AN APPLICATION OF GAUSSIAN PROCESS MODELS FOR CONTROL DESIGN

Kristjan Ažman [*,1] **Juš Kocijan** [*,**]

[*] *Jozef Stefan Institute, Ljubljana, Slovenia*
[**] *University of Nova Gorica, Nova Gorica, Slovenia*

Abstract:
This paper describes a method of modelling nonlinear dynamical systems from measurement data for control design purpose. The method merges linear local model blending approach with Bayesian Gaussian process modelling. Fixed-Structure Gaussian Process model can be interpreted as linear model structure with varying and probabilistic parameters, which are represented with Gaussian process models. It can be applied for extended local linear equivalence class of nonlinear systems. The obtained nonlinear system model can be used for control system design based on parametric process model. The modelling and control design will be illustrated with a simple example.

Keywords: Nonlinear systems identification, Gaussian Process models, velocity-based linearisation, nonlinear control

1. INTRODUCTION

A number of identification methods exist for modelling nonlinear dynamic systems from data. One possible approach to modelling are local model networks (LMN) which are attractive for so called *divide and conquer* control design (Murray-Smith and Johansen, 1997). In this approach global behaviour is represented with the network of simple local models where each local model describes some particular operating region and global behaviour is achieved by blending the dynamics of the local models. An important issue of this approach is the realisation of a model blending which has a strong impact on blended model transparency in off-equilibrium regions described in (Murray-Smith et al., 1999; Johansen et al., 2000; Leith and Leithead, 1999). LMN approaches, regardless of the blending realisation, also encounter the issue of scheduling vector selection. Scheduling vector is a vector defining the current region of operation and assists the blending mechanism to properly match the nonlinear dynamics.

A way to deal with the issue of model's off-equilibrium behaviour and retaining the transparency of system is representing the nonlinear system in the velocity based linearisation (VBL) form (Leith and Leithead, 1999). VBL in contrast to conventional series expansion approach enables the representation of system in every operating point. Nevertheless, the blending and scheduling mechanisms need to be determined for every nonlinear system. Prior knowledge about the system itself is used for determination of these two mechanisms whenever it is available.

There exist alternative approaches. It was shown in (Murray-Smith et al., 1999) that Gaussian process (GP) models can be effectively used to deal with modelling of dynamics in off-equilibrium

[1] The authors gratefully acknowledge the contribution of the Ministry of Higher Education, Science and Technology of Republic of Slovenia, Grant No. P2-0001.

regions even when the data there is relatively sparse.

GP models can be used for modelling of dynamic systems in various ways, e.g. as regression model describing nonlinear system (Kocijan et al., 2005), regression model with incorporated prior knowledge of linear local models (Solak et al., 2002; Leith et al., 2002; Ažman and Kocijan, 2005) or GP local model network (Gregorčič and Lightbody, 2005). However, since the GP model is a probabilistic nonparametric model, the only possible control approach with such model available is predictive control.

Another application of a GP modelling is its use for the identification of a nonlinear structure (Leithead et al., 2003) important for determination of scheduling variables.

In this paper an approach is proposed where an unknown nonlinear system is modelled using a linear model with varying parameters. These parameters are predicted using the GP models according to the current operating region. GP models are useful for modelling for several reasons.

- They tend to achieve acceptable modelling results even with relatively small training data sets.
- Beside predicting the output the GP model also gives the measure of confidence in prediction dependant on the training data density.
- When local models are blended with the GP models as the local models' parameters these GP models incorporate also the information about the dependance of parameters on individual regressors.

The resulting model, named Fixed-Structure Gaussian process (FSGP) model, can be seen as a parametric model, namely a fixed linear model structure with GP models representing varying parameters. This kind of nonlinear system model offers possibility for application of various control design methods.

The paper is organized as follows. In Section 2, the Gaussian process model is briefly reviewed. In Section 3, the Fixed-Structure Gaussian Process model is introduced and in Section 4 a possible control design is discussed. An illustrating example is given in Section 5 and the conclusions are given in Section 6.

