UNIVERSITAT POLITÈCNICA DE CATALUNYA

Ph.D. Programme:
SOFTWARE

Doctoral Thesis

QUALITATIVE MODELING AND SIMULATION OF BIOMEDICAL SYSTEMS USING FUZZY INDUCTIVE REASONING

Angela Nebot Castells

Institut de Cibernètica

November 1994

Advisors: François E. Cellier Rafael M. Huber

Acknowledgments

This doctoral thesis has been developed in the framework of two CICYT projects, CICYT TIC 90–0711 and CICYT TIC 93–0447, that have provided infrastructure support for this work. This support is gratefully acknowledged.

I would like to express my thankfulness to all the people who have contributed in one way or another to the work presented in this doctoral thesis. In particular, I would like to mention:

Dr. François E. Cellier, Associate Professor at the University of Arizona, for his continuous advice and guidance along the realization of this dissertation. He gave me his support in every single moment, something that I will never forget. He knew not only how to awake my interest and eagerness for the research, but also showed me to enjoy it. In his own words: *research is so much fun.* My sincere acknowledgments and admiration to him.

Dr. Rafael Huber, Professor at the Universitat Politècnica de Catalunya, for his interest and support during the evolution of this dissertation. He gave me courage to continue with this research during the hard moments that all too often exist during the development of a doctoral thesis. His guidance also in the professional area is truly appreciated.

Dr. Derek A. Linkens, Professor at the University of Sheffield, for giving me the opportunity to spend some time at that university and for providing me with the anesthesiology application data.

I would like also to thank *Dr. Roser García* and *Dr. Jaume Canet* of the Hospital de la Santa Creu i Sant Pau in Barcelona for introducing me to the anesthesiology science, as well as for their invaluable comments and opinions during the numerous meetings that we held together.

Dr. Montserrat Vallverdú, researcher at the Institut de Cibernètica of the Universitat Politècnica de Catalunya, for providing me with the cardiovascular control application data. Her guidance in the cardiology field as well as in the development of the related text has been very valuable for me.

I would also like to thank my companions and friends, *Francisco Mugica* and *Alvaro de Albornoz*, for the very hard but also happy moments that we shared during the development of our respective theses. I will miss them a lot in a few months when they will be back in their home in Mexico.

I also want to thank the rest of the SAPS group, *Sebastián Medina*, *Josefina López* and *Jesús Sardá*, as well as all the members of the control area of the Institut de Cibernètica for their highly appreciated support of my efforts.

My sincere acknowledgments go to my friend and desk neighbor *Robert Griñó*. His stoic patience and deep knowledge of control helped me a lot on many occasions, not only in technical but also in personal discussions.

I also want to thank for the support and courage that my friends of the administration department as well as the librarian of the Institut de Cibernètica gave me constantly. Their help with the, always too bureaucratic, administrative aspects of this work is sincerely appreciated.

Finally, I want to mention all the people that provided moral support in every single moment, especially my husband, parents, sisters, brothers in law, and my nephew, David, not to forget my unconditional friends *Jordi Castells* and *Oriol Maynés*.

Summary of the doctoral thesis

QUALITATIVE MODELING AND SIMULATION OF BIOMEDICAL SYSTEMS USING FUZZY INDUCTIVE REASONING

Author: *Angela Nebot Castells*
Advisors: *François E. Cellier* *Rafael M. Huber*

Biomedical engineering is a discipline that addresses medical and biological problems through the use of theories borrowed from the physical sciences, and technologies inherited from engineering.

As a consequence of the imprecision of knowledge available in biomedicine in general, one would expect that qualitative reasoning techniques, as they have been developed during recent decades by researchers working in the area of artificial intelligence, would be ideally suited to tackle problems stemming from biomedical fields. Yet, the application of artificial intelligence to biomedical sciences has not advanced rapidly in the past. Artificial intelligence techniques have been proliferated much more rapidly to and within other areas of science and technology.

Several difficulties inherent to the biomedical fields have restrained the progress in modeling and simulation of this type of systems in the past, problems that make these systems much more difficult to tackle than practically all other types of systems met anywhere in science and engineering.

The aim of the work developed in this dissertation was to address some of these difficulties, of soft sciences in general and of biomedical engineering in particular, and to come up with a methodology that would make optimal use of the limited knowledge available to the modeler, a modeling methodology that

would not get confused by the inevitable incompleteness and even inconsistency of information generally available for these types of systems.

To this end, a qualitative modeling and simulation methodology called *Fuzzy Inductive Reasoning (FIR)* has been employed, a modeling technique of fairly recent vintage that looked promising for the task at hand, and for which a prototypical implementation was available to us. The methodology has been refined, and a second generation of FIR software was implemented that would allow us to work with biomedical and other soft science systems.

This thesis is structured in 10 chapters. The first chapter describes the motivation for this doctoral thesis research. It also discusses the selection of the methodology, and introduces the aims and scope of this dissertation. The state of the art of qualitative analysis applied to biomedical applications is presented in Chapter 2, providing an insight into the different qualitative methodologies used to this day for dealing with the problems inherent in biomedical applications.

Chapter 3 describes in detail the *Fuzzy Inductive Reasoning* methodology. The foundations of this methodology were laid by Prof. George Klir of the *State University of New York (SUNY) at Binghamton.* It has been further developed and brought to maturity in a joint effort by three Ph.D. students of the *Universitat Politècnica de Catalunya* under the guidance of Prof. François Cellier and Prof. Rafael Huber. The author of this dissertation was one among these three students.

The methodology used for mixed quantitative/qualitative modeling and simulation is presented in Chapter 4 of the dissertation, where a discussion of its usefulness for soft sciences is also given. This concept is described by means of a hydraulic control system example, where the validity of the proposed approach to mixed modeling and simulation is clearly demonstrated.

The next chapters are centered on biomedical systems, starting with Chapter 5 where the difficulties inherent in dealing with these types of systems are enumerated and explained in detail.

Chapter 6 focuses on the qualitative control of biomedical systems. Two vastly different types of biomedical system controllers are described: the control of the depth of anaesthesia and the central nervous system that controls the hemodynamical system. The former application represents a technical (external) controller of a biomedical system, whereas the latter represents an electro-chemical (internal) control mechanism built into a biomedical system. The purpose of the former application is the external control of a biomedical

system, whereas the purpose of the latter is that of modeling a part of the central nervous system control, of understanding its functioning, and of predicting its future behavior. The difficulties encountered when modeling and simulating these two controllers are presented, and the solutions obtained by applying the fuzzy inductive reasoning approach to these systems demonstrate the strength of this methodology.

In Chapter 7, some limitations and weaknesses of the FIR methodology, as it had originally been devised, are presented. These problems are related to the limits of predictability of behavior in biomedical applications. To this end, a conceptual barrier limiting the predictability of future states of the systems under investigation is introduced. This conceptual barrier has been called *causality horizon*, and two systems, a linear state–space model and a biomedical system, served to demonstrate the concept.

Chapter 8 describes an enhancement of the FIR methodology for dealing with incomplete data records. To this end, a technique called *missing data option* is implemented. The missing data feature enables the researcher to work with incomplete data records and extract as much information from them as they contain. The feature makes it possible to convert incomplete quantitative data sets to reduced qualitative data sets in order to derive the best possible qualitative model for prediction of future system behavior.

Chapter 9 uses the missing data option to tackle one important problem, described already in Chapter 7, namely the elimination of patient–specific behavior. This work is a first attempt at tackling this difficult problem. The technique has been applied, until now, to a single anesthesiology system only. However, the results obtained are encouraging enough to, in a near future, apply this technique to other systems for which more patient data are available.

The research effort documented in this dissertation has focused on methodological issues. The main emphasis of the dissertation was to enhance the FIR methodology to be able to apply it to soft science systems in general and to biomedical science systems in particular. Therefore, the different biomedical applications presented in this doctoral thesis serve only to demonstrate the feasibility and validity of the chosen approach. The dissertation does not claim to have solved any medical problem in its full complexity. This task is left to future, more application–oriented, research projects that will need to be conducted in a joint effort of researchers with engineering and life science backgrounds.

An extensive performance comparison between the FIR methodology and

other inductive modeling techniques, such as the neural network approach and the NARMAX method, is an important aspect of this thesis that confirms the capability of the FIR methodology to deal with this type of systems. This thesis demonstrates that the FIR methodology is indeed a very powerful tool for the identification of biomedical system models, and for predicting their future development.

Index

List of Figures

List of Tables

Chapter 1

Introduction

Qualitative simulation has become a fashionable branch of research in artificial intelligence. Human reasoning has been understood as a process of mental simulation, and qualitative simulation has been introduced as an attempt to replicate, in the computer, facets of human reasoning.

Qualitative simulation can be defined as evaluating the behavior of a system in qualitative terms (Cellier, 1991a). To this end, the states that the system can be in are lumped together to a finite (discrete) set. For example, instead of dealing with temperature as a real–valued quantity with values such as 22.0^{o}C, or 71.6^{o}F, or 295.15 K, qualitative temperature values may be characterized as 'cold,' 'warm,' or 'hot.'

Qualitative variables are variables that assume qualitative values. Variables of a dynamical system are functions of time. The behavior of a dynamical system is a description of the values of its variables over time. The behavior of qualitative variables is commonly referred to as *episodical behavior*. Qualitative simulation can thus be defined as the process of inferring the episodical behavior of a qualitative dynamical system or model.

Qualitative variables are frequently interpreted as an ordered set without distance measure (Babbie, 1989). It is correct that 'warm' is "larger" (warmer) than 'cold,' and that 'hot' is "larger" (warmer) than 'warm.' Yet, it is not true that

$$\text{'warm'} - \text{'cold'} = \text{'hot'} - \text{'warm'} \tag{1.1}$$

or, even more absurdly, that

$$\text{'hot'} = 2 * \text{'warm'} - \text{'cold'} \tag{1.2}$$

The approaches to qualitative modeling proposed in the literature vary in how they derive qualitative rules from either structural knowledge (modeling from first principles) or behavioral knowledge (modeling from observations). The approaches to qualitative simulation proposed in the literature vary in how they propagate qualitative facts through the qualitative rules, i.e., in how they infer new qualitative knowledge from existing knowledge.

The qualitative modeling and simulation methodology of choice in this thesis is the so–called **Fuzzy Inductive Reasoning (FIR)** technique. It is entirely based on behavioral knowledge. Therefore, the approach is well suited for dealing with soft sciences, such as biomedical engineering, where the structure of systems to be modeled is usually either totally or at least partially unknown.

Biomedical Engineering is a research area of fairly recent vintage. According to Rushmer (Rushmer, 1972), it can be defined as the discipline that applies principles of science and engineering, as well as technological practices and methods to characterizing, understanding, and solving medical and biological problems.

During recent years, the use of analysis and dynamical system modeling methods to solve biological and medical problems has considerably increased. These techniques can supply accurate descriptions of complex dynamical processes, provide ideas for how to improve experimental designs, and offer means for testing hypotheses. The application of these techniques to physiological systems can generate a deep and diversified understanding of the general nature of these systems and of the complex processes that govern them.

Biomedical engineering encompasses three major facets:

- **Clinical Engineering** deals with the application of engineering and technology concepts to the improvement of the operation of hospitals and other health care systems. It is concerned, for example, with questions of resource management and equipment maintenance.

- **Medical Engineering** studies the application of engineering and technology concepts to the development of instrumentation, materials, diagnostic and therapeutic devices, artificial organs, and other medical equipment.

- **Medical Computer Science** focuses on the application of computer science concepts and methods to the development of equipment for diagnosis and therapeutic aid, as well as for routine daily work in hospitals and nursing homes.

Of course, the borders of the division between the three subareas of biomedical engineering are not crisp, and there is no reason either why they should be. Information must flow freely between these three subareas to obtain the most useful results.

It is the aim of this thesis to illuminate inherent characteristics, limitations, and problems of soft sciences in general, and of biomedical engineering in particular. A few of the known problems in biomedical engineering technology have been solved, and some of the previous limitations of biomedical engineering technology have been removed. To this end, the fuzzy inductive reasoning (FIR) methodology has been applied to tackling some practical biomedical applications of realistic complexity. The solutions that are being exemplified in this thesis by means of these applications are of a fairly general nature, and it will be possible to apply the same approaches to a large variety of other biomedical applications in the future.

This is the first account ever of applying the FIR methodology to biomedical engineering problems, and we are convinced that we have made a significant contribution to biomedical engineering research by doing so. We predict that many more biomedical applications of this technology will be reported in the future.

In the next sections of this report, the motivation for this research is being presented as well as a justification of why the FIR methodology was chosen to accomplish the goals of this work. The aims and scope of the doctoral thesis are also outlined later in this chapter. Finally, in order to simplify the reading of this thesis, its overall structure is provided in the last section of the present chapter.

1.1 Motivation for the Research

The dissertation focuses on the application of fuzzy inductive modeling and simulation methodologies to the analysis of biomedical systems.

Since qualitative simulation, due to its inherent coarseness, is able to

deal with *uncertainty*, it was hypothesized that this methodology should be well suited for dealing with an application area in which different types of uncertainties are inevitable.

Biomedical systems exhibit uncertainties with respect to:

- **the functional relationships** among the variables of the system, which may be unknown or at least incompletely known,
- **the model parameters**, which may exhibit a large variation from one patient to another, and
- **the measurement data**, which are mostly scarce and often incomplete, since the patient's needs and comfort are more important than the obtaining of reliable data records.

Biomedical systems were studied under two different scenarios:

- **Purely qualitative modeling and simulation**: The internal structure of the system is either entirely unknown or at least partially unknown.
- **Mixed quantitative/qualitative modeling and simulation**: The mathematical models of some subsystems are well understood, whereas the internal structure of other subsystems is unknown. A mixed model can describe those portions of the overall system that are well understood by quantitative differential equation models, while other aspects that are less well understood may be representable in qualitative terms.

Several reasons can be brought forward that may justify the use of qualitative modeling and simulation in biomedical systems. These can be summarized as follows:

- The internal structure of biomedical systems is commonly found to be totally or partially unknown, making it impossible to apply analytical models. For instance, it is well known that a blood volume increase within a vascular compartment causes an increase in blood pressure. However, formulating an exact mathematical model of this relationship requires a wealth of information that is often not available, such as the dimensions of different vessels, the visco–elastic characteristics of the walls, etc.

- The knowledge of the physiological and biomedical processes is often too scanty and fragmentary for an accurate mathematical description of their internal dynamics.

- Knowledge, encoded in a mathematical model, e.g. formulated as a set of differential equations, cannot easily be translated into useful explanation and justification mechanisms as they are needed by the end users of these models, i.e., by the medical personnel.

- Biomedical models, in order to be useful, must have interfaces to the human users of these models, i.e., the medical personnel, both at the input side and the output side. It is therefore of vital importance that these models, at least at their interface points, are compatible with human reasoning mechanisms. It does not make sense at all to formulate models that require input data that the medical personnel is unable to provide, or that generate outputs that are not translatable into meaningful and applicable biomedical recommendations.

- In biomedical systems, it is often very useful to explore all possible system behaviors. This requirement can only be satisfied by qualitative methodologies.

Therefore, qualitative modeling and simulation techniques with their inherent tolerance for uncertainty and ambiguity were expected to provide an excellent platform for the simulation of biomedical systems that may be difficult to model in quantitative terms, and even where quantitative models are available, qualitative models may constitute an important complement to the more classical quantitative models.

1.2 Selection of the Methodology

Several qualitative methodologies have been applied to biomedical systems in the past including expert systems, neural networks, qualitative physics, and fuzzy systems.

Some of these techniques are *knowledge–based*, i.e., they rely on structural knowledge about the internal functioning of the system under study. Qualitative aspects are brought into these methods by relaxing the quantitative details of this structural knowledge. For example, a particular quantitative formula relating two internal variables to each other may be replaced by

a generalized (qualitative) formula stating that variable y is monotonically increasing with variable x. In the artificial intelligence literature, these approaches are often referred to as "modeling from first principles." They are essentially deductive modeling approaches.

Other techniques are *pattern–based*, i.e., they rely on behavioral knowledge about the external relationships between inputs and outputs of the system under study. Qualitative aspects are brought into these methods by relaxing details of the behavioral knowledge. For example, rather than looking at precise (microscopic) relationships between two quantitative variables, they look at the more global (macroscopic) relationships between class values of qualitative variables.

The most widely advocated among the qualitative simulation techniques are the knowledge–based approaches that were originally derived from the *Naïve Physics Manifesto* (Hayes, 1979). Several dialects of these types of qualitative models exist (de Kleer and Brown, 1984), (Forbus, 1984), (Kuipers, 1986). They are best summarized in (Bobrow, 1984).

Although in the past two decades some expert systems were successfully developed to function as automated consultation systems for the diagnosis of diseases (Shortliffe, 1976; Bratko, 1988), artificial intelligence was not able to make a significant impact on the health care domain (Uckun, 1992). One reason for the deficiency of expert systems to deal with many aspects of biomedical problems is their knowledge–based foundation. Whereas static (statistical) knowledge about relationships among biomedical variables can be obtained easily, precise knowledge about the time constants that govern the dynamic behavior of biomedical systems is rarely available. Consequently, knowledge–based approaches to biomedical system modeling often fail when dealing with time–varying phenomena. Also, these approaches are usually in trouble when confronted with missing data, as they are all too common in biomedical applications. Although first attempts have been reported that may indeed, in the future, enable knowledge–based systems to deal with incomplete and even inconsistent information (Genesereth and Nilsson, 1987), the proposed non–monotic reasoning techniques are very difficult to apply in practice, and no large scale engineering applications of these techniques have yet been reported (Sarjoughian, 1994).

Among the pattern–based approaches to modeling, neural networks are the most widely acclaimed techniques. Neural networks are more likely to be able to come up with predictive models. However, also neural networks haven't made much of an impact on the health care domain yet. The reasons for their

inability to deal with biomedical systems adequately are manifold:

- Most neural networks operate on quantitative (real–valued) variables, rather than on qualitative (enumerated) variables. They are not naturally compatible with the higher levels of human reasoning, and it is therefore difficult to provide meaningful interfaces between a neural network and its medical users.
- Neural networks don't usually provide much insight into their internal reasoning processes. A medical doctor will not usually trust a recommendation made by a program, unless he or she can obtain a printout of the qualitative reasoning steps that led to this recommendation. It is thus insufficient to just provide a qualitative interface while performing the internal computations in a quantitative fashion.
- Neural networks provide for a natural interpolation/extrapolation capability. While this is usually a desirable feature, it is quite dangerous when applied to biomedical systems. Neural networks may predict behavior or issue recommendations that are not justifiable on the basis of the available facts. Any model should carry with it a self–assessment capability that is able to check whether the inherent assumptions behind the model are met, e.g. whether the model variables are operating in a range for which the model has been validated. This is even more important in qualitative modeling. Neural networks don't usually offer any self–assessment capabilities at all.

The fuzzy inductive reasoning methodology originally proposed by George Klir (Klir, 1985), was chosen for developing this dissertation, since it overcomes many of the problems mentioned earlier both with respect to the knowledge–based and the pattern–based approaches (Nebot, 1994). Some of the advantages of this methodology are the following:

- The technique can be applied to any system available to experimentation and observation. Inductive reasoning is fully pattern–based, thus, there is no need for knowing the internal structure of the system under study. In this respect, inductive reasoners are similar to neural networks.
- Inductive reasoners allow the otherwise qualitative models to treat time as a continuous (quantitative) variable. This is of primary importance if we wish to model and simulate mixed quantitative/qualitative systems.

- The methodology contains an inherent model validation mechanism inside the simulation method, which will prevent it from reaching conclusions that are not justifiable on the basis of the available facts. In this respect, inductive reasoners are similar to knowledge–based systems.

- Inductive reasoning operates in a qualitative fashion just like the knowledge–based reasoners. Although the inductive reasoner is not able, at this moment, to offer a complete trace back of the full reasoning process, as expert systems do, it does provide information about the subset of variables selected for the reasoning process, and it can at least provide a justification for the predicted output based on the qualitative states of the selected input variables. Another student of the *Universitat Politècnica de Catalunya* is currently working on an extension of the methodology that will permit, in the future, to provide the user with a more complete and better comprehensible trace back of the reasoning mechanisms.

