

“© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

# System Identification of Static Nonlinear Elements: A Unified Approach of Active Learning, Over-fit Avoidance, and Model Structure Determination

Nalika Ulapane

*Electrical and Electronic Engineering  
The University of Melbourne  
Parkville VIC 3010, Australia  
nalika.ulapane@unimelb.edu.au*

Karthick Thiyagarajan

*UTS Robotics Institute  
University of Technology Sydney  
Ultimo NSW 2007, Australia  
karthick.thiyagarajan@uts.edu.au*

Sarath Kodagoda

*UTS Robotics Institute  
University of Technology Sydney  
Ultimo NSW 2007, Australia  
sarath.kodagoda@uts.edu.au*

**Abstract**—Systems containing linear first-order dynamics and static nonlinear elements (i.e., nonlinear elements whose outputs depend only on the present value of inputs) are often encountered; for example, certain automobile engine subsystems. Therefore, system identification of static nonlinear elements becomes a crucial component that underpins the success of the overall identification of such dynamical systems. In relation to identifying such systems, we are often required to identify models in differential equation form, and consequently, we are required to describe static nonlinear elements in the form of functions in time domain. Identification of such functions describing static elements is often a black-box identification exercise; although the inputs and outputs are known, correct mathematical models describing the static nonlinear elements may be unknown. Therefore, with the aim of obtaining computationally efficient models, calibrating polynomial models for such static elements is often attempted. With that approach comes several issues, such as long time requirements to collect adequate amounts of measurements to calibrate models, having to test different models to pick the best one, and having to avoid models over-fitting to noisy measurements. Given that premise, this paper proposes an approach to tackle some of those issues. The approach involves collecting measurements based on an uncertainty-driven Active Learning scheme to reduce time spent on measurements, and simultaneously fitting smooth models using Gaussian Process (GP) regression to avoid over-fitting, and subsequently picking best fitting polynomial models using GP-regressed smooth models. The principles for the single-input-single-output (SISO) static nonlinear element case are demonstrated in this paper through simulation. These principles can easily be extended to MISO systems.

**Index Terms**—Active Learning, Gaussian Process, nonlinear systems, over-fitting, system identification, uncertainty

## I. INTRODUCTION

Accurate identification of static elements is crucial for identifying effective models for control applications related to dynamical systems, especially when it comes to the domain of automobile engine subsystems [1]. Such systems commonly include linear dynamics and static nonlinear elements, and the system architecture comprising such linear dynamics and static elements has been specifically studied over the years [2]. When it comes to identifying static elements, a well-known challenge encountered is the requirement of a long

time for measurements [1], especially in relation to systems like engines as measurements typically have to be taken when system responses have reached a steady state. Furthermore, on instances where the model structure is unknown and black-box identification is required (which is often the case in real-world system identification tasks), computation time has to be spent for model structure determination and over-fit avoidance. It is therefore desirable to reduce the time required for measurements by enabling the collection of only an amount of data which is optimal and informative enough in some sense, and also reducing the computation time by developing algorithms that can systematically determine suitable model structures. This paper contributes to this space by presenting a unified approach, by combining Active Learning [3], [4] to systematically collect informative measurements, and model structure determination in a way that over-fitting is avoided.

Optimal design of experiments (DoE) in relation to system identification has been extensively studied [5]–[7] with the aim of reducing time spent on experiments, or maximizing the information gained within a time budget. DoE often relies on Fisher Information [7], [8] and thus comes the limitation of having to assume a model for the system beforehand. Assuming a model can become challenging in some real-world system identification tasks, and as an alternative, the approach of Mean Value Engine Modeling (MVEM) [9] is sometimes adopted. MVEM is usually physics-based, and the resulting models can be complicated, making them difficult to be coupled with DoE since mathematically describing ‘Information’ can become tedious. Active Learning which is greatly analogous to DoE, in its conventional use adopts techniques like ‘query by committee’ [10] to collect measurements guided by some means of model uncertainty. However, such approaches used for system identification [4] too often rely on an assumed model, or a set of models, which can be identified as a limitation, especially since on the other hand, more modern machine learning approaches such as Gaussian Processes (GP)s [11] enable similar uncertainty-based approaches on a non-parametric paradigm avoiding the requirement of having to assume parametric models.

In this paper, the authors propose a greedy Active Learning scheme to systematically collect measurements to identify static nonlinear elements. The scheme starts with a set of given prior measurements and decides the next point to take measurements using maximum uncertainty produced by GP regression as the criteria for deciding. The scheme operates recursively until reaching some termination criteria. Since GP is non-parametric, the proposed Active Learning scheme is model-free, and therefore is different from parametric model-based approaches presented in works such as [4], [7]. On termination of the Active Learning-based measurement scheme, a polynomial-based model structure is determined and calibrated. Some important literature about polynomial structure determination include the works [12], [13]; however, they are not coupled with Active Learning schemes unlike in this paper, and revolve around determining polynomial structures using a given set of measurements. We demonstrate the approach on identification of a single-input-single-output (SISO) static nonlinear element. The approach can easily be extended to the MISO case.

