Key words - nonlinear systems, model identification, Gaussian process model

Juš Kocijan[1,2], Kristjan Ažman[1]

GAUSSIAN PROCESS MODEL IDENTIFICATION: A PROCESS ENGINEERING CASE STUDY

Gaussian process prior models, although well known in the Bayesian statistics community, are a relatively new approach for modelling of dynamic systems and consequently a novelty in systems and control community. The Gaussian process prior model is a probabilistic nonparametric model. The complexity of such model depends on the amount of input data used for identification and contained in the model. This complexity adds to the computational load necessary for multiple-step-ahead prediction and for model simulation. The combination of local linear models and sparse data in off-equilibrium regions can be utilised to reduce the amount of data in Gaussian process model. This approach is used in the paper for the modelling of a gas-liquid separator process plant, where substantial amount of measurement data in equilibrium region and sparse measurement data in distance from equilibrium region are combined in one Gaussian process model.

1. INTRODUCTION

Owing to operating and safety constraints, the available measured data from which we are often required to construct an empirical model of process plant is often concentrated mainly around equilibrium points with only relatively sparse data measured far from them. A common approach in this situation is to build local models using the data in the vicinity of equilibrium points and then blend these models so as to obtain a nonlinear model covering the operating envelope, e.g. [3],[6]. In paper [6] the use of Gaussian processes (GP) for modelling dynamic systems has been proposed as an alternative approach. Further developments of this probabilistic, nonparametric approach for modelling dynamic systems led to incorporation of local models, e.g [4] and to the local model network version of GP modelling [2].

Gaussian process prior model [7] is a probabilistic, non-parametric model used mainly in the Bayesian statistics community. It can be used to tackle many of the standard problems usually solved by artificial neural networks and is applied on a range of static nonlinear regression problems. However, it is only until recently that they have been also used for dynamic systems modelling and in the context of iterative multiple-step ahead prediction, e.g. [5]. Gaussian process model describes both the dynamic characteristics of nonlinear system, and at the same time provides information about the

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

confidence in its predictions. The model can highlight the areas of the input space where prediction quality is poor due to the lack of data, by indicating the higher variance around the predicted mean.

A key issue in nonparametric GP models is that in their simplest form the computational burden is relatively large. The computational burden is associated with training data covariance matrix inversion and can be reduced by employing approximate inverses. An alternative approach considered in this paper is to summarise the measured data in the vicinity of the equilibrium points into the local models which are then incorporated into the GP model. This not only accords well with engineering practice but has the potential to directly reduce the computational burden.

Despite the current challenges, such as the computational burden associated with the number of training data, the Gaussian process approach has a number of exciting advantages. The simple model structure, reduced problems associated with the choice of the model structure and the uncertainty information on the predictions are attractions of the Gaussian process approach. The model also offers the possibility to incorporate various kinds of prior knowledge into the model, e.g. linear local models, statical characteristic, hysteresis, etc. Furthermore, Gaussian Process models could be used for investigating the potential benefits of a nonparametric approach to modelling.

The purpose of this paper is to present an identification case study with measurements from a gas-liquid separator pilot plant using Gaussian Process model with incorporated linear models (LMGP model). The study will show that the use of this model:

- solves issues of unbalanced distribution of data points for modelling,
- reduces the large computational burden due to the large amount of contained data in ordinary Gaussian Process model and
- provides not only the information about plant behaviour but also information about the model prediction confidence.

2. MODELLING WITH GAUSSIAN PROCESSES

2.1. GAUSSIAN PROCESS MODEL

A detailed presentation of Gaussian processes can be found, e.g., in [7]. The Gaussian process is a collection of random variables with a joint multivariate Gaussian distribution, $f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, and is fully described by the mean and covariance matrix. For simplicity, a zero-mean process is assumed. The elements Σ_{ij} of the covariance matrix $\mathbf{\Sigma}$ are covariances between the values of the function $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ and are functions of the corresponding arguments $\mathbf{x}_i$ and $\mathbf{x}_j$: $\Sigma_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$. 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 Gaussian process prior directly over functions instead of parameterising $f(\mathbf{x})$. In the following, a stationary process is assumed, where the stationarity assumption implies that the covariance between two points depends only on the distance between them and is invariant with a translation in the input space. Then a common choice of covariance function is the squared exponential or Gaussian function:

$$\text{cov}[f(\mathbf{x}_p), f(\mathbf{x}_q)] \;=\; C(\mathbf{x}_p, \mathbf{x}_q) \;= v_1 \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d (x_p^d - x_q^d)^2\right] \tag{1}$$

where x_p^d denotes the d^{th} component of the D-dimensional input vector $\mathbf{x}_p$, and $v_1, w_1, \ldots, w_D$ are free parameters called hyperparameters in the GP model framework. This covariance function is such that the points that are close together in the input space lead to more correlated outputs than the points further apart (a smoothness assumption). The parameter v_1 controls the vertical scale of the variation, 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 the variance v_0, $\epsilon \sim \mathcal{N}(0, v_0)$, and put a Gaussian process prior with the covariance function (1) and the unknown parameters on $f(.)$. Within this probabilistic framework, we can write $y_1, \ldots, y_N \sim \mathcal{N}(0, \mathbf{K}_N)$, with $\mathbf{K} = \mathbf{\Sigma} + v_0\mathbf{I}$ where $\mathbf{I}$ is an $N \times N$ identity matrix. Making a prediction with this model means that we are searching for the predictive distribution y_{N+1} corresponding to a new given input $\mathbf{x}_{N+1}$ based on a set of N training data pairs $\mathbf{x}_i, y_i{}_{i=1}^N$.

If we split $y_1, \ldots, y_{N+1}$ into two parts, $\mathbf{y} = [y_1, \ldots, y_N]$ and y^*, we can write

$$\mathbf{y}, y^* \sim \mathcal{N}(0, \mathbf{K}_{N+1}) \quad \text{with} \quad \mathbf{K}_{N+1} = \begin{bmatrix} \begin{bmatrix} & \mathbf{K} & \end{bmatrix} & \begin{bmatrix} \mathbf{k}(\mathbf{x}^*) \end{bmatrix} \\ \begin{bmatrix} \mathbf{k}(\mathbf{x}^*)^T \end{bmatrix} & \begin{bmatrix} \kappa(\mathbf{x}^*) \end{bmatrix} \end{bmatrix} \tag{2}$$

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 the joint probability (2) can be divided 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 the training targets and $\mathbf{X}$ is the $N \times D$ matrix of the corresponding training inputs. The unknown hyperparameters of the covariance function and the noise variance v_0 can then be estimated via the maximization of the log-likelihood. The conditional part of (2) provides 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 the mean and variance

$$\mu(\mathbf{x}^*) = \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{y} \tag{3}$$

$$\sigma^2(\mathbf{x}^*) = \kappa(\mathbf{x}^*) - \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}^*) \tag{4}$$

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

The above modelling procedure was developed for modelling static nonlinearities, but it can be readily applied for modelling dynamic systems, as shown in [5]. Consider a dynamic system in the ARX representation, where the output at time step k depends on the delayed outputs and the exogenous control inputs:

$$y(k) = f(y(k-1), \ldots, y(k-L), u(k-1), \ldots, u(k-L)) + \epsilon(k) \tag{5}$$

where $\epsilon(k)$ is white noise and the output $y(k)$ depends on the state vector $\mathbf{x}(k) = [y(k-1), y(k-2), \ldots, y(k-L), u(k-1), u(k-2), \ldots, u(k-L)]^T$ at time step k.

Multiple-step-ahead predictions of a system modelled by (5) can be achieved by iteratively making repeated one-step-ahead predictions, up to the desired horizon. A naive way of doing so is, at each time-step, to feed back the mean of the predictive distribution (the estimate of the output) by considering $\mathbf{x}(k+n) = [\hat{y}(k+n-1), \ldots, \hat{y}(k+n-L), u(k+n-1), \ldots, u(k+n-L)]^T$, where $\hat{y}(k+n-i)$ is the point estimate of $y(k+n-i)$. Although this approach, also called output error approach, is approximate, as the variance of the lagged outputs on the right-hand side of equation (5) is neglected, it is commonly used when modelling dynamic systems with neural networks or fuzzy models. Another possibility for multiple-step-ahead prediction is presented in [1] where the Monte Carlo simulation is used.

