A utility-based traffic sensor network optimization for data-driven traffic prediction

Yuxing Cheng, Marco Rinaldi, Serge Hoogendoorn, Bart van Arem

Department of Transport and Planning, Delft University of Technology, The Netherlands

ABSTRACT

Traffic sensors collect real-time data on current traffic conditions, and much research has focused on determining optimal sensor locations for various monitoring and information-inference tasks. Meanwhile, data-driven models for time-series traffic prediction increasingly rely on these sensor inputs. However, existing sensor-location studies rarely account for the specific role of sensor data in data-driven prediction models. This work aims to reduce the size of the sensor network with only a minor impact on model performance. We proposed a genetic algorithm method for traffic sensor network optimization, with prediction accuracy and quantified epistemic uncertainty as the objectives. The results show that sensors distributed in various locations have different utilities for prediction performance. In the case study of Amsterdam highway ring roads (A10), we can obtain a reduction of 60% of the total sensor budget while preserving adequate prediction model performance.

Keywords: Traffic sensor location problem, traffic speed prediction, uncertainty quantification, heuristic algorithm.

1. INTRODUCTION

Real-time information from traffic sensor networks is crucial for traffic operators. The Traffic Sensor Location Problem (TSLP) addresses two fundamental issues: how many sensors are needed, and where they should be placed (Owais 2022). Current research on TSLP has examined the problem under various objectives, such as network observability (Castillo et al. 2012; Viti et al. 2014) and information entropy (Lo 1991). The applications based on the sensor network mentioned in the existing literature are mostly focused on traffic monitoring and information inference, such as OD estimation (Zhou and List 2010), link flow inference, or travel time estimation. However, we found little research exploring how sensor networks can be optimized for data-driven prediction models.

Meanwhile, studies on data-driven traffic prediction often overlook how the sensor network design affects predictive uncertainty, even in settings with sparse sensor coverage (Liu et al. 2024). Data-driven methods are argued to be data-hungry, which hinders them from wider real-world applications. Beyond that, existing studies usually ignore the uncertainty of the prediction model, which could have a significant impact on the follow-up traffic management application, such as predictive traffic signal control (Poelman et al. 2023).

These gaps prompt the question: How much sensor data do data-driven models truly need for robust and accurate traffic predictions? And what if certain sensors are turned off—does the model's performance deteriorate significantly?

This study aims to optimize the traffic sensor network according to its utility in data-driven prediction models regarding prediction accuracy and uncertainty. To achieve this goal, we reveal the relationship between sensor network design, including the number (budget) and location of sensors, and the prediction performance based on the sensor data. This process will discern the most informative samples from the sensor network in the context of the data-driven prediction model. These samples contain critical information regarding different prediction performance indicators, regarding the accuracy and uncertainty of the prediction model.

In the data-driven prediction models, two types of uncertainty are considered: aleatoric and epistemic uncertainty. Aleatoric uncertainty describes the information that the model cannot interpret and can only be reduced by improving the accuracy of the collected data, up to an irreducible threshold related to the uncertainty inherent in the physical phenomena of traffic. Epistemic Uncertainty, on the other hand, describes how much the chosen predictive model accurately matches the modeled phenomenon, which can be improved by better modeling or by collecting more data. Thus, by quantifying the epistemic uncertainty of the prediction models, we can investigate the impact of incomplete/missing/partial input data on the prediction model. In this study, we employ Deep Ensembles (DE) (Lakshminarayanan, Pritzel, and Blundell 2017) to quantify both aleatoric and epistemic uncertainty associated with network-level, multi-step highway speed forecasting.

By systematically reducing the number of sensors—modeled as graph nodes in the input—we show that our proposed approach can maintain prediction accuracy and minimize uncertainty despite fewer sensors. This result highlights the method's potential to lower sensing costs without sacrificing reliable, robust traffic predictions.

2. METHODOLOGY

Due to the nonlinearity of data-driven prediction models, each sensor's data collectively contributes to the model output, making it impractical to assign an absolute utility to any single sensor. Consequently, a realistic way to optimize a large-scale sensor network is to seek a near-optimal sensor configuration. To avoid early convergence on a local optimum, which is a common pitfall of gradient-based methods, we employ a heuristic Genetic Algorithm (GA) to explore different sensor arrangements under various objective functions.

In this study, we assume historical time series data of the target prediction variable are available for all potential sensor locations. Under different constraints on the number of sensors, we then optimize which sensors to select, with either prediction accuracy or quantified uncertainty of pre-trained data-driven models as the fitness in GA.

