Journal of Manufacturing Systems

Deep active-learning based model-synchronization of digital manufacturing stations using human-in-the-loop simulation

--Manuscript Draft--

Manuscript Number:	SMEJMS-D-22-01279
Article Type:	Full Length Article
Keywords:	Human-centric manufacturing; Digital twin; Virtual reality (VR); AI and machine learning; Virtual manufacturing; Digital transformation
Corresponding Author:	Namhun Kim, Ph.D. Ulsan National Institute of Science and Technology Ulsan, KOREA, REPUBLIC OF
First Author:	Clint Alex Steed, M.S.
Order of Authors:	Clint Alex Steed, M.S.
	Namhun Kim, Ph.D.
Abstract:	The effective and accurate modeling of human performance is one of the key technologies in virtual and smart manufacturing systems. In this paper, we propose a framework to simplify human assembly task modeling to obtain informative and significant data of human behavioral patterns with less trials. This is achieved by using the virtual reality (VR) to prototype data-acquisition systems for human manufacturing tasks. An active learning model is employed to reduce the number of experiments conducted by intelligently selecting the experimental conditions that will yield the most informative result. The resulting system requires less experimental trials and is automated. In VR experiments involving throughput rate, a deep active learning model significantly reduces the amount of data required, thereby speeding up the experiment and modeling process. The proposed method can quickly generate human performance models in virtual systems and improve experiment scalability. Previous data from a similar assembly task may be required for parameter tuning and design choices.
Opposed Reviewers:	

Dr. Lui Wang, Editor-in-Chief

Journal of Manufacturing Systems (JMS)

November 4, 2022

Dear Dr. Wang,

We wish to submit an original research article, titled "Deep active-learning based model-synchronization of digital manufacturing stations using human-in-the-loop simulation."

In this paper, we use a digital twin to prototype a human assembly station with the aim of developing a model to predict the station's throughput rate. The number of experimental trails required is reduced by having a deep learning model actively select the experimental conditions at runtime. Thus, this framework can (1) reduce the cost of workstation development through virtual prototyping and (2) reduce risk by creating performance models before investing in hardware. We believe this work provides a practical upgrade path to including humans into Industry 5.0 when applied to retrofitting existing workstations.

The direction of this work was inspired by the recent special issues in JMS (Human-centric Smart Manufacturing and Digital Twin towards Smart Manufacturing and Industry 4.0) among other things. Therefore, we believe the journal's audience will find this work interesting. We attempt to describe a holistic solution by implementing a vertical slice, mixing (1) high-level concepts like Digital Twin, (2) low-level concepts like uncertainty-based sampling, and (3) human experiments, all while trying not to alienate readers by supplying sufficient background information. Naturally, it is challenging to prepare a brief and good-quality manuscript for multi-disciplinary work like this and we welcome any improvements suggested by the JMS Editorial team.

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere. We have no conflicts of interest to disclose. All authors approve the final article.

Please address all correspondence to nhkim@unist.ac.kr.

Sincerely,

Namhun Kim, Ph.D.

Professor, Department of Mechanical Engineering

Ulsan National Institute of Science and Technology

50, UNIST-gil, Ulsan 44919, Republic of Korea, (+82)-52-217-2715,

nhkim@unist.ac.kr.

Highlights

- A novel framework combining virtual reality and deep active-learning results in rapid performance modeling of human manufacturing processes.
- Automated modeling, design of experiments, and virtual experimental conduction reduces labor and increases scale of experiments.
- Significantly less experimental trials by using an active-learning model.

Deep active-learning based model-synchronization of digital manufacturing stations using human-in-the-loop simulation.

Clint Alex Steed [a], Namhun Kim[*, a]

[a] Department of Mechanical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, 44919

[*] Corresponding author: nhkim@unist.ac.kr

Deep active-learning based model-synchronization of digital manufacturing stations using human-in-the-loop simulation.

Abstract

The effective and accurate modeling of human performance is one of the key technologies in virtual and smart manufacturing systems. In this paper, we propose a framework to simplify human assembly task modeling to obtain informative and significant data of human behavioral patterns with less trials. This is achieved by using the virtual reality (VR) to prototype data-acquisition systems for human manufacturing tasks. An active learning model is employed to reduce the number of experiments conducted by intelligently selecting the experimental conditions that will yield the most informative result. The resulting system requires less experimental trials and is automated. In VR experiments involving throughput rate, a deep active learning model significantly reduces the amount of data required, thereby speeding up the experiment and modeling process. The proposed method can quickly generate human performance models in virtual systems and improve experiment scalability. Previous data from a similar assembly task may be required for parameter tuning and design choices.

Keywords – Human-centric manufacturing; Digital twin; Virtual reality (VR); AI and machine learning; Virtual manufacturing; Digital transformation.

Date: 2022-11-04

1. Introduction

As modern manufacturing systems include physical, data-acquisition, and simulation components, human integration in the manufacturing system implementation has been identified as a key factor impeding adoption [1], [2] ⍰. There has been a desire to move towards human-centric production for years [3]–[5] , but modeling human performance is complex. Recently interest in this area has seen a dramatic increase, with several special issues [6]–[8] dedicated to including humans in manufacturing systems, motivated by an EU report [9] placing human-centric production as a core value of Industry 5.0.

Human workstations are becoming increasingly "smart" in order to improve the productivity and effectivity of operators. [10] identifies Human-centric assembly and Mixed reality as key areas for future development. Developing these stations can be costly as they need to consider ergonomic and cognitive load, while employing sensors, human-machine interfaces, VR/AR systems, etc. Thus, virtual manufacturing and digital twin for human process is an active research topic [11]–[13]. This work investigates prototyping sensor-based human workstations for assembly tasks using Virtual Reality (VR). One issue that arises with virtual human workstation experiments is the cost of human labor in human-in-the-loop (HIL) simulation can be significant, particularly for skilled artisans. Therefore, this work goes further by investigating reducing the amount of human labor required to model the workstation's performance using an active learning model. Combined these two advancements allow rapid prototyping and modelling of manufacturing workstation performance, while reducing risk due to sensor complexity and initial cost, allowing systems designers to gain performance metrics early in the design process.

The structure of this paper is as follows. The literature review (section 2) provides background on active learning, mentioning related work. Section 3 covers the experimental design, where: a VR-simulation generates data, and a data-sampling experiment illustrates data-efficiency. Section 4 presents the theory where sample selecting is formulated as a search problem. Section 5 reports the results of the simulation and shows that utility-based significantly reduces the number of trials. Finally, in Section 6 we discuss and conclude the investigation. The remainder of this chapter serves to briefly introduce the framework and its' components.

1.1 Framework overview

As show in Figure1, the main idea of proposed work is to implement a digital twin prototype before implementing the physical workstation, thereby reducing financial risk, and resulting in a better final implementation. To fully exploit this virtual workstation, we employ deep active learning to reduce the number of experimental trails required to model the station's performance. This results in a smart digital twin of the process, consisting of a virtual workstation and a model for predicting the workstation's performance.

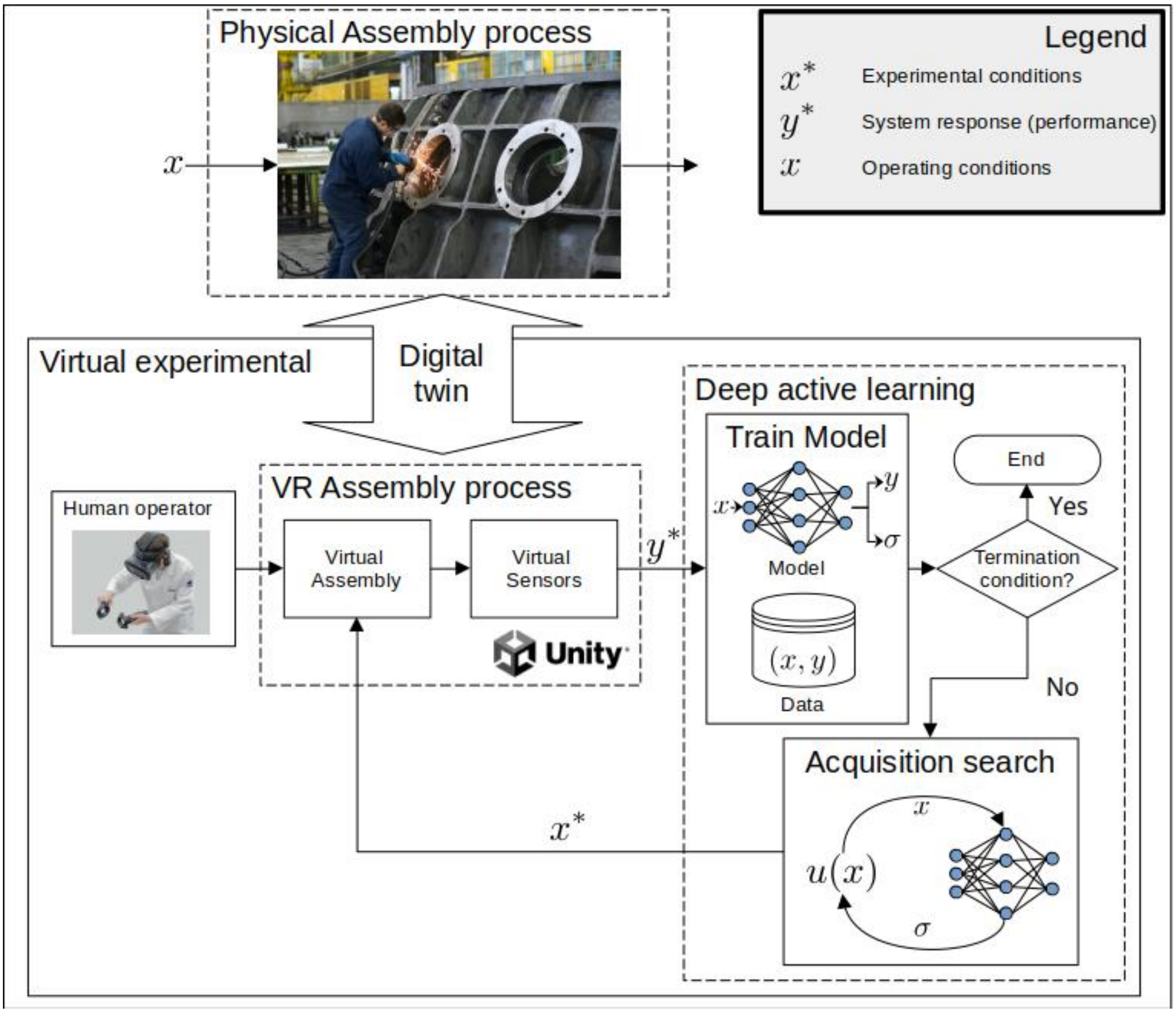


Figure 1 The proposed framework uses a virtual manufacturing process and deep-active learning model to perform HIL simulation. Note that these virtual workstations produce input (operating condition) and output (response/performance) data and therefore require sensors.

The virtual workstation prototype produces data. It records the systems performance (y) and measures and controls the operating conditions (x) using VR and HIL simulation. This allows investigating processes, sensors, and producing data before the physical implementation. The active learning model iteratively designs experiments by intelligently selecting the next trail. The models' objective is to acquire informative samples, reducing the number of trials needed to model the system. Sample acquisition involves selecting the operating conditions (x) and measuring the system response (y). The data is typically acquired though sensors. Once running, the model is trained in every iteration until it satisfies the termination condition, concluding the experiment. The resulting model can be used for prediction.