1.3 Aims and Scope of the Thesis

The work developed in this dissertation establishes a contribution to modeling and simulation efforts in soft sciences, such as biomedical engineering. To this end, a qualitative methodology based on induction and fuzziness is proposed that addresses some of the difficulties inherent in dealing with these types of systems. To illustrate the results obtained in this research effort, two kinds of biomedical signals are being used:

- Anaesthesia signals used to control the amount of Isoflurane dosage to be applied to patients during surgical operations. These data records were measured at the *University Hospital of the University of Glasgow* (Scotland) during two different types of surgery: a renal transplant operation and an abdominal–perineal resection operation.

- Cardiovascular signals from a patient with coronary disease obtained by means of cardiac catheterization in the *Hospital de la Santa Creu i Sant Pau* (Barcelona).

Describing in more detail the purpose of this work, the following aspects can be mentioned:

1. **Modeling and simulation of mixed quantitative/qualitative biomedical systems**. The main objective here is to be able to simulate biomedical systems with partial structural knowledge. Subsystems for which quantitative structural knowledge is available should be modeled by traditional quantitative techniques, such as differential equation models, whereas subsystems for which such detailed knowledge is lacking should be modeled by qualitative techniques.

   A position control system involving a hydraulic motor with a four–way servo valve demonstrates, for the first time, the process of mixed quantitative and qualitative simulation using fuzzy inductive reasoning (Cellier *et al.*, 1992, 1994). With this example, it has been possible to tackle the theoretical challenges of the mixed modeling methodology without mingling them at once with practical issues of poorly understood biomedical processes. This work is described in Chapter 4 of this dissertation.

   Subsequently, a biomedical system representing the cardiovascular control system is being studied. The cardiovascular control system is composed of the *hemodynamical system* with a well known structure comprising the heart and the blood vessels, and the *central nervous system control*, with a partially unknown structure, responsible for the beating of the heart. This work is presented in Chapter 6 of this dissertation.

2. **Qualitative control of biomedical systems**. This research, described also in Chapter 6, has been centered around a biomedical control application. Fuzzy inductive reasoning is applied to a system for predicting the right value of an anaesthetic agent to be administered to patients during surgery. It replicates a human decision making process (Nebot *et al.*, 1993a).

   The main objective here is to address some of the problems that arise when working with this kind of systems, such as dealing within a single qualitative model with time constants that are one order of magnitude apart. It turns out that stiff systems pose difficulties not only to quantitative simulation attempts, but create serious obstacles also when dealt with in qualitative terms.

3. **Dealing with incomplete measurement data**. Modelers of biomedical systems are often plagued by incomplete data records. When a set of data is being recorded from a patient in the hospital, a variety of unexpected circumstances can produce gaps of data not registered by the computer.

Qualitative methodologies that cannot deal with missing data values are therefore quite useless for biomedical applications. The main objective of this work, explained in detail in Chapter 8 of the dissertation, is the development of a missing data feature, that makes the qualitative modeling methodology resilient to missing data values, thereby enabling the researcher to successfully work with incomplete data sets and to extract as much information from them as they contain. The feature, called *missing data option*, makes it possible to convert incomplete quantitative data sets to reduced qualitative data sets in order to derive the best possible qualitative model for prediction of future system behavior (Nebot and Cellier, 1994b).

4. **Elimination of patient–specific behavior**. This research focuses on preconditioning biomedical data to eliminate patient–specific behavior. In most biomedical applications, such as for instance the control of depth of anaesthesia of a patient undergoing surgery, it is extremely useful to have available a model that identifies not only the behavior of one concrete patient on a specific day during a specific operation, but one that is able to capture the behavior of a class of similar patients undergoing similar operations.

   In Chapter 9, a systematic way for taking advantage of medical information obtained from different patient/operation pairs is being developed. The measured data streams stemming from different such pairs are being combined in such a way as to allow the synthesis of a common model. It is expected that this model can be used for an entire class of similar patients undergoing similar types of surgery (Nebot and Cellier, 1994a).

5. **Limitations to predictability of behavior**. In this research, the concept of a *causality horizon* is introduced, a conceptual barrier limiting the predictability of future states of a system under investigation.

   Experiences with several applications have shown that the quality of predictions is not always the same. In particular, it was much more difficult to obtain even half–way decent predictions for biomedical applications, whereas the predictions in technical applications were accurate far beyond our original expectations. This part of the research, described in Chapter 7 of the dissertation, illuminates and explains this discrepancy by means of the causality horizon concept.

   This concept is directly related to the shape of the cross–correlation functions between each input with the output of a given multi–input single–output system and the autocorrelation function of its output. The

correlation functions can be viewed as *measures of causality* (Nebot *et al.*, 1994).

1.4 Structure of the Thesis

This thesis is structured in 10 chapters. The state of the art of qualitative reasoning applied to biomedical applications is presented in Chapter 2, providing an insight into the different qualitative methodologies used to this day for dealing with the problems inherent in biomedical applications.

Chapter 3 describes in detail the *Fuzzy Inductive Reasoning* methodology. The foundations of this methodology were laid by Prof. George Klir of the *State University of New York (SUNY) at Binghamton*. It has been further developed and brought to maturity in a joint effort by three Ph.D. students of the *Universitat Politècnica de Catalunya* under the guidance of Prof. François Cellier and Prof. Rafael Huber. The author of this dissertation was one among these three students. In the interest of completeness and closedness of this dissertation, the methodology is explained here in full, although the credit for this development must be shared by all three Ph.D. students. A corresponding chapter will consequently be present in all three dissertations.

The methodology used for mixed quantitative/qualitative modeling and simulation is presented in Chapter 4 of the dissertation, where a discussion of its usefulness for soft sciences is also given. This concept is described by means of a hydraulic control system example, where the validity of the proposed approach to mixed modeling and simulation is clearly demonstrated. Also this research represents a joint effort of all three Ph.D. students.

The next chapters are centered on biomedical systems, starting with Chapter 5 where the difficulties inherent in dealing with these types of systems are enumerated and explained in detail.

Chapter 6 focuses on the qualitative control of biomedical systems. Two vastly different types of biomedical system controllers are described: the control of the depth of anaesthesia and the part of the central nervous system that controls the hemodynamical system. The former application represents a technical (external) controller of a biomedical system, whereas the latter represents an electro–chemical (internal) control mechanism built into a biomedical system. The purpose of the former application is the external control of a biomedical system, whereas the purpose of the latter is that of modeling a portion of the central nervous system, of understanding its

functioning, and of predicting its future behavior. The difficulties encountered when modeling and simulating these two controllers are presented, and the solutions obtained by applying the fuzzy inductive reasoning approach to these systems demonstrate the strength of this methodology.

In Chapter 7, some limitations and weaknesses of the FIR methodology, as it had originally been devised, are presented.

Chapter 8 describes an enhancement of the FIR methodology for dealing with incomplete data records. To this end, a technique called *missing data option* is implemented.

Chapter 9 uses the missing data option to tackle one important problem, described already in Chapter 7, namely the elimination of patient–specific behavior.

Finally, Chapter 10 provides a summary of the applicability of this methodology, lists the major contributions of the dissertation, and presents an outlook of open problems and possible future research efforts extending the work presented in this thesis.

Chapter 2

The Use of Qualitative Reasoning in Biomedicine

2.1 Introduction

The aim of this chapter is to provide a brief state–of–the–art survey of ongoing research efforts in the use of qualitative reasoning techniques as they are applied to biomedical processes, and to show how the research presented in this dissertation fits into the overall picture of related research activities.

First, a historical overview of the early days of this research area is given. Subsequently, the chapter offers a classification of the different qualitative reasoning methodologies that have contributed significantly to this application area, and that are therefore considered to be of major interest here. For each of these techniques, the most important results obtained are summarized with reference to the most relevant publications in each case. A list of those application areas within the larger field of biomedicine that have most benefitted from these research efforts is also provided.

The qualitative approach to modeling and simulation has several advantages over conventional numerical methods when dealing with medical problems. A full description of its advantages has already been presented in Chapter 1 of this dissertation, therefore, only the two most relevant ones are remembered here. First, in order to obtain a quantitative model, exact numerical values of all parameters characterizing the system have to be obtained. This is a true obstacle to quantitative modeling in medicine as well as in biomedical engineering, since many parameters of these types of systems are difficult or

impossible to measure, and even the values of those that can be measured may differ vastly from one patient to another. Second, the qualitative point of view is closer to human reasoning mechanisms, and therefore, if the entire chain of thoughts rather than only a final result obtained in obscure ways, is reported back to the human users of the tool, i.e., the medical personnel, people who are often characterized by an almost instinctive fear of computers and a deep–rooted distrust in their functionings, these users will be much more at ease, and the findings may be met with more trust, making them much more effective on the long run.

These considerations suggest that artificial intelligence techniques might offer good perspectives for these kinds of systems. It is indeed a fact that meanwhile many medical instruments rely on artificial intelligence in order to:

- provide a more accurate interpretation of patient status (**diagnosis**),
- provide a better understanding of causal mechanisms (**explanation**), and
- suggest modifications in patient management (**therapy**).

This was not the case two decades ago when qualitative modeling and simulation methodologies were first investigated by researchers and applied to a wide range of disciplines with significant success, yet failed to make a true impact on biomedical sciences. Several biomedical expert systems were developed during the seventies and the beginning of the eighties, triggered by the advocacy of researchers in the academic community recommending the use of artificial intelligence concepts, such as expert systems and causal reasoning. The MYCIN diagnostic system (Shortliffe, 1976) is the earliest well–publicized example of a biomedical expert system. Several other expert systems were developed around the same time, some of which are still being worked on at the current time. However, few of these expert systems were of extended practical use, maybe with the exceptions of INTERNIST-I/QMR described in (Miller *et al.*, 1986) and DXplain (Barnett *et al.*, 1987).

There is a principal reason why artificial intelligence research failed when applied to biomedical sciences. The main idea of the early artificial intelligence researchers working on such applications was to come up with automated diagnosis systems, an idea that was vehemently rejected by the health care professionals. Most physicians are much more inclined to rely on their colleagues' opinions than on computers. This had already been predicted by Szolovits who wrote that this kind of *programs will be clinically accepted*

only when their indispensability is established, in other words, when it can be successfully demonstrated that medical personnel does perform better when assisted by these programs (Uckun, 1992).

As it is described by Uckun, some other reasons contributed to the lack of success, such as:

- the relatively shallow level of knowledge researchers have about most disease processes,
- the complexity and the inherent variability of human anatomy and physiology,
- lack of understanding of the cognitive processes that underlie medical decision making,
- the sheer size of the required knowledge bases,
- lack of familiarity with computers on the part of most health care personnel, and
- the legal implications of making use of computer–assisted advice in the case of health care applications.

Recent advances in biomedical instrumentation have led to more and more digital and analog data being obtained from the patient and presented to the medical personnel in real time, e.g. during surgery or in an intensive care unit. Practitioners like this wealth of data, but feel often overwhelmed, and live with the constant fear that one day they might overlook something important in the sea of information presented to them, something they could have noticed if only they had paid attention to the right monitor at the right moment of time.

They therefore yearn for *intelligent data monitors*, i.e., artificial intelligence systems that monitor the incoming data, screen it, look out for anomalies, and point out potential problems to the practitioner. Consequently, the focus is no longer on automated decision making, but rather on smart sensors, intelligent monitors, data fusion techniques, data filtering methods, and other tools that can assist the human practitioner in coming up with the right decision within a minimum amount of time.

If the scope of artificial intelligence research efforts is focused on satisfying the expressed needs of medical personnel rather than satisfying their own

desires and aspirations while ignoring the realities of the application area at hand, then the perspectives of success of qualitative methodologies applied to biomedical systems are in fact quite high. This realization led to a second generation of artificial intelligence tools for biomedical processes.

In the next section, a general classification of qualitative methodologies that have been applied in the past to address questions in biomedicine is provided.

2.2 Qualitative Methodologies for Reasoning with Biomedical Systems

In the previous section, a general review of the early days of qualitative methodologies applied to biomedical systems was presented. The objective of the current section is to provide an insight into some of the results obtained in the biomedical field using qualitative techniques. Qualitative methodologies that have been successfully applied to biomedical systems can be grouped into the following five categories: expert systems, neural networks, qualitative physics, fuzzy systems, and inductive reasoners.

2.2.1 Expert Systems

As has been pointed out earlier, expert system technologies have been applied to biomedical systems already in the seventies. This is the reason why a lot of publications can be found in this area. Most of the research efforts in biomedicine that make use of expert systems are focused on diagnosis. Major authors of papers contributing to these efforts are: Shortliffe (MYCIN), Bratko (KARDIO), Miller (INTERNIST–I), Marine (CAEMF), Levy (SESAM–DIABETE), and Linkens (RESAC).

MYCIN was the first rule–based expert system developed in the biomedical field (Shortliffe, 1976). A major shortcoming of most early medical expert systems, such as MYCIN, is their total ignorance of the *passing of time*. Yet, time is a major factor that ought to have an influence on the diagnosis process. Taking the flow of time into consideration is crucial for proper assessment of the patient's evolution. Several theories have been developed that discuss how *time* can be incorporated as a factor in expert systems, and indeed, many of the newer biomedical expert systems are time–dependent. One such system is ONCOCIN (Kahn *et al.*, 1985). Its function is to aid the chemotherapy of

cancer. In that system, the patient is associated with a temporal graph whose nodes represent important clinical events. These nodes are *open* if the event is currently taking place, and they are *closed* if the event has already ended.

A later version of MYCIN is NEOMYCIN developed in the early eighties (Clancey and Letsinger, 1984). The idea behind NEOMYCIN was a reconfiguration of the rule–based expert system MYCIN for applications to teaching. The resistance of practitioners immediately vanished as soon as no real patients were at stake. The purpose of NEOMYCIN is to raise the confidence level of young medical doctors in their capabilities to come up with correct diagnoses of diseases. This is definitely useful.

ONYX is an extension of the ONCOCIN system (Langlotz *et al.*, 1987). The modeling tasks are enhanced by means of an object–oriented simulation environment that facilitates the use of the system. The simulation strategy used is similar to the *device* concept advocated in (de Kleer and Brown, 1984).

Another example of a time–dependent expert systems is CAEMF–2, offering a language for the description and representation of medical temporal knowledge (Marín *et al.*, 1990). The application area of CAEMF–2 is obstetrics. CAEMF–2 represents a further development of an earlier biomedical expert system, CAEMF, used for patient monitoring during pregnancy.

A recent development in the area of temporal reasoning is the HyperLipid system developed by (Rucker *et al.*, 1990). Its main purpose is to combine expert systems and temporal databases in the domain of lipid management.

Although most of the early biomedical expert systems concentrated on various issues related to diagnosis, the variety of different diseases or types of applications to which these expert systems were applied is extremely diversified. They have been applied to almost all the areas within the biomedical domain. An important focus is primary care. Fox and Frost developed in the early eighties an expert systems capable of assisting primary care personnel with screening, diagnosis, and patient management (Fox and Frost, 1985). The idea was to offer a second opinion during and after a consultation with a medical doctor. As another example, Binik developed an expert system for the diagnosis of sexual dysfunctions (Binik *et al.*, 1988).

Another diagnostic expert system is CADIAG–2/PANCREAS (Adlassnig and Scheithauer, 1989). This system was developed for the diagnosis of pancreatic diseases. The CADIAG–2 system is integrated into the medical information system of the Vienna General Hospital. This integration allows

CADIAG–2 to collect patients' data directly from the information system, and to infer diagnoses from this data. CADIAG–2/PANCREAS contains the profiles of 10 different pancreatic diseases.

Apart from applications dedicated to computer-assisted diagnosis of pancreatic diseases alone, there are several other systems that include pancreatic diagnoses among others. This is the case of INTERNIST–ICADUCEUS, aimed at medical diagnoses in the entire field of internal medicine. INTERNIST–ICADUCEUS contains profiles of eight pancreatic diseases (Miller *et al.*, 1986).

Kinney and co–workers developed an expert system for the diagnosis of ascites (Kinney *et al.*, 1988). The authors found that rules produced from a small set of non–redundant examples seemed to capture most of the currently available knowledge about diagnosis of ascites. This system offered a correct clinical diagnosis in 82% of all cases when it was tested against a new set of patient data.

Another widely publicized expert system developed in the eighties is KARDIO (Bratko *et al.*, 1989; Bratko, 1988; Mozetič, 1990). KARDIO has been developed for electrocardiogram (ECG) diagnosis of disorders in the heart, known as cardiac arrhythmia, and the main task of the system is the interpretation of electrocardiograms. This expert system is implemented in Prolog, and it is composed of 35 rules. Its inference mechanism is a forward chaining engine. KARDIO does not offer a temporal dimension. KARDIO is probably the most used artificial intelligence system for ECG interpretation and one of the most successful and complete qualitative models in biomedicine to date (Hunter *et al.*, 1991).

Hunter and co–workers, strongly influenced by the KARDIO project, developed a qualitative approach based on the definition of a formalism that represents the state of the system at a point in time or over a time interval, and that defines a set of constraints (or rules) that specify what transitions between states are possible (Hunter *et al.*, 1991). Consequently, this approach is more closely related to qualitative physics methodologies than to expert systems, yet, it was decided to include it here due to its direct relation with KARDIO. The objective of this work, as in the case of KARDIO, is the interpretation of electrocardiograms.

Several expert systems, mostly developed during the late eighties and the early nineties, have focused on other than diagnosis tasks. RESAC is a system designed for controlling the depth of anaesthesia during surgery (Linkens *et*

al., 1986; Linkens *et al.*, 1990). RESAC comprises a rule–based backward chaining inference engine with about 400 rules, and makes use of fuzzy logic and Bayesian reasoning.

Another example is SESAM–DIABETE, an interactive educational expert system for controlling the amount of insulin dosage required by diabetic patients (Levy *et al.*, 1989). It was developed from a previous expert system called SESAM, and has been implemented in MBX, a Lisp–based general–purpose expert system design software. The SESAM–DIABETE system is currently accessible through the French Teletel network by means of a Minitel, and is under clinical evaluation.

Prokosch and co–workers developed an expert system for the human genetics area (Prokosch *et al.*, 1989). This work presents an object–oriented fact–based model for the representation of genealogical information, and can be used as a genetic counseling system. The application is of interest since this field has otherwise been largely ignored by the artificial intelligence community.

Ursino and colleagues present a method for qualitative simulation of dynamic physiological models that, in some aspects, is derived from QSIM (cf. section on Qualitative Simulation in this chapter). As in QSIM, the status of a physiological quantity is described by "magnitude" and the "rate of change." The main difference is that Ursino adopts a simplified representation of magnitude that is in closer agreement with that commonly used by physicians (Ursino *et al.*, 1992). In contrast to QSIM, the inference engine is an expert system developed within the KEE environment using a class of production rules that the authors call "certainty rules." This methodology has so far been tested by means of simple submodels of the cardiovascular system.

The AI/MM system designed by Kunz (Kunz, 1983), is an uncommon example of an expert system based on empirical rules that also uses the knowledge of a physiologic mathematical model of the system, in that case the renal system. The aim of this work is prediction of future states of the system, as well as to offer causal explanations of observed behavior.

The aim of PNEUMON–IA is to assess the etiology of community–acquired pneumonia from clinical, radiological, and laboratory data obtained at the beginning of an epidemic (Verdaguer *et al.*, 1992). PNEUMON–IA uses an inference engine named MILORD that operates on fuzzy logic and linguistic variables to express uncertainty (Sierra, 1989).

A medical diagnostic expert system in the domain of post–menopausal osteoporosis has been developed by Binaghi and co–workers (Binaghi *et*

al., 1993). The fundamental aims of the expert system are to standardize knowledge about and support physicians in the early detection of post–menopausal osteoporosis. This system has been generated within the AMDIS fuzzy expert system shell.

