

CUQDS: Conformal Uncertainty Quantification Under Distribution Shift for Trajectory Prediction

Huiqun Huang¹, Sihong He², Fei Miao^{1*}

¹University of Connecticut

²University of Texas at Arlington

{huiqun.huang, fei.miao}@uconn.edu, sihong.he@uta.edu

Abstract

Trajectory prediction models that can infer both future trajectories and their associated uncertainties of the target vehicles are crucial for safe and robust navigation and path planning of autonomous vehicles. However, the majority of existing trajectory prediction models have neither considered reducing the uncertainty as one objective during the training stage nor provided reliable uncertainty quantification during inference stage, especially under potential distribution shift. Therefore, in this paper, we propose the **Conformal Uncertainty Quantification under Distribution Shift** framework, CUQDS, to quantify the uncertainty of the predicted trajectories of existing trajectory prediction models under potential data distribution shift, while improving the prediction accuracy of the models and reducing the estimated uncertainty during the training stage. Specifically, CUQDS includes 1) a learning-based Gaussian process regression module that models the output distribution of the base model (any existing trajectory prediction neural networks) and reduces the estimated uncertainty by an additional loss term, and 2) a statistical-based Conformal P control module to calibrate the estimated uncertainty from the Gaussian process regression module in an online setting under potential distribution shift between training and testing data. Experimental results on various state-of-the-art methods using benchmark motion forecasting datasets demonstrate the effectiveness of our proposed design.

Introduction

Accurately and efficiently predicting the future trajectories of the target vehicles is pivotal for ensuring the safety of path planning and navigation of autonomous systems in dynamic environments (Hu et al. 2023; Kedia, Zhou, and Karumanchi 2023). However, trajectory uncertainty due to the changing external surrounding environments (e.g., road networks, neighborhood vehicles, passengers, and so on) or intrinsic intention changes of drivers, can lead to both trajectory *distribution shift* that change over time and *overconfidence* in the output of trajectory prediction models. It remains challenging to predict the distribution of the future trajectories of the target vehicles (rather than focusing solely on point estimates), and to quantify the predicted trajectory uncertainty under potential distribution shift.

*Corresponding author

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Various uncertainty quantification methods have been proposed for trajectory prediction. One popular approach is to estimate the distribution (e.g., Laplace distribution or Gaussian distribution) of future trajectories of target vehicles by direct modeling methods (Zhou et al. 2022; Mao et al. 2023; Zhu et al. 2023; Salzmann et al. 2020). However, this method often overlooks the impact of *model limitation* (the restricted ability of models to represent the real-time trajectories data) and the consequence overconfident uncertainty estimation in the inference stage. Another popular approach is to calibrate the preliminary estimated uncertainty in the inference stage by the statistical-based methods. Specifically, methods such as split conformal prediction (CP) (Shafer and Vovk 2008; Lindemann et al. 2023) provide confidence intervals that guarantee to contain the ground truth target data with a predefined probability. However, they are not applicable to the situations under distribution shift.

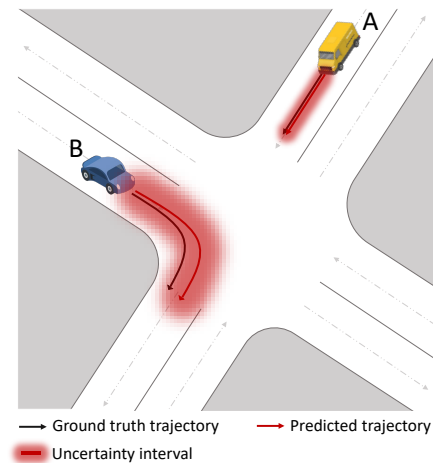


Figure 1: Illustration of the importance of a good uncertainty interval estimation for the predicted trajectories of the target vehicles.

In this paper, we propose CUQDS, i.e., **Conformal Uncertainty Quantification under Distribution Shift** framework, to quantify the output uncertainty of existing trajectory prediction models under distribution shift, while improving the prediction accuracy of the models and reducing the estimated uncertainty. The proposed CUQDS frame-

work integrates a learning-based Gaussian process regression (Rasmussen, Williams et al. 2006) module with a statistics-based conformal P control module (Angelopoulos, Candes, and Tibshirani 2023). Specifically, in the training stage, CUQDS adopts the Gaussian process regression module to estimate the output distribution of base model. The variance of the output distribution quantifies the uncertainty of the predicted trajectories. We further use the output distribution to construct uncertainty interval that guarantees to cover the true target trajectory with a predefined probability in the long-run. As shown in Fig. 1, a good uncertainty interval should tend to be narrow when the predicted trajectory closely aligns with the ground truth trajectory (e.g., vehicle A), whereas it should be widen when the output trajectory deviates (e.g., vehicle B). To achieve this, we introduce additional loss term to reduce the estimated uncertainty in the training stage, ensuring narrow uncertainty interval while covering the true target trajectory with a predefined probability. In the inference stage, CUQDS calibrates the estimated output uncertainty upon potential distribution shift by a statistics-based conformal P control module. Different from standard conformal prediction method and its variants who use fixed conformal quantile in the inference stage and violate the distribution shift assumption, we first initialize the conformal quantile using the validation data and update it after every prediction step during the inference stage.

The key contributions of this work are summarized as follows:

1. We propose CUQDS framework that integrates both the learning-based and statistical-based modules to provide uncertainty quantification for trajectory prediction under distribution shift. The proposed learning-based module is trained alongside the base model to quantify its output uncertainty. In the training stage, the main objectives are to enhance prediction accuracy of base model and reduce the estimated uncertainty by incorporating an additional loss term.
2. We introduce the statistical-based conformal P control module to calibrate the estimated output uncertainty from the learning-based module under potential distribution shift during the inference stage. To alleviate the impacts of data distribution shift on uncertainty estimation, the conformal quantile is first initialized using the validation data and keep updating after each prediction step in the inference stage.
3. We validate the effectiveness of our proposed framework on the Argoverse 1 motion forecasting dataset (Chang et al. 2019) and five state-of-the-art baselines. Compared to base models without our CUQDS, the experiment results show that our approach improves prediction accuracy by an average of 7.07% and reduce uncertainty of predicted trajectory by an average of 25.41%.

