Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

GREY BOX MODELLING – BRANCHES AND EXPERIENCES

Sohlberg B.*, Jacobsen E.W.**

**Dalarna University Borlänge, Sweden, Bjorn.Sohlberg@du.se*
***Royal Institute of Technology, Stockholm, Sweden, Jacobsen@ee.kth.se*

Abstract: This paper deals with methods and experiences of incorporating a priori knowledge into mathematical models of industrial processes and systems. Grey box modelling has been developed in several directions and can be grouped into branches depending on the way a priori knowledge is handled. In this paper we divide grey box modelling into the following branches; constrained black box identification, semi-physical modelling, mechanistic modelling, hybrid modelling and distributed parameter modelling. Experiences from case studies demonstrate the different branches of grey box modelling procedures. In the applications, the grey box models have been used for model based control, soft sensors, process supervision and failure detection. Further, distributed parameter modelling presents a specific challenge in that it is difficult to distinguish model reduction errors from model-data discrepancies. By estimating the model reduction error and forming hypothesis tests based on the estimate, the problem can be dealt with effectively. *Copyright © 2008 IFAC*

Keywords: Grey box modelling, Identification and modelling, Nonlinear systems, Advanced process control, Industrial processes.

1. INTRODUCTION

Modelling of industrial processes is traditionally based on white box modelling or black box identification. White box modelling implies that the model is constructed using scientific relations that completely describe the process. Black box identification employs a parametric model, which is adapted to measured data obtained from experiments on the process.

For many industrial processes, there exists some, but incomplete knowledge concerning the system. This implies that between the white box models and the black box models, there is a grey zone which gives a third way of making models of engineering systems, (Bohlin, 1991). In this approach a priori knowledge concerning the process is used and unknown parts of the model are estimated from measured data. The idea behind grey box modelling can be formulated as “don’t estimate what you already know but test your knowledge against experimental data”.

The choice of modelling method depends on several factors. Besides the type of a priori process knowledge available, the purpose of the model decides what kind of modelling approach that should be used. For example, the objective of the model can be model based control, simulation, supervision or failure detection. Further, the possibility to make designed experiments may also influence the choice of modelling approach, since black box identification needs more “informative” measurement data than the grey box modelling method.

This paper focus on in what way a priori knowledge can be incorporated into a grey box model and is divided into five main branches. The first branch, *constrained black box identification*, emanates from the black box identification frame, where a priori knowledge is incorporated as constraints on the model parameters. A second approach, *semi physical modelling*, makes use of case specific nonlinear transformations of measured input/output process signals. These new transformed signals are then used to estimate unknown parameters of a conventional black box model.

A third branch, *mechanistic modelling*, starts with a basic model originating from mathematical relations, given by first principles equations. The model is expanded to a wider structure mainly by hypothesis testing and using information about the residuals. A fourth branch, *hybrid modelling*, separates the model into a white box part, modelled by means of first principles equations, and a black box part modelled using neuro-fuzzy or other methods. A fifth branch considers *distributed parameter systems*. A specific challenge in calibrating and validating PDEs is that of distinguishing between model reduction errors and model-data discrepancies.

2. GREY BOX MODELLING BRANCHES

The distinction between the grey box modelling branches, in terms of the way a priori knowledge is included in the model, is not crisp. The branches defined here are considered to cover the main directions using a priori knowledge to form model structures of physical systems.

2.1 Constrained Black Box Identification

Constrained black box identification uses a black box model where specific parameters are constrained based on physical relations. The reason is that a linear black box model may give inconsistent parameter estimation due to noisy and few measurements.

A basis for this approach is that a simple continuous model can be transformed into a corresponding discrete time model. Known restrictions of the continuous model such as process stability and step response can be used to define limits on the static gain and the time constants, which are imposed on the parameters of the discrete model, (Tulleken, 1991) and (Murakemi and Seborg, 2000). This grey box approach allows a relatively straightforward extension to also include stochastic aspects.

For nonlinear processes, neuro-fuzzy models can be used to model nonlinearities. Lindskog and Ljung (2000) incorporate a priori knowledge as static monotonic gain curves where the model is given by fuzzy model structures. The authors show how a fuzzy model can be parameterized by using triangular membership functions. Aguirre et al. (2004) present a procedure that permits the use of steady state information to constrain the identification procedure of nonlinear dynamic black box models. The procedure includes a general framework that relates the steady state function to the corresponding model terms and parameters. It is shown that the resulting model always will have the specified static nonlinearity.

