Learning to Learn the Macroscopic Fundamental Diagram using physic informed and meta Machine Learning techniques

Amalie Roark,*[1], Guido Cantelmo[2], and Francisco Camara Pereira[3]

[1]M.Sc, Technical University of Denmark, Denmark
[2]Dr., Division of Transportation Science, Technical University of Denmark, Denmark
[3]Prof. Dr., Division of Transportation Science, Technical University of Denmark, Denmark

Short summary

The Macroscopic Fundamental Diagram (MFD) is a popular tool used to describe traffic dynamics in an aggregated way, with applications ranging from traffic control to incident analysis. However, estimating the MFD for a given network requires large numbers of loop detectors, which is not always available in practice.
This research proposes using Meta-Learning to alleviate this problem. Specifically, we propose using data from multiple different cities to train a Machine Learning (ML) model that can more accurately learn the MFD using limited data. First, we compare the traditional bi-parabolic model from the literature with a non-parametric Multi-Task Physics-Informed Neural Network (MT-PINN). Then, a Model-Agnostic Meta-Learning (MAML) framework is implemented to estimate MFDs when limited data is available. Results show that MAML successfully generalizes across diverse urban settings and improves performance on cities with limited data, and demonstrate the potential of using Meta-Learning when a limited number of detectors is available.
Keywords: MFD, Estimation, Multi-Task Physics-Informed Neural Network, Meta Learning, Transfer Learning, data scarcity.

1 Introduction

Cities are plagued by various degrees of congestion, with many negative impacts such as air pollution (Batista et al., 2022). A significant amount of research has been done to understand the nature and dynamics of urban congestion and how it propagates in the attempt to reduce its impact on the economy, health, and the environment. The Macroscopic Fundamental Diagram (MFD) extends the principles of the link-level fundamental diagram to an entire network, describing the relationship between the aggregated flow and density across a city or urban area (Daganzo, 2007). This concept was first introduced to capture the collective traffic dynamics at a macroscopic scale, and has since become a powerful tool for analyzing and managing urban traffic networks (Geroliminis & Daganzo, 2008).

Nevertheless, deriving the MFD from empirical data is a challenging exercise since (i) it requires data from a large number of loop detectors and (ii) detectors should cover the entirety of the network (Aghamohammadi & Laval, 2022). Several analytical and simulation-based methods have been proposed to address this problem. The Method of Cuts (Daganzo & Geroliminis, 2008), for example, allows estimating upper bounds for the MFD parameters. Laval & Castrillón (2015) and Aghamohammadi & Laval (2022) extended the method to consider bias in the data and stochasticity. However, these methods were proposed for estimating signalized corridors rather than the critical accumulation density or the parameters of the congested and uncongested branches of the MFD. The literature does not agree on one shape for the MFD. In practice, a bi-parabolic interpolation of the data is often used as an analytical approximation. Machine Learning holds the potential to reveal the true MFD shape, but data scarcity, especially for the congested branch, poses challenges. Researchers suggest mitigating this issue with Physic-Informed Machine Learning to embed constraints in the model designs (Yuan et al., 2021).

This paper builds upon and expands these findings. Meta-Learning - often referred to as "learning to learn" - is a paradigm aimed at training Machine Learning models that can generalize across multiple tasks or datasets. The idea is that, as we have an analytical understanding of the MFD, as well as data sets from different cities, this knowledge can be used to estimate the MFD in cities where only limited data is available. The contribution of this research is twofold. First, we propose a Multi-Task Physics-Informed Neural Network (MTPINN) framework for learning the shape of the MFD. The Multi-Task component considers co-dependency between traffic dynamics, while the Physics-Informed part ensures that the MFD shows well-known properties. Then, we combine this approach with Meta-Learning algorithms, with the aim of developing a model that can more easily generalize across different cities and hence be trained on a reduced amount of data.

2 Data

As the meta-Machine Learning model will require data from multiple cities, we first introduce the adopted dataset. This research uses the UTD19 dataset (Loder et al., 2020), which contains data from 39 cities. Among the 39 cities, however, only a few have speed data, and only three cities have both occupancy, flow, and speed data available. After data clearing, only 29 cities were kept out of the original 39. Data were pre-processed using the same steps discussed in (Aghamohammadi & Laval, 2022), which analyzed the same cities. The MFD plots, in terms of occupancy and flow, are depicted in Figure 1.