Related Work

Modeling the distribution of the trajectory of the target vehicles, instead of solely focusing on the point estimation of the future trajectory, proves to be an efficient approach (Chai et al. 2019; Deo and Trivedi 2018; Phan-Minh

et al. 2020; Zeng et al. 2021) to avoid missing potential behavior in trajectory prediction methods. Early works sample multiple potential future trajectories (Liang et al. 2020; Ngiam et al. 2021; Gupta et al. 2018; Rhinehart, Kitani, and Vernaza 2018; Rhinehart et al. 2019) to approximate the predicted trajectory distribution. However, these approaches still rely on limited point estimation and suffer from overconfidence upon trajectory prediction in real-world settings. To mitigate these issues, recent works directly estimate the distribution of future trajectories by fitting the (mixture) distribution models based on the embedded features of input trajectories (Zhou et al. 2022; Varadarajan et al. 2022; Cui et al. 2019; Shi et al. 2023). However, these works mostly train the (mixture) distribution models by the MLP-based predictor in the last layer of the model. Such designs fail to estimate an accurate enough distribution of future trajectories under data distribution shift.

Uncertainty quantification in trajectory prediction is challenging and usually solved by two categories of methods: direct modeling and statistical-based methods. Specifically, direct modeling assumes the target data follows specific distribution, designs the learning-based neural networks for distribution estimation, and introduces the corresponding loss function to model the uncertainty directly. To achieve this, existing works usually assume the data follows Gaussian distribution (Mao et al. 2023) or Laplace distribution (Gu et al. 2024) and model the corresponding distribution by learning-based designs. However, estimating rigorous uncertainty by direct modeling can be challenging, as the model may easily overfit the training dataset such that invalid in the inference stage under data distribution shift. **Conformal prediction** (Shafer and Vovk 2008; Angelopoulos, Candes, and Tibshirani 2023) stands out as a method in statistical-based inference that has proven effective in constructing predictive sets, ensuring a certain probability of covering the true target values (Ivanovic et al. 2022; Xu et al. 2014; Sun et al. 2024; Xu et al. 2023; Stankeviciute, M Alaa, and van der Schaar 2021). However, the assumption of identical distribution across all datasets invalidates the standard conformal prediction under distribution shift. To address this, **adaptive conformal prediction** methods (Gibbs and Candes 2021; Stanton, Maddox, and Wilson 2023) have been proposed to achieve the desired coverage frequency under distribution shift by leveraging the distance between new queries and the training dataset. However, these approaches either lack base model uncertainty estimation or fail to incorporate mechanisms for reducing uncertainty during the training process, as achieved by our proposed CUQDS.

In this study, we propose the CUQDS framework to provide output distribution for the predicted trajectories of base model under distribution shift, while improving the prediction accuracy of base model and reducing the estimated uncertainty by introducing additional loss term. In particular, CUQDS adopts a Gaussian process regression module to estimate the output uncertainty of base model, then utilizes a statistical-based conformal P control module to calibrate this output uncertainty by taking into account the model performance on recent trajectory data. Moreover, different from

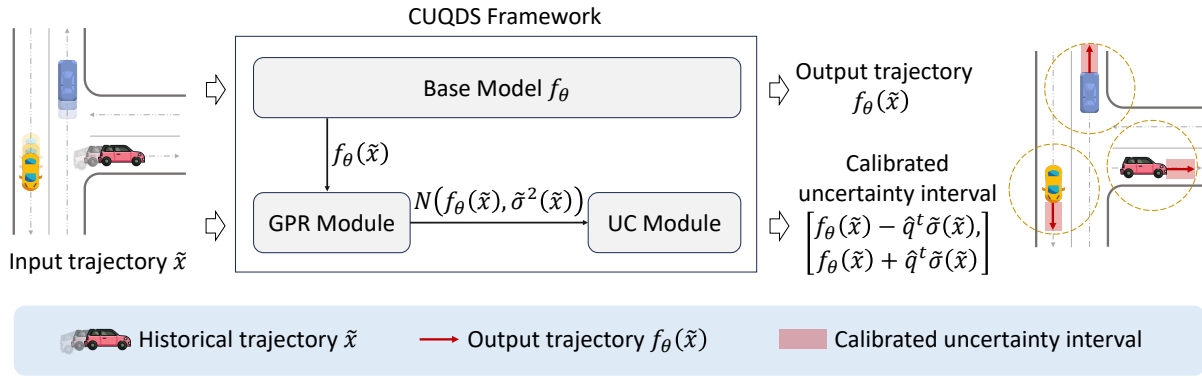


Figure 2: Our CUQDS models the conditional output distribution $\tilde{Y}|\tilde{x}, \mathcal{D} \sim \mathcal{N}(f_\theta(\tilde{x}), \tilde{\sigma}^2(\tilde{x}))$ of base model f_θ by the Gaussian process regression (GPR) module, and provides the correspond calibrated uncertainty interval $[f_\theta(\tilde{x}) - \hat{q}^t \tilde{\sigma}(\tilde{x}), f_\theta(\tilde{x}) + \hat{q}^t \tilde{\sigma}(\tilde{x})]$ for the predicted trajectories by the uncertainty calibration (UC) module.

standard conformal prediction methods who use the fixed conformal quantile in the inference stage, we first initialize the conformal quantile using the validation data and update it after every prediction in the inference stage to alleviate the data distribution shift problem.