2. GAUSSIAN PROCESS MODEL

A detailed presentation of Gaussian processes can be found e.g. in (Rasmusen and Williams, 2006). A Gaussian process is a random function fully characterized by its mean and covariance functions. For simplicity, we assume a zero-mean process. Given $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, the corresponding $f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)$ can be viewed as a collection of random variables which have a joint multivariate Gaussian distribution: $f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n) \sim \mathcal{N}(0, \mathbf{\Sigma})$, where Σ_{pq} gives the covariance between $f(\mathbf{x}_p)$ and $f(\mathbf{x}_q)$ and is a function of the corresponding $\mathbf{x}_p$ and $\mathbf{x}_q$: $\Sigma_{pq} = C(\mathbf{x}_p, \mathbf{x}_q)$. The covariance function $C(.,.)$ can be of any kind, provided that it generates a positive definite covariance matrix $\mathbf{\Sigma}$. The Gaussian Process model fits naturally in the Bayesian modelling framework, as it places a prior directly over functions, instead of parameterizing $f(\mathbf{x})$. In the following, we assume a stationary process, where the stationarity assumption implies that the covariance between two points depends only on the distance between them and is invariant by translation in the input space. A common choice of covariance function is the squared exponential or Gaussian one:

$$\begin{aligned}\mathrm{Cov}[f(\mathbf{x}_p), f(\mathbf{x}_q)] \;&=\; C(\mathbf{x}_p, \mathbf{x}_q) \;= \\ &= v_1 \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d (x_p^d - x_q^d)^2\right] + v_0 \delta_{pq}\end{aligned}$$

where x_p^d denotes the d^{th} component of the D-dimensional input vector $\mathbf{x}_p$, and $v_1, v_0, w_1, \ldots, w_D$ are free parameters and δ_{pq} is Kronecker operator. This covariance function is such that points close together in the input space lead to more correlated outputs than points further apart (a smoothness assumption). The parameter v_1 controls the vertical scale of variation, v_0 is noise variance and the w_d's are inversely proportional to the horizontal length-scale in dimension d.

Let the input/target relationship be $y = f(\mathbf{x}) + \epsilon$. We assume an additive white noise with variance v_0, $\epsilon \sim \mathcal{N}(0, v_0)$, and put a GP prior on $f(.)$, with covariance function (1) and unknown parameters. Within this probabilistic framework, we can write $y_1, \ldots, y_{N+1} \sim \mathcal{N}(0, \mathbf{K}_{N+1})$, with $K_{N+1_{pq}} = \Sigma_{pq}$. If we split $y_1, \ldots, y_{N+1}$ into two parts, $\mathbf{y} = [y_1, \ldots, y_N]$ and y^*, we can write

$$\mathbf{y}, y^* \sim \mathcal{N}(0, \mathbf{K}_{N+1}) \tag{1}$$

with

$$\mathbf{K}_{N+1} = \begin{bmatrix} \begin{bmatrix} & \\ & \mathbf{K} \\ & \end{bmatrix} & \begin{bmatrix} \\ \mathbf{k}(\mathbf{x}^*) \\ \\ \end{bmatrix} \\ \begin{bmatrix} \mathbf{k}(\mathbf{x}^*)^T \end{bmatrix} & \begin{bmatrix} \kappa(\mathbf{x}^*) \end{bmatrix} \end{bmatrix} \tag{2}$$

where $\mathbf{K}$ is an $N \times N$ matrix giving the covariances between y_p and y_q, for $p, q = 1 \ldots N$, $\mathbf{k}(\mathbf{x}^*)$ is an $N \times 1$ vector giving the covariances between y^* and y_p ($k_p(\mathbf{x}^*) = C(\mathbf{x}_p, \mathbf{x}^*)$, for $p = 1 \ldots N$), and $\kappa(\mathbf{x}^*) = C(\mathbf{x}^*, \mathbf{x}^*)$ is the covariance between the test output and itself.

