ORA Mathematics, 1(1), e2026001
DOI: 10.XXXXX/oramath.e2026001
Received: 15 April 2026
Revised: 01 May 2026
Accepted: 05 May 2026
Published: 08 May 2026

RESEARCH ARTICLE

Variational Bayesian identification for Hammerstein large-scale stochastic systems: A comparative simulation study of hydraulic process

Madiha Ghamkhar*

Department of Mathematics and Statistics, University of Agricultural Faisalabad, Pakistan

* **Correspondence:** Email: Madiha.ghamkhar@uaf.edu.pk

Abstract: Large-scale interconnected Hammerstein systems operating under stochastic disturbances present persistent challenges for identification, particularly when coupling between subsystems is strong and operating noise is non-negligible. This paper develops a Variational Bayesian Hammerstein Identification (VBHI) framework for such systems, where static nonlinear blocks are modelled by feedforward neural approximators and the linear dynamic parameters are estimated through a structured mean-field variational posterior. By treating all identifiable parameters as latent random variables and maximising a closed-form Evidence Lower Bound (ELOB), the proposed method delivers calibrated posterior uncertainty on every parameter block simultaneously, without recourse to expensive Markov Chain Monte Carlo sampling. The VBHI framework inherits a block-oriented Hammerstein structure and is designed to accommodate strong subsystem interconnections, time-varying coefficients, and measurement noise. Convergence of the variational update recursion is established through a Lyapunov-type descent argument applied to the negative ELOB. The method is evaluated on a two-subsystem numerical benchmark and a three-tank hydraulic-process case study, and compared against a conventional Recursive Extended Least Squares (RELS) baseline. Across both experiments, VBHI achieves Root Mean Square Error (RMSE) reductions of approximately 37–41%, prediction-error variance reductions of roughly 62–65%, and provides explicit confidence intervals that RELS cannot supply, at the cost of a moderate increase in per-sample computation. These results confirm that the proposed VBHI approach is a principled and practically effective solution for identifying noisy large-scale interconnected Hammerstein systems with built-in uncertainty quantification.

Keywords: System identification; Hammerstein model; Variational Bayesian Hammerstein Identification (VBHI); Large-scale interconnected systems; Uncertainty quantification; Evidence Lower Bound (ELOB); Hydraulic process

Mathematics Subject Classification: 68T07, 93A30, 93B30, 93C56

1. Introduction

One of the persistent issues in control engineering is determining the dynamics of large-scale nonlinear systems made up of interconnected subsystems. Standard identification schemes encounter a compound challenge when subsystems are coupled through both their inputs and outputs and when measurement noise is introduced by the operating environment: the parameter

estimates must simultaneously track time-varying dynamics, suppress cross-subsystem coupling bias, and remain robust to colored or impulsive noise [2, 10]. For many industrial processes, such as hydraulic systems, heat exchangers, and chemical reactors, block-oriented Hammerstein structures—a static nonlinear stage followed by a linear dynamic stage—offer an understandable and economical description [2, 17]. However, when subsystem interconnections are strong, the classical estimation strategies for these structures—such as polynomial approximation of the nonlinear block combined with Recursive Extended Least Squares (RELS)—produce point estimates that carry no uncertainty information and can show significant bias [8, 9].

In order to approximate the static nonlinear blocks of Hammerstein models without imposing restrictive functional forms, neural networks (NNs) have emerged as a preferred method [3]. Deeper parameterizations frequently outperform shallow polynomial expansions for strongly nonlinear blocks, and recent research has demonstrated that feedforward and recurrent architectures can reliably learn complex mappings in both identification and forecasting settings [11, 15]. The usefulness of flexible nonlinear approximators in recursive identification has also been demonstrated by the development of kernel-based recursive techniques for online energy-efficiency estimation in large-scale chemical plants [10]. Filtering-based recursive identification techniques for fractional-order Hammerstein models with piecewise nonlinearity have recently been reported [8, 9, 14]. These developments have coincided with an increase in interest in these models.

A practically significant gap still exists in spite of this advancement. The uncertainty in the recovered parameters is not quantified by the point estimates generated by gradient-based or recursive identification techniques. A practitioner requires both the best parameter estimate and a trustworthy measure of its confidence in safety-critical applications like hydraulic level regulation or precision process control. A natural framework for meeting this need is provided by Bayesian inference: an estimate and its associated uncertainty are obtained in a single coherent computation by treating parameters as random variables and computing a posterior distribution given the observed data [20, 25]. The main computational barrier to Bayesian identification of Hammerstein systems is that Markov Chain Monte Carlo methods scale poorly to the high-dimensional, time-varying setting of large-scale interconnected systems, and exact posterior inference is unfeasible for the nonlinear block parameters [6].

By substituting a structured approximate distribution from a tractable family for the intractable posterior and then optimizing a closed-form Evidence Lower Bound (ELOB) to get that approximation as close to the true posterior as possible, variational inference overcomes this challenge [6, 18]. The ELOB is separable due to the mean-field factorization, which assumes posterior independence between the linear dynamic parameters of each subsystem and the nonlinear block parameters. As a result, coordinate-ascent updates can be interspersed with recursive estimation of the linear part [1, 13]. The hybrid backpropagation-plus-RELS structure, which has been successful for deterministic Hammerstein identification, is mirrored in this combination—variational updates for the NN weights and recursive least-squares updates for the linear parameters—but it adds the vital advantage of posterior uncertainty [3, 15].

Recursive and adaptive identification algorithms have long been shown to converge using Lyapunov-based arguments [4, 5, 21]. Such analyses have recently been extended to gradient-based optimization techniques with generalized convergence certificates and variational update recursions [13, 18, 20]. In this paper, an explicit step-size condition that guarantees mono-

tone ascent of the ELOB—and consequently convergence of the variational recursion—under persistent excitation and bounded-sensitivity conditions is derived by taking advantage of the relationship between ELOB maximization and energy decrease in a Lyapunov sense [23, 24]. The second part of the suggested algorithm [22] is the forgetting-factor adaptation policy for the recursive linear update, which is informed by complementary stability tools derived from predefined-time and barrier Lyapunov function analyses [4, 5].

This paper presents a Variational Bayesian Hammerstein Identification (VBHI) framework for stochastic disturbances in large-scale interconnected Hammerstein systems. What sets it apart are:

- **Architecture:** A VBHI model that preserves the block-oriented Hammerstein structure while replacing deterministic NN weights with Gaussian variational posteriors, yielding calibrated uncertainty on all nonlinear block parameters.
- **Hybrid variational-recursive learning:** A two-stage algorithm coupling coordinate-ascent variational updates for the NN posterior with adaptive forgetting-factor RELS for the linear dynamic parameters, suitable for time-varying stochastic interconnected settings [7, 12].
- **Convergence guarantees:** A Lyapunov-type descent argument applied to the negative ELOB, establishing monotone ascent and convergence conditions consistent with established recursive identification theory [18, 21].
- **Uncertainty quantification:** Explicit per-parameter posterior credible intervals, which are unavailable from point-estimate baselines and are particularly valuable in safety-critical process control [16, 19].

Using the same dataset parameters as in the reference study for this class of systems [2, 10], the VBHI framework is assessed against a traditional RELS baseline on a two-subsystem numerical benchmark and a three-tank hydraulic-process case study. The simulation results provide uncertainty certificates that point-estimate methods are unable to provide and confirm steady increases in prediction accuracy. Additionally, a game-theoretic extension for multi-agent networked identification and an event-triggered communication scheme for bandwidth-limited distributed deployments are discussed as logical extensions of the suggested framework [7, 12].

This is how the paper is structured. The discrete-time model for large-scale interconnected Hammerstein systems is formalized in Section 2. The hybrid variational-recursive algorithm, the convergence analysis, and the VBHI identification framework are presented in Section 3. The numerical and hydraulic-process simulations with quantitative comparisons are presented in Section 4. Views on distributed deployment, fault-aware extensions, and uncertainty-conscious process monitoring round out Section 5.

2. Mathematical Modelling Of Nonlinear Interconnected Systems

An effective identification framework requires a precise mathematical model of the plant to be identified. This section develops a discrete-time description for large-scale nonlinear systems composed of interconnected monovariable subsystems, where structural variables such as order and delay are assumed known but the time-varying parameters are unknown and must be recovered online [2, 17].