Methodology

Problem Formulation

Suppose we have a training dataset $\mathcal{D}_1 = \{(x_i, y_i)\}_{i=1}^{N_1} = \{(x^t, y^t)\}_{t=L-1}^{T_1}$, a validation dataset $\mathcal{D}_2 = \{(x_i, y_i)\}_{i=1}^{N_2} = \{(x^t, y^t)\}_{t=L-1}^{T_2}$, and a testing dataset $\tilde{\mathcal{D}} = \{(\tilde{x}_i, \tilde{y}_i)\}_{i=1}^{N_3} = \{(\tilde{x}^t, \tilde{y}^t)\}_{t=L-1}^{T_3}$. N_1 , N_2 , and N_3 are the number of data samples, and T_1 , T_2 , and T_3 are the time periods of each dataset. $T_1 \cap T_2 \cap T_3 = \emptyset$. In this paper, we assume that training/validation datasets are drawn from the same distribution, while the distribution of testing dataset shifts over time. As shown above, we use two ways to represent each dataset, where (x_i, y_i) (or $(\tilde{x}_i, \tilde{y}_i)$) denotes the random input-output trajectory pair within each dataset indexed by i , and (x^t, y^t) (or $(\tilde{x}^t, \tilde{y}^t)$) is time series data and represents the input-output trajectory pair of the current time step t . Both x_i and x^t denote the input historical trajectory during the past L historical time steps and are sampled from domain $\mathcal{X} \in \mathbb{R}^{L \times D}$, and y_i and y^t denote the corresponding target trajectory during the J following time steps from domain $\mathcal{Y} \in \mathbb{R}^{J \times D}$. D is the dimension of target features. We will omit the index i or t when there is no conflict.

We assume that we have a trajectory prediction model f with parameters θ . We call this model f_θ a base model and it can be implemented as different structures of deep neural network (Zhou et al. 2022; Liu et al. 2021; Zhou et al. 2023; Liang et al. 2020; Zhong et al. 2022). The model f_θ is trained using the dataset \mathcal{D} , where the target trajectory, denoted as Y , is conditioned on input x and the dataset \mathcal{D} . We model the conditional distribution of Y given x and \mathcal{D} as a Gaussian distribution, where $Y|x, \mathcal{D} \sim \mathcal{N}(\mu(x), \sigma^2(x))$. In this formulation, $\mu(x)$ and $\sigma^2(x)$ are functions that map the input x to the mean and variance of the Gaussian distribution, respectively. We treat the predicted trajectory of

the base model f_θ as the mean $\mu(x)$ directly. Our main task in this study is to estimate the variance $\sigma^2(x)$ to denote the output uncertainty.

We propose the Conformal Uncertainty Quantification framework, CUQDS, as illustrated in Fig. 2, to quantify the uncertainty of the predicted trajectory of base model under potential distribution shift. The main objectives of this framework are to improve prediction accuracy of base model and reduce the output uncertainty. The framework introduces two modules, 1) In particular, in the training and validation stages (Alg. 1), we propose a learning-based Gaussian process regression module to approximate the conditional output distribution ($Y|x, \mathcal{D} \sim \mathcal{N}(\hat{\mu}(x), \hat{\sigma}^2(x))$) based on f_θ . 2) In the testing or inference stage (Algorithm. 2), we propose a statistical-based conformal P control module to calibrate the output distribution $\tilde{Y}|\tilde{x}, \theta \sim \mathcal{N}(\tilde{\mu}(\tilde{x}), \tilde{\sigma}^2(\tilde{x}))$ by considering the potential distribution shift between the training and testing datasets and the performance limitation of f_θ on current time series trajectory input. The conformal quantile of the module is first initialized using the validation data and keep updating after each prediction during the inference stage. We build the calibrated uncertainty interval and ensure it covers the true target trajectory with a predefined probability in long-run.

To summarize, during training and validation stage, our goal is to find the parameters $[\theta, \omega]$ such that minimizing the loss function \mathcal{L} on training data:

$$[\theta, \omega] = \underset{\theta, \omega}{\operatorname{argmin}} \mathcal{L}(\theta, \omega | \mathcal{D}). \quad (1)$$

The loss \mathcal{L} is a weighted combination of base model's loss $\mathcal{L}_1(\theta)$ and Gaussian processing regression model's loss $\mathcal{L}_2(\omega, \theta)$.

$$\mathcal{L} = w_1 \mathcal{L}_1(\theta) + w_2 \mathcal{L}_2(\omega, \theta), \quad (2)$$

where $w_1 \in \mathbb{R}$ and $w_2 \in \mathbb{R}$ are the weights adjusting the influence of two loss terms, respectively. During inference stage, we will calibrate the estimated uncertainty of the predicted trajectory.

Preliminary

• **Gaussian Process Regression** often acts as the surrogate model to estimate the output distribution of existing models (Erlygin et al. 2023). It introduces additional loss term to guild the learning process of existing models and reduce the output uncertainty. Gaussian process regression assumes any combinations of data samples follows different join distribution and thus good at capturing the nonlinear relationship among time series data samples.

Given an existing deep neural network model f_θ , Gaussian process regression views this model as a black box and aims to estimate the output distribution ($Y|x, \theta \sim \mathcal{N}(\hat{\mu}(x), \hat{\sigma}^2(x))$) corresponding to the inputs x . We treat the output of model $f_\theta(\cdot)$ on x as the mean value $\hat{\mu}(x)$ of the estimated output distribution and the correspond variance $\hat{\sigma}^2(x)$ as the uncertainty of $f_\theta(x)$. Gaussian process regression also assumes that the observed values y differ from the function values $f_\theta(x)$ by additive noise ϵ , and further assumes that this noise follows an independent, identically distributed Gaussian distribution with zero mean and variance of σ_ϵ^2 . Now the prediction process is formulated as $Y|x = \hat{\mu}(x) + \epsilon$, where (x, y) is any data samples from \mathcal{D} . Then we have:

$$Y|x \sim N\left(\hat{\mu}(x), \frac{\sum_{i=1}^{N_1} k(x, x_i)}{N_1} + \sigma_\epsilon^2 I\right), \quad (3)$$

where $k(\cdot, \cdot)$ is the user-defined kernel function or the covariance function. For example, $k(x, x')$ quantifies the similarity between trajectory input data x and x' .

• **Conformal Prediction** (Shafer and Vovk 2008; Angelopoulos, Candes, and Tibshirani 2023) is a statistical-based method to construct predictive sets for any model, ensuring a certain probability of covering the true target values. It assumes identical distribution across all datasets. Given the base model f_θ whose output distribution follows ($Y|x, \theta \sim \mathcal{N}(\hat{\mu}(x), \hat{\sigma}^2(x))$), conformal prediction converts the heuristic notion of uncertainty (e.g. the estimated standard deviation $\hat{\sigma}(x)$) into rigorous prediction intervals of the form $(\hat{\mu}(x) \pm q\hat{\sigma}(x))$, where q is the estimated conformal quantile from conformal prediction using validation dataset.