The structure of the paper is as follows: Section II describes the principles of GP-based regression; Section III presents the system identification problem formulation; Section IV presents the methodology detailing the Active Learning scheme and the polynomial model structure determination approach; Section V presents the results; and Section VI concludes the paper by discussing the implications of results and potential avenues for future work.

## II. RELATED KNOWLEDGE: GAUSSIAN PROCESS FORMULATION FOR REGRESSION

GPs are a strong tool used for solving multi-dimensional nonlinear regression problems as well as classification problems. Reference [11] is useful to study GPs in detail, while this section summarizes the GP regression formulation for the one-dimensional nonlinear regression case for a smooth function.

Suppose there is a noisy process  $y = f(x) + \epsilon$  where  $x \in \mathfrak{R}$  is a one dimensional independent variable,  $y \in \mathfrak{R}$  is a one dimensional dependent variable, and  $\epsilon \in \mathfrak{R}$  is noise, and it is required to learn a regression model describing the process. Firstly, it is required to collect some training data  $X$  and  $Y$  where  $X = [x_1, x_2, x_3, \dots, x_n]^T$  is a vector containing measurement points and  $Y = [y_1, y_2, y_3, \dots, y_n]^T$  is a vector containing corresponding noisy process responses;  $n \in \mathbb{Z}^+$ . Secondly, a kernel  $K(X, X)$  whose elements are given by  $k_{i,j} = k(x_i, x_j)$  where  $k(\cdot, \cdot)$  is some kernel function and  $x_i, x_j \in X$ , has to be selected. Choices for kernel functions and some rationale behind the suitability for some kernel functions for some specific applications are available in [11]. The work of this paper focuses on smooth functions, and therefore has selected the squared exponential kernel function (described in (1);  $\alpha, \eta \in \mathfrak{R}$  in (1) are hyper-parameters) as it has been known to work well with smooth functions and has been commonly used in previous works to learn smooth

functions [14]–[20].

$$k(x_i, x_j) = \alpha^2 \exp \left\{ -\frac{1}{2\eta^2} (x_i - x_j)^2 \right\} \quad (1)$$

When the squared exponential kernel function is used,  $\alpha$  and  $\eta$  along with an additional hyper-parameter (i.e., noise variance  $\sigma$ ) construct the set  $\theta = \{\alpha, \eta, \sigma\}$  which is the set of hyper-parameters. Following a maximum likelihood estimation formulation (details given in [11]), a regression model describing the noisy process can be obtained by minimizing (with respect to  $\theta$ ) the cost function of negative log marginal likelihood given by:

$$-\log[p(Y|X, \theta)] = \frac{1}{2} X^T \Sigma^{-1} X + \frac{1}{2} \log |\Sigma| + \frac{n}{2} \log(2\pi); \quad (2)$$

where  $p(Y|X, \theta)$  is the conditional probability of vector  $Y$  given the vector  $X$ ,  $|\cdot|$  is the determinant of a matrix, and  $\Sigma$  is the covariance matrix given by:

$$\Sigma = K(X, X) + \sigma^2 I \quad (3)$$

where  $I$  is the corresponding identity matrix.

After finding  $\theta$  which minimizes the negative log marginal likelihood, that solution can be used in (4) and (5) to predict process outputs  $y^*$  for arbitrary inputs  $x^*$ . These results can be interpreted as a Gaussian distribution in the form of  $y^*|x^* \sim \mathcal{N}(y^*, \sigma_y^*)$ .  $\sigma_y^*$  in (5) is the predicted uncertainty, which is used later in this paper as the criteria for selecting informative measurement points.

$$y^* = K(x^*, X) \Sigma^{-1} Y \quad (4)$$

$$(\sigma_y^*)^2 = \alpha^2 + \sigma^2 - K(x^*, X) \Sigma^{-1} K(X, x^*) \quad (5)$$

## III. PROBLEM FORMULATION

This section presents the system identification problem formulation. Suppose there is a static nonlinear SISO element in the form of  $y = f(x) + \epsilon$  where  $x \in \mathfrak{R}$  is the input,  $y \in \mathfrak{R}$  is the output and  $\epsilon \in \mathfrak{R}$  is noise. The element is subject to the following assumptions:

