ENGINEERS AUSTRALIA
ACAM 10
10th AUSTRALASIAN CONGRESS ON APPLIED MECHANICS
1 – 3 December 2021
VIRTUAL CONGRESS
ACAMCONFERENCE.COM.AU

ISBN number 978-1-925627-59-6

Auto-tuning Bayesian Filtering for Model Identification and Updating Using Reinforcement Learning

Xin Zhang[1], Yuguang Fu[2,*], Sunny Sharma[1], Shirley Dyke[1, 3]

[1]*School of Civil Engineering, Purdue University, West Lafayetter, IN, USA*

[2]*School of Civil and Environmental Engineering, Nanyang Technological University, Singapore*

[3]*School of Mechanical Engineering, Purdue University, West Lafayetter, IN, USA*

[*]*Corresponding author E-mail: fu303@purdue.edu*

ABSTRACT

With the development of dynamic theory and earthquake engineering, different devices have been introduced to control structure systems in vibration. Numerically identifying the parameters of these systems become the focus in structure engineering field. Researchers have developed Bayesian inference method to develop system parameters, which is able to predict robust results. However, choosing a set of suitable prior distributions in Bayesian inference method is a challenging task. Currently, researchers will preset a range for the prior distributions and then incrementally try the values in this range until find the prior distributions that achieve optimum parameters. This method is time-consuming because researchers need to generate prior distribution' value as many as possible and then test all of these values to find the most optimum ones. And the number of prior distribution values tried in this process is limited due to the limitation of human workforce, which then causes the result is not the most optimum one. To solve this problem, a reinforcement learning-based tool is developed in this work to generate suitable prior distribution for Bayesian inference method. Reinforcement learning is an algorithm which is utilized to optimize the strategies and decisions. In this work, choosing optimum prior distribution will be seen as a decision-making problem and then reinforcement learning will be used to optimize the decision of choosing prior distribution. Settings in each step of reinforcement learning are adjusted in this work for this problem. With this method, researchers can be enslaved from trying too many prior distribution values and the values of the prior distribution are searched in a continuous space rather than incrementally, which makes the results more accurate.

Keywords: Bayesian filtering, auto tuning, reinforcement learning, hyperparameter optimization, model identification.

INTRODUCTION

Civil engineers utilize predictive modelling to conclude the information from the observation of structural experiment and predict the behavior of the structure in future hazardous events. However, identifying suitable prediction models is a difficult and time-consuming work. Previous researchers developed Bayesian inference methods to quantify the behavior of the structure and then use it to predict.

Based on (Lund et al. 2020 a), researchers have typically focused on three different approaches including analytical inference techniques inspired by Kalman filter (Stepanov 2011), sampling techniques such as particle filters (Gordon et al. 1993), optimization-based approach (Blei et al. 2016, Bishop, 2006) to implement Bayesian inference on practical model identification. New approaches like automatic differentiation vatiational inference (Lund et al. 2020 b) has also been studied in past few years. However, there is a common problem in these methods that, researchers need to identify optimal prior distribution of variables for Bayesian inference. Currently, researchers mainly use brute force method to find suitable set of prior distribution of variables (Lund et al. 2020 b). In this method, researchers need to evaluate the range of prior distribution of each variable and then iterately select a limited number of values for each variable from the range. By trying different value groups of these variables, researchers could find one group of variables that offer the best model performance. The model performance obtained from this method highly relies on the evaluation of range of prior distribution for each variable and the iteration density of selecting variable values. Larger range and higher density could improve the possibility finding better prior distribution but the calculation time will be improved. Another problem for this method is that iterately selecting values from the evaluated range has the limitation of accuracy of prior distribution value. This is because the accuracy of prior distribution value is determined by the length of iteration.

To address the challenges, reinforcement learning (hereby, RL) method has been introduced in this work. From above statement, finding optimal prior distributions of variables can be generalized as a optimization problem and the final goal is finding optimal prior distributions from the evaluated range. Theoretically, RL method is a suitable way to solve such problem. Thus, a RL-based model prior distribution identifier has been developed in this paper. In section 'Technical Approach', theoretical base and general workflow for this identifier will be introduced. In section 'Case Study', the data used for testing the RL-based identifier will be introduced and then the prior distribution identifier will be tested through real experimental data. In the section 'Results', the comparison of the results from previous method and RL-based method will be exhibited and discussed.

