

An Evaluation of Methods for Assessing Robot Kinematic Model Accuracy in the Presence of Noise

Mitchell R. Woodside, Patrick Bazzoli, Philip A. Olubodun, J. Adam Nisbett, Guixiu Qiao, and Douglas A. Bristow¹

Abstract—While there are many works developing methods for modeling and calibrating robot kinematics, assessing the accuracy of those models has received little attention. However, accuracy assessment is critically important for applications where the robot must operate with absolute accuracy over a large region of workspace, such as in robotic machining. When the model of such a system is well calibrated, the remaining deterministic error can be quite complex, owing to complicated gearing errors, deformations, and quasi-static thermal changes. Locating the largest deterministic error requires an exploration over the workspace, but assessing the largest error is complicated by repeatability error and measurement noise. How then to assess the largest error from such a measurement set? This paper evaluates the efficacy of two conventional methods, maximum measured error and outlier rejection, and a novel method based on model invalidation that uses a hypothesis testing framework. A machining robot is used to develop a numerical study for evaluation of these methods under differing magnitude of measurement noise. A high-order kinematic model of the robot is constructed as used as the true robot kinematics, and the workspace for that system is used as the region of interest. A best-fit Denavit-Hartenberg (DH) model is used as the model whose accuracy is to be measured. The study shows that the largest deterministic error can be difficult to locate with just a few percent of points approaching the defining accuracy limit. As expected, the largest measured error provides a poor (over)estimate of the error as noise is increased, but outlier rejection can be equally as bad as rare large deterministic errors can be easily mistaken for low-probability large random error. The novel model invalidation method, however, performs well across all noise levels.

I. INTRODUCTION

Industrial robots are increasingly being used in applications where absolute accuracy is required across large portions of their workspace. One such application is in robotic machining [1],[2] where the large range of motion and flexibility of a robotic setup are primary drivers, while the complexity of motion paths precludes the manual setup and alignment of local frames that are typically used to achieve accurate motion.

*This work was performed under the following financial assistance award 70NANB23H177 from U.S. Department of Commerce, National Institute of Standards and Technology.

Mitchell R. Woodside, Patrick Bazzoli, Philip A. Olubodun, J. Adam Nisbett, and Douglas A. Bristow are with the Department of Mechanical and Aerospace Engineering, Missouri University of Science and Engineering, Rolla, MO, USA.

Guixiu Xiao is with the Cognition and Collaboration Systems Group, National Institute of Standards and Technology, Gaithersburg, MD, USA

¹corresponding author: dbristow@mst.edu

Achieving good absolute accuracy is highly dependent on the quality of the kinematic calibration, which has long been a focus of study [3],[4]. Many metrology tools have been considered for measurement of the robot during calibration, although laser trackers are commonly used today due to their ability to rapidly measure the robot over large regions of workspace. Likewise, a multitude of calibration optimization methods have been utilized including least-squares minimization and maximum likelihood methods in which stochastic measurement error and robot repeatability can be incorporated [5]. While most calibrated kinematic models use corrections to the robot DH parameters, newer methods have begun to consider higher-order descriptions of kinematic error including position-dependent terms [6],[7] inspired by calibration strategies utilized in higher accuracy machine tools [8].

Methods to assess robot accuracy, on the other hand, have received little attention in the literature. The ISO 9283 standard [9] provides a method for measuring points on a constant-orientation, inclined plane, but these points explore a very limited region of the workspace. Other methods use random, or quasi-random, points to obtain a better distribution over the workspace [6]. Accuracy is often assessed as the mean magnitude or maximum magnitude of the measured error. While the maximum error may be most relevant to applications requiring absolute accuracy, it is highly dependent on both locating the region of the workspace where the largest errors exist, and the random robot repeatability error and measurement noise when that measurement is acquired. None-the-less, the largest measured error may be frequently cited in practice, despite this sensitivity problem. Likewise, it may also be commonplace to discard large outliers. In [6], the authors fit a gamma distribution to the acquired measurements and extract the 99% percentile of the distribution to reduce these sensitivities. However, such an approach assumes that the true deterministic accuracy follows a normal distribution, which is unlikely to be true. Thus, the challenges of assessing the largest absolute accuracy error of a robot over a workspace is twofold: locating the region where the deterministic error is largest and decoupling the deterministic accuracy from the random robot repeatability and measurement noise. While a straightforward method of isolating the deterministic error is to repeat each measurement multiple times,

doing so reduces the number of unique locations to measure across the workspace when measurement time is constrained.

In this work we explore the problem of assessing the maximum deterministic accuracy error of the robot over a region. The study will focus on a particular machining robot and its workspace using differing magnitudes of robot repeatability and measurement noise. In order to efficiently evaluate this scenario, the study is performed numerically. A high-order kinematic error model [6] is identified for the robot and used as the true robot kinematics, while best-fit DH parameter model is used as the nominal model whose accuracy is to be assessed. Simulated measurements of a set of points over the workspace are performed and then evaluated using two conventional assessment methods, maximum measured error and outlier rejection, and one novel invalidation method. The invalidation method uses a hypothesis rejection test at each measurement to optimally determine whether a new location should be measured or whether a previous location should be remeasured. In doing so, the method optimally navigates the dual challenges of exploring the workspace for the largest error and decoupling the deterministic accuracy from random repeatability and noise error.

The remainder of the paper is organized as follows. Background on kinematic models and conventional accuracy assessment methods are presented in Section II. The novel model invalidation assessment method is presented in Section III. A numerical study is presented in Section IV and concluding remarks are given in Section V.

