73	4.08	5.52
	Periodic	2.63	4.54	7.63	3.23	5.25	7.26	7.53	4.91	6.38	7.26
	Cosine	2.44	4.46	3.42	3.28	2.97	5.20	7.04	4.77	2.32	5.20
	Matern32	3.20	5.57	4.34	2.86	3.69	5.36	6.97	4.73	3.98	5.36
	RBF + Periodic	3.70	5.57	4.32	3.18	3.72	5.43	7.03	4.24	3.97	5.43
December	Linear	5.19	6.03	3.21	3.79	3.98	6.71	6.03	4.36	4.72	6.71
	RBF	4.94	6.00	3.26	3.66	4.09	6.32	5.51	4.10	4.25	6.32
	Periodic	4.95	5.31	3.63	3.45	4.53	5.51	7.63	5.57	5.61	5.51
	Cosine	5.79	7.86	2.50	3.72	3.59	6.50	6.27	4.48	4.23	6.50
	Matern32	4.98	5.99	3.21	3.66	4.03	6.13	5.32	4.01	4.18	6.13
	RBF + Periodic	5.25	6.00	4.16	3.71	4.08	5.75	5.67	11.25	5.74	5.75

Table 23 Performance evaluation of (RBF + Per. Exponential + Periodic) kernel on an EWMA (α = 0.3) smoothed dataset for departure delay prediction across different flight routes.

Month	Destination Airport and Departure Delay MAE									
	ATL	CLT	DEN	JFK	LAS	LAX	LGA	MIA	ORD	SEA
January	4.45	3.35	5.22	5.65	5.44	6.41	4.43	5.90	4.85	6.41
February	7.22	4.97	1.99	6.94	5.33	3.14	6.67	7.30	5.88	3.14
March	5.77	5.44	7.83	3.18	12.77	5.13	6.06	5.74	6.87	5.13
April	5.39	11.54	8.59	4.83	2.62	9.83	9.65	8.49	4.71	9.83
May	6.84	10.56	9.21	7.99	4.23	4.54	7.08	6.39	4.65	4.54
June	7.42	9.64	5.34	6.66	6.59	8.36	8.31	10.84	5.72	8.37
July	4.94	17.69	4.24	5.50	11.04	10.17	15.94	7.34	6.18	10.17
August	8.18	2.50	11.59	3.56	4.26	7.49	9.57	4.11	7.30	7.49
September	7.11	4.47	3.46	4.64	3.29	6.75	12.18	5.08	5.42	6.75
October	4.56	4.27	4.66	5.04	7.06	4.53	3.75	6.98	4.88	4.53
November	4.29	6.04	17.97	5.54	8.22	3.49	11.94	9.54	6.01	3.49
December	10.41	8.05	2.38	4.76	10.88	3.75	12.51	7.96	5.46	3.75

of flights to various destinations. However, in March, the prediction for flights flying to LAS had an MAE of 12.77 minutes. Subsequently, flights to CLT showed a notable MAE of 11.54 minutes in April and 10.56 minutes in May. In June, the MAE exceeded 10 minutes for flights from DFW to MIA which can be due to increased passenger volume on the DFW-MIA route during the summer months, which potentially introduces noise into the training data. In July, we

observed larger prediction errors for flights to several airports, including CLT, LAS, LAX, LGA, and SEA. In the month of August, increased prediction errors were observed for flights en route to DEN. This trend shifted in September when flights destined for New York airport exhibited significant discrepancies. Both these routes persisted in demonstrating prediction anomalies in November. In December, the prediction error was larger for flights heading to ATL and LAS. From January to December, almost 89% of the predictions exhibited an MAE of less than 10 minutes. Based on these observations, it can be concluded that the proposed kernel, which integrates RBF, Periodic Exponential, and Periodic kernel, is capable of predicting departure delays for American Airlines across a variety of routes. However, through optimal hyperparameter adjustments, the model's predictive accuracy can be further refined for particular flight conditions.