- 1)  $x$  and  $y$  are measurable.
- 2)  $y$  is a continuous and smooth function of  $x$ , but the structure of the function  $f(\cdot)$  is unknown.
- 3)  $x$  has a range  $x_{min} < x < x_{max}$ , where  $x_{min}, x_{max} \in \mathfrak{R}$ , and the values of  $x_{min}$  and  $x_{max}$  are known.
- 4) The element shows open loop input-to-output stability  $\forall x$  in  $x_{min} < x < x_{max}$ ; implying that  $|y| < y_{max}, \forall y$ , where  $|\cdot|$  denotes absolute value and  $y_{max} \in \mathfrak{R}^+$ .
- 5) Noise  $\epsilon$  is zero mean white noise where  $|\epsilon| < \epsilon_{max}, \forall \epsilon$ , and  $\epsilon_{max} \in \mathfrak{R}^+$ .
- 6)  $\frac{\epsilon_{max}}{y_{max}} < \gamma$  where  $\gamma \in \mathfrak{R}^+, \gamma \ll 1$ .

The objective would now be to produce a computationally efficient model  $\hat{y} = \hat{f}(x)$  which satisfactorily describes the element while not over-fitting the measurements. The approach proposed to achieve this objective, starts by collecting some prior measurements  $X = [x_1, x_2, x_3, \dots, x_n]^T$  and  $Y = [y_1, y_2, y_3, \dots, y_n]^T$  and estimating a GP model

$y_{gp}|x \sim \mathcal{N}(y_{gp}, \sigma_y)$  from those measurements by minimizing the negative log marginal likelihood objective function given in (2). Further measurements  $\{x_{n+1}, y_{n+1}\}$  will be collected via a maximum  $\sigma_y$ -based Active Learning scheme (described in Section IV); those measurements will be added to the sets  $X$  and  $Y$  and recursive estimation of the GP model and data collection will take place until some termination criteria (described in Section IV) indicating the convergence of the GP model is reached. The aim of Active Learning as discussed before, is to reduce the number of measurements collected as opposed to collecting all possible measurements.

In the interest of producing a computationally efficient model, we impose a polynomial structure to  $\hat{f}$  given by  $\hat{y} = \sum_{k=0}^m \phi_k x^k$  and construct the set  $\phi = [m, \phi_1, \phi_2, \dots, \phi_m]$ . Eventually, the solution this paper seeks to find is:

$$\phi^* = \arg \min_{\phi} \frac{1}{N} \sum_{i=1}^N \frac{|y_i - y_{gp_i}|}{|y_{gp_i}|}, \quad (6)$$

which minimizes the Mean Absolute Relative Error (*MARE*) between the polynomial model and the GP model over a user specified  $N$ ,  $N \in \mathbb{Z}^+$  number of function evaluation points where  $y_{gp_i} \neq 0$ , with the aim of the solution  $\phi^*$  providing a polynomial model describing the nonlinear static element whilst not over-fitting to the collected measurements. Section IV details the methodology proposed to solve the problem formulated in this section.

## IV. METHODOLOGY

The methodology is broken down and detailed in the subsections coming within this section. Fig. 1 and 2 at the end of this section depict the methodology in the form of two flowcharts; Fig. 1 depicts the Active Learning scheme proposed to obtain a GP model, and Fig. 2 depicts the subsequent polynomial model identification scheme which relies on the GP model.

### A. The Static Nonlinear Element

The methodology is presented in this section via a demonstrative example using the static nonlinear element given by:

$$y = \exp(-2x) \sin(7\pi x) + \exp(-3x) + 0.5 + \epsilon. \quad (7)$$

Assumptions (1) to (6) listed in Section III are valid for this element;  $x_{min}$  and  $x_{max}$  are set to be 0 and 1 respectively following Assumption (3), and  $\gamma$  is set to be 0.05 following Assumption (6). Readers have the freedom to experiment the proposed approach with different static elements subject to the assumptions.

### B. Performing Measurements

The Active Learning scheme operates using some given measurements  $X = [x_1, x_2, \dots, x_n]^T$ ,  $Y = [y_1, y_2, \dots, y_n]^T$ , and by recursively collecting measurements  $\{x_{n+1}, y_{n+1}\}$  and adding them to the sets  $X$  and  $Y$ . Capturing noise becomes necessary, therefore any measurement instance  $\{x_i, y_i\}$  would have to be repeated multiple times (say  $M$  occasions). As a

result,  $x_i \in X$ ,  $\forall i$  would be given by  $x_i = \frac{1}{M} \sum_{j=1}^M x_{i,j}$ , and

$y_i \in Y$ ,  $\forall i$  would be given by  $y_i = \frac{1}{M} \sum_{j=1}^M y_{i,j}$ . In addition,

variance of  $y_i$  (i.e.,  $\sigma_i^2 = \frac{1}{M} \sum_{j=1}^M (y_{i,j} - y_i)^2$ ) would have to be

recorded  $\forall i$  and be arranged as  $\sigma_{me} = [\sigma_1, \sigma_2, \dots, \sigma_n]^T$  along with sets  $X$  and  $Y$ .  $M$  is set to be 10 for the demonstrative example.