2. Related Works

Humans are essential for manufacturing systems to remain competitive in environments with mass personalization and creative decision [14]. Human wellbeing was recently noted as a strategic component of sustainable economic growth in manufacturing [15].

2.1 Human data acquisition

Data acquisition for human processes is useful for safety, prediction, and diagnosis [16]. Industries where human operators are static (e.g., long distance drivers and pilots) have seen commercial success in applying sensor-based systems, some examples are sensors placed in chairs, steering wheels, operator facing cameras, etc. [17]–[24].

Manufacturing typically requires operators to work autonomously in unstructured environments and therefore has seen less commercial sensor system adoptions despite attempts. Some work using medical equipment, such as EEG or EKG, is promising but prohibitive due to sensor practicality and cost in unstructured manufacturing environments [25]. Oral swabs can be used to predict human fatigue [26], but biological sampling methods are impractical and invasive. Investment in sensors can be a risk as it is not clear what the data will reveal.

[27] mentions that modeling can address this by providing sample data, but also says these models must be compared against empirical human-in-the-loop data, creating the chicken and egg dilemma. Thus, Virtual reality (VR) has beneficial properties that address many of these issues. Using head mounted displays and controllers we can simulate rich interactions between humans and the manufacturing environment without the need for additional hardware and sensors.

VR has been used extensively for mass development [28] and visualizing and planning manufacturing systems and layouts [29]. Here, the interest is in planning data acquisition systems for human assembly tasks. [30], [31] developed VR experimental frameworks for human trails. [31] provided a framework for planning experimental trails based on the selected factors, with the additional functionality of conducting remote experiments by storing data on a remote database. [30] developed a user-friendly framework for experimental design, requiring little knowledge of computer programming. Where both of this work simplifies experimental design, this work investigates outsourcing this responsibility to an active model.

2.2 Human performance modeling and Wright's learning curve

Human performance models (HPM) predict human behavior in a task or system. Since human behavior is complex, simplified models are developed based on specific requirements. Historically, proposed models were used to identify factors affecting performance, thereby enabling ergonomists to design systems that optimize human performance [32]. Computational models like ACT-R [33] have shown to be successful in predicting human performance but require years of programming experience, making analytical models more common. [34] identifies machine learning as an important technology for future human smart manufacturing systems. [35] used K-nearest neighbors to classify tasks by skill level requirement, allowing hiring managers to select operators with the appropriate skills. [36] used an Artificial Neural Network (ANN) to model the relationship between the work environment, worker personalities, and their subsequent performance, but did not provide a quantifiable measure of the accuracy. Assembly models are typically more concerned with modeling sequential steps in order to plan for robotic collaboration [37] or standardize operating procedures.

Recently, the human factors methods community has shown interest to incorporate technologies like artificial intelligence and big data due to the changing nature of work [8]. [38] highlights automatic extraction of practical information from heterogeneous data (a capability of deep learning) and real-time data collection as key enablers of including digital human models into manufacturing Cyber Physical Systems (CPSs).

Wright's law predicts the falling cost due to cumulative production [39] using logarithmic/exponential functions ($y = a\,x^b$). In [40] Wright's law was found to be more accurate than Moore's law at predicting the cost of silicon. In the private sector, it has also successfully been used to predict the cost of lithium-ion technologies for investment [41], [42]. Wright learning has been applying to human production schedules [39] , where it predicts the reduction in a task duration. It later received numerous additions/modifications considering work induced fatigue, and rest schedules [4], [43]–[45]. To the author's knowledge the variance of Wright's law was not previously recorded in literature and was observed in this work.

2.3 Intelligent sampling techniques

In this section we mention previous works investigating data efficient experiments and modeling. These typically work by selecting the next experimental sample data-point intelligently to reduce the number of experimental trials required.

2.3.1 Design of experiments

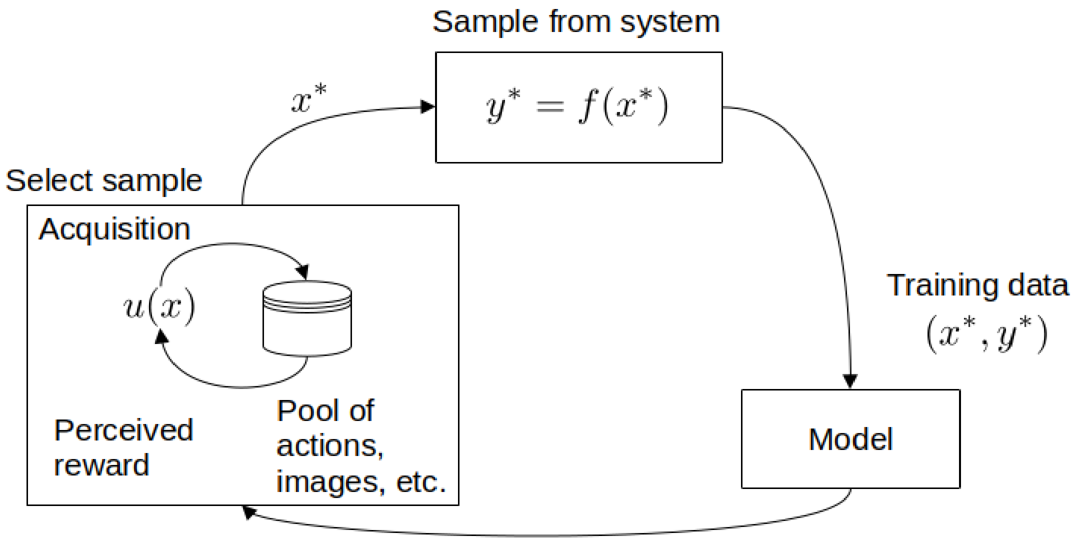
Where experimental design is concerned with selection and blocking of factor-level combinations, Design of experiments (DOE) builds on this by including a protocol for analyzing the data and selecting the next experimental sample [46]. DOE methods like factorial design and response surface are well known and have their roots in statistics. These methods typically use linear regression models and are, therefore, limited to situations with few factors, few factor-levels, and constant random error. Recently, iterative design [47] has shown promising results in optimizing the process operating parameters.

2.3.2 Reinforcement learning

The methods in this section can be classified as reinforcement learning. However, the emphasis is usually on stringing together a sequence of decisions to escape non-trivial local minima in complex environments. For example, [48] investigated a tightly related topic, "The design of experiment using reinforcement learning" and in one example shows a car escaping a bowl by driving around to build up momentum and catapult out. Here the aim is simply to acquire the next informative sample. However, reinforcement learning is more general and not often related to experimental. Some unrelated work uses reinforcement learning to search for neural-network architecture. This is relevant here because as the data set grows larger (1) the model may need to change as more data becomes available, and (2) the model is retrained in every iteration so selecting an architecture with shorter training times is advantages. Where [49] highlights the ability to continue training as a desirable feature of some algorithms, [50]–[53] show that reinforcement learning can be used to optimize the model architecture at runtime.

2.3.3 Pool-based active learning

Active learning refers to having the model choose the next action. In pool-based active learning, samples from an unlabeled pool are labeled for classification. The motivation resembles this work, since gathering unlabeled images from the internet is easy, but the act of labeling the data is "expensive" as it requires human effort. The goal is to maximize a model's performance with the fewest labeled samples possible [49] . Two pools of data are used; the training data pool consists of labeled data, and a data-bank pool consists of unlabeled data. The unlabeled pool is searched, selecting a sample that will most improve the model's performance. The acquisition function returns the samples usefulness. A (human) oracle labels the selected sample.



	Acquisition	Sample
Reinforcement learning	State + Reward action selection	Perform action in environment
Design of experiments	Regression, optimization, etc.	Conduct experiment
Pool-based Active learning (classification)	Search through pool for sample	Human oracle labels

Figure 2 The related works reinforcement learning, design of experiments, and active learning are similar. In the discrete case the pool stepped through before selecting samples, in the real-valued case optimization/search is performed.

As shown in Figure 2, pool-based active learning for classification problems is by far more popular than experimental intervention, with modern literature equating active learning with classification[49], [50], [54]. There is comparatively less work considering regression [53]. One should note that regression problems can fit into pool-based active learning, where labeling data refers to populating real-valued targets (y) and the unlabeled databank would be an infinite pool in the input space (x). Hence, the distinction between this work and pool-based active learning is that we formulate the acquisition of the next sample as a search/optimization problem. Instead of searching through a batch of pre-existing data to be classified, we search a surrogate space that describes the samples informativeness.

2.3.4 System identification

System identification uses statistical methods to model dynamical systems [55] and are therefore temporal in nature. Naturally, experimental design has also been used here to reduce required data. Dynamic Data Driven Application Systems techniques [56] strongly resemble this work but emphasizes sensor characteristics and control. Here we do not assume temporal systems and ignore sensor characteristics due to the high accuracy of VR systems. We do not ignore system noise.

All these frameworks are similar, sometimes differing only in terminology. Yet, they emphasize different things. Design of experiments offers runtime optimization of processes by adjusting the system response, reinforcement learning investigates how agents can escape local minima, and active learning tries to reduce the amount of data required by selecting data intelligently. We use the terminology from active learning because our emphasis is on data reduction.

2.4 Acquisition strategies

In pool based active classification, an acquisition function returns a samples' usefulness and is used for selecting informative samples from the pool. Optimal experimental design [57] formalizes informative samples for multi-variate response, by selecting the next sample that equivalently minimizes the variance of estimators and contains the most information-content. For example, A-optimality selects samples with the maximizes average information across dimensions, where C-optimality selects samples with the single maximum information.

The acquisition functions balance several concerns when selecting the next sample, these concerns are equally valid in classification and real-valued regression problems. One acquisition concern is selecting the sample with the highest expected model change [58]. Since the value of the sample selected is not known a common heuristic is the variance computed from query by committee [59]. Another concern is diversity, as samples congregated around a small area will likely not be indicative of the general behavior. Evidential sampling [60] considers the geometric position of samples as a measure of uncertainty. [53] applies this to regression prediction of driver drowsiness by selecting the next sample based on the centroid of the previous samples. [61] proposes passive sampling where the acquisition function is based on a separate (non-learning) model, not requiring re-training at each iteration, and achieving more stable performance by avoiding fluctuations from selecting samples with the highest regression errors.

2.5 Research objectives

This work contributes to the next generation of human-centric cyber physical systems [12], [16], [62], [63] by making experimentation and implementation of process modeling practical. Naturally, it also synergizes well with the digital twin paradigm by simulating and modeling workstation performance.

2.6 Preliminaries

This section mentions the preliminaries necessary for understanding this work.

2.6.1 Model training and uncertainty

For the effective sample selection the model needs to predict the mean response (y) and epistemic uncertainty (σ), where epistemic uncertainty quantifies sparsely represented data [64] .

$$(y, \sigma) = g(x)$$

Several techniques estimate uncertainty in regression models, Query by committee (ensembles) [59], [65] is the de facto and was used here. Other methods include Bayesian neural-networks [66], [67] , bagging/bootstrap [68] , dropout-techniques [69] and Gaussian process regression [66], [70].