VI. Discussion of Results

From the above results, it is evident that there is no single choice kernel to accurately predict delays for all the flights using GPR. A composite kernel formed by combining RBF, Periodic Exponential, and Periodic kernel can perform well for most flights if we need to understand the seasonal pattern. The combined kernel can capture non-periodic trends using RBF, strictly periodic characteristics using the Periodic kernel, and periodic behavior that decays with time using the Periodic Exponential kernel. This is the reason the combined kernel outperforms in most cases. However, when the delay pattern is erratic, and training data is too noisy, the MAE for predicted delay can be very large. In such a case, a non-stationary kernel needs to be introduced.

The predicted arrival delay MAE was particularly high in those months where the volume of passengers changes significantly and months with extreme weather. For east coast airports, including ATL and MIA, the prediction MAE increased during the summer months, which is obvious as this airport has inconsistent passenger volume in the summer months because Florida is a popular summer destination, which makes the training data noisy. A similar event was seen in MSP, ORD, and SEA prediction during the month of January and February. During this time, these airports have large amounts of snowfall, which makes flight delays more frequent. As a result, the model can't capture the actual pattern due to outliers in the training data. Furthermore, the prediction MAE for most airports is small for the late spring and fall. This is evident because, during the months of late spring and fall, there is no extreme weather that can affect the training data. The MAE of MSP airport in January was 64.88 minutes, which indicates that the combined kernel completely failed to capture the arrival delay pattern for the MSP-DFW route. Meanwhile, the MAE for the ATL-DFW flight was 3.46 minutes in November, which shows how effective the combined kernel can be with less noisy training data.

In the case of departure delay prediction from DFW, the predicted delay MAE was significantly large in the month of July. As per statistics, July is one of the peak months with an increased number of passengers for summer travel. This increased MAE makes sense because we are using limited training data, from which the model captured peak delays of

the past month, which may not be the case for the month of July, thus causing a difference in predicted and actual delay. Although the MAE for arrival delay was larger for ORD and MSP during January and February, the MAE for departure delay is below 6 minutes for both airports because DFW doesn't have snowfall during this month, and it doesn't affect the departure delay training data. In addition, the predicted MAE of departure delays is less than 10 min for most of the flights throughout the year, which points out that DFW airport and American Airlines are functioning effectively on the departure to keep the delays as low as possible.

VII. Conclusion

In this paper, we presented a machine-learning method for flight delay prediction. The framework, developed using GPR, demonstrated high efficacy in its predictions. The training data for particular flights were obtained from the Bureau of Transportation Statistics (BTS) website. We implemented the proposed method to predict arrival and departure delays at Dallas/Fort Worth International Airport on flights arriving from or departing to ten of the busiest airports in the USA. The MAE of the prediction, when assessed across various flights, was within the threshold of 10 minutes. The method introduced can also be applied to forecast the time series outcomes. Specifically, our findings suggest that flight delays display both trend and periodic characteristics, which can be effectively captured using combined kernels made of RBF, Periodic Exponential, and Periodic kernels. Furthermore, the prediction's confidence interval can serve as an indicator of its reliability. These results underscore the considerable importance of GPR in forecasting time series data. The prediction model can be beneficial for airlines if they require prior information about the pattern of future delays. The airlines can work on mitigating the possible causes that may trigger this delay if operational factors.

VIII. Limitations and Future Work

The proposed method has few limitations in application to flight delay prediction. Considering numerous airlines frequently adjust their schedules for certain routes, our dataset for the examined flights was somewhat limited. This limitation made using complex kernels suboptimal, leading to a wider confidence interval in our predictions. Similarly, the complex kernel can fail due to numerical precision during matrix inversion, which can be avoided by adding jitter to the diagonal element of the covariance matrix. Also, the kernel selected for a particular flight might not perform well for another specific flight. In addition, the training data are assumed to be stationary, so the selected kernel can fail to capture peak delays if the actual training data is non-stationary. Since we have used the Exponentiated Weighted Moving Average (EWMA) approach to enhance data smoothness, if important information is smoothed out during this smoothing process, we may get a poor fit. Furthermore, the proposed GPR method has a computational cost of $O(N^3)$. Thus, the computational time increases significantly with an increase in training data.