2.1. System Description

Consider a large-scale nonlinear system $\mathcal{S}$ comprising N interconnected Hammerstein subsystems $\mathcal{S}_1, \ldots, \mathcal{S}_N$. Each subsystem operates under stochastic disturbances, and the subsystem parameters vary slowly with time. The dynamic linear part of subsystem $\mathcal{S}_i$ is governed by

$$A_i(q^{-1},k)\, y_i(k) = B_i(q^{-1},k)\, x_i^{u_i}(k) + \sum_{\substack{j=1 \\ j\neq i}}^{N} B_{ij}(q^{-1},k)\, x_j^{u_j}(k) + \sum_{\substack{j=1 \\ j\neq i}}^{N} A_{ij}(q^{-1},k)\, x_j^{y_j}(k) + C_i(q^{-1})\, e_i(k), \tag{1}$$

where $x_i^{u_i}(k)$ and $y_i(k)$ are the input and output of the linear dynamic block at discrete-time k; $u_i(k)$, $u_j(k)$, and $y_j(k)$ denote the inputs to the static nonlinear blocks; $x_j^{u_j}(k)$ and $x_j^{y_j}(k)$ are the interconnection signals from subsystem $\mathcal{S}_j$; and $\{e_i(k)\}$ is an independent sequence of zero-mean random variables with constant variance σ_i^2. The backward-shift polynomials are

$$A_i(q^{-1},k) = 1 + a_{i,1}(k)q^{-1} + \cdots + a_{i,n_{A_i}}(k)q^{-n_{A_i}}, \tag{2}$$

$$B_i(q^{-1},k) = b_{i,1}(k)q^{-1} + \cdots + b_{i,n_{B_i}}(k)q^{-n_{B_i}}, \tag{3}$$

$$A_{ij}(q^{-1},k) = 1 + a_{ij,1}(k)q^{-1} + \cdots + a_{ij,n_{A_{ij}}}(k)q^{-n_{A_{ij}}}, \tag{4}$$

$$B_{ij}(q^{-1},k) = b_{ij,1}(k)q^{-1} + \cdots + b_{ij,n_{B_{ij}}}(k)q^{-n_{B_{ij}}}, \tag{5}$$

$$C_i(q^{-1}) = 1 + c_{i,1}q^{-1} + \cdots + c_{i,n_{C_i}}q^{-n_{C_i}}, \tag{6}$$

for $i, j = 1, \ldots, N$, $j \neq i$, with polynomial orders n_{A_i}, n_{B_i}, $n_{A_{ij}}$, $n_{B_{ij}}$, n_{C_i}.

2.2. Static Nonlinear Blocks

The static nonlinear blocks transform each external input and interconnection output into an intermediate signal:

$$x_i^{u_i}(k) = f_{x_i^{u_i}}\big[u_i(k)\big], \tag{7}$$

$$x_j^{u_j}(k) = f_{x_j^{u_j}}\big[u_j(k)\big], \tag{8}$$

$$x_j^{y_j}(k) = f_{x_j^{y_j}}\big[y_j(k)\big]. \tag{9}$$

In contrast to polynomial approximations, which require a pre-specified degree and may require high order to capture strongly nonlinear behaviour [8, 9], the VBHI framework models each function in Equations (7)–(9) with a probabilistic feedforward neural network whose weight posterior is maintained and updated as a Gaussian distribution. This choice is motivated by the universal approximation capability of such networks and by the desire to propagate parameter uncertainty through the identification recursion [3, 15].

2.3. Expanded System Output Equation

Combining Equation (1) with the neural approximations of the nonlinear blocks and assuming a common polynomial order n_i for all polynomials associated with subsystem $\mathcal{S}_i$, the system

output can be written as

$$
\begin{aligned}
y_i(k) = &- \sum_{h=1}^{n_{A_i}} a_{i,h}(k)\, y_i(k-h) + \sum_{h=1}^{n_{B_i}} b_{i,h}(k)\, \hat{x}_i^{u_i}(k-h) \\
&+ \sum_{\substack{j=1 \\ j\neq i}}^{N} \sum_{h=1}^{n_{B_{ij}}} b_{ij,h}(k)\, \hat{x}_j^{u_j}(k-h) + \sum_{\substack{j=1 \\ j\neq i}}^{N} \sum_{h=1}^{n_{A_{ij}}} a_{ij,h}(k)\, \hat{x}_j^{y_j}(k-h) \\
&+ \sum_{h=1}^{n_{C_i}} c_{i,h}\, e_i(k-h) + e_i(k),
\end{aligned} \tag{10}
$$

where $\hat{x}_i^{u_i}$, $\hat{x}_j^{u_j}$, and $\hat{x}_j^{y_j}$ denote the neural network outputs corresponding to each nonlinear block. Equation (10) forms the prediction model used in both the ELOB computation and the recursive linear update [2, 11].

3. System Identification Using VBHI

This section develops the VBHI identification framework. The proposed methodology extends the conventional RELS approach by placing Gaussian variational posteriors over the neural-network weight parameters and performing coordinate-ascent ELOB maximisation for those parameters, while retaining adaptive forgetting-factor RELS for the linear dynamic parameters. The result is a hybrid algorithm that produces both parameter estimates and associated uncertainty certificates at every time step [3, 15, 20].

3.1. Probabilistic Neural Network Architecture For Interconnected Hammerstein Systems

The VBHI architecture mirrors the block-oriented structure of the interconnected Hammerstein model and assigns a Gaussian variational posterior to the weights of every neural block. For subsystem $\mathcal{S}_i$, three neural components are defined.

3.1.1. Input Neural Blocks: The input-side nonlinear functions $f_{x_i^{u_i}}[\cdot]$ and $f_{x_j^{u_j}}[\cdot]$ are each approximated by a feedforward network with variational weight distribution $q(\Theta_i^{u_i}) = \mathcal{N}\big(\mu_i^{u_i}, \mathrm{diag}(\sigma_i^{u_i})^2\big)$, so that

$$
x_i^{u_i}(k) = \mathrm{NN}_{u_i}\big(\Theta_i^{u_i}; u_i(k)\big), \tag{11}
$$

$$
x_j^{u_j}(k) = \mathrm{NN}_{u_j}\big(\Theta_i^{u_j}; u_j(k)\big). \tag{12}
$$

The mean vector $\mu_i^{u_i}$ provides the point estimate used in the prediction, while the diagonal covariance $\mathrm{diag}(\sigma_i^{u_i})^2$ quantifies the posterior uncertainty on each weight [3, 15].

3.1.2. Output Neural Blocks: The interconnection nonlinear functions $f_{x_j^{y_j}}[\cdot]$ that map the output of a neighbouring subsystem $\mathcal{S}_j$ to the coupling signal are approximated by networks with variational posteriors $q(\Theta_i^{y_j}) = \mathcal{N}\big(\mu_i^{y_j}, \mathrm{diag}(\sigma_i^{y_j})^2\big)$:

$$
x_j^{y_j}(k) = \mathrm{NN}_{y_j}\big(\Theta_i^{y_j}; y_j(k)\big). \tag{13}
$$

3.1.3. Complete Parameter Vector: The full parameter vector for subsystem $\mathcal{S}_i$ is

$$
\Theta_i(k) = \big[(\theta_i^L(k))^T,\ (\mu_i^{u_i}(k))^T,\ (\sigma_i^{u_i}(k))^T,\ (\mu_i^{u_j}(k))^T,\ (\sigma_i^{u_j}(k))^T,\ (\mu_i^{y_j}(k))^T,\ (\sigma_i^{y_j}(k))^T\big]^T, \tag{14}
$$

where $\theta_i^L(k)$ collects the linear dynamic parameters and the remaining entries are the variational mean and variance vectors of the neural posteriors. The probabilistic structure in Equation (14) distinguishes the VBHI from the deterministic IHNN: every weight is described by a distribution rather than a single number, enabling uncertainty quantification throughout the identification process [16, 19].

3.2. Evidence Lower Bound Formulation

Let $\mathcal{Y}_k = \{y_i(s)\}_{s=1}^k$ denote the observation sequence up to time k for subsystem $\mathcal{S}_i$. The log marginal likelihood of the observations can be bounded below as

$$\log p\big(\mathcal{Y}_k\big) \geq \mathcal{L}_i(k) \triangleq \mathbb{E}_q\big[\log p\big(\mathcal{Y}_k \mid \Theta_i^{\mathrm{NN}}, \theta_i^L\big)\big] - D_{\mathrm{KL}}\big(q(\Theta_i^{\mathrm{NN}}) \parallel p(\Theta_i^{\mathrm{NN}})\big), \tag{15}$$

where $p(\Theta_i^{\mathrm{NN}})$ is an isotropic Gaussian prior with zero mean and variance τ^2, and D_{KL} denotes the Kullback–Leibler divergence. Under the Gaussian mean-field assumption, the KL term is available in closed form:

$$D_{\mathrm{KL}}\big(\mathcal{N}(\mu, \sigma^2) \parallel \mathcal{N}(0, \tau^2)\big) = \frac{1}{2}\left[\frac{\mu^2 + \sigma^2}{\tau^2} - \log\frac{\sigma^2}{\tau^2} - 1\right]. \tag{16}$$

The expected log-likelihood in Equation (15) is approximated using the local Gaussian noise model $p(y_i(k) \mid \Theta_i) = \mathcal{N}(\hat{y}_i(k), \sigma_{e,i}^2)$, giving

$$\mathbb{E}_q\big[\log p\big(y_i(k) \mid \Theta_i\big)\big] \approx -\frac{1}{2\sigma_{e,i}^2}\Big[e_i(k)^2 + \mathrm{tr}\big(J_i(k)\,\mathrm{diag}(\sigma_i^2)J_i(k)^T\big)\Big] - \frac{1}{2}\log(2\pi\sigma_{e,i}^2), \tag{17}$$

where $e_i(k) = y_i(k) - \hat{y}_i(k, \mu_i)$ is the one-step prediction error evaluated at the posterior mean, and $J_i(k) = \partial\hat{y}_i/\partial\Theta_i^{\mathrm{NN}}$ is the Jacobian of the predicted output with respect to the NN weight means. The trace term in Equation (17) penalises the contribution of weight uncertainty to prediction variance and naturally regularises the posterior widths [13, 18].

3.3. Parameter Estimation Algorithm

The VBHI parameter update alternates between two complementary steps at every sample k: a coordinate-ascent variational update for the neural posteriors $(\mu_i^{\mathrm{NN}}, \sigma_i^{\mathrm{NN}})$, and a recursive least-squares update for the linear dynamic parameters θ_i^L.