II. BACKGROUND

A. Kinematic Models for Industrial Robots

Kinematics are used to relate a robot's joint positions to a Cartesian position and orientation of its end effector [10]. The output of a forward kinematic model, i.e., the kinematic model that computes the position and orientation of the end effector from the robot's joint position, is often represented by a 4x4 homogeneous transformation matrix. This transformation matrix for a standard six degree of freedom robot arm is,

$$\mathbf{T}_6^0(\mathbf{q}) = \begin{bmatrix} \mathbf{R}_6^0(\mathbf{q}) & \mathbf{p}_6^0(\mathbf{q}) \\ \mathbf{0} & 1 \end{bmatrix} \quad (1)$$

where $\mathbf{R}_6^0 \in \mathbb{R}^{3 \times 3}$ and $\mathbf{p}_6^0 \in \mathbb{R}^{3 \times 3}$ are, respectively, a rotation matrix and position vector that describe the position and orientation of the robot's end effector. The vector, \mathbf{q} , is a vector of the robot's joint positions,

$$\mathbf{q} = [q_1, q_2, \dots, q_6]^T \quad (2)$$

where q_i for $i = 1, 2, \dots, 6$ is the angle of each joint.

There are many forward kinematic models that can be used to compute the transformation in (1). In this paper two different models are discussed, the Denavit-Hartenberg (DH) model [11] and a high-order

joint-dependent model [6], [12]. The high-order joint-dependent model will be referred to as the Volumetric Error Compensation (VEC) model.

The DH model of a robot with six rotary joints is,

$$\mathbf{T}_6^0(\mathbf{q}) = \mathbf{T}_1^0(q_1) \mathbf{T}_2^1(q_2) \dots \mathbf{T}_6^5(q_6) \quad (3)$$

which can be described as a product of six homogeneous transformation matrices, \mathbf{T}_i^{i-1} , that define the relative transformation between coordinate frames attached to joint i and joint $i-1$. Each \mathbf{T}_i^{i-1} is a product of four elementary transformations,

$$\mathbf{T}_i^{i-1}(q_i) = \mathbf{Rot}_z(\theta_i(q_i)) \mathbf{Trans}_z(d_i) \mathbf{Trans}_x(a_i) \mathbf{Rot}_x(\alpha_i) \quad (4)$$

where $\mathbf{Rot}_x(\cdot)$ and $\mathbf{Rot}_z(\cdot)$ describe rotations about the X and Z axes of a coordinate frame, and $\mathbf{Trans}_x(\cdot)$ and $\mathbf{Trans}_z(\cdot)$ describe translations along these axes. The DH rotation parameters, $\theta_i(\cdot)$ and α_i define the orientation of the relative transformation, and the DH translation parameters, a_i and d_i , define its translation.

The DH convention is a simple model which assumes that, apart from the simple joint rotations, the robot is perfectly rigid. It is commonly used to represent the nominal forward kinematics of a robot due to its simple formulation; however, it does not account for the nonlinear link deflections and joint imperfections of a real industrial robot.

To account for these nonlinearities, the VEC model places additional error matrices, \mathbf{E}_i^{i-1} , between each DH joint transformation, \mathbf{T}_i^{i-1} . Thus, the VEC model of a robot with six rotary joints is,

$$\tilde{\mathbf{T}}_6^0(\mathbf{q}, \mathbf{s}) = \mathbf{E}_1^0(q_1, s_1) \mathbf{T}_1^0(q_1) \dots \mathbf{E}_6^5(q_6, s_6) \mathbf{T}_6^5(q_6) \quad (5)$$

where \mathbf{s} is a boolean function selector vector that describes the travel direction of each joint (used to account for joint backlash), and the error matrices, \mathbf{E}_i^{i-1} , are of the form,

$$\mathbf{E}_i^{i-1}(q_i, s_i) = \begin{bmatrix} 1 & -\varepsilon_z(q_i, s_i) & \varepsilon_y(q_i) & \delta_x(q_i) \\ \varepsilon_z(q_i, s_i) & 1 & -\varepsilon_x(q_i) & \delta_y(q_i) \\ -\varepsilon_y(q_i) & \varepsilon_x(q_i) & 1 & \delta_z(q_i) \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6)$$

Each element of the error matrix is computed using an n^{th} order Chebyshev polynomial, which is mapped to a selected range of the robot's joint space. The polynomials are parameterized using the methods discussed in [6], [12].

B. Conventional Accuracy Assessment Methods

Let j denote the index of a set of m operating points that span a desired region of a multi-dimensional system's operating space. This region will be referred to as the evaluation space and the true accuracy of a system

model for the set of points spanning this subspace is defined as,

$$a = \max_S \|\mathbf{e}_j\| \quad (7)$$

where the vector, $\mathbf{e}_j \in \mathbb{R}^n$, is the deterministic error between the true and modeled system outputs, n denotes the dimension of the system's output, and S is the set of all error magnitudes at each operating point, that is,

$$S = \{\|\mathbf{e}_j\| : j = 1, 2, \dots, m\} \quad (8)$$

Note that when assessing the position accuracy of a multi-dimensional system, e.g., an industrial robot, the elements of \mathbf{e}_j would correspond the position error in each spatial direction, i.e., the errors in the X, Y, and Z axes of the common inertial frame.