TECHNICAL APPROACH

Reinforcement Learning Overview

Reinforcement learning is an area of machine learning among three basic machine learning categories, along with supervised learning and unsupervised learning. The idea of reinforcment learning is inspired by the trial-and-error process of human learning in psychology (Wei et al. 2020). The concept of reinforcement learning algorithm focuses on finding a balance between exploration of uncharted terriroty and exploitation of current knowledge (Miura et al. 2003). Basic reinforcement is modeled as a Markov decision process (Howard 1960) including a set of environment and agent states, *S*, a set of actions, *A*, of the agent, probability of transition from state *S* to *S'*, *P(S, S')* and immediate reward after transition from state *S* to *S'* with action *A*. The agent in reinforcement learning will take an action A_t at time t based on current state S_t and current reward r_t. This action will interact with environment and

then the state of agent will change to S_{t+1}. A reward function will be utilized for giving a reward r_{t+1} based on state S_{t+1}. Then agent will take S_{t+1} and r_{t+1} into consider and take another action A_{t+1}. Above loop will be repeated many times until satisfactory cumulative reward or step reward obtained. Basic process of reinforcement learning is shown in **Figur**.

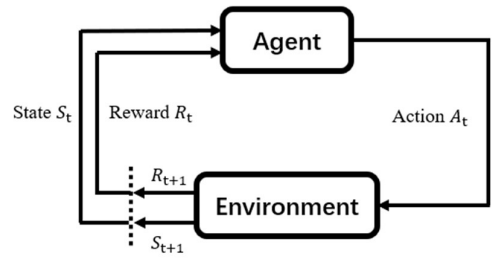


Figure 1. Basic process of reinforcement learning.

Reinforcement learning aims to find optimal policy which could obtain maximum reward (Zoph et al. 2016). Many researchers apply this method in many fields like robot control. Similarly, in model identification problem, researchers also need to find optimal prior distribution (policy) to get the good prediction performance (reward), which inspires us to consider it as a typical reinforcment learning problem. Thus, this study wants to focus on building reinforcement learning framework for the automated policy making of model prior distribution identification.

Dataset General Description

The reinforcement learning method setting need to be designed for specific problem. Thus, in this work, we utilized the structure response data obtained from previous experiment (Sharma 2021) for training and testing reinforment learning method. This experiment is a traditional structure dynamic test. However, researchers added the magnetorheological damper as the semi-active control for energy dissipation in the structural system. This kind of damper has nonlinear responses during the energy dissipation, which will cause the nonlinear model identification study. In this experiment, different input signals were utilized by researchers to get different responses. According to (Sharma 2021), there are two kinds of behaviors under these input signals. For structural displacements that are less than 2 mm one behavior has been observed and for displacements more than 2 mm another behavior has been observed. In Sharma's work, he mainly focused on the tractable behavior of the MR damper observed for structural displacements of more than 2 mm. Thus, in this work, we mainly utilize the response data recording structural displacement of more than 2 mm. **Figure 2** shows one of the example of force vs displacement response.

The basic work of this paper is developing an automatic reinforcement-learning based tool that could identify the model which could predict the structure behavior under certain input signals. Thus, in this work, some input signals and their corresponding responses will be used for training and others will be used for validating the model identification performance. Among all of the non-linear models, the Bouc Wen model will be picked as the base model (as expressed below) and its coefficients will be identified through Unscented Kalman filter (UKF), named as UKF-BW model hereafter.

$$\begin{aligned} m\ddot{u}(t) + c\dot{u}(t) + F(t) &= f(t) \\ f &= k_s x + k_z z \\ \dot{z} &= \rho(\dot{x} - \sigma|\dot{x}||z|^{n-1}z + (\sigma - 1)\dot{x}|z|^n) \end{aligned} \tag{1}$$

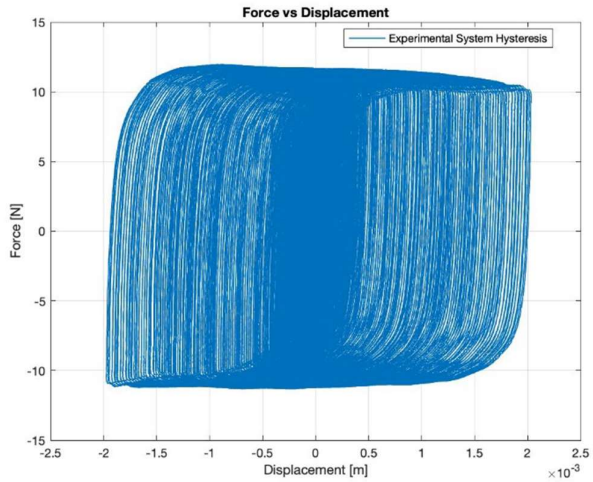


Figure 2. Force vs displacement plot for structure under input signal II.

