Gaussian processes for real-time urban traffic estimation using a fleet of drones

Batuhan Avci*[1], Marko Susnjar*[1], Marko Maljkovic[1], and Nikolas Geroliminis[1]

[1]Urban Transport Systems Laboratory (LUTS) École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.
{batuhan.avci,marko.susnjar,marko.maljkovic,nikolas.geroliminis}@epfl.ch

Short summary

Urban traffic monitoring requires accurate network-wide estimates even though only a limited portion of the road network can be observed in real time. At the same time, the growing deployment of Unmanned Aerial Vehicles (UAVs) in everyday applications is creating new opportunities for high-resolution, real-time traffic monitoring. Compared with traditional fixed loop detectors, fleets of UAVs acting as mobile sensing networks offer substantially greater flexibility, improved accuracy, and broader spatial coverage. In this paper, we propose a Gaussian-process-based framework for real-time estimation and short-term prediction of urban traffic variables from partial drone observations. The area of interest is represented as a weighted graph, and the proposed covariance structure combines a graph Matérn kernel, which captures the topology of the road network, with a temporal kernel that models the evolution of traffic over time. In addition to providing network-wide estimates and predictions, the proposed framework also yields a principled quantification of predictive uncertainty at unobserved locations. The model is trained and evaluated using micro-simulations of Barcelona's urban road network under multiple demand scenarios, demonstrating the effectiveness of Gaussian processes for traffic parameter estimation and prediction.

Keywords: Gaussian processes, Traffic variable estimation, Urban traffic monitoring, Unmanned aerial vehicles, Spatio-temporal modeling, Mobile sensing

1 Introduction

The ability to observe traffic conditions across an urban road network in real time is a fundamental requirement for modern traffic management systems. Such conditions are typically described by traffic variables such as speed, density, and flow, and estimating them over a road network becomes crucial for various applications like adaptive signal control (Tsitsokas et al. (2023), Zoabi & Haddad (2026)), perimeter control (Kouvelas et al. (2017), Moshahedi & Kattan (2023)), and urban emission estimation (Espadaler-Clapés et al. (2023)). However, directly observing traffic variables across an entire urban road network is often infeasible, as it is not possible to monitor an entire city simultaneously. As a result, traffic conditions over the full network must be inferred from partial observations collected at the current and previous time steps. These networks include thousands of road segments whose states are tightly coupled through intersections, and traffic conditions on each segment evolve dynamically, driven by signal cycles, demand profiles, and incidents.

In traffic monitoring, conventional sensing approaches typically rely on loop detectors and cameras installed at fixed positions along the roads. Although widely used in practice, these sensors cannot be repositioned in response to changing traffic conditions and as such provide low spatial coverage. Moreover, they are susceptible to measurement noise and have high deployment costs, while providing only limited information, even at the link level (Xiong et al. (2025)). Recently, drones have emerged as a promising approach for traffic monitoring. As illustrated in Figure 1, a fleet of drones can act as a network of mobile sensors to monitor the area of interest, and high-resolution information can be gathered for all the road segments within their sensing range (Elloumi et al. (2018), Barmpounakis & Geroliminis (2020), Maljkovic & Geroliminis (2025)). Nevertheless, even

*These authors contributed equally to this work.

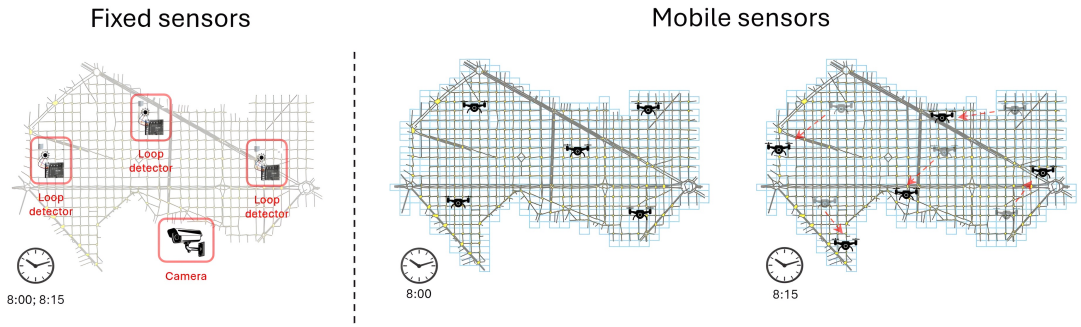


Figure 1: Illustration of the traffic network in the city of Barcelona. Left: deployment of fixed traffic sensors, including loop detectors and cameras. Right: spatial discretization of the same network into rectangular cells of 160 m × 90 m, where each cell represents the field of view of a single drone.

a fleet of drones can only cover a fraction of large networks at any given time, leading to partial and time-varying spatial coverage as the drones move along their trajectories.

This partial observability poses an essential challenge in traffic monitoring settings: how can traffic variables be estimated across the entire road network when only noisy measurements from a small and dynamically changing subset of road segments are available? Addressing this question essentially requires a method that satisfies three key criteria: (i) it should infer traffic variables based on only a limited set of observations, (ii) it should quantify the uncertainty of the estimates at unobserved road segments, and (iii) it should incorporate the physical structure of the road network in such a way that the spatial correlations of the estimated state conform to the underlying topology of the road network.