3 Methodology

The key variables used in the MFD are typically obtained by aggregating data from all road segments in the network. The network-wide flow is the average flow across all segments in the network, while the density is the weighted average density across all road segments. While density is commonly used in the formulation of the MFD, in practice, occupancy often serves as a surrogate for density. This also applies to this study, where we will analyze the MFD as a function of the aggregated flow and occupancy. Hence, in the next section, we estimate parametric and nonparametric models that approximate the plots reported in Figure 1. The average occupancy is the independent variable x, while the average flow is the dependent variable $f(x)$.

Bi-parabolic hybrid model

The bi-parabolic hybrid model is designed to use two parabolas that share the same vertex to approximate the MFD in each city. These functions correspond to a shape founded in theory and allow a physical interpretation of their parameters. The model encodes this relationship as two piecewise parabolic functions that meet at a shared vertex, representing the critical occupancy (CD) at which traffic flow reaches its maximum. It is implemented as a fully interpretable white-box model, with its parameters estimated using backpropagation. The first parabola describing the uncongested regime is required to pass through the origin, $(0,0)$, and then ends at the vertex (x_{cd}, f_{vertex}), and is defined as:

$$f_1(x) = -a_1(x - x_{cd})^2 + f_{vertex},\ x \leq x_{cd} \tag{1}$$

Where $a_1 = \frac{f_{vertex}}{x_{cd}^2}$ ensures that the parabola passes through the origin. The second parabola describes the congested regime and extends from the vertex to a second intersection with the x-axis:

$$f_2(x) = -a_2(x - x_{cd})^2 + f_{vertex},\ x \geq x_{cd} \tag{2}$$

where $a_2 > 0$. The negativity of a_2 is enforced through a logit transformation, ensuring that the parabola opens downward. The model is trained using a composite loss function, which can be summarized by:

$$\mathcal{L}_{total} = \beta(\mathcal{L}_1 + \mathcal{L}_2) + \alpha\mathcal{L}_\lambda \tag{3}$$

In this expression, $\mathcal{L}_1$ and $\mathcal{L}_2$ are the MSE losses for the two regimes given the two fits described in equations 1 and 2. The main feature of this model is that the parameters of the two parabolas and the critical occupancy are jointly estimated. $\alpha\mathcal{L}_\lambda$ is a penalty function that ensure that the estimated CD fits the data while being consistent with the theory. Finally, weights are applied to compensate for the uneven data distribution.

