

IDENTIFICATION OF WIENER MODELS USING OPTIMAL LOCAL LINEAR MODELS

Martin Kozek¹, Sabina Sinanović¹

¹Vienna University of Technology,
1040 Vienna, Gußhausstraße 27-29 / 325 A5, Austria

kozek@impa.tuwien.ac.at (Martin Kozek)

Abstract

The Wiener model is a versatile nonlinear block oriented model structure for miscellaneous applications. In this paper a method for identifying the parameters of such a model using optimal local linear models is presented. The linear model part is represented by a discrete-time transfer function and the non-linear characteristic is represented by piece-wise linear functions. Parameter estimation as well as partitioning of the local linear models is simultaneously accomplished by the identification procedure. The optimality of the proposed algorithm is threefold: First, each local model is linear in the parameters and therefore optimal parameter estimation methods like Recursive Least-Squares can be applied, thus leading to a robust solution. Second, the region of validity of each local model is adaptively optimized using the Chi-squared distribution of the estimated residual. This approach not only enables an automatic choice of the model size but it also incorporates the measurement noise level of the output variable into the result. And third, the resulting global model has a minimum of local models while guaranteeing optimal performance. A simulation of the pharmacological Propofol model is included, which documents the ability of the algorithm to balance the output noise with the systems nonlinearity.

Keywords: nonlinear identification, Wiener model, optimality, local linear models.

Presenting Author's biography

Martin Kozek was born in Vienna, Austria. He received the Diplom-Ingenieur in mechanical engineering from Vienna University of Technology, graduating with distinction in 1994. From 1995 on he was University Assistant at the Institute for Machine and Process-Automation. Between 1997 and 1998 he was visiting researcher at the University of Utah, USA, with a Kurt Gödel scholarship. He received his Doctorate from the Vienna University of Technology in 2000.



He is currently Assistant Professor at the Institute for Mechanics and Mechatronics at the Vienna University of Technology. His theoretical areas of research interests are primarily non-linear systems modeling and identification together with active control of structural vibrations. He has application emphasis on biomedical problems, vehicle dynamics, and process control.

1 Introduction

The identification of Wiener models is of great importance for different types of applications. Acoustics, meteorology, and pharmacological models of drug interaction [1] are but a few examples of successful application areas. The model structure is versatile due to its serial combination of a linear dynamic transfer function $G(z^{-1})$ and a nonlinear static characteristic (see fig.1).

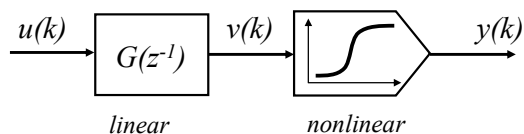


Fig. 1 Wiener model structure

In a typical experimental modeling setup only the input to the system $u(k)$ and the output $y(k)$ are measured. The intermediate variable $v(k)$ is not known, and moreover, the output $y(k)$ is corrupted by measurement noise. The complexity of the identification procedure strongly depends on the parameterization of the nonlinear static characteristic.

Several methods have been proposed for this purpose: Cubic splines [2], polynomials, linear static maps [3], and radial basis functions, e.g. [4]. Recently, local linear models have been successfully incorporated in a Recursive Least-Squares (RLS) estimation algorithm [5]. However, most of these structures require a proper a priori partitioning c_i of the unknown nonlinearity together with a predefined number of local models for a satisfying performance, see fig. 2.

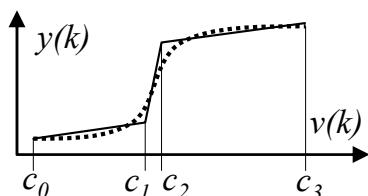


Fig. 2 Local linear models with partitioning c_i

It is obvious that a poor choice of the partitioning c_i will lead to a poor performance of the overall model, especially if the region of validity of the local models becomes too large and strong nonlinear effects lead to large local errors.

In this paper an algorithm is presented which automatically chooses the number of models and develops an optimal partitioning c_i . The algorithm utilizes a reconstruction of the intermediate variable $v(k)$ to assess the hypothesis of a linear model. This is achieved by means of the Chi-squared (χ^2) distribution of the residual for each local model. Using a ramp-like input signal with super-positioned noise for identification an optimal partitioning c_i can be found by a recursive method: The region of validity of each local model is increased until the hypothesis of

linearity tested by the variance of the local model fails. Only then a new local model is added.