However, there is potential for enhancing the output of the proposed method. The accuracy can be further improved by refining the confidence interval of the predictions. Also, other input features can be examined for feature importance

scores to evaluate if they significantly affect the delay prediction. The GPR can improve performance if other significant features are incorporated into the training process. In addition, the future scope of this research shall focus on identifying more effective methods for noise reduction in the training dataset. Also, a more computationally efficient Sparse Gaussian Process Regression can be implemented to reduce computational cost. The arrival delay of a particular flight can be affected by the departure delay of the same flight at the origin airport. The proposed application considered flight delay as a single-output problem; however, it can also be implemented as a multi-output problem where we can simultaneously predict the departure and arrival delay for a specific flight. The Multi-Output Gaussian Process can increase prediction accuracy if departure and arrival delays are correlated.

Acknowledgments

This work is supported by the National Aeronautics and Space Administration under Grant/Cooperative Agreement Number: 80NSSC21K1508.

References

[1] Britto, R., Dresner, M., and Voltes, A., "The impact of flight delays on passenger demand and societal welfare," *Transportation Research Part E: Logistics and Transportation Review*, Vol. 48, No. 2, 2012, pp. 460–469. doi:https://doi.org/10.1016/j.tre.2011.10.009, URL https://www.sciencedirect.com/science/article/pii/S1366554511001347.

[2] "On-Time Arrival Performance Dallas/Fort Worth, TX: Dallas/Fort Worth International (January - December, 2022) [Data File]," , 2023. URL https://www.transtats.bts.gov/ot_delay/OT_DelayCause1.asp.

[3] Freitas, C. J., "The issue of numerical uncertainty," *Applied Mathematical Modelling*, Vol. 26, No. 2, 2002, pp. 237–248. doi:https://doi.org/10.1016/S0307-904X(01)00058-0, URL https://www.sciencedirect.com/science/article/pii/S0307904X01000580.

[4] Tang, Y., Kurths, J., Lin, W., Ott, E., and Kocarev, L., "Introduction to Focus Issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 30, No. 6, 2020, p. 063151. doi:10.1063/5.0016505, URL https://doi.org/10.1063/5.0016505.

[5] Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., and Zhao, D., "Flight Delay Prediction Based on Aviation Big Data and Machine Learning," *IEEE Transactions on Vehicular Technology*, Vol. 69, No. 1, 2020, pp. 140–150. doi:10.1109/TVT.2019.2954094.

[6] Kim, Y. J., Choi, S., Briceno, S., and Mavris, D., "A deep learning approach to flight delay prediction," *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, 2016, pp. 1–6. doi:10.1109/DASC.2016.7778092.

[7] Yu, B., Guo, Z., Asian, S., Wang, H., and Chen, G., "Flight delay prediction for commercial air transport: A deep learning approach," *Transportation Research Part E: Logistics and Transportation Review*, Vol. 125, 2019, pp. 203–221.

doi:https://doi.org/10.1016/j.tre.2019.03.013, URL https://www.sciencedirect.com/science/article/pii/S1366554518311979.

[8] Esmaeilzadeh, E., and Mokhtarimousavi, S., "Machine Learning Approach for Flight Departure Delay Prediction and Analysis," *Transportation Research Record*, Vol. 2674, No. 8, 2020, pp. 145–159. doi:10.1177/0361198120930014, URL https://doi.org/10.1177/0361198120930014.

[9] Mueller, E., and Chatterji, G., "Analysis of aircraft arrival and departure delay characteristics," *AIAA's Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical Forum*, 2002, p. 5866. doi:10.2514/6.2002-5866.

[10] Yufeng Tu, M. O. B., and Jank, W. S., "Estimating Flight Departure Delay Distributions—A Statistical Approach With Long-Term Trend and Short-Term Pattern," *Journal of the American Statistical Association*, Vol. 103, No. 481, 2008, pp. 112–125. doi:10.1198/016214507000000257, URL https://doi.org/10.1198/016214507000000257.