2.2.2 Neural Networks

This technology is currently receiving major attention from the medical community. A large number of research efforts are currently ongoing that make use of this methodology for the purpose of dealing with medical systems. Contrary to the previously discussed expert systems that are predominantly knowledge–based, neural networks are almost exclusively pattern–based. Thus, they deal with a distinct set of applications. The principal tasks dealt with by this methodology are patient monitoring, prognosis (forecasting patient progress), medical control, and classification. Cheung, Navabi, Linkens, Ciarroca, and Orr are some of the major researchers in this field.

The majority of neural networks deals with quantitative information directly, and therefore, such neural networks are not qualitative techniques in a puristic sense. Yet, they are able to function meaningfully in the presence of incomplete or imprecise knowledge, and they tackle the same kinds of problems that some of the qualitative techniques address. Thus, it makes sense to include them in the list of tools used for qualitative reasoning. Also, this approach comes in fact closest to the qualitative reasoning technique that is advocated in this dissertation, and therefore, it is even more important to include neural networks in the discussion. When assessing the success or failure of our own qualitative reasoning technique, we will frequently do so by comparing our results with those obtainable using neural networks.

In the area of patient monitoring, Artificial Neural Networks (ANNs) can play an important role in the classification of waveforms. Navabi and his colleagues have developed a computer–based, integrated monitoring system that collects and interactively manages physiological data from six routinely used operating room monitors (Navabi *et al.*, 1991). ANNs were used in this system for two purposes: first, for correctly classifying CO_2 waveforms into spontaneous, mechanical, and mechanical mixed with spontaneous breathing; second, for the detection of elevated and depressed so–called ST segments in the ECG signals. Two three–layer feed–forward ANNs were used, in the first case with 50 input nodes, six hidden layer nodes, and one output node, and in the second case with 80 input nodes, six hidden layer nodes, and two output nodes. This system also uses other kinds of reasoning tools, such as an expert

system for the purpose of intubation detection.

Other experimental operating room monitors have used ANNs as well and have obtain very good results, such as those described in (Cheung and Hull, 1989; Ciarroca, 1989; Orr and Westenskow, 1989).

Artificial neural networks are also used in the control area. As an alternative to the expert system RESAC, Linkens and Rehman have designed a neural network, ANNAD, to control the anaesthetic depth of patients during surgical operations (Linkens and Rehman, 1992a; Rehman *et al.*, 1993). The idea behind this development was that a neural network could make recommendations much faster than a 400–rule expert system. This is important in a real–time environment. A back–propagation algorithm was used to train a three–layered network with 10 input nodes, three hidden layer nodes, and four outputs nodes. This controller was able to forecast the amount of anaesthetic agent to be administered to a patient during surgery.

Pedrycz developed a scheme of electrocardiogram pattern classification using fuzzy sets for the categorization of the space of parameters for the recognition process, while using a single–layer ANN for the classification process (Pedrycz *et al.*, 1991).

Another area of application of neural networks inside the medical domain is as a predictive instrument for optimizing the scheduling of cardiac surgery patients in times of limited intensive care unit (ICU) resources. This is an important issue in countries such as Canada, where cardiovascular intensive care resources are limited and where waiting lists for cardiac surgery exist (Tu and Guerriere, 1993). The network developed uses the standard back–propagation algorithm, and it is composed of three layers of 15 processing input units, 12 hidden units, and a single output unit.

Schaltenbrand and co–workers have applied a new method based on a neural network model for discovering sleep stage patterns from data. The method is based on a simultaneous anaysis of three electrophysiological signals: electroencephalogram, electrooculogram, and electromyogram (Schaltenbrand *et al.*, 1993). The system contains two separate ANNs, an automatic sleep stage classifier consisting of a multilayer feedforward ANN, and a supervision unit watching over the automatic decision made that uses a non–supervised ANN.

Although ANNs had been known since the early forties (McCulloch and Pitts, 1943), only the advent of the backpropagation algorithm popularized as recently as 1986 (Rumelhart *et al.*, 1986) made ANNs applicable to more

than mere toy problems. Consequently, all the neural network applications in biomedicine are of a fairly recent vintage, and new results appear in the press daily. An end to this development is not yet in sight, and it is to be predicted that more, and exciting, research results will be published shortly.

2.2.3 Qualitative Physics

Some qualitative techniques that have their origin in the more traditional modeling methods have been advocated quite heavily in the eighties and have received a lot of attention. Their main emphasis is on explanation, diagnosis, and prognosis. The main advocates of these techniques are Kuipers, de Kleer, and Forbus.

2.2.3.1 Qualitative Simulation

Kuipers' approach, implemented in QSIM, can be viewed as an abstraction of ordinary differential equations (Kuipers, 1986). A model is defined by a set of "functions" of time, and a set of "constraints" on these functions. The values that the functions can take are either a *landmark* (a symbolic value), or an interval between two adjacent landmarks. Usual mathematical operations, such as arithmetic operations, time derivatives, and monotonically increasing and decreasing functions, are used to relate constraints and functions. QSIM treats time as a sequence of symbolic time points that are created when a function crosses its landmark value or when the derivative of a function becomes equal to zero. Kuipers' approach allows new landmarks to be discovered in the course of the qualitative simulation.

QSIM is the most rigorous of all the qualitative physics approaches due to its solid mathematical basis, and it is also the least structurally constraining among these methods. Therefore, this methodology is well suited for the representation of a wide variety of domains reaching from renal physiology to population dynamics.

Kuipers and Kassier applied QSIM to physiological processes, such as the renal control of the salt and water balance in the body, and the regulation of arterial blood pressure (Kuipers and Kassier, 1985). Later on, Coiera used QSIM as a basis for a diagnostic architecture applied to acid–base disorders and their regulation (Coiera, 1990).

Another diagnostic system that uses QSIM is NEOANEMIA. The domain

of application of NEOANEMIA and its predecessor ANEMIA is the iron metabolism in the human body, and it is used to recognize disorders causing anaemia. This system takes advantage of a qualitative simulation based on the QSIM constraint language for deducing expected parameter values, deriving a hybrid expert system (Stefanelli *et al.*, 1988; Lanzola *et al.*, 1990). NEOANEMIA has been implemented using KEE and Common Lisp, and is able to explain why a particular disease has been taken into consideration and to show the conditions that led to its pursuance.

There are two major limitations to QSIM. The first is related to the close relationship between differential equation models and QSIM models. It is almost impossible to model, in QSIM, systems that are not well suited to be modeled with differential equations as well. Second, QSIM can unfortunately generate spurious behaviors that are not physically feasible, which is a highly undesirable property of this methodology. This is referred to as *excessive branching*.

Another system that uses the QSIM technique is MIMIC, developed by Dvorak and co–workers (Dvorak *et al.*, 1990). MIMIC is a method for monitoring continuous–time dynamic systems, and in order to test its performance, MIMIC has been applied to an electric water heater.

2.2.3.2 Qualitative Physics Based on Confluences

This is the qualitative physics approach of de Kleer and Brown (de Kleer and Brown, 1984). It is centered around the concept of "confluence," i.e., a qualitative differential equation. The structure of the model is described through its components and their interconnections.

It is assumed that the behavior of a physical system can be derived from the behavior of its physical components. Unfortunately, this is not always the case. Therefore, this methodology can only be applied to systems were this assumption holds. This is the reason why qualitative physics based on confluences has been predominantly applied to electronic circuits, whereas it is not evident that this technology can be successfully applied to biomedical modeling.

Although at least one attempt to develop a scheme for qualitative representation of behavior in the medical domain using qualitative physics based on confluences has been reported in the literature (Bylander *et al.*, 1988), it has not been shown in any detail, how this could be applied to physiological

processes.

Weinberg has developed a prototype that models the electrical subsystem of the heart, and simulates normal and deviant cardiac rhythms (Weinberg *et al.*, 1989). This has been accomplished by means of a methodology that combines the device–centered methodology of de Kleer and Brown, and the process theory approach of Forbus (cf. next subsection of this chapter). The authors claim that this prototype could be the kernel of a decision–making tool for assisting medical personnel, or might be used as an instructional tool for teaching students about cardiac functions.

2.2.3.3 Qualitative Process Theory

The Qualitative Process Theory (QPT) was developed by Forbus (Forbus, 1984). In this methodology, all changes in the physical world are described in terms of "processes." As Forbus writes in his article, processes usually start and stop when orderings between quantities change. In QPT, the value of a number is represented by a "quantity space," a partial ordering of quantities determined by the domain physics and the analysis being performed. A quantity consists of two parts, an amount and a derivative, each of the two comprised of a sign value and a magnitude value. QPT does not contain the qualitative simulation algorithm. This is implemented separately, and is called Qualitative Process Engine (QPE).

Compared with the previously discussed approach of de Kleer and Brown, this methodology has the advantage that models can be dynamically restructured when an object no longer exists or when a new object is born to the system. Its main inconveniences are that it is the most complicated of all three approaches in terms of designing programs implementing the methodology, and that the computational complexity of the method is larger than in the other two cases.

QPT has been applied to realistic biomedical problems such as patient monitoring. One intelligent monitoring framework using this methodology is SIMON (Uckun *et al.*, 1993). SIMON has been applied to ventilator management in premature infants with respiratory distress syndrome. SIMON uses a methodology called YAQ. Its modeling method, prediction mechanism, and qualitative algebra are derived from QPT.

Another system based on QPT is QMI (Qualitative Measurement Interpretation algorithm) developed by Todd (Todd, 1988). QMI uses the

quantity spaces generated by the envisionment process to interpret numerical data taken across time. The application is based on the interpretation of blood glucose data using an envisionment generated from a model of the physiology of human glucose regulation.

2.2.4 Fuzzy Systems

Fuzzy systems constitute yet another qualitative reasoning paradigm. Fuzzy controllers have successfully been applied to various medical systems, and they therefore deserve to be mentioned in this context. Some of the more important contributions to the field of fuzzy systems as related to medical systems have been obtained and reported by: Meier, Isaka, Ying, and Czogala.

Isaka and Sebald propose a procedure for designing a fuzzy controller using a high–dimensional numerical optimization algorithm (Isaka and Sebald, 1992). In order to design a fuzzy controller, it is required to understand very well the dynamics of the process to be controlled and how they are influenced by the control variables, and it is also important to understand what effect the fuzzy membership functions exert on the performance of the fuzzy controller. This method, by means of an optimization strategy, selects the membership functions such that they optimize the control performance. The method has been applied to blood pressure regulation.

A fuzzy controller for the cultivation of microorganisms has been described by Czogala and Rawlik (Czogala and Rawlik, 1989). These authors applied two different kinds of controllers, a classical PID controller and a fuzzy controller, in order to compare their performance. The results obtained by means of fuzzy control were considerably better than those obtained by PID control due to the fact that the biological mechanisms involved are not completely known and highly non–linear.

A fuzzy controller for anaesthesia has been presented by Meier (Meier *et al.*, 1992). The mean arterial pressure is used as a parameter for estimating the amount of anaesthetic agent, Isoflurane, to be administered to a patient during surgery. The design of this fuzzy controller was iterative, and the reference points of the membership functions as well as the rules were chosen by trial and error. A major attraction of this controller is that, at the time of reporting, it had already been tested during 11 real surgical operations, where the anaesthetists who supervised the controller never had to intervene or override its recommendations. This is not the case of the fuzzy controller developed by Ying (Ying *et al.*, 1990), that still needs some improvements

before it can be used in a real clinical situation. Ying's fuzzy controller is used to regulate the mean arterial pressure of critically ill patients through sodium nitroprusside infusions.

Linkens and Mahfouf introduced another fuzzy controller for anaesthesia, this time to control muscle relaxation rather than the awareness level (Linkens and Mahfouf, 1988). In this work, the fuzzy rules were obtained using human operator control recorded manually during some operations. The recordings of the anesthetist's commands were then used to fill in the rule base. In order to solve the problem of knowledge acquisition in the design of fuzzy controllers, Linkens and Hasnain developed a technique for self–organization of fuzzy controllers (Linkens and Hasnain, 1991). This technique was capable of generating a meaningful set of rules for highly nonlinear dynamic systems.

Although the greater part of fuzzy set research in biomedicine has focused on control issues, fuzzy sets have been used for other purposes as well. This is the case of the work reported by Feng (Feng *et al.*, 1991), where fuzzy sets were used for the detection of the epicardial boundary. The resulting automated system is able to detect the endocardial and epicardial boundaries in a 2–D echocardiography. In this research, the high–level knowledge of global intensity change in the image is acquired from experts, and is then represented in the form of fuzzy linguistic descriptions and relations. The knowledge of local intensity change is then deduced from the knowledge of global intensity change through fuzzy reasoning.

The work of Cios (Cios *et al.*, 1991) is another example of a fuzzy biomedical system used for other than control purposes. This research focuses on the diagnosis of coronary artery stenosis. The authors claim that their tool could be used as a stand–alone system for the incorporation of automatically generated production rules into the knowledge base of an expert system.

Mira and his colleagues (Mira *et al.*, 1991) developed another strategy using the fuzzy paradigm for data acquisition to automatically build a database and a knowledge representation scheme. This acquisition methodology has been used in the expert system ONCOGAL for advise in diagnosis and chemotherapy treatment of cancer.

Barro and co–workers developed a fuzzy classifier of cardiac beats. The classification of each beat is performed applying fuzzy conditional statements that represent the knowledge of the cardiologist expert and use a set of descriptions of temporal and morphological attributes of the analyzed beat (Barro *et al.*, 1990a). This signal classifier has been integrated in a system

for the diagnosis of arrhythmia that has been implemented in a real–time monitoring system for physiological signals (SUTIL), developed for patients interned in coronary care units (Barro *et al.*, 1990b).

It can be noticed that fuzzy systems have already been employed in a large variety of real clinical situations for many different purposes. This is a very active research area, and more results are expected to be reported shortly.

2.2.5 Inductive Reasoning

Fuzzy Inductive Reasoning (FIR) is a fairly new paradigm that combines some attributes of knowledge–based systems with others of pattern–based approaches, is similar in its applicability to that of artificial neural networks, yet contains facets of fuzzy systems as well. Hence, FIR is a very versatile modeling paradigm that raises high expectations. Its major aims in the context of biomedical research are control, modeling, prognosis, diagnosis, and monitoring.

The FIR methodology has been derived from Klir's General Systems Problem Solving (GSPS) approach (Klir, 1985). Its main focus is on general systems analysis, to study the conceptual modes of behavior of systems. A crisp implementation of this theory was developed in 1987 (Cellier and Yandell, 1987) as an application library for CTRL–C. It was called SAPS–II, since it represented a reimplementation of an earlier tool called SAPS, which stands for Systems Approach Problem Solver (Uyttenhove, 1979). Later on, a fuzzy extension of SAPS–II was created by Li and Cellier (Li and Cellier, 1990). This was the basis for subsequent enhancements of this methodology with the purpose of obtaining a tool that can be used for qualitatively studying the behavior of biomedical systems (Nebot *et al.*, 1993a; Nebot and Cellier, 1994a, 1994b). Currently, several versions of SAPS–II exist. One is configured as a toolbox for Matlab. Another version (implementing only a subset of SAPS tools) can be used together with ACSL for mixed quantitative and qualitative simulations.

A reimplementation of the RESAC expert system (Linkens *et al.*, 1986; Linkens *et al.*, 1990) has recently been realized by means of the FIR methodology (Nebot *et al.*, 1993a). The system, called FIRAD, was able to control the amount of anaesthetic agent to be given to a patient during surgery. It compares favorably with the earlier neural network–based reimplementation called ANNAD (Linkens and Rehman, 1992a; Rehman *et al.*, 1993).

Another application of this methodology inside the biomedical field is the modeling and simulation of the cardiovascular control system. A combined quantitative/qualitative modeling and simulation technique was used in that case, allowing the representation of the hemodynamical system by means of quantitative methodologies using a differential equation model, and that of the central nervous system by means of a FIR model. The FIR model here replaces a previously developed NARMAX model (Vallverdú, 1993). The FIR model compares very favorably to the NARMAX model in many respects.

Since the FIR methodology is a very new approach in its early development phase that is still undergoing frequent modifications and enhancements, it was preferred to use the tool for reimplementing existing applications rather than for tackling new problems. In this way, the previous implementations can be used as a gauge to measure the success or failure of the FIR methodology against.

Each of the five methodologies described in this chapter is characterized by its own advantages and limitations when dealing with biomedical applications. Therefore, it depends on the characteristics of the biomedical system to be modeled and simulated, which of them should best be used. A description of the major advantages and disadvantages of the different methodologies presented here was already provided in the first chapter of this dissertation.

2.3 Aims of Qualitative Research

Each of the methodologies presented in the previous section is suitable for tackling a different subset of biomedical problems, and therefore the objectives vary from one approach to another. Table 2.1 presents the most frequently reported aims and scopes of these methodologies in a tabular form together with a description of their significance.

As it has been mentioned in previous sections, *diagnosis* is the most frequent focus of qualitative research within the biomedical engineering domain. Expert systems were the methodology most used for diagnostic purposes, especially during the early years. This is the case of many systems developed during the late seventies, the eighties and the early nineties, such are: MYCIN (Shortliffe, 1976), (Fox and Frost, 1985), (Binik *et al.*, 1988), CADIAG–2/PANCREAS (Adlassnig and Scheithauer, 1989), INTERNIST–I/CADUCEUS (Miller, 1986), (Kinney *et al.*, 1988), and KARDIO (Bratko *et al.*, 1989; Bratko, 1988; Mozetič, 1990).

Aim	**Significance**	**Methodology**
Diagnosis	Patient state	Mostly expert systems
Prognosis	Prediction of patient progress	ANNs and FIRs, Qualitative physics
Therapeutic Management	Therapy suggestion or modification	Mostly expert systems
Monitoring	Integration of monitors and data processing	ANNs and FIRs, Fuzzy systems
Modeling	Many applications of different kinds	All kinds of methodologies
Decision Support	Provide suggestion for treatment plan	So far rejected by most physicians
Control	Behavior optimization	FIRs, Fuzzy systems
Consultation System	Monitoring and management of patients records	Mostly expert systems
Causal Explanation	Description of causal mechanisms	Expert systems, Qualitative physics
Imaging	Boundary detection of an image	ANNs, Fuzzy systems, Statistical methods
Classification	To choose between different options	Usually recurrent neural networks
Training	To educate medical personnel	All kinds of methodologies

Table 2.1: *Principal aims of qualitative research*

A very important aim that is currently receiving major attention by the medical community is *prognosis*. Prognosis is based on the prediction of patient progress, and is extremely valuable in the medical context. Especially in those circumstances, where an advance notice of a potential problem of as little as five minutes may mean the difference between life and death of the patient, an early warning at practically any cost will be most welcome. Since the purpose of prognosis systems is only to set off an alarm at as early a time as possible, such systems are accepted without reservation by the medical practitioners, especially for use in intensive care units where no medical doctor is constantly present, but needs to be notified by the resident nurse as soon as a potential problem has been observed. Prognosis systems have been less accepted for the operating theater so far, since there is less of a need (the specialist is around at all times) and since too frequent alarms can be a major hassle. The medical field that has so far benefitted most from automated prognosis systems is cardiology. An early example of a biomedical prognosis system is AI/MM (Kunz, 1983).

Another primary focus of qualitative techniques is *control*. Automated feedback control can be useful both in the operating theater and in all kinds of care units, relieving the medical personnel of high–frequency low–level interventions. Several research efforts have been reported in which single–sensor/single–actuator (SSSA) controllers were designed to regulate a unique parameter of the body by use of a single physiological signal. A well studied application is the regulation of blood pressure by means of a drug. This is the case of the following works: (Kuipers and Kassier, 1985), (Ying *et al.*, 1990), (Isaka and Sebald, 1992), and (Meier *et al.*, 1992). In some of these cases, the blood pressure was controlled as an indicator of another signal that cannot directly be measured: the depth of anaesthesia during surgery.

Other qualitative controllers dealing with the regulation of anaesthetic depth during surgery are RESAC (Linkens *et al.*, 1986; Linkens *et al.*, 1990), ANNAD (Linkens and Rehman, 1992a; Rehman *et al.*, 1993), and FIRAD (Nebot *et al.*, 1993a). These systems are multi–sensor/single–actuator (MSSA) controllers. Qualitative control efforts have been reported for all of the qualitative methodologies presented in the previous section.

