An example of Gaussian process model identification

Kristjan Ažman[†] and Juš Kocijan[†,‡]
†Department of System and Control
Jožef Stefan Insitute
Jamova 39, 1000 Ljubljana, Slovenia
Phone: +386 1 4773 707   E-mail: kristjan.azman@ijs.si
‡Nova Gorica Polytechnic, Nova Gorica, Slovenia

Abstract— **The paper describes the identification of nonlinear dynamic systems with a Gaussian process prior model. This approach is an example of a probabilistic, non-parametric modelling. Gaussian process model can be considered as the special case of radial basis function network and as such an alternative to neural networks or fuzzy black box models. An attractive feature of Gaussian process model is that the variance, associated with the model response, is readily obtained. Variance can be seen as uncertainty of the model and can be used to obtain more accurate multi-step ahead prediction. The method is demonstrated on laboratory pilot plant identification.**

I. INTRODUCTION

While there are numerous methods for identification of linear dynamical systems from measured data (e.g. [1]), nonlinear systems are more difficult to tackle. This paper presents the modelling of dynamic systems with Gaussian processes prior model, which is probabilistic, non-parametric black-box model and is comparable to neural networks (NN) or Takagi-Sugeno fuzzy models (TSFM). The output of Gaussian processes (GP) model is Gaussian distribution, expressed in terms of mean and variance. Mean value represents the predicted output and the variance can be viewed as the measure of confidence in the predicted output. Obtained variance is important result, which distinguishes the GP method from NN or TSFM. It can be used to improve multi-step ahead prediction, which will be presented in this paper, and in the control of identified system [2]. The number of GP model's parameters which need to be optimized is much smaller than in the comparable NN or TSFM. This reduces the problem of optimizing algorithm getting stuck in local minima. Another potentially useful attribute of GP model is the possibility to include various kinds of prior knowledge into the model, e.g. local models [3], [4], static characteristic, etc.

GP model was first used for regression problem by O'Hagan [5], but it gained popularity among machine learning community later through works of Rasmussen [6] and Neal [7]. Research of a possible use of the GP model for the identification of dynamical systems was initiated recently with European project MAC (2000-04), which resulted in first publications on dynamical system identification with GP models [8], [9], [10]. The use of the model with uncertain input, which enables the propagation of uncertainty through the model, is presented in [11], [12]. GP model was already used to identify some simulated first order examples, e.g. [8] [9], but to our knowledge never used to identify dynamical system on measurement data.

The purpose of this paper is to present GP model identification of two tank pilot plant. Simulation with and without propagation of the uncertainty is used to validate the trained GP model.

The paper is organized as follows. In the next section the problem to be solved is stated. In Section 3 basic principles of GP modelling are described. Dynamical system identification with GP model with propagation of variance through the GP model is outlined. In Section 4 the results of identification of laboratory pilot plant are presented. In the last section main conclusions are exposed.

II. PROBLEM STATEMENT

Consider the following autoregressive model of a L-th order dynamical system, where the current output $y(k) \in \mathcal{R}$ depends on delayed outputs and control inputs:

$$y(k) = f(\mathbf{x}(k)) + \epsilon(k) \qquad (1)$$

and $\epsilon(t)$ is a white noise with variance v_0. Vector of regressors $\mathbf{x}(k)$

$$\mathbf{x}(t) = [y(k-1), \ldots, y(k-L), u(k-1), \ldots, u(k-L)]^T \qquad (2)$$

from operating space $\mathcal{R}^D, D = 2L$, is composed of the previous values of outputs y and inputs u up to a given lag L, where $(k-i)$ denotes corresponding time sample.

We wish to model this dynamic system using a Gaussian process model and make n-step ahead predictions for validation. Gaussian processes model provides us with output value and its uncertainty. Simulation with and without propagation of uncertainty will be presented and used to evaluate the identified GP model on laboratory pilot plant measurement data.

III. INTRODUCTION TO DYNAMICAL SYSTEMS IDENTIFICATION WITH GAUSSIAN PROCESSES MODEL

A. Modelling with Gaussian processes

Identification with Gaussian processes method is a statistical, non-parametric method for identification. In more detail it is presented for example in [13], [14], [6].