2.6.2 Ensemble models

Ensembles are based on Query by committee [59] where multiple models predict the same value. The intuition is that if the data sufficiently describes the behavior, all models will predict the same outcome, the magnitude of the discrepancy in the prediction can be interpreted as uncertainty. Figure 3 illustrates this and indicates that sampling at regions of high uncertainty can benefit the model. The results are then combined to obtain the mean and variance [71], [72] See [73] for a recent review.

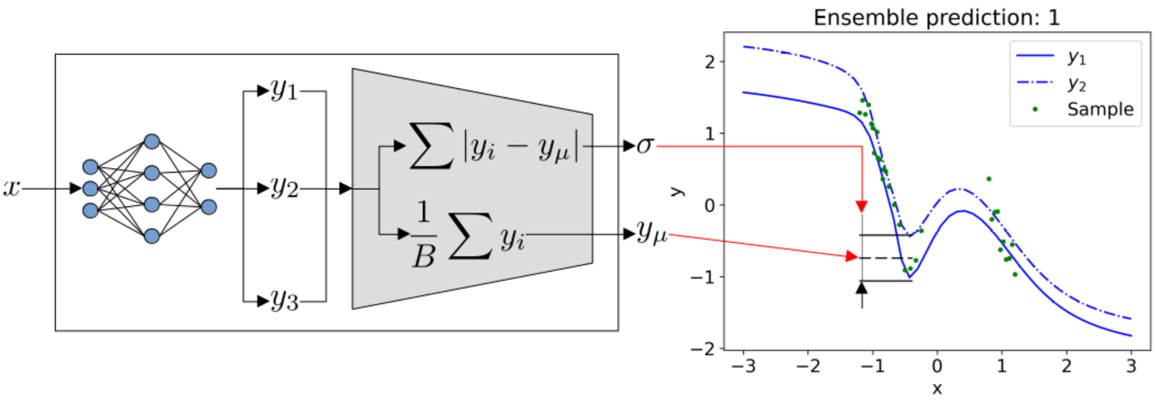


Figure 3 Ensemble uncertainty estimation. Here the ensembles attempt to predict the same value. The uncertainty is the distance between predictions. The mean prediction is the average of predictions.

The mean and uncertainty can be determined using the equations below, where B is the number of ensembles. The predicted-mean (μ) being the average of the ensemble-predictions (g_i), and the uncertainty being the distance between the mean and the ensemble-prediction.

$$\mu(x) = \frac{1}{B}\sum g_i(x)$$

$$\sigma^2(x) = \frac{1}{B-1}\sum \mu(x) - g_i(x)$$

2.6.3 Loss function

The loss function must now quantify the mean error and uncertainty. The well-known negative log likelihood [74] loss function was used with minor modifications. Gaussian noise is assumed.

$$L = \frac{n}{2}\left[log(\sigma^2) + \left|\frac{y_\mu - y_t}{\sigma^2}\right|^2\right]$$

3. Proposed Methodology

The aim of the experiment was to show that active sampling will require less data than random sampling. To this end, VR simulations were conducted to gather data where human operators completed a series of assembly tasks. Next, a sampling experiment compares active and random sampling using the acquired data. Ideally, one would conduct one simulation-experiment using active sampling and another using random sampling. Instead, the data is reused. Wright-learning was selected as the case study for the experimental task, where the goal is predicting the duration of an assembly task.

3.1 Virtual manufacturing simulation design

The experiment involved human operators performing common manufacturing assembly tasks. The task duration and other data was also recorded but not used here. As shown in Figure 3, subjects are required to place components in specific locations and an audio-visual prompt informs them whether the task was completed correctly. The task is performed for several repetitions. The task sequence culminates in the operator selecting components from magazines and welding them together according to a diagram.

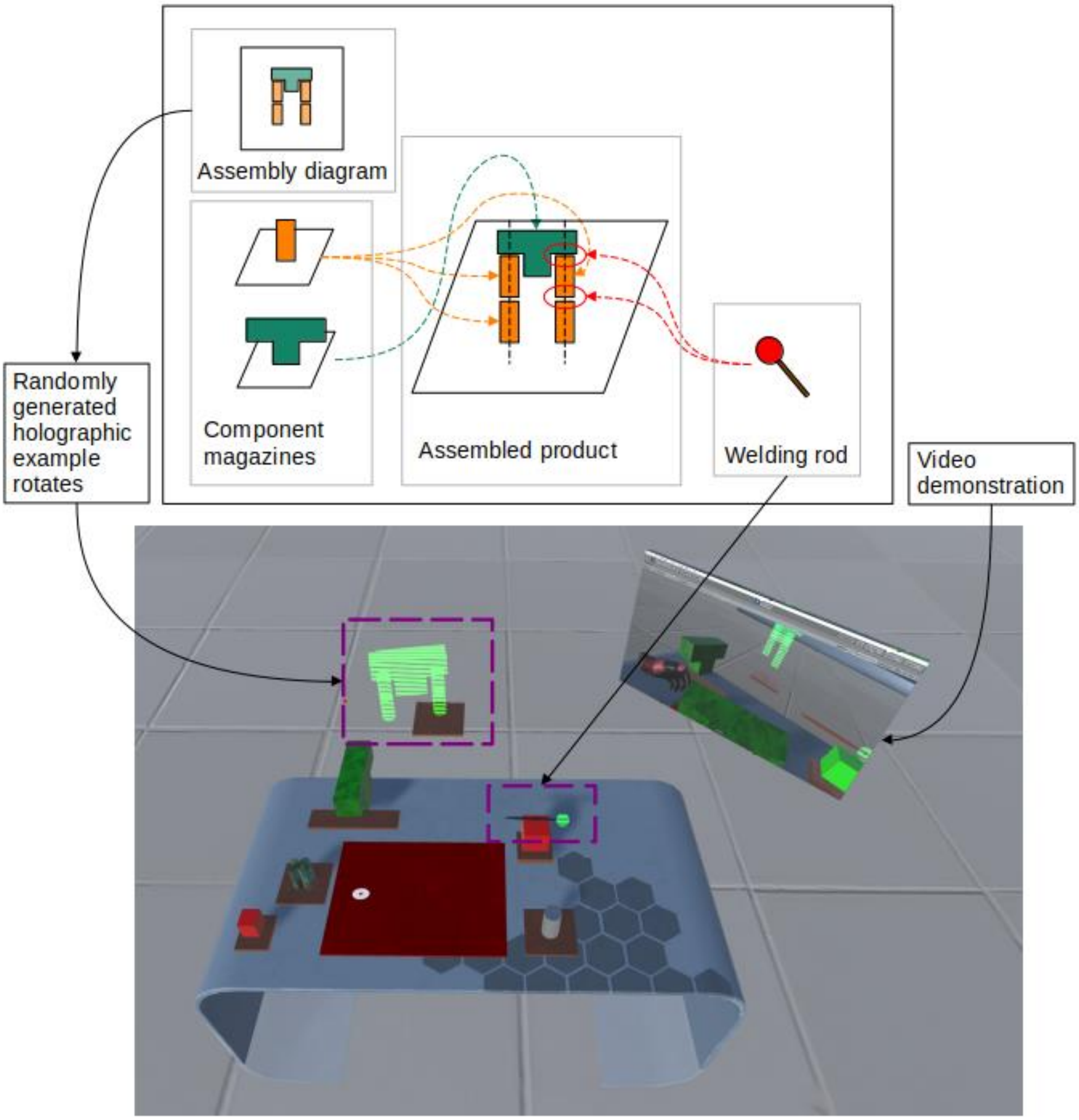


Figure 4 The final assembly task joining is shown schematically (above) and in VR (below). The operator selects components from magazines and welds them together. The final assembly is illustrated via a hologram.

The set of tasks are defined as 1) placement, 2) stacking, 3) sorting, and 4) joining. 'Placement' simulated pick and place operations. 'Stacking' involved users placing one component on the next, simulating accumulated errors. In 'sorting,' subjects are given a randomly generated schematic, where they must select the correct components from magazines and place them in the illustrated position. 'Joining' has the subjects weld components into a predefined assembly.

3.1.1 Simulated task description

Placement simulates repeatable pick-and-place operations and was scheduled first due to its simplicity. Stacking involved users placing one component on the next, simulating accumulated errors. These are considered repeatable actions. In sorting subjects are given a randomly generated schematic, where they must select the correct components from magazines and place them in the illustrated position. Joining has the subjects weld components into an assembly from a small library. These two tasks simulate cognitive load.

3.1.2 Experimental procedure

In a calibration phase the workstation table-height was adjusted based on the individual's limb length. Individuals assumed a series of poses allowing the calculation of the limb length. This blocked ergonomic factors between subjects which was particularly evident for taller individuals. This hints that VR can be used to design ergonomic workstations without the hardware but is outside the scope of this work and so was not explored further. Similarly, during development, the workstation layout was configured to place components within comfortable reach of operators. In trials for placement, stacking, and sorting (tasks 1-3) subjects completed ten repetitions. In task 4, subjects only assembled five components due to the duration and challenge of the final task. Errors did not count toward the repetitions in a trial, so if two errors occurred in a ten-repetition trial, the two error reps are repeated, totaling twelve. Subjects completed up to seven trials over the span of five days. Trials were spaced randomly, with a minimum of 5-hour space between trials. Subjects were between the age of 20-30[1]. There were eleven males and one female who took part in the study[2]. Not all subjects completed six sequential trails. Table A1 shows the frequency of trails completed. All recipients had little previous exposure to using VR. The tasks were not explicitly explained to the subject, save for a video and an introductory tutorial. No monetary compensation was given for this experiment. The VR environment was developed in Unity3D® using the SteamVR plugin, custom C# code, and an HTC-Vive Cosmos head mounted display with controllers.

3.2 Sampling data experiment

The main objective is to reduce the number of experiments conducted. To do this we compare random sampling and active sampling. In active sampling, the model selects the next experiment sample x^* based on an acquisition function, in random sampling x^* is selected randomly within the experimental range.

