Proceedings of the ASME 2025
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC-CIE2025
August 17-20, 2025, Anaheim, CA

DETC2025-169050

INTEGRATING IMAGE SEGMENTATION AND DIGITAL HUMAN MODELING THROUGH MULTI-FIDELITY GAUSSIAN PROCESS FOR DRIVER VISIBILITY ANALYSIS

M. Amin Firouzi[1], Yitong Bu[1], H. Onan Demirel [1,2], Christopher Hoyle[1,*],

[1]School of Mechanical, Industrial and Manufacturing Engineering, Oregon State University
[2]Department of Design, Manufacturing and Engineering Management, University of Strathclyde

ABSTRACT

Accurate assessment of driver visibility is crucial in automotive design and safety enhancement, particularly in situations where A-pillars obstruct the driver's field of view. To address this challenge, this research develops a multi-fidelity Gaussian Process (MF-GP) modeling framework to enhance visibility prediction by integrating low-fidelity (LF) image segmentation data with high-fidelity digital human modeling (DHM) simulations. By leveraging a limited set of high-fidelity samples, the proposed MF-GP framework systematically calibrates low-fidelity data to improve predictive accuracy while reducing computational costs. Two A-pillar cutout designs (3.75 cm and 5 cm) were analyzed under varying HF sampling densities of 3%, 7%, and 10%. Results indicate that the 3.75 cm cutout is more sensitive to sparse HF sampling, requiring a denser HF dataset to achieve stable calibration. In contrast, the 5 cm cutout, benefiting from improved LF-HF alignment, achieves comparable accuracy with fewer HF samples. Model validation using root mean square error (RMSE) and coefficient of determination (R^2) confirms that increasing HF sampling enhances surrogate model accuracy, with the effect being more pronounced in cases where model performance is susceptible to high-fidelity data. The proposed framework provides a computationally efficient methodology for driver visibility prediction and human-in-the-loop design applications. Future research could explore adaptive HF sampling strategies and ensemble surrogate modeling techniques to further enhance multi-fidelity learning efficiency.

Keywords: Multi-fidelity modeling, Bayesian calibration, Digital human modeling (DHM), Image segmentation, Product Design, Simulation-based Design, Metamodeling, Human-Centered Design

*Corresponding author
Documentation for `asmeconf.cls`: Version 1.38, May 19, 2025.

NOMENCLATURE

d	Euclidean distance between points
k	Kernel function
l	Kernel length scale
n	Number of data points
ρ	Scaling factor for low-fidelity model
r	Distance between inputs for the kernel
y_{HF}	High-fidelity model output
y_{LF}	Low-fidelity model output
$\mathbf{x}$	Input vector
$\mathbf{h}$	Input height vector
$\mathbf{K}$	Covariance matrix
$\boldsymbol{\theta}$	Hyperparameters of the Gaussian process
σ^2	Kernel variance
$\delta(\mathbf{x})$	Discrepancy term in the high-fidelity model
$\log p(\mathbf{y}\|\mathbf{x},\boldsymbol{\theta})$	Log marginal likelihood
R^2	Coefficient of determination
RMSE	Root mean square error

1. INTRODUCTION

Computational modeling has become an indispensable element of engineering and scientific research, providing researchers with an effective means of exploring complex physical systems without extensive physical experimentation. High-fidelity (HF) computational simulations may provide accurate predictions but are generally resource-intensive for large applications. Multifidelity modeling (MFM), however, offers an alternative by integrating high-fidelity resource-intensive models with lower precision but more cost-efficient low-fidelity (LF) approximation models to provide both cost efficiency and predictive accuracy in one approach; MFM has quickly become an indispensable method used across countless fields as a key methodology to tackle challenges associated with large applications and large-scale applications alike [1].

Kennedy and O'Hagan [2] pioneered a Bayesian Gaus-

sian Process (GP)-based framework to systematically incorporate multiple fidelity levels while simultaneously decreasing dependence on expensive high-fidelity evaluations, while increasing computational efficiency by using lower-fidelity simulations as sources for informing and refining high-fidelity model predictions. Although initially applied to computer experiments, this methodology has since been extended for uncertainty quantification and model calibration in flow measurement [3]. Furthermore, extensions to refinements have been introduced, such as recursive co-kriging by Le Gratiet [4] and Le Gratiet and Garnier [5], which have further increased computational efficiency. Beyond computational modeling, MFM has gained significant traction in engineering design, particularly in automotive safety, where computational efficiency plays a crucial role. Adaptive multi-fidelity sampling frameworks have been developed in connected and automated vehicle (CAV) systems to enhance safety analysis while reducing the computational burden of HF evaluations [6]. This approach is particularly relevant in enhancing safety assessments in vehicle design, where MFM can reduce computational costs in evaluating accident risks. However, driver vision has not been addressed in the safety area, and this remains a gap in this research field and has become a motivation for this article. A safety issue comes with an obstruction area due to the automobile A-pillar, which is the vertical support structure located on either side of the windshield, as shown in Figure 1.

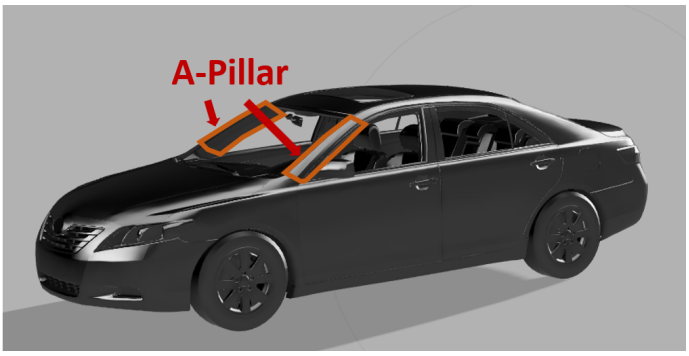

FIGURE 1: ILLUSTRATION OF THE A-PILLAR STRUCTURE LINKING THE WINDSHIELD, ROOF, AND FRAME

A-pillars provide vital structural support yet may obscure near-field vision, leading drivers to miss pedestrians or other hazards in their field of vision [7]. Increasing the thickness of the A-pillar, its coverage angle must remain between 6 ° and 12 ° [8] to ensure that the driver can adequately control the environment. While previous research explored A-pillar modifications, such as cutouts or voids, to increase driver visibility with Digital Human Modeling (DHM) simulations [9], the computational aspect of this simulation method remains the go-to approach. Firouzi et al. [10, 11] employed Gaussian Process (GP) surrogate modeling applied to DHM simulations and design optimization techniques, which are well-suited for Bayesian optimization due to their ability to achieve accurate results with a minimal number of iterations. However,DHM uses ray casting and iterative CAD modeling techniques, which are time-intensive and impractical for rapid evaluations [12]. To overcome these computational limitations, we developed an alternative image segmentation-based approach that offers a significantly faster method for evaluating driver visibility.

To bridge this gap, we employed Kennedy and O'Hagan's [1] multi-fidelity calibration framework, integrating a limited number of DHM samples as HF data with LF image segmentation data. This multi-fidelity (MF) approach leverages the complementary strengths of both fidelity levels to enable rapid yet accurate visibility predictions, addressing a critical gap in vehicle safety assessment. This framework was applied to two different A-pillar cutout designs with diameters of 5 cm and 3.75 cm. As an example, a GP surrogate model was trained using data from both DHM and image segmentation simulations for the 5 cm cutout design. This surrogate model effectively captures the broader design space, as illustrated in Figures 2 and 3. It generalizes across key input variables, including driver height and object position, based on a training distribution that spans the operational domain. As shown in the Figures 2 and 3, the model produces smooth, continuous visibility surfaces over this domain, demonstrating effective interpolation and generalization beyond the training points.