Reinforcement Learning Process and Settings in This Work

As described in section 'Reinforcement Learning Overview', reinforcement learning contains five basic sections: agent, action, environment, state and reward. To build the automatic model identification tool, these five sections need to be defined properly with the corresponding sections in model identification process.

The agent item in reinforcement learning-based automatic model identifier is defined as the prior distribution generator. This is because we hope that this agent could help to find the suitable prior distribution used in Bayesian inference. In this work, the prior generator is set as a RNN controller. Through this controller, we can generate a sequence of prior distributions that used in Bayesian inference. **Figure 3** shows the process that how the controller recurrent neural network genenrate the prior distributions.

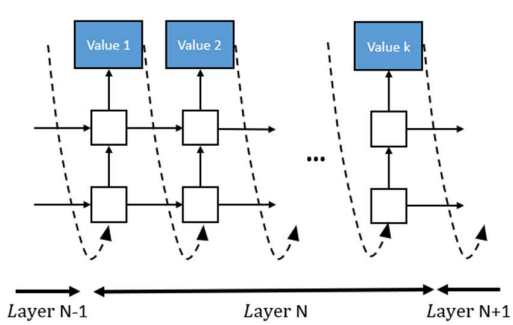


Figure 3. RNN controller for generating Bayesian inference prior distributions.

The action item is the behavior that the agent produce. Thus in this work, the action item is a series of number that represent the predicted prior distributions. The length of these series of number is same as the length of prior distributions needed for model identification problem. The environment item is the one that interact with action item and then make the change of agent state. Thus, in this work, the

environment is defined as the preset non-linear model. The prior distributions generated in action item will react with this environement and produce the model with predicted coefficients. Later, this model will be used to predict the response of the structure under certain input signals and this response is set as the state item in reinforcement learning algorithm. This state item (response of the structure) will be compared with the real structure response obtained from experiment. Since the purpose of this work is to find the model that could predict the structure response as precisesly as possible, the reward function will be set to indicate the model prediction accuracy to the real response. The value obtained in reward function will be used to update the the prior distribution generator. The basic process of the reinforcement learning is designed as **Figure 4**.

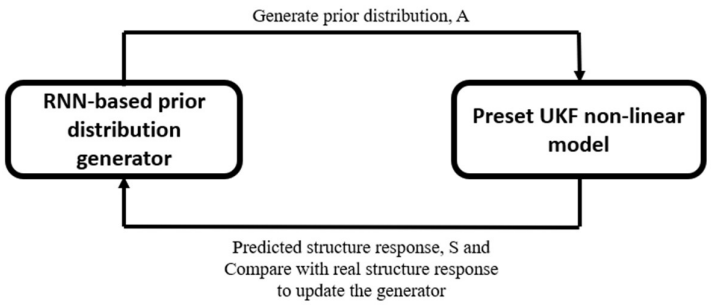


Figure 4. Basic workflow for reinforcement learning-based model identifier.

CASE STUDY

Case Introduction

The case utilized for this work is from (Sharma 2021). In his work, he tested multiple structures with MR damper including single degree of freedom frame and multi-degree (3 degree) of freedom frame. All of these structures are equipped with magnetorheological (MR) damper and thus, these structures will present non-linear response under certain excitation signals. Seven signals with eight different settings are shown in table 1. According to Sharma's study, structural displacment that is less than 2 mm has different behavior with structural dispalcement that is more than 2 mm so he mainly focused on structural displacement that more than 2 mm. Thus, in our work, we also only focus on the experiments that make the structural displacement more than 2 mm. The corresponding response of the structure to these input signals are also recorded by Sharma (2021). To simplify our work, we will use the experiment data from input signal I, II, III, VI in Sharma (2021) thesis to training and validate the reinforcement learning-based model in our work. In previous work, researhcers identify the model using only experiment data from one input signal once upon a time due to the limitation of computation. But with the help of computer, all of these data will be used together in the training rather than separate usage like in previous work. In this way, we can get a model with strong generalization ability to fit all of these data rather than just one type of data. Sharma (2021) also tried to identify the model but the model he identified only has good performance when tested in the data which he used for identifying the model. This means this model could have bad performance in other different type of data. Thus, reinforcement learning method can fix this kind of limitation by training all data together.