### C. The Active Learning Scheme

1) *Initializing Hyper-Parameters*: Suppose the Active Learning scheme is to be started with a given set of measurements  $X$ ,  $Y$ , and  $\sigma_{me}$  having  $n$  measurements. The first step authors propose is to arrange  $X$  to be in ascending order, i.e., set  $X = [x_1, x_2, \dots, x_n]^T$ , such that  $x_i \leq x_{i+1}$ ,  $\forall i, i \in \mathbb{Z}^+, i \leq n - 1$ , and set  $Y = [y_1, y_2, \dots, y_n]^T$  to be corresponding to  $X$ . Thereafter, initializing hyper-parameters  $(\alpha, \eta, \sigma)$  as follows is proposed:

$$\alpha_{ini} = \frac{1}{n-1} \sum_{i=1}^{n-1} |y_{i+1} - y_i| \quad (8)$$

$$\eta_{ini} = \frac{1}{n-1} \sum_{i=1}^{n-1} |x_{i+1} - x_i| \quad (9)$$

and setting  $\sigma_{ini}$  to have the mean value of elements of  $\sigma_{me}$ .

Since overlapping  $x$  values would be deliberately avoided,  $\eta_{ini}$  will naturally be nonzero. However, since  $y$  is a system response, such a guarantee of being nonzero would not naturally be imposed on  $\alpha_{ini}$ . Should a rare instance resulting  $\alpha_{ini} = 0$  occur for some set  $Y$ , an exception handling routine should be in place to catch such instances and proceed to add one or a few more random measurements to the sets  $X$ ,  $Y$  and  $\sigma_{me}$  in order to capture some degree of non-linearity of the static element in question, and enforce  $\alpha_{ini} \neq 0$ , prior to using the collected measurements to estimate GP models.

2) *Collecting Prior Measurements*: Since  $n$  prior measurements arranged as  $X = [x_1, x_2, \dots, x_n]^T$ ,  $Y = [y_1, y_2, \dots, y_n]^T$ , and  $\sigma_{me} = [\sigma_1, \sigma_2, \dots, \sigma_n]^T$  are required to begin the Active Learning scheme, collecting any  $n$  random measurements would be adequate as long as the condition  $\alpha_{ini} \neq 0$  is satisfied. For the purpose of the demonstrative example though, authors do not resort to random measurements, and go on to set  $n = 2$  and impose just two prior measurements such that  $x_1 = x_{min}$  and  $x_2 = x_{max}$ , since  $x_{min}$  and  $x_{max}$  values will be known beforehand according to Assumption (3). When starting with just two measurements, should the event  $y_1 = y_2$  occur resulting in  $\alpha_{ini} = 0$ , more measurements would have to be added to the prior measurement set as said earlier, before using the prior measurements to estimate GP models.

3) *Estimating GP Models:* Estimating GP models (i.e., models in the form:  $y_{\hat{g}p}|x \sim \mathcal{N}(y_{\hat{g}p}, \sigma_y)$ ) is to be done by using any available measurement sets  $X$  and  $Y$ , and minimizing the negative log marginal likelihood cost function in (2). Starting from the initial values of the hyper-parameters (i.e.,  $\theta_{ini} = \{\alpha_{ini}, \eta_{ini}, \sigma_{ini}\}$ ), gradient-based optimization can be used to reach a nearest local minimum in the cost function [11]. The solution  $\theta^* = \{\alpha^*, \eta^*, \sigma^*\}$  resulting from the local minimum is taken as the solution yielding the GP model from the available set of measurements  $X$  and  $Y$ .

