2015 IEEE 54th Annual Conference on Decision and Control (CDC)
December 15-18, 2015. Osaka, Japan

A Bayesian Approach for Model Identification of LPV Systems with Uncertain Scheduling Variables*

Farshid Abbasi[1], Javad Mohammadpour[1], Roland Tóth[2], Nader Meskin[3]

Abstract— **This paper presents a Gaussian Process (GP) based Bayesian method that takes into account the effect of additive noise on the scheduling variables for identification of linear parameter-varying (LPV) models in input-output form. The proposed method approximates the noise-free coefficient functions by a local linear expansion on the observed scheduling variables. Therefore, additive noise on the scheduling variables is reconstructed as a corrective term added to the output noise that is proportional to the squared gradient obtained from the posterior of the Gaussian Process. An iterative procedure is given so that the obtained solution converges to the best estimation of the coefficient functions according to the given measure of fitness. Moreover, the expectation and covariance functions estimated by GP are modified for the noisy scheduling variable case to include the noise contribution on the estimated expectation and covariance functions. The model training procedure identifies noise level in the measurements including outputs and scheduling variables by estimating the noise variances, as well as other defined hyperparameters. Finally, the performance of the proposed method is compared to the standard GP approach through a numerical example.**

I. Introduction

Most of the existing methods for model identification of *linear parameter-varying* (LPV) systems consider the scheduling variables to be noise free. However, the presence of uncertainty, i.e., noise, in the measured data including the scheduling variables is inevitable and can lead to an inaccurate model identification. Hence, the precise knowledge of scheduling variables in the presence of uncertainties is a critical issue in both LPV model identification and LPV control design.

Several identification methods have been recently proposed to cope with noisy scheduling variables corresponding to the so-called error-in-variables problem in the context of *linear time-invariant* (LTI) systems [1]. Unlike the LTI framework, nonlinear dependency of the *linear parameter-varying* LPV model coefficients on the scheduling variables is considered to be the main source of complexity in coping with the noise corrupted scheduling variables. There are very few works examining the model identification of LPV systems considering noise corrupted scheduling variables. The previous approaches of [4], [5] have focused on the identification of LPV *input/output* (LPV-IO) models using set-membership and *instrumental variable* (IV) based methods. More specifically, a convex relaxation approach is proposed in [5] under the assumption that all the noisy observations including outputs and scheduling variables are bounded. Moreover, the IV-based method presented in [4] is capable of coping with noisy scheduling variables assuming that the instrument is uncorrelated with the scheduling variable noise and the scheduling dependency is linear. More recently, a bias-corrected, IV-based method has been developed for the identification of LPV models from noise corrupted measurements of the outputs and the scheduling variables [6]. While, the recent works have offered significant improvement for the identification of LPV systems, they, however, have assumed that the dependency on the scheduling variables is *a priori* known. The present work introduces a Bayesian-based approach assuming *a priori* unknown dependency, only characterized in terms of a prior distribution, on the noise corrupted scheduling variables. The Bayesian-based approaches provide a rich variety of *a priori* kernels that can effectively characterize such distributions and hence identify structural characteristics of the systems under study [7]. The Bayesian formulation is based on the expression of the beliefs about the prior information or measurements through specification of *a priori* knowledge before observing new data. In this paper, an extension of one of such approaches, namely a Gaussian Process (GP) based approach is formulated to identify the dependency of LPV model coefficients on the scheduling variables while they are corrupted with a Gaussian noise process.

Throughout this paper, notation $A \odot B$ is used to represent the Hadamard product of the matrices A and B of the same dimension such that $[A \odot B]_{ij} = [A_{ij}] \cdot [B_{ij}]$. In addition, I_N, $\mathbb{R}$, $\mathbb{Z}$ and $\mathbb{R}^n$ denote the $N \times N$ identity matrix, the set of real numbers, the set of integer numbers and the set of n-dimensional vector space with real elements, respectively, and $(\cdot)^\top$ represents the transpose of the associated vector or matrix.

The rest of the paper is organized as follows. Section II describes the LPV model formulation. Section III explains the principles of Gaussian processes and the corresponding Bayesian identification framwork. The formulation of the error in the scheduling variables problem is given in Section IV. Finally, simulation results are shown in Section VI, and concluding remarks are provided in Section VII.

*This work was made possible by the NPRP grant # NPRP 5-574-2-233 from Qatar National Research Fund (QNRF), a member of Qatar Foundation. The statements made here are solely the responsibility of the authors.

[1]Complex Systems Control Laboratory (CSCL), College of Engineering, The University of Georgia, Athens, GA 30602, E-mail: javadm@uga.edu.

[2]Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands, E-mail: r.toth@tue.nl.

[3]Dept. of Electrical Engineering, Qatar University, Doha, Qatar, E-mail: nader.meskin@qu.edu.qa.