Based on the RLS-algorithm presented in [5], this method balances the local model size with respect to noise level and model fit.

The remainder of the paper is structured as follows: In section 2 the parameterization together with the model structure and the RLS-algorithm are defined. In section 3 the criterion for optimal model size is explained in detail and the identification algorithm is presented. The statistical properties of the algorithm and their implications for practical implementations are discussed in section 4. Some simulation results for a pharmacological model are given and discussed in section 5, and a brief summary of the main contents concludes the paper.

2 RLS Identification

In this section the structure and parameterization of the Wiener model are defined such that each local model is linear in the parameters and therefore least squares methods for parameter estimation can be applied.

2.1 Linear dynamic model

The globally valid linear dynamic model is represented by a discrete-time transfer function $G(z^{-1})$ with input $u(k)$ and output $v(k)$:

$$G(z^{-1}) = \frac{b_m z^{-m} + \dots + b_1 z^{-1}}{a_n z^{-n} + \dots + a_1 z^{-1} + 1} = \frac{V(z)}{U(z)} \quad (1)$$

This representation can be easily transformed into a linear difference equation

$$v(k) + a_1 v(k-1) + \dots + a_n v(k-n) = b_1 u(k-1) + \dots + b_m u(k-m) \quad (2)$$

where the a_p and b_q are the parameters of the denominator and numerator polynomials, respectively. The intermediate variable $v(k)$ is therefore given by

$$v(k) = -a_1 v(k-1) - \dots - a_n v(k-n) + b_1 u(k-1) + \dots + b_m u(k-m) \quad (3)$$

2.2 Local model structure

The i -th local linear static model is defined by (see fig. 2)

$$y(k) = K_i v(k) + d_i, \quad (4)$$

with K_i being the gain and d_i being the intercept of the line. Rearranging (4) for $v(k)$ yields

$$v(k) = \frac{1}{K_i} y(k) - \frac{d_i}{K_i}. \quad (5)$$

Inserting (5) into (3) to eliminate the intermediate variable $v(\cdot)$ and solving for the output $y(k)$ gives

$$y(k) = -a_1 y(k-1) - \dots - a_n y(k-n) + b_1 K_i u(k-1) + \dots + b_m K_i u(k-m) + d_i [1 + a_1 + \dots + a_n]. \quad (6)$$

Equation (6) constitutes the local linear model

$$y(k) = \mathbf{x}(k) \boldsymbol{\theta}_i, \quad (7)$$

where the regressor vector is defined as

$$\mathbf{x}(k) = [y(k-1) \dots y(k-n) \quad u(k-1) \dots u(k-m) \quad 1] \quad (8)$$

and the parameter vector $\boldsymbol{\theta}_i$ is given by

$$\boldsymbol{\theta}_i = [-a_1 \dots -a_n \quad K_i b_1 \dots K_i b_m \quad d_i (1 + a_1 + \dots + a_n)]^T. \quad (9)$$

Eq. (7) defines a local linear dynamic model with an additional bias term. Obviously, the model is linear in the parameters $\boldsymbol{\theta}_i$ and depending on the noise spectrum a suitable linear parameter estimation algorithm can be used for local optimal estimation. The RLS-algorithm in a formulation without the need of a matrix inversion is given by the set of coupled recursive equations

$$\gamma(k) = \frac{P(k)x(k+1)}{1 + x^T(k+1)P(k)x(k+1)}, \quad (10)$$

$$\boldsymbol{\theta}_i(k+1) = \boldsymbol{\theta}_i(k) + \gamma(k)[y(k+1) - x^T(k+1)\boldsymbol{\theta}_i(k)], \quad (11)$$

$$P(k+1) = P(k) - \gamma(k)x^T(k+1)P(k). \quad (12)$$

It is important to note that the parameters of the linear dynamic transfer functions are globally valid for all local models. Since the individual parameters θ_{ij} in (9) contain the original parameters a_p , b_q , K_i , and d_i in a redundant way, the transfer function may be reconstructed assuming unity gain [5]. Therefore, the static gain between input $u(k)$ and output $y(k)$ is exclusively assigned to the local linear static model given by (4) and

$$G(z=0) = \frac{b_m + \dots + b_1}{a_n + \dots + a_1 + 1} = 1 \quad (13)$$

holds. In the remainder of this paper it is assumed that the RLS-algorithm is applied for the estimation of the unknown local parameter vector $\boldsymbol{\theta}_i$.

