12th IFAC International Symposium on
Advanced Control of Chemical Processes
July 14-17, 2024. Toronto, Canada

Machine learning multi-step-ahead modelling with uncertainty assessment

Erbet Almeida Costa [*] Carine Menezes Rebello [*]
Vinícius Viena Santana [*] Idelfonso B. R. Nogueira [*]

[*] *Department of Chemical Engineering, Norwegian University of Science and Technology, Gløshaugen, Trondeim, Norway (e-mail: erbet.a.costa@ntnu.no, carine.m.rebello@ntnu.no, vinicius.v.santana@ntnu.no, idelfonso.b.d.r.nogueira@ntnu.no).*

Abstract: This study presents a strategy for multi-step-ahead identification of robust machine learning (ML). Hence, we focus on the disparity between standard single-step prediction models and the requirement for multi-step forecasting, which is crucial for Model Predictive and Optimization schemes. This work explores how the proposed muti-step-ahead strategy can diminish the prediction uncertainty compared to the traditional single-step approach. The paper evaluates the multi-step identification with uncertainty assessment in different model architectures, including those based on recursive neural networks. A key aspect of the analysis is the application of these models to a polymerization reactor, a standard benchmark in algorithm evaluation. The results reveal that multi-step recursive models significantly reduce prediction uncertainty compared to single-step models, particularly when feedback mechanisms are involved. This study highlights the advantages of multi-step models and their potential benefits for control and optimization schemes.

Keywords: Machine Learning Assisted Modeling; Estimation and Robust Estimation; Model Predictive Control.

1. INTRODUCTION

Dynamic modeling of chemical processes is usually related to control and optimization schemes. Hence, it is an important topic for the proper and efficient operation of these processes. On the other hand, using rigorous models for such scenarios is computationally expensive, raising the opportunity to apply surrogate models.

However, applying surrogate models in such schemes can lead to several limitations, as these models usually have good accuracy for predicting but not in simulation scenarios, where multi-step-ahead predictions are necessary. Due to these limitations, it is usual to find works in the literature that employ machine learning-based models in control and optimization schemes using a single-step approach, like Shin et al. (2020); Entezari et al. (2023); Ren et al. (2022); Norouzi et al. (2023); Afram et al. (2017); Wang et al. (2022).

Furthermore, the efficacy of these models in dynamic prediction — in terms of accuracy, robustness, and efficiency — is intrinsically linked to the model's prediction uncertainty. In environments with stringent operational constraints, high uncertainty can compromise the prediction feasibility and, therefore, the model application. Conversely, using a model with excessive uncertainty may result in overly conservative actions in more lenient settings, hindering the system's performance and objective attainment (Costa et al., 2024, 2023).

Addressing this gap, Park et al. (2023) introduces an innovative network architecture that facilitates multi-step forecasting, enabling the model to predict over a horizon matching the Model predictive control (MPC) prediction horizon in a single iteration. This approach, merging a nonlinear autoregressive exogenous model (NARX) data architecture with recurrent networks, presents numerous benefits for dynamic prediction.

Our study proposed a systematic strategy for multi-step-ahead modeling and uncertainty identification of ML-based prediction based on Park's results. We employ Bayesian inference to quantify prediction uncertainty, tackling the inference challenge through Markov chain Monte Carlo simulations. Furthermore, it studies the application of different model architectures and their influence on prediction. The developments proposed in the work are compared with the traditional single-step approach.

The paper is structured as follows: Section 2 elaborates on the methodology, Section 3 discusses a case study and presents the results, and Section 4 concludes and presents the prospects and implications of our findings.

2. METHODOLOGY

The internal prediction in model-based optimization (e.g., model predictive control, real-time optimization) is illustrated in Figure 1. The simulation is done based on a sequence of inputs $U = [u_k,\ u_{k+1},\ u_{k+2}, \dots\ u_{k+Hp}]$

⋆ This paper has been sponsored by the Norwegian Research Council.

