The Parameter Signature Isolation Method and Applications

THE PARAMETER SIGNATURE ISOLATION METHOD AND APPLICATIONS

A Dissertation Presented

by

JAMES R. MCCUSKER

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2011

Mechanical and Industrial Engineering

THE PARAMETER SIGNATURE ISOLATION METHOD AND APPLICATIONS

A Dissertation Presented

by

JAMES R. MCCUSKER

Approved as to style and content by:

Kourosh Danai, Chair

Yossi Chait, Member

Christopher V. Hollot, Member

Donald Fisher, Department Chair
Mechanical and Industrial Engineering

ACKNOWLEDGMENTS

This dissertation would not have been possible without the support and guidance of a number of people.

I would like to express my sincere gratitude to my advisor, Dr. Kourosh Dania. Without his guidance throughout this process, none of this work would have been possible. His sound advice, in both research as well as in life, has shaped me into the engineer, researcher, and person that I am today. I am truly thankful for the mentoring that was provided by Dr. Danai, and will be forever grateful for his selfless approach to my education.

I would also like to thank Dr. Christopher Hollot and Dr. Yossi Chait for helping to guide my research as the members of my committee. Both Dr. Hollot and Dr. Chait have provided the resources necessary to gain the deep understanding in simulation modeling and system identification that was necessary for me to complete this work. I also would like to acknowledge the appreciation that I have for their diligence and patience throughout this process.

Significant contributions were also made to this research by Jeffery Simmons and the members of the Systems Modeling and Analysis group at Pratt and Whitney. As our contact at Pratt and Whitney, Mr. Simmons facilitated the ability to obtain the FANJETPW simulation model which was a basis for much of the research. He was also able to provide a perspective that contributed greatly to the application of this method to jet engine models.

I would also like to thank my friends and family that have provided the encouragement and support that made me the person that I am today. Without them,

none of this would have been possible. My mother and father, Patricia and Richard McCusker, furnished me, my brother, and sister with the most stable and supportive environment a child could grow up in. Throughout the years that were spent on this research, they have always provided encouragement and understanding that has given me the confidence to continue pursuing my dreams. Although words cannot express my gratitude, I would like to sincerely thank them for everything.

While in pursuit of this degree and throughout my life, both my brother and sister, Michael and Catherine McCusker, have set the examples for which I continually strive for. From my early childhood on, I always looked to my brother Michael and seen the prime example of how I would hope to be able to live my life. Regardless of how wrapped up I was in something, his uncanny ability to to put all of life's troubles, big and small, into perspective would always bring me back to earth. Although he may not have realized, this as well as the example that he set, helped me immensely over the last couple of year.

Ever since my sister and I were children, I tried to imitate her. Although, for our childhood, this was potentially annoying to her at times, I found that as we grew older, my drive to be like my big sister was more about the quality of the example she set rather than the thrill a younger brother gets from irritating his sister. While we were in the pursuit of our individual graduate studies, she set the shining example of how to balance research, family, friends and life. She helped me to focus when it was necessary while always being there for a much needed escape should either of us need it. Without her, the 'marathon' that is graduate school would not have been nearly as enjoyable as it was. Although we are far removed from our childhood in Carver, MA, I still see both of my siblings as my big brother and sister and I still strive to live by the example they set.

Finally, I would like to express my sincere gratitude to my wife, Paskalia McCusker, who has been endlessly supportive throughout this process. Her patience,

diligence, and drive have provided me with the solid grounding necessary to complete this work. Although our life together began while I was in pursuit of this degree, I feel that the greatest thing to come out of these last few years is the fact that I met her and been welcomed into the Papa family. Thank you for all your support and I am looking forward to everything that is to come.

ABSTRACT

THE PARAMETER SIGNATURE ISOLATION METHOD AND APPLICATIONS

MAY 2011

JAMES R. MCCUSKER

B.Sc., WENTWORTH INSTITUTE OF TECHNOLOGY

M.Sc., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Kourosh Danai

The aim of this research was to develop a method of system identification that would draw inspiration from the approach taken by human experts for simulation model tuning and validation. Human experts are able to utilize their natural pattern recognition ability to identify the various shape attributes, or signatures, of a time series from simulation model outputs. They can also intelligently and effectively perform tasks ranging from system identification to model validation. However, the feature extraction approach employed by them cannot be readily automated due to the difficulty in measuring shape attributes. In order to bridge the gap between the approach taken by human experts and those employed for traditional iterative approaches, a method to quantify the shape attributes was devised.

The method presented in this dissertation, the *Parameter Signature Isolation Method (PARSIM)*, uses continuous wavelet transformation to characterize specific

aspects of the time series shape through surfaces in the time-scale domain. A salient characteristic of these surfaces is their enhanced delineation of the model outputs and/or their sensitivities. One benefit of this enhanced delineation is the capacity to isolate regions of the time-scale plane, coined as *parameter signatures*, wherein individual output sensitivities dominate all the others. The parameter signatures enable the estimation of each model parameter error separately with applicability to parameter estimation. The proposed parameter estimation method has unique features, one of them being the capacity for noise suppression, wherein the feature of relying entirely on the time-scale domain for parameter estimation offers direct noise compensation in this domain. Yet another utility of parameter signatures is in measurement selection, whereby the existence of parameter signatures is attributed to the identifiability of model parameters through various outputs.

The effectiveness of PARSIM is demonstrated through an array of theoretical models, such as the Lorenz System and the Van der Pol oscillator, as well as through the real-world simulation models of an injection molding process and a jet engine.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

INTRODUCTION

Nonlinear dynamic models are the essential components of virtual environments that drive today's design, optimization, control, and automation practice. They are the natural choice for characterizing the behavior of biological, ecological, social, and economic systems, as well as artifacts like aircraft and manufacturing systems. But to be effective, models must have a high degree of accuracy to reliably represent the process. This underscores the significance of model validation concurrent with parameter estimation as part of the system identification process. Model validation entails ensuring the fidelity of the model in representing the process. Parameter estimation pertains to adjusting the model parameters (i.e., model coefficients and exponents) so as to maximize the accuracy of the model outputs relative to the experimental data available from the process. The focus of this research is to present an alternative to traditional approaches, which rely on a lumped representation of the simulation error, to instead utilize a feature extraction-based approach that is inspired by the expert's approach.

Traditional Approach

In the traditional approach to model validation the fidelity of the model is evaluated by how closely the model outputs $\hat{\mathbf{y}}(t, \mathbf{u}(t), \bar{\mathbf{\Theta}}) = M_{\Theta}(t, \mathbf{u}(t))$ match the process observations: $\mathbf{Y}^N = [\mathbf{y}(t_1), \ldots, \mathbf{y}(t_N)]^T$ sampled at $t_k \in [t_1, t_N]$. In the above formulation, M_{Θ} denotes the model structure, $\bar{\mathbf{\Theta}}$ is the nominal parameter vector used to simulate the model, and $\mathbf{u}(t)$ is the input applied to the process and used for

simulation. The traditional time-based measures such as the Akaike and Schwarz criteria [1, 37] and/or the whiteness of the prediction error $\epsilon(t)$:

$$\epsilon(t) = \mathbf{y}(t, \mathbf{u}(t)) - \hat{\mathbf{y}}(t, \mathbf{u}(t)) \tag{1}$$

and its correlation with inputs [41], focus on a cumulative comparison of the outputs as time series. As such, they ignore the localized similarities (in shape and size) of the model outputs to observations [60], that are the most revealing aspects of dynamic systems commonly considered by human experts in model development.

Like model validation, parameter estimation of nonlinear dynamic models [41] has traditionally been based on lumped characterization of the prediction error, regardless of the domain of representation (i.e., time, frequency, or time-scale (time-frequency)). Parameter estimation generally entails searching the parameter space $\mathbf{\Theta}$ to minimize a cost function V as

$$\hat{\mathbf{\Theta}} = \hat{\mathbf{\Theta}}(\mathbf{Y}^N) = \arg\min_{\mathbf{\Theta}} V(\mathbf{\Theta}, \mathbf{Y}^N) \tag{2}$$

where

$$V(\mathbf{\Theta}, \mathbf{Y}^N) = \sum_{k=1}^{N} L(\epsilon(t_k)) \tag{3}$$

is a scalar-valued (typically positive) function, defined by L, of the prediction error. In cases where the process can be accurately represented by a model that is linear-in-parameters [69, 26], the parameters can be estimated by linear regression [73]. Otherwise, parameter estimation becomes analogous to a multi-objective (due to multiplicity of outputs) nonlinear optimization [2] that can be solved by gradient-based methods [32, 70], genetic algorithms [27], convex programming [24], Monte Carlo optimization [67], or adaptive estimation techniques [68, 59]. Even though the above methods employ different strategies for search of the parameter space, they all rely on a lumped representation of the prediction error.

The Expert's Approach

In contrast to the traditional approach, the expert's approach relies on feature extraction. Features such as shapes are easily compared by human experts, due to the natural pattern recognition ability of humans. However, these features are not as readily quantifiable due to the difficulty to define measures that can quantify the shape attributes of time series [11]. It is recalled here that the correlation coefficient provides a measure of shape similarity between the time series. However, this measure which represents the overall shape similarity of time series pairs is deficient in considering the various shape attributes considered in visual inspection.

Traditionally, it is not an easy task to characterize the shape of the model outputs and observations. To illustrate the two aspects of comparison, consider the pressure measurements from an injection molding operation shown in Figure I.1, together with their estimated counterparts by two different models. The sum of the absolute prediction errors by Models A and B in Figure I.1 are 2198 and 1400, respectively, indicating that Model B provides a closer estimate of the experimental data. On the other hand, most of the error by Model A is associated with the last few data points of its output, and a decision on the quality of the model based solely on the magnitude of the prediction error may not be prudent. If judged by an expert, the conformity of the shape of models' estimates with the shape of observations would also be a factor in model validation. As is shown in this research, the transformation to the time-scale domain by continuous wavelet transforms provides the framework for assessing the shape similarities, and proves to be an effective method for model validation.

Similarly, for parameter estimation, experts rely predominantly on two types of information: (1) the level of uncertainty of parameters, and (2) the expected effects of these parameters on the outputs. They usually select a parameter, among the uncertain parameters, which is expected to most effectively reduce the simulation error

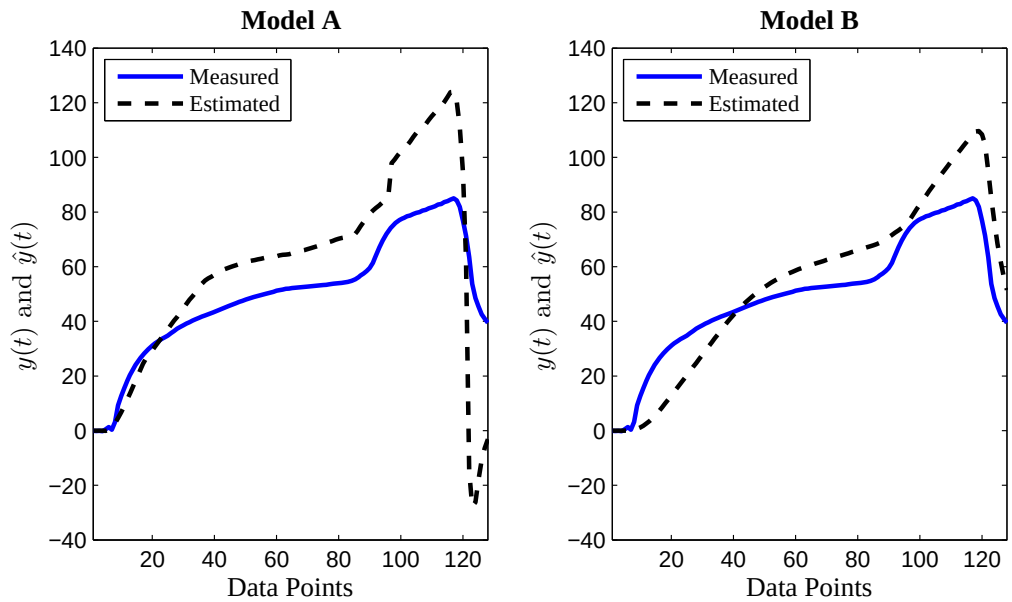

Figure I.1: Measured pressure from the mold during an injection molding operation shown with its estimates by two different models.

(i.e., the difference between the measured and simulated outputs) and make adjustments to it within its range. They then use this trial and error exercise to establish (learn) mental models of the effects of various model parameters on simulation errors.

To illustrate the expert's approach to parameter estimation and the formulation of mental models, the example of a cylindrical plunge grinding model is used. In this model, there are two uncertain parameters, the effective wear flat area (A_{eff}) and system stiffness (k_e). The expert's approach to parameter estimation entails adjusting one parameter at a time and observing the effect on the prediction error. A sample of such exploration is shown in Figure I.2, which illustrates the effect of perturbations in A_{eff} and k_e on the simulated power error. As observed from these results, the two parameters A_{eff} and k_e distinctly and uniquely affect the simulated power. Observed by a human expert, the change in the shape of the error caused by each parameter in Figure I.2 would be a component of the repository of mental models that constitute the effect of parameters. The expert would then identify the suspect parameter by contrasting the simulation error with the various effects in this

repository. In following the expert's approach, one can tune the parameter values until a desired output is achieved.

There are several advantages to the expert's approach. First, is its selectivity as to which model parameters it adjusts, so it is computationally simple. Second, is its applicability to linear and nonlinear models alike. Third, is its ability to consider the numerical range of model parameters during adaptation, thereby avoiding unrealistic adjustments. Fourth, is its sole reliance on the available measurements of the physical process from its normal operation, which obviates superfluous process runs solely for model adaptation through regression. The experts approach, however, lacks a well-defined procedure. Experts can usually consider a few parameters at a time and may not be able to consider all of the parameters collectively to identify the most erroneous parameter(s). For large systems with potentially many parameters to adjust, the haphazardness of the method could lead to impatience and possibly wrong conclusions about the validity of the model. However, if one could automate an approach inspired by the expert's approach, one could feasibly be able to benefit from its strengths while minimizing its weaknesses.

Problem Formulation

Let the model $M_{\Theta}(t, \mathbf{u}(t))$ accurately represent the process. Then the model outputs $\hat{\mathbf{y}}(t, \mathbf{u}(t)) = M_{\Theta}(t, \mathbf{u}(t))$ generated with the same input $\mathbf{u}(t)$ applied to the process will match the measured process outputs $\mathbf{y}(t, \mathbf{u}(t))$ (in mean-square sense) if the model parameters $\mathbf{\Theta} = [\theta_1, \ldots, \theta_Q]^T \in \Re^Q$ match the true parameters $\mathbf{\Theta}^* = [\theta_1^*, \ldots, \theta_Q^*]^T$; i.e.,

$$\mathbf{y}(t, \mathbf{u}(t)) = \hat{\mathbf{y}}(t, \mathbf{u}(t), \mathbf{\Theta}^*) + \nu \tag{4}$$

where ν represents unbiased measurement noise. If the model is identifiable [42]; i.e.,

$$\hat{\mathbf{y}}(t, \mathbf{u}(t)|\mathbf{\Theta}^{'}) \equiv \hat{\mathbf{y}}(t, \mathbf{u}(t)|\mathbf{\Theta}^{''}) \Longrightarrow \mathbf{\Theta}' = \mathbf{\Theta}'' \tag{5}$$

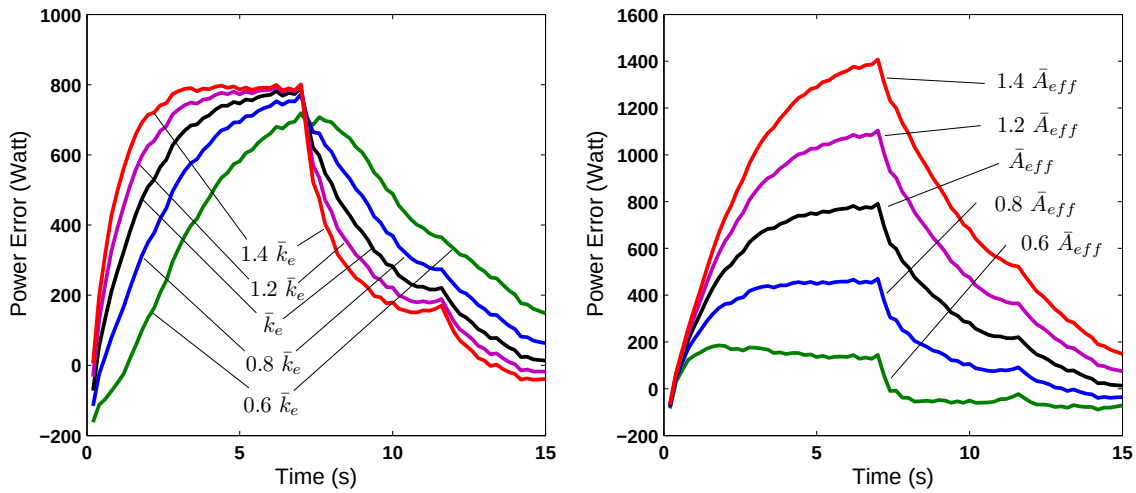

Figure I.2: Simulation error plots of grinding power obtained with perturbed values of effective wear flat area (A_{eff}) and system stiffness (k_e), with $\bar{A}_{eff}$ and $\bar{k}_e$ representing the nominal values of these parameters.

then

$$\mathbf{y}(t, \mathbf{u}(t)) \equiv \hat{\mathbf{y}}(t, \mathbf{u}(t), \bar{\mathbf{\Theta}}) + \nu \Longrightarrow \bar{\mathbf{\Theta}} = \mathbf{\Theta}^* \tag{6}$$

under the ideal conditions of adequately exciting input $\mathbf{u}(t)$ and reasonable signal-to-noise ratio.

Parameter estimation becomes necessary when the model parameters $\bar{\mathbf{\Theta}}$ are inaccurate, as represented by a nonzero prediction (simulation) error between the measured outputs $\mathbf{y}(t, \mathbf{u}(t))$ and model outputs $\hat{\mathbf{y}}(t, \mathbf{u}(t), \bar{\mathbf{\Theta}})$, as

$$\epsilon(t, \mathbf{u}(t), \mathbf{\Theta}^*, \bar{\mathbf{\Theta}}) = \mathbf{y}(t, \mathbf{u}(t)) - \hat{\mathbf{y}}(t, \mathbf{u}(t), \bar{\mathbf{\Theta}}) \tag{7}$$

For parameter estimation, PARSIM, like the gradient-based methods of parameter estimation, relies on a first-order Taylor series approximation of the model as

$$\mathbf{y}(t, \mathbf{u}(t)) \approx \hat{\mathbf{y}}(t, \mathbf{u}(t), \bar{\mathbf{\Theta}}) + \sum_{i=1}^{Q} \Delta\theta_i \left. \frac{\partial \hat{\mathbf{y}}(t, \mathbf{u}(t), \mathbf{\Theta})}{\partial \theta_i} \right|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}} + \nu \tag{8}$$

where $\Delta\theta_i = \theta_i^* - \bar{\theta}_i$ denotes the deviation of each parameter from its true value (parameter error) and $\partial\hat{\mathbf{y}}(t, \mathbf{u}(t), \mathbf{\Theta})/\partial\theta_i$ represents the vector of output sensitivity with respect to each parameter θ_i. By substituting (8) into (7), the prediction error can be approximated as the weighted sum of the output sensitivities as

$$\epsilon^N(t, \mathbf{u}(t), \mathbf{\Theta}^*, \bar{\mathbf{\Theta}}) \approx \mathbf{\Phi}\Delta\mathbf{\Theta} + \nu \tag{9}$$

where $\Delta\mathbf{\Theta} = [\Delta\theta_1, \ldots, \Delta\theta_Q]^T$ is the vector of parameter errors and $\mathbf{\Phi}$ denotes the matrix of output sensitivities with respect to the model parameters at individual sample points t_k, as

$$\mathbf{\Phi} = \begin{bmatrix} \partial\hat{y}(t_1, \bar{\mathbf{\Theta}})/\partial\theta_1 & \ldots & \partial\hat{y}(t_1, \bar{\mathbf{\Theta}})/\partial\theta_Q \\ \vdots & \ddots & \vdots \\ \partial\hat{y}(t_N, \bar{\mathbf{\Theta}})/\partial\theta_1 & \ldots & \partial\hat{y}(t_N, \bar{\mathbf{\Theta}})/\partial\theta_Q \end{bmatrix} \tag{10}$$

Each column of $\mathbf{\Phi}$, which characterizes the sensitivity of the output with respect to a model parameter in the interval $[t_1, t_N]$, comprises a component of the *parameter sensitivity* in traditional sensitivity analysis [25]. In PARSIM, these output sensitivities are approximated empirically, and, as such, they are referred to as *parameter effects* to underscore their potential differences with parameter sensitivities. The parameter effects are obtained empirically (in simulation) by perturbing one parameter at a time, as defined below.

Definition 1 The *parameter effect*, $\mathcal{E}$, is the change in the model's output due to a perturbation in model parameter. Formally, given input $\mathbf{u}(t)$ and a perturbation $\delta\theta_i$ to parameter θ_i, the parameter effect $\mathcal{E}_i$ is defined as

$$\mathcal{E}_i(t, \mathbf{u}(t), \bar{\mathbf{\Theta}}) = \frac{\hat{y}(t, \mathbf{u}(t), \bar{\mathbf{\Theta}} + \delta\theta_i) - \hat{y}(t, \mathbf{u}(t), \bar{\mathbf{\Theta}})}{\delta\theta_i} \tag{11}$$

Given the above definition, a parameter effect may be considered an approximation to the model output sensitivity with respect to the parameter, as

$$\left.\frac{\partial \hat{y}(t, u(t), \mathbf{\Theta})}{\partial \theta_i}\right|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}} = \lim_{\delta\theta_i \to 0} \frac{\hat{y}(t, u(t), \bar{\mathbf{\Theta}} + \delta\theta_i) - \hat{y}(t, u(t), \bar{\mathbf{\Theta}})}{\delta\theta_i} \approx \mathcal{E}_i(t, u(t), \bar{\mathbf{\Theta}}, \delta\theta_i) \tag{12}$$

and substituted in (9) to define the Taylor Series approximation of the prediction error in terms of the parameter effects, as

$$\epsilon(t, \mathbf{u}(t), \mathbf{\Theta}^*, \bar{\mathbf{\Theta}}) \approx \sum_{i=1}^{Q} \Delta\theta_i \; \mathcal{E}_i(t, \mathbf{u}(t), \bar{\mathbf{\Theta}}, \delta\theta_i) + \nu \tag{13}$$

It should be noted here that due to the dependence of the parameter effects, $\mathcal{E}_i$, on the nominal parameter vector $\bar{\mathbf{\Theta}}$ and parameter perturbations $\delta\theta_i$, the above formulation is nonlinear-in-parameter and, as such, it cannot be used for linear regression.

To illustrate the concept of parameter effects and their utility in approximating the prediction error, let us consider the error in the displacement of a nonlinear mass-spring-damper:

$$m\ddot{x}(t) + c|\dot{x}(t)|\dot{x}(t) + kx^3(t) = u(t) \tag{14}$$

where $x(t)$ denotes displacement, m the system's lumped mass, c its damping coefficient, k its spring constant, and $u(t)$ an external excitation force. Consider the prediction error, $\epsilon(t) = x(t) - \hat{x}(t)$, to be caused by a mismatch between the nominal parameters $\bar{\mathbf{\Theta}} = [\bar{m},\, \bar{c},\, \bar{k}]^T = [340,\, 10500,\, 125 \times 10^3]^T$ used to obtain $\hat{x}(t, u(t), \bar{\mathbf{\Theta}})$ and the true parameters $\mathbf{\Theta}^* = [375, 9800, 130 \times 10^3]^T$ used to obtain $x(t) = \hat{x}(t, u(t), \mathbf{\Theta}^*) + \nu$. The simulated prediction error due to an impulse input ($u(t) = \delta(t)$) with $\nu = 0$ is shown in the top plot of Figure I.3 (solid line) together with its approximation by (13) (dashed line). The parameter effects were computed according to (11) with the perturbations at 1% of the parameter values; i.e., $\delta\theta_i = 0.01\ \theta_i$. Also shown in the lower plots of this figure are the parameter effects of m, c, and k, which are the

constituents of the error approximation in (13). The results indicate that the error is closely approximated by the weighted sum of the parameter effects in the time-span of simulation.

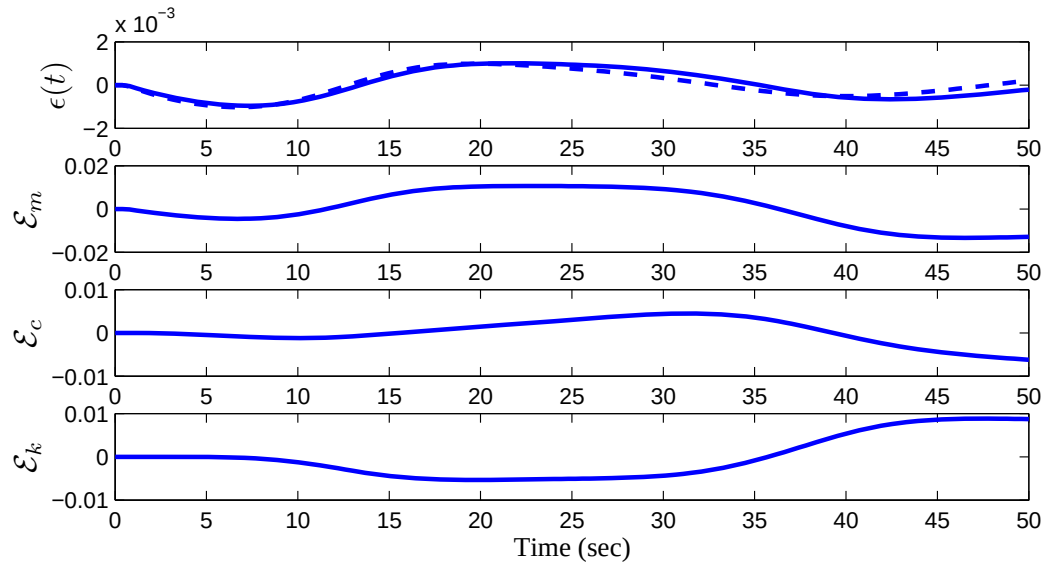

Figure I.3: Impulse response prediction error of the nonlinear mass-spring-damper system in (14) (top plot solid line) and its approximation by the weighted sum of parameter effects according to (13) (top plot dashed line). The lower plots are the parameter effects of m, c and k in (14) at the nominal parameter values $\bar{\mathbf{\Theta}} = [\bar{m},\, \bar{c},\, \bar{k}]^T = [340,\, 10500,\, 125 \times 10^3]^T$.

For cases where parameter estimation does not yield the true parameters, the source of the estimation difficulty could be the inaccuracy of the model, the inadequacy of excitation by the input, and/or the poor signal-to-noise ratio of observations. This leads to the necessity to analyze the validity of the simulation model. The presence of a nonzero prediction error would necessitate a validation metric to assess the closeness of various models to the process despite their erroneous model parameters for selection of the closest model. The discussion of the various validation metrics inspired by the expert's approach is presented in Chapter 2.

Preface

The innovation of this research is to expand the time series that represent the model outputs and observations into surfaces, via continuous wavelet transforms, so as to take advantage of the transparency afforded in the time-scale domain for system identification. We demonstrate here the methodology that can take advantage of the transparency afforded by this domain for model validation and parameter estimation [51, 13, 52, 50, 53]. This method which is coined as the *Parameter Signature Isolation Method (PARSIM)* is described in detail in Chapter 1.

Briefly, we have discovered that regions in the time-scale plane can be isolated wherein the prediction error can be attributed to the error of a single model parameter. At these regions, which are coined as *parameter signatures* in this research, one can estimate the error of each model parameter separately, for iterative parameter estimation [13] (see Chapter 3). The parameter estimation solution discovered has been shown, albeit anecdotally, that is less prone to local minima entrapments [52] and less sensitive to inadequate excitation [13] (see Section 3.4). The ability to perform parameter estimation directly in the time-scale domain also makes possible denoising in this domain without the need to reconstruct the signal back in the time domain. To demonstrate this feature, we have devised a technique that introduces a bias in parameter estimation according to the estimate of noise at each pixel (see Chapter 4).

Yet another benefit of operating in the time-scale domain is the transparency afforded for parameter identifiability. The ability to extract a parameter signature depends on the independence of the corresponding output sensitivity. This, in turn, informs the identifiability of each model parameter by the considered output. The final component of this research is the demonstration of the use of this concept in measurement selection of turbo-jet engines (see Chapter 5).

CHAPTER 1

THE PARAMETER SIGNATURE ISOLATION METHOD (PARSIM)

The Parameter Signature Isolation Method benefits from the enhanced delineation of transient signals in the time-scale domain. This enhanced delineation provides the capacity to isolate regions of the time-scale plane, coined as *parameter signatures*, wherein individual output sensitivities dominate the others. Due to this dominance, the prediction error can be attributed to the error of a single parameter at each parameter signature so as to enable estimation of each model parameter error separately. The focus of this chapter is to illustrate the underlying concepts of the Parameter Signature Isolation Method. The subsequent chapters discuss the applications of PARSIM.

1.1 Transformation to the Time-Scale Domain

The transformation to the time-scale domain is achieved through wavelet transform. A wavelet transform (WT) is obtained by the convolution of a wavelet function $\psi(t)$ with a time signal. The wavelet function can be manipulated in two ways: (i) it can be moved sideways (translated), as $\psi(t-\tau)$, to coincide with the different segments of the signal, and (ii) it can be widened (dilated) or narrowed (constricted), as

$$\psi_s(t) = \frac{1}{s}\,\psi(\frac{t}{s}) \tag{1.1}$$

where s denotes dilation (scale) parameter, to align with a larger or smaller segment of the signal at its current location. The transformation of a time signal $f(t)$ via the WT has the form [45]

$$W\{f\}(t,s) = \int_{-\infty}^{\infty} f(\tau)\psi_s^*(t-\tau)d\tau \tag{1.2}$$

where $\psi^*(t)$ is the complex conjugate of $\psi(t)$ and t denotes the translation (time) parameter. Since the wavelet has two degrees of freedom, namely translation and dilation, it can be used to analyze the signal in a narrow time window and with respect to a vast range of scales (frequencies) [78]. For illustration purposes, the Gauss WT of the prediction error in Figure I.3 is shown in Figure 1.1 where $W\{\epsilon\}(t,s)$ is comprised of a (128×72) matrix of wavelet coefficients. In analogy to images, hereafter the location of each wavelet coefficient will be refereed to as a "pixel" and represented by its coordinates: (t_k, s_l).

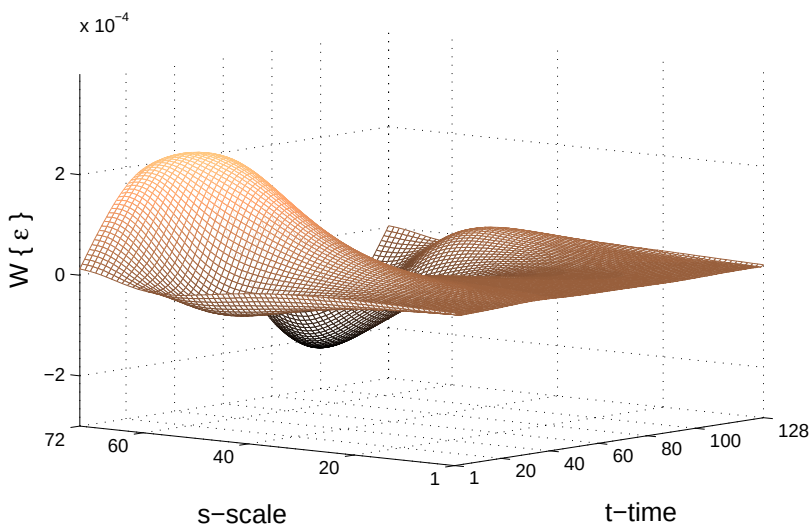

Figure 1.1: Gauss WT of the prediction error in Figure I.3.

Wavelet functions have been used in parameter estimation before, but mostly as basis functions to transform the original model into a function series that is linear in

parameters. In those cases, the wavelet coefficients, $W\{f\}(t,s)$, which represent the weights of individual functions; e.g., [26, 8] are estimated to minimize an objective function as in (3) in terms of the WT of the prediction error, $W\{\epsilon\}$. The original model parameters are then determined as a function of all the wavelet coefficients. The approach used in this research for parameter estimation contrasts the traditional approach in that small subsets of wavelet coefficients of the error are separately associated with individual model parameters so as to directly provide an estimate of the parameter's deviation from its true value in the time-scale domain. The key to this association between subsets of wavelet coefficients and model parameters is the differential nature of continuous wavelet functions [45], as described below.