Uncertainty quantification in prediction

We adopt variance-based and differential entropy metrics to assess uncertainty in traffic speed predictions. In the variance-based quantification, we note that $(\mu_{m,t}, \sigma^2_{m,t})$ are the mean and variance of the probabilistic distribution prediction of each model m at the time step t, then the law of total variance decomposes the total uncertainty into the following components (Lakshminarayanan, Pritzel, and Blundell 2017):

$$Var(\hat{y}_{m,t}) = \mathbb{E}(\sigma^2_{m,t}) + Var(\mu_{m,t}) \tag{1}$$

Where $\mathbb{E}(\sigma^2_{m,t})$ represents aleatoric uncertainty, and $Var(\mu_{m,t})$ represents the epistemic uncertainty. Aleatoric uncertainty originates from the inherent unpredictability or limited observability of critical factors influencing the traffic process—such as sensor types, measurement errors, or the inherent randomness of traffic dynamics. Epistemic uncertainty, by contrast, reflects gaps in knowledge that could be filled with more complete data or improved models. It stems from the modeling process itself and limitations in data coverage. High epistemic uncertainty implies that more data are needed, making it especially relevant to sensor network design.

Genetic Algorithm

The process of heuristic sensor network optimization begins with initializing an initial population $P_0 = \{x_1^0, x_2^0, \ldots, x_n^0\}$, which is a set of candidate solutions of sensor combination under the same sensor amount constraints. Each individual x_n^i represents a potential combination of sensors, where i indicates the counter of specific amount constraints, and n indicates the number of individuals in the population.

As the algorithm iterates, a new population P_i is generated in each iteration. If i is not zero, P_i is formed by randomly selecting i individuals from the best individual found in the previous iteration, denoted as x_{best}^{i-1}. This selective approach helps concentrate the search around promising regions of the solution space, leveraging the knowledge gained from earlier iterations. Within each population P_i , the algorithm evaluates the fitness for each sensor selection set x_n^i. Subsequently, the best individual x_{best}^i is updated to reflect the most effective sensor configuration found in that iteration.

Algorithm 1. Pseudocode for genetic algorithm-based sensor selections

Input initial population: P_0, i=0 {iteration counter}
IF i!=0:
    P_i= random select *m* individuals in x_{best}^{i-1}
End if
For each generation in P_i:
    CALCULATE fitness for $\{x_1^i, x_2^i \ldots, x_n^i\}$ in P_i
    IF no significant improvement over several generations:
        BREAK early
    ADJUST crossover and mutation rates based on improvement
    SELECT top-performing individuals (elites) to carry over
    GENERATE a new population P_i through **crossover** and **mutation**
    UPDATE the best individual x_{best}^i
END FOR
Save x_{best}^i
i +=1

Prediction model

In this paper, we used some classical and competitive traffic speed prediction models in the state of the art. We aim to expand this testing to other prediction models in future research. Considering the graph structure of the sensor network, we employ two graph neural network (GNN) approaches for this case study: Spatio-Temporal Graph Convolutional Networks (STGCN) (STGCN) (Yu, Yin, and Zhu 2018) and Spatial-Temporal Graph Attention Networks STGAT(Kong et al. 2020). Although these models share certain features, their distinct architectures help us assess whether different structural choices lead to varying optimal sensor distributions.

For computational efficiency, we pre-train all models using the full sensor network data over the same time period. We then apply the proposed GA to optimize which sensors feed into each pre-trained model. For quantifying the epistemic uncertainty, we assume the prediction output of each model m at the time step t as $\mu_{m,t}$ referred in Eq. (1).

Our case study is drawn from the Amsterdam highway ring roads (A10) in 2021, capturing the speed data. This network comprises nine highways connecting downtown Amsterdam, nearby towns, and Schiphol International Airport. Each highway is uniformly divided into 400-meter segments, and each segment hosts a single sensor, which leads to a total of 193 loop detectors. Data are provided by the Dutch National Data Warehouse (NDW, www.ndw.nu). In Figure 1 sensors on different roads are distinguished by color-coded nodes.

Figure 1. The network with 193 sensors on the Amsterdam highway ring roads (A10)

3. RESULTS AND DISCUSSION

This section presents our findings on the selected case study. We begin by applying the method proposed by (Yang and Zhou 1998) to establish a lower bound for sensor selection that ensures full observation of the highway network. Here, "full observation" means every vehicle passing through the network must encounter at least one sensor.