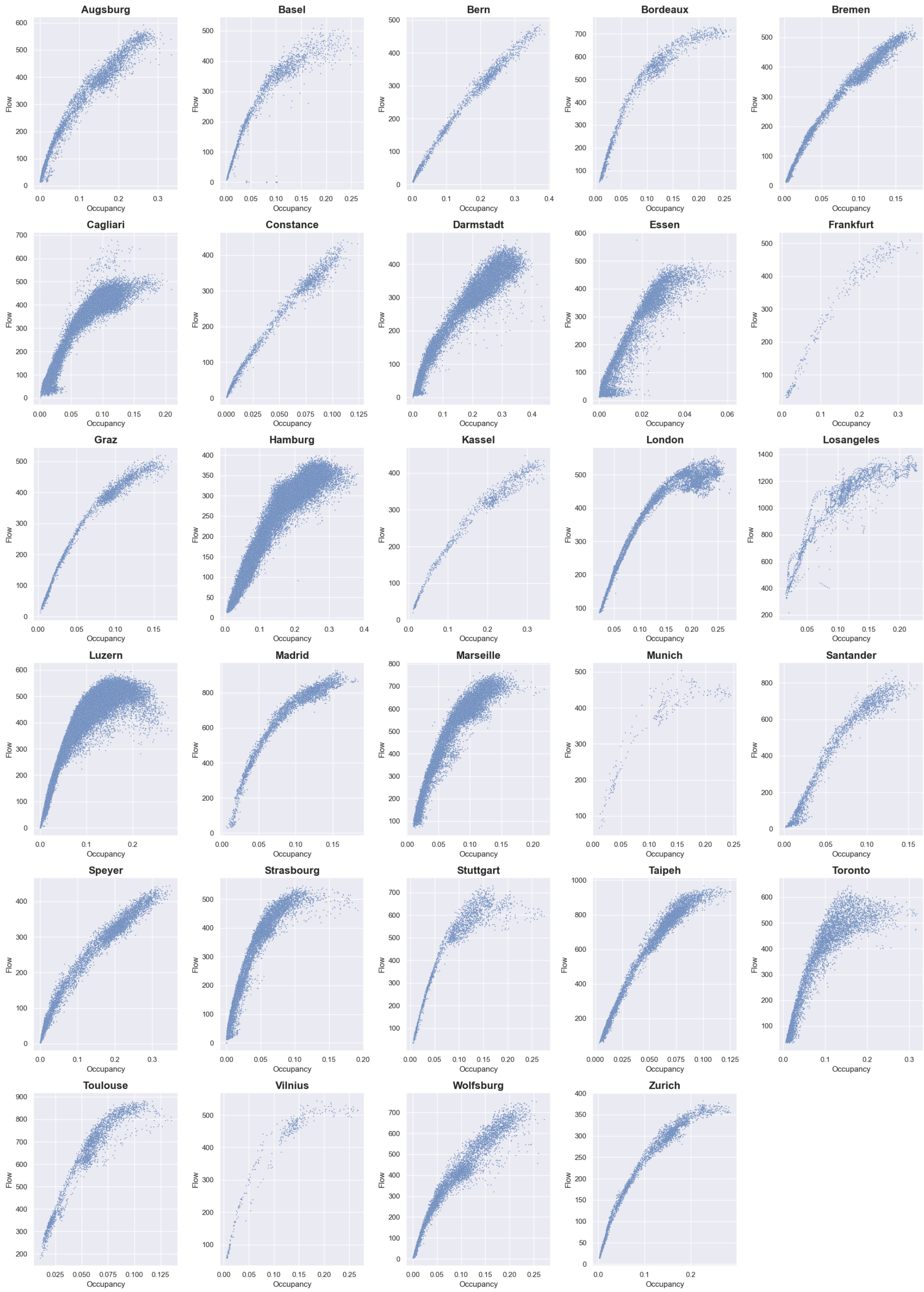


Figure 1: The plots show the average occupancy versus average flow for 29 cities based on data from UTD19

.

Multi-Task Physics-Informed NN

The Multi-Task Physics-Informed Neural Network (MTPINN) estimates three key parameters of the MFD: the critical occupancy, the maximum flow, and the flow-occupancy relationship. Multi-Task Learning (MTL) is a paradigm where multiple related tasks are learned simultaneously within a shared framework. A shared base includes common layers that learn a unified representation of the data, capturing common dynamics across different tasks. Each task has a separate output branch tailored to predict a specific target variable. By leveraging a shared representation (in our case, between critical occupancy, maximum flow, and flow-occupancy relationship), MTPINN improves learning efficiency and generalization.

While powerful, MTL is unlikely to capture the difference between congested and uncongested branch, especially with limited data. Therefore, we apply a Physic Informed regularization to the Loss Function to force the model respecting analytical properties of the MFD. Hence, the loss function is composed of two components:

$$\mathcal{L}_{total} = \mathcal{L}_{MSE} + \mathcal{L}_{Physics} \tag{4}$$

$\mathcal{L}_{MSE}$ represents the loss of the purely data driven model, while $\mathcal{L}_{Physics}$ computes the Physics-Informed loss, designed to encourage the predictions to follow a bi-parabolic shape. $\mathcal{L}_{Physics}$ is comprised of the following terms:

$$\mathcal{L}_{Physics} = w_1 \cdot \mathcal{L}_1 + w_2 \cdot \mathcal{L}_2 + \lambda_{offset} + \lambda_{scale} + \lambda_{max} \tag{5}$$

That is, two loss terms $\mathcal{L}_1$ and $\mathcal{L}_2$, weighted using w_1 and w_2, and three penalty terms. $\mathcal{L}_1$ and $\mathcal{L}_2$ guide the model predictions are designed to share the vertex. The first parabola describes the observations from 0 to the critical occupancy, x_{cd}. It is defined as follow:

$$f_1(x) = a_1 + (x - x_{cd})^2 + (f_{max} - f_{offset}) \tag{6}$$

where the value a_1 given by:

$$a_1 = -\frac{(f_{max} - f_{offset})}{x_{cd}^2} \tag{7}$$

The second parabola describes the observations to the right of the critical occupancy and is defined as:

$$f_2(x) = a_2 + (x - x_{cd})^2 + (f_{max} - f_{offset}) \tag{8}$$

where the value a_2 given by:

$$a_2 = -\frac{(f_{max} - f_{offset})}{((x_{scaler} - 1) \cdot x_{cd})^2} \tag{9}$$

The denominators in the expressions for a_1 and a_2 control the width of the parabolas, i.e., the horizontal distance between either intersection with the x-axis and the vertex point. For a_1, the desired distance is x_{cd}.