Modeling denotes the process of finding qualitative relationships between the variables of a system. In other words, it is the process of identifying a model that represents the dynamics of the system. Therefore, it is always necessary to obtain a model of the system before any other aim or scope can be accomplished. For this reason, it can be argued that modeling does not constitute an aim in its own right. In biomedical applications, the structure

of the systems under investigation are often at least partially unknown. Hence, it is of crucial importance for experimentation and in order to increase the knowledge that the doctors possess about these systems, to be able to model them. Therefore, qualitative modeling of biomedical systems, the inner functionings of which are either not well known or even totally unknown, is a truly important task that therefore deserves to be mentioned as a goal of its own.

Monitoring comprises several purposes, from the integration of a variety of signals coming from the patient or from other monitors (sensor fusion), to the incorporation of "smart alarm" algorithms that reduce the number of false alarms produced by the monitors, which is one of the major problems of actual monitors (Navabi *et al.*, 1991). Other monitors are enhanced by means of the incorporation of algorithms that permit the real–time scanning of one or more monitor signals detecting anomalies in these signals and/or are able to predict problems in the immediate future by some sort of extrapolation algorithms (Navabi *et al.*, 1991). Other related publications in this general area are: (Cheung and Hull, 1989), (Ciarroca, 1989), (Orr and Westenskow, 1989), (Uckun *et al.*, 1993), and (Barro *et al.*, 1990b).

There are aims that are usually more closely related to a specific qualitative methodology. This is the case of *therapeutic management*, *consultation systems*, and *decision support*, functions that are all usually supported by expert systems. Therapeutic management helps the clinician with suggesting a therapy to be followed by the patient, or changing a therapy that has been followed in the past. This is the next step after diagnosis. One such expert system is ONCOGAL (Mira *et al.*, 1991). An example of a decision support and consultation system is presented in (Fox and Frost, 1985), a system that is being used by primary care doctors for obtaining a second opinion.

In Table 2.1, the *imaging* aim is also included. This aim is not much related to the primary goals of this state–o–the–art survey, because it is basically focused on static images and therefore on edge detection and pattern matching problems. An example is the work by Feng (Feng *et al.*, 1991).

Classification is the process to decide, given a new input set, which output set is concerned. Recurrent neural networks are perfect for this kind of purpose (Navabi *et al.*, 1991; Schaltenbrand *et al.*, 1993), however, some results have also been reported using fuzzy sets (Barro *et al.*, 1990a; Pedrycz *et al.*, 1991). Signal classification has most frequently been applied to cardiology for the detection of a specific kind of anomaly in the ECG signal. The proper classification of such anomalies is very useful for diagnostic purposes. However,

signal classification has also been applied to other application areas, such as for the diagnosis of respiratory diseases (Navabi *et al.*, 1991), or the identification of sleep stages (Schaltenbrand *et al.*, 1993).

Current trends are also focused on an area that has been coined *causal explanation*. This means the ability of a given methodology to explain causal mechanisms, i.e., to be able to offer a complete description of the different steps trough which the reasoner has passed to obtain the final conclusion. This is very useful when dealing with biomedical applications since it provides more trust in the recommendation made by the system, as medical doctors are not usually inclined to blindly follow a recommendation made by an obscure algorithm that he or she doesn't really understand, and moreover, it may provide additional useful information. For example, an anaesthetist may like to know *why* he or she is supposed to increase the amount of Isuflurane in a given situation (e.g., because the systolic arterial pressure of the patient is way too high), rather then only being told by how much to increase it. The ability to offer causal explanations is directly dependent on the methodology in use. Expert systems, for instance, have the possibility to offer high–level explanations. This is not the case of neural networks, whose structure does not permit to give any kind of explanation. Inductive reasoning methodologies are somewhere in between. They offer the possibility to provide a general explanation of what is happening, but at a considerably lower level than expert systems.

Training is another area with lots of potential. Yet, educational systems have so far mostly been by–products of other research efforts. The reason is that, whereas hospitals have almost unlimited resources to conduct clinical research, the medical colleges fight with the same funding limitations as all the other colleges. It is thus much more difficult to attract research funding for propaedeutical purposes than for clinical needs. Yet, this is a very attractive area for future cooperation between educators in medicine and researchers in engineering and computer science.

2.4 Biomedical Applications

A wide range of biomedical problems has been tackled by means of qualitative reasoning methodologies. However, as has been shown in a previous section, one of the major application areas, in which a variety of different reasoning methodologies have been successfully used, is *cardiology*. The intensity of research in the field of cardiology can be explained by the fact that the heart

- Acid–Base Disorders	- Diabetes
- Anaesthesia	- Muscles
- Anaemia	- Nervous System
- Arterial Pressure	- Obstetrics
- Assisted Ventilation	- Primary Care
- Bacterium Infection	- Psychiatry
- Cancer Chemotherapy	- Renal Physiology
- Cardiology	- Sexual Dysfunctions
- Cirrhosis	- Sleep

Table 2.2: *Applications most frequently cited in literature*

is one of the most central organs of the human body. Heart problems are usually fulminant in the sense that death comes quickly, and there is often not much time to react, once the symptoms of the disease have set in. This is not the case in many of the other aspects of human health. It is therefore fairly easy to get funding for sensible clinical research in cardiology, and this is of course another motivating factor for research in this area. Yet another factor that helps to explain the large number of ongoing research activities in this area is the relative ease with which large amounts of data can be obtained from patients using non–intrusive measurement techniques.

The application areas that have been mentioned in the literature most frequently have been listed in Table 2.2 in alphabetic order.

Arterial Blood Pressure is another field that has been frequently studied using practically all of the qualitative reasoning methodologies presented in this chapter. The reason is that this application is simple enough to be used as a benchmark problem for all kinds of control techniques (Linkens, 1992), be it quantitative, e.g. of the PID type, or be it qualitative, e.g. using fuzzy techniques such as those proposed in (Ying *et al.*, 1990; Meier *et al.*, 1992; Isaka and Sebald, 1992), or using qualitative simulation (Kuipers and Kassier, 1985).

Another application area that has been widely studied using qualitative methodologies is *anaesthesia*. The anaesthetic field, as well as almost all the other applications from Table 2.2, can be treated from different points of view and with different scopes in mind, and depending on those, the complexity of the work to be accomplished may vary drastically. It is much simpler to regulate the mean arterial pressure through the anaesthetic dose than to try

to truly predict the anaesthetic depth of a patient using all of the available sensory information.

An attraction of this area is that a large number of results have been reported using different methodologies, such as expert systems (Linkens *et al.*, 1986; Linkens *et al.*, 1990), neural networks (Linkens and Rehman, 1992a), fuzzy inductive reasoning (Nebot *et al.*, 1993a), and fuzzy systems (Meier *et al.*, 1992). Therefore, it is possible to compare the performance of these methodologies and to assess their relative advantages and disadvantages at least for this one application area.

Qualitative physics methodologies have been applied to a fairly small subset of biomedical problems related to *acid–base disorders* (Coiera, 1990), to *assisted ventilation* (Uckun *et al.*, 1993), and to *cardiology* (Kuipers and Kassier, 1985). In a previous section, the major limitations of the three types of qualitative physics methodologies have been shown, limitations that hamper their applicability in the case of biomedical systems.

2.5 Conclusions

In this chapter, a general view of the use of qualitative analysis in biomedical engineering has been offered. It was shown that the first generation of research efforts relating to the use of qualitative modeling and simulation methodologies in biomedical systems did not make a significant impact on the biomedical sciences. This fact was explained to have been caused by several factors that were discussed in some detail in this chapter. In summary:

- The first generation tools primarily focused on automated diagnosis, leading to a rejection by the medical personnel.
- The level of knowledge that researchers have about most disease processes is relatively shallow.
- The human anatomy and physiology are highly complex and exhibit a large variability from one individual to the next.
- The cognitive processes that underlie medical decision making are still poorly understood.
- The knowledge bases required for meaningful decision making would be formidable in their size, yet knowledge is very hard to come by.

- Few clinicians are sufficiently familiar with computers to contribute to these research efforts in any significant way, or even to use the systems adequately once they have been designed.
- The legal implications of computer–assisted advice in the health care domain form a considerable barrier to its acceptance.

Current research in this area can be classified as *second generation* efforts, and hopefully a major subset of the causes that led to the rejection of the first generation tools can be overcome. In fact, several of these factors have already disappeared. For example, this is the case of the scope of automated diagnosis systems. Current research on the use of qualitative methodologies for diagnosis is much more in line with the needs of the medical personnel than this was the case with the first generation tools. Rather than trying to replace human decision making, the new systems are designed to aid the medical personnel in their tasks in an optimal manner. This has been outlined in the section on the "aims of qualitative research."

During the current decade, both the researchers, most of whom hold advanced degrees in either engineering or computer science, and the medical doctors, who are envisaged to be the end users of the tools to be developed, have become acutely aware that a more intensive level of communication and knowledge transfer between them is absolutely essential, if these research efforts are to amount to anything. This awareness has lead to in–depth studies of the needs and requirements of the physicians. One such study is for instance the work by Forsythe and co–workers (Forsythe *et al.*, 1992). The hope is therefore substantiated that some of the previous obstacles can be eliminated or at least alleviated in the second generation set of tools.

It can be seen already now that the second generation research efforts on qualitative modeling and simulation methodologies applied to biomedical systems are indeed making a significant impact on the biomedical domain, and this impact is expected to increase even further in the near future.

This doctoral thesis contributes to overcome several of the limitations that have previously prevented the use of qualitative reasoning in biomedicine from being successful and that have hampered the development of analytical tools for use in biomedicine as a whole. It does so by advocating the use of a new qualitative reasoning paradigm, the fuzzy inductive reasoning methodology.

Chapter 3

Fuzzy Inductive Reasoning Methodology

3.1 Introduction

Up to this point, we have looked at the state–of–the–art of various qualitative modeling and simulation techniques advocated in the literature. From here on, we shall concentrate on one particular technique, *Fuzzy Inductive Reasoning* (FIR), for the remainder of this dissertation.

The FIR methodology is still fairly new and has not yet been widely used. Therefore, it is presented in this chapter, in order to ensure the understandability of this thesis. The examples presented in the subsequent chapters of the thesis will demonstrate that the FIR methodology is indeed a very promising technique for dealing with biomedical applications.

The FIR methodology is composed of four main processes, namely *fuzzification*, *qualitative modeling*, *qualitative simulation*, and *defuzzification*. These four modules form the basis of the FIR methodology. However, other functions have been implemented as well, functions that offer additional features to the user.

The aim of this chapter is to provide a detailed account of the basic tool set within the overall FIR methodology. In recent years, several new techniques, tools, and features have been developed and added to the methodology, in order to enhance its reasoning power in the contexts of qualitatively dealing with complex non–linear technical systems, the automatic design of fuzzy

controllers, and the qualitative analysis of systems from the soft sciences.

Three Ph.D. students of the *Universitat Politècnica de Catalunya* (UPC) were involved in generating these additional tools for enhancing the FIR methodology. In particular, the author of this dissertation signs responsible for the research efforts related to the qualitative analysis of systems stemming from the soft sciences in general, and from biomedicine in particular.

The advances and new techniques developed to support this endeavor will be introduced along the chapters of this thesis, and their effectiveness will be demonstrated by means of realistically complex biomedical examples. The focus of the current chapter is on presenting a detailed description of the *FIR kernel*, i.e., those functions that are common to all three research directions.

Whereas a rudimentary FIR kernel had already been implemented by other researchers before either of the three dissertations even began (Cellier and Yandell, 1987; Li and Cellier, 1990; Cellier, 1991a), also the kernel functions needed many modifications and enhancements, before they could adequately and in a robust fashion be used to tackle complex problems. Three Ph.D. students of the *Universitat Politècnica de Catalunya* sign jointly responsible for upgrading and maintaining the FIR kernel modules.

3.2 The Methodology

The inductive reasoning methodology had originally been developed by George Klir (Klir, 1985) as a tool for general system analysis, to study the conceptual modes of behavior of systems. The inductive reasoning set of methods and algorithms forms a subset of Klir's *General System Problem Solving* (GSPS) framework, facets of which have been described in numerous of his publications starting in the seventies. Klir's research efforts along the lines of his GSPS framework are still ongoing.

The GSPS methodology distinguishes between different types of "systems" characterized by different levels of abstraction. In our own terminology, however, a *system* is the physical entity from which mathematical descriptions of varying abstraction, so–called *models*, can be derived. Using Bernard Zeigler's words, *the system is nothing but a potential source of data* (Zeigler, 1976). Consequently, Klir's "systems" will be called "models" in the remainder of this thesis.

The GSPS methodology distinguishes between an infinity of abstraction

levels. The most abstract model is the *base model*. It simply encodes knowledge about which facets of the real system are to be captured in the mathematical description. For most practical purposes, this knowledge consists in a declaration of a *set of variables* to be contained in the model.

The next higher (i.e., more refined, or less abstract) level along Klir's "epistemological hierarchy" is the *data model*. In order to climb the epistemological hierarchy ladder from the level of the base model to that of the data model, one or several experiments must be performed on the real system, whereby sensors are to be attached to physical quantities that are represented by some or all of the variables to be included in the model. The result is a bunch of data streams, or trajectories, which, at this point, still do not contain a description of any logical or causal relationship connecting these data streams to each other. Thus, the data model is characterized by a set of variables measured and recorded over time. Their only known relationship so far is their common time stamp.

In our own terminology, we shall call this the *raw data model*, because the data have not yet been processed in any way. The raw data model, in our implementation of the methodology, is represented by a real–valued matrix, whereby each column denotes one variable trajectory, i.e., the recording of the values of one variable as a function of time, whereas each row denotes one data record, i.e., a collection of the values of all variables with identical time stamp.

In order to proceed to higher levels along the epistemological hierarchy ladder, it will prove useful to preprocess these data. In our implementation of the methodology, the raw data model will be preprocessed into a *qualitative data model*, whereby each raw (quantitative) data value is being replaced by a qualitative triple. The details of this process will be explained in due course. Since the raw data model and the qualitative data model contain exactly the same information, GSPS does not distinguish between the two. Both are located at the same epistemological hierarchy level. Climbing up the hierarchy ladder inevitably implies, according to Klir, adding more information to the model. Since the transformation from the raw (quantitative) data model to its qualitative counterpart does not add any information to the model, Klir places them at the same hierarchy level.

Climbing up the ladder one rung further, we end up with the *behavior model*. The behavior model adds logical or causal relationships to subsets of the variables. Whereas before we did not know anything about the causal relationship between the recorded variables, they might even stem from entirely different physical objects for that matter, at the new level, this is no longer

possible. Now, we know which set of variables we must consult to infer knowledge about one or several other variables. All so–called input/output models are located at this hierarchical level.

The FIR methodology, a subset of the GSPS methodology, is located entirely at the hierarchical levels of the data and behavioral models. It deals with transformations within each of these levels, and with transitions between the two levels. The *fuzzification module* describes a transformation within the data model level, namely from the quantitative (raw) data model to its qualitative counterpart. The *qualitative modeling module* describes the step up the ladder from the data model to the behavioral model. This is accomplished by *induction*. The term induction is synonymous with climbing up the epistemological ladder, while deduction means descending it. The *qualitative simulation module* denotes the transition back down the ladder to the previous level, and the *defuzzification module* performs another transformation at the data model level.

Climbing the epistemological hierarchy ladder even further, we reach the rung of the *structural models*. Most deductively derived (classic) differential equation models are located at that level. Here, the causal relationships of the former behavioral models are concretized to explicit structural relationships between variables, i.e., formulae replacing mere tabulations. Finally, GSPS defines infinitely many rungs of so–called *meta–models* that are not further qualified in the GSPS architecture. In some publications, the first meta model level is characterized by variable structure models, i.e., by models that abruptly change their behavior as a consequence of a discrete event taking place (Uyttenhove, 1979). However, since the FIR methodology does not deal at all with these higher elevated rungs of the GSPS methodology, there is no need to explore their properties any further in this dissertation.

In the late seventies, a Ph.D. student of George Klir's: *Hugo Uyttenhove*, went about to implement a significantly large subset of the GSPS methodology under the name *Systems Approach Problem Solver*, abbreviated as SAPS (Uyttenhove, 1979). Unfortunately, the limited computer science tools available to Uyttenhove at that time did not lend themselves to a sufficiently flexible implementation of the GSPS concepts with the limited manpower resources available to a single Ph.D. student, and consequently, SAPS could never be used for anything but mere toy problems.

In the mid eighties, Cellier and his students went about to reimplement SAPS as a CTRL–C function library. The new implementation was called SAPS–II (Cellier and Yandell, 1987). CTRL–C provided all the matrix

manipulation capabilities necessary for processing the GSPS data structures in an elegant and convenient fashion, and the interactive programming environment of CTRL–C provided sufficient flexibility to enable the user to combine individual SAPS modules to ever more powerful building blocks.

Fuzzy measures were introduced into the GSPS methodology in the late eighties (Klir and Folger, 1988; Klir, 1989; Wang and Klir, 1992), and were incorporated into the SAPS–II toolkit by *DongHui Li*, a student of Francois Cellier (Li and Cellier, 1990).

Even more recently, SAPS–II has been demonstrated to be an effective tool for qualitatively studying the behavior of highly complex non–linear technical systems (Cellier *et al.*, 1992, 1994; de Albornoz and Cellier, 1993a, 1993b; Cellier and Mugica, 1992), as well as biomedical systems (Nebot *et al.*, 1993a, 1993b).

SAPS–II is currently available as either a CTRL–C library or a Matlab toolbox. It runs on any platform offering an implementation of either CTRL–C or Matlab and a Fortran compiler. No other facilities are needed.

The fuzzy inductive reasoning methodology is composed of four basic functions: *fuzzification* (fuzzy recoding), *qualitative modeling* (fuzzy optimization), *qualitative simulation* (fuzzy forecasting), and *defuzzification* (fuzzy regeneration).

The *fuzzy recoding module* converts quantitative values into qualitative triples. The first element of the triple is the class value, the second element is the fuzzy membership value, and the third element is the side value. The class value represents a coarse discretization of the original real–valued variable. The fuzzy membership value denotes the level of confidence expressed in the class value chosen to represent a particular quantitative value. Finally, the side value tells us whether the quantitative value is to the left or to the right of the peak value of the membership function. The side value, which is a specialty of our methodology since it is not commonly introduced in fuzzy logic, is responsible for preserving the complete knowledge in the qualitative triple that had been contained in the original quantitative value.

The *fuzzy optimal mask function* realizes the process of qualitative modeling. It is able to establish qualitative relationships between different variables of the model. It does so by a process of exhaustive search in the discrete search space of the class values.

Fuzzy simulation is performed by means of the fuzzy forecasting function,

which is able to predict future qualitative outputs (qualitative triples) from past similar experiences. Fuzzy simulation interpolates between previous occurrences of similar behavioral patterns, and uses the interpolated values to extrapolate the output variable across time.

Finally, the *fuzzy regeneration facility* implements the inverse process of the fuzzy recoding module. It converts qualitative triples back to quantitative values. Since fuzzy recoding preserves the complete information of the original quantitative value, an immediate cascade of a fuzzy recoding operation followed by a fuzzy regeneration operation restores the original signal without any error. This is a special feature of our particular dialect of fuzzy logic. Most fuzzy logic signals lose information in the process of fuzzification, information that cannot be retrieved by means of defuzzification.

In the following sections, an accurate description of each of these processes is provided. Also, a simple example is carried through this chapter, in order to explain how these modules are being used in practice.

3.2.1 Fuzzification

It had previously been explained that a transformation from quantitative values into qualitative triples is very useful for the purpose of inductive modeling. Any data fitting algorithm (and this is what inductive modeling is all about) invariably involves some sort of optimization procedure. Thus, inductive modeling applied to the original quantitative, i.e., real–valued, variables involves a search across an n–dimensional continuous search space. Such a search is invariable very time–consuming. By converting the quantitative values to qualitative triples, the search is simplified dramatically, since the search space gets reduced to the n–dimensional discrete search space of the class values. Using this approach, the class values are used for a fairly coarse optimization, whereas the fuzzy membership values are then used for the fine interpolation between neighboring class values, once the optimal class value has been found.