The system's true output is often unknown and typically measured by an external measurement device. Unfortunately, the uncertainties in both the system and the measurement device will make the measured output and consequently the measured error, $\tilde{\mathbf{e}}_j \in \mathbb{R}^n$, stochastic. In this work it will be assumed that the uncertainties are independent, Gaussian, zero-mean, and have small magnitude. Although these assumptions may or may not affect the practical accuracy of the system model, they will still allow for a relative comparison of the accuracy assessment methods. Thus, when $\tilde{\mathbf{e}}_j$ is linearized to the first order, it is a single instance of the random variable, \mathbf{Y}_j , that follows the multi-dimensional normal distribution,

$$\mathbf{Y}_j \sim \mathcal{N}(\mathbf{e}_j, \Sigma_j) \quad (9)$$

where Σ_j is the covariance matrix of all uncertainties at the j^{th} operating point. Since the uncertainties are assumed to be uncoupled, Σ_j will have the form,

$$\Sigma_j = \begin{bmatrix} \sigma_{1,j}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{2,j}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{n,j}^2 \end{bmatrix} \quad (10)$$

where the diagonals of Σ_j are the variances of each uncoupled element in \mathbf{Y}_j .

A conventional method for assessing a model's accuracy is to apply (7) directly to the set of error measurements. This is simply,

$$\tilde{a} = \max_{\tilde{S}} \|\tilde{\mathbf{e}}_j\| \quad (11)$$

where \tilde{S} is the set of all error measurement magnitudes at each operating point, defined as,

$$\tilde{S} = \{\|\tilde{\mathbf{e}}_j\| : j = 1, 2, \dots, m\} \quad (12)$$

Although this method is a direct application of (7) it will often produce a worse accuracy (larger maximum error magnitude) than the true accuracy due to the stochasticity of $\tilde{\mathbf{e}}_j$.

To combat this issue, another conventional method is to remove outliers from \tilde{S} . The accuracy for this method is,

$$\tilde{a} = \max_{\tilde{S} \setminus O} \|\tilde{\mathbf{e}}_j\| \quad (13)$$

where O is the set of outliers, defined as,

$$O = \{O \subseteq \tilde{S} : \|\tilde{\mathbf{e}}_j\| > c \text{ for } j = 1, 2, \dots, m\} \quad (14)$$

In (14), c is the selected confidence interval for the distribution of the measurements in \tilde{S} . While this method removes seemingly erroneous measurements, it often removes important measurements and the quantified accuracy can differ from the system's true accuracy significantly. The shortcomings of both of these methods will be demonstrated through simulation in Section IV.

III. MODEL INVALIDATION

This accuracy assessment method, developed in [13] and re-presented here in its algorithmic form, stems from the concept that since $\tilde{\mathbf{e}}_j$ is stochastic, there is always a chance that a new measurement at the same operating point that will produce a larger error. Thus, it is difficult if not impossible to quantify the accuracy and validate a system model for all possible instances [13]. For this reason, this method focuses on a more probabilistic scenario, that is, if enough error measurements have been collected, a model can be invalidated with respect to a desired accuracy bound, η , with statistical confidence [13]. This is why this method is referred to as Model Invalidation (MI).

Fundamentally, MI can be described as a hypothesis test that is performed recursively as more measurements of the model's error are acquired. At the k^{th} iteration of the algorithm, the null and alternate hypotheses, H_0 and H_1 , are,

$$\begin{cases} H_0 & \forall \|\mathbf{e}_j\| \in S_k, \|\mathbf{e}_j\| \in [0, \eta] \\ H_1 & \exists \|\mathbf{e}_j\| \in S_k : \|\mathbf{e}_j\| \notin [0, \eta] \end{cases} \quad (15)$$

where $\eta \geq 0$ is a predetermined accuracy bound that the true errors of the model are compared against, and $S_k \subseteq S$ is the current set of true error magnitudes, which have been measured up to and including the current iteration of the algorithm. Let \tilde{S}_k be the current set of these measurements, then the goal of the hypothesis test is to determine if the measurements in \tilde{S}_k provide sufficient "evidence" to reject the null hypothesis, which will invalidate the model. More precisely, the criteria for rejecting the null hypothesis is,

$$\begin{cases} \mathbf{P}_k(\tilde{S}_k|H_0) < 1 - \alpha & \text{Reject } H_0 \\ \mathbf{P}_k(\tilde{S}_k|H_0) \geq 1 - \alpha & \text{Fail to reject } H_0 \end{cases} \quad (16)$$

where \mathbf{P}_k is the probability that all future measurements will have a larger magnitude than the current measurements in \tilde{S}_k , given that H_0 is true. This probability (evidence) is compared to the preselected confidence

interval, $\alpha \in [0, 1]$, to determine if it sufficient to reject H_0 , and is computed by,

$$\mathbf{P}_k(\bar{S}_k|H_0) = \prod_{j=1}^{m_k} (1 - \mathbf{p}_k^*(\bar{\mathbf{e}}_j|H_0)) \quad (17)$$

where m_k is the total number of measurements that have been collected at the current iteration, and \mathbf{p}_k^* is probability that all future measurements at the j^{th} operating point will have a smaller magnitude than the current measurement at the same operating point, $\bar{\mathbf{e}}_j$, given that H_0 is true. This probability is computed using the Cumulative Distribution Function (CDF) of the $\chi^2(n)$ distribution by,

$$\mathbf{p}_k^*(\bar{\mathbf{e}}_j|H_0) = \int_0^{\mathbf{R}(\bar{\mathbf{e}}_j)} \frac{x^{n/(2-1)} e^{-x/2}}{2^{n/2} \Gamma(n/2)} dx \quad (18)$$