Let $\beta(t)$ be any real smoothing function that has a nonzero integral and

$$\forall t \in \Re, \;\; |\beta(t)| \leq \frac{C_m}{1+t^m} \tag{1.3}$$

for some decay exponent m and constant C_m. For instance, one may consider this smoothing function to be the impulse response of a low-pass filter or the Gaussian function [46]. For function $\beta(t)$, $\beta_s(t)$ denotes its dilation by the scale factor s, as in (1.1), and $f * \beta_s(t)$ its convolution with $f(t)$ as

$$f * \beta_s(t) = \int_{-\infty}^{\infty} f(\tau)\beta_s(t-\tau)d\tau \tag{1.4}$$

Now if $\psi(t)$ is the nth order derivative of the smoothing function $\beta(t)$ as

$$\psi(t) = (-1)^n \frac{d^n(\beta(t))}{dt^n} \tag{1.5}$$

which has a zero average:

$$\int_{-\infty}^{\infty} \psi(\tau)d\tau = 0 \tag{1.6}$$

then it is called a wavelet, and its convolution with $f(t)$ is called the wavelet transform, $W\{f\}$, of $f(t)$, as defined in (1.2). Mallat and Hwang have shown that this WT is

a multiscale differential operator of the smoothed function $f * \beta_s(t)$ in the time-scale plane [45]; i.e.,

$$W\{f\}(t,s) = s^n \frac{d^n}{dt^n} \left(f * \beta_s(t)\right) \tag{1.7}$$

For instance, the Gauss wavelet which is the first derivative of the Gaussian function will result in a WT that is the first derivative of the signal $f(t)$ smoothed by the Gaussian function at different scales, and orthogonal to it. Similarly, the Mexican Hat wavelet will produce a WT that is the second derivative of this smoothed signal and the first derivative of the Gauss WT.

The above differential nature of continuous wavelet transforms can be used to delineate the differences between the parameter effects, due to the fact that differentiation accentuates the differences between signals. This is achieved by considering the parameter effects as time signals and transforming them to the time-scale domain by a continuous wavelet function. It is shown that given sufficient difference between the parameter effects, one can find regions in the time-scale plane wherein the WT of a parameter effect exceeds all the others.

To illustrate the enhanced delineation achieved in the time-scale domain, let us examine the singular values of the parameter effects of the nonlinear mass-spring-damper model in (14) at the nominal parameter vector $\bar{\mathbf{\Theta}} = [\bar{m}, \bar{c}, \bar{k}] = [383, 10290, 132600]$. In the time domain, the singular values, λ_{it}, are:

$$[\lambda_{1t}\ \lambda_{2t}\ \lambda_{3t}] = [2.8442\ 0.1414\ 0.0144]$$

but in the time-scale domain the singular values of the transformed parameter effects, λ_{iw}, will be different for each scale and the transformation function used. According to Principle Component Analysis [33], the more separated the characteristic roots (singular values) are, the more correlated the data. This separation of the singular values can be characterized by the larger value of the product index $\prod_{i=1}^{3} \lambda_i$ or the

smaller value of the ratio index λ_1/λ_3 [33]. For the above time-domain singular values, these indices are

$$\prod_{i=1}^{3} \lambda_{it} = 0.0058 \quad \text{and} \quad \frac{\lambda_{1t}}{\lambda_{3t}} = 197$$

and for the singular values in the time-scale domain, the above indices for the average singular values across all scales from transformations by the Gaussian smoothing function and the Gauss and Mexican Hat wavelets are shown in Table 1.1. Although the sum of each set is the same in both the time and time-scale domains; i.e.,

$$\sum_{i=1}^{3} \lambda_{it} = \sum_{i=1}^{3} \lambda_{iw} = 3$$

the product index of the average singular values obtained from transformation by the Gauss and Mexican Hat wavelets are larger than their time-domain counterpart, indicating less separation of the singular values in the time-scale domain with these transformations. It is also noteworthy that the ratio index continually decreases from transformation by the Gaussian smoothing function to those of the Gauss and Mexican Hat wavelets, which are the first and second derivatives of the Gaussian function, respectively. Equally noteworthy is the smaller separation of the singular values in the time-scale domain by the Gaussian function due to the smoothing effect of this function on the parameter effects.

Table 1.1: The indices of the average singular values of the transformed parameter effects in the time-scale domain by the Gaussian function and Gauss and Mexican Hat wavelets.

Function	$\prod_{i=1}^{3} \bar{\lambda}_{iw}$	$\bar{\lambda}_{1w}/\bar{\lambda}_{3w}$
Gaussian function	0.004	278
Gauss wavelet	0.0089	207
Mexican Hat wavelet	0.0202	162

1.2 Parameter Signatures

As mentioned earlier, PARSIM relies on identifying regions in the time-scale domain wherein a single parameter effect is much larger than all the others. These regions consist of pixels characterized by their coordinates (t_k, s_l), and the union of all pixels wherein a single parameter effect dominates the others is referred to as a *parameter signature*, as formally defined below.

Definition 2 If a pixel (t_k, s_l) exists which satisfies the following condition

$$|W\{\mathcal{E}_i\}(t_k, s_l)| >> |W\{\mathcal{E}_j\}(t_k, s_l)| \;\; \forall j = 1, \ldots, Q \neq i \tag{1.8}$$

then it is labeled as (t_k^i, s_l^i) and included in Γ_i, the parameter signature associated with θ_i. □

As a side note, analogous to the above in the time domain would be finding a component $\partial\hat{y}(t_k)/\partial\theta_i$ in the sensitivity matrix $\mathbf{\Phi}$ in (10) that would be larger than all the other components in that row. If such a single row with such characteristic could be found, it would be considered 'lucky'. Yet observations indicate that such isolated regions can be consistently found within the time-scale plane with different wavelets for all but parameters with collinear parameter effects (discussed in the next section).

The availability of parameter signatures Γ_i provides significant transparency to the parameter estimation problem by making it possible to attribute the WT of the prediction error

$$W\{\epsilon\}(t, s) \approx \sum_{i=1}^{Q} \Delta\theta_i \, W\{\mathcal{E}_i\}(t, s) + W\{\nu\} \tag{1.9}$$

to a single parameter error $\Delta\theta_i$ for all the pixels $(t_k^i, s_l^i) \in \Gamma_i$, as

$$W\{\epsilon\}(t_k^i, s_l^i) \approx \Delta\theta_i \, W\{\mathcal{E}_i\}(t_k^i, s_l^i) + W\{\nu\} \tag{1.10}$$

The above equation, which represents one of the Q single-parameter replacement equations to the multi-parameter error equation in (1.9), is the key to decoupling the prediction error into several single-parameter equations. Using the above single-parameter approximation of the error over the pixels $(t_k^i, s_l^i) \in \Gamma_i$, one can obtain the estimate of each parameter error as

$$\widehat{\Delta\theta}_i(\bar{\mathbf{\Theta}}) \approx \frac{1}{N_i} \sum_{k=1}^{N} \sum_{l=1}^{M} \frac{W\{\epsilon\}(t_k^i, s_l^i)}{W\{\mathcal{E}_i\}(t_k^i, s_l^i)} \quad \forall\, (t_k^i, s_l^i) \in \Gamma_i \tag{1.11}$$

where N_i denotes the number of pixels (t_k^i, s_l^i) included in Γ_i, M denotes the number of scales, and N is the number of time samples included in the data. The above estimate of parameter errors then provides the basis for estimating each parameter separately.

It is worth noting here that (1.11) can be regarded as a single-parameter gradient-based estimate in the time-scale domain. In Newton-Raphson method, for instance, that uses a gradient-based estimate for a model of the form $y = f(\theta)$, the parameter error is estimated as [14]

$$\widehat{\Delta\theta} = \frac{f(\bar{\theta})}{f'(\bar{\theta})} \tag{1.12}$$

where $\bar{\theta}$ denotes the current parameter value and f' the derivative of f with respect to θ. The parameter error estimate in (1.11) is identical to (1.12) except that it uses the average of the WT of f divided by WT of f' at the pixels included in the signature wherein a single-parameter model scenario holds.

Prior to analyzing the details of parameter estimation using PARSIM, the methods of parameter signature extraction must be discussed. Two different techniques have been developed for extracting the parameter signatures. One approach uses thresholds and the other a dominance factor.

1.2.1 Parameter Signature Extraction by a Threshold

The simplest technique of parameter signature extraction entails applying a common threshold to the WT of each parameter effect $W\{\mathcal{E}_i\}$ across the entire time-scale plane, and then identifying those pixels wherein only one $W\{\mathcal{E}_i\}$ is nonzero. The threshold operation takes the form

$$\overline{W\{\mathcal{E}_i\}}(t_k, s_l) = \begin{cases} 0 & |W\{\mathcal{E}_i\}(t_k, s_l)| < \eta_t \ \max_{(t,s)} |W\{\mathcal{E}_i\}| \\ W\{\mathcal{E}_i\}(t_k, s_l) & \text{otherwise} \end{cases} \tag{1.13}$$

where $0 < \eta_t < 1$ is the threshold value. Parameter signature extraction then entails searching in the time-scale plane for those pixels (t_k, s_l) where only one $\overline{W\{\mathcal{E}_i\}}$ is non-zero. The pixels labeled as $(t_k^i, s_l^i) \in \hat{\Gamma}_i$ then satisfy the following:

$$|\overline{W\{\mathcal{E}_i\}}(t_k^i, s_l^i)| > 0, \ \ \overline{W\{\mathcal{E}_j\}}(t_k^i, s_l^i) = 0 \ \ \forall j = 1, \ldots, Q \neq i \tag{1.14}$$

which is equivalent to:

$$\begin{aligned} |W\{\mathcal{E}_i\}(t_k^i, s_l^i)| &> \eta_t \max_{(t,s)} |W\{\mathcal{E}_i\}|, \\ |W\{\mathcal{E}_j\}(t_k^i, s_l^i)| &< \eta_t \max_{(t,s)} |W\{\mathcal{E}_j\}| \ \ \forall j = 1, \ldots, Q \neq i \end{aligned} \tag{1.15}$$

1.2.2 Parameter Signature Extraction by a Dominance Factor

Signature extraction based on a dominance factor complies more closely with the definition of parameter signatures. This entails a search of the time-scale plane to identify pixels that satisfy the notion of parameter signature by a dominance factor η_d. Such a search for the individual pixels (t_k, s_l) takes the form

$$\text{If } (t_k, s_l) \exists : \ \left|\overline{W\{\partial \hat{\mathbf{y}}_j / \partial \theta_i\}}(t_k, s_l)\right| > \eta_d \left|\overline{W\{\partial \hat{\mathbf{y}}_j / \partial \theta_m\}}(t_k, s_l)\right| \ \ \forall \, m = 1, \ldots, Q \neq i$$

$$\Longrightarrow (t_k, s_l) \in \Gamma_i \tag{1.16}$$

where

$$\overline{W\{\partial\hat{\mathbf{y}}_j/\partial\theta_i\}} = \frac{W\{\partial\hat{\mathbf{y}}_j/\partial\theta_i\}}{\max_{(t,s)} |W\{\partial\hat{\mathbf{y}}_j/\partial\theta_i\}|}$$

It is clear from (1.13) and (1.16) that both the threshold η_t and the dominance factor η_d affect the location as well as the size of the parameter signatures. However, they have different effects on the parameter signatures. This is illustrated via the parameter signatures shown in Figures 1.2 and 1.3 for each of the nonlinear mass-spring-damper model's parameters.

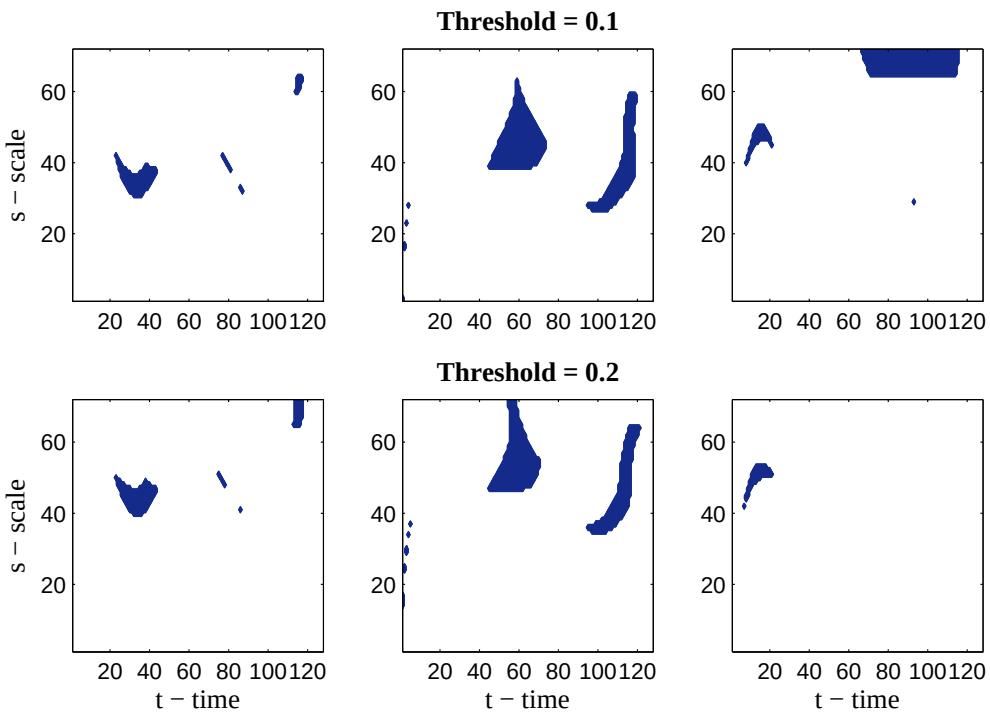

Figure 1.2: The location of pixels included in the parameter signatures of m, c and k (left to right) of the nonlinear mass-spring-damper model using Gauss WT and $\eta_t = 0.1$ (top) and $\eta_t = 0.2$ (bottom) in (1.13).

In Figure 1.2 we can see that as η_t increases from 0.1 to 0.2, the locations of the parameter signatures move even though for parameters m and c the overall size, in terms of the number of pixels, does not. We also see that for parameter k, there is a

large region present for $\eta_t = 0.1$. This is common in threshold-based signature extraction where slight changes in the threshold can allow for regions to either be included or excluded. This is due to the fact that using a threshold of 0.1, the magnitude at a pixel of $W\{\partial\hat{\mathbf{y}}_j/\partial\theta_i\}$ can exceed the threshold by an extremely small value whereas it may fall below the threshold by a minute value for another threshold. In these instances, these pixels do not adhere closely to the definition for a parameter signature. A possible strategy to ensure quality pixels is to apply a secondary threshold wherein a pixel is considered part of a parameter signature only if it exceeds η_t by a prescribed amount. Another strategy is to intelligently select the threshold based on the quality of the resulting parameter signature. While this is addressed further in Section 3.3.1, for the remainder of the dissertation, the secondary threshold approach is utilized unless otherwise noted.

A benefit as well as a potential drawback of threshold-based parameter signature extraction is the inherent bias towards selecting pixels in the higher scale (lower frequency) regions. This is due to basing the threshold on the maximum absolute value in the time-scale plane. Most often in transient signals, the highest magnitudes are predominantly in the high scale regions and, as a result, the lower scale regions are often discarded in parameter signature extraction. Where this is a benefit is in simulations that suffer from noise contamination due to noise being most prominent in the low scale (high frequency) regions. As a result, threshold-based signature extraction has an inherent capacity for noise suppression.

The alternative to thresholding is the dominance factor-based method. A sample of the parameter signatures of the nonlinear mass-spring-damper model parameters by the dominance factors of $\eta_d = 2$ and $\eta_d = 4$ are shown in Figure 1.3. Here we can see that as the dominance factor is raised, fewer pixels are included in the parameter signatures. If the dominance factor is increased substantially, it is possible that no parameter signatures could be extracted, and as it pertains to parameter estimation, it

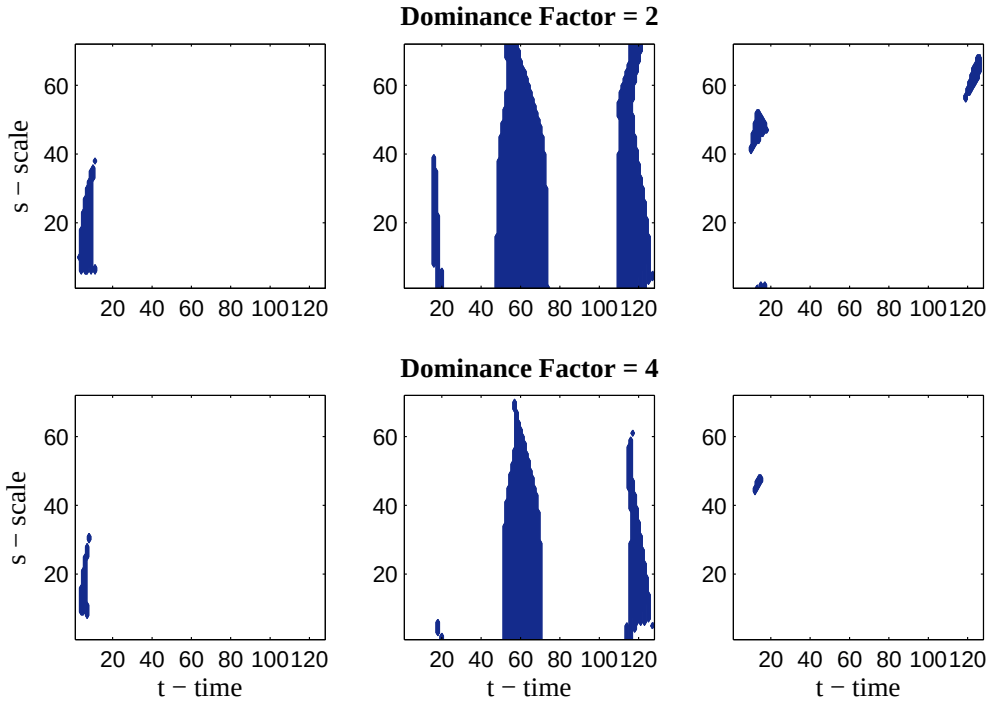

Figure 1.3: The location of pixels included in the parameter signatures of m, c and k (left to right) of the nonlinear mass-spring-damper model using Gauss WT and $\eta_d = 2$ (top) and $\eta_d = 4$ (bottom) in (1.13).

would lead to dormancy of parameter estimation. Although this method of parameter signature extraction adheres more closely to the definition of the parameter signature, it has a bias toward the lower scale regions. As stated previously, pixels in this region are more susceptible to the effects of noise and, therefore, this method is not advisable for systems experiencing a high degree of measurement noise. For the remainder of this dissertation, parameter estimation is demonstrated with the threshold-based method, whereas model validation and output selection are demonstrated via the dominance factor-based method of parameter signature extraction.

1.3 Constraint on Parameter Signature Extraction

Before proceeding to the utility of parameter signatures, it is important to identify circumstances in which parameter signatures cannot be extracted. In general, the

uniqueness of the parameter estimation solution is contingent upon the posterior identifiability of the model [75] which is a function of the input conditions and noise as well as the structural identifiability of the model [79]. But the ability to extract parameter signatures depends solely on the collinearity of parameter effects; i.e., $\mathcal{E}_i = \gamma\mathcal{E}_j$, which is synonymous with a unity correlation coefficient between pairs of parameter effects; i.e., $|\rho| = 1$. This constraint is stated in the following conjecture.

Conjecture Parameter signatures cannot be extracted for collinear parameter effects.

Rationale Collinear parameter effects will result in identical nonzero regions for $\overline{W\{\mathcal{E}_i\}}$ and $\overline{W\{\mathcal{E}_j\}}$ according to the threshold operation in (1.13), thus making it impossible to extract unique signatures for the corresponding parameters according to (1.14). □

One way to explain the above conjecture is to consider the WT of a time signal $\zeta_i(t)$:

$$W\{\zeta_i\} = \int_{-\infty}^{\infty} \zeta_i(\tau)\frac{1}{s}\psi\left(\frac{t-\tau}{s}\right) d\tau \tag{1.17}$$

as the cross-correlation of $\zeta_i(t)$ with $\psi_s(t)$ at different times t and scales s. The wavelet coefficients, which represent the cross-correlation values, will depend upon the magnitude of $\zeta_i(t)$ as well as the conformity of $\zeta_i(t)$ to the shape of the dilated $\psi(t)$ at different scales. The normalization of the wavelet coefficients according to $\max|W\{\zeta_i\}|$ in (1.13) nullifies the dependence of the wavelet coefficients on the amplitude of $\zeta_i(t)$ and leaves the correlation between the shapes of $\zeta_i(t)$ and $\psi_s(t)$ as the only factor affecting the magnitude of the WT at different times and scales. Accordingly, a signal $\zeta_1(t)$ that is only slightly different from $\zeta_2(t)$ will correlate more than $\zeta_2(t)$ with $\psi_s(t)$ at some combination of times and scales and less at some others.

To provide further insight into the above concept, one may consider the two signals ζ_1 and ζ_2 in Figure 1.4 as representing the parameter effects of two hypothetical

parameters θ_1 and θ_2. These two parameters have nearly collinear parameter effects with a correlation coefficient ρ of 0.9997. Yet if we consider the difference between their absolute normalized wavelet transforms,
$(|W\{\zeta_1\}|/\max|W\{\zeta_1\}|) - (|W\{\zeta_2\}|/\max|W\{\zeta_2\}|)$, also shown in Figure 1.4 by the Gauss and Mexican Hat WTs, one observes that they consist of both positive and negative values. This indicates that for each signal, there are regions of the time-scale plane wherein the absolute value of the signal's normalized wavelet coefficient exceeds the other's, albeit by a small margin. Therefore, some threshold η can be found to satisfy (1.15). It is also worth noting that because of the small difference margins between the normalized wavelet coefficients in the time-scale plane, the quality of the parameter signatures associated with θ_1 and θ_2 would be quite poor, hence, yielding unreliable parameter error estimates. One can extrapolate these results to multiple signals, with the expectation that the regions included in individual parameter signatures will become smaller with the overlap from the other signals' wavelet coefficients. However, given the independence of the parameter effects and adequate resolution in the time-scale plane (i.e., number of pixels), there will always be at least a pixel wherein the wavelet coefficient of each parameter effect will exceed all the others.

As a counterpoint to the highly correlated signals in Figure 1.4, one may consider the two uncorrelated signals ζ_3 and ζ_4 ($\rho = 0.08$) in Figure 1.5, associated with the hypothetical parameters θ_3 and θ_4. The $(|W\{\zeta_3\}|/\max|W\{\zeta_3\}|)-(|W\{\zeta_4\}|/\max|W\{\zeta_4\}|)$ by the Gauss and Mexican Hat WTs in Figure 1.5 not only are similar in trend to those in Figure 1.4 but are also more exaggerated in magnitude, ensuring much more reliable parameter signatures.

In order to extend these results to parameter signature extraction, the parameter signatures of the hypothetical parameters θ_1, θ_2, θ_3, and θ_4 were extracted with the Gauss WT and $\eta_t = 0.1$, as shown in Figure 1.6. The results clearly indicate the sparseness of the parameter signatures $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$, relative to $\hat{\Gamma}_3$ and $\hat{\Gamma}_4$, as the direct

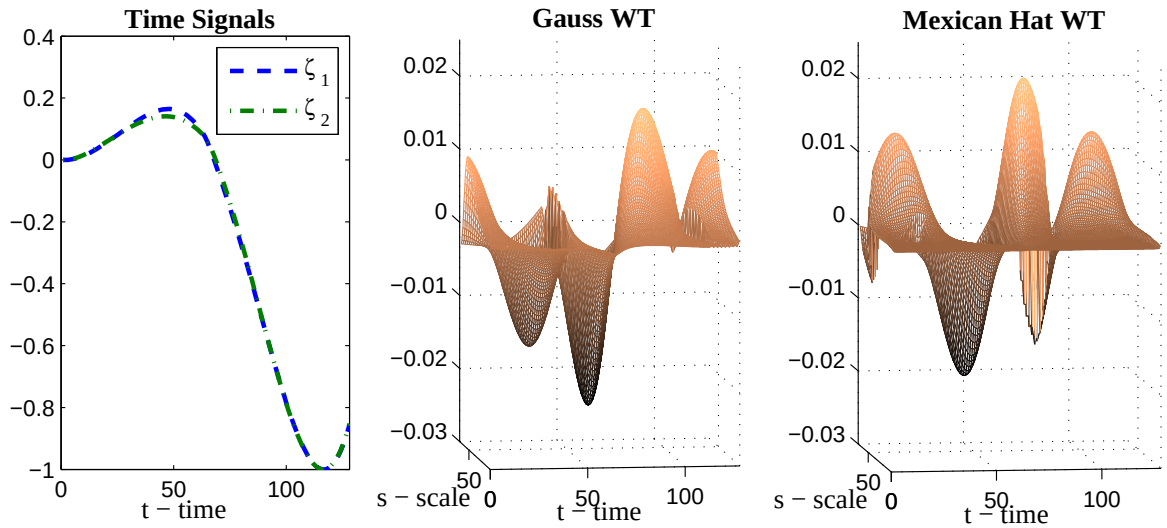

Figure 1.4: Two highly correlated signals and the difference between the absolute values of their Gauss WT.

reflection of near collinearity of their corresponding parameters effects. This reaffirms the conjecture that highly correlated parameter effects impede parameter signature extraction.

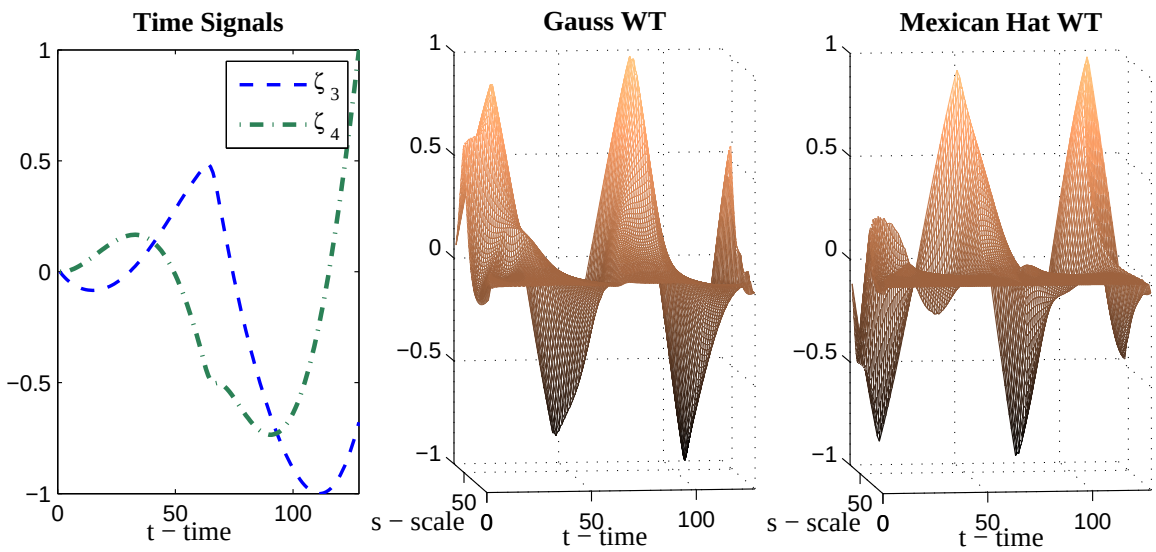

Figure 1.5: Two uncorrelated signals and the difference between the absolute values of their Gauss WT.

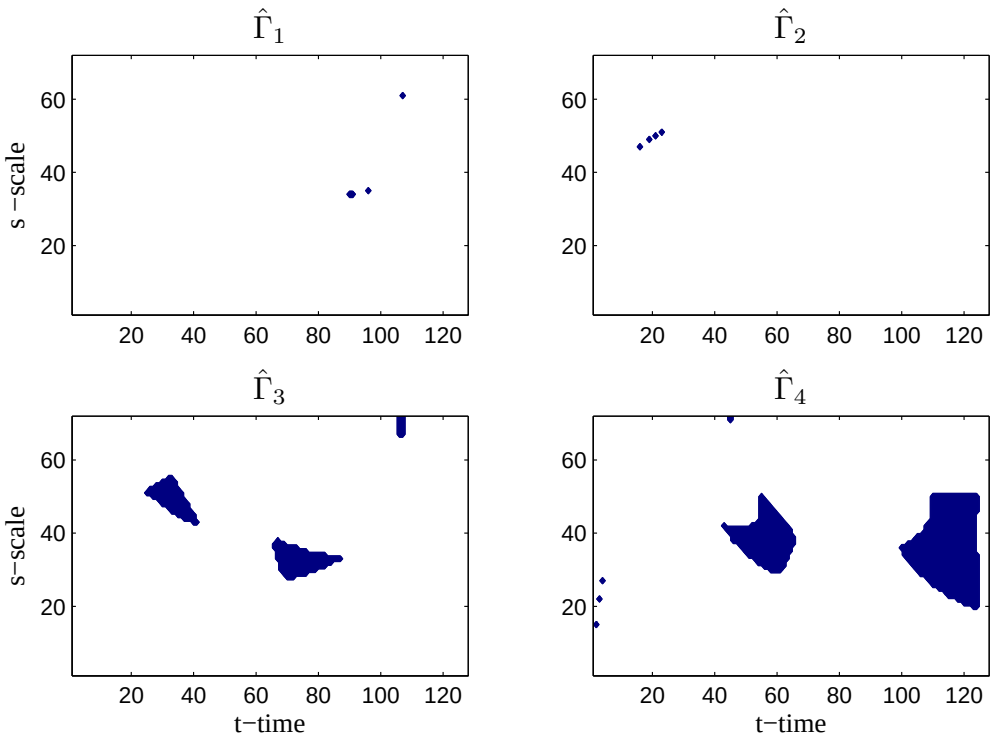

Figure 1.6: The parameter signatures of hypothetical parameters corresponding to the parameter effects ζ_1 and ζ_2 in Figure 1.4 and ζ_3 and ζ_4 in Figure 1.5, shown, respectively, in the top and bottom sub-figures. As expected, there are very few pixels included in the top parameter signatures due to the high correlation of ζ_1 and ζ_2. In contrast, the extracted parameter signatures associated with the uncorrelated signals ζ_3 and ζ_4 comprise many pixels.

CHAPTER 2

MODEL VALIDATION

Scientists are increasingly turning to sophisticated computer-based simulation models to predict and optimize process behavior in virtual environments. Yet to be effective, simulation models must have a high degree of accuracy to be reliable. An important part of model development, therefore, is model validation whereby the fidelity of the model is evaluated in representing the process.

A comprehensive survey of model validation metrics that quantify the closeness of model outputs with process observations is provided in [60]. The survey separates the metrics into the two categories of *verification* and *validation*. Verification addresses the fidelity of the simulation code in realizing the model's conceptual content. Validation is concerned with the model's ability to represent the process. This chapter is concerned with model validation and assumes that the model's conceptual content is adequately realized by the simulation code.

Given that process observations have the quality and breadth to provide a comprehensive view of the dynamics of the process [64, 65, 20], the fidelity of the model in representing the process can be evaluated by how closely the model outputs match the observations. One of the most common measures of closeness is the magnitude of the prediction error between the process observations and model outputs. Some other measures are the Akaike and Schwarz criteria [1, 37], the whiteness of the prediction error, the correlation of the error with the process inputs, and the certainty of parameter estimates [41, 19]. But for dynamic systems, as revealing as the magnitude, is the similarity of shape of the model outputs to the observations.

Although shapes are easily compared by human experts, due to the natural pattern recognition ability of humans, they are not as readily quantifiable due to the difficulty to define measures that can quantify the shape attributes of time series [11]. This research addresses this void by relying on the continuous wavelet transforms to represent the shape attributes of time series in both time and frequency (scale).

As discussed in Chapter 1, the wavelet coefficients obtained by the Gauss wavelet transform (WT) of a time series represent the first derivative of a time series smoothed by the Gaussian function, and the wavelet coefficients obtained by the Mexican hat WT denote the second derivative of this smoothed signal. Given that the first derivative of a signal represents its slope and the second derivative represents its rate of slope change, these wavelet coefficients quantify the slope and slope changes of the time signal and can, in turn, be used for similarity assessment of time series shapes.