---

1 There tends to be a substantial mature population in manufacturing, not represented in this study.

2 This is not an unusual distribution of sexes in manufacturing environments.

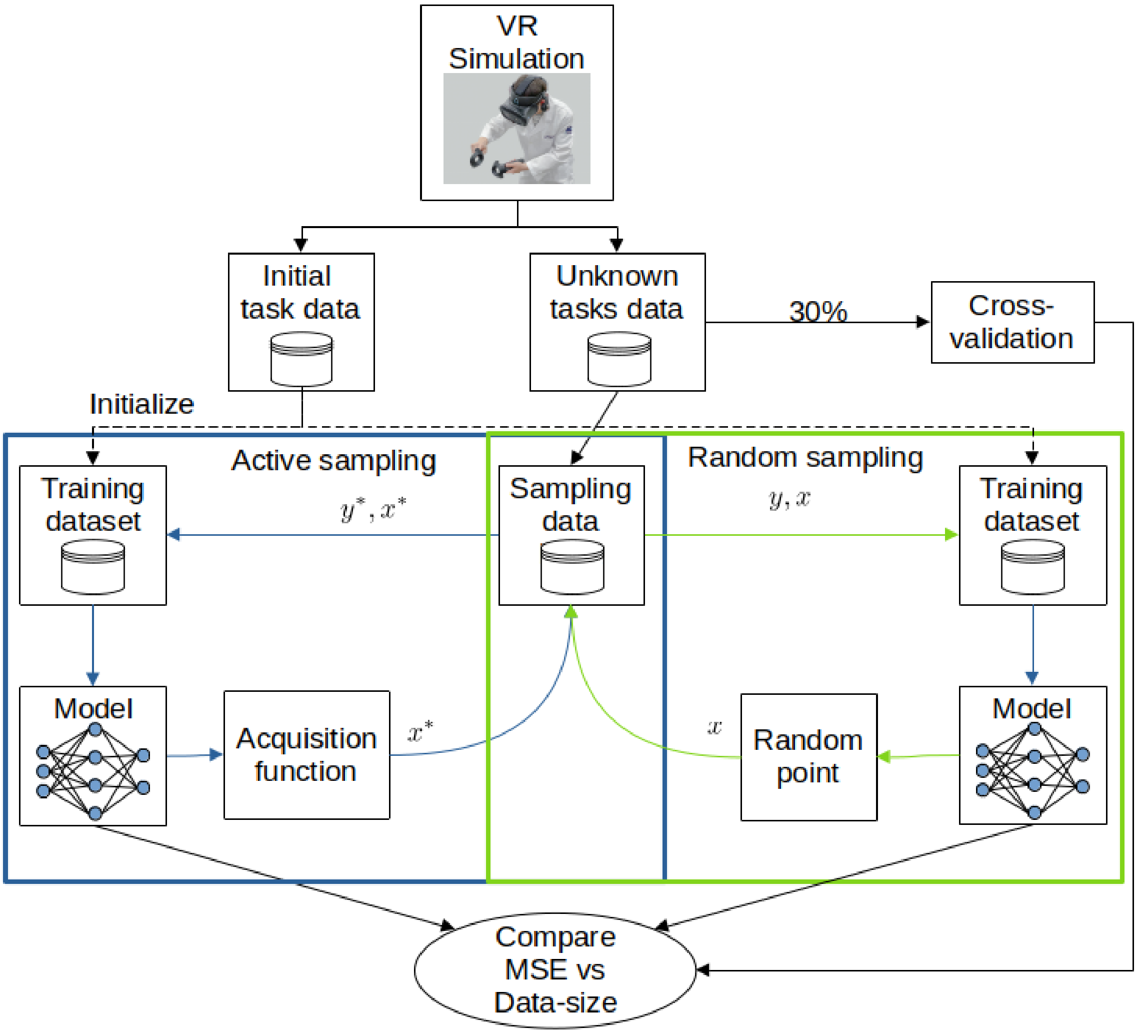


Figure 5 The sampling experiment design. The data is acquired from VR experiments and split into initialization, cross-validation, and data-bank sets, to compare active and random sampling's MSE error.

The data acquired from the VR simulation is split into the initial task-data, which is used to initialize the model as shown in Figure 5. The three remaining tasks data is split into a cross validation dataset and the sampling data bank, which is used by both random and active sampling. In the experimental loop shown above, both models select data from the databank and append it to their training data. The difference is the active model uses informed acquisition, where the random model selects the sample arbitrarily using a unform distribution. Finally, the MSE and training dataset size are compared showing which model performs better. Notice that the one tasks data is initially used to train the model. It is assumed that some information of a similar manufacturing process is available. In order to visualize the results in a 3D graph, the operating conditions represent the task complexity and number of repetitions ($x \in \mathbb{R}^2$), where the system response represents the task duration ($y \in \mathbb{R}$). Additional conditions resulted in nominal model accuracy improvements and obscured the model progress. The Pytorch package was used for these experiments. For runtime deployment, Unity provides the 'Barracuda' module.

4. Theory

4.1 Active learning model

As shown in Figure 6, the active learning algorithm used follows a straightforward loop. Starting at a selected initial point x^*, its samples from the unknown system. Where $f(x)$ is the system being observed, $g(x)$ is our function approximating the system which also predicts the uncertainty, $(y, \sigma) = g(x)$. Appendix A illustrates snapshots of this for an example function.

The sequence is as follows:

1. Conduct an experiment at x^*, resulting in $y^* = f(x^*)$.
2. Appended the data to the training dataset.
3. The model is trained using the current training pool.
4. The next sampling point x^* is determined.
5. This process is repeated until we reach a termination condition.

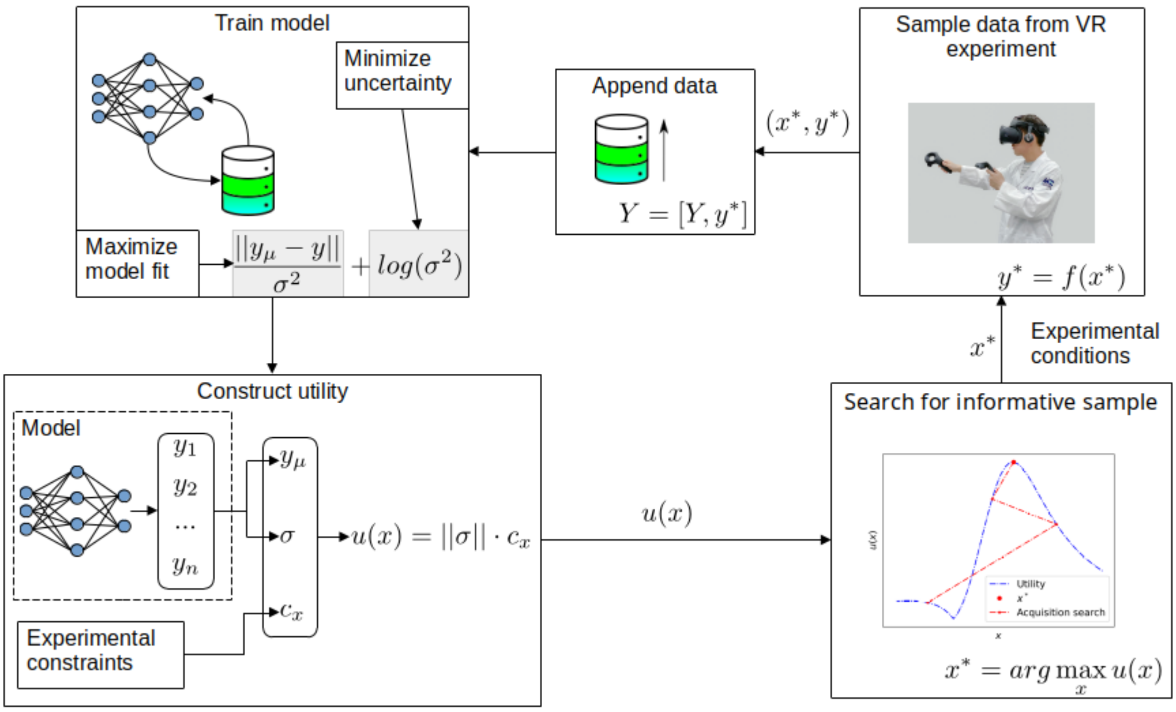


Figure 6 The main loop of the algorithm as applied in this work. The utility represents the acquisition function and is used to search for informative experimental conditions. It is constructed from the model uncertainty prediction and user designed functions.

4.2 Uncertainty estimation

In this work, a convenient (local) ensemble method inspired by [65] was used. It has two main differences to global ensemble methods as presented in Figure 7. Firstly, instead of using multiple networks, a single neural network has multiple mean-predictions. Secondly, there is no need to split the data. This model is suitable as an entry point for ease of use, but we suggest other ensemble methods for practical applications. The figure below shows the two configurations.

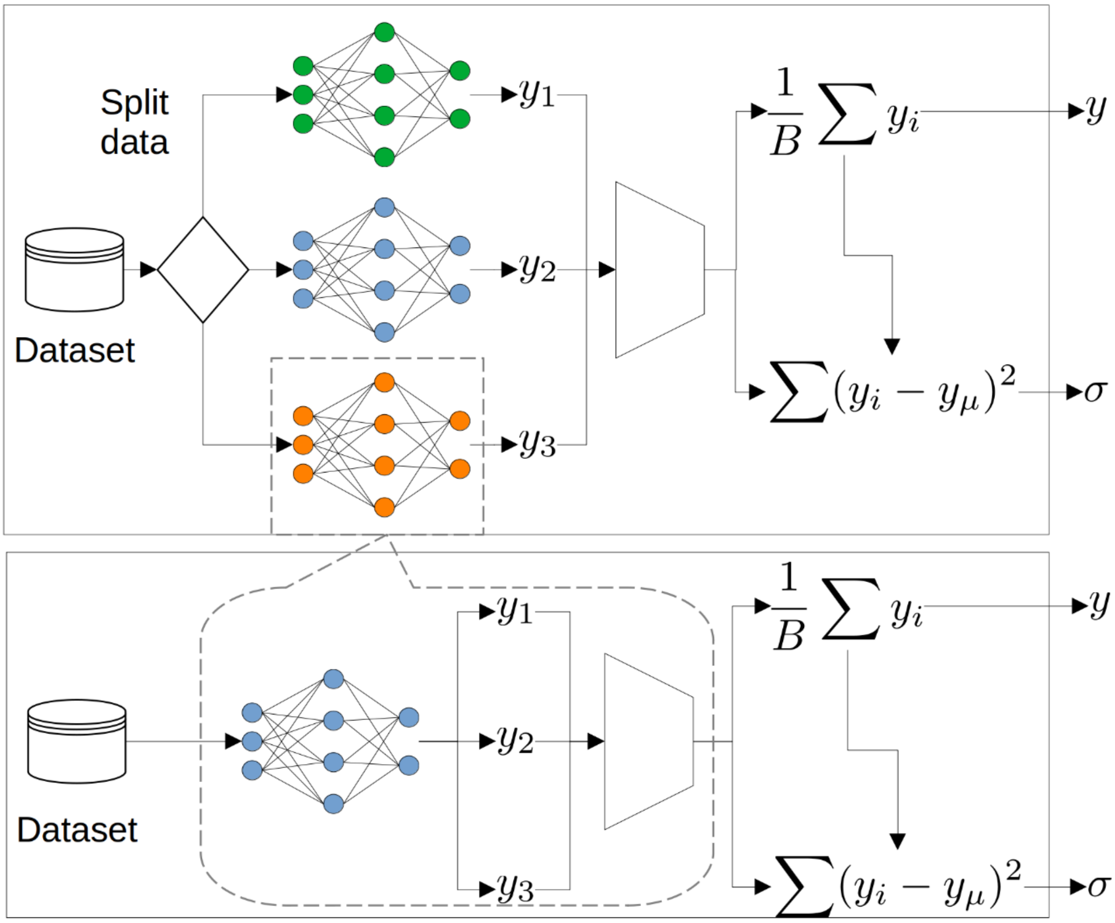


Figure 7 Ensembles predicting the mean response and uncertainty. Conventional (Global) ensemble methods (top) train multiple models on random subsets of the data. Below, the local ensemble method is simpler and more convenient.

Note that local ensemble can be used within Global ensemble methods, hence the two are not exclusive. We do not claim that this method yields better results and suspect this model will not scale well with dimensionality, number of ensembles, inputs, and outputs due to the interaction between neurons.

4.3 Utility based acquisition and experimental constraints

We formulate the selection of the next sample as a search problem, with the acquisition function acting as the objective function. Recall that the uncertainty σ is a real value given by a norm of $\sigma(x)$ or $g_\sigma(x)$ and $x \in \mathbb{R}^n$, where n is the input dimensionality. We also assume that all inputs are controllable. The optimization/search tools (Standard Gradient descent) are already present in the heuristic deep learning framework (Pytorch), making this a natural solution.