2.3 Global model structure

The global model structure can be written as a superposition of the individual local models

$$y(k) = \sum \phi_i \mathbf{x}(k) \boldsymbol{\theta}_i, \quad (14)$$

with ϕ_i being a validity function for each local model. The validity function can be chosen to be discrete (e.g. zero or one) or continuously differentiable [5]. It

should be noted that the validity functions are defined by the partitioning c_i described earlier. This means that each local model is valid from a certain point on the v -axis c_i up to the neighboring point c_{i+1} . This partitioning also permits nonlinear characteristics where a unique inverse does not exist.

The problem of an optimal global model can be stated as follows: How many local models should be chosen, and how to choose the partitioning of the nonlinearity (see fig.2)? These questions are addressed and answered in the following section.

3 Criterion for local model size

After introducing the basic idea for optimal model choice the algorithm will be presented in form of a flow-chart.

3.1 Fundamental idea

The fundamental idea of choosing the optimal local model size is to use the variance σ_{ei}^2 of the residual e_i of each local model. The residual σ_{ei}^2 of each local model is composed of the error covariance σ_v^2 (due to measurement noise) and the bias covariance σ_b^2 (due to model mismatch):

$$\sigma_{ei}^2 = \sigma_v^2 + \sigma_b^2 \quad (15)$$

For a given confidence level α (e.g. $\alpha=95\%$) the variance of the residual for a linear model is χ^2 -distributed with respect to the degrees of freedom (DOF) [6]. In this case the bias covariance σ_b^2 will be zero and only the measurement noise contributes to the residual. If there is a significant deviation from this distribution, the assumption of a linear model must be abandoned. In fig. 3 this optimal model size occurs at the sample denoted by c_1 (at DOF=10). A larger model will certainly be worse due to a bias error (the nonlinear characteristic is significantly different from the local model) and a smaller model will be worse due to a variance error (less DOFs result in a larger covariance of parameters).

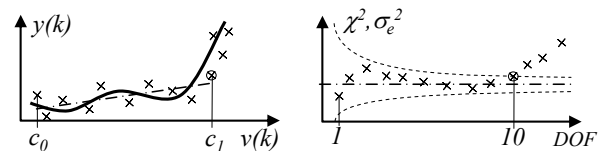


Fig. 3 Left: Local linear model. Right: Variance of the residual σ_{ei}^2 and χ^2 -distribution over model DOF.

The variance error is acceptable for a local linear model since it is inside the respective χ^2 -distribution (see fig. 3, right). The hypothesis of a linear model is still valid. In this case the region of validity for that specific local linear model can be extended until the bias error becomes dominant and the residual will come to lie outside the χ^2 -distribution.

If σ_v^2 is a priori known or estimated the respective χ^2 -distribution can be utilized as an upper bound for testing the linearity of the local model.

3.2 Algorithm for optimal model choice

In the case of a ramp-like input signal with additional white noise the algorithm proceeds in two nested loops.

In the inner loop the RLS-algorithm is employed for parameter estimation and recursive computation of the current residual. This residual is tested for linear model hypothesis using the χ^2 -distribution. If the hypothesis holds (meaning that a linear model is appropriate to explain the variation in the output), new measurement data are processed and the region of validity for the local model is automatically increased.

In the outer loop the RLS-algorithm is initialized and in the case of a rejected linear model hypothesis the parameters of the local model are stored together with the information on the region of validity (see fig.4).

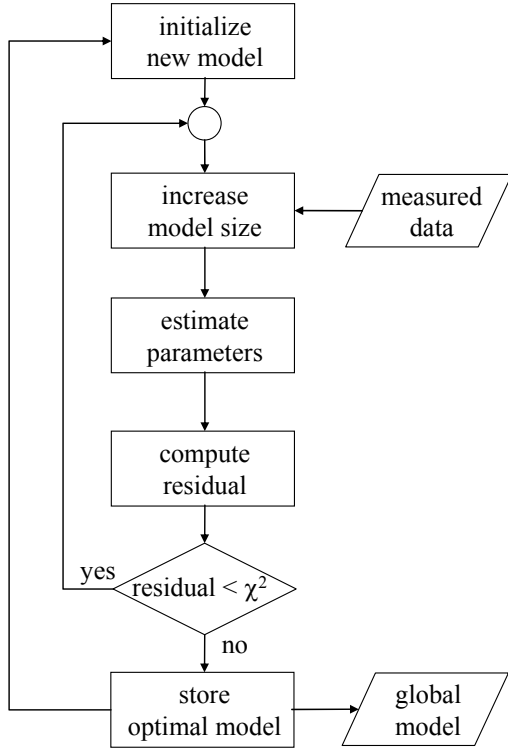


Fig. 4 Algorithm for optimal choice of local linear models using the χ^2 -criterion.