Gaussian Process (GP) regression provides a plausible probabilistic framework for this task, as it yields both a predictive mean and a measure of predictive uncertainty, thereby inherently satisfying requirements (i) and (ii) of the problem formulation (Rasmussen & Williams (2006)). Within such a framework, the covariance function (which is completely determined by the choice of its kernel) plays a central role in determining how the GP generalizes from observed to unobserved locations. However, standard kernels that are defined based on Euclidean distance are not well suited to road networks, where two segments may be close in Euclidean space but disconnected in the network topology (Borovitskiy et al. (2021)).

In this work, we propose a spatio-temporal GP framework building on the Graph Matérn kernel (Borovitskiy et al. (2021)) to infer traffic variables from partial observations collected by a fleet of drones. The road network is represented as an undirected, weighted graph, and the traffic variables over this graph are jointly modeled as a sample from a GP. At each discrete time step, observations are available only at a subset of nodes determined by the drones' instantaneous positions. Based on these partial measurements, the GP produces posterior estimates of the traffic variables over the entire network, together with associated uncertainty quantification even for unobserved segments. We train and evaluate the proposed model in a realistic simulation environment based on the city of Barcelona, where microscopic traffic simulations generated in Aimsun provide ground-truth traffic variables over the full network.

The paper is outlined as follows: Section 2 introduces the GP framework in detail. Section 3 then introduces the GP training procedure and the data used. Finally, Section 4 presents and discusses the numerical results before concluding remarks are presented in Section 5.

2 Methodology

Preliminaries

GPs extend the concept of a multivariate Gaussian distribution to stochastic processes, providing a probabilistic framework for learning unknown functions while quantifying predictive uncertainty (Rasmussen & Williams (2006)). A GP can be viewed as a distribution over functions $f : \mathcal{X} \to \mathbb{R}$, fully characterized by its mean function $m : \mathcal{X} \to \mathbb{R}$ and covariance (i.e., kernel) function $k : \mathcal{X}^2 \to$

$\mathbb{R}$. In this framework, the latent function f is typically modeled as a sample from a GP denoted as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \tag{1}$$

where $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ represent d-dimensional feature vectors encapsulating specific attributes within the input domain. For any finite collection of inputs $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, the corresponding function evaluations $\mathbf{f} = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N)]^\top$ follow a joint Gaussian distribution $\mathbf{f} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$, where the covariance matrix elements are given by $\mathbf{K}_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$. This formulation enables the model to provide not only point estimates but also a principled measure of predictive uncertainty through the variance of the distribution.

In many applications, the input space is typically assumed to be a continuous Euclidean domain $\mathbb{R}^d$, and conventional kernels are defined using the Euclidean distance. However, when modeling urban traffic, the spatial domain is constrained to a complex road network. In this context, modeling the city as a continuous plane may not reflect the true structure of the network because two road segments can be close in Euclidean distance while still being topologically disconnected due to physical constraints or traffic regulations (e.g., prohibited turns). To guarantee a valid covariance structure, we employ a modeling approach that explicitly encodes the underlying graph topology. The details of this covariance formulation will be elaborated in the subsequent parts of this section. First, we formally define our urban road network model.

Road network model

We represent the urban road network as an undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V} = \{v_1, \ldots, v_n\}$ is a set of n nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. Each node $v_i \in \mathcal{V}$ represents the midpoint of a road segment. An edge $(v_i, v_j) \in \mathcal{E}$ exists if and only if the corresponding road segments are physically connected in the network. This graph representation reflects the fact that traffic states evolve over a structured topology, where interactions arise primarily between adjacent segments linked through intersections or direct continuation of the roadway. The strength of these spatial interactions is encoded by the symmetric weight matrix $\mathbf{W} \in \mathbb{R}_{\geq 0}^{n \times n}$, whose entries W_{ij} quantify the pairwise coupling strength between nodes v_i and v_j. Specifically, we define

$$W_{ij} = \begin{cases} \omega(v_i, v_j) > 0 & \text{if } (v_i, v_j) \in \mathcal{E}, \\ 0 & \text{otherwise,} \end{cases}$$

with $W_{ii} = 0$ for all i (i.e., no self-loops) and where $\omega : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_{>0}$ is a weighting function. In the context of urban traffic, this weighting is intended to capture the intuition that directly connected segments that are geographically close tend to exhibit stronger dependence, since congestion, and flow propagation are more likely to affect neighboring parts of the network than distant ones. A natural choice for ω could be, for example, the inverse of the distance between segment midpoints.

This graph construction serves as the spatial backbone of the GP-based traffic model introduced in the following subsection. By defining $\mathcal{G}$ over road-segment midpoints, the model captures how local traffic states interact through the underlying road topology. Each realization of traffic on $\mathcal{G}$ thus corresponds to a network-wide traffic state, whose spatial correlations are shaped by the connectivity and geometry of the underlying traffic infrastructure.