Given the validation dataset \mathcal{D}_2 which has N_2 data samples, conformal prediction follows the following steps: 1) Define a conformal score function $s(x, y) \in \mathbb{R}$ (smaller scores encode better agreement between x and y). 2) Compute the conformal quantile q as the $\frac{\lceil (N_2+1)(1-\alpha) \rceil}{N_2}$ quantile of the validation scores $S = \{s(x_i, y_i)\}_{i=1}^{N_2}$. $\alpha \in [0, 1]$ is the user-defined error rate. 3) Then for every unseen data sample (\tilde{x}, \tilde{y}) , conformal prediction forms the prediction set $\mathcal{C}(\tilde{y}|s(\tilde{x}, \tilde{y}) \leq q)$. Conformal prediction guarantees that the prediction set contains the true target value with a probability at least $1 - \alpha$.

Gaussian Process Regression Module

Existing literature of trajectory prediction models (Zhou et al. 2022, 2023; Liang et al. 2020; Zhong et al. 2022) focuses on providing point estimates of future trajectories.

Algorithm 1: Training & validation stage

- 1 **Input:** a base model $f(\cdot)$ with an initialized parameter θ and a covariance function $k(\cdot)$ with an initialized parameter ω , error rate $\alpha \in [0, 1]$, score function $s(\cdot)$, total number of training epochs epo .
 - 2 **Data:** training dataset $D_1 = \{x_i, y_i\}_{i=1}^{N_1}$, validation dataset $D_2 = \{x^t, y^t\}_{t=1}^{T_2}$.
 - 3 **Output:** well-trained base model $f_\theta(\cdot)$, updated statistic \hat{q}^1 , score set $S = \{s(x^t, y^t)\}_{t=L-1}^{T_2}$ and error set $E = \{e^t\}_{t=L-1}^{T_2}$ estimated from the validation data.
 - 4 Initialization: $\hat{q}^0 = 1$.
 - 5 **for** $epoch$ in epo **do**
 - 6 Training $\text{CUQDS}_{\theta, \omega}$ with loss \mathcal{L} and training dataset D_1 .
 - 7 Update $\{\theta, \omega\} \leftarrow \underset{\theta, \omega}{\text{argmin}} \mathcal{L}(\theta, \omega|D)$
 - 8 Initialize score set $S = \{\}$, error set $E = \{\}$.
 - 9 **for** $t = 1 : T_2$ **do**
 - 10 $f_\theta(x^t), \hat{\sigma}(x^t) \leftarrow \text{CUQDS}_{\theta, \omega}(x^t)$
 - 11 Compute score $s^t = s(x^t, y^t)$.
 - 12 Compute conformal prediction set \mathcal{C}^t .
 - 13 Compute $e^t = \mathbf{1}_{\tilde{y}^t \notin \mathcal{C}^t}$.
 - 14 $S \leftarrow s^t, E \leftarrow e^t$.
 - 15 $\eta = \beta \max(S)$, then $\hat{q}^t = \hat{q}^{t-1} + \eta(\bar{E} - \alpha)$.
 - 16 Update $\hat{q}^1 \leftarrow \hat{q}^{T_2}$.
-

However, the uncertainty of the future trajectories due to the changing environment or the intrinsic intention changes of drivers can lead to significant distribution shift and overconfident trajectory prediction. Such distribution shift and overconfident prediction can greatly impact the subsequent decision-making processes, such as robust path planning (Hu et al. 2023; Kedia, Zhou, and Karumanchi 2023). Hence, we propose to consider output uncertainties from both the modeling limitation of the model f_θ upon current trajectory inputs and the noise inherent in the trajectory data.

In this study, we introduce to utilize the Gaussian process regression method as introduced in preliminary to estimate the preliminary output distribution of existing trajectory prediction models. In our case, the base model f_θ can be any existing deep neural network structure in the literature, examples include (Zhou et al. 2022; Liu et al. 2021; Zhou et al. 2023; Liang et al. 2020; Zhong et al. 2022)). Moreover, in real-world settings of autonomous vehicles, the collected historical trajectories of the target vehicles inevitably contain noisy data during perception and object tracking steps. Such noise usually hardly captured by the trajectory prediction models such that greatly impacts the output trajectory and the corresponding uncertainty. To mitigate this problem, we model such noisy impact as $\epsilon \in \mathbb{R}^D$ as in the standard Gaussian process regression method.

Learnable Kernel Function In this paper, we define the covariance function as the radial basis function

$$k(x, x') = l_1^2 \exp\left(-\frac{(x - x')^2}{2l_2^2}\right), \quad (4)$$

where $l_1 \in \mathbb{R}$ and $l_2 \in \mathbb{R}$ are the parameters to be trained. The closer x and x' are, the higher the value of $k(x, x')$, reaching its maximum value of l_1^2 .

Eigenvector Inducing Variables As shown in Eq. 3, the covariance $\hat{\sigma}^2(x)$ of standard Gaussian process regression method is typically optimized based on the whole training data. However, this can be computationally expensive when the number of training samples (N_1) is large, given that the entire Gaussian process regression model requires $O(N_1^3)$ computational complexity and $O(N_1^2)$ memory complexity. To alleviate such complexity while maintaining the effectiveness of the Gaussian process regression module, we propose a modification to approximate the covariance $\hat{\sigma}^2(x)$ in Eq. 3 by extracting $M \in \mathbb{R}$, $M \ll N_1$, inducing variables $\{v_i\}_{i=1}^M$, $v_i \in \mathbb{R}^{L \times D}$ by Principal Component Analysis (PCA). These M inducing variables summarize the key information of training dataset. More specifically, we first standardize each input trajectory data sample $x_i \in \mathbb{R}^{L \times D}$ in training dataset \mathcal{D}_1 by its own mean and standardization along time dimension. Then we build the $N_1 \times N_1$ covariance matrix by $\text{cov}(x, x') = \mathbf{E}[(x - \mathbf{E}(x))(x' - \mathbf{E}(x'))]$. The further steps follow the standard PCA processes.