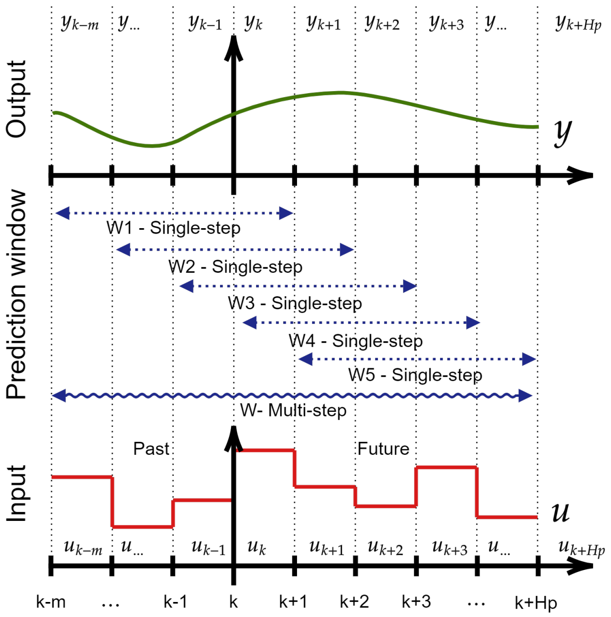

Fig. 1. Prediction window of single-step-ahead and multi-step-ahead neural networks.

up to the horizon generates respective predictions $Y = [y_k,\ y_{k+1}, \cdots\ y_{k+Hp}]$.

Incorporating neural networks into schemes as an alternative to linear models requires recurrent predictions, i.e., a simulation scenario. This process is executed multiple times in isolation from real system measurements, requiring predictions to be system-independent. Figure 1 demonstrates two approaches for constructing these networks.

Both approach utilizes a NARX architecture to predict a single step. As shown in Figure 1, prediction windows (W1 to W5) represent the predictors, constructed using past data to ascertain the system's current state. The inputs are then sequentially incorporated into the model for future state prediction, requiring the advance of the prediction windows by a one-time step for each prediction. Hence, the single-step-ahead network requires iterative prediction, where network outputs are recursively fed back into subsequent predictions. This feedback loop can lead to error propagation throughout the predictions.

The multi-step approach proposed is based on generating a sequence of predictions over a prediction horizon, Hp, preventing error propagation through feedback. This design considers both historical data and multiple forward-step predictions. Here, control actions over the prediction horizon are inserted once, allowing the model to return the complete prediction set.

The methodology posed here is depicted in Figure 2, which follows similar construction processes for single- and multi-step networks proposed by Costa et al. (2023) and Costa et al. (2024). The training data is generated using a nonlinear model with Latin Hypercube Sampling (LHS)-type input. Post-generation data is divided into training, validation, and testing sets. Networks are then identified using the Hyperband algorithm, which optimizes network architecture, including layer count, neuron number, activation functions, and other parameters. This phase encompasses training and evaluating the networks with unseen data. The subsequent phase involves assessing network uncertainty via a Markov chain Monte Carlo (MCMC) algorithm. This assessment yields network parameters' joint probability density function (PDF), enabling uncertainty propagation in network predictions. In the final stage, the methodology diverges for the two network types. For multi-step networks, uncertainty propagation is conducted through a single network query, yielding predictions for the entire horizon. In contrast, single-step-ahead networks employ feedback, where each prediction is integrated into the next time step's prediction.

Intrinsically, single-step and multi-step networks differ in the construction of predictors. In the first case, the single-step-ahead feedforward neural network (SS-FNN) is built with dense layers without internal dynamics. In turn, single-step-ahead Long short-term memory (SS-LSTM) networks have LSTM layers with internal dynamics; in this case, the construction of the network is such that the prediction is obtained from the evolution of the internal states of the network, and only the last state is returned. In the last case, multi-step-ahead long short-term memory (MS-LSTM), the network is adjusted to produce the evolution of Hp data samples. A TimeDistributed layer is added to the network output and coupled with a lambda layer that returns the desired prediction states. Then, the construction of MS-LSTM networks takes advantage of the dynamics of the LSTM network and adjusts the weights so that the internal dynamics of the network match the desired dynamics. Therefore, the difference between the neural networks relies on the predictions and uncertainty once the SS-FNN and SS-LSTM are one-step-ahead neural networks, and the MS-LSTM is a multi-step-ahead neural network that will have less computational effort and uncertainty.