Reinforment Learning Setting

For this special case study, we designed the reinforment learning algorithem as following description.

Agent

As described in seciton 'Reinforcement Learning Process and Settings in This Work', the agent in this work is set as the RNN-based prior distributions generator and this controller is able to generate the prior distritions used later in UKF-BW model. According to the requirement from UKF-BW model, there are 13 unknowns need to be identified including 6 prior mean values, 6 prior covariance values and 1 alpha value in UKF. The design of agent is shown in **Figure 5**:

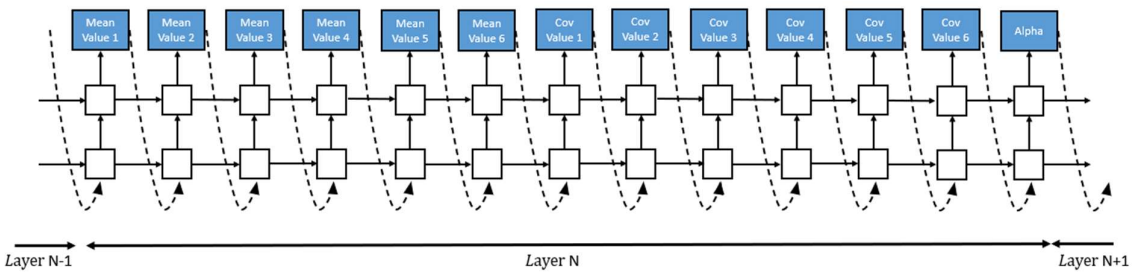


Figure 5. Agent item designed for this work

Action

As described, action item will be an array which contains the values of prior distribtuions generated from agent item. Thus, in this work, the action item is defined as a series number of which the length is 13. Action item is designed as **Figure 6** shows:

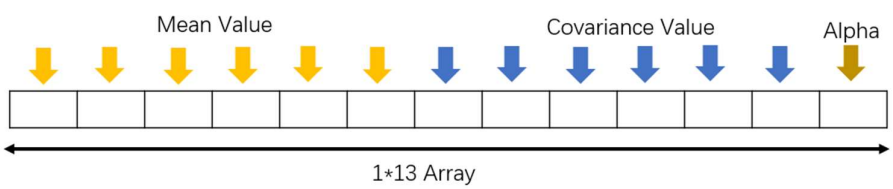


Figure 6. Action item designed for this work

Environment

As described, environment item is designed as UKF-BW model. In this work, we build such model with matlab libraries.

State

The state item is the prediction structure response and in this work, we will record the structure displacement, structure acceleration and internal force in damper as the structure response. Then the prediction of these values will be compared with experimental data. The state item is also an array composed of the arrays of predicted structure displacement, predicted structure acceleration and predicted internal force. The length of state array is the summation of the length of these three arrays.

Reward

The reward function is an index which is used for evaluate the performance of current action, which is reflected by the predicted state under current action. Thus, in this work, the predicted state of each step will be compared with experiment data and the predicted responses are supposed to be as close to experiment data as possble. Thus, we used r square as the reward functions. Then functions are designed as Equation 2:

$$r^2 = 1 - \sum \frac{(S_{exp,i,t} - S_{pre,i,t})^2}{(S_{exp,i,t} - \bar{S}_{exp,i,t})^2} \tag{2}$$

In this equation, $S_{exp,i,t}$ is i-th data point of experimental state at step t while $S_{pre,i,t}$ is i-th data point of prediction state at step t. The reward function r square will constrain the reward in the interval $(-\infty, 1]$. If the value of reward function is close to 1, the performance of prediction is better. Thus, we can also utilize a threshold for reward function to stop the reinforcement learning training. For example, if the prediction one is supposed to obtain beyond 0.9 r square, the threshold can be set as 0.9. when the r suqare is beyond this threshold, the training will be stopped. The prior distribution generated in this iteration is the answer we need for model identification. For UKF-BW model, sometimes it could generate inefficient model which will cause the error become very large, which will influence the analysis of the training process plot. To solve this problem, the the reward will be set as -1 in this situation. To build the relationship between performance and stop training threshold, the final reward function is set as Equation 3:

$$Reward = -(threshold - r^2) \tag{3}$$

With Equation 3, if the performance is closer to threshold, then the reward will become larger.

Agent update policy

According to description in section 'Reinforcement Learning Process and Settings in This Work', Once the new state and reward have been obtained in one iteration, the RNN-based prior distribution generator need to be updated. There are several different methods and policies that can be utilized for updating the agent. In this work, proximal policy optimization algorithms from (Schulman 2017) is utilized. The r square value from reward function will be used as the index of proximal policy optimization algorithms. With this update policy, agent will automatically adjust the weights in RNN of agent.

