



Prediction of Bus Passenger Traffic using Gaussian Process Regression

Vidya G S¹ · Hari V S¹

Received: 24 September 2021 / Revised: 28 April 2022 / Accepted: 5 May 2022 / Published online: 4 June 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The paper summarizes the design and implementation of a passenger traffic prediction model, based on Gaussian Process Regression (GPR). Passenger traffic analysis is the present day requirement for proper bus scheduling and traffic management to improve the efficiency and passenger comfort. Bayesian analysis uses statistical modelling to recursively estimate new data from existing data. GPR is a fully Bayesian process model, which is developed using PyMC3 with Theano as backend. The passenger data is modelled as a Poisson process so that the prior for designing the GP regression model is a Gamma distributed function. It is observed that the proposed GP based regression method outperforms the existing methods like Student-t process model and Kernel Ridge Regression (KRR) process.

Keywords Gaussian process regression · Bayesian analysis · PyMC3 · Poisson process · Gamma prior · Student-t

1 Introduction

Public transport system is the mode of travel for the common man. With the urbanization and population explosion, traffic management is becoming a challenging task. It undergoes a lot of changes in the recent times towards automation and intelligent fleet management. Artificial Intelligence and Machine learning based systems will revolutionize future transportation. One key aspect of automation is the prediction of passenger traffic to reduce congestions, accident rates, waiting time and to increase passenger comfort. Accurate passenger prediction helps in proper bus and crew scheduling. Thus the constrained services can be effectively managed, thereby improving the quality of the service.

The real time management of bus services enhances the profitability of the service providers and reduces the waiting time of passengers. The more efficient the public transport, the more the common man gets attracted towards it, reducing the usage of private cars, traffic congestion and pollution. The passenger traffic is a non stationary stochastic

process. The morning traffic is mostly contributed by working persons and the student community. The noon traffic includes the common man who is travelling for their daily needs and the post noon session covers all those who are returning home after their work. The bus accessibility, type of traffic, land use, trip timing are all the influential factors affecting passenger flow. The analysis of public transport systems include geospatial data analysis, ticket data analysis and crew analysis. Often, such data are stochastic which requires statistical modelling and characterization.

In the present study, a set of Pune Mahanagar Parivahan Mahamandal Limited (PMPML) bus routes are analyzed with the help of Gaussian Process Regression (GPR) with the objective to maximize the earning per kilometer (*epk*) and passenger comfort. The GPR is a powerful class of Machine learning algorithm that relies on a few parameters to make predictions making it an ideal candidate for predicting the passenger traffic with limited data.

2 Literature Review

A set of models including Auto Regressive Moving Average models (ARMA), error correction models (VECM and EC-VARMA), space time ARMA (STARMA), and multivariate autoregressive space state models (MARSS) are used in the United States [1] for traffic analysis and prediction. GPS data collected from private cars is used to model a multivariate state space ARIMA model [2, 3] to forecast travel times in

✉ Vidya G S
vidyahari.gs@gmail.com

Hari V S
hariivs@ceconline.edu

¹ Department of Electronics, College of Engineering
Chengannur, A P J Abdul Kalam Technological University,
Kerala 689121 Thiruvananthapuram, India

many countries. Short term passenger data prediction, [4] based on time series data, achieved better prediction performance. Neural networks [5] are employed to predict traffic data with the trend towards intelligent systems.

GPR has been found to be efficient on several multi-dimensional regression tasks [6]. GPR is one of the new machine learning method in the field of Bayesian Statistics [7]. Gaussian Processes (GPs) have been applied successfully to the problem of learning predictive state models [8]. GPs are now being used in modelling traffic data and prediction [9] as the traffic pattern exhibits a random behaviour. The wide range of applications in which GPR is used includes controlling MLS of Maglev train [10], rainfall prediction [11], solar PV panel power prediction [12], predicting covid-19 pandemic confirmed cases in South Korea [13] and assessing the energy behaviour of RC shear walls [14]. Several technologies are being incorporated to improve data collection [15], energy management [16] and traffic efficiency [17]. The Gaussian process regression technique is now widely used in many areas like rural traffic prediction [18], traffic modelling in high speed networks [19], Stock price forecasting [20], energy measurement and verification [7], soil moisture modelling [21], predicting multi-step ahead SPEI drought index [22], weather forecasting [23] and even in big data analysis [24].

The Gaussian process regression is currently being used in the prediction of both passenger and data traffic. The GPR technique is used in air traffic for quantifying data driven trajectory accuracy and uncertainty in Flight trajectory predictions [25, 26]. In maritime traffic, GPR is used for traffic probabilistic prediction based on ship motion pattern extraction [27] that includes both destination and trajectory prediction. In rail transport, the average speed is one of the most critical factors affecting travel time between two access points. GPR technique is used for the estimation of the average speed for a railway signalling system [28]. Considering the various aspects of traffic management, the global speed prediction [29] and individual longitudinal driving behaviour description of drivers [30], based on historical data, is done using multiple Gaussian process regression for different driving styles.

3 Theory

GPR is a supervised machine learning technique that provides a mapping from input to output [31]. The GPR techniques are more application specific than general regression models. The training procedure adapted by GPR process is that it considers the whole training data set at each time the model makes a prediction. Thus the adversarial problems are completely eliminated when compared with traditional machine learning approaches. The response of GPR model

is robust against slight variations in the passenger data. The main benefit of GPR is that it works well with small datasets and provide uncertainty measurements on the predictions. The training process using GPR model is relatively simpler compared to other ML models, since the hyperparameters of GPR are adaptively obtained by maximum likelihood estimation process. The GPR based models can easily overcome the overlearning or other data driven issues, since the training process considers the whole set of training data. The predictive accuracy, that is quantitatively constructed is least affected by any Gaussian noise [32]. Thus GPR is selected for the prediction of passenger traffic.

3.1 Passenger Arrival Model

It is required to maintain uniform number of passengers in all stages to increase the passenger comfort and revenue. The passenger arrival is modelled to accomplish this, so that proper bus scheduling is done. The arrival of passengers is conjectured as a sum [33] of several time varying Poisson process [3], where the probability of x passengers in a given period given by

$$P(x, \lambda_i) = \sum_{i=1}^n \frac{e^{-\lambda_i} \lambda_i^x}{x!} \quad (1)$$

Here $\lambda(t)$ is the average passengers per interval which is a random process. The passengers arriving at various stops choose buses for specific routes accordingly, that account for the bus passenger traffic, whose stochastic nature is analyzed below.