In our case study, the network comprises 37 possible origins and destinations, including on- and off-ramps as well as connections to external roads. As illustrated by the red nodes in Figure 2, 69 sensors—35.8% of all sensors—are identified to meet this lower-bound requirement. We use this result as a baseline for subsequent comparisons.

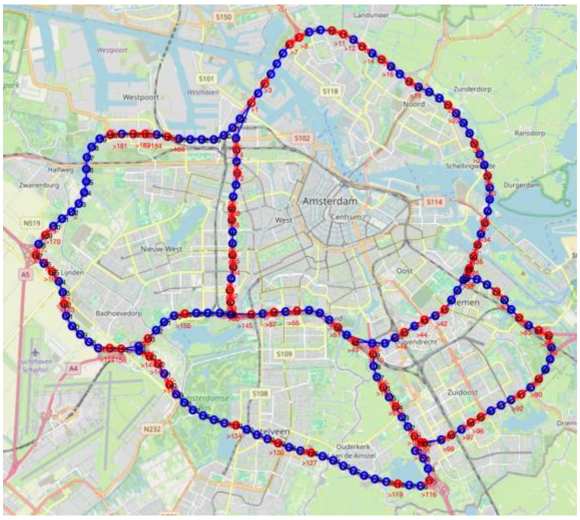
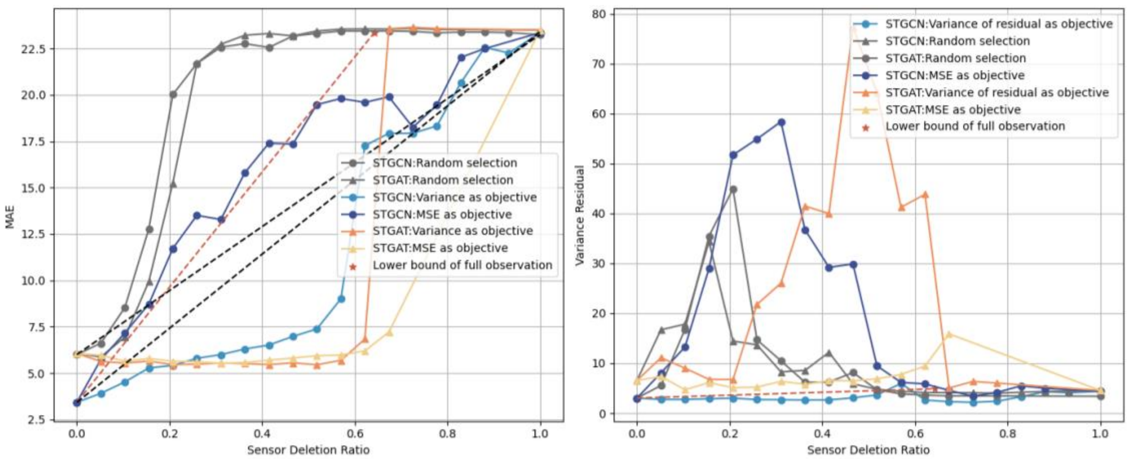


Figure 2. The lower bound of sensors with full observation of the Amsterdam highway rings.

Performance of prediction model under sensor budget constraints

We use two objectives, MSE and epistemic uncertainty on all the sensor locations, which are chosen for GA optimization under different sensor amounts as constraints. Figure 3 illustrates how both metrics evolve as the number of sensors decreases. The red star marks the model performance under the sensor configuration depicted in Figure 2, which reflects the lower bound of full observation. The black dash lines as references reflect the linear decay of metrics with sensor deletion.



a) Accuracy (MSE) b) Epistemic Uncertainty

Figure 3. Relationship between prediction performance (accuracy, uncertainty) and sensor deletion ratio

