Dynamical systems identification using Gaussian processes models with incorporated local models

K. Ažman[a,*], J. Kocijan[a,b]

[a] *Jožef Stefan Institute, Department of System and Control,*
Jamova 39, 1000 Ljubljana, Slovenia
[b] *University of Nova Gorica, Nova Gorica, Slovenia*

Abstract

Gaussian process (GP) models form an emerging methodology for modelling nonlinear dynamic systems which tries to overcome certain limitations inherent to traditional methods such as e.g. neural networks (ANN) or local model networks (LMN).

The GP model seems promising for three reasons. First, less training parameters are needed to parameterize the model. Second, the variance of the model's output depending on data positioning is obtained. Third, prior knowledge e.g. in the form of linear local models can be included into the model. In this paper the focus is on GP with incorporated local models as the approach which could replace local models network.

Much of the effort up to now has been spent on the development of the methodology of the GP model with included local models, while no application and practical validation has yet been carried out. The aim of this paper is therefore twofold. The first aim is to present the methodology of the GP model identification with emphasis on the inclusion of the prior knowledge in the form of linear local models. The second aim is to demonstrate practically the use of the method on two higher order dynamical systems, one based on simulation and one based on measurement data.

1. Introduction

Identification of nonlinear dynamical systems from measured data has received significant attention in the last two decades. Artificial neural networks (ANN) and fuzzy logic models can be viewed as universal approximators. The main practical disadvantages of these methods are (Johansen et al. (2000)):

- the lack of transparency (model structure does not reflect the physical properties of the system) and

[*] Corresponding author. tel: +386 1 477 3637, fax: +386 1 477 3994
Email addresses: kristjan.azman@ijs.si (K. Ažman), jus.kocijan@ijs.si (J. Kocijan)

- the curse of dimensionality.

Rather than struggling with a global model of the system one can employ a network of (simple) local models wherein each model describes some particular operating region. Such an approach is referred to as the Local Models Network (LMN, Murray-Smith and Johansen (1997)). The global description of the process is then obtained by blending the responses of the local models.† The relative importance of each local model is expressed through the values of validity functions and depends on the current system's state in the form of scheduling variables – usually a subset of model's regressors.

The number of unknown parameters in such an LMN is typically smaller than in an ANN with comparable quality of fit. However, some of the inconveniences still related to LMN are (Gregorčič and Lightbody (2003b); Murray-Smith et al. (1999); Johansen et al. (2000)):

- the problem of describing off-equilibrium dynamics,
- global/local optimization issue and
- scheduling vector selection.

The problem of modelling off-equilibrium dynamics (Johansen et al. (2000); Murray-Smith et al. (1999); Gregorčič and Lightbody (2003b)) with local models originates in a system's "tendency" towards equilibrium. As a consequence there is usually a lack of measured data in the regions away from equilibrium, which makes the construction of valid local models for those regions rather difficult. This problem can be highlighted e.g. in process industries, where a lot of data can be taken only in particular operating regions of the system so that only those regions can be satisfactorily modelled with local models. For the rest of the tentative operating regions the lack of data prevents the construction of valid local models.

Two approaches to LMN parameters optimisation are possible (Gregorčič and Lightbody (2003b)):

- global learning approach, where all the LMN parameters are optimised using all identification data in a single regression operation and
- local learning approach, where the parameters for each local model representing corresponding subspace are estimated separately using corresponding data.

LMN optimisation with global learning approach usually provides globally better fit as the local model parameters in off-equilibrium regions are used to increase the level of validity of associated local models, but these parameters no longer represent the system's local dynamics (Murray-Smith et al. (1999)).

†Another possible approach is to blend the parameters of the local models instead, see e.g Gregorčič and Lightbody (2003b), but here we will resort to LMN with blended outputs.

When in contrast a local learning approach is used, the local models' parameters do represent local dynamical behaviour which results in more transparent local models. Such models are more applicable to use in analysis and control design. Their drawback is that they are valid in smaller operating regions, which results in non-modelled regions of the system, leading again to the problem of describing off-equilibrium dynamics.

The problem of smaller validity regions can be eluded using a parsimonious scheduling vector, see e.g. Johansen et al. (2000). This extends the validity of the local models but the result of blending two local models far away from equilibrium can be non-smooth or even discontinuous (Gregorčič and Lightbody (2003b)), possibly leading to stability problems in control applications (Gregorčič and Lightbody (2003b)). Also, transparency of individual local models is lost. More on the LMN approaches to system identification can be found e.g. in Murray-Smith and Johansen (1997) and more on the problems accompanying this approach in Johansen et al. (2000), Murray-Smith et al. (1999) and Gregorčič and Lightbody (2003b).

* * * * *

In this paper the focus will be on Gaussian process models (Williams (1998); Gibbs (1997); MacKay (2003); Rasmussen (1996)), an emerging branch of black-box system identification aimed at alleviating the need for models with large parametrization, e.g. ANN.

In order to avoid the problems above, GP models rely on an entirely different approach. The GP model (Williams (1998)) smooths the information given as the training data. The output is predicted by weighting targets with respect to distance between training and test input.

The result of prediction for any given vector input[‡] is a Gaussian probability distribution of the output, expressed by its mean and variance. This is a result of assuming prior knowledge on the distribution of individual model predictions. It means that the most probable value of the output is given as the mean value but with smaller possibilities the value of the output can be on either side of the mean value, with variance describing this output probability distribution. This assumption is not valid generally, but it can be applied often as an approximation of the usually unknown distribution of individual model predictions. In this context the variance is the measure of confidence in the predicted mean value of the output. A higher variance value represents smaller confidence in the predicted output and vice versa. In the GP model this confidence measure depends not only on the variance of the parameters, as e.g. in ANN, but also on the position of training data in the input space.

Optimization of the GP model's parameters is the most time consuming part of identification. Although the number of hyperparameters that need to be optimized in the GP model is small compared to ANN, the computational burden

[‡] Input here means the values of corresponding input regressors, i.e. data points not signals.

associated with optimization rises fast with the number of training points (Gibbs (1997)). Much of the computational burden can be saved with the introduction of local models in the GP model. A local model, typically parameterized with only few parameters, can successfully describe a subset of training points reflecting the local dynamics of the system (Solak et al. (2003)). Thus introduction of local models into the GP model can result in reduced computational burden associated with the optimization of hyperparameters. Some other methods to reduce computational burden are discussed in Kocijan et al. (2005).

When comparing the GP model with incorporated local models to LMN several benefits of the GP model are noticed. Some typical LMN problems, e.g. off-equilibrium dynamics, global/local optimization issue and scheduling vector selection, are avoided and also confidence in the model's prediction is given using the GP model.

Most of the research related to dynamic systems identification with the GP model with incorporated LM so far has been focused on algorithm synthesis. Surprisingly no application to plant data obtained by measurements has been reported so far. The examples used to illustrate the performance are restricted to the simulated first order dynamic systems (Solak et al. (2003); Leith et al. (2002); Kocijan and Leith (2004)) and there are no published results on identification of higher order dynamic systems.

The purpose of this paper is twofold; firstly, to review the methodology of the dynamic systems identification with the GP model with incorporated local models in a consistent manner and to illustrate it with an example. Secondly, the first attempt to apply the methodology on a real process is described. The encouraging results are believed to promote further applications.

The paper is organized as follows. In the Section 2 the fundamentals of GP modelling are presented. The problem is stated and dynamic system identification is sketched in Section 3, where also linear local models are introduced and the evolution of the GP model with incorporated local models is described. The identification method is first demonstrated on a simulated second order nonlinear dynamic system example in Section 4. Finally, in Section 5 the identification method is applied on measurement data when identifying the laboratory pilot plant. The last section emphasises the main results and concludes the paper.

2. Modelling with Gaussian processes

The Gaussian process is a Gaussian random function, fully described by its mean and variance. Gaussian processes can be viewed as a collection of random variables $f(\mathbf{x}_i)$ with joint multivariate Gaussian distribution: $f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n) \sim \mathcal{N}(0, \mathbf{\Sigma})$. Elements Σ_{ij} of covariance matrix $\mathbf{\Sigma}$ are covariances between values of the function $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ and are functions of corresponding arguments $\mathbf{x}_i$ and $\mathbf{x}_j$: $\Sigma_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$. Any function $C(\mathbf{x}_i, \mathbf{x}_j)$ can be a covariance function, providing it generates a nonnegative definitive covariance matrix $\mathbf{\Sigma}$.

Certain assumptions about the process are made implicitly with the covariance function selection. The *stationarity of the process* results in the value of covariance function $C(\mathbf{x}_i, \mathbf{x}_j)$ between inputs $\mathbf{x}_i$ and $\mathbf{x}_j$ depending only on their distance and being invariant to their translation in the input space, see e.g. MacKay (2003). *Smoothness of the output* reflects in outputs $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ having higher covariance when inputs $\mathbf{x}_i$ and $\mathbf{x}_j$ are closer together. The common choice (e.g. Williams (1998); Neal (1996); Kocijan et al. (2005)) for covariance function, representing these assumptions, is:

$$C(\mathbf{x}_i, \mathbf{x}_j) \;=\; \mathrm{Cov}[f(\mathbf{x}_i), f(\mathbf{x}_j)] \;=\; v \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d (x_i^d - x_j^d)^2\right] \tag{1}$$

where D is a length of vector $\mathbf{x}$ and $\boldsymbol{\Theta} = [w_1, \ldots, w_D, v]^T$ is a vector of parameters called hyperparameters.[§] Hyperparameter v controls the magnitude of the covariance and hyperparameters w_d represent the relative importance of each component x^d of vector $\mathbf{x}$.

The GP model fits nicely into the Bayesian modelling framework. The idea behind GP modelling is to place the prior directly over the space of functions instead of parameterizing the unknown function $f(\mathbf{x})$ (MacKay (2003)). The simplest type of such prior is Gaussian. Consider the system

$$y(k) = f(\mathbf{x}(k)) + \epsilon(k) \tag{2}$$

with white Gaussian noise $\epsilon(k) \sim \mathcal{N}(0, v_0)$ with variance v_0 and the vector of regressors $\mathbf{x}(k)$ from operating space $\mathcal{R}^D$. The GP prior with covariance function (1) with unknown hyperparameters is put on the space of functions $f(.)$. Within this framework we have $y_1, \ldots, y_N \sim \mathcal{N}(0, \mathbf{K})$ with $\mathbf{K} = \boldsymbol{\Sigma} + v_0\mathbf{I}$, where $\mathbf{I}$ is $N \times N$ identity matrix.[¶]

Our task is to find the predictive distribution of y_{N+1} corresponding to a new given input $\mathbf{x}_{N+1}$ based on a set of N training data pairs $\{\mathbf{x}_i, y_i\}_{i=1}^N$. For the collection of random variables $(y_1, \ldots, y_N, y_{N+1})$ it can be written:

$$\begin{pmatrix} \mathbf{y} \\ y_{N+1} \end{pmatrix} \sim \mathcal{N}(0, \mathbf{K}_{N+1}) \tag{3}$$

with covariance matrix

$$\mathbf{K}_{N+1} = \begin{bmatrix} \begin{bmatrix} & \\ & \mathbf{K} & \\ & \end{bmatrix} & \begin{bmatrix} \\ \mathbf{k}(\mathbf{x}_{N+1}) \\ \\ \end{bmatrix} \\ \\ \begin{bmatrix}\mathbf{k}(\mathbf{x}_{N+1})^T\end{bmatrix} & \begin{bmatrix}k(\mathbf{x}_{N+1})\end{bmatrix} \end{bmatrix} \tag{4}$$

[§] The parameters of a Gaussian process are called hyperparameters due to their close relationship to the hyperparameters of a neural network (Gibbs (1997)).