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

Gaussian process is a Gaussian random function, fully described by its mean and variance. We can view on Gaussian process as on a collection of random variables, which have joint multivariate Gaussian distribution: $f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$. Elements Σ_{ij} of covariance matrix $\boldsymbol{\Sigma}$ 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 covariance function, providing it generates nonnegative definitive covariance matrix $\boldsymbol{\Sigma}$.

Certain assumptions about the process are made implicitly with the selection of covariance function. First such assumption is stationarity of the process. This 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. [15]. Second assumption is smoothness of the output. This results in the higher covariance for inputs which are closer together in the input space. Common choice for covariance function, representing these assumptions, is:

$$C(\mathbf{x}_i, \mathbf{x}_j) = v \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d (x_i^d - x_j^d)^2\right] \quad (3)$$

where D is length of vector $\mathbf{x}$ and $\boldsymbol{\Theta} = [w_1, \ldots, w_D, v]^T$ are parameters called hyperparameters, as they define probability distribution of the functions and do not parameterize the function itself. Hyperparameter v controls the magnitude of the covariance and hyperparameters w_i are representing relative importance of each component x^d of vector $\mathbf{x}$. Neal showed [7] that GP model with covariance function (3) is analogous to radial basis neural network (RBNN) with one hidden layer and infinite number of hidden units when assuming certain RBNN's weights distribution.

Consider system (1) with additive white noise with variance v_0, $\epsilon \sim \mathcal{N}(0, v_0)$. The GP prior with covariance function (3) with unknown hyperparameters is put on space of functions $f(.)$. Within this probabilistic 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.

Based on a set of N training data pairs $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$ we wish to find the predictive distribution of y^* corresponding to a new given input $\mathbf{x}^*$. For collection of random variables $(y_1, \ldots, y_N, y^*)$ we can write:

$$\begin{pmatrix} \mathbf{y} \\ y^* \end{pmatrix} \sim \mathcal{N}(0, \mathbf{K}_{N+1}) \quad (4)$$

with covariance matrix

$$\mathbf{K}_{N+1} = \begin{bmatrix} \begin{bmatrix} \mathbf{K} \end{bmatrix} & \begin{bmatrix} \mathbf{k}(\mathbf{x}^*) \end{bmatrix} \\ \begin{bmatrix} \mathbf{k}(\mathbf{x}^*)^T \end{bmatrix} & \begin{bmatrix} k(\mathbf{x}^*) \end{bmatrix} \end{bmatrix} \quad (5)$$

where $\mathbf{y} = [y_1, \ldots, y_n]^T$ is $N \times 1$ vector of training targets. We can divide this joint probability 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 $\mathbf{X}$ the $N \times D$ matrix of training inputs.

We need to estimate the unknown parameters of the covariance function (3), as well as the noise variance v_0. This is done via maximization of the log-likelihood

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Theta}) &= \log(p(\mathbf{y}|\mathbf{X})) \\ &= -\frac{1}{2}\log(|\mathbf{K}|) - \frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{N}{2}\log(2\pi) \end{aligned} \quad (6)$$

with vector of hyperparameters $\boldsymbol{\Theta} = [w_1 \ldots w_D, \ v, \ v_0]^T$ 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:

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial \Theta_i} = -\frac{1}{2}\text{trace}\left(\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \Theta_i}\right) + \frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \Theta_i}\mathbf{K}^{-1}\mathbf{y} \quad (7)$$

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 prior on the parameters and compute their posterior probability [6].

Now that the hyperparameters are optimized, we can obtain prediction of the GP model at the input $\mathbf{x}^*$. The conditional part of (4) provides the predictive distribution of y^*:

$$p(y^*|\mathbf{y}, \mathbf{X}, \mathbf{x}^*) = \frac{p(\mathbf{y}, y^*)}{p(\mathbf{y}|\mathbf{X})} \quad (8)$$

It can be shown [14] that this distribution is Gaussian with mean and variance:

$$\begin{aligned} \mu(\mathbf{x}^*) &= \mathbf{k}(\mathbf{x}^*)^T\mathbf{K}^{-1}\mathbf{y} \quad (9) \\ \sigma^2(\mathbf{x}^*) &= k(\mathbf{x}^*) - \mathbf{k}(\mathbf{x}^*)^T\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}^*) + v_0 \quad (10) \end{aligned}$$