4) *Selecting the Next Best Measurement:* After estimating a GP model (i.e.,  $y_{\hat{g}p}|x \sim \mathcal{N}(y_{\hat{g}p}, \sigma_y)$ ) using measurements  $X = [x_1, x_2, \dots, x_n]^T$ , a new set  $X_t$  is constructed such that  $X_t = [x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_N]^T$  where  $N > n$ . The specialty of  $X_t$  is that it would contain  $N$  equally spaced values, whose minimum and maximum would be  $x_{min}$  and  $x_{max}$  respectively. The set of measurements  $X$  would then be removed from  $X_t$  to construct a new set  $X_t^* = X_t - X$ ,  $X_t^* = [x_{n+1}, x_{n+2}, \dots, x_i, \dots, x_N]^T$ . Uncertainty (i.e.,  $\sigma_y$ ) values will be generated for all points in  $X_t^*$  using (5), and the set  $\sigma_y^* = [\sigma_{n+1}, \sigma_{n+2}, \dots, \sigma_i, \dots, \sigma_N]^T$  will be constructed. The index  $i^*$  corresponding to the maximum uncertainty value in  $\sigma_y^*$  will be selected, and thus, index  $n + 1 = i^*$  will be considered as the next best (i.e.,  $x_{n+1}$ ) measurement to be taken. Suppose  $i^*$  becomes an adjacent point to any of the available measurements, an exception is then made to come to effect to enable  $i^*$  to become a random index not adjacent to or overlapping with any available measurements. This exception allows for greater spatial distribution for the measurements. Subsequently, the next best measurement  $\{x_{n+1}, y_{n+1}\}$  will be added to sets  $X$  and  $Y$ , and the process will follow recursively by re-initializing hyper-parameters, re-estimating the GP model, and relocating the maximum uncertainty from subsequent sets of  $\sigma_y^*$ , until the termination criteria specified in the next subsection is reached. The flowchart in Fig. 1 will be helpful to grasp the Active Learning scheme.

For the purpose of the demonstrative example,  $N$  is set to be 1000.

5) *Termination Criteria:* Since the Active Learning scheme iteratively takes new measurements, the foremost termination criteria imposed is a maximum number of iterations. Function evaluation is done on at most  $N$  points when selecting the next best measurement, and as a result, limiting the maximum number of iterations too to  $N$  is proposed;  $N$  iterations would result in  $N$  equally spaced points covering the whole measurement range according to the method in subsection IV-C4. Therefore, if  $i$  is the number of iterations,  $i = N$  will be one termination criteria.

Next, suppose the hyper-parameters resulting after  $i$  iterations  $2 \leq i < N$  are  $\alpha_i^*$ ,  $\eta_i^*$ , and  $\sigma_i^*$ . The following measures are defined:

$$\gamma_\alpha = \frac{|\alpha_i - \alpha_{i-1}|}{|\alpha_{i-1}|} \quad (10)$$

$$\gamma_\eta = \frac{|\eta_i - \eta_{i-1}|}{|\eta_{i-1}|} \quad (11)$$

$$\gamma_\sigma = \frac{|\sigma_i - \sigma_{i-1}|}{|\sigma_{i-1}|} \quad (12)$$

Ideally,  $\gamma_\alpha$ ,  $\gamma_\eta$ , and  $\gamma_\sigma$  should all tend to zero to imply model convergence. Therefore, the termination criteria is set with respect to a threshold  $\gamma_{hyp}$ ,  $\gamma_{hyp} \in \mathfrak{R}^+$ ,  $\gamma_{hyp} \ll 1$ , such that termination would occur if the three conditions  $\gamma_\alpha < \gamma_{hyp}$ ,  $\gamma_\eta < \gamma_{hyp}$ , and  $\gamma_\sigma < \gamma_{hyp}$  will be satisfied simultaneously, and would remain satisfied for  $N_{hyp}$  consecutive instances (iterations). For the demonstrative example,  $\gamma_{hyp} = 0.05$  and  $N_{hyp} = 10$  values are set.

#### D. Estimating Polynomial Models

The GP model (i.e.,  $y_{\hat{g}p}|x \sim \mathcal{N}(y_{\hat{g}p}, \sigma_y)$ ) resulting from the Active Learning scheme, would itself serve as a model describing the static nonlinear element of interest. However, GP models as such are known to be computationally expensive [11], and therefore are not very likely to be usable as control-oriented computationally efficient models at present. As such, the ultimate solution this paper endeavors to obtain is a model of polynomial architecture in the form of:  $\hat{y} = \sum_{k=0}^m \phi_k x^k$ . Unknowns of this architecture can be included into a set  $\phi = [m, \phi_1, \phi_2, \dots, \phi_m]^T$ , where  $m$  is the polynomial order, and the set  $\phi_c = [\phi_1, \phi_2, \dots, \phi_m]^T$  contains the polynomial coefficients.