[¶] When assuming different kind of noise the covariance function should be changed appropriately, e.g. Gibbs (1997).

where $\mathbf{y} = [y_1, \ldots, y_n]^T$ is an $N \times 1$ vector of training targets. This joint probability can be divided into a marginal and a conditional part. The marginal term gives us the likelihood of the training data: $\mathbf{y}|\mathbf{X} \sim \mathcal{N}(0, \mathbf{K})$, where $\mathbf{X}$ is the $N \times D$ matrix of training inputs.

We need to estimate the unknown hyperparameters $v, w_1 \ldots w_n$ of the covariance function (1), as well as the noise variance v_0. This is done via maximization of the log-likelihood

$$\begin{aligned} \mathcal{L}(\mathbf{\Theta}) &= \log(p(\mathbf{y}|\mathbf{X})) = \\ &= -\frac{1}{2}\log(\mid \mathbf{K} \mid) - \frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{N}{2}\log(2\pi) \end{aligned} \tag{5}$$

with the vector of hyperparameters $\mathbf{\Theta} = [w_1 \ldots w_D,\ v,\ v_0]^T$ (parameter v_0 is included in the vector of hyperparameters $\mathbf{\Theta}$ from now on) and $N \times N$ training covariance matrix $\mathbf{K}$. The optimization requires the computation of the derivative of $\mathcal{L}$ with respect to each of the parameters:

$$\frac{\partial \mathcal{L}(\mathbf{\Theta})}{\partial \Theta_i} = -\frac{1}{2}\text{trace}\left(\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \Theta_i}\right) + \frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \Theta_i}\mathbf{K}^{-1}\mathbf{y} \tag{6}$$

Here, it involves the computation of the inverse of the $N \times N$ covariance matrix $\mathbf{K}$ at every iteration, which can be computationally demanding for large N. An alternative method for parameters optimization is to put a Gaussian prior on the parameters and compute their posterior probability (Rasmussen (1996)).

Given that the hyperparameters are known, the prediction of the GP model at the input $\mathbf{x}_{N+1}$ can be obtained. The conditional part of (3) provides the predictive distribution of y_{N+1}:

$$p(y_{N+1}|\mathbf{y}, \mathbf{X}, \mathbf{x}_{N+1}) = \frac{p(\mathbf{y}, y_{N+1})}{p(\mathbf{y}|\mathbf{X})} \tag{7}$$

It can be shown (e.g. Gibbs (1997)) that this distribution is Gaussian with mean and variance:

$$\begin{aligned} \mu(y_{N+1}) &= \mathbf{k}(\mathbf{x}_{N+1})^T\mathbf{K}^{-1}\mathbf{y} & (8) \\ \sigma^2(y_{N+1}) &= k(\mathbf{x}_{N+1}) - \mathbf{k}(\mathbf{x}_{N+1})^T\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}_{N+1}) + v_0 & (9) \end{aligned}$$

where $\mathbf{k}(\mathbf{x}_{N+1}) = [C(\mathbf{x}_1, \mathbf{x}_{N+1}), \ldots, C(\mathbf{x}_N, \mathbf{x}_{N+1})]^T$ is the $N \times 1$ vector of covariances between training inputs and the test input and $k(\mathbf{x}_{N+1}) = C(\mathbf{x}_{N+1}, \mathbf{x}_{N+1})$ is the autocovariance of the test input.

Vector $\mathbf{k}(\mathbf{x}_{N+1})^T\ \mathbf{K}^{-1}$ in (8) can be interpreted as a vector of smoothing terms which weights training outputs $\mathbf{y}$ to make a prediction at the test point $\mathbf{x}_{N+1}$. If the new input is far away from the data points, the term $\mathbf{k}(\mathbf{x}_{N+1})^T\ \mathbf{K}^{-1}\ \mathbf{k}(\mathbf{x}_{N+1})$ in (9) will be small, so that the predicted variance $\sigma^2(\mathbf{x}_{N+1})$ will be large. Regions of the input space, where there are few data or where the data have high complexity or are corrupted with noise, are in this way indicated through higher variance.

3. Dynamical system identification

3.1. Problem statement

Consider the autoregressive model of a L-th order dynamical system (2), where the vector of regressors $\mathbf{x}(k)$ is composed of previous values of outputs y and control inputs u up to a given lag L, $D = 2L$:

$$\mathbf{x}(k) = [y(k-1), y(k-2), \dots, y(k-L), u(k-1), u(k-2), \dots, u(k-L)]^T \quad (10)$$

where $(k - i)$ denotes the appropriate time sample.

The use of the GP method for dynamical systems identification can be found in (Kocijan et al. (2005); Leith et al. (2002)), here only the main results will be recalled.

One way to do n-step ahead prediction is to make iterative one-step ahead predictions up to the desired step n whilst feeding back the mean of the predicted output. This approach is similar to that used in modelling dynamic systems with ANN and is illustrated in Figure 1. The drawback of this approach is that the information about the uncertainty of the output is neglected. The simulation of the GP model, where complete output distribution is fed back, is presented in Kocijan and Girard (2005) and Girard (2004).

[Insert Figure 1 about here]

To avoid problems using the general GP dynamic models cited in the Introduction, we wish to concentrate here on the modelling of dynamic system (2) using a Gaussian process model with incorporated linear local models. With this model we would like to be able to make n-step ahead prediction.

3.2. Local models

Let us assume that the point $\mathbf{x}_i$ from (10) is the equilibrium point of the stable, generally nonlinear system (2). We would like to present the system's dynamical behaviour in the vicinity of point $\mathbf{x}_i$ with approximation in the form of a linear local model $\mathcal{M}_i$:

$$y(\mathbf{x}) = f(\mathbf{x}_i) + \boldsymbol{\theta}_i^T \left(\mathbf{x} - \mathbf{x}_i\right) \quad (11)$$

where

$$\begin{aligned} \boldsymbol{\theta}_i^T &= \left[\mathbf{a}_i^T, \mathbf{b}_i^T\right] && (12)\\ \mathbf{a}_i &= \left[\frac{\partial f}{\partial y_{k-1}}, \cdots, \frac{\partial f}{\partial y_{k-L}}\right]_i^T && (13)\\ \mathbf{b}_i &= \left[\frac{\partial f}{\partial u_{k-1}}, \cdots, \frac{\partial f}{\partial u_{k-L}}\right]_i^T && (14) \end{aligned}$$

$\boldsymbol{\theta}_i$ are parameters of linear local model $\mathcal{M}_i$ centered in $\mathbf{x}_i$.

Two different types of information are used to construct linear local model $\mathcal{M}_i$:

- functional values (functional observation in Solak et al. (2003)) – values of the system's output $f(\mathbf{x}_i)$ in the centre of the model $\mathbf{x}_i$ and
- derivatives (derivative observation in Solak et al. (2003)) – vector of partial derivatives of system's output $f(\mathbf{x})$ with respect to components $\mathbf{x}_i^k$ of the vector of regressor $\mathbf{x}_i$:
$\boldsymbol{\theta}_i = \left[\frac{\partial f}{\partial y_{k-1}}, \cdots, \frac{\partial f}{\partial y_{k-L}}, \frac{\partial f}{\partial u_{k-1}}, \cdots, \frac{\partial f}{\partial u_{k-L}}\right]_i^T.$

Local models can be derived using any standard linear regression method that gives consistent and unbiased solution, see e.g. Ljung (1999).

3.3. Incorporation of local models into the GP model

Since differentiation is a linear operation, the derivative of a GP remains a GP (Solak et al. (2003)). Consequently, within the Gaussian process modelling framework, the derivatives can be used together with functional values, thus providing the way to include linear local models into the GP model. Some results on this topic can be found in Solak et al. (2003), Leith et al. (2002), Kocijan and Leith (2004). The GP model with incorporated local models will be referred to as the LMGP (Local Models incorporated into Gaussian Processes) model.

To include derivatives into the GP model only the covariance function must be changed appropriately. When using the covariance function (1) between two data points, one can find that the covariance function between a data point and derivative is:

$$\mathrm{Cov}[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^d}, f(\mathbf{x}_j)] = -v w_d (x_i^d - x_j^d) \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d (x_i^d - x_j^d)^2\right] \tag{15}$$

In the same manner the covariance function between two derivatives reads:

$$\mathrm{Cov}[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^d}, \frac{\partial f(\mathbf{x}_j)}{\partial x_j^e}] = v w_e (\delta_{e,d} - w_d (x_i^e - x_j^e)(x_i^d - x_j^d)) \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d (x_i^d - x_j^d)^2\right] \tag{16}$$

where $\delta_{e,d}$ is the Kronecker operator between indices d and e:

$$\delta_{e,d} = \begin{cases} 1, & e = d \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

3.4. Composition of the LMGP model

The problem of off-equilibrium dynamics dictates the following approach to LMGP model composition. Regions of the system, where enough data is given for local models identification — usually in the vicinity of the equilibrium curve — are modelled with local models. Regions of the system, lacking enough data to construct local models, are modelled with individual samples of the system's response. This knowledge is together incorporated into the GP model as illustrated in Figure 2 for the first order example. The GP model "smooths" this information and is able to make robust prediction of the system's response even where the data describing the system is sparse. Another possible approach is to partition the operational space $\mathcal{R}^{\mathcal{D}}$ into several regions and model each region with its own GP model (Gregorčič and Lightbody (2003a)).

[Insert Figure 2 about here]

With the introduction of the local models into the GP model the derivatives are added to the vector of targets $\mathbf{y}$ in equation (8), constituted only of function observations. The values of the regressors corresponding to included derivatives are added to the input matrix $\mathbf{X}$.