3.2 Passenger Traffic Model

The passenger traffic at any instant shows random behaviour and does not depend on the previous passenger history, making it a Markovian process. Linear estimation models like Kalman filter [34], Bayesian network based filters like particle filter [35] are used to model passenger traffic. The theory behind these bayesian filters (Kalman and Particle) is the estimation of posterior distribution of the passenger data by iteratively computing over the state of a dynamical system as

$$p(y_k | z_{1:k}, u_{1:k-1}) \propto p(z_k | y_k) \times \int p(y_k | x_{k_1}, u_{k_1}) p(y_{k_1} | z_{1:k}) dy_{k_1} \quad (2)$$

Here y represents the true state and z represents the measurements of a Hidden Markov Model (HMM). The true state y is an unobserved Markov process obtained from the passenger data, and the measurements z are observations of HMM, represent the passenger count. The parameter $u_{1:k-1}$ are the initial control measures or hyperparameters. The prediction model described by Eq. 2, $p(z_k | y_k)$ represents the likelihood

of the passenger data and $\int p(y_k|y_{k1}, u_{k1})p(y_{k1}|z_{1:k})dy_{k1}$ indicates the prior, whose product results in the posterior distribution. Generally, these filter models are parametric representation of the processes involved [36]. The major challenge that is faced in parametric models is that, they are not intelligent enough to reproduce all the aspects of a dynamical system. So the need of the hour is a non-parametric Bayesian model. Gaussian process regression(GPR) overthrows these parametric models and learns the prediction and observation models dynamically.

3.3 Gaussian Process Regression

Gaussian processes are non-parametric models that define a distribution over the mapping function $f(x)$. The mapping function is a random variable of the input passenger data X . The distribution [37] is over infinite dimensions, but the computations are done using limited finite resources. This property makes GPs suitable for many spatial and temporal prediction tasks. Let x be the variable associated with the mapping function and x' be its time shifted version. A GP prior on the mapping function $f(x)$ is completely specified by its mean and covariance as

$$f(x) \sim GP(m(x), k(x, x')) \quad (3)$$

Here GP represents a Gaussian process distribution with mean $m(x)$ and $k(x, x')$ being the kernel or covariance function. $k(x, x')$ gives the covariance of $f(x)$ and $f(x')$, and $k(x, x)$ gives the variance of $f(x)$. The regression task aims at predicting the target $f(x_*)$ for new passenger data x_* , that does not belong to the input data set D (where $D \in X$). To predict this value, GPR model uses the GP prior on the mapping function $f(x)$, along with the knowledge provided by the data set D to calculate the posterior distribution over the mapping function. Finally, it makes inferences about the conditional distribution of the function value at x_* .

The generalization properties of GPs rest almost entirely within the choice of kernels. The kernel specifications are computed using the passenger dataset and not inbuilt ones. The passenger data, modelled as a discernible time series, is used for modelling the prior which in turn is employed for computing the kernel parameters. The joint distribution of $f(x)$ and $f(x_*)$ is multivariate normal represented as

$$\begin{bmatrix} f(x) \\ f(x_*) \end{bmatrix} \sim N\left(\begin{bmatrix} m(x) \\ m(x_*) \end{bmatrix}, \begin{bmatrix} k(x, x')k(x_*, x) \\ k(x_*, x)k(x_*, x') \end{bmatrix}\right) \quad (4)$$

The conditional distribution is then calculated for predicting the function values $f(x_*)$ at new test points x_* , given by

$$f(x_*)|f(x) \sim N(k(x_*, x)k(x, x)^{-1}|f(x) - m(x)| + m(x_*), k(x_*, x_*) - k(x, x_*)k(x, x)^{-1}k(x, x_*)) \quad (5)$$

The methodology for analysing the bus passenger traffic using the above model is detailed in the next section.

4 Methodology

The methodology of work is as shown in Fig. 1. The different steps are

- Data mining and modelling
- Modelling of prior distribution
- Gaussian process modelling
- Comparison with existing methods
- Interpretation and analysis of results

4.1 Data Mining and Modelling

The data related to the Pune city bus traffic is provided by the PMPML authorities in various formats. The data with their specifications from various stakeholders in the organization is given in Table 1.

The Geospatial data gives the latitude and longitude values of different stops in the Pune city for the selected routes. The tripsheet report gives the information regarding the number of trips assigned in each route, direction of travel (up or down trip), the allotted bus number and crew members in each trip, trip timings and the assigned depot name. The passenger profile reports include the details of the number of passengers (including children) boarded from different stops, their start and end stages, the ticket rates (concession allotted if needed), ticket timing, total count of passengers and the revenue collected in each trip.

4.2 Modelling of Prior Distribution

The steps in modelling the data as prior distribution is shown in Fig. 2.

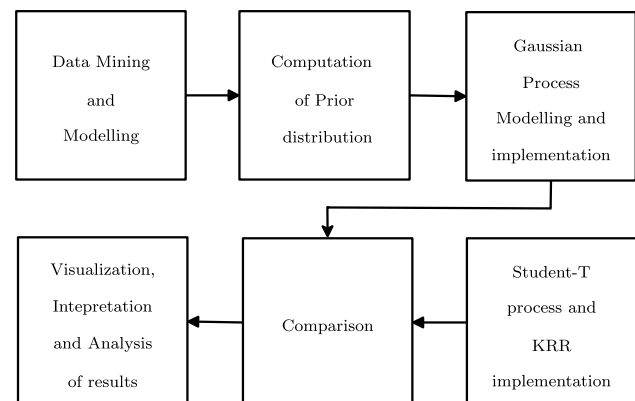


Figure 1 Methodology of work.

Table 1 Data collected from PMPML.

No.	Data	Specifications
1	Geospatial Data	Latitude and Longitude Data for these 47 routes
2	Tripsheet Report	Detailed view of trips and schedules in all routes
3	Passenger Profile Report	Complete ticketing data of all routes

The passenger arrival is modelled as a Poisson process as given in Eq. 1. The λ_i values in the Eq. 1 are calculated, based on the minimization of mean square error (MMSE) criterion between the actual statistics and the desired one. If the actual passenger data $P(x)$ and the desired is $P_d(x)$, the mean square error between $P(x)$ and $P_d(x)$ is.

$$\begin{aligned}\zeta &= E[|P_d(x) - P(x)|]^2 \\ &= \text{tr}\{E\{(P_d - P)(P_d - P)^T\}\}\end{aligned}\quad (6)$$

The values of the parameters that minimizes the error is obtained by

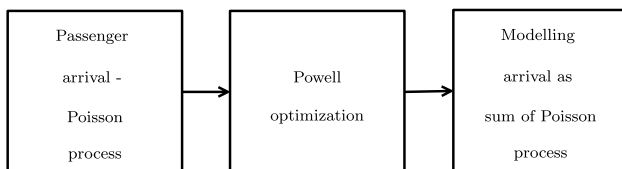
$$\frac{\partial \zeta}{\partial \lambda_1} = 0, \frac{\partial \zeta}{\partial \lambda_2} = 0, \frac{\partial \zeta}{\partial \lambda_3} = 0, \dots, \frac{\partial \zeta}{\partial \lambda_n} = 0, \quad (7)$$

The optimization using Powell algorithm [38], yielded 4 significant λ values. After calculating these values, the arrival process is modelled as a sum of Poisson processes. Then the passenger data is modelled to compute the prior distribution.