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

Vector $\mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1}$ in (9) can be interpreted as a vector of smoothing terms which weights training outputs $\mathbf{y}$ to make a prediction at the test point $\mathbf{x}^*$. If the new input is far away from the data points, the term $\mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}^*)$ in (10) will be small, so that the predicted variance $\sigma^2(\mathbf{x}^*)$ will be large. Regions of the input space, where there are few data or where the data has high complexity or noise, are in this way indicated through higher variance.

B. Dynamical system identification

The identification method above was originally used for identification of static functions, but it can be extended to model dynamical systems as well [8].

Our task is to model the dynamical system (1) and be able to make n-step ahead prediction. One way to do n-step ahead prediction is to make iterative one-step ahead predictions up to desired step n whilst feeding back the predicted output.

Two approaches to iterated one-step ahead prediction are possible using GP model. In the first approach only the mean values of the predicted output are feed back to input. Vector of regressors $\mathbf{x}(k)$ at time sample k from (2) for GP model is composed from L past predicted outputs and L past inputs:

$$\mathbf{x}(k) = [\hat{y}(k-1), \ldots, \hat{y}(k-L), u(k-1)), \ldots, u(k-L)]^T \tag{11}$$

where $\hat{y}(k-i)$ and $u(k-i)$ are past prediction $\hat{y}$ and input u at time sample $(k-i)$ respectively.

This "naive" approach neglects information about the uncertainty of the output and is similar to that used in modelling dynamic systems with NNs. The obtained variance is still a rough indicator of the model's local one-step ahead accuracy, but the values of predicted mean and variance are not correct for the n-step ahead prediction.

C. Propagation of the uncertainty through the GP model

In second approach to n-step ahead prediction not only the mean value of the predicted output is feed back to input of the GP model, but complete output distribution of the GP model is feed back instead [16], [12]. This way GP model not only models dynamical behaviour of the system but also provides the information about confidence in its prediction. Two realizations of this approach are possible:

- numerical integration over input distribution, e.g. with MCMC method and
- analytical solution, where assumption about output having Gaussian probability distribution, is made [12], [11].

Latter approach, referred to as "exact" approach, is depicted in Fig. 1. Output distribution $N(m(k), v(k))$ with mean value $m(.)$ and variance $v(.)$ of the model in time step k is presumed to be Gaussian. Note that distribution in Fig. 1 is denoted with N and not $\mathcal{N}$, as it is Gaussian only by presumption.

In "exact" approach input to GP model vector of past outputs $\mathbf{x}_{\hat{y}}(k) = [\hat{y}(k-1), \ldots, \hat{y}(k-L)]^T$ is presented as:

$$\mathbf{x}_{\hat{y}}(k) \sim \mathcal{N}(\boldsymbol{\mu}_{\hat{y}}, \boldsymbol{\Sigma}_{\hat{y}}) \tag{12}$$

where the vector of mean past predictions $\boldsymbol{\mu}_{\hat{y}}$ is

$$\boldsymbol{\mu}_{\hat{y}} = [m(\mathbf{x}_{\hat{y}}(k-1)) \ldots m(\mathbf{x}_{\hat{y}}(k-L))]^T \tag{13}$$

and elements $\Sigma_{i,j}$ of covariance matrix $\boldsymbol{\Sigma}_{\hat{y}}$ between past predicted outputs are

