To be presented at the 13th IFAC Symposium on System Identification (SYSID 2003)
Rotterdam, The Netherlands, August 27-29, 2003

Session: Education and Training in System Identification

TEACHING SEMIPHYSICAL MODELING TO CHEMICAL ENGINEERING STUDENTS USING A BRINE-WATER MIXING TANK EXPERIMENT

D. E. Rivera [1]

Control Systems Engineering Laboratory
Department of Chemical and Materials Engineering
Arizona State University, Tempe, Arizona 85287

Abstract: The Chemical Engineering program at Arizona State offers an integrated series of core courses that teach students how conservation and accounting principles can be applied to describe engineering phenomena across disciplines. A brine-water mixing tank experiment was introduced in the third course in the series (ECE 394C: Understanding Engineering Systems Via Conservation) as a capstone modeling project for the recitation portion of the course. The experiment provides students with "hands-on" experience on a real-life system incorporating process, electrical, and mechanical components, as well as real-time data acquisition and control. A major feature of the brine-water tank project is that students apply a comprehensive system identification procedure relying on semiphysical (a.k.a. "grey box") models to complement their understanding of first-principles modeling. This paper describes the brine-water tank experiment, presents the formulation of the semiphysical parameter estimation problem, and describes the comprehensive procedure that students undertake to go from process data to validated plant models.

Keywords: system identification education, semiphysical modeling

1. INTRODUCTION

ECE 394 Systems, Understanding Engineering Systems via Conservation, is the third in an experimental core curriculum developed at Texas A&M which has been part of the Chemical Engineering curriculum since the fall of 1992. Students traditionally take ECE 394 Systems in the spring semester of their junior year. This four credit hour course is structured with three lecture hours a week and one weekly 2-hour recitation. The course stresses the broad-based use of accounting and conservation principles to model systems involving process, electrical, and mechanical components (separately and in combination). Another principal course objective is the use of computer-based tools to model engineering systems of practical interest.

[1] to whom all correspondence should be addressed; phone:(480) 965-9476, email:daniel.rivera@asu.edu

In ECE 394 Systems, students are confronted with the "reality" of engineering systems from the very first lecture. Students are made aware that real systems are:

- dynamic/unsteady-state ("steady-state is a figment of the imagination"),
- nonlinear,
- multivariable (i.e., possess multiple inputs and outputs),
- uncertain (i.e., models of real systems lack accuracy),
- stochastic (i.e., real systems are subject to random behavior, and as such cannot be always described by deterministic models). Precision errors will always be present in models.

Students are also presented in the first lecture (and frequently reminded thereafter) of the saying attributed to famous statistician Professor G.E.P. Box of the University of Wisconsin, "*all models are wrong, but* *some models are useful*." Students work in recitation as part of three-person teams. Two individual reports and three group presentations are required as part of the modeling project.

The brine-water tank experiment (Figure 1) is used in ECE 394 Systems as an ongoing project to bring students to reconcile the abstraction of mathematical modeling with the realities of a practical system. The main objective of this experiment is to develop, via first principles and semiphysical modeling techniques, *useful* mathematical models of the tank behavior displaying good predictive ability. Specifically, the students are asked to model the dynamic response of salt concentration in the outlet stream (c) and level in the tank (h) to changes in the inlet brine flowrate (q_c), the fresh water flowrate (q_w), and outlet flow (q_F). The tank is interfaced to a industrial-scale real-time computing platform, namely a Honeywell TotalPlant Solution System (previously known as the TDC3000, Figure 3). The engineer is capable of adjusting all three tank flows via the TDC3000 regulatory control points FIC100, FIC101, and LIC100 (see Figure 1). The experiment also requires students to generate a suitable calibration between the signal generated from an on-line conductivity sensor and the salt concentration (in g/ℓ) for the outlet stream in the tank.