It should be noted that this criterion automatically adapts to the measurement noise level of the output. Since the bias of the linear model due to the nonlinearity σ_b^2 is masked by the covariance of the output noise σ_v^2 , small output noise levels will lead to more localized and accurate models while large output noise levels will yield more extended and less accurate models as depicted in fig.5. For a given nonlinearity the model size depends on the noise and confidence levels only.

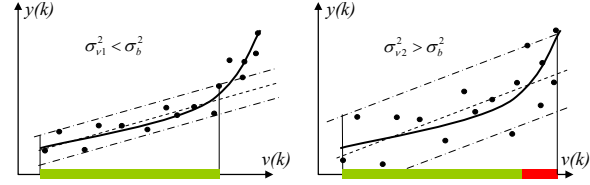


Fig. 5 Comparison of model size depending on noise level. Left: Small output noise leads to small and accurate models. Right: Large output noise leads to more extended less accurate models.

One of the main advantages of the proposed algorithm is the fully adaptive choice of model size regarding output noise level. There is only one scalar parameter for tuning of the models left to the user: The confidence level α for hypothesis testing. For a small α smaller and more accurate local models will result, while for a large α larger models with a smaller overall model complexity emerge.

3.3 Recursive formulation of ARX residual

For the ARX model with a transfer function of

$$G(z^{-1}) = \frac{b_m z^{-m} + \dots + b_1 z^{-1}}{a_n z^{-n} + \dots + a_1 z^{-1} + 1} = \frac{B(z^{-1})}{A(z^{-1})} \quad (16)$$

the residual $e(k)$ can be conveniently defined by

$$e(k) = A(z^{-1})y(k) - B(z^{-1})u(k). \quad (17)$$

Using the elements of the parameter vector defined in eq. (9) and allowing for the constant term $\theta_{i(n+m+1)}$ the residual of the individual local models $e_i(k)$ results in

$$e_i(k) = y(k) + \theta_{i1}y(k-1) + \dots + \theta_{in}y(k-n) - \theta_{i(n+1)}u(k-1) - \dots - \theta_{i(n+m)}u(k-m) - \theta_{i(n+m+1)}. \quad (18)$$

Based on the definition of the covariance of the residual

$$\sigma_{ei}^2(k) = \frac{1}{k-1} \sum_{j=1}^k [e_i(j) - \mu_{ei}]^2 = \frac{1}{k-1} \sum_{j=1}^k e_i^2(j) \quad (19)$$

the recursive formulation for the covariance of the residual σ_{ei}^2 in the i -th model is finally

$$\sigma_{ei}^2(k+1) = \frac{1}{k+1} [k\sigma_{ei}^2(k) + e_i^2(k+1)]. \quad (20)$$

Eqs. (18) and (20) can be easily integrated in existing recursive identification schemes (e.g. RLS) with a minimum of computational effort. It should be noted that for other than ARX models only eq. (18) has to be adjusted accordingly while eq. (20) is valid regardless of the noise model structure.

3.4 Input signals

The most favorable input signal $u(k)$ for applying the proposed criterion is a ramp with superimposed white noise (in order to guarantee persistent excitation). In that case the local models will almost always increase

from small to large values on the $v(k)$ -axis and each local model is optimized in a sequential procedure.

Nevertheless, algorithms have been proposed where arbitrary input signals can be applied [4]. These approaches require the additional incorporation of on-line adjustment of already existing model boundaries. Although the main idea of such an approach is already available it is beyond the scope of this paper to discuss such an algorithm here.

4 Statistical properties

The statistical properties of the Wiener model and the estimated parameters are important not only from a theoretical point of view, but they also give to the user additional information on the reliability of the identified model. One important feature is the prediction interval of the non-linear characteristic since it contains the output uncertainty of the identified model in the most relevant way.