For our modelling purposes, we can then divide this joint probability into a marginal and a conditional part. Given a set of N training data pairs, $\{\mathbf{x}_p, y_p\}_{p=1}^N$, the marginal term gives us the likelihood of the observed data: $\mathbf{y}|\mathbf{X} \sim \mathcal{N}(0, \mathbf{K})$, where $\mathbf{y}$ is the $N \times 1$ vector of training targets and $\mathbf{X}$ the $N \times D$ matrix of the corresponding training inputs. We can then estimate the unknown parameters of the covariance function, as well as the noise variance v_0, via maximisation of the log-likelihood. The conditional part of (1) provides us with the predictive distribution of y^* corresponding to a new given input $\mathbf{x}^*$. We only need to condition the joint distribution on the training data and the new input $\mathbf{x}^*$, $p(y^*|\mathbf{y}, \mathbf{X}, \mathbf{x}^*) = \frac{p(\mathbf{y}, y^*)}{p(\mathbf{y}|\mathbf{X})}$. It can be shown that this distribution is Gaussian with mean and variance

$$\begin{aligned} \mu(\mathbf{x}^*) &= \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{y} \\ \sigma^2(\mathbf{x}^*) &= \kappa(\mathbf{x}^*) - \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}^*) \end{aligned} \tag{3}$$

This way, we can use the predictive mean $\mu(\mathbf{x}^*)$ as an estimate for y^* and the predictive variance or standard deviation $\sigma(\mathbf{x}^*)$ as the uncertainty or measure of confidence attached to the estimate.

3. MODELLING OF THE SYSTEM

In this paper the system is assumed to be discrete with sampling time T, represented as a pseudo-continuous one with delayed signals. Such system has same behaviour as the discrete one in the points of sampling.

Consider the delayed nonlinear system

$$\begin{aligned} \mathbf{x}(t+T) &= \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t)), \\ \mathbf{y}(t) &= \mathbf{G}(\mathbf{x}(t), \mathbf{u}(t)), \end{aligned} \tag{4}$$

which may be reformulated, without loss of generality, in form denoted as extended local linear equivalence (ELLE)(Leith and Leithead, 1999)

$$\begin{aligned} \mathbf{x}(t+T) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{f}(\boldsymbol{\rho}), \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) + \mathbf{g}(\boldsymbol{\rho}), \end{aligned} \tag{5}$$

where $\mathbf{x}(t) \in \mathcal{R}^n$, $\mathbf{u}(t) \in \mathcal{R}^m$ and $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are appropriately dimensioned constant matrices, $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are nonlinear functions and $\boldsymbol{\rho} = \boldsymbol{\rho}(\mathbf{x}(t), \mathbf{u}(t)) \in \mathcal{R}^q, q \leq m + n$, embodies the nonlinear dependence of the dynamics on the state and input with $\nabla_x \boldsymbol{\rho}$, $\nabla_u \boldsymbol{\rho}$ constant (Leith and Leithead, 1999). Differentiating (5) an alternative representation of the nonlinear system is

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{w}(t), \\ \mathbf{w}(t+T) &= \mathbf{A}(\boldsymbol{\rho})\mathbf{w}(t) + \mathbf{B}(\boldsymbol{\rho})\dot{\mathbf{u}}(t), \\ \dot{\mathbf{y}}(t) &= \mathbf{C}(\boldsymbol{\rho})\mathbf{w}(t) + \mathbf{D}(\boldsymbol{\rho})\dot{\mathbf{u}}(t). \end{aligned} \tag{6}$$

where

$$\begin{aligned} &\mathbf{A}(\boldsymbol{\rho}) = \mathbf{A} + \nabla\mathbf{f}(\boldsymbol{\rho})\nabla_{\mathbf{x}}\boldsymbol{\rho}, \quad \mathbf{B}(\boldsymbol{\rho}) = \mathbf{B} + \nabla\mathbf{f}(\boldsymbol{\rho})\nabla_{\mathbf{u}}\boldsymbol{\rho}, \\ &\mathbf{C}(\boldsymbol{\rho}) = \mathbf{C} + \nabla\mathbf{g}(\boldsymbol{\rho})\nabla_{\mathbf{x}}\boldsymbol{\rho}, \quad \mathbf{D}(\boldsymbol{\rho}) = \mathbf{D} + \nabla\mathbf{g}(\boldsymbol{\rho})\nabla_{\mathbf{u}}\boldsymbol{\rho}, \end{aligned} \tag{7}$$