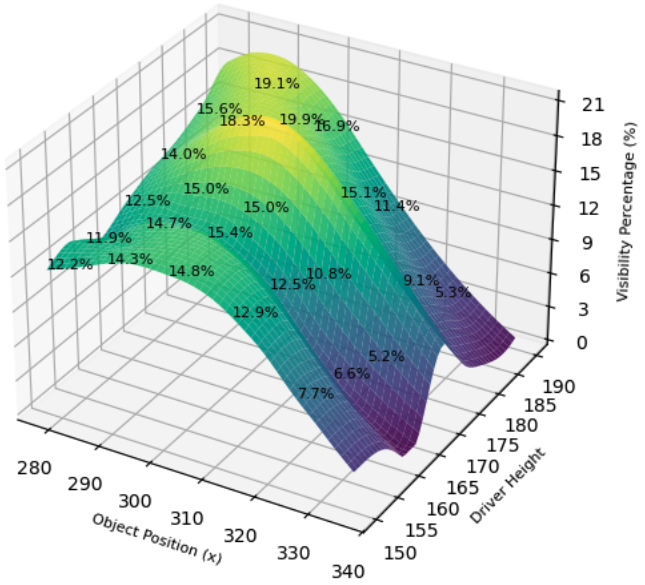

FIGURE 2: SURROGATE MODEL OF HIGH FIDELITY MODEL - DHM SIMULATION.

2. MULTI-FIDELITY BACKGROUND

Multi-fidelity modeling (MFM) has quickly emerged as an indispensable technique in computational science, combining high-fidelity (HF) and low-fidelity (LF) models to balance computational efficiency with predictive accuracy. Kennedy and O'Hagan [1] pioneered MFM by devising a Bayesian Gaussian Process (GP) framework that integrates various fidelity levels systematically while decreasing dependence on costly simulations by taking advantage of cheaper approximation methods; their pioneering work inspired subsequent studies which led to advanced multi-fidelity methodologies used for optimization, uncertainty quantification, model calibration [13], [14]).

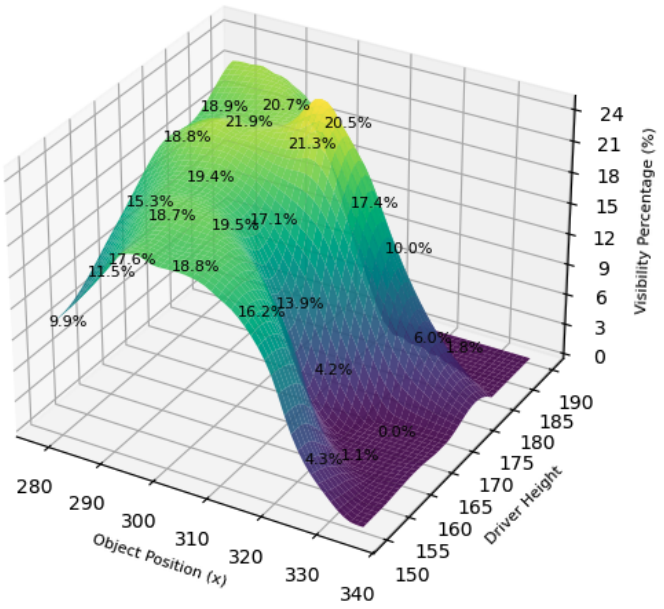

FIGURE 3: SURROGATE MODEL OF LOW FIDELITY MODEL-IMAGE SEGMENTATION.

2.1 Bayesian Inference and Multi-Fidelity Calibration

MFM relies heavily on Bayesian inference for probabilistic modeling of system uncertainties and systematic calibration of multi-fidelity models. This is demonstrated by Seshadri et al. [15], who used Bayesian inference to reconstruct engine temperature distributions from sparse sources. Their work accounted for measurement errors and insufficient sampling as separate sources. Bayesian assimilation techniques have also proven invaluable in computational fluid dynamics. Examples include using Bayesian assimilation techniques to infer boundary conditions [16] and to align turbulence models with direct numerical simulation data [17].

Goh et al. [18] developed a hierarchical Bayesian model based on Kennedy and O'Hagan's autoregressive structure [1], refining MF calibration by modeling discrepancy terms and improving the predictive performance of Gaussian Processes. Recursive co-kriging was then utilized by Le Gratiet [4] and Le Gratiet and Garnier [5] to increase computational efficiency by iteratively refining LF predictions prior to integrating HF data - leading to improved model fidelity alignment while simultaneously decreasing costs through taking account of discrepancies more effectively. These approaches significantly improved model fidelity alignment while simultaneously lowering computational costs by accounting for discrepancies more effectively.

2.2 Gaussian Process Surrogate Models in Multi-Fidelity Frameworks

GPs are frequently employed in MFM due to their flexibility in function approximation and uncertainty quantification. Kennedy and O'Hagan's autoregressive model [1], popularly referred to as AR1, remains one of the primary approaches used for structured fidelity management [19] in engineering applications with structured fidelity requirements; using it allows LF approximations as baselines while corrections via discrepancy functions can also be incorporated.

Co-kriging extends this paradigm by explicitly modeling cross-correlations among fidelity levels, thus improving the predictive accuracy of engineering simulations [20]. Multi-output GP by Qian et al. [21] integrating qualitative and quantitative factors in computer experiments requires building correlation functions, which can model complex interactions among different factor types. This approach has potential applications in MFM. Ghosh et al. [22] introduce Max IF-UCR, a cost-aware adaptive sampling strategy for multi-fidelity GP modeling. Unlike conventional two-step methods, it selects design points and fidelity levels simultaneously, maximizing uncertainty reduction per unit cost. Enhanced by the GP Believer Strategy, it pre-assesses the impact of new points, minimizing unnecessary high-fidelity evaluations. Their approach outperforms Max MF-UCR, achieving significant cost savings while maintaining accuracy, as demonstrated on analytical benchmarks and a thermodynamic fluidized bed process.

2.3 Discrepancy Correction and Model Refinement in Multi-Fidelity Modeling

One of the most significant challenges in MFM is addressing discrepancies between LF and HF models. Kennedy and O'Hagan [1] formulated a Bayesian discrepancy model that applies a GP-based bias function to systematically correct LF predictions, forming the foundation for modern calibration techniques.

To refine this process, alternative approaches have emerged, including hierarchical Bayesian calibration [23], which introduce probabilistic structures to account for systematic inconsistencies across fidelity levels. Co-kriging techniques [20] further enhance this alignment by capturing cross-correlations between LF and HF datasets, reducing reliance on extensive HF evaluations while preserving model accuracy. Beyond static corrections, adaptive sampling strategies have been proposed to selectively refine LF models, ensuring that HF evaluations are applied only in regions of high LF error [22]. This uncertainty-aware approach prioritizes computational resources where they are most needed, avoiding unnecessary HF simulations while maintaining predictive accuracy.

Recursive inference techniques, such as those developed by Le Gratiet [4], further optimize computational efficiency by iteratively refining LF predictions before incorporating HF data. Unlike direct fidelity correction, this approach gradually adjusts LF outputs through incremental updates, leading to faster convergence and improved computational efficiency.