For an assumed order  $m$ ,  $\phi_c$  can be found by minimizing the norm  $\|\hat{y} - y_{\hat{g}p}\|$ , where  $\hat{y}$  would be a vector containing  $N$  polynomial model output points, and  $y_{\hat{g}p}$  would be a corresponding vector containing corresponding GP model outputs. When the polynomial model is expressed in its matrix form:  $\hat{y} = X_p \phi_c$ ,  $\phi_c$  can be found by linear least squares as:

$$\phi_c = (X_p^T X_p)^{-1} X_p^T y_{\hat{g}p}. \quad (13)$$

Thus, by assuming a polynomial order  $m$ , and solving for  $\phi_c$  using (13) would yield a polynomial model describing the static nonlinear element, solving the problem which this paper focused on. It is proposed to improve the polynomial model by starting from  $m = 1$ , and recursively increasing the polynomial order while solving for coefficients  $\phi_c$ . Such an approach aligns with the Weierstrass approximation theorem [21] and can produce polynomial models having a low residual error. The termination criteria in (14) is proposed for the recursion to stop. As a result of this termination criteria, recursion would stop when the *MARE* (expressed in (15)) between the polynomial model and the GP model falls below a threshold  $\gamma_p$ ,  $\gamma_p \in \mathfrak{R}^+$ , and  $\gamma_p \ll 1$ . The flowchart in Fig. 2 will be helpful to grasp the polynomial model fitting scheme.

$$MARE < \gamma_p \quad (14)$$

$$MARE = \frac{1}{N} \sum_{i=1}^N \frac{|\hat{y}_i - y_{\hat{g}p_i}|}{|y_{\hat{g}p_i}|} \quad (15)$$

For the purpose of the demonstrative example,  $\gamma_p$  is set to be 0.01.

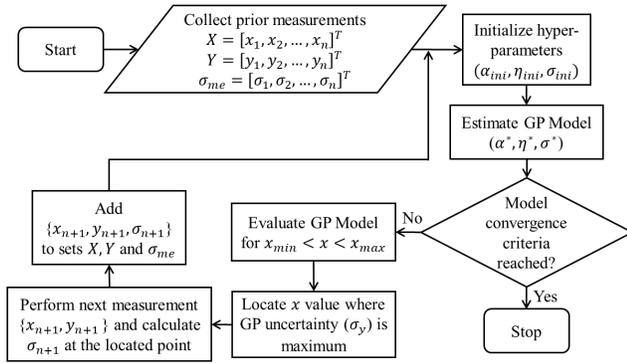


Fig. 1. Flowchart depicting the proposed Active Learning scheme.

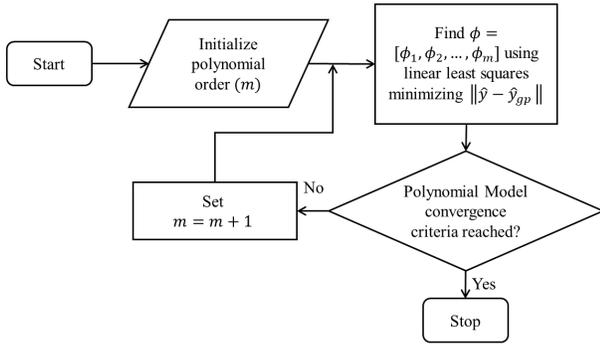


Fig. 2. Flowchart depicting the polynomial fitting scheme.

## V. RESULTS

The parameter values used for the demonstrative example involving the static nonlinear element in (7), are summarized in Table I, and the proposed termination criteria can be summarized as follows:

TABLE I  
PARAMETERS USED FOR THE DEMONSTRATIVE EXAMPLE

Parameter	Symbol	Value
Minimum of $x$	$x_{min}$	0
Maximum of $x$	$x_{max}$	1
Noise Threshold	$\gamma$	0.05
Repetitions per measurement (to start the Active Learning scheme)	$M$	10
Number of prior measurements	$n$	2
Number of function evaluation points	$N$	1000
Convergence threshold for hyper-parameters	$\gamma_{hyp}$	0.05
Number of consecutive instances $\gamma_{hyp}$ has to be met	$N_{hyp}$	10
Convergence threshold for polynomial models	$\gamma_p$	0.01

- The Active Learning scheme terminates when either the number of iteration  $i$  becomes equal to the maximum number of function evaluation points  $N$ , or when  $\gamma_\alpha$ ,  $\gamma_\eta$  and  $\gamma_\sigma$  (described in (10), (11), and (12)) satisfy the

conditions  $\gamma_\alpha < \gamma_{hyp}$ ,  $\gamma_\eta < \gamma_{hyp}$ , and  $\gamma_\sigma < \gamma_{hyp}$ , for an  $N_{hyp}$  number of consecutive instances (iterations).

- The polynomial fitting terminates when  $MARE$  (described in (15)) satisfies  $MARE < \gamma_p$ .

The identification exercise was repeated 50 times to assess the performance of the methodology. Fig. 3 shows the histogram of the number of iterations taken for the Active learning scheme to converge. Fig. 4 shows measurements taken in a trial which converged in 164 Active Learning iterations and Fig. 5 shows the performance of the GP and polynomial models estimated from this data, alongside the noisy ( $\gamma = 0.05$ ) static element. How termination criteria for both the Active Learning scheme and the polynomial fitting scheme are met for this example are illustrated in Fig. 6 and Fig. 7.