The paper is organized as follows. Section 2 presents a brief description of the experimental apparatus. Section 3 discusses the first-principles model for the tank and the corresponding derivation of a semiphysical model for this system. Section 4 describes the steps involved in developing a comprehensive semiphysical modeling procedure, beginning from experimental design and concluding with model validation. Section 5 presents a summary and conclusions.

2. EXPERIMENTAL DESCRIPTION

Figure 1 shows both the process and the instrumentation used in this experiment. The flow of tap water to the process is regulated by measuring the flow with an orifice meter and changing the valve position on the water line according to an algorithm in a regulatory control point in the TDC3000. This control loop is assigned the tagname FIC100. Similarly, the flow of a concentrated salt solution is controlled with loop FIC101. The level in the tank is measured with a differential pressure cell (d/p) with one leg connected to the bottom of the tank and the other leg open to the atmosphere. The regulatory control point LIC100 compares this level with a desired level and manipulates the flow through the drain line. The salt concentration leaving and entering the tank is measured with conductivity cells and read into the system via analog input points CI100 and CI102, respectively. The conductivity measurements are displayed as the PVs (Process Values) of CI100 and CI102. By setting the appropriate instrument range limit parameters in the system (PVEUHI and PVEULO) the students are able to implement a linear correlation relating the raw 4-20ma signal from the conductivity cells to a sensible value for concentration in units of g/ℓ. CIC100 is a regulatory control point used in a subsequent course (ChE 461 Introduction to Process Control) (Rivera *et al.*, 1996) which adjusts the salt inlet flowrate setpoint (FIC101.SP) to keep exit stream salt concentration at setpoint (CI100.PV); students are asked to leave this point on MANUAL throughout the experiment.

3. BRINE-WATER TANK MODELING

3.1 *First-principles modeling*

During lecture and through homework assignments, students use MATLAB with SIMULINK to develop a first-principles dynamical model describing the effect of the various system inputs on the level and salt concentration. The principles of conservation of total mass and accounting of the salt species in the tank are used to derive this model. The level dynamics of the system are described by a differential equation arising from the conservation of total mass in the system

$$\frac{dh}{dt} = \frac{1}{A}(q_w - q_F + \frac{\rho_c}{\rho} q_c) \tag{1}$$

while the dynamics of salt in the tank are modeled by accounting for this species in the system:

$$\frac{dc}{dt} = \frac{q_c}{V}(c_c - \frac{\rho_c}{\rho} c) - \frac{q_w c}{V}, \quad V = hA \tag{2}$$

A, the crossectional area of the tank, ρ_c and ρ, the inlet brine and inlet water/outlet stream densities (respectively), and c_c, the inlet brine concentration, are constant-valued parameters in the model. An example of the SIMULINK window built by students is shown in Figure 2. Furthermore, Matlab with SIMULINK can be used compare the results of the first-principles nonlinear model with the responses obtained from its linearized equivalent at an operating condition; this enables students to evaluate the modeling errors associated with linearization.

3.2 *Semiphysical modeling*

The derivation of the semiphysical model follows along the line of the analysis presented in Lindskog (1996). Assuming constant volume in the tank (as the result of tight level control in the system) and constant densities for all streams, the first-principles model per Equations 1 and 2 reduces to:

$$\frac{dc}{dt} = \frac{q_c\, c_c}{V} - \frac{(q_c + q_w)\, c}{V} \tag{3}$$

Using a forward-difference approximation on the derivative (for a sampling time T) leads to

$$\frac{c(t+1)-c(t)}{T} = \frac{q_c(t)\,c_c(t)}{V} - \frac{(q_c(t)+q_w(t))\,c(t)}{V} \tag{4}$$

which solving for $c(t+1)$ yields

$$c(t+1) = c(t) + \frac{q_c(t)\,c_c(t)\,T}{V} - \frac{(q_c(t)+q_w(t))\,c(t)\,T}{V} \tag{5}$$