Given n_{eq} local models and n_{oeq} samples of the system's response, describing the system's behaviour, one of the possible ways to compose input/target data $\mathbf{X}/\mathbf{y}$ for training of hyperparameters is:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{oeq} \\ \mathbf{X}_{eq} \\ \mathbf{X}_{eq} \\ \vdots \\ \mathbf{X}_{eq} \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} \mathbf{Y}_{oeq1} \\ \mathbf{Y}_{eq1} \\ \boldsymbol{\theta}^1 \\ \vdots \\ \boldsymbol{\theta}^D \end{bmatrix} \tag{18}$$

$$\mathbf{X}_{oeq} = [\mathbf{Y}_{oeq}\ \mathbf{U}_{oeq}] \tag{19}$$

$$\mathbf{X}_{eq} = [\mathbf{Y}_{eq}\ \mathbf{U}_{eq}] \tag{20}$$

where

$\mathbf{Y}_{oeq1}$ is a $n_{oeq} \times 1$ target vector of the system's out-of-equilibria response points;

$\mathbf{X}_{oeq}$ is a $n_{oeq} \times D$ input matrix of appropriate regressors corresponding to target vector $\mathbf{Y}_{oeq1}$;

$\mathbf{Y}_{eq1}$ is a $n_{eq} \times 1$ target vector of the system's response points in the centres of local models;

$\mathbf{X}_{eq}$ is a $n_{eq} \times D$ input matrix of appropriate regressors corresponding to target vector $\mathbf{Y}_{eq1}$;

$\boldsymbol{\theta}^1$ is a $n_{eq} \times 1$ vector of derivatives $\frac{\partial f}{\partial y_{k-1}}$ at input regressor matrix $\mathbf{X}_{eq}$ (vector of derivatives $\frac{\partial f}{\partial y_{k-1}}$ for all n_{eq} incorporated local models);

$\boldsymbol{\theta}^L$ is a $n_{eq} \times 1$ vector of derivatives $\frac{\partial f}{\partial y_{k-L}}$ at input regressor matrix $\mathbf{X}_{eq}$;
$\boldsymbol{\theta}^{L+1}$ is a $n_{eq} \times 1$ vector of derivatives $\frac{\partial f}{\partial u_{k-1}}$ at input regressor matrix $\mathbf{X}_{eq}$;
$\boldsymbol{\theta}^D$ is a $n_{eq} \times 1$ vector of derivatives $\frac{\partial f}{\partial u_{k-L}}$ at input regressor matrix $\mathbf{X}_{eq}$;

Again as in (10), L is the order of the system and $D = 2L$ is the number of regressors. Subscript *oeq* denotes data representing out-of-equilibrium behaviour (response) and subscript *eq* denotes data representing equilibrium behaviour of the system in the form of local models.

Let $n = n_{eq} + n_{oeq}$ be the number of functional data (input-target data points) where n_{eq} is the number of identified local models to be incorporated into the GP model. For the L-th order system there exist in total D vectors of derivatives $\boldsymbol{\theta}^k$, $k = 1, \ldots, D$ with length n_{eq} – one for each input regressor $\mathbf{x}_i$. Thus the size of the input matrix $\mathbf{X}$ is $(n + D \cdot n_{eq}) \times D$ and the length of the target vector $\mathbf{y}$ is $(n + D \cdot n_{eq})$.

When training data $\mathbf{X}/\mathbf{y}$ is composed as presented, the covariance matrix $\mathbf{K}$, vector of covariances between test input and training inputs $\mathbf{k}(\mathbf{x})$ and autocovariance of test input $k(\mathbf{x})$ need to be:

$$\mathbf{K} = \begin{bmatrix} \Big[C(\mathbf{x}_i, \mathbf{x}_j)\Big] & \Big[\mathrm{Cov}[f(\mathbf{x}_i), \frac{\partial f(\mathbf{x}_j)}{\partial x_j^e}]\Big]_{e=1} & \cdots & \Big[\mathrm{Cov}[f(\mathbf{x}_i), \frac{\partial f(\mathbf{x}_j)}{\partial x_j^e}]\Big]_{e=D} \\ \Big[\mathrm{Cov}[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^d}, f(\mathbf{x}_j)]\Big]_{d=1} & \Big[\mathrm{Cov}[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^d}, \frac{\partial f(\mathbf{x}_j)}{\partial x_j^e}]\Big]_{d=1,e=1} & \cdots & \Big[\mathrm{Cov}[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^d}, \frac{\partial f(\mathbf{x}_j)}{\partial x_j^e}]\Big]_{d=1,e=D} \\ \vdots & \vdots & \vdots & \vdots \\ \Big[\mathrm{Cov}[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^d}, f(\mathbf{x}_j)]\Big]_{d=D} & \Big[\mathrm{Cov}[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^d}, \frac{\partial f(\mathbf{x}_j)}{\partial x_j^e}]\Big]_{d=D,e=1} & \cdots & \Big[\mathrm{Cov}[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^d}, \frac{\partial f(\mathbf{x}_j)}{\partial x_j^e}]\Big]_{d=D,e=D} \end{bmatrix} \tag{21}$$

$$\mathbf{k}(\mathbf{x}) = \begin{bmatrix} \Big[C(\mathbf{x}_i, \mathbf{x})\Big] \\ \Big[\mathrm{Cov}[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^d}, f(\mathbf{x})]\Big]_{d=1} \\ \vdots \\ \Big[\mathrm{Cov}[\frac{\partial f(\mathbf{x}_i)}{\partial x_i^d}, f(\mathbf{x})]\Big]_{d=D} \end{bmatrix} \tag{22}$$

$$k(\mathbf{x}) = \Big[C(\mathbf{x}, \mathbf{x})\Big] = v \tag{23}$$

respectively.

The information about the system's behaviour in the vicinity of the point $\mathbf{x}_i$, that was presented as a (large) set of functional values as described in Section 2 and in Kocijan et al. (2005), Kocijan et al. (2003), is now compressed in the parameters of linear local model $\mathcal{M}_i$. This way the behaviour of the system around each equilibrium point is represented with fewer data points, which can effectively reduce the computational burden.

We have to be careful to ensure that all local models are formed using the same state representation. Values of regressors (10) are used as state coordinates in our case, but other choices are possible as well.

Input data can contain noise information. Where this information is available it is added to corresponding elements of covariance matrix (Solak et al. (2003)); where not, the parameter describing noise variance is trained as stated in Section 2. Prediction of the LMGP model is given with equations (8) and (9).

Dynamic response of the LMGP model in off-equilibrium regions is represented with data points (response samples) and therefore represents global and not local dynamical behaviour in these regions. On the other hand the incorporated local models on the equilibrium curve encapsulate the system's local dynamics and as the parameters of these local models do not change with optimisation, the dynamics of the system remains well modelled. The LMGP model does not have a scheduling variable and also does not suffer very much from partitioning as the local models need only be put over the equilibrium curve in the necessary density.

Beside not suffering from some of the problems of the LMN approach, the *confidence* in the LMGP model's output, depending on the input data, is also provided. This confidence can be seen as the criterion for model quality in corresponding region of the system.

4. Illustrative example

In this section the GP model with incorporated LM approach is presented on the identification of the following discrete nonlinear second-order dynamical system (Matko et al. (2000)):

$$\begin{aligned} y(k) &= 0.893y(k-1) + 0.0371y^2(k-1) - 0.05y(k-2) - \\ &\quad -0.05u(k-1)y(k-1) + 0.157u(k-1) + \epsilon(k) \end{aligned} \tag{24}$$

where the output is corrupted with white Gaussian noise $\epsilon(k) \sim \mathcal{N}(0, \sigma^2)$ with variance $\sigma^2 = 4 \cdot 10^{-4}$.

Our task will be to model the region bounded with input spanning between $u_{\min} = -2$ in $u_{\max} = 4$ for the purpose of n-step ahead prediction. The statical characteristic of the system (24) in the interested region is depicted in Figure 3. The nonlinearity of the system is also shown in Figure 4, where the system's response to the alternating step signal with growing magnitude is presented.

[Insert Figure 3 about here]

[Insert Figure 4 about here]

4.1. Identification

As stated in Section 3.4, two different types of data represent the unknown system in the LMGP model:

- local models, describing the system's dynamics in their centers and vicinity, with centers lying on the equilibrium curve and
- samples of the system's response, which describe the system regions not described with LM (usually transient regions between equilibrium states).

These two different representations require two different measurement types.

For obtaining the off-equilibrium data, which describe the system's behaviour in transient regions, the system must be excited with such input that samples of regressors cover as much operating space $\mathcal{R}^D$ as possible. In our example a pseudo-random binary signal (PRBS) was used as an excitation signal, except that the magnitude of the input could occupy any random value between $u_{\min}$ and $u_{\max}$ when changed. Training points for the LMGP model were later sampled from input and system's response.

Behaviour in the vicinity of the equilibrium curve is modelled with local models. To obtain the local model's parameters, the system is first driven into the equilibrium point with a static input signal. After the settlement of the system's response, PRBS with small magnitude ΔU is added to the input to stimulate the system's dynamic response around the equilibrium point.

To obtain equilibrium dynamics of the system (24), five approximately evenly distributed local models on the equilibrium curve were identified. Their centers on the equilibrium curve can be seen in Figure 3. PRBS signal with switching time $T_{sw} = 4$ steps and the magnitude of perturbation $\Delta U = 0.3$ was selected so that local models can be identified despite the noise. The models were identified using the *Instrumental Variables(IV)* algorithm (Ljung (1999)).

An example of the identified local model response in equilibrium point $(U_{eq}, Y_{eq}) = (0.415, 0.4)$ is presented in Figure 5, and it can be seen that the model perfectly captures the dynamics of the system. It should be however taken into account that here the identified system is ideal and of known order.

[Insert Figure 5 about here]

Each of five local models contributed one functional value (value of system's response at equilibrium point) and four derivatives (one for each regressor) to the training data. Thus together with fourteen points, sampled out of the system's off-equilibrium response, the LMGP model was formed using 39 training points. The estimates of local models' parameters variances, gained through identification, were added to corresponding elements of the covariance matrix (21).

4.2. Validation

The acquired LMGP model was validated with data not used in identification. Two different inputs for simulation were used, both with the same general properties as the sampled identification input, i.e. switching time step $T_{sw} = 4$ and the range between $u_{\min} = -2$ and $u_{\max} = 4$:

- validation signal 1, where the maximum change of the input signal was limited, so that the model was moving through better modelled regions and
- validation signal 2, where the change of the input was not limited and the model moved through whole operating region.