3. RESULTS

The styrene polymerization reactor is an important component in the polymer production process. This paper uses a styrene polymerization reactor as a case study. The reactor mathematical model was proposed by Hidalgo and Brosilow (1990) and Alvarez and Odloak (2012). The process initiates with the decomposition of an initiator, leading to the formation of radicals. These radicals interact with monomer molecules, creating new live polymer chains. The growth of these polymer chains proceeds through a series of propagation steps, where monomers are continuously added. This growth phase ends when the propagating radicals become inactive, forming dead polymer chains. The reactor's operation is based on a phenomenological model that accounts for the short lifetime of polymer radicals, the dominant role of monomer consumption in chain propagation, and the exclusion of chain transfer reactions to monomers and solvents. Thermal initiation of the monomer is not a concern at the operational temperatures of this reactor. Moreover, the model considers the chain termination rate and the relative insignificance of initiation and termination heat compared to the heat of polymerization. The model is integral to understanding the reactor's dynamics and is pivotal in developing effective control strategies for the styrene polymerization process as given by (Alvarez and Odloak, 2012):

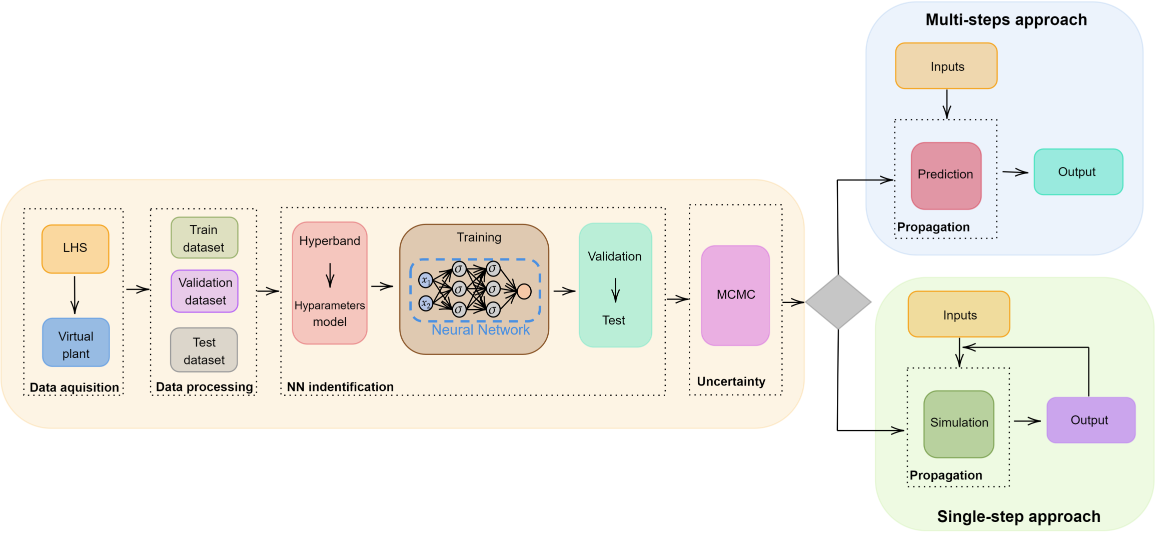

Fig. 2. Methodology chart.