Rearranging and consolidating terms leads to the semiphysical structure

$$c(k) = c(k-1) + \theta_1 q_c(k-1)c_c(k-1) + \theta_2 q_c(k-1)c(k-1) + \theta_3 q_w(k-1)c(k-1) \tag{6}$$

Estimates of θ_1, θ_2, and θ_3 can be obtained from the first-principles model

$$\theta_1 = \frac{T}{V} \qquad \theta_2 = -\frac{T}{V} \qquad \theta_3 = -\frac{T}{V} \tag{7}$$

or alternatively, they can be estimated from plant data by recognizing that θ_1, θ_2, and θ_3 are linear in the parameters and hence linear regression can be readily applied.

4. A COMPREHENSIVE SEMIPHYSICAL MODELING EXPERIENCE

Having recognized that parameter estimation in semiphysical modeling constitutes a regression problem, students are then asked to perform a series of tasks that comprise a comprehensive identification procedure. These include:

(1) *Experimental Design*. Students are asked to use the first-principles MATLAB/SIMULINK model to design an informative experiment on the system. The design consists of a series of step changes of varying magnitude and duration that are intended to highlight the nonlinear behavior of the system and take into account the dominant time dynamics. The experiment must not exceed a 2 hr time period (the length of a recitation session) and must avoid taking the sensors and actuators past their limits. Figure 4 shows a TDC3000 data screen for a typical experimental run designed by the students. Various experimental runs are performed during the course of two weeks in the semester, and these are used to serve as estimation and validation data sets for the ensuing parameter estimation problem.

(2) *Model structure selection and parameter estimation*. Students are then asked to develop a Matlab program that uses regression analysis to estimate parameters of the semiphysical model. In addition to the three-parameter model structure shown in Equation 6, the program must also estimate parameters for the following difference equation model structures:

Four parameter model (Version A):

$$\begin{aligned} c(k) = {} & \theta_4 c(k-1) \\ & +\theta_1 q_c(k-1)c_c(k-1) \\ & +\theta_2 q_c(k-1)c(k-1) \\ & +\theta_3 q_w(k-1)c(k-1) \end{aligned} \tag{8}$$

Four parameter model (Version B):

$$\begin{aligned} c(k) = {} & c(k-1) \\ & +\theta_1 q_c(k-1)c_c(k-1) \\ & +\theta_2 q_c(k-1)c(k-1) \\ & +\theta_3 q_w(k-1)c(k-1)+\theta_4 \end{aligned} \tag{9}$$

Five parameter model:

$$\begin{aligned} c(k) = {} & \theta_4 c(k-1) \\ & +\theta_1 q_c(k-1)c_c(k-1) \\ & +\theta_2 q_c(k-1)c(k-1) \\ & +\theta_3 q_w(k-1)c(k-1)+\theta_5 \end{aligned} \tag{10}$$

The "four parameter" and "five parameter" models have more degrees of freedom and therefore allow greater flexibility in improving the goodness-of-fit as compared to the "three-parameter" difference equation.

(3) *Model Validation*. Ultimately, the goal of model validation is to determine the model structure and parameters leading to predictions that are both physically meaningful and result in lower errors when compared on a validation data set (i.e., a data set other than the one used for estimation). The semiphysical model estimates are compared against each other and against the responses obtained from the first-principles model (in both continuous-time and difference equation form). In addition, students are asked to compute, display, and plot the maximum and Root-Mean-Square (RMS) errors for both the estimation and crossvalidation data sets. The RMS and maximum errors are determined on the basis of the residual time series

$$e_{resid}(k) = c(k) - \hat{c}(k) \qquad k = 1,\cdots,N \tag{11}$$

which is the difference between the measured concentration ($c(k)$) and that estimated from a model ($\hat{c}(k)$). N is the total number of observations in the data set. The RMS error is computed as

$$RMSerr = \left(\frac{1}{N}\sum_{k=1}^{N} e^2_{resid}(k)\right)^{1/2} \tag{12}$$

while the maximum error consists of the largest absolute magnitude in the residuals, determined by

$$MAXerr = \max_k |e_{resid}(k)| \qquad (13)$$

$$k = 1, \cdots, N$$

(4) *Reflection*. Determining which model (semi-physical or first-principles) is "best" is not enough. Students are asked to examine their experience with the system and list all possible sources of error and prioritize them in order of importance. The inquisitive student will recognize problems related with the calibration of measurements, the relative effect of the simplifying assumptions, and similar circumstances. Ultimately, the students realize the importance of semiphysical modeling and of working with data as a valuable tool in modeling.