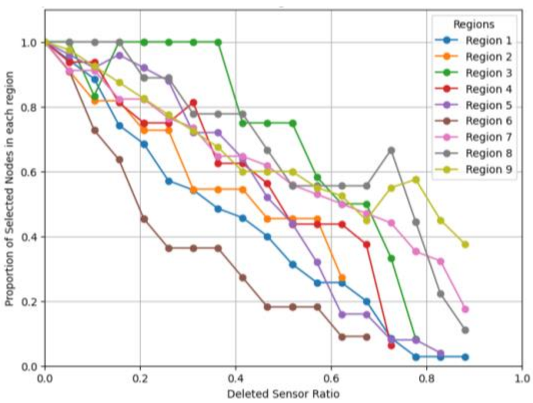
Comparing the results with random sensor selection under increasing deletion ratios, our proposed heuristic algorithm significantly mitigates declines in both accuracy and uncertainty. Notably, when about 60% of sensors are removed, there is a sharp drop in accuracy (illustrated in Figure 3 a) accompanied by a spike in uncertainty (illustrated in Figure 3 b)—signaling the collapse of the model's accuracy. Crucially, the proposed method helps avoid this collapse, and the sensor count

at the threshold of a sharp drop in accuracy (around 60%) is close to the number of sensors required for full observation.
Interestingly, the sensor configuration meeting the lower bound of full observation underperforms even a random selection. This outcome suggests that ensuring every vehicle is recorded at least once on each section (sensor-to-sensor) it traverses (i.e., full observability) does not necessarily align with providing the most valuable data for predictive modeling. This is not surprising, considering that the full observability problem is non-unique, and therefore multiple sensor location solutions might exhibit different performance on prediction tasks, as was shown in earlier works (Rinaldi and Viti 2017).

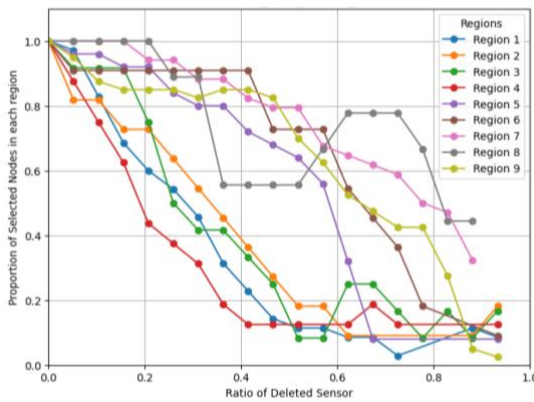
Moreover, the significant increase in epistemic uncertainty underscores the rapid decline in accuracy, indicating that our optimized sensor network is pushed to its limit in preserving model performance.

Distribution of sensor selection
We analyzed how sensors are distributed across various segments of the ring roads under different sensor budget constraints.

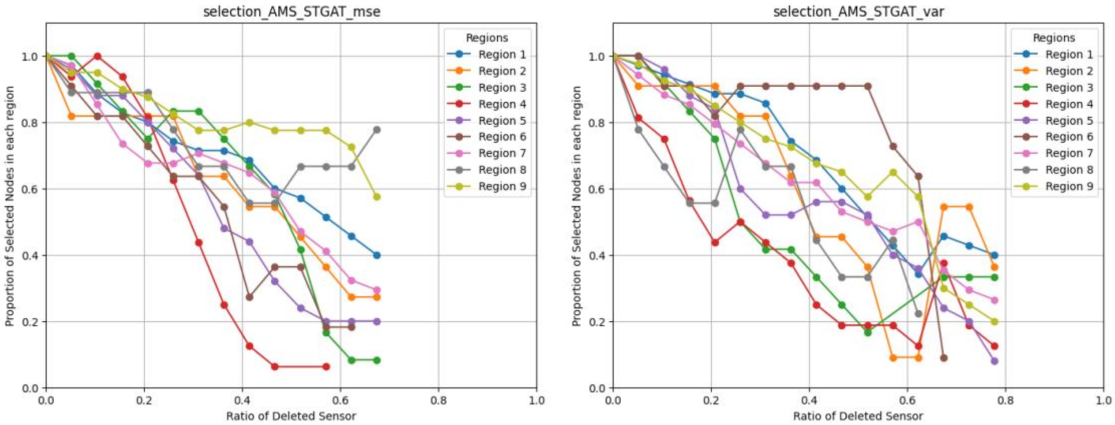
Figure 4 illustrates how the proportion of selected sensors shifts among different regions. When accuracy is the objective, the rate at which sensors are omitted follows a similar pattern across all regions (evidenced by the relatively narrower spread of curves in Figure 4 a) and Figure 4 c), compared to Figure 4 b) and Figure 4 d)). However, when the uncertainty is regarded as objective, sensors in the inner rings around central Amsterdam (Regions 1–4) are more likely to being excluded. This phenomenon suggests that while sensors across different roads contribute roughly equally to predictive accuracy, their impact on model uncertainty is more uneven—particularly when comparing the inner rings (Regions 1–4) with the outer rings (Regions 5–9).