978-1-4799-7886-1/15/$31.00 ©2015 IEEE

II. LPV Input-Output Models

We consider a *single-input single-output* (SISO) linear parameter-varying (LPV) system defined in the auto-regressive form with exogenous input (ARX) as

$$y(k) = -\sum_{i=1}^{n_a} a_i(p(k))y(k-i) + \sum_{i=0}^{n_b} b_i(p(k))u(k-i) + e(k), \tag{1}$$

where $k \in \mathbb{Z}$ denotes the discrete time, $y : \mathbb{Z} \to \mathbb{R}$ is the system output and $u : \mathbb{Z} \to \mathbb{R}$ is the system input. Also, $p : \mathbb{Z} \to \mathbb{P}$ is the so-called *scheduling variable* with $\mathbb{P} \subseteq \mathbb{R}^{n_p}$ and $e(k)$ is an *independent and identically distributed* (i.i.d.) white stochastic noise process that is independent of u and p. The coefficients $a_i(p(k)), b_i(p(k))$ are assumed to be bounded possibly nonlinear functions over $\mathbb{P}$ that fully characterize the LPV model (1). To estimate the structure of the model coefficient functions usually requires the parametrization of a_i and b_i in terms of *a priori* known basis functions. To avoid difficulties arising from an inappropriate selection of the basis functions, nonparametric approaches have been proposed in the literature [8]. This can favor the LPV modeling of time-varying or nonlinear systems specially in case of the dynamic dependencies of the model coefficients on the scheduling variables, i.e., dependency on $p(k),\ p(k-1), \ldots$.

Both parametric and nonparametric approaches for LPV model identification aim at describing the underlying dependencies of the model coefficients on the scheduling variables. However, often only a measured version of $p(k)$ is available, polluted by noise. As a result, estimation of the dependencies of the coefficients on the scheduling variables often leads to bias of the estimated coefficient functions, referred to as an *error-in-variables* problem. The present work is an effort to identify and compensate for the error in the scheduling variables by a modified Bayesian approach. The model (1) can be represented in a more compact form by introducing the following notations:

$$x_i(k) = -y(k-i),\ i = 1, \ldots, n_a, \tag{2a}$$

$$x_{n_a+j+1}(k) = u(k-j),\ j = 0, \ldots, n_b, \tag{2b}$$

$$\mathbf{x}(k) = [x_1(k)\ x_2(k)\ \ldots\ x_{n_a+n_b+1}(k)]^\top. \tag{2c}$$

Additionally, if the coefficient function vector is defined as

$$\mathbf{g} = [\mathbf{g}_1\ \ldots\ \mathbf{g}_{n_g}] = [a_1\ \ldots\ a_{n_a}\ b_0\ \ldots\ b_{n_b}], \tag{3}$$

with $n_g = n_a + n_b + 1$, then

$$y(k) = \mathbf{g}(p(k))\mathbf{x}(k) + e(k). \tag{4}$$

Equation (4) can be rewritten as

$$y(k) = \sum_{i=1}^{n_g} \mathbf{g}_i(p(k))x_i(k) + e(k), \tag{5}$$

where $x_i(k)$ indicates the i-th entry of the vector $\mathbf{x}(k)$ and $\mathbf{g}_i$ is the i-th entry of the vector function $\mathbf{g}$.

III. Introduction to Gaussian Processes

Gaussian process (GP) has been introduced to capture functional maps from observations and find the posterior distributions of the underlying functional dependencies over the observed data. The GP regression model in the dynamic case can be described as

$$y(k) = \mathcal{F}(\mathcal{D}(k)) + e(k), \tag{6}$$

where $\mathcal{D}(k)$ is the vector of the observations, and e is an i.i.d. noise process with $e(k) \sim \mathcal{N}(0, \sigma_e^2)$, denoting a normal distribution with zero mean and variance σ_e^2. In the LPV context, the GP method is adopted to estimate the model coefficients $\mathbf{g}$ and their dependencies on p assuming that $\mathcal{F}$ is describable as a particular realization of a Gaussian process with a zero-mean prior distribution with a previously chosen symmetric positive definite covariance function $\mathcal{K}(\cdot,\cdot)$ as

$$\mathcal{F}(\cdot) = \mathcal{GP}\big(0, \mathcal{K}(\cdot,\cdot)\big), \tag{7}$$

where $\mathcal{GP}$ denotes the Gaussian process [9]. Accordingly, the joint distribution of the output data y (conditioned) w.r.t. a given data set $\mathcal{D}$ and a test output data $\mathcal{F}_*$ is

$$\begin{bmatrix} y \\ \mathcal{F}_* \end{bmatrix} = \mathcal{N}\left(0, \begin{bmatrix} \mathcal{K}(\mathcal{D},\mathcal{D}) + \sigma_e^2 I_N & \mathcal{K}(\mathcal{D},\mathcal{D}_*) \\ \mathcal{K}(\mathcal{D}_*,\mathcal{D}) & \mathcal{K}(\mathcal{D}_*,\mathcal{D}_*) \end{bmatrix}\right), \tag{8}$$

where $\mathcal{D}_*$ is a given set of test points and $\mathcal{D}$ is the given set of observations. It should be noted that if there are N_* test points and N training data, then the covariance matrix $\mathcal{K}(\mathcal{D},\mathcal{D}_*)$ would be an $N \times N_*$ matrix. Hence, to obtain the posterior distribution over functions, the joint distribution is conditioned on the observations. The following predictive equations can be obtained by deriving the conditional distributions [9]

$$\mathcal{F}_*|(\mathcal{D}, y, \mathcal{D}_*) \sim \mathcal{N}(\bar{\mathcal{F}}_*, \mathrm{Cov}(\mathcal{F}_*)), \tag{9a}$$

$$\bar{\mathcal{F}}_* \triangleq \mathbb{E}[\mathcal{F}_*|(\mathcal{D}, y, \mathcal{D}_*)] = \mathcal{K}(\mathcal{D}_*,\mathcal{D})[\mathcal{K}(\mathcal{D},\mathcal{D}) + \sigma_e^2 I_N]^{-1} y, \tag{9b}$$

$$\begin{aligned}\mathrm{Cov}(\mathcal{F}_*) = {} & \mathcal{K}(\mathcal{D}_*,\mathcal{D}_*) \\ & - \mathcal{K}(\mathcal{D}_*,\mathcal{D})[\mathcal{K}(\mathcal{D},\mathcal{D}) + \sigma_e^2 I_N]^{-1}\mathcal{K}(\mathcal{D},\mathcal{D}_*).\end{aligned} \tag{9c}$$

The mean and covariance functions obtained from (9b)-(9c) can statistically characterize the coefficients of the LPV model (5).