$$\Sigma_{i,j} = \begin{cases} v(\mathbf{x}_{\hat{y}}(k-i)) + v_0, & i = j \\ \text{cov}(\hat{y}(k-i), \hat{y}(k-j)) & \text{otherwise.} \end{cases} \tag{14}$$
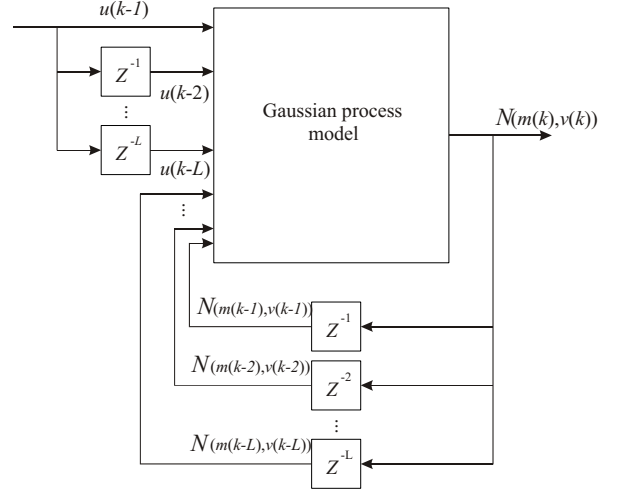

Fig. 1. Simulation with repeated one step-ahead prediction of dynamical model where uncertainty is propagated

The regressor $\mathbf{x}$ from (11) is completed with vector $\mathbf{x}_u = [u(k-1)), \ldots, u(k-L)]^T$, where $u(k-i)$ is the value of the input at time sample $k-i$. As they are known precisely, we can view on each of these values as on distribution with variance $\sigma^2(u(k-i)) = 0$.

Thus the input of the GP model $\mathbf{x}$, at which we wish to predict the output, is a normally distributed random variable with mean vector $\boldsymbol{\mu}_{\mathbf{x}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$:

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\hat{y}} \\ \mathbf{x}_u \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\hat{y}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right) \tag{15}$$

New output $\hat{y}(k) \sim \mathcal{N}(m(\mathbf{x}(k)), v(\mathbf{x}(k)) + v_0)$ at time sample k is computed using (16) and (17), see [11], [12] for details.

$$m(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) = E_{\mathbf{x}}[\mu(\mathbf{x})] \tag{16}$$

$$v(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) = E_{\mathbf{x}}[\sigma^2(\mathbf{x})] + E_{\mathbf{x}}[\mu(\mathbf{x})^2] - (E_{\mathbf{x}}[\mu(\mathbf{x})])^2 \tag{17}$$

where $E_{\mathbf{x}}[f(\mathbf{x})] = \int_{-\infty}^{+\infty} f(\mathbf{x}) \, p(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) \, d\mathbf{x}$ and the expressions for mean $\mu(\mathbf{x})$ and variance $\sigma(\mathbf{x})$ are from (9) and (10) respectively.

Due to space limitation only final expressions for mean and variance are given, refer to [12] for detailed calculations. Predicted mean $m(.)$ and predicted variance $v(.)$ at time sample k for normally distributed random variable $\mathbf{x}$ with mean vector $\boldsymbol{\mu}_{\mathbf{x}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$ are:

$$m(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) = \sum_{i=1}^{N} \beta_i \, C(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{x}_i) \, C_{corr1}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{x}_i) \tag{18}$$

and

$$v(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) = v - m(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}})^2 - \\ - \sum_{i,j=1}^{N} (K_{ij}^{-1} - \beta_i \beta_j) \, C(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{x}_i) C(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{x}_j) C_{corr2}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{x}_b) \tag{19}$$

where

$$C_{corr1}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{x}_i) = |\mathbf{I} + \mathbf{W}^{-1}\boldsymbol{\Sigma}_{\mathbf{x}}|^{-1/2} \\ \exp\left[\frac{1}{2}(\boldsymbol{\mu}_{\mathbf{x}} - \mathbf{x}_i)^T \boldsymbol{\Delta}^{-1}(\boldsymbol{\mu}_{\mathbf{x}} - \mathbf{x}_i)\right] \tag{20}$$

$$\boldsymbol{\Delta}^{-1} = \mathbf{W}^{-1} - (\mathbf{W} + \boldsymbol{\Sigma_x})^{-1} \qquad (21)$$

$$C_{corr2}(\boldsymbol{\mu_x}, \mathbf{x}_b) = \left| 2\mathbf{W}^{-1}\boldsymbol{\Sigma_x} + \mathbf{I} \right|$$
$$\exp\left[\frac{1}{2}(\boldsymbol{\mu_x} - \mathbf{x}_b)^T \boldsymbol{\Lambda}^{-1}(\boldsymbol{\mu_x} - \mathbf{x}_b) \right] \quad (22)$$