2.2. INCORPORATION OF DERIVATIVE OBSERVATIONS

In this paper we are only summarising the main modelling results of various authors, because the focus of the paper is the modelling case study. The topic is explained in e.g. [1],[4].

The main idea behind the incorporation of derivative observations is to join two different sorts of data in a single model. This comes in very useful for dynamic systems modelling, as we will see later. Since differentiation is a linear operation, the derivative of a GP remains a GP. This enables the Gaussian process modelling framework to be extended to include observed derivatives of the function as well as (or instead of) the values of the function itself. The output (target) vector $\mathbf{y}$, which before consisted solely of output measurements, now also contains the derivatives of regressors. The corresponding inputs remain the values of the regressors associated with each function and derivative observation.

If the function derivatives are treated in the same way and in the same set of training pairs as function observations, then due to the different nature of the data one must appropriately change the covariance function (which changes the covariance matrix), reflecting the covariance between the data. In the case of the Gaussian covariance function (1), the covariance between the functional and derivative data becomes:

$$\mathrm{cov}[\frac{\partial y_p}{\partial x_p^d}, y_q] = -v_1 w_d (x_p^d - x_q^d) \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d (x_p^d - x_q^d)^2\right] \tag{6}$$

and in the case of two derivative data the covariance function becomes:

$$\mathrm{cov}[\frac{\partial y_p}{\partial x_p^d}, \frac{\partial y_q}{\partial x_q^e}] = v_1 w_e (\delta_{d,e} - w_d (x_p^d - x_q^d)(x_p^e - x_q^e) \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d (x_p^d - x_q^d)^2\right], \tag{7}$$

where $\frac{\partial y_p}{\partial x_p^d}$ denotes the first derivative of y_p in the direction of the d^{th} component of the D-dimensional input vector $\mathbf{x}_p$.

When the covariance matrix $\mathbf{K}$ and the vector $\mathbf{k}^*$ are expanded with covariance functions for derivative observations (6) and (7), the predictive distribution of a function output corresponding to a new $\mathbf{x}$ has a mean and variance that can be calculated by equations (3) and (4), as is the case when using only one sort of data, i.e., function observations.

Derivative observations around a selected equilibrium point can, for dynamic systems, be interpreted as the parameters of the local linear model about this particular point and can be synthesized using standard linear regression. With such a synthesis the information carried by (many) training

points in the vicinity of equilibrium points can be reduced to the local model, thereby effectively reducing the number of data points in the model for computational purposes.

The data may or may not contain information about the noise. For function observations, assuming white noise, the noise information is added to the diagonal elements of the covariance matrix corresponding to these points. If no information is available, the noise variance v_0 is learned as in the case without derivative observations. When standard identification methods are used the noise for every derivative observation is also obtained. The acquired covariance matrices of each identified linear local model are added to the overall covariance matrix for the corresponding derivative component.

3. CASE STUDY

3.1. PROCESS PLANT DESCRIPTION

The semi-industrial process plant used for the case study in the paper is the unit for separating the gas from liquid that forms part of a larger pilot plant positioned at Department of systems and control, Jozef Stefan Institute. The screen shot from SCADA presenting the scheme of this plant is given in Figure 1.

The role of the separation unit is to capture flue gases under low pressure from the effluent channels by means of water flow, to cool them down and then supply them under high-enough pressure to other parts of the pilot plant.

The flue gases coming from the effluent channels are absorbed by the water flow into the water circulation pipe through injector.

The water flow is generated by the water ring pump. The speed of the pump is kept constant. The pump feeds the mixture of water and gas into the tank, where gas is separated from water. Hence the accumulated gas in tank forms a sort of 'gas cushion' with increased internal pressure. Owing to this pressure, the flue gas is blown out from tank into the neutralization unit. On the other side, the 'cushion' forces water to circulate back to the reservoir. The quantity of water in the circuit is constant.