4.3 Gaussian Process Modelling

The advantage of Gaussian process regression technique is that it fits best for those data sets with no definite pattern. The traffic data being random in nature, Gaussian process prediction method is the most suitable one. A Gaussian Process is completely specified by a mean and covariance function. The Gaussian Processes is used to perform regression via supervised learning, called Gaussian Process Regression. This regression process is modelled as kernelized Bayesian linear regression, where the kernel parameterization is determined by the choice of covariance or kernel function, as well as the passenger data used to make predictions. Given a training dataset consisting of N observations:

$$D_{train} = (X, y) = \{x_i, y_i\}_{i=1}^N, x_i, y_i \in \mathbf{R} \quad (8)$$

**Figure 2** Steps in modelling the prior.

As well as a test dataset consisting of N' points:

$$D_{test} = (X_*) = \{x_{*,i}\}_{i=1}^{N'}, x_{*,i} \in \mathbf{R} \quad (9)$$

GPR predicts a posterior Gaussian distribution for targets over test points X_* by computing the parameters of this Gaussian distribution given observed training data. This GPR model is validated and compared with the existing technologies like Student-t process and KRR and the final results are analyzed. The experimental steps for implementing the above methodology are detailed below.

5 Experiment

The experimental steps in passenger prediction along with the tools in performing it are presented in Fig. 3. These steps are implemented using different python modules. The data is read using python *pandas* module, an open source easy to use tool that is used for handling large data sets. The data is converted as pandas dataframes, which are easily

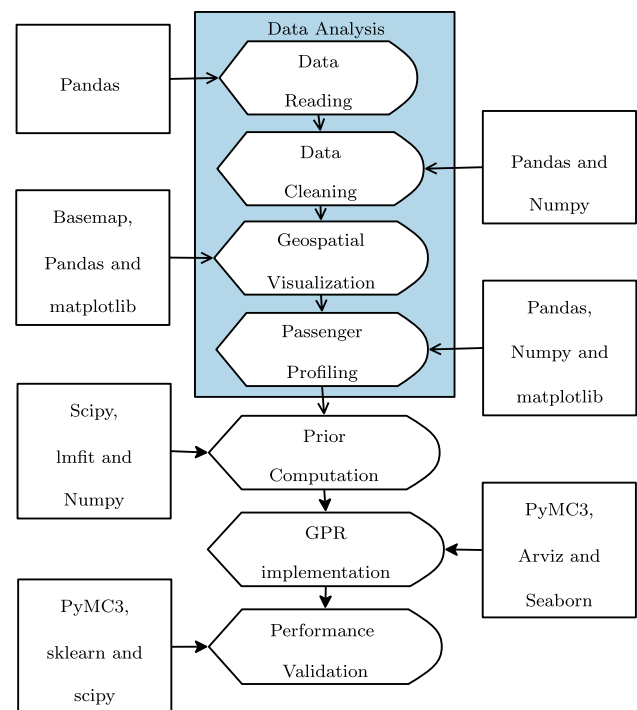
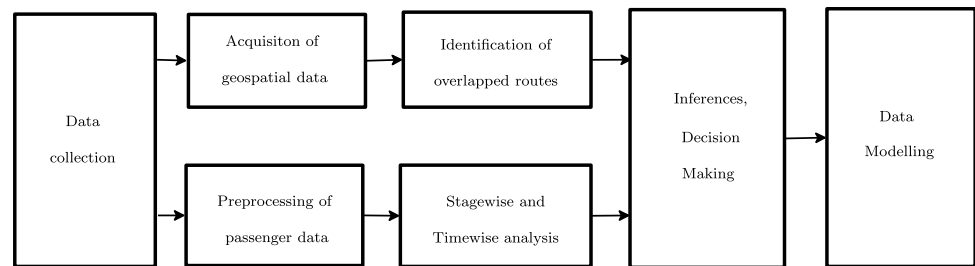
**Figure 3** Experimental steps with tools.

Figure 4 Steps in data analysis.

manipulated and cleaned using *numpy* module in python. The geospatial data is visualized using *Basemap*, *pandas* and *matplotlib* modules. For passenger profiling, tripsheet and passenger reports are read using *pandas*, analyzed using *numpy* and plotted using *matplotlib* module. The prior distribution is computed using *numpy*, *scipy* and *lmfit* modules. After the computation of the prior, GPR is designed and implemented in *PyMC3* environment with the support of *arviz* and *seaborn* modules. *PyMC3*, an open source probabilistic programming framework with *Theano* as backend, allows flexible implementation of Bayesian statistical models in python. The performance of the model is validated using the modules in *PyMC3*, *scipy* and *sklearn* modules.

5.1 Data Analysis

A total of 47 bus routes in Pune city are taken for the study. The complete data related to these set of routes is taken into consideration. The detailed steps in data analysis are shown in Fig. 4. For proper modelling, analysis is categorized into Geospatial analysis and Passenger analysis.

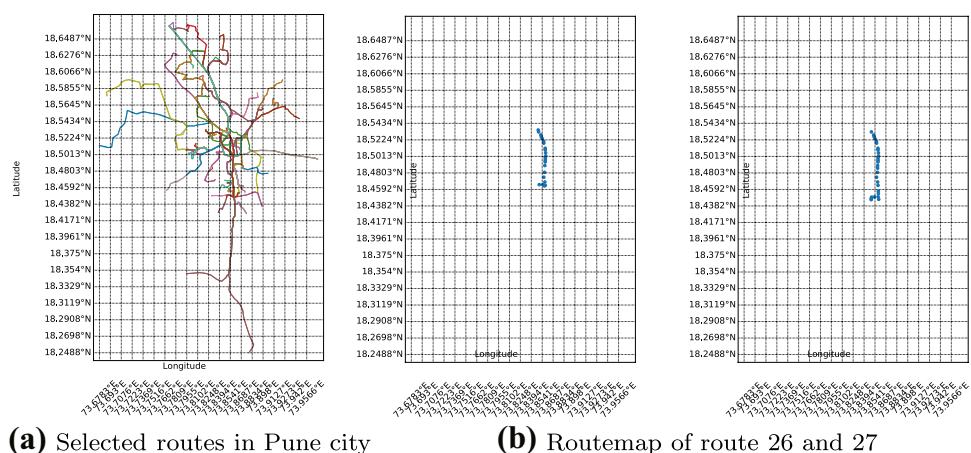
5.1.1 Geospatial Analysis

The geospatial analysis aims in concentrating the land use pattern of the selected routes. The geospatial data of 47 routes include the detailed report of different stops in all routes with their latitude and longitude values. It is plotted to study the geographical extent of the existing routes