$$\frac{d[M]}{dt} = \frac{Q_m[M_f] - Q_t[M]}{V} - k_p[M][P], \tag{1}$$

$$\frac{d[I]}{dt} = \frac{Q_i[I_f] - Q_t[I]}{V} - k_d[I], \tag{2}$$

$$\frac{dD_0}{dt} = 0.5k_t[P]^2 - \frac{Q_t D_0}{V}, \tag{3}$$

$$\frac{dD_1}{dt} = M_m k_p[M][P] - \frac{Q_t D_1}{V}, \tag{4}$$

$$\frac{dD_2}{dt} = 5M_m k_p[M][P] + M_m \frac{k_p^2}{k_t}[M]^2 - \frac{Q_t D_2}{V}, \tag{5}$$

$$\frac{dT}{dt} = \frac{Q_i[T_f - T]}{V} + \frac{-\Delta H_r}{\rho C_p} k_p[M][P] - \frac{hA}{\rho C_p V}(T - T_c), \tag{6}$$

$$\frac{dT_c}{dt} = \frac{Q_c(T_{cf} - T_c)}{V_c} + \frac{hA}{\rho C_p V}(T - T_c), \tag{7}$$

$$k_j = A_j exp\left(\frac{-E_j}{T}\right), \; j = d,\ p,\ t, \tag{8}$$

$$[P] = \left[\frac{2f_i k_d[I]}{k_t}\right]^{0.5}, \tag{9}$$

$$Q_t = Q_i + Q_s + Q_m, \tag{10}$$

$$\bar{M}_w = M_m \frac{D_2}{D_1}, \tag{11}$$

$$\eta = 0.0012(\bar{M}_w)^{0.71}. \tag{12}$$

This system of differential equations describes the dynamics of monomer, initiator, dead-polymer chains, and live-polymer chains within the styrene polymerization reactor. Each equation represents the rate of change of concentration of these components over time, accounting for the effects of reaction kinetics, feed rates, and volume changes. The constants k_p, k_i, k_t, and k_d are the rate constants for propagation, initiation, termination by combination, and termination by disproportionation, respectively. The variables $[M]$, $[I]$, $[P]$, and $[T]$ denote the concentrations of monomer, initiator, live-polymer chains of length n, and reactor temperature. M_f, I_f, and T_f are the inlet concentrations of monomer, initiator, and temperature, respectively. V represents the reactor volume. D_0, D_1, and D_2 are the dead polymer's respective zero, first, and second-order moments.

Viscosity (Equation 12) is a crucial variable in styrene production for several reasons. It defines the grade of the polymer produced. On the other hand, from an operational standpoint, viscosity impacts the flow and mixing within the reactor. Excessive viscosity can lead to operational challenges, such as difficulties in stirring and transferring the polymer solution. However, viscosity is a property that is hard to measure online. Thus, predicting the viscosity is vital to ensure product quality and maintain efficient and safe reactor operations. In this work, we have chosen the polymer melt viscosity of the produced styrene as a variable to be modeled, as suggested by (Alvarez and Odloak, 2012).

Identifying the artificial neural networks followed the methodology outlined in Figure 2. For each network type, we defined a specific hyperparameter search space. These hyperparameters were optimized using the Hyperband algorithm, as proposed by Li et al. (2016). The hyperparameter spaces for each network and the corresponding optimization results are detailed in Table 3. Furthermore, Figure 3 illustrates the training progression of the networks.

An early stopping strategy was employed during the training process to prevent overfitting. The training process was also configured to ensure that the optimal weights achieved during the training phase were retained when the training stopped. As observed in Figure 3, despite a decreasing trend in training loss, there was an increasing trend in validation loss. This divergence triggered the activation of the early stopping mechanism, leading to the termination of the training. This approach ensures the networks are well-tuned and generalize effectively to new data, avoiding overfitting the training dataset.

Figure 4 presents a parity plot that compares the network's performance across training, testing, and validation datasets. This parity plot is instrumental in visualizing

Table 1. Model parameters and initial conditions (Alvarez and Odloak, 2012).