In order to understand the basic relations among variables and to illustrate the nonlinearity of the process a mathematical model is introduced. The gas-liquid separation pressure sub-system of interest can be described by a set of two equations.

$$\begin{aligned}
\frac{dp_1}{dt} &= \frac{1}{S_1(h_{T_1}-h_1)}(p_0(\alpha_0+\alpha_1 p_1+\alpha_2 p_1^2-k_1 R_1^{u_1-1}\sqrt{p_1}) \\
&+ (p_0+p_1)(\Phi_w-k_2 R_2^{u_2-1}\sqrt{p_1+k_w(h_1-h_{T_2})})), \\
\frac{dh_1}{dt} &= \frac{1}{S_1}(\Phi_w-k_2 R_2^{u_2-1}\sqrt{p_1+k_w(h_1-h_{T_2})}),
\end{aligned} \tag{8}$$

where u_i is the command signal of valve Vi, $i=1,2$, where V1 is the valve on output from tank R4.1 to other subprocess and V2 is the valve between tanks R4.2 and R4.1, h_i is the level in tank R4.i, $i=1,2$, p_1 is the relative air pressure in tank R4.1, S_i is the section area of tank R4.i, p_0 is atmospheric pressure, h_{Ti} is height of tank R4.i, $i=1,2$, R_i is the open-close flow ratio of valve Vi, $i=1,2$, k_i is the flow coefficient of valve Vi, $i=1,2$, Φ_w is the known constant water flow through pump, $\alpha_i; i=1,2,3$ are constant parameters.

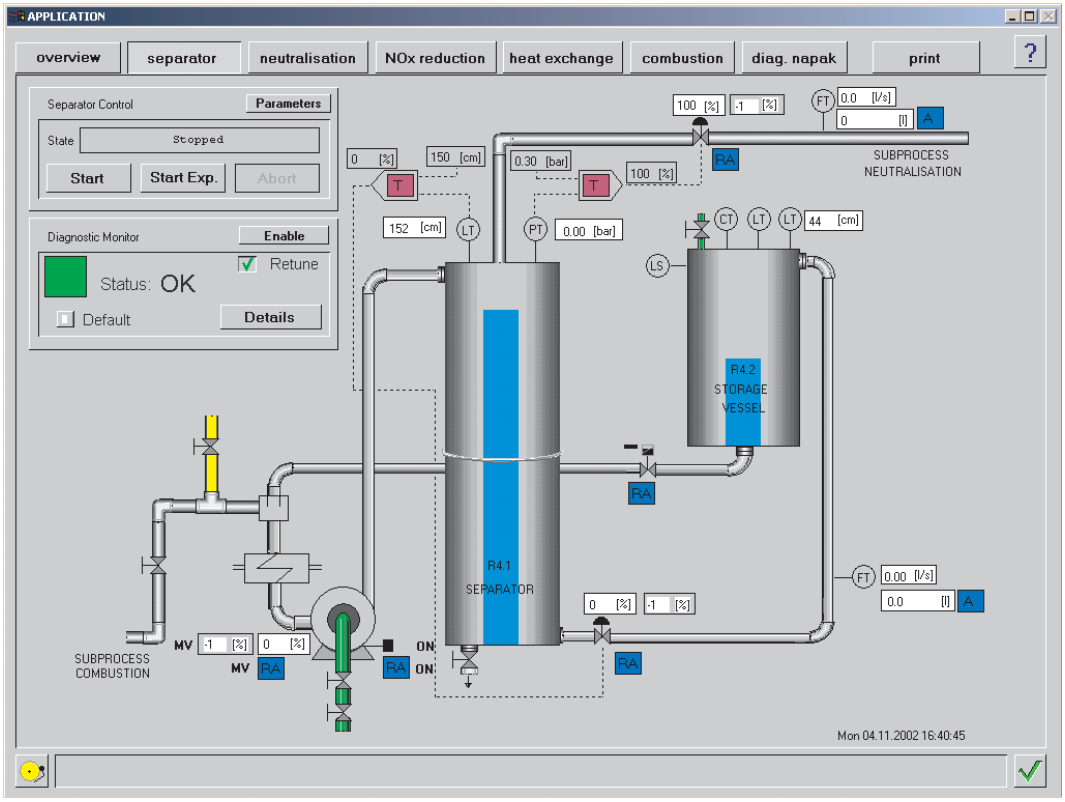

Fig. 1: SCADA scheme of the gas-liquid separator

From the model presented, it can be seen that the nonlinear process is of a multivariable nature (two inputs and two outputs with dynamic interactions between the channels). In our case a level feedback control was implemented. Consequently the dynamic system could be approached as a single-input single-output dynamic system with the command signal of valve V_1 as the input and the pressure in tank R4.1 as the output. As can be seen from Eqs. (8) pressure is nonlinearly related to level and input flow which results in different dynamic behaviour depending on the operating region.