Problem definition

Let $\mathcal{X} = \mathcal{V} \times \mathcal{T}$ denote the spatio-temporal input domain, where $\mathcal{V}$ is the node set of the road network graph $\mathcal{G}$ and $\mathcal{T} \subseteq \mathbb{R}_{\geq 0}$ is the temporal horizon. Each input $\mathbf{x} = (v, t) \in \mathcal{X}$ indexes a spatial location $v \in \mathcal{V}$ at time $t \in \mathcal{T}$.

For a particular traffic variable (e.g. flow, density, speed, etc.), we model it on a network-wide level as a function $f : \mathcal{X} \to \mathbb{R}$, drawn from a spatio-temporal GP

$$f \sim \mathcal{GP}(m,\ k), \qquad m : \mathcal{X} \to \mathbb{R}, \quad k : \mathcal{X} \times \mathcal{X} \to \mathbb{R},$$

where $m(\mathbf{x}) = m((v, t))$ is the expected mean value of the traffic variable on road segment $l \in \mathcal{V}$ at time step $t \in \mathcal{T}$, and $k(\mathbf{x}, \mathbf{x}')$ is the kernel encoding spatio-temporal dependencies induced by the network topology and temporal dynamics.

We consider a mobile sensing framework comprising a fleet of P drones that collectively monitor the road network over a set of L discrete observation time steps $\{t_j\}_{j=1}^{L} \subset \mathcal{T}$. In practice, recent

advances in computer vision enable accurate real-time estimation of vehicle trajectories, which in turn support a wide range of downstream applications depending on the application of interest, such as traffic congestion and noise emission estimation (Espadaler-Clapés et al. (2023)), outdoor parking monitoring (Kim et al. (2024)), real-time traffic flow parameter estimation (Ke et al. (2019)), and driving behavior analysis (Krajewski et al. (2018)). We therefore assume that each drone can estimate the traffic variable of interest using onboard computation.

At each time step t_j, drone $p \in \{1, \dots, P\}$ observes a subset of road links $\mathcal{V}_j^{(p)} \subseteq \mathcal{V}$ determined by its instantaneous position. The aggregate set of observed nodes at time t_j is then

$$\mathcal{V}_j = \bigcup_{p=1}^{P} \mathcal{V}_j^{(p)} \quad \text{with} \quad n_j = |\mathcal{V}_j|.$$

Both $\mathcal{V}_j^{(p)}$, and therefore $\mathcal{V}_j$, may vary across time steps depending on the sensing policy. For instance, a hovering drone yields a static $\mathcal{V}_j^{(p)}$, whereas a cruising trajectory produces a time-varying coverage pattern. The observation of a road segment $v_i \in \mathcal{V}_j$ at time t_j is modeled as

$$y_{ij} = f(v_i, t_j) + \varepsilon_{ij}, \qquad \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_n^2)$$

where σ_n^2 denotes the noise variance, capturing both sensor measurement noise and the uncertainty arising from the reconstruction and identification of individual vehicle trajectories. Across the monitoring horizon comprising all L time steps, the total number of collected observations is $N = \sum_{j=1}^{L} n_j$. We enumerate these as $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathcal{X}$, where each $\mathbf{x}_s = (v_i, t_j)$ corresponds to a specific node $v_i \in \mathcal{V}_j$ observed at time t_j, and denote

$$\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset \mathcal{X}, \qquad \mathbf{y} = [y_1, \dots, y_N]^\top \in \mathbb{R}^N$$

where y_s is the noisy measurement associated with input $\mathbf{x}_s$.

Since f is modeled as a GP and the observation noise is assumed to be Gaussian, the posterior $f \mid \mathbf{X}, \mathbf{y}$ is also a GP. Hence, the posterior mean and covariance for an arbitrary query point $\mathbf{x}_* = (v_*, t_*) \in \mathcal{X}$ are given by

$$\begin{aligned} m(\mathbf{x}_* \mid \mathbf{X}, \mathbf{y}) &= m(\mathbf{x}_*) + \mathbf{k}_*^\top \left(\mathbf{K}_{\mathbf{XX}} + \sigma_n^2 \mathbf{I}_N\right)^{-1} (\mathbf{y} - \mathbf{m}) \\ \operatorname{cov}(\mathbf{x}_*, \mathbf{x}_*' \mid \mathbf{X}, \mathbf{y}) &= k(\mathbf{x}_*, \mathbf{x}_*') - \mathbf{k}_*^\top \left(\mathbf{K}_{\mathbf{XX}} + \sigma_n^2 \mathbf{I}_N\right)^{-1} \mathbf{k}_{*'} \end{aligned} \tag{2}$$

where $\mathbf{m} = [m(\mathbf{x}_s)]_{s=1}^N \in \mathbb{R}^N$ is the mean vector over the observed inputs, $[\mathbf{K}_{\mathbf{XX}}]_{ss'} = k(\mathbf{x}_s, \mathbf{x}_{s'})$ is the $N \times N$ covariance matrix over the observed inputs, $\mathbf{k}_* = [k(\mathbf{x}_*, \mathbf{x}_s)]_{s=1}^N \in \mathbb{R}^N$ is the cross-covariance vector between the query point and the observations, and $\mathbf{I}_N$ is the N-dimensional identity matrix. The predictive variance at $\mathbf{x}_*$ is obtained by setting $\mathbf{x}_*' = \mathbf{x}_*$ in (2).