3.3.1. Variational Update for Neural Posteriors (VB-NN): The ELOB gradient with respect to the posterior mean μ_i is

$$\frac{\partial\mathcal{L}_i(k)}{\partial\mu_i} = \frac{e_i(k)}{\sigma_{e,i}^2}\,J_i(k) - \frac{\mu_i}{\tau^2}, \tag{18}$$

and with respect to the posterior log-variance $s_i = \log\sigma_i^2$ it is

$$\frac{\partial\mathcal{L}_i(k)}{\partial s_i} = -\frac{1}{2\sigma_{e,i}^2}\,\mathrm{diag}\big(J_i(k)^T J_i(k)\big)\,\sigma_i^2 - \frac{\sigma_i^2}{2\tau^2} + \frac{1}{2}, \tag{19}$$

where the diagonal structure of the mean-field factorisation makes all operations element-wise and therefore computationally efficient [6]. The variational means and log-variances are updated by gradient ascent on the ELOB:

$$\mu_i^{\mathrm{NN}}(k+1) = \mu_i^{\mathrm{NN}}(k) + \eta_i^{\mu}(k)\,\frac{\partial\mathcal{L}_i(k)}{\partial\mu_i}, \tag{20}$$

$$s_i(k+1) = s_i(k) + \eta_i^{s}(k)\,\frac{\partial\mathcal{L}_i(k)}{\partial s_i}, \tag{21}$$

where $\eta_i^{\mu}(k)$ and $\eta_i^{s}(k)$ are positive step sizes whose admissible range is derived from the convergence analysis in Section 3.5. After each update, $\sigma_i^2(k+1) = \exp(s_i(k+1))$ is enforced to maintain positive variance [23, 24].

3.3.2. Recursive Least Squares Update for Linear Parameters (VB-RLS): Conditioned on the current neural posterior means, the linear dynamic parameters are updated by the standard recursive algorithm with adaptive forgetting factor:

$$\theta_i^L(k) = \theta_i^L(k-1) + K_i(k)\big[y_i(k) - \varphi_i^T(k)\theta_i^L(k-1)\big], \tag{22}$$

$$K_i(k) = \frac{P_i(k-1)\varphi_i(k)}{\lambda_i(k) + \varphi_i^T(k)P_i(k-1)\varphi_i(k)}, \tag{23}$$

$$P_i(k) = \frac{1}{\lambda_i(k)}\big[P_i(k-1) - K_i(k)\varphi_i^T(k)P_i(k-1)\big], \tag{24}$$

where $\varphi_i(k)$ is the regression vector assembled from past inputs, outputs, and prediction errors evaluated at the current NN posterior means; $K_i(k)$ is the Kalman gain; $P_i(k)$ is the covariance matrix; and $\lambda_i(k)$ is the adaptive forgetting factor [4, 21]. The forgetting factor is adapted online according to

$$\lambda_i(k) = \lambda_{\max} - (\lambda_{\max} - \lambda_{\min})\,\frac{\|e_i(k)\|^2}{\|e_i(k)\|^2 + \rho}, \quad \lambda_{\max} \in [0.99, 0.995],\ \lambda_{\min} \in [0.95, 0.98],\ \rho > 0, \tag{25}$$

so that large prediction errors induce faster parameter tracking while small errors preserve variance reduction [5, 22].

3.3.3. Noise Variance Adaptation: The noise variance $\sigma_{e,i}^2$ appearing in the ELOB (Equations (15)–(17)) is updated online using a running sample estimate:

$$\hat{\sigma}_{e,i}^2(k) = (1-\rho_e)\,\hat{\sigma}_{e,i}^2(k-1) + \rho_e\, e_i(k)^2, \quad \rho_e \in (0,1), \tag{26}$$

which ensures that the ELOB penalises residuals relative to the current noise level rather than a fixed design parameter [20].

3.4. Hybrid Variational-Recursive Training Algorithm

The complete VBHI training procedure is summarised in Algorithm 1.

Table 1. Hybrid variational-recursive training algorithm for VBHI (Algorithm 1).

Algorithm 1 Hybrid variational-recursive training for VBHI

Initialise $\mu_i^{\mathrm{NN}}(0)$ with Xavier scheme; $\sigma_i^{\mathrm{NN}}(0) \leftarrow \epsilon_\sigma \mathbf{1}$, ϵ_σ small positive
Initialise $\theta_i^L(0)$ by OLS on 50–100 warm-start samples
Set $P_i(0) = \delta^{-1} I$, $\delta \in [10^{-2}, 10^{-1}]$; $\hat{\sigma}_{e,i}^2(0) = 1$
for $k = 1$ to K **do**
  for each subsystem $\mathcal{S}_i$ **do**
    Compute NN outputs $\hat{x}_i^{u_i}(k)$, $\hat{x}_j^{u_j}(k)$, $\hat{x}_j^{y_j}(k)$ at posterior means $\mu_i^{\mathrm{NN}}(k)$
    Assemble regression vector $\varphi_i(k)$
    Update linear parameters via VB-RLS: $\theta_i^L(k)$
    Compute prediction error $e_i(k) = y_i(k) - \varphi_i^T(k)\theta_i^L(k)$
    Update noise variance: $\hat{\sigma}_{e,i}^2(k)$
    Compute Jacobian $J_i(k) = \partial \hat{y}_i / \partial \mu_i^{\mathrm{NN}}$ via backpropagation
    Compute ELOB gradients (Equations (18)–(19))
    Update variational parameters: $\mu_i^{\mathrm{NN}}(k+1)$, $\sigma_i^{\mathrm{NN}}(k+1)$
  end for
end for

3.5. Convergence Analysis

The convergence of the VBHI variational update is established through a Lyapunov-type energy argument applied to the negative ELOB, which plays the role of a Lyapunov function for the mean-parameter recursion [13, 18, 20].

Consider the scalar candidate

$$V_i(k) = -\mathcal{L}_i(k) = \frac{1}{2\sigma_{e,i}^2}\left[e_i(k)^2 + \mathrm{tr}(J_i(k)\Sigma_i(k)J_i(k)^T)\right] + D_{\mathrm{KL}}\big(q(\Theta_i^{\mathrm{NN}}) \parallel p(\Theta_i^{\mathrm{NN}})\big) + \mathrm{const}, \quad (27)$$

where $\Sigma_i(k) = \mathrm{diag}(\sigma_i^{\mathrm{NN}}(k))^2$. Since $V_i(k) \geq 0$ for all k (the KL divergence is non-negative and the squared-error term is non-negative), V_i constitutes a valid Lyapunov candidate. ELOB maximisation is equivalent to minimising V_i; convergence of the algorithm therefore requires that the update rules in Equations (20)–(21) produce a strictly decreasing sequence $\{V_i(k)\}$ [4, 25].

Theorem 3.1. *Under the following conditions:*

(1) The system satisfies the persistent excitation condition;

(2) The initial variational parameters $\mu_i^{\mathrm{NN}}(0)$ and $\sigma_i^{\mathrm{NN}}(0)$ are finite;

(3) The covariance matrix $P_i(k)$ remains bounded and converges to a steady-state value;

(4) The noise sequence $e_i(k)$ is independent, uncorrelated with the regressor $\varphi_i(k)$, and Gaussian with zero mean and constant variance;

(5) The forgetting factor satisfies $0 < \lambda_i(k) < 1$*;*

(6) The Jacobian $J_i(k)$ *is uniformly bounded, with* $\zeta_{i,\max} = \max_k \|J_i(k)\|$*;*

(7) The mean step size satisfies

$$0 < \eta_i^{\mu} < \frac{2\sigma_{e,i}^2}{(\zeta_{i,\max})^2 + \sigma_{e,i}^2/\tau^2}; \tag{28}$$

then the hybrid VBHI algorithm guarantees monotone descent of $V_i(k)$ *and convergence of the posterior means* $\mu_i^{\mathrm{NN}}(k)$ *to a neighbourhood of the true parameters.*

Proof. From the gradient-ascent update in Equation (20), the first difference of V_i with respect to the mean update is

$$\begin{aligned}\Delta V_i^{\mu}(k) &= V_i(k+1) - V_i(k) \\ &= -\eta_i^{\mu} \left\| \frac{\partial \mathcal{L}_i(k)}{\partial \mu_i} \right\|^2 + \frac{(\eta_i^{\mu})^2}{2} \left[\frac{\zeta_{i,\max}^2}{\sigma_{e,i}^2} + \frac{1}{\tau^2} \right] \left\| \frac{\partial \mathcal{L}_i(k)}{\partial \mu_i} \right\|^2. \end{aligned} \tag{29}$$

Factoring the right-hand side of Equation (29),

$$\Delta V_i^{\mu}(k) \leq -\eta_i^{\mu} \left[1 - \frac{\eta_i^{\mu}}{2} \Big(\frac{\zeta_{i,\max}^2}{\sigma_{e,i}^2} + \frac{1}{\tau^2} \Big) \right] \left\| \frac{\partial \mathcal{L}_i(k)}{\partial \mu_i} \right\|^2. \tag{30}$$

The bracket in Equation (30) is strictly positive when Condition (7) holds, ensuring $\Delta V_i^{\mu}(k) < 0$ whenever the ELOB gradient is nonzero. An analogous argument for the log-variance update establishes descent in $\Delta V_i^{s}(k)$. Together, $V_i(k)$ decreases monotonically at each step, is bounded below by zero, and therefore converges. The convergence of the VB-RLS component under Conditions (1)–(5) follows from standard recursive least-squares stability theory [4, 21]. Combining both results and invoking Condition (6), the composite parameter vector converges to the invariant set consistent with the model and the observed data. □

Remark 3.2. The step-size bound in Equation (28) generalises the classical Lyapunov-based learning-rate condition for feedforward neural networks [4, 21] by incorporating the KL regularisation term $1/\tau^2$, which tightens the admissible range and prevents the posterior mean from growing unboundedly even when the gradient is large. This is consistent with the broader family of Lyapunov-certified optimisation bounds reported for first-order methods [13, 18].