4.1 Parameter covariance from RLS

The parameter covariance of the RLS-algorithm is automatically delivered by

$$\text{cov}(\boldsymbol{\theta}_i(k)) = \sigma_{ei}^2 P(k), \quad (21)$$

where $P(k)$ is the matrix defined in eq.(12). Therefore, the RLS-algorithm includes an on-line estimate of the parameter uncertainties. A problem associated with eq.(21) is that the parameter vector $\boldsymbol{\theta}_i$ contains the original system parameters in a lumped formulation. Only the covariance of the numerator coefficients a_p is explicitly contained. It is therefore not possible to compute prediction intervals for the non-linear characteristic from the parameter covariance alone.

4.2 Estimation of the intermediate variable

Using an estimate of the unknown intermediate variable $v(k)$ the above outlined problem can be overcome. If an estimate of $v(k)$ is available the prediction interval for each local model may be simply computed by a statistic based on the estimated data.

In principle, two possibilities for estimating $v(k)$ exist: Filtering the input $u(k)$ through the identified model of the transfer function (using eq.(3)) or using the measured output $y(k)$ as an input to the inverse local static model (utilizing eq.(5)). For two reasons the first mentioned approach is favorable. First, the input is assumed to be exactly known whereas the output is corrupted by measurement noise. This output noise will certainly corrupt the reconstructed intermediate variable $v(k)$ as well. This will be especially problematic in the case of a small gain K_i , since the inverse will amplify the noise level. Second, the transfer function model is globally valid. By means of eq.(21) the parameter set with the smallest error covariance can be utilized for estimation of $v(k)$.

The best choice for computing an estimate of the intermediate variable is given by

$$\hat{v}(k) = -\hat{a}_1 \hat{v}(k-1) - \dots - \hat{a}_n \hat{v}(k-n) + \hat{b}_1 u(k-1) + \dots + \hat{b}_m u(k-m), \quad (22)$$

which constitutes a simulated output using the identified transfer function with the most reliable parameter set.

4.3 Computation of the prediction intervals

The prediction intervals $p_i(v)$ are readily computed using a standard formulation for a data set with a linear regression model:

$$p_i(v) = K_i v + d_i \pm t_{\%_2, n-2} s_i \sqrt{1 + \frac{1}{n} + \frac{(v - \mu_v)^2}{\sum_{j=1}^n (v_j - \mu_v)^2}} \quad (23)$$

Here, K_i and d_i are the parameters of the linear local model, $t_{\%_2, n-2}$ is the $\%_2$ percentile of Student's t -distribution with $n-2$ degrees of freedom, s_i is the sample covariance, and μ_v is the sample mean of the intermediate variable. Extensive Monte-Carlo simulations have shown that the prediction intervals defined by eq.(23) do contain the expected percentage of data regardless of data amount or model parameters.

5 Simulation results

5.1 Pharmacological model

The Wiener model of the pharmacological Propofol model is studied [1]. The linear dynamic part represents a multi-compartment model for the distribution, elimination, and metabolism of the drug inside the body tissue (pharmacokinetics) and is given by the transfer function

$$G(z^{-1}) = \frac{0.37z^{-1} + 0.3z^{-2}}{1 - 0.83z^{-1} + 0.5z^{-2}}. \quad (24)$$

The nonlinear static part represents the effect of the drug concentration at the central nervous system (pharmacodynamics) and is defined by the Hill-equation

$$y(v) = 100 \frac{v^5}{400^5 + v^5}. \quad (25)$$

The Hill-Equation poses a strong nonlinear (sigmoid) relation between the drug concentration at the effect compartment (effect concentration) and the patient's reaction (BIS – bi-spectral index, a measure derived from EEG signals).

The input signal $u(k)$ consists of 200 data points and is defined by a ramp with superimposed white noise of variance $\sigma_u^2 = 15^2$. The output $y(k)$ is corrupted by measurement noise (an ARX noise model is assumed). Two cases are studied: 1) The variance of the

measurement noise is $\sigma_m^2=1.2$ resulting in 5 small and accurate local models. 2) $\sigma_m^2=48$ resulting in 3 larger and less accurate models. The nonlinearities for both cases are depicted in figs. 6 and 7.

The black dots represent the data from simulation (observe that the actual v -coordinates are not available for identification), the red crosses are the reconstructed data from the simulation, and local linear models together with prediction intervals for a confidence level of 90% are plotted in blue. The true nonlinearity is given by the red line.

