Identification of room temperature models using k-step PEM for Hammerstein systems

Fabian Paschke
Division Engineering of Adaptive Systems (EAS)
Fraunhofer Institute for Integrated Circuits (IIS)
01069 Dresden, Germany
fabian.paschke@eas.iis.fraunhofer.de

Tobias Zaiczek
Division Engineering of Adaptive Systems (EAS)
Fraunhofer Institute for Integrated Circuits (IIS)
01069 Dresden, Germany
tobias.zaiczek@eas.iis.fraunhofer.de

Klaus Röbenack
Faculty of Electrical and Computer Engineering, Institute of Control Theory
Dresden University of Technology
01062 Dresden, Germany
klaus.roebenack@tu-dresden.de

***Abstract*—The contribution focuses on the identification of short term room temperature prediction models using k-step ahead prediction error minimization (PEM). In the first part of the paper we describe the general identification problem and summarize the k-step PEM-algorithm for discrete time Hammerstein systems. In the second part the k-step ahead approach is applied to measured data of a conference room and compared to the standard 1-step PEM approach. The results indicate, that k-step ahead approach can indeed lead to better prediction performance in this particular practical application.**

***Index Terms*—system identification, building, hammerstein, model predictive control**

I. INTRODUCTION

In the recent years model predictive control (MPC) has gained a lot of interest in the building automation sector [1]–[5]. Especially predictive control of the integrated room automation (IRA) has been of constant interest [6], [7], which requires an accurate dynamic model for the short-term temperature prediction of the considered room.

A purely physically motivated modeling approach using energy balance equations is often problematic due to the unavailability of accurate parameters, such as heat transfer coefficients and air exchange rates. In contrast system identification has proven to be an adequate technique for estimating accurate dynamic models from actual process data [8]–[10].

Identification of dynamic systems for the purpose of control design has a long history in the field of control theory. In the past a variety of different identification methods such as subspace identification, instrumental variable methods, or prediction error minimization (PEM) had been developed. Due to the availability of appropriate algorithms and especially because of its favorable statistical properties, PEM is one of the most frequently applied identification methods for linear time-invariant (LTI) systems.

In most PEM applications the parameter estimation is performed by minimization of the mean square of 1-step ahead prediction errors. It has been shown that 1-step PEM is an efficient estimator and thus in theoretical analysis leads to best 1-step and k-step prediction performance [11]. However in many practical situations this will not hold true, since the efficiency statement relies on the assumptions of gaussian distributed innovations and further it is assumed that the description of the actual process is within considered model set [12]. From a practical standpoint these assumptions can be more or less restrictive, thus PEM identification using a k-step or a multi-step criterion can be desirable in practice.

In this contribution we summarize PEM identification using k-step criterion for general Hammerstein systems (see e. g. [13], [14], and [15] for other identification methods). We apply this method to actual data gathered from a conference room and show that k-step PEM can indeed be beneficial in terms of prediction accuracy. Further we demonstrate the importance of the initialization problem (see [16], [17]) of the optimal predictor in our particular application. The article is organized as follows: In Sections 2 and 3 we outline the modeling assumptions and illustrate the algorithm for estimating the system parameters on the basis of a k-step mean square error (MSE) criterion. In the following section the algorithm is applied to data gathered from a conference room. Further the results are compared to a standard 1-step PEM approach, while in the last section a summary of our results and the focus of next research topics are given.

II. MODELING ASSUMPTIONS

We assume a general single-input single-output (SISO) Hammerstein system

$$A(q,\theta)y[t] = q^{-d}\frac{B(q,\theta)}{F(q,\theta)}g(u[t],\eta) + \frac{C(q,\theta)}{D(q,\theta)}e[t], \quad (1)$$

where $t = 1, 2, ..., N$ denotes the discrete time and $g(u[t],\eta)$ is a continuously differentiable static nonlinearity parameterized by an unknown parameter vector η. Furthermore q denotes the time shift operator, thus $qy[t] = y[t+1]$ and $q^{-1}y[t] = y[t-1]$

respectively, and the innovation $e[t]$ is a sequence of uncorrelated random numbers with zero mean and unknown variance σ_e^2. The polynomials $A(q,\theta)$, $C(q,\theta)$, $D(q,\theta)$ and $F(q,\theta)$ are assumed to be monic, whereas $B(q,\theta)$ is of no specific structure. The corresponding polynomial degrees are denoted by n_a, n_b, n_c, n_d, and n_f, where the parameters a_i, b_i, c_i, d_i, and f_i are summarized in the parameter vector θ. Furthermore, due to the stability of the optimal prediction (see Section III), the polynomials $C(q,\theta)$ and $F(q,\theta)$ are allowed to have roots inside the unit circle only (see Section III). The structure of the considered system (1) is depicted in Figure 1.