An illustration of the various steps with some representative test data sets is shown in Figures 5 through 12. These plots are generated using the Matlab/SIMULINK files developed by the students throughout the course of the semester. Estimation data (consisting in this case of one step change each for the inlet brine and fresh water flows) is shown in Figures 5 and 6. The relative agreement between the first principles and 3-parameter semiphysical model results can be seen in Figure 5. Simulation results that include the two four-parameter models and the five-parameter model are shown in Figure 7. All semiphysical models closely agree, and as reflected in the RMS values (Figure 8), increasing the number of parameters yields improved goodness of fit in the estimation data set. Parameter estimates are presented on the Matlab command window and compared to first-principles coefficients; Figure 9 shows the values obtained for the four-parameter model (Version A). Simulation results on the validation data set (Figure 10) indicate that this model has the best predictive ability over all the evaluated models. This is reflected in both a better visual fit in the simulation as well as superior RMS and MAX errors (Figures 11 and 12, respectively).

5. SUMMARY AND CONCLUSIONS

The brine-water mixing tank is a relatively simple experiment that, while originating from the field of chemical engineering, can be readily taught to students across disciplines. The experiment described in this paper exposes students to significant concepts in modeling, identification, and numerical computing in a challenging experimental and real-time information setting. Semiphysical modeling is introduced in a meaningful way while demanding only a modest mathematical background from students: knowledge of differential equations, basic numerical methods, and regression analysis. Copies of the Matlab/SIMULINK files implementing this procedure (as well as some sample data files) can be obtained by request from the author at (`daniel.rivera@asu.edu`).

6. REFERENCES

Lindskog, P. (1996). Methods, Algorithms, and Tools for System Identification Based on Prior Knowledge. PhD thesis. Linköping University, Sweden. Dept. of Electrical Engineering.

Rivera, D.E., K.S. Jun, V.E. Sater and M.K. Shetty, "Teaching Process Dynamics and Control Using an Industrial-Scale Real-Time Computing Environment," *Computer Applications in Engineering Education*, Computer-Aided Chemical Engineering Education Special Issue, Vol 4., No. 3, pp 191-205, 1996.

