Доклади на Българската академия на науките
Comptes rendus de l'Académie bulgare des Sciences

Tome 63, No 4, 2010

SCIENCES ET INGENIERIE

Automatique et informatique

GAUSSIAN PROCESS MODELLING CASE STUDY WITH MULTIPLE OUTPUTS

Juš Kocijan*,, Alexandra Grancharova*****

(*Submitted by Corresponding Member Ch. Roumenin on December 11, 2009*)

Abstract

The Gaussian-process (GP) model is an example of a probabilistic, non-parametric model with uncertainty predictions. It can be used for the modelling of complex nonlinear systems and also for dynamic systems identification. The output of the GP model is a normal distribution, expressed in terms of the mean and variance. At present it is applied mostly for the modelling of dynamic systems with one output. A possible channel structure for multiple-input multiple-output model and a case study for the modelling of a system with more than one output, namely a gas-liquid separator, is given in this paper.

Key words: dynamic systems modelling, systems identification, Gaussian process model, multivariable systems

1. Introduction. Nonlinear dynamic systems are more difficult for the identification than linear dynamic systems, especially systems with more than one output. Usual approaches for the identification of nonlinear black-box models are artificial neural networks, fuzzy models and others.

Gaussian process (GP) models form new, emerging complementary method for nonlinear system identification. GP model is a probabilistic non-parametric black-box model. It differs from most of the other black-box identification approaches as it does not try to approximate the modelled system by fitting the parameters of the selected basis functions but rather searches for the relationship among measured data. Gaussian processes models are closely related to

This work was financed by the Slovenian Research Agency, grants Nos P2-0001 and J2-2099 and contract No. BI-BG/09-10-005 and the National Science Fund of the Ministry of Education, Youth and Science of Republic of Bulgaria, contract No DO02-94/14.12.2008.

approaches such as Support Vector Machines and specially Relevance Vector Machines [1]. The output of Gaussian process model is a normal distribution, expressed in terms of mean and variance. Mean value represents the most likely output and the variance can be viewed as the measure of its confidence. Obtained variance, which depends on amount of available identification data, is important information distinguishing the GP models from other methods. Another potentially useful attribute of GP model is the possibility to include various kinds of prior knowledge into the model, e.g. local models, static characteristic, etc. Disadvantage of the method is potential computational burden for optimization that increases with number of data and number of regressors.

Applications of the GP model for the identification of dynamic systems are presented in e.g. [2–4]. While the published research results considered only identification of dynamic systems with one output, the purpose of this contribution is the identification of nonlinear dynamic system with more than one output, namely identification of nonlinear multivariable dynamic system.

2. Modelling of dynamic systems with Gaussian processes. A Gaussian process is an example of the use of a flexible, probabilistic, non-parametric model with uncertainty predictions. Its use and properties for modelling are reviewed in [1,5].

A Gaussian process is a collection of random variables which have a joint multivariate Gaussian distribution. The mean $\mu(\mathbf{x})$ and the covariance function $C(\mathbf{x}_p, \mathbf{x}_q)$ fully specify the Gaussian process. Covariance function $C(\mathbf{x}_p, \mathbf{x}_q)$ can be interpreted as a measure of distance between input points $\mathbf{x}_p$ and $\mathbf{x}_q$. For systems modelling it is usually composed from two main parts:

$$C(\mathbf{x}_p, \mathbf{x}_q) = C_f(\mathbf{x}_p, \mathbf{x}_q) + C_n(\mathbf{x}_p, \mathbf{x}_q), \tag{1}$$

where C_f represents a functional part and describes the unknown system we are modelling and C_n represents a noise part and describes the model of noise.