These are approximated with $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}$:

$$\boldsymbol{\alpha} \approx \mathbf{A}(\boldsymbol{\rho}) = \nabla_{\mathbf{x}}\mathbf{x}(t+T) \tag{8}$$

$$\boldsymbol{\beta} \approx \mathbf{B}(\boldsymbol{\rho}) = \nabla_{\mathbf{u}}\mathbf{x}(t+T) \tag{9}$$

$$\boldsymbol{\gamma} \approx \mathbf{C}(\boldsymbol{\rho}) = \nabla_{\mathbf{x}}\mathbf{y}(t) \tag{10}$$

$$\boldsymbol{\delta} \approx \mathbf{D}(\boldsymbol{\rho}) = \nabla_{\mathbf{u}}\mathbf{y}(t) \tag{11}$$

Each of the $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}$ elements is approximated with its own Gaussian Processes model. The training data for the GP models are coefficients of individual linearisations obtained from local linear models identified in various equilibrium as well as off-equilibrium regions.

This paper will not focus on details of how the local models are obtained. More on the identification of the local models and the associated issues can be found e.g. in (Murray-Smith and Johansen, 1997), (Murray-Smith et al., 1999) and (Johansen et al., 2000) and references in there. Nevertheless, the identified linear local models need to be of the same order, must describe corresponding region satisfactorily well and must be located in equilibrium as well as off-equilibrium points. The off-equilibrium models are necessary for two reasons:

- they uniquely define the system (Leith and Leithead, 2002) and
- they provide the data necessary for training the GP models describing entire operating region.

It is important to note that a local linear input-output model only specifies parameters up to a co-ordinate transformation (Leith and Leithead, 2002). In this paper we always use lagged inputs and outputs as our state co-ordinates for simplicity but of course other choices are possible.

The FSGP system modelling consist roughly of two stages. The first stage is the identification of local linear models in equilibrium and off-equilibrium points. The results of the first stage are sets of derivatives of nonlinear function which are at the same time coefficients of linear local models.

In the second stage the GP models are learned for each set of identified coefficients or parameters. Each set has a functional dependency on the states and the inputs. GP model can be used as a relevance detector and via GP model training necessary regressors, namely the states and the

inputs, to which the parameters are functionally linked are revealed.

As the local models usually can not be identified for every point of the operating region the smoothing property of GP models can be used to acquire the values of the local model parameters for the points lying between the points where linear local models were identified. This blending of the parameters is in our case realised with modelling each local model parameter with its own GP model. The input into GP models is the vector of regressors, which can be interpreted as the vector of scheduling variables $\boldsymbol{\rho}(t)$. The output of each GP model is a corresponding most likely value of the parameter θ_k and associated variance var θ_k, where k denotes k-th parameter from vector of parameters $\boldsymbol{\theta} = \left[[\alpha_{ij}]^T [\beta_{ij}]^T [\gamma_{ij}]^T [\delta_{ij}]^T \right]^T$. Variances $\mathbf{var}\ \boldsymbol{\theta}^t$ are expressing the confidence in the predicted values of parameters, which depends on the amount of information available for the modelling. The better the region is modelled – more local models are used – the lower is the variance.

The realisation of the model is similar to that in the velocity based linearisation approach (Leith and Leithead, 1999), where second order time derivative $\ddot{\mathbf{x}}(t)$ is replaced with first order time derivative of delayed $\dot{\mathbf{x}}(t+T)$ like in equation (6).

The obtained nonlinear model can be viewed as a parametric model with probabilistic and variable parameters $\boldsymbol{\theta}(\boldsymbol{\rho}(t))$ – FSGP model. Each parameter depends on the vector of scheduling variables $\boldsymbol{\rho}(t)$, which consists of all states and inputs or their subset.

4. CONTROL DESIGN

What are the benefits of a FSGP model? Number of nonlinear identification methods including GP model identification methods provide models that can be used only with a model based predictive control. FSGP model is a parametric model with probabilistic parameters and can be used for wider range of control design methods but model based predictive control.