2.2 Semi Physical Modelling

When the process exhibits significant nonlinear behaviour, linear black box models give poor correspondence between the process and the model behaviour. For cases where first principle modelling is not feasible, it can be possible to make use of physical insight to transform the input and outputs variables to new variables which are used as regressor to develop a linear black box model.

The name semi physical modelling is in some papers referred as a modelling procedure based on first order principles, but we follow the definition given by Lindskog and Ljung (1994), where physical insight is used for nonlinear transformation of measured data.

For cases when the nonlinear steady state characteristic is known it is possible to use Hammerstein models, Wiener models or feedback block oriented models. For such cases, simple identification procedures have been developed by Pearson and Pottman (2000).

2.3 Mechanistic Modelling

For cases where a fundamental insight into the mechanisms that underlie the behavior of a process exists, relevant balance equations can be formulated as a set of first order equations (Stephanopolus, 1984). In the simplest case only some parameters in the model are unknown and have to be estimated from measured data. In the general case, the initial model structure formulated based on physical insight must be refined to match the experimental data.

A systematic approach to grey box mechanistic modelling is given in Sohlberg (1998). The activity to develop a mechanistic grey box model is considered as an object oriented procedure, which consists of basic modelling, experiments, estimation, expanded modelling and model appraisal.

An approach to derive a designer's guide for iterative development of grey box models is given in Bohlin (1994). The guide focuses on the situation when the uncertainty of a priori knowledge and the quality of the measured data are such that a model contains other stochastic elements than measurement errors.

2.4 Hybrid Modelling

Hybrid modelling separates the model into a white or a grey box part, which is modelled by means of first principles equations, and a completely black box part, to be identified using measured data. The black box part is entered either as a serial or parallel part. The serial method typically aims at modelling a specific physical part of the process model, see Fig. 1.

The parallel approach is used to extend an approximate first principle model with a general black box part, which compensate for the simplifications in the approximated model, see Fig. 2.

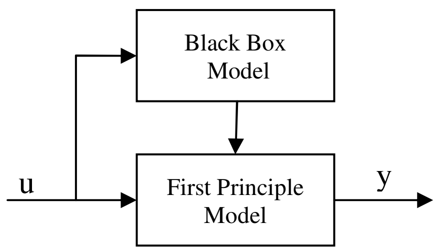

Fig. 1 Serial approach

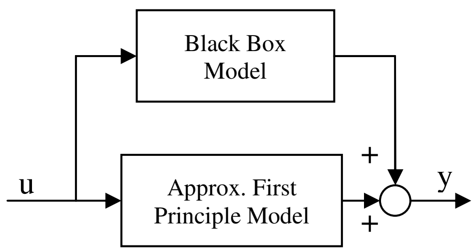

Fig. 2 Parallel approach

Thompson and Kramer (1994) employ neural networks to model the black box part. They conclude that the a priori knowledge is important when the white/grey box part is formulated. If this part of the model misfits, the hybrid model will perform poorly.

In addition, Lith et al. (2002) make use of first principles relations where unknown black box parts are described by fuzzy logic models. Wang et al. (2002) incorporate spoken/written information from experts or operators, local linear models and measured data into a modelling framework, which they describe as fuzzy hybrid modelling.

2.5 Distributed Parameter Modelling

Spatially distributed phenomena are important in many chemical and biochemical processes, and include for instance any process involving mass or energy transport by convection or diffusion. For such processes, a priori knowledge will result in a model structure involving partial differential equations (PDE).

A common approach to calibrate and validate PDE models is to perform an a priori spatial discretization so as to obtain an ODE model, (Funkquist, 1997) and (Liu, 2005). However, as pointed out in Liu (2005) and (Liu and Jacobsen, 2004) there are several potential pitfalls with this approach. First, if the physically derived PDE model structure is completely replaced by an approximate ODE model, then important information about the underlying process has been discarded. On the other hand, if the ODE model is used only for calibration and validation, then the model error introduced by spatial discretization can easily lead to falsely validated PDE models. As a simple example of the latter, (Liu and Jacobsen, 2004) show how finite difference discretization introduces artificial diffusion effects in the ODE model, artefacts that can not be distinguished from the physical diffusion modelled in the PDE structure.

Liu (2005) propose a grey-box identification scheme for distributed parameter systems based on integrating the spatial discretization with calibration and validation of the ODE model. The integration is made feasible by estimating the model error imposed by discretization and formulating simple hypothesis tests relating the parameters of the discretization mesh to the estimated error. The method ensures that there is no interference between model reduction errors and model-data discrepancies.