Meta-Learning

This section combines Meta-Learning with the MTPINN architecture. The Meta-Learning framework is based on the Model-Agnostic Meta-Learning (MAML) formulation (Finn et al., 2017) and is used to estimate the MFD diagram when only a limited number of detectors is equipped with sensors.
Meta-Learning is a Machine Learning paradigm aimed at training models that can generalize across multiple tasks or datasets and seeks to develop models that can quickly adapt to new tasks with minimal retraining (Finn et al., 2017). This is particularly valuable in scenarios where task-specific data is limited, but data about similar tasks is abundant. Meta-Learning typically involves two nested optimization loops:

- **Inner loop:** The model is trained on a specific task or dataset. This involves task-specific updates to the model's parameters to optimize performance for that task.
- **Outer loop:** The Meta-Learning algorithm adjusts the model's initial parameters or learning strategy based on feedback from multiple tasks. The goal is to identify a set of parameters that enables the model to perform well across a variety of tasks with minimal fine-tuning.

Dataset Construction: Traffic data is organized into city-wide groups, with separate training (support set) and validation (query set) datasets for each city. Each task in the ML model corresponds to predicting flow based on occupancy data for a specific city. The support set is used for inner-loop adaptation, while the query set is used to evaluate the performance in the outer loop. The support set consists of the MFD calculated from a subset of randomly selected detectors for each city. The chosen number of detectors for sampling is n = 75, 50, 25, and 10. Cities with less than 100 detectors were removed, reducing the number of cities from 29 to 21. For each city, this number of detectors is randomly sampled 30 times. This imitates the situation in which we cannot observe the true MFD in a new city due to limited data resulting from few detectors. The query set consists of the MFDs calculated from the complete set of detectors available for each city, thus corresponding to what is considered the true MFDs.

Meta-training Procedure: The structure of the meta-training is built around two optimization objectives. The objective of the inner loop is to minimize the task-specific loss $\mathcal{L}_{task}$ by adapting the model parameters θ for a given task $\mathcal{T}_i$:

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{task}(\mathcal{D}_i^{support}; \theta) \tag{10}$$

where θ are the shared meta-parameters that are cloned for each new task, θ_i' are the task-adapted parameters for $\mathcal{T}_i$, α is the inner loop learning rate, and $\mathcal{D}_i$ is the support set for $\mathcal{T}_i$, the sampled task i. The meta-objective is to minimize the expected loss over the query sets of all tasks after performing the inner-loop adaptation:

$$\min_\theta \mathbb{E}_{\mathcal{T}_i \sim p(\mathcal{T})}[\mathcal{L}_{task}(\mathcal{D}_i^{query}; \theta_i')] \tag{11}$$

where $\mathcal{D}_i^{query}$ is the query set for $\mathcal{T}_i$, θ_i' is computed from the inner loop, and $p(\mathcal{T})$ is the distribution over tasks.
By differentiating through the inner-loop updates, the meta-gradient is computed to optimize θ using the outer-loop learning rate β:

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_i \mathcal{L}_{tassk}(\mathcal{D}_i^{query}; \theta_i') \tag{12}$$