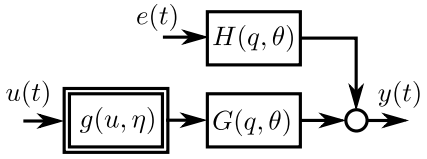

Fig. 1. Structure of considered Hammerstein system with $G(q,\theta) = \frac{q^{-d}B(q,\theta)}{A(q,\theta)F(q,\theta)}$ and $H(q,\theta) = \frac{C(q,\theta)}{A(q,\theta)D(q,\theta)}$

III. IDENTIFICATION

A. Optimal k-step prediction

As common in PEM the parameters θ and η of (1) will be estimated by minimization of the mean squared error of the optimal k-step ahead prediction $\hat{y}[t|t-k,\theta,\eta]$[1]

$$(\hat{\theta},\hat{\eta}) = \arg\min_{\theta,\eta} \frac{1}{N-t_s+1}\sum_{t=t_s}^{N} \varepsilon^2[t|t-k,\theta,\eta], \quad (2)$$

where the prediction error sequence is defined by

$$\varepsilon[t|t-k,\theta,\eta] = y[t] - \hat{y}[t|t-k,\theta,\eta] \quad (3)$$

and $t_s > 1$ is an appropriately selected integer value (see Section III-C). To derive an expression for the (MSE)-optimal k-step prediction $\hat{y}$, it is useful to rewrite the stochastic part of (1) such that[2]

$$\begin{aligned} y[t] &= \frac{q^{-d}B(q)}{A(q)F(q)}g(u[t]) + \frac{C(q)}{A(q)D(q)}e[t] \\ &= \frac{q^{-d}B(q)}{A(q)F(q)}g(u[t]) + q^{-k}\frac{R(q)}{A(q)D(q)}e[t] \\ &\quad + S(q)e[t], \end{aligned} \quad (4)$$

where

$$S(q) = 1 + s_1 q^{-1} + \ldots + s_{k-1}q^{-k+1} \quad (5)$$

contains the first k elements of the series expansion of $C(q)/A(q)D(q)$ in q^{-1} and $q^{-k}R(q)/A(q)D(q)$ is the remainder (see [11], [12], [18]). The polynomial R can be expressed as

$$R(q) = r_0 + r_1 q^{-1} + \ldots + r_{n_r}q^{-n_r}, \quad (6)$$

where $n_r = \max(n_a + n_d - 1, n_d - k)$. From (1) we can substitute

$$e[t-k] = \frac{A(q)D(q)}{C(q)}\Big(y[t-k] - \frac{q^{-d}B(q)}{A(q)F(q)}g\big(u[t-k]\big)\Big) \quad (7)$$

into (4), which yields

$$\begin{aligned} y[t] &= \frac{q^{-d}B(q)}{A(q)F(q)}g(u[t]) + \frac{R(q)}{C(q)}\Big(y[t-k] \\ &\quad - \frac{q^{-d}B(q)}{A(q)F(q)}g\big(u[t-k]\big)\Big) + S(q)e[t] \\ &= \frac{q^{-d}B(q)D(q)S(q)}{C(q)F(q)}g(u[t]) + \frac{R(q)}{C(q)}y[t-k] \\ &\quad + S(q)e[t], \end{aligned} \quad (8)$$

where the relation $C(q) - q^{-k}R(q) = A(q)D(q)S(q)$ has been used (cf. (4)). The term $S(q)e[t]$ depends only on the values of e from t to $t-k+1$ (cf. (5)) and is the only unpredictable part in Equation (8) at time $t-k$. We thus conclude that

$$\hat{y}[t|t-k] = \frac{q^{-d}B(q)D(q)S(q)}{C(q)F(q)}g(u[t]) + \frac{R(q)}{C(q)}y[t-k] \quad (9)$$

is the expression for the optimal k-step ahead prediction[3] at time $t-k$. Putting (9) into (3) we obtain

$$\begin{aligned} C(q)F(q)\varepsilon[t|t-k,\theta,\eta] &= A(q)D(q)F(q)S(q)y[t] \\ &\quad - q^{-d}B(q)D(q)S(q)g(u[t]), \end{aligned} \quad (10)$$

which is the filter that needs to be implemented in order to estimate the k-step ahead prediction error sequence in (2).