$$\boldsymbol{\Lambda}^{-1} = \left(\frac{1}{2}\mathbf{W} \right)^{-1} - \left(\frac{1}{2}\mathbf{W} + \boldsymbol{\Sigma_x} \right)^{-1} \qquad (23)$$

and $\mathbf{x}_i$ and $\mathbf{x}_j$ are i-th and j-th input regressors with their mean $\mathbf{x}_b = \frac{1}{2}(\mathbf{x}_i + \mathbf{x}_j)$, $\boldsymbol{\beta} = \mathbf{K}^{-1}\mathbf{y}$, $\mathbf{I}$ is $D \times D$ identity matrix and $\mathbf{W}^{-1}$ is diagonal matrix of hyperparameters w_i: $\mathbf{W}^{-1} = \mathrm{diag}[w_1, \dots, w_D]$.

IV. IDENTIFICATION OF A TWO TANK SYSTEM

A. Laboratory pilot plant and chosen subsection for identification

The process scheme of laboratory pilot plant's subsystem is presented in Fig. 2. Subsystem consist of two tanks, $R1$ and $R2$, connected with flow paths, which serve to supply liquid from the reservoir $R0$.
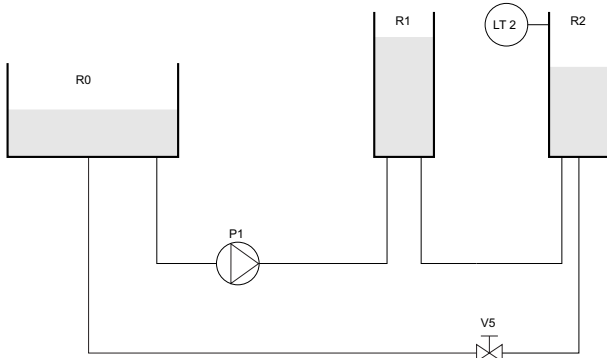

Fig. 2.   Process scheme of chosen subsystem of the plant

Flow path from reservoir $R0$ to tank $R1$ has built-in pump $P1$, driven with 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.

Flow is generated by varying the angular speed of the pump $P1$. The other interesting part is 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 $R0$. Capacity of the reservoir $R0$ is much greater than 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 $P1$. 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.

B. Operating region, measurements and results

The represented second order system is 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. Static characteristic of the system's response is given in Fig. 3. Besides the static nonlinearity, observed in Fig. 3, the system's dynamics is nonlinear too.
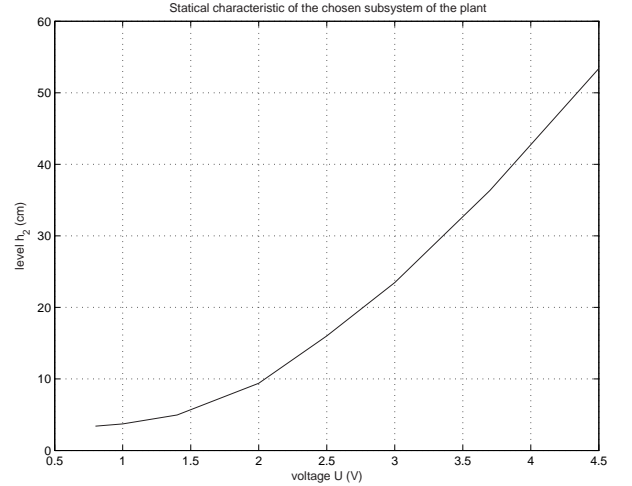

Fig. 3.   Static characteristic of the chosen subsystem of the plant

Working region is restricted with height $h_{max} = 60$ centimeters of the tanks $R1$ and $R2$. The maximum voltage on $P1$ was fixed to $U_{max} = 4$ Volts with trial and error method, preventing the liquid level h_1 to reach the top h_{max} of the tank $R1$. Sample time $T_s = 10$ seconds was chosen experimentally, so that dynamics of the system was satisfactory captured.