Some possible choices for C_f are: square exponential or Gaussian covariance function, which is most frequently used in functional part

$$C_f(\mathbf{x}_p, \mathbf{x}_q) = v_1 \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d (x_{dp} - x_{dq})^2\right], \tag{2}$$

linear covariance function

$$C_f(\mathbf{x}_p, \mathbf{x}_q) = \sum_{d=1}^{D} w_d x_{dp} x_{dq}, \tag{3}$$

rational quadratic covariance function

$$C_f(\mathbf{x}_p, \mathbf{x}_q) = v_1 \left[1 + \frac{1}{2\alpha}\sum_{d=1}^{D} w_d (x_{dp} - x_{dq})^2\right]^{-\alpha}, \tag{4}$$

constant covariance function, which is frequently used in the noise part, when it is considered to be a white noise

$$C_n(\mathbf{x}_p, \mathbf{x}_q) = v_0, \tag{5}$$

$\mathbf{\Theta} = [w_1 \ldots w_D \ \alpha \ v_1 \ v_0]^T$ are the 'hyperparameters' of the covariance functions, x_{dp} and x_{dq} are d^{th} components of input vectors $\mathbf{x}_p, \mathbf{x}_q$ and D is the input dimension. For a given problem, the parameters are identified (learned) using the data at hand. After the learning, one can use the w parameters as relevance indicators of the corresponding input regressors: if w_d is zero or near zero it means that the inputs in dimension d contain little information and could possibly be removed.

For a new test input $\mathbf{x}^*$, the predictive distribution of the corresponding output is $y^*|(\mathbf{X}, \mathbf{y}), \mathbf{x}^*$ and is Gaussian, with mean and variance

$$\mu(\mathbf{x}^*) = \mathbf{k}(\mathbf{x}^*)^T \ \mathbf{K}^{-1} \ \mathbf{y}, \tag{6}$$

$$\sigma^2(\mathbf{x}^*) = k(\mathbf{x}^*) \ - \ \mathbf{k}(\mathbf{x}^*)^T \ \mathbf{K}^{-1} \ \mathbf{k}(\mathbf{x}^*), \tag{7}$$

where $\mathbf{K}$ is the $N \times N$ training covariance matrix, $\mathbf{k}(\mathbf{x}^*) = [C(\mathbf{x}^1, \mathbf{x}^*), \ldots, C(\mathbf{x}^N, \mathbf{x}^*)]^T$ is the $N \times 1$ vector of covariances between the test and training cases, and $k(\mathbf{x}^*) = C(\mathbf{x}^*, \mathbf{x}^*)$ is the covariance between the test input and itself.

Gaussian processes can, like neural networks, be used to model static nonlinearities and can therefore be used for modelling of dynamic systems[2–4] if lagged samples of input and output signals are fed back and used as regressors. In such cases an autoregressive model is considered, such that the current output depends on previous outputs, as well as on previous control inputs.

$$\begin{aligned} \mathbf{x}(k) &= [y(k-1), y(k-2), \ldots, y(k-L), u(k-1), \\ &\quad u(k-2), \ldots, u(k-L)]^T, \\ \hat{y}(k) &= f(\mathbf{x}(k)) + \epsilon, \end{aligned} \tag{8}$$

where k denotes the consecutive number of data sample. Let $\mathbf{x}$ denote the state vector composed of the previous outputs y and inputs u up to a given lag L, and ϵ is white noise.

The cross-validation response fit is usually evaluated by performance measures. Beside commonly used performance measures such as *e.g.* mean relative square error, $\text{MRSE} = \sqrt{\dfrac{\sum_{i=1}^{N} e(i)^2}{\sum_{i=1}^{N} y(i)^2}}$, where $y(i)$ and $e(i) = \hat{y}(i) - y(i)$ are the system's output and prediction error in i-th step of simulation, the performance measures such as log predictive density error [1–3]

$$\text{LPD} = \frac{1}{2}\log(2\pi) + \frac{1}{2N}\sum_{i=1}^{N}\left(\log(\sigma^2(i)) + \frac{e(i)^2}{\sigma^2(i)}\right), \tag{9}$$

where $\sigma^2(i)$ is the prediction variance in i-th step of simulation, can be used for evaluating GP models. It takes into account not only mean prediction but the entire predicted distribution. Another possible performance measure is the negative log-likelihood of the training data [1] $\mathrm{LL} = \frac{1}{2}\log \mid \mathbf{K} \mid + \frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} + \frac{N}{2}\log(2\pi)$. LL is the measure inherent to the hyperparameter optimisation process and gives the likelihood that the training data is generated by given, *i.e.* trained model. Therefore it is applicable for validation on identification data only. The smaller the MRSE, LPD and LL are, the better the model is.