a) Accuracy as objective (STGCN)



b) Uncertainty as objective (STGCN)



c) Accuracy as objective (STGAT)   d) Uncertainty as objective (STGAT)

Figure 4. The change in the proportion of selected sensor in different regions, with different optimization objective

We finally examined whether partially deleting sensors from the prediction model input would change the distribution of prediction performance across different highway segments. Figure 5 displays the distribution of prediction metrics at various sensor deletion ratios, using a color scale (red to green) centered on the average value of corresponding metric in each selection scenario. These uneven distributions of accuracy and uncertainty across different roadway locations illustrate the model's trade-off between global and local prediction performance under certain budgets. However, as shown in Figure 5, only marginal shifts in prediction performance are observed when sensor locations vary under different sensor constraints.

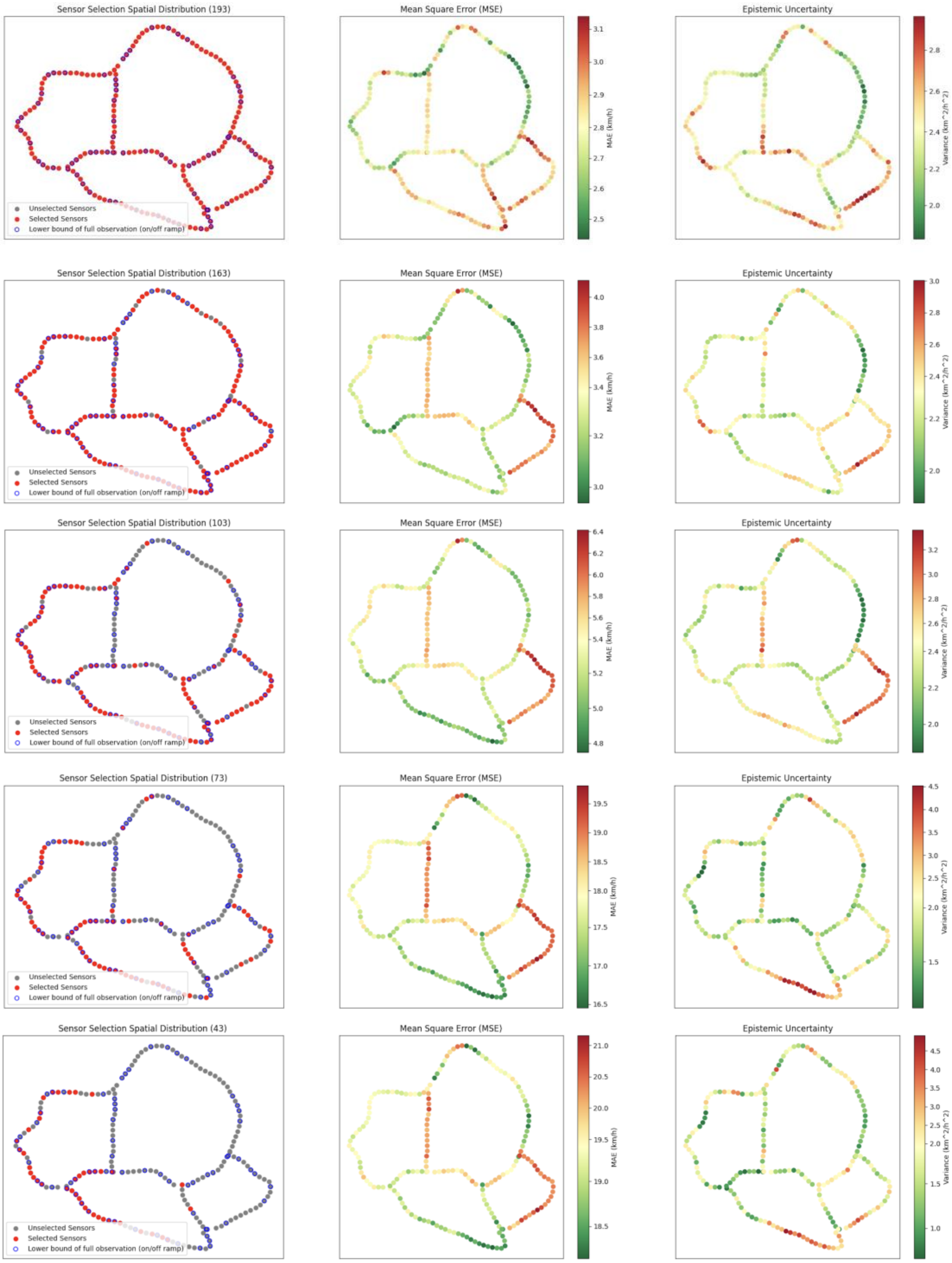


Figure 5 The distribution of prediction performance with different sensor deletion ratios (STGCN model, accuracy as objective)