Spatio-temporal covariance structure

The kernel k must jointly encode spatial correlations governed by the road network topology and temporal correlations capturing the dynamics of intra-day traffic fluctuations. We use a separable (product) kernel, factoring the joint covariance into independent spatial and temporal components

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \, k_{\mathcal{G}}(v, v') \, k_{\mathcal{T}}(t, t') \tag{3}$$

where $\mathbf{x} = (v, t)$, $\mathbf{x}' = (v', t')$, and parameter $\sigma_f^2 > 0$ is the signal variance (i.e., output scale). Note that the separability assumption implies that the spatial correlation structure is time-invariant, and, conversely, that the temporal correlation structure does not depend on the spatial location.

For the spatial component $k_{\mathcal{G}}$ in (3), we employ the Graph Matérn kernel (Borovitskiy et al. (2021)), a discrete analogue of continuous Matérn fields defined by the graph Laplacian $\mathbf{\Delta} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D} = \operatorname{diag}(d_1, \dots, d_n)$ with $d_i = \sum_j W_{ij}$ is the degree matrix. The Laplacian admits the eigendecomposition $\mathbf{\Delta} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, with orthogonal eigenvectors $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$ and eigenvalues $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Following the characterization of Matérn kernels on graphs, the spatial covariance matrix is defined as

$$\mathbf{K}_{\mathcal{G}} = \left(\frac{2\nu_g}{\kappa_g^2}\mathbf{I}_n + \mathbf{\Delta}\right)^{-\nu_g} = \mathbf{U} \operatorname{diag}\left(\left\{\left(\frac{2\nu_g}{\kappa_g^2} + \lambda_l\right)^{-\nu_g}\right\}_{l=1}^{n}\right) \mathbf{U}^\top$$

where $\nu_g > 0$ and $\kappa_g > 0$ are the smoothness and lengthscale parameters, respectively. The element-wise covariance between nodes v and v' is

$$k_{\mathcal{G}}(v, v') = \sum_{l=1}^{n} \left(\frac{2\nu_g}{\kappa_g^2} + \lambda_l \right)^{-\nu_g} u_l(v)\, u_l(v')$$

where $u_l(v)$ denotes the v-th component of the l-th eigenvector. By construction, the kernel assigns high correlation to nodes that are topologically close on $\mathcal{G}$, ensuring that information propagates strictly through valid network connections.

For the temporal component $k_{\mathcal{T}}$ in (3), we want to capture the dynamics of traffic evolution along the time axis. In urban traffic, abrupt changes arise from congestion waves, signal cycles, and incidents. To model this behavior, we use the standard Matérn kernel with smoothness parameter $\nu = 1/2$, which reduces to the exponential kernel

$$k_{\mathcal{T}}(t, t') = \exp\left(-\frac{|t - t'|}{\ell_{\mathcal{T}}} \right) \tag{4}$$

where $\ell_{\mathcal{T}} > 0$ is the temporal lengthscale governing the rate of correlation decay. Larger values of $\ell_{\mathcal{T}}$ encode stronger persistence, while smaller values allow the model to track fluctuations.

3 Data and Training

Data overview

The data used in this paper was generated by (Xiong et al. (2025)) using the microscopic traffic simulator Aimsun (Casas et al. (2010)), which provides trajectories for each vehicle, together with complete information on the road network topology and geometry of road segments. The simulated network corresponds to the urban area in the city of Barcelona with 1570 road segments and 721 intersections, modeled as an undirected graph with 1570 nodes and 2803 edges connecting the road segments that have an intersection in common (as described in Section 2).

A total of 101 time series were generated from simulation runs with varying traffic demand levels. Each simulation was initialized using historically calibrated origin-destination data for a specific time period of the day, which was then randomly perturbed and multiplicatively scaled to produce augmented, but realistic demand profiles. The original historic demand corresponds to the scale 1.0, and 20 additional series were generated for each demand scale in $\{1.20, 1.35, 1.50, 1.65, 1.80\}$, thereby representing a range of increasingly congested traffic scenarios in the urban area. Every time series captures a four-hour interval of a single day, consisting of a 15-minute warm up period, 1 hour and 45 minutes of main demand, followed by two hours without new vehicle generation to allow the network to unload. Only the two-hour interval containing realistic traffic conditions, beginning immediately after the warm-up phase and together with the first 15 minutes of the unload phase, is retained for analysis.

The simulator provides precise vehicle trajectories, characterized by the vehicle positions at each simulation time step and the exact times at which vehicles enter and exit the road segments. From these trajectories, segment-level traffic parameters such as Vehicle Kilometers Traveled (VKT), Vehicle Hours Traveled (VHT), speed, and related quantities can be computed, as described in (Xiong et al. (2025)). In addition, (Fonod et al. (2025)) shows that, in well-performed stable drone flights, it is possible to accurately extract the trajectories of the majority of vehicles within the monitored area using computer vision techniques. Accordingly, we assume that whenever a road segment is observed by a drone, the ground-truth value of the traffic variable of interest on that segment is available. With this in mind, we focus on a dataset containing ground-truth VKT values for all $n = 1570$ road segments, computed at $L = 39$ time steps (corresponding to a 3-minute aggregation over the selected 2-hour interval) for each of the 101 simulated time series.