A naïve solution would sample where the uncertainty is the highest $x^* = \underset{x}{argmax}(\sigma(x))$. This turns out to be a reasonable approach but does not account for experimental design constraints. Instead, we propose utility as a means of combining the multiple concerns of the acquisition function by biasing the model's selection, in turn retaining some control over the model.

Consider $Util(x) = \sigma(x) * c_1(x) * c_2(x) * \ldots$, where $c_1, \ldots, c_n$ are constraint functions we design. The utility is shaped by multiplying these functions. We typically design these functions to be in a range of 0 to 1, e.g., $c_i(x) \epsilon [0,1]$. We now select the point with the maximum utility (instead of uncertainty) using optimization $x^* = \max\left(Util(x^*)\right)$.

When designing functions for experimental constraints the result is typically binary (1 or 0). Figure 8 below shows the use of step functions. One can see that due to our constraints the maximum utility, hence our sampling point, will always occur within our experimental range[3].

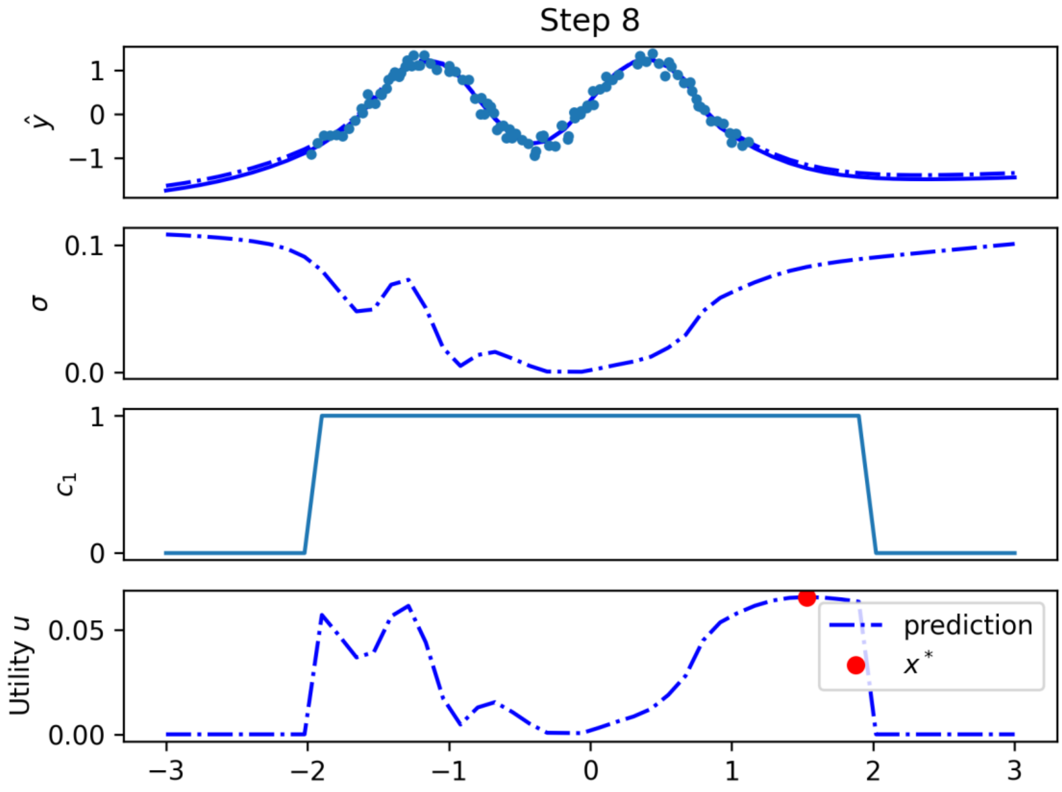


Figure 8 The construction of utility $u(x) = c_1(x) * \sigma(x)$ for an example $f(x) = sin(4x) + q$. The step function c_1 constrains the selection of x^* within the experimental design range. The sample is selected by maximizing the utility, $x^* = \underset{x}{argmax}(u(x))$.

An issue involving the algorithm stalling by sampling in the same region was overcome using these methods. Appendix 8.2 mention additional acquisition constraints.

3 Excluding when the utility function or constraints are zero functions.

5. Results

The question we attempt to answer here is "Does active sampling reduce the number of experiments required to model the throughput rate of an unknown human assembly process, given data from a similar task?" After conducting simulations using the data acquired from VR experimental trials, random and active sampling were compared to see which converges to the minimum error with fewer data-points.

5.1 Virtual manufacturing experiment

When examining the data gathered from the VR experiments (Figure 9), as expected we see the mean duration resembles Wrights' learning curve (exponential decay). The resulting variance is interesting and to our knowledge has not yet been observed. The variance is largest in the beginning (when the subject is first learning the task) and decreases with repetitions. This is true for both inter-subject and intra-subject (inter-repetition) variance and can be interpreted as "operators start at different levels, but after practice tend towards the same performance" and "operator tend to take more risks at the start, but find reliable methods with experience," respectively. Also note that the data is not normally distributed (skew) as a task is more likely to take longer. Therefore, assuming a Gaussian distribution is not accurate, we previously made this assumption for convenience.

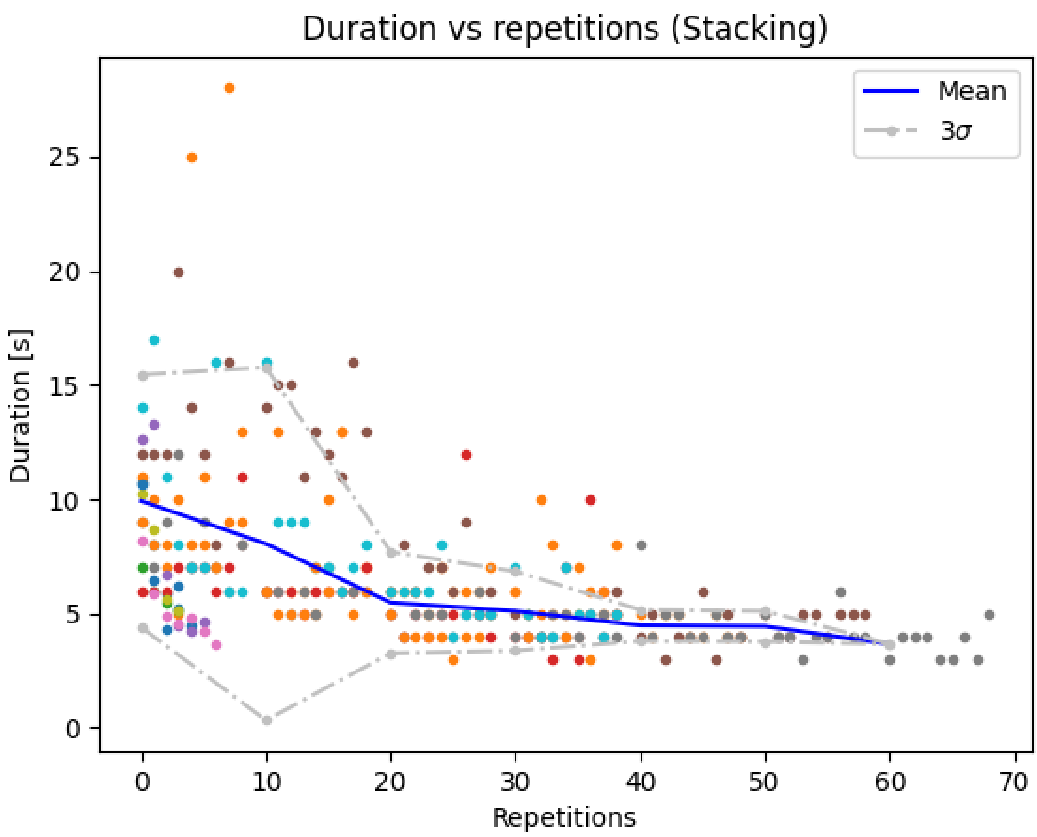


Figure 9 The durations for the stacking, the mean, and variance are taken across each trial (10 repetitions). The assorted color dots correspond to individuals' measurements. An operator is likely to have a task take longer than shorter, resulting in a skew distribution.

A similar trend was observed for the different tasks. In the final task the decrease in variance is not as pronounced. This could be due to less exposure (half-repetitions) or being a more complex task. From these results, we can confirm that Wrights' learning curve can predict the task duration in all four cases. Having observed the variance, an initial large value decaying describes it. This magnitude of the initial value is significant because it affects the performance of the epistemic uncertainty-based acquisition. Specifically, the larger the data variance, the more samples are required to model that specific region.

5.2 Sampling experiment results

When comparing active sampling and random sampling we found active sampling converged to a lower error with less data than random sampling, as expected. The figure that follows illustrates this. The performance of active learning is significantly better than random sampling. Additionally, we see that active learning has a lower variance between multiple runs. It is not clear whether this is the result of a lower mean MSE or more stable learning.

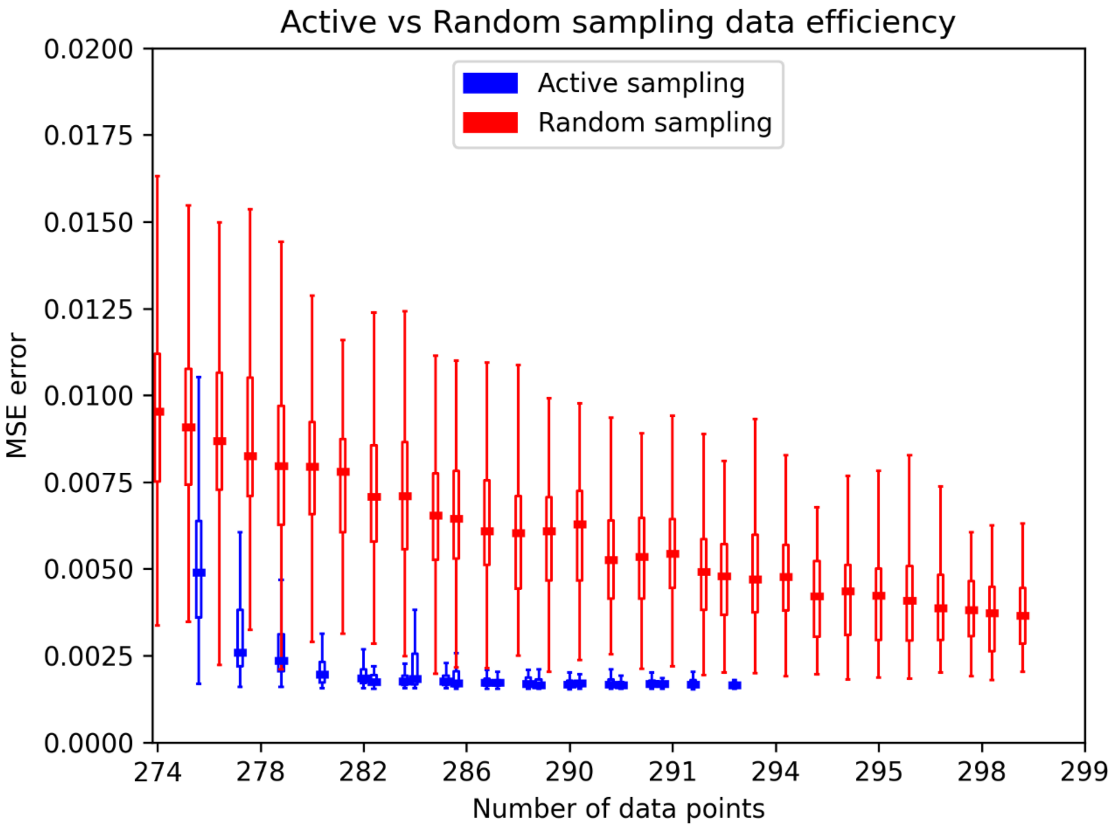


Figure 10 The experimental results of cross validation error of the new tasks for random vs active data-point sampling for one hundred runs. Note how active sampling converges to a low error quicker than random sampling.