Inference under Distribution Shift Through the Gaussian process regression with a learnable kernel function, we are able to estimate the target trajectory's distribution that in the training/validation dataset. However, under potential distribution shift between training and testing data in real-world settings, solely keep using the information learn from the training/validation datasets will lead to high generalization error. To take into account the potential distribution shift between the dataset $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2\}$ and the testing dataset $\tilde{\mathcal{D}}$, we instead estimating the conditional output distribution of base model in the testing stage by $\tilde{Y}|\tilde{x}, \mathcal{D} \sim \mathcal{N}(\tilde{\mu}(\tilde{x}), \tilde{\sigma}(\tilde{x}))$, where $(\tilde{x}, \tilde{y}) \in \tilde{\mathcal{D}}$. Then we compute the variance of the conditional output distribution of base model in the testing stage by

$$\tilde{\sigma}^2(\tilde{x}) = k(\tilde{x}, \tilde{x}) - K_{\tilde{x}M} [K_{MM} + \sigma_\epsilon^2 I]^{-1} (K_{\tilde{x}M})^\top, \quad (5)$$

where $K_{\tilde{x}M} = (k(\tilde{x}, v_1), \dots, k(\tilde{x}, v_M))^\top$ and $[K_{MM}]_{ij} = k(v_i, v_j)$. We then construct the uncertainty interval as $[f_\theta(\tilde{x}) - \tilde{\sigma}(\tilde{x}), f_\theta(\tilde{x}) + \tilde{\sigma}(\tilde{x})]$, indicating there is a high probability that it will cover the true target trajectories.

To find the optimal parameters for the covariance function and the noise ϵ , and reduce the estimated uncertainty, we introduce a new loss term \mathcal{L}_2 :

$$\mathcal{L}_2(\omega, \theta) = \frac{1}{N_1} \sum_{i=1}^{N_1} \left(-\frac{1}{2} e_i^\top [\bar{K}_{x_i M} + \sigma_\epsilon^2 I]^{-1} e_i \right) \quad (6)$$

$$-\frac{1}{2} \log |\bar{K}_{x_i M} + \sigma_\epsilon^2 I| + \frac{1}{2N_1} \log 2\pi, \quad (7)$$

where $\omega = [l_1, l_2, \sigma_\epsilon]$ contains all trainable parameters in the Gaussian process regression module, $\bar{K}_{x_i M} = \frac{1}{M} \sum_{j=1}^M k(x_i, v_j)$, and $e_i = y_i - f_\theta(x_i)$.

Uncertainty Quantification under Distribution Shift through Calibration

The learning-based Gaussian process regression module is prone to overfit the training data and provide overconfident uncertainty estimation under distribution shift. To solve this, we propose a statistical-based conformal P control module to calibrate the output uncertainty from the Gaussian process regression module by considering the performance of base model under distribution shift. The idea of this method is based on both the conformal prediction (Shafer and Vovk 2008) and the P control in (Angelopoulos, Candes, and Tibshirani 2023). However, different from standard conformal prediction, we assume the training/validation data and testing data follow different distributions, and the distribution of testing data shifts over time. Instead of using the fixed conformal quantile q as in standard conformal prediction, our proposed module updates the conformal quantile q^t in time step t based on model performance under distribution shift.

In the inference stage, we aim to achieve a long-run average coverage rate in time, ensuring that the calibrated uncertainty interval covers the true target trajectories with an average probability of $1 - \alpha$, namely $\sum_{t=1}^T \text{err}^t / T \rightarrow \alpha$ as $T \rightarrow \infty$. $\text{err}^t = \mathbf{1}(y^t \notin \mathcal{C}^t)$ and $\alpha \in (0, 1)$ is a predefined target error rate threshold.

Uncertainty Calibration in CUQDS Framework To update the conformal quantile q^t under distribution shift, we first initialize the conformal quantile using the validation data as shown in Alg. 1. For each data sample (x^t, y^t) in validation data \mathcal{D}_2 , we first compute the corresponding conformal score $s(x^t, y^t)$ and the conformal prediction set \mathcal{C}^t as standard conformal prediction. We define the score function $s(x, y)$ to evaluate the model performance for each data sample

$$s(x, y) = \frac{|y - f_\theta(x)|}{\sigma(x)}, \quad (8)$$

where y is the ground truth, $f_\theta(x)$ and $\sigma(x)$ are respectively the mean and std of the output distribution of the base model.

Moreover, we introduce an additional error set E to record the coverage errors of uncertainty intervals. The error e^t is set to 1 if y^t is outside of \mathcal{C}^t , indicating a prediction error, and 0 otherwise. The estimated conformal score $s(x^t, y^t)$ is added to the score set S , and the error e^t is added to the error set E . The conformal quantile \hat{q}^t is then updated based on both the existing conformal scores and coverage errors. The updating rule involves a learning rate $\eta = \beta \max(S)$ and adjusts \hat{q}^t according to the formula

$$\hat{q}^t = \hat{q}^{t-1} + \eta(\bar{E} - \alpha),$$

where \bar{E} is the average coverage error in the coverage error set and α is the predefined error rate. This step allows the model to adapt its uncertainty estimation based on the observed performance during validation.

During the inference stage, the conformal P control module inherits the conformal quantile, the score set S , and the error set E from the final validation iteration. The conformal quantile updating steps in the inference stage as shown in Alg. 2 are similar as in the validation stage. For each time