Note that in case of $k = 1$, i. e. $S(q) = 1$, we obtain standard 1-step PEM. We can thus regard k-step PEM as 1-step PEM by filtering $y[t]$ and $g(u[t])$ through the MA-prefilter $S(q)$. Of course, from the estimation point of view, prefiltering of the data is not possible, since $S(q)$ depends on the unknown coefficients of $C(q)$, $D(q)$ and $A(q)$.

B. Gradient

In general the optimization problem (2) defines a nonlinear least squares (NLS) problem[4]. In order to solve the problem efficiently, analytic expressions for the gradients of the residuals $\varepsilon[t|t-k,\theta,\eta]$ in (10) w. r. t. the parameter vectors θ and η are needed.

Since the derivatives of $\varepsilon[t|t-k]$ will depend on the coefficients of the filter $S(q)$, it is necessary to derive explicit

[1]The term $\hat{y}[t|t-k]$ denotes the prediction of the system output at time t given the output measurements up to time $t-k$. This means that the prediction $\hat{y}[t|t-k]$ at time t is allowed to depend on measured output values up to $t-k$ only, whereas all values of the input u up to the current time t can be used.

[2]The parameter vectors θ and η will be suppressed for clarity.

[3]The formal proof is conceptually similar to the proof for ARMA Models given in [18].

[4]Precisely speaking Eq. (2) is a NLS problem with nonlinear inequality constraints, since the roots of $C(q)$ and $F(q)$ are allowed to have roots inside the unit circle only. In case of $n_c \leq 2$ and $n_f \leq 2$ these restrictions can be expressed as a set of linear inequality constraints.

expressions for the coefficients s_j ($j = 0, \ldots, k-1$), which leads to a system of nonlinear equations that can be solved iteratively:

$$s_j = c_j - \sum_{i=1}^{\min(j,n)} p_i s_{j-i} \quad j = 1, \ldots, k-1;\ s_0 = 1 \tag{11}$$

The partial derivatives of s_j w. r. t. the coefficients of $A(q)$, $C(q)$ and $D(q)$ can then be written as

$$\begin{aligned}
\frac{\partial s_j}{\partial a_l} &= -\sum_{i=1}^{\min(j,n)} \frac{\partial p_i}{\partial a_l} s_{j-i} + \frac{\partial s_{j-i}}{\partial a_l} p_i \\
\frac{\partial s_j}{\partial d_l} &= -\sum_{i=1}^{\min(j,n)} \frac{\partial p_i}{\partial d_l} s_{j-i} + \frac{\partial s_{j-i}}{\partial d_l} p_i \\
\frac{\partial s_j}{\partial c_l} &= \underbrace{1}_{\substack{\text{if } j=l,\\ 0 \text{ otherwise}}} - \sum_{i=1}^{\min(j,n)} \frac{\partial s_{j-i}}{\partial c_l} p_i.
\end{aligned} \tag{12}$$

Here p_i are the coefficients of $P(q) = A(q)D(q)$ and thus $\frac{\partial p_i}{\partial a_l} = d_{i-l}$ and $\frac{\partial p_i}{\partial d_l} = a_{i-l}$ if $0 \leq i-l \leq n$ and 0 otherwise. Next we define the derivatives of $S(q)$ w. r. t. its parameters

$$\frac{\partial S(q)}{\partial a_i} = \frac{\partial s_1}{\partial a_i} q^{-1} + \ldots + \frac{\partial s_{k-1}}{\partial a_i} q^{-k+1}, \tag{13}$$

where $\partial S(q)/\partial c_i$ and $\partial S(q)/\partial d_i$ are defined similarly.

Finally we can derive expressions for the gradients $\partial\varepsilon/\partial\theta$ and $\partial\varepsilon/\partial\eta$ of the sequence $\varepsilon[t]$

$$\begin{aligned}
CF\frac{\partial \varepsilon[t]}{\partial a_i} &= \Big(ADF\frac{\partial S}{\partial a_i} + q^{-i}DFS\Big)y[t] \\
&\quad - q^{-d}BD\frac{\partial S}{\partial a_i}g(u[t]) \\
CF\frac{\partial \varepsilon[t]}{\partial b_i} &= -q^{-d-i}DSg(u[t]) \\
CF\frac{\partial \varepsilon[t]}{\partial c_i} &= ADF\frac{\partial S}{\partial c_i}y[t] - q^{-d}BD\frac{\partial S}{\partial c_i}g(u[t]) \\
&\quad - F\varepsilon[t-i] \\
CF\frac{\partial \varepsilon[t]}{\partial d_i} &= \Big(ADF\frac{\partial S}{\partial d_i} + Aq^{-i}FS\Big)y[t] \\
&\quad - q^{-d}B\Big(D\frac{\partial S}{\partial d_i} + q^{-i}S\Big)g(u[t]) \\
CF\frac{\partial \varepsilon[t]}{\partial f_i} &= ADSy[t-i] - C\varepsilon[t-i] \\
CF\frac{\partial \varepsilon[t]}{\partial \eta_i} &= -q^{-d}BDS\frac{\partial g(u[t])}{\partial \eta_i},
\end{aligned} \tag{14}$$