3. Modelling of systems with more than one output. Most of the literature considering modelling with Gaussian process models is concentrated on the problem of predicting a single output variable from an input. When more than one output is to be predicted and cross correlations are not neglected, then the correlation between multiple channels needs to be incorporated in the model. An overview of some published works is given in [1].

In the case of dynamic systems modelling with multiple correlated channels the work presented in [6] gives a possible solution for modelling of linear filters with multiple outputs with Gaussian process models.

A lot of literature on dynamic systems deals with multiple-input multiple-output systems, or so called multivariable systems, e.g. [7,8]. Following this experience and experience from identification of nonlinear dynamic systems based on Gaussian process models a model structure, or prior on model structure, is suggested in this paper as follows. Beside the influence of the input corresponding to channels output also interactions of other inputs and correlations with other channels' outputs are to be considered. Each model output can be predicted from a separate sub-model with inputs that consist of input signal samples and their lagged values of all channels and lagged samples from all output signals. No non-lagged samples from output signals are included into vector of inputs due to causality reason.

This principle is graphically shown in Fig. 1, where arrows denote the flow of lagged sampled values from input and output signals. This, relatively complex structure can be simplified later in the identification procedure when the redundant interactions or cross-correlations are removed via relevance detection of input regressors as mentioned in the previous section. The procedure of identification and validation of separate sub-models should follow the standard procedure as described in [2].

The semi-industrial process plant used for the case study in the paper is the unit for separating the gas from liquid [4,9] that forms part of a larger pilot plant.

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.

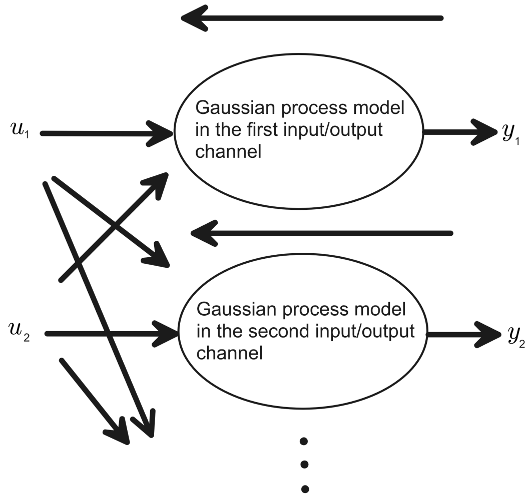

Fig. 1. Principial scheme of modelling multiple-input multiple-output scheme with flow of data used for modelling is indicated

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 the tank into the neutralization unit. On the other hand, the 'cushion' forces water to circulate back to the reservoir. The quantity of water in the circuit is constant. A detailed nonlinear mathematical model can be found in [4]. The nonlinear process is of a multivariable nature (two inputs and two outputs with dynamic interactions between the channels).

The aim of model identification in our case is a potential model for control design or prediction of output variables based on input signals. The GP model is composed of two parts according to structure proposed in the previous section: one is the sub-model that predicts pressure and the other is the sub-model that predicts liquid level. We limited the selection of possible covariance functions to functions described in Section 2. A systematical iterative procedure of comparing modelling results with measures introduced in Section 2 for various combinations of covariance functions for the functional part and of different input regressors for two output models was pursued. The backward approach from the higher number of regressors towards the lower number was used. A user-friendly experimentation with the process plant is made possible by an interface with the Matlab/Simulink environment [10]. 727 input-output data pairs were sampled