The idea behind this choice is to better present the properties of the LMGP model. The results of simulation where the input was validation signal 1 is presented in Figure 6 and the corresponding simulation error with accompanying 95% confidence band of the the model's prediction is shown in Figure 7. Note that for illustration purposes these two figures represent only a segment of the whole simulation result. The autocorrelation of the simulation error Φ_{ee} and the cross-correlation between input and simulation error Φ_{ue} for simulation on validation signal 1 can be seen in Figure 8. From Figures 6 to 8 it can deduced that the description of system behaviour in off-equilibrium regions is satisfactory enough, even though only fourteen samples of the system's response were used. The model could be further improved by adding more samples of the system's response to training data.

[Insert Figure 6 about here]

[Insert Figure 7 about here]

[Insert Figure 8 about here]

The segment of the result of simulation on validation signal 2 are depicted in Figures 9 and 10, where predicted output together with 95% confidence band is compared to system's output and simulation error together with 95% confidence measure are shown correspondingly. These results show another useful attribute of the LMGP model. That is, the confidence in the prediction of the output decreases where the particular region of the system is not modelled adequately, which can be seen from Figure 10 when comparing the error and corresponding variance of the prediction.

[Insert Figure 9 about here]

[Insert Figure 10 about here]

Also two quality measures (Kocijan et al. (2005)) were used on the results of validation:

- mean squared error

$$\mathrm{SE} = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \tag{25}$$

and

- log-predictive density error (Girard (2004))

$$\mathrm{LD} = \frac{1}{2N}\sum_{i=1}^{N}\left(\log(2\pi) + \log(\sigma_i^2) + \frac{e_i^2}{\sigma_i^2}\right) \tag{26}$$

where $e_i = \hat{y}_i - y_i$ is an error of the model's simulation and σ_i^2 a predicted variance of the model in i-th time step. The results of both quality measures applied to validation signals can be seen from Table 1.

[Insert [Table] 1 about here]

In the LMGP model framework the three exposed problems of the LMN approach have diminished influence. The system in off-equilibrium regions is represented with data points and the interpolation is smooth as one of the attributes of the GP model. The problem of changing the model's local dynamics properties to provide a better global fit was solved within the Gaussian processes framework, i.e. the information describing system's local dynamics does not change with optimisation. The problem of scheduling vector selection drops out as there is no scheduling vector. The problem of region partitioning is reduced as the local models are put only on the equilibrium curve and not over the whole operating region as in the case of LMN. Also the values of covariance function hyperparameters can be used as an indication of the influence along the corresponding regressor components.

5. Identification of a two tank system

In the previous section the identification of a discrete second-order system with the LMGP model was presented. Here the results of the same identification method will be presented for a laboratory pilot plant. Again the purpose of the model is n-step ahead prediction.

5.1. Laboratory pilot plant and chosen subsection for identification

The flowsheet of the laboratory pilot plant's unknown subsystem is presented in Figure 11. The subsystem consist of two tanks, $R1$ and $R2$, connected with flow paths, serving to supply liquid from the reservoir $R0$. The flow path from reservoir $R0$ to tank $R1$ has a built-in pump $P1$, driven with a DC motor with permanent magnet. The angular speed of the motor is controlled by the analog controller. The time constant of the angular speed is very short compared to the time constants of the dynamics of the levels in the tanks, i.e. we can consider no lag between the reference speed and the real one.

[Insert Figure 11 about here]

Flow is generated by varying the angular speed of the pump $P1$. The other interesting part is the manual valve $V5$, which is positioned on the path from tank $R2$ to reservoir $R0$. It is partly open, so it enables liquid flow from tank $R2$ back to reservoir $R0$. The capacity of the reservoir $R0$ is much greater than the capacity of the tanks so that its level can be considered constant during the operation.

Voltage on the motor, which represents the input into the system, drives the pump $P1$. The pump generates flow from reservoir $R0$ to tank $R1$. Liquid flows from tank $R1$ to tank $R2$ and from there back to reservoir $R0$ through ventil $V5$. The liquid level in tank $R2$ represents the output of the system and was measured using capacity sensor.

5.2. Operating region, measurements and identification

The represented second order system is a single-input single-output system, where voltage U on the DC motor of pump $P1$ is the input, and liquid level h_2 in tank $R2$ is the output of the system. The static characteristic of the system's response is given in Figure 12. Besides the static nonlinearity, observed in Figure 12, the system's dynamics is nonlinear too.

[Insert Figure 12 about here]

The working region is restricted by height $h_{max} = 60$ cm of the tanks $R1$ and $R2$. The maximum voltage on $P1$ was fixed to $U_{max} = 4$ V by the trial and error method, preventing the liquid level h_1 from reaching the top h_{max} of the tank $R1$. Sample time $T_s = 10$ seconds was chosen experimentally, so that the dynamics of the system were satisfactorily modelled.

As in the previous section example, the LMGP model demands two different types of measurements for two different types of incorporated information:

- measurements providing data for linear local model identification in different points on the statical characteristic (equilibrium curve) and
- measurements providing data to model the system in off-equilibrium regions.

When identifying linear local models, the system was first brought to the desired equilibrium point by the corresponding static input signal U_{eq}. Then PRBS with magnitude ΔU (see Table 2 for values) and switching time $T_{sw} =$ 20 s was added to it. An example of input and measured output signal for working point $(U_{eq}, Y_{eq}) =$ (2.5 V, 16 cm) is given in Figure 13. Local models were, as in Section 4, identified using the *Instrumental Variables* algorithm

(Ljung (1999)). They were validated on the PRBS with the same characteristics as, but a different sequence to, the signal for identification. An example of identified local model's response to the validation signal is given in Figure 13.

[Insert Figure 13 about here]

Three local models were identified. Their parameters, together with the estimates of their standard deviations, are presented in Table 2. The second order was chosen for the local models as the underlying first principle model is of the second order. Higher order models were also tried, but increasing the order did not contribute to an improvement of the model's response. A discrete linear local model, modelling the behaviour of the system in the vicinity of working point (U_{eq}, Y_{eq}), is described by parameters a_1, a_2, b_1 and b_2:

$$y(k) = -a_1 y(k-1) - a_2 y(k-2) + b_1 u(k-1) + b_2 u(k-2) \qquad (27)$$

where y is liquid level h_2 and u is voltage U at appropriate time samples.

[Insert [Table] 2 about here]

A second type of measurement provided us with data describing the system in off-equilibrium regions. Again input signals were realized as PRBS signals where magnitude could occupy any value between $U_{\min} = 0.8$ V and $U_{\max} = 4$ V when changed. Two such signals were used, one for training and one for validation. A part of the signal for validation can be seen in Figure 14.

[Insert Figure 14 about here]

Three local models obtained in equilibrium points and 68 samples from off-equilibrium system's response were incorporated into the LMGP model, resulting in 83 input/target training pairs. The estimates of local models' parameters variances, gained through identification, were added to corresponding elements of the covariance matrix (21).

5.3. Validation

The simulation input signal for validation was different from the one for identification. The results of the simulation of the trained LMGP model on the part of the validation signal can be seen in Figure 15. The signals over the whole time scale are again not shown for illustration purposes. Absolute error of the model simulation together with predicted 95% confidence band is depicted in Figure 16. In Figure 17 the autocorrelation of the simulation error and cross-correlation on the whole validation signal between input and simulation error is shown. The performance measure values for validation are SE = 0.26 and LD = 3.5. These results show that the LMGP performance is moderately fine.

[Insert Figure 15 about here]

[Insert Figure 16 about here]

[Insert Figure 17 about here]

6. Conclusions

The basic concepts of dynamical systems identification with the GP model with incorporated local models were introduced in this paper. The method was demonstrated on two dynamical systems — a nonlinear discrete second order system and a laboratory pilot plant. To our knowledge these are the first published results of the LMGP method used to identify a higher order system and also the first results where the method was used to identify a system using measured data. The result of the illustrative example identification showed that method can be successful and the result of the plant identification proved the method's practical capability.

The advantages of the Gaussian process modelling method with incorporated local models (LMGP method) over LMN concepts are:

- Equilibrium and off-equilibrium information describing the system is joined transparently in the LMGP model.
- The description of the system's behaviour does not change with optimization as the data describing the system (local models, system's response samples) does not change.
- There is no problem of scheduling vector determination.

Beside avoiding some of the LMN problems, the additional bonus of the method is that it gives confidence in its prediction based on the variance of the parameters and the distribution of the training data in the input space. Confidence can be viewed as the indicator of model validity region and used, for example, in control of the identified system.

When comparing the LMGP model to the GP model, the advantage of the former is that it provides the possibility to include prior knowledge in the form of local models. This also results in possible reduction of the computational burden, as local models can effectively substitute a larger set of training points describing a nearby region of the system. The drawback of the method is the partitioning of the operating space, though not to the same extend as in LMN, because the local models are to be obtained in the equilibrium points only. This problem originates in an unknown system to be identified, and is common to all approaches where local models are used.

The future plans are to include the propagation of uncertainty into the LMGP model and exploitation of the various potential applications of models obtained with the presented method. Another interesting problem is systematic data preprocessing and data fusion from large amounts of documented signals available in industrial environments, e.g. the process industry.

Acknowledgements

We thank Dr. Đani Juričić and Prof. Dr. Roderick Murray-Smith for helpful discussion. The support of the Ministry of Higher Education, Science and Technology, Grant No. P2-0001, is also gratefully acknowledged.

References

Gibbs, M., 1997. Bayesian Gaussian processes for regression and classification. Ph.D. thesis, Cambridge University.

Girard, A., 2004. Approximate methods for propagation of uncertainty with Gaussian process models. Ph.D. thesis, University of Glasgow, Glasgow.

Gregorčič, G., Lightbody, G., 2003a. An affine Gaussian process approach for nonlinear system identification. Systems Science Journal 29 (2), 47–63.

Gregorčič, G., Lightbody, G., 2003b. From multiple model networks to the Gaussian processes prior model. In: Proceedings of IFAC ICONS conference. Faro, pp. 149–154.

Johansen, T., Shorten, R., Murray-Smith, R., 2000. On the interpretation and identification of dynamic Takagi-Sugeno fuzzy models. IEEE Transactions on Fuzzy Systems 8 (3), 297–313.

Kocijan, J., Banko, B., Likar, B., Girard, A., Murray-Smith, R., Rasmussen, C., 2003. A case based comparison of identification with neural networks and Gaussian process models. In: Proceedings of IFAC ICONS conference. Faro, pp. 137–142.

Kocijan, J., Girard, A., 2005. Incorporating linear local models in Gaussian process model. In: Proceedings of 16th IFAC world congress. Prague.

Kocijan, J., Girard, A., Banko, B., Murray-Smith, R., December 2005. Dynamic systems identification with Gaussian processes. Mathematical and Computer Modelling of Dynamic Systems 11 (4), 411–424.

Kocijan, J., Leith, D., 2004. Derivative observations used in predictive control. In: Proceedings of IEEE Melecon. Dubrovnik.