IV. Formulation of the Error in Scheduling Variables Problem

Gaussian processes have been successfully applied to a variety of applications in the context of dynamic systems and proven that they can accurately capture the underlying mapping of the input space to the output space. However, there are some limitations due to the assumptions made about the noise conditions. The standard GP algorithm is based on the assumption that the input data are free of measurement errors and also independent from the noise process which is a white (stationary) Gaussian noise. However, it is very likely – specially in industrial processes – that the input data are also corrupted by signal-independent sensor noise. As described earlier, in the present work, the scheduling variables available in the data set are assumed to be corrupted with an i.i.d. Gaussian noise. Let $\breve{p}$ denote the n_p-dimensional scheduling variable vector defined as

$$\breve{p}_k = p_k + \varepsilon_p(k), \tag{10}$$

where $\varepsilon_p \sim \mathcal{N}(0, \Sigma_p)$ is a white Gaussian noise that is independent of u and e. p is the noise-free scheduling variable that actually affects the underlying system. To simplify the notation, p_k is used instead of $p(k)$. It is assumed that the scheduling variables are independently corrupted by noise, and hence the noise variance Σ_p is a diagonal matrix. In the LPV model (5), the coefficients are functions of the noisy scheduling variables and calculating the posterior distribution is intractable using the standard GP framework. We employ the first order approximation of the model coefficients obtained using Taylor's series expansion on the observed data as

$$\mathbf{g}_i(\breve{p}_k - \varepsilon_p(k)) \approx \mathbf{g}_i(\breve{p}_k) - \varepsilon_p^\top \frac{\partial \mathbf{g}_i(\breve{p}_k)}{\partial p} + \cdots. \tag{11}$$

Since the derivative of a Gaussian process is itself another Gaussian process [10], the previous assumptions still hold true for the Taylor's expansion of the coefficients. However, one might argue that these functions are not available and need to be identified, which will be later examined in this section. Approximation in (11) gives a good estimation of the effect of scheduling variables noise on the function evaluation. We note that additional terms can also be kept beyond the affine approximation in (11) at the expense of more complexity eventually leading to a much higher computational load. Substituting (11) into (5) and considering the derivative of the expectation (mean) of the LPV model coefficients obtained by Gaussian process, we have

$$y(k) = \sum_{i=1}^{n_g} \mathbf{g}_i(\breve{p}_k) x_i(k) - \sum_{i=1}^{n_g} \varepsilon_p^\top \frac{\partial \bar{\mathbf{g}}_i(\breve{p}_k)}{\partial p} x_i(k) + e(k), \tag{12}$$

where $\bar{\mathbf{g}}_i(\breve{p}_k)$ represents the mean value of the LPV model coefficients at an observed scheduling variable. The obtained heteroscedastic model considers the errors in both noisy output and noisy scheduling variables. Hence, the new error term can be considered as

$$\tilde{e}(k) = -\sum_{i=1}^{n_g} \varepsilon_p^\top \frac{\partial \bar{\mathbf{g}}_i(\breve{p}_k)}{\partial p} x_i(k) + e(k) \tag{13}$$

According to (12) and (13), the probability of the output y given the functions $\mathbf{g}_i, i = 1, \ldots, n_g$, and data set $\mathcal{D} = \{\breve{p}(k), y(k), u(k)\}_{k=1}^N$ can be obtained as

$$P(y|\ (\mathbf{g}, \mathcal{D})) = \mathcal{N}\big(\mathbb{E}(y), \sum_{i=1}^{n_g} x_i(m) \frac{\partial \bar{\mathbf{g}}_i(\breve{p}_m)}{\partial p}^\top \Sigma_p \frac{\partial \bar{\mathbf{g}}_i(\breve{p}_n)}{\partial p} x_i(n) + {\sigma_e}^2\big), \tag{14}$$

$$\mathbb{E}(y) = \sum_{i=1}^{n_g} \mathbf{g}_i(\breve{p}_k) x_i(k). \tag{15}$$

This can be seen as an equivalent formulation to considering the given scheduling variables as deterministic and adding a corrective term to the output error term. To obtain the posterior distribution, the prior is considered as the standard GP (6)

$$P\big(\mathbf{g}_i(p_k) x_i(k) | \mathcal{D}\big) = \mathcal{N}\big(0, \mathcal{K}^i(\mathcal{D}, \mathcal{D})\big), \tag{16}$$

where $\mathcal{K}^i(\mathcal{D}, \mathcal{D})$ is the $N \times N$ symmetric covariance matrix defined as

$$\mathcal{K}(\mathcal{D}_m, \mathcal{D}_n) = \sum_{i=1}^{n_g} x_i(m) k^i(\breve{p}_m, \breve{p}_n) x_i(n), \tag{17}$$

$$k^i(\breve{p}_m, \breve{p}_n) = \lambda_i \exp\Big((\breve{p}_m - \breve{p}_n)^\top \mathcal{W}_i^{-1} (\breve{p}_m - \breve{p}_n)\Big), \tag{18}$$