T a b l e 1

Comparison of models fit for pressure output (the left part of Table 1) and liquid level output (the right part of Table 1). The upper part consists of models with rational quadratic function with the following regressors: model M_{p1} with $p(k-1)$, $p(k-2)$, $u_1(k-1)$, $u_1(k-2)$, $u_2(k-1)$, $u_2(k-2)$, $h(k-1)$, $h(k-2)$, model M_{p2} with $p(k-1)$, $p(k-2)$, $u_1(k-1)$, $u_2(k-1)$, $h(k-1)$, model M_{p3} with $p(k-1)$, $u_1(k-1)$, $h(k-1)$, model M_{h1} with $h(k-1)$, $h(k-2)$, $u_1(k-1)$, $u_1(k-2)$,$u_2(k-1)$, $u_2(k-2)$, $p(k-1)$, $p(k-2)$, model M_{h2} with $h(k-1)$, $h(k-2)$, $u_2(k-1)$, $p(k-1)$, model M_{h3} with $h(k-1)$, $u_2(k-1)$, $p(k-1)$. Lower part consists of models with winning set (marked with $\clubsuit$) of regressors M_{p2} (left) and M_{h2} (right) with different covariance functions: square exponential – Gaussian, linear and rational quadratic

Regressors	Ident. LL	Valid. MRSE	Valid. LPD	Regressors	Ident. LL	Valid. MRSE	Valid. LPD
M_{p1}	−2501	0.080	8.152	M_{h1}	−4067	0.084	5978
$M_{p2}^{\clubsuit}$	−2500	0.043	−1.515	$M_{h2}^{\clubsuit}$	−4011	0.062	−0.391
M_{p3}	−2383	0.049	−0.926	M_{h3}	−3948	0.083	3880
Cov. fun. M_{p2}	Ident. LL	Valid. MRSE	Valid. LPD	Cov. fun. M_{h2}	Ident. LL	Valid. MRSE	Valid. LPD
Gaussian	−2496	0.045	−1.508	Gaussian	−4020	0.087	−1.95
Linear	−2063	0.194	−0.91	Linear	−3408	0.419	38.53
RQ$^{\clubsuit}$	−2500	0.043	−1.515	RQ$^{\clubsuit}$	−4011	0.062	−0.391

uniformly with sampling time of 20 s from input and identification output signal and used as estimation data for Gaussian process models. The same number of samples from input and output signals for validation was used as validation data for Gaussian process models' predictions.

Validation measures for a sample of cases at the end of systematic selection of model structure is given in Table 1. The table gives logarithm of likelihood LL that is used as a validation measure for optimisation of estimation data and MSRE and LPD measures for validation of predictions on validation input signal. The small part of backward regressor selection for models with rational quadratic covariance function (4) is given in the upper part of Table 1, while the comparison of the winning regressors' selection in the model with different covariance functions for functional parts are given in the lower part of Table 1.

Higher weight was put on measure LPD for a model selection in the validation procedure, because LPD measure is more appropriate for Bayesian models as it incorporates also the variance of predictions. The winning models are of the second order with regressors as described in the caption of Table 1 and contain rational quadratic covariance function (4) for the functional part and constant covariance function (5) for the noise part. These two models give better validation results with validation signals than models with other covariance functions and other regressors. Simulation responses on the validation signal, which was different