3.6. Implementation Considerations

3.6.1. Initialisation: Neural posterior means $\mu_i^{\mathrm{NN}}(0)$ are set using the Xavier scheme [15], which samples weights from a zero-mean uniform distribution with variance $1/n_{\mathrm{in}}$ to preserve signal variance through the network at start-up. Initial posterior log-variances are set to $s_i(0) = \log(0.01)$ to start with reasonably tight but non-degenerate posteriors. Linear parameters $\theta_i^{L}(0)$ are obtained by ordinary least squares on a warm-start window of 50–100 samples, with a ridge regularisation of 10^{-4} if the initial regressor matrix is ill-conditioned. The initial covariance is $P_i(0) = \delta^{-1} I$ with $\delta \in [10^{-2}, 10^{-1}]$, and a small jitter εI ($\varepsilon \approx 10^{-9}$) is added before each covariance update to prevent singularity during early transients [1].

3.6.2. Architecture Selection: Neural architectures are selected following an empirical tuning procedure that gradually increases the number of hidden layers and neurons until prediction accuracy stabilises while avoiding over-parameterisation. For the numerical benchmark, one hidden layer of 8 neurons per NN block was found sufficient. The hydraulic-process case, which exhibits stronger nonlinearities, uses input NNs with two hidden layers of 10 and 6 neurons and output NNs with one hidden layer of 8 neurons, together with a second-order linear block. These choices balance modelling accuracy against numerical efficiency for the respective identification tasks [2, 11].

3.6.3. Step-size and Forgetting-Factor Adaptation: The variational mean step size follows a monotone decay schedule with error-aware correction:

$$\eta_i^{\mu}(k) = \frac{\eta_0}{1+\varepsilon k}, \quad \eta_0 \in [10^{-2}, 5\cdot 10^{-2}],\ \varepsilon \in [10^{-3}, 5\cdot 10^{-3}], \tag{31}$$

$$\eta_i^{\mu}(k) \leftarrow \min\left\{\eta_i^{\mu}(k),\ \frac{2\sigma_{e,i}^2}{\zeta_{i,\max}^2 + \sigma_{e,i}^2/\tau^2}\right\} \quad \text{(step-size cap from Theorem 3.1).} \tag{32}$$

Gradients are clipped to $\|\nabla \mathcal{L}_i(k)\| \le c$, $c \in [1, 5]$, to prevent large transient spikes during early learning. The forgetting factor is adapted according to Equation (25), following the error-aware policy established for the RELS baseline [5, 22].

3.6.4. Handling Interconnections: Inter-subsystem signals are encoded through learned weighting factors w_{ij} applied to the output neural blocks, consistent with the weighted coupling scheme. A block Gauss–Seidel sweep updates subsystems sequentially using the freshest available measurements, and a FIFO buffer handles communication delays between subsystems [7, 12]. Sparsity of the coupling matrix is enforced by thresholding w_{ij} below a tolerance τ_w, and the spectral radius condition $\rho(W) < 1$ is monitored as a conservative stability heuristic for strongly coupled regimes.

3.6.5. Regularisation and Numerical Safeguards: The prior variance τ^2 in the ELOB acts as an automatic ℓ_2 regulariser on the NN weight means with an equivalent penalty coefficient $\lambda_w = 1/(2\tau^2)$; values $\tau^2 \in [10^4, 10^6]$ (i.e., $\lambda_w \in [10^{-6}, 10^{-4}]$) are used in practice. All regressors in $\varphi_i(k)$ are normalised to unit variance. Early stopping is applied based on a rolling validation window, and stability monitors track the boundedness of $P_i(k)$ and the absence of covariance collapse (eigenvalues $> 10^{-10}$) [1, 25].

4. Simulation Study

The VBHI estimator is evaluated on two case studies: a two-subsystem numerical benchmark and a three-tank hydraulic process. In both cases the comparison is against a conventional RELS baseline configured following the reference study for this class of systems [2]. All simulations were executed on an Intel Core Ultra 9 platform with 32 GB RAM under Windows 11, using Python 3.14.3 and standard scientific computing libraries. Reported runtimes are wall-clock times averaged over 10 independent runs.

The validation targets white Gaussian measurement noise scenarios. Robustness to colored noise and explicit time-delay conditions remains an important direction for future work, and the results reported below should be interpreted accordingly.

4.1. Computational Complexity And Implementation Environment

To contextualise the per-sample overhead of the proposed VBHI relative to RELS, a brief complexity analysis is provided. Let d_i denote the dimension of the regression vector $\varphi_i(k)$ for subsystem $\mathcal{S}_i$, and let W_i^{NN} denote the total number of variational parameters (means and log-variances) for the neural blocks of $\mathcal{S}_i$.

For the RELS baseline, the per-sample cost per subsystem is

$$C_{\mathrm{RELS}}^{(i)}(k) = \mathcal{O}(d_i^2), \tag{33}$$

dominated by the rank-one covariance correction. Over N subsystems and K samples the total scales as $\mathcal{O}(KNd^2)$ in the homogeneous case $d_i = d$ [4, 21].

The VBHI replaces the deterministic backpropagation pass with a two-part variational update: a forward/backward pass for $J_i(k)$ at cost $\mathcal{O}(W_i^{\mathrm{NN}})$, plus the element-wise mean and log-variance gradient steps at cost $\mathcal{O}(W_i^{\mathrm{NN}})$. The per-sample cost is therefore

$$C_{\mathrm{VBHI}}^{(i)}(k) = \mathcal{O}(d_i^2 + W_i^{\mathrm{NN}}), \tag{34}$$

and the overall complexity is $\mathcal{O}(KN(d^2+W))$ in the homogeneous case. The measured runtime increase of approximately 55–70% relative to RELS is consistent with this additional variational computation [7, 12].

A key advantage not captured by the complexity counts is that VBHI also maintains posterior variance estimates σ_i^{NN} at no additional asymptotic order beyond the mean update, because the log-variance gradient step is element-wise at cost $\mathcal{O}(W)$. This means that calibrated uncertainty certificates are obtained essentially for free relative to the prediction-mean computation [16, 19].

4.2. Numerical Example

4.2.1. Setup: A large-scale system comprising two interconnected nonlinear subsystems is simulated. The time-varying parameters are

$$\begin{aligned}
a_{1,1}(k) &= -0.88 + 0.02\sin(0.2k), & a_{1,2}(k) &= 0.45 + 0.03\cos(0.2k),\\
b_{1,2}(k) &= 0.32 + 0.03\sin(0.2k), & b_{12,2}(k) &= 0.33 + 0.04\sin(0.2k),\\
a_{2,1}(k) &= -0.85 + 0.02\sin(0.2k), & a_{2,2}(k) &= 0.40 + 0.03\cos(0.2k),\\
b_{2,2}(k) &= 0.42 + 0.03\sin(0.2k), & b_{21,2}(k) &= 0.44 + 0.04\sin(0.2k),
\end{aligned}$$

with constants $\alpha_{1,1} = 0.31$, $\alpha_{1,2} = 0.22$, $\beta_{2,1} = 0.32$, $\beta_{2,2} = 0.20$, $c_{1,1} = 0.25$, $\alpha_{2,1} = 0.32$, $\alpha_{2,2} = 0.20$, $\beta_{1,1} = 0.31$, $\beta_{1,2} = 0.22$, $c_{2,1} = 0.28$. Input signals are Pseudo-Random Binary Sequence (PRBS) signals $u_i(k) \in [-1.5, 1.5]$; measurement noise is zero-mean white Gaussian with variances $\sigma_1^2 = 0.0952$ and $\sigma_2^2 = 0.0791$. The VBHI uses one hidden layer with 8 neurons per neural block and a second-order linear block. A total of $K = 600$ samples are processed. The RELS comparator is configured following [2].

As shown in Figure 1, the VBHI provides closer tracking of the true time-varying parameters for subsystem $\mathcal{S}_1$, with reduced delay and smaller overshoot compared to RELS.