The availability of the magnitude of the slopes and/or their rate changes at different locations of the time-scale plane provides not only the framework for comparing the shapes of time series at different frequencies but also the capacity to consider different measures of comparing the shapes. In this chapter we consider the slope of time series, as represented by its Gauss WT, and define three different measures to assess the closeness of the shape attributes. The effectiveness of these measures is first tested in a scenario where the true form of the model is known. These measures are then implemented in validation of two models of injection molding where the slopes of model outputs are compared with the slopes of measured pressures from the mold during injection molding cycles.

2.1 Time-Scale Domain Insight

The representation of slope and slope changes by the Gauss and Mexican hat WTs is illustrated in Figure 2.1. Shown in this figure are the surfaces in the time-scale domain of the Gaussian-smoothed measured pressure of Figure I.1 (top left), of the

Gauss WT of the measured pressure (top middle), and of the Mexican Hat WT of measured pressure (top right). Since it is hard to visualize the derivative relationship between the surfaces, also shown in this figure are slices of the surfaces at the first scale. The first slice (bottom left), denoted as $S_G\{y\}(t,1)$, is very close in shape to the measured pressure in Figure I.1, due to the minimal effect of smoothing in low scales (high frequencies). The second slice (bottom middle), denoted as $W_G\{y\}(t,1)$, represents the slope of the measured pressure by its first derivative, as characterized by the correspondence between the times associated with the spikes in this figure and the slope changes in the measured pressure. The last slice, $W_M\{y\}(t,1)$, represents the rate of slope changes of the measured pressure, which is another aspect of the shape of the signal that can be compared with its estimate by various models. The wavelet coefficients obtained by the continuous WTs thus represent the magnitudes of shape attributes of signals and can be used for comparison of their shapes.

2.2 Similarity Measures

In an effort to emulate the feature extraction approach of human experts, three different measures are considered here to represent the difference of surfaces obtained by wavelet transforms. The first measure is the Euclidean distance that provides an overall measure of magnitude difference between the wavelet coefficients. The second is a weighted Euclidean distance to discount the magnitude differences of wavelet coefficients that do not correspond to the same location in the time-scale plane. The third measure is the magnitude difference of wavelet coefficients at locations where the two surfaces markedly differ from each other.

The first measure is the Euclidean distance, d_E, which analogous to the sum of squared prediction error, represents the sum of squared difference between the wavelet coefficients of the process observations, $W\{y\}$ and the model output, $W\{\hat{y}\}$, as

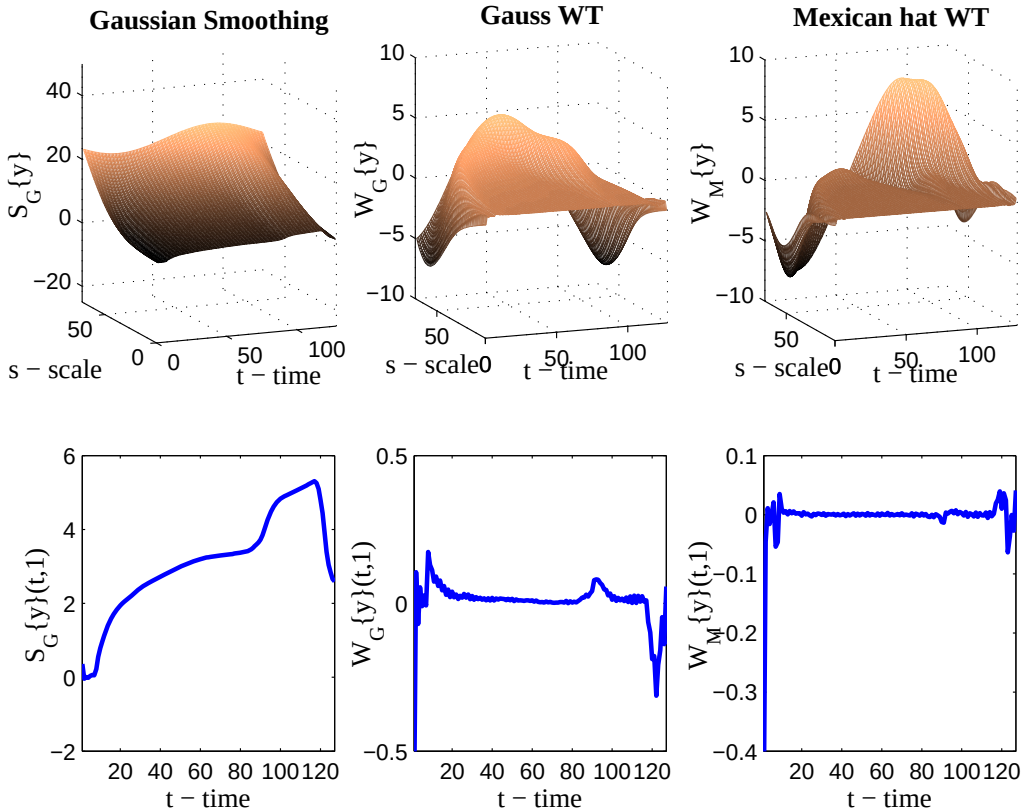

Figure 2.1: Surfaces in the time-scale domain of the Gaussian-smoothed measured pressure in Figure I.1 shown in the top left, and the Gauss and Mexican Hat WTs of the measured pressure in the top middle and top right plots, respectively. The bottom plots represent the time series associated with a slice of each surface at the first scale to represent the differential relationship between the surfaces.

$$d_E^2(y, \hat{y}) = \sum_{m=1}^{MN} \left(W\{y\}_m - W\{\hat{y}\}_m\right)^2 \tag{2.1}$$

where $W\{y\}_m$ or $W\{\hat{y}\}_m$ represents the magnitude of the wavelet coefficient at the pixel m. As is clear from the above equation, the Euclidean distance provides a cumulative (lumped sum) difference between the wavelet coefficients and, as such, it does not account for pixel by pixel dissimilarity of the surfaces representing the wavelet coefficients. A drawback to the lack of discrimination between positional differences, for instance, is the inability to detect shape differences due to time delays.

An alternative to the Euclidean distance is the weighted Euclidean distance (also known as image Euclidean distance [81]) that assigns more weight to pixel by pixel dissimilarities between surfaces. The weighted Euclidean distance, d_I, discounts the difference in magnitudes of wavelet coefficients according to the mutual distance between their locations on the time-scale plane, as [81]

$$d_I^2(y, \hat{y}) = \frac{1}{2\pi\sigma^2} \sum_{k,l=1}^{NM} \exp\{-|P_k - P_l|^2/2\sigma^2\}(W\{y\}_k - W\{\hat{y}\}_k)(W\{y\}_l - W\{\hat{y}\}_l) \quad (2.2)$$

where σ is a width parameter that represents the discount rate associated with the pixel distance, k and l denote the coordinates of each pixel on the time-scale plane, P_k and P_l denote pixel locations, and $|P_k - P_l|$ represents the distance between two pixels on the image lattice. According to (2.2), the weighted Euclidean distance fully incorporates the difference in magnitude of wavelet coefficients with identical locations and discounts by the weight "$\exp\{-|P_k - P_l|^2/2\sigma^2\}$" the magnitude difference when the two locations do not coincide on the time-scale plane (i.e., image lattice).

The characteristic difference between the Euclidean and weighted Euclidean distances is illustrated through the top view of the binary surfaces shown in Figure 2.2. The distances between Images 1, 2 and 3 are shown in Table 2.1. By visual inspection of these binary contours (referred to as images), it is clear that Images 1 and 2 are closer to each other than either is to Image 3. However, the Euclidean distance (d_E) between Images 1 and 2 is 112 and larger than the distance of 95 between Images 1 and 3, indicating more similarity between Images 1 and 3. In contrast, the weighted Euclidean distance (d_I) between Images 1 and 2 is 1.64 and smaller than the distance 4.17 between Images 1 and 3, providing a better agreement with the visual similarity of these images.

The third measure considered is in accord with the expert's approach of focusing on the pronounced shape differences between the process observations and their estimates. Accordingly, this measure is constructed to focus on the wavelet coefficients

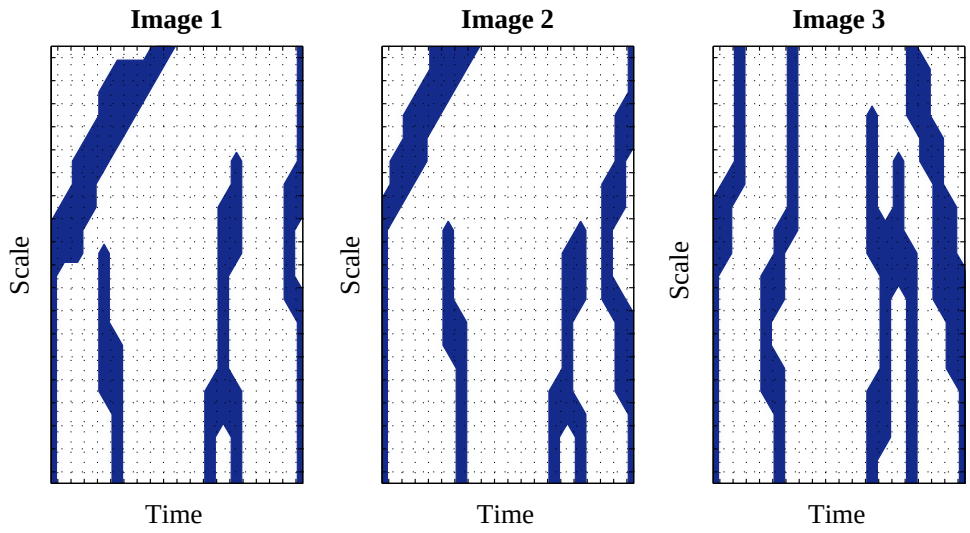

Figure 2.2: Top view of binary wavelet coefficients shown as images to illustrate the characteristic difference between the Euclidean distance and weighted Euclidean distance.

Table 2.1: Euclidean distance and weighted Euclidean distance between the images in Figure 2.2.

Image Distance	Images		
	1 & 2	1 & 3	2 & 3
d_E	112	95	109
d_I	1.64	4.17	3.97

of regions of marked deviation between the WTs of the two time series. The identification of these regions is particularly facilitated by the enhanced delineation of the wavelet transforms in the time-scale domain.

Similar to the parameter signature extraction approach discussed in Chapter 1, the identification of regions of marked deviation between the surface of process observations and model outputs can be conducted by isolating those pixels of the time-scale plane wherein the WT of one time series exceeds the other's by a dominance factor

of η_d. These pixels which are characterized by their coordinates (t_k^d, s_l^d) and marked by the superscript d can be identified as

$$\left|\overline{W\{y\}}(t_k, s_l)\right| > \eta_d \left|\overline{W\{\hat{y}\}}(t_k, s_l)\right| \quad \text{or} \quad \left|\overline{W\{\hat{y}\}}(t_k, s_l)\right| > \eta_d \left|\overline{W\{y\}}(t_k, s_l)\right| \Longrightarrow (t_k, s_l) \in (t_k^d, s_l^d) \tag{2.3}$$

where

$$\overline{W\{y\}} = \frac{W\{y\}}{\max_{(t,s)} |W\{y\}|}, \quad \overline{W\{\hat{y}\}} = \frac{W\{\hat{y}\}}{\max_{(t,s)} |W\{\hat{y}\}|}$$

For illustration purposes, the regions of marked difference between the Gauss WTs of the process observations and model outputs in Figure I.1 are shown in Figure 2.3. The time coordinates of the regions clearly coincide with those times in Figure I.1 where the slopes of the time series pairs drastically differ, hence, highlighting the locations of considerable difference between the shapes of the two time series.

Based on these identified regions of dominance, the third measure which will be hereafter referred to as the dominance distance, d_D, is defined as

$$d_D = \sum |W\{y\} - W\{\hat{y}\}|(t_k^d, s_l^d) \tag{2.4}$$

to denote the marked difference between the wavelet coefficients of the process observations and model outputs. For illustration purposes, the differential wavelet coefficients $|W\{y\} - W\{\hat{y}\}|$ associated with the regions of dominance in Figure 2.3 are shown in Figure 2.4.

2.3 A Controlled Validation Test

The validity of the proposed measures can only be evaluated in a controlled validation scenario where the structural form of the process producing the observations is known. Such a scenario is depicted in a simulation-based study described below.

Consider drug kinetics in human body as depicted in Figure 2.5, where the drug injected into the blood (compartment 1) exchanges linearly with the tissues (com-

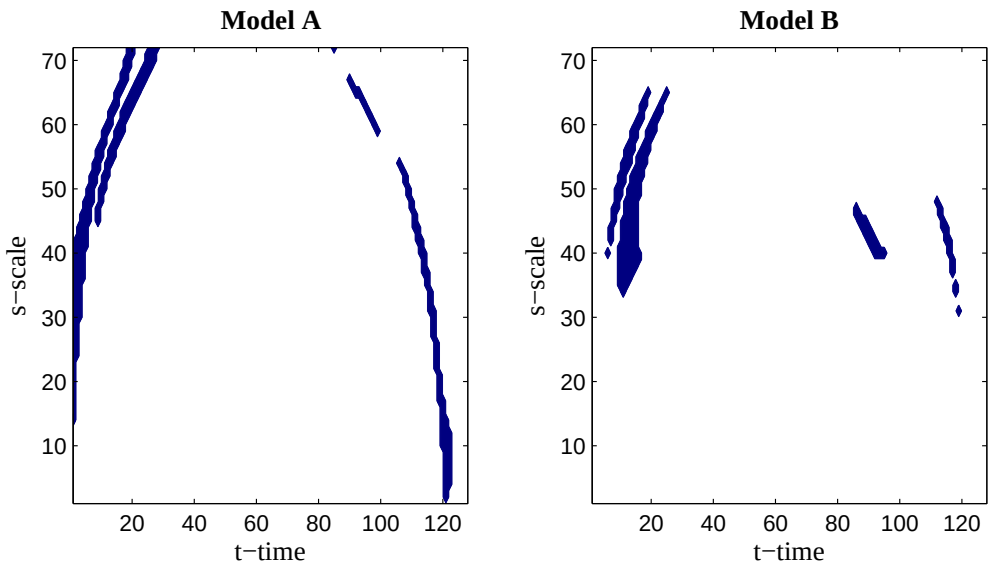

Figure 2.3: Regions of marked deviation between the Gauss wavelet coefficients of the measured pressure and its estimates in Figure I.1.

partment 2). Two different models can be considered to represent the process. The nonlinear model assumes that the drug is irreversibly removed with a nonlinear saturation characteristic (Michaelis-Menten dynamics) from compartment 1 and linearly from compartment 2. The linear model considers the transformation to be linear from both compartments. The drug is injected into compartment 1. The state variables x_1 and x_2 represent drug masses in the two compartments, $u(t)$ denotes the drug input, $y(t)$ is the measured drug, k_{12}, k_{21}, and k_{02} denote constant parameters, V_M and K_m are classical Michaelis-Menten parameters, and b_1 and c_1 are input and output parameters. The two models that can be considered to represent this system are:

$$\begin{aligned} \text{Linear Model}: \quad \dot{x}_1(t) &= -k_{21}x_1(t) - k_{12}x_2(t) + b_1u(t) \\ \dot{x}_2(t) &= k_{21}x_1(t) - (k_{02} + k_{12})x_2(t) \\ y(t) &= c_1x_1(t) \end{aligned} \tag{2.5}$$

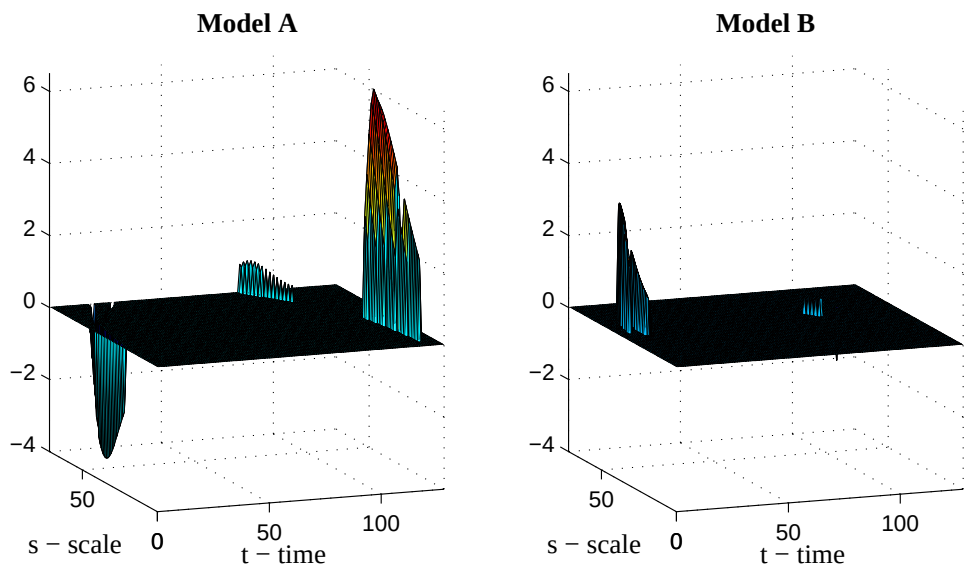

Figure 2.4: Differential wavelet coefficients, $|W\{y\} - W\{\hat{y}\}|$, of the measured pressure and its estimates at the marked regions in Figure 2.3.

$$\begin{aligned} \text{Nonlinear Model}: \quad \dot{x}_1(t) &= -\left(k_{21} + \frac{V_M}{K_m + x_1(t)}\right) x_1(t) + k_{12}x_2(t) + b_1 u(t) \\ \dot{x}_2(t) &= k_{21}x_1(t) - (k_{02} + k_{12})x_2(t) \\ y(t) &= c_1 x_1(t) \end{aligned} \tag{2.6}$$

For simulation purposes, the 'true' parameter values were obtained from Carson et al. [6] to represent galactose intake per kg body weight (kg B W) as

$$\mathbf{\Theta}^* = [k_{21}^*, k_{12}^*, V_M^*, K_m^*, k_{02}^*, c_1^*, b_1^*] = [3.033, 2.4, 378, 2.88, 1.5, 1.0, 1.0]$$

Given the designed format of this simulation-based study, one could presume to know the true parameter values and estimate the model output at the true parameter values. However, such a scenario would not be realistic since even in cases where the model closely represents the process-form the true parameters are generally not known and need to be estimated. In reality, not only the model mismatches the process-form, but also the input conditions having produced the observations are

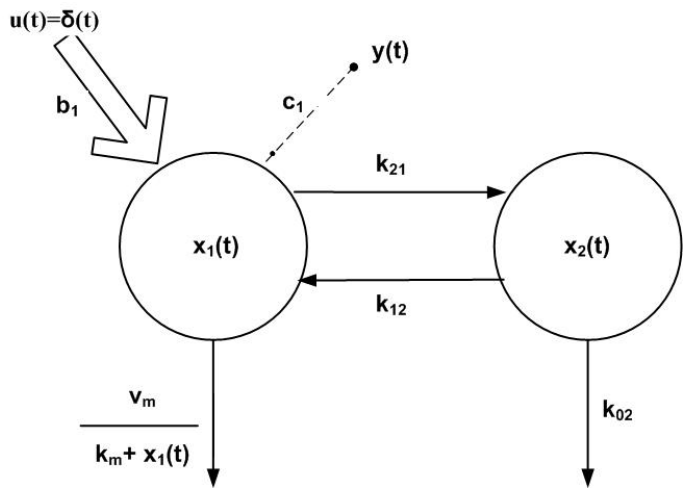

Figure 2.5: A nonlinear two compartment model of drug kinetics in human body. The circles in this figure depict the compartments.

not fully known. All these discrepancies, hence, preclude the estimation of the model parameters and force model validation to be performed with nominal parameters that deviate from the true parameter values. In this study, an up to 30% parameter error is incorporated by estimating the model outputs at nominal parameters that deviate from the true parameter values.

The study was performed according to the following format. Each one of the linear or nonlinear model in (2.5) and (2.6) was used, one at a time, to represent the process-form and to produce the observations as $y(t, \mathbf{\Theta}^*) + \nu$. These models were also used to produce $\hat{y}(t, \bar{\mathbf{\Theta}})$ as the model, except at some nominal parameter $\bar{\mathbf{\Theta}}$, where $\bar{\mathbf{\Theta}}$ was randomly selected within $\pm 30\%$ of $\mathbf{\Theta}^*$. Both the process observations $y(t)$ and the model output $\hat{y}(t)$ were generated to represent the system response to a single dose of drug $x_1(0) = 0.1$ mg / 100 ml kg B W, $x_2(0) = 0$, with ν incorporated through a sequence of normally distributed numbers. The proposed study, therefore, included two cases where the observations and model outputs were produced with the same model-form, albeit at different parameter values, and two other cases where the

process observations and model outputs were produced with different model-forms as well as different parameter values.

One hundred different model outputs, $\hat{y}(t)$, were obtained each at a random nominal parameter value in the nature of the Monte Carlo Simulation. First, the average magnitudes of the sum of absolute prediction error between each set of "process observations" and the one hundred model outputs were obtained to verify the validity of the study format. The average sums of absolute prediction error are shown in Table 2.2 for the various cases with different levels of signal-to-noise ratio (SNR). The signal-to-noise ratio was estimated empirically according to the relationship [4]

$$\text{SNR} = \gamma_s^2/\gamma_n^2$$

where γ_s^2 and γ_n^2 denote the mean squared values of the signal and noise, respectively. Also shown in this table are the ratios of the error sums in each column to indicate the degree of separation achieved by the prediction error for model/process mismatch. The results in Table 2.2 indicate that indeed the absolute prediction error sums are lower when the model matches the process-form (as indicated by smaller diagonal numbers shown in bold), and that these magnitudes are affected by the signal-to-noise ratio. But the prediction error is an established measure of model validation, so the fact that it indicates the match between the model and process-form is not a verdict on the effectiveness of this measure. It rather ascertains the reasonableness of the model validation scheme considered.

Having validated the reasonableness of the model validation test, this test can be used to evaluate the utility of the proposed measures of shape assessment. For this, the effectiveness of the three measures was evaluated in matching the slopes of the process observations to the model outputs by using their Gauss WTs. It should be noted here that we only suffice to a comparison of the slopes because the test is for the effectiveness of the similarity measures (distances). This test could be identically

Table 2.2: Confirmation of the effectiveness of the model validation test by the smaller average sums of the absolute prediction error corresponding to the match between the model and process form across different levels of signal-to-noise ratio (SNR).

Model Form	Process Form									
	Linear Model					Nonlinear Model				
	SNR					SNR				
	∞	17	11	7	5	∞	17	11	7	5
	$\sum_t \lvert\epsilon(t)\rvert$									
Linear Model	**0.48**	**0.66**	**1.03**	**1.45**	**1.88**	4.23	4.16	4.13	4.17	4.26
Nonlinear Model	3.84	3.91	4.01	4.14	4.30	**0.87**	**0.99**	**1.29**	**1.66**	**2.06**
Ratio (per column)	8.00	5.92	3.89	2.86	2.29	4.86	4.20	3.20	2.51	2.07

performed for the distances with the Mexican Hat WT to compare the rate of slope changes. The similarity measures are shown in Table 2.3 along with the ratios of the measures in each column, to again indicate the degree of separation provided by each measure for model mismatch. The results in this table indicate that (i) all three measures are effective in matching the model with the process-form through the slope of process observations and model outputs, as highlighted by the smaller diagonal distances (shown as bold) between the like model and process-form, and (ii) they also provide as good a resolution for the model mismatch as the sums of absolute prediction error in Table 2.2. As with the prediction error in Table 2.2, the ratios associated with the three measures are affected by the signal-to-noise ratio, even though they continue to indicate the mismatch between the model and process-form. The results also indicate that the dominance distance, d_D, seems to degrade more readily than the other distances at higher levels of signal-to-noise ratio. As was alluded to in Section 1.2, this is possibly due to the potentially higher influence of noise on the wavelet coefficients of observations in the lower scale regions of the time-scale plane.

Table 2.3: Evaluation of the effectiveness of the proposed distances in matching the model to the process form for the drug kinetics process across different levels of signal-to-noise ratio (SNR). The values in this table correspond to the mean distance for 100 different model outputs generated at random nominal parameter values within 30% of the true parameter.

Model Form	Process Form									
	Linear Model					Nonlinear Model				
	SNR					SNR				
	∞	17	11	7	5	∞	17	11	7	5
					d_E					
Linear Model	**0.02**	**0.02**	**0.03**	**0.04**	**0.05**	0.11	0.12	0.12	0.12	0.12
Nonlinear Model	0.11	0.11	0.12	0.12	0.12	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**
Ratio (per column)	5.5	5.5	4.0	3.0	2.4	5.5	6.0	4.0	3.0	2.4
					d_I					
Linear Model	**9.8e-4**	**15e-4**	**23e-4**	**32e-4**	**42e-4**	15e-3	15e-3	16e-3	16e-3	16e-3
Nonlinear Model	14e-3	14e-3	14e-3	15e-3	15e-3	**21e-4**	**24e-4**	**31e-4**	**39e-4**	**48e-4**
Ratio (per column)	14	9.3	6.1	4.7	3.6	7.3	6.3	5.2	4.1	3.3
					d_D					
Linear Model	**0.23**	**0.33**	**0.61**	**1.01**	**1.45**	2.32	3.53	3.28	3.11	3.58
Nonlinear Model	3.16	3.06	3.14	3.32	3.56	**0.22**	**0.72**	**1.29**	**1.94**	**2.65**
Ratio (per column)	13.7	9.3	5.2	3.3	2.5	10.6	4.9	2.5	1.6	1.4

2.4 A Practical Application

The proposed similarity measures were next evaluated in assessing the suitability of different models in representing an injection molding process. For this, the instrumented ASTM test mold shown in Figure 2.6 was considered with three cavities. For model validation, the actual pressures were measured at five locations inside the mold during the molding cycle. Here each portion of the mold can be thought of as an "element" with a unique pressure gradient modeled as a rod or strip with two end-nodes. By assembling the element conductance matrix and the element flow rate vector, a global conductance equation is formed. The mold is instrumented such that

the inlet pressure (P_1), runner pressure (P_2), and cavity entrance pressures (P_3, P_4, and P_5) are measured at the nodal locations shown in Figure 2.6.

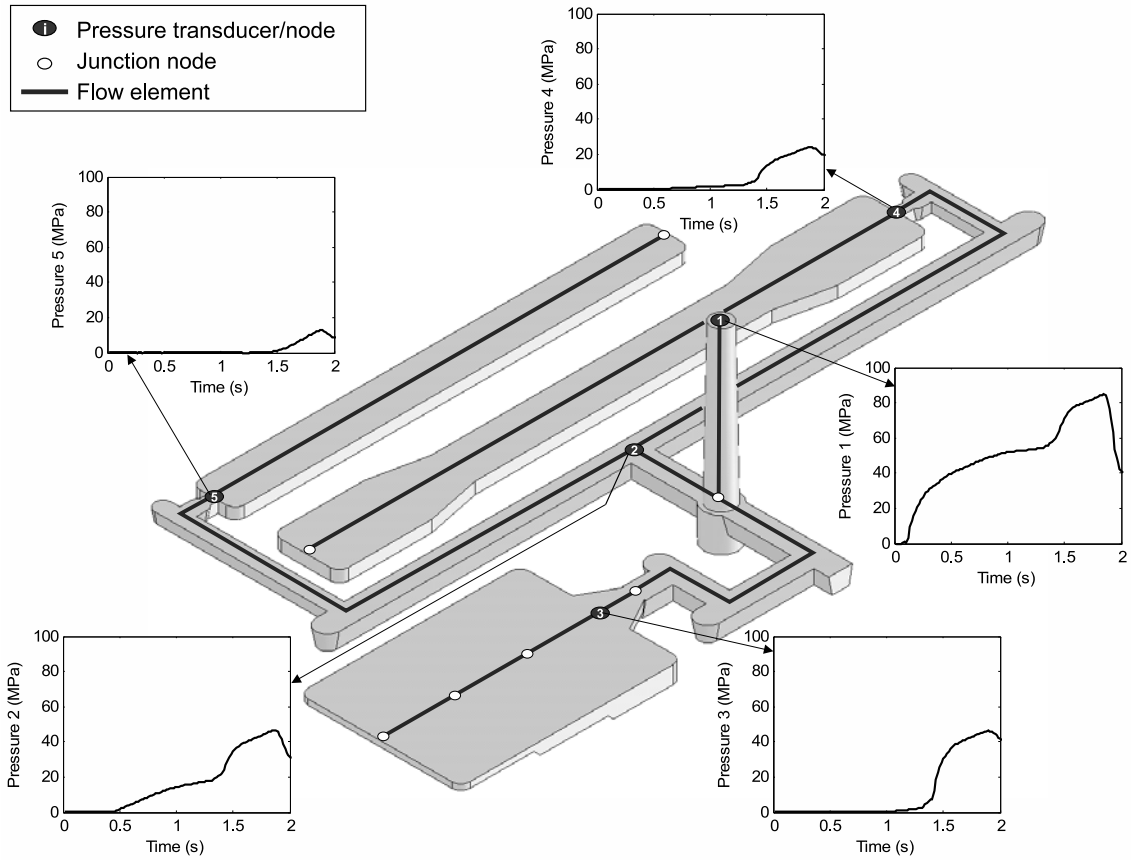

Figure 2.6: Instrumented ASTM test mold and a set of measured pressures from one of the experiments.

The mold geometry, melt rheology, and molding conditions are generally, but not precisely, known. The solution of the mass, momentum, and energy equations should yield a vector of pressure predictions that is consistent with the pressures observed by implanted transducers. However, variances in the model parameters and the inaccuracy of the model will lead to errors between the observed and simulated pressures throughout the molding cycle.

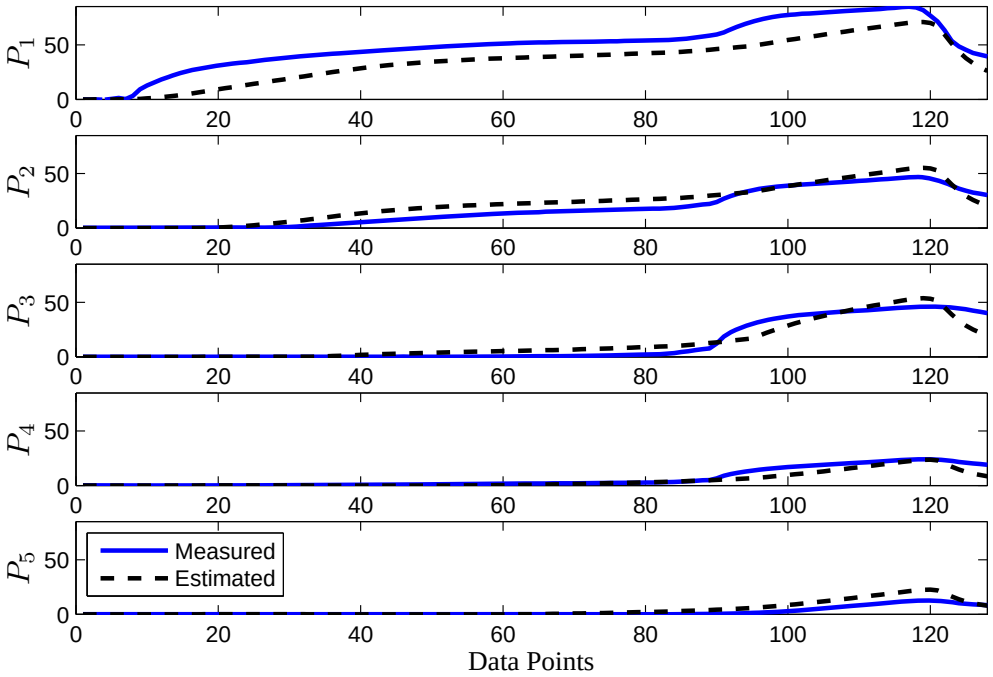

Figure 2.7: Pressure values obtained at five different locations of the mold shown together with their estimated counterparts by Model 3.