3. APPLICATIONS

Four different applications are presented to illustrate the different branches of grey box modelling. A more comprehensive and detailed presentation of grey box modelling for the corresponding systems is given in the references. The purpose of this chapter is to exemplify the stipulated branches.

3.1 Pickling Process

3.1.1 Process Description

After hot rolling in the steel industry, the oxide scale on the surface of the steel strip has to be removed before further processing of the steel strip, for instance cold rolling. Removal of oxide is carried out within the pickling process, where the steel strip passes through a bath of hydrochloric acid. Ferrous salt and water is formed when the oxide is dissolved in the acid. This means that hydrochloric acid is consumed.

The pickling process consists of four pickling tanks and there is a cascade flow of pickling solution counter-current to the direction of strip movement, see Fig. 3. The consumption of hydrochloric acid to dissolve the oxide is unknown but depends on the concentration of acid, strip velocity, strip width and strip thickness.

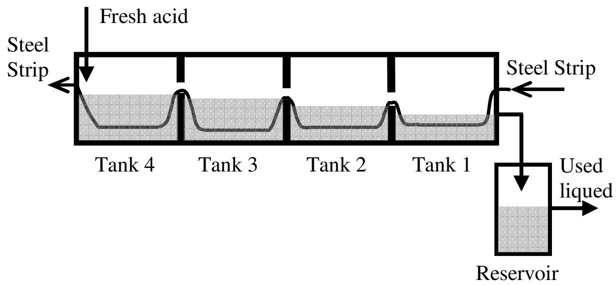

Fig. 3 Pickling process

3.1.2 Grey Box Modelling

The process consists of flow dynamics, mixing of liquids with different concentrations and chemical reactions where hydrochloric acid is consumed.

Application of the principle of mass balance for the liquid gives a mechanistic grey box model of the volume in each tank. Further, according to the principle of mass balance, the hydrochloric acid in each tank is modelled as mixing of different flows and consumption of hydrochloric acid. This model will be a hybrid grey box model where the consumption of hydrochloric acid is modelled as a black box part given by general continuous functions $s(x,u,\theta)$. The grey box model is presented in (Sohlberg, 2005).

$$\begin{aligned}\frac{dx}{dt} &= f(x,u,s(x,u,\theta)) + v \\ y &= g(x,u) + w\end{aligned} \tag{1}$$

where x is the state vector consisting of the concentrations and volumes of the tanks. The input vector u consists of the input flow of fresh acid, strip velocity, strip width and strip thickness. The vector θ consists of unknown parameters and y is the output vector consisting of the measured concentration in the tanks and the volume of the reservoir. The variables v and w are process and measurement noise.

The unknown parts $s(x,u,\theta)$ is approximated by means of Taylor series, where the series consist of the constant $s(x_0,u_0)$ and the partial derivatives calculated at the stationary point, (x_0, u_0). Note that x and u are vectors and that the series includes higher order terms to make the parts valid within a wider range than a linear approximation. In equation (2) below the unknown partial derivatives are represented by θ.

$$s(x,u,\theta) = s(x_0,u_0) + \left.\frac{\partial s}{\partial x}\right|_{\substack{x_0\\u_0}} \Delta x + \left.\frac{\partial s}{\partial u}\right|_{\substack{x_0\\u_0}} \Delta u + \left.\frac{\partial^2 s}{\partial x^2}\right|_{\substack{x_0\\u_0}} \Delta x^2 + \left.\frac{\partial^2 s}{\partial u^2}\right|_{\substack{x_0\\u_0}} \Delta u^2 + \left.\frac{\partial^2 s}{\partial x \partial u}\right|_{\substack{x_0\\u_0}} \Delta x \Delta u + .. \quad (2)$$

To find the relevant parameters of the Taylor series, a model which consists of only the constant terms is used to form a basic model. The Taylor series is then expanded by one parameter separately. The parameter which decreases the loss function most during parameter estimation is included in the model, provided that the decrease is significant according to the Likelihood test, (Rao, 1973).

3.1.3 Outcome of the Grey Box Model

The main advantages of using the Taylor series expansion method compared to other nonlinear black box models, for example neural networks, is the possibility to keep the number of estimated parameters low. The model is also expanded to incorporate stochastic parts besides conventional noise acting on the states and measurements.