One possible control design approach is a gain-scheduling control design. In this case the local controllers are designed for selected local models of the process.

Selection of the process local models for control design depends on the region where closed-loop dynamics is expected and is in general not the same as the selection of the local models used for process modelling. It is sensible to keep the system in the well modelled region, i.e. where the variances of the local models' parameters predictions are small. Parameters of the local controllers depend on the same scheduling variables as the associated process local model parameters.

The basic idea of modelling and control design is illustrated in the next simple example.

5. EXAMPLE

In this section a first order nonlinear system:

$$y(t+T) = \frac{y(t)}{1+y^2(t)} + u^3(t), \tag{12}$$

in discrete points identical to (Narendra and Parthasarathy, 1990) will be used to illustrate modelling a FSGP model convenient for a control design.

The operating region of the system, defined with $0 < u < 2.35$ and corresponding values of y in equilibrium $0 < y < 15.7$, is divided in two regions for better illustration of the FSGP model properties:

- well modelled region defined with $1.25 < u < 2.35$ and corresponding $2.3 < y < 13$ in equilibrium and
- badly modelled region, which scoops the rest of the working region.

Our aim is to model this system with FSGP model i.e. to model the derivatives of the system $\frac{\partial y(t+T)}{\partial y(t)}$ and $\frac{\partial y(t+T)}{\partial u(t)}$ with two GP models and use this model first for multi-step ahead prediction (simulation) and then control design.

5.1 Modelling

5.1.1. Modelling local model parameters with the GP model The parameters of the local models describing the system (12)

$$\boldsymbol{\theta}(\boldsymbol{\rho}) = [a(\boldsymbol{\rho})\ b(\boldsymbol{\rho})] \tag{13}$$

$$a(\boldsymbol{\rho}) = \frac{\partial y(t+T)}{\partial y(t)} = \frac{1-y^2(t)}{(1+y^2(t))^2} \tag{14}$$

$$b(\boldsymbol{\rho}) = \frac{\partial y(t+T)}{\partial u(t)} = 3u^2(t) \tag{15}$$

depend on scheduling vector $\boldsymbol{\rho} = [y(t)\ u(t)]^T$, but due to the simplicity of the system the parameter a depends only on $y(t)$ and the parameter b only on $u(t)$. This simplifies the modelling of each parameter as there is no distinction between equilibrium and off-equilibrium points in representation of the training data for the GP model.

Training points for the two GP models describing the parameters a and b are in general obtained

with the identification of linear local models. Any algorithm for identification can be used for example instrumental variable (IV) algorithm. The true values of parameters a and b and corresponding predictions $\tilde{a}$ and $\tilde{b}$ together with training points can be seen in Figure 1. We can distinguish between well and badly modelled regions by observing the error and the variance of the prediction in Figure 1.

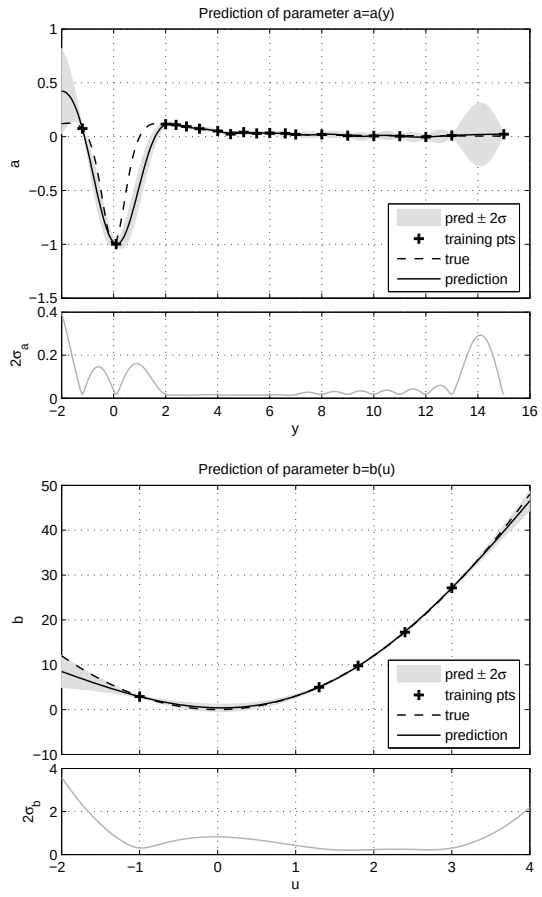

Fig. 1. Comparison of the Gaussian process model output and true value of the system's parameter for parameter a (upper figure) and parameter b (lower figure)