Nominal Process Parameters	Value
Frequency factor for initiator decomposition, $A_d(h^{-1})$	2.142×10^{17}
Activation energy for initiator decomposition, $E_d(K)$	14897
Frequency factor for propagation reaction, $A_p(L \cdot mol^{-1} \cdot h^{-1}$	3.81×10^{10}
Activation temperature for propagation reaction, $E_p(K)$	3557
Frequency factor for termination reaction, $A_t(Lmol^{-1}h^{-1})$	4.50×10^{12}
Activation temperature for termination reaction, $E_t(K)$	843
Initiator efficiency, f_i	0.6
Heat of polymerization, $-\Delta H_r(J \cdot mol^{-1})$	6.99×10^4
Overall heat transfer coefficient, $hA(J \cdot K^{-1} \cdot L^{-1})$	1.05×10^6
Mean heat capacity of reactor fluid, $\rho C_p(JK^{-1}L^{-1})$	1506
Heat capacity of cooling jacket fluid, $\rho_c C_p c(JK^{-1}L^{-1})$	4043
Molecular weight of the monomer, $M_m(g \cdot mol^{-1})$	104.14
Initial conditions	Value
Reactor volume, $V(L)$	3000
Volume of cooling jacket fluid, $Vc(L)$	3312.4
Concentration of initiator in feed, $I_f(mol \cdot L^{-1})$	0.5888
Concentration of monomer in feed, $M_f(mol \cdot L^{-1})$	8.6981
Temperature of reactor feed, $T_f(K)$	330
Inlet temperature of cooling jacket fluid, $Tcf(K)$	295

Table 2. Steady-state inputs conditions and region of sampling for the LHS algorithm.

Variable	Steady-state	Minimum	Maximum
Flow rate of initiator, $Q_i(L \cdot h^{-1})$	108.0	91.8	124.2
Flow rate of solvent, $Q_s(L \cdot h^{-1})$	459.0	367.2	550.8
Flow rate of monomer, $Q_m(L \cdot h^{-1})$	378.0	302.4	453.6
Flow rate of cooling jacket fluid, $Q_c(L \cdot h^{-1})$	471.6	377.3	565.9

Table 3. Hyperparameters settings and results.
MS-LSTM - Multi-Step with 50 past samples and 50 future samples; SS-FNN - Single step network with dense layers; and SS-LSTM - Single Step network with LSTM layers.

		Hyperspace	
Network	MS-LSTM	SS-FNN	SS-LSTM
Num. Layers	1-4	1-3	1-3
Type of Layers	LSTM	DENSE	LSTM
Num. Neurons	4 to 20	5 to 30	2 to 32
Learning rate	1×10^{-3}	log sampling from 1×10^{-4} to 1×10^{-2}	log sampling from 1×10^{-4} to 1×10^{-2}
Activation function	tanh and sigmoid	relu or tanh orlinear]	tanh and sigmoid
Optimization algorithm	Adam	Adam	Adam
Loss and Metrics	[MSE, MAE]	[MSE, MAE]	[MSE, MAE]
		Best hyperparameters	
Num. Layers	2	2	2
Num. Neurons	[20,16]	[30,25]	[26,6]
Activation function	tanh and sigmoid	linear	tanh and sigmoid
Learning rate	1×10^{-3}	5.6×10^{-4}	8.8×10^{-3}
Trainables parameters	4465	1581	4301

the alignment between the network's predictions and the actual data across different phases of the model's evaluation. The plot demonstrates a high correlation between the predicted values and the respective datasets used for training, testing, and validation. Such a close fit signifies that the network has learned the underlying patterns effectively and can generalize well to unseen data. This level of fit is crucial for validating the model's accuracy and potential applicability in real-world scenarios.

The next step is to identify the uncertainties of the models, which, as mentioned before, is essential for the reliable application of these models. This step was done by employing each network's Markov chain Monte Carlo (MCMC) method, generating a probability distribution from 20,000 samples. The prediction uncertainty at each sampling instant was then determined based on the probability density functions (PDFs) derived from these distributions. This process involves solving the inference problem to build the PDF of the weight and biases for each network and obtain the uncertainty of the neural network parameters. Subsequently, this uncertainty propagated to the prediction and gives the prediction uncertainty.