In complex multiphysics systems, fidelity mismatches often exhibit nonlinear behavior, requiring nonlinear discrepancy modeling rather than the traditional linear formulations assumed in the Kennedy-O'Hagan framework. Recent studies [19] explore hierarchical Bayesian approaches that treat the discrepancy as a latent variable, capturing nonlinear fidelity gaps more effectively. This is particularly relevant in computational fluid dynamics and structural mechanics, where high-fidelity variations can be irregular. Together, these advancements reinforce the growing sophistication of MF calibration, demonstrating how adaptive sampling, recursive inference, and nonlinear discrepancy correction

can significantly improve predictive reliability while minimizing computational cost.

3. SIMULATION SETUP

This section outlines our approach to high- and low-fidelity simulations, implementing the Digital Human Modeling (DHM) approach to evaluate driver visibility obstructed by the A-pillar. We will investigate both commercial DHM software and our implementation of the image segmentation method.

3.1 Overview

Prior research by Joffe et al. [24] has shown, that visibility percentages obtained through simulations using the DHM approach and those derived via image segmentation generally exhibit similar percent visibility outcomes. However, significant discrepancies occur when objects are very close or far from the vehicle. Image segmentation offers faster results but lacks the accuracy of DHM simulation data. In this study, we designate the image segmentation data as low-fidelity and the DHM simulation data as high-fidelity.

This research does not involve optimization in the conventional sense of selecting between LF and HF data across the design space. Instead, LF data is assumed to be freely available throughout the design space, while HF data is employed selectively in regions where significant discrepancies between the two models occur. This targeted usage of HF data allows for computational efficiency without sacrificing accuracy in critical areas.

This study includes a primary simulation scenario: a pedestrian crossing in front of a sedan. The pedestrian is used as a traffic object and has fixed heights, while the driver's height varies from 150 to 190 cm in increments of 5 cm per simulation. Each simulation tests two A-pillar cutout sizes: 3.75 cm and 5 cm. As shown in Figure 4, the pedestrian moves from left to right across the driver's field of view. This scenario was selected because pedestrian interactions represent a common and critical safety concern in urban driving, particularly in areas of near-field vision that the A-pillar frequently blocks. Designing with the DHM approach is widely utilized to evaluate early-stage prototypes regarding ergonomics, including visibility and obstruction. For example, Marshall et al. assessed the field of view of truck drivers to measure the blind zone [25]. Karmakar et al. evaluated the field of view of pilots in jet aircraft with various postures, positions, and populations to demonstrate the effectiveness of DHM in terms of ergonomic assessment [26]. However, DHM visualization methods, which primarily focus on percent visibility measurements, rely heavily on static scene analysis, limiting their effectiveness for ergonomic assessments in dynamic environments.

To address this limitation, we use computer vision techniques—specifically object detection and image segmentation—to analyze dynamic simulations and automate visual obscuration assessments. We set up simulations in Siemens Jack, using its visual obscuration toolkit for HF analysis and simulation videos for LF assessments [27]. In the LF simulation, we integrated YOLOv8 for object detection and Meta SAM for image segmentation [28] [29]. Both HF and LF simulations are implemented in Jack to generate scenarios for visual obscuration assessments.

3.2 High Fidelity Simulation

Because Jack's visual obscuration toolkit requires a rectangular target plane to represent traffic objects, we created separate planes for pedestrians and cyclists, each sized to match their height and width. In the pedestrian crossing scenario, the rectangle starts at the left edge of the A-pillar and moves 2 cm to the right at each step. At each position, we recorded visibility from the driver's field of view through the A-pillar cut-out until the pedestrian reaches the right edge.

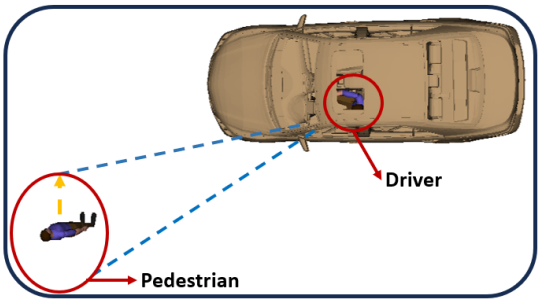

FIGURE 4: THE PEDESTRIAN CROSSING THE SEDAN SCENARIO. BLUE DASHED LINES ARE THE A-PILLAR BLIND ZONE AND THE PEDESTRIAN'S WALKING AREA FOR BOTH HIGH-FIDELITY SIMULATION AND LOW-FIDELITY SIMULATION. THE YELLOW DASHED ARROW INDICATES THE DIRECTION OF THE PEDESTRIAN'S WALKING MOTION.

3.3 Low Fidelity Simulation

For the computer vision-based visual obscuration assessment, we first assigned a walking motion to the pedestrian in Jack's Task Simulation Builder to generate walking videos. This walking motion followed the same path used in the HF simulation. For each scenario, we generated two videos: (1) an unobscured video that shows only the traffic object (with the sedan and driver hidden) and (2) an obscured video that includes the full simulation scene with all objects. YOLOv8 was used to detect the traffic object (pedestrian) in the unobscured video and pass its location to Meta SAM for image segmentation. The location information constrains the segmentation to a specific region, allowing the segmentation masks to obtain all pixels of the traffic object. The area visual obscuration percent(AVO%) computed as Equation (1)[30]. The number of pixels in this segmentation mask is the denominator in Equation (1). Next, the location data from the unobscured video was passed to the obscured video to obtain obscured traffic objects. We assigned a green color to differentiate the A-pillar from traffic objects. All green pixels were excluded from the obscured traffic object's segmentation mask during the segmentation. The number of pixels collected in this segmentation mask served as the numerator in Equation(1). An image segmentation result is shown in Figure 5.

$$\text{AVO } \% = \frac{\text{Obscured Pixels}}{\text{Total Pixels}} \times 100, \tag{1}$$

where "obscured pixels" are those blocked by the A-pillar and "total pixels" represent the full visible area of the object.

In Jack, the visual obscuration toolkit used ray-casting to a target plane representing the traffic object from the driver's

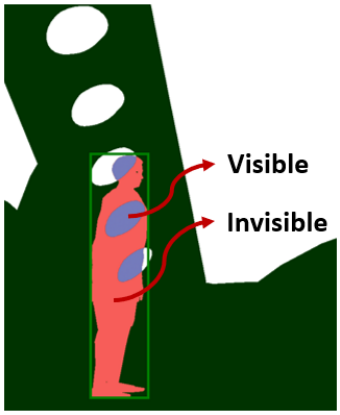

FIGURE 5: THIS IS IMAGE SEGMENTATION FROM THE DRIVER'S EYE VIEW WINDOW. BLUE PARTS ARE VISIBLE TO THE DRIVER, AND THE A-PILLAR OBSCURES RED PARTS.

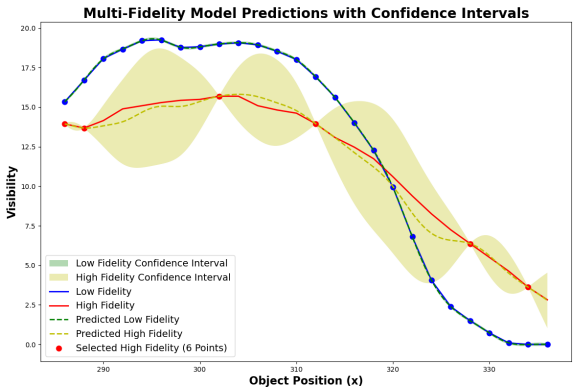

FIGURE 6: 1D MODEL PREDICTION WITH CO-KRIGING SURROGATE MODEL.

eye through the A-pillar cutout. This process involved analyzing the sedan model to determine which areas blocked rays, so the processing time may increase with the complexity of the 3D model. In contrast, the image segmentation method assessed visual obscuration using 2D videos instead of 3D scenes, improving computational efficiency.

