Chapter 25
Multiscale Filtering and Applications to Chemical and Biological Systems

Mohamed N. Nounou
Texas A&M University at Qatar, Qatar

Hazem N. Nounou
Texas A&M University at Qatar, Qatar

Muddu Madakyaru
Texas A&M University at Qatar, Qatar

ABSTRACT

Measured process data are a valuable source of information about the processes they are collected from. Unfortunately, measurements are usually contaminated with errors that mask the important features in the data and degrade the quality of any related operation. Wavelet-based multiscale filtering is known to provide effective noise-feature separation. Here, the effectiveness of multiscale filtering over conventional low pass filters is illustrated though their application to chemical and biological systems. For biological systems, various online and batch multiscale filtering techniques are used to enhance the quality of metabolic and copy number data. Dynamic metabolic data are usually used to develop genetic regulatory network models that can describe the interactions among different genes inside the cell in order to design intervention techniques to cure/manage certain diseases. Copy number data, however, are usually used in the diagnosis of diseases by determining the locations and extent of variations in DNA sequences. Two case studies are presented, one involving simulated metabolic data and the other using real copy number data. For chemical processes it is shown that multiscale filtering can greatly enhance the prediction accuracy of inferential models, which are commonly used to estimate key process variables that are hard to measure. In this chapter, we present a multiscale inferential modeling technique that integrates the advantages of latent variable regression methods with the advantages of multiscale filtering, and is called Integrated Multiscale Latent Variable Regression (IMSLVR). IMSLVR performance is illustrated via a case study using synthetic data and another using simulated distillation column data.

DOI: 10.4018/978-1-4666-4450-2.ch025

INTRODUCTION

With the advancements in computing and sensing technologies, large amounts of data are continuously collected from various engineering systems or processes. These data are a rich source of information about the systems they are collected from. Unfortunately, real data are usually contaminated with errors (or noise) that mask the important features in the data and affect their usefulness in practice. Therefore, measured process data need to be filtered to enhance their quality and usefulness. For example. In biological systems, measured genomic data are used to construct genetic regulatory network models that describe the interactions among different genes within the cells (Jong, 2002; Chou et al., 2006; Gonzalez et al., 2007; Kutalik et al., 2007; Wang et al.,2010; Meskin et al., 2011b). These models are used not only to understand and predict the behavior of the biological system, but also to design intervention techniques that can be ultimately used to manage and cure major phenotypes (Ervadi-Radhakrishnan & Voit, 2005; Meskin et al., 2011a). The presence of measurement noise in the data, however, degrades the accuracy of estimated genetic regulatory network models and the effectiveness of any intervention technique in which these model are used (Kutalik et al., 2007; Wang et al., 2010). Also, Copy Number (CN) data are experimental biological data that are usually used in the diagnosis of diseases by determining the locations and extent of aberrations in DNA sequences. CN data are usually very noisy, which makes it difficult to define the abnormal regions in the DNA (Alqallaf & Tewfik, 2007). Thus, it is important to filter biological data to improve their accuracy and the effectiveness of the applications in which they are used. In chemical processes, on the other hand, measured process data are usually used to develop empirical models, especially when fundamental models are difficult to obtain. An important example is inferential models, which are used to estimate key process variables, which are difficult to measure online from other variables that are easier to measure (Frank & Friedman, 1993; Stone & Brooks, 1990; Kano et al., 2000; Wold, 1982). Unfortunately, the measured data used in estimating empirical models are usually contaminated with errors that degrade the quality of the models and their ability to predict the process behavior (Bakshi, 1999; Palavajjhala et al., 1996; Nounou & Nounou, 2005). Filtering these data will not only enhance the accuracy of estimated models, but also improve any operation (e.g., control, monitoring, etc.) in which these models are used.

In general, filtering techniques can be classified into three main categories: filtering with a model, filtering with an empirical model, and filtering without a model. Model-based filtering techniques minimize the error between the measured and filtered data while requiring the filtered data to satisfy the available model. Methods in this category include Kalman filtering (Sorenson, 1985), Moving Horizon Estimation, and particle filtering (Rawlings & Bakshi, 2006). Of course, the quality of the filtered data depends on the accuracy of the models used. In practice, however, models are not usually available a priori. In the absence of a fundamental model and in the case of multivariate filtering, an empirical model that is extracted from the relationship between the measured variables can also be used in data filtering. Methods in this category include Principal Component Analysis (PCA) (Kramer & Mah, 1994). Since accurate process models usually are not easily obtained, the most widely used filtering methods do not rely on fundamental or empirical models, instead, they rely on information about the nature of the errors or the smoothness of the underlying signal. Examples of model-free filters include the well-known low pass filters, such as Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) filters (Tham & Parr, 1994). Examples of FIR and IIR filters include the Mean Filter (MF) and the Ex-