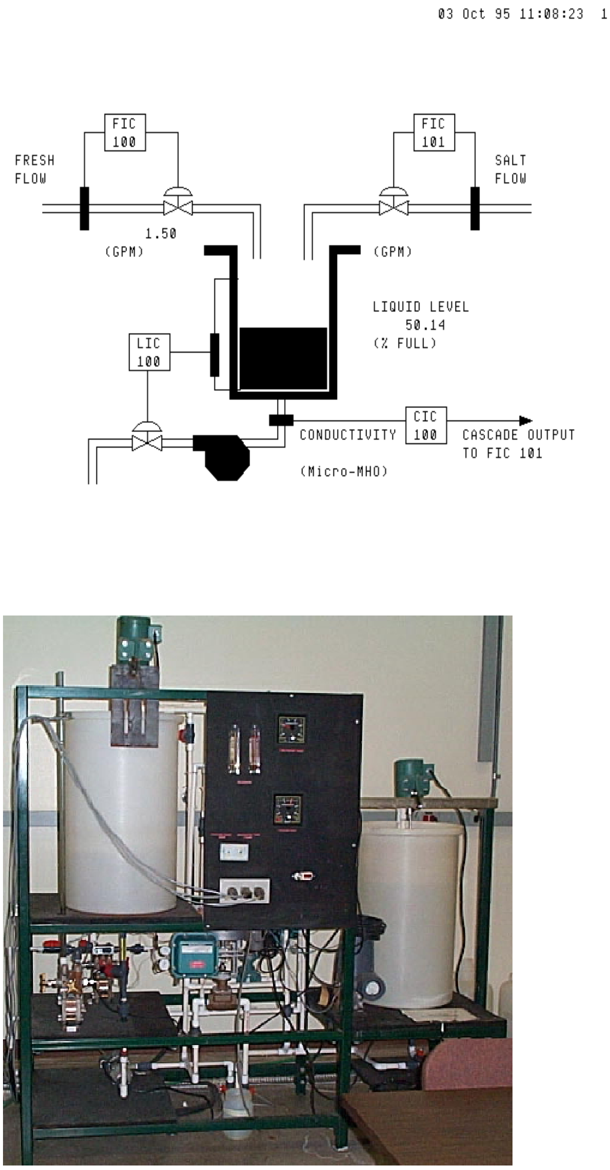

Fig. 1. Brine-water mixing tank schematic (top) and photograph (bottom).

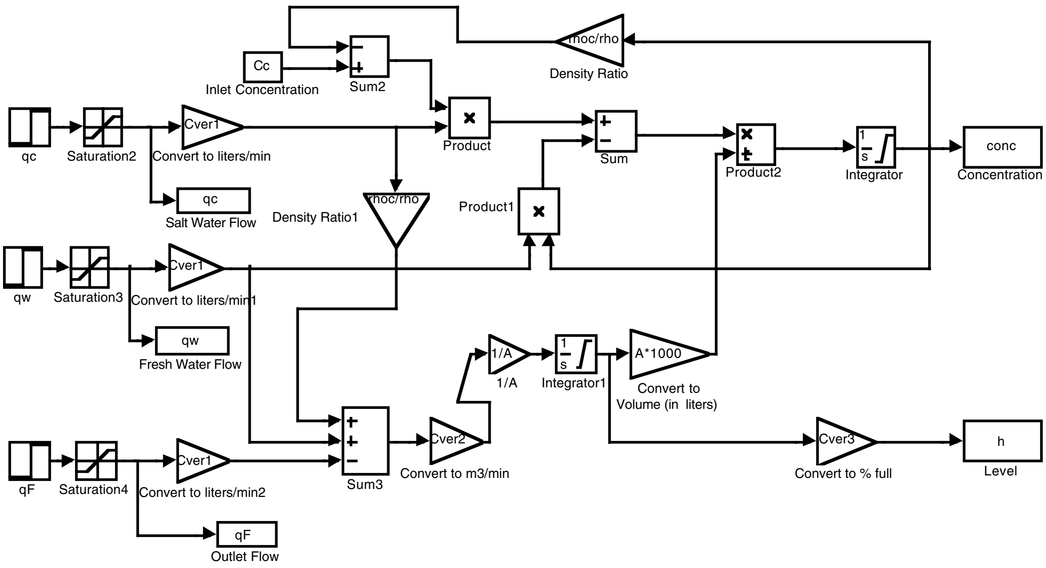

Fig. 2. SIMULINK window for Brine-Water Mixing Tank First-Principles Model

Fig. 3. Representative cluster of Universal and Global User Stations for ASU's TotalPlant Solution System.