where $\mathcal{W}_i$ is the diagonal matrix of characteristic length-scale, and λ_i is a positive scalar factor representing the value of the covariance function when $\breve{p}_m$ and $\breve{p}_n$ are very close. Using the (approximation) LPV model in (12), the gradient term can be considered as a secondary error term to compensate for the error in the scheduling variables and its effect on the output. Hence, similar to the variance of the output error, only the elements on the diagonal are kept for calculating the joint covariance matrix. The associated diagonal matrix is defined as

$$\mathcal{Q}_i(n, n) = x_i(n) \frac{\partial \bar{\mathbf{g}}_i(\breve{p}_n)}{\partial p}^\top \Sigma_p \frac{\partial \bar{\mathbf{g}}_i(\breve{p}_n)}{\partial p} x_i(n), \tag{19}$$

for $n = 1, \ldots, N$. The calculated probabilities (14) and (16) are combined to obtain the following posterior distribution

$$\begin{bmatrix} y \\ \mathbf{g}_i(\mathcal{P}^*) \end{bmatrix} = \mathcal{N}\left(0, \begin{bmatrix} \mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I_N + \mathcal{Q} & \kappa^i(\mathcal{D}, \mathcal{P}^*) \\ \kappa^i(\mathcal{P}^*, \mathcal{D})^\top & k^i(\mathcal{P}^*, \mathcal{P}^*) \end{bmatrix}\right), \tag{20}$$

where $\mathcal{P}^* = \breve{p}_i^*,\ i = 1, \ldots, N_*$, is a test point and $\kappa^i(\mathcal{D}, \mathcal{P}^*)$ is defined as follows

$$\kappa^i(\mathcal{P}^*, \mathcal{D}) = \kappa^i(\mathcal{D}, \mathcal{P}^*) = \begin{bmatrix} x_i(1) k^i(\breve{p}_1, \mathcal{P}^*) \\ x_i(2) k^i(\breve{p}_2, \mathcal{P}^*) \\ \vdots \\ x_i(N) k^i(\breve{p}_N, \mathcal{P}^*) \end{bmatrix}, \tag{21}$$

and $\mathcal{Q}$ is the $N \times N$ matrix of the derivatives calculated as

$$\mathcal{Q} = \sum_{i=1}^{n_g} \mathcal{Q}_i, \tag{22}$$

and $\mathrm{Cov}(\mathbf{g}_i(\breve{p}_m), \mathbf{g}_i(\breve{p}_n)) = k^i(\breve{p}_m, \breve{p}_n)$ is the covariance or *kernel function* given by (18). According to (20), the posterior mean and covariance are obtained as

$$\bar{\mathbf{g}}_i = \mathbb{E}[\mathbf{g}_i(\mathcal{P}^*) | \mathcal{D}, \mathcal{P}^*] = \kappa^i(\mathcal{P}^*, \mathcal{D}) \big[\mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I_N + \mathcal{Q}\big]^{-1} \mathcal{Y}, \tag{23}$$

$$\mathrm{Cov}[\mathbf{g}_i(\mathcal{P}^*)] = k^i(\mathcal{P}^*, \mathcal{P}^*) - \kappa^i(\mathcal{P}^*, \mathcal{D}) \big[\mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I_N + \mathcal{Q}\big]^{-1} \kappa^i(\mathcal{D}, \mathcal{P}^*), \tag{24}$$

where $\mathcal{Y} = [y(1), y(2), \ldots, y(N)]^\top$. To simplify the notation we define α as

$$\alpha = \Big(\mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I_N + \mathcal{Q}\Big)^{-1} \mathcal{Y}. \tag{25}$$

As observed, $\bar{\mathbf{g}}_i$ is dependent on its derivative, and hence an analytical solution does not exist to the resulting equations. Hence, an iterative procedure is proposed here. To this purpose, we first calculate α using standard GP, i.e., from (25) without the derivative term $\mathcal{Q}$. Then, the $\mathcal{Q}_i$'s are obtained by substituting α into (26). $\mathcal{Q}$ would then be computed from (22) and replaced in (25) to find α. This procedure is repeated until it converges to the best estimation of the system output through estimating the coefficient functions of the LPV model. The measure of the fitness along with more technical details are explained later in this section. After calculating the derivatives, substituting them in (19) and defining $\mathcal{P} = \{\breve{p}_1, \breve{p}_2, \ldots, \breve{p}_N\}$, we obtain

$$\mathcal{Q}_i(n,n) = x_i(n)\alpha^\top \big(2\Delta_n \odot \kappa^i(\breve{p}_n,\mathcal{P})^\top\big)^\top \mathcal{W}_i^{-\top}\Sigma_p\mathcal{W}_i^{-1} \big(2\Delta_n \odot \kappa^i(\breve{p}_n,\mathcal{P})^\top\big)\alpha x_i(n), \quad (26)$$

where Δ_n is defined as $\Delta_n = [\breve{p}_n - \breve{p}_1, \breve{p}_n - \breve{p}_2, \ldots, \breve{p}_n - \breve{p}_N]^\top$ and $\kappa^i(\breve{p}_n,\mathcal{P})$ is a vector defined as

$$\kappa^i(\breve{p}_n,\mathcal{P}) = \Big[x_i(1)k^i(\breve{p}_n,\breve{p}_1)) \,, \cdots ,\, x_i(N)k^i(\breve{p}_n,\breve{p}_N)\Big]. \quad (27)$$