User-friendly experimentation with the process plant is enabled through interface with the Matlab/Simulink environment. This interface enables PLC access with the Matlab/Simulink using OPC protocol via TCP/IPv4 over Ethernet IEEE802.3. Control algorithms for experimentation can be prepared in Matlab code or as Simulink blocks and extended with functions/blocks, which access PLC. In our case all schemes for data acquisition were put together as Simulink blocks.

3.2. IDENTIFICATION AND VALIDATION

One of the reasons for difficult identifiability of nonlinear dynamic systems in general, and this system is no exception, is in the distribution of data used for model identification. Except for the case when the system input is a random signal with magnitudes uniformly and densely distributed over the entire region — which is very unrealistic input — data is distributed so that the majority of the samples are concentrated close to the equilibrium curve and only sparse data is measured on a distance from it.

Gaussian process model is proposed as a possible solution that handles this problem caused by unbalanced distribution of the data well due to the possibility of joining different sorts of data.

Since the process to be identified is characterised as predominantly the first order system, a model of the form (9) is identified

$$p_1(k+1) = f(p_1(k), u_1(k), h_1(k)), \tag{9}$$

which means that pressure $p_1(k)$, valve signal $u_1(k)$ and liquid level $h_1(k)$ are selected as regressors. The mean value of the predicted pressure $p_1(k)$ is feed back, and the predicted mean and variance are calculated as described in the previous section. Attempts have been made to identify the system with a higher order model, but the results were not better.

Based on the response and iterative cut-and-try procedure, a sampling time of 15 seconds is selected. Seven equilibrium points at seven different liquid levels spanning the operating region of interest were selected. At each equilibrium point a small-scale random binary signal with zero mean and magnitude 0.02 was applied. A linear approximation to the local dynamics at each selected equilibrium point was obtained with linear systems identification of ARX model. In addition to this equilibrium information, a small, sparse set of off-equilibrium input-output data consisting of 60 data points was selected. Larger numbers of off-equilibrium observations were also studied and this number was chosen as a compromise between the accuracy of achieved fit and number of data points used. These data points were taken from the input signal given in Figure 2. They were selected based on the Euclidean distance from equilibria with equation (10).

$$(p_1(k) - p_1(k-1))^2 + (u_1(k) - u_1(k-1))^2 + (h_1(k) - h_1(k-1))^2 > 1.6 \cdot 10^{-3} \tag{10}$$

To summarise, following information was used to train the LMGP model:

- seven equilibrium input-output values spanning the operating region of interest,
- the set of coefficients of the linear models representing partial derivatives of the output in the selected equilibrium points,
- sixty input-output values sampled from off-equilibrium points.

The response of the LMGP model on the validation input signal together with the process response is given in Figure 3. The corresponding standard deviation and the level of liquid which predominantly characterises the nonlinear dynamics are given in Figure 4. The goodness of fit of both models was also compared by computing following two cost functions:

- average squared test error

$$SE = \frac{1}{N}\sum(\hat{y}_i - y_i)^2 = 5.6033 \cdot 10^{-4} \tag{11}$$

and

- log-predictive density error

$$LD = \frac{1}{2N}\sum(\log(2\pi) + \log(\sigma_i^2) + \frac{(\hat{y}_i - y_i)^2}{\sigma_i^2}) = -1.97 \tag{12}$$