5.1.2. Putting the local models together — blending the model System described with equation (12) can be written in a form similar to velocity based linearisation as in equation (6).

The approximations of the parameters $\tilde{a}(k)$ and $\tilde{b}(k)$ are predicted using corresponding GP models.

The resulting model was first validated. The response of the system and the FSGP model to a random input signal is depicted in Figure 2. Standard deviations associated with each of the predicted parameters are given in Figure 3.

The crossing between well and badly modelled regions can be clearly seen from simulation result

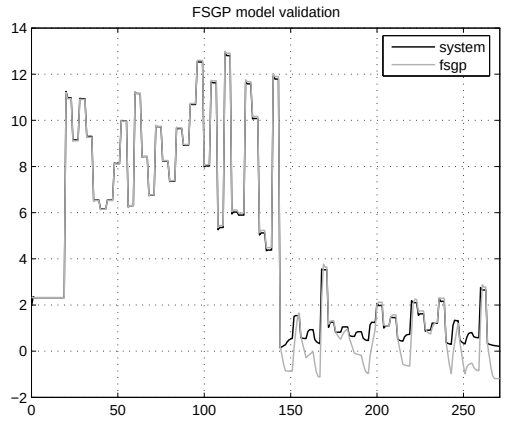

Fig. 2. Responses of dynamic system and its model

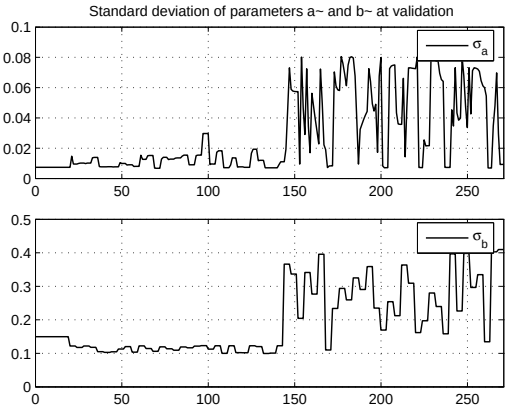

Fig. 3. Standard deviations of parameter predictions $\tilde{a}$ (upper figure) and parameter predictions $\tilde{b}$ (lower figure)

depicted in Figure 2 as well as from the values of parameters' standard deviation in Figure 3.

5.2 Control design

A controller that compensates the process dynamics is designed. Since the process is of the first order a nonlinear proportional integral controller is selected. Mean values of predictions $\tilde{a}$ and $\tilde{b}$ were used in the controller design. The targeted closed-loop dynamics was of the first order with predetermined time constant, which is to be uniform in the entire operating region. The obtained closed-loop response is given in Figure 4.

The closed-loop performance can be assessed with the comparison of obtained response to the specified reference closed-loop response. It can be seen from Figure 4 that closed-loop response does not follow the closed-loop requirements well, when y decreases below value 2, where the GP models of linearised parameters exhibit low confidence.