It should be noted that for calculating the covariance matrix, the coefficient functions are assumed to be mutually independent and hence their associated derivatives are also mutually independent [10]. The added diagonal matrix $\mathcal{Q}$ to the output noise variance in (20) is a corrective term that compensates for the error in the scheduling variables by taking into account the effect of the gradient of the mean function as a measure of sensitivity to noise-corrupted scheduling variables. Since, the corrective term $\mathcal{Q}$ needs to be found to calculate the expectation of the LPV model coefficients, an iterative procedure is defined. First, the gradient of the estimated coefficients $\bar{\mathbf{g}}_i(p_k)$ by the standard GP are calculated and substituted in (20). In fact, we find the derivative of the coefficients from the mean function obtained via the standard GP at the training points. The obtained gradient is used to calculate the corrective additive term to update the probability distribution (20). Next, the updated distribution is used to estimate the coefficient functions and system output accordingly. Then, the hyperparameters including the noise variance of n_p scheduling variables and that of the output are tuned through trial and error to maximize the so-called *best fit ratio* (BFR) defined by

$$\text{BFR} := 100\% \cdot \max\left(1 - \frac{\|y(k) - \hat{y}(k)\|_{l_2}}{\|y(k) - \bar{y}\|_{l_2}}, 0\right), \quad (28)$$

which is considered to be the fitness score [11]. In (28), $\hat{y}$ is the simulated output of the estimated model, y is the true output and $\bar{y}$ represents the mean of the true output y. Next, the gradient of the estimated posterior mean $\bar{\mathbf{g}}_i(p_k)$ is used to update the corrective term and retrain the process. The procedure is iterative and continues until the maximum BFR is achieved.

Learning with Uncertain Scheduling Variables

The expectation (23) and covariance (24) of the coefficient functions obtained by GP are modified to include the noise contribution in the scheduling variables. To this aim, the expectation of the modified mean and covariance are obtained by integrating over the distribution of the scheduling variables [12]. In the present work, the test points are assumed to be a set of Gaussian distributions, and hence, the integral is analytically tractable. It should be noted that the true scheduling variables are not observable; however, we have access to their distribution $\mathcal{N}(\mathcal{P}^*, \Sigma_p)$, where $\mathcal{P}^*$ is the observed test point [13]. Therefore, the noise-free scheduling variables are assumed to be Gaussian distributed $\tilde{\mathcal{P}}^* \sim \mathcal{N}(\mathcal{P}^*, \Sigma_p)$, where $\tilde{\mathcal{P}}^* = p_i^*, \ i = 1, \ldots, N_*$. According to the given distribution on the scheduling variables, the expectation of the covariance function (18) is obtained as

$$\begin{aligned} k_*^i(\mathcal{P}^*, \breve{p}_k) &= \mathbb{E}_{\tilde{\mathcal{P}}^*}[k^i(\tilde{\mathcal{P}}^*, \breve{p}_k)|\mathcal{P}^*, \Sigma_p] \\ &= \int_{-\infty}^{+\infty} k^i(\tilde{\mathcal{P}}^*, \breve{p}_k)P(\tilde{\mathcal{P}}^*|\mathcal{P}^*, \Sigma_p)\mathrm{d}\tilde{\mathcal{P}}^*. \end{aligned} \quad (29)$$

Eventually, we have

$$k_*^i(\mathcal{P}^*, \breve{p}_k) = \lambda_i \mid I + W_i^{-1}\Sigma_p \mid^{-\frac{1}{2}} \exp\Bigg(- (\mathcal{P}^* - \breve{p}_k)^\top(\mathcal{W}_i + \Sigma_p)^{-1}(\mathcal{P}^* - \breve{p}_k)\Bigg). \quad (30)$$

The expected value of the LPV model coefficients given the observed scheduling variables is obtained from (20) and (29) as

$$\begin{aligned} &\mathbb{E}_{\tilde{\mathcal{P}}^*}[\mathbf{g}_i(\tilde{\mathcal{P}}^*) \mid (\mathcal{P}^*, \mathcal{D})] = \bar{\mathbf{g}}_i(\mathcal{P}^*) \\ &\quad = \kappa_*^i(\mathcal{P}^*, \mathcal{D})^\top \Bigg(\mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I_N + \mathcal{Q}\Bigg)^{-1} \mathcal{Y}, \end{aligned} \quad (31)$$

where $\mathcal{Y} = [y(1) \ \ y(2) \ \ \ldots \ \ y(N)]$ and $\kappa_*^i(\mathcal{P}^*, \mathcal{D})$ is as defined in (21) considering the expectation of the covariance function $k_*^i(\mathcal{P}^*, \breve{p}_k)$ instead of the previously defined $k^i(\mathcal{P}^*, \breve{p}_k)$. To calculate the predictive covariance, the total covariance law is implemented (see [14]), where

$$\begin{aligned} \text{Cov}[\mathbf{g}_i(\tilde{\mathcal{P}}^*) \mid (\mathcal{P}^*, \mathcal{D})] &= \mathbb{E}_{\tilde{\mathcal{P}}^*}[\text{Cov}[\mathbf{g}_i(\tilde{\mathcal{P}}^*) \mid (\mathcal{P}^*, \mathcal{D})]] \\ &\quad + \text{Cov}_{\tilde{\mathcal{P}}^*}[\bar{\mathbf{g}}_i(\tilde{\mathcal{P}}^*)]. \end{aligned} \quad (32)$$