Identification and validation data for GP model was obtained with measurements. Pseudo–random binary signal (PRBS) was used for input, except that the value of the magnitude of the signal could occupy any value between $U_{min} = 0.8$ and $U_{max} = 4$ Volts when changed. Two different signals were used, one for training of the model and one for model validation. The signal for training is depicted in Fig. 4 and the signal for validation in Fig. 5.

The second order model was assumed as the underlying system is of second order too. Training points were sampled arbitrary from training signal. 92 training points were used for training of the GP model. The number of training points was selected as a tradeoff between quality of prediction and training time. As this is nonparametric model, where all training inputs are constituting the model via covariance matrix $\mathbf{K}$, merely increasing the number of training points would not necessary substantially improve the prediction of the model, as is the case in our example.

With the training the following values were gained for hyperparameters: $v = 218$, $v_0 = 0.03$, $\mathbf{w} = [2.0\mathrm{E}-4,\ 2.5\mathrm{E}-3,\ 0.155,\ 1.28\mathrm{E}-2]^T$.

The result of the simulation of the GP model without propagation of variance — the "naive" method — can be seen in Fig. 6 and the result of simulation with propagation of the variance — the "exact" method — in Fig. 7. From Fig. 6 we can deduce, that the mean value of the model's
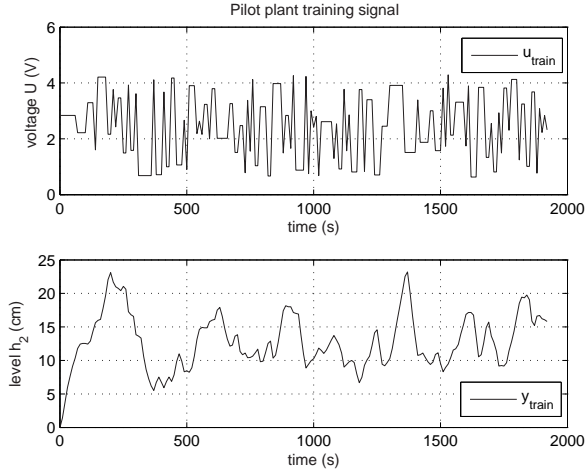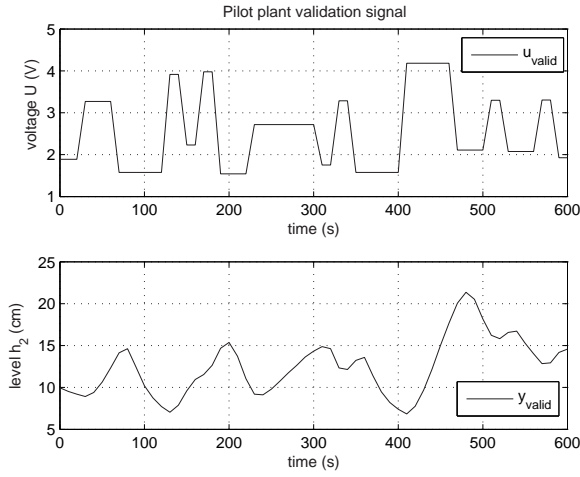
Fig. 4. Training signal for LMGP model of the plant

Fig. 5. Validation signal for LMGP model of the plant

Fig. 6. Validation of the GP model with simulation – "naive" method
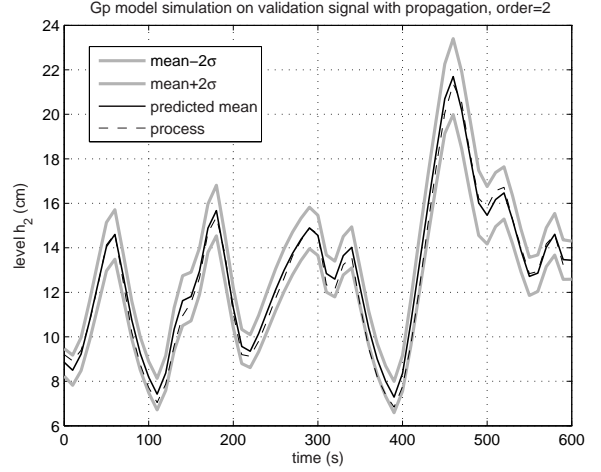
Fig. 7. Validation of the GP model with simulation – "exact" method