where the arguments θ, η and q have been discarded for clarity.

C. Initial conditions

While implementing the filters (10) and (14) the question of choice of suitable initial conditions arises. If the initial values are not chosen carefully, the prediction error sequence $\varepsilon[t]$ in (10) can have large transients, which can lead to a significant bias and/or a decrease of accuracy of the parameter estimates $\hat{\theta}$ and $\hat{\eta}$ [16], [17].

Denoting the sequence on the right hand side of Equation (10) by $w[t]$ we can express (10) by

$$\underbrace{C(q)F(q)}_{L(q)}\varepsilon[t] = w[t], \tag{15}$$

where $n_l = n_c + n_f$. Because the sequence $w[t]$ is obtained by MA-filtering of $y[t]$ and $g(u[t])$ it is easy to verify that $n_a+n_d+n_f+k-1$ and $d+n_b+n_d+k-1$ initial conditions will be needed to compute $w[t]$. Since the indexing of y and u starts at 1 by definition, it is advisable to evaluate the cost function (2) starting from time $t_s = \max(n_a+n_d+n_f, d+n_b+n_d)+k$.

Due to the fact that (15) is an AR-filter with input sequence $w[t]$, $n_l = n_c + n_f$ initial conditions of $\varepsilon[t]$ for $t = t_s - n_l, \ldots, t_s - 1$, need to be chosen. For 1-step PEM there have been proposed different methods for the initialization, which can be used for k-step PEM as well:

1) All initial values of $\varepsilon[t]$ are set to 0 for $t < t_s$.
2) Compute the initial conditions such that the first n_c+n_f samples of the predictor $\hat{y}[t]$ match the measured data $y[t]$.
3) Include initial values into the parameter vector and estimate them together with θ and η.
4) Estimate initial conditions using backforecasting procedure described in [16], [17].
5) Compute the initial conditions during the evaluation of the cost function using least squares (LS) approach.

The first and the second method are conceptually simple to implement, but can lead to bad results if the predictor filter (9) has poles close to the unit circle. In practice this problem often appears for output error models or if a k-step criterion is used in order to estimate the model. From the theoretical point of view the third procedure is preferable, because it obviously should lead to the lowest value of the cost function (2). However it can increase the dimension of the parameter vector significantly, especially if multiple datasets are used for the estimation, since for each set independent initial conditions need to be defined. Thus, in practice, the methods 4 and 5 are preferable. The backforecasting procedure is discussed in [16], [17] can be used similarly for a k-step criterion as well, whereas the last method will be discussed below.

In order to estimate the initial conditions using LS method, it is helpful to transform (15) to state space form. Defining the state vector to $x[t] = \big(\varepsilon[t-1], \ \ldots, \ \varepsilon[t-n_l]\big)^\top$ we can express (15) as

$$\begin{aligned}
x[t+1] &= \underbrace{\begin{pmatrix} -l_1 & -l_2 & \cdots & -l_{n_l} \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}}_{A} x[t] + \underbrace{\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{B} w[t] \\
\varepsilon[t] &= \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}}_{C} x[t] + \underbrace{1}_{D} w[t].
\end{aligned} \tag{16}$$

Propagating (16) forward in time yields a system of linear equations

$$\varepsilon = \Psi\varepsilon_0 + \Omega w, \tag{17}$$

where $\varepsilon = (\varepsilon[t_s], \ldots, \varepsilon[N])^\top$, $\varepsilon_0 = (\varepsilon[t_s-1], \ldots, \varepsilon[t_s-n_l])^\top$, $w = (w[t_s], \ldots, w[N])^\top$ and

$$\Psi = \begin{pmatrix} CA \\ CA^2 \\ \vdots \\ CA^{\tilde{N}} \end{pmatrix} \quad \Omega = \begin{pmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ CA^{\tilde{N}-2}B & \cdots & CB & D \end{pmatrix}. \quad (18)$$

Here Ψ is called the (extended) observability matrix, Ω is the Toeplitz matrix containing the Markov parameters CB, CAB, CA^2B, ... of the system [19] and $\tilde{N} = N - t_s + 1$.