Training procedure

Out of the available data, 75 time series are used during the training process and the remaining 26 series are used for evaluation, with proportional distribution of different demand scales in both datasets. The training data is used both to construct the mean function of the Gaussian

process in (1) and to learn the parameters of the kernel capturing the spatio-temporal correlations introduced in (3)-(4).

Recall that the Gaussian process is defined over inputs of the form $\mathbf{x} = (v_i, t_j)$, where v_i denotes a road segment and t_j a discrete time step. Since repeated realizations of the traffic process are available from simulation, we define the mean function $m(\mathbf{x})$ in a nonparametric manner as an empirical average over historical observations at each spatio-temporal location. We assume that complete ground-truth information is available for 30 days, represented by 30 time series collected in the set $\mathcal{H}_{\text{mean}}$. Then, if we let $y^{(\ell)}(\mathbf{x})$ denote the traffic value observed at location $\mathbf{x}$ in the ℓ-th series, the historical mean can be defined as

$$m(\mathbf{x}) = \frac{1}{|\mathcal{H}_{\text{mean}}|} \sum_{\ell \in \mathcal{H}_{\text{mean}}} y^{(\ell)}(\mathbf{x}).$$

The average demand scale of the series included in the set $\mathcal{H}_{\text{mean}}$ and used to calculate the historical mean is approximately 1.27, which is close to the original historical demand scale.

The remaining part of the training set, denoted by $\mathcal{H}_{\text{ker}}$, is averaged into a single representative time series and used to learn the parameters of the kernel $k(\mathbf{x}, \mathbf{x}')$, describing the correlations in urban traffic in case of greater congestion. More precisely, rather than estimating the kernel parameters directly from the raw traffic observations $y^{(\ell)}(\mathbf{x})$, we introduce an averaging step that yields the smoothed observations

$$\overline{y}(\mathbf{x}) = \frac{1}{|\mathcal{H}_{\text{ker}}|} \sum_{\ell \in \mathcal{H}_{\text{ker}}} y^{(\ell)}(\mathbf{x}),$$

thereby reducing the noise induced by stochastic fluctuations in traffic observations and providing a cleaner representation of the underlying spatio-temporal traffic dynamics. Furthermore, during this learning stage, we assume that only partial observations are available at each time step. In particular, since estimating spatio-temporal correlations is more sensitive than estimating the mean, and to reflect the limited observability of the road network throughout the day, we restrict the available data to averaged ground-truth values on the road segments covered by 100 randomly distributed drones.

The GP hyperparameters and the noise variance $\theta = \left[\sigma_f, \nu_g, \kappa_g, \ell_{\mathcal{T}}, \sigma_n\right]$ are estimated by maximizing the marginal log-likelihood of the training data. Let $\mathbf{y}$ denote the smoothed observations and $\mathbf{X}$ denote the training inputs (i.e., road segments and time steps) corresponding to those observations. Since the posterior distribution is Gaussian, the optimal parameters are obtained by maximizing $L(\theta) = \log p(\mathbf{y} \mid \mathbf{X}, \theta)$, i.e.,

$$\theta^* = \underset{\theta}{\operatorname{argmax}}\, L(\theta) = \underset{\theta}{\operatorname{argmax}} \left(-\frac{1}{2}(\mathbf{y} - \mathbf{m})^T \mathbf{K}_\theta^{-1} (\mathbf{y} - \mathbf{m}) - \frac{1}{2} \log\det(\mathbf{K}_\theta) - \frac{N}{2} \log(2\pi) \right),$$

where N denotes the number of available observations and $\mathbf{K}_\theta = \mathbf{K}_{\mathbf{XX}} + \sigma_n^2 \mathbf{I}_N$. Training proceeds for a fixed number of gradient-based iterations, after which the optimized hyperparameters define the trained GP model used for posterior inference. All computations, including likelihood evaluation, gradient backpropagation, and parameter updates, were implemented using the GPyTorch library (Gardner et al. (2018)).

4 Results and discussion

Case study

In this case study, a fleet of P drones is distributed in the urban area of the city of Barcelona. It is assumed that each drone can monitor an area of 160 m × 90 m with its camera from an altitude of approximately 100 m. Hence, the urban area is divided into rectangular subareas of the specified size, as shown in Figure 1. The simulation runs for 2 hours, starting at the beginning of the main demand, divided into 39 equidistant time steps.

At each time step, each of the P drones is located in one of the cells of the grid, and a single cell is never monitored by two or more drones at the same time. We assume that a drone can observe all road segments that are (even partially) inside the cell under monitoring, so that it is capable of providing complete and accurate information about the VKT in those road segments.

We also assume that drones can change their positions by moving to one of the four adjacent cells in a way that they synchronously take the new positions at the beginning of the next time step. The interval between two consecutive time steps is 3 minutes, which is enough for a drone to take precise observations in the current cell and move to an adjacent cell.