4. CONCLUSIONS

Not all locations where data may be collected provide equal predictive information. In this study, we seek a reduction in sensor network costs for data-driven time-series traffic prediction tasks, considering this underlying hypothesis. By selecting the most valuable sensors in the network under budget constraints, the proposed utility-based sensor network optimization method shows its potential to preserve much of the value of the sensor to the predictive models, at a lower sensor budget. Beyond this, the method is proven to identify the threshold as a reference for the sensor network cost that may prevent significant prediction model performance failure caused by limited data coverage.

Future work could explore additional TSLP approaches, particularly those incorporating network structures—beyond the observability-based strategy presented here. Such comparisons may clarify whether network structure or historical data predominantly determines sensor-network utility within data-driven prediction models. With further investigation, the proposed method may also offer counterfactual explanations by pinpointing the most informative data samples in a dataset, thereby enhancing model interpretability.

ACKNOWLEDGEMENTS

We would like to express our gratitude for the financial support provided by the Netherlands National Science Foundation (NWO) through the PERSPECTIEF Program P21-08'XCARCITY.

REFERENCES

Castillo, Enrique, Ana Rivas, Pilar Jiménez, and José María Menéndez. 2012. “Observability in Traffic Networks. Plate Scanning Added by Counting Information.” *Transportation*. doi:10.1007/s11116-012-9390-0.

Kong, Xiangyuan, Weiwei Xing, Xiang Wei, Peng Bao, Jian Zhang, and Wei Lu. 2020. “STGAT: Spatial-Temporal Graph Attention Networks for Traffic Flow Forecasting.” *IEEE Access* 8. doi:10.1109/ACCESS.2020.3011186.

Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. 2017. “Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles.” In *Advances in Neural Information Processing Systems*,.

Lo, H. P. 1991. “Estimation Of Origin-Destination Matrix From Traffic Counts: A Comparison Of Entropy Maximizing And Information Minimizing Models.” *Transportation Planning and Technology* 16(2). doi:10.1080/03081069108717474.

Owais, Mahmoud. 2022. “Traffic Sensor Location Problem: Three Decades of Research.” *Expert Systems with Applications* 208. doi:10.1016/j.eswa.2022.118134.

Poelman, MC, A Hegyi, A Verbraeck, JWC van Lint, and jwc vanlint. 2023. “Structure-Free Model-Based Predictive Signal Control: A Sensitivity Analysis on a Corridor with Spillback.” doi:10.1016/j.trc.2023.104174.

Rinaldi, Marco, and Francesco Viti. 2017. “Exact and Approximate Route Set Generation for Resilient Partial Observability in Sensor Location Problems.” *Transportation Research Part B: Methodological* 105. doi:10.1016/j.trb.2017.08.007.

Viti, Francesco, Marco Rinaldi, Francesco Corman, and Chris M J Tampère. 2014. “Assessing Partial Observability in Network Sensor Location Problems.” *Transportation Research Part B* 70: 65–89. doi:10.1016/j.trb.2014.08.002.

Viti, Francesco, Wouter Verbeke, and Chris M.J. Tampère. 2008. “Sensor Locations for Reliable Travel Time Prediction and Dynamic Management of Traffic Networks.” *Transportation Research Record* (2049). doi:10.3141/2049-12.

Yang, Hai, and Jing Zhou. 1998. “Optimal Traffic Counting Locations for Origin-Destination Matrix Estimation.” *Transportation Research Part B: Methodological* 32(2). doi:10.1016/s0191-2615(97)00016-7.

Yu, Bing, Haoteng Yin, and Zhanxing Zhu. 2018. “Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting.” In *IJCAI International Joint Conference on Artificial Intelligence,* doi:10.24963/ijcai.2018/505.

Zhou, Xuesong, and George F. List. 2010. “An Information-Theoretic Sensor Location Model for Traffic Origin-Destination Demand Estimation Applications.” *https://doi-org.tudelft.idm.oclc.org/10.1287/trsc.1100.0319* 44(2): 254–73. doi:10.1287/TRSC.1100.0319.