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx \quad (19)$$

where $\Gamma(\cdot)$ is the gamma function, and \mathbf{R} is the Mahalanobis distance [14],

$$\mathbf{R}(\bar{\mathbf{e}}_j) = \sqrt{(\bar{\mathbf{e}}_j - \mathbf{e}_j)^T \bar{\Sigma}_j^{-1} (\bar{\mathbf{e}}_j - \mathbf{e}_j)} \quad (20)$$

In (20), $\bar{\Sigma}_j$ is the covariance matrix that corresponds to $\bar{\mathbf{e}}_j$ at the current iteration of the algorithm. Based on the assumptions stated at the beginning of Section II-B, $\bar{\Sigma}_j$ will follow the form of (10).

To understand how the probability in (17) can be used to reject H_0 , consider the one-dimensional example, show in Fig. 1, for a single operating point. In the figure, the measurement, \bar{e} , follows a distribution that does not satisfy H_0 , but the probability computed from the measurement, \mathbf{P}_k , is based on a distribution that does satisfy H_0 (as would be done for the hypothesis test). As shown in the top plot of the figure, when the measurement occurs farther away from its true value, e , (which is less likely) \mathbf{P}_k is relatively large. In contrast, when the measurement occurs closer to e (which is more likely) \mathbf{P}_k is relatively small, as shown in the bottom plot of the figure. To summarize, measurements that follow distributions that do not satisfy H_0 , will be more likely to compute smaller \mathbf{P}_k during the hypothesis test.

Unfortunately, the probability as defined by (17) - (20) cannot be computed since \mathbf{e}_j is unknown. For this reason, a candidate of the true error, $\hat{\mathbf{e}}_j$, is used in its place. The candidate is carefully constructed to be,

$$\hat{\mathbf{e}}_j = \begin{cases} \bar{\mathbf{e}}_j & \|\bar{\mathbf{e}}_j\| \leq \eta \\ \frac{\bar{\mathbf{e}}_j}{\|\bar{\mathbf{e}}_j\|} \eta & \text{o.w.} \end{cases} \quad (21)$$

This ensures that the probability computed using the candidate, $\hat{\mathbf{P}}_k$, satisfies the inequality $\mathbf{P}_k \leq \hat{\mathbf{P}}_k$; therefore, a rejection based off $\hat{\mathbf{e}}_j$ will also cause a rejection based on \mathbf{e}_j .

Using $\hat{\mathbf{P}}_k$ the rejection criteria in (16) can be reformulated for easier interpretation of the results. This

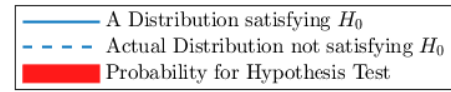


Fig. 1. Graphic depiction of probability that may reject H_0 , computed from a measurement that follows a distribution that does not satisfy H_0 . The top figures shows a less likely scenario, where the measurement occurs further away from its true value. The bottom figure shows a more likely scenario, where the measurement occurs closer to its true value.

reformulation is,

$$\begin{cases} \hat{\mathbf{P}}_k^*(\bar{S}_k|H_0) > \alpha & \text{Reject } H_0 \\ \hat{\mathbf{P}}_k^*(\bar{S}_k|H_0) \leq \alpha & \text{Fail to reject } H_0 \end{cases} \quad (22)$$

where $\hat{\mathbf{P}}_k^*$ is referred to as the rejection probability. This probability is defined as,

$$\hat{\mathbf{P}}_k^*(\bar{S}_k|H_0) = 1 - \hat{\mathbf{P}}_k(\bar{S}_k|H_0); \quad (23)$$

where $\hat{\mathbf{P}}_k$ is computed by replacing \mathbf{e}_j with $\hat{\mathbf{e}}_j$ in (17) - (20). The rejection probability, $\hat{\mathbf{P}}_k^*$, which is the opposite of $\hat{\mathbf{P}}_k$, will be large for cases where rejection should occur, making the results more easy to interpret.

As mentioned previously, the probabilities and inequalities in (15) - (23) are conducted at every iteration of the MI algorithm, but only after a new measurement has been acquired. The results from the hypothesis test conducted at the previous iteration are used to decide which operating point the new measurement should be sampled at. If $\hat{\mathbf{P}}_{k-1}^* \leq \alpha$, then the new measurement will be sampled at an unexplored operating point. If $\hat{\mathbf{P}}_{k-1}^* > \alpha$, then the new measurement will be resampled at a previously explored operating point which had the worst measured error, i.e., the smallest value of $\hat{\mathbf{p}}_{k-1}^*$. When multiple measurements are sampled at the same operating point, these measurements are averaged and the average is used to replace the current (averaged) measurement in \bar{S}_k . Additionally, the covariance matrix at this operating point is also modified to reflect the variance of the averaged measurement. That is,

$$\bar{\Sigma}_j = \frac{1}{\sqrt{r_k}} \Sigma_j \quad (24)$$

where r_k is the number of resampled measurements and Σ_j is the covariance matrix for a single measurement, defined in (9).