In the FIR methodology, the fuzzification process is accomplished by means of the *fuzzy recoding function*. Recoding denotes the process of converting a quantitative variable to a qualitative variable.

In most transformations from a quantitative to a qualitative space, some information is lost in the process. Obviously, a temperature value of 97^{o}F contains more information than the value 'hot.' Our fuzzy recoding technique

avoids this problem.

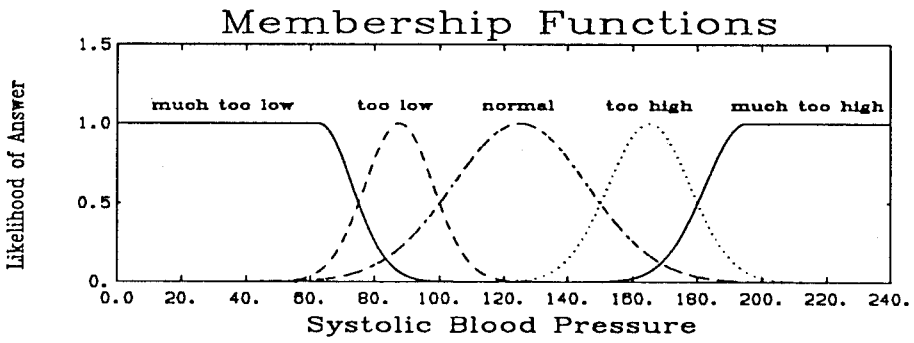


Figure 3.1: Membership Functions of the *Systolic Blood Pressure.*

Figure 3.1 shows the fuzzy recoding of a variable called "systolic blood pressure." For example, a quantitative systolic blood pressure of 135.0 is recoded into a qualitative class value of 'normal' with a fuzzy membership function value of 0.895, and a side function value of 'right' (since 135.0 is to the *right* of the maximum of the bell–shaped membership function that characterizes the class 'normal'). Thus, a single quantitative value is recoded into a *qualitative triple.* Any systolic blood pressure with a quantitative value between 100.0 and 150.0 will be recoded into the qualitative class value 'normal.' The fuzzy membership function denotes the value of the bell–shaped fuzzy membership curve that is associated with the selected class, read out at the point of the quantitative value. It is always a value between 0.5 and 1.0. Other fuzzification techniques make use of the tails of the membership functions to resolve ambiguity. They assign multiple class values and multiple membership values to a single quantitative value. Our own dialect of fuzzy logic handles the ambiguity issue differently. The *side value* is introduced as a third piece of information to eliminate the ambiguity, and the tails of the membership functions can thus be ignored.

Evidently, no information is lost in the process of fuzzy recoding. The qualitative triple contains exactly the same information as the original quantitative value, and it is thus possible to *regenerate* the quantitative value

from the qualitative triple precisely, i.e., without any error or uncertainty, at any point in time.

In the first implementation of fuzzy measures within SAPS–II (Li and Cellier, 1990), bell–shaped rather than the more commonly used triangular fuzzy membership functions had been implemented, just as those shown in Figure 3.1. However in the mean time, triangular membership functions have been added as an additional option.

The bell–shaped membership functions can be expressed mathematically using the equation:

$$Memb_i = \exp(-\tau_i \cdot (x - \mu_i)^2) \tag{3.1}$$

where x is the continuous variable to be recoded, μ_i is the algebraic mean between two neighboring landmarks (*landmarks* being the points where the class value changes), and τ_i is determined such that the membership function, $Memb_i$, degrades to a value of 0.5 at both of these landmarks.

At this point, the question needs to be raised, how many discrete levels (i.e., classes) should be selected for each state variable, and where the borderlines (landmarks) that separate neighboring regions from each other are to be drawn.

From statistical considerations, it is known that in any cluster analysis, each legal discrete state should be recorded at least five times (Law and Kelton, 1990). Thus, a relation exists between the total number of legal states and the number of data points required to base the modeling effort upon:

$$n_{\text{rec}} \geq 5 \cdot n_{\text{leg}} = 5 \cdot \prod_{\forall i} k_i \tag{3.2}$$

where n_{rec} denotes the total number of recordings, i.e., the total number of observed states, n_{leg} denotes the total number of distinct legal states, i is an index that loops over all variables in the state, and k_i denotes the number of levels that the i^{th} variable can assume. The number of variables is usually given, and the number of recordings is frequently predetermined. In such a case, the optimum number of levels, n_{lev}, of all variables can be found from the following equation:

$$n_{\text{lev}} = \text{round}\ (\ \sqrt[n_{\text{var}}]{n_{\text{rec}}/5}\ ) \tag{3.3}$$

assuming that all variables are classified into the same number of levels. For reasons of symmetry, an odd number of levels is often preferred over an even number of levels. Abnormal states ('too low,' 'too high,' and 'much too low,' 'much too high') are grouped symmetrically about the 'normal' state.

The number of levels chosen for each variable determines the expressiveness and predictiveness of the qualitative model. The *expressiveness* of a qualitative model is a measure of the information content that the model provides. The *predictiveness* of a qualitative model is a measure of its forecasting power, i.e., it determines the length of time over which the model can be used to forecast the future behavior of the underlying system (Li and Cellier, 1990).

If all variables are recoded into exactly one level, the qualitative model exhibits only one legal state. Such a model is called a *null model*. It is able to predict the future behavior of the underlying system perfectly over an infinite time span (within the framework of its model resolution). Yet the prediction does not provide any useful information. Thus, the null model is characterized by an infinitely high predictiveness and a zero expressiveness.

On the other hand, if every variable is recoded into 1000 levels, the system exhibits a high number of legal states. The expressiveness (i.e., resolution) of such a model will be excellent. Each state contains a large amount of valuable information about the real system. Yet the predictiveness of this model will be miserable unless an extremely large base of observed data is available. In all likelihood, this model cannot be used to predict the behavior of the real system for even a single time step into the future. Consequently, a compromise must be reached.

The number of levels chosen for each variable influences directly the computational complexity of the inference stage. Traditional fuzzy systems usually require between seven and 13 classes for each variable (Aliev *et al.*, 1992; Maiers and Sherif, 1985). An exhaustive search in such a high–dimensional discrete search space would be very expensive, and the number of classes should therefore be reduced, if possible, to help speed up the optimization. It was shown in (Mugica and Cellier, 1993) that the selected fuzzy inferencing technique makes it possible to reduce the number of levels down to usually three or five, a number confirmed by several practical applications (de Albornoz and Cellier, 1993a, 1993b; Cellier, 1991c; Vesanterä and Cellier, 1989).

Once the number of levels of each variable has been selected, the landmarks must be chosen to separate neighboring regions from each other. There are

several ways to find a meaningful set of landmarks. The most effective way is based on the idea that the expressiveness (or information contents) of the model will be maximized if each level is observed equally often. In order to distribute the observed trajectory values of each variable equally among the various levels, they are sorted into ascending order, the sorted vector is then split into n_{lev} segments of equal length, and the landmarks are chosen anywhere between the extreme values of neighboring segments, e.g., using the arithmetic mean values of neighboring observed data points in different segments.

Let us now introduce a simple example to show how fuzzification can be accomplished using the SAPS–II toolbox in the Matlab environment. Our example is a linear system described by the following equations:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\cdot\mathbf{x}+\mathbf{b}\cdot u\\ &= \begin{pmatrix} 0 & 1 & 0\\ 0 & 0 & 1\\ -2 & -3 & -4\end{pmatrix}\cdot\mathbf{x}+\begin{pmatrix}0\\0\\1\end{pmatrix}\cdot u\end{aligned}$$

$$\begin{aligned}y &= \mathbf{C}\cdot\mathbf{x}+\mathbf{d}\cdot u\\ &= \begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\cdot\mathbf{x}+\begin{pmatrix}0\\0\\0\end{pmatrix}\cdot u\end{aligned}$$

Computing the step response, it can be concluded that the settling time is $t_l \approx 6$ sec (Cellier, 1991a). Therefore, if it is decided to use a "mask depth" of three covering the slowest time constant in the system, the communication interval should be $\delta t \approx 3$ sec. The term "mask" has not yet been introduced, but this will happen shortly.

A binary random sequence is used as input signal to the system, in order to excite it in an optimal manner at all relevant frequencies. It was decided to recode each of the output states into three levels (the input is already binary), and therefore, the number of legal states can be computed as:

$$n_{\text{leg}} = \prod_{\forall i} k_i = 2\cdot 3\cdot 3\cdot 3 = 54 \tag{3.4}$$

and the required number of recordings is:

$$n_{\text{rec}} = 5 \cdot n_{\text{leg}} = 270 \tag{3.5}$$

The following Matlab code simulates the system over 300 communication intervals, in order to obtain data to work with.

```
≫  t = 0 : 3 : 897;
≫  u = round(rand(t));
≫  x0 = zeros(3, 1);
≫  [y, x] = lsim(a, b, c, d, u, t, x0);
```

The input/output data obtained from the quantitative simulation run constitute the raw data model, *meas*.

```
≫  meas = [u', y];
```

The raw data model is a matrix with four columns (four variables) and 300 rows (300 data records). The first variable (column) is already binary, but the other three variables need to be recoded into three levels each. The landmarks are computed using the sorting algorithm outlined earlier:

```
≫  for i = 2 : 4,
≫     [mi, indx] = sort(meas(:, i));
≫     m(:, i) = mi;
≫  end
≫  lm = [m(1, :)
≫          0.5 * (m(100, :) + m(101, :))
≫          0.5 * (m(200, :) + m(201, :))
≫          m(300, :)];
```

The matrix m contains the sorted variable vectors, whereby sorting was done for each variable separately. The first column of m is a dummy column. The landmarks, lm, are stored in a 4×4 matrix, where each column denotes a variable, and each row denotes a landmark. The four landmarks for each variable are: the smallest variable value ever recorded, the boundary between the classes 'low' and 'medium,' the boundary between the classes 'medium' and 'high,' and the largest variable value ever recorded. Again, the first column of lm is a dummy column.

Once the landmarks have been computed, we are ready to use the SAPS–II recoding algorithm to fuzzify (discretize) the three output variables.

$\gg$ $to = 1:3;$
$\gg$ **for** $i = 2:4,$
$\gg$ $from = [lm(1:3,i), lm(2:4,i)]';$
$\gg$ $[c, m, s] = \text{recode}(meas(:,i), 'fuzzy', from, to);$
$\gg$ $class(:,i) = c; \quad memb(:,i) = m; \quad side(:,i) = s;$
$\gg$ **end**

The first argument of the *recode* function is a column vector containing a quantitative trajectory to be recoded into a qualitative episode. The second argument denotes the method of recoding, in our case *fuzzy*, and the third and fourth arguments contain the mapping information. The desired shape of the membership functions (bell–shaped or triangular) is stored in a global variable, rather than being passed on to the *recode* function as an argument.

The *from* matrix is a $2 \times n_{\text{lev}}$ matrix, where n_{lev} denotes the number of classes. It contains the rearranged landmarks. Each column of the *from* matrix selects a range in the original quantitative domain to be mapped into the discrete value stored in the corresponding position of the *to* vector. Thus, the *from* matrix and *to* vector for the systolic blood pressure of Figure 3.1 would be:

$$from = \begin{pmatrix} 0.0 & 75.0 & 100.0 & 150.0 & 180.0 \\ 75.0 & 100.0 & 150.0 & 180.0 & 999.9 \end{pmatrix}$$

$$to = (1 \quad 2 \quad 3 \quad 4 \quad 5)$$

In the linear system example presented above, the mapping information is obtained from the previously computed landmarks.

Notice that the class values, in SAPS–II, are integers rather than mnemonics. Thus, we use the values '1,' '2,' and '3' to represent the classes 'low,' 'medium,' and 'large.'

The same holds for the side function values. Rather than using the mnemonics 'left,' 'center,' and 'right,' SAPS–II encodes the side values as '-1', '0', and '$+1$.'

Since the input signal of the linear example is already binary, it is only necessary to add its membership function value (always 1.0) and its side

function value (always '0'). This is accomplished using the following Matlab code:

```
≫  [nrec, nvar] = size(meas);
≫  for j = 1 : nrec,
≫    if meas(j, 1) == 0
≫      class(j, 1) = 1;
≫      memb(j, 1) = 1.0;
≫      side(j, 1) = 0;
≫     else
≫      class(j, 1) = 2;
≫      memb(j, 1) = 1.0;
≫      side(j, 1) = 0;
≫    end
≫  end
```

This completes the description of recoding as applied to the linear system example.

3.2.2 Qualitative Modeling

By now, the quantitative trajectory behavior has been recoded into a qualitative episodical behavior. In SAPS–II, the episodical behavior is stored in a *qualitative data model*. It consists of three matrices of identical size, one containing the class values, the second storing the membership information, and the third recording the side values. Each column represents one of the observed variables, and each row denotes one time point, i.e., one recording of all variables, or one recorded state.

The class values are in the set of legal levels that each variable can assume. They are all positive integers, usually in the range from '1' to '5,' as SAPS–II uses integers in place of symbolic values to represent qualitative levels.

In the FIR methodology, the fuzzy modeling process is performed by means of the *fuzzy optimal mask function*. It optimizes the predictiveness of the model by performing an exhaustive search in the discrete space of the class values. The details of how this is accomplished are presented in this section.

3.2.2.1 Masks as Qualitative Models

How does the episodical behavior support the identification of a qualitative model of a given system for the purpose of forecasting its future behavior for any given input stream?

A continuous trajectory behavior has been recorded and is available for modeling. The inputs of the real system and a set of measurable outputs have been recorded as functions of time and are stored in the trajectory behavior. The trajectory behavior can be separated into a set of input trajectories, u_i, concatenated from the right with a set of output trajectories, y_i, as shown in the following example containing two inputs and three outputs:

$$\begin{array}{l} time \\ 0.0 \\ \delta t \\ 2 \cdot \delta t \\ 3 \cdot \delta t \\ \vdots \\ (n_{\text{rec}} - 1) \cdot \delta t \end{array} \begin{array}{c} \begin{array}{ccccc} u_1 & u_2 & y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \end{array} \tag{3.6}$$

where n_{rec} is the number of data records, and δt is the sampling interval.

In order to avoid possible ambiguities, it is defined that the terms "input" and "output," when used in this chapter without further qualifier, shall always refer to the input and output variables of the subsystem to be modeled by the qualitative reasoner.

In the process of modeling, it is desired to discover finite automata relations among the recoded variables that make the resulting state transition matrices as deterministic as possible. If such a relationship is found for every output variable, the behavior of the system can be forecast by iterating through the state transition matrices. The more deterministic the state transition matrices are, the higher is the likelihood that the future system behavior will be predicted correctly.

A possible relation among the qualitative variables for this example could be of the form:

$$y_1(t) = \tilde{\mathrm{f}}(y_3(t - 2\delta t), u_2(t - \delta t), y_1(t - \delta t), u_1(t)) \tag{3.7}$$

where $\tilde{f}$ denotes a qualitative relationship. Notice that $\tilde{f}$ does not stand for any (known or unknown) explicit formula relating the input arguments to the output argument, but only represents a generic causality relationship that, in the case of the FIR methodology, will be encoded in the form of a tabulation of likely input/output patterns, i.e., a state transition table. In SAPS–II, Equation 3.7 is represented by the following matrix:

$$\begin{array}{c} {}_t\backslash^x \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{ccccc} u_1 & u_2 & y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & -1 \\ 0 & -2 & -3 & 0 & 0 \\ -4 & 0 & +1 & 0 & 0 \end{pmatrix} \end{array} \tag{3.8}$$

The negative elements in this matrix are referred to as m–inputs. M–inputs denote input arguments of the qualitative functional relationship. They can be either inputs or outputs of the subsystem to be modeled, and they can have different time stamps. The above example contains four m–inputs. The sequence in which they are enumerated is immaterial. They are usually enumerated from left to right and top to bottom. The single positive value denotes the m–output. The terms m–input and m–output are used in order to avoid a potential confusion with the inputs and outputs of the plant. In the above example, the first m–input, i_1, corresponds to the output variable y_3 two sampling intervals back, $y_3(t-2\delta t)$, whereas the second m–input refers to the input variable u_2 one sampling interval into the past, $u_2(t-\delta t)$, etc.

In the FIR methodology, such a representation is called a *mask*. A mask denotes a dynamic relationship among qualitative variables. A mask has the same number of columns as the episodical behavior to which it should be applied, and it has a certain number of rows, the *depth* of the mask.

The mask can be used to flatten a dynamic relationship out into a static relationship. It can be shifted over the raw data matrix, the selected m—inputs and m–output can be extracted from the raw data, and they can be written next to each other in one row of the so–called *input/output matrix*. The selected m—inputs and m–output are those that are visible through the holes of the mask, when it is in a specific position. This is the reason for the terminology: the *mask*. Figure 3.2 illustrates this process.

After the mask has been applied to the qualitative data model, the formerly dynamic episodical behavior has become static, i.e., the relationships are now contained within single rows.

Each row of the input/output matrix is called a *state* of the system. A

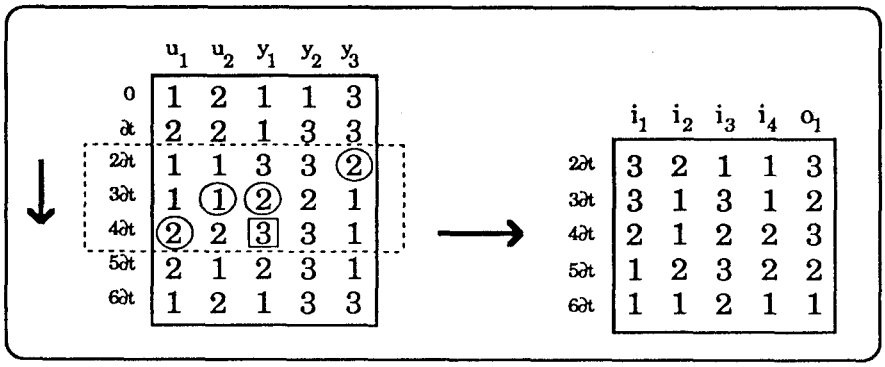


Figure 3.2: Flattening of Dynamic relationships by Use of a *Mask*.

state consists of an *input state* and an *output state*. The input state denotes the vector of values of all the m–inputs belonging to the state, and the output state is the value of the single m–output of the state. The set of all possible states is referred to as the set of *legal states* of the qualitative model.

The number of rows of the mask matrix is called the *depth* of the mask. Neighboring rows of the mask represent neighboring data records. Thus, a mask of depth n covers a time interval of $\Delta t = (n-1) \cdot \delta t$ time units.

How should the time distance between two logged entries of the trajectory behavior, δt, and the time span to be covered by the mask, Δt, be chosen? Experience has shown that the mask should cover the largest time constant, t_l, and that the sampling rate, δt, should be no larger than half the shortest time constant, t_s, of the system to be captured by the model, thus:

$$\Delta t \geq t_l \quad ; \quad \delta t \leq \frac{t_s}{2} \tag{3.9}$$

The *depth* of the mask can then be computed as follows:

$$depth = \text{round}(\frac{\Delta t}{\delta t}) + 1 \tag{3.10}$$

but this ratio should not be much larger than three or four, otherwise the inductive reasoner won't work very well, since the computing effort grows factorially with the size of the mask. In later chapters of the dissertation, it will be shown how this limitation can be overcome. An extensive analysis of the effects that an incorrect selection of these parameters have on the prediction is provided in Chapter 7.

In order to determine the two time constants of interest, several techniques can be used. If the physical system itself is available for experimentation, a Bode diagram of the system to be modeled can be obtained. This enables to determine the eigen frequencies of the system, and in particular, the smallest and largest eigen frequencies. The smallest eigen frequency ω_{low} is the smallest frequency, at which the tangential behavior of the amplitude of the Bode diagram changes by -20 dB/decade, and the largest eigenvalue ω_{high} is the highest frequency where this happens. The largest time constant, t_l, and the shortest time constant, t_s, of the system can then be computed as follows:

$$t_l = \frac{2\pi}{\omega_{low}} \quad ; \quad t_s = \frac{2\pi}{\omega_{high}} \tag{3.11}$$

If an analytical (quantitative) model of the system under investigation is available, the relevant time constants can be read out from it. If only time series are available that may have been measured earlier by someone else, or if the physical system is not open to free experimentation, such as in the case of a biomedical system involving humans, spectra of the input and output signals can be determined, but the information obtained from those may be deceiving. Ultimately, the modeler may have to rely on expert opinion as to what these time constants may be. In biomedicine, this is frequently the case.