Training Setting

The training stop criterial is set by a threshold, when the performance of current iteration has achieved this threshold, the training will automatically stops. In this work, the we set three thresholds for structure displacement, structure accleration and internal force of damper as 0.9, 0.9 and 0.9, respectively. This setting is able to guarantee that model could perform well on all of predicted displacement, acceleration and internal force. Learning rate in this work is set as 0.01. A RTX 2080 Ti GPU is utilized for training the model.

RESULTS AND DISCUSSION

Results Obtained through Reinforcement Learning Method

Figure 7 shows the training process. The horizontal axis represents the training episodes and the vertical axis represents the reward of each epoch. From the plot, the phenomena can be observed that at the begninning of trainning, the agent generated many inefficient UKF-BW model which makes the episode reward keeps in -22.8. This is because the threshold set is 0.9. There are 4 experimental records used and for each experimental recors, there are 3 structure responses recorded. Thus totally we have 12 structure responses. If the UKF-BW model is inefficient in all of these predictions, the reward will be 12 * (-(0.9-(-1))) = -22.8. Thus, the more -22.8 values appear, the worse agent capacity is. From this analysis, the agent's performance is not efficient. However, with the training processing, the inefficient predicted prior distribution become less. Particularly after 450 episode, the agent is able to generate

only efficient models. The training will stop when all of 12 predicted responses achieve the thresholds. The prediction results are shown in **Figure 8** and **Figure 9**.

The results shows that the prediction responses are qualitatively close to experimental records. In next section, Quantitative analysis of the prediction performance will be compared with the performance obtained in (Sharma 2021).