(Fig. 5a). The latitude and longitude values of each route are read using *pandas* and converted into individual dataframes. These dataframes are read by the *Basemap* package to visualize the extent and coverage of individual routes. These routes are plotted in a canvas with the longitude and latitude values of Pune city along horizontal and vertical directions. The visualization is achieved using *Basemap* and *matplotlib* module. The plot reveals that these routes cover almost the entire Pune city. To figure out the various statistics and to study the feasibility of different routes, the routes with more than 80% overlap are identified (Fig. 5b). The identified routes are further studied on the basis of the number of schedules, trip timings and passenger traffic which is detailed in the next section.

5.1.2 Passenger Analysis

The tripsheet report and passenger profile reports are used for passenger modelling. The steps involved in cleaning and analysis are illustrated in Fig. 6. The data from the ticketing machines needed cleaning since it contained non ASCII characters, missing entries and misplaced values. Approximately more than 500 files are corrected. Python modules are developed for correcting them iteratively. The files of each route are read as *pandas* dataframes, divided into uptrips and downtrip dataframes using the data obtained from the Tripsheet report. In cases where trips are wrongly numbered as up and down, stages are extrapolated from the passenger profile report to separate the dataframes correctly

Figure 5 Geospatial visualization of routes.

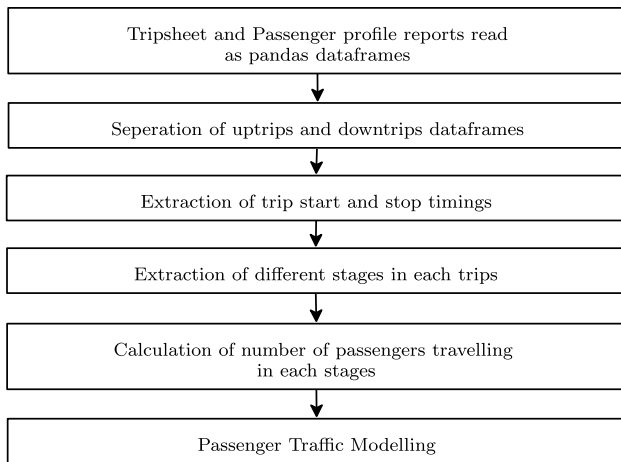


Figure 6 Steps in passenger traffic analysis.

into uptrips and downtrips. After separating them, the number of passengers travelling in different stages are calculated for detailed analysis.

The passenger data analysis is broadly divided into time-wise analysis and stagewise analysis of passenger data. The timewise analysis is performed to interpret the travel behaviour of passengers. The travel pattern of passengers vary depending on the day as well as the time of travel. The detailed analysis gives inferences that it shows peak during office hours (Fig. 7a). The day wise analysis (Fig. 7b) is performed to understand passenger flow behaviour and came to a conclusion that it shows peak on working days.

In order to draw more inferences, the stagewise analysis of passengers is also studied. The Fig. 8 shows the passenger history in route 24, one among the longest and busiest traffic route in Pune city. There are twelve fare stages in this bus route and the graph shows the average number of passengers retained in the bus as the trip proceeds through the different fare stages. It reveals that schedules are running with more than fifteen

passengers in almost all the stages and the bus utilization is good. All these analyses leads to the requirement of modelling the passenger arrival.

5.2 Prior Computation

As the arrival of passengers follows a Poisson process, the prior distribution selected is a Gamma Distribution [39]. The Gamma distribution, expressed as

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad (10)$$

where $\Gamma(\alpha)$ is given by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad (11)$$

The optimization using Powell algorithm calculated four significant λ values. The gamma function parameters α and β are calculated using the Poisson rate λ and variance of the data σ_x^2 as

$$\alpha = \frac{\lambda^2}{\sigma_x^2} \quad (12)$$

$$\beta = \frac{\sigma_x^2}{\lambda} \quad (13)$$

Thus these statistics arrives in choosing Gamma distribution as the prior model in GPR.

5.3 Implementation of GPR

The GPR model is implemented by using *PyMC3* environment as it is very flexible and Gamma distribution is the default prior in this environment for GPs. In this environment, GPs are designed and implemented by choosing the

Figure 7 Passenger analysis.

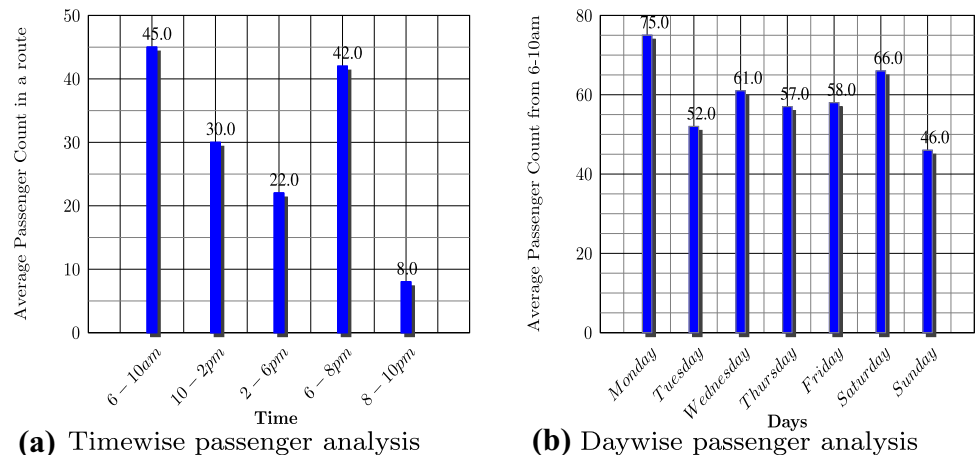
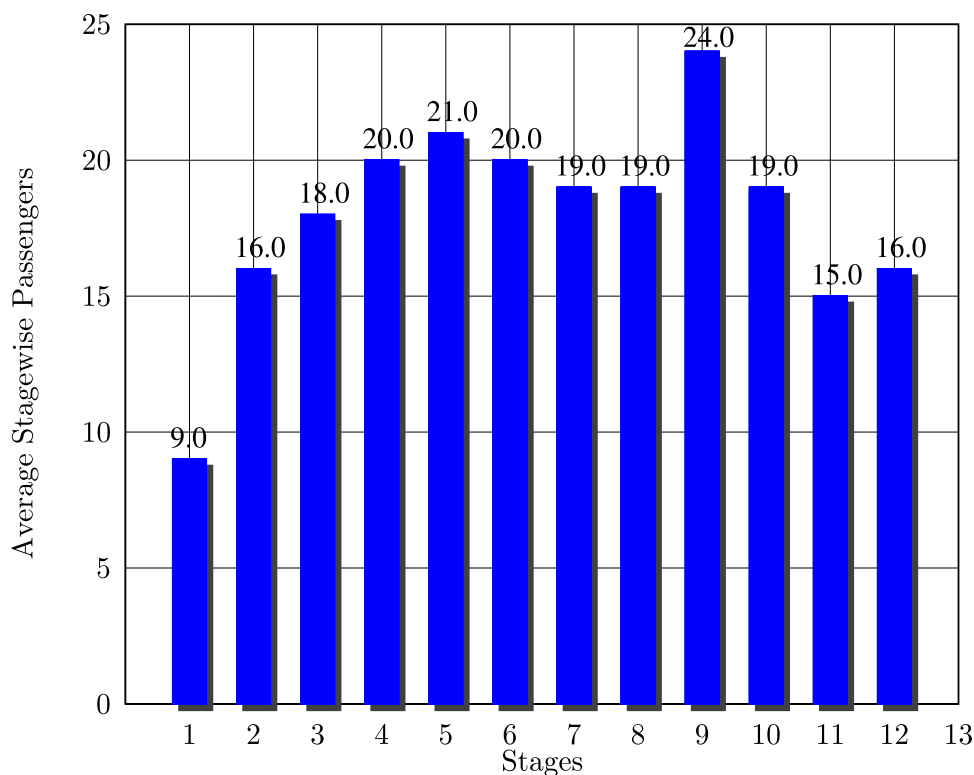