As Figure 10 shows, the resulting MSE is randomly distributed. Since a model's uncertainty will not decrease monotonically between iterations, a mean-threshold error condition will not perform well. A predetermined number of iterations was sufficient for this work, but another measure may be useful.

6. Discussion

6.1 Experimental findings

In the experimental results, the VR trials were able to observe Wright learning across multiple tasks in human assembly processes. Moreover, the observation of variance indicates that higher fidelity data can be captured and may be useful. Recall that this simulation was done without investing in hardware or specific sensors. Therefore, VR can serve as a vendor neutral prototyping tool.

The sampling results showed that active sampling can reduce the number of experimental trials conducted required to model a system's performance. When comparing active sampling and random sampling, active sampling achieved lower prediction errors with a similar number of data-samples. Therefore, the objective of achieving sample-efficiency has been met. This did require that the active model be able to control the experimental conditions, is easily achieved in software defined systems like VR, but will require more work in physical systems. The Wright learning curve possessed a high variance region which caused the model to stall, sampling at the same point. Constraints were introduced to address this issue and successfully resolved it. This solution does however require that prior knowledge of or data of a similar process be available. The resulting MSE error was a distribution and there is a growing interest in Bayesian deep learning models. This should be investigated further but is outside of the scope of this work.

The scalability of experiments is greatly extended by combining active learnings automated efficient sampling with VRs software defined operating conditions and portability. Adding remote databases [30] allows multiple concurrent and decentralized simulations to be conducted, producing large amounts of data quickly. This can be applied to performance modeling or workstation design.

6.2 Limitations and future work

One of the main concerns with the proposed work is that as the number of samples and dimensionality of the system increase the training time increases. To address this, local ensembles were used here, but other (non-deep learning) models can also be used. More practical applications are needed to investigate the tradeoffs involved in training time, model complexity, and implementation practicality. This work assumed utility to only consider operating conditions $u(x)$. We are currently exploring using active learning for simultaneous runtime modeling and control/optimization by formulating our utility to also consider system performance/response $u(x, y)$.

6.3 Conclusions

This work firstly aims to investigate whether VR was suitable to prototype human assembly process data-acquisition systems. To this end, a VR environment was built to simulate the manual assembly process and record the task duration. The data acquired follows Wrights learning curve, illustrating that the simulation was able to observe the operator performance. A novel finding was that Wright's learning curve also possesses (predictable) variance.

The second objective of this research was to reduce the number of experimental trials required by using active learning. Using a sampling experiment, we showed that active learning can reduce the number of experimental samples required by selecting samples based on epistemic uncertainty.

In conclusion, the experimental results present the possibility to including humans into modern manufacturing systems by (1) reducing investment risk by allowing prototyping of new and existing workstations with data-acquisition functionality, (2) making modeling of human behavior more practical by automating the experimentation process and reducing the amount of human trials required, and (3) facilitating the adoption of data-based human operator systems by lowering the barrier of entry.

Human trial statement

The VR trials were conducted in accordance with the UNIST Institutional Review Board of ethics (UNISTIRB-2203-A) and the Declaration of Helsinki. Subjects' identities will remain private. Moreover, the conductors took liberties to make the experience comfortable for subjects without sacrificing the accuracy of the experiment/simulation.

Acknowledgments

This is part of a larger body of work investigating the place of humans in industry and was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government (MOTIE, 20193310100030), and the Nuclear Safety Research Program through the Korea Foundation Of Nuclear Safety (KoFONS) using the financial resource granted by the Nuclear Safety and Security Commission (NSSC) of the Republic of Korea. (No. 2003020).

7. References

[1] P. Leitão, J. Barbosa, A. Pereira, J. Barata, and A. W. Colombo, “Specification of the PERFoRM architecture for the seamless production system reconfiguration,” *IECON Proc. Ind. Electron. Conf.*, pp. 5729–5734, 2016, doi: 10.1109/IECON.2016.7793007.

[2] S. Karnouskos and P. Leitao, “Key Contributing Factors to the Acceptance of Agents in Industrial Environments,” *IEEE Trans. Ind. Inform.*, vol. 13, no. 2, pp. 696–703, 2017, doi: 10.1109/TII.2016.2607148.

[3] A. Kolus, R. Wells, and P. Neumann, “Production quality and human factors engineering: A systematic review and theoretical framework,” *Appl. Ergon.*, vol. 73, no. October 2017, pp. 55–89, 2018, doi: 10.1016/j.apergo.2018.05.010.

[4] F. Fruggiero, S. Riemma, Y. Ouazene, R. Macchiaroli, and V. Guglielmi, “Incorporating the Human Factor within Manufacturing Dynamics,” *IFAC-Pap.*, vol. 49, no. 12, pp. 1691–1696, 2016, doi: 10.1016/j.ifacol.2016.07.825.

[5] M. Yung, A. Kolus, R. Wells, and W. P. Neumann, “Examining the fatigue-quality relationship in manufacturing,” *Appl. Ergon.*, vol. 82, no. August 2018, p. 102919, 2020, doi: 10.1016/j.apergo.2019.102919.

[6] Wang Baicun, Peng Tao, Xi Vincent Wang, Thorsten Wuest, David Romero, and Lihui Wang, Eds., “Human-centric Smart Manufacturing: Trends, Issues and Challenges,” *J. Manuf. Syst.*, 2021.

[7] F. Sgarbossa, E. Grosse, W. P. Neumann, and C. Berlin, “Call for Papers: Human-centric production and logistics systems,” *Int. J. Prod. Res.*, 2022, [Online]. Available: https://www.callforpapers.co.uk/human-factors-i50

[8] N. A. Stanton, “Special issue on human factors and ergonomics methods,” *Hum. Factors Ergon. Manuf.*, vol. 32, no. 1, pp. 3–5, 2022, doi: 10.1002/hfm.20943.

[9] “Industry 5.0: Towards more sustainable, resilient and human-centric industry.” https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/industry-50-towards-more-sustainable-resilient-and-human-centric-industry-2021-01-07_en (accessed Sep. 20, 2022).

[10] L. Wang, “A futuristic perspective on human-centric assembly,” *J. Manuf. Syst.*, vol. 62, pp. 199–201, Jan. 2022, doi: 10.1016/J.JMSY.2021.11.001.

[11] M. Zhang, F. Tao, and A. Y. C. Nee, “Digital Twin Enhanced Dynamic Job-Shop Scheduling,” *J. Manuf. Syst.*, vol. 58, pp. 146–156, Jan. 2021, doi: 10.1016/J.JMSY.2020.04.008.

[12] D. K. Baroroh and C.-H. Chu, “Human-centric production system simulation in mixed reality: An exemplary case of logistic facility design,” *J. Manuf. Syst.*, vol. 65, pp. 146–157, Oct. 2022, doi: 10.1016/J.JMSY.2022.09.005.

[13] B. Wang, P. Zheng, Y. Yin, A. Shih, and L. Wang, “Toward human-centric smart manufacturing: A human-cyber-physical systems (HCPS) perspective,” *J. Manuf. Syst.*, vol. 63, pp. 471–490, Apr. 2022, doi: 10.1016/J.JMSY.2022.05.005.

[14] Y. Lu, X. Xu, and L. Wang, “Smart manufacturing process and system automation – A critical review of the standards and envisioned scenarios,” *J. Manuf. Syst.*, vol. 56, pp. 312–325, Jul. 2020, doi: 10.1016/J.JMSY.2020.06.010.

[15] Y. Lu *et al.*, “Outlook on human-centric manufacturing towards Industry 5.0,” *J. Manuf. Syst.*, vol. 62, pp. 612–627, Jan. 2022, doi: 10.1016/J.JMSY.2022.02.001.

[16] Q. Liu, M. Liu, H. Zhou, F. Yan, Y. Ma, and W. Shen, “Intelligent manufacturing system with human-cyber-physical fusion and collaboration for process fine control,” *J. Manuf. Syst.*, vol. 64, pp. 149–169, Jul. 2022, doi: 10.1016/J.JMSY.2022.06.004.

[17] S. Kerick, J. Metcalf, T. Feng, A. Ries, and K. McDowell, “Review of Fatigue Management Technologies for Enhanced Military Vehicle Safety and Performance,” *Tech. Rep. US Army Res. Lab.*, no. September, 2013.

[18] C. C. Liu, S. G. Hosking, and M. G. Lenné, “Predicting driver drowsiness using vehicle measures: Recent insights and future challenges,” *J. Safety Res.*, vol. 40, no. 4, pp. 239–245, 2009, doi: 10.1016/j.jsr.2009.04.005.

[19] H. B. Kang, “Various approaches for driver and driving behavior monitoring: A review,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 616–623, 2013, doi: 10.1109/ICCVW.2013.85.

[20] A. Sahayadhas, K. Sundaraj, and M. Murugappan, “Detecting driver drowsiness based on sensors: A review,” *Sens. Switz.*, vol. 12, no. 12, pp. 16937–16953, 2012, doi: 10.3390/s121216937.

[21] K. Ransikarbum, N. Kim, S. Ha, R. A. Wysk, and L. Rothrock, “A Highway-Driving System Design Viewpoint Using an Agent-Based Modeling of an Affordance-Based Finite State Automata,” *IEEE Access*, vol. 6, pp. 2193–2205, 2017, doi: 10.1109/ACCESS.2017.2782257.

[22] D. F. Dinges, G. Maislin, R. M. Brewster, G. P. Krueger, and R. J. Carroll, “Pilot test of fatigue management technologies,” *Transp. Res. Rec.*, no. 1922, pp. 175–182, 2005, doi: 10.3141/1922-22.

[23] G. Yang, Y. Lin, and P. Bhattacharya, “A driver fatigue recognition model based on information fusion and dynamic Bayesian network,” *Inf. Sci.*, vol. 180, no. 10, pp. 1942–1954, 2010, doi: 10.1016/j.ins.2010.01.011.

[24] H. J. Baek, G. S. Chung, K. K. Kim, and K. S. Park, “A smart health monitoring chair for nonintrusive measurement of biological signals,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 1, pp. 150–158, 2012, doi: 10.1109/TITB.2011.2175742.

[25] Z. Sedighi Maman, M. A. Alamdar Yazdi, L. A. Cavuoto, and F. M. Megahed, “A data-driven approach to modeling physical fatigue in the workplace using wearable sensors,” *Appl. Ergon.*, vol. 65, pp. 515–529, 2017, doi: 10.1016/j.apergo.2017.02.001.

[26] E. Bal, O. Arslan, and L. Tavacioglu, “Prioritization of the causal factors of fatigue in seafarers and measurement of fatigue with the application of the Lactate Test,” *Saf. Sci.*, vol. 72, pp. 46–54, 2015.