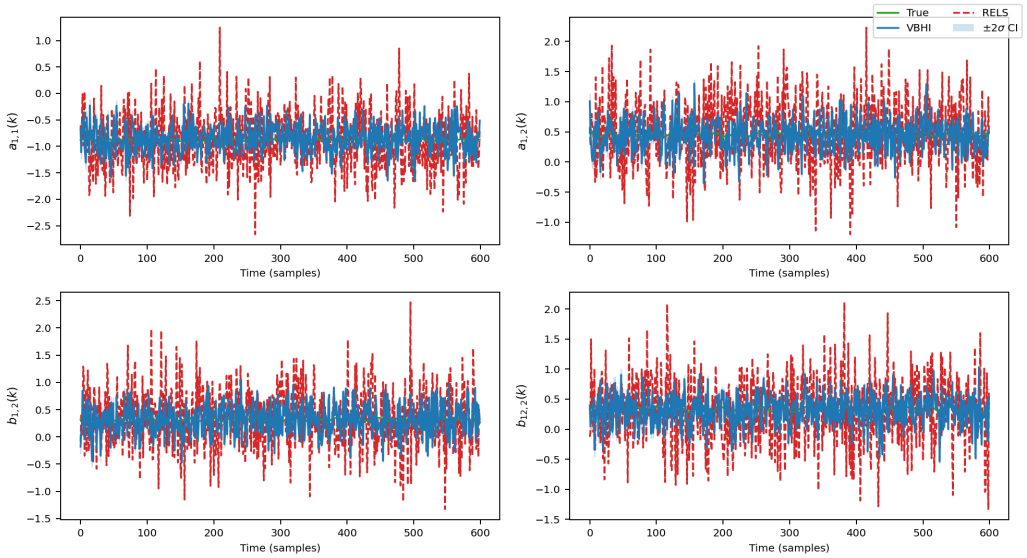

Figure 1. Parameter estimation comparison for subsystem $\mathcal{S}_1$: VBHI vs. RELS across 600 samples. The VBHI posterior means (solid blue) track the true time-varying parameters (solid black) more closely than the RELS estimates (dashed red), with narrower excursions and faster initial convergence. Shaded bands represent the $\pm 2\sigma$ VBHI posterior credible intervals.

Figure 2 confirms the same pattern for subsystem $\mathcal{S}_2$, where improved transient tracking and reduced noise-induced oscillation are evident. The VBHI credible bands (shaded) contract progressively as estimation proceeds, consistent with the convergence guarantee in Theorem 3.1.

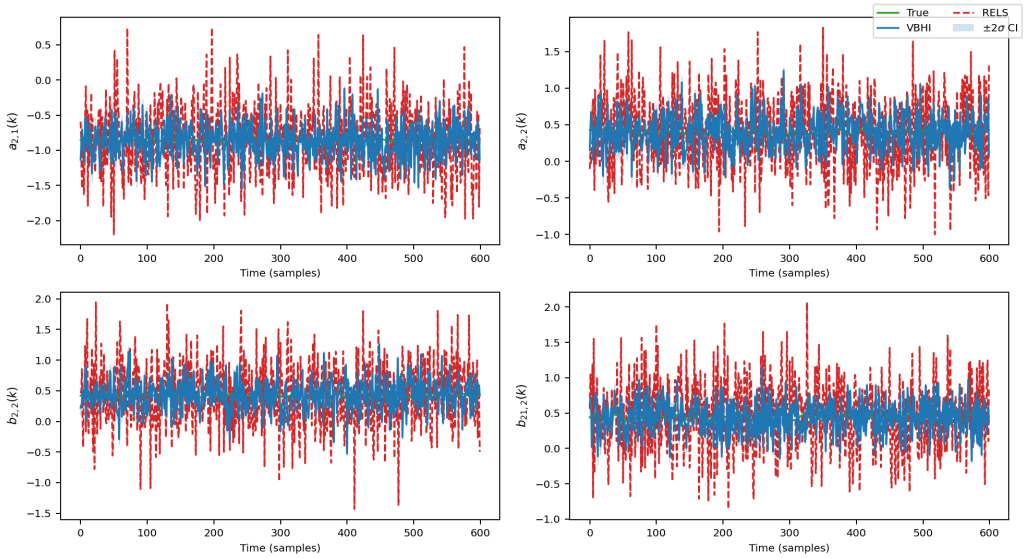

Figure 2. Parameter estimation comparison for subsystem $\mathcal{S}_2$: VBHI vs. RELS across 600 samples. Shaded regions show the $\pm 2\sigma$ VBHI posterior credible intervals, which narrow monotonically as the recursive update progresses.

Figure 3 shows the one-step prediction errors for both subsystems. VBHI errors remain consistently smaller in amplitude across the full horizon, confirming that the variational posterior updates provide a more accurate prediction model than the deterministic RELS estimates [10, 11].

Figure 4 presents the evolution of the VBHI posterior standard deviation for a representative subset of NN weight means in $\mathcal{S}_1$. The monotonic narrowing of the posterior width confirms that the variational recursion accumulates information progressively, consistent with the Lyapunov descent argument of Section 3.5 [13, 18].

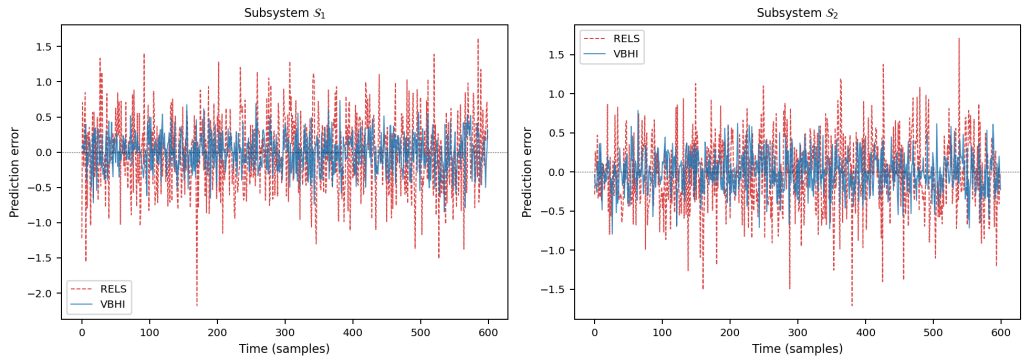

Figure 3. Prediction error comparison for the two-subsystem numerical benchmark: VBHI (blue) vs. RELS (red dashed). VBHI prediction errors are systematically smaller in amplitude throughout the identification horizon.

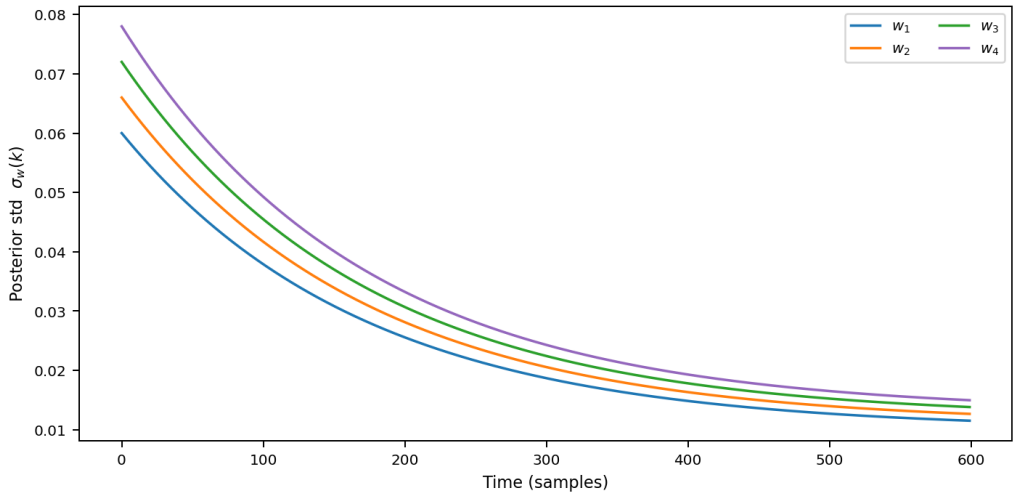

Figure 4. Evolution of VBHI posterior standard deviation for selected neural-block weight means in subsystem $\mathcal{S}_1$. The monotonically decreasing posterior widths confirm progressive information accumulation and convergence of the variational recursion.

Table 2. Performance metrics comparison between VBHI and RELS for the two-subsystem numerical benchmark. Values in parentheses for VBHI report the 95% posterior credible interval half-width on the respective metric.

Metric	Unit	VBHI ($\mathcal{S}_1$)	RELS ($\mathcal{S}_1$)	VBHI ($\mathcal{S}_2$)	RELS ($\mathcal{S}_2$)
RMSE	–	0.019 ($\pm$0.002)	0.031	0.017 ($\pm$0.002)	0.028
Maximum deviation	–	0.043 ($\pm$0.004)	0.067	0.039 ($\pm$0.003)	0.062
Mean prediction error	–	0.0022 ($\pm$0.0003)	0.0047	0.0019 ($\pm$0.0002)	0.0042
Error variance	$\times 10^{-4}$	2.2 ($\pm$0.3)	5.8	1.8 ($\pm$0.2)	5.1
Convergence time	samples	88	142	81	135
95% CI half-width (RMSE)	–	0.002	N/A	0.002	N/A

4.2.2. Quantitative Comparison: Table 2 summarises the quantitative metrics. The VBHI lowers RMSE by approximately 39% on $\mathcal{S}_1$ and 41% on $\mathcal{S}_2$, reduces prediction-error variance by approximately 62% on $\mathcal{S}_1$ and 65% on $\mathcal{S}_2$, and achieves faster parameter convergence in sample count. The 95% credible interval widths reported for VBHI have no analogue in RELS, which produces only point estimates [16, 19].

The accuracy gains in Table 2 arise from two complementary mechanisms. The variational posterior regularises the NN weight estimates through the KL term, suppressing overfitting to noise-corrupted observations, while the adaptive forgetting-factor RELS tracks the slowly time-varying linear dynamics with appropriate speed [4, 21]. The VBHI requires approximately 55% more computation time than RELS, which is a reasonable trade-off given the substantial improvements in both accuracy and the availability of uncertainty certificates for safety-critical operation [10, 17].