Figure 8 Stagewise passenger data.

covariance function and then calculating the marginal likelihood for prediction. The covariance matrices are often described using functions known as kernels. The one used here is the Exponential quadratic kernel as it results in a smooth prior on functions sampled from Gaussian process. For an input function $f(x)$, it is represented as

$$K(x, x') = \eta^2 \exp\left(-\frac{\|x - x'\|^2}{2l^2}\right) \quad (14)$$

Here $\|x - x'\|^2$ is the squared Euclidean distance given as

$$\|x - x'\|^2 = (x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \dots + (x_n - x'_n)^2 \quad (15)$$

η and l represents the hyperparameters to interpret and analyze data. The term l is called the characteristic length scale (or bandwidth or variance) and controls the width of the kernel along a particular axis and η represents the length along other axes or the amplitude. These hyperparameters represent the relevance of each value in the input data. If the l value is large, the covariance becomes almost independent of that particular input data. Hence this data is invalidated and removed, thus implements Automatic Relevance Determination (ARD) process [40]. In our prediction problem, l is specified using Gamma distribution and η using Half normal distribution. The GP is instantiated using this covariance function and marginal likelihood is computed to predict the passenger data. The mean of this distribution is taken as the prediction. To evaluate the model performance, traceplot,

posterior plot and autocorrelation of the covariance function parameters are examined.

The route 24 that connects Katraj to Maharashtra Housing Yerwada, covering a distance of 16 km is selected due to its heavy traffic and it represents the reliable samples of Pune population since it connects the major parts of Pune city. The passenger traffic data for a period of three months (October, November and December 2019) is taken.

5.4 Performance Validation

The performance of GP model is validated using

- RMSE
- Kernel density estimation
- Energy level distribution

The RMSE for the predicted data $f(x_*)$ is given by

Table 2 Parameters calculated using Powell method.

No.	Parameters	Value
1	λ_1	69.09
2	λ_2	54.27
3	λ_3	85.71
4	λ_4	67.08

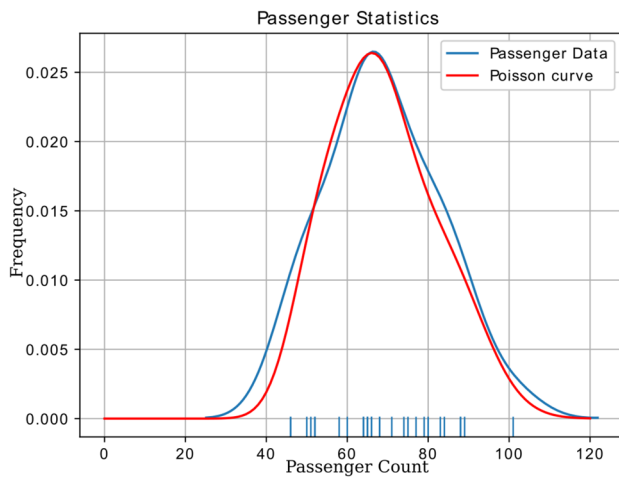


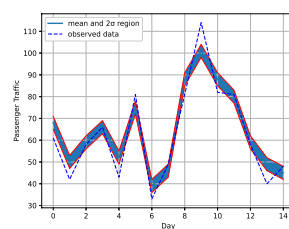
Figure 9 Passenger statistics.

$$RMSE = \sqrt{\frac{1}{n} \sum_n [f(x_*) - f(x)]^2} \quad (16)$$

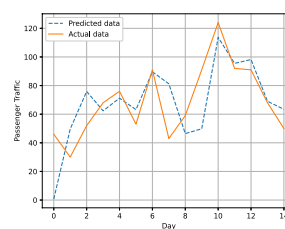
It indicates the numerical reliability of the model in predicting the traffic data. The kernel density estimation plot is used to compare the pdf of predicted one with the actual one. It is used in validating the predicted passenger data models for choosing proper bus scheduling algorithms. In the energy distribution graph, the difference in energy between the start and the end of the trajectory indicates the efficiency of the sampler. If the distribution shows longer tails, the efficiency of the sampler will deteriorate quickly. The energy levels of the samples are used to identify posteriors with problematically long tails.

The Gaussian Regression model is compared with the existing elliptical processes [41] such as Student-t processes and Kernel Ridge Regression process, predicting the random traffic data in terms of the above parameters. The Student-t has been used in variety of applications [42–45]. It is derived from the Gaussian process by placing an inverse Wishart process prior on the kernel. The KRR process is used in predicting wind speed [46], molecular orbital energy predictions [47] and in many other applications.

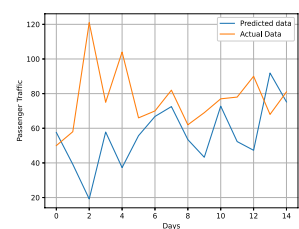
Figure 10 Comparison of prediction.