Algorithm 1 Adapted MAML for Few-Shot Supervised Learning

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters (learning rates)
  randomly initialize θ
  while not done **do**
    Sample batch of tasks $\{\mathcal{T}_i\} \sim p(\mathcal{T})$
    for all $\mathcal{T}_i$ **do**
      Sample $K \cdot N$ obs. $\mathcal{D}_i^{\text{support}} = \{x^{(j)}, y^{(j)}\}_{j=1}^K$ from $\mathcal{T}_i$
      Sample M obs. $\mathcal{D}_i^{\text{query}} = \{x^{(j)}, y^{(j)}\}_{j=1}^M$ from $\mathcal{T}_i$
      Initialize task-specific parameters $\theta_i' \leftarrow \theta$
      for N **inner loop steps do**
        Compute task-specific loss:
  1.5em$\mathcal{L}_{\text{task}}(\mathcal{D}_i^{\text{support}}; \theta_i')$
        Update task-specific parameters:
  1.5em$\theta_i' \leftarrow \theta_i' - \alpha \nabla_{\theta_i'} \mathcal{L}_{\text{task}}(\mathcal{D}_i^{\text{support}}; \theta_i')$
      end for
      Compute query loss: $\mathcal{L}_{\text{task}}(\mathcal{D}_i^{\text{query}}; \theta_i')$
    end for
    Accumulate meta-gradient: $\nabla_\theta \sum_i \mathcal{L}_{\text{task}}(\mathcal{D}_i^{\text{query}}; \theta_i')$
    Update meta-parameters: $\theta \leftarrow \theta - \beta \nabla_\theta \sum_i \mathcal{L}_{\text{task}}(\mathcal{D}_i^{\text{query}}; \theta_i')$
  end while
  return θ

Algorithm 1 describes the steps in the model's training procedure in pseudo-code and is closely inspired by Finn et al. (2017). However, several differences with the original model exist. The primary difference is that tasks are constructed from structured traffic datasets, divided into support and query sets with sequential adaptation. Hence, each task represent a city and a (random) set of n detectors. This means that for each task $\mathcal{T}_i$ in a batch of tasks, $K \cdot N$ observations are sampled, as described in the algorithm 1. This modification addresses the type of data scarcity typically faced when working with MFD estimation.

Meta-testing Procedure: The Meta-testing phase imitates the situation in which we use the MAML model pre-trained on multiple cities to estimate the MFD for a city where only a limited number n of detectors are equipped with sensors. After the model has been trained, it is tested on three random cities, and the statistics, such as the RMSE and the correlation coefficient, are stored for later evaluation.

The meta-testing procedure mimics the inner loop in that the trained MAML model takes the same number of n gradient steps as the number of inner steps and uses the same number, K, of observations from the support test set for each of these in the inner loop.

4 Results

MFD Estimation

First, we tested the bi-parabolic hybrid model and the Multi-Task Physics-Informed model assuming that all data is available. Results for two cities are shown in Figure 3. In short, the results show that the MTPINN model provides better fits to the data. This is shown by a lower value of the Loss function, which suggests a better fit of the data. On the other hand, the resulting function may be challenging to interpret. If the regularization is reinforced, the model would return a bi-parabolic shape, collapsing on the same results as the bi-parabolic hybrid model. In summary, the results show that backward propagation is a powerful tool to estimate all parameters (CD, congested and uncongested branches of the MFD) and that ML models can better fit the data at the cost of having a nonparametric function that may be more difficult to interpret.

Using MAML to cope with data scarcity

This subsection shows the results for the Meta-Model. We built several dataset for training, each time leaving three random cities out of the 21 available for meta-testing. The results are shown in Figure 2. Results clearly show that the error (RMSE) obtained using the MAML is systematically lower than the one using MTPINN, for the same set of detectors. The correlation coefficients, showing the correlation between estimated and real average occupancy and flow values, are also higher. 'Real' in this case refers to the unknown MFD shape, assuming that all detectors are available. This suggests that MAML successfully facilitates the learning of general MFD properties across the cities, as the pre-trained MAML model provides overall better predictions for unseen cities with limited detector data.

5 Conclusions

This research develops a new Meta Machine Learning Model based on MAML designed to estimate the MFD parameters. The model inputs data from multiple cities for training. This allows the model to learn general MFD properties and better generalize. Then, the pre-trained model can be used to estimate the MFD on a new city where only limited sensors are equipped with loop detectors. Results show the potential of this new approach in mitigating potential biases arising from the availability of limited detectors.