The pickling line is a bottle neck in the production line. The system is simulated using a model predictive controller where both the flow of fresh acid and strip velocity are used as control variables, (Sohlberg, 2007). The simulation using a grey box model shows that it is possible to increase the production by approximately 15% and still keep the concentrations in the tanks within acceptable limits. The simulation also shows the importance of using the strip velocity as a control variable in addition to the currently used fresh acid.

Further, the relevant parts of the Taylor series can be used to explain the behaviour of the process and provide ideas for further investigations concerning design and control of the pickling process, (Sohlberg, 2005).

3.2 Rinsing Process

3.2.1 Process Description

The rinsing process is also a part of a steel strip pickling line. After the pickling process, the strip passes a rinsing process to rinse the strip from hydrochloric acid. Via the steel strip, the acid is transferred from the pickling process into the first rinse tank, see Fig. 4. Within the rinsing process, the strip is rinsed by a circulated flow of rinse water.

After each rinse tank, the steel strip passes between a pair of squeezer rolls. They are there to reduce the amount of liquid transferred via the strip. The rolls are rubber-surfaced iron rolls and are the parts that are liable to wear out. If the rolls become worn out, the risk of an insufficiently clean strip is impending. Clean water is fed into the last rinsing tank to dilute the rinse liquid in this tank; otherwise the concentration of hydrochloric acid would become too high. An equivalent amount of rinse liquid flows to the next tank.

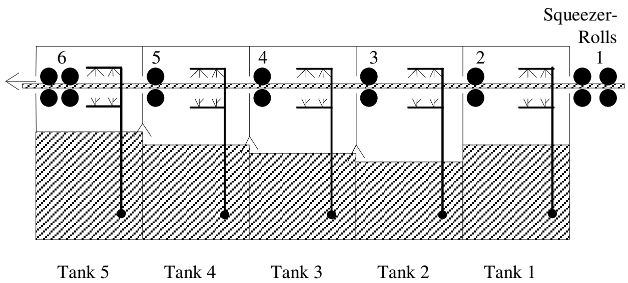

Fig. 4 Rinsing process

3.2.2 Grey Box Modelling

The grey box mechanistic modelling procedure is applied and is based on the principle of mass balance where flows with different concentration are mixed in each rinsing tanks. The unknown parts of the system consist of the flow via the strip after passing the squeezer rolls. A systematic way of finding a model describing the flow via the strip is presented in Sohlberg (1998). To find the model, we make use of the knowledge that the steel strip makes an imprint on the squeezer rolls surface of rubber. This gives a feasible model of the flows as:

$$Fb_i = Kb_i \cdot u_v u_w + Kt_i \cdot u_v (u_t - Btoff_i)\alpha \quad (3)$$

α=1 if Bt > $Btoff_i$ else α=0

Where u_v, u_w and u_t are the strip velocity, strip width, strip thickness and Kb_i, Kt_i, $Btoff_i$ are unknown estimated parameters for the flow after squeezer roll No i. Equation (3) says that there is a considerable flow besides the strip which is dependent on the strip thickness.

3.2.3 Outcome of the Grey Box Model

An Extended Kalman Filter based on the mechanistic grey box model is used to estimate both the process states and the unknown parameters describing the flow passing the squeezer rolls. The flow is a measure of the condition of the squeezer rolls and is used to advice the process operators which rolls are worn and should be exchanged at the next planned stop for maintenance.

Besides of using the grey box model for supervision of wearing parts, the grey box model shows that the process depends on another variable, strip thickness, which was not known before the project started.

3.3 River Flow System

3.3.1 Process Description

The system is a part of a river between two power stations with a distance about 12 km, see Fig. 5. The water flows through the turbines are not measured but is a result of generation of electrical energy, which depends on the demands for energy during day and night. The water level is measured at five places along the route. The river flow is also influenced by inflow from small rivers along the river, rainfall, evaporation and melting snow.

One of the difficulties in controlling the water levels is that there is no basin between the two power stations. Another difficulty is the backwater effect due to the slope of the river being less than 0.01 percent, (Chow et al., 1988). Consequently, upstream propagation must be considered when the river is controlled and the water levels considered along the river cannot be controlled independently of each other; a change in the water flow will influence all levels considered in a complex way.