Since $\varepsilon[t]$ is assumed to be a white noise sequence with zero mean the LS-estimate

$$\varepsilon_0 = -\left(\Psi\Psi^\top\right)^{-1}\Psi^\top\left(\Omega w\right) \quad (19)$$

is an efficient estimator of the initial conditions ε_0.

In order to obtain the initial conditions $\partial\varepsilon[t]/\partial\theta$ and $\partial\varepsilon[t]/\partial\eta$ ($t = t_s-n_l, \ldots, t_s-1$), it is necessary to differentiate (15), which leads to

$$L(q)\frac{\partial\varepsilon[t]}{\partial\theta_i} = \frac{\partial w[t]}{\partial\theta_i} - \frac{\partial L(q)}{\partial\theta_i}\varepsilon[t], \quad (20)$$

where $\partial w[t]/\partial\theta_i$ denotes the derivative of $w[t]$ w. r. t. an element of the parameter vector θ. Similarly, by transformation to state space form, one can obtain the vector of initial conditions

$$\frac{\partial\varepsilon_0}{\partial\theta_i} = -\left(\Psi\Psi^\top\right)^{-1}\Psi^\top\left(\Omega\frac{\partial w}{\partial\theta_i}\right), \quad (21)$$

where $\partial\varepsilon_0/\partial\theta_i = (\partial\varepsilon[t_s-1]/\partial\theta_i, \ldots, \partial\varepsilon[t_s-n_l]/\partial\theta_i)^\top$.

IV. APPLICATION

In this section we apply the k-step PEM to actual data gathered from a conference room. First a brief setup description is given (see [10] for a thorough description), whereas in the subsequent part the actual identification results are presented and compared to the standard 1-step PEM approach.

A. Experimental Setup

The considered conference room is located in the ground floor of Fraunhofer IIS/EAS in Dresden and is part of a massive office building, which was constructed in the late 1950's. A scheme of the considered room is depicted in Figure 2.

In order to estimate a model for the room temperature $y[t] := \vartheta_r[t]$ the following measured input signals were available:

1) Heating power $\dot{Q}_h[t]$ [kW] supplied by the radiator heating
2) Cooling power $\dot{Q}_c[t]$ [kW] supplied by the fancoils
3) Outside temperature $\vartheta_{out}[t]$ [°C]
4) Solar radiation $\dot{Q}_{sol}[t]$ [W/m^2] on a horizontal plane, azimuth angle $\alpha_{az}[t]$ [°] and height $\alpha_h[t]$ [°] of the sun

Figure 3 illustrates an 18h long exemplary batch of data recorded in November 2015. Observe that the room temperature sensor shows significant quantization noise with a step size of approx. 0.32K. Notice further that the measurements of the heating power show erroneous peaks at the beginning of each heating period (see Fig. 3), which have been ignored throughout the paper and are caused by the measurement principle. The reason for this behavior is, that the difference of the supply and return temperature, which are measured before and after the radiators, is comparably high at the beginning of each heating period, because cool water is stored in the radiators. Once they had been flowed completely with "fresh" heating medium, the measurement of the heating power drops instantly and displays reliable values.