4. MULTI-FIDELITY FRAMEWORK

This research used simulation results for various heights of the drivers and positions of objects to simulate pedestrian crossing street scenarios. Two distinct datasets were utilized: image segmentation results as LF data while digital human modeling serves as the ground truth, representing HF model. Notably, discrepancies between low—and high-fidelity data became most apparent near extreme points within the design space, highlighting that model calibration can be an approach to utilizing the benefits of both models to enhance prediction.

Before adopting the MFM approach in full force, an initial calibration test was carried out on one dataset to establish whether calibration could be an appropriate technique and assess the level of agreement between datasets. The driver height category tested (155 cm) showed that the calibrated model was closely aligned with the driver HF data. As depicted in Figure 6, this proved the validity of the calibration approach as both datasets displayed similar patterns, with the calibrated model providing reasonable approximations of ground truth when the calibrated model was close in alignment.

Figure 7 reveals the MF framework employed in this study, designed to systematically combine LF and HF data sources in order to enhance predictive accuracy while reducing computational costs. The framework consists of two fidelity classes to facilitate image segmentation data calibration against DHM samples using a structured multi-fidelity model. A Multi-fidelity Gaussian Process (MF-GP) model is employed to refine LF predictions to match HF data more closely. Calibration is accomplished with gradient-based optimizers such as multi-start BFGS [31] to increase training efficiency. The implementation leverages the Emukit libraries [32, 33], which utilizes GPy [34] for Gaussian Process modeling for robust and scalable multi-fidelity learning.

4.1 High-Fidelity Data Selection and Sampling Points

Since HF simulations are computationally expensive, a two-step sampling approach was implemented to ensure efficient and diverse coverage of the design space while minimizing redundancy.

1. **Latin Hypercube Sampling (LHS)**: Latin Hypercube Sampling was selected over other methods because it provides uniform input space coverage with fewer samples, reducing clustering and improving sampling efficiency in high-dimensional settings [35]. Each sampled design point, $\mathbf{x}$, was defined by driver height and object position parameters.

2. **Nearest Neighbor Selection**: After generating LHS points, the closest matching points were selected from the available HF dataset using the Euclidean distance. This ensured that the sampled HF data aligned closely with the intended design while respecting the existing data constraints.

$$d(\mathbf{x}, \mathbf{x}') = \sqrt{\sum_{i=1}^{n} (x_i - x_i')^2} \tag{2}$$

where $\mathbf{x}$ is design sampled point, and $\mathbf{x}'$ is the nearest data point in the existing dataset.

This process was repeated in 50 independent experiment designs to assess the consistency and performance of the calibration model systematically. Since LF models tend to be computationally inexpensive, we utilized the entire LF dataset. MFM was tested using **3%**, **7%**, and **10%** of the high-fidelity dataset to ensure efficient calibration with minimal reliance on HF data. These sampling rates were selected to reflect scenarios where HF evaluations are limited, and are consistent with previous findings that demonstrate that multi-fidelity frameworks can outperform single-fidelity models when using only a small number of

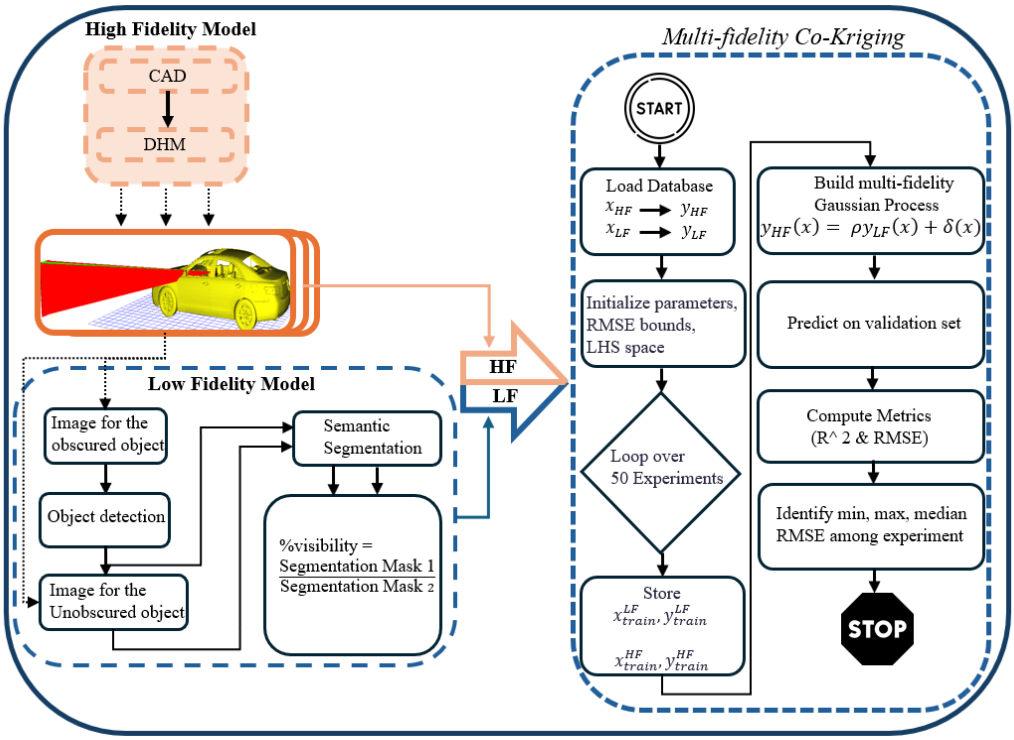

FIGURE 7: MULTI-FIDELITY FRAMEWORK.

HF samples [36]. This approach enabled an assessment of the model's sensitivity to data availability and its predictive accuracy under varying fidelity constraints.

4.2 Multi-Fidelity Model Formulation

To integrate LF and HF data within a framework, an MF-GP model was implemented based on the Kennedy and O'Hagan (KOH) formulation [1]. This model assumes that HF data can be expressed as a transformation of LF predictions with an additive discrepancy term, capturing systematic deviations across fidelity levels. The relationship between fidelity levels is given by:

$$y_{\text{HF}}(\mathbf{x}) = \rho y_{\text{LF}}(\mathbf{x}) + \delta(\mathbf{x}) + \epsilon, \tag{3}$$

where:

- $y_{\text{HF}}(\mathbf{x})$: **High-fidelity output**,
- $y_{\text{LF}}(\mathbf{x})$: **Low-fidelity prediction**,
- ρ: **Scaling factor** aligning LF predictions with HF data,
- $\delta(\mathbf{x})$: **Discrepancy function**, accounting for systematic differences between fidelity levels,
- $\epsilon \sim \mathcal{N}(0, \sigma^2)$: **Independent Gaussian noise term**.

To model the correlation between fidelity levels, a multi-fidelity covariance function is constructed, enabling information transfer between LF and HF data. The covariance structure is defined as:

$$\mathbf{K}_{\text{multi-fidelity}} = \begin{bmatrix} \mathbf{K}_{\text{LF,LF}} & \mathbf{K}_{\text{LF,HF}} \\ \mathbf{K}_{\text{HF,LF}} & \mathbf{K}_{\text{HF,HF}} \end{bmatrix}, \tag{4}$$

where:

- $\mathbf{K}_{\text{LF,LF}}$ and $\mathbf{K}_{\text{HF,HF}}$: **Covariance matrices** of the LF and HF Gaussian Processes, respectively,
- $\mathbf{K}_{\text{LF,HF}} = \rho \mathbf{K}_{\text{LF,LF}}$: **Cross-covariance** between fidelity levels, ensuring alignment of LF and HF data,
- $\delta(\mathbf{x})$: **Modeled as an independent Gaussian Process (GP)** with its own covariance function.