3.2.2.2 Determination of the Optimal Mask

How is a mask found that, within the framework of all allowable masks, represents the most deterministic state transition matrix? This mask will optimize the predictiveness of the model. In SAPS–II, the concept of a *mask candidate matrix* has been introduced. A mask candidate matrix is an ensemble of all possible masks from which the best is chosen by a mechanism of exhaustive search. The mask candidate matrix contains '-1' elements where

the mask has a potential m–input, a '+1' element where the mask has its m–output, and '0' elements to denote forbidden connections. Thus, a good mask candidate matrix to determine a predictive model for variable y_1 in the previously introduced five–variable example might be:

$$\begin{array}{c} {}_t\backslash^x \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{ccccc} u_1 & u_2 & y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & +1 & 0 & 0 \end{pmatrix} \end{array} \tag{3.12}$$

Corresponding mask candidate matrices are used to find predictive models for y_2 and y_3. In all three mask candidate matrices, the instantaneous values of the other two output variables are blocked out in order to prevent *algebraic loops* to occur between the output variables that are to be estimated.

In SAPS–II, the *foptmask* routine determines the optimal mask from the class value matrix, the fuzzy membership matrix, a mask candidate matrix, and a parameter that limits the maximum tolerated *mask complexity*, i.e., the largest number of non–zero elements that the mask may contain. *Foptmask* searches through all legal masks of complexity two, i.e., all masks with a single m–input and finds the best one; it then proceeds by searching through all legal masks of complexity three, i.e., all masks with two m–inputs and finds the best of those; and it continues in the same manner until the maximum complexity has been reached. In all practical examples, the quality of the mask will first grow with increasing complexity, then reach a maximum, and then decay rapidly. A good value for the maximum complexity is usually five or six.

Each of the possible masks is compared to the others with respect to its potential merit. The optimality of the mask is evaluated with respect to the maximization of its forecasting power.

The *Shannon entropy measure* is used to determine the *uncertainty* associated with forecasting a particular output state given any legal input state. The Shannon entropy relative to one input state is calculated from the equation:

$$H_i = \sum_{\forall o} \mathrm{p}(o|i) \cdot \log_2 \mathrm{p}(o|i) \tag{3.13}$$

where $\mathrm{p}(o|i)$ is the "conditional probability" of a certain m–output state o to occur, given that the m–input state i has already occurred. The term

probability is meant in a statistical rather than in a true probabilistic sense. It denotes the quotient of the observed frequency of a particular state divided by the highest possible frequency of that state.

The overall entropy of the mask is then computed as the sum:

$$H_m = -\sum_{\forall i} \mathrm{p}(i) \cdot H_i \tag{3.14}$$

where $\mathrm{p}(i)$ is the probability of that input state to occur. The highest possible entropy $H_{\max}$ is obtained when all probabilities are equal, and a zero entropy is encountered for relationships that are totally deterministic.

A normalized overall entropy reduction H_r is defined as:

$$H_r = 1.0 - \frac{H_m}{H_{\max}} \tag{3.15}$$

H_r is obviously a real number in the range between 0.0 and 1.0, where higher values usually indicate an improved forecasting power. The masks with highest entropy reduction values generate forecasts with the smallest amounts of uncertainty.

The fuzzy membership associated with the value of a qualitative variable is a *measure of confidence*. In the computation of the input/output matrix, a confidence value can be assigned to each row. The confidence of a row of the input/output matrix is the joint membership of all the variables associated with that row (Li and Cellier, 1990).

The joint membership of i membership functions is defined in SAPS–II as the smallest individual membership:

$$Memb_{\text{joint}} = \bigcap_{\forall i} Memb_i = \inf_{\forall i}(Memb_i) \stackrel{\text{def}}{=} \min_{\forall i}(Memb_i) \tag{3.16}$$

The confidence vector indicates how much confidence can be expressed in the individual rows of the input/output matrix.

The *basic behavior* of the input/output model can now be computed. It is defined as an ordered set of all observed distinct states, together with a measure of confidence of each state. Rather than counting the observation frequencies (as would be done in the case of a probabilistic measure), the

individual confidences of each observed state are accumulated. If a state has been observed more than once, more and more confidence can be expressed in it. Thus, the individual confidences of each observation of a given state are simply added together to compute the cumulative membership.

This is a peculiar and rather uncommon choice. Most fuzzy algorithms either use the *smallest* individual membership value to compute joint membership (as we do in SAPS–II), and the *largest* individual membership value to denote cumulative membership:

$$Memb_{\text{joint}} = \bigcap_{\forall i} Memb_i = \inf_{\forall i}(Memb_i) \stackrel{\text{def}}{=} \min_{\forall i}(Memb_i)$$

$$Memb_{\text{cumul}} = \bigcup_{\forall i} Memb_i = \sup_{\forall i}(Memb_i) \stackrel{\text{def}}{=} \max_{\forall i}(Memb_i)$$

or alternatively, they use the sum of the individual membership functions to denote accumulation (as we do in SAPS–II), and then they use the product to denote intersection (or joining):

$$Memb_{\text{joint}} = \prod_{\forall i} Memb_i$$

$$Memb_{\text{cumul}} = \sum_{\forall i} Memb_i$$

There is no deep reason behind our unconventional choice. It simply resulted from experimentation. It led to the best mask selections in a fairly large number of experiments.

In order to be able to use the Shannon entropy, which is a *probabilistic* measure of information content, in the computation of the fuzzy optimal mask, the accumulated confidences must first be converted back to values that can be interpreted as conditional probabilities. To this end, the confidences of all states containing the same m–input state are added together, and the confidence of each of them is then divided by this sum. The resulting normalized confidences can be interpreted as conditional probabilities.

Application of the Shannon entropy to a confidence measure is a somewhat questionable undertaking on theoretical grounds, since the Shannon entropy

was derived in the context of probabilistic measures only. For this reason, some scientists prefer to replace the Shannon entropy by other types of performance indices (Klir, 1989; Shafer, 1976), which have been derived in the context of the particular measure chosen. However, from a practical point of view, numerous simulation experiments have shown that the Shannon entropy works satisfactorily also in the FIR context.

One problem still remains. The size of the input/output matrix increases as the complexity of the mask grows, and consequently, the number of legal states of the model grows quickly. Since the total number of observed states remains constant, the frequency of observation of each state shrinks rapidly, and so does the predictiveness of the model. The entropy reduction measure does not account for this problem. With increasing complexity, H_r simply keeps growing. Very soon, a situation is encountered where every state that has ever been observed has been observed precisely once. This obviously leads to a totally deterministic state transition matrix, and H_r assumes a value of 1.0. Yet the predictiveness of the model will be dismal, since in all likelihood already the next predicted state has never before been observed, and that means the end of forecasting. Therefore, this consideration must be included in the overall quality measure.

It was mentioned earlier that, from a statistical point of view, every state should be observed at least five times (Law and Kelton, 1990). Therefore, an *observation ratio*, O_r, is introduced as an additional contributor to the overall quality measure (Li and Cellier, 1990):

$$O_r = \frac{5 \cdot n_{5\times} + 4 \cdot n_{4\times} + 3 \cdot n_{3\times} + 2 \cdot n_{2\times} + n_{1\times}}{5 \cdot n_{\text{leg}}} \tag{3.17}$$

where:

n_{leg} = number of legal m–input states;
$n_{1\times}$ = number of m–input states observed only once;
$n_{2\times}$ = number of m–input states observed twice;
$n_{3\times}$ = number of m–input states observed thrice;
$n_{4\times}$ = number of m–input states observed four times;
$n_{5\times}$ = number of m–input states observed five times or more.

If every legal m–input state has been observed at least five times, O_r is equal to 1.0. If no m–input state has been observed at all (no data are available), O_r is equal to 0.0. Thus, O_r can also be used as a quality measure.

The overall *quality of a mask*, Q_m, is then defined as the product of its uncertainty reduction measure, H_r, and its observation ratio, O_r:

$$Q_m = H_r \cdot O_r \tag{3.18}$$

The *optimal mask* is the mask with the largest Q_m value.

Notice that the observation ratio does not influence the *quality* of a forecast if it is possible to make a forecast at all. It only influences the *likelihood* that a forecast can indeed be made. In other words, higher complexity masks with a large entropy reduction value but with smaller overall quality (due to their high complexity) will usually provide excellent forecasts ... if they are able to produce forecasts at all.

Let us use the linear example described in the fuzzification section to show how the optimal mask function is performed using SAPS–II in the Matlab environment. In that example, three optimal masks should be computed, one for each output variable. The following code shows the optimal mask analysis for the first output. As has been mentioned earlier, the first 270 rows of the raw data matrix are used for model identification. The remaining 30 rows will be use for model validation.

$\gg$ $cclass = class(1:270,:);$
$\gg$ $mmemb = memb(1:270,:);$
$\gg$ $mcan = -\text{ones}(3,4);$
$\gg$ $mcan(3,2:4) = [1,0,0];$
$\gg$ $[mask, hm, hr, qm, mhis] = \text{foptmask}(cclass, mmemb, mcan, 5);$

In SAPS–II, the *foptmask* function returns: the overall best mask found in the optimization, *mask*; the row vector that contains the Shannon entropies of the best masks for every considered complexity, H_m; the row vector containing the corresponding uncertainty reduction measures, H_r; and yet another row vector listing the quality measures, Q_m, of these suboptimal masks. Finally, *foptmask* also returns the *mask history matrix*, m_{his}, a matrix that consists of a horizontal concatenation of all suboptimal masks. One of these masks is the optimal mask, which, for reasons of convenience, is also returned separately in the first output argument, *mask*.

3.2.3 Qualitative Simulation

Once an optimal mask has been determined, it can be applied to the given class value matrix resulting in a particular input/output matrix. Since the input/output matrix contains functional relationships within single rows, the rows of the input/output matrix can now be sorted in alphanumerical order. The result of this operation is called the *behavior matrix* of the system. The behavior matrix is a finite state machine. For each input state, it shows which output is most likely to be observed.

Forecasting has now become a straightforward procedure. The mask is simply shifted further down beyond the end of the raw data matrix, the values of the m–inputs are read out from the mask, and the behavior matrix is used to determine the future value of the m–output, which can then be copied back into the raw data matrix. In fuzzy forecasting, it is essential that, together with the class value of the output, also fuzzy membership and side values are forecast. Thus, fuzzy forecasting predicts an entire qualitative triple from which a quantitative variable can be regenerated whenever needed.

In fuzzy forecasting, the membership and side functions of the new input state are compared with those of all previous recordings of the same input state contained in the behavior matrix. The one input state with the most similar membership and side functions is identified. For this purpose, a *normalization function*:

$$p_i = class_i + side_i * (1 - Memb_i) \tag{3.19}$$

is computed for every element of the new input state. The p_i values are quantitative (real–valued) variables that can be used to represent the relative magnitude of a particular qualitative triple. However, they are *not* regenerations of the original quantitative signals. They are normalized variables. Irrespective of whether an original signal was very small, ranging from -10^{-15} to $+10^{-14}$, or very large, ranging from 10^6 to 10^{12}, the corresponding p_i signal ranges *exactly* from 0.5 to 1.5 for values in class '1,' from 1.5 to 2.5 for values in class '2,' etc. Consequently, different p_i signals can be compared to each other or can be summed up, without weighing them relative to each other, something that would not be meaningful using the original or regenerated signals. The normalization function *is* a transformation from a qualitative triple to a quantitative variable, but this variable lives in a different space from the original quantitative variable.

The p_i values corresponding to the different variables of an input state are then concatenated to form the vector:

$$\mathbf{p} = [p_1, p_2, \ldots, p_j] \tag{3.20}$$

assuming, the state contains j m–inputs. We call the vector $\mathbf{p}$ the *norm image* of the original input state.

The analysis proceeds by computing norm images for every previous recording of the same input state. Let us call these vectors $\mathbf{p_k}$. Every $\mathbf{p_k}$ vector is a little different, since only the class values of the recorded input states are identical, but not their membership or side function values.

Finally, the $\mathcal{L}_2$ norms of the differences between the $\mathbf{p}$ vector representing the new norm image and the $\mathbf{p_k}$ vectors representing all previous recordings of the same input state are computed:

$$d_k = \|\mathbf{p} - \mathbf{p_k}\|_2 \tag{3.21}$$

and the previous recording with the smallest $\mathcal{L}_2$ norm is identified. Its *output* and *side* values are then used as forecasts for the *output* and *side* values of the current state.

Forecasting of the new membership function value is done a little differently. Here, the five previous recordings with the smallest $\mathcal{L}_2$ norms are used (if at least five such recordings are found in the behavior matrix), and a distance–weighted average of their fuzzy membership functions is computed and used as the forecast for the fuzzy membership function of the current state.

This is done in the following manner. Absolute weights are computed using one of two formulae. If none of the five smallest distance functions, d_k is exactly equal to zero, we use the equation:

$$w_{\mathrm{abs}_k} = \frac{(d_{\max}^2 - d_k^2)}{d_{\max} \cdot d_k} \tag{3.22}$$

where the index k loops over the five closest neighbors, and $d_i \leq d_j \quad , \quad i < j$; $d_{\max} = d_5$.

Evidently, the above formula will not work if any of the d_k values is zero, since this leads to a singularity. In this situation, the modified equation:

$$w_{\text{abs}_k} = \begin{cases} 0.0 \ ; & d_k \neq 0.0 \\ \\ 1.0 \ ; & d_k = 0.0 \end{cases} \tag{3.23}$$

is being used instead.

The idea behind these formulae is that, if one of the previous observations leads to a very small distance function, its weight should dominate the computation, yet if all distance functions are equally large, we should make use of an arithmetic mean between the previous distance functions.

Using the sum of the five absolute weights:

$$s_w = \sum_{\forall k} w_{\text{abs}_k} \tag{3.24}$$

it is possible to compute relative weights:

$$w_{\text{rel}_k} = \frac{w_{\text{abs}_k}}{s_w} \tag{3.25}$$

The relative weights are numbers between 0.0 and 1.0, and their sum always equals 1.0. Thus, the relative weights can be interpreted as percentages. Using this idea, the membership function of the new output can be computed as a weighted sum of the membership functions of the outputs of the previously observed five nearest neighbors:

$$Memb_{out_{new}} = \sum_{\forall k} w_{rel_k} \cdot Memb_{out_k} \tag{3.26}$$

In most of the experiments made in the past, it could be noticed that the fuzzy forecasting function computed using this algorithm generated a more accurate forecast than when using a probabilistic forecasting function (another option available within SAPS–II).

Let us now check how the simulation process for the linear example would be encoded:

$\gg$ $oc = class(1:270,:);$
$\gg$ $om = memb(1:270,:);$
$\gg$ $os = side(1:270,:);$

```
≫  nc = class(271 : 300, 1);
≫  nm = memb(271 : 300, 1);
≫  ns = side(271 : 300, 1);
≫  for i = 1 : 30,
≫     j = 270 + i − 1;   k = j + 1;
≫     ic = [oc; nc(i), 0, 0, 0];
≫     im = [om; nm(i), 1.0, 1.0, 1.0];
≫     is = [os; ns(i), 0, 0, 0];
≫     [fc1, fm1, fs1] = fforecast(ic, im, is, mask1, j);
≫     [fc2, fm2, fs2] = fforecast(ic, im, is, mask2, j);
≫     [fc3, fm3, fs3] = fforecast(ic, im, is, mask3, j);
≫     oc = [oc; nc(i), fc1(k, 2), fc2(k, 3), fc3(k, 4)];
≫     om = [om; nm(i), fm1(k, 2), fm2(k, 3), fm3(k, 4)];
≫     os = [os; ns(i), fs1(k, 2), fs2(k, 3), fs3(k, 4)];
≫  end
```

The qualitative simulation operates in the following way: We loop over the 30 steps of the desired forecast. In each step, the SAPS–II routine *fforecast* is called thrice, once with each of the three optimal masks to predict one new qualitative triple at a time. At the end of the loop, the three predicted triples are copied back into the qualitative model, and the pointer is shifted down one row. After the 30 steps are completed, 30 new qualitative triples have been forecast for each of the three output variables.

Initially, the *class*, *membership*, and *side* values of the first 270 steps are copied as past data into the matrices *oc*, *om*, and *os*, where the character "*o*" stands for *old*. These three matrices together make up the past history data base. The last 30 rows of the first column are then copied into the vectors *nc*, *nm*, and *ns*, where the letter "*n*" stands for *new*. These three vectors constitute the future inputs.

After these preparatory steps, we start with the loop. The matrices *ic*, *im*, and *is* are the three past data matrices concatenated from below with one new row containing the next input value in the first column, and arbitrary values in the second to fourth columns. These matrices are the first three input arguments of routine *fforecast*. The fourth input argument is the optimal mask to be used during the prediction, and the fifth and last input argument denotes the number of past data rows within the three data matrices. Upon return, fc_i is the same matrix as *ic*, but with the new prediction filled into one of the previously unused spots of the last row of the class value matrix, fm_i is the updated *im* matrix, and fs_i is the augmented *is* matrix.

At the end of the loop, the three old data matrices are updated to include one new row at the bottom, and the same procedure begins again one sampling interval into the future.

One of the most significant advantages of SAPS–II in comparison with other qualitative methodologies is its *intrinsic model validation mechanism* that ensures that SAPS–II will decline to predict anything that is not justifiable given the available facts. This is accomplished by computing a *measure of forecasting quality*, Q_f. The measure of forecasting quality is defined as the *cumulative output probability obtained during the forecasting process*.

SAPS–II not only maintains a qualitative model of the system under study, which enables it to estimate the value of the output variable of that system at any point in time, but it also maintains a statistical model of the qualitative model of that system, which enables it to estimate the accuracy of its own predictions, i.e., it maintains an *error model*. Thus, with every prediction, SAPS–II also generates an estimate of the probability of correctness of that prediction.

If SAPS–II is asked to forecast more than one step at a time, Q_f is computed by multiplying the individual probabilities of correctness of the predictions of each step with each other. Since every probability is a number ≤ 1.0, the cumulative probability of a forecast can only decline with time, never increase. This is meaningful, since errors committed during one step must reduce the probability of correctness of the subsequent prediction even more, since already the initial value used in that step is uncertain. The user can specify a value for $Q_{f_{\min}}$ below which no further prediction is meaningful. As soon as the cumulative Q_f value has decreased to a level smaller than $Q_{f_{\min}}$, the forecasting process will come to a halt.

Notice that the probabilistic error model used by SAPS–II assumes the statistical independence of each forecast. This assumption is evidently never justified. However, it is a conservative assumption, i.e., forecasting may stop too early because of it, but never too late. Notice further that the error model is a probabilistic one rather than a possibilistic one. It has thus been designed for use with crisp inductive reasoning rather than fuzzy inductive reasoning.

In fuzzy inductive reasoning, we are currently using a pseudo–probabilistic approach. Rather than computing the probability of correctness of a prediction, we compute the confidence that we have in that prediction. We can also compute the accumulated confidences of all possible predictions during the same step, defined as the sum of all individual confidences for each possible

prediction, and divide the individual confidence of the prediction in question by this sum. In this way, we can obtain a number between 0.0 and 1.0 that can again be reinterpreted as a "probability of correctness" of that prediction.

Of course, it would also be possible to design directly a possibilistic error models, i.e., an error model that is better in line with the possibilistic reasoning approach taken by the FIR methodology as a whole. For example, it would be reasonable to define the confidence of a prediction as the largest of the confidences of each previous observations of that outcome (currently, the individual confidences of previous observations are accumulated), and then define the cumulative confidence over several steps as the smallest of all the confidences in each individual step, and terminate forecasting as soon as this value has decreased below $Q_{f_{\min}}$. However, it has been our experience that the pseudo-probabilistic approach works nicely.