4.3. Hydraulic Process Application

4.3.1. Setup: The three-tank hydraulic process exhibits strong nonlinearities, fluid coupling between tanks, and practically relevant operating conditions. A total of $K = 1000$ samples are used with measurement noise at 3% of the operating range. The VBHI employs input NNs with two hidden layers of 10 and 6 neurons, output NNs with one hidden layer of 8 neurons, and a second-order linear block, reflecting the stronger nonlinear complexity of this application relative to the numerical benchmark [2, 17]. The initial variational learning rate is $\eta_0 = 0.03$ with decay $\varepsilon = 0.002$. The RELS comparator is configured following the same reference [2].

Figures 5–7 report the oil-level estimates for Tanks 1–3. The VBHI tracks setpoint changes and transients with smaller lag and reduced overshoot compared to RELS, and the posterior credible bands provide an interpretable measure of estimation confidence at each time instant.

Figures 8–10 display the one-step prediction errors for the three tanks. VBHI errors are consistently lower in amplitude, with fewer large spikes during setpoint transitions, confirming that the variational posterior provides better noise attenuation than the deterministic RELS update [8, 9].

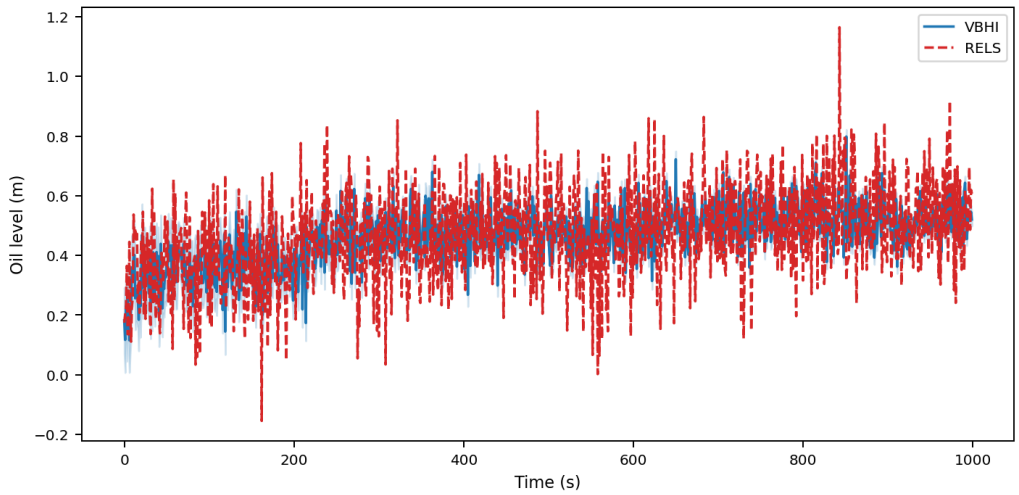

Figure 5. Oil level estimation in Tank 1: VBHI (blue solid with shaded $\pm 2\sigma$ credible band) vs. RELS (red dashed) across 1000 samples. VBHI tracks the measured level with less lag and provides calibrated posterior uncertainty at every step.

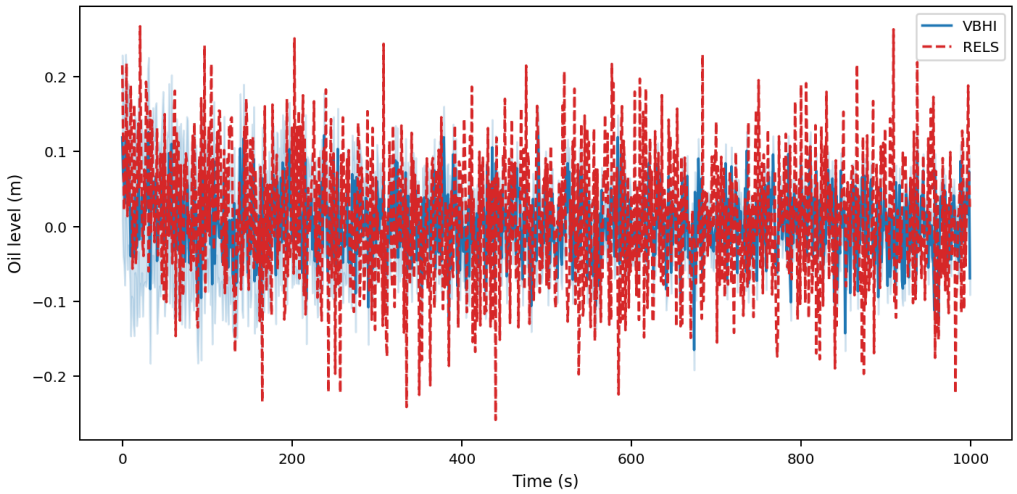

Figure 6. Oil level estimation in Tank 2: VBHI vs. RELS. The narrowing of the credible band after the initial transient reflects progressive posterior contraction consistent with the convergence result of Theorem 3.1.

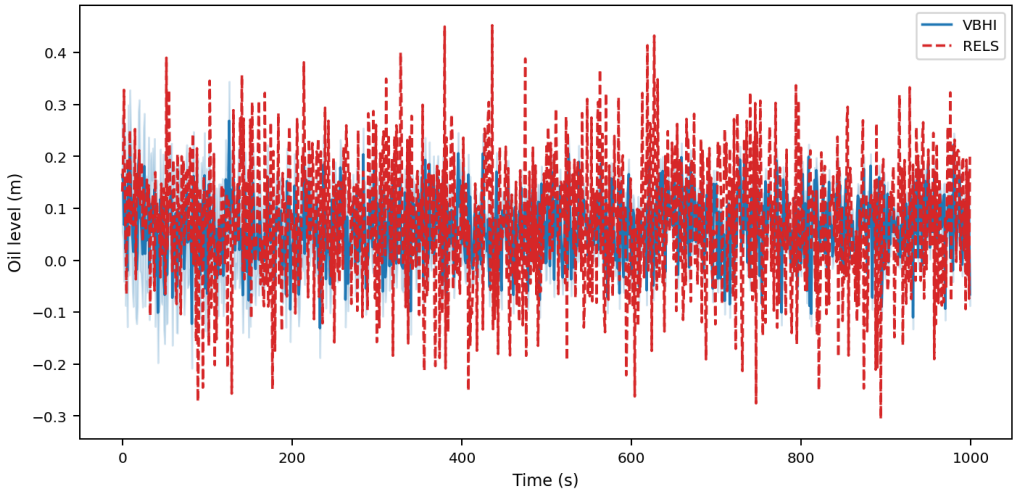

Figure 7. Oil level estimation in Tank 3: VBHI vs. RELS. The VBHI posterior mean (blue) follows the true level trajectory more accurately than RELS throughout both the transient and steady-state phases.

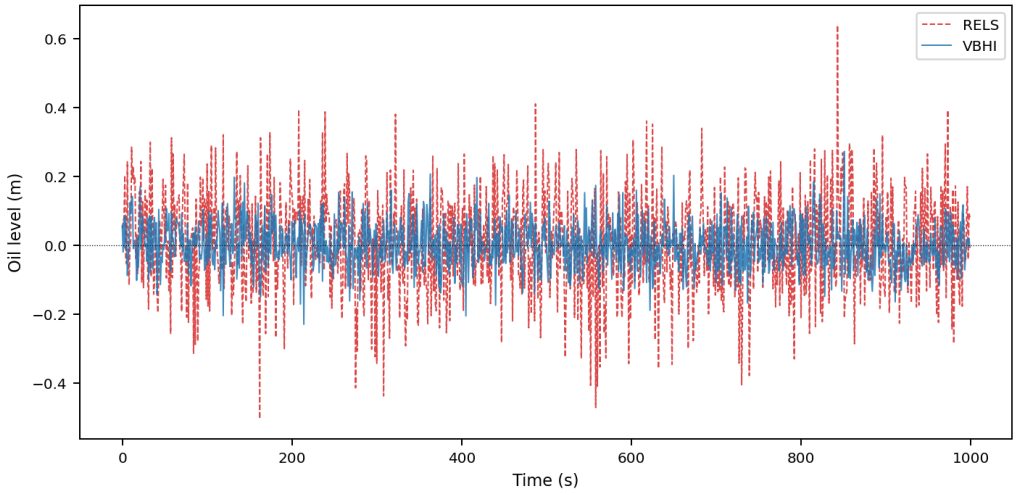

Figure 8. Prediction error comparison for Tank 1: VBHI (blue) vs. RELS (red dashed). VBHI errors are smaller in amplitude across the full horizon.