(a) GPR prediction



(b) Student-t process prediction



(c) KRR process prediction

6 Results

The significant λ values calculated, using Powell optimization algorithm are presented in Table 2.

Using these parameters, the passenger data is modelled (Fig. 9) and obtained a Poisson response. The Gamma parameters, computed using these Poisson parameters yielded 10.67 as the α value and 6.93 as the β value for designing the GPR model.

The GPR model is used to predict the passenger travel pattern for fifteen days. The gaussian model prediction is shown in Fig. 10a and that by Student-t model in Fig. 10b and by KRR model in Fig. 10c. Unlike the case of Student-t distribution model, the predicted value is almost following the actual one in the case of GPR.

The trace plot of the gaussian process model is given by (Fig. 11). The posterior density of noise decreases monotonically. The posterior distribution spread is indicated using Highest Posterior Density (HPD) interval or the bayesian credible interval. It is the shortest interval indicating a given portion of the probability density. The posterior plot, that indicates the distribution of hyperparameter, η and length function, l is shown in Fig. 12. It indicates mean value 45 and HPD of 94% for the hyperparameter η .

The precision of the model is benchmarked by a sharply decaying autocorrelation function. An ideal sample from the posterior distribution has autocorrelation value similar to an impulse function. Figure 13 shows that the samples from posterior distribution shows rapidly decaying autocorrelation values, validating the model.

The performance of the model is analyzed, validated and compared in terms of its RMSE value, kde and sampler efficiency.

6.1 Computation of Root Mean Square Error

The mean square error of the predicted data for the GPR model is found to be 5.33, while that for the existing model like Student-t, the RMSE value is 13.36 and for KRR it is 24.61. The performance of the model is much better as the error is decreased by a factor of 8 compared to Student-t and by 19.28 while comparing with KRR model, validating the suitability of the GPR model for predicting passenger traffic.

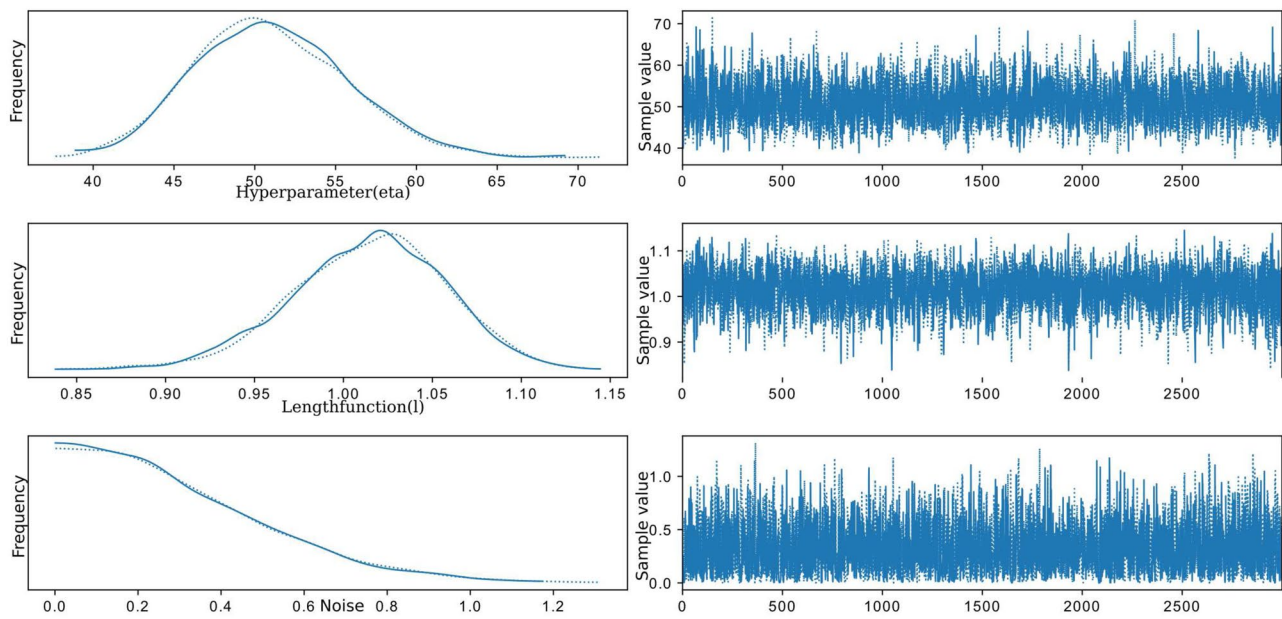


Figure 11 Traceplot of gaussian model.

Figure 12 Posterior plots.

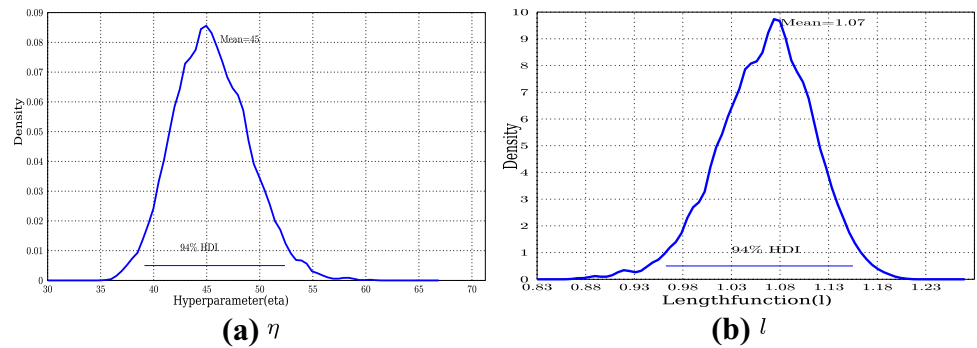


Figure 13 Autocorrelation plot.

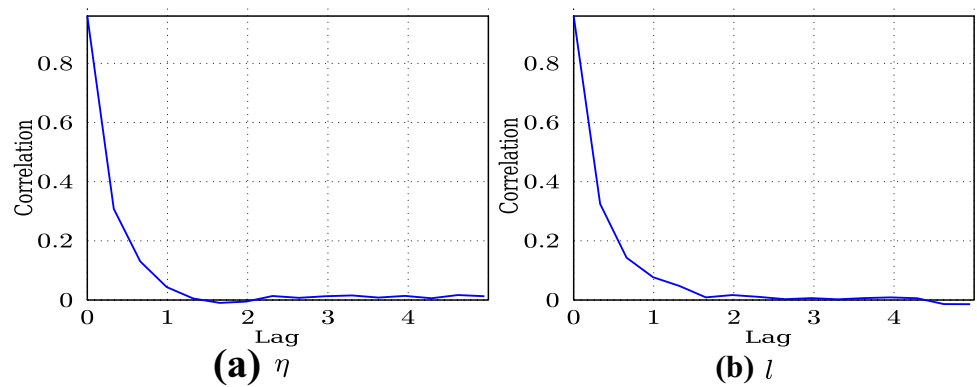


Figure 14 Kernel density estimation.