Algorithm 2: Testing stage

- 1 **Input:** the well-trained framework CUQDS $_{\theta, \omega}$, the conformal quantile \hat{q}^1 , the error set E , the conformal score set S , the error rate α from the training stage, score function $s(\cdot)$.
 - 2 **Data:** testing data $\tilde{D} = \{(\tilde{x}^t, \tilde{y}^t)\}_{t=L-1}^{T_3}$.
 - 3 **Output:** predicted trajectory and the correspond calibrated uncertainty interval.
 - 4 **for** $t = 1 : T_3$ **do**
 - 5 $f_{\theta}(\tilde{x}^t), \hat{\sigma}(\tilde{x}^t) \leftarrow \text{CUQDS}_{\theta, \omega}(\tilde{x}^t)$
 - 6 Calibrate std $\tilde{\sigma}(\tilde{x}^t)$ by $\hat{q}^{t-1} \tilde{\sigma}(\tilde{x}^t)$.
 - 7 Compute score $s(\tilde{x}^t, \tilde{y}^t)$ by Eq. 8.
 - 8 Compute conformal prediction set C^t by Eq. 9.
 - 9 Compute $e^t = \mathbf{1}_{\tilde{y}^t \notin C^t}$.
 - 10 $S \leftarrow s^t, E \leftarrow e^t$.
 - 11 $\eta = \beta \max(S)$, then $\hat{q}^t = \hat{q}^{t-1} + \eta(\bar{E} - \alpha)$.
-

step t in the inference stage, we calibrate the output uncertainty and construct the uncertainty interval by using the updated conformal quantile \hat{q}^t ,

$$C^t = [f_{\theta}(\tilde{x}^t) - \hat{q}^t \tilde{\sigma}(\tilde{x}^t), f_{\theta}(\tilde{x}^t) + \hat{q}^t \tilde{\sigma}(\tilde{x}^t)]. \quad (9)$$

Experiment

Experimental Setups

Dataset & Key Setups: We use the Argoverse 1 motion forecasting dataset (Chang et al. 2019) to verify the efficacy of our approach. This dataset collects trajectory data from Miami and Pittsburgh, with a sample rate of 10 Hz. Given that the ground truth future trajectories are not provided in this official test sequences but are essential to our CUQDS in the testing stage to update the conformal quantile \hat{q}^t , we repartition the sequences. In particular, we split the 205,942 official training sequences into 166,470 training data and 39,472 validation data, and use the official 39,472 validation sequences as testing data in this study.

In this study, $\alpha = 0.1$. We implement existing base models by using their default settings, unless otherwise specified. The host machine is a server with IntelCore i9-10900X processors and four NVIDIA Quadro RTX 6000 GPUs.

Prediction Accuracy Evaluation Metrics: Following the standard evaluation protocol, we utilize metrics including minimum Average Displacement Error ($\min \text{ADE}_k$), minimum Final Displacement Error ($\min \text{FDE}_k$), and Miss Rate (MR_k) to evaluate the prediction accuracy of the model. As common, k is selected as 1 and 6.

Uncertainty Evaluation Metrics: To further verify the efficacy of our CUQDS in reducing the predicted uncertainty to provide narrow while accurate uncertainty interval, we adopt the Negative Log-Likelihood (NLL) (Feng et al. 2021) to assess the level of uncertainty in the predicted distribution. Lower values connote a higher degree of precision in uncertainty estimation and narrower uncertainty interval.

Main Results

We adopt HiVT (Liang et al. 2020), LaneGCN (Liang et al. 2020), LBA (Zhong et al. 2022), LBF (Zhong et al. 2022), and Trajectron++ (Salzmann et al. 2020) as base models and apply our CUQDS on them to verify the efficacy of CUQDS. To compare our CUQDS with state-of-the-art methods, we also apply SPCI (Xu and Xie 2023) into the above five base models to serve as the alternative of our CUQDS. In specifically, we implement additional module whose structure is the same as the estimator in SPCI to predict the variance of the predicted trajectory and add the KLD (Meyer and Thakurdesai 2020) loss term to reduce the estimated uncertainty. As presented in Table. 1 and Table. 2, our CUQDS improves the prediction accuracy 7.07% on average. Compared with incorporating SPCI in the base models, our CUQDS reduces the output uncertainty which quantified by NLL by 25.41% on average, and improves the average coverage rate of the estimated uncertainty interval by 21.02%. The results indicate that our CUQDS is capable of providing the trust worthy uncertainty quantification for the output trajectory of base model under potential distribution shift. Our estimated uncertainty intervals are narrow and achieve high coverage rate of covering the true target trajectories.

Compared with the transformer-based model HiVT, applying our CUQDS in HiVT provides the output uncertainty of HiVT by considering the distribution difference between the training and testing data. Compared with the base models of LaneGCN, LBA, LBF, and Trajectron++ who provide a confidence score or the estimated distribution for each output trajectory, our CUQDS provides the output distribution of base model instead of point estimates and calibrates the correspond uncertainty by taking into account the model performance on recent inputs. SPCI try to predict the conformal quantile by the learning-based estimator to provide the prediction interval. However, such estimator is prone to overfit the training data and invalid under distribution shift.

Ablation Study

• **Conformal Prediction VS conformal P control module:** We conduct ablation study on replacing the P control uncertainty calibration module with the standard split conformal prediction (Shafer and Vovk 2008). More specificity, the standard split conformal prediction (CP) estimates the fixed conformal quantile \hat{q} as the $\frac{(1-\alpha)(N_2+1)}{N_2}$ smallest element in the conformal score set $\mathcal{S} = \{s(x_i, y_i)\}_{i=1}^{N_2}$ from validation dataset.

We implement the standard split conformal prediction setting in base model LaneGCN. The average coverage rates of the uncertainty intervals estimated by our P control module and the split conformal prediction are 0.746 and 0.402, respectively. Our P control module outperforms the standard split conformal prediction. This helps to prove the effectiveness of our uncertainty calibration module in adapting to potential distribution shift.

• **With VS without the P control uncertainty calibration module:** We further verify the efficacy of the P control uncertainty calibration module by removing the whole calibration process in the testing stage. As shown in Table. 3,

Scheme	minADE ₆ ↓	minFDE ₆ ↓	MR ₆ ↓	minADE ₁ ↓	minFDE ₁ ↓	MR ₁ ↓
HiVT	0.692	1.043	0.106	1.291	2.895	0.499
HiVT+SPCI	0.694	1.047	0.107	1.302	2.921	0.501
HiVT+CUQDS	0.682	1.034	0.097	1.218	2.696	0.455
LaneGCN	0.719	1.094	0.104	1.375	3.023	0.505
LaneGCN+SPCI	0.739	1.169	0.114	1.585	3.573	0.568
LaneGCN+CUQDS	0.704	1.033	0.097	1.255	2.744	0.482
LBA	0.717	1.094	0.103	1.395	3.100	0.503
LBA+SPCI	0.719	1.098	0.104	1.402	3.100	0.523
LBA+CUQDS	0.705	1.044	0.092	1.251	2.801	0.483
LBF	0.720	1.098	0.104	1.530	3.467	0.549
LBF+SPCI	0.721	1.099	0.105	1.533	3.469	0.550
LBF+CUQDS	0.714	1.091	0.097	1.455	3.188	0.522
Trajectron++	0.764	1.121	0.123	1.602	3.713	0.589
Trajectron+++SPCI	0.732	1.079	0.119	1.593	3.572	0.571
Trajectron+++CUQDS	0.693	1.101	0.112	1.491	3.404	0.548

Table 1: Prediction results and performance comparison on testing dataset when with and without our CUQDS.