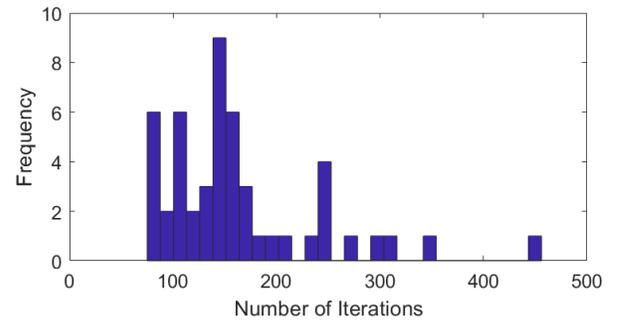


Fig. 3. Histogram over 50 trials, showing number of iterations taken for the Active Learning scheme to converge, mean = 164 iterations, std = 76 iterations.

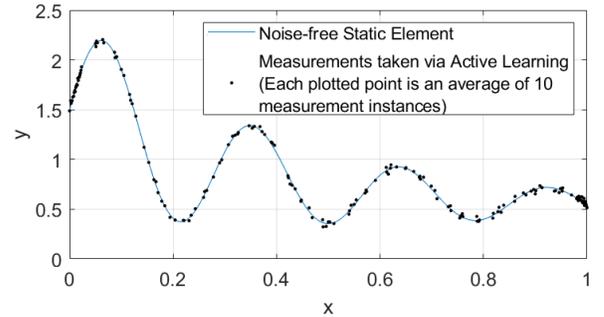


Fig. 4. Measurements taken in a trial which converged in 164 Active Learning iterations.

## VI. CONCLUSIONS

An approach unifying Active Learning, over-fit avoidance, and model structure determination (polynomial architecture) for system identification of static nonlinear elements was proposed. The approach includes two steps: (1) Using a GP-based Active Learning scheme to estimate a smooth nonlinear model while collecting an optimal set of measurements in a sense—such a method constitutes a framework for data collection which guides users about the sufficiency or insufficiency of measurements collected; (2) A polynomial model

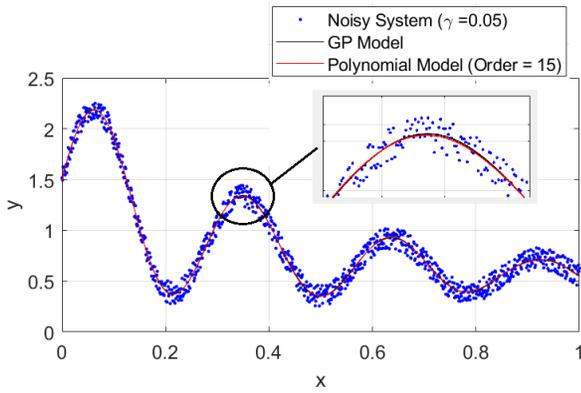


Fig. 5. Estimated GP and polynomial models alongside the noisy ( $\gamma = 0.05$ ) system.

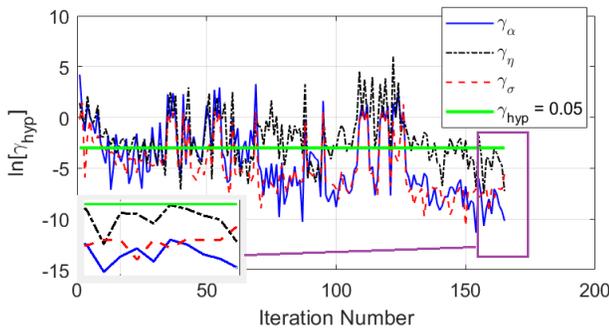


Fig. 6. Variation of the natural logarithm of  $\gamma_\alpha$ ,  $\gamma_\eta$ , and  $\gamma_\sigma$  values against iterations with zoomed view of convergence region satisfying the termination criteria for the Active Learning scheme described in subsection IV-C5.

fitting scheme using the estimated GP model. Estimating the smooth GP model avoids over-fitting; the subsequent polynomial fitting delivers a model that is computationally efficient and control-oriented than the GP model. Techniques relevant for parameter initialization and algorithm termination were introduced and the effectiveness of the approach was demonstrated for the SISO case via an example. Since this work focused only on static elements, future work can focus on extending similar schemes to be generalized over dynamical systems as well as MIMO cases. Another interesting question to focus on will be to eliminate the use of GP (as it can be computationally expensive for large data sets) and propose Active Learning schemes based on more computationally efficient means while being model-free (i.e., not relying on an assumed model or a set of assumed models). Relaxing some assumptions made in this paper will also be interesting.