output corresponds quite well to target. The variance on the other hand can not be used as the proper measure of confidence, as it is not taking the uncertainty of the input into account, which results in too firm confidence in prediction. That is not the case when "exact" method is used. The confidence in the prediction loosens as the uncertainty of previous output is taken into consideration. This is particulary highlighted in the regions, modelled with smaller number of data, as can be seen from Fig. 8 in time interval from 400 to 550 seconds. We can observe, that during most of the simulation system's output y lies within the 95% measure of confidence interval 2σ around the predicted output of the GP model $\hat{y}$: $|\hat{y}(t) - y(t)| < 2\sigma(t)$, see Fig. 8.

Two quality measures [8] were used for results of validation:

- mean squared error (SE)
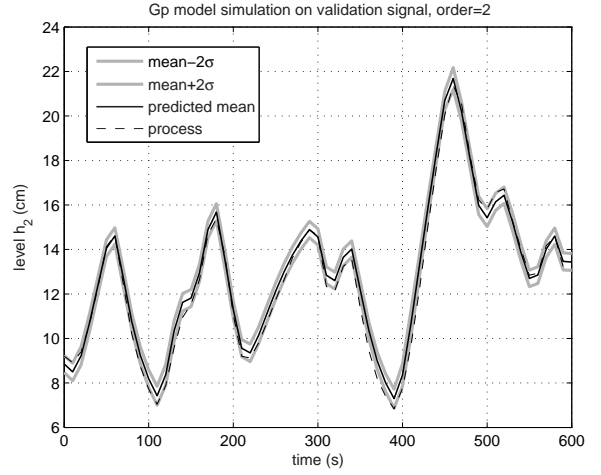
$$\text{SE} = \frac{1}{N} \sum_{i=1}^{N} e_i^2 \qquad (24)$$

and

- log-predictive density error (LD) [12]

$$\text{LD} = \frac{1}{2N} \sum_{i=1}^{N} \left(\log(2\pi) + \log(\sigma_i^2) + \frac{e_i^2}{\sigma_i^2} \right) \qquad (25)$$

where $e_i = \hat{y}_i - y_i$ is error of model's prediction and σ_i^2 predicted variance of the model in i-th time step.

Their evaluation for both simulation methods is given in Table I. It can be seen from comparison of SE, that the mean predicted value does not improve much with the use of "exact" method. More important is, that the variance can be used as the measure of confidence when using "exact" method, which can be seen from comparison of LD values in Table I and simulation results in Figs. 6, 7 and 8.

TABLE I
VALUES OF QUALITY MEASURES ON GP MODEL SIMULATION RESULTS

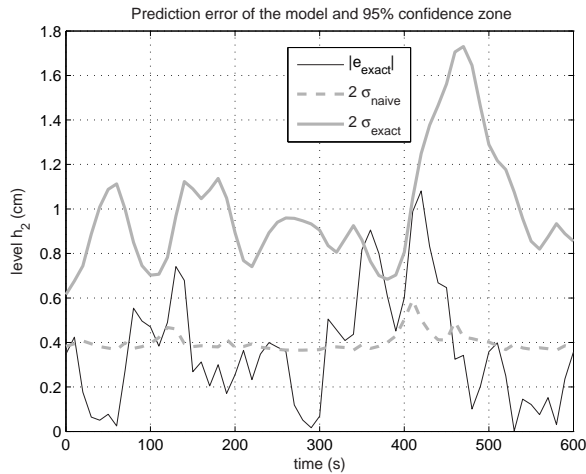| method	"naive"	"exact"
SE	0.2039	0.2019
LD	1.66	0.65

Fig. 8. Absolute error of model with propagation compared to 95% confidence values of "naive" and "exact" method

V. Conclusions

In this paper the identification of dynamical systems with GP model is presented and used on two-tank system example. The identified GP model was validated with simulation. Two simulation methods were used — "naive" method without and "exact" method with propagation of variance.