The Matérn 5/2 covariance function was chosen for its balance between smoothness and flexibility, producing twice-differentiable sample paths that are better suited for modeling moderately rough functions than the overly smooth squared exponential kernel [37]

$$k(\mathbf{x}, \mathbf{x}\prime) = \sigma^2 \left(1 + \sqrt{5}r + \frac{5r^2}{3}\right) \exp(-\sqrt{5}r), \tag{5}$$

where:

- $r = \frac{||\mathbf{x}-\mathbf{x}'||}{l}$: **Scaled Euclidean distance** between input points,
- l: **Length scale**, controlling smoothness,
- σ^2: **Signal variance**, defining function variability.

This covariance function ensures that both the LF and HF Gaussian processes maintain continuity and adaptability, capturing complex variations across fidelity levels. By incorporating the discrepancy function within this MF-GP formulation, the model is capable of correcting LF predictions to better approximate HF

outputs, while quantifying uncertainty in fidelity discrepancies. This approach allows for computationally efficient surrogate modeling, particularly beneficial in scenarios where HF data is scarce or expensive to obtain. The model's hyperparameters—including length scales, noise variances, the scaling factor ρ, and the cross-covariance terms—are jointly optimized by maximizing the log marginal likelihood of the observed data across fidelity levels. Specifically, the log marginal likelihood function accounts for the full MF covariance matrix, ensuring that all relevant parameters, including those governing the interaction between LF and HF data, are tuned simultaneously. This joint optimization process enables the model to effectively learn the relationships and uncertainties across fidelity levels, thereby improving predictive accuracy and robustness. This optimization was performed using gradient-based methods, which efficiently tuned the model parameters to support robust and scalable predictive inference across fidelity levels.

4.3 Implementation with Emukit

The MF-GP model was implemented using the Emukit [32] toolbox, a modular framework for constructing and optimizing different models across multiple fidelity levels. Emukit provides efficient tools for handling multi-fidelity learning based on GPy, ensuring seamless integration of low-fidelity and high-fidelity data while optimizing model performance.

Key components of the implementation included:

- **LinearMultiFidelityKernel**: Defines the correlation between LF and HF data by constructing a joint covariance function. This kernel follows the autoregressive structure of the Kennedy and O'Hagan framework, ensuring that HF predictions leverage LF information while accounting for discrepancies.

- **GPyMultiOutputWrapper**: Manages the multi-output Gaussian Process model, allowing for the simultaneous training of LF and HF data within a unified Gaussian Process framework. This wrapper enables efficient optimization and predictive inference by leveraging the shared structure between fidelity levels.

- **Hyperparameter Optimization**: The model hyperparameters, including the length scale, variance, and noise terms, were optimized by maximizing the log marginal likelihood function:

$$\log p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}| - \frac{n}{2}\log(2\pi), \quad (6)$$

  where:

  - $\mathbf{K}$ is the covariance matrix, encapsulating the dependencies between fidelity levels,
  - $\mathbf{y}$ represents the observed data across all fidelity levels,
  - $\boldsymbol{\theta}$ denotes the set of hyperparameters to be optimized,
  - n is the number of observations.

4.4 Model Validation and Performance Metrics

The accuracy and generalization capability of the calibrated MF model were assessed using HF data points that were excluded from the calibration process. To quantify the model's predictive performance, the following statistical metrics were employed:

- **Root Mean Square Error (RMSE)**: This metric evaluates the average magnitude of prediction errors by computing the standard deviation of residuals. A lower RMSE value indicates higher predictive accuracy.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}, \quad (7)$$

  where:

  - y_i represents the actual HF observations,
  - $\hat{y}_i$ denotes the predicted HF values from the multi-fidelity model,
  - n is the number of HF validation points.

- **Coefficient of Determination** R^2: The R^2 metric measures the proportion of variance in the HF data that is captured by the model. A R^2 value close to 1 signifies a strong correlation between predictions and actual HF values, while a value close to 0 indicates weak predictive capability.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \quad (8)$$

  where:

  - y_i and $\hat{y}_i$ denote the actual and predicted HF values, respectively,
  - $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the mean of the HF observations.

- **Actual vs. Predicted Plot**: These plots provide a visual representation of the degree of agreement between the true model and the calibrated model where:

  - *The True Model* is the model estimated on 180 HF only samples,
  - *The Calibrated Model* is the MFM estimated using the HF samples to calibrate the LF samples.

These metrics provide a comprehensive assessment of the model's predictive accuracy and reliability. RMSE captures the absolute magnitude of errors, while R^2 quantifies how well the model explains variance in the HF data. Together, these indicators validate the effectiveness of the multi-fidelity framework in leveraging LF data to enhance HF predictions.

5. RESULT

The MF-GP model was evaluated for two different cutout sizes, 3.75 cm and 5 cm, across 6, 12, and 18 high-fidelity samples that represent the **3%**, **7%**, and **10%** of HF data. Figures 8 and 9 present the model's predictive performance under each condition.

This section provides a detailed analysis of both cases, emphasizing *multi-fidelity modeling principles*, *sampling strategies*, and *model quality*, specifically in terms of root mean square error and the coefficient of determination. Additionally, the impact of HF sample density on the surrogate model's accuracy is discussed to evaluate the effectiveness of fidelity-aware calibration.

Figure 8 clearly demonstrates that the MF surrogate model is highly susceptible to high-fidelity sample density. When only six HF samples are available, predictive versus actual plots exhibit substantial deviations from the ideal fit line, and RMSE values increase significantly, signaling major prediction errors. Furthermore, the R^2 score drops drastically, indicating an inadequate representation of variance within the dataset. The three-dimensional surface reconstructions appear coarse due to insufficient calibration data, and R^2 continues to decrease as the model does not sufficiently capture the variance in the dataset.

From an MFM perspective, the reduced accuracy observed in the six HF sample cases is expected, as the model heavily relies on LF predictions, which inherently contain bias. Furthermore, due to the limited availability of HF data, the model is unable to effectively refine its discrepancy function, leading to poor fidelity alignment. Differences in data density exacerbate misalignments between the LF and HF models. While LF data are obtained from simulation videos at a higher temporal resolution, HF data are sampled at discrete 2 cm intervals. This disparity contributes to the misalignment of fidelity, mainly when the HF sample density is low.

As the number of HF samples increases to 12 and 18, RMSE values decrease systematically, and R^2 improves. This indicates that the MF surrogate model captures an increasing proportion of variance associated with HF behavior. MF fidelity vs HF plots exhibit tighter clustering around the ideal fit line, and the 3D surface reconstructions become progressively more refined, suggesting that the model's approximation of HF behavior improves with additional calibration data.

A key challenge associated with the 3.75 cm cutout is the discrepancy in visibility perception between the driver's eye view and the DHM-based obscuration test. The driver's eye view often obscures the lower portions of nearby traffic objects, while the DHM test registers these objects as fully visible. This inconsistency introduces systematic bias in near-field object integration within the MF framework, which requires a higher HF sample density to be effectively corrected.