The prediction output of each network as a function of input steps is shown in Figure 5. This figure also highlights 50 sampling instants, the predetermined number for constructing the multi-step network. All the variables are shown dimensionless in Figure 5 to make it possible to compare, given the difference in the order of magnitude of the original units. In contrast, the single-step-ahead prediction networks employed a feedback mechanism with predicted values to emulate the operation within a predictive controller, where predictions are made without mea-

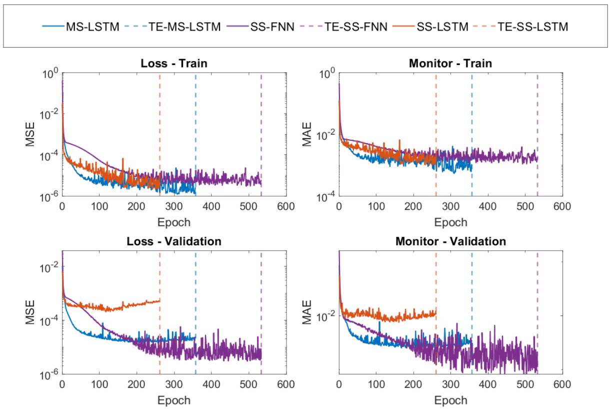

Fig. 3. Loss and monitor training and validation comparison.
MS-LSTM - Multi-Step with 50 past samples and 50 future samples; SS-FNN - Single step network with dense layers; and SS-LSTM - Single Step network with LSTM layers.

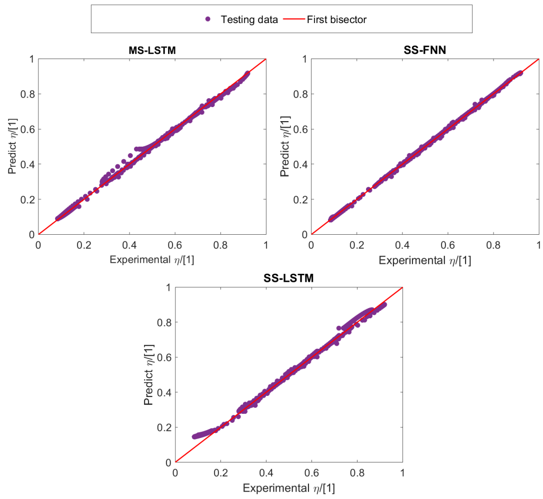

Fig. 4. Parity plot for testing data.
MS-LSTM - Multi-Step with 50 past samples and 50 future samples; SS-FNN - Single step network with dense layers; and SS-LSTM - Single Step network with LSTM layers.

suring the predicted variable directly. In this prediction, a set of 1000 weights and bias values were randomly selected for each network. Predictions were then calculated for each set, and uncertainty was calculated using the Haario et al. (2006) approach.

The analysis reveals a notable distinction in uncertainty levels. Networks using feedback, SS-FNN and SS-LSTM, exhibit significantly higher uncertainty throughout the prediction horizon than the multi-step network. The uncertainty and the most likely predicted values in the multi-step network closely align with the actual values.

Regarding the Single-Step Feedforward Neural Network (SS-FNN) and Single-Step Long Short-Term Memory (SS-LSTM) networks, the predictions for subsequent instants are made recursively, resulting in each network undergoing 50 evaluations. Conversely, the Multi-Step 50-50 (MS-LSTM) network can predict 50 future instants in a single query. This attribute substantially reduces the computational load in control applications when a multi-step prediction network is used. Figure 6 shows each network's computational time histogram during uncertainty propagation. Time was calculated for each trajectory. It is possible to observe that for the multi-step network, the computational time is more than 80% lower than that of the other networks.