The main advantage of GP model over NN or fuzzy models is, that its output is predicted Gaussian distribution instead of predicted value. Gaussian distribution is defined with its mean and variance and in terms of GP model we can use the mean as prediction value and variance as the measure of confidence in prediction. It was shown that this uncertainty measure can be effectively used with "exact" realization to improve simulation results of the GP model. Second advantage of GP model over neural networks or fuzzy logic models is the smaller number of parameters that need to be optimized, as this reduces the problem of local minima.

Another potential benefit of GP model is possibility to include prior knowledge in elegant matter. Disadvantage of the GP model is high computational burden, associated with training of hyperparameters and propagation of the uncertainty.

In this paper two-tank system was successfully identified with nonparametric GP model. Only six hyperparameters of the model needed to be optimized. The model was validated with simulation, where gained uncertainty was effectively used to obtain confidence regions around predicted output.

Ongoing activities include incorporation of prior knowledge into GP model and improvement of the algorithms as well as various applications of GP models.

References

[1] L. Ljung, *System Identification – Theory for the User,* 2nd Edn., Prentice Hall, New Jersey, 1999.
[2] J. Kocijan, R. Murray-Smith, C.E. Rasmussen and A. Girard, "Gasussian process model based predictive control", Proc.: American Control Conference, Boston, pp. 2214-2219, 2004.
[3] E. Solak, R. Murray-Smith, W.E. Leithead, D.J. Leith and C.E. Rassmusen, "Derivative observations in Gaussian Process Models of Dynamic Systems", S. Becker, S. Thrun and K. Obermayer (Eds) *Advances in Neural Information Processing Systems 15*, MIT Press, pp. 529–536, 2003.
[4] J. Kocijan and D.J. Leith, "Derivative Observations used in Predictive Control". Proc.: Melecon 2004, Dubrovnik, 2004.
[5] A. O'Hagan, "On curve fitting and optimal design for regression (with discussion)", *Journal of the Royal Statistical Society B*, vol. 40, pp.1-42, 1978.
[6] C.E. Rasmussen, *Evaluation of Gaussian Processes and Other Methods for NonLinear Regresion*, PhD thesis, University of Toronto, Toronto, 1996.
[7] R.M. Neal, *Bayesian learning for neural networks*, Springer Verlag, New York, 1996.
[8] J. Kocijan, A. Girard, B. Banko and R. Murray-Smith, "Dynamic Systems Identification with Gaussian Processes", Proc.: 4th Mathmod, Wien, 2003, extended version in *Mathematical and Computer Modelling of Dynamic Systems* (in press).
[9] J. Kocijan, B. Banko, B. Likar, A. Girard, R. Murray-Smith and C.E. Rasmussen, "A case based comparison of identification with neural networks and Gaussian process models", Proc.: IFAC Intelligent Control Systems and Signal Processing Conference, Faro, pp. 137–142, 2003.
[10] G. Gregorčič and G. Lightbody: "Gaussian Processes for Modelling of Dynamic Non-linear Systems", Proc.: Irish Signals and Systems Conference, Cork, 2002.
[11] A. Girard, C.E. Rasmussen, J. Quiñonero Candela and R. Murray-Smith, "Multi-step ahead prediction for nonlinear dynamic sytems - A Gaussian Process treatment with propagation of the uncertainty", S. Becker, S. Thrun and K. Obermayer (Eds) *Advances in Neural Information Processing Systems 15*, MIT Press, pp. 1033–1040, 2003.
[12] A. Girard, *Approximate Methods for Propagation of Uncertainty with Gaussian Process Models*, Ph.D. Thesis, University of Glasgow, Glasgow, 2004.
[13] C.K.I. Williams, "Prediction with Gaussina processes: from linear regression and beyond", M.I. Jordan (Ed) *Learning and Inference in Graphical models*, Kluwer Academic Press, pp. 599–621, 1998.
[14] M.N. Gibbs, *Bayesian Gaussian Processes for Regression and Classification*, Ph.D. Thesis, Cambridge University, Cambridge, 1997.
[15] D.J. Mackay, "Gaussian Processes", *Information theory, inference and learning algorithms*, Cambridge University Press, 2003.
[16] J. Kocijan and A. Girard, "Incorporated linear local models in Gaussian process model", accepted for presentation at 16th IFAC World Congress, Prague, 2005.