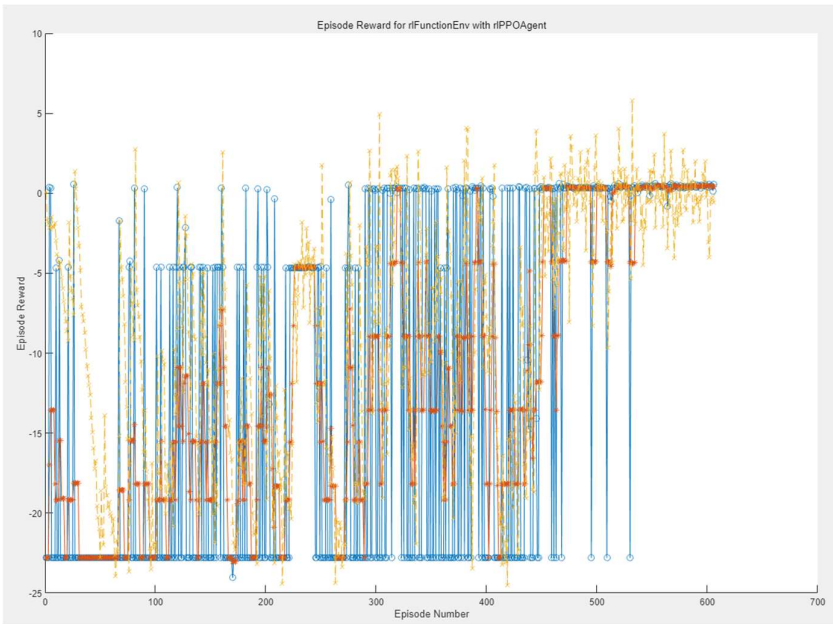
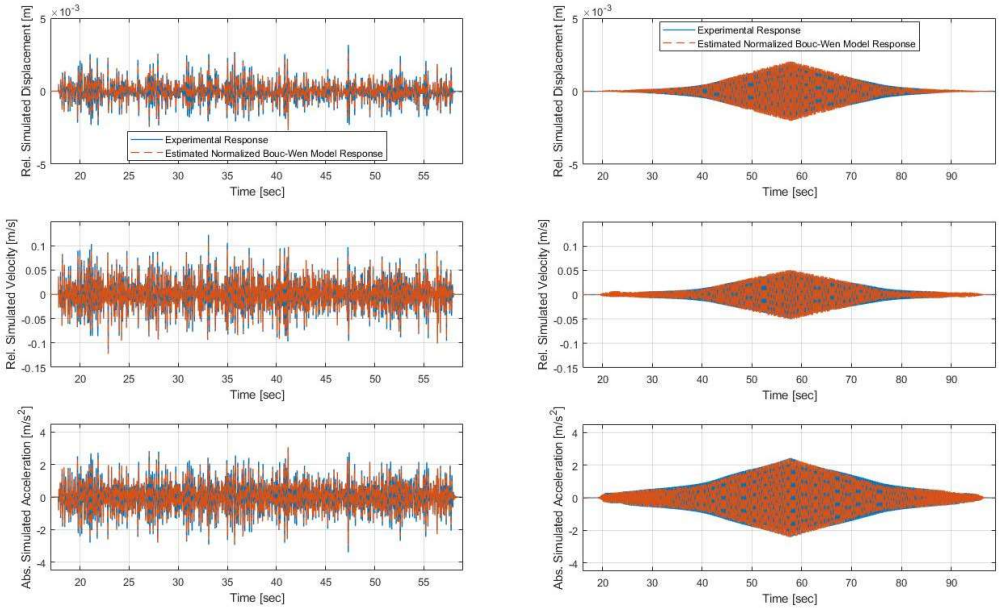
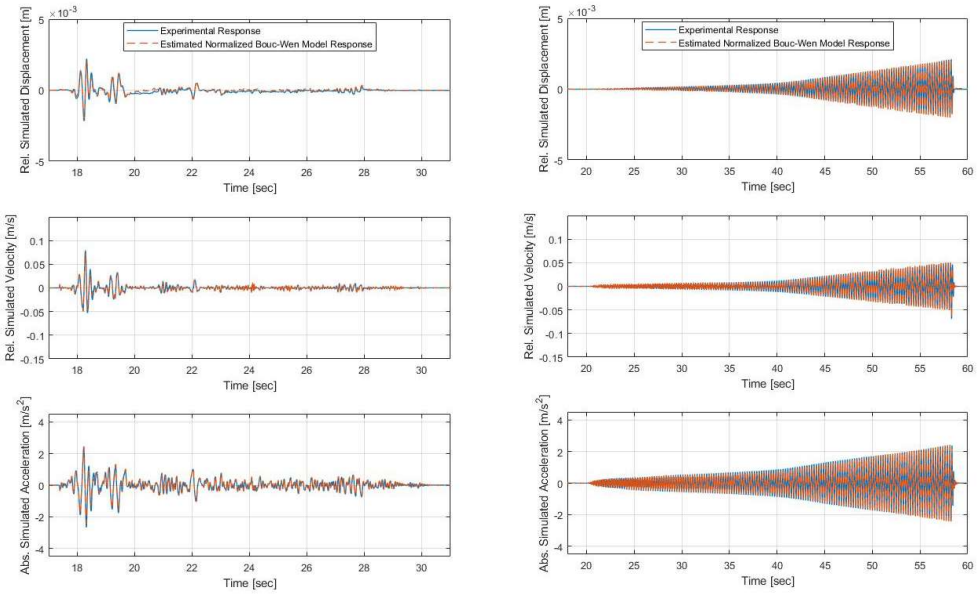