The covariance and mean are calculated from the distribution described by (20). Then, the total covariance law results in the following

$$\begin{aligned} &\text{Cov}[\mathbf{g}_i(\tilde{\mathcal{P}}^*) \mid (\mathcal{P}^*, \mathcal{D})] = \mathbb{E}_{\tilde{\mathcal{P}}^*}[k^i(\tilde{\mathcal{P}}^*, \tilde{\mathcal{P}}^*)] \\ &- \mathbb{E}_{\tilde{\mathcal{P}}^*}[\kappa^i(\tilde{\mathcal{P}}^*, \mathcal{D})^\top \big(\mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I + \mathscr{D}\big)^{-1}\kappa^i(\mathcal{D}, \tilde{\mathcal{P}}^*)] \\ &+ \mathbb{E}_{\tilde{\mathcal{P}}^*}[\bar{\mathbf{g}}_i(\tilde{\mathcal{P}}^*)\bar{\mathbf{g}}_i(\tilde{\mathcal{P}}^*)^\top] - \mathbb{E}_{\tilde{\mathcal{P}}^*}[\bar{\mathbf{g}}_i(\tilde{\mathcal{P}}^*)]\mathbb{E}_{\tilde{\mathcal{P}}^*}[\bar{\mathbf{g}}_i(\tilde{\mathcal{P}}^*)]^\top. \end{aligned} \quad (33)$$

After substituting the predictive mean in (33), the predictive covariance function for a given test point $\mathcal{P}^*$ is obtained as

$$\begin{aligned} &\text{Cov}[\mathbf{g}_i(\tilde{\mathcal{P}}^*) \mid (\mathcal{P}^*, \mathcal{D})] = \mathbb{E}_{\tilde{\mathcal{P}}^*}[k^i(\tilde{\mathcal{P}}^*, \tilde{\mathcal{P}}^*)] \\ &- \mathbb{E}_{\tilde{\mathcal{P}}^*}[\kappa^i(\tilde{\mathcal{P}}^*, \mathcal{D})^\top \big(\mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I_N + \mathscr{D}\big)^{-1}\kappa^i(\mathcal{D}, \tilde{\mathcal{P}}^*)] \\ &+ \mathbb{E}_{\tilde{\mathcal{P}}^*}[\kappa^i(\tilde{\mathcal{P}}^*, \mathcal{D})^\top \Bigg(\mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I_N + \mathscr{D}\Bigg)^{-1} \mathcal{Y}\mathcal{Y}^\top \\ &\Bigg(\mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I_N + \mathscr{D}\Bigg)^{-1} \kappa^i(\tilde{\mathcal{P}}^*, \mathcal{D})] \\ &- \mathbb{E}_{\tilde{\mathcal{P}}^*}[\kappa^i(\tilde{\mathcal{P}}^*, \mathcal{D})^\top \Bigg(\mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I_N + \mathscr{D}\Bigg)^{-1} \mathcal{Y}] \\ &\times \mathbb{E}_{\tilde{\mathcal{P}}^*}[\kappa^i(\tilde{\mathcal{P}}^*, \mathcal{D})^\top \Bigg(\mathcal{K}(\mathcal{D}, \mathcal{D}) + \sigma_e^2 I_N + \mathscr{D}\Bigg)^{-1} \mathcal{Y}]. \end{aligned} \quad (34)$$

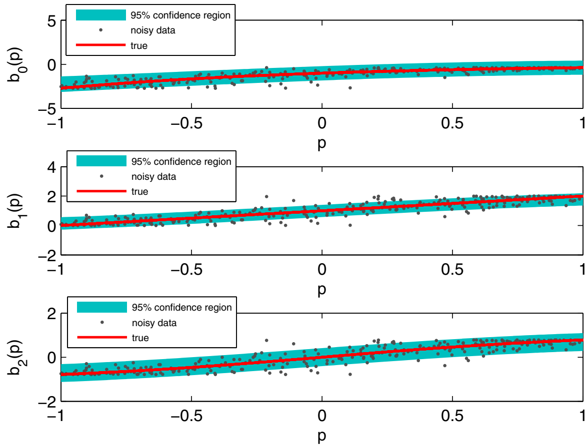

Fig. 1. The estimated covariance functions by employing the proposed method.

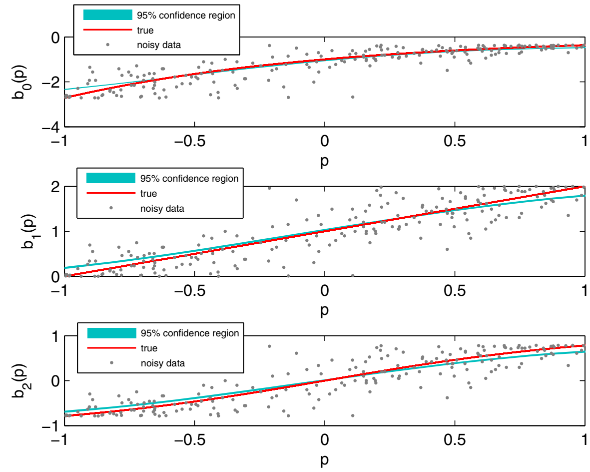

Fig. 2. The estimated covariance functions using the standard GP as elaborated in [15].