Leith, D., Leithead, W., Solak, E., Murray-Smith, R., 2002. Divide and conquer identification using Gaussian processes. In: Proceedings of the 41st Conference on Decision and Control. Las Vegas, AZ, pp. 624–629.

Ljung, L., 1999. System identification – theory for the user, 2nd Edition. Prentice Hall, New Jersey.

MacKay, D., 2003. Information theory, inference and learning algorithms. Cambridge University Press, Cambridge.

Matko, D., Škrjanc, I., Mušič, G., 2000. Robustness of fuzzy control and its application to a thermal plant. Mathematics and Computer in Simulation 51, 245–255.

Murray-Smith, R., Johansen, T. (Eds.), 1997. Multiple model approaches to modelling and control. Taylor and Francis, London.

Murray-Smith, R., Johansen, T., Shorten, R., 1999. On transient dynamics, off-equilibrium behaviour and identification in blended multiple model structures. In: Proceedings of European Control Conference. BA-14. Karlsruhe.

Neal, R., 1996. Bayesian learning for neural networks. Vol. 118 of Lecture Notes in Statistics. Springer-Verlag, New York.

Rasmussen, C., 1996. Evaluation of Gaussian processes and other methods for nonlinear regresion. Ph.D. thesis, University of Toronto.

Solak, E., Murray-Smith, R., Leithead, W., Leith, D., Rasmussen, C., 2003. Derivative observations in Gaussian process models of dynamic systems. In: Becker, S., Thrun, S., Obermayer, K. (Eds.), Proceedings of Advances in Neural Information Processing Systems conference. Vol. 15. MIT Press, pp. 529–536.

Williams, C., 1998. Prediction with Gaussian processes: from linear regression and beyond. In: Jordan, M. (Ed.), Learning in graphical models. Vol. 89 of NATO Science Series D. Springer, Berlin, pp. 599–621.

Figure 1: Simulation with repeated one-step-ahead prediction of a dynamical model

Figure 2: Target data of LMGP model consist of local models and data points – an illustration of the approach for first order system. Ovals represent local models' proximity regions and dots represent samples of off-equilibrium system's response.

Figure 3: Static characteristic of the second-order dynamical system

Figure 4: Response of the second-order dynamical system to a growing alternate step input signal

Figure 5: Example of the identified local model response in equilibrium point $Y_{eq} = 0.4$

Figure 6: Simulation of identified LMGP model on validation signal 1

Figure 7: Absolute error of the LMGP model simulation on the validation signal 1 together with 95% confidence band

Figure 8: Autocorrelation of the simulation error Φ_{ee} (top) and cross-correlation between simulation input and error Φ_{ue} (bottom) on validation signal 1

Figure 9: Simulation of identified LMGP model on validation signal 2

Figure 10: Absolute error of the LMGP model simulation on the validation signal 2 together with 95% confidence band

Figure 11: Process scheme of chosen subsystem of the plant

Figure 12: Static characteristic of the chosen subsystem of the plant

Figure 13: Example of the identified local model response at equilibrium point $U_{eq} = 2.5$ V, $Y_{eq} = 16.0$ cm

Figure 14: Validation signal for LMGP model of the plant

Figure 15: Comparison between LMGP two tanks model simulation and plant response to the same input signal for validation

Figure 16: Absolute error of the two tanks LMGP model simulation on the validation signal together with 95% confidence

Figure 17: Autocorrelation of the simulation error Φ_{ee} (top) and cross-correlation between simulation input and error Φ_{ue} (bottom) on validation signal for LMGP model of the plant

Table 1: Two quality measures of simulation for validation for illustrative example (24), namely mean square error SE (25) and log-predictive density error LD (26)

Valid. signal	SE	LD
1	1.8E-3	-1.00
2	3.2E-3	-0.47

Table 2: (U_{eq}, Y_{eq}) define i-th local model working point in (cm, V), $-a_1, -a_2, b_1, b_2$ are the coefficients of the linear local model together with their standard deviations and ΔU is the magnitude of PRBS in V in working point for LM estimation

U_{eq}	Y_{eq}	$-a_1$	$-a_2$	b_1	b_2	ΔU
1.0	3.7	0.7573 ± 0.0699	-0.0580 ± 0.0532	0.2903 ± 0.0051	0.0704 ± 0.0248	0.4
2.0	9.2	0.8550 ± 0.0362	-0.0462 ± 0.0338	1.0203 ± 0.0348	0.9716 ± 0.0574	0.4
2.5	16.0	1.1594 ± 0.0186	-0.2912 ± 0.0174	0.7792 ± 0.0247	0.8767 ± 0.0384	0.2

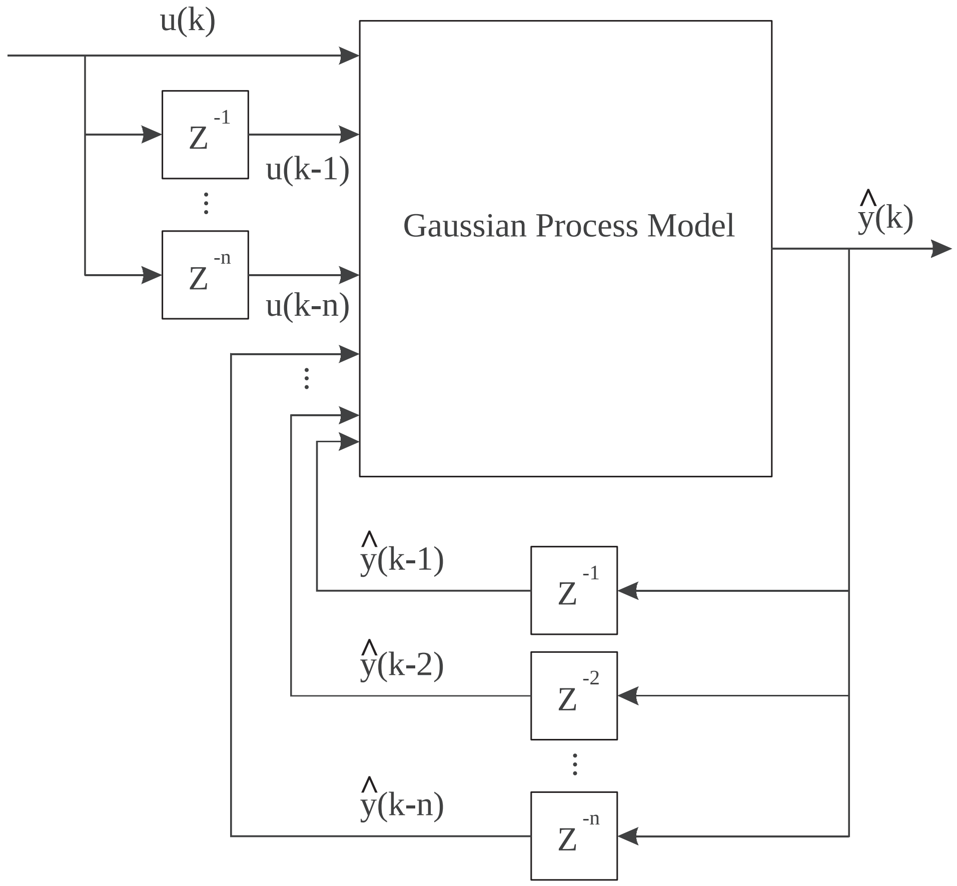

Figure 1:

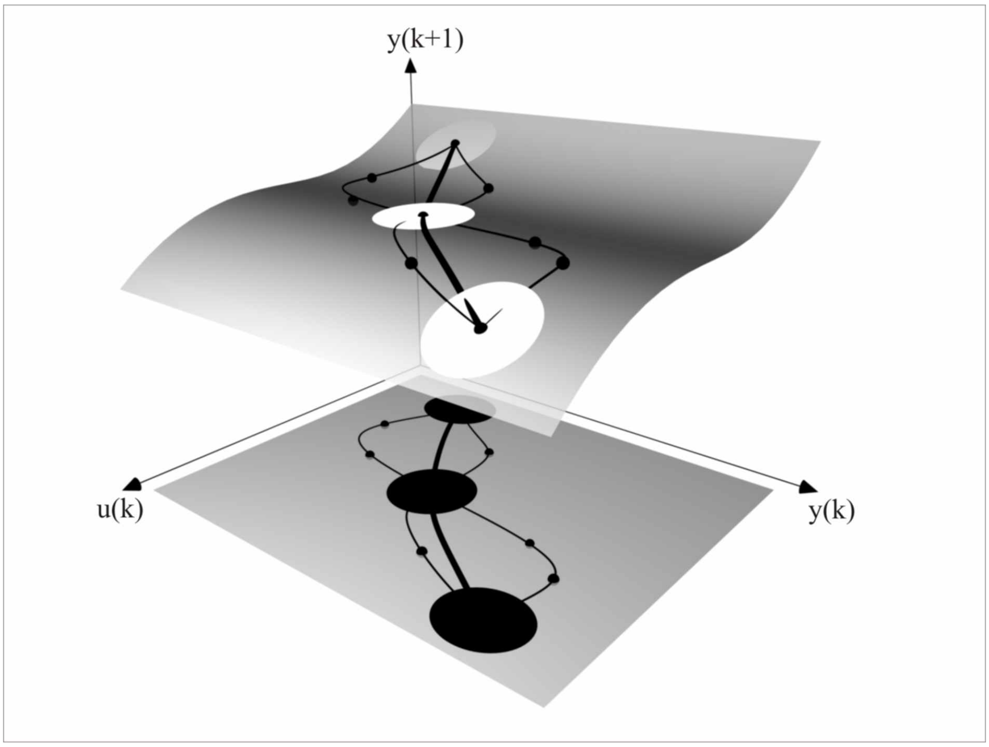

Figure 2:

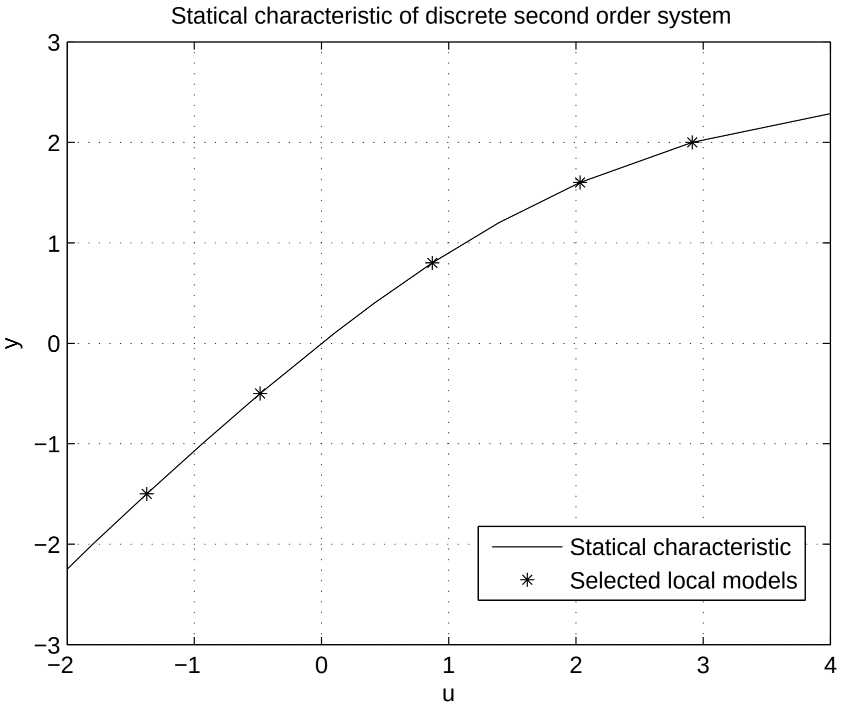
Statical characteristic of discrete second order system
y
u
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
4
Statical characteristic
Selected local models

Figure 3:

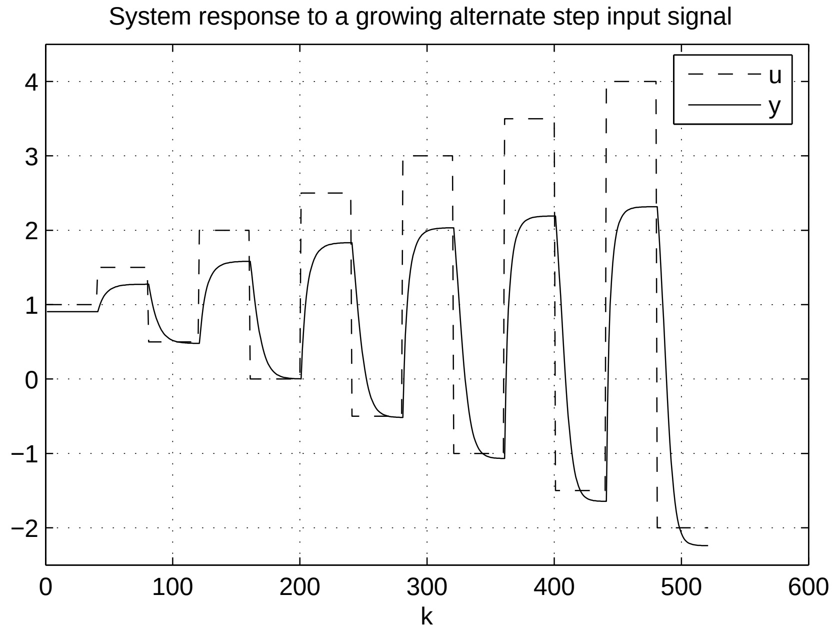
System response to a growing alternate step input signal
u
y
4
3
2
1
0
−1
−2
0
100
200
300
400
500
600
k

Figure 4:

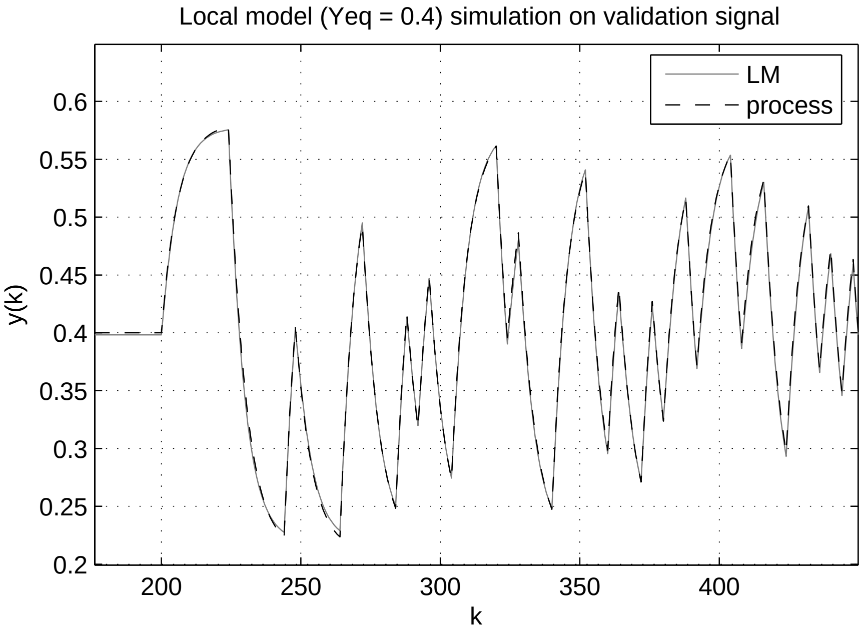
Local model (Yeq = 0.4) simulation on validation signal
LM
process
y(k)
k
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
200
250
300
350
400

Figure 5:

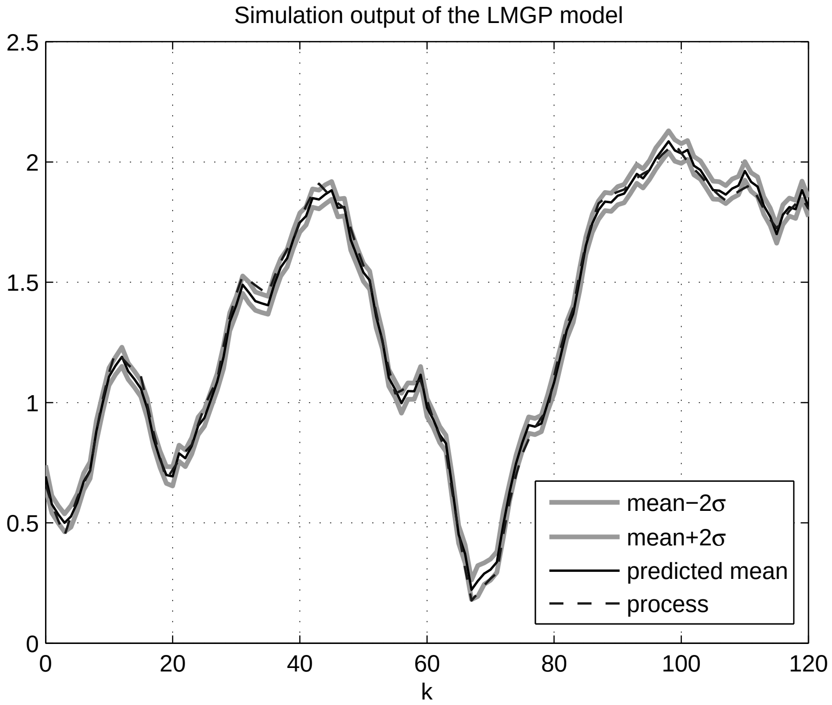
Simulation output of the LMGP model
2.5
2
1.5
1
0.5
0
0
20
40
60
80
100
120
k
mean−2σ
mean+2σ
predicted mean
process

Figure 6:

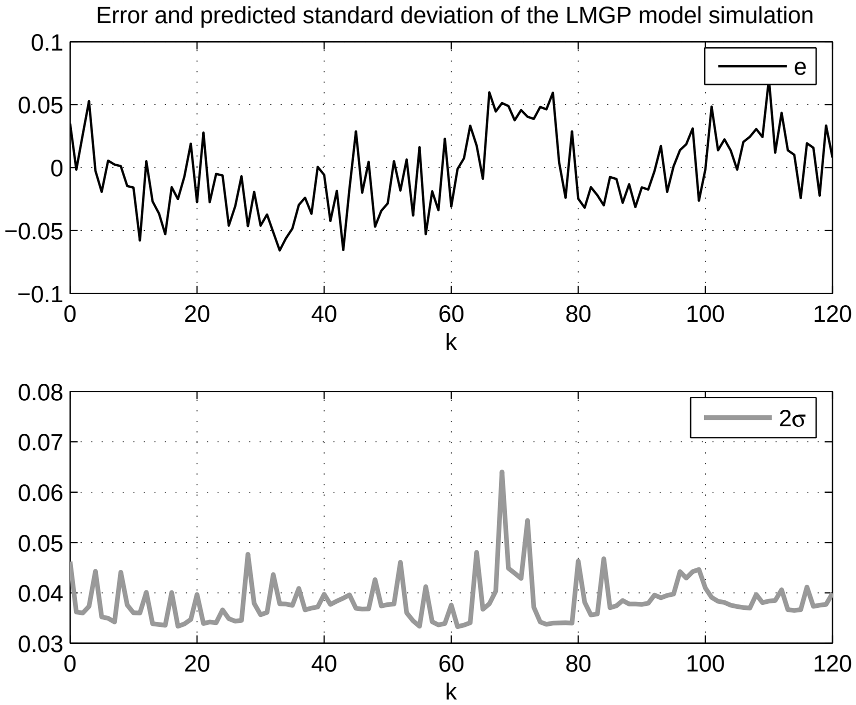
Error and predicted standard deviation of the LMGP model simulation
0.1
0.05
0
−0.05
−0.1
e
0
20
40
60
80
100
120
k
0.08
0.07
0.06
0.05
0.04
0.03
2σ
0
20
40
60
80
100
120
k

Figure 7:

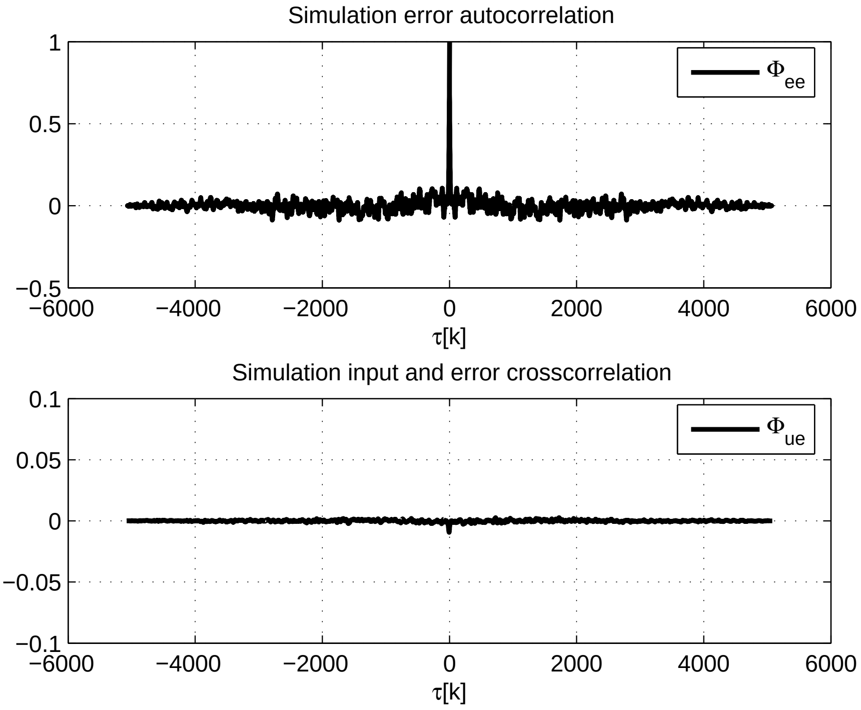
Simulation error autocorrelation
1
0.5
0
−0.5
−6000
−4000
−2000
0
2000
4000
6000
τ[k]
Φee
Simulation input and error crosscorrelation
0.1
0.05
0
−0.05
−0.1
−6000
−4000
−2000
0
2000
4000
6000
τ[k]
Φue