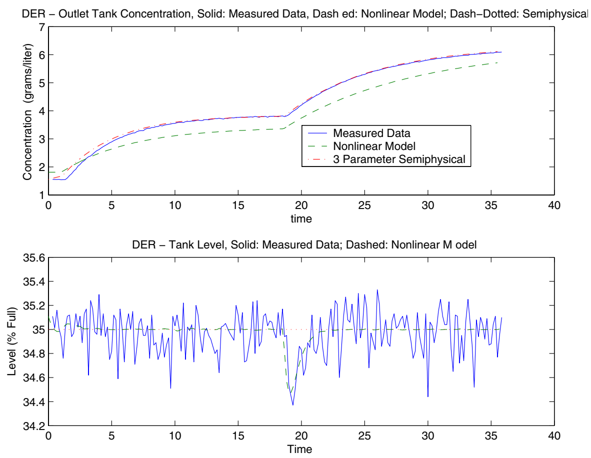

Fig. 5. Output time series for the estimation data set.

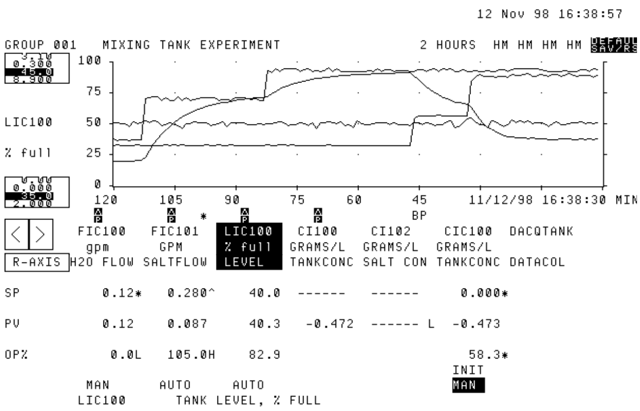

Fig. 4. Estimation data collected from the mixing tank, shown on a Honeywell TPS group display

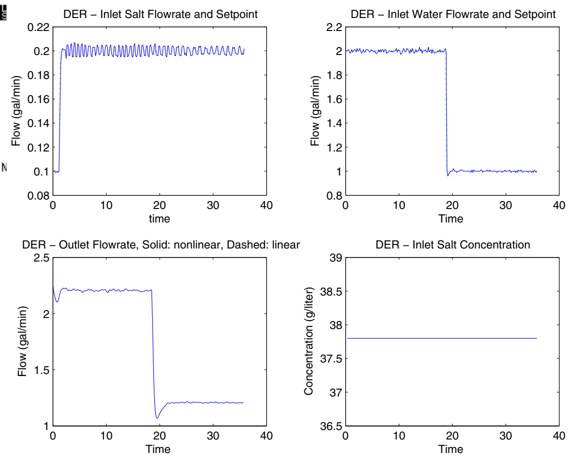

Fig. 6. Input time series for the estimation data set.

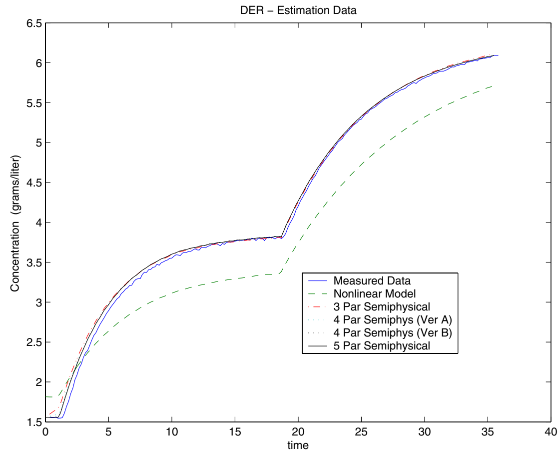

Fig. 7. Simulation results on the estimation data set for the first-principles and semiphysical models.