The Gaussian process parameters are first determined from the training data using the described procedure and fixed during the evaluation process. A time series from the test set $\mathcal{H}_{\text{test}}$, which contains 26 time series in total, is then used to evaluate the quality of the GP estimates on a previously unseen day. The results reported below are the average over all 26 time series in $\mathcal{H}_{\text{test}}$. Using the GP parameters learned from historical data together with the observations collected by the drone fleet at the current time step t and k preceding time steps $t-1, t-2, \ldots, t-k$, we can compute the posterior distributions of the traffic variables for all road segments. When the objective is to estimate the traffic variables at the current time step t, we refer to the task as *traffic state estimation*. When the objective is instead to infer the traffic variables one step ahead, at time $t+1$, we refer to it as *traffic state prediction*.

The estimation error of the Gaussian process for an arbitrary query point $\mathbf{x}_* = (v_*, t_*)$ is given by $(\hat{y}_* - y_*)^2$, where y_* is a ground-truth value of the traffic parameter in question for a road segment v_* at a time step t_*, and $\hat{y}_*$ is the GP estimate. To evaluate the estimation quality across the entire network of n road segments at a fixed time step t_*, we use the normalized root mean square error given by formula

$$\text{err} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} y_i^2}}$$

The performance of the GP is compared against a baseline based solely on historical data, namely the error obtained when using the historical mean at time step t_* as the estimate for the current day. In addition, we also consider a second benchmark that combines the historical mean with current observations: for all road segments covered by drones, the observed values are used directly, while for unobserved segments, the historical mean is used.

Finally, we also investigate how the number of drones, which is directly related to the number of observations available, improves the historical mean estimate and influences the quality of the GP estimate. We perform this analysis for two deployment scenarios: static and moving drones. In the former case, drones do not change positions between time steps, and observations are always available for the same subset of road links. In the latter case, drones move to one of the adjacent cells at every time step. In both cases, the positions of the drones are initialized randomly, without overlapping. Since some cells are more informative, the result depends on the stochastic distributions of drones. Therefore, each analysis is averaged over 100 runs.

Traffic state estimation

We deploy a fleet of $P \in \{10, 20, 40, 60, 80, 100, 150, 200\}$ randomly distributed drones on the grid, and report the results of the VKT estimation for each road segment at the current time step t using partial observations at the current and two previous time steps, applying the methods described above. The procedure is repeated for all 26 time series from the test dataset, and the distribution of the average estimation error (normalized RMSE) versus the number of active drones is shown in Figure 2. The results are compared between four different methods: historical mean, historical mean with observations, Gaussian process with fixed drones (static case), and Gaussian process with moving drones (dynamic case). In addition to the average error, the variation of the estimation error across the time series is illustrated in Figure 3 in the form of boxplots, providing the range between 25th and 75th percentile of the error, as well as the maximum and minimum error values, for all the test series and for each of the four estimation methods.

By using current drone observations in combination with the historical mean, the estimation error can be significantly reduced, as shown in Figure 2. When the GP framework is applied, the error is reduced additionally, achieving noticeably better performance, in both the static and the dynamic case. In the latter, when drones monitor connected cells in the neighborhood during consecutive time steps by making random movements to adjacent fields, the best results are achieved. The improvement of the GP is particularly valuable in the case of series with larger demand scale than in the historic case, which results in more consistent results and narrower confidence intervals in Figure 3. Again, the dynamic GP approach achieves the best performance in terms of robustness.

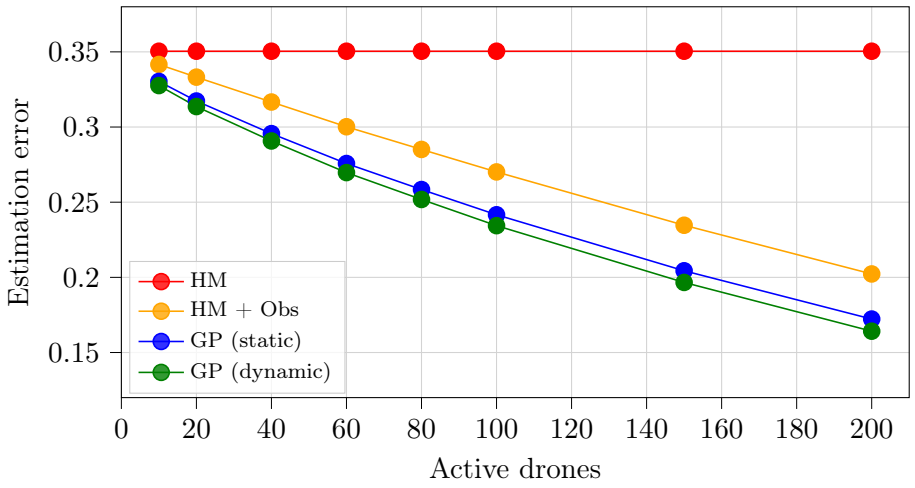


Figure 2: Mean estimation error across different numbers of active drones. Partial observations from time steps $t = 24, 25, 26$ are used to predict the full-graph values at $t = 26$. Each curve represents the average performance over all test time sequences, where for each sequence the results are first averaged over 100 simulation runs with different initial drone placements, and then averaged across all sequences.