2.4.1 Experimental and Model Adaptation

Molding trials were conducted with polypropylene on a 50 metric ton Milacron Ferromatic molding machine. The ASTM test mold was instrumented with piezoelectric pressure transducers at the locations shown in Figure 2.6. A full factorial design of experiments was conducted to vary the melt temperature, coolant temperature, and injection velocity. The measured pressures are shown in Figure 2.7 along with their simulated counterparts by a model. Three different models were considered to represent the process. The first model (Model 1) is a *Newtonian* or *non-isothermal* model with the assumption of incompressibility. The second model (Model 2) incorporates a *power law viscosity model* instead, and the third model (Model 3) uses a *power law* with a first order delay to account for the melt compressibility of β/dt. The simulation results included in Figure 2.7 were obtained with the third model, and

these results would be different if instead the *Newtonian* or *non-isothermal* model was used with the assumption of incompressibility. As such, a major concern in model validation is to determine if the prediction error is mostly due to the error in the model parameters or it is due to qualitative deficiency of the model and assumptions used in simulating the outputs.

In an attempt to eliminate the parameter error, the rheological parameters of the models were adapted using the Gauss-Newton method. The other twenty six model parameters which were associated with the mold geometry were assumed to be accurate. The prediction error used for parameter estimation included all the five measured pressures and their estimates. Parameters were estimated for each of the eight input conditions pertaining to different temperatures, injection velocities, and other conditions. The sum of the absolute prediction errors at the five mold locations in Figure 2.6 normalized relative to the smallest prediction error for each molding trial before adaptation (BA) and after adaptation (BA) of the parameters are shown in Table 2.4 for different models and input conditions.

Table 2.4: Normalized sum of absolute errors at the five locations of the mold for each input condition before and after parameter estimation by the Gauss-Newton method.

Input Conditions	$\sum \lvert\epsilon(t)\rvert$					
	Model 1		Model 2		Model 3	
	Before	After	Before	After	Before	After
1	2.07	1.70	1.45	1.37	1.04	**1**
2	7.15	1.28	1.26	**1**	1.27	1.16
3	1.98	1.69	1.41	1.39	1.05	**1**
4	6.58	1.22	1.24	**1**	1.25	1.14
5	2.47	1.60	1.17	1.12	1.07	**1**
6	6.79	1.26	1.25	**1**	1.26	1.14
7	2.65	1.63	1.22	1.14	1.08	**1**
8	6.82	1.21	1.18	**1**	1.19	1.09

The results in Table 2.4 illustrate the conjugation between the quality of the model on one hand and effectiveness of parameter estimation on the other, which is the *Achilles heel* of simulation model development. In comparing the prediction error in each column before and after estimation, one can observe that although the errors are reduced by regression, they never approach zero. Moreover, some of the lower errors may be achieved at the expense of unrealistic (even negative) parameter estimates that ensure *model failure* at other processing conditions. Indeed, the statistics indicate that while the addition of model complexity and related parameters reduces the mean average error it actually increases the standard deviation of the error. For a model to be sufficiently robust for process or quality control, both a low mean and standard deviation of the error are required. The question then arises as when one would decide that the complexity of the model is sufficient and suitable for adaptation. For instance, the results in Table 2.4 indicate that although there is clear improvement in the simulation results due to the use of *Power Law* (Models 2 and 3) instead of Newtonian (Model 1), the improvements are not as pronounced when considering compressibility in place of incompressibility, especially after adaptation. According to the prediction error results, Model 2 (incompressible melt) provides better estimates for input conditions 2, 4, 6, and 8 (shown as bold), whereas Model 3 seems to be the better model for the other input conditions 1, 3, 5, and 7 (also shown as bold). Furthermore, the errors vary from run to run, indicating that the model's accuracy varies with the molding conditions. So, when does one stop adding to the model complexity and concentrate on adaptation? As is shown here, the continuous WT allows comparison of the shape of model outputs with the process observations' as a complement to the already commonplace magnitude comparison.

2.4.2 Validation Based on Both Magnitude and Shape

The closeness of model outputs to process observations was next evaluated for Models 2 and 3 according to the magnitude and slope difference of measured pressures and their estimates by the models. The measure of magnitude difference was the absolute prediction error sum. The shape difference was obtained by the three distances introduced in Section 2.2 applied to the Gauss WT of the measured pressures and model outputs. As in the previous validation case, in order to average out the effect of parameter error, the model outputs were obtained at one hundred different nominal parameter values within $\pm$ 25% of $\bar{\mathbf{\Theta}} = [200,000\ 0.25\ 0.1]$. The mean values of the absolute prediction error sum and the three distances of the Gauss wavelet coefficients between the measured pressures and model outputs were obtained for both Model 2 and Model 3. The normalized values of all these measures relative to the smallest corresponding measure at each input condition are listed in Table 2.5. The results indicate that according to the absolute prediction error sum, the weighted Euclidean distance, d_I, and the dominance distance, d_D, Model 3 provides a closer estimate of the pressures. The Euclidean distance, on the other hand, indicates Model 2 as the more suitable for two of the input conditions. As was illustrated through the plots in Figure 2.2, the Euclidean distance, d_E, is not as good a measure as the weighted Euclidean distance, d_I, in image comparison, therefore its verdicts should be discounted in contrast with the weighted Euclidean distance. Also noteworthy among these results is the consistency of the results provided by the dominance distance, d_D, which provide the best resolution in comparing the two models. Although Model 3 could be readily determined as the more suitable representation of the process-form according to the prediction error, this determination is now reinforced by the three distances between the slopes of model outputs and observations.

Table 2.5: Normalized average sum of the absolute prediction error and the three distances for the two models of injection molding at 100 random nominal parameter values.

Input Conditions	$\sum \lvert\epsilon(t)\rvert$		d_E		d_I		d_D	
	Model 2	Model 3	Model 2	Model 3	Model 2	Model 3	Model 2	Model 3
1	1.34	**1**	12.56	**1**	13.61	**1**	4.44	**1**
2	1.03	**1**	1.01	**1**	1.02	**1**	1.28	**1**
3	1.29	**1**	9.17	**1**	10.94	**1**	4.50	**1**
4	1.02	**1**	**1**	1.04	1	1	1.13	**1**
5	1.11	**1**	1.28	**1**	1.32	**1**	1.24	**1**
6	1.02	**1**	**1**	1.03	1.02	**1**	1.17	**1**
7	1.14	**1**	1.28	**1**	1.32	**1**	1.29	**1**
8	1.03	**1**	1.03	**1**	1.12	**1**	1.26	**1**

2.5 Discussion

The results presented indicate the utility of continuous wavelet transforms in representing the shape attributes of time series. This makes possible formulation of similarity measures to represent the closeness of the shapes of model outputs to the observations' in order to assess the suitability of the model in characterizing the process. Some of the other points to be considered are:

- *Other Similarity Measures.* The distances considered here are only three of the similarity measures that can be implemented in the time-scale domain. A fourth measure is the time warping distance [7] that can characterize time delays between the measured and model outputs. But there can also be image distances that would represent the other image aspects considered by human experts in visual inspection. This study only points to the potential of distances as model validation metrics and leaves the development of more customized distance measures to future studies.

- *Composite Validation Index.* A reliable validation metric for dynamic models would be ideally based on more than just the similarity of output pairs. Such a metric, would not only consider the similarity between the outputs of the model and observations, but also the time-based validation metrics already proven relevant such as the whiteness of the prediction error, its cross-correlation with inputs, and/or the certainty of parameter estimates. Toward this end, one could consider combining various image distances and time-based measures into a composite validation index to provide a comprehensive measure of model closeness across different operating conditions.

CHAPTER 3

PARAMETER ESTIMATION

The parameter error estimates obtained by (1.11) are not exact for the following reasons: (1) the estimates are local due to the dependence of $\mathcal{E}_i(t)$ on the nominal parameter vector $\bar{\mathbf{\Theta}}$; (2) the extracted parameter signatures $\hat{\Gamma}_i$ by (1.13) or (1.16) are only approximations of the ideal parameter signatures Γ_i; and (3) the first-order approximation of the model in (13) ignores the higher-order terms. As such, parameter estimation by PARSIM needs to be conducted iteratively.

The estimated parameter errors $\widehat{\Delta\theta_i}$ can be potentially used with any adaptation rule. Here we explore their utility in Newton-type methods to test their fidelity in parameter estimation. Following the general form of adaptation in Newton-type methods, parameter estimation by PARSIM takes the form

$$\hat{\theta}_i(q+1) = \hat{\theta}_i(q) + \mu\, \widehat{\Delta\theta}_i(q) \tag{3.1}$$

where q is the adaptation step, $\widehat{\Delta\theta}_i$ is estimated according to (1.11) and μ is the size of adaptation per iteration. The remainder of this section is devoted to evaluation of PARSIM's estimation performance according to (3.1) for a variety of different cases. Specifically, PARSIM is evaluated first in a noise-free condition to test the fidelity of parameter error estimates in iterative parameter estimation. Next, PARSIM is applied to two challenging nonlinear models to test its breadth of applicability in single output cases. Finally, a hybrid approach to parameter estimation is demonstrated that concurrently analyzes solutions from both the time and time-scale domains. This approach is demonstrated by tuning a complex jet engine simulation model.

Throughout this study PARSIM's performance is compared with that of the Gauss-Newton method to provide a basis for evaluating its performance vis-*à*-vis regression. The Gauss-Newton method [70] used here has the same form as (3.1) except that $\widehat{\Delta\boldsymbol{\Theta}}$ is obtained according to:

$$\widehat{\Delta\boldsymbol{\Theta}} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\epsilon^N \tag{3.2}$$

As a first test, PARSIM was applied to the estimation of the nonlinear mass-spring-damper model parameters in (14) using the model's impulse response as output. One hundred estimation runs were performed with random initial parameter values within 25% of $\boldsymbol{\Theta}^*$. A step size of $\mu = 0.50$ was used for both PARSIM and the Gauss-Newton method with a threshold level of $\eta = 0.10$ in (1.13) for extracting the parameter signatures in PARSIM. The mean values of the parameter estimates from the 100 estimation runs of PARSIM and the Gauss-Newton method after 50 iterations are listed in Table 3.1. Along with the parameter estimates are the mean values of the precision error ϵ_Θ obtained as

$$\epsilon_\Theta^2 = \sum_{i=1}^{3}\left((\theta_i^* - \hat{\theta}_i)/\theta_i^*\right)^2 \tag{3.3}$$

which although not available in practical applications, because of unknowable true parameters, is a valuable measure for its representation of the accuracy of estimates. The results in Table 3.1 indicate that PARSIM provides less precise estimates than the Gauss-Newton method using the Gauss WT and better estimates with the Mexican Hat WT. Although anecdotal at this point, it is worth noting that these results are consistent with the level of delineation the above wavelet transforms provide for the parameter effects of this model in the time-scale domain, as indicated by the singular values in Table 1.1. As a measure of the convergence effectiveness of the two methods, the sums of absolute prediction error during the estimation runs of PARSIM using

the Mexican Hat WT are compared with those from the Gauss-Newton method in Figure 3.1. The results clearly indicate that PARSIM with the Mexican Hat WT provides a faster convergence for this model than the Gauss-Newton Method.

Table 3.1: Fiftieth-iteration mean of one hundred estimation runs of the nonlinear mass-spring-damper model parameters by PARSIM and the Gauss-Newton method. Random initial parameter values within 25% of the true values were used for each estimation run.

True Parameters	**Parameter Estimates**					
	PARSIM				**Gauss-Newton**	
	Gauss WT		**Mexican Hat WT**			
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
$m = 375$	374.98	0.0824	375	$7.6925e^{-12}$	375	$1.3578e^{-5}$
$c = 9800$	9800.90	4.0381	9800	$6.2681e^{-10}$	9800	$7.9358e^{-4}$
$k = 130 \times 10^3$	129988	51.0358	130×10^3	$1.0591e^{-8}$	130×10^3	0.0069
Precision Error, ϵ_θ	$5.1492e^{-4}$	$3.5312e^{-4}$	$9.8972e^{-14}$	$3.9090e^{-14}$	$9.3542e^{-8}$	$4.9567e^{-8}$

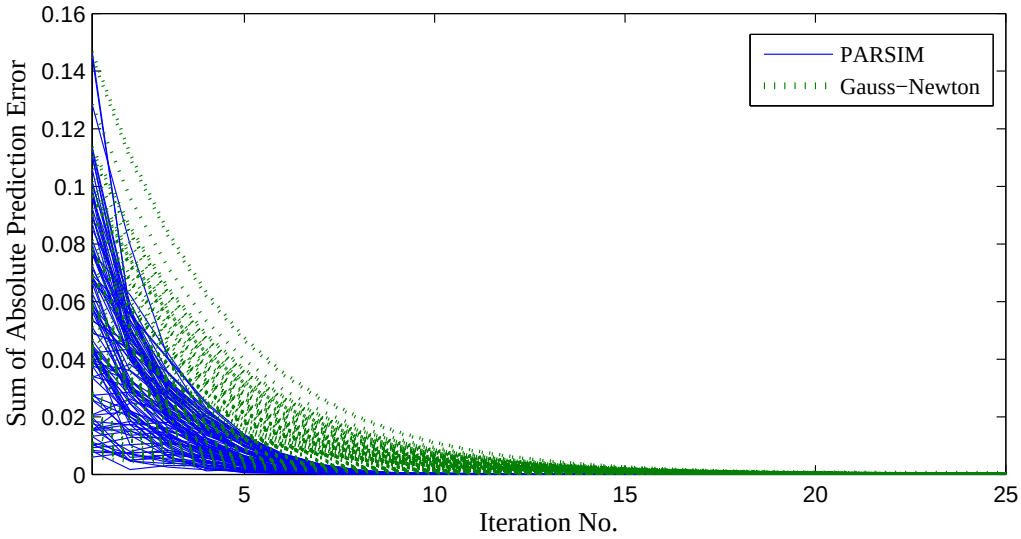

Figure 3.1: Prediction error during one hundred estimation runs of the nonlinear mass-spring-damper model parameters by PARSIM and the Gauss-Newton method.

3.1 Input Conditions

Another point of interest in parameter estimation is the effect of input conditions. To test the performance of PARSIM with a different input condition, parameter estimates were obtained using the free response, due to an initial displacement, of the nonlinear mass-spring-damper model in (14). For this, $x(t)$ was simulated in response to an initial displacement of $x(0) = 0.20$ cm. The mean and standard deviation of the adapted parameters from one hundred estimation runs of PARSIM with the Mexican Hat WT and $\eta = 0.1$ together with those from the Gauss-Newton method are shown in Table 3.2. As before, random initial parameter values within 25% of the actual parameter values in $\mathbf{\Theta}^*$ were used for each estimation run. The results indicate that the estimated parameters by PARSIM are considerably more accurate and consistent than those by the Gauss-Newton method. Although anecdotal, the results point to a potentially lesser sensitivity of PARSIM to the input conditions, and at the very least, motivate a study of PARSIM's requirements for the input conditions.

Table 3.2: Twentieth-iteration mean and standard deviation values of one hundred estimation runs of the nonlinear mass-spring-damper model parameters obtained from the free response of the system to an initial displacement. As before, the initial parameter values were randomly selected within 25% of their true values.

True Parameter Values	**Parameter Estimates Nonlinear mass-spring-damper**			
	PARSIM		**Gauss-Newton**	
	Mean	St. Dev.	Mean	St. Dev.
$m = 375$	374	13	437	80
$c = 9800$	9777	352	11419	2084
$k = 130 \times 10^3$	129697	4668	151479	27642
Precision Error, ϵ_θ	0.0491	0.0331	0.2973	0.3594

3.2 Nonlinear Ill-Conditioned Models

To test its versatility, PARSIM was also applied to two ill-conditioned models. The first model is the nonlinear two-compartment model of drug kinetics in (2.6) which has with near nonidentifiable parameters. The second case is the Van der Pol oscillator, which exhibits bifurcation characteristics [80] that challenge its first-order approximation by (13).

For the first case, the collinearity of the parameter effects of k_{21}, k_{12}, and k_{02} was evaluated by their correlation coefficients as

$$\rho_{k_{21}k_{12}} = 0.9946 \qquad \rho_{k_{21}k_{02}} = -0.9985 \qquad \rho_{k_{12}k_{02}} = -0.9888$$

All the three coefficients are near unity, which indicate the difficulty to extract reliable parameter signatures for them. To verify this point, the parameter signatures of the three parameters were extracted using the Gauss WT. Based on these parameter signatures, the parameter errors were estimated according to (1.11) as $\widehat{\Delta\Theta} = [\widehat{\Delta k_{21}}, \widehat{\Delta k_{12}}, \widehat{\Delta k_{02}}] = [0.1942, 0, 0]$ which are null for k_{12} and k_{02} due to inability to extract parameter signatures for these two parameters at the current nominal parameters.

Next, parameter estimation was tried. For estimation purposes, the output $\hat{y}(t, \hat{\mathbf{\Theta}})$ of the drug kinetics model in (2.6) was simulated. Both PARSIM and the Gauss-Newton method were used to estimate the parameters k_{21}, k_{12}, and k_{02}, which deviated from their true values. The adapted parameters, shown in Figure 3.2, indicate that the near nonidentifiability of the parameters of this model impedes estimation by either method. However, the results reveal another inherent characteristic of the two methods. In PARSIM's case, the near nonidentifiability of the parameters precludes parameter signature extraction for two of the parameters, so these parameters remain unchanged from their initial values. With the Gauss-Newton method, on the

other hand, all three parameters are adapted to minimize the error, but due to near nonidentifiability, the parameter estimates diverge from their true values.

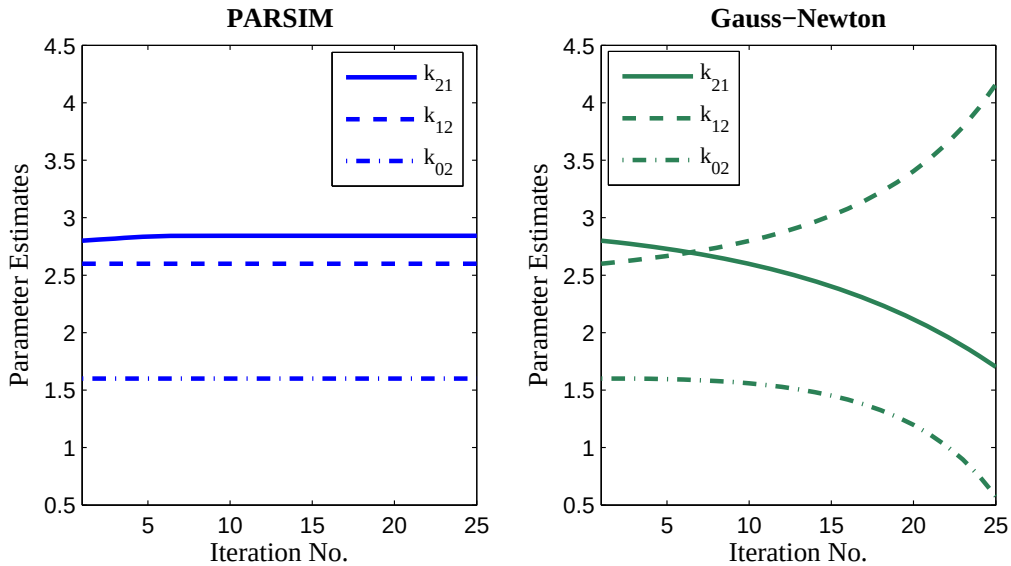

Figure 3.2: Estimated parameters of the drug kinetics model in (2.6) by PARSIM (left) and the Gauss-Newton method (right).

In the second estimation case, the Van der Pol oscillator, had the form

$$m\ddot{x} - c(1 - x^2)\dot{x} + kx = 0 \tag{3.4}$$

with its true parameters defined as $\mathbf{\Theta}^* = [m^*, c^*, k^*] = [375, 9800, 130 \times 10^3]^T$. The Van der Pol oscillator was simulated with the initial condition $x(0) = 0.02$ and $\dot{x}(0) = 0$, and as before one hundred estimation runs were performed with different initial parameter values within 10% of $\mathbf{\Theta}^*$. Both PARSIM using the Gauss WT and the Gauss-Newton method were applied to this model with a step size of $\mu = 0.50$. The threshold value for PARSIM was $\eta_t = 0.20$. The mean value of the adapted parameters and their standard deviations at the twenty-fifth iteration of the two methods are listed in Table 3.3. As observed from the results, the two methods are similar in that they do not consistently converge to the true parameters despite

minimizing the prediction error. PARSIM, however, provides a more accurate overall estimate of this model's parameters, in part due to its more frequent convergence to the true parameter values.

Table 3.3: Twenty fifth-iteration mean and standard deviation values of the adapted Van der Pol oscillator parameters from one hundred estimation runs of PARSIM and the Gauss-Newton method. Random initial parameter values were used for each estimation run within 10% of the true values.

True Parameter Value	**Parameter Estimates**			
	PARSIM		**Gauss-Newton**	
	Mean	St. Dev.	Mean	St. Dev.
$m = 375$	380	16.17	385	17.87
$c = 9800$	9921	422.32	10062	467.11
$k = 130 \times 10^3$	131.6 $\times 10^3$	5.605 $\times 10^3$	133.5 $\times 10^3$	6.196 $\times 10^3$
Precision Error, ϵ_θ	0.0638	0.0439	0.0785	0.0526

3.3 Variable Threshold and Adaptation Size

The performance of PARSIM in parameter estimation as previously described depends on thresholds and adaptation steps that were chosen by trial and error and fixed for all adaptation iterations. The transparency afforded by the parameters signatures, however, does provide measures for autonomous selection of the threshold η_t in (1.13) and the adaptation step μ in (3.1). The criteria and strategies devised for these measures are discussed below. To illustrate this concept, the Lorenz model is utilized which has the form [63]:

$$\begin{aligned} \dot{x}_1 &= \sigma(x_2 - x_1) \\ \dot{x}_2 &= p_1 x_1 - p_2 x_2 - x_1 x_3 + p_3 \\ \dot{x}_3 &= x_1 x_2 - b x_3 \end{aligned} \tag{3.5}$$

with the true parameter values

$$\theta^* = \begin{bmatrix} \sigma^* & p_1^* & p_2^* & p_3^* & b^* \end{bmatrix} = \begin{bmatrix} 10 & 28 & 1 & 8/3 & 8/3 \end{bmatrix}$$

For this illustration, and throughout the dissertation, for this model, $y(t) = x_1(t, \bar{\theta})$ where

$$\bar{\theta} = \begin{bmatrix} \bar{\sigma} & \bar{p_1} & \bar{p_2} & \bar{p_3} & \bar{b} \end{bmatrix} = \begin{bmatrix} 12 & 25 & 1.5 & 2 & 2 \end{bmatrix}$$

The prediction error due to the mismatch between the nominal and true parameter values with $\nu = 0$ is shown in Figure 3.3 along with the approximation of the error according to (8). It is clear that the first-order approximation deviates from the true error after about 0.2 s of simulation. This is one of the challenging aspects of the Lorenz model which causes inaccuracy in the gradient estimates by NLS and the parameter error estimates by PARSIM. Regardless, the Lorenz model provides a good platform to illustrate the concepts of *Threshold Selection* and *Adaptation Step Size Selection*.

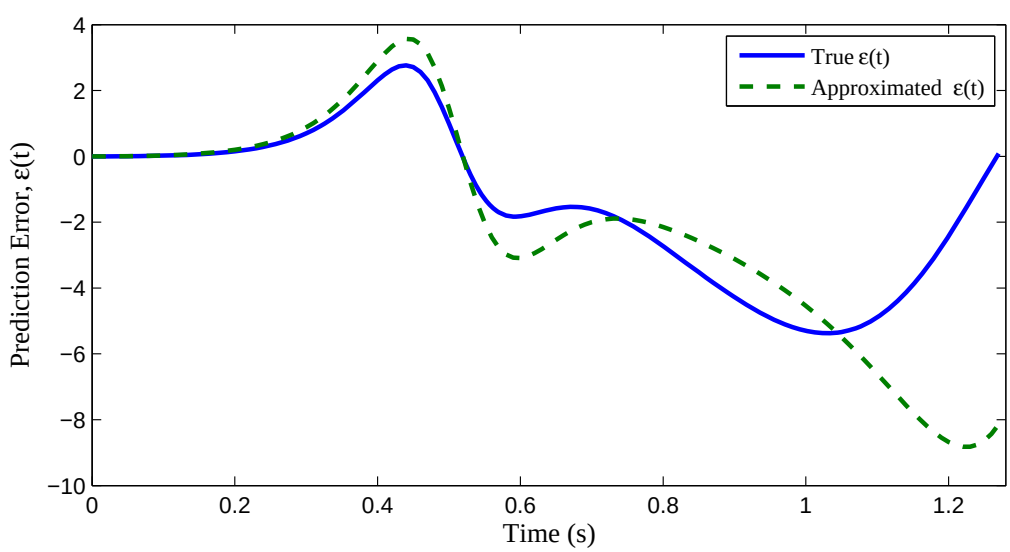

Figure 3.3: Approximation of the prediction error in x_1 of the Lorenz system by (8). The solid line represents the prediction error and the dashed line is its approximation.

3.3.1 Threshold Selection

As noted above, PARSIM relies on the threshold η_t to extract the parameter signatures according to (1.13). As such, the threshold level can have a significant effect on the quality of the extracted parameter signatures as well as their locations. This is illustrated in Figure 3.4 for the parameter signature of p_3 of the Lorenz model extracted with two different threshold levels. It is, therefore, important to devise a strategy whereby a suitable threshold value is selected for extracting quality signatures for each $\bar{\mathbf{\Theta}}$.

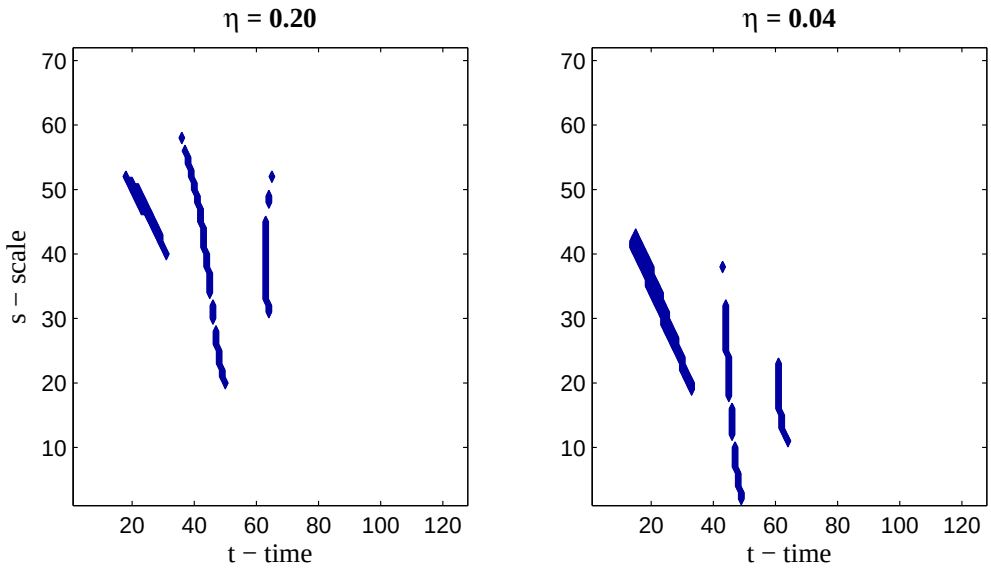

Figure 3.4: The signatures of p_3 in the Lorenz model obtained with the two threshold levels of $\eta = 0.20$ (left) and $\eta = 0.04$ (right) using a Mexican Hat wavelet transform.

however, assessing the quality of the parameter signatures is not a straightforward task. Explicit to the definition of the parameter signature Γ_i is that the $W\{\mathcal{E}_i\}$ be much larger than all the other $W\{\mathcal{E}_j\}$ $\forall\ j \neq i$. But the strategy used in (1.13) only ensures (1.15) which does not necessarily satisfy the condition of dominance explicit to the definition of parameter signatures. Given that the notion of dominance is associated with the magnitude of $W\{\mathcal{E}_i\}$, one can potentially consider as a criterion the closeness of the mean of $W\{\mathcal{E}_i\}$ at the pixels (t_k^i, s_l^i) to the $\max_{(t,s)} |W\{\mathcal{E}_i\}|$.

Another possible criterion is the number of pixels included in the parameter signature. However, up until this point, none of the discussed criteria adequately assess the quality of parameter signatures. The measure of quality that corresponds the best to parameter estimation performance is the consistency of the parameter error estimates obtained from the individual pixels of the parameter signature, quantified by the variance of the parameter error estimates, as

$$\sigma^2_{\hat{\theta}_i} = \frac{1}{N_i - 1} \sum_{k=1}^{N} \sum_{l=1}^{M} \left(\frac{W\{\epsilon\}(t^i_k, s^i_l)}{W\{\mathcal{E}_i\}(t^i_k, s^i_l)} - \widehat{\Delta\theta}_i \right)^2 \quad \forall\, (t^i_k, s^i_l) \in \Gamma_i \tag{3.6}$$

The reasoning for using the parameter error variance as the measure of parameter signature quality is that ideally every pixel included in the parameter signature ought to provide the same parameter error estimate. Accordingly, large discrepancies between these estimates would indicate a deficiency in the parameter signature extraction process, which may be corrected by the better selection of the threshold level η_t in (1.13).

As an illustration of the effectiveness of the above criterion, the parameter error estimates of the parameter b in the Lorenz model are shown at each pixel of the corresponding parameter signature in Figure 3.5 obtained with two different threshold levels. Also shown in the figure, is the variance of the estimates for each parameter signature. The larger variance in the left plot clearly indicates the notably larger scatter among the parameter error estimates relative to those on the right. Ordinarily, if the notion of the parameter signature is satisfied, then all the parameter error estimates should be equal ($\sigma^2_{\hat{\theta}_i} = 0$) and there should be no need for averaging them as is performed in (1.11). In this light, the more scattered the parameter error estimates are (i.e., the higher their variance), the less confidence can be ascribed to the quality of the extracted parameter signature.

A factor that can potentially improve the quality of the extracted parameter signatures is the threshold level η_t in (1.13). A threshold level, however, affects all the parameter signatures, and each parameter signature corresponds to a parameter

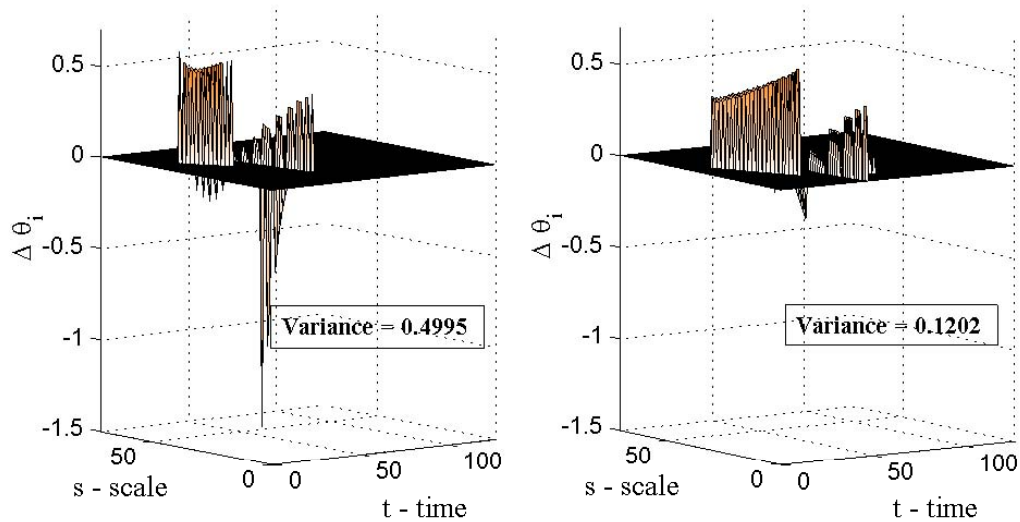

Figure 3.5: The $\widehat{\Delta\theta_i}$ of p_3 at each pixel of its parameter signatures extracted with the two threshold levels of $\eta = 0.20$ (left) and $\eta = 0.04$ (right).

error variance. Here we focus on the largest variance, which is associated with the lowest quality, as the weakest link. Therefore, the search for the threshold level is performed so as to minimize the largest variance among all the parameter error estimates, as

$$\eta_t^*(q) = \arg\min_{\eta_t} \max_i \sigma^2_{\hat{\theta}_i}(q, \eta_t), \quad \eta_{min} \leq \eta_t \leq \eta_{max} \tag{3.7}$$

where η_t^* is the selected threshold for the iteration number q within the range $[\eta_{min}, \eta_{max}]$. A reasonable range is $\eta_{min} = 0.02$ and $\eta_{max} = 0.20$. According to this strategy, the threshold level can be determined for each adaptation step separately, with separate threshold levels considered for each output in multi-output adaptation.

This approach to threshold selection would not be viable if it were to be applied to selecting an optimal dominance factor. This is due to the fact that by increasing the dominance factor, the number of pixels comprising a parameter signature would decrease. As a result, a variance-based method would simply select the largest possible dominance factor that would produce a single pixel and effectively reduce the variance

to zero. Considering this, it is noted that to this point, there is no effective approach to selecting the optimal dominance factor for parameter adaptation.

3.3.2 Adaptation Step Size Selection

The magnitude of the adaptation step size $\mu \in (0, 1]$ in (3.1) represents the confidence given to the parameter error estimate $\widehat{\Delta\theta}_i(q)$ in leading the parameter estimate, $\hat{\theta}_i(q)$ to its correct value, θ_i^*. Lower values for μ tend to be more stable, but they prolong the estimation. In time-based estimation, like NLS, the magnitude of μ is selected according to the convexity of the problem and is generally constant at every iteration. In PARSIM, on the other hand, in addition to convexity, the quality of the parameter signature can be a factor in selection of μ, and since the quality of parameter signatures depends on $\bar{\mathbf{\Theta}}$ which is different at each iteration, a different μ can be selected for each adaptation iteration. Using a different adaptation step size per iteration would then lead to the adaptation rule:

$$\hat{\theta}_i(q+1) = \hat{\theta}_i(q) + \mu_i(q)\, \widehat{\Delta\theta}_i(q) \tag{3.8}$$

The selection of threshold level at each iteration was discussed above as a way of improving the quality of parameter signature. Another factor that also affects this quality is the uniqueness of the parameter effects. As described previously, the ability to extract parameter signatures is contingent upon the level of correlation between the parameter effects, computed as [4]

$$|\rho_{ij}| = \frac{|C_{ij}|}{\sigma_i \sigma_j} = \frac{\sum_{k=1}^{N} (\mathcal{E}_i(t_k) - \bar{\mathcal{E}}_i)(\mathcal{E}_j(t_k) - \bar{\mathcal{E}}_j)}{(\sum_{k=1}^{N} \mathcal{E}_i(t_k)^2 - N\bar{\mathcal{E}}_i^2)(\sum_{k=1}^{N} \mathcal{E}_j(t_k)^2 - N\bar{\mathcal{E}}_j^2)} \tag{3.9}$$

where $|\rho_{ij}|$ is the absolute value of the correlation coefficient between pairs of parameter effects, k represents the sample point and $\bar{\mathcal{E}}$ is the mean value of the parameter effect. According to the conjecture in Chapter 1, it will be impossible to extract pa-

rameter signatures when $|\rho_{ij}| = 1$ and it will be difficult to extract quality parameter signatures when $|\rho_{ij}|$ is close to 1.

Using the above observation, another factor in the quality of the parameter signature is the level of correlation between a parameter effect and all the other parameter effects. This correlation value can then be factored into the magnitude of μ as

$$\mu_i(q) = 1 - \max |\rho_{ij}(q)| \qquad \forall j \neq i \tag{3.10}$$

3.3.3 Validation

To evaluate the advantages of the above selection strategies, parameter estimation results were obtained with and without selective thresholding and variable adaptation sizes for the Lorenz model. The prediction and precision errors for each case are shown in the left and right plots of Figure 3.6, respectively. The results in Figure 3.6 indicate that both selection strategies enhance the performance of PARSIM and that together they improve the convergence of PARSIM considerably.

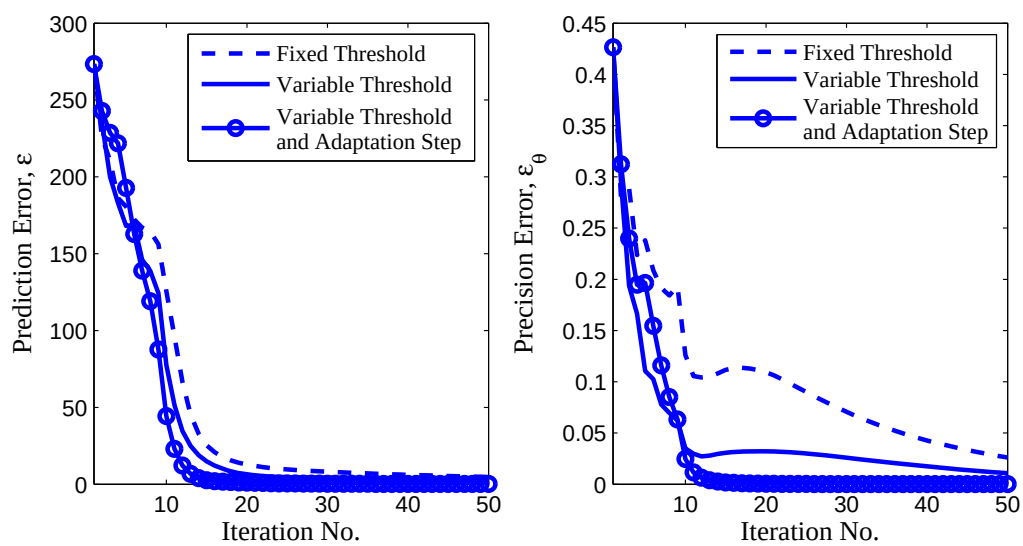

Figure 3.6: Prediction error for the adaptation results using both fixed and variable thresholding as well as variable thresholding and output scaling for Lorenz.

3.4 Integrating Solutions from the Time and Time-Scale Domain

It is understood that an important attribute of PARSIM is its separate estimation of parameters. Although this attribute is responsible for its effective convergence characteristics, it could also lead to dormancy of estimation when parameter signatures cannot be extracted for single parameters. The inability to extract parameter signatures may be due to the overcrowding of the time-scale plane by the multitude of output sensitivities competing for dominance or lack of parameter identifiability by the output. This concept is illustrated in this section using Chua's oscillator.

An important attribute of PARSIM is the potential to avoid local minima due to its independence from the gradient of the prediction error contour. This motivates the idea of a competitive strategy wherein concurrent solutions are considered from PARSIM and nonlinear least squares (NLS), to remedy NLS's susceptibility to local minima entrapment while benefiting from its relative immunity to dormancy of parameter estimation. In the strategy shown here, the two methods are compared frequently (e.g., every p iterations of parameter adaptation) and the better solution is selected according to its effectiveness in reducing the prediction error. Tests that confirm the effectiveness of this competitive strategy in eluding local minima, in parameter estimation using the Van der Pol oscillator, are also presented.

The dormancy of the parameter estimation in the absence of parameter signatures is best illustrated with Chua's circuit which is introduced in the next example.

Chua's oscillator is described by a set of three ordinary differential equations called Chua's equations [34]:

$$\begin{aligned}
\frac{dI_3}{dt} &= -\frac{R_0}{L}I_3 - \frac{1}{L}V_2 \\
\frac{dV_2}{dt} &= \frac{1}{C_2}I_3 - \frac{G}{C_2}(V_2 - V_1)
\end{aligned}$$

$$\frac{dV_1}{dt} = \frac{G}{C_1}(V_2 - V_1) - \frac{1}{C_1}f(V_1) \tag{3.11}$$

where

$$f(V1) = G_b V_1 - (G_a - G_b)(|V_1 + E| - |V_1 - E|)$$

and

$$\begin{aligned} \mathbf{\Theta}^* &= \begin{bmatrix} L^* & R_0^* & C_2^* & G^* & G_a^* & G_b^* & C_1^* & E^* \end{bmatrix} \\ &= \begin{bmatrix} -9.7136 & 4.75 & -1.0837 & 33.932813 & -0.5 & .0064 & 1 & 1 \end{bmatrix} \end{aligned}$$

For this illustration and throughout the thesis for this model,

$$\begin{aligned} \bar{\mathbf{\Theta}} &= \begin{bmatrix} \bar{L} & \bar{R}_0 & \bar{C}_2 & \bar{G} & \bar{G}_a & \bar{G}_b & \bar{C}_1 & \bar{E} \end{bmatrix} \\ &= \begin{bmatrix} 0.98L^* & 1.02R_0^* & 0.98C_2^* & 1.02G^* & 0.98G_a^* & 1.02G_b^* & 0.98C_1^* & 1.02E^* \end{bmatrix} \end{aligned}$$

The correlation matrix for the parameter effects based on the first output; i.e., $y_1 = x_1$, yields

$$R = \begin{bmatrix} 1.0000 & -0.1462 & 0.6368 & 0.1481 & 0.3840 & -0.3813 & -0.5898 & 0.3839 \\ -0.1462 & 1.0000 & 0.2325 & -0.9606 & -0.9655 & 0.9656 & -0.1170 & -0.9657 \\ 0.6368 & 0.2325 & 1.0000 & -0.0675 & -0.0032 & -0.0041 & -0.9823 & -0.0042 \\ 0.1481 & -0.9606 & -0.0675 & 1.0000 & 0.9515 & -0.9517 & -0.0782 & 0.9512 \\ 0.3840 & -0.9655 & -0.0032 & 0.9515 & 1.0000 & -0.9997 & -0.1018 & 1.0000 \\ -0.3813 & 0.9656 & -0.0041 & -0.9517 & -0.9997 & 1.0000 & 0.1085 & -0.9997 \\ -0.5898 & -0.1170 & -0.9823 & -0.0782 & -0.1018 & 0.1085 & 1.0000 & -0.1007 \\ 0.3839 & -0.9657 & -0.0042 & 0.9512 & 1.0000 & -0.9997 & -0.1007 & 1.0000 \end{bmatrix}$$

which indicates collinearity $|\rho_{ij} = 1|$ between the parameter effects of G_a G_b and E. Parameter estimation was, therefore, performed on only five of the parameters.

With only the first output used; i.e., $y = x_1$, the estimates by PARSIM are shown in Figure 3.7 along with the estimates from the Gauss-Newton method (NLS). The estimation results indicate that two of the parameters, C_2 and C_1, remain completely

unchanged by PARSIM from their initial values. In contrast, the Gauss-Newton method continues to adapt the parameters at each iteration, albeit for 300 iterations before they reach their correct values. These results are representative of the tendency of the Gauss-Newton method to continually adapt the parameters even when the gradient is quite small. Therefore, one advantage of integration of PARSIM with NLS is continual adaptation of parameters. The other advantage is PARSIM's propensity to evade local minima. This is illustrated through the non-convex contour of a Van der Pol oscillator.

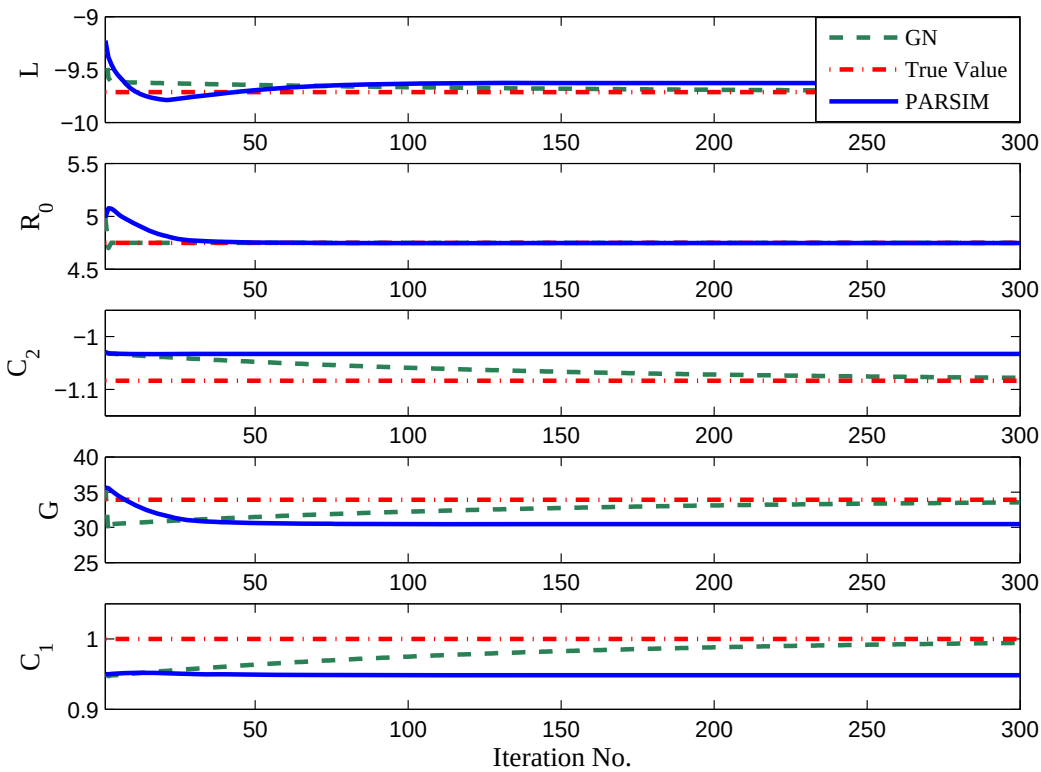

Figure 3.7: Estimates of five parameters of the Chua's circuit obtained with the first output, $y_1 = x_1$ by both PARSIM and the Gauss-Newton method.

The Van der Pol oscillator is used with its true parameters defined as $\mathbf{\Theta}^* = [m^*, c^*, k^*] = [375, 9800, 130 \times 10^3]^T$, and was simulated with the initial condition

$x(0) = 0.02,\ \dot{x}(0) = 0$. Both PARSIM and the Gauss-Newton method were applied to this model for estimation of the parameters c and k. Only these two parameters were considered to enable graphical representation of the error contour. Two starting points were then selected on the non-convex region of the error surface and both PARSIM (using the Gauss wavelet transform) and the Gauss-Newton method were applied to the estimation of parameters from these two starting points. The trajectory of estimates by the two methods obtained with an adaptation step of $\mu = 0.75$ are shown in Figure 3.8. The results indicate that PARSIM, because of its independence from the gradient of the contour, can lead the estimates to their correct values (the bottom of the bowl) whereas the Gauss-Newton method misses them due to the unfavorable location of the starting points.

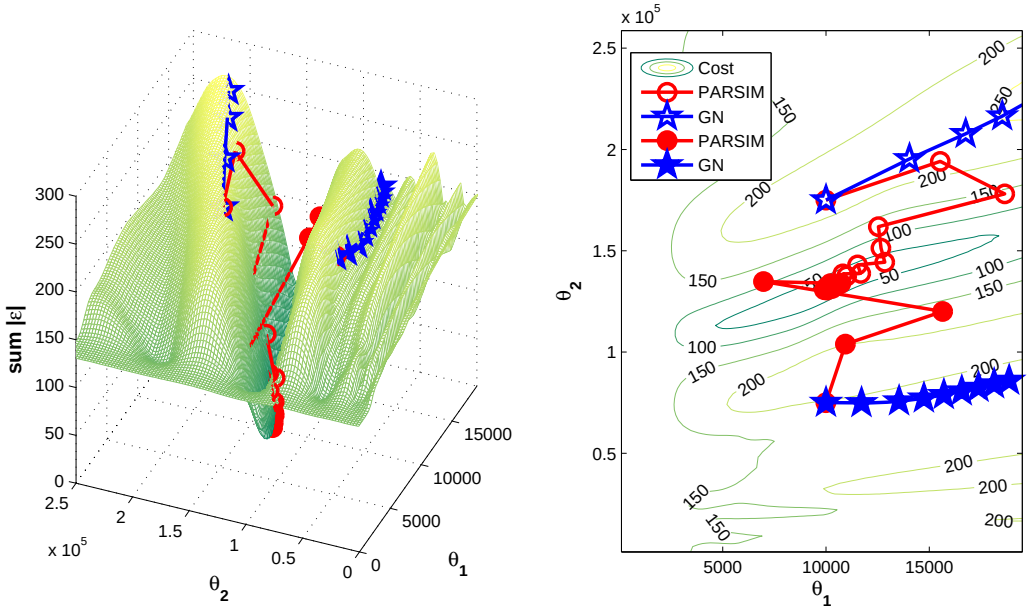

Figure 3.8: Two cases where nonlinear least squares would have inferior performance relative to PARSIM due to the location of the initial parameter values.

In order to demonstrate the proposed integration of estimation solutions, a hybrid approach is devised to integrate the solutions from PARSIM and NLS. For this, a competitive mechanism is designed whereby the two solutions are evaluated concur-

rently after every so many iterations to evaluate the effectiveness of each solution in reducing the prediction error. Similar techniques like this have of course been devised between nonlinear least squares and other algorithms, like genetic search, that are equally immune to local minima entrapments. The advantage of PARSIM, however, is that it is as efficient as NLS in finding the global minimum, so the cost associated with the added search is reduced considerably.

For illustration purposes, the competitive scheme was designed according to the magnitude of the prediction error at each iteration. The results from the application of this competitive approach to the two starting points in Figure 3.8 are shown in Figure 3.9. The results indicate that the solution trajectory is primarily from PARSIM, especially at the beginning of the search, when NLS is incapable of delivering a good trajectory toward the global minimum. For a better perspective of the results from this competitive approach, also shown are the prediction errors in Figure 3.10. Here we see that, for these two starting points, the hybrid approach was able to consistently reduce the prediction error and improve upon the parameter estimation results from either NLS or PARSIM being used independently.

3.5 Discussion

The adaptation results are to only demonstrate the viability of the estimated parameter errors. In that light, the results from the Gauss-Newton method are only meant as a reference and not for a comparative study. Nevertheless, the results represent some important aspects of PARSIM. An obvious drawback of PARSIM is its considerable computational expense, almost ten times that of regression on a modern PC. As such, the added computation needs to be justified by superior performance. We believe PARSIM's potential for integration of multiple outputs, due to added transparency in the time-scale domain, and the capacity to cope with noise, by taking advantage of the noise suppression techniques of time-scale domain, are two of the

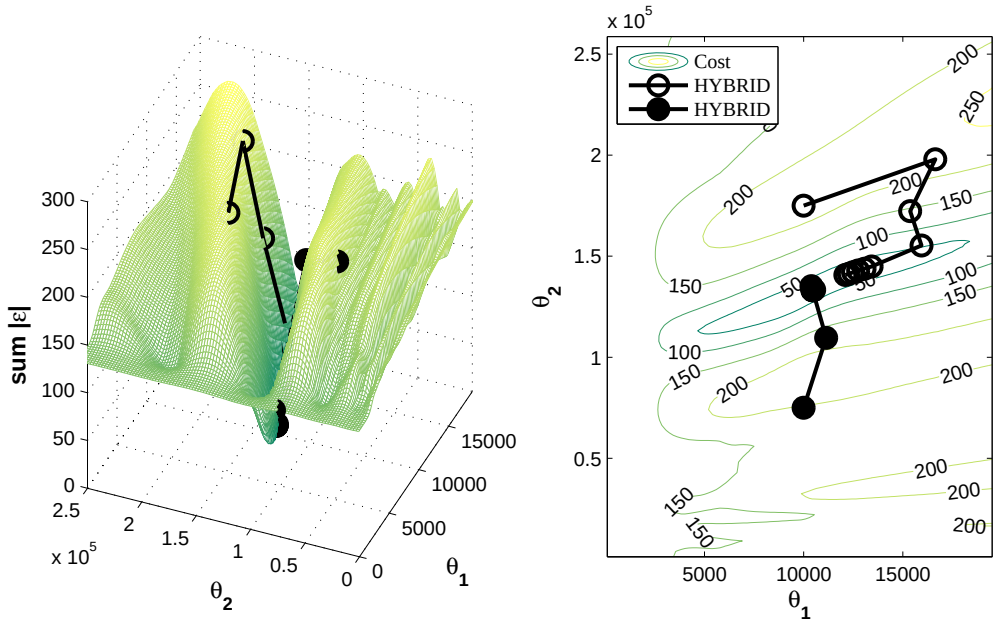

Figure 3.9: Convergence behavior of the competitive approach to parameter estimation for the starting points in Figure 3.9 where nonlinear least squares would get trapped in local minima.

features of PARSIM that could compel its use in practical applications. To this end, some of the research issues that need further attention are briefly discussed below.

- *Effect of Thresholds*

  A methodical study of the effect of the threshold level, η_t in (1.13), or the dominance factor, η_t in 1.16, on the extracted signatures would entail investigating its effect on the quality of the estimated parameter signatures. One measure of this quality is the consistency of parameter error estimates obtained. In this chapter, the variance of the parameter estimates at each pixel of a parameter signature was presented as a measure of quality for threshold selection. Another potential measure is how well the extracted parameter signatures satisfy the notion of parameter dominance, which can be assessed by the distance of the wavelet coefficients from the threshold level at the pixels included in the parameter signature. Using this measure one could search among the possible

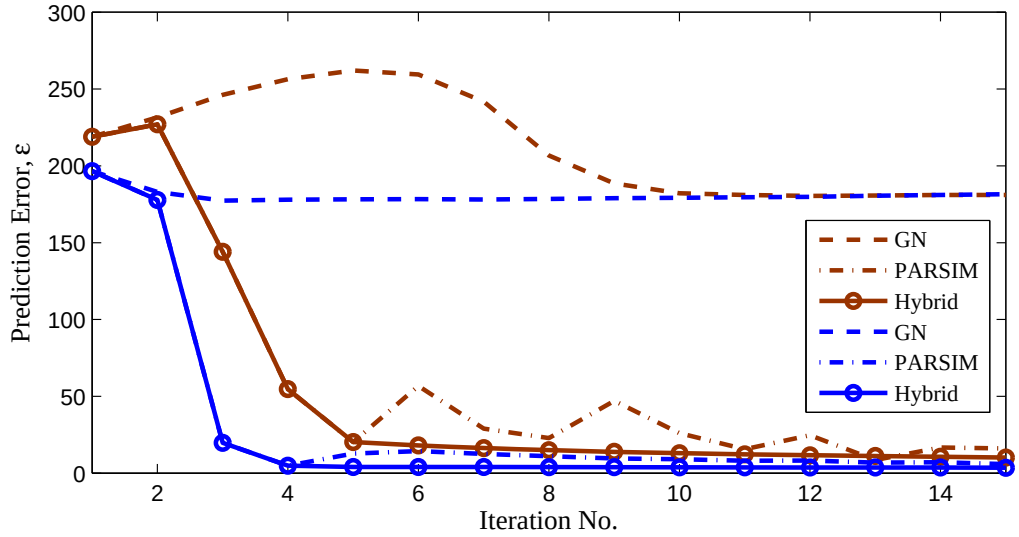

Figure 3.10: The prediction error of the Van der Pol oscillator by the three parameter estimation methods for the starting points in Figure 3.8.

threshold levels in the range, say $\eta_t = (0,\ 0.2]$, to define the best η_t for each adaptation iteration.

- *Type of Wavelet Transform*

  There is considerable literature on the smoothing effect of different wavelets and their multi-scale differentiation capacity [46, 45, 47]. Our preliminary results indicate that the better delineation of the parameter effects by a WT in the time-scale domain translates into more reliable parameter signatures. But further study is needed before such a sole measure can be used for selecting the wavelet.

- *Adaptation Step*

  Ideally the adaptation step μ in (3.1) should be selected according to the convexity of the model which is an important factor in the global accuracy of the parameter error estimates [43]. In PARSIM, the adaptation step size has been selected based on the identifiability of model parameter and this has produced promising results. Experimentation with adjusting the adaptation step μ associ-

ated with each parameter adaptation according to the quality of the parameter signatures has also been conducted, but the results have been mixed. More research is therefore needed before the quality of the parameter signatures can be effectively factored into the adaptation step.

CHAPTER 4

NOISE COMPENSATION BY PARSIM

PARSIM's capacity to perform parameter estimation in the time-scale domain obviates the need to reconstruct the signal in the time domain. This removes an important constraint that has so far impeded development of more effective denoising techniques for improving parameter estimation.

The common approach to improving the signal-to-noise ratio is to low-pass filter the measurements [31, 12]. Among them, particularly noteworthy is the method of filtering introduced by Donoho and co-workers [16, 18, 17] which transforms the signal to the time-scale domain, reduces the high frequency noise by thresholding the wavelet coefficients in the lower scales (associated with the higher frequencies) and then reconstructs the wavelet coefficients back in the time domain [16]. Similar to the above approach, is the wavelet shrinkage method [9] that uses Bayesian priors to associate the noise level with the distortion of the wavelet coefficients for their shrinkage. Even though the reconstructed signal has been shown to be minimax [18], it is not necessarily suitable for improving the parameter estimates, due to the disconnect between denoising and the parameter estimation process. The Parameter Signature Isolation Method (PARSIM) [13] not only provides the missing link between denoising and parameter estimation but also obviates the need to reconstruct the signal in the time domain due to its sole reliance on the time-scale domain for parameter estimation.

In PARSIM, each model parameter error is estimated separately in isolated regions of the time-scale domain wherein the parameter is speculated to be dominantly

affecting the prediction error. Since the parameter error estimates depend on the prediction error in isolated regions, they can benefit from a method that discounts the parameter error estimates according to the estimated distortion of the prediction error at each pixel of the time-scale domain. Such a noise compensation is introduced in this chapter with results that indicate improvement in the parameter estimates beyond the other filtering/denoising techniques considered here.

4.1 Noise Compensation Technique

The noise-compensation method shown in this thesis estimates the distortion by noise of the wavelet coefficients of the prediction error, $W\{\epsilon\}$. Noise distortion is estimated by smoothing the wavelet coefficients in the time-scale domain and is based on the assumption that an estimate of the signal distortion due to noise can be obtained from the difference between the noisy signal and its smoothed version. For an illustration of this assumption, one can refer to the three plots in Figure 4.1 that show the real and noisy impulse responses of the mass-spring-damper model along with its smoothed version by low-pass filtering. For reference, the mass-spring-damper model has the form:

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = u(t) \tag{4.1}$$

where the model parameters are m, c, and k. It is clear that even though the smoothed signal in Figure 4.1 does not match the true signal, especially in the more choppy segments of the signal, its does provide an estimate of the signal's distortion by noise.

The proposed noise-compensation method first estimates the distortion by noise of the wavelet coefficients of the prediction error, $W\{\epsilon\}$. It then uses this estimate to assign confidence to the parameter error estimates obtained at each pixel. This notion is utilized in the time-scale domain for estimation of noise distortion of the wavelet coefficients of the prediction error.

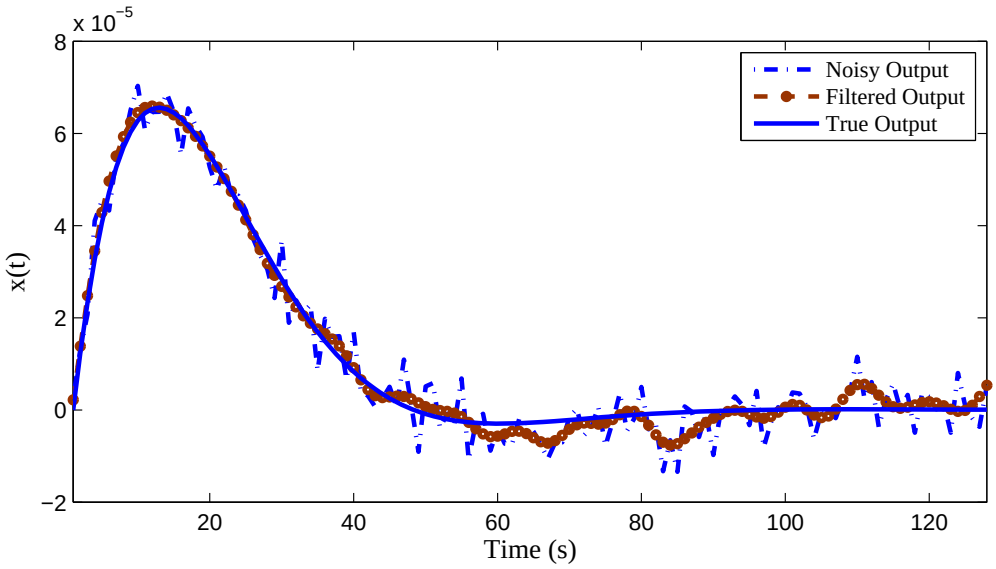

Figure 4.1: The real impulse response of the mass-spring-damper model with and without noise, and its low-pass filtered version.

In order to estimate the level of distortion by noise of the wavelet coefficients in different regions of the time-scale domain, the time data at each scale is considered as a signal like the noisy output in Figure 4.1. In this light, we can define time smoothing as

$$S_{s_l}\left(W\{f\}(t_k)\right) = S\left(W\{f\}(t_k, s_l)\right) \quad t_1 \leq t_k \leq t_N \tag{4.2}$$

where S denotes the smoothing function and S_{s_l} the time-smoothed wavelet coefficients along the time samples at scale s_l. For illustration purposes, the smoothed $W\{\epsilon\}$ by an 8th order polynomial fit along time for a sample prediction error $\epsilon(t)$ is shown in Figure 4.2. It is clear from the results that time-smoothing is effective in reducing the rapid changes in the wavelet coefficients. Using the time-smoothed wavelet coefficients, $S_s(W\{\epsilon\})$, the distortion by noise at each pixel can be estimated as

$$\widehat{W\{\nu\}} = W\{\epsilon\} - S_s(W\{\epsilon\}) \tag{4.3}$$

where $\widehat{W\{\nu\}}$ denotes the estimate of the wavelet coefficients of noise and $S_s(W\{\epsilon\})$ represents the time-smoothed wavelet coefficients of the prediction error as defined

in (4.2) for each pixel. It should be noted here that one could potentially smooth along the scale axis (scale smoothing) as well as the time axis. This, however, does not produce a significant improvement in the estimation of noise, due to an already smooth profile of the scales at each time sample.

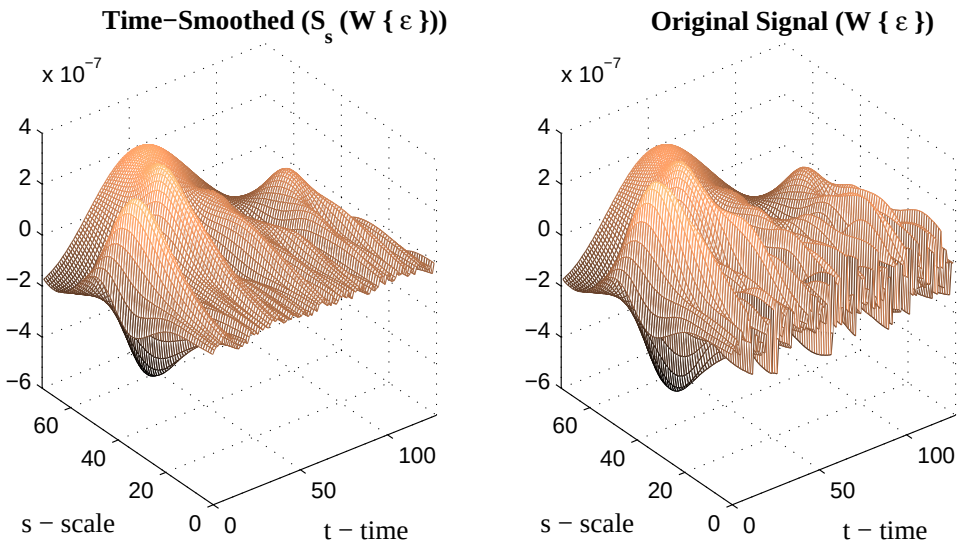

Figure 4.2: The noise affected wavelet coefficients smoothed along time (left) and the wavelet transform of the original signal (right).

To illustrate this point, let us consider the wavelet transform of the noise in Figure 4.1 and its estimate according to (4.3), shown in Figure 4.3. The two sets of wavelet coefficients indicate that the estimate in (4.3) is very similar to reality, particularly in the lower scales where the distortion by noise is the most pronounced. The relatively poor estimate of noise at the higher scales is due to absence of rapid variations of $W\{\epsilon\}$ at the higher scales which precludes any difference between $W\{\epsilon\}$ and its time-smoothed version, $S_s(W\{\epsilon\})$.

The above approximation is, of course, too coarse to be used for denoising the prediction error. Instead, it can be used as a confidence factor in the range $[0, 1]$ to discount the parameter error estimates according to (1.11) from Chapter 1. The confidence factor, w_{kl}, which is defined as

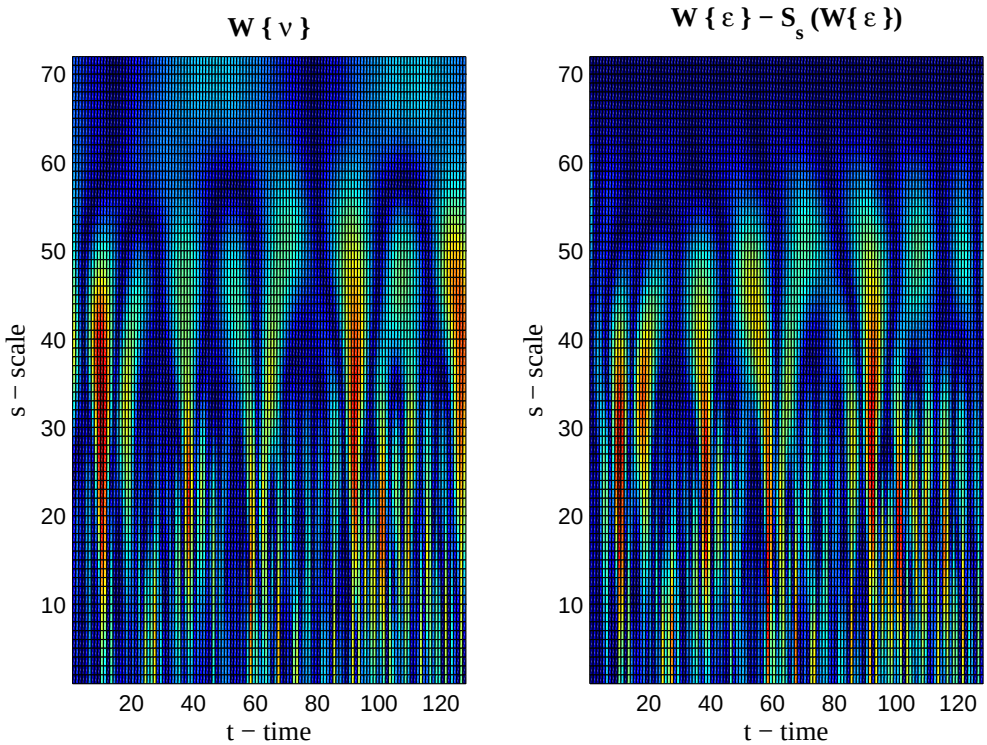

Figure 4.3: Wavelet transform of a noise sample and the difference between the wavelet transform of the prediction error and its smoothed version.

$$w_{kl} = 1 - \left| \frac{\widehat{W\{\nu\}}(t_k, s_l)}{max_{(t,s)}\widehat{W\{\nu\}}} \right| \tag{4.4}$$

can then be incorporated as the weight of the prediction error in the estimation of the parameter error in (1.11) to yield the biased parameter estimate as:

$$\widehat{\Delta\theta}_{ib} = \frac{1}{N_i} \sum_{k=1}^{N} \sum_{l=1}^{M} \frac{w_{kl} W\{\epsilon\}(t_k^i, s_l^i)}{W\{\mathcal{E}_i\}(t_k^i, s_l^i)} \quad \forall\, (t_k^i, s_l^i) \in \Gamma_i \tag{4.5}$$

where the subscript b denotes the bias in the parameter error estimate.

For illustration, the confidence factors obtained for the sample prediction error in Figure 4.1 are shown in Figure 4.4. Using confidence factors such as those in (4.5) to bias the parameter error estimates yields the parameter estimates in Table 4.1 which are shown together with those obtained without the confidence factors. The results

show significant improvement in the parameter estimates due to the biased estimates in (4.5). What is even more appealing about this noise compensation method is that it can also be used in conjunction with time-filtering. To demonstrate the potential benefit of this two-pronged approach, also shown in Table 4.1 are the parameter estimates obtained according to (4.5) after the prediction error had been filtered with a low-pass filter (*Filter 1*). The results are clearly more precise than before.

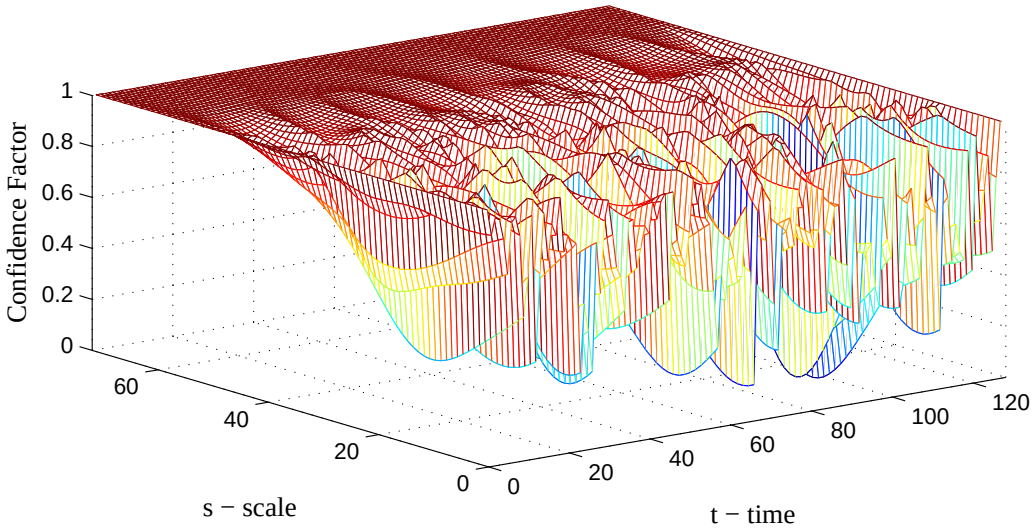

Figure 4.4: Estimated confidence factor at each pixel used as weights w_{kl} in the estimation of $\widehat{\Delta\theta}_{ib}$ by (4.5).

Table 4.1: Twenty-fifth iteration mean estimates of the mass-spring-damper parameters by PARSIM without and with the confidence factors before and after filtering the time signal.

True Parameter Value	**Parameter Estimates**		
	PARSIM	PARSIM (Biased)	PARSIM (*Filter 1* + Biased)
$m = 375$	358	361	371
$c = 9800$	9606	9593	9690
$k = 130 \times 10^3$	128×10^3	130×10^3	132×10^3
Precision Error	0.0518	0.0439	0.0240

4.2 Noise Level Analysis

The results reported to this point, although not comprehensive, demonstrate the effectiveness of the noise compensation method. The improvement in the parameter estimates depends not only on the smoothing method used but also the convexity of the model, the wavelet transform used, the resolution of the time-scale plane (number of time samples and scales used for transformation), as well as the level of noise present in the data. Among these, the level of noise and its effect on the parameter estimates requires further attention. For this, parameter estimates were obtained with different noise levels by both PARSIM and the Gauss-Newton method. The estimation results obtained with zero mean Gaussian noise of different magnitudes are shown in Figure 4.5. The parameter estimates from the Gauss-Newton method were obtained with noisy output, and filtered outputs by a low-pass filter, *Filter 1*, and a denoising filter based on hard wavelet thresholding [16] of lower frequencies, *Filter 2*. The estimation results from PARSIM were obtained with the noisy output according to both (1.11) and (4.5) and with a filtered output, by *Filter 1*, using (4.5). The results indicate that, as expected, all the estimates are adversely affected by the noise level. They also confirm that the Gauss-Newton method benefits from filtering in the time domain and that the most benefit is attained from *Filter 2* which performs thresholding of the wavelet coefficients in the time-scale domain. The best overall estimates, however, are still provided by PARSIM using biased estimates with low-pass filtered outputs. Here it should be noted that these results are not necessarily the best that could be obtained with PARSIM. For instance, the smoothing in the time-scale domain, which was obtained with a polynomial fit of the same order at all the noise levels, could be changed to more effectively estimate the noise distortion. It would also be possible to take advantage of a denoising measure like thresholding to better estimate the noise distortion.

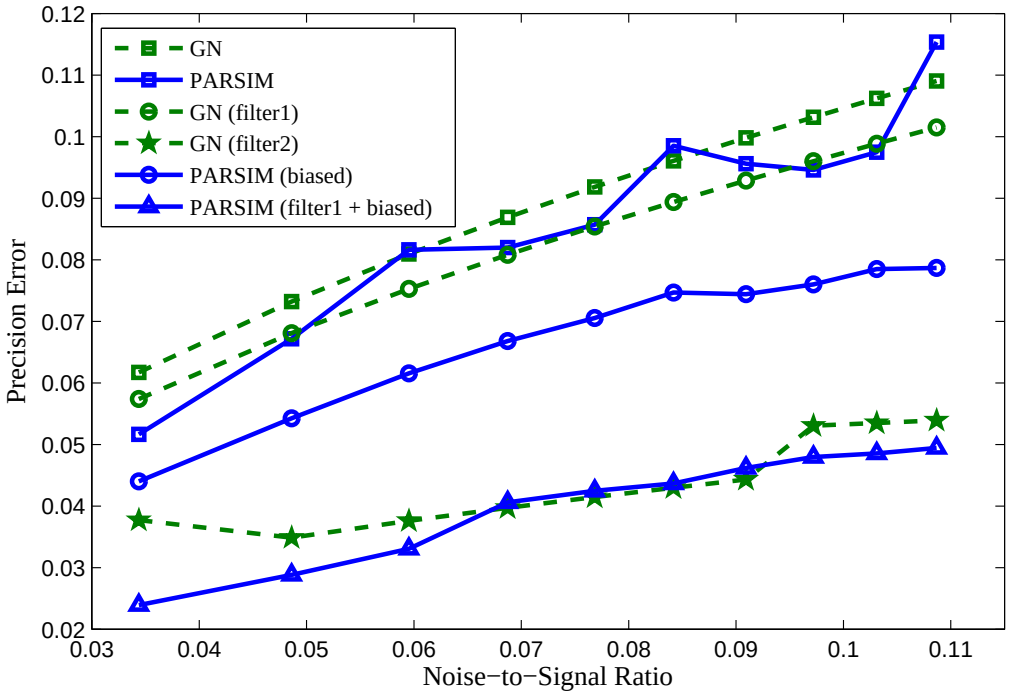

Figure 4.5: The precision error obtained with PARSIM and Gauss-Newton method at different levels of signal-to-noise ratio.

Another issue worth evaluating is the type of noise. All the results obtained so far are with additive zero-mean Gaussian noise, so the question arises as whether the proposed method would be as effective with another type of noise, say, one with a uniform distribution. This point was evaluated by repeating the estimates with an output contaminated with additive uniformly distributed noise. For brevity, only shown in Figure 4.6 are the estimates from the Gauss-Newton method, PARSIM and biased PARSIM with the low-pass filtered time signal, which show that the biased estimates from PARSIM with the low-pass filtered signal are equally as improved, in this case, as their counterparts obtained with Gaussian noise.

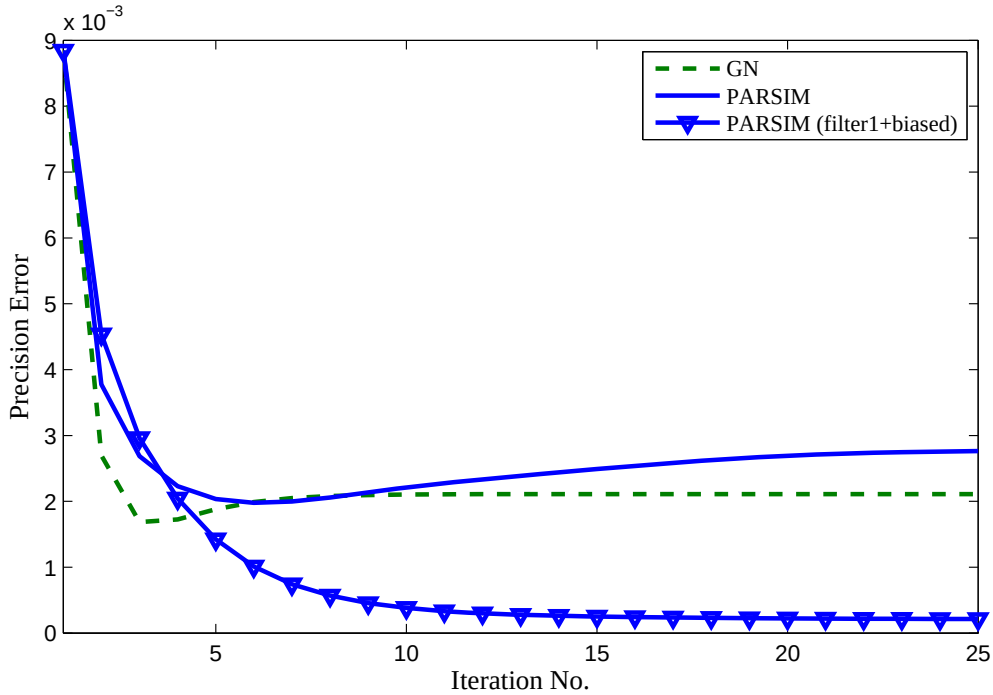

Figure 4.6: The improvement in the precision error of the proposed method when noise has a uniform distribution.

4.3 Discussion

Insofar as the focus of this thesis describing the notion of direct noise compensation in the time-scale domain and the potential improvements it can provide for parameter estimation, the results sufficiently support the chapter's focus. In view of this focus, we do not claim to be presenting the best method of noise compensation in the time-scale domain, nor do we propose this method as superior to the other types of denoising and/or filtering. The level of improvement achieved in parameter estimation by PARSIM depends not only on the smoothing method used for time-smoothing in (4.3) but also on the convexity of the estimation model, the wavelet transform used, the resolution of the time-scale plane (number of time samples and scales used for transformation), as well as the level of noise present in the data. As such, further work is required to truly develop the proposed noise compensation method and realize its

potentials. Some of the issues to be considered in the implementation of this method are:

- *Significance of noise distortion estimates.* A major contribution of this research is the connection between the noise distortion estimates and the confidence factors used to bias parameter estimation by PARSIM. In this chapter, we have only sufficed to a rudimentary method of noise distortion estimation by 8th degree polynomial smoothing of the time signals at individual scales, but we believe more effective methods of noise distortion estimation are possible that will inevitably improve the proposed noise compensation method.

- *Computation cost.* Another issue associated with this method is its computational expense. Relative to the methods of denoising, which reconstruct the signal back in time domain, the proposed method provides the advantage of precluding inverse wavelet transformation demanded by reconstruction. On the other hand, PARSIM requires the wavelet transformation of the prediction error as well as the output sensitivities. It also requires estimation of noise distortion, which is performed here through the laborious smoothing of individual time series at each scale. As such, the proposed method is appropriate for cases where the precision of parameter estimates outweighs the inherent cost of computation involved.

- *Evaluation criterion.* A luxury afforded by simulation-based studies is the availability of the precision error. In practical cases, however, the true parameter values are unknown and the only criterion for parameter estimation effectiveness is the prediction error which is not necessarily indicative of the precision of estimates (see Table 4.1). Accordingly, one does not have the freedom to choose the type of time-smoothing, for instance, to determine the best method of noise distortion estimation. So one will have to generalize the noise compensation

method and all its constituents *a priori*. Such an effort would require a certain level of simulation-based experimentation that can only be justified when the precision of estimates is of particular importance.

CHAPTER 5

MEASUREMENT SELECTION FOR JET ENGINE TRANSIENTS

In environments that are conducive to sensory measurements of various types and/or at different locations, there is often a need for measurement economy to improve computational performance and efficiency and/or to reduce sensor cost and maintenance. To address this need, sensors and their locations are selected to eliminate redundancy. A case in point is the turbo-jet engine for which pressures, temperatures, and rotational speeds can be measured at various locations [72], yet only a limited number of them can be realistically used for monitoring purposes and fault diagnosis [22].

Sensors and their locations may be selected according to practical considerations such as ease of measurement, sensor reliability and cost. But these considerations are secondary to the observability the measurements provide to the process. One aspect of this observability concerns parameter identifiability, as it pertains to whether "*the parameters of [the] model can be uniquely (globally or locally) determined from data*"[42]. The formal definition of parameter identifiability is provided in (5).

In turbo-jet engines, a predominant use of measurements is for on-board performance monitoring [77, 15], to estimate a set of 'health parameters' that represent the efficiency and flow capacity of individual components [44, 35, 49, 5]. Due to practical considerations, however, the measurements used for in-flight health monitoring are acquired during steady-state operation and, as such, they only characterize the static aspect of the engine behavior [36]. Under this scenario, since the information content

of each steady-state measurement is confined to a static gain, as many measurements are required as the number of parameters to be estimated [22]. As to which measurements to be included in the measurement suite, parameter identifiability is determined according to the correlation coefficients between the output sensitivities [66].

However, sensory measurements contain much richer content when their transients are included to represent the dynamics of the engine [40]. Even though such transient measurements may not be readily available in-flight, they are accessible in the form of transient decks from test stands and are particularly valuable for fault diagnosis of faulty engines [56, 57, 39, 29, 30] or tuning the models of newly developed engines. Selection of transient measurements, however, is more challenging because of the need to account for the higher level of observability provided by the transients.

Selection of dynamic measurements is generally model-based were as the potential process measurements are represented by the outputs of the model in relation to the process states and parameters. According to the model, the value of combinations (suites) of measurements can be evaluated for the observability they provide to the states or parameters to be estimated. For state estimation, the selection criteria is often a scalar measure such as the trace, the determinant, or the smallest singular value of the observability matrix [61] or the state prediction error covariance matrix [55]. Similarly, for parameter estimation, the selection criteria are related to the Fisher information matrix [54] as those widely used for chemical plants [38, 3, 82, 10, 58], civil structures [71, 76, 62, 83, 48], and others [21]. The above criteria, however, need to be computed for each combination of measurements and even then they represent only an assessment of the overall parameter identifiability provided by each measurement combination (suite). As is shown in this thesis, transforming the output sensitivities into the time-scale domain circumvents the above limitations. The method outlined here specifies the identifiability of each parameter by individual outputs; as such, it provides considerable resolution to the measurement selection process.

5.1 Measurement Selection Background

Measurement selection that has traditionally involved dynamic data entails searching for the suite of outputs that provide maximum observability for the parameters. The different criteria to assess the observability of outputs are all associated with the Fisher information matrix, $\mathbf{\Omega}_\theta$, which is the lower bound of the Cramer-Rao inequality [28]:

$$\operatorname{cov} \widehat{\mathbf{\Theta}} \geq \mathbf{\Omega}_\theta^{-1} \tag{5.1}$$

where $\operatorname{cov} \widehat{\mathbf{\Theta}}$ denotes the covariance matrix of the parameter estimate $\widehat{\mathbf{\Theta}}$. The rationale for using the Fisher information matrix is that the more identifiable the parameters by the outputs, the smaller the variance of the "minimum variance unbiased estimator" that equates this lower bound for linear systems [28]. Formally, output selection entails selecting the optimal M-dimensional output suite from a total of P outputs, where $M \leq P$. This selection can be performed by any of the following strategies: (i) minimizing the trace of $\mathbf{\Omega}_\theta^{-1}$, (ii) maximizing its largest eigenvalue, or (iii) minimizing its determinant [54].

The direct applicability of the above criteria, however, is limited in practical applications. Limitations stem from the assumption of linearity associated with the Fisher information matrix as well as the need to estimate the criterion for individual suites of measurements [48]. However, before discussing the limitations, let us study the underlying concept. To this end, let us consider the complete suite of P outputs where the vector of sampled data $\mathbf{Y}$ would consist of the P outputs stacked after each other as

$$\mathbf{Y} = [y_1(t_1) \dots y_1(t_N), \dots, y_P(t_1) \dots y_P(t_N)]^T$$

For the sake of discussion, let us first consider the ideal scenario of $\mathbf{Y}$ being in linear-in-parameter form:

$$\mathbf{Y} = \mathbf{\Phi\Theta} + \nu \tag{5.2}$$

where $\mathbf{\Phi} \in \Re^{PN \times Q}$ represents a known matrix that is independent of $\mathbf{\Theta}$ and $\nu \sim N(0, \sigma^2 \mathbf{I})$ denotes measurement noise. For this case, the minimum variance unbiased estimator is the least-squares estimator:

$$\widehat{\mathbf{\Theta}} = (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T \mathbf{Y} \tag{5.3}$$

and the Fisher information matrix has the form

$$\mathbf{\Omega}_\theta^{-1} = (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \sigma^2 \tag{5.4}$$

Minimizing the Fisher information matrix for this case will correspond to maximizing the determinant of $\mathbf{\Phi}$. This will, in turn, be synonymous with increasing the spread of columns of this matrix corresponding to more compactly distributed singular values (eigenvalues) [33].

But given that the linear-in-parameter form is unrealistic, one can resort to a more generic form of $\mathbf{Y}$, obtained by its first-order approximation, as

$$\mathbf{Y} \approx \widehat{\mathbf{Y}}(\bar{\mathbf{\Theta}}) + \mathbf{\Phi}(\bar{\mathbf{\Theta}}) \Delta \mathbf{\Theta} + \nu \tag{5.5}$$

at the nominal parameter vector $\bar{\mathbf{\Theta}}$. Note that the form of this model is similar to that of (5.2), with $\Delta \mathbf{\Theta} = \mathbf{\Theta}^* - \bar{\mathbf{\Theta}} = [\Delta\theta_1, \ldots, \Delta\theta_Q]^T$ representing the vector of parameter errors and $\mathbf{\Phi}(\bar{\mathbf{\Theta}})$ denoting the Jacobian matrix of output sensitivities, as

$$\mathbf{\Phi}(\bar{\mathbf{\Theta}}) = \begin{bmatrix} \partial\hat{y}_1(t_1,\bar{\mathbf{\Theta}})/\partial\theta_1 & \dots & \partial\hat{y}_1(t_1,\bar{\mathbf{\Theta}})/\partial\theta_Q \\ \vdots & \ddots & \vdots \\ \partial\hat{y}_1(t_N,\bar{\mathbf{\Theta}})/\partial\theta_1 & \dots & \partial\hat{y}_1(t_N,\bar{\mathbf{\Theta}})/\partial\theta_Q \\ \vdots & \ddots & \vdots \\ \partial\hat{y}_P(t_1,\bar{\mathbf{\Theta}})/\partial\theta_1 & \dots & \partial\hat{y}_P(t_1,\bar{\mathbf{\Theta}})/\partial\theta_Q \\ \vdots & \ddots & \vdots \\ \partial\hat{y}_P(t_N,\bar{\mathbf{\Theta}})/\partial\theta_1 & \dots & \partial\hat{y}_P(t_N,\bar{\mathbf{\Theta}})/\partial\theta_Q \end{bmatrix} \tag{5.6}$$

This model, however, differs from (5.2) in that it is nonlinear-in-parameter because of the dependence of the Jacobian matrix $\mathbf{\Phi}(\bar{\mathbf{\Theta}})$ on $\Delta\mathbf{\Theta}$. As such, the parameters of this model need to be estimated iteratively as

$$\hat{\mathbf{\Theta}}(q+1) = \hat{\mathbf{\Theta}}(q) + \mu\widehat{\Delta\mathbf{\Theta}}(q) \tag{5.7}$$

with the parameter error estimates at each iteration q obtained by, say, nonlinear least-squares [70], as

$$\widehat{\Delta\mathbf{\Theta}} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\epsilon^N \tag{5.8}$$

where ϵ denotes the prediction error:

$$\epsilon^N = \mathbf{Y} - \widehat{\mathbf{Y}}(\bar{\mathbf{\Theta}}) \tag{5.9}$$

What is revealing about the above solution is that, like the solution to the linear-in-parameter model in (5.3), it also benefits from maximizing the spread of columns of the Jacobian matrix $\mathbf{\Phi}$. As such, the strategy of using measurements that yield the maximum spread between the columns of $\mathbf{\Phi}$ is also relevant for the nonlinear-in-parameter model of (5.5). It leads to maximal separation between the output sensitivities and improves local estimation performance by NLS at every iteration of (5.8). However, choosing outputs that yield the maximum spread of $\mathbf{\Phi}$ faces two

concerns: (i) the local nature of $\mathbf{\Phi}$ that stems from its dependence on the nominal parameter vector $\bar{\mathbf{\Theta}}$, and (ii) the case-specificity of $\mathbf{\Phi}$ that is associated with individual suites of outputs. The first concern; i.e., the local nature of $\mathbf{\Phi}$, can be addressed by estimating the average spread of $\mathbf{\Phi}$ at different nominal parameter vectors. But the second concern cannot be readily addressed. Output selection based on the matrix $\mathbf{\Phi}$ would entail using a criterion, such as the determinant, to provide a lumped assessment of the distinctness of the columns of $\mathbf{\Phi}$ provided by the output suite. The suite with the largest determinant will then be determined as optimal in providing the maximum overall observability to the model parameters. What is needed to address this second concern is a one-to-one account of observability between the parameters and outputs, which is shown in this thesis to be possible by transforming the output sensitivities; i.e., columns of $\mathbf{\Phi}$, into the time-scale domain.

5.2 The Engine Model

To demonstrate the measurement selection strategy, The engine model FANJETPW provided by Pratt & Whitney was used. FANJETPW is a simplified representation of the NPSS model in Matlab/Simulink^{TM} form. The FANJETPW model, which is in state-variable form, provides a low fidelity representation of a turbofan engine with basic control. As such, heat transfer effects between the component surfaces and the flow stream are ignored. Moreover, the model does not include any sensor or actuator dynamics. In this study, the neglect of sensor dynamics does not affect the selected measurements so long as these dynamics are independent of the parameters.

The engine model, characterizes the steady-state and dynamic performance of five major engine modules: the low and high pressure compressors and turbines, and the fan. Figure 5.1 depicts the stations where outputs are simulated. Hereafter the health parameters are referred to as as model parameters and, for brevity, are numbered

them in the following order: HPC_{Nc}, HPC_{eff}, HPT_{Nc}, HPT_{eff}, LPC_{Nc}, LPC_{eff}, LPT_{Nc}, LPT_{eff}, fan_{Nc}, and fan_{eff}. Overall there are 10 parameters considered for identifiability analysis in this model. The outputs considered are temperatures at stations 2.5 (T25), 3.0 (T30), and 5.0 (T50), pressures at stations 2.5 (P25) and 3.0 (P30), and the rotational speeds of both the core (N2) and the fan (N1). Similar to the parameters, the 7 practical outputs from this model are numbered in the following order: N2, N1, T25, T50, P25, T30, and P30.

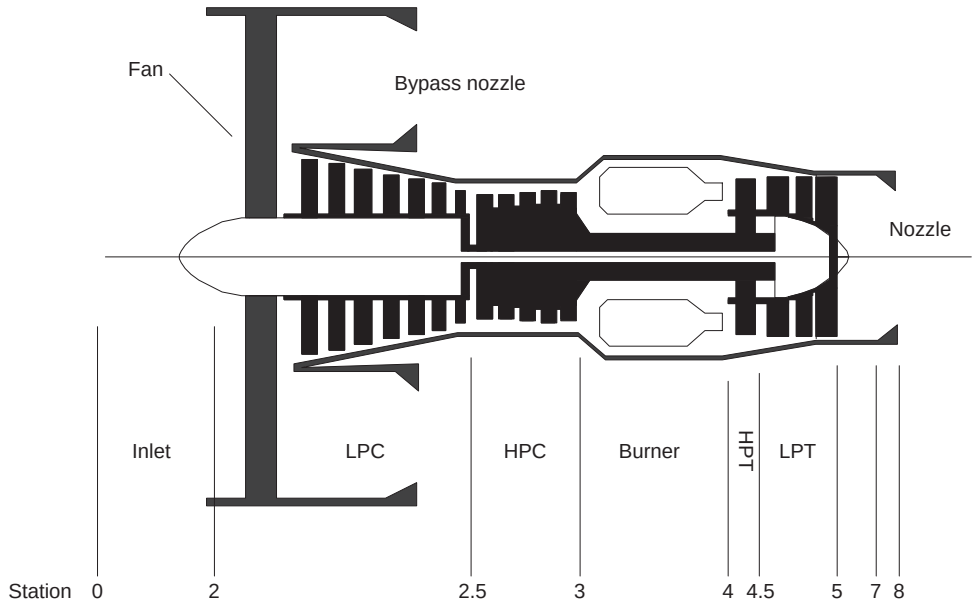

Figure 5.1: Schematic diagram of the high-bypass-ratio turbofan engine represented by the FANJETPW simulation model, together with the station and primary component locations.

For analysis purposes, the transient deck was simulated at the sea level static conditions: Mach # 0, T_{amb} = 518 R, P_{amb} = 14.7 psia, and Altitude = 0 ft, with the Power Lever Angle (PLA) varied according to the top plot of Figure 5.2 to excite engine dynamics. Here the PLA was held constant at 88.5 degrees for 10 seconds to allow for all startup transients to subside, it was then ramped down to 50 degrees over 5 seconds, held constant at 50 degrees for 5 seconds, and then ramped back up to 88.5 degrees over 5 seconds where it was held constant for another 5 seconds. As

is customary in engine analysis, the first 10 seconds of data associated with engine startup transients were excluded from analysis, marked by the dashed vertical line in Figure 5.2. For illustration purposes, four of the simulated outputs (that will be selected later as essential for parameter identifiability) are also shown in Figure 5.2.

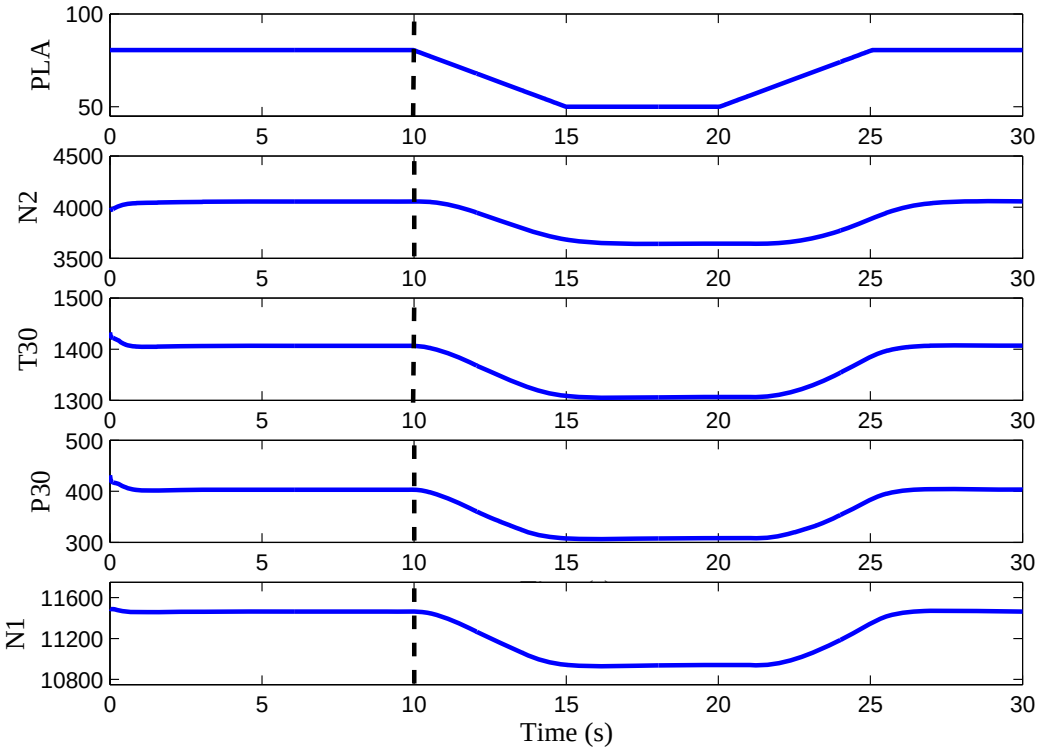

Figure 5.2: The input (Power Lever Angle in degrees) applied to the model to generate the transient outputs along with with 4 of the 7 simulated outputs.

5.3 PARSIM-Based Output Selection

The approach presented here capitalizes on the link between the existence of parameter signatures and the uniqueness of output sensitivities as the basis for parameter identifiability. For this, we rely on the fact that the more correlated an output sensitivity with the other output sensitivities, the less likely it is to extract the corresponding parameter signature. As such, the ability to extract a parameter signature

can be used as an indicator of observability for that parameter by this output.For this study, we used the parameter signatures extracted via a dominance factor in (1.16).

As was stated earlier, the magnitude of dominance factor η_d affects the parameter signatures. This is illustrated via the parameter signatures extracted at three different dominance factors in Figure 5.3 for one of the engine model parameters. The parameter signatures extracted at higher dominance factors in this figure include fewer pixels, but the more consequential effect of the dominance factor is in the consistency of the parameter error estimates across the parameter signatures. As such, of the higher quality the parameter signature is the more consistent parameter error estimates its pixels yield. This consistency can be quantified by the variance of the parameter error estimates in (3.6). To illustrate the above concept, the parameter error estimates obtained at individual pixels of the parameter signatures in Figure 5.3 are shown in Figure 5.4 along with the variance of the estimates across each parameter signature. The results indicate that the variance of the parameter error estimates is lower at higher dominance factors. With a dominance factor of $\eta_d = 2$, for instance, there is a significant scatter among the parameter error estimates in both positive and negative directions, but at the higher dominance factor of $\eta_d = 3$, the negative estimates diminish considerably toward the target estimate. In accordance with these results, the dominance factor will hereafter be used as the measure of quality of the parameter signature.

5.4 Connection to Parameter Identifiability

The key to the utility of parameter signatures in parameter identifiability analysis is the link between the quality of the extracted parameter signatures and the uniqueness of the corresponding output sensitivities; i.e., the columns of Jacobian $\mathbf{\Phi}$. This point was illustrated by the parameter signatures in Figure 1.5. Here one could not estimate the parameter signatures $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ with a dominance factor of, say, larger

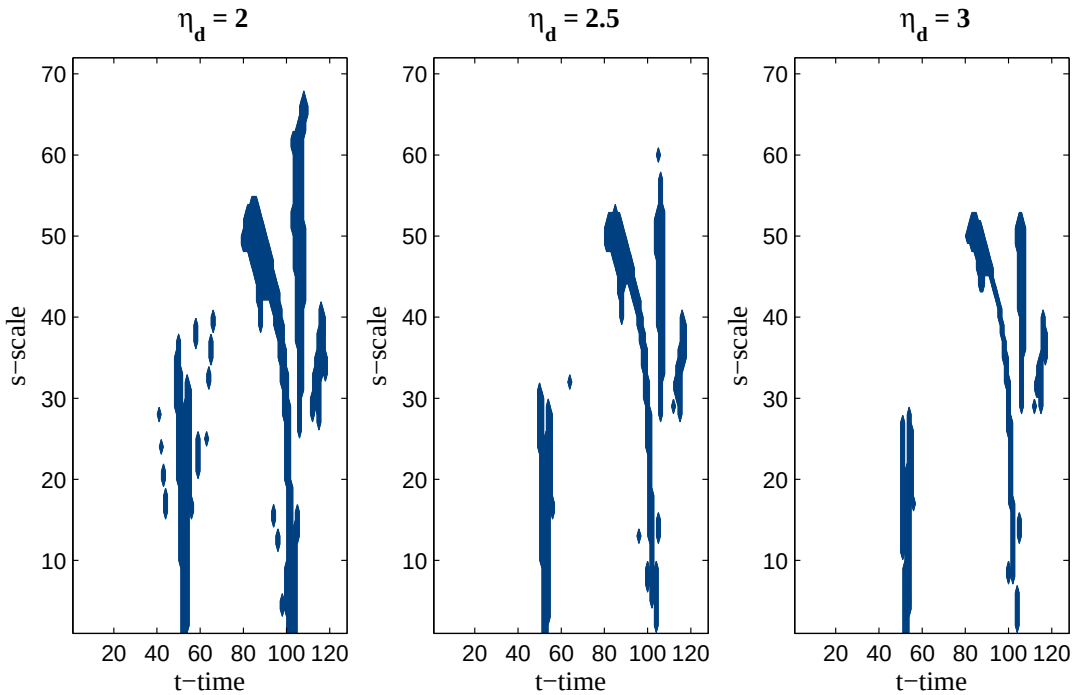

Figure 5.3: Parameter signatures of an engine parameter extracted at the three different dominance factors of 2, 2.5, and 3.

than 1.03, according to the normalized wavelet coefficient difference in Figure 1.4, whereas $\hat{\Gamma}_3$ and $\hat{\Gamma}_4$ could be obtained with very large dominance factors according to the wavelet coefficients in Figure 1.4. As such, the existence of a parameter signature extracted by a decent dominance factor of, say, larger than 2 will ensure the corresponding output sensitivity to be adequately uncorrelated with the rest, hence the identifiability of the parameter by the output. We hereafter use the existence of a parameter signature as the criterion for parameter identifiability by the output.

The second advantage of parameter signatures concerns the delineation they provide to the localized dissimilarities of output sensitivities. To illustrate this point, let us consider the N2 output sensitivities with respect to parameters LPT_{eff} and fan_{eff} in the left plot of Figure 5.5, which have a correlation coefficient of 0.9361. Although the two output sensitivities have a similar overall shape, as represented by their high correlation coefficient, they have distinct local differences that can be

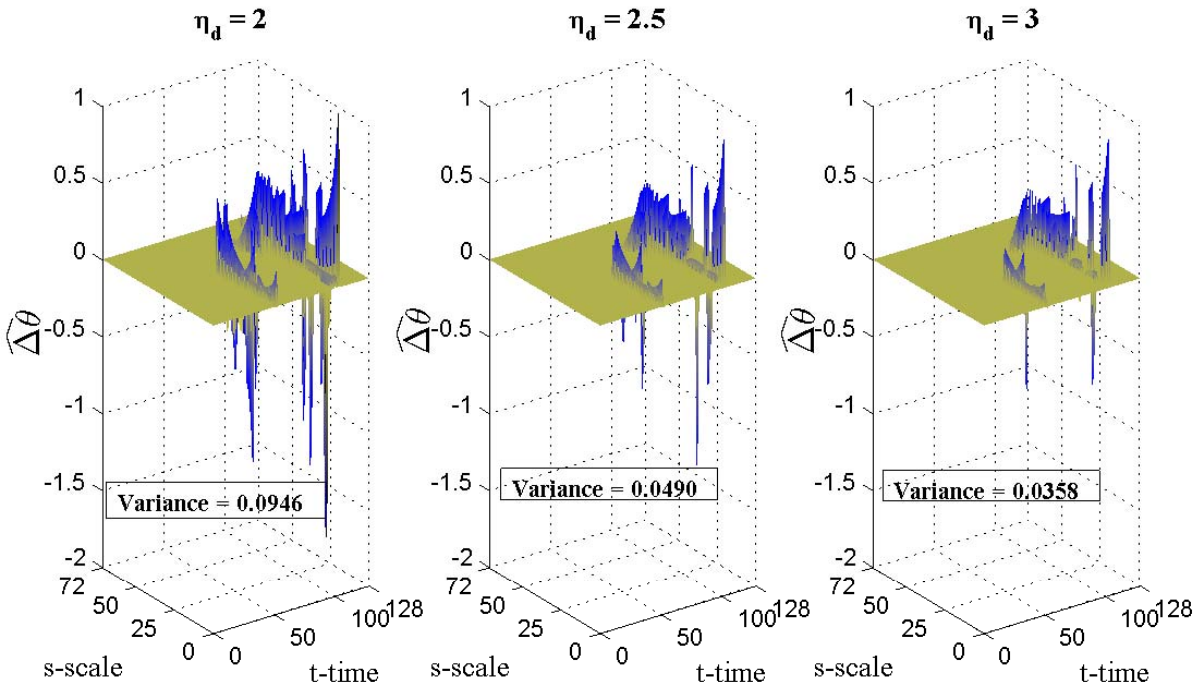

Figure 5.4: Estimated value of the parameter error at individual pixels of the parameter signatures in Figure 5.3.

accentuated by their wavelet coefficients. A case in point is the time window of 15 to 17.5 seconds, showed by the dotted lines in this figure. The localized dissimilarity of the output sensitivities is accentuated in the differential wavelet coefficients, $|\bar{W}(\partial N1/\partial LPT_{eff})| - |\bar{W}(\partial N2/\partial fan_{eff})|$, in the right plot of Figure 5.5 in the corresponding range of (96-112 time samples) indicated by the dotted lines in the time-scale plane. As is shown in the right-hand plot, the ridge that leads down toward the lower scales in this time window is a reflection of the difference in the higher frequency components of the output sensitivities in the left-hand plot in the same time window. Whereas such local dissimilarity is masked in a lumped measure such as the correlation coefficient, the pixels associated with this ridge will be included in the parameter signatures as indication of the localized dissimilarity between the output sensitivities.

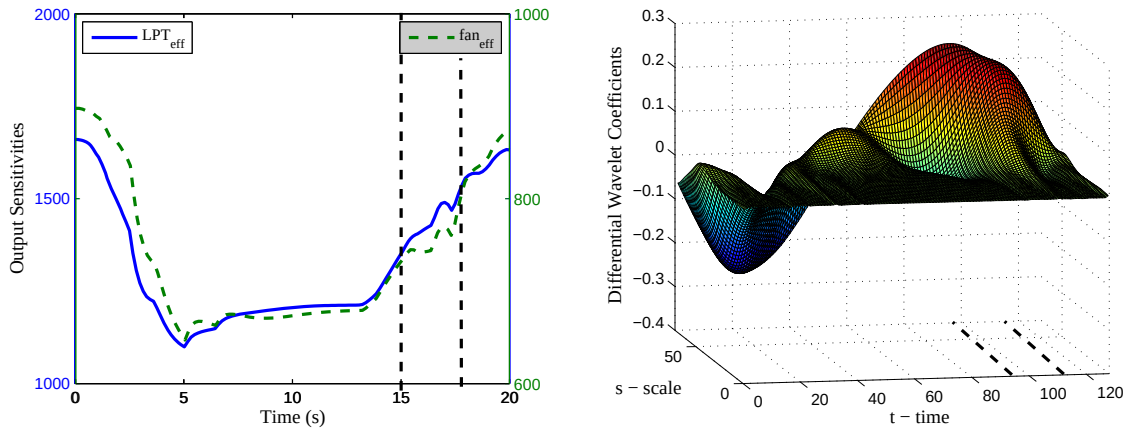

Figure 5.5: Illustration of the added resolution provided by wavelet coefficients in delineating the local dissimilarities between output sensitivities (between the dotted lines).

5.5 Application to Jet Engines

The FANJETPW engine model discussed in Section 5.2 provides seven outputs that can be potentially measured. There are also ten parameters that would need to be estimated. Therefore, the objective of output selection is to select the smallest set of outputs that can provide full observability to the parameters.

Using the outputs partially shown in Figure 5.2, the first step for output selection was to obtain for each output its sensitivities to individual parameters. For this application, given the seven outputs and ten parameters, a total of seventy output sensitivities were obtained. For illustration purposes, a sample of core speed (N2) output sensitivities are shown in Figure 5.6. While several output sensitivities seem to be unique in shape, indicating the observability of the corresponding parameters by this output, there are also some that are similar, for instance, the output sensitivities with respect to LPT_{eff} and fan_{eff}.

The customary approach to evaluating the similarity between the output sensitivities is the correlation matrix, such as the one shown in Table 5.1 for the output sensitivities in Figure 5.6. It is interesting to note that not only is the similarity between

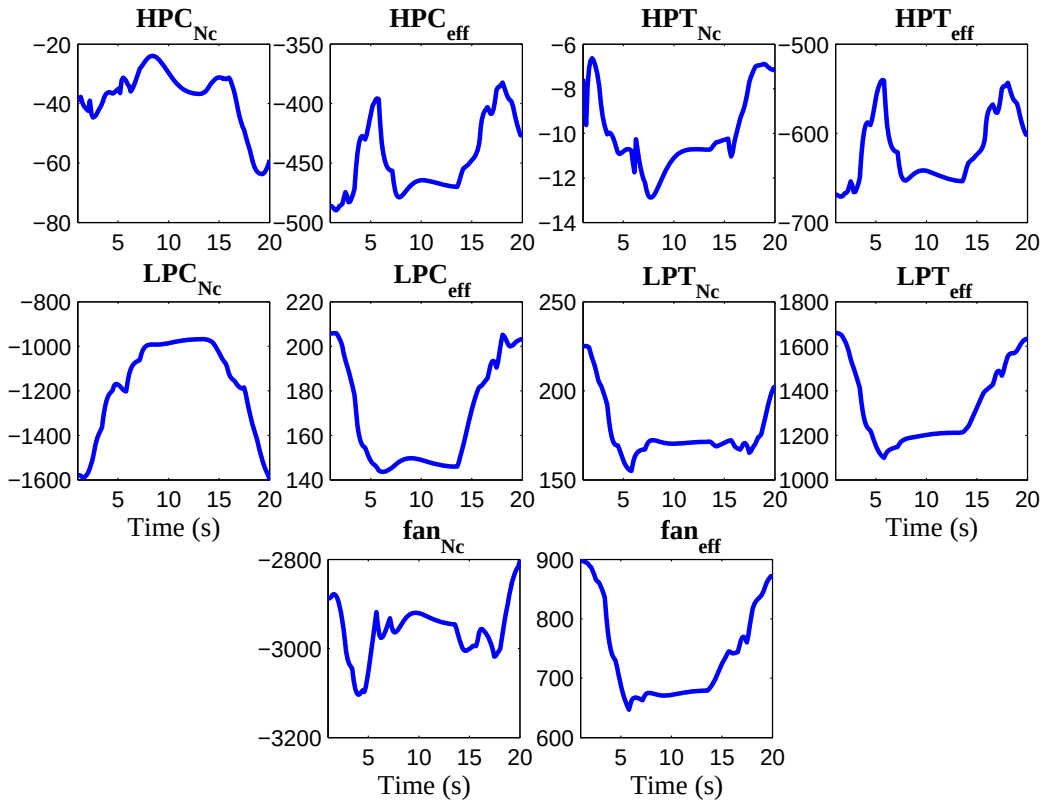

Figure 5.6: The output sensitivities of output N2 with respect to the model parameters.

the output sensitivities of LPT_{eff} and fan_{eff} verified in this matrix ($\rho = 0.9361$), but also are the similarities between several others, shown by boxes in Table 5.1. In fact, upon inspection, it seems like the sensitivity of this output to almost every parameter is highly correlated with some other output sensitivity, except for HPT_{Nc} and LPT_{Nc}. Therefore, according to the correlation coefficients associated with this set of output sensitivities of N2, it appears that this output provides observability to only parameters HPT_{Nc} and LPT_{Nc}. But as it will be shown later, the above identifiability results do not match the conclusions made by parameter signatures, which reveal its identifiability of not only HPT_{Nc} but also LPC_{Nc}, LPC_{eff}, and fan_{Nc} by this output. Two reasons can be cited for the disagreement between the two conclusions: (i) the limited scope of correlation coefficients, as pairwise analysis

tools, for evaluating the linear dependence of the output sensitivities (i.e., rank of the Jacobian matrix) and (ii) the enhanced delineation provided by wavelet transforms to the dissimilarity between the output sensitivities (see Figure 5.5). Of course, the level of linear independence of the output sensitivities can be assessed by the smallest singular value of the Jacobian matrix Φ that encompasses them. But such a measure, such as any other of its kind, can only provide a measure of overall identifiability of the parameters by the output. As it is shown below, the parameter signatures provide instead a one-to-one assessment of observability between the parameters and outputs, hence, they elucidate the level of identifiability of each parameter by each output.

Table 5.1: The correlation coefficients between the output sensitivities in Figure 5.6.

	HPC_{Nc}	HPC_{eff}	HPT_{Nc}	HPT_{eff}	LPC_{Nc}	LPC_{eff}	LPT_{Nc}	LPT_{eff}	fan_{Nc}	fan_{eff}
HPC_{Nc}	1	-0.215	-0.2432	-0.1471	0.6798	-0.117	-0.5666	-0.431	$\boxed{-0.892}$	-0.4665
HPC_{eff}		1	0.3208	$\boxed{0.9943}$	-0.5937	$\boxed{0.8648}$	0.1824	$\boxed{0.8043}$	0.411	0.7229
HPT_{Nc}			1	0.2897	-0.3726	0.5361	0.4076	0.5885	0.2138	0.4921
HPT_{eff}				1	-0.5681	$\boxed{0.8418}$	0.1238	0.7597	0.3538	0.6934
LPC_{Nc}					1	-0.5838	-0.7648	-0.7758	-0.6529	$\boxed{-0.9158}$
LPC_{eff}						1	0.4663	$\boxed{0.9264}$	0.2232	$\boxed{0.8368}$
LPT_{Nc}							1	0.6853	0.4069	0.7836
LPT_{eff}								1	0.4572	$\boxed{0.9361}$
fan_{Nc}									1	0.4501
fan_{eff}										1

The next step in the implementation of this method was extraction of the parameter signatures according to (1.16). For illustration purposes, the parameter signatures obtained from the output sensitivities in Figure 5.6 at the dominance factor of 2 are shown in Figure 5.7. It is interesting to note that contrary to the indications by the correlation coefficients in Table 5.1, it is possible to extract parameter signatures for HPC_{Nc}, HPT_{Nc}, LPC_{Nc}, LPC_{eff}, LPT_{eff}, and fan_{Nc}. If the existence of these parameter signatures could be ascertained under all conditions, then one could conclude that this output provides observability to the corresponding parameters.

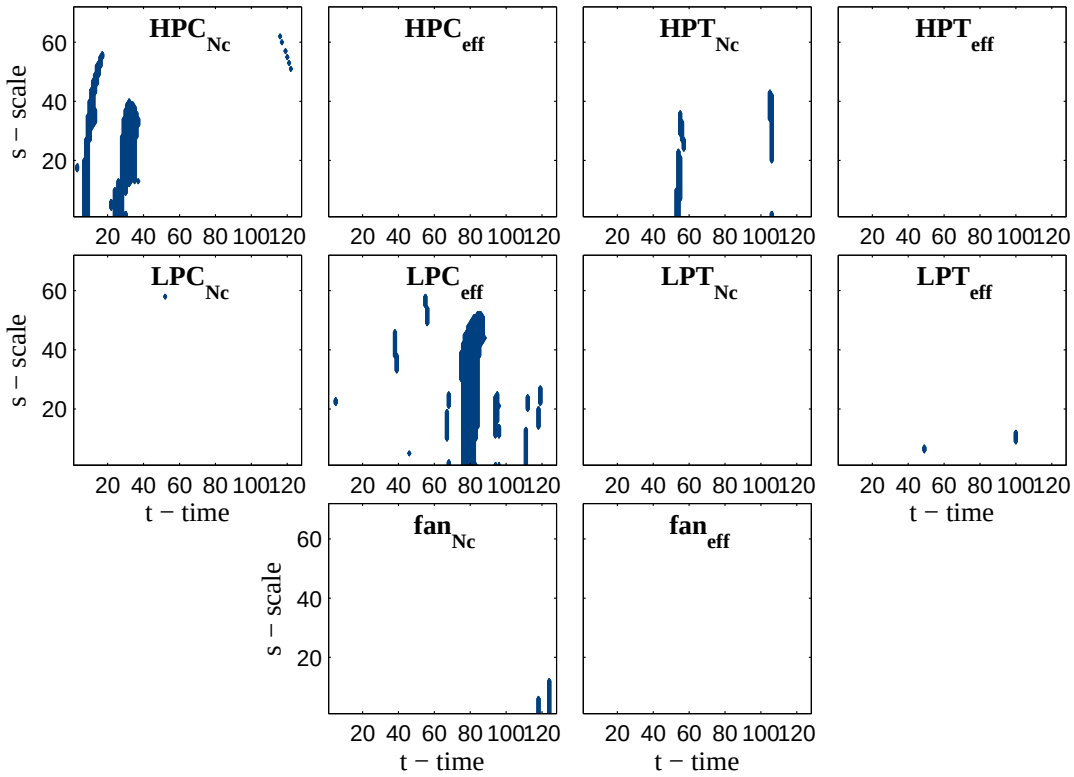

Figure 5.7: The parameter signatures for the output N2 extracted at a dominance factor of 2.

The parameter signatures in Figure 5.7 have two dependencies. One is the dominance factor η_d in (1.16). The other is the nominal parameter vector $\bar{\mathbf{\Theta}}$ at which the output sensitivities in (5.6) are computed. To study the influence of the dominance factor on identifiability results, one can consider the binary matrices in Figure 5.8 representing the identifiability of the engine parameters by the outputs at the dominance factors of 2, 2.5, and 3. A dark block in each matrix indicates the existence of a parameter signature for that output. The results indicate that each of these matrices provides a different verdict for parameter identifiability according to the dominance factor used. For example, referring back to the parameter signatures in Figure 5.7, obtained from output N2 (output 1), the dark blocks in the first row of the left matrix, associated with output 1, indicate the existence of parameter signatures for parameters 1, 3, 5, 6, 8 and 9 at a dominance factor of 2. However, as expected, increasing the

dominance factor, while improving the quality of the parameter signature, will reduce its size (i.e., number of pixels) to the point of its diminishment (e.g., see Figure 5.3). This point is reflected in the other binary matrices in Figure 5.8. For example, for the same output 1 (N2), parameter identifiability is reduced to parameters 1, 3, 6, 8 and 9 for the dominance factor of 2.5, and to parameters 3, 6, 8 and 9 for the dominance factor of 3. Setting a suitable dominance factor that will yield an accurate parameter identifiability assessment is, therefore, an important part of this method. One approach is to perform a search of the highest dominance factor that will yield complete parameter identifiability with all the outputs. This dominance factor can then be used to evaluate the quality of parameter signatures. For instance, if one were to approach a dominance factor of, say, 1.2 in order to achieve full parameter identifiability by the complete output suite, then one can conclude that there is not adequate observability by the outputs given that η_d =1.2 would hardly satisfy the notion of dominance.

The second dependency of the parameter signatures is the nominal parameter vector, $\bar{\mathbf{\Theta}}$, at which the output sensitivities are computed. As such, the robustness of parameter signatures needs to be evaluated across a wide range of nominal parameter vectors. One measure of this robustness is the percentage of times parameter signatures can be extracted over a range of nominal parameter values. As a test of this measure, output sensitivities were produced at 25 random parameter vectors within the range of $\pm 4\%$ of the original vector used for extracting the parameter signatures in Figure 5.7. The percentage of times parameter signatures could be extracted for each output are illustrated in Figure 5.9 and listed in Table 5.2 at the dominance factor of 2.

For demonstration purposes, let us use the parameter signatures obtained at the dominance factor of 2 as the basis for identifiability analysis. According to the results in Table 5.2 there is ample identifiability for some parameters, e.g., parameter 3

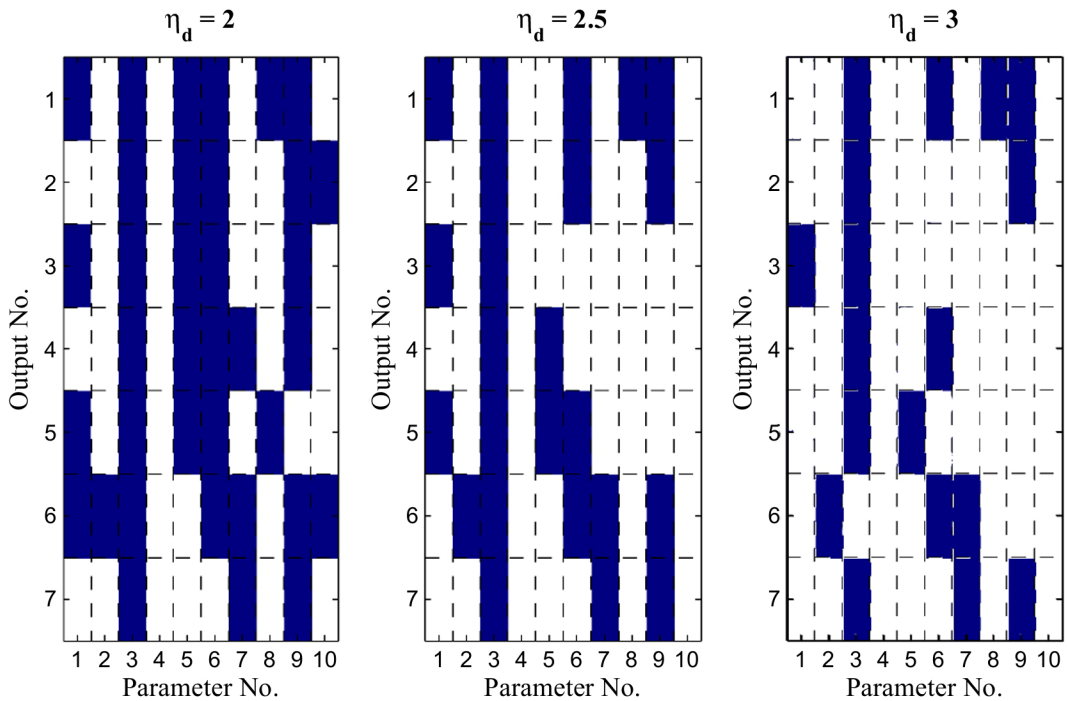

Figure 5.8: The parameter identifiability provided by engine outputs at the dominance factors of 2, 2.5, and 3. Parameters are numbered according to the following order: HPC_{Nc}, HPC_{eff}, HPT_{Nc}, HPT_{eff}, LPC_{Nc}, LPC_{eff}, LPT_{Nc}, LPT_{eff}, fan_{Nc}, and fan_{eff}; the outputs are ordered as: N2, N1, T25, T50, P25, T30, and P30. The dark blocks indicate the presence of parameter signatures.

(HPT_{Nc}). But there are also those that are identifiable through only one output. For instance, parameter 2 (HPC_{eff}) is identifiable by only output 6 (T30). According to this analysis, one can define the outputs that are necessary for identifiability of parameters that are identifiable by only one output. For this engine model, outputs N2, N1, T30, and P30 are necessary for identifiability of LPT_{eff}, HPT_{eff}, HPC_{eff}, and fan_{eff}, respectively. Given that this output suite provides observability to the remaining six parameters as well, it is determined as the critical suite for this transient deck

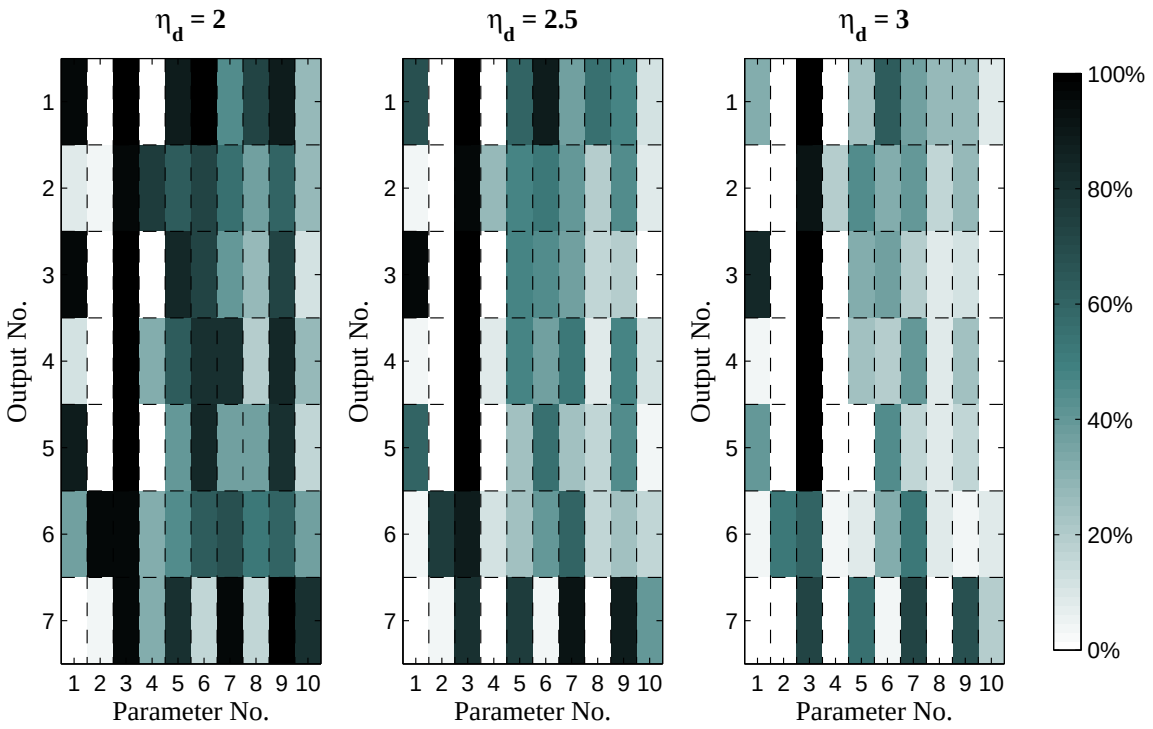

Figure 5.9: Graphical representation of the percentage of parameter signatures that remain present for each measurement across 25 different nominal parameter values using the dominance factors of 2, 2.5 and 3.

5.6 Validation by Parameter Estimation

The criterion for output selection is the ability to estimate all of the parameters by the output suite. It is, therefore, befitting to test the validity of the selected output suite in parameter estimation. We can, of course, use PARSIM for parameter estimation which relies on the parameter signatures for this purpose. However, in order to provide a level of independence between the output selection approach and the parameter estimation method, the selected output suite was tested by nonlinear least-squares (NLS) [70]. This test was performed in separate stages. First, the critical suite of four outputs consisting of outputs N2, N1, T30, and P30 was used for estimation of all ten parameters. The parameter estimates are shown in Figure 5.10. They indicate that the parameter estimates indeed converge to their true values,

Table 5.2: The percentage of parameter signatures extracted for 25 different nominal parameter values.

	HPC_{Nc}	HPC_{eff}	HPT_{Nc}	HPT_{eff}	LPC_{Nc}	LPC_{eff}	LPT_{Nc}	LPT_{eff}	fan_{Nc}	fan_{eff}
N2	96	0	100	0	88	100	44	72	88	28
N1	8	4	96	76	64	72	56	36	60	28
T25	96	0	100	0	84	72	40	28	72	12
T50	12	0	100	32	64	80	80	20	84	28
P25	88	0	100	0	40	84	36	36	80	16
T30	36	96	96	32	44	64	68	52	60	36
P30	0	4	96	32	80	16	96	16	100	80

shown by the dashed line, thus confirming the adequate observability these outputs provide for all the parameters.

Next, the selected outputs were tested by performing parameter estimation with smaller suites of three outputs, each missing one of the allegedly critical outputs. For this test, five sets of initial parameter values within ±4% of the true parameter values were used for the estimation runs. Parameter estimation runs were then performed by NLS with each of the 3-output suites and all five initial parameter values. Of the twenty estimation runs performed in total, most failed after the first iteration due to unacceptable parameter estimates. The parameter estimates from the estimation runs that lasted more than one adaptation iteration are shown in Figure 5.11, with the 'x' indicating simulation failure due to drastically erroneous parameter estimates outside the maps of the simulation model. The results indicate that all parameter estimation tests failed, except for one, thus further confirming the necessity of the four outputs: N2, N1, T30, and P30 for complete parameter identifiability. The one successful estimation run is consistent with the percentages shown in Table 5.2.

In the last stage, the robustness of the estimation solution with the suite of 4 critical outputs was tested with the same initial parameter values as in Figure 5.11. The parameter estimates from these estimation runs, shown in Figure 5.12, again confirm the earlier results in Figure 5.10 that the 4-output suite identified by the parameter signatures is indeed sufficient for parameter estimation of all ten parameters in the

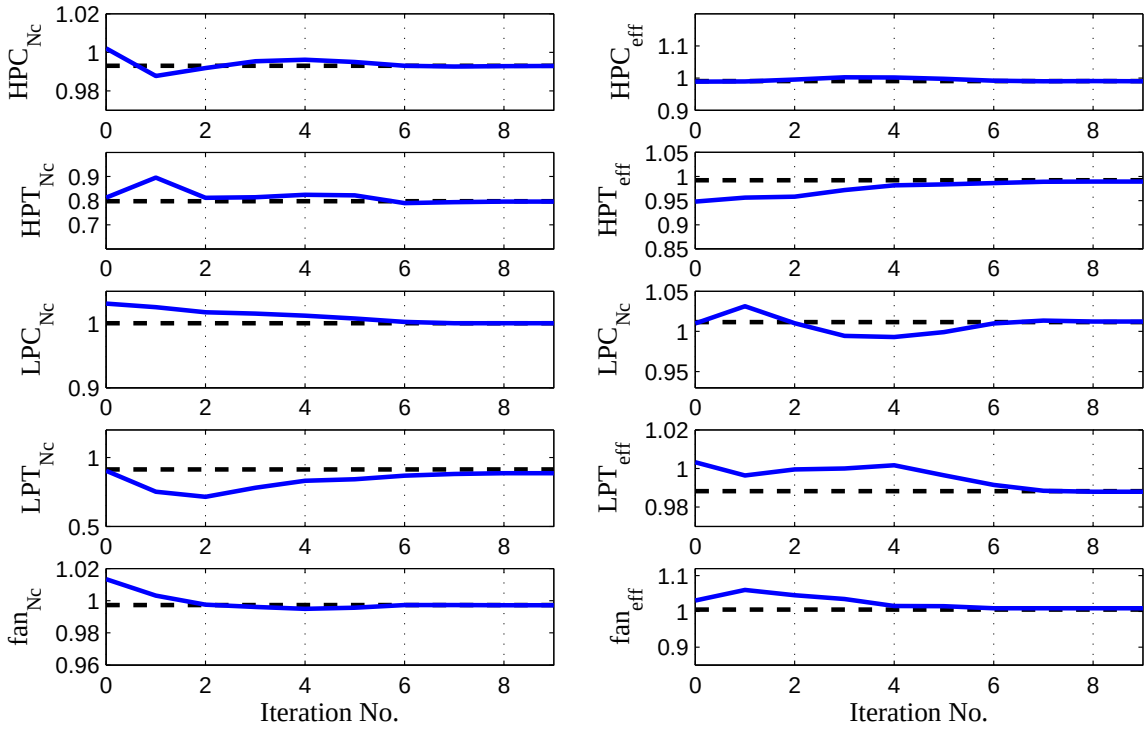

Figure 5.10: Paramter estimation results by NLS using the four outputs N2, N1, T30, and P30. For reference, the true parameter values are shown by the dashed line.

engine model. It should be noted that the small error in the estimates of HPT_{eff} and LPT_{Nc} at the end of one of the estimation runs is likely due to inadequate adaptation, which should be corrected with further iterations.

5.7 Discussion

The results obtained from the application of the proposed sensor/measurement selection method to the engine model demonstrate the advantage of the proposed method over the traditional time-based measures. Although no limitations are foreseen for practical application of this method, it behooves us to consider some of the issues that may arise in practice.

- *Accuracy of the Model.* The method of measurement selection introduced in this paper, such as like any other counterpart method, is model-based. As such, the

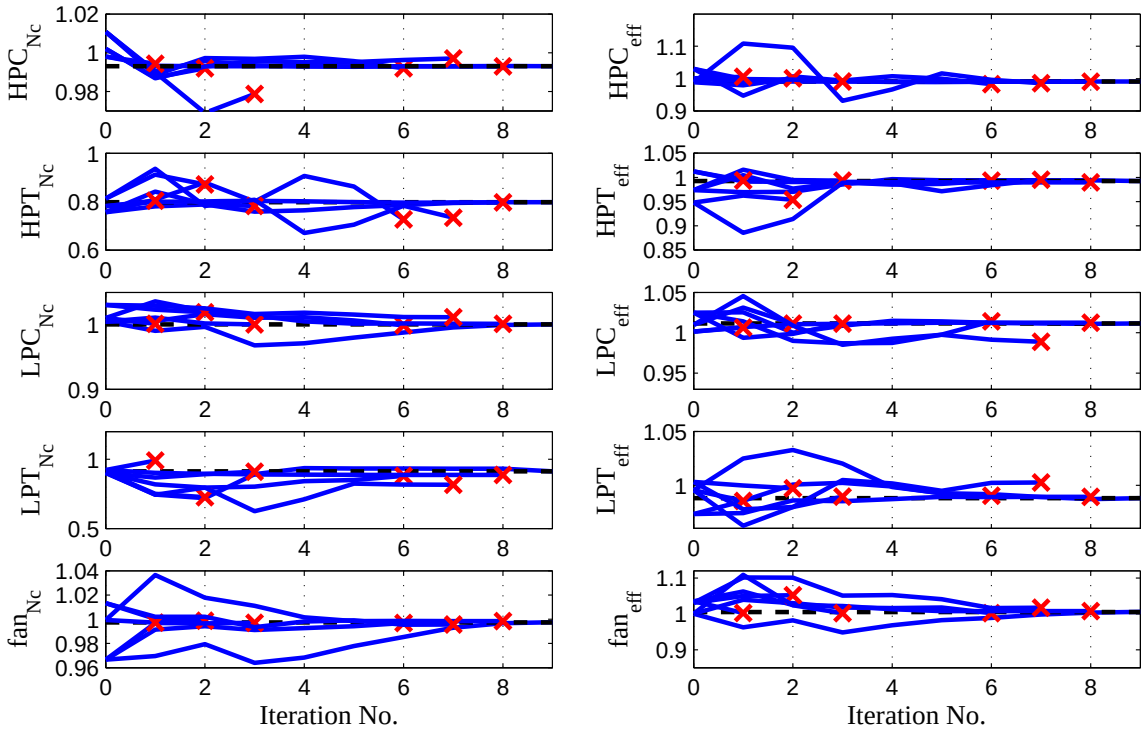

Figure 5.11: Parameter estimation results by NLS using output suites of three outputs, each missing one of the outputs N2, N1, T30 and P30 that were deemed necessary for parameter estimation. The 'x' at the end of runs denotes a failed simulation due to a drastically erroneous parameter estimate.

accuracy of the model is of concern, as is the fidelity of its first order approximation by (5.9). In order to account for some of this uncertainty, we adopted an averaging strategy of considering the parameter signatures at a multitude of model parameters. However, this strategy does not compensate for modeling error and the question remains as to what extent the inevitable inaccuracies of the model would influence the output selection results. Although the answer to this question is not trivial and remains to be investigated, our expectation is that modeling inaccuracies should not be detrimental to the analysis so long as the model correctly represents the qualitative sensitivity of the outputs relative to the model parameters. Related to modeling inaccuracy, is the completeness

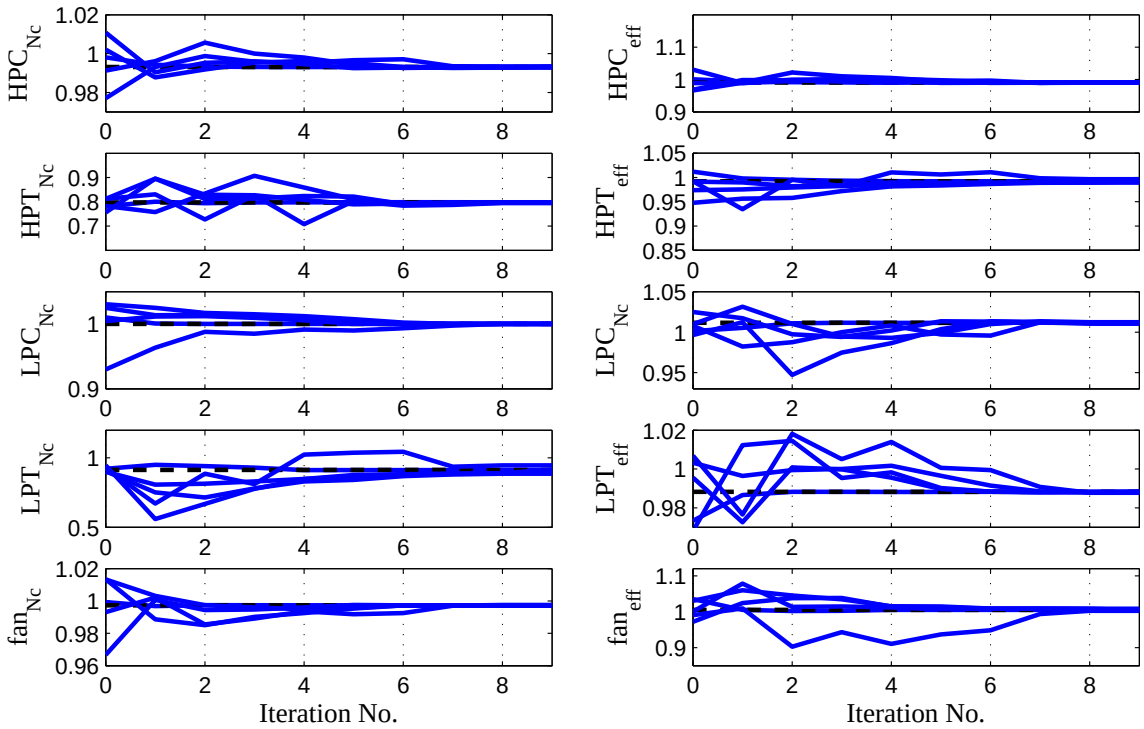

Figure 5.12: Several sets of parameter estimation results by NLS using the suite of 4 outputs: N2, N1, T30, and P30 with the starting parameter values that already failed in Figure 5.11. The true parameter values as before are shown by the dashed line.

of the model in representing the various dynamics, such as actuator and sensor dynamics. Again, in advance of a methodical study to address this issue, it can be speculated that these dynamics should not affect the output sensitivities, so long as they are known and that no model parameters are associated with these dynamics.

- *Noise.* A common concern in practical applications is measurement noise. To the extent that the proposed method, such as its counterparts is model-based, measurement noise does not factor into the analysis, even though it is of concern in reality. As a recourse, one could consider evaluating the effect of noise on the parameter signatures by adding noise to the simulated outputs. Although not

investigated yet, it can be speculated that the addition of noise will likely reduce the robustness of parameter signatures beyond those observed in its absence.

- *Computation time.* The proposed method requires transformation to the time-scale domain via continuous wavelet transforms. It also relies on averaging the parameter signatures at different model parameters to account for nonlinearity. These steps would add to the computation effort, but in the present study the computation time associated with parameter signature extraction/output selection constitutes only a small fraction of the time of engine simulation. Given the batch nature of output selection and its inevitable reliance on engine simulation, regardless of the method used, it is doubtful that the added cost of computation associated with the proposed method will be a deterrent in its application.

- *Input conditions.* The method of output/measurement selection introduced here is a posterior identifiability method [74] so far as it relies on the simulated outputs of the model. As such, this method also depends on the input conditions that generate the transients and produce the output sensitivities. In this chapter, we assume the input conditions to be dictated by the test procedure. As such, we do not consider the input as a control variable, despite its influence on the parameter signatures. In practice, however, the inputs used to generate the engine output transients can be designed ('optimized'), within the engine constraints, so as to maximize the identifiability of the parameters by the output transients [28, 54, 23].

- *Selection of Dominance Factor* As is clear from the results; e.g., Fig. 5.8, a crucial factor in the proposed method is the dominance factor η_d in (1.16). For the results in this chapter, we arbitrarily chose $\eta_d = 2$ as being adequate. However, it is possible to adopt a selection strategy whereby the appropriate dominance

factor is determined according to the quality of the parameter signatures. As discussed already and shown in Fig. 5.4, the quality of parameter signatures can be defined by the consistency of the parameter error estimates, as formulated in (3.6).

CHAPTER 6

CONCLUSIONS

The contribution of this work is to capitalize on the representation of shape attributes of time series by continuous wavelet transforms as well as their capacity to delineate minute differences between time series. We have taken advantage of the enhanced delineation of the surfaces created by wavelet transforms to identify regions in the time-scale domain wherein the wavelet coefficients of one time series dominates the others. These regions which are called parameter signatures in this work have been instrumental in a variety of system identification scenarios.

In model validation, we have demonstrated the utility of continuous wavelet transforms in representing the shape attributes of time series. The availability of the signal's slopes and/or its rate of slope changes at different locations of the time-scale plane provides the framework for comparing the shapes of model outputs with their measured counterparts at different times and frequencies. This approach also allows for the capacity to consider different measures of comparison. The benefit of comparing the slopes of outputs via Gauss WT by three different measures of closeness based on image distances was demonstrated and the results illustrate the promise of this approach to model validation.

In parameter estimation, the availability of parameter signatures allows decoupling of the first-order prediction error equation into multiple single-parameter equations. Our research demonstrates that this expansion into the time-scale domain of the prediction error can be a viable alternative to the traditionally exercised compaction of the prediction error as a scalar. Our anecdotal observations indicate that for convex

error surfaces, PARSIM's results are comparable to those by NLS, though PARSIM is not as agile with low-quality parameter signatures extracted from gradually sloped error surfaces. On the other hand, PARSIM has been observed to evade local minima entrapments on non-convex surfaces as discussed in Section 3.4. The results presented in this section for the Van der Pol oscillator indicate that PARSIM, because of its independence from the gradient of the contour, can lead the estimates to their correct values (the bottom of the bowl) whereas the GN method misses them due to the unfavorable location of the starting points.

PARSIM also provides the capacity to perform direct noise compensation in the time-scale domain. The common approach for improving the precision of parameter estimates with noisy measurements is to filter the measurements, and among such filters particularly noteworthy is one which transforms the signal to the time-scale domain, reduces the high-frequency noise by thresholding the wavelet coefficients in the lower scales (higher frequencies), and then reconstructs the wavelet coefficients back in the time domain. PARSIM, due to its capacity to perform parameter estimation in the time-scale domain, obviates this need to reconstruct the signal in the time domain. This utility of PARSIM has been demonstrated by incorporating confidence factors to account for the distortion of the prediction error when estimating the parameter errors. Such confidence factors, which represent the estimates of noise distortion at different pixels of the time-scale plane, are then incorporated as weights to yield the biased parameter estimates. The results from this approach were shown to significantly improve the precision of the parameters estimates.

Parameter signatures have also been shown to be applicable to measurement selection. Traditionally, the concept of measurement selection is directly linked to parameter identifiability analysis which has been investigated extensively by the research community. The key to identifiability analysis is the uniqueness of the corresponding output sensitivities; i.e., the columns of the Jacobian matrix. Since the existence of

parameter signatures is also contingent upon the uniqueness of output sensitivities, the parameter signatures can be readily used as a comprehensive measure of identifiability analysis in order to select outputs and/or input profiles. In this research, the mere existence of parameter signatures has been used as evidence of parameter identifiability by the output, but the robustness of measurement selection and its methodical formulation can be significantly improved by using a uniform index of parameter signature quality. We have demonstrated the utility of this method in output selection of aircraft engines and expect it to be widely applicable to other systems such as civil structures and chemical plants.

This work describes the creation of a new approach to system identification and the several methods that have so far resulted from it, but much more can be done towards the development of this approach. As is expected of any new approach, there are many unanswered questions ranging from theoretical to practical that need to be addressed. The theoretical questions stem from the uniqueness of the time-scale domain, in providing added transparency to parameter estimation, to the uniqueness of the Newton-type method as the platform to implement the parameter error estimates for parameter estimation. Some of these questions are discussed in the following sections.

6.1 Model Validation

The distance measures considered in this research for model validation are only three of the similarity measures that can be implemented in the time-scale domain. Other potential measures include the time warping distance [7] that can characterize time delays between the measured and model outputs. But there can also be image distances that would represent the other image aspects considered by human experts in visual inspection. A reliable validation metric for dynamic models would be ideally based on more than just the similarity of output pairs as well. Toward this end, one

could consider combining various image distances and time-based measures into a composite validation index to provide a comprehensive measure of model closeness across different operating conditions. This study only points to the potential of distances as model validation metrics and leaves the development of more customized distance measures to future studies.

6.2 Parameter Estimation

PARSIM was developed primarily as a method of parameter estimation for simulation model tuning. Although this was extensively analyzed, various topics remain for the subject of future studies.

6.2.1 Convergence

By comparing PARSIM's single-parameter estimation of the model parameters with a single-parameter Newton-Raphson method, it becomes clear that PARSIM implements this method for individual parameters at the individual pixels of the corresponding parameter signatures. PARSIM's convergence behavior, therefore, parallels Newton-Raphson's except for the nuances of operation in the time-scale domain pertaining to: (i) the shape attribute(s) of the output sensitivities represented by the wavelet transform(s), (ii) the uniqueness of output sensitivities which correspond to the existence and quality of the parameter signatures, and (iii) the contour of the error surface. Among the above factors, the uniqueness of the parameter signatures is synonymous with the Jacobian being full-ranked, which is also important to NLS, as is the fidelity of the first-order approximation of the prediction error. As such, the conditions for PARSIM's convergence relate to those that influence (i) the existence and quality of parameter signatures, (ii) the consistency of the parameter error estimates obtained from the parameter signatures, and (iii) convergence characteristics of these single-parameter estimates to their true values. In order to determine the

strengths as well as the limitations of PARSIM, the convergence characteristics of the estimation strategy need to be studied, ranging from the mechanism for error minimization to the interaction between the estimated parameters. These convergence characteristics remain an integral topic for future studies.

6.2.2 Multi-output Parameter Estimation

Another point of interest is the utility of PARSIM in multi-output cases. Traditionally in multi-objective optimization, the various outputs are weighted and incorporated into a cost function, similar to the single-output case. Although not discusses in this thesis, a preliminary investigation into parameter estimation of multi-output simulation models was conducted with mixed results. In this investigation each of the simulation outputs were given equal weight in the objective cost function and, although in most cases the parameters were effectively estimated, conflicting estimates for the model parameters by various outputs made the estimation laborious and unstable. We believe the transparency available in the time-scale plane ought to provide more sophisticated ways of integrating the outputs. Toward this objective, issues associated with identifiability of parameters via individual outputs will need to be considered as well.

6.2.3 Parameter Signature Quality

One of the unique aspects of PARSIM is the transparency it offers to parameter estimation through the quality of parameter signatures. One benefit of being able to assess the quality of parameter signatures will be to evaluate the reliability of the parameter error estimate. In this research, the parameter estimates produced by PARSIM were improved through quantifying the quality of the parameter signatures for adaptation step size selection. For this, the quality of the parameter signatures was ascertained though the variance of the estimates among all pixels of a parameter signature. Although this approach effectively improved convergence, it is thought

that through developing a comprehensive quantification of the quality of a parameter signatures, a more detailed picture of a simulation model can be developed and can further improve the method. It is also expected that through quantifying the quality of parameter signatures weights could be assigned to individual parameter error estimates obtained from individual outputs, in multi-output cases, and/or from individual wavelet transforms of the same output (when multiple wavelet transforms are used to represent different shape attributes). It is also expected that the development of a comprehensive parameter signature quality measure will have significant implications to measurement and input selection for evaluating parameter identifiability.

6.2.4 Noise Compensation

In the approach taken in this thesis for noise compensation, the approximation of noise was shown to be less accurate in the higher scale regions. This is predominantly due to the higher frequency content of noise. Although it can be seen that noise affects the higher scales, the approximation technique developed in this research does not effectively approximate the noise in this region. Various approaches were investigated to improve the approximation in these regions, however the results were mixed and the most stable approach was the one presented in this thesis. We believe that through an investigation into the affects of noise on the higher scales, we may be able to improve on an already effective approach to parameter estimation of noisy systems.

6.3 Filter Design

Selection of the wavelet transform is a key component to the application of PARSIM to any system. In this thesis, the wavelet was selected according to its effectiveness in parameter estimation and did not utilize a priori knowledge of the system. We are attracted to continuous wavelet transforms because of their differential capacity, however we feel that this effect is not unique and can be replicated by customized fil-

ters [45]. Developing wavelets and/or filters that represent specific aspects of the time signal's shape, beyond those provided by its derivatives, is another exciting aspect of this research. To this end, one can consider designing customized filters that would also provide filtering capacity for noise suppression. Although, there has not been an extensive investigation into the feasibility of designing model or noise-specific filters, the prospect of designing such filters remains an intriguing topic for future studies.

BIBLIOGRAPHY

[1] Akaike, H. A new look at the statistical model identification. *IEEE Trans. on Automatic Control 19*, 6 (1974), 716–723.

[2] Astrom, K., and Eykhoff, P. System identification - a survey. *Automatica 7* (1971), 123–162.

[3] Bagajewicz, M. J. Design and retrofit of sensor networks in process plants. *AIChE 43*, 9 (1997), 2300–2306.

[4] Bendat, J. S., and Piersol, A. G. In *Random Data Analysis and Measurement Procedures, 3rd Edition* (2000), Wiley Intersciences.

[5] Brotherton, T., Volponi A. Luppold R., and Simon, D.L. estorm: Enhanced self tuning on-board real- time engine model. In *Proceedings of the 2003 IEEE Aerospace Conference* (2003).

[6] Carson, E. R., Cobelli C., and Finkelstein, L. In *The Mathematical Modeling of Metabolic and Endocrine Systems* (1983), John Wiley and Sons.

[7] Chan, K. P., Fu A. Yu C. Haar wavelets for efficient similarity search of time-series: with and without time warping. *IEEE Trans Knowledge Data Eng 15*, 3 (2003), 686–705.

[8] Chen, S. L., and Ho, K. C. Identification of nonlinear systems by haar wavelet. In *Proc. of IMECE04* (2004), Paper No. IMECE2004-62417.

[9] Chipman, H. A., Kolaczyk E. D. McCulloch R. E. Adaptive bayesian wavelet shrinkage. *J. of the American Statistical Association 92* (1997), 1413–1421.

[10] Chmielewski, D. J., Palmer T., and Manousiouthakis, V. On the theory of optimal sensor placement. *AIChE 48*, 5 (2002), 1001–1012.

[11] Cole, R., Shasha D., and Zhao, X. Fast window correlations over uncooperative time series. In *Proc. of KDD 2005* (Chicago, Illinois, USA, August 2005).

[12] Comte, F., and Rozenholc, Y. A new algorithm for fixed design regression and denoising. *Ann. Inst. Statist. Math. 56*, 3 (2004), 449–473.

[13] Danai, K., and McCusker, J. R. Parameter estimation by parameter signature isolation in the time-scale domain. *ASME J. of Dynamic Systems, Measurement and Control 131*, 4 (2009).

[14] Deuflhard, P. In *Newton Methods for Nonlinear Problems* (2004), Springer.

[15] Doel, D. L. Temper - a gas-path analysis tool for commercial jet engines. *ASME J. of Eng. for Gas Turbines and Power 116*, 1 (1994), 82–89.

[16] Donoho, D. L., and Johnstone, I. M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika 81*, 3 (1994), 425–455.

[17] Donoho, D. L., and Johnstone, I. M. Adapting to unknown smoothness via wavelet shrinkage. *J. of the American Statistical Association 90*, 432 (1995), 1200–1224.

[18] Donoho, D. L., Johnstone I. M., Kerkyacharian, D., and Picard, D. Wavelet shrinkage, asymptopia? *J. of the Royal Statistical Society, Series B (Methodological) 57*, 2 (1995), 301–369.

[19] Dunstan, W. J., and Bitmead, R. R. Model confidence for nonlinear systems. In *Proc. of the 15th IFAC Triennial World Congress* (Barcelona, Spain, 2002).

[20] Dunstan, W. J., and Bitmead, R. R. Nonlinear model validation using multiple experiments. In *Proc. of the 42nd IEEE Conf. on Decision and Control* (Maui, Hawaii, USA, 2003).

[21] Ehlers, J., and Diop, A. Sensor selection and state estimation for wind turbine controls. In *Proc. of the 45th AIAA Aerospace Science Meeting and Exhibit* (Reno. Mevada, 2007).

[22] Espana, M. D., and Gilyard, G. B. On the estimation algorithm used in adaptive performance optimization of turbofan engines. In *NASA TM 4551* (1993).

[23] Fedorov, V. V., and Khabarov, V. Duality of optimal designs for model discrimination and parameter estimation. *Biometrika 73*, 1 (1986), 183–190.

[24] Fletcher, R. In *Practical Methods of Optimization, 2nd Ed* (1987), Wiley.

[25] Frank, P. M. In *Introduction to System Sensitivity Theory* (New York, 1978), Academic Press.

[26] Ghanem, R., and Romeo, F. A wavelet-based approach for model and parameter identification of non-linearsystems. *Int'l Journal of Nonlinear Mechanics 36* (2001), 835–859.

[27] Goldberg, D. E. In *Genetic Algorithms* (1989), Addison Wesley.

[28] Goodwin, G. C., and Payne, R. L. In *Dynamic System Identification: Experiment Design and Data Analysis* (1977), Academic Press.

[29] Gupta S., Ray A., Sarkar S., and Yasar, M. Fault detection and isolation in aircraft gas turbine engines. part 1: Underlying concept. *Proc. of Institute of Mechanical Engineers, Part G: J. of Aeropace Engineering 22*, 10.1243/09544100JAERO311 (2008).

[30] Gupta S., Ray A., Sarkar S., and Yasar, M. Fault detection and isolation in aircraft gas turbine engines. part 2: Validation on a simulation test bed. *Proc. of Institute of Mechanical Engineers, Part G: J. of Aeropace Engineering 222*, 10.1243/09544100JAERO312 (2008).

[31] Gustafsson, F. Determining the initial states in forward-backward filtering. *IEEE Transactions on Signal Processing 44*, 4 (April 1996), 988–992.

[32] Hertz, J., Krogh A., and Palmer, R. G. In *Introduction to the Theory of Neural Computation* (1991), Addison Wesley.

[33] Jackson, J. E. In *A User's Guide to Principle Components* (1991), John Wiley & Sons, Inc.

[34] Kennedy, M. P. Abc - adventures in bifurication & chaos: A program for studying chaos. *J. Franklin Inst. 331B*, 6 (1994), 631–658.

[35] Kerr, L. J., Nemec T. S., and Gallops, G. W. Real- time estimation of gas turbine engine damage using a control based kalman filter algorithm. In *ASME Paper 91-GT-216* (1991).

[36] Kobayashi, I., Simon D. L., and Litt, J. S. Application of a constant gain extended kalman filter for in-flight estimation of aircraft engine performance parameters. In *Army Research Laboratory Report: ARL-TR-2955* (2005).

[37] Koehler, A. B., and Murphree, E. S. A comparison of the akaike and schwarz criteria for selecting model order. *Applied Statistics 37*, 2 (1988), 187–195.

[38] Kumar, S., and Seinfeld, J. H. Optimal location of measurements in tubular reactors. *Chemical Engineering Science 33* (1978), 1507–1516.

[39] Leonard, O., Borguet S., and Dewallef, P. Adaptive estimation algorithm for aircraft engine performance monitoring. *J. of Propulsion and Power 24*, 4 (2008), 763–769.

[40] Li, Y. G. A gas turbine diagnostic approach with transient measurements. *Proc. of the Institute of Mechanical Engineers, Part A: J. of Power and Energy 217*, 2 (2003), 169–117.

[41] Ljung, L. In *System Identification, 2d Ed.* (1999), Prentice Hall.

[42] Ljung, L., and Glad, T. On global identifiability for arbitrary model parametrizations. *Automatica 30*, 2 (1994), 265–275.

[43] Loh, A.-P., Annaswamy A. M., and Skantze, F. P. Adaptation in the presence of a general nonlinear parameterization: An error model approach. *IEEE Trans. on Automatic Control 44*, 9 (1999), 1634–1652.

[44] Luppold, R. H., Roman J. R. Gallops G. W., and Kerr, L. J. Estimatingin-flight engine performance variations using kalman filter concepts. In *AIAA Paper, AIAA-89-2584* (1989).

[45] Mallat, S. In *A Wavelet Tour of Signal Processing, 2nd Ed.* (1999), Academic Press.

[46] Mallat, S., and Hwang, W. L. Singularity detection and processing with wavelets. *IEEE Trans. on Information Theory 38*, 2 (1992), 617–643.

[47] Marr, D. *Vision*. W. H. Freeman and Co., San Francisco, CA, USA, 1982.

[48] Mathioudakis, K., and Kamboukos, Ph. Assessment of the effectiveness gas path diagnostic schemes,. *ASME J. of Eng. for Gas Turbine and Power 128* (2006), 57–63.

[49] Mathioudakis, K., Kamboukos Ph. Stamatis A. Gas turbine component fault detection from a limited number of measurements. In *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy* (2004), pp. 609–618.

[50] McCusker, J.R., Currier T., and Danai, K. Improved parameter estimation by noise compensation in the time-scale domain. *Signal Processing 91*, 1 (2011), 72–84.

[51] McCusker, J.R., Danai K., and Kazmer, D.O. Validation of dynamic models in the time-scale domain. *ASME J. of Dynamic Systems, Measurement and Control 132*, 6 (2010).

[52] McCusker, J.R., and Danai, K. Integrating parameter estimation solutions from the time and time- scale domains. In *Proc. Of 2009 Dynamic Systems and Control Conference* (Hollywood, California, October 2009), ASME.

[53] McCusker, J.R., McKinley M.G., and Danai, K. Iterative controller tuning in the time-scale domain. In *Proc. Of 2009 Dynamic Systems and Control Conference* (Hollywood, California, October 2009), ASME.

[54] Mehra, R. K. Optimal input signals for parameter estimation in dynamic systems - survey and new results. *IEEE Trans. Automatic Control 19*, 6 (1974), 753–768.

[55] Mehra, R. K. Optimization of measurement schedules and sensor designs for linear dynamic systems. *IEEE Trans. Autom. Control 21*, 1 (1976), 55–64.

[56] Merrington, G. L. Fault diagnosis of gas turbine engines from transient data. *ASME J. of Eng. for Gas Turbine and Power 111* (1989), 237–243.

[57] Merrington, G. L. Fault diagnosis in gas turbine using a model-based technique. *ASME J. of Eng. for Gas Turbine and Power 116* (1994), 374–380.

[58] Muske, K. R., and Georgakis, C. Optimal measurement system design for chemical processes. *AIChE 49*, 6 (2003), 1488–1494.

[59] Narendra, K. S., and Annaswamy, A. M. In *Stable Adaptive Systems* (1989), Prentice Hall.

[60] Oberkampf, W. L., and Barone, M. F. Measures of agreement between computation and experiment: validation metrics. *Journal of Computational Physics 217*, 1 (September 2006), 5–36.

[61] P., Muller., and Webber, H. Analysis and optimization of certain qualities of controllability and observability of linear dynamic systems. *Automatica 8* (1972), 237–246.

[62] Papadimitriou, C., Beck J. L., and Au, S.-K. Entropy-based optimal sensor location for structural model updating. *J. of Vibration and Control 6* (2000), 781–799.

[63] Parlitz, U. Estimating model parameters from time-series by auto-synchronization. *Physical Review Letters 76*, 8 (1996), 1232–1235.

[64] Popper, K. In *The Logic of Scientific Discovery, Volume 1, Fourth Ed.* (Great Portland Street, London W1., 1959), Hutchinson & Co. LTD, pp. 178–202.

[65] Popper, K. In *The Myth of the Framework, Volume 1* (29 West 35th Street, New York, NY 10001., 1994), Routledge.

[66] Provost, M. J. Observability analysis for successful diagnosis of gas turbine faults. In *Von Karman Institute Lecture Series* (January 2003), VKI LS.

[67] Rubinstein, R. Y. In *Monte Carlo Optimization, Simulation, and Sensitivity of Queueing Networks* (1986), Wiley.

[68] Sastry, S., and Bodson, M. In *Adaptive Control* (1989), Prentice Hall.

[69] Schetzen, M. In *The Volterra and Wiener Theories of Nonlinear Systems* (1980), Wiley.

[70] Seber, G. A. F., and Wild, C. J. In *Nonlinear Regression* (1989), John Wiley and Sons, Inc.

[71] Shi, Z. Y., Law S. S., and Zhang, L. M. Optimum sensor placement for structural damage detection. *J. of Engineering Mechanics* (2000), 1173–1179.

[72] Simon, D. L., Gang S. Hunter G. W. Guo T.-H., and Semega, K. J. Sensor needs for control and health management of intelligent aircraft engines. In *Proc. of Turbo Expo 2004* (2004).

[73] Soderstrom, T., and Stoica, P. In *System Identification* (1989), Prentice Hall.

[74] Thomaseth, K., and Cobelli, C. Parameter information content during model identification experiments. In *Proc. of IFAC Symposium on Modeling and Control in Biomedical Systems* (Warwick, United Kingdom, 1997).

[75] Thomaseth, K., and Cobelli, C. Generalized sensitivity functions in physiological system identification. *Annals of Biomedical Engineering 27* (1999), 607–616.

[76] Udwalia, F. E. Methodology for optimum sensor locations for parameter identification in dynamic systems. *J. of Eng. Mechanics 120*, 2 (1994), 368–390.

[77] Urban, L. A., and Volponi, A. J. Mathematical methods of relative engine performance diagnostics. *SAE Transactions, Journal of Aerospace 101* (1992), 2025–2050.

[78] Walker, J. Fourier analysis and wavelet analysis. *Notices of AMS 44*, 6 (1997), 658–670.

[79] Walter, E., and Pronzato, L. On the identifiability and distinguishability of nonlinear parametric models. *Mathematics and Computers in Simulation 42* (1996), 125–134.

[80] Wang, N., Dayawansa W. P., and Martin, C. F. Van der pol oscillator networks. In *Proc. of the 38th Conference on Decision and Control* (Phoenix, Arizona, USA, 1999), pp. 393–398.

[81] Wang, L., Zhang Y., and Feng, J. On the euclidean distance of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence 27*, 8 (2005), 1334–1339.

[82] Wouwer, A. V., Point N. Porteman S., and Remy, M. An approach to the selection of optimal sensor locations in distributed parameter systems. *J. Process Control, 10* (2000), 291–300.

[83] Xia, Y., and Hao, H. Measurement selection for vibration-based structural damage identification. *J. of Sound and Vibration 236*, 1 (2000), 89–104.