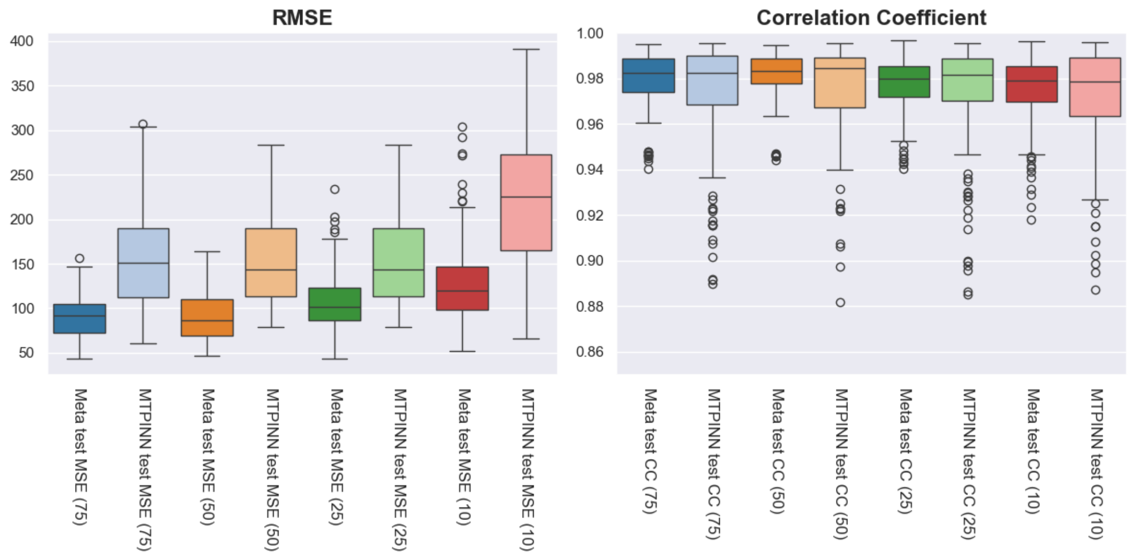


Figure 2: Performance metrics for the meta-model trained and tested on different types of datasets and the MTPINN comparison model.

References

Aghamohammadi, R., & Laval, J. A. (2022). Parameter estimation of the macroscopic fundamental diagram: A maximum likelihood approach. *Transportation Research Part C: Emerging Technologies*, *140*, 103678.

Batista, S., Tilg, G., & Menéndez, M. (2022). Exploring the potential of aggregated traffic models for estimating network-wide emissions. *Transportation Research Part D: Transport and Environment*, *109*, 103354.

Daganzo, C. F. (2007). Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, *41*(1), 49–62.

Daganzo, C. F., & Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological*, *42*(9), 771–781.

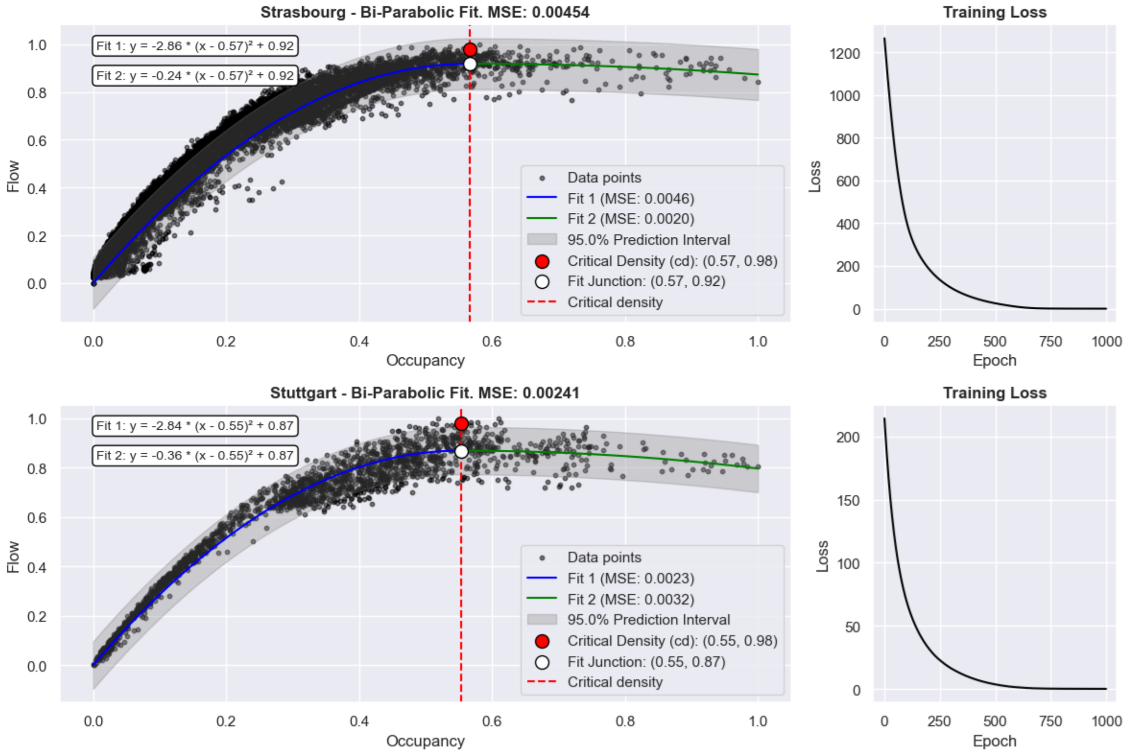
Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning* (pp. 1126–1135).

Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, *42*(9), 759–770.

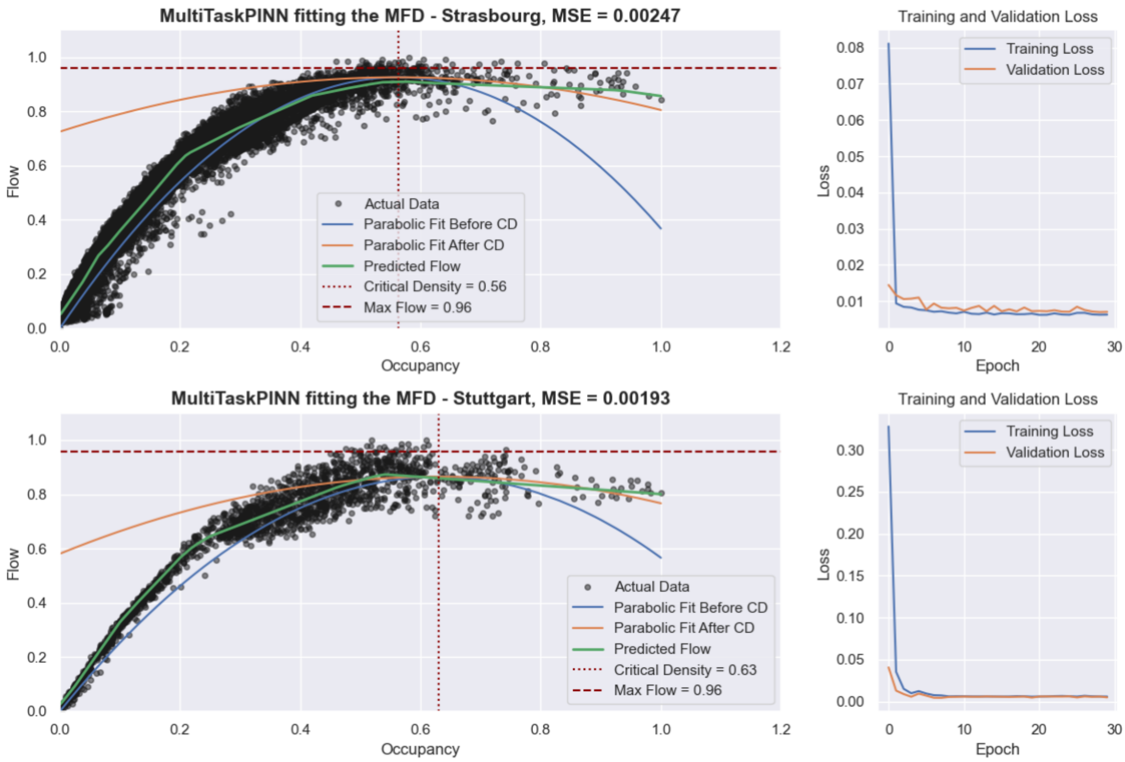
Laval, J. A., & Castrillón, F. (2015). Stochastic approximations for the macroscopic fundamental diagram of urban networks. *Transportation Research Procedia*, *7*, 615–630.

Loder, A., Axhausen, K. W., Ambühl, L., & Menendez, M. (2020). *Utd19: Urban traffic data from 40 cities.* (Unpublished)

Yuan, Y., Zhang, Z., Yang, X. T., & Zhe, S. (2021). Macroscopic traffic flow modeling with physics regularized gaussian process: A new insight into machine learning applications in transportation. *Transportation Research Part B: Methodological*, *146*, 88–110.



(a)



(b)

Figure 3: The figure shows the results of the (a) bi-parabolic hybrid model and for the (b) MTPINN for two different cities - Strasbourg, Stuttgart. The shaded area shows the 95% prediction interval.