The MI algorithm, as defined by (15) - (24), will iterate indefinitely if $\|\bar{\mathbf{e}}_j\| > \eta$ and enough resampled

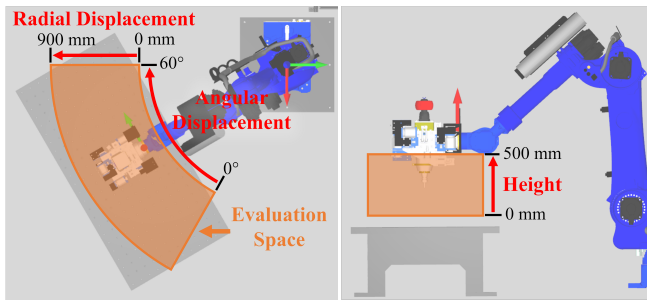


Fig. 2. A graphic depiction of the evaluation space used for assessing kinematic model accuracy in the simulation.

measurements have been collected at this operating point such that $\dot{P}_k^* > \alpha$. Thus, a practical limitation on resampling is necessary to deploy the algorithm on a real system. In [13], a measurement strategy was presented that balances the concept of exploring the evaluation space and performing an adequate resampling at "bad" operating points. In practice, this measurement strategy is preferred; however, in this work the algorithm will be terminated after a predetermined number of measurements (both at new and old operating points) have been collected. This is done to provide a more equitable comparison between MI and the other methods presented in Section II-B.

IV. ACCURACY ASSESSMENT SIMULATION

A. Evaluation Space and Kinematic Models

The industrial robot investigated in this work is used for both machining and grinding; therefore, the accuracy of its kinematic model will be assessed in an evaluation space over the machining table in which those processes are performed. For simplicity, the evaluation space, seen in Fig. 2, was selected to follow a cylindrical coordinate system, having a total angular displacement of 0° to 60° , a total radial displacement of 0 to 900 mm and a height of 500 mm.

Before evaluating the accuracy assessment methods, it was desired to better understand the potential complexities of a kinematic model's error within the evaluation space. Since an unrealistic number of measurements would be required to create a dense map of these errors, the robot's actual kinematics were simulated using a VEC model and a DH model was used as the kinematic model for accuracy assessment.

The VEC model was fitted to the physical robot over a space that encapsulates the evaluation space. The residual errors between the VEC model and the physical robot can be seen in Table I. To emulate conventional practices for kinematic calibration, the DH model was "calibrated" to the VEC model using a least-squares best fit over the same points used to identify the VEC model. The fitted DH parameters can be seen in Table II, and the residual error between the fitted DH model and VEC model is shown in Table III.

TABLE I
Residual Error Between VEC Model and Actual Robot

	Mean	Max	99% Gamma
Position (mm)	0.705	1.635	1.731
Orientation (mrad)	1.487	3.149	3.248

TABLE II
Best Fit DH Parameters

θ (rad)	d (mm)	a (mm)	α (rad)
$q_1 + 0.001$	-1.157	323.453	-1.573
$q_2 - 1.568$	-0.898	1151.356	3.140
$q_3 + 0.001$	0.066	299.501	-1.571
$q_4 - 3.142$	-1226.154	0.029	-1.571
$q_5 - 3.114$	-0.104	0.026	-1.571
$q_6 + 0.002$	-224.698	0.028	3.142

TABLE III
Residual Error Between DH Model and VEC Model

	Mean	Max	99% Gamma
Position (mm)	0.405	0.847	0.918
Orientation (mrad)	1.336	2.851	2.866

B. Analysis of True Accuracy over Evaluation Space

The error of the calibrated DH model will change throughout the evaluation space due to a variety of factors, such as, the position and orientation of its end effector and its wrist configuration. Thus, at several radial slices of the evaluation space, a grid of Tool Center Point (TCP) positions was generated, and at each position in the grid the error was calculated for a combination of $\pm 45^\circ$ rotations about X, Y, and Z axis of the robot's flange and its two wrist configurations. These orientations and wrist configurations were selected to encapsulate potential use cases of the robot when performing a machining or grinding operation in the evaluation space.

Once all errors were calculated, the maximum position and orientation error magnitudes over all evaluated orientations and wrist configurations were found at each TCP position in the grid. To better interpret the results, the error grid for each radial slice is depicted by the heat maps shown in Fig. 3. These heat maps, when compared to statistical metrics of the error distribution, provide a more detailed and intuitive description of how the DH model's error can change within the evaluation space.

As shown in Fig. 3, the maximum position and orientation error magnitudes are highly nonlinear and their largest values change location in each radial slice. More specifically, the largest position errors occur toward the outer reaches of the space at the radial slices of 0° and 60° ; however, between 15° and 45° the largest position error is much less predictable. The largest orientation errors are even more unpredictable and tend to be located near the various discontinuities observed in the heat