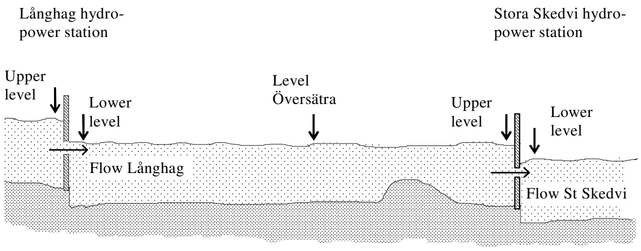

Fig. 5 Schematic outline of the river section

3.3.2 Grey Box Modelling

There is a nonlinear relationship between the produced energy, P(k), height of fall, $\Delta y(k)$ and water flow the turbines, u(k).

$$u(k) = a_1\left(\frac{P(k)}{\Delta y(k)}\right) + a_2\left(\frac{P(k)}{\Delta y(k)}\right)^2 + a_3\left(\frac{P(k)}{\Delta y(k)}\right)^3 \tag{4}$$

Applying the principle of mass balance gives an approximate linear relation between the water flows and water levels. This model is transformed to a discrete version, which corresponds to a linear ARMAX model given by equation (5) below. Restrictions on time constant and steady state gain on the continuous model are imposed on the discrete version. The grey box model is presented in (Sohlberg and Sernfält, 2002).

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})e(k) \tag{5}$$

In eq. (5), y(k) is a vector of the measured water levels at three locations along the river. The input u(k) is a vector consisting of the water flows trough the turbines at the corresponding water power stations, calculated from eq (4). The variable e(k) represents white noise.

Hence, the grey box modelling procedure for the river system is based on both the semi physical modelling procedure and restricted black box identification.

3.3.3 Outcome of the Grey Box Model

The river system is characterised by constraints on the water levels and water flows, time delays and measurable disturbances. Simulation based on grey box model shows that it is possible to control the river so the water levels vary less than during manual control. Further; the water level at the power stations be can be maintained at a higher level than during manual control, which means that about 5% more electric power can be produced with the same amount of water flow.

3.4 Chromatography

3.4.1 Process Description

Chromatography is a separation method widely used for analysis and product purification in the chemical and biochemical industry. A chromatography column consists of a column packed with solid material. A liquid containing the components to be separated is injected at one end of the column, and as the liquid moves through the column the components get adsorbed and diffuse into the packing material. Since different molecules are adsorbed differently, the effective speed through the column will differ significantly between the different components, resulting in a separation at the outlet.

3.4.2 Grey Box Modelling

A mechanistic dynamic model for a component in the liquid can be derived from a mass balance in the axial direction.

$$\frac{\partial c_L}{\partial t} = -v_{int}\frac{\partial c_L}{\partial z} + D_{ax}\frac{\partial^2 c_L}{\partial z^2} - \frac{1-\varepsilon_T}{\varepsilon_T}\frac{\partial q}{\partial t} \tag{6}$$

Standard Robin and Neuman conditions are employed at the boundaries. The adsorption to the stationary phase is described by so-called Langmuir sorption kinetics

$$\frac{\partial q(z,t)}{\partial t} = k_{ads}c_L(q_m - q) - k_{des}q \tag{7}$$

Model (6) - (7) is known as a *kinetic dispersive model* in the chromatography literature. A more frequently employed model structure, the *equilibrium dispersive model,* is obtained by assuming that the adsorption process is in equilibrium at all times. The assumption is that the effect of the sorption dynamics is essentially identical to that of diffusion, and hence can be modelled by the diffusion term in (6).

The grey-box modelling in Liu (2005) employs a finite element discretization of (6) and the mesh density is used to ensure that the model reduction errors do not to interfere with possible model-data discrepancies. Since the model reduction error depends on the parameter fit, the mesh selection is integrated with the calibration and performed in an iterative manner.

3.4.3 Outcome of the Grey-Box Model

The grey-box modelling reveals that the commonly employed *equilibrium dispersive model* can not be used to describe the effects of the sorption dynamics in chromatography. Thus, the common assumption that the impact of sorption dynamics on the component separation is similar to that of diffusion is incorrect. This has important consequences both for modelling and design of chromatography processes.

4. CONCLUSIONS

Grey box modelling is based on a priori knowledge about the process. The type of process knowledge available is case dependent and it is therefore difficult to generalize. This paper divides the grey box methodology into five branches.

Constrained black box identification uses physical insight to constrain the model parameters which are estimated using measured data.

Semi-physical modelling makes use of physical insight to transform the input and outputs variables to new variables which are used as regressor in a linear black box model.