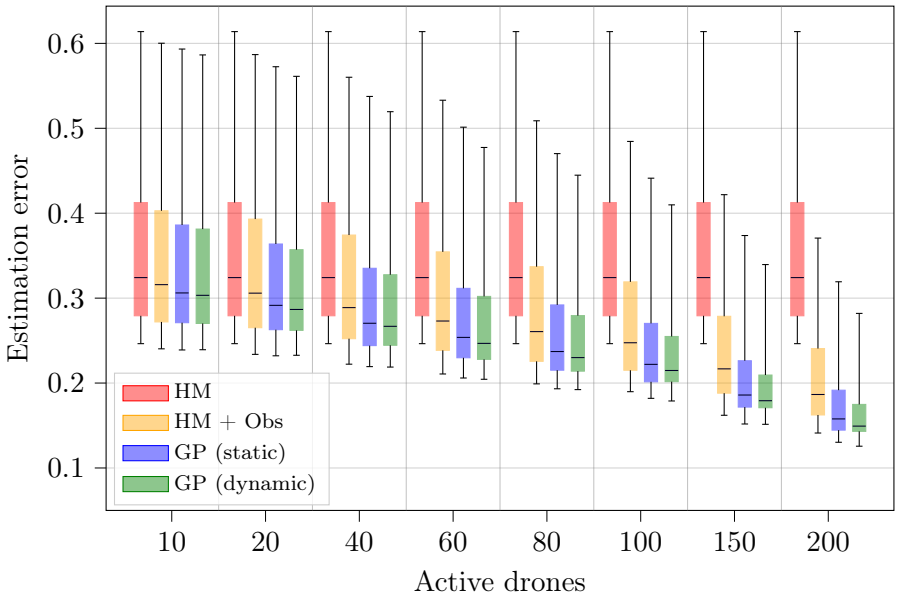


Figure 3: Estimation error comparison across different numbers of active drones. Partial observations from time steps $t = 24, 25, 26$ are used to predict the full-graph values at $t = 26$. For each time sequence, results are averaged over 100 simulation runs with different initial drone placements, and the boxplots therefore reflect the variability of these averaged performances across time sequences.

Traffic state prediction

In the second task, we evaluate the ability of the previously introduced methods to forecast VKT values one time step ahead of the latest available observations. As in the estimation task, a fleet of $P \in \{10, 20, 40, 60, 80, 100, 150, 200\}$ randomly distributed drones is deployed on the grid, and partial observations collected at the two preceding time steps are used to predict VKT values for all road segments at the next time step. We report results for two different time windows within the simulation horizon, as shown in Figure 4. Each curve represents the average error over 26 test time sequences, where for each sequence results are first averaged over 100 simulation runs with different initial drone placements and then averaged across sequences.

Since the prediction target lies in the future with respect to the latest measurements, no drone

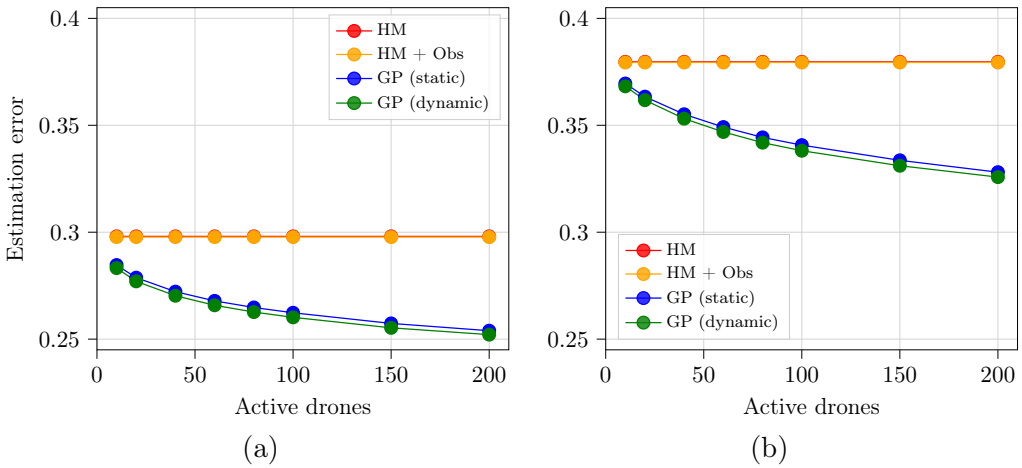


Figure 4: Mean prediction error across different numbers of active drones. In (a), partial observations from $t = 5, 6$ are used to predict the full-graph values at $t = 7$, while in (b), observations from $t = 30, 31$ are used to predict the values at $t = 32$.

observations are available at the target time step itself. Thus, the historical mean with observations reduces to the plain historical mean method. The two curves therefore overlap in both plots. The GP based methods outperform the historical mean methods, though the magnitude of the error depends on the time window. In Figure 4b, all methods exhibit a higher error compared to Figure 4a. In both experiments, the relative ordering of the methods is preserved and the GP provides a moderate improvement already with a small number of drones. The GP with static drones and the dynamic variant achieve very similar performance, in contrast to the estimation experiment where the dynamic case showed a clearer advantage. Overall, these results indicate that the proposed spatio-temporal GP can be used for short-horizon forecasting of network-wide traffic parameters from sparse drone observations, and that its advantage over the historical mean is preserved across different phases of the daily demand profile.