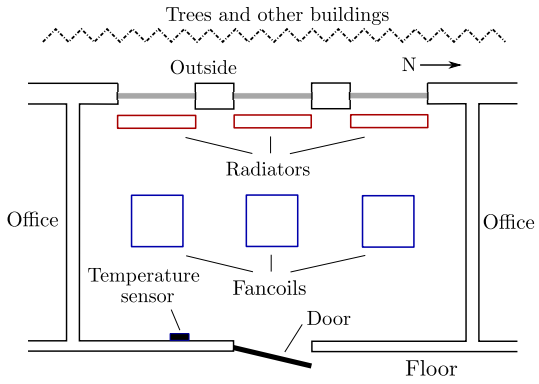

Fig. 2. Schema of considered conference room

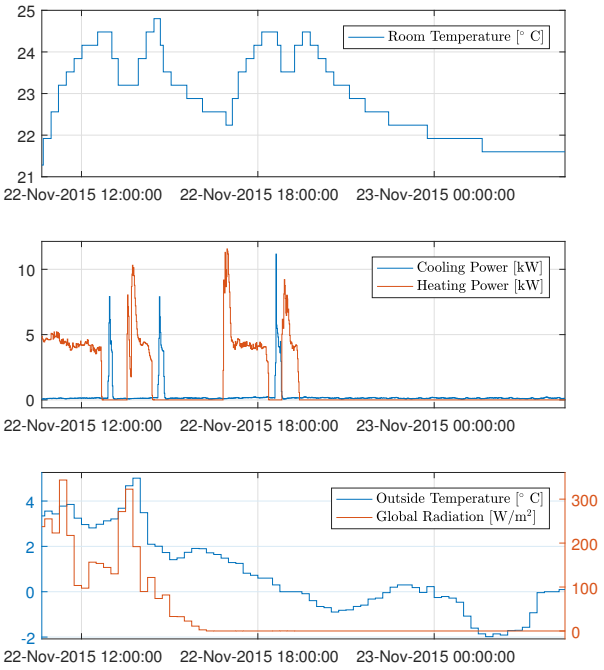

Fig. 3. Recorded data from November 2015

B. Model setup

Since a linear relation between the signals $\dot{Q}_h$, $\dot{Q}_c$, ϑ_{out} and the room temperature can be assumed, (1) - (3) will be used directly as inputs to the LTI block $G(q, \theta)$, whereas the signals

(4) are used to calculate the solar gain in the room[5] (see [10] for further discussion):

$$g_4(u,\eta) = \begin{cases} \dot{Q}_{sol}[t]\mathrm{sgm}(t,\eta)\frac{\cos\Theta[t]}{\sin\alpha_h[t]}, & \text{for } \Theta \leq \frac{\pi}{2}, \alpha_h > 0 \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

Here Θ denotes the angle between the normal vector of the window and the vector pointing into the center of the sun, that can be calculated by

$$\cos\Theta(t,\eta) = \begin{array}{l} \cos\alpha_h[t]\sin\beta_w\cos(\alpha_{az}[t]-\gamma_w) \\ +\sin(\alpha_h[t])\cos\gamma_w \end{array}, \tag{23}$$

where the known angles γ_w and β_w specify orientation of the window surfaces with respect to geographic north and the surface of the earth. Further the sigmoid function

$$\mathrm{sgm}(t,\eta) = \frac{1}{1+e^{-\delta(\alpha_h[t]-\alpha_{h0})}} \tag{24}$$

is used to describe the shading of the sun by the buildings and trees in front of the conference room (see Fig. 2), where δ and α_{h0} are scaling/shifting parameters of the sigmoid function.

C. Identification

In order to estimate a model of the room recorded data from the time period of September 2015 to September 2016 was considered. Due to obvious sensor faults (stuck sensor values) it was necessary to sort out corrupted data. Further, since the room has an binary occupancy sensor, we were able to preselect data of the unoccupied room (see [10] for accurate description of the room), resulting in 19176 samples with sampling time $T_s = 5\text{min}$ (≈ 66.6 days). The data was distributed on 98 datasets with minimum length of 4h, were $2/3$ of the datasets were utilized for estimation of the parameters θ and η, while the remaining data was used for cross-validation. The solution of the optimization problem (2) for the prediction horizon $k = 12$ (1h) was performed with the LevenbergMarquardt algorithm, were the initial guess for θ was computed by pseudo-linear regression algorithm (ARX/ARMAX/BJ) and instrumental variable (OE) method, respectively (see [11] and [20]).

Figure 4 illustrates the mean absolute error (MAE) of k-step ahead prediction errors on the crossvalidation datasets for $k = 1$ and $k = 12$ as a function of the model complexity n for different parameterizations of $G(q,\theta)$ and $H(q,\theta)$. Here n denotes the equal orders of the numerator and denominator degrees of $G(q)$ and $H(q)$. Thus for the ARMAX structure $n = n_a = n_{b,i} = n_c$ and $n_f = n_d = 0$ were assumed ($i = 1,\ldots,4$). Notice that the Box-Jenkins ($n = n_{b,i} = n_f = n_c = n_d$ and $n_a = 0$) and the output error ($n = n_{b,i} = n_f$ and $n_a = n_c = n_d = 0$) structures were parametrized by equal denominator polynomials $F(q)$ for all 4 inputs of $G(q,\theta)$.

[5]The distinction of cases in Equation (22) is due to the fact, that for angles $\Theta(t) > \pi/2$ the sun is not visible from the windows of the room and thus no direct radiation can enter through the windows. Notice furthermore that in some references the global radiation is split up to direct and diffuse radiation, which was ignored throughout the paper because of its small expected influence.