Figure 9 presents 5 cm cutout design results showing similar trends but notable variations in model robustness and sampling efficiency. Comparing 6 HF samples with 3.75 cm samples leads to significantly lower RMSE values and higher R^2 scores than their 3.75 cm counterparts, showing that MF surrogate models perform better under sparse HF conditions; MF model versus HF plots show variations that seem less extreme, while 3D reconstructions reveal smoother transitions indicating stronger correlation between LF-HF relations.

From a multi-fidelity modeling perspective, the 5 cm cutoff offers superior LF-HF alignment due to its wider field of view, which reduces near-field visibility discrepancies. This improved LF-HF alignment arises because the broader field of view in the 5cm cutout reduces partial occlusions in LF views, making them more consistent with DHM results. In contrast, the narrower 3.75cm cutout more frequently obscures near-field objects in LF data while they remain visible in HF simulations, increasing the discrepancy. The wider cutoff reduces this mismatch, leading to more stable calibration with fewer HF samples.

As more HF samples increase to 12 and then 18 samples, the RMSE gradually decreases while the R^2 scores improve, similar to those with cases of 3.75 cm. This could indicate that a 5 cm cutout offers an inherently stronger relationship between low- and high-frequency samples and thus allows smoother calibration with fewer samples; MF predictive model versus HF plots show improved alignment with ideal-fit lines while 3D surrogate reconstruction is refined similarly but less sensitively due to sparsity issues.

From an efficiency perspective, the 5 cm cutout requires significantly less HF sampling density to achieve accurate calibration than its 3.75 cm counterpart; its model achieved relatively stable $RMSE$ and R^2 performance with only six samples collected - showing how larger cutouts reduce the need for extensive data collection to achieve comparable accuracy levels compared with 3.75 cm cases which require much denser datasets to reach comparable accuracy levels.

6. CONCLUSION AND FUTURE WORK

This study evaluated the predictive performance of a multi-fidelity Gaussian Process (MF-GP) model for driver visibility assessment across two cutout sizes, 3.75 cm and 5 cm, using the MF simulation framework. The HF data were obtained from DHM simulations, while the LF data were derived from image segmentation-based simulations. This MF approach leverages the complementary strengths of both fidelity levels to improve predictive accuracy while minimizing computational costs. Results demonstrated that the quality of an MF surrogate model can be strongly affected by HF sample availability and alignment sensitivity, particularly for a 3.75 cm cutout, which required denser HF sample distribution. In contrast, the 5 cm cutout achieved comparable accuracy with fewer HF samples, due to improved LF-HF alignment and reduced visibility discrepancies.

Preliminary timing comparisons were conducted to provide a perspective on computational efficiency. Although a complete breakdown of the simulation overhead is beyond the scope of this paper, the average processing time for a single DHM simulation, excluding the initial model setup, was approximately 45 seconds. In contrast, the average processing time for a single image segmentation operation was about 8 seconds. Although these figures vary depending on hardware and implementation details, they illustrate the significant computational advantage of incorporating LF data within the MFM framework.

These results demonstrate how image segmentation methods can serve as reliable LF approximations when combined with DHM simulations in an MF-GP framework. This approach effectively captures visibility trends and has great promise for human-in-the-loop design studies; however, issues related to near-field object detection remain and may adversely affect calibration quality under sparse HF sampling.

Although this study did not employ adaptive sampling, its findings indicate that future work could focus on developing MF-

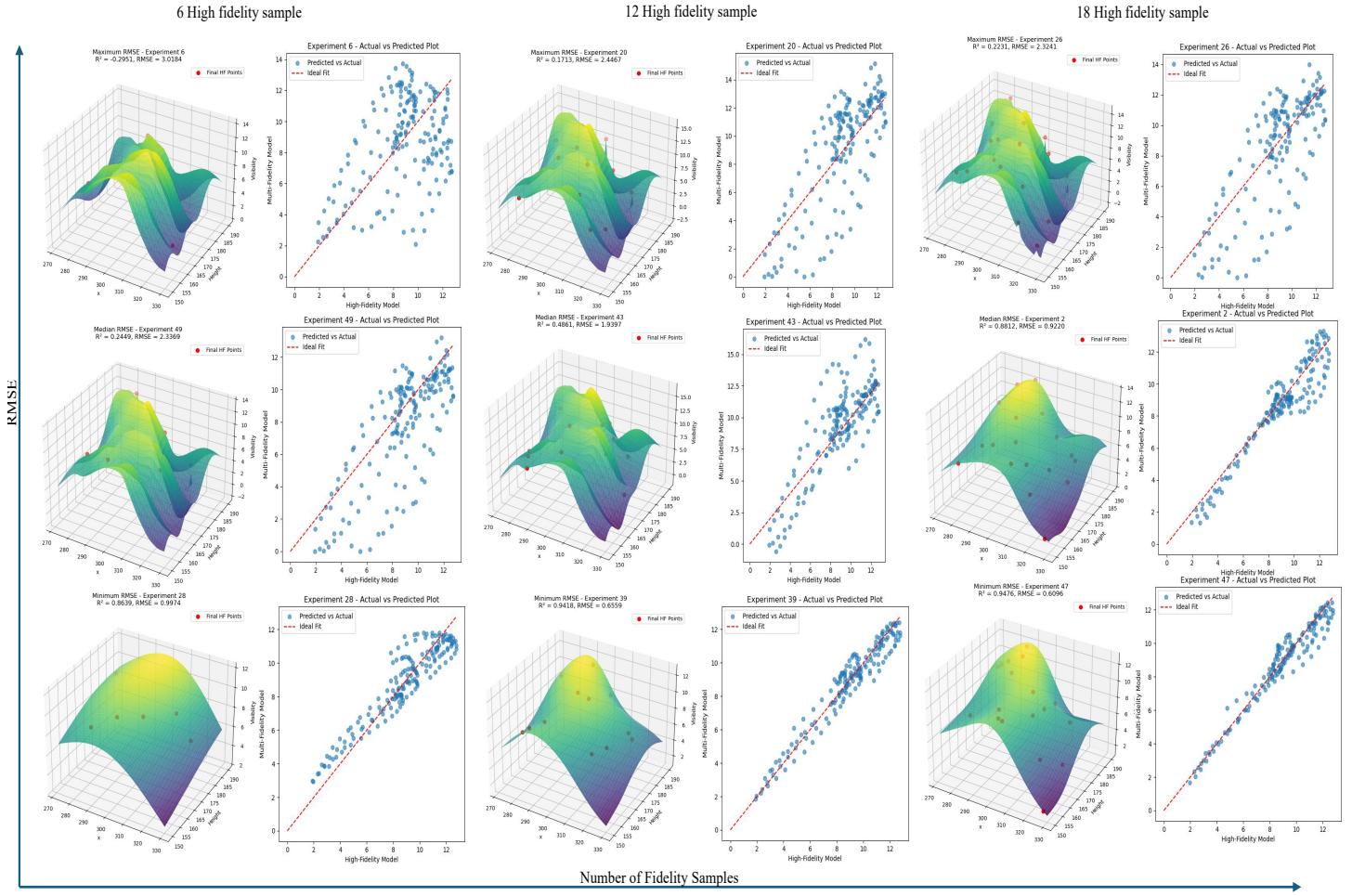

FIGURE 8: MF SURROGATE MODELS AND THEIR PERFORMANCES UNDER VARYING HF SAMPLE DENSITIES FOR THE 3.75 CM CUTOUT.

aware sampling strategies in which points are dynamically allocated depending on uncertainty estimation to optimize data efficiency, reduce redundant evaluations while maintaining predictive accuracy of the models, and enhance sample selection by decreasing redundant evaluations while still achieving predictive accuracy of models. Such strategies would ensure optimal sample selection while maintaining accurate predictions from MF models. Continued refinement of these methodologies within the MF-GP framework may further reduce the dependence on large HF datasets while maintaining robust predictive performance, ultimately improving computational efficiency in visibility evaluation and broader engineering design contexts.