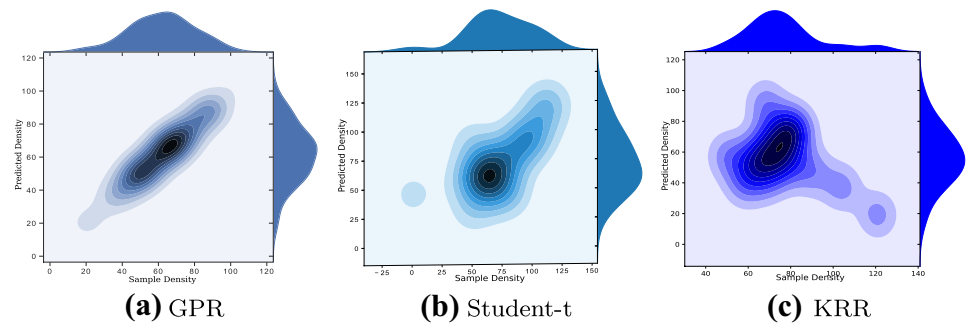
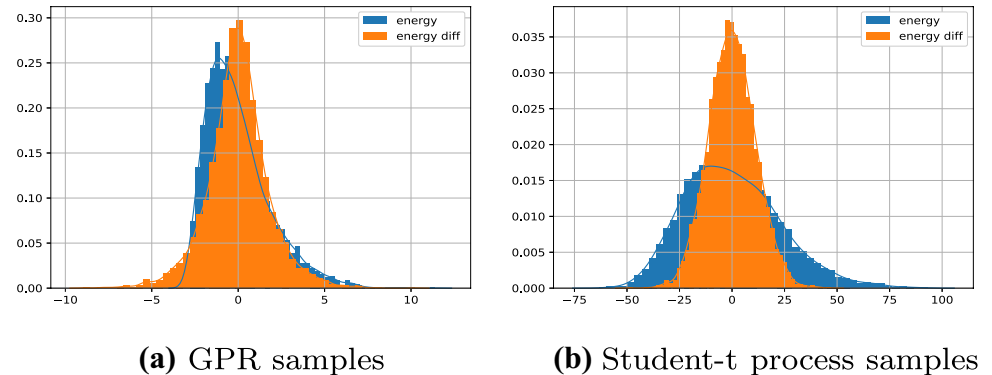


Figure 15 Energy of the samples.



6.2 Kernel Density Estimation

The statistical properties are analyzed by plotting the joint probability densities (samples and the predicted values) for both models (Fig. 14). In the case of student-t and KRR, it is visible from the figure that the probability density plots of actual and predicted ones are different (Fig. 14b, c) while that with the GPR model (Fig. 14a), both the samples and predicted one follows similar distribution.

6.3 Sampler Efficiency

In order to assess the efficiency of the sampler, overall energy level distribution of the samples, with the change of energy between successive samples, are plotted. The sampler efficiency in the case of Student-t is low, as the distribution has longer tails. From Fig. 15, it is clear that GPR sampler (Fig. 15a) is highly efficient than Student-t sampler (Fig. 15b).

7 Inferences and Conclusion

Gaussian process regression model is designed to predict the bus passenger traffic which shows random nature and high parametric dependence. The GPR model is observed

to perform well in terms of prediction accuracy even with limited data. The arrival of passengers at different bus stops follows Poisson distribution whose statistics links with the bus passenger traffic. The passenger data being modelled as a Gamma distribution, along with the likelihood, GPR computes the posterior distribution. Finally the conditional distribution is computed to evaluate the predicted values. The training and prediction process are much simpler and faster than the traditional machine learning techniques. The GPR method took only 170 seconds to complete the training and prediction process compared to 626.46 seconds for the Student-t process. The present work is done with the objective of developing scheduling algorithms. It is also instrumental in designing Intelligent traffic management systems. Such intelligent systems that are adapted to the passenger traffic will improve the financial performance of transportation systems and passenger comfort.

Declarations

Ethical Approval The authors did not receive support from any organization for the submitted work.