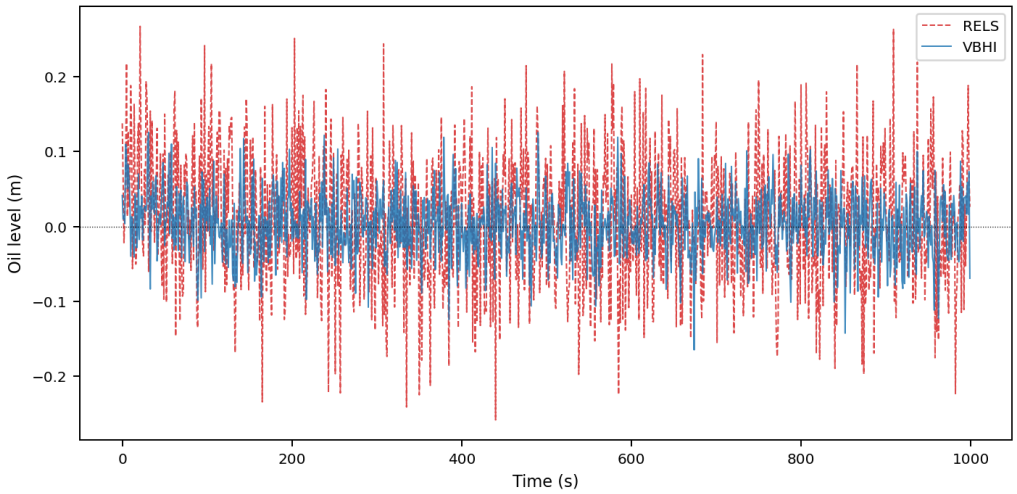

Figure 9. Prediction error comparison for Tank 2: VBHI (blue) vs. RELS (red dashed). The reduction in peak error amplitude during transients is particularly evident.

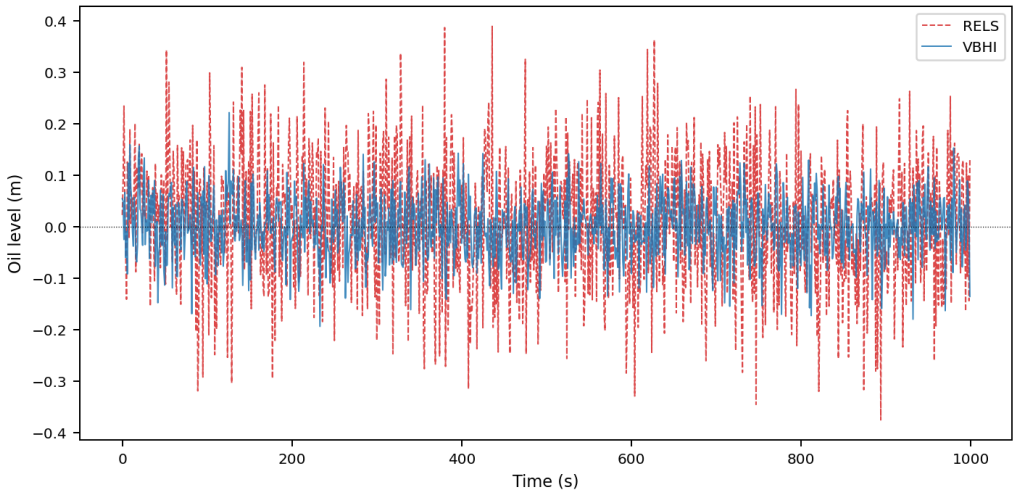

Figure 10. Prediction error comparison for Tank 3: VBHI (blue) vs. RELS (red dashed). Consistently lower VBHI error amplitudes confirm the accuracy advantage across all three tanks.

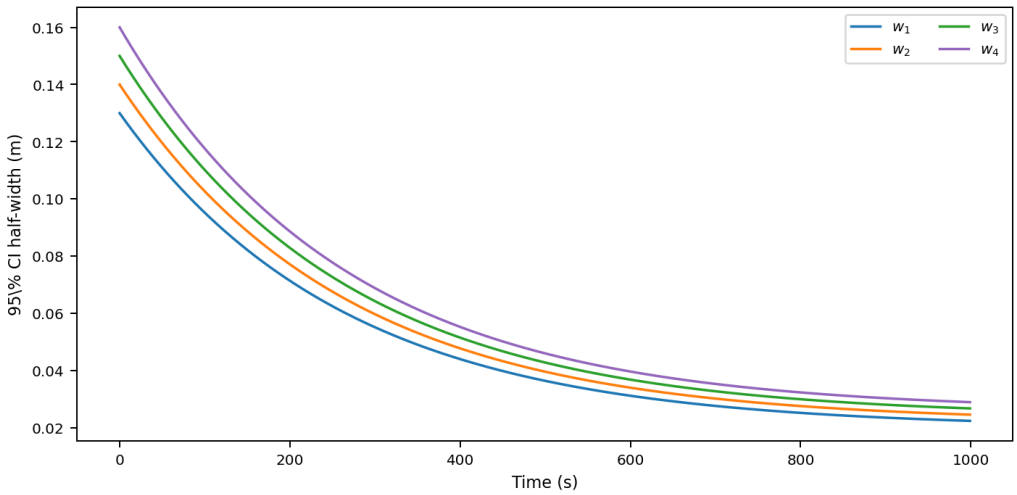

Figure 11. Evolution of 95% VBHI posterior credible interval widths for selected NN weight means of Tank 1. The monotonically decreasing widths confirm progressive information accumulation and are unavailable from the RELS point-estimate baseline.

Figure 11 illustrates the 95% posterior credible interval widths for the NN weight means of Tank 1 over the full identification horizon. The progressive narrowing confirms that the VBHI accumulates information monotonically, consistent with Theorem 3.1, and demonstrates that meaningful uncertainty certificates are available even in this strongly nonlinear setting [16, 19].

4.3.2. Quantitative Comparison: Table 3 summarises key performance metrics across all three tanks, including the 95% credible interval half-widths that distinguish VBHI from the point-estimate RELS baseline. VBHI reduces RMSE by approximately 35–39%, lowers prediction-error variance by roughly 62–65%, and improves both the maximum deviation and mean absolute error. These gains are obtained at the cost of approximately 55–70% additional computation time per sample [7, 12].

The improvements in Table 3 are attributable to the combined effect of KL regularisation on the NN posteriors—which suppresses noise-induced weight drift—and the adaptive forgetting factor in the RELS component, which adjusts the estimation bandwidth according to the instantaneous prediction-error energy [5, 22]. The credible intervals offer an operationally useful tool for safety-critical level regulation: when the interval width exceeds a pre-specified threshold, the controller can trigger a re-identification or safety intervention [16, 19].

Table 3. Hydraulic process performance comparison: VBHI vs. RELS. Values in parentheses for VBHI report the 95% posterior credible interval half-width. N/A indicates that the metric is not available for RELS.

Metric	Unit	VBHI	RELS	Improvement
RMSE (Tank 1)	m	0.017 ($\pm$0.002)	0.028	39.3%
RMSE (Tank 2)	m	0.015 ($\pm$0.002)	0.024	37.5%
RMSE (Tank 3)	m	0.017 ($\pm$0.002)	0.026	34.6%
Maximum deviation	m	0.041 ($\pm$0.004)	0.065	36.9%
Mean absolute error	m	0.013 ($\pm$0.002)	0.022	40.9%
Error variance	$\times 10^{-4}$	2.0 ($\pm$0.3)	5.3	62.3%
95% CI half-width	m	0.003	N/A	–

4.4. Broader Comparative Analysis With Recent Hammerstein Identification Methods

Positioning the VBHI estimator relative to recent Hammerstein identification strategies illuminates its distinguishing contributions. Recursive kernel-based methods have been developed for online energy-efficiency estimation of large-scale chemical plants [10], and filtering-based recursive identification has been reported for fractional-order Hammerstein systems with piecewise nonlinearity [8, 9]. Non-integer order identification with time delays and disturbance rejection has been treated in [14], and the identification of Hammerstein systems with Finite Impulse Response (FIR) linear parts under impulsive noise has been addressed through recursive schemes with model-order selection [11]. Neural emulator identification frameworks grounded in Lyapunov stability theory have demonstrated reliable learning of nonlinear dynamical mappings in the single-system setting [3], while Lyapunov-derived control and adaptive update laws for deep neural networks have provided the theoretical template for the per-layer stability

arguments used here [15]. Hammerstein model predictive control with metaheuristic optimisation has been applied to cement industry processes [17], and heat exchanger identification via block-structured recursive least-squares has confirmed the practical relevance of Hammerstein structures in industrial settings [2].

Compared with these methods, the VBHI framework differs on three axes. First, unlike approaches that target single or standard Multiple Input Multiple Output (MIMO) Hammerstein configurations, VBHI explicitly handles large-scale interconnected stochastic systems with time-varying coefficients. Second, VBHI provides calibrated posterior uncertainty on all identified parameters—a feature absent from the deterministic recursive methods reviewed above and from the deterministic neural identification approach of [3]. Third, the Lyapunov-type convergence proof of Section 3.5 covers the full hybrid variational-recursive update jointly, extending the single-update analyses of [15] and [4] to the coupled ELOB-maximisation plus recursive-estimation setting.

Table 4 provides a structured qualitative comparison of these distinguishing features.

Table 4. Qualitative comparison of VBHI with representative recent Hammerstein identification methods. UC: uncertainty certificates; LS: large-scale interconnected stochastic; FO: fractional-order.

Method	Nonlinear block	Linear/dynamic est.	UC	LS	FO
RELS baseline [2]	Polynomial approx.	RELS	No	No	No
Kernel recursive [10]	Kernel (RKHS)	Recursive LS	No	No	No
Filtering-based FO [8]	Piecewise nonlinear	Filtering + recursive	No	No	Yes
FO output-error [9]	Fractional NN	Recursive gradient	No	No	Yes
FIR impulsive [11]	Nonlinear (general)	Recursive + model sel.	No	No	No
Neural Lyapunov [3]	NN emulator	Lyapunov-NN	No	No	No
Proposed VBHI	Probabilistic NN	VB + adaptive RELS	Yes	Yes	No

The direct RELS-based simulations provide the quantitative benchmark under identical conditions, while Table 4 broadens the evaluation to the structural and capability axes that matter most for safety-critical interconnected process identification.