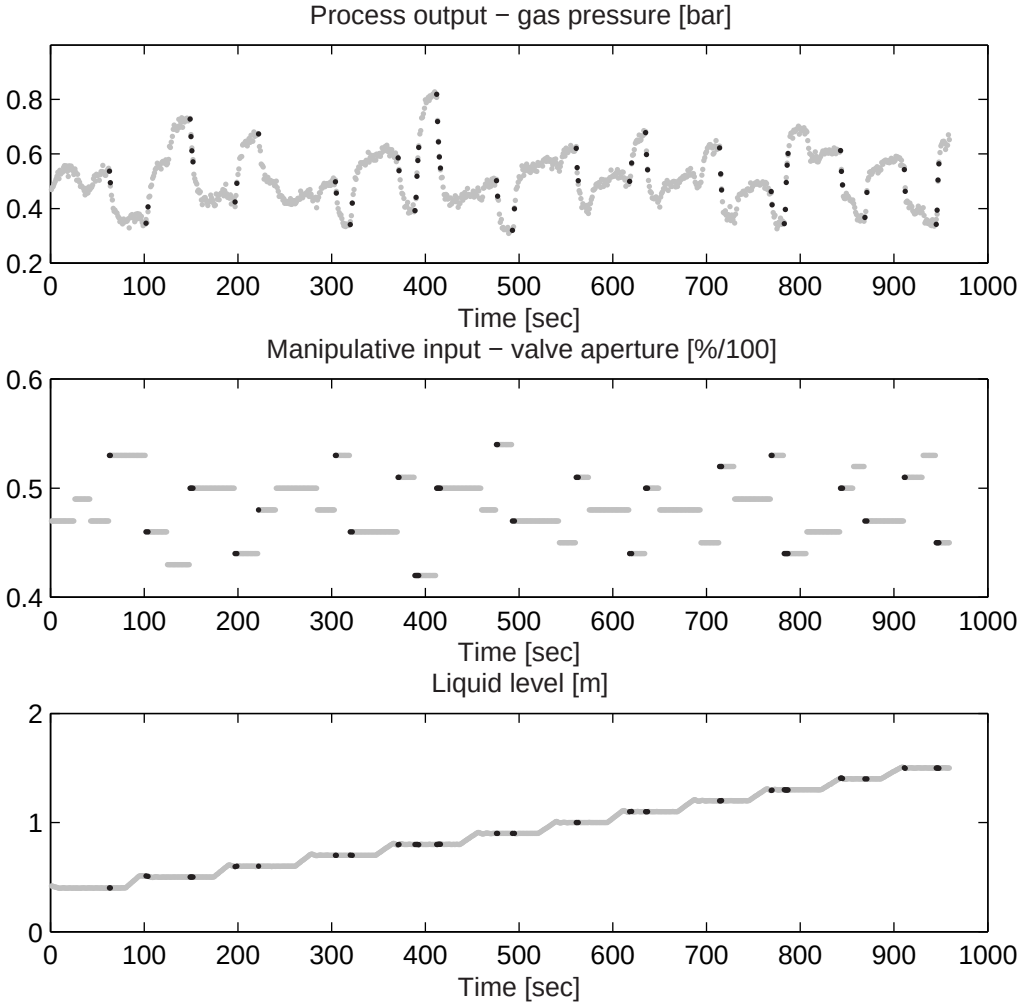

Fig. 2: Signals from which samples for identification (black dots) are taken

where N is the number of validation points, y_i is the process response (target), $\hat{y}_i$ is the model output (predictive mean) and σ_i^2 is the predictive variance at the time step i. SE, or one of its versions, is commonly used performance measure, which compares only the mean values of model predicted distribution of the model to the output of the process. LD on the other hand does not take into account only the mean value of predictive distribution but also its variance and is therefore providing more competent measure for GP model.

4. CONCLUSIONS

A case study where Gaussian process model with incorporated linear local models is used for the dynamic system identification of gas-liquid separator is presented in the paper.

The problem of system identifiability from the available data exist in practice as a result of an uneven data distribution. Concentrated data in the vicinity of the equilibrium points and sparse data far away from them make the task of identification very complex. It is very difficult to obtain a model