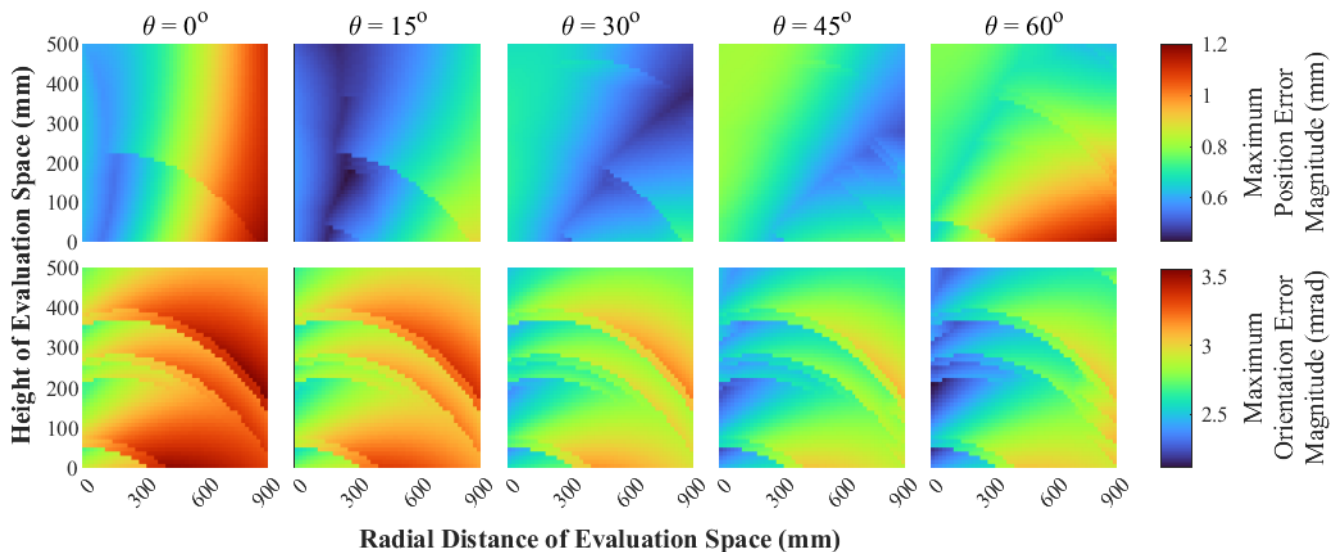


Fig. 3. Heat maps of maximum position and orientation error magnitudes over radial slices of evaluation space.

maps. At first glance, these discontinuities may appear due to simulation errors; however, they are primarily caused by the change in the robot's wrist configuration, which is likely to occur on the physical system. Overall, these results demonstrate the complexity of the DH model's errors in the evaluation space, which can make the assessment of its accuracy much more difficult conduct.

C. Evaluation of Accuracy Assessment Methods

A simulation was constructed to evaluate the efficacy of the presented accuracy assessment methods when deployed on a complex evaluation space, as the one shown in Fig. 3, and to investigate their sensitivity to different magnitudes of noise. Other accuracy assessment methods may exist; however, the ones evaluated in this work were selected primarily for their conventional nature and intended use case. Additionally, to maintain conciseness this simulation will only focus on the assessment of the DH model's position accuracy.

To construct the simulation, a set of 200 joint configurations were selected from a quasi-random sequence spanning the workspace. Once the subset was selected, measurements were simulated by generating white noise with a normally distributed random number generator and summing the noise to the model's true position error. The simulated noise was generated for six different magnitudes, ranging from $50 \mu\text{m}$ to $300 \mu\text{m}$. This range was selected to encompass the noise magnitudes of many spatial measurement devices such as laser trackers and metrology cameras.

The Maximum Measurement and Outlier Rejection methods were deployed directly to the sets of 200 error measurements for each noise magnitude. The MI algorithm was also applied to these measurement sets; however, since the algorithm produces a binary output,

i.e., reject or fail to reject, the smallest accuracy bound, η , that did not reject the model with a selected confidence interval of $\alpha = 60\%$, was used. Additionally, if the criteria for re-measurement occurred, i.e., $P_k^* > \alpha$, a new error measurement at the same joint configuration was generated and the mean measurement was calculated as described in Section III. This was permitted for a total of 200 iterations of the algorithm to facilitate a fair comparison between it and the other accuracy assessment methods, which could not perform re-measurement.

To ensure that the results were indicative of the accuracy assessment methods and not a specific instance of the randomly generated noise, the simulation was repeated 100 times for each noise magnitude and the results of all methods were averaged to show their general trends. Additionally, the largest and smallest assessed accuracies of each method were recorded to quantify their range of variability between the repeated simulations. These results are shown in Fig. 4, where the positive and negative side of the error bars show the largest and smallest assessed accuracies over all 100 simulations.

As shown in Fig. 4, all methods are affected by the magnitude of the noise; however, some are more significantly affected than others. The Maximum Measurement method tended to produced an assessed accuracy that was much larger than the model's true accuracy in the subset. On average, the assessed accuracy increased with the magnitude of the noise and the range of variations between simulations were skewed to be larger than the average value. Similarly, the assessed accuracy produced by the Outlier Rejection method also increased with the magnitude of the noise; however, at lower noise magnitudes it tended to be smaller than the true accuracy. Although this method does both under and over assess the robot's position accuracy, it does have a much smaller and more evenly distributed

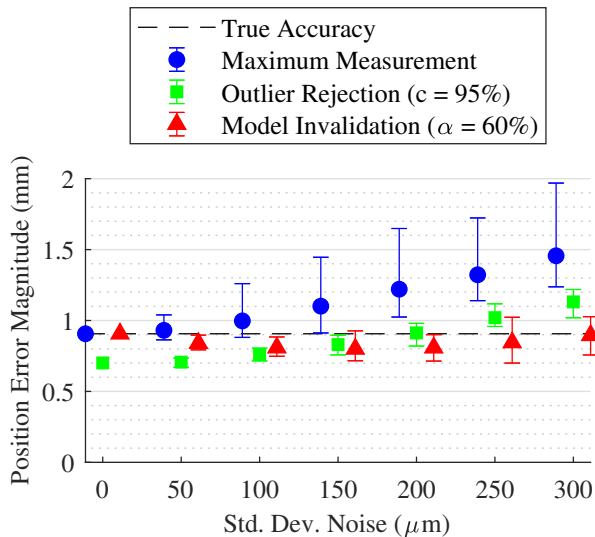


Fig. 4. Comparison of accuracy assessment methods at different noise magnitudes for 100 separate simulations. Error bars of Maximum Measurement and Outlier Rejection show largest and smallest determined accuracy over all simulations. Model invalidation shows smallest η that failed to reject the model over all simulations.

range of variations between simulations. Overall, these observations demonstrate the noise sensitivity of the conventional accuracy assessment methods.