This can be rewritten in the following form

$$\begin{aligned}\mathrm{Cov}[\mathbf{g}_i(\mathcal{P}^*)] = \\ \lambda_i - \sum_{n=1}^{N}\sum_{m=1}^{N} \mathcal{S}_{mn}\mathbb{E}_{\tilde{\mathcal{P}}^*}[\kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D}_m)\kappa^i(\mathcal{D}_n,\tilde{\mathcal{P}}^*)] \\ + \mathcal{Y}^\top \mathcal{S}\mathbb{E}_{\tilde{\mathcal{P}}^*}[\kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D})\kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D})^\top]\mathcal{S}\mathcal{Y} - \bar{\mathbf{g}}_i(\mathcal{P}^*)^2, \quad (35)\end{aligned}$$

where

$$\mathcal{S} = \big(\mathcal{K}(\mathcal{D},\mathcal{D}) + \sigma_e^2 I_N + \mathcal{Q}\big)^{-1}, \quad (36)$$

and

$$\begin{aligned}&\mathbb{E}_{\tilde{\mathcal{P}}^*}[\kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D})\kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D})^\top] \\ &= \int_{-\infty}^{+\infty} \kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D})\kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D})^\top P(\tilde{\mathcal{P}}^*|\mathcal{P}^*,\Sigma_p)\mathrm{d}\tilde{\mathcal{P}}^*. \quad (37)\end{aligned}$$

The integration over the given distribution leads to the following expression for the corresponding elements of (37)

$$\begin{aligned}&\mathbb{E}_{\tilde{\mathcal{P}}^*}[\kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D})\kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D})^\top]_{m,n} = \\ &\mathbb{E}_{\tilde{\mathcal{P}}^*}[\kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D}_m)\kappa^i(\tilde{\mathcal{P}}^*,\mathcal{D}_n)] = \lambda_i \mid 2\mathcal{W}_i^{-1}\Sigma_p + I \mid^{-\frac{1}{2}} \\ &\quad x_i(m)k^i(\mathcal{P}^*,\mathcal{D}_m)x_i(n)k^i(\mathcal{P}^*,\mathcal{D}_n) \\ &\times \exp\Big(-(\mathcal{P}^* - \frac{\breve{p}_m + \breve{p}_n}{2})^\top(\mathcal{W}_i + \frac{1}{2}\mathcal{W}_i\Sigma_p^{-1}\mathcal{W}_i)^{-1} \\ &\qquad (\mathcal{P}^* - \frac{\breve{p}_m + \breve{p}_n}{2})\Big). \quad (38)\end{aligned}$$

Substituting (38) back into (35), the predictive covariance of the LPV model coefficients is obtained. Therefore, (31) together with (35) form the basis for *one-step ahead prediction* of the system output by the obtained predictive distribution over the given uncertain scheduling variables.

A. Example

An LPV system described by a finite impulse response (FIR) model and a nonlinear dynamic dependency on the scheduling variables is considered here. The model is described as

$$y(k) = \sum_{i=0}^{2} b_i(p_{k-i})u(k-i) + e_0(k), \quad (39)$$

$$b_0(p_k) = -\exp(-p_k), \quad b_1(p_{k-1}) = 1 + p_{k-1},$$
$$b_2(p_{k-2}) = \tan^{-1}(p_{k-2}),$$

where e_0 is a zero mean stochastic noise process with a Gaussian distribution $\mathcal{N}(0,\sigma^2)$, $\sigma = 0.05$. The scheduling variable is generated by $p_k = \sin(\frac{\pi}{30}k)$ and an additive noise ε_p with a Gaussian distribution $\mathcal{N}(0,\Sigma_p)$, $\Sigma_p = 0.1$ is simulated to corrupt the scheduling variable resulting in $\breve{p}_k = p_k + \varepsilon_p(k)$. A data set $\mathcal{D} = \{\breve{p}_k, y(k), u(k)\}_{k=1}^N$ with $N = 400$ snapshots is collected from the system (39) by considering a periodic input $u(k)$ as

$$u(k) = \begin{cases} 1 & \text{if } k = 1 \\ 0 & \text{if } k = 2,3. \end{cases} \quad (40)$$

As mentioned before, the training data set contains the noise corrupted scheduling variables and noisy output as the measurement data. The robustness of the proposed approach to the noise in variables is examined here and the results are compared to the standard GP. The hyperparameters including the RBF kernel parameters are obtained through the training process as $\mathcal{W}_1 = 1.17$, $\mathcal{W}_2 = 1.17$, $\mathcal{W}_3 = 1.17$ and $\lambda_1 = 1.2$, $\lambda_2 = 4.5$, $\lambda_3 = 5.6$; the output and scheduling variable's noise variance are also estimated as $\sigma_e = 0.048$ and $\Sigma_p = 0.98$, respectively. The estimated variance from previous step is used to slightly reduce the noise on the measured scheduling signal. Moreover, the evaluated covariance function for every coefficient using the proposed method and the standard GP are shown in Figures 1 and 2, respectively. As observed from the figures, the proposed approach estimates the LPV coefficients with a higher BFR in compared to the standard GP-based LPV model identification developed in [15]. It should be mentioned that to calculate the BFR, the estimated and true coefficient functions $\hat{\mathbf{g}}_i$ and $\mathbf{g}_i$ are used instead of the $\hat{y}$ and y in (28). The proposed approach

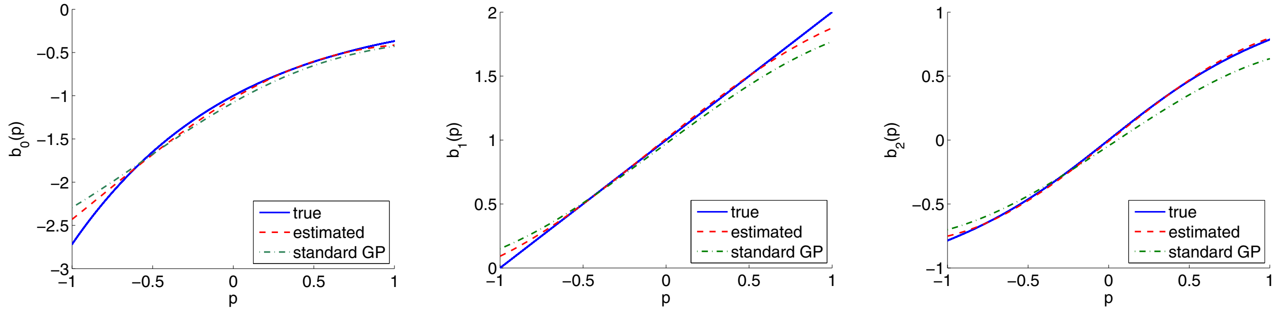

Fig. 3. Estimated coefficient functions by the proposed method and standard GP.