[11] Wesonga, R., Nabugoomu, F., and Jehopio, P., "Parameterized framework for the analysis of probabilities of aircraft delay at an airport," *Journal of Air Transport Management*, Vol. 23, 2012, pp. 1–4. doi:https://doi.org/10.1016/j.jairtraman.2012.02.001, URL https://www.sciencedirect.com/science/article/pii/S0969699712000555.

[12] Pérez–Rodríguez, J., Pérez–Sánchez, J., and Gómez–Déniz, E., "Modelling the asymmetric probabilistic delay of aircraft arrival," *Journal of Air Transport Management*, Vol. 62, 2017, pp. 90–98. doi:https://doi.org/10.1016/j.jairtraman.2017.03.001, URL https://www.sciencedirect.com/science/article/pii/S0969699716304392.

[13] Vandal, T., Livingston, M., Piho, C., and Zimmerman, S., "Prediction and Uncertainty Quantification of Daily Airport Flight Delays," *Proceedings of The 4th International Conference on Predictive Applications and APIs*, Proceedings of Machine Learning Research, Vol. 82, edited by C. Hardgrove, L. Dorard, and K. Thompson, PMLR, 2018, pp. 45–51. URL https://proceedings.mlr.press/v82/vandal18a.html.

[14] Zoutendijk, M., and Mitici, M., "Probabilistic Flight Delay Predictions Using Machine Learning and Applications to the Flight-to-Gate Assignment Problem," *Aerospace*, Vol. 8, No. 6, 2021. doi:10.3390/aerospace8060152, URL https://www.mdpi.com/2226-4310/8/6/152.

[15] Wang, Z., Liao, C., Hang, X., Li, L., Delahaye, D., and Hansen, M., "Distribution Prediction of Strategic Flight Delays via Machine Learning Methods," *Sustainability*, Vol. 14, No. 22, 2022. doi:10.3390/su142215180, URL https://www.mdpi.com/2071-1050/14/22/15180.

[16] Palar, P. S., Zakaria, K., Zuhal, L. R., Shimoyama, K., and Liem, R. P., "Gaussian processes and support vector regression for uncertainty quantification in aerodynamics," *AIAA Scitech 2021 Forum*, 2021, p. 0181. doi:10.2514/6.2021-0181, URL https://arc.aiaa.org/doi/abs/10.2514/6.2021-0181.

[17] Cao, G., Lai, E. M.-K., and Alam, F., "Gaussian process model predictive control of an unmanned quadrotor," *Journal of Intelligent & Robotic Systems*, Vol. 88, 2017, pp. 147–162. doi:https://doi.org/10.1007/s10846-017-0549-y.

[18] Brahim-Belhouari, S., and Bermak, A., "Gaussian process for nonstationary time series prediction," *Computational Statistics & Data Analysis*, Vol. 47, No. 4, 2004, pp. 705–712. doi:https://doi.org/10.1016/j.csda.2004.02.006, URL https://www.sciencedirect.com/science/article/pii/S0167947304000301.

[19] Idé, T., and Kato, S., "Travel-Time Prediction using Gaussian Process Regression: A Trajectory-Based Approach," *Proceedings of the 2009 SIAM International Conference on Data Mining (SDM)*, ????, pp. 1185–1196. doi:10.1137/1.9781611972795.101, URL https://epubs.siam.org/doi/abs/10.1137/1.9781611972795.101.

[20] Amer, A. W., Roy, S., and Kopsaftopoulos, F., "Probabilistic SHM under varying loads via the integration of Gaussian Process Regression and physics-based guided-wave propagation models," *AIAA Scitech 2021 Forum*, 2021, p. 0434. doi:10.2514/6.2021-0434, URL https://arc.aiaa.org/doi/abs/10.2514/6.2021-0434.