REFERENCES

[1] Kennedy, Marc C and O'Hagan, Anthony. "Predicting the output from a complex computer code when fast approximations are available." *Biometrika* Vol. 87 No. 1 (2000): pp. 1–13.

[2] Kennedy, Marc C and O'Hagan, Anthony. "Bayesian calibration of computer models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* Vol. 63 No. 3 (2001): pp. 425–464.

[3] Cruz, Gonçalo G, Babin, Cedric, Ottavy, Xavier and Fontaneto, Fabrizio. "A Bayesian Data Driven Multi-Fidelity Modelling Approach for Experimental Under-Sampled Flow Reconstruction." *Turbo Expo: Power for Land, Sea, and Air*, Vol. 87080: p. V13AT29A043. 2023. American Society of Mechanical Engineers.

[4] Le Gratiet, Loic. "Multi-fidelity Gaussian process regression for computer experiments." Ph.D. Thesis, Université Paris-Diderot-Paris VII. 2013.

[5] Le Gratiet, Loic and Garnier, Josselin. "Recursive co-kriging model for design of computer experiments with multiple levels of fidelity." *International Journal for Uncertainty Quantification* Vol. 4 No. 5 (2014).

[6] Gong, Xianliang, Feng, Shuo and Pan, Yulin. "An adaptive multi-fidelity sampling framework for safety analysis of connected and automated vehicles." *IEEE Transactions on Intelligent Transportation Systems* Vol. 24 No. 12 (2023): pp. 14393–14405.

[7] Sivak, M., Schoettle, B., Reed, M. P. and Flannagan, M. J. "Body-pillar vision obstructions and lane-change crashes." *Journal of Safety Research* Vol. 38 No. 5 (2007): pp. 557–561. DOI 10.1016/j.jsr.2007.06.003.

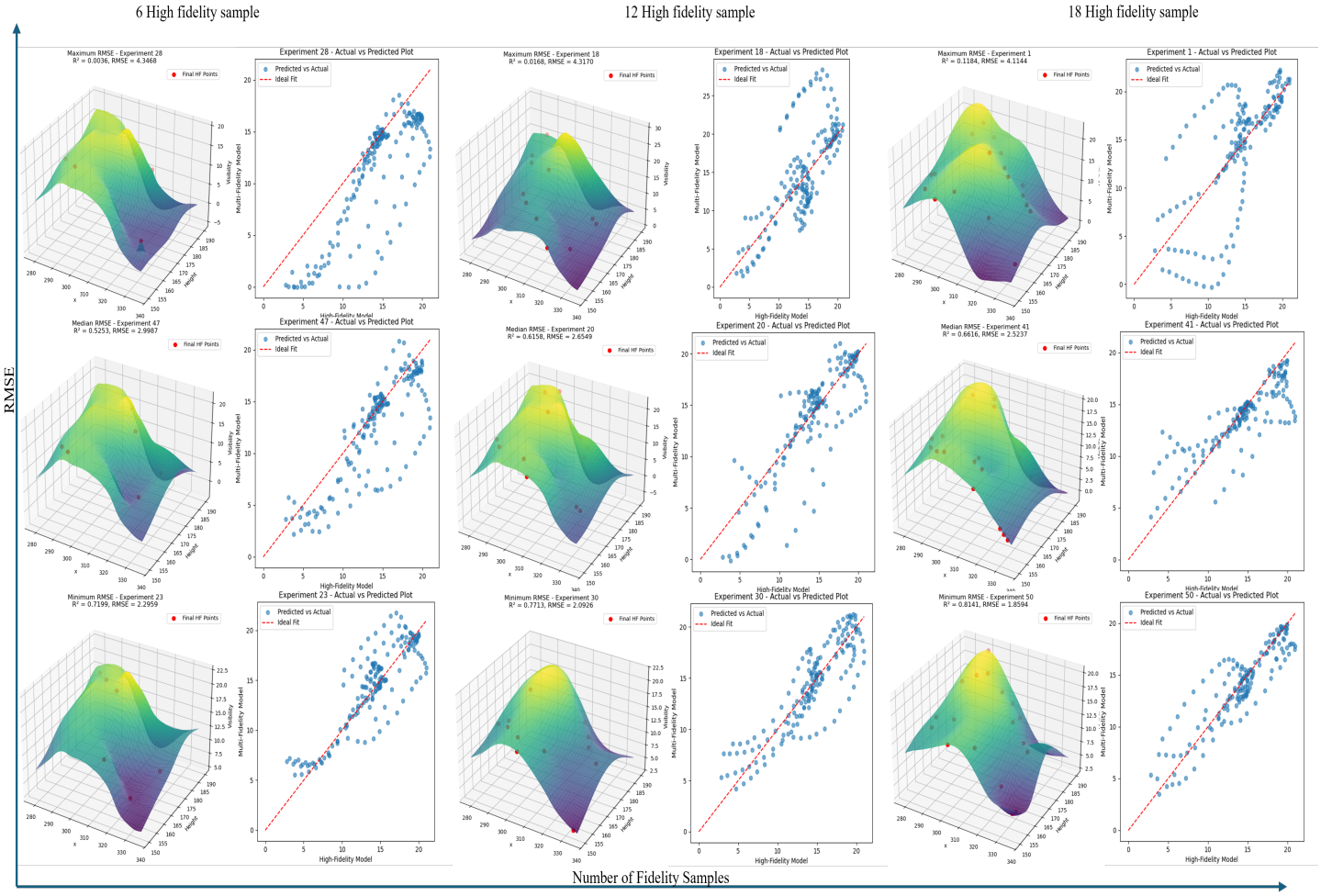

FIGURE 9: MF SURROGATE MODELS AND THEIR PERFORMANCES UNDER VARYING HF SAMPLE DENSITIES FOR THE 5 CM CUTOUT.

[8] Cao, C., Wei, J., Wang, X. and Tan, H. "Driver's Perception of A-Pillar Blind Area: Comparison of Two Different Auditory Feedback." Rau, P.-L. P. (ed.). *Cross-Cultural Design. Applications in Cultural Heritage, Tourism, Autonomous Vehicles, and Intelligent Agents*, Vol. 12773: pp. 171–181. 2021. Springer International Publishing. DOI 10.1007/978-3-030-77080-8_15.

[9] Srinivasan, S. and Demirel, H. O. "Quantifying Vision Obscuration of A-Pillar Concept Variants Using Digital Human Modeling." *Volume 2: 42nd Computers and Information in Engineering Conference (CIE)*: p. V002T02A049. 2022. American Society of Mechanical Engineers. DOI 10.1115/DETC2022-89781.

[10] Firouzi, M. Amin, Bhaskaran, Vignesh, Demirel, H. Onan and Hoyle, Christopher. "Metamodels and Vision Obstruction: A New Lens on Driver Visibility." *Proceedings of the ASME 2024 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. Volume 2A: 44th Computers and Information in Engineering Conference (CIE): p. V02AT02A056. 2024. DOI 10.1115/DETC2024-143657. URL https://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2024/88346/V02AT02A056/7412659/v02at02a056-detc2024-143657.pdf, URL https://doi.org/10.1115/DETC2024-143657.