TABLE I
EXAMPLE: THE MSE AND BFR OF THE ESTIMATED LPV MODEL COEFFICIENT FUNCTIONS USING THE PROPOSED LPV IDENTIFICATION APPROACH AND THE ONE IN [15].

	Standard GP [15]		Proposed GP	
Coefficient	MSE	BFR	MSE	BFR
b_0	0.09	62.35%	0.0132	85.56%
b_1	0.0608	64.96%	0.0086	86.85%
b_2	0.0396	65.89%	0.0062	86.52%

offers promising results in estimating the uncertainty level by presenting a wider confidence region that contains the uncertain data, as well as the true coefficients, whereas the confidence regions obtained by the standard GP do not include the true coefficients in various sections of the plots. This error can be justified due to the inability of standard GP in adapting to the presence of uncertainty in the scheduling variables. Furthermore, the estimated coefficient functions using the two approaches are shown in Figure 3.

The *best fit ratio* (BFR) and *mean square error* (MSE) are used to quantify the estimated model accuracy by evaluating it for each coefficient function and the results are shown in Table I. The results illustrate that the proposed method in this paper can effectively provide an accurate estimation of the LPV model coefficients to cope with the uncertainty in the data.

V. CONCLUDING REMARKS

A new system identification approach for input-output LPV models is presented in this paper based on Gaussian Process (GP) to compensate for the errors in the scheduling variables. The proposed approach uses a linear approximation to capture the effect of scheduling variables noise on the evaluated coefficient functions on the observed scheduling variables. This leads to acquiring a better understanding of the uncertainties in data through more accurate formulation of the noise effect on the LPV model coefficients compared to the standard GP. The results indicate that the proposed method gives a more accurate estimation of the LPV model coefficient functions in the presence of both noisy measurement outputs and erroneous scheduling variables. The simulation results demonstrate that the proposed approach can effectively cope with uncertainties in the scheduling variables.

REFERENCES

[1] M. Hong, T. Söderström, and W. X. Zheng, "Accuracy analysis of bias eliminating least squares estimates for errors-in-variables systems," *Automatica*, vol. 43, pp. 1590-1596, 2007.

[2] T. Söderström, "Errors-in-variables methods in system identification," *Automatica*, vol. 43, pp. 939-958, 2007.

[3] W. X. Zheng, "Transfer function estimation from noisy input and output data," *Automatica*, vol. 12, pp. 365-380, 1998.

[4] M. Butcher and R. Longchamp, "On the consistency of certain identification methods for linear parameter varying systems," in *Proc. of the 17th IFAC world congress*, Seoul, South Korea, 2008, pp. 4018-4023.

[5] V. Cerone, D. Piga, and D. Regruto, "A convex relaxation approach to set membership identification of LPV systems," *Automatica*, vol. 49, pp. 2853-2859, 2013.

[6] D. Piga, P. Cox, R. Tóth, and V. Laurain, "LPV system identification under noise corrupted scheduling and output signal observations," *Automatica*, vol. 53, pp. 329-338, 2015.

[7] A. Chiuso and G. Pillonetto, "Bayesian and nonparametric methods for system identification and model selection," in *Proc. of the European Control Conference*, June 2014, pp. 2376-2381.

[8] B. Schölkopf and A. Smola, *Learning with kernels*, Cambridge MA: MIT Press, 2002.

[9] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. London, England: Massachusetts Institute of Technology, 2006, vol. 403.

[10] E. Solak, R. Murray-Smith, W.E. Leithead, D.J. Leith, and C.E. Rasmussen, "Derivative observations in Gaussian process models of dynamic systems," in *Proc. of the Neural Information Processing Systems Conference*, Dec. 2003, pp. 1033-1040.

[11] S. Rizvi, J. Mohammadpour, R. Tóth, and N. Meskin, "A Kernel-based Approach to MIMO LPV State-space Identification and Application to a Nonlinear Process System," in *Proc. of the 1st IFAC Workshop on Linear Parameter varying Systems*, Grenoble, France, 2015, to appear.

[12] A. Girard, C. E. Rasmussen, J. Quinonero-Candela, and R. Murray-Smith, "Gaussian process priors with uncertain inputs: Application to multiple-step ahead time series forecasting," in *Advances in Neural Information Processing Systems*, Cambridge, MA: MIT Press, vol. 15, pp. 171-194, 2003.

[13] P. Dallaire, C. Besse, and B. Chaib-draa, "Learning Gaussian process models from uncertain data," in *Proc. of the 16th International Conference on Neural Information Processing*, 2008, pp. 344-351.

[14] K. L. Chung, *A Course in Probability Theory*, San Diego CA: Academic Press, 2001.

[15] A. Golabi, N. Meskin, R. Tóth, and J. Mohammadpour, "A Bayesian approach for estimation of linear-regression LPV models," in *Proc. of the IEEE Conf. Decision and Control*, Dec. 2014, pp. 2555-2560.