Notice that SAPS-II, being an experimental research tool, currently offers a number of different forecasting routines implementing slightly different inferencing algorithms and employing slightly different input/output formats. Since these routines are experimental and may change over time, it does not make sense to describe each of them in detail in this thesis. The SAPS-II user is asked to consult the *help* function provided with each of these routines to learn about the subtle differences between them.

3.2.4 Defuzzification

The fuzzy forecasting method allows to retrieve pseudo-continuous output signals with a high quality using the *regenerate* function. This means that also a forecast of the continuous-time signals can be obtained (Cellier, 1991a). The *regenerate* function implements the inverse process to the *recode* function, and therefore, the same equation used for recoding is valid here as well (cf. Equation 3.1).

It is important to notice that qualitative reasoning itself only comprises the qualitative modeling and simulation processes. The fuzzification and defuzzification modules are not part of the reasoning process, but they are essential in order to enable the FIR methodology to operate in a mixed quantitative and qualitative modeling and simulation environment. Chapter 4 demonstrates the implementation of such a mixed environment. An example will be shown where some subsystems of the overall process are modeled by means of differential equations, whereas other subsystems are modeled using the FIR approach to qualitative modeling. The overall system is being

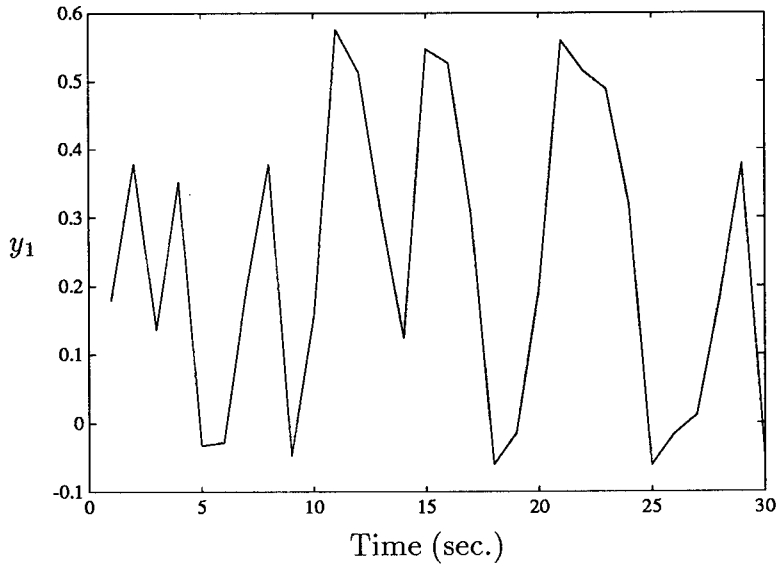


Figure 3.3: Original and Regenerated Continuous–Time Signal y_1.

simulated in a mixed qualitative/quantitative simulation environment.

Let us conclude the linear system example by showing the code that accomplishes the regeneration of quantitative outputs from the qualitative triples that were previously obtained using the *fforecast* function.

```
≫  from = 1 : 3;
≫  for i = 2 : 4,
≫    to = [lm(1 : 3, i), lm(2 : 4, i)]';
≫    fc = oc(271 : 300, i);
≫    fm = om(271 : 300, i);
≫    fs = os(271 : 300, i);
≫    r = regenerate(fc, fm, fs, from, to);
≫    y(:, i − 1) = r;
≫  end
```

In the *regenerate* function, the class vector, fc, the membership vector, fm, and the side vector, fs, containing the qualitative forecast are converted back into a quantitative trajectory, r. The *from* and *to* parameters exchange their role between *recode* and *regenerate*. The signal regeneration process is repeated thrice, once for every one of the three predicted outputs.

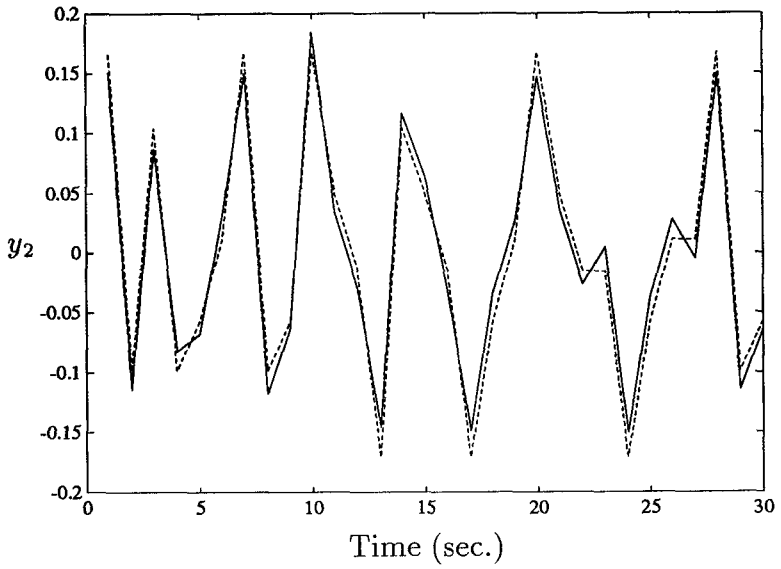


Figure 3.4: Original and Regenerated Continuous–Time Signal y_2.

Now that we have obtained the quantitative forecasts, y_i, we can compare them to the remainder of the original data to validate our three qualitative models. A graphical comparison is given in Figures 3.3, 3.4, and 3.5. The solid lines represent the original data, whereas the dashed lines represent the regenerated signals. In order to compute the error of the previous forecasts, the normalized mean square error (in percentages) has been used. This error is described in Equation 3.27.

$$MSE = \frac{E[(y(t) - \hat{y}(t))^2]}{y_{\text{var}}} \cdot 100\% \tag{3.27}$$

where y_{var} is the variance defined as:

$$y_{\text{var}} = E[y^2(t)] - \{E[y(t)]\}^2 \tag{3.28}$$

The normalized mean square error (in percentages), is $4.1 \times 10^{-4}\%$ for signal y_1, 3.02% for signal y_2, and 0.57% for signal y_3, respectively [1].

[1]The original errors obtained with the programs shown in this chapter were considerably larger, namely 4.5% for signal y_1, 7.1% for signal y_2, and 6.6% for signal y_3. These errors

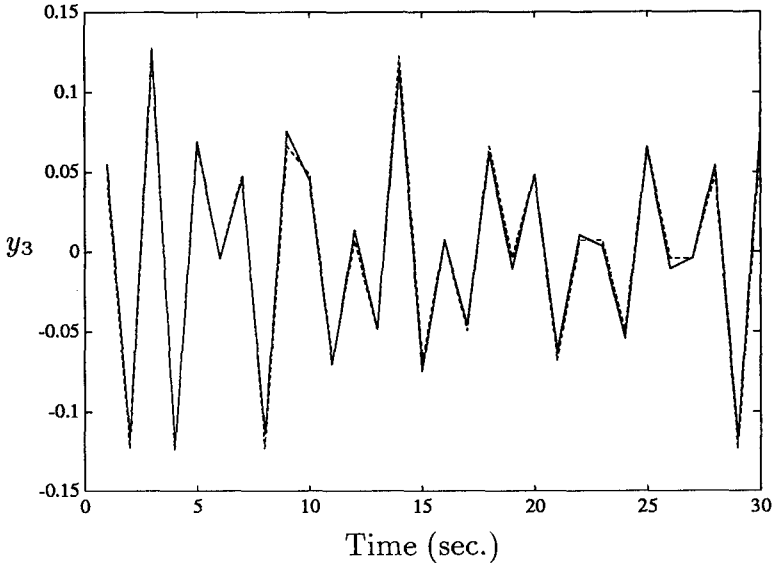


Figure 3.5: Original and Regenerated Continuous–Time Signal y_3.

These results look extremely promising indeed. From knowledge of the input/output behavior of the system alone (no structural knowledge had been used in the modeling process), we were able to come up with a qualitative model with impressive forecasting power, a model that reasons qualitatively about the system under study, but that furthermore can provide quantitative estimates of the output signals as well.

3.3 Conclusions

In this chapter, the qualitative modeling methodology of choice, *Fuzzy Inductive Reasoning*, has been introduced with a high level of detail. The four cornerstones of the FIR methodology: *fuzzification*, *qualitative modeling*, *qualitative simulation*, and *defuzzification* have been minutely discussed, and the utility of this technology has been demonstrated by means of a simple

were, for all three variables, caused mainly by a poor prediction of values near the lower end of the range of values. Therefore, we decided to introduce a fourth class with a narrow range at the bottom for each of the three output variables, and to compensate for the relative loss of data (the qualitative model has now a larger number of legal states) by increasing the number of data records from 300 to 600, while still predicting the last 30 points.

example. The recoding function allows to convert quantitative signals into their qualitative counterparts. This process is the first step on the way of obtaining qualitative knowledge about the system under study. The optimal mask analysis allows us to determine qualitative causality relations among a set of causally related variables. An optimal mask can be viewed as a sort of feature extractor. Future outputs of the system can be predicted by means of the forecasting process. The forecasting function predicts not only the class value of the new output, but also its corresponding membership and side values. Finally, the estimated qualitative output triples can be converted back to quantitative signals by means of the regeneration function.

The present chapter establishes the basis of the fuzzy inductive reasoning methodology. Our implementation of this methodology is the SAPS–II software, which is available as either a CTRL–C library or a Matlab toolbox.

The qualitative reasoning process itself is composed of the qualitative modeling and simulation engines. The fuzzification and defuzzification modules are not directly related to the reasoner itself. However, they are essential as well, since they enable the user to work with mixed quantitative and qualitative simulation models, whereby quantitative subsystems are modeled using differential equations and qualitative subsystems are modeled using the FIR technology.

The next chapter of this dissertation explains how such a mixed modeling and simulation environment can be constructed, and it demonstrates the use of the newly introduced concepts and tools by means of a generic example.

Chapter 4

Mixed Quantitative/Qualitative Modeling and Simulation

4.1 Introduction

One serious difficulty with many qualitative modeling and simulation techniques is that they have a tendency of *overgeneralizing* knowledge. Thereby, too much ambiguity is introduced into the qualitative results, making them practically worthless. Detailed knowledge should be incorporated into the model wherever it is available in order to reduce the ambiguity of the results. However, many qualitative modeling and simulation approaches do not permit to incorporate quantitative knowledge at all, even where it is available.

There exist many examples of systems with partial structural knowledge available to the modeler. These are systems that contain facets or components, usually called subsystems, that are well understood and for which quantitative structural knowledge is available, whereas other facets or components of the overall system are not well understood, the functioning of which may even be totally unknown.

Subsystems for which quantitative structural knowledge is available should be modeled using traditional quantitative techniques, such as sets of ordinary or partial differential equations, whereas those subsystems for which such detailed knowledge is lacking should be modeled by qualitative techniques.

For example, whereas the mechanical properties of a human heart are well understood and can be described by differential equation models fairly well,

the effects of chemical substances on the behavior of the heart are poorly understood and cannot easily be quantified. A mixed model should be used to describe those portions of the overall system that are well understood by quantitative differential equation models, whereas other aspects that are less well understood are being represented in qualitative terms.

It is, in such cases, desirable to have available a mixed modeling methodology that allows to encode each type of subsystem in the manner most appropriate for it, and that makes it possible to simulate the ensemble of models in an efficient way and with minimal ambiguity.

The main objective of this chapter is to show that mixed quantitative and qualitative modeling and simulation is feasible within the FIR methodology, and how the two worlds: the differential equation models to describe quantitative knowledge and the FIR models to encode qualitative information, are reconcilable with each other.

4.2 Importance of a Mixed Technique

Mixed modeling and simulation is needed for all types of systems that are composed of well–known and poorly–understood subsystems. This characteristic is common to a large variety of different applications.

Fuzzy control systems are of that nature. If it is desired to simulate a fuzzy control system, it is necessary to generate models of the plant to be controlled and of the controller itself, and simulate both together in one simulation program. Usually, the plant is a real physical system for which detailed quantitative knowledge is available to the modeler. Thus, the model of the plant can be and should be described by a differential equation model. On the other hand, the fuzzy controller enacts a qualitative control upon the system, and should be dealt with accordingly. Therefore, when a fuzzy controller is used to govern a physical plant, a system with partial structural knowledge is created. It is thus essential to have available a mixed modeling and simulation technique that enables the user to model each subsystem separately in the most effective way, and then simulate the combined system using a mixed simulation environment (Cellier and Mugica, 1992).

Fault monitoring in complex systems is another area where the use of a mixed modeling and simulation technique is highly desirable. During the design of a fault monitoring system, it is often useful to be able to simulate the

real plant for which the fault monitor is to be developed, together with faults that are scheduled to occur frequently, and together with the fault monitor that tries to detect these faults and then discriminate between them. In the real system, one may have to wait for a long time before a particular fault occurs, and therefore, it may be hard to debug the fault monitor using the real system. As in the previous example, the plant to be monitored is usually a physical plant with lots of detailed knowledge available to the modeler. Thus, it is appropriate to model the plant by use of differential equations. The fault monitor, on the other hand, mimics the behavior of a human plant operator. It contains one or several qualitative models of facets or modes of functioning of the very same plant, and reasons about discrepancies between expected and observed plant behavior (Cellier and de Albornoz, 1993b).

In biomedical engineering, it is frequently necessary to deal with systems with partial structural knowledge. Many subsystems of the human body contain facets, e.g. mechanical or electrical aspects, that have been well studied and that can therefore be represented reliably by differential equation models. Yet, these very same subsystems usually contain other facets, often related to chemical processes, the functioning of which is known in an intuitive manner only. Therefore, as in the other two cases, it is essential to be able to have available a mixed quantitative and qualitative modeling and simulation methodology that can be used to analyze such systems.

The development of the mixed modeling and simulation methodology presented in this chapter is common to three doctoral theses under simultaneous development at the *Universitat Politècnica de Catalunya*, focusing on the three application areas that were briefly outlined in this section. All of them require a mixed quantitative/qualitative modeling and simulation environment to work with (Cellier *et al.*, 1992, 1994).

The design of such an environment is not a straightforward task. Several problems arise that need to be addressed. How should a mixed quantitative and qualitative simulation deal with the fact that the quantitative subsystems treat the independent variable, *time*, as a quantitative variable, whereas the qualitative subsystems treat the same variable qualitatively? When does a particular qualitative event occur in terms of quantitative time? How are the explicit experimental conditions that are needed by the quantitative subsystems accounted for in the qualitative subsystems?

Quite obviously, a number of incompatibility issues exist between quantitative and qualitative subsystems that must be settled before mixed simulations can be attempted. In a mixed simulation, also the qualitative

subsystems must treat *time* as a quantitative variable. Furthermore, the purpose of qualitative models in the context of mixed simulations is revised. It is no longer their aim to enumerate episodical behaviors. Instead, also the qualitative models are now used to determine a single episodical behavior in response to a single set of qualitative experimental conditions. However, whereas quantitative simulation of a deterministic system generates the one and only *true* trajectory behavior, qualitative simulation only is able to generate the *most likely* episodical behavior and attach a measure of likelihood to each prediction made.

It is thus necessary to devise qualitative modeling and simulation capabilities that are compatible with their quantitative counterparts and that can be used to represent qualitative subsystems, such as those mentioned above, appropriately and in terms of knowledge available to the system designer at the time of modeling.

4.3 Mixed Quantitative/Qualitative Technique

The two main components of the FIR methodology that enable the combination of quantitative and qualitative subsystems are the *fuzzification* (recode) and *defuzzification* (regeneration) modules.

The fuzzification module performs the functions of a fuzzy A/D converter. It converts quantitative signals (analog signals) into qualitative signals (discrete signals). However, and contrary to the regular A/D converters, it uses a very coarse discretization that is augmented by fuzzy membership functions to be used for the fine interpolation between neighboring discrete states. In contrast, the defuzzification module operates as a fuzzy D/A converter. It converts qualitative signals back into quantitative signals. Consequently, the interfaces between the quantitative and qualitative subsystems are achieved by means of the FIR recoding and regeneration functions.

Let us sketch a generic example that demonstrates how the problem is tackled. Let us assume that we have a system consisting of four subsystems, as presented in Figure 4.1.

It is furthermore assumed that the internal structure of subsystems S_1, S_3, and S_4 is well known, and therefore, quantitative models for each of these subsystems can be constructed. However, subsystem S_2 is modeled qualitatively, because its internal structure is unknown. The only fact that is known about subsystem S_2 is that there exist (possibly dynamic) input/output

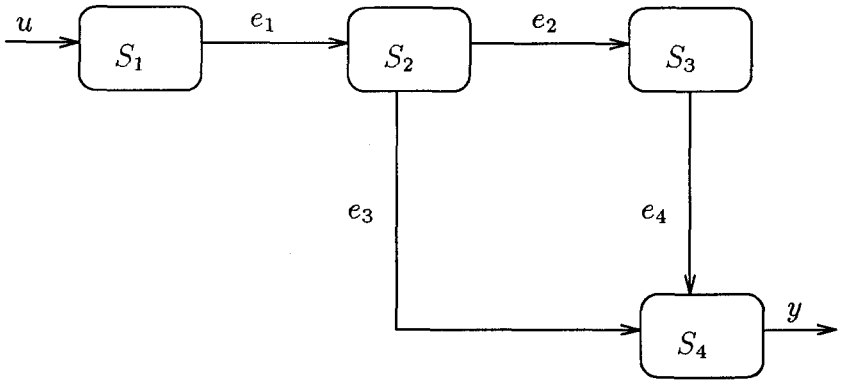


Figure 4.1: Example of a System Composed of 4 Subsystems

relationships between the signals e_1 and e_2 on the one hand and between e_1 and e_3 on the other. These qualitative relationships can be represented by two separate FIR models.

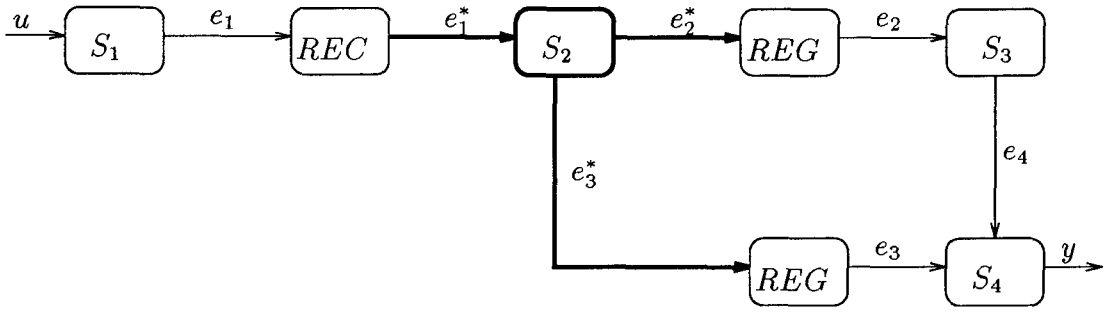


Figure 4.2: Example of a Mixed Simulation Process

In such a system, mixed simulation can be performed as shown in Figure 4.2. The *REC* box represents the fuzzification process. It converts the quantitative signal e_1 into a qualitative triple called e_1^*, allowing the qualitative model of subsystem S_2 to simulate (predict) new output values e_2^* and e_3^*. The *REG* boxes denote the defuzzification processes that convert the predicted qualitative outputs of subsystem S_2, e_2^* and e_3^*, into the quantitative signals e_2 and e_3. These are then ready to be used by the quantitative subsystems S_3 and S_4.

The above generic example does not contain any feedback loops, but the realistic engineering example that is presented in detail later in this chapter

will show that feedback loops do not present any insurmountable barriers to the methodology.

Quantitative subsystems are modeled using a mixture of differential and algebraic equations and are simulated using the Advanced Continuous Simulation Language, abbreviated as ACSL (MGA, 1991). Qualitative subsystems are modeled by means of the fuzzy inductive reasoning methodology, and therefore, they are represented as masks. The simulation of those subsystems is accomplished using a subset of the SAPS–II modules. In order to simulate both types of subsystems together, interfaces had to be created between ACSL and SAPS–II.