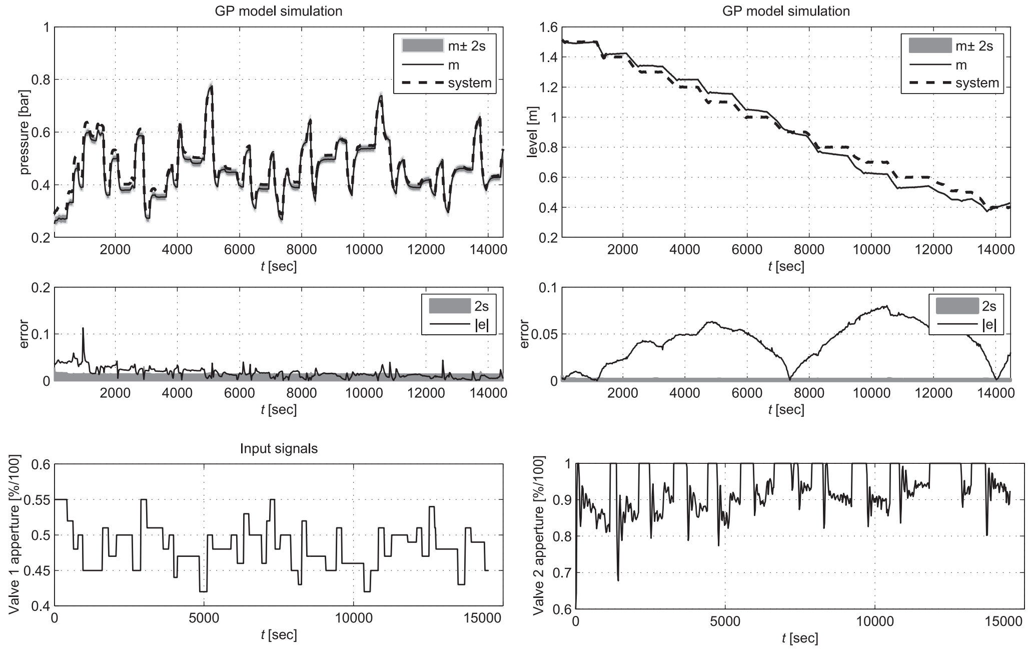

Fig. 2. Comparison of pressure measurement and Gaussian process model response (upper left figures), comparison of liquid level measurement and Gaussian process model response (upper right figures) and input signals into multi-input multi-output system (bottom figures)

from the identification one, of the finally selected models, denoted as models M_{p2} and M_{h2} and corresponding input signals are given in Fig. 2. Models M_{p2} and M_{h2} have different input regressors, but are of the same order and they use the same covariance function which is not surprising considering cross-interactions in the multivariable model.

It can be seen from Fig. 2 that model responses on validation input signals match measured process responses relatively well for most of the regions. This proves that applied procedure and selected structure of multivariable model were selected correctly.

4. Conclusions. The paper describes the identification of nonlinear dynamic systems with multiple-inputs and multiple-outputs by Gaussian process models. Selection of model inputs, the structure of channels between input data and predicted outputs and covariance function for calculation of covariances between estimation data are determined. A method for selecting of the possible structure of channels in the model is proposed and illustrated on a case study of modelling a gas-liquid separator process which has two signal inputs and two signal outputs.

The obtained results in the case study show that the applied procedure was selected correctly. This kind of model can be applied for control design or prediction of selected output variable, based on known input signals.

REFERENCES

[1] Rasmussen C. E., C. K. I. Williams. Gaussian Processes for machine learning. Cambridge, MIT Press, MA, 2006.

[2] Ažman K., J. Kocijan. ISA Transactions, **46**, 2007, No 4, 443–457.

[3] Kocijan J., A. Girard, B. Banko, R. Murray-Smith. Mathematical and Computer Modelling of Dynamic Systems, **11**, 2005, No 2, 411–424.

[4] Kocijan J., B. Likar. Simulation Modelling Practice and Theory, **16**, 2008, No 8, 910–922.

[5] Williams C. K. I. In: Learning in Graphical Models (ed. M. I. Jordan), Dordrecht, Kluwer Academic, 1998, 599–621.

[6] Boyle P., M. Frean. In: Advances in Neural Information Processing Systems 17 (eds L. K. Saul, Y. Weiss, L. Bottou), Cambridge, MIT Press, MA, 2005, 217–224.

[7] Maciejowski J. M. Multivariable Feedback Design, Addison-Wesley, Wokingham, 1989.

[8] Skogestad S., I. Postlethwaite. Multivariable feedback control: analysis and design, John Wiley, New York, 2005.

[9] Kocijan J., G. Žunič, S. Strmčnik, D. Vrančić. International Journal of Control, **75**, 2002, No 14, 1082–1091.

[10] Likar B. Gaussian process based nonlinear model predictive control, MSc. Thesis, University of Ljubljana, Ljubljana, 2004 (in Slovene).

**Jozef Stefan Institute*
39, Jamova Str.
SI-1000 Ljubljana, Slovenia
e-mail: jus.kocijan@ijs.si

***University of Nova Gorica*
13, Vipavska Str.
SI-5000 Nova Gorica, Slovenia

****Institute of Control and System Research*
Bulgarian Academy of Sciences
Acad. G. Bonchev Str., Bl. 2
1113 Sofia, Bulgaria
e-mail: alexandra.grancharova@abv.bg