Figure 7. Training Process of Reinforcement Learning

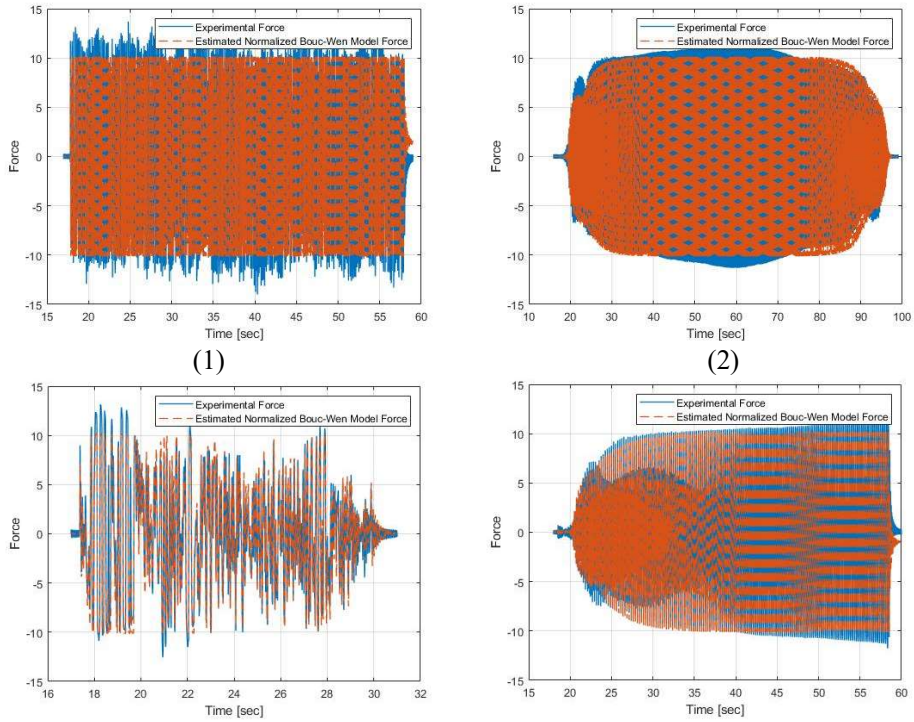


(1) (2)



(3) (4)

Figure 8. Displacement, velocity and acceleration prediction for (1) Input signal I; (2) Input signal II; (3) Input signal III; (4) Input signal VI.



(1) (2)

(3) (4)

Figure 9. Internal force in damper prediction for (1) Input signal I; (2) Input signal II; (3) Input signal III; (4) Input signal VI.

Conparison with Results of a Conventional Method

As stated in (Sharma 2021), he also tried to identify the model corresponding to the SDOF with MR damper. The method he used is very similar to (Lund et al., 2020b). He firstly iterately select 100 sets of prior distributions in a reasonable range and then compare the performances of model generated through these 100 sets using Bayesian inference method. In this section, quantitiative analysis of prediction performance through reinforcement learning will be processed. In (Sharma 2021), author used root mean square error (denoted as RMSE) and envelope area of displacement vs internal force (denoted as Ex) plot as the index to evaluate the performance of prediction. Thus, besides r square for displacement, acceleration and force, RMSE and Ex will also be calculated for comparison. Basically, we prefer larger r square, smaller RMSE and Ex. The quantitative comparison with the results from (Sharma 2021) are shown in following **Table 1** to **Table 4**.

Table 1. Comparison of prediction results for input signal I.

	$r^2{}_{disp}$	$r^2{}_{acc}$	$r^2{}_{force}$	$RMSE_{disp}$	$RMSE_{acc}$	$RMSE_{force}$	Ex
RL Method	0.9190	0.9011	0.9128	0.3413	0.3616	0.3027	0.0297
Previous	0.9200	0.9008	0.9092	0.3386	0.3629	0.3067	0.0293

Table 2. Comparison of prediction results for input signal II.

	$r^2{}_{disp}$	$r^2{}_{acc}$	$r^2{}_{force}$	$RMSE_{disp}$	$RMSE_{acc}$	$RMSE_{force}$	Ex
RL Method	0.9941	0.9882	0.9606	0.0749	0.1118	0.1956	0.0149
Previous	0.9910	0.9869	0.9587	0.0924	0.1175	0.2009	0.0139

Table 3. Comparison of prediction results for input signal III.

	$r^2{}_{disp}$	$r^2{}_{acc}$	$r^2{}_{force}$	$RMSE_{disp}$	$RMSE_{acc}$	$RMSE_{force}$	Ex
RL Method	0.9023	0.9552	0.9003	0.3340	0.2246	0.3134	0.0184
Previous	0.9016	0.9565	0.9061	0.3353	0.2230	0.3061	0.0179

Table 4. Comparison of prediction results for input signal VI.

	$r^2{}_{disp}$	$r^2{}_{acc}$	$r^2{}_{force}$	$RMSE_{disp}$	$RMSE_{acc}$	$RMSE_{force}$	Ex
RL Method	0.9913	0.9869	0.9546	0.0913	0.1177	0.2119	0.0144
Previous	0.9910	0.9869	0.9587	0.0924	0.1175	0.2009	0.0133

From above tables, the reinforcement learning method is able to achieve similar or even better results than previous method. Besides, from (Sharma 2021), the prediction from UKF-BW model itself has the limitation, which then constrains the further ability of reinforcement learning method. Besides, previous method train through experiment record under one input signal and validate in the same experiment record, which reduce the model's generalization ability. It is highly possible that model has satisfying prediction in one kind of experiment record but perform badly on another kind of experiment record. But reinforcement learning can train with all kinds of experiment records and thus it guarantees that it could have satifying prediction for all of these experiments. The prior distributions generated through

above process can be precise to 32 decimal places. Thus, through this work, reinforcement learning method is validated as a advanced tool for Bayesian inference model identification process.