Differing from the conventional methods, the MI algorithm produced an assessed accuracy that was much more consistent across all noise magnitudes than the other methods. The range of variations between simulations were evenly distributed around the mean; however, they tended to be larger than the Outlier Rejection method at larger noise magnitudes. To better understand how the MI algorithm operates and why it tended to have more consistent results across the evaluated noise magnitudes, the iteration history of measurements and rejection probabilities were investigated for two simulations conducted at noise magnitudes of $300 \mu\text{m}$.

Fig. 5 shows the iteration history of the measurements and rejection probabilities of a simulation that invalidated the DH model. As shown in the top plot of the figure, the accuracy bound ($\eta = 0.834 \mu\text{m}$) was smaller than the true accuracy ($0.906 \mu\text{m}$); however, large measurement noise produced measurements (blue and orange dots) that were often larger than the robot's true accuracy and consequently larger than η . This caused the rejection probability (purple dots) to quickly increase and approach α after only 20 iterations. Eventually, the algorithm acquired measurements that resulted in a rejection probability above α and resampling occurred. The orange dots show the averaged measurements after resampling and the yellow dots show their corresponding rejection probabilities.

The true errors between 50 to 176 iterations were primarily below η with the exception of one true error which was only slightly above the bound. Thus, the resampling that occurred in this region was primarily due

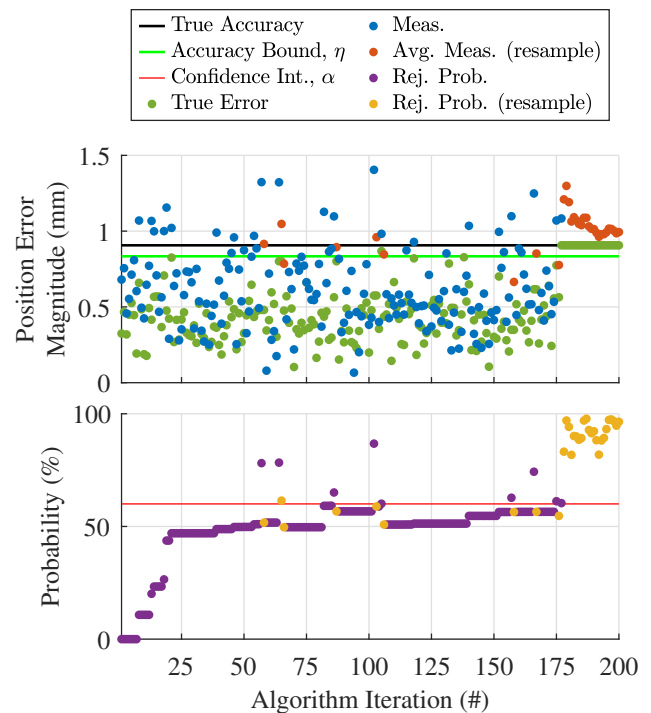


Fig. 5. Iteration history of simulation that invalidated the DH model, where the noise magnitude was $300 \mu\text{m}$, $\eta = 0.834 \mu\text{m}$ and $\alpha = 60\%$.

to measurement noise. After 176 iterations, the true error at the evaluated joint configuration was the largest error in the set and consequently the error that defined the true accuracy of the model. Measurements of this error quickly produced a rejection probability above α and further resampling produced an average measurement that continued to remain above the bound. Thus, the algorithm did not explore new joint configurations and continued resampling the "bad" joint configuration for the remainder of its allocated measurements. Since the rejection probability was greater than α at the final iteration, the model was correctly rejected. These results show the general working principle of the MI algorithm and more specifically its ability to detect and prioritize resampling of true errors at "bad" joint configurations of the kinematic model.

It should be noted that when η was increased to $0.844 \mu\text{m}$ and the simulation was repeated, the rejection probability calculated at the first measurement of the largest true error did not exceed α , no resampling occurred, and the model was not rejected. Thus, proper selection of η and α are important to achieve accurate results.

After investigating the MI algorithm's iteration history for a simulation that invalidated the DH model, the iteration history of a different simulation was investigated where the DH model could not be invalidated. This simulation's iteration history is shown in Fig. 6. As shown in the figure, the accuracy bound ($\eta = 1.023 \mu\text{m}$) was larger than the true accuracy ($0.906 \mu\text{m}$),