Figure 4 clearly shows that ARX structure is only suitable for long prediction horizons if the model order is chosen high enough, but generally performs well for 1-step ahead prediction. This is an expected behavior, since the ARX structure typically attenuates high frequencies [21]. Box-Jenkins and ARMAX structures perform good for 1- and for 12-step ahead predictions, where BJ structure has clear advantages for $n \leq 4$ (This is not surprising since BJ structures contain more adjustable parameters for equal n). Notice that ARMAX and BJ Models of degree ≥ 4 lead to a MAE of ≈ 0.136K for an prediction horizon of 1h. Since the output error structure has an constant frequency-weighting it shows almost similar prediction performance for $k = 1$ and $k = 12$. Further we'd like to mention, that OE and BJ structures often lead to numerical problems during the optimization procedure, whereas for ARX and ARMAX models the optimization procedure is less problematic[6].

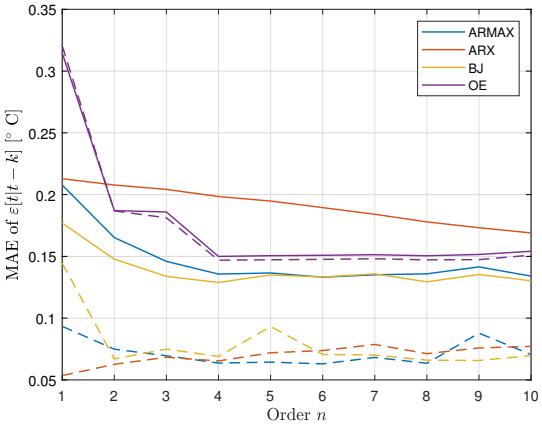

Fig. 4. Mean absolute error (2) on the cross-validation datasets as a function of order n of $G(q,\theta)$ and $H(q,\theta)$ for $k = 1$ (dashed) and $k = 12$ (solid).

Figure 5 illustrates the prediction performance on the cross-validation datasets for 6 estimated ARMAX-models with $n = 4$. It is clearly visible that k-step ahead criterion can indeed be beneficial in terms of prediction accuracy, especially if the initial values of the residual filter (10) are chosen to be 0 (initialization method (1)). The plot further demonstrates the high relevance of the initialization procedure of the residual filter (10), which is due to the fact, that multiple relatively short datasets ($4 - 60$h) had been used for identification. It can also be seen, that comparing to the LS-initialization, the backforcasting procedure leads to very similar results.

V. Conclusion

In this paper we demonstrated an approach for k-step identification of discrete-time Hammerstein models. Additionally we applied the algorithm to actual measured data of an unoccupied room and showed that the estimated models achieve a

[6]This statement basically corresponds to the reported behavior in [21] and [22] for 1-step PEM.

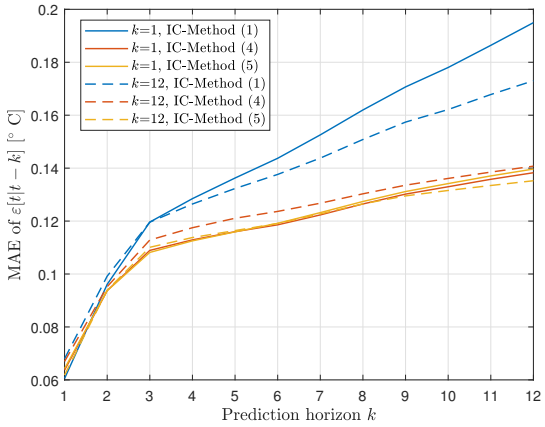

Fig. 5. MAE (2) on the cross-validation datasets as a function of prediction horizon for 6 different ARMAX models with $n = 4$, identified with 1-step and 12-step criterion, and different initialization methods of (10) (see Section III-B).

reasonable prediction performance. We further compared the k-step criterion to standard 1-step PEM and saw that a k-step criterion can indeed be beneficial in terms of prediction accuracy.

However, for actual MPC applications of the integrated room automation, still important issues need to be solved. First, due to the lack of heat meters, the heating and cooling powers $\dot{Q}_h[t]$ and $\dot{Q}_c[t]$ are usually not directly measurable in practical setups and can not be used as actual control variables for the process. Instead the supply temperature of the heating circuit and the valve positions of the thermostats should be used, which technically leads to bilinear identification and control problems.