Conflict of Interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

- Pavlyuk, D. (2017). Short-term traffic forecasting using multivariate autoregressive models. *Procedia Engineering*, 178, 57–66.
- Rodriguez-Deniz, H., Jenelius, E., & Villani, M. (2017). Urban network travel time prediction via online multi-output gaussian process regression. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1–6). IEEE.
- Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., & Damas, L. (2013). Predicting taxi-passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems*, 14, 1393–1402.
- Xue, R., Sun, D. J., & Chen, S. (2015). Short-term bus passenger demand prediction based on time series model and interactive multiple model approach. *Discrete Dynamics in Nature and Society*, 2015.
- Feng, G. (2015). Network traffic prediction based on neural network. In *2015 International Conference on Intelligent Transportation, Big Data and Smart City* (pp. 527–530). IEEE.
- Terry, N., & Choe, Y. (2021). Splitting gaussian processes for computationally-efficient regression. *PLOS ONE*, 16, 1–17.
- Maritz, J., Lubbe, F., & Lagrange, L. (2018). A practical guide to gaussian process regression for energy measurement and verification within the bayesian framework. *Energies*, 11.
- Quinero Candela, J., & Rasmussen, C. (2005). A unifying view of sparse approximate gaussian process regression. *Journal of Machine Learning Research*, 6, 1935–1959.
- Xu, Y., Xu, W., Yin, F., Lin, J., & Cui, S. (2017). High-accuracy wireless traffic prediction: A gp-based machine learning approach. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference* (pp. 1–6).
- Sun, Y., Wang, S., Lu, Y., & Xu, J. (2022). Gaussian process dynamic modeling and backstepping sliding mode control for magnetic levitation system of maglev train. *Journal of Theoretical and Applied Mechanics*, (pp. 49–62).
- Subrahmanyam, K. V., Ramsenthil, C., Girach Imran, A., Chakravorty, A., Sreedhar, R., Ezhilrajan, E., Bala Subrahmanyam, D., Ramachandran, R., Kumar, K. K., Rajasekhar, M. et al. (2021). Prediction of heavy rainfall days over a peninsular indian station using the machine learning algorithms. *Journal of Earth System Science*, 130, 1–9.
- Zazoum, B. (2021). Solar photovoltaic power prediction using different machine learning methods. *Energy Reports*, .
- Pooja, W., Snehal, N., Sonam, K., Wagh, S., & Singh, N. (2021). Effect of increased number of covid-19 tests using supervised machine learning models. In *2021 Australian & New Zealand Control Conference (ANZCC)* (pp. 131–136). IEEE.
- Topaloglu, B., Kaya, G. T., Sutcu, F., & Deger, Z. T. (2021). Machine learning-based assessment of energy behavior of rc shear walls. *arXiv preprint arXiv:2111.08295*, .
- Goudarzi, S., Kama, N., Anisi, M. H., Zeadally, S., & Mumtaz, S. (2019). Data collection using unmanned aerial vehicles for internet of things platforms. *Computers & Electrical Engineering*, 75, 1–15.
- Mumtaz, S., Lundqvist, H., Huq, K. M. S., Rodriguez, J., & Radwan, A. (2014). Smart direct-lte communication: An energy saving perspective. *Ad Hoc Networks*, 13, 296–311.
- Duan, W., Gu, J., Wen, M., Zhang, G., Ji, Y., & Mumtaz, S. (2020). Emerging technologies for 5g-iov networks: Applications, trends and opportunities. *IEEE Network*, 34, 283–289.
- Hu, J., Li, X., & Ou, Y. (2014). Online gaussian process regression for time-varying manufacturing systems. In *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)* (pp. 1118–1123). IEEE.
- Bayati, A., Asghari, V., Nguyen, K., & Cheriet, M. (2016). Gaussian process regression based traffic modeling and prediction in high-speed networks. In *2016 IEEE Global Communications Conference (GLOBECOM)* (pp. 1–7).
- Hoque, K. E., & Aljamaan, H. (2021). Impact of hyperparameter tuning on machine learning models in stock price forecasting. *IEEE Access*, 9, 163815–163830.
- Andugula, P., Durbha, S. S., Lokhande, A., & Suradhaniwar, S. (2017). Gaussian process based spatial modeling of soil moisture for dense soil moisture sensing network. In *2017 6th International Conference on Agro-Geoinformatics* (pp. 1–5).
- Ghasemi, P., Karbasi, M., Zamani Nouri, A., Sarai Tabrizi, M., & Azamathulla, H. M. (2021). Application of gaussian process regression to forecast multi-step ahead spei drought index. *Alexandria Engineering Journal*, 60, 5375–5392.
- Cai, H., Jia, X., Feng, J., Li, W., Hsu, Y.-M., & Lee, J. (2019). Gaussian process regression for numerical wind speed prediction enhancement. *Renewable Energy*, 146.
- Raissi, M., Babaei, H., & Karniadakis, G. E. (2019). Parametric Gaussian process regression for big data. *Computational Mechanics*, 64, 409–416.
- Graas, R., Sun, J., & Hoekstra, J. (2021). Quantifying accuracy and uncertainty in data-driven flight trajectory predictions with gaussian process regression. In *11th SESAR Innovation Days*.
- Xie, G., & Chen, X. (2021). Efficient and robust online trajectory prediction for non-cooperative unmanned aerial vehicles. *Journal of Aerospace Information Systems*, (pp. 1–11).
- Rong, H., Teixeira, A., & Guedes Soares, C. (2022). Maritime traffic probabilistic prediction based on ship motion pattern extraction. *Reliability Engineering and System Safety*, 217, 108061.
- Akçay, M. T., Akgundogdu, A., & Tiryaki, H. (2021). Estimation of the average speed for a railway signaling system by using gaussian process regression methods with bayesian optimization. *Demiryolu Mühendisliği*, 14, 274–286.
- Xie, H., Hu, D., & Song, K. (2021). An iterative optimization algorithm for vehicle speed prediction considering driving style and historical data effects. In *2021 40th Chinese Control Conference (CCC)* (pp. 6094–6100).
- Soldevila, I. E., Knoop, V. L., & Hoogendoorn, S. (2021). Car-following described by blending data-driven and analytical models: a gaussian process regression approach. *Transportation research record*, 2675, 1202–1213.
- Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer school on machine learning* (pp. 63–71). Springer.
- Lin, C., Li, T., Chen, S., Liu, X., Lin, C., & Liang, S. (2019). Gaussian process regression-based forecasting model of dam deformation. *Neural Computing and Applications*, 31.
- Church, K., & Gale, W. (1995). Poisson mixtures. *Natural Language Engineering*, 1.
- Mihaylova, L., Boel, R., & HEGYI, A. (2006). An unscented kalman filter for freeway traffic estimation. In H. Van Zuylen, & F. Middelham (Eds.), *Proceedings of 11th IFAC Symposium on Control in Transportation Systems* (pp. 31–36).
- Mihaylova, L., & Boel, R. (2004). A particle filter for freeway traffic estimation. In *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)* (pp. 2106–2111 Vol.2), volume 2.
- Thrun, S., Burgard, W., Fox, D., & Arkin, R. (2005). *Probabilistic Robotics*. Intelligent Robotics and Autonomous Agents series. MIT Press.
- Osvaldo, M., & Safari (2018). *Bayesian Analysis with Python*. Intelligent Robotics and Autonomous Agents series. Packt Publishing Ltd.
- Fletcher, R., & Powell, M. J. D. (1963). A Rapidly Convergent Descent Method for Minimization. *The Computer Journal*, 6, 163–168.

39. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.
40. Wipf, D., & Nagarajan, S. (2007). A new view of automatic relevance determination. *Advances in neural information processing systems*, 20.
41. Fang, K.-T., Kotz, S., & Ng, K. W. (2018). *Symmetric multivariate and related distributions*. Chapman and Hall/CRC.
42. Archambeau, C., & Bach, F. (2011). Multiple gaussian process models. *arXiv preprint arXiv:1110.5238*.
43. Xu, Z., Yan, F., & Qi, Y. (2011). Sparse matrix-variate t process blockmodels. *Proceedings of the AAAI Conference on Artificial Intelligence*, 25.
44. Yu, S., Tresp, V., & Yu, K. (2007). Robust multi-task learning with t-processes. In *Proceedings of the 24th International Conference on Machine Learning* (p. 1103–1110). Association for Computing Machinery.
45. Zhang, Y., & Yeung, D. (2010). Multi-task learning using generalized t process. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (pp. 964–971). PMLR volume 9 of *Proceedings of Machine Learning Research*.
46. Douak, F., Melgani, F., & Benoudjit, N. (2013). Kernel ridge regression with active learning for wind speed prediction. *Applied energy*, 103, 328–340.
47. Stuke, A., Todorović, M., Rupp, M., Kunkel, C., Ghosh, K., Himanen, L., & Rinke, P. (2019). Chemical diversity in molecular orbital energy predictions with kernel ridge regression. *The Journal of chemical physics*, 150, 204121.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.