5 Conclusions

In this paper, we show how drones can be used as mobile sensors to monitor and predict traffic states in urban regions in real time. To this end, we introduce a spatio-temporal GP framework designed to estimate and make short-term predictions of traffic variables using partial observations collected by a fleet of drones. The proposed model integrates a graph-based spatial kernel that accounts for the structure of the road network with a temporal kernel that captures how traffic changes over time. We evaluate the framework through a simulated case study based on Barcelona's urban road network. The results show that our method outperforms historical mean approaches, which are generally strong baselines due to their dependence on extensive historical data. Building on these findings, future work will investigate the design of drone trajectories that explicitly account for predictive uncertainty and aim to reduce it, as well as on validating the proposed framework using real-world traffic data.

Acknowledgements

This work was supported as a part of NCCR Automation, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 51NF40_225155).

References

Barmpounakis, E., & Geroliminis, N. (2020). On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment. *Transportation Research Part C: Emerging Technologies*, *111*, 50-71. doi: https://doi.org/10.1016/j.trc.2019.11.023

Borovitskiy, V., Azangulov, I., Terenin, A., Mostowsky, P., Deisenroth, M., & Durrande, N. (2021, 13–15 Apr). Matérn Gaussian processes on graphs. In A. Banerjee & K. Fukumizu (Eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics* (Vol. 130, pp. 2593–2601). PMLR.

Casas, J., Ferrer, J. L., Garcia, D., Perarnau, J., & Torday, A. (2010). Traffic simulation with Aimsun. In *Fundamentals of traffic simulation* (pp. 173–232). Springer.

Elloumi, M., Dhaou, R., Escrig, B., Idoudi, H., & Saidane, L. A. (2018). Monitoring road traffic with a UAV-based system. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)* (p. 1-6). doi: 10.1109/WCNC.2018.8377077

Espadaler-Clapés, J., Barmpounakis, E., & Geroliminis, N. (2023). Traffic congestion and noise emissions with detailed vehicle trajectories from UAVs. *Transportation Research Part D: Transport and Environment*, *121*, 103822. doi: https://doi.org/10.1016/j.trd.2023.103822

Fonod, R., Cho, H., Yeo, H., & Geroliminis, N. (2025). Advanced computer vision for extracting georeferenced vehicle trajectories from drone imagery. *Transportation Research Part C: Emerging Technologies*, *178*, 105205. doi: https://doi.org/10.1016/j.trc.2025.105205

Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., & Wilson, A. G. (2018). GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. In *Advances in Neural Information Processing Systems.*

Ke, R., Li, Z., Tang, J., Pan, Z., & Wang, Y. (2019). Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow. *IEEE Transactions on Intelligent Transportation Systems*, *20*(1), 54-64. doi: 10.1109/TITS.2018.2797697

Kim, S., Tak, Y., Barmpounakis, E., & Geroliminis, N. (2024). Monitoring outdoor parking in urban areas with unmanned aerial vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *25*(10), 13393-13406. doi: 10.1109/TITS.2024.3397588

Kouvelas, A., Saeedmanesh, M., & Geroliminis, N. (2017). Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. *Transportation Research Part B: Methodological*, *96*, 26-45. doi: https://doi.org/10.1016/j.trb.2016.10.011

Krajewski, R., Bock, J., Kloeker, L., & Eckstein, L. (2018). The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In *2018 21st International Conference on Intelligent Transportation Systems* (p. 2118-2125). doi: 10.1109/ITSC.2018.8569552

Maljkovic, M., & Geroliminis, N. (2025). On learning-based traffic monitoring with a swarm of drones. In *2025 European Control Conference (ECC)* (p. 1842-1847). doi: 10.23919/ECC65951.2025.11186821

Moshahedi, N., & Kattan, L. (2023). Alpha-fair large-scale urban network control: A perimeter control based on a macroscopic fundamental diagram. *Transportation Research Part C: Emerging Technologies*, *146*, 103961. doi: https://doi.org/10.1016/j.trc.2022.103961

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning.* The MIT Press.

Tsitsokas, D., Kouvelas, A., & Geroliminis, N. (2023). Two-layer adaptive signal control framework for large-scale dynamically-congested networks: Combining efficient max pressure with perimeter control. *Transportation Research Part C: Emerging Technologies*, *152*, 104128. doi: https://doi.org/10.1016/j.trc.2023.104128

Xiong, W., Fonod, R., Alahi, A., & Geroliminis, N. (2025). Multi-source urban traffic flow forecasting with drone and loop detector data. *IEEE Transactions on Intelligent Transportation Systems*, *26*(11), 18637-18652. doi: 10.1109/TITS.2025.3605014

Zoabi, R., & Haddad, J. (2026). Enhanced queue-based max-pressure traffic signal control. *Transportation Research Part C: Emerging Technologies*, *186*, 105538. doi: https://doi.org/10.1016/j.trc.2026.105538