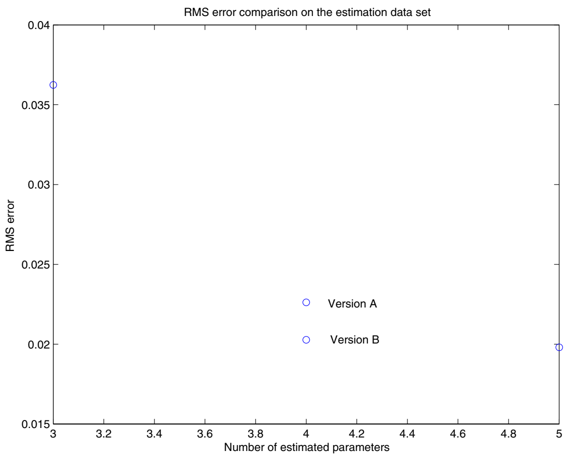

Fig. 8. RMS error comparison on the estimation data set.

```
Four Parameter Model - Ver A
Estimated Theta1  = 0.019993
First principles, Theta1 = 0.013557

Estimated Theta2  = 0.047278
First principles, Theta2 = -0.014069

Estimated Theta3  = -0.015284
First principles, Theta3 = -0.013557

Estimated Theta4  = 0.98189
First principles, Theta4 = 1

Estimation Data Results
RMS error = 0.022619
MAX error = 0.056326

Validation Data Results
RMS error = 0.22734
MAX error = 0.46546
```

Fig. 9. Parameter estimates for the four-parameter semiphysical model (Version A), compared with coefficients obtained from first-principles.

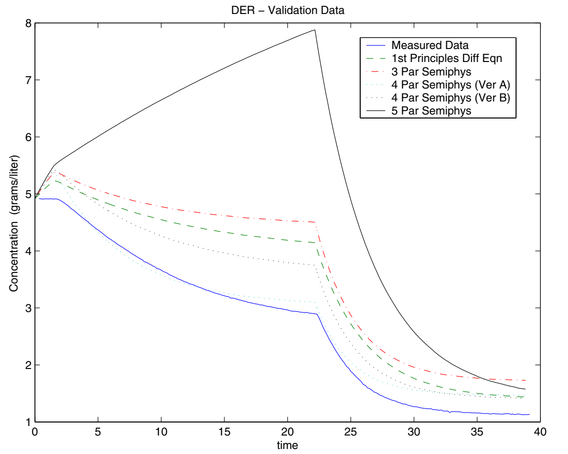

Fig. 10. Simulation results for the validation data set.

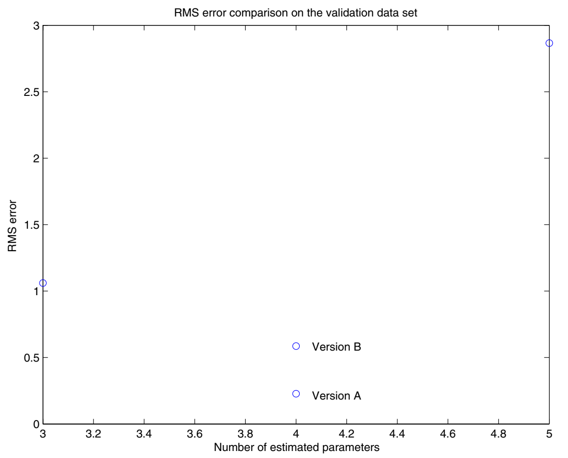

Fig. 11. RMS error comparison on the validation data set.

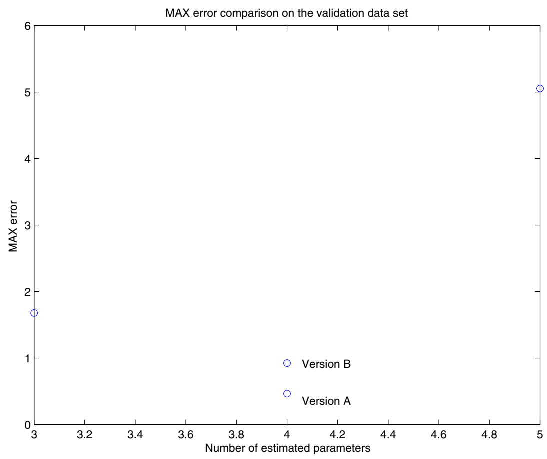

Fig. 12. MAX error comparison on the validation data set.