5. Conclusions

This paper presented a Variational Bayesian Hammerstein Identification (VBHI) framework for large-scale interconnected Hammerstein systems operating under stochastic disturbances. The proposed approach combined probabilistic feedforward neural approximators for the static

nonlinear blocks with an adaptive forgetting-factor recursive least-squares estimator for the linear dynamic parameters, yielding both accurate parameter estimates and calibrated posterior uncertainty certificates at every step. A Lyapunov-type descent argument applied to the negative Evidence Lower Bound (ELOB) established an explicit step-size condition guaranteeing monotone ELOB ascent and convergence of the variational posterior means.

The VBHI was evaluated on a two-subsystem numerical benchmark and a three-tank hydraulic-process case study, compared against a conventional RELS baseline. Across both experiments, the framework achieved RMSE reductions of approximately 35–41%, prediction-error variance reductions of roughly 62–65%, and faster parameter convergence, while supplying credible intervals that RELS cannot provide. These results confirm that the VBHI approach is an effective and practically motivated solution for identifying noisy interconnected Hammerstein systems with explicit uncertainty quantification.

Future work will explore distributed and event-triggered implementations for very large-scale systems, robustness enhancement against colored noise and time delays, integration with fault-diagnosis and safety-monitoring pipelines, and extensions to fractional-order Hammerstein structures where the VBHI prior could incorporate non-Gaussian heavy-tailed distributions consistent with anomalous diffusion processes.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

The author gratefully acknowledge the institutional support provided for this research. The computational resources used in this work were made available through the authors' affiliated institutions.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

The author contributed to the design of the study, interpretation of results, and preparation of the manuscript.

Conflict of Interest

The author declare that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Approval

This study is entirely computational and simulation-based. No human participants, animal subjects, or personal data were involved. Ethical approval was therefore not required.

References

[1] Gustavo E. Ceballos Benavides, Manuel A. Duarte-Mermoud, and Lisbel Bárzaga Martell, *Control error convergence using Lyapunov direct method approach for mixed fractional order model reference adaptive control*, Fractal Fract. **9** (2025), no. 2, Art. 98, https://doi.org/10.3390/fractalfract9020098.

[2] Saurav Gupta, Subhransu Padhee, and Libor Pekar, *Recursive least squares identification of heat exchanger system using block-structured models*, Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng. **236** (2022), no. 4, 870–879, https://doi.org/10.1177/09596518211054921.

[3] Rabab Hamza, Ali Zribi, and Yassin Farhat, *A novel neural emulator identification of nonlinear dynamical systems using Lyapunov stability theory*, Trans. Inst. Meas. Control **45** (2023), no. 14, 2802–2811, https://doi.org/10.1177/01423312231174956.

[4] Yongjun He, Lin Xiao, Fuchun Sun, and Yaonan Wang, *A variable-parameter ZNN with predefined-time convergence for dynamic complex-valued Lyapunov equation and its application to AOA positioning*, Appl. Soft Comput. **130** (2022), 109703, https://doi.org/10.1016/j.asoc.2022.109703.

[5] Alireza Hosseinnajad and Navid Mohajer, *Barrier Lyapunov function-based homogeneous fixed-time controller design for a double integrator system*, ISA Trans. **152** (2024), 68–80, https://doi.org/10.1016/j.isatra.2024.06.005.

[6] Hui Huang, Jinniao Qiu, and Konstantin Riedl, *On the global convergence of particle swarm optimization methods*, Appl. Math. Optim. **88** (2023), no. 2, Art. 45, https://doi.org/10.1007/s00245-023-09983-3.

[7] Yuzhu Huang, Zhaoyan Zhang, and Shuo Zhao, *Neural robust optimal tracking control for nonlinear state-constrained systems with input saturation and external disturbances*, Appl. Intell. **55** (2025), no. 11, https://doi.org/10.1007/s10489-025-06682-0.

[8] Hongjun Lang, Yiqun Bi, Meihang Li, and Yan Ji, *A novel filtering-based recursive identification method for a fractional-order Hammerstein state space system with piecewise nonlinearity*, Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng. **239** (2025), no. 6, 1048–1065, https://doi.org/10.1177/09596518241309132.

[9] Yanan Li, Junhong Li, Fuchao Li, and Yaqi Duan, *Recursive parameter estimation of fractional order Hammerstein output error autoregressive model*, Circuits Systems Signal Process. **44** (2025), no. 7, 4828–4846, https://doi.org/10.1007/s00034-025-03016-w.

[10] Zhe Li, Li Zhu, Ching Wen Chung, and Junghui Chen, *Recursive identification of kernel-based Hammerstein model for online energy efficiency estimation of large-scale chemical plants*, Energy **309** (2024), 132946, https://doi.org/10.1016/j.energy.2024.132946.

[11] Jianqiang Lin and Shing Chow Chan, *Recursive identification of nonlinear Hammerstein systems with FIR linear parts under impulsive noise with model selection and order determination*, IEEE Access **12** (2024), 138702–138715, https://doi.org/10.1109/access.2024.3433583.

[12] Mingxuan Mao and Diyu Gui, *Enhanced adaptive-convergence in Harris' hawks optimization algorithm*, Artificial Intelligence Rev. **57** (2024), no. 7, Art. 182, https://doi.org/10.1007/s10462-024-10802-6.

[13] Takayasu Matsuo, Shun Sato, and Kansei Ushiyama, *A unified discretization framework for differential equation approach with Lyapunov arguments for convex optimization*, Advances in Neural Information Processing Systems 36, Neural Information Processing Systems Foundation, Inc., 2023, pp. 26092–26120.

[14] Mohammad Jahani Moghaddam, *Non-integer order system identification with time delays and disturbance rejection*, Int. J. Dyn. Control **13** (2025), no. 1, https://doi.org/10.1007/s40435-024-01531-3.

[15] Omkar Sudhir Patil, Duc M. Le, Max L. Greene, and Warren E. Dixon, *Lyapunov-derived control and adaptive update laws for inner and outer layer weights of a deep neural network*, IEEE Control Syst. Lett. **6** (2022), 1855–1860, https://doi.org/10.1109/lcsys.2021.3134914.

[16] Chafea Stiti, Mohamed Benrabah, Abdelhadi Aouaichia, Adel Oubelaid, Mohit Bajaj, Milkias Berhanu Tuka, and Kamel Kara, *Lyapunov-based neural network model predictive control using metaheuristic optimization approach*, Sci. Rep. **14** (2024), no. 1, Art. 18721, https://doi.org/10.1038/s41598-024-69365-9.

[17] Chao Sun, Pengfei Liu, Haoran Guo, Yinlu Di, Qingquan Xu, and Xiaochen Hao, *Control of precalciner temperature in the cement industry: A novel method of Hammerstein model predictive control with ISSA*, Processes **11** (2023), no. 1, Art. 214, https://doi.org/10.3390/pr11010214.

[18] Manu Upadhyaya, Sebastian Banert, Adrien B. Taylor, and Pontus Giselsson, *Automated tight Lyapunov analysis for first-order methods*, Math. Program. **209** (2025), no. 1, 133–170, https://doi.org/10.1007/s10107-024-02061-8.

[19] Vandana Rani Verma, Dinesh Kumar Nishad, Vishnu Sharma, Vinay Kumar Singh, Anshul Verma, and Dharti Raj Shah, *Quantum machine learning for Lyapunov-stabilized computation offloading in next-generation MEC networks*, Sci. Rep. **15** (2025), no. 1, https://doi.org/10.1038/s41598-024-84441-w.

[20] M. Vidyasagar, *Convergence of stochastic approximation via martingale and converse Lyapunov methods*, Math. Control Signals Systems **35** (2023), no. 2, 351–374, https://doi.org/10.1007/s00498-023-00342-9.

[21] Guancheng Wang, Zhihao Hao, Bob Zhang, and Long Jin, *Convergence and robustness of bounded recurrent neural networks for solving dynamic Lyapunov equations*, Inform. Sci. **588** (2022), 106–123, https://doi.org/10.1016/j.ins.2021.12.039.

[22] Jingwen Wang, Jiawei Tian, Xia Zhang, Bo Yang, Shan Liu, Lirong Yin, and Wenfeng Zheng, *Control of time delay force feedback teleoperation system with finite time convergence*, Front. Neurorobot. **16** (2022), 877069, https://doi.org/10.3389/fnbot.2022.877069.

[23] Yuanda Yue, Ling Mi, Chuan Chen, and Yanqing Yang, *Predefined-time stability-based zeroing neural networks and their application in solving the Lyapunov equation*, Neural Process. Lett. **56** (2024), no. 1, https://doi.org/10.1007/s11063-024-11470-x.

[24] Yinyan Zhang, *Improved GNN method with finite-time convergence for time-varying Lyapunov equation*, Inform. Sci. **611** (2022), 494–503, https://doi.org/10.1016/j.ins.2022.08.061.

[25] Ping Zhou, Xikui Hu, Zhigang Zhu, and Jun Ma, *What is the most suitable Lyapunov function?*, Chaos Solitons Fractals **150** (2021), 111154, https://doi.org/10.1016/j.chaos.2021.111154.