Figure 8:

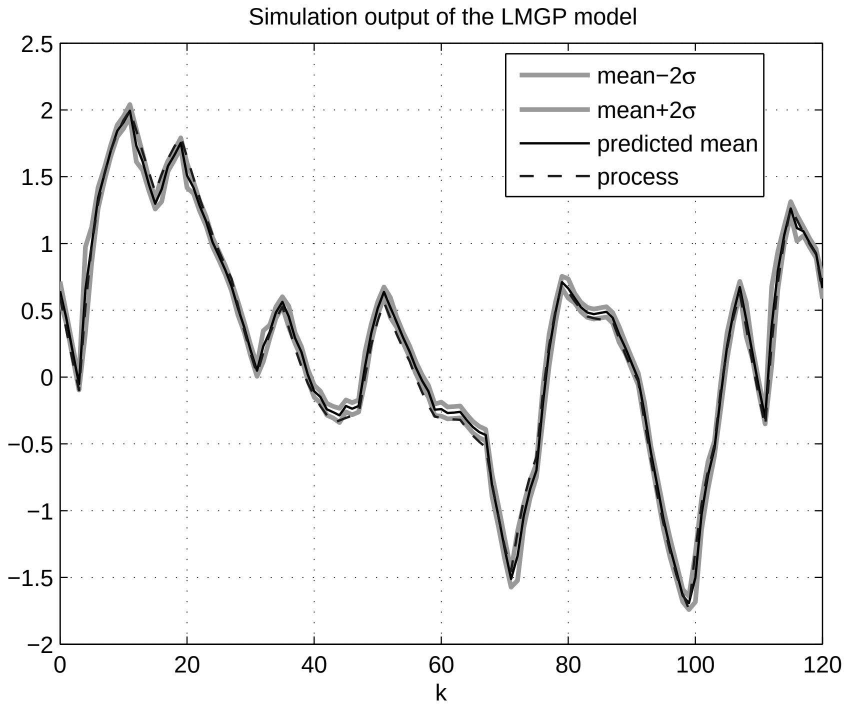
Simulation output of the LMGP model
2.5
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
0
20
40
60
80
100
120
k
mean−2σ
mean+2σ
predicted mean
process

Figure 9:

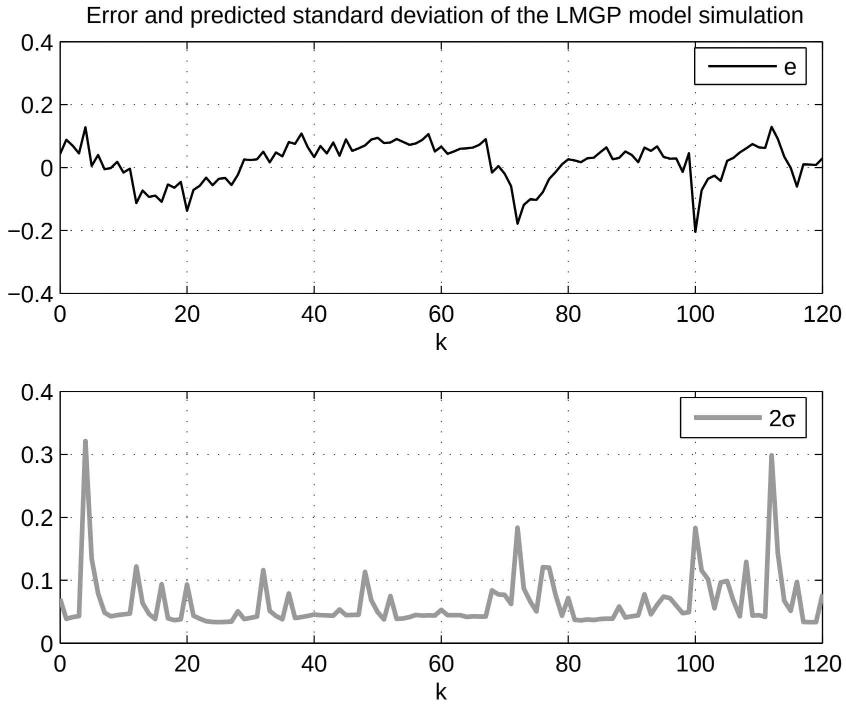
Error and predicted standard deviation of the LMGP model simulation
0.4
0.2
0
−0.2
−0.4
0
20
40
60
80
100
120
k
e
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
120
k
2σ

Figure 10:

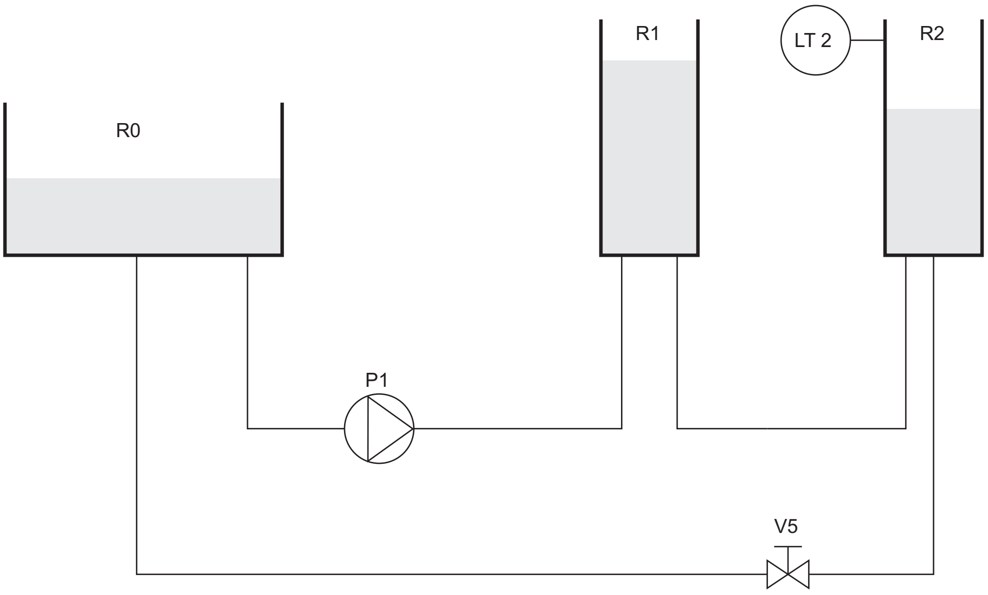

Figure 11:

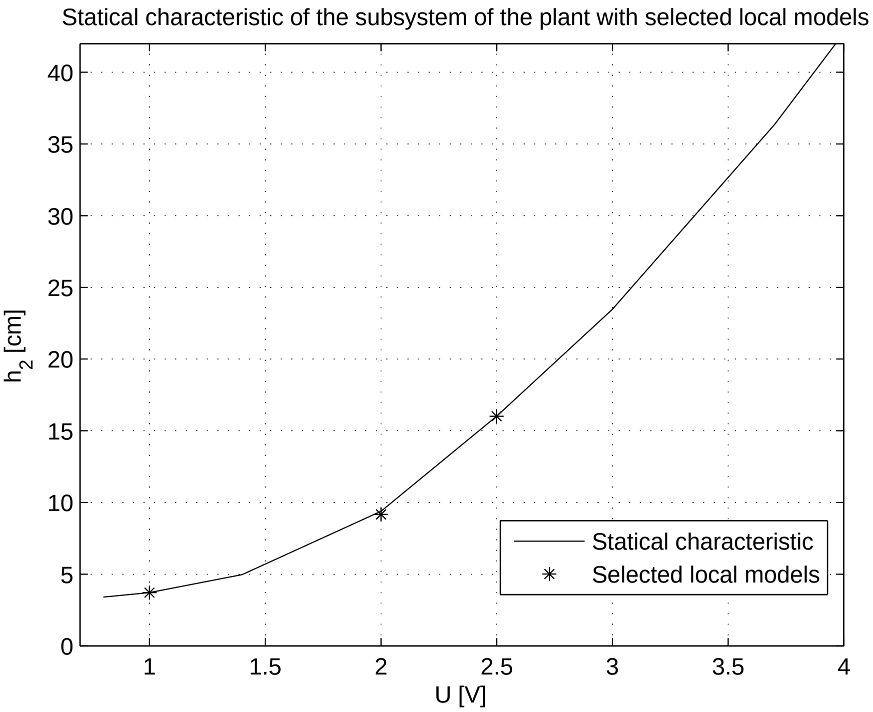
Statical characteristic of the subsystem of the plant with selected local models
h_2 [cm]
U [V]
0
5
10
15
20
25
30
35
40
1
1.5
2
2.5
3
3.5
4
Statical characteristic
Selected local models

Figure 12:

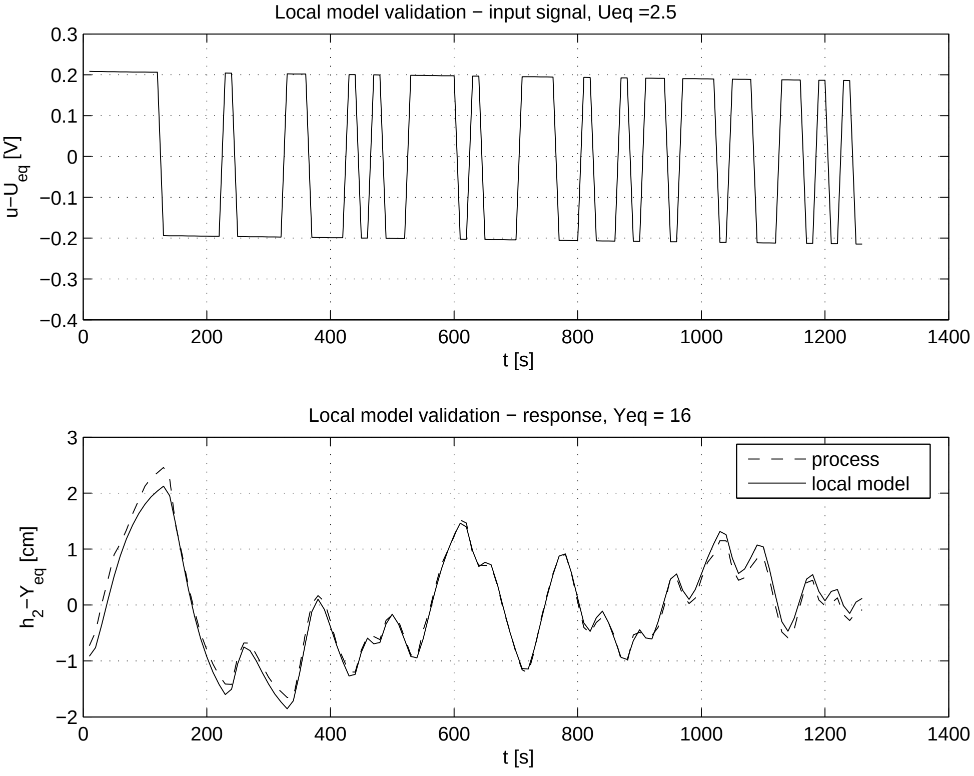
Local model validation − input signal, Ueq =2.5
u−U$_{eq}$ [V]
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
0
200
400
600
800
1000
1200
1400
t [s]
Local model validation − response, Yeq = 16
h$_2$−Y$_{eq}$ [cm]
3
2
1
0
−1
−2
process
local model
t [s]

Figure 13:

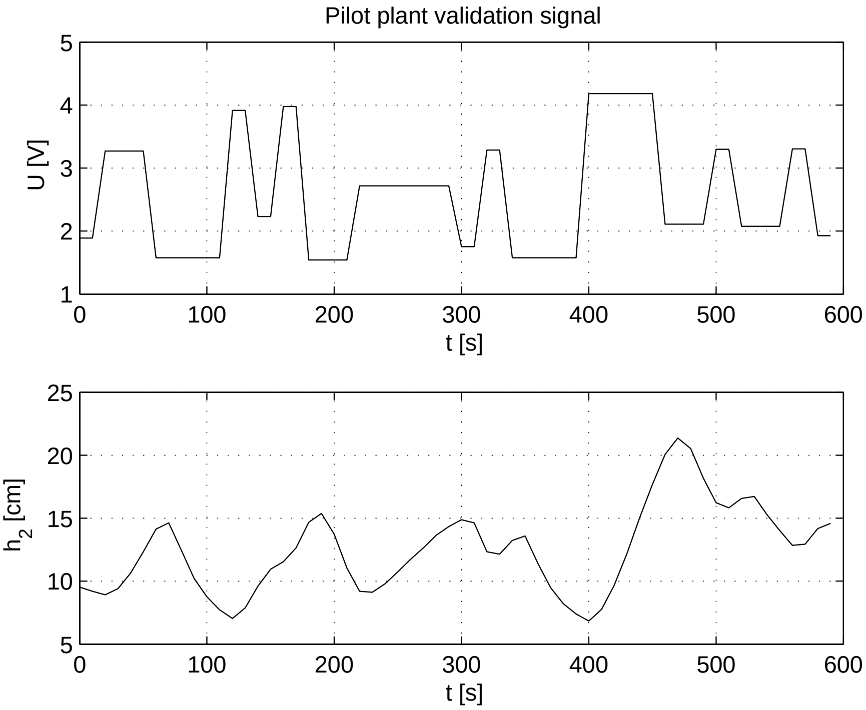
Pilot plant validation signal
U [V]
1
2
3
4
5
0
100
200
300
400
500
600
t [s]
h_2 [cm]
5
10
15
20
25
0
100
200
300
400
500
600
t [s]

Figure 14:

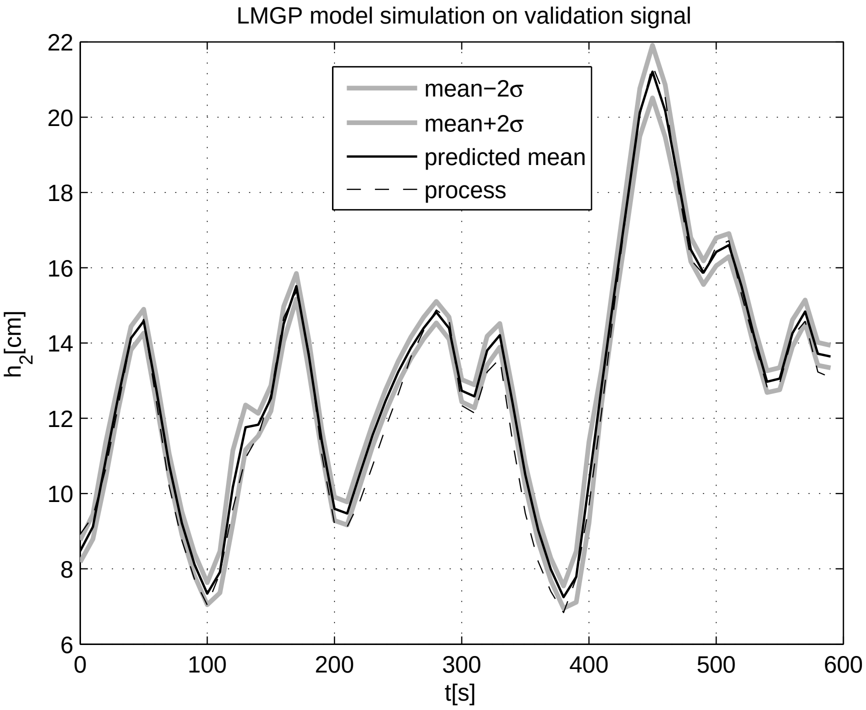
LMGP model simulation on validation signal
22
20
18
16
14
12
10
8
6
h2[cm]
0
100
200
300
400
500
600
t[s]
mean−2σ
mean+2σ
predicted mean
process

Figure 15:

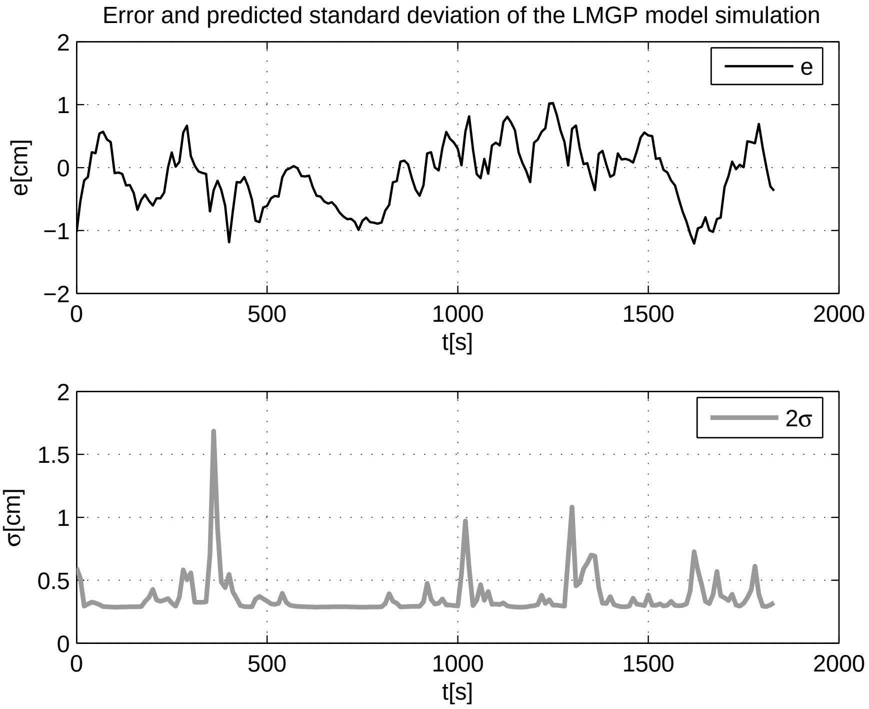
Error and predicted standard deviation of the LMGP model simulation
e[cm]
e
2
1
0
−1
−2
0
500
1000
1500
2000
t[s]
σ[cm]
2σ
2
1.5
1
0.5
0
0
500
1000
1500
2000
t[s]

Figure 16:

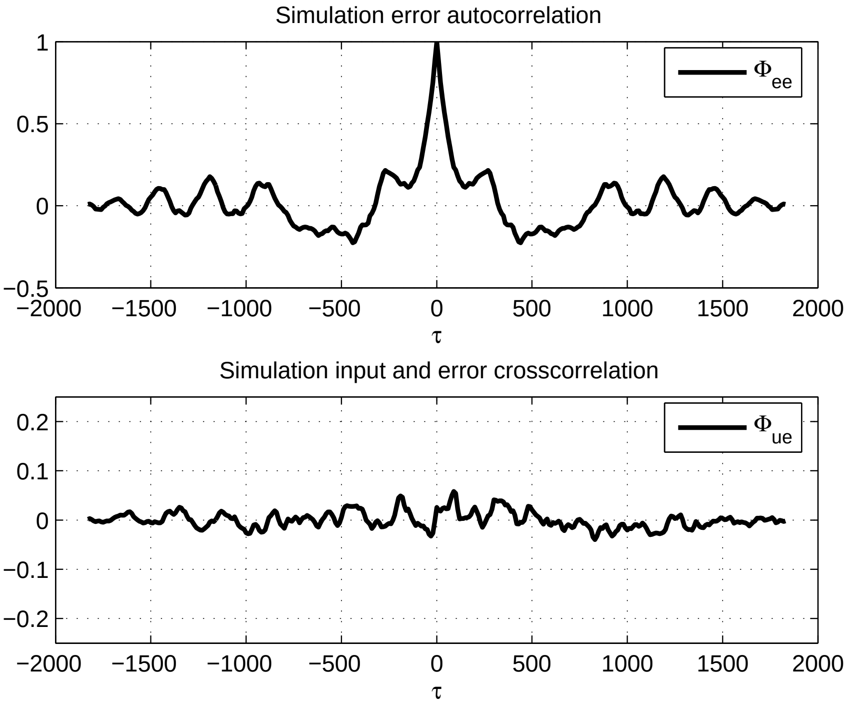
Simulation error autocorrelation
1
0.5
0
−0.5
−2000
−1500
−1000
−500
0
500
1000
1500
2000
τ
Φ_{ee}
Simulation input and error crosscorrelation
0.2
0.1
0
−0.1
−0.2
−2000
−1500
−1000
−500
0
500
1000
1500
2000
τ
Φ_{ue}

Figure 17: