14th IFAC Symposium on System Identification, Newcastle, Australia, 2006
SAND2006-0050C

Input-Output Data Scaling for System Identification*

Seth L. Lacy[†] **Vit Babuška**[‡]

[†]Air Force Research Laboratory, seth.lacy@kirtland.af.mil
[‡] General Dynamics AIS, now with Sandia National Laboratories, vbabusk@sandia.gov

Abstract: System identification, or modelling from data, is used to develop dynamic models for control design, performance prediction, and system analysis. Some system identification algorithms are sensitive to the relative scaling of different input and output channels. Ideally, the identified model would be insensitive to arbitrary scaling of the input and output channels. For system identification algorithms that are scale-sensitive, re-scaling the input and output data can alleviate poor identification results due to initial inappropriate relative scaling of the data. In this paper we describe several methods for choosing weighting matrices for system identification with application to a laboratory experiment.

Keywords: Modelling, System Identification, Frequency Domain Identification, Subspace Methods, Structure Systems, Multivariable Systems, Discrete Time Systems, Control

1 Introduction

Models produced by some system identification algorithms depend on the scale of the data used to create them. If one sensor measures displacement in nanometers and another sensor measures displacement in parsecs, the identified model may differ from one identified where both sensors measure displacement in the same units. Not all system identification methods are scale-dependent. For example, maximum likelihood methods and the method of [10] are scale-insensitive. Since no synthesis methods exist for maximum likelihood identification and other non-convex optimization approaches, the modeler generally must synthesize a model using another method to be used as the initial condition for further optimization. Time domain scale-independent methods such as [10] may not be suitable for the identification of large precision space structures [9].

We categorize system identification algorithms into two types. The first category contains algorithms for model *synthesis*. The second category contains algorithms for model *tuning*. Model synthesis techniques are characterized by using only data to create a model. No initial condition or other information is required for the algorithm to estimate a model.

Once a model has been synthesized using a method such as [3, 10, 11, 13], an optimization problem can be developed based on the modeler's goals and constraints to solve for the model. This is model tuning. These model tuning techniques are iterative and require a good initial condition to guarantee convergence to a reasonable solution. Thus the system identification process generally consists of two steps, a synthesis step followed by a tuning step.

Typically, the goal of input/ouput data scaling is to reduce the disparity in the range of signal levels so that the model synthesis identification step considers all input-output channel pairs equally [12]. In this work we describe several methods for selecting input-output weighting matrices for model synthesis. Input/output scaling may also be used in the model tuning step through a judicious choice of model parameters (e.g., see [4]); however those types of methods do not apply to the model synthesis step where the initial model is created.

2 Problem Description and Notation

We are given input-output data, and are to develop a model describing the dynamics relating the commanded input to the measured output. In the time domain, the data appear as

$$U = \begin{bmatrix} u(1) & \cdots & u(N) \end{bmatrix}, \quad (2.1)$$
$$Y = \begin{bmatrix} y(1) & \cdots & y(N) \end{bmatrix} \quad (2.2)$$

where $U \in \mathbb{R}^{m\times N}$, $Y \in \mathbb{R}^{p\times N}$, and N is the number of samples available. The variables $u(k) \in \mathbb{R}^m$ and

*This work was supported in part by the AFOSR under laboratory research initiative 00VS17COR.

$y(k) \in \mathbb{R}^p$ are given by

$$u(k) = \begin{bmatrix} u_1(k) \\ \vdots \\ u_m(k) \end{bmatrix} \quad \text{and} \quad y(k) = \begin{bmatrix} y_1(k) \\ \vdots \\ y_p(k) \end{bmatrix} \tag{2.3}$$

where $u_i(k)$ is the input applied at time k to input channel i. Similarly, $y_i(k)$ is the measured response at time k for output channel i. In the frequency domain, the complex Frequency Response Function (FRF) data $G \in \mathbb{C}^{p \times m \times N_f}$ is written such that $G_{i,j}(f_k)$ is the response at output i, to input j, at frequency f_k.

In the time domain case, the weighting matrices $W_u \in \mathbb{R}^{m \times m}$ and $W_y \in \mathbb{R}^{p \times p}$ are applied as

$$\hat{U} = W_u U \tag{2.4}$$

$$\hat{Y} = W_y Y, \tag{2.5}$$

where $\hat{U} \in \mathbb{R}^{m \times N}$ and $\hat{Y} \in \mathbb{R}^{p \times N}$ represent the scaled data. In the frequency domain, the weighting matrices are applied at each frequency as

$$\hat{G}(f_k) = W_y G(f_k) W_u \tag{2.6}$$

where $\hat{G}(f_k), G(f_k) \in \mathbb{C}^{p \times m}$.

Once we have selected W_u and W_y, and synthesized an initial model based on the scaled data, we must invert the data scaling to return the model to the initial scaling. For a state space model, this is done as

$$A = \hat{A} \tag{2.7}$$

$$B = \hat{B} W_u^{-1} \tag{2.8}$$

$$C = W_y^{-1} \hat{C} \tag{2.9}$$

$$D = W_y^{-1} \hat{D} W_u^{-1}, \tag{2.10}$$

where $\hat{A}, \hat{B}, \hat{C}$, and $\hat{D}$ represent the model in scaled coordinates and A, B, C, and D represent the model in as-measured coordinates.

3 RMS Data Scaling in the Time Domain

One approach to selecting W_u and W_y for time domain data is to choose them such that

$$RMS(\hat{U}) = R_U \tag{3.1}$$

$$RMS(\hat{Y}) = R_Y \tag{3.2}$$

where $R_U \in \mathbb{R}^m$ and $R_Y \in \mathbb{R}^p$ are selected by the modeler and

$$RMS(X) \triangleq \begin{bmatrix} \frac{\sqrt{\sum_{i=1}^{N} \left(x_1(i) - \frac{1}{N}\sum_{j=1}^{N} x_1(j)\right)^2}}{N-1} \\ \vdots \\ \frac{\sqrt{\sum_{i=1}^{N} \left(x_n(i) - \frac{1}{N}\sum_{j=1}^{N} x_n(j)\right)^2}}{N-1} \end{bmatrix}. \tag{3.3}$$

This problem can be solved with positive definite diagonal matrices W_u and W_y

$$W_u = \text{diag}(R_U./RMS(U))) \tag{3.4}$$

$$W_y = \text{diag}(R_Y./RMS(Y))) \tag{3.5}$$

where the symbol "./" represents element-by-element division, and the output of the function $\text{diag}(u) : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal elements are the elements of the argument

$$\text{diag}(u) \triangleq \begin{bmatrix} u_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & u_n \end{bmatrix}. \tag{3.6}$$

One strategy for choosing R_U and R_Y is to set each element to 1. Another is to set them inversely proportional to the standard deviation of each channel, if these data are available.

4 RMS Data Scaling in the Frequency Domain with Condition Number Optimization

In the frequency domain the goal is to choose W_u and W_y such that

$$\sqrt{\frac{1}{N_f} \sum_{k=1}^{N_f} \left| \hat{G}_{i,j}(f_k) \right|^2} = \hat{R}_{i,j} \tag{4.1}$$

where $\hat{R} \in \mathbb{R}^{p \times m}$ is selected by the modeler. We also require that W_u and W_y both be invertible. To enforce the condition (4.1) and the invertibility condition, we define the optimization problem

Problem 1: Minimize $J_c(W_u, W_y)$ such that (4.1).

The cost function $J_c(W_u, W_y)$ is defined as

$$J_c(W_u, W_y) \triangleq \frac{\text{cond}(W_u)}{m} + \frac{\text{cond}(W_y)}{p}, \tag{4.2}$$

the sum of the condition numbers of W_u and W_y. In this case, W_u and W_y are generally fully populated matrices and $J_c \geq \frac{1}{m} + \frac{1}{p}$. Problem 1 can be solved using gradient search based methods.

One strategy for choosing $\hat{R}$ is to set $\hat{R} = 1_{p,m}$ where $1_{p,m} \in \mathbb{R}^{p \times m}$ is the $p \times m$ ones matrix. Another strategy is to set each element inversely proportional to the channel-by-channel standard deviation, in an attempt to synthesize a model suitable for maximum likelihood tuning. A third approach is to use this in an iterative fashion, at each iteration choosing $\hat{R}$ equal to the un-weighted RMS error from the last iteration.

5 RMS Data Scaling in the Frequency Domain with Shape Control

As in Section 4 we select W_u and W_y such that (4.1). However, instead of choosing W_u and W_y that satisfy (4.1) and have minimum condition numbers, we elect to minimize the difference in "shape" between the scaled and unscaled frequency response functions [1]. To do this, we define the inner product

$$G_{i,j} \cdot \hat{G}_{i,j} \triangleq \sum_{k=1}^{N_f} G_{i,j}(f_k)\text{conj}(\hat{G}_{i,j}(f_k)), \quad (5.1)$$

the norm

$$\|G_{i,j}\| \triangleq \sqrt{G_{i,j} \cdot \overline{G}_{i,j}} = \sqrt{\sum_{k=1}^{N_f} |G_{i,j}(f_k)|^2}, \quad (5.2)$$

and then the shape matrix $s(G, \hat{G}) \in [-1, 1]^{p\times m}$ is

$$s_{i,j}(G, \hat{G}) \triangleq \frac{\text{real}(G_{i,j} \cdot \hat{G}_{i,j})}{\|G_{i,j}\| \left\|\hat{G}_{i,j}\right\|}. \quad (5.3)$$

$s_{i,j}$ takes on values in $[-1, 1]$, with 0 representing orthogonal frequency response functions, 1 representing parallel SISO frequency response functions, and -1 representing anti-parallel SISO frequency response functions. This shape matrix s is a measure of how close two frequency response functions are to each other, ignoring scale. More precisely, note that $s_{i,j}(aG, b\hat{G}) = s_{i,j}(G, \hat{G})$ for any $0 \neq a, b \in \mathbb{R}^+$. The optimization problem is then

Problem 2: Minimize $J_{\text{s}}(W_u, W_y)$ such that (4.1). The cost function $J_{\text{s}}(W_u, W_y)$ is defined as

$$J_{\text{s}}(W_u, W_y) \triangleq \frac{\|1_{p,m} - s(G, W_y G W_u)\|_{\text{F}}}{2\sqrt{mp}} + \alpha J_{\text{c}}(W_u, W_y) \quad (5.4)$$

where $\alpha \in \mathbb{R}$ is chosen to weigh the relative importance of maintaining shape relative to minimizing the condition numbers of W_u and W_y. In this context, W_u and W_y are generally fully populated matrices. Problem 2 can be solved using gradient search based methods.

6 Non-iterative RMS Data Scaling in the Frequency Domain

A variation on the method described in Section 4 is to constraint the weighting matrices to be diagonal and have positive elements. Under this condition,

$$W_y R W_u = \hat{R} \quad (6.1)$$

where $R \in \mathbb{R}^{p\times m}$ is given by

$$R_{i,j} \triangleq \sqrt{\frac{1}{N_f} \sum_{k=1}^{N_f} |G_{i,j}(f_k)|^2} \quad (6.2)$$

This allows us to formulate an optimization problem for which a closed form solution exists.

Problem 3: Minimize $J_{\text{r}}(W_u, W_y)$ where

$$J_{\text{r}}(W_u, W_y) \triangleq \|V - \epsilon\|_{\text{F}} \quad (6.3)$$

$$V_{i,j} \triangleq \log \hat{R}_{i,j} \quad (6.4)$$

$$(6.5)$$

subject to $W_u \in diag(\mathbb{R}^{+m\times m})$, $W_y \in diag(\mathbb{R}^{+p\times p})$. The matrix $\epsilon \in \mathbb{R}^{p\times m}$ is a modeler selected full rank matrix with entries such that $\epsilon_{i,j} << 1$ and

$$\frac{max_{i,j}(\epsilon_{i,j})}{min_{i,j}(\epsilon_{i,j})} \approx 1. \quad (6.6)$$

The solution is

$$\begin{bmatrix} \text{vec}(W_y) \\ \text{vec}(W_u) \end{bmatrix} = \log^{-1}(M^* \text{vec}(\epsilon_{p,m})) \quad (6.7)$$

where * means pseudo-inverse, $\log^{-1}$ is the element-by-element exponential, vec is the matrix vectorization function and

$$M = \begin{bmatrix} I_p & 1_{p,1} & 0_{p,1} & \dots & 0_{p,1} \\ I_p & 0_{p,1} & 1_{p,1} & & \vdots \\ \vdots & \vdots & & \ddots & 0_{p,1} \\ I_p & 0_{p,1} & \dots & 0_{p,1} & 1_{p,1} \end{bmatrix}. \quad (6.8)$$

Because W_y, W_u are constrained to be diagonal, there are only $m + p - 1$ independent weighting parameters. The problem is underdetermined, so the *a-priori* effectiveness of this approach is not guaranteed. However, the benefit is that no numerical optimization is required.

7 Input-Output Normalization Data Scaling in the Frequency Domain

Another option is to compute the input and output scaling matrices independently in the spirit of the time domain scaling method of Section 3. The output weighting can be selected so that RMS levels of all the outputs are equal for equal input levels.

$$W_y = \text{diag}(1./S_c) \quad (7.1)$$

$$S_c \triangleq 1_{1,p} R \quad (7.2)$$

$$W_u = \text{diag}(1./S_r) \quad (7.3)$$

$$S_r \triangleq R\, 1_{m,1} \quad (7.4)$$

8 Experimental Example: Deployable Optical Telescope Plant

Here we apply the methods of Sections 4 to 7 to a laboratory example. The Deployable Optical Telescope (DOT) is a space traceable sparse-aperture telescope used to develop and evaluate technologies critical to the fielding of future large space telescopes [2, 5–9, 14–16]. The telescope has 10 actuators and 9 sensors. Frequency response function

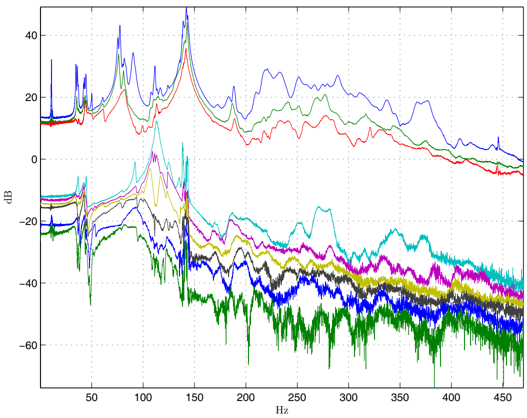

Figure 1: DOT Principal Gains

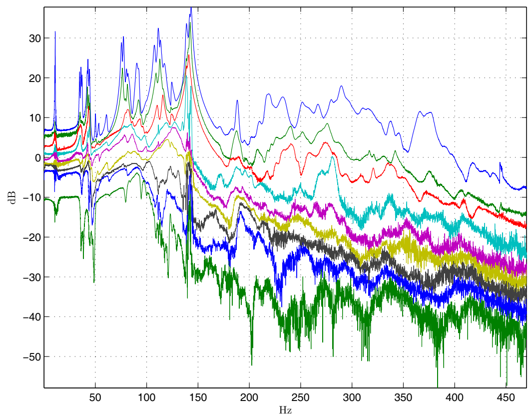

Figure 2: DOT Scaled Principal Gains

data were collected using sine dwell data collection methods. The singular values of the data are plotted in Figure 1. Note the large dynamic range and "gap" apparent in the principal gains.

First, scaling matrices were selected according the method of Section 5 with $\alpha = 0.01$. The original RMS matrix is

$$RMS(G) = \begin{bmatrix} 7 & 2 & 9 & 3 & 3 & 3 & 1 & 3 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 2 & 10 & 9 & 9 & 1 & 3 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 3 & 4 & 3 & 3 & 2 & 7 & 7 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{8.1}$$

where three sensors dominate the response of the system, and the ratio of the maximum to the minimum RMS value is

$$r(RMS(G)) \triangleq \frac{\max(RMS(G))}{\min(RMS(G))} \tag{8.2}$$

$$= 1,265.8, \tag{8.3}$$

and the scaled RMS matrix is

$$RMS(W_y G W_u) = 1_{9,10}. \tag{8.4}$$

The scaling matrices have

$$\text{cond}(W_u) = 7.7404 \tag{8.5}$$

and

$$\text{cond}(W_y) = 33.9753. \tag{8.6}$$

The principal gains of the scaled system are shown in Figure 2. Note that the gap between singular values 3 and 4 has disappeared. The shape matrix is depicted in Figure 3.

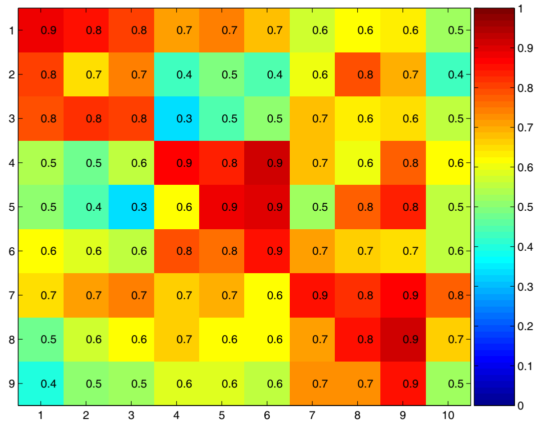

Figure 3: DOT Shape Matrix

The utility of the scaling methods of Sections 6 and 7 were also investigated with this system. All three data scaling methods reduced the ratio of maximum-to-minimum SISO RMS response levels, r.

Scaling Method	r	$\|G - H\|$	n
NONE	1265.81	2.59	200
Shape	1.00	3.44	194
Analytical Log	8.36	4.05	192
Analytical Sums	10.91	3.12	194

H is the unscaled identified model. The shape optimization method reduced the ratio maximally, but the non-iterative methods were effective also. The reduction in the dynamic range of the principal gains was comparable for all the scaling methods to that shown in Figure 2.

Now we will discuss the effects of the different scaling approaches on the properties of the identified model. The DynaMod [3] system identification software was used to synthesize models from the scaled and unscaled FRF data. The model order was selected to be 200 arbitrarily. The "ideal" model order was not obvious from the singular value plot. Each model was synthesized with 256 block rows and 400 was the upper bound on the model order

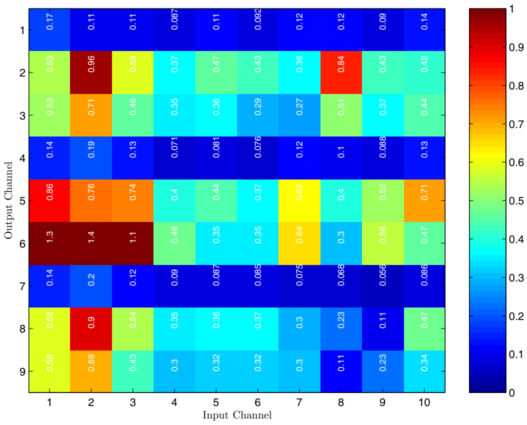

Figure 4: E for model identified from unscaled data

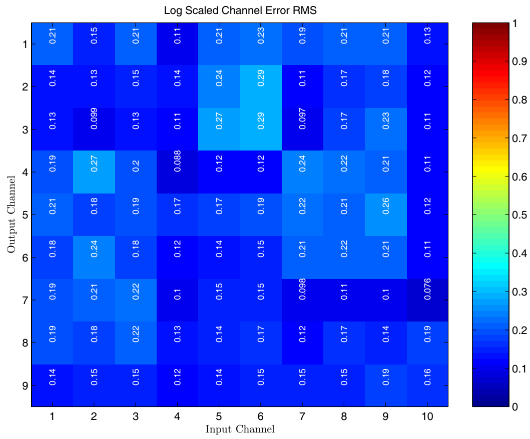

Figure 6: E for model identified from scaled data, see Section 6.

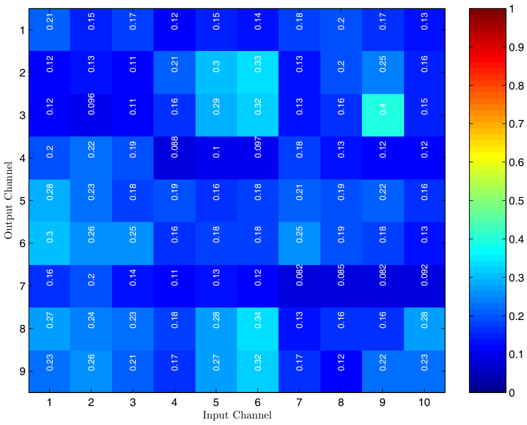

Figure 5: E for model identified from scaled data, see Section 5.

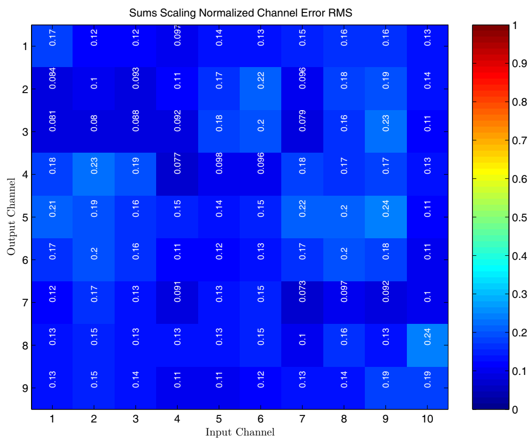

Figure 7: E for model identified from scaled data, see Section 7.

in the synthesis algorithm. All models synthesized with scaled FRFs had unstable poles which were truncated. The model identified from the unscaled FRF data was stable. The quality of the identified models was assessed with a normalized SISO error metric

$$E_{i,j} = \sqrt{\frac{\sum_{k=1}^{N_f} |H_{i,j}(f_k) - G_{i,j}(f_k)|^2}{\sum_{k=1}^{N_f} |H_{i,j}(f_k)|^2}}, \qquad (8.7)$$

the RMS of the error between the measured and model generated FRFs in each input/output pair, scaled by the RMS level of the measured FRF. H is the unscaled identified model. This gives a normalized measure of the level of the model error in each channel. See Figures 4 - 7.

While the RMS error increased for all of the models identified using scaled data compared to the model identified using the unscaled data, the normalized SISO error E was significantly decreased using all of the scaling approaches.

9 Conclusion

Some system identification algorithms are sensitive to input-output scaling. Different channels can have different ranges associated with them for several reasons: the units of the data with respect to the quantity being measured, the disturbance affecting each channel, and the system response to the input excitation. For system identification algorithms that are scale-sensitive, applying weighting matrices to the input and output data may alleviate poor identification results due to initial inappropriate relative scaling of the data. We described several methods for choosing weighting matrices for frequency domain system identification. We also presented experimental results from a structural control experiment.

The net effect of RMS input-output scaling may be to underweight the response of some paths in a frequency band as much as other paths may be overweighted in a frequency band without scaling. This means that applying the weighting matrices and synthesizing models for different frequency bands separately and then stitching the resulting models together to create the final identified model may improve the quality of the final, stitched, model.

Input-output data scaling must be used judiciously because it essentially warps the space from which model parameters are extracted. The scaling methods described in this paper effectively amplify low response level paths and diminish the high response paths. The net effect on the final unscaled model may be smaller errors in the low response paths at the expense of larger errors in the high response paths.

Based on our experiences, input/output data scaling may be beneficial when r is very large (e.g., $>$ 1000). Data scaling was very effective on the DOT system that satisfied this criterion, but had less dramatic results on a system with small r. The choice of scaling method depends on the problem. One should choose the method that leads to the best identified model; however we know of no theory to determine this a-priori. From a practical perspective, one should always start with the simplest approaches.

References

[1] V. Babuška, D. Carter, S. Lane, and S. L. Lacy. FRF correlation and error metrics for plant identificiation. In *AIAA Structures, Structural Dynamics, and Materials Conference*, Austin, TX, April 2005.

[2] K. D. Bell, R. L. Moser, M. K. Powers, and S. R. Erwin. Deployable optical telescope ground demonstration. In *Proceedings of the SPIE*, pages 559–567, Munich Germany, July 2000.

[3] C. Blaurock. *Dynamod Users Manual.* Midé Technology Corporation, 200 Boston Ave, Suite 1000, Medford, MA 02155, 2003.

[4] C.-T. Chou and J. Maciejowski. System identification using balanced parameterizations. *IEEE Transactions on Automatic Control*, 42(7):956–974, July 1997.

[5] R. S. Erwin, K. N. Schrader, R. L. Moser, and S. F. Griffin. Experimental demonstration of precision control of a deployable optics structure. *Journal of Vibration and Acoustics*, 124(3):441–450, July 2002.

[6] R. J. Fuentes, K. N. Schrader, M. J. Balas, and R. S. Erwin. Direct adaptive disturbance rejection and control for a deployable space telescope, theory and application. In *American Control Conference*, pages 3980–3985, Arlington VA, June 2001.

[7] J. B. Hoagg, S. L. Lacy, and D. S. Bernstein. Broadband adaptive disturbance rejection for a deployable optical telescope testbed. In *Proceedings the American Control Conference*, pages 4953–4958, Portland OR, June 2005.

[8] S. Huybrechts, P. Wegner, A. Maji, B. Kozola, S. Griffin, and T. Meink. Structural design for deployable optical telescopes. In *2000 IEEE Aerospace Conference Proceedings*, pages 367–372, Big Sky, MT, March 2000.

[9] S. L. Lacy, V. Babuška, K. N. Schrader, and R. J. Fuentes. System identification of space structures. In *Proceedings of the American Control Conference*, pages 2335–2340, Portland OR, June 2005.

[10] W. E. Larimore. ADAPT$_{\mathrm{X}}$ *Users Manual.* Adaptics, Inc, McLean VA, 1999.

[11] K. Liu, R. N. Jacques, and D. W. Miller. Frequency domain structural system identification by observability range space extraction. *Journal of Dynamic Systems Measurement and Control*, 118:211–220, 1996.

[12] G. Mallory. *Development and Experimental Validation of Direct Controller Tuning for Spaceborne Telescopes.* PhD thesis, Massachusetts Institute of Technology, 2000.

[13] T. McKelvey, H. Akçay, and L. Ljung. Subspace-based identification of infinite-dimensional multivariable systems from frequency-response data. *Automatica*, 32(6):885–902, June 1996.

[14] K. N. Schrader, R. H. Fetner, M. J. Balas, and S. R. Erwin. Sparse-array phasing algorithm based on recursive estimation of fringe contrast. In *Proceedings of the SPIE*, pages 146–157, Waikoloa, HI, August 2002.

[15] K. N. Schrader, R. H. Fetner, J. H. Donaldson, R. J. Fuentes, and S. R. Erwin. Integrated control system development for phasing and vibration suppression for a sparse-array telescope. In *Proceedings of the SPIE*, pages 134–135, Waikoloa, HI, August 2002.

[16] K. N. Schrader, R. H. Fetner, S. F. Griffin, and S. R. Erwin. Development of a sparse-aperture testbed for opto-mechanical control of space-deployable structures. In *Proceedings of the SPIE*, pages 384–395, Waikoloa, HI, August 2002.