[11] Firouzi, M. A., Jetton, C., Bhaskaran, V., Hoyle, C. and Demirel, H. O. "Improving Road Safety With Human-In-The-Loop Bayesian Optimization Using Driver Vision Obstruction Simulations." *Proceedings of the ASME 2024 International Mechanical Engineering Congress and Exposition*, Vol. 1: Acoustics, Vibration, and Phononics; Advanced Design and Information Technologies: p. V001T02A017. 2024. ASME, Portland, Oregon, USA. DOI 10.1115/IMECE2024-145790.

[12] H. Onan Demirel, Salman Ahmed and Duffy, Vincent G. "Digital Human Modeling: A Review and Reappraisal of Origins, Present, and Expected Future Methods for Representing Humans Computationally." *International Journal of Human–Computer Interaction* Vol. 38 No. 10 (2022): pp. 897–937. URL https://doi.org/10.1080/10447318.2021.1976507.

[13] Sarkar, Soumalya, Mondal, Sudeepta, Joly, Michael, Lynch, Matthew E, Bopardikar, Shaunak D, Acharya, Ranadip and Perdikaris, Paris. "Multifidelity and multiscale Bayesian

framework for high-dimensional engineering design and calibration." *Journal of Mechanical Design* Vol. 141 No. 12 (2019): p. 121001.

[14] Fernández-Godino, M Giselle. "Review of multi-fidelity models." *arXiv preprint arXiv:1609.07196* (2016)DOI 10.3934/acse.2023015.

[15] Seshadri, Pranay, Duncan, Andrew B, Thorne, George, Parks, Geoffrey, Diaz, Raul Vazquez and Girolami, Mark. "Bayesian assessments of aeroengine performance with transfer learning." *Data-Centric Engineering* Vol. 3 (2022): p. e29. DOI 10.1017/dce.2022.29.

[16] Sousa, Jorge and Gorlé, Catherine. "Computational urban flow predictions with Bayesian inference: Validation with field data." *Building and Environment* Vol. 154 (2019): pp. 13–22. DOI 10.1016/j.buildenv.2019.02.028.

[17] Foures, Dimitry PG, Dovetta, Nicolas, Sipp, Denis and Schmid, Peter J. "A data-assimilation method for Reynolds-averaged Navier–Stokes-driven mean flow reconstruction." *Journal of Fluid Mechanics* Vol. 759 (2014): pp. 404–431. DOI 10.1017/jfm.2014.566.

[18] Goh, Joslin, Bingham, Derek, Holloway, James Paul, Grosskopf, Michael J, Kuranz, Carolyn C and Rutter, Erica. "Prediction and computer model calibration using outputs from multifidelity simulators." *Technometrics* Vol. 55 No. 4 (2013): pp. 501–512. DOI 10.1080/00401706.2013.838910.

[19] Peherstorfer, Benjamin, Willcox, Karen and Gunzburger, Max. "Survey of multifidelity methods in uncertainty propagation, inference, and optimization." *SIAM Review* Vol. 60 No. 3 (2018): pp. 550–591. DOI 10.1137/16M1082469.

[20] Forrester, Alexander IJ, Sóbester, András and Keane, Andy J. "Multi-fidelity optimization via surrogate modelling." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* Vol. 463 No. 2088 (2007): pp. 3251–3269. DOI 10.1098/rspa.2007.1900.

[21] Qian, Peter Z G, Wu, Huaiqing and Wu, CF Jeff. "Gaussian process models for computer experiments with qualitative and quantitative factors." *Technometrics* Vol. 50 No. 3 (2008): pp. 383–396.

[22] Ghosh, Sayan, Kristensen, Jesper, Zhang, Yiming, Subber, Waad and Wang, Liping. "A strategy for adaptive sampling of multi-fidelity Gaussian processes to reduce predictive uncertainty." *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 59193: p. V02BT03A024. 2019. American Society of Mechanical Engineers. DOI 10.1115/DETC2019-98418.

[23] Tuo, Rui, Wu, CF Jeff and Yu, Dan. "Surrogate modeling of computer experiments with different mesh densities." *Technometrics* Vol. 56 No. 3 (2014): pp. 372–380.

[24] Joffe, Gabrielle B., Bu, Yitong and Demirel, H. Onan. "Assessing Vision Obstruction and Safety in E-Scooter Accidents: An Integrated Approach Using Digital Human Modeling and Computer Vision (Accepted Paper)." *Proceedings of the Human-Computer Interaction International (HCII)*. 2025.

[25] Marshall, Russell, Summerskill, Steve, Paterson, Abby and Eland, Anthony. "The use of digital human modeling for the definition and contextualisation of a direct vision standard for trucks." *DHM2020*. IOS Press (2020): pp. 99–107.

[26] Karmakar, Sougata, Pal, Madhu Sudan, Majumdar, Deepti and Majumdar, Dhurjati. "Application of digital human modeling and simulation for vision analysis of pilots in a jet aircraft: a case study." *Work* Vol. 41 No. Supplement 1 (2012): pp. 3412–3418.

[27] Inc., Siemens Product Lifecycle Management Software. "Jack Fact Sheet: A Premier Human Simulation Tool." (2011). URL https://www.siemens.com/plm. Accessed: 2025-03-08.

[28] Jocher, Glenn, Chaurasia, Ayush and Qiu, Jing. "Ultralytics YOLOv8." (2023). URL https://github.com/ultralytics/ultralytics.

[29] Kirillov, Alexander, Mintun, Eric, Ravi, Nikhila, Mao, Hanzi, Rolland, Chloe, Gustafson, Laura, Xiao, Tete, Whitehead, Spencer, Berg, Alexander C., Lo, Wan-Yen, Dollár, Piotr and Girshick, Ross. "Segment Anything." *arXiv:2304.02643* (2023).

[30] Demirel, H Onan, Jennings, Alex and Srinivasan, Sriram. "An early design method to quantify vision obstruction: formula one (F1) halo case study." *International Conference on Human-Computer Interaction*: pp. 32–44. 2022. Springer.

[31] Fletcher, Roger. *Practical methods of optimization*. John Wiley & Sons (2000).

[32] Paleyes, Andrei, Pullin, Mark, Mahsereci, Maren, McCollum, Cliff, Lawrence, Neil and González, Javier. "Emulation of physical processes with Emukit." *Second Workshop on Machine Learning and the Physical Sciences, NeurIPS*. 2019.

[33] Paleyes, Andrei, Mahsereci, Maren and Lawrence, Neil D. "Emukit: A Python toolkit for decision making under uncertainty." *Proceedings of the Python in Science Conference* (2023).

[34] GPy. "GPy: A Gaussian Process Framework in Python." http://github.com/SheffieldML/GPy (2012). Accessed: March 2025.

[35] McKay, M. D., Beckman, R. J. and Conover, W. J. "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code." *Technometrics* Vol. 21 No. 2 (1979): pp. 239–245.

[36] Pennati, Giacomo, Zhang, Jun, Durocher, Martine and Cacace, Silvio. "A comparative review of multi-fidelity surrogate modeling techniques for engineering applications." *Aerospace Science and Technology* Vol. 106 (2020): p. 106339. DOI 10.1016/j.ast.2020.106339.

[37] Rasmussen, Carl Edward and Williams, Christopher K. I. *Gaussian processes for machine learning*, 3rd ed. Adaptive computation and machine learning, MIT Press, Cambridge, Mass. (2008).