[27] A. Sebok, C. Wickens, and R. Sargent, “Using meta-analyses results and data gathering to support human performance model development,” *Proc. Hum. Factors Ergon. Soc.*, pp. 783–787, 2013, doi: 10.1177/1541931213571171.

[28] G. Lawson, D. Salanitri, and B. Waterfield, “Future directions for the development of virtual reality within an automotive manufacturer,” *Appl. Ergon.*, vol. 53, pp. 323–330, Mar. 2016, doi: 10.1016/J.APERGO.2015.06.024.

[29] W. Dangelmaier, M. Fischer, J. Gausemeier, M. Grafe, C. Matysczok, and B. Mueck, “Virtual and augmented reality support for discrete manufacturing system simulation,” *Comput. Ind.*, vol. 56, no. 4, pp. 371–383, May 2005, doi: 10.1016/J.COMPIND.2005.01.007.

[30] J. Grübel, R. Weibel, M. H. Jiang, C. Hölscher, D. A. Hackman, and V. R. Schinazi, “EVE: A Framework for Experiments in Virtual Environments,” 2017, pp. 159–176. doi: 10.1007/978-3-319-68189-4_10.

[31] J. Brookes, M. Warburton, M. Alghadier, M. Mon-Williams, and F. Mushtaq, “Studying human behavior with virtual reality: The Unity Experiment Framework,” *Behav. Res. Methods*, vol. 52, no. 2, pp. 455–463, Apr. 2020, doi: 10.3758/s13428-019-01242-0.

[32] J. L. Harbour, “Human performance modeling: A case study,” *Perform. Improv.*, vol. 49, no. 8, pp. 36–41, Sep. 2010, doi: 10.1002/PFI.20171.

[33] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, “An integrated theory of the mind,” *Psychol. Rev.*, vol. 111, no. 4, pp. 1036–1060, Oct. 2004, doi: 10.1037/0033-295X.111.4.1036.

[34] B. Wang, Y. Li, and T. Freiheit, “Towards intelligent welding systems from a HCPS perspective: A technology framework and implementation roadmap,” *J. Manuf. Syst.*, vol. 65, pp. 244–259, Oct. 2022, doi: 10.1016/j.jmsy.2022.09.012.

[35] N. Li, H. Kong, Y. Ma, G. Gong, and W. Huai, “Human performance modeling for manufacturing based on an improved KNN algorithm”, doi: 10.1007/s00170-016-8418-6.

[36] T. S. Baines and J. M. Kay, “Human performance modelling as an aid in the process of manufacturing system design: A pilot study,” *Int. J. Prod. Res.*, vol. 40, no. 10, pp. 2321–2334, Jul. 2002, doi: 10.1080/00207540210128198.

[37] "Human-robot collaboration empowered by hidden semi-Markov model for operator behaviour prediction in a smart assembly system | Elsevier Enhanced Reader." https://reader.elsevier.com/reader/sd/pii/S0278612521002508?token=4728B25D0B8603B879EC8CEE5E156CFC8B63BDFD781019D2682EF667156C50959F97718BAD4D4751D022F30D1AD7B19A&originRegion=us-east-1&originCreation=20220922125312 (accessed Sep. 22, 2022).

[38] G. Paul and L. Briceno, "A Conceptual Framework of DHM Enablers for Ergonomics 4.0," in *Lecture Notes in Networks and Systems*, vol. 223 LNNS, 2022, pp. 403–406. doi: 10.1007/978-3-030-74614-8_50.

[39] T. P. Wright, "Factors Affecting the Cost of Airplanes," *J. Aeronaut. Sci.*, vol. 3, no. 4, pp. 122–128, Feb. 1936, doi: 10.2514/8.155.

[40] B. Nagy, J. D. Farmer, Q. M. Bui, and J. E. Trancik, "Statistical Basis for Predicting Technological Progress," *PLoS ONE*, vol. 8, no. 2, p. 52669, 2013, doi: 10.1371/journal.pone.0052669.

[41] Sam Korus, "Wright's Law Predicted 109 Years of Autos Gross Margin, and Now Tesla's," *Ark-invest Analyst Research*, 2019. https://ark-invest.com/articles/analyst-research/wrights-law-predicts-teslas-gross-margin/ (accessed Sep. 16, 2022).

[42] Brett Winton, "Moore's Law Isn't Dead: It's Wrong - Long Live Wright's Law," *Ark-invest Analyst Research*, 2019. https://ark-invest.com/articles/analyst-research/wrights-law-2/ (accessed Sep. 16, 2022).

[43] E. M. Dar-Ei, *HUMAN LEARNING: From Learning Curves to Learning Organizations*, vol. 29. Boston, MA: Springer US, 2000. doi: 10.1007/978-1-4757-3113-2.

[44] G. Mummolo, S. Digiesi, and G. Mossa, "Learning and Tirediness Phenomena in Manual Operation Performed in Lean Automated Manufacturing Systems: a Reference Model," 2004.

[45] N. Asadayoobi, M. Y. Jaber, and S. Taghipour, "A new learning curve with fatigue-dependent learning rate," *Appl. Math. Model.*, vol. 93, pp. 644–656, May 2021, doi: 10.1016/j.apm.2020.12.005.

[46] R. A. Fisher, *The design of experiment*. Edinburgh, Scotland: Oliver and Boyd, 1935.

[47] M. Wantawin, W. Yu, S. Dachanuwattana, and K. Sepehrnoori, "An Iterative Response-Surface Methodology by Use of High-Degree-Polynomial Proxy Models for Integrated History Matching and Probabilistic Forecasting Applied to Shale-Gas Reservoirs," *SPE J.*, vol. 22, no. 06, pp. 2012–2031, Dec. 2017, doi: 10.2118/187938-PA.

[48] Christopher Gatti, *Springer Theses Recognizing Outstanding Ph.D. Research*. Accessed: Apr. 20, 2022. [Online]. Available: http://www.springer.com/series/8790

[49] P. Ren *et al.*, "A Survey of Deep Active Learning," *ACM Comput. Surv.*, vol. 54, no. 9, Dec. 2022, doi: 10.1145/3472291.

[50] M. Haußmann, F. Hamprecht, and M. Kandemir, "Deep active learning with adaptive acquisition," in *IJCAI International Joint Conference on Artificial Intelligence*, 2019, vol. 2019-Augus, pp. 2470–2476. doi: 10.24963/ijcai.2019/343.

[51] Y. Geifman and R. El-Yaniv, "Deep active learning with a neural architecture search," 2019.

[52] P. Ren *et al.*, "A Comprehensive Survey of Neural Architecture Search," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–34, May 2022, doi: 10.1145/3447582.

[53] D. Wu, C. T. Lin, and J. Huang, "Active learning for regression using greedy sampling," *Inf. Sci.*, vol. 474, pp. 90–105, Feb. 2019, doi: 10.1016/J.INS.2018.09.060.

[54] S. Ren, Y. Deng, W. J. Padilla, and J. Malof, "Hyperparameter-free deep active learning for regression problems via query synthesis," Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.12632

[55] Lennart. Ljung, *System identification : theory for the user*. Prentice Hall, 1999.

[56] E. Blasch, S. Ravela, and A. Aved, "Handbook of Dynamic Data Driven Applications Systems," *Handb. Dyn. Data Driven Appl. Syst.*, pp. 1–750, Nov. 2018, doi: 10.1007/978-3-319-95504-9/COVER.

[57] A. N. Donev and A. C. Atkinson, *Optimum Experimental Designs*. Clarendon Press, Oxford Statistical Science Series, 1992.

[58] W. Cai, Y. Zhang, and J. Zhou, "Maximizing expected model change for active learning in regression," *Proc. - IEEE Int. Conf. Data Min. ICDM*, pp. 51–60, 2013, doi: 10.1109/ICDM.2013.104.
[59] R. Burbidge, J. J. Rowland, and R. D. King, "Active Learning for Regression Based on Query by Committee".
[60] A. P. Soleimany, A. Amini, S. Goldman, D. Rus, S. N. Bhatia, and C. W. Coley, "Evidential Deep Learning for Guided Molecular Property Prediction and Discovery," *ACS Cent. Sci.*, vol. 7, no. 8, pp. 1356–1367, Aug. 2021, doi: 10.1021/acscentsci.1c00546.
[61] H. Yu and S. Kim, "Passive sampling for regression," *Proc. - IEEE Int. Conf. Data Min. ICDM*, pp. 1151–1156, 2010, doi: 10.1109/ICDM.2010.9.
[62] Y. Feng *et al.*, "Human-cyber-physical system for operation in nuclear reactor possessing asymmetric multi-task learning-based predicting framework," *J. Manuf. Syst.*, vol. 64, pp. 443–453, Jul. 2022, doi: 10.1016/J.JMSY.2022.07.008.
[63] C. A. Steed, "A simulation-based approach to develop a holonic robotic cell," *Ind. Robot*, vol. 46, no. 1, p. IR-07-2018-0149, Mar. 2019, doi: 10.1108/IR-07-2018-0149.
[64] T. Pearce, M. Zaki, A. Brintrup, and A. Neely, "High-Quality Prediction Intervals for Deep Learning: A Distribution-Free, Ensembled Approach," presented at the International Conference on Machine Learning, Feb. 2018. [Online]. Available: http://proceedings.mlr.press/v80/pearce18a.html
[65] Y. Lai, Y. Shi, Y. Han, Y. Shao, M. Qi, and B. Li, "Exploring uncertainty in regression neural networks for construction of prediction intervals," *Neurocomputing*, vol. 481, pp. 249–257, Apr. 2022, doi: 10.1016/j.neucom.2022.01.084.
[66] Y. Li *et al.*, "Deep Bayesian Gaussian processes for uncertainty estimation in electronic health records," *Sci. Rep.*, vol. 11, no. 1, p. 20685, Dec. 2021, doi: 10.1038/s41598-021-00144-6.
[67] V. Mullachery, A. Khera, and A. Husain, "Bayesian Neural Networks," Jan. 2018, [Online]. Available: http://arxiv.org/abs/1801.07710
[68] T. Hothorn and B. Lausen, "Double-bagging: combining classifiers by bootstrap aggregation," *Pattern Recognit.*, vol. 36, no. 6, pp. 1303–1309, Jun. 2003, doi: 10.1016/S0031-3203(02)00169-3.
[69] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," presented at the Proceedings of The 33rd International Conference on Machine Learning, Jun. 2015. [Online]. Available: https://proceedings.mlr.press/v48/gal16.html
[70] C. Fiedler, C. W. Scherer, and S. Trimpe, "Practical and Rigorous Uncertainty Bounds for Gaussian Process Regression," 2021. [Online]. Available: www.aaai.org
[71] T. Heskes, "Practical confidence and prediction intervals," in *Proceedings of the 9th International Conference on Neural Information Processing Systems*, Cambridge, MA, USA, Dec. 1996, pp. 176–182.
[72] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Trans. Neural Netw.*, vol. 22, no. 9, pp. 1341–1356, Sep. 2011, doi: 10.1109/TNN.2011.2162110.
[73] M. Abdar *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, Dec. 2021, doi: 10.1016/J.INFFUS.2021.05.008.
[74] J. Quiñonero-Candela, C. E. Rasmussen, F. Sinz, O. Bousquet, and B. Schölkopf, "Evaluating Predictive Uncertainty Challenge," 2006, pp. 1–27. doi: 10.1007/11736790_1.