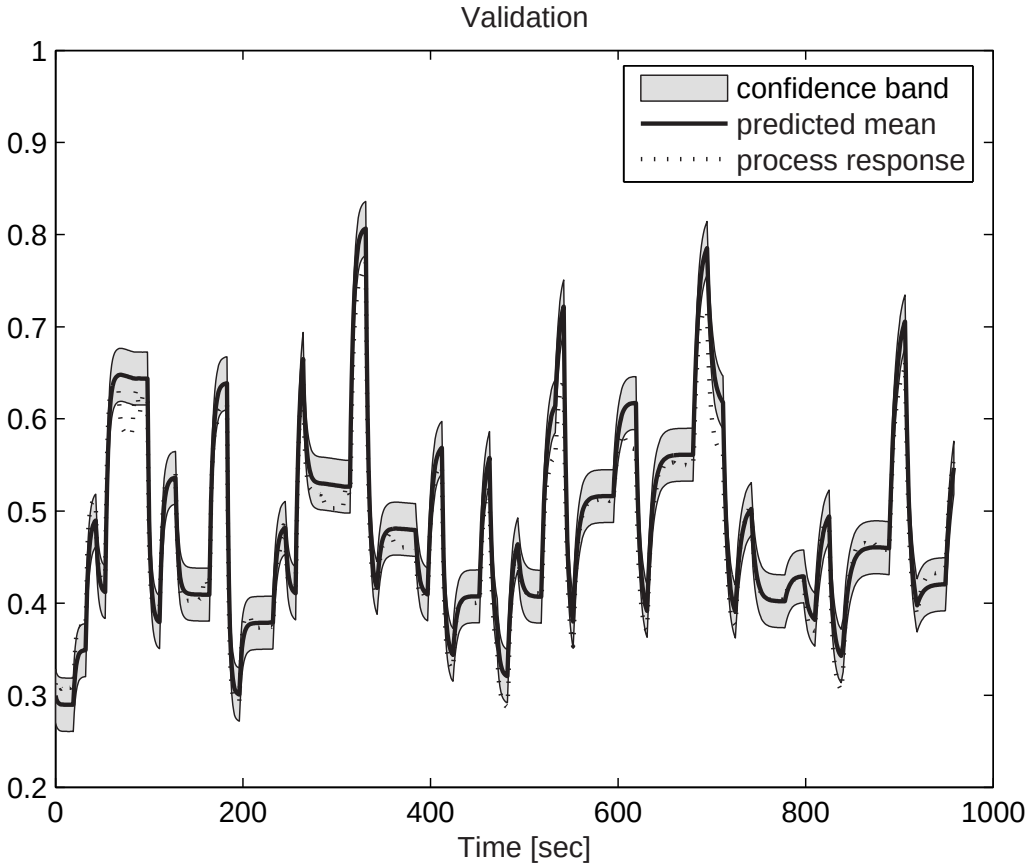

Fig. 3: Response on the validation data

from such data that will describe the dynamics far away from equilibrium points as well as it does close to them. A modelling approach that handles this situation well is GP modelling. However, any intensive calculation with such a model could represent a significant computational burden.

GP model with incorporated linear local models represents a form of probabilistic model that is computationally less demanding than GP model due to equilibrium data compression and reduction. Such model can be effectively used for modelling of nonlinear dynamic systems.

The gas-liquid separator plant is a representative of nonlinear process plants where uneven measurement data distribution takes place. A model based on these measurement data was developed in two stages. Firstly, local linear models were obtained around seven equilibrium points. These data were combined with sixty off-equilibrium data points and seven equilibrium points into Gaussian process model with incorporated linear local models. The model validation of obtained nonparametric model with relatively low number of data points is given. Obtained model can be used for various purposes e.g. process control or prediction of plant performance.

The contribution of this paper is the application of system identification with GP models with incorporated linear local models on a gas-liquid separator case study.

The used method is not meant as a replacement of the existing nonlinear system identification methods. Rather it is a useful complementary approach when uneven measurement data distribution is pronounced and when the system designer finds nonparametric model useful.