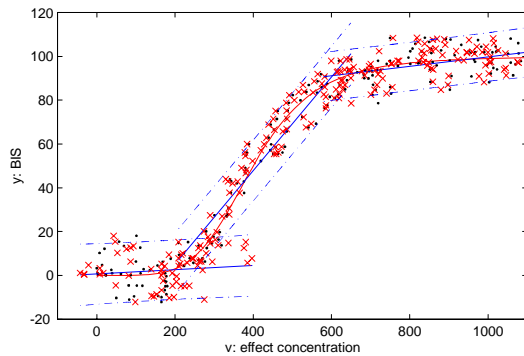


Fig. 6 Static nonlinearity with $\sigma_m^2=48$ resulting in 3 local models.

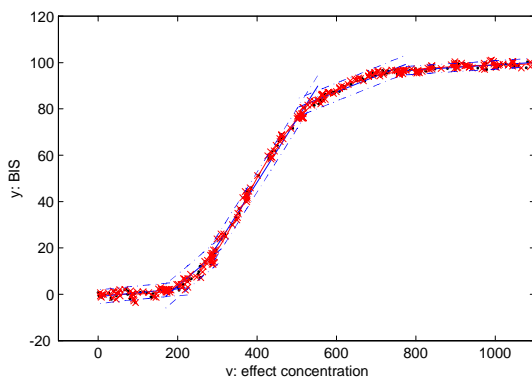


Fig. 7 Static nonlinearity with $\sigma_m^2=1.2$ resulting in 5 local models.

5.2 Identification results

From figs. 6 and 7 the performance of the proposed criterion is apparent: Using the χ^2 -distribution the information contained in the data is extracted in an optimal way. When a small output noise level is present (fig. 7) the nonlinearity can be detected earlier by the criterion and the models are automatically chosen smaller. In the presence of large output noise (fig. 6) nonlinearities are masked by the variance in the output variable $y(k)$ resulting in local models with a much larger region of validity. These models are obviously less accurate, however, they fully exploit the information in the data and the overall global model is significantly smaller.

Another important fact is the dependence of the parameter covariance on the slope of the local model. As can be seen in fig. 6, the first model with a region of validity in the variable $v(k)$ from approximately -70

to 400 obviously gives a poor estimation of the v -variable. Nevertheless, this does not affect the overall model performance, since both slope and gain are estimated correctly. In order to extract a partitioning c_i out of the local models the intersections of the local linear models can be easily computed.

6 Conclusion

In this paper a parameter estimation method for the identification of a Wiener model based on local linear models has been proposed. The model structure consists of a linear dynamic transfer function followed by a static map which is parameterized by piece-wise linear models. The region of validity of these models is simultaneously optimized with the parameter estimation utilizing the χ^2 -distribution of the covariance of the residual for a given confidence level. The local model is allowed to grow until the covariance of the residual is significantly exceeding the expected value for a linear model. A simulation of the pharmacological Propofol model clearly demonstrates the performance of the new criterion. It enables an automatic and optimal determination of the number of models, their respective region of validity, and guarantees optimal performance of the overall global model.

7 References

- [1] M. Kozek and N. Jovanović. Efficient Modelling of Propofol Pharmacokinetics: A Novel 2-Compartment Model. *BioMedSim03*, 113-118, Lebanon, May 28-30, 2003.
- [2] Y. Zhu. Parametric Wiener Model Identification for Control. *14th World Congress of IFAC*, Paper H-3a-02-1, 1999, 37-42, P.R. China.
- [3] T. H. Pelt and D. S. Bernstein, "Nonlinear system identification using hammerstein and nonlinear feedback models with piecewise linear static maps - part 1: Theory," in Proc. American Control Conference, (Chicago, Illinois), pp. 225-229, IEEE, June 2000.
- [4] S. Jakubek, C. Hametner, N. Keuth, and A. Voigt: Neue Methoden für die dynamische Identifikation mit Neuro-Fuzzy Netzen; *Automatisierungstechnik (at)*, 10 (2006), S. 486 - 494.
- [5] M. Kozek and Ch. Hametner. Block-oriented identification of Hammerstein/Wiener-models using the RLS-algorithm. *International Journal of Applied Electromagnetics and Mechanics*, **25**, 2007, 529-535.
- [6] D. Freedman, R. Pisani, and R. Purves. *Statistics*. Fourth Edition, W.W. Norton, 2007.