## REFERENCES

- [1] L. Eriksson and L. Nielsen, *Modeling and control of engines and drivelines*. John Wiley & Sons, 2014.
- [2] S. A. Billings and S. Fakhouri, "Identification of systems containing linear dynamic and static nonlinear elements," *Automatica*, vol. 18, no. 1, pp. 15–26, 1982.
- [3] K. Yu, J. Bi, and V. Tresp, "Active learning via transductive experimental design," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 1081–1088.

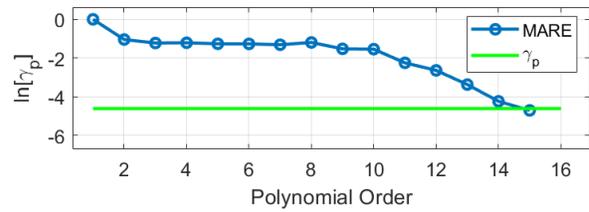


Fig. 7. Variation of  $\ln[MARE]$  against polynomial order showing the termination criteria (for polynomial fitting) crossing point described in (14).

- [4] T. RayChaudhuri and L. G. Hamey, "Active learning for nonlinear system identification and control," in *Proc. IFAC World Congress*, 1996, pp. 193–197.
- [5] L. Ljung, *System Identification: Theory for the User*, ser. Prentice Hall information and system sciences series. Prentice Hall PTR, 1999. [Online]. Available: <https://books.google.com.au/books?id=nHFoQgAACAAJ>
- [6] C. R. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer, "Robust optimal experiment design for system identification," *Automatica*, vol. 43, no. 6, pp. 993–1008, 2007.
- [7] M. Deflorian and S. Zaglauer, "Design of experiments for nonlinear dynamic system identification," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 13 179–13 184, 2011.
- [8] J. J. Rissanen, "Fisher information and stochastic complexity," *IEEE transactions on information theory*, vol. 42, no. 1, pp. 40–47, 1996.
- [9] T. J. Broomhead, "A model based approach to the control of diesel generators," Ph.D. dissertation, 2016.
- [10] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992, pp. 287–294.
- [11] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.
- [12] S. Billings, S. Chen, and M. Korenberg, "Identification of mimo nonlinear systems using a forward-regression orthogonal estimator," *International journal of control*, vol. 49, no. 6, pp. 2157–2189, 1989.
- [13] Y. Guo, L. Z. Guo, S. A. Billings, and H.-L. Wei, "Identification of nonlinear systems with non-persistent excitation using an iterative forward orthogonal least squares regression algorithm," *International Journal of Modelling, Identification and Control*, vol. 23, no. 1, pp. 1–7, 2015.
- [14] N. Ulapane, A. Alempijevic, T. Vidal-Calleja, J. V. Miro, J. Rudd, and M. Roubal, "Gaussian process for interpreting pulsed eddy current signals for ferromagnetic pipe profiling," in *2014 9th IEEE Conference on Industrial Electronics and Applications*. IEEE, 2014, pp. 1762–1767.
- [15] N. N. Ulapane and S. G. Abeyratne, "Gaussian process for learning solar panel maximum power point characteristics as functions of environmental conditions," in *2014 9th IEEE Conference on Industrial Electronics and Applications*. IEEE, 2014, pp. 1756–1761.
- [16] A. M. N. N. B. Ulapane, "Nondestructive evaluation of ferromagnetic critical water pipes using pulsed eddy current testing," Ph.D. dissertation, University of Technology Sydney, 2016.
- [17] K. Thiyagarajan, S. Kodagoda, and N. Ulapane, "Data-driven machine learning approach for predicting volumetric moisture content of concrete using resistance sensor measurements," in *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2016, pp. 1288–1293.
- [18] N. Ulapane, A. Alempijevic, T. Vidal Calleja, and J. Valls Miro, "Pulsed eddy current sensing for critical pipe condition assessment," *Sensors*, vol. 17, no. 10, p. 2208, 2017.
- [19] N. Ulapane, L. Nguyen, J. V. Miro, and G. Dissanayake, "A solution to the inverse pulsed eddy current problem enabling 3d profiling," in *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2018, pp. 1267–1272.
- [20] J. Valls Miro, N. Ulapane, L. Shi, D. Hunt, and M. Behrens, "Robotic pipeline wall thickness evaluation for dense nondestructive testing inspection," *Journal of Field Robotics*, vol. 35, no. 8, pp. 1293–1310, 2018.
- [21] M. H. Stone, "The generalized weierstrass approximation theorem," *Mathematics Magazine*, vol. 21, no. 5, pp. 237–254, 1948.