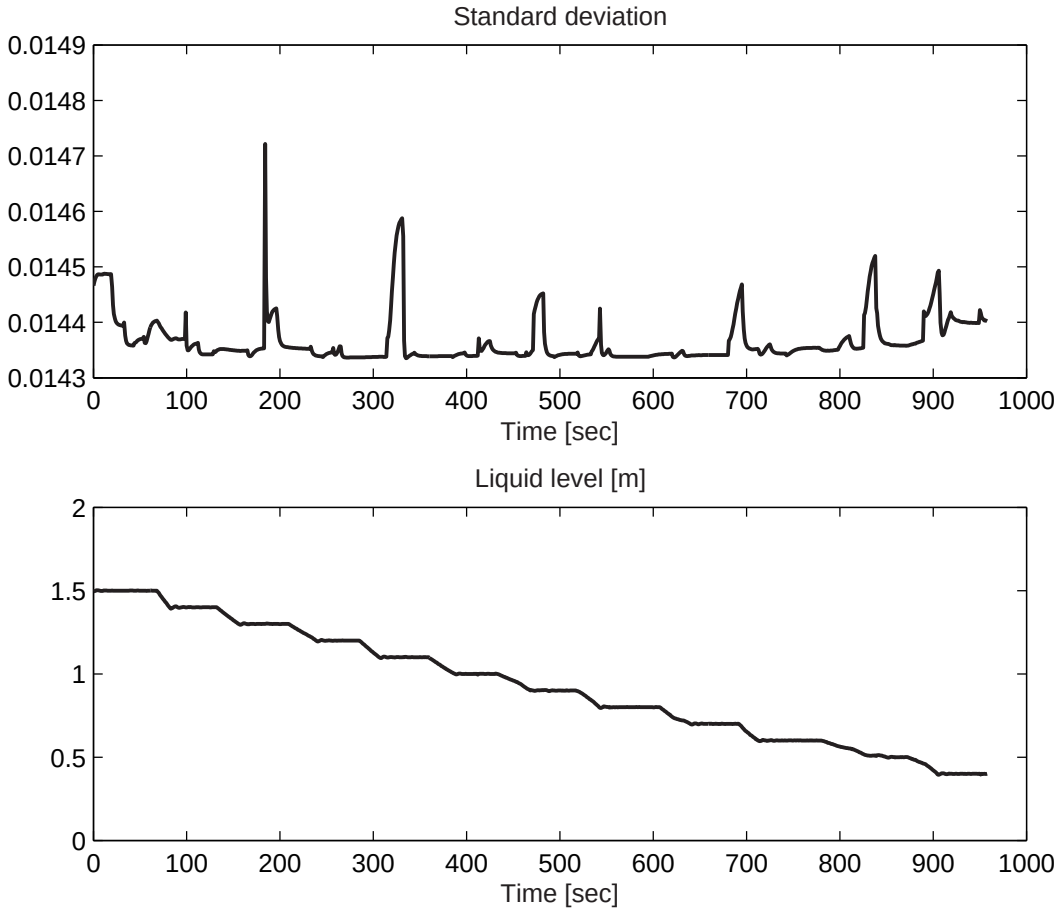

Fig. 4: Standard deviation and the level of liquid in tank R4.1

References

[1] AŽMAN K., KOCIJAN J., *Comprising Prior Knowledge in Dynamic Gaussian Process Models*, In: B. Rachev, A. Smrikarov (eds.), Proceedings of the International Conference on Computer Systems and Technologies - CompSysTech 2005, 16.-17.6., Varna, Bulgaria, sec IIIB.2, 2005.

[2] GREGORČIČ G., LIGHTBODY G., *An affine Gaussian process approach for nonlinear system identification*, System Science, 29, No. 2.,2003

[3] JOHANSEN, T.A., SHORTEN R. and MURRAY-SMITH R., *On the Interpretation and Identification of Dynamic Takagi-Sugeno Fuzzy Models*, IEEE Transactions on Fuzzy Systems, 8, 297-313, 2002.

[4] KOCIJAN, J. and GIRARD A., Incorporating linear local models in Gaussian process model. Proceedings of IFAC World Congress 2005, Prague, 2005.

[5] KOCIJAN J., GIRARD A., BANKO B. and MURRAY-SMITH R. *Dynamic Systems Identification with Gaussian Processes*, Mathematical and Computer Modelling of Dynamical Systems, 11, 411-424, 2005.

[6] MURRAY-SMITH, R., JOHANSEN T. A. and SHORTEN R., *On transient dynamics, off-equilibrium behaviour and identification in blended multiple model structures*, European Control Conference, Karlsruhe, BA-14, 1999.

[7] RASMUSSEN C.E., WILLIAMS C.K.I., *Gaussian Processes for Machine Learning*, Cambridge, MA: MIT Press, 2005.