For cases when there exists a fundamental insight into the mechanisms of the underlying behaviour of a process, a mechanistic grey-box modelling procedure consists of basic modelling, experiments, estimation, expanded modelling and model appraisal.

Hybrid modelling separates the model into a white or a grey box part, modelled by means of first principles equations, and a completely black box part, to be identified using measured data.

Distributed parameter modelling presents a specific challenge in that it is difficult to distinguish model reduction errors from model-data discrepancies. By estimating the model reduction error and forming hypothesis tests based on the estimate, the problem can be dealt with effectively.

REFERENCES

Agirre L.A., Barroso M.F.S, Saldanha R.R. and Mendes E.M.A.M (2004). Imposing Steady State Performance on Identified Nonlinear models by means of constrained Parameter Estimation. *IEE Proc.-Control Theory Appl.*, **Vol. 151**, No. 2.

Bohlin T. (1991). Interactive System Identification: Prospects and Pitfalls. *Springer-Verlag*. Berlin

Bohlin T. (1994). Derivation of a Designers Guide for Iterative Grey Box Identification of Nonlinear Stochastic Objects. *International Journal Control*, **Vol. 59**, No 6, pp. 1505 – 1524

Chow V.T., Maidment D. R. and Mays L. W. (1988). Applied Hydrology. *MacGraw-Hill.*

Funkquist J. (1997). Grey Box Identification of a Continuous Digester – A Distributed Parameter Process. *Control Engineering Practice*, **Vol 5**. pp. 919 - 930

Lindskog P. and Ljung L. (2000). Ensuring Monotic Gain Characteristic in Estimated Models by Fuzzy Model Structures. *Automatica.* **Vol. 36**. pp. 311 – 317.

Forsell U. and Lindskog P. (1997). Combining Semiphysical and Neural Network Modeling: An Example of its Usefulness. *11th IFAC Symposium on System Identication*. **Vol. 2**. pp. 795 - 798

Lindskog P. and Ljung L. (1994). Tools for Semi Physical Modelling. *10th IFAC Symposium on System Identication*. pp. 1199 – 1204

Lith van P. F. Betlem B. H. L. and Roffel B. (2002). A Structured Modelling Approach for Dynamic Hybrid Fuzzy-First Principles Models. *Journal of Process Control.* **Vol. 13**. pp. 605 – 615.

Liu Y. (2005). Grey Box Identification of Distributed Parameter Systems. Doctoral Thesis. *Signals, Sensors and Systems, KTH*, Stockholm, Sweden.

Liu Y. and Jacobsen E.W. (2004). Error Detection and Control in Grey Box Modelling of Distributed Parameter Processes. *Proceedings IFAC DYCOPS7*, Boston.

Murakami K., Seborg D. E. (2000). Constrained parameter estimation with application to blending operations. *Journal of Process Control*, **Vol. 10**, pp. 195 – 202

Pearson R.K. and Pottman M. (2000). Gray-box Identification of Block-oriented Nonlinear Models. Journal of Process Control, Vol. 10, pp. 301 – 315

Rao C.R. (1973). Linear Statistical Inference and Its Applications, *John Wiley*, London.

Sohlberg B. (1998). Supervision and Control for Industrial Processes. Advances in Industrial Control. *Springer verlag*. London.

Sohlberg B. and Sernfält M. (2002). Grey Box Modelling for River Control. *Journal of Hydroinformatics*. **Vol. 4** pp 265 – 280.

Sohlberg B. (2003). Grey Box Modelling for Model Predictive Control of a Heating Process. *Journal of Process Control*. **Vol. 13**, No. 3, pp. 225 – 238.

Sohlberg B. (2005). Hybrid Grey Box Modelling of a Pickling Process. *Control Engineering Practice*. **Vol. 13**, Issue 9, , pp 1093-1102

Sohlberg B. (2007). Control of a Pickling Process using a Grey Box Model. *IFAC Symposium on Automation in Mining, Mineral and Metal Processing*. pp. 249 – 254.

Stephanopoulos G. (1984). Chemical Process Control. An Introduction to Theory and practice. *Prentice Hall*. New Jersey.

Tulleken H. J. A. F. (1991) Application of the Grey-Box Approach to Parameter Estimation in Physicochemical Models. *The 30th Conference on Decision and Control*, Brighton, England. Pp. 1177 – 1183.

Thompson M. L. and Kramer M. A. (1994). Modelling Chemical Processes Using Prior Knowledge and Neural Networks. *AIChE Journal*, **Vol. 40** No. 8 pp. 1328 – 1340.