CONCLUSION

Through the study of this work, some conclusions can be made:

- Reinforcement learning method is an advanced method that can be applied in Bayesian inference model identification process. In this work, it proves its ability to replace previous work and improve the generalization of the model to be identified.
- The case selected for validation in this work is not an ideal case. This is because UKF-BW model has its limitation to get more precise prediction results, which limits the reinforcement learning method to search better prior distributions to get better prediction. However, even the initial model is not ideal, reinforcement learning method still performs strong capacity to get relatively satisfying results. And after developing code, all of the identification process will be automatically finished. Besides, the prior distributions predicted are more precise than previous work.
- Further work should focus on using different reward function, e.g., involving RMSE and Ex into the reward function to obtain the model that meets the researcher's goals.

Reference List

Bishop, Christopher M. "Pattern recognition." *Machine learning* 128, no. 9 (2006).

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe. "Variational Inference: A Review for Statisticians. arXiv 2016." *arXiv preprint arXiv:1601.00670* (2016).

Howard, Ronald A. "Dynamic programming and markov processes." (1960).

Gordon, Neil J., David J. Salmond, and Adrian FM Smith. "Novel approach to nonlinear/non-Gaussian Bayesian state estimation." In *IEE Proceedings F-radar and signal processing*, vol. 140, no. 2, pp. 107-113. IET, 1993.

Lund, Alana, Ilias Bilionis, and Shirley J. Dyke. "Variational Inference for Nonlinear Structural Identification." *Journal of Applied and Computational Mechanics (2020)*.

Lund, Alana, Shirley J. Dyke, Wei Song, and Ilias Bilionis. "Identification of an experimental nonlinear energy sink device using the unscented Kalman filter." *Mechanical Systems and Signal Processing* 136 (2020): 106512.

Miura, Toru, Christian Braendle, Alexander Shingleton, Geoffroy Sisk, Srinivas Kambhampati, and David L. Stern. "A comparison of parthenogenetic and sexual embryogenesis of the pea aphid Acyrthosiphon pisum (Hemiptera: Aphidoidea)." *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 295, no. 1 (2003): 59-81.

Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. "Proximal policy optimization algorithms." *arXiv preprint arXiv:1707.06347* (2017).

Sharma, Sunny Ambalal. 2021. "STOCHASTIC MODEL GENERATION AND SELECTION FOR DEVICE EMULATING STRUCTURAL MATERIAL NONLINEARITY". Purdue University Graduate School. doi:10.25394/PGS.14449845.v1.

Stepanov, O. A. "Kalman filtering: Past and present. An outlook from Russia.(On the occasion of the

80th birthday of Rudolf Emil Kalman)." *Gyroscopy and Navigation* 2, no. 2 (2011): 99-110.

Wei, Shiyin, Yuequan Bao, and Hui Li. "Optimal policy for structure maintenance: A deep reinforcement learning framework." *Structural Safety* 83 (2020): 101906.

Zoph, Barret, and Quoc V. Le. "Neural architecture search with reinforcement learning." arXiv preprint arXiv:1611.01578 (2016).

BIOGRAPHY

Xin Zhang is a third-year civil engineering PhD at Purdue University. He received a bachelor's degree in civil engineering from Tongji University and a master's dgree from Northwestern University. His current research placement is studying the application of artificial intelligence techniques in structrue health monitoring field. His resaerch insterests include AI-based bridge inspection, reinforcement learning-based model identification and updating.

Yuguang Fu is an assistant professor in the school of civil and envionmental engineering at Nanyang Technological University. He received his PhD degree of civil engineering from University of Illinois at Urbana-Champaign and both master and bachlor degree from Tongji University. His laboratory (Laboratory for Intelligent Infrastructure Technology) is primarily focused on smart sensing and diagnostic technologies, advanced experimental techniques and their applications in infrastructure modeling, monitoring and managemenet.

Sunny Sharma is the graduate engineer in Walter P Moore company. He received his master's degree in civil engineering from Purdue University and his Bachelor's degreee in civil engineering from Veermata Jijabai Technological Institute. His research place is studying stochastic model generation and selection for device emulating structural material nonlinearity.

Shirley Dyke is the professor of mechanical engineering and civil engineering at Purdue University. She received her PhD degree in civil engineering from University of Notre Dame. Her Intelligent Infrastructure Systems Laboratory is focused on bring technological advances to infrastructure design and implementation. One of primary field is the development of the Real Time Hybrid Simulation technique. Her research insterest also includes structural dynamics and control, cyber-physical systems, machine vision, real time hybrid simulation, damage detection and structural condition monitoring and cyberinfrastructure development.