[21] Mohanty, S., Chattopadhyay, A., Peralta, P., Das, S., and Willhauck, C., "Fatigue life prediction using multivariate gaussian process," *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 16th AIAA/ASME/AHS Adaptive Structures Conference, 10th AIAA Non-Deterministic Approaches Conference, 9th AIAA Gossamer Spacecraft Forum, 4th AIAA Multidisciplinary Design Optimization Specialists Conference*, 2008, p. 1837. doi:10.2514/6.2008-1837.

[22] Liu, M., Chowdhary, G., da Silva, B. C., Liu, S.-Y., and How, J. P., "Gaussian Processes for Learning and Control: A Tutorial with Examples," *IEEE Control Systems*, Vol. 38, No. 5, 2018, pp. 53–86. doi:10.1109/mcs.2018.2851010, URL https://doi.org/10.1109/mcs.2018.2851010.

[23] Duvenaud, D., "Automatic model construction with Gaussian processes," Ph.D. thesis, Apollo - University of Cambridge Repository, 2014. doi:10.17863/CAM.14087, URL https://www.repository.cam.ac.uk/handle/1810/247281.

[24] Williams, C. K., and Rasmussen, C. E., *Gaussian processes for machine learning*, Vol. 2, MIT press Cambridge, MA, 2006.

[25] Wang, J., "An Intuitive Tutorial to Gaussian Process Regression," *Computing in Science & Engineering*, Vol. 25, No. 4, 2023, pp. 4–11. doi:10.1109/MCSE.2023.3342149.

[26] Yadav, A., Kumar, A., Rana, R. P. S., Chandrakar, M., Pazoki, M., and El Sehiemy, R. A., "An Efficient Monthly Load Forecasting Model Using Gaussian Process Regression," *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, 2021, pp. 1–8. doi:10.1109/GUCON50781.2021.9574008.

[27] Liu, B., Kiskin, I., and Roberts, S., "An Overview of Gaussian process Regression for Volatility Forecasting," *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 681–686. doi:10.1109/ICAIIC48513.2020.9065045.

[28] Ebden, M., "Gaussian processes: A quick introduction," *arXiv preprint arXiv:1505.02965*, 2015.

[29] Murphy, K. P., *Machine learning: a probabilistic perspective*, MIT press, 2012.

[30] Wakefield, J., *Bayesian and frequentist regression methods*, Vol. 23, Springer, 2013. doi:https://doi.org/10.1007/978-1-4419-0925-1.

[31] Schulz, E., Speekenbrink, M., and Krause, A., "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of Mathematical Psychology*, Vol. 85, 2018, pp. 1–16. doi:https://doi.org/10.1016/j.jmp.2018.03.001, URL https://www.sciencedirect.com/science/article/pii/S0022249617302158.

[32] Khanal, A., Bhusal, R., Subbarao, K., Chakravarthy, A., and Okolo, W., "Open Source Toolbox for Generalized Polynomial Chaos Expansion-based Uncertainty Quantification for sUAS Traffic Operations," *AIAA SCITECH 2023 Forum*, ???? doi: 10.2514/6.2023-0151, URL https://arc.aiaa.org/doi/abs/10.2514/6.2023-0151.

[33] Paciorek, C., and Schervish, M., "Nonstationary Covariance Functions for Gaussian Process Regression," *Advances in Neural Information Processing Systems*, Vol. 16, edited by S. Thrun, L. Saul, and B. Schölkopf, MIT Press, 2003. URL https://proceedings.neurips.cc/paper_files/paper/2003/file/326a8c055c0d04f5b06544665d8bb3ea-Paper.pdf.

[34] GPy, "GPy: A Gaussian process framework in python," http://github.com/SheffieldML/GPy, since 2012.

[35] Khanal, A., Bhusal, R., Subbarao, K., Chakravarthy, A., and Okolo, W., "Learning Stochastic Processes Using Gaussian Processes: An Application to Flight Delay Prediction," *AIAA SCITECH 2023 Forum*, ???? doi:10.2514/6.2023-1257, URL https://arc.aiaa.org/doi/abs/10.2514/6.2023-1257.