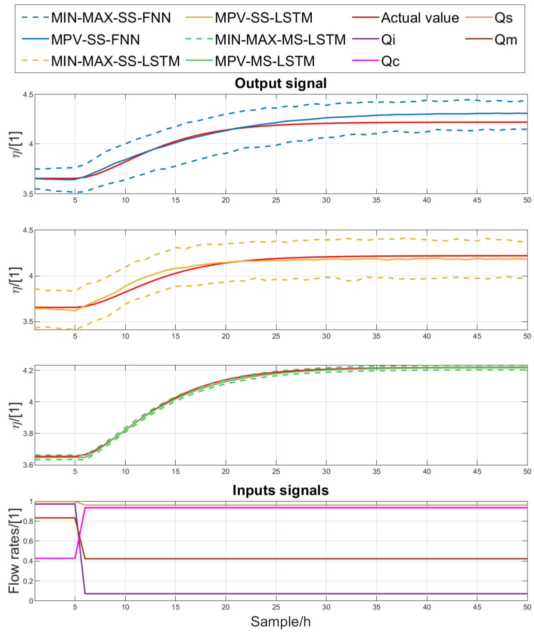

Fig. 5. Uncertainty assessment of network prediction.
MS-LSTM - Multi-Step with 50 past and 50 future samples, SS-FNN - Single step network with dense layers, and SS-LSTM - Single Step network with LSTM layers. MIN-MAX are the respective uncertainty boundaries. Qs, Qm, Qi and Qc are the inputs signals.

Comparative analysis with the results in Table 3 indicates that, despite its complexity, the MS-LSTM network, with an equivalent number of layers, neurons, and parameters, as the SS-LSTM network, exhibits lower prediction uncertainty. This finding underscores the MS-LSTM network's efficiency in terms of computational cost and predictive reliability.

4. CONCLUSION

This article presented a methodology for modeling and uncertainty analysis of multi-step-ahead machine learning models for dynamic modeling. The proposed methodology is based on training and validating data-based models with algorithms and evaluating model uncertainty using Bayesian inference and Markov Chain Monte Carlo algorithms to solve the inference problem.

The case study presented shows the dynamic modeling of a polymerization reactor commonly used as a benchmark for control algorithms. The results show that the models

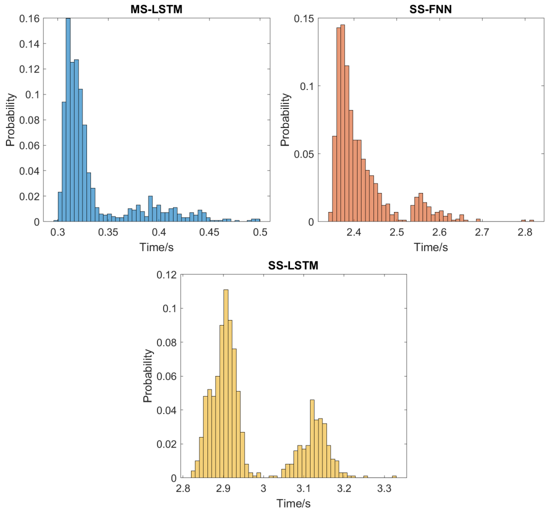

Fig. 6. Computational time used for prediction. MS-LSTM - Multi-Step with 50 past samples and 50 future samples, SS-FNN - Single step network with dense layers, and SS-LSTM - Single Step network with LSTM layers.

built to predict a sampling instant have considerably more significant uncertainty than those built to predict multiple steps. Multi-step networks predict HP steps are performed in a single model query. The computational time of the multi-step networks is more than 80 % lower than the single-step networks. On the other hand, networks with NARX architecture have more significant computational effort, as the prediction needs to be performed recursively. From an uncertainty point of view, multi-step networks presented lower uncertainty than single-step networks.