Scheme	+SPCI (NLL)	+CUQDS (NLL)	+SPCI (CR)	+CUQDS (CR)
HiVT	21.354	16.354	0.705	0.832
LaneGCN	25.532	18.896	0.603	0.746
LBA	23.634	17.453	0.602	0.721
LBF	23.723	17.535	0.614	0.716

Table 2: Compare the NLL distance between ground truth and predicted distribution, and the average cover rate (CR) of the estimated uncertainty interval in settings of 1) SPCI+base models, 2) CUQDS+base models.

Scheme	Without UC	With UC
HiVT+CUQDS	0.603	0.832
LaneGCN+CUQDS	0.497	0.746
LBA+CUQDS	0.570	0.721
LBF+CUQDS	0.569	0.716

Table 3: The coverage rates of the uncertainty interval with and without the uncertainty calibration (UC) module.

by calibrating the output uncertainty of the base model by the P control uncertainty calibration module, the coverage rate of the calibrated uncertainty interval improves 35.10% on average comparing with without calibration.

• **Gaussian process regression module VS direct modeling:** To validate the effectiveness of our Gaussian process regression module, we replace the Gaussian process regression module with the self-attention based design (Mao et al. 2023) to predict the variance of the output distribution and add the KLD loss term to reduce the estimated uncertainty. We apply this setting on the base model HiVT. As shown in Table. 4, both of the methods achieve better prediction accuracy than without considering uncertainty in the base model. Our CUQDS slightly exceed the direct modeling method in all prediction accuracy evaluation metrics. The results prove that our CUQDS is sufficient in providing output uncertainty estimation for the trajectory prediction base model and im-

Scheme	minADE ₆ ↓	minFDE ₆ ↓	MR ₆ ↓
HiVT+CUQDS	0.682	1.034	0.097
HiVT+DM	0.694	1.047	0.107

Table 4: Results of our CUQDS and replacing the Gaussian process regression module with the self-attention based design for uncertainty estimation.

proving the prediction accuracy.

Conclusion & Discussion

In this study, we present the framework CUQDS to estimate the output distribution for any existing trajectory prediction models under distribution shift, while improving the prediction accuracy and reducing the output uncertainty. CUQDS adopts the Gaussian process regression module to model the output distribution of the base model. CUQDS then utilizes a statistical-based P control module to calibrate the estimated uncertainty under distribution shift. The experiment results demonstrate the efficacy of CUQDS in improving the prediction accuracy and reducing the prediction uncertainty. Our findings highlight the importance of quantifying the uncertainty of output trajectory in trajectory prediction under distribution shift. In the future work, we plan to extend our work to more state-of-the-art models and provide uncertainty of output trajectory in each time step.

Acknowledgments

The work is supported by the National Science Foundation under Grants CNS-2047354 (CAREER), and the New England University Transportation Center (NEUTC). Funding for the NEUTC Program is provided by the Office of Assistant Secretary for Research and Innovation (OST-R) of the United States Department of Transportation. The recommendations of this study are those of the authors and do not represent the views of NEUTC.