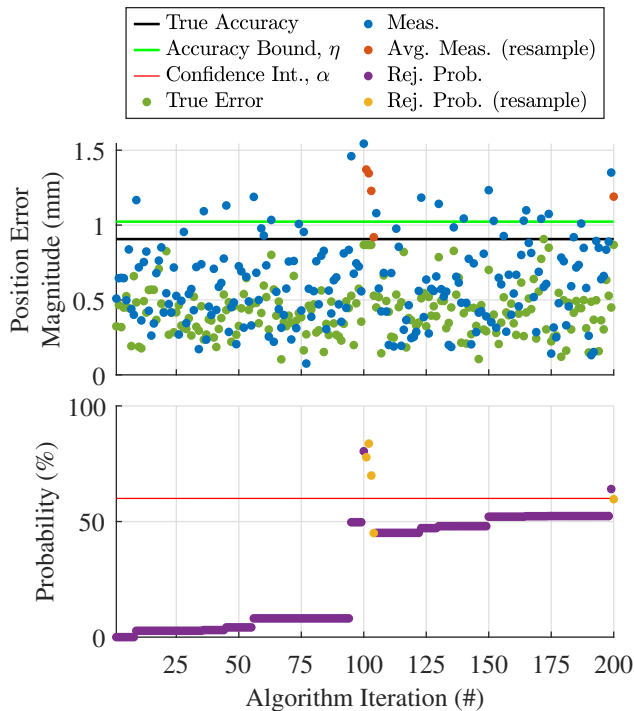


Fig. 6. Iteration history of simulation that could not invalidate the DH model, where noise magnitude was $300\ \mu\text{m}$, $\eta = 1.023$ and $\alpha = 60\%$.

and all true errors were below η . Thus, the resampling that occurred was due solely to measurement noise. In particular, the true errors measured at the 100th and 198th iteration were resampled several times before the rejection probability went back below α and the DH model could not be invalidated. This simulation provides a good demonstration of how the MI algorithm operates well in noisy environments.

V. CONCLUSIONS

In this work, three methods used to assess the accuracy of an industrial robot's kinematic model were presented and evaluated through simulation. Two of these methods, i.e., Maximum Measurement and Outlier Rejection, are commonly used for their simplicity; however, these methods were found to be sensitive to uncertainties in the measurement of the kinematic model's output, such as, noise or the system's repeatability. The third and proposed method, referred to as Model Invalidation, was found to be less sensitive to these uncertainties due to its stochastic framework, which can detect and prioritize the resampling of potentially "bad" operating points that produced large model errors. Its ability to detect and prioritize "bad" operating points is not only beneficial for large measurement uncertainties, but also for this specific application of assessing the accuracy of a robot's kinematic model, which was shown to exhibit complex error behavior throughout the robot's workspace.

While this method works well in the presence of noise, it can only operate as expected if its parameters, η and α , are selected appropriately. Additionally, the MI method

in its current form, is only valid for uncoupled covariance matrices, i.e., covariance matrices following the form of (10). Future work should develop more rigorous methods for selecting η and α , and investigate ways to modify the MI algorithm's probability calculations so that covariance matrices with cross terms can be used. This will allow the method to be applied to a more general set of systems.

DISCLAIMER

Certain commercial entities, equipment, or materials may be identified in this document in order to illustrate a point or concept. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology (NIST), nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

REFERENCES

- [1] L. Cen and S. Melkote, "Effect of robot dynamics on the machining forces in robotic milling," *Procedia Manufacturing*, vol. 10, pp. 486–496, 2017.
- [2] P. Olubodun, M. Woodside, and D. Bristow, "Procedures and performance of a robotic machining system with metrology-in-the-loop feedback control," *Manufacturing Letters*, vol. 36, pp. 361–370, 2023.
- [3] M. Sklar, "Geometric calibration of industrial manipulators by circle point analysis," 1989, pp. 178–202.
- [4] D. Bennett, S. Member, and J. Hollerbach, "Autonomous calibration of single-loop closed kinematic chains formed by manipulators with passive endpoint constraints," *IEEE Transactions on Robotics and Automation*, vol. 7, no. 5, pp. 597–606, 1991.
- [5] C. Wampler, J. Hollerbach, and T. Arai, "An implicit loop method for kinematic calibration and its application to closed-chain mechanisms," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 5, pp. 710–724, 1995.
- [6] L. Ma, P. Bazzoli, P. M. Sammons, R. G. Landers, and D. A. Bristow, "Modeling and Calibration of High-Order Joint-Dependent Kinematic Errors for Industrial Robots," *Robotics and Computer-Integrated Manufacturing*, vol. 50, pp. 153–167, 2018.
- [7] M. Alam, S. Ibaraki, K. Fukuda, S. Morita, H. Usuki, N. Otsuki, and H. Yoshioka, "Inclusion of bidirectional angular positioning deviations in the kinematic model of a six-dof articulated robot for static volumetric error compensation," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 6, pp. 4339–4349, 2022.
- [8] S. Ibaraki and W. Knapp, "Indirect measurement of volumetric accuracy for three-axis and five-axis machine tools: A review," *International Journal of Automation Technology*, vol. 6, no. 2, pp. 110–124, 2012.
- [9] ISO 9283, "Manipulating industrial robots-performance criterion and related test methods," International Organization for Standardization, 1998.
- [10] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot Modeling and Control*. Wiley & Sons, Inc., 2006.
- [11] J. Denavit and R. S. Hartenberg, "A Kinematic Notation for Lower-Pair Mechanisms Based on Matrices," *Journal of Applied Mechanics*, vol. 22, no. 2, pp. 215–221, 1955.
- [12] L. Ma, "Volumetric error compensation for industrial robots and machine tools," Ph.D. dissertation, Missouri University of Science and Technology, 2019.
- [13] P. Bazzoli, "Model and control for precision robotic machining," Ph.D. dissertation, Missouri University of Science and Technology, 2023.
- [14] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1, pp. 49–55, 1936.