ACKNOWLEDGEMENTS

The present work contributes to completing a sub-project at SUBPRO, a research-based innovation center within Subsea Production and Processing at the Norwegian University of Science and Technology. The authors would like to express their gratitude for the financial support from SUBPRO, funded by the Research Council of Norway through grant number 237893, major industry partners, and NTNU.

REFERENCES

Afram, A., Janabi-Sharifi, F., Fung, A.S., and Raahemifar, K. (2017). Artificial neural network (ann) based model predictive control (mpc) and optimization of hvac systems: A state of the art review and case study of a residential hvac system. *Energy and Buildings*, 141, 96–113. doi:https://doi.org/10.1016/j.enbuild.2017.02.012.

Alvarez, L.A. and Odloak, D. (2012). Optimization and control of a continuous polymerization reactor. *Brazilian Journal of Chemical Engineering*, 29(4), 807–820. doi:10.1590/S0104-66322012000400012. URL https://doi.org/10.1590/S0104-66322012000400012.

Costa, E.A., Rebello, C.d.M., Fontana, M., Schnitman, L., and Nogueira, I.B.d.R. (2023). A robust learning methodology for uncertainty-aware scientific machine learning models. *Mathematics*, 11(1). doi:10.3390/math11010074.

Costa, E.A., Rebello, C.M., Schnitman, L., Loureiro, J.M., Ribeiro, A.M., and Nogueira, I.B. (2024). Adaptive digital twin for pressure swing adsorption systems: Integrating a novel feedback tracking system, online learning and uncertainty assessment for enhanced performance. *Engineering Applications of Artificial Intelligence*, 127, 107364. doi:https://doi.org/10.1016/j.engappai.2023.107364.

Entezari, A., Aslani, A., Zahedi, R., and Noorollahi, Y. (2023). Artificial intelligence and machine learning in energy systems: A bibliographic perspective. *Energy Strategy Reviews*, 45, 101017. doi:https://doi.org/10.1016/j.esr.2022.101017.

Haario, H., Laine, M., Mira, A., and Saksman, E. (2006). Dram: efficient adaptive mcmc. *Statistics and computing*, 16, 339–354.

Hidalgo, P. and Brosilow, C. (1990). Nonlinear model predictive control of styrene polymerization at unstable operating points. *Computers & Chemical Engineering*, 14(4), 481–494. doi:https://doi.org/10.1016/0098-1354(90)87022-H.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., and Talwalkar, A. (2016). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18, 1–52. URL http://arxiv.org/abs/1603.06560.

Norouzi, A., Heidarifar, H., Borhan, H., Shahbakhti, M., and Koch, C.R. (2023). Integrating machine learning and model predictive control for automotive applications: A review and future directions. *Engineering Applications of Artificial Intelligence*, 120, 105878. doi:https://doi.org/10.1016/j.engappai.2023.105878.

Park, J., Babaei, M.R., Munoz, S.A., Venkat, A.N., and Hedengren, J.D. (2023). Simultaneous multistep transformer architecture for model predictive control. *Computers & Chemical Engineering*, 178, 108396. doi:https://doi.org/10.1016/j.compchemeng.2023.108396.

Ren, Y.M., Alhajeri, M.S., Luo, J., Chen, S., Abdullah, F., Wu, Z., and Christofides, P.D. (2022). A tutorial review of neural network modeling approaches for model predictive control. *Computers & Chemical Engineering*, 165, 107956. doi:https://doi.org/10.1016/j.compchemeng.2022.107956.

Shin, Y., Smith, R., and Hwang, S. (2020). Development of model predictive control system using an artificial neural network: A case study with a distillation column. *Journal of Cleaner Production*, 277, 124124. doi:https://doi.org/10.1016/j.jclepro.2020.124124.

Wang, D., Shen, Z.J., Yin, X., Tang, S., Liu, X., Zhang, C., Wang, J., Rodriguez, J., and Norambuena, M. (2022). Model predictive control using artificial neural network for power converters. *IEEE Transactions on Industrial Electronics*, 69(4), 3689–3699. doi:10.1109/TIE.2021.3076721.