Additionally in this contribution we only investigated the prediction performance of the model in case of an unoccupied room. The question how well the model performs in case of occupancy, particularly depends on the size of the room and the number of people, for which in practice usually no data is available. From the identification point of view the room occupancy represents an unknown input to the system, thus the application of blind system identification techniques might be investigated [23]. Furthermore CO2 and/or humidity sensors could be used to estimate the number of occupants, which then could be utilized as an additional input to the model.

ACKNOWLEDGMENT

The authors acknowledge the financial support by the Federal Ministry for Economic Affairs and Energy of Germany in the project FMopt (support code 03ET1372A).

Supported by:

Federal Ministry for Economic Affairs and Energy

on the basis of a decision by the German Bundestag

REFERENCES

[1] Y. Ma, F. Borrelli, B. Hencey, A. Packard, and S. Bortoff, "Model predictive control of thermal energy storage in building cooling systems," *Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, pp. 392–397, 2009.

[2] M. Gwender, D. Gyalistras, F. Oldewurtel, K. Wirth, B. Lehmann, V. Stauch, and S. Carina. (2010) Prädiktive Gebäuderegelung mithilfe von Wetter- und Anwesenheitsvorhersagen: Resultate des Projekts OptiControl.

[3] F. Oldewurtel, A. Parisio, C. N. Jones, M. Morari, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and K. Wirth, "Energy efficient building climate control using stochastic model predictive control and weather predictions," *American Control Conference*, pp. 5100–5105, 2010.

[4] E. Žáčeková, P. Samuel, and V. Zdeněk, "Model predictive control relevant identification using partial least squares for building modeling," *Australian Control Conference*, 2011.

[5] M. Y. Lamoudi, M. Alamir, and B. Patrick, "Model predictive control for energy management in buildings, part 2: Distributed model predictive control," in *IFAC Proceedings*, vol. 45, no. 17, 2012, pp. 226–231.

[6] D. Gyalistras, M. Gwerder, F. Oldewurtel, C. N. Jones, M. Morari, B. Lehmann, K. Wirth, and V. Stauch, "Analysis of energy savings potentials for integrated room automation," *Clima 2010 - REHVA World Congress*, 2010.

[7] F. Paschke, M. Franke, and J. Haufe, "Modellprädiktive Einzelraumregelung auf Basis empirischer Modelle," *Central European Symposium on Building Physics 2016 / Bausim 2016*, 2016.

[8] S. Privara, "Building modeling and identification for predictive control," *Energy and Buildings*, 2013.

[9] D. Sturzenegger, D. Keusch, L. Muffato, D. Kunz, and R. S. Smith, "Frequency-domain identification of a ventilated room for model based control," *Preprints of the 19th World Congress, The International Federation of Automatic Control*, pp. 593–598, 2014.

[10] F. Paschke and T. Zaiczek, "Identification of dynamic models for the short-term temperature prediction in a single room," *IFAC-PapersOnLine*, vol. 51, no. 2, pp. 79 – 84, 2018, 9th Vienna International Conference on Mathematical Modelling.

[11] T. Söderström and P. Stroica, *System Identification*. Prentice Hall International, 1989.

[12] T. Söderström and P. Stroica, "Uniqueness of estimated k-step prediction models of ARMA processes," *Systems & Control Letters*, vol. 4, no. 6, pp. 325 – 331, 1984.

[13] I. Goethals, K. Pelckmans, S. J. A. K., and B. De Moor, "Subspace identification of hammerstein systems using least squares support vector machines," *IEEE Transactions on Automatic Control*, 2005.

[14] F. Giri and E.-W. Bai, Eds., *Block-oriented Nonlinear System Identification*. Springer, 2010.

[15] A. Wills, T. B. Schön, L. Ljung, and B. Ninnes, "Identification of hammerstein-wiener models," *Automatica*, 2013.

[16] T. Knudsen, "A new method for estimating ARMAX models," *IFAC Proceedings Volumes*, 1994.

[17] T. Knudsen, "The initialization problem in parameter estimating for general SISO models," *IFAC Proceedings Volumes*, 1996.

[18] K. J. Aström, *Introduction to Stochastic Control Theory, Volume 70*. Academic Press, 1970.

[19] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems*. Kluwer Academic Publishers, 1996.

[20] P. Stoica, T. Söderström, A. Ahlén, and G. Solbrand, "On the convergence of pseudo-linear regression algorithms," *International Journal of Control*, 1985.

[21] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice Hall, 2012.

[22] O. Nelles, *Nonlinear System Identification*. Springer, 2001.

[23] C. Yu and M. Verhaegen, "Blind multivariable ARMA subspace identification," *Automatica*, vol. 66, pp. 3–14, April 2016.