8. Appendix supporting plots and figures

8.1 Toy problem

In the figure below we show the results for a toy problem. On the left we see the response surface, on the right the uncertainty surface. We sample at the point of highest uncertainty. After several iterations, the response surface resembles the generating function.

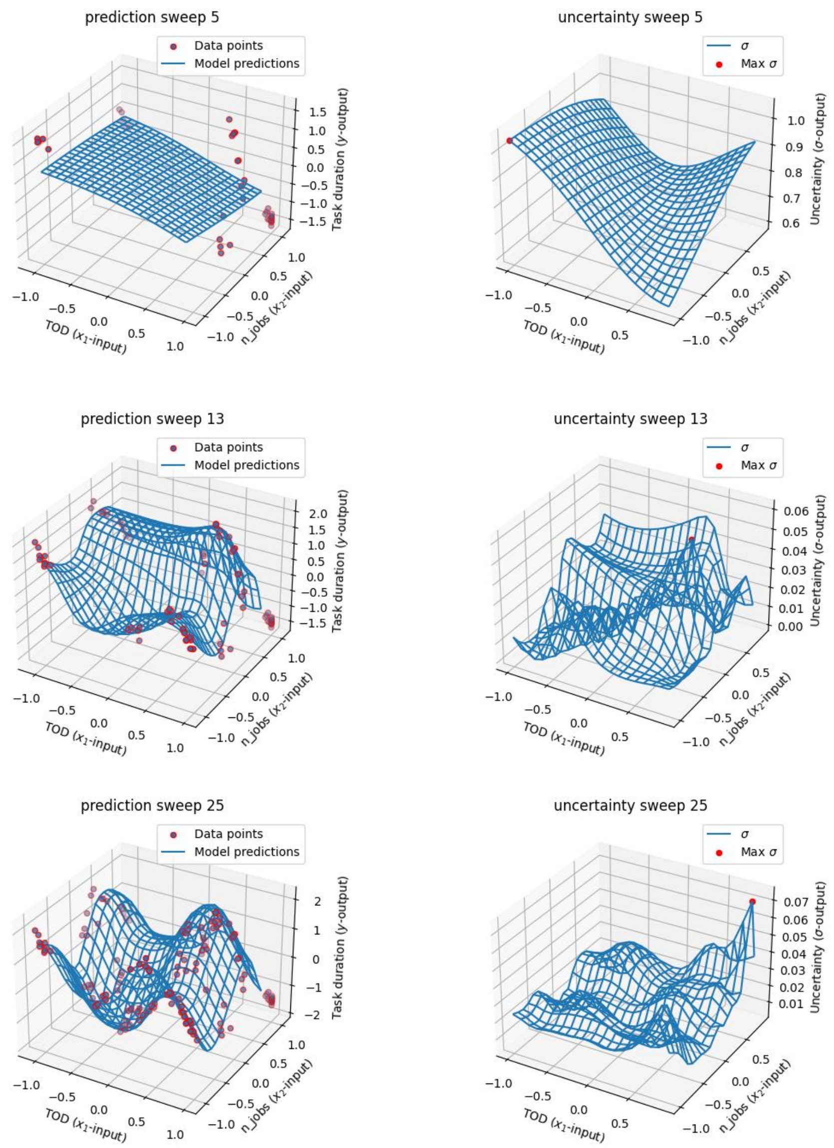


Figure A1: A toy problem with two input-features ($y = \sin(x_1) + \sin(x_2) + N(0, q^2)$). On the left the mean predictions [gif here]. On the right the uncertainty [gif here]. Note how the plot axis limit the selection of the maximum σ.

8.2 Acquisition function extras

8.2.1 Limiting the change

There may be cases where we would like to bias the next sample to be geometrically near to the previous sample. Consider the example of a process involving varying the temperature or feed speed to find ideal operating conditions. One would prefer to change these factors only slightly to avoid complications because of irregular throughput, heating energy costs, and temperature fluctuations. In this case we encourage points that are near to the current operating point. e.g., $c(x) = \frac{\beta^2}{\beta^2+(x_c-x)^2}$.

8.2.2 Diversity inclusion

On the other hand, we may want to sample points that are further away to encourage data diversity. In this case we penalize points that are close together. e.g., $c(x) = C^{-\frac{\beta^2}{\beta^2+(x_c-x)^2}}$.

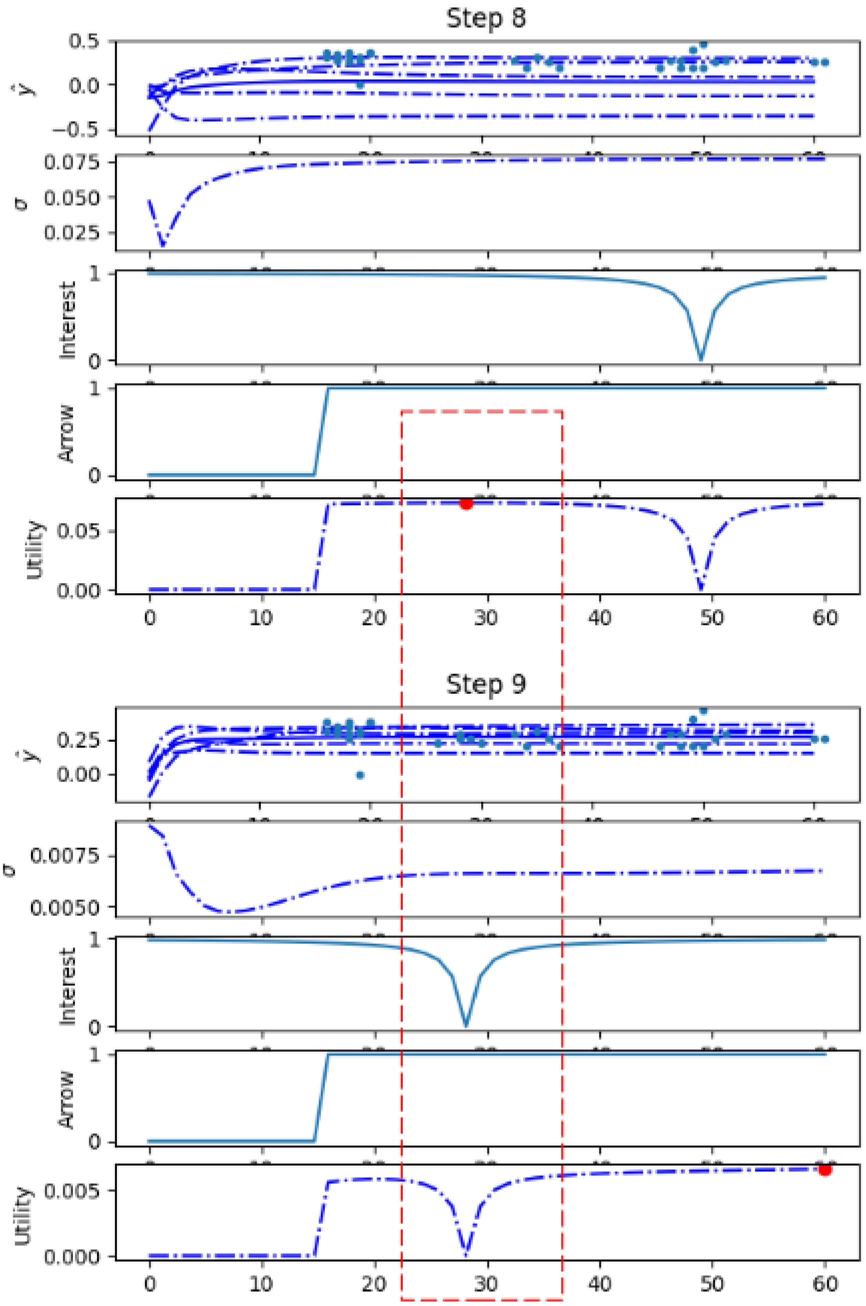


Figure A2: Shows how we can encourage diversity in samples by encouraging distance between consecutive points. The red dot represents the point of highest utility and the next sample.

8.2.3 Other uses for utility

Often, we want to bias the acquisition of x^* near the region of interest. During testing we found this useful for biasing x^* away from regions of high uncertainty. A candidate function is shown below, where p is the point of interest. The region need not be a point but could be a higher-dimensional geometric feature.

$$c(x) = \frac{\beta^2}{\beta^2 + (p - x)^2}$$

.

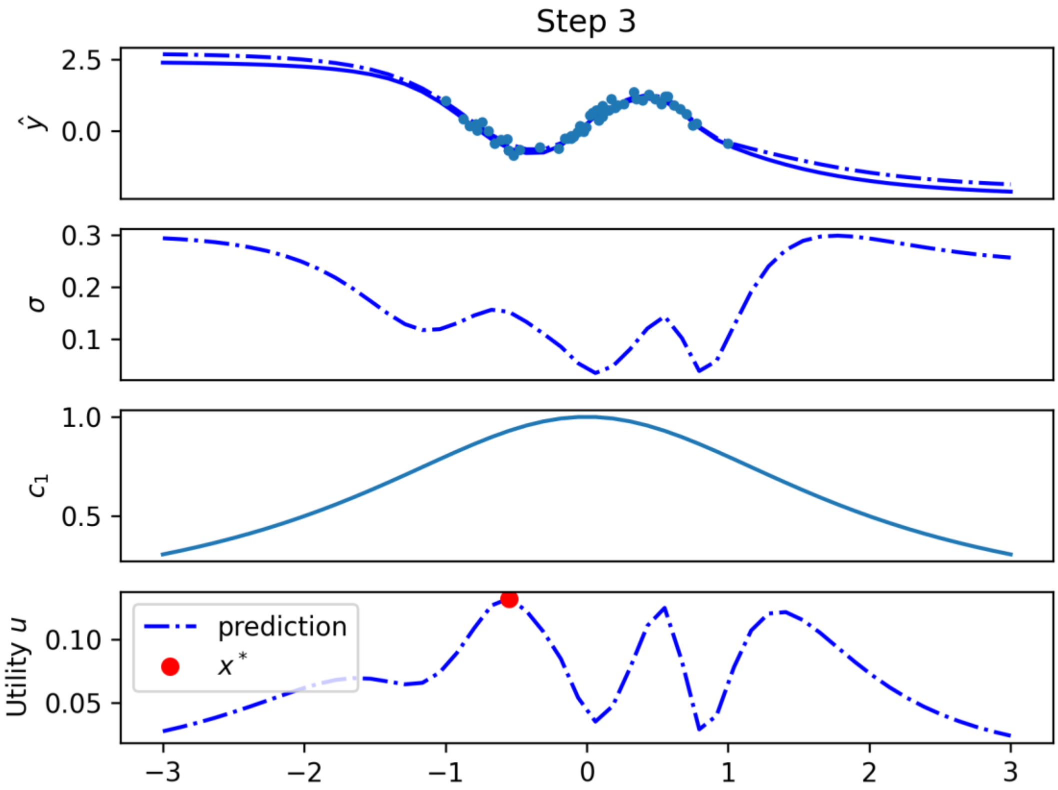


Figure A3: The utility is biased to sample nearer to the region of interest (0). Points far from this area will not be considered.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

2022/11/04

Clint Alex Steed

Namhun Kim