References

- Angelopoulos, A. N.; Candes, E. J.; and Tibshirani, R. J. 2023. Conformal pid control for time series prediction. *arXiv preprint arXiv:2307.16895*.
- Chai, Y.; Sapp, B.; Bansal, M.; and Anguelov, D. 2019. Multipath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction. *arXiv preprint arXiv:1910.05449*.
- Chang, M.-F.; Lambert, J.; Sangkloy, P.; Singh, J.; Bak, S.; Hartnett, A.; Wang, D.; Carr, P.; Lucey, S.; Ramanan, D.; et al. 2019. Argoverse: 3d tracking and forecasting with rich maps. In *Proc. of IEEE/CVF*, 8748–8757.
- Cui, H.; Radosavljevic, V.; Chou, F.-C.; Lin, T.-H.; Nguyen, T.; Huang, T.-K.; Schneider, J.; and Djuric, N. 2019. Multimodal trajectory predictions for autonomous driving using deep convolutional networks. In *2019 ICRA*, 2090–2096. IEEE.
- Deo, N.; and Trivedi, M. M. 2018. Convolutional social pooling for vehicle trajectory prediction. In *Proc. of IEEE CVPR workshops*, 1468–1476.
- Erlygin, L.; Zholobov, V.; Baklanova, V.; Sokolovskiy, E.; and Zaytsev, A. 2023. Uncertainty estimation for time series forecasting via Gaussian process regression surrogates. *SIGKDD 2023*.
- Feng, D.; Harakeh, A.; Waslander, S. L.; and Dietmayer, K. 2021. A review and comparative study on probabilistic object detection in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(8): 9961–9980.
- Gibbs, I.; and Candes, E. 2021. Adaptive conformal inference under distribution shift. *NeurIPS 2021*, 34: 1660–1672.
- Gu, X.; Song, G.; Gilitschenski, I.; Pavone, M.; and Ivanovic, B. 2024. Producing and Leveraging Online Map Uncertainty in Trajectory Prediction. *Proc. of the IEEE/CVF CVPR*.
- Gupta, A.; Johnson, J.; Fei-Fei, L.; Savarese, S.; and Alahi, A. 2018. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proc. of IEEE CVPR*, 2255–2264.
- Hu, Y.; Yang, J.; Chen, L.; Li, K.; Sima, C.; Zhu, X.; Chai, S.; Du, S.; Lin, T.; Wang, W.; et al. 2023. Planning-oriented autonomous driving. In *Proc. of the IEEE/CVF*, 17853–17862.
- Ivanovic, B.; Lin, Y.; Shrivastava, S.; Chakravarty, P.; and Pavone, M. 2022. Propagating state uncertainty through trajectory forecasting. In *2022 ICRA*, 2351–2358. IEEE.
- Kedia, S.; Zhou, Y.; and Karumanchi, S. H. 2023. Integrated Perception and Planning for Autonomous Vehicle Navigation: An Optimization-Based Approach. In *Proc. of the IEEE/CVF*, 3205–3214.
- Liang, M.; Yang, B.; Hu, R.; Chen, Y.; Liao, R.; Feng, S.; and Urtasun, R. 2020. Learning lane graph representations for motion forecasting. In *Computer Vision–ECCV 2020*, 541–556. Springer.
- Lindemann, L.; Cleaveland, M.; Shim, G.; and Pappas, G. J. 2023. Safe planning in dynamic environments using conformal prediction. *IEEE Robotics and Automation Letters*.
- Liu, Y.; Zhang, J.; Fang, L.; Jiang, Q.; and Zhou, B. 2021. Multimodal motion prediction with stacked transformers. In *Proc. of the IEEE/CVF Conference on CVPR*, 7577–7586.
- Mao, W.; Xu, C.; Zhu, Q.; Chen, S.; and Wang, Y. 2023. Leapfrog Diffusion Model for Stochastic Trajectory Prediction. In *Proc. of IEEE/CVF*, 5517–5526.
- Meyer, G. P.; and Thakurdesai, N. 2020. Learning an uncertainty-aware object detector for autonomous driving. In *2020 IEEE/RSJ International Conference on IROS*, 10521–10527. IEEE.
- Ngiam, J.; Caine, B.; Vasudevan, V.; Zhang, Z.; Chiang, H.-T. L.; Ling, J.; Roelofs, R.; Bewley, A.; Liu, C.; Venugopal, A.; et al. 2021. Scene transformer: A unified architecture for predicting multiple agent trajectories. *arXiv preprint arXiv:2106.08417*.
- Phan-Minh, T.; Grigore, E. C.; Boulton, F. A.; Beijbom, O.; and Wolff, E. M. 2020. Covernet: Multimodal behavior prediction using trajectory sets. In *Proc. of IEEE/CVF*, 14074–14083.
- Rasmussen, C. E.; Williams, C. K.; et al. 2006. *Gaussian processes for machine learning*, volume 1. Springer.
- Rhinehart, N.; Kitani, K. M.; and Vernaza, P. 2018. R2p2: A reparameterized pushforward policy for diverse, precise generative path forecasting. In *Proc. of ECCV*, 772–788.
- Rhinehart, N.; McAllister, R.; Kitani, K.; and Levine, S. 2019. Precog: Prediction conditioned on goals in visual multi-agent settings. In *Proc. of IEEE/CVF*, 2821–2830.
- Salzmann, T.; Ivanovic, B.; Chakravarty, P.; and Pavone, M. 2020. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *Computer Vision–ECCV 2020*, 683–700. Springer.
- Shafer, G.; and Vovk, V. 2008. A Tutorial on Conformal Prediction. *Journal of Machine Learning Research*, 9(3).
- Shi, S.; Jiang, L.; Dai, D.; and Schiele, B. 2023. Mtr++: Multi-agent motion prediction with symmetric scene modeling and guided intention querying. *arXiv preprint arXiv:2306.17770*.
- Stankeviciute, K.; M Alaa, A.; and van der Schaar, M. 2021. Conformal time-series forecasting. *NeurIPS 2021*, 34: 6216–6228.
- Stanton, S.; Maddox, W.; and Wilson, A. G. 2023. Bayesian optimization with conformal prediction sets. In *AISTATS*, 959–986. PMLR.

- Sun, J.; Jiang, Y.; Qiu, J.; Nobel, P.; Kochenderfer, M. J.; and Schwager, M. 2024. Conformal prediction for uncertainty-aware planning with diffusion dynamics model. *NeurIPS 2024*, 36.
- Varadarajan, B.; Hefny, A.; Srivastava, A.; Refaat, K. S.; Nayakanti, N.; Cornman, A.; Chen, K.; Douillard, B.; Lam, C. P.; Anguelov, D.; et al. 2022. Multipath++: Efficient information fusion and trajectory aggregation for behavior prediction. In *2022 ICRA*, 7814–7821. IEEE.
- Xu, C.; and Xie, Y. 2023. Sequential predictive conformal inference for time series. In *International Conference on Machine Learning*, 38707–38727. PMLR.
- Xu, R.; Chen, W.; Xiang, H.; Xia, X.; Liu, L.; and Ma, J. 2023. Model-agnostic multi-agent perception framework. In *2023 IEEE ICRA*, 1471–1478. IEEE.
- Xu, W.; Pan, J.; Wei, J.; and Dolan, J. M. 2014. Motion planning under uncertainty for on-road autonomous driving. In *2014 IEEE ICRA*, 2507–2512. IEEE.
- Zeng, W.; Liang, M.; Liao, R.; and Urtasun, R. 2021. Lanercnn: Distributed representations for graph-centric motion forecasting. In *2021 IEEE/RSJ International Conference on IROS*, 532–539. IEEE.
- Zhong, Y.; Ni, Z.; Chen, S.; and Neumann, U. 2022. Aware of the history: Trajectory forecasting with the local behavior data. In *European Conference on Computer Vision*, 393–409. Springer.
- Zhou, Z.; Wang, J.; Li, Y.-H.; and Huang, Y.-K. 2023. Query-Centric Trajectory Prediction. In *Proc. of the IEEE/CVF CVPR*, 17863–17873.
- Zhou, Z.; Ye, L.; Wang, J.; Wu, K.; and Lu, K. 2022. Hivt: Hierarchical vector transformer for multi-agent motion prediction. In *Proc. of IEEE/CVF*, 8823–8833.
- Zhu, D.; Zhai, G.; Di, Y.; Manhardt, F.; Berkemeyer, H.; Tran, T.; Navab, N.; Tombari, F.; and Busam, B. 2023. IPCC-TP: Utilizing Incremental Pearson Correlation Coefficient for Joint Multi-Agent Trajectory Prediction. In *Proc. of IEEE/CVF*, 5507–5516.