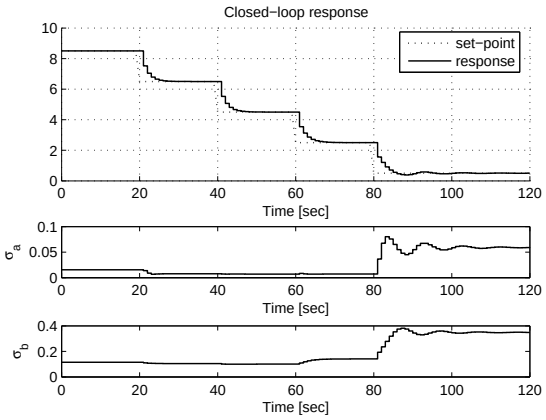

Fig. 4. Closed-loop response of the system controlled by compensator

6. CONCLUSIONS

The paper presents the idea of the Fixed-Structure Gaussian Process (FSGP) model and its application in the control design. FSGP model can be interpreted as a linear model structure with varying and probabilistic parameters, where GP models are used to model each of the varying parameters. The model is applicable for the extended local linear equivalence class of nonlinear models. The modelling and the control of such model were briefly presented and shown on a simple example.

The contributions of this paper is a probabilistic approach to model nonlinear system for a control design based on the parametric models. The advantages of presented approach can be as follows:

- FSGP model provides the measure of confidence in the predicted parameters based on the training data. Together with information of the parameter sensitivity the measure of confidence can be useful information for closed-loop system operation.
- FSGP model works well with relatively small but reasonably selected number of linear local models, which form training data for the GP models of the model parameters.
- The selection of scheduling variables, i.e. inputs of the GP models, is based on the relevance detection capability of GP models.
- The mechanisms for blending and scheduling of local models are joined together.
- FSGP model enables a use of some control design methods based on the parametric process models.

REFERENCES

Ažman, K. and J. Kocijan (2005). Comprising Prior Knowledge in Dynamic Gaussian Process Models. In *Proc.: International Conference on Computer Systems and Technologies – CompSysTech 2005* (B. Rachev and A. Smrikarov (Ed)), IIIB.2. Varna, Bulgaria.

Gregorčič G. and G. Lightbody (2005). Gaussian process approaches to nonlinear modelling for control, In: *Intelligent Control Systems using Computational Intelligence Techniyues*, IEE Control Series 70, IEE, London.

Johansen, T.A., R. Shorten and R. Murray-Smith (2000). On the Interpretation and Identification of Dynamic Takagi-Sugeno Fuzzy Models. *IEEE Transactions on Fuzzy Systems,* **8**, 297–313.

Kocijan J., A. Girard, B. Banko and R. Murray-Smith (2005). Dynamic Systems Identification with Gaussian Processes, *Mathematical and Computer Modelling of Dynamical Systems*, **11**, 411–424.

Leith, D.J. and W.E. Leithead (1999). Analytic framework for Blended Multiple Model Systems Using Linear Local Models. *Int. J. Control,* **72**, 605–619.

Leith, D.J. and W.E. Leithead (2002). Global reconstruction of Nonlinear Systems From Families of Linear Systems. In: *Proceedings of IFAC World Congress*, Barcelona.

Leith D.J., W.E. Leithead, E. Solak, R. Murray-Smith (2002). Divide & conquer identification using Gaussian process priors. In: *Proceedings of the 41st IEEE Conference on Decision and Control*, Las Vegas, Nevada, 624–629.

Leithead, W.E., E. Solak and D.J. Leith (2003). Direct Identification of Nonlinear Structure Using Gaussian Process Prior Models. In: *Proc.: European Control Conference ECC 2003*, Cambridge.

Murray-Smith, R., T.A. Johansen and R. Shorten (1999). On transient dynamics, off-equlibrium behaviour and identification in blended multiple model structures. In: *Proc.: ECC European Control Conference*, BA–14. Karsruhe, Germany.

Murray-Smith R. and T.A. Johansen (Ed) (1997). *Multiple Model approaches to Modelling and Control*. Taylor and Francis, London.

Narendra, K.S. and K. Parthasarathy (1990). Identification and Control of Dynamical Systems Using Neural Networks, *IEEE Transactions on NN*, **1**, 4–27.

Rasmusen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA.

Solak, E., R. Murray-Smith, W.E. Leithead, D.J. Leith and C.E. Rasmussen (2002). Derivative observations in Gaussian Process models of dynamic systems, In: *Advances in Neural Information processing Systems* (S. Becker, S. Thrun and K. Obermayer (Ed)), **15**, 529–536. MIT Press, Cambridge, MA.