These interfaces will subsequently be described. Notice that, of all the SAPS–II modules, only three are needed during simulation: the *fuzzy recoding* module, the *fuzzy regeneration* module, and the *fuzzy forecasting* model. All other SAPS modules, such as the optimal mask generation module, are used in an off–line fashion only, and need not be accessible from within the mixed simulation environment. Therefore, the masks and the recoded raw data matrices to be used in the mixed simulation are either introduced as local variables in the ACSL program, or they are imported from the Matlab or CTRL–C environment.

All three interfaces are encoded as ACSL macros. The *recode* interface takes the following form:

macro $recode\ (class, memb, side, signal, from, nlev)$
  procedural $(class, memb, side = signal, from, nlev)$
    call $recode\ (class, memb, side, signal, from, 1, nlev)$
  end ! $of\ procedural$
macro end

ACSL calls the same Fortran–encoded subroutine that resides inside the SAPS–II software when called from Matlab. It packs it into a macro for easier use from within the ACSL software environment. The *recode* macro has three output parameters, the class value, *class*, the fuzzy membership value, *memb*, and the side value, *side*. It also has three input parameters, the analog input to be recoded, *signal*, the *from* matrix, with the same definition as provided in Chapter 3, and the number of discrete levels, n_{lev}, into which the continuous signal is to be recoded, corresponding to the number of columns of the *from* matrix. The call to the Fortran routine contains one more parameter (with value '1'), which denotes the number of data records to be recoded. Contrary

to Chapter 3, where *recode* had always been used to fuzzify the entire raw data matrix in a single function call, in the ACSL interface, we only fuzzify the current value of the analog variable, *signal*.

The *regen* macro performs the inverse operation. It has the same formal arguments as the *recode* macro, just arranged in a different sequence.

macro $regen\ (signal, class, memb, side, to, nlev)$
  procedural $(signal = class, memb, side, to, nlev)$
    call $regen\ (signal, class, memb, side, to, 1, nlev)$
  end ! $of\ procedural$
macro end

It converts a qualitative triple, <*class*,*memb*,*side*>, into an analog variable, *signal*. The *to* matrix assumes here the role of the former *from* matrix. As in the case of the *recode* macro, *regen* converts a single record of a single variable at any one time.

The forecasting interface is a bit more complex. It looks as follows:

macro $forcst\ (class, memb, side, mask, cmat, mmat, smat, k)$
  procedural $(class, memb, side = mask, cmat, mmat, smat, k)$
    call $forcst\ (class, memb, side, mask, cmat, mmat, smat,$
      $k - 1, nvar, depth, nrec)$
  end ! $of\ procedural$
macro end

The *forcst* macro forecasts one qualitative triple, <*class*,*memb*,*side*>, using the optimal mask, *mask*. However, the *forcst* routine must also access the experience data base (the past behavior) consisting of the class value matrix, *cmat*, the membership value matrix, *mmat*, and the side value matrix, *smat*. Also these variables must thus be passed through the parameter list. Due to constraints of the Fortran language, we must also provide the subroutine with the dimensions of all these matrices by letting it know the number of variables involved, n_{var}, the depth of the mask, *depth*, and the number of records for which the experience data base is dimensioned, n_{rec}. In order to keep the run–time calling list short, the dimensions are treated as constants, rather than passing them along as formal arguments of the macro. The parameter k is a pointer variable that points to the data record, for which a new value is to be forecast. The *forcst* routine wants to know how many records the experience data base currently contains, i.e., one row less.

We need two additional macros in order to be able to use the ACSL/SAPS interface. Remember from Chapter 3 that, before being able to forecast, we must pack the new input triples into the experience data base and preset the to be predicted outputs to arbitrary legal values. This is accomplished using the *prefrc* macro:

```
macro prefrc (cmat, mmat, smat, class, memb, side, k)
  procedural (cmat, mmat, smat = class, memb, side, k)
    cmat(k, 1) = class
    mmat(k, 1) = memb
    smat(k, 1) = side
    cmat(k, 2) = 0
    mmat(k, 2) = 1.0
    smat(k, 2) = 0
    cmat(k, 3) = 0
    mmat(k, 3) = 1.0
    smat(k, 3) = 0
    cmat(k, 4) = 0
    mmat(k, 4) = 1.0
    smat(k, 4) = 0
  end ! of procedural
macro end
```

assuming that we are dealing with a four–variable system with one input and three outputs.

Finally, we need to pack the forecasts back into the experience data base. This is done using the *pack* macro:

```
macro pack (dat, u, y1, y2, y3, k)
  procedural (dat = u, y1, y2, y3, k)
    dat(k, 1) = u
    dat(k, 2) = y1
    dat(k, 3) = y2
    dat(k, 4) = y3
  end ! of procedural
macro end
```

assuming again the same four–variable system.

We are now ready to describe the qualitative simulation to be performed within the ACSL environment. Let us continue to assume a four–variable system with a single input, u, and the three outputs, y_1, y_2, and y_3. The sampling rate is assumed to be $\delta t = 3$ seconds.

discrete *saps*
  interval $tsaps = 3.0$
  integer $k, cu, su, cy1, sy1, cy2, sy2, cy3, sy3$
  procedural
    $k =$ifix$(t/tsaps + 1.5)$
    $cu =$ifix$(u) + 1$
    $mu = 1.0$
    $su = -$ifix$(2.0 * (u - 0.5))$
    prefrc$(cmat, mmat, smat = cu, mu, su, k)$
    forcst$(cy1, my1, sy1 = mask1, cmat, mmat, smat, k)$
    forcst$(cy2, my2, sy2 = mask2, cmat, mmat, smat, k)$
    forcst$(cy3, my3, sy3 = mask3, cmat, mmat, smat, k)$
    pack$(cmat = cu, cy1, cy2, cy3, k)$
    pack$(mmat = mu, my1, my2, my3, k)$
    pack$(smat = su, sy1, sy2, sy3, k)$
    $y1 =$regen$(cy1, my1, sy1, ty1, 3)$
    $y2 =$regen$(cy2, my2, sy2, ty2, 3)$
    $y3 =$regen$(cy3, my3, sy3, ty3, 3)$
  end ! *of procedural*
end ! *of discrete saps*

The qualitative simulation is encoded in a discrete block of ACSL to be executed once every 3.0 seconds. Inside this block, we first determine the current value of the pointer variable, k, i.e., we need to know what row of the experience data base corresponds to the current time, t. Recoding of the input was done manually in the above example, since the input was assumed to be already binary, thus no fuzzy recoding was needed. The qualitative input triples are immediately stored in the experience data base using the *prefrc* macro. Forecasting can now begin. Three separate forecasts are made, one for each output variable. The predicted qualitative output triples are then packed back into the experience data base, and quantitative outputs are regenerated from them.

The ACSL/SAPS interface is considerably less comfortable in its utilization

than the Matlab/SAPS interface. The reasons for this inconvenience are speed requirements. It is assumed that SAPS routines will be called from within the Matlab environment sparingly. Thus, a slow interface using files to transfer variables to and fro between Matlab and SAPS–II is acceptable. Each time a SAPS module is called from within Matlab or CTRL–C, a new process is spawned. This makes the interface even slower. The same luxury is not acceptable when calling SAPS from within ACSL. This interface is written for complex nonlinear applications, and it is foreseen that hundreds or maybe even thousands of steps are to be forecast. This requires a very efficient interface between the two languages, which, in turn, forces us down to the level of Fortran subroutine calls.

In the next section, the mixed simulation technique is addressed by means of a simple, yet non–trivial, example from the technical domain. A position control system involving a hydraulic motor with a four–way servo valve demonstrates the process of mixed quantitative and qualitative simulation using fuzzy inductive reasoning.

In this example, the hydraulic subsystem is being modeled by means of qualitative techniques, whereas the mechanical and electrical parts are modeled using differential equation models. This system has been chosen, since complete quantitative knowledge is available for this system, so that the results obtained can be compared with the results obtainable by purely quantitative approaches. In this way, it was possible to attack the theoretical challenges of the mixed modeling methodology first, without mingling them at once with practical issues of poorly understood biomedical processes.

A realistic and meaningful biomedical application of the proposed technique has been developed also. It is presented in Chapter 6 of this thesis. This biomedical application represents aspects of the cardiovascular system of the human body.

4.4 Position Control System Application

The example was chosen simple enough to be presented in full, yet complex enough to demonstrate the generality and validity of the approach. However, it is not suggested that the chosen example represent a meaningful application of mixed quantitative and qualitative simulation. After all, if complete knowledge about a process indeed *is* available, why bother to design a qualitative model for it. The example was chosen to prove the concept and to clearly present

the methodology, not as a realistic and meaningful application of the proposed technique.

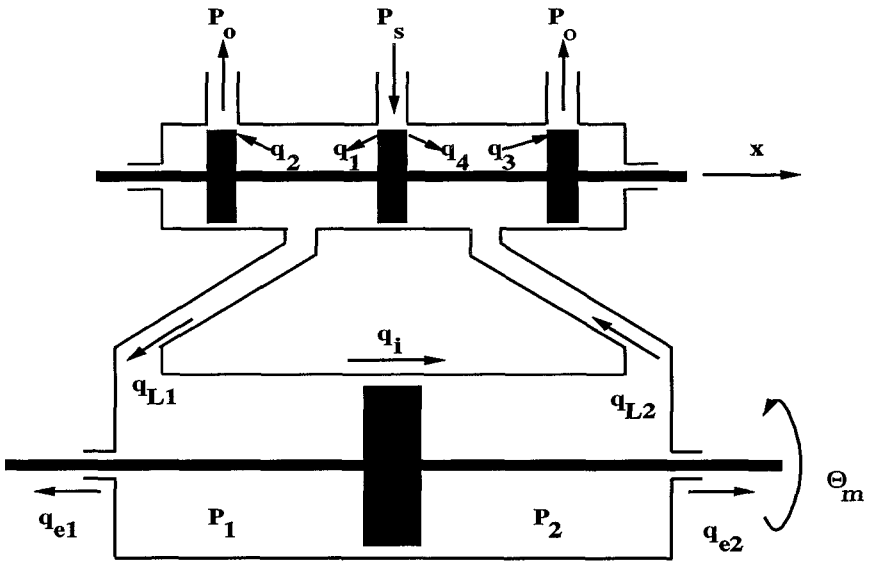


Figure 4.3: Hydraulic Motor with a Four-way Servo Valve

Figure 4.3 shows a hydraulic motor with a four-way servo valve. The flows from the high-pressure line into the servo valve and from the servo valve back into the low-pressure line are turbulent. Consequently, the relation between flow and pressure is quadratic.

$$\begin{aligned} q_1 &= k(x_0 + x)\sqrt{P_S - p_1} \\ q_2 &= k(x_0 - x)\sqrt{p_1 - P_0} \\ q_3 &= k(x_0 + x)\sqrt{p_2 - P_0} \\ q_4 &= k(x_0 - x)\sqrt{P_S - p_2} \end{aligned} \tag{4.1}$$

The chosen parameter values are $P_S = 0.137 \times 10^8$ Nm^{-2}, $P_0 = 1.0132 \times 10^5$ Nm^{-2}, $x_0 = 0.05$ m, and $k = 0.248 \times 10^{-6}$ kg$^{-1/2}$m$^{5/2}$.

The change in the chamber pressures is proportional to the effective flows in the two chambers.

$$\begin{aligned} \dot{p}_1 &= c_1(q_{L1} - q_i - q_{e1} - q_{ind}) \\ \dot{p}_2 &= c_1(q_{ind} + q_i - q_{e2} - q_{L2}) \end{aligned} \tag{4.2}$$

with $c_1 = 5.857 \times 10^{13}$ kgm^{-4}sec^{-2}. The internal leakage flow, q_i, and the external leakage flows, q_{e1} and q_{e2}, can be computed as,

$$\begin{aligned} q_i &= c_i \cdot p_L = c_i(p_1 - p_2) \\ q_{e1} &= c_e \cdot p_1 \\ q_{e2} &= c_e \cdot p_2 \end{aligned} \tag{4.3}$$

where $c_i = 0.737 \times 10^{-13}$ kg^{-1}m^{4}sec, and $c_e = 0.737 \times 10^{-12}$ kg^{-1}m^{4}sec.

The induced flow, q_{ind}, is proportional to the angular velocity of the hydraulic motor, ω_m:

$$q_{ind} = \psi \cdot \omega_m \tag{4.4}$$

with $\psi = 0.575 \times 10^{-5}$ m^3, and the torque produced by the hydraulic motor is proportional to the load pressure, p_L:

$$T_m = \psi \cdot p_L = \psi(p_1 - p_2) \tag{4.5}$$

The mechanical side of the motor has an inertia of $J_m = 0.08$ kgm^2, and a viscous friction of $\rho = 1.5$ kgm^{2}sec^{-1}.

The hydraulic motor is embedded in the control circuitry shown on Figure 4.4. In the mixed quantitative and qualitative simulation, the mechanical and electrical parts of the control system are represented by differential equation models, whereas the hydraulic part is represented by a fuzzy inductive reasoning model.

For this purpose, it was assumed that no knowledge exists that would permit a description of the hydraulic equations by means of a differential equation model. All that is known is that the mechanical torque, T_m, of the hydraulic motor somehow depends on the control signal, u, and the angular velocity, ω_m.

For validation purposes, the mixed simulation results are compared with previously obtained purely quantitative simulation results. The purely

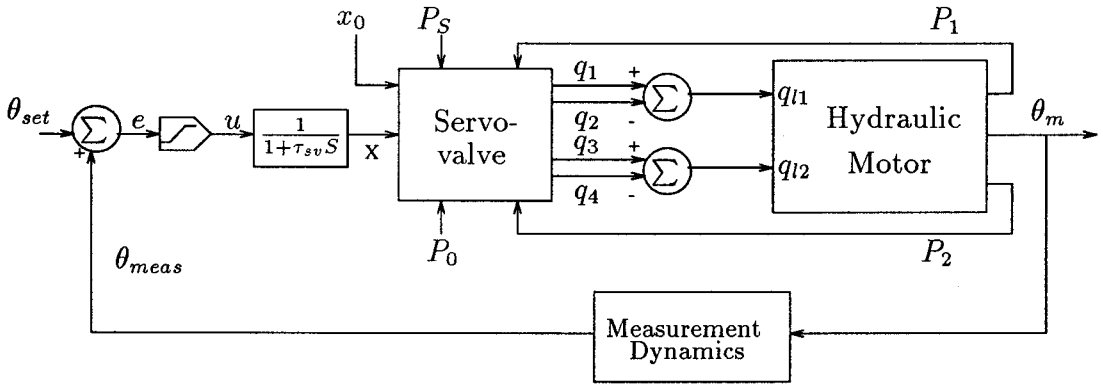


Figure 4.4: Hydraulic Motor Position Control Circuit

quantitative simulation of the overall system (Figure 4.4) was implemented using an ACSL program, simulating the system across 2.5 seconds. A binary random input signal was applied to the input of the system, θ_{set}.

The values of the control signal, u, the angular velocity, $\omega_m = \dot{\theta}_m$, and the torque, T_m, stemming from the first 2.25 seconds of the quantitative simulation were recoded to generate the fuzzy inductive model of the hydraulic motor.

The values of the last 0.25 seconds of quantitative simulation were stored for validation purposes. Validation is accomplished by comparing the simulation results of the new mixed model with those of the purely quantitative model, which is being used in place of "measurement data."

4.4.1 Building the Fuzzy Inductive Model

As described in the previous sections, the fuzzy inductive model is constructed in two steps. In the first step, the quantitative data are recoded, and in the second step, the fuzzy optimal mask is determined from the recoded data.

4.4.1.1 Fuzzy Recoding of the Hydraulics

The first question to be addressed in the recoding process is the selection of the appropriate sampling rate (communication interval) for the continuous variables to be recorded (either from measurements or, as in this example, from a quantitative simulation study). In the given example, this value can

be deduced (cf. Equation 3.10) once the mask depth has been chosen and the longest time constant of the system determined. The eigenvalue is at -20 sec^{-1}, and therefore, the longest time constant is 0.05 seconds. In order to capture the longest time constant, the three variables u, ω_m, and T_m must be sampled once every 0.025 seconds if a mask depth of $depth = 3$ is chosen.

Unfortunately, fuzzy inductive forecasting predicts only one value of T_m per sampling interval. Thus, the mixed qualitative and quantitative simulation behaves like a sampled–data control system with a sampling rate of 0.025 seconds. Thereby, the stability of the control system is lost because the sampling rate is too slow to keep up with the changes in the system. From a control system perspective, it is necessary to sample the variables considerably faster. An ACSL program was coded to study different sampling rates in order to obtain a stable control performance. This program introduces into the quantitative simulation a delay in the computation of the torque. The largest delay time that can be introduced without losing stability of the control system was identified. It was determined that the longest tolerable delay is 0.0025 seconds. Consequently, the mask depth must be increased from three to 21.

The next step is to find the number of discrete levels into which each of these variables should be recoded. For the given example, it was decided that all three variables can be sufficiently well characterized by three levels. A discretization of the variables in this manner implies that the number of legal states is 27 ($3 \times 3 \times 3$).

As explained before, it is desirable to record each state at least five times. Consequently, a minimum of 130 recordings, corresponding to a total simulation time of 0.325 seconds, is needed. However, due to the mismatch between the sampling rate required by fuzzy forecasting and the actually used sampling rate that is required due to the control system characteristics, considerably more data are needed. It was decided to choose a total simulation time of 2.5 seconds with 2.25 seconds being used for model identification, and the remaining 0.25 seconds being used for validation. This provides the optimal mask module with 900 recordings used for model identification, while fuzzy forecasting is carried out over the final 100 steps.

4.4.1.2 Fuzzy Optimal Mask of the Hydraulics

With the data recoded as previously described, it is possible to build the qualitative model of the hydraulics by means of the fuzzy optimal mask module

inside SAPS–II. To combine the qualitative and quantitative simulation models, it was necessary to observe the limitation imposed by the dynamic stability problem, while covering with the mask the longest time constant, $\Delta t = 0.05$ seconds, of interest to the qualitative model. This meant that, as mandated by control theory, a sampling interval of $\delta t = 0.0025$ seconds had to be chosen. Consequently, the mask depth can be computed as:

$$depth = \text{round}(\frac{\Delta t}{\delta t}) + 1 = 21 \tag{4.6}$$

Filling such a large mask candidate matrix up with -1 elements to denote potential inputs, even while limiting the search to masks of complexities up to six only, would be painfully slow. Therefore, the following approach was taken. From the point of view of fuzzy reasoning, a mask depth of three is usually sufficient. Consequently, it was decided to consider only inputs in the first, the 11^{th}, and the 21^{st} row of the mask, blocking all other rows out by setting the corresponding elements of the mask candidate matrix equal to 0. In this way, the search can proceed quickly, and yet, the resulting "optimal" mask will still be very close to the truly optimal mask. Thus, the following mask candidate matrix of depth 21 was chosen:

$$\begin{array}{l|ccc} {}_t\backslash^x & u & \omega_m & T_m \\ t-20\delta t & -1 & -1 & -1 \\ t-19\delta t & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ t-11\delta t & 0 & 0 & 0 \\ t-10\delta t & -1 & -1 & -1 \\ t-9\delta t & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ t-\delta t & 0 & 0 & 0 \\ t & -1 & -1 & +1 \end{array} \tag{4.7}$$

This mask candidate matrix indicates that the mechanical torque, T_m, at time t may depend on the current values of u and ω_m, as well as on past values of u, ω_m, and T_m at times $t - 0.025$ seconds and $t - 0.05$ seconds.

The following optimal mask has been found for this example:

$$
\begin{array}{c|ccc}
t\backslash x & u & \omega_m & T_m \\
t-20\delta t & 0 & -1 & -2 \\
t-19\delta t & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
t-11\delta t & 0 & 0 & 0 \\
t-10\delta t & 0 & 0 & 0 \\
t-9\delta t & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 & 0 \\
t & -3 & 0 & +1
\end{array}
\tag{4.8}
$$

In other words:

$$T_m(t) = \tilde{\mathrm{f}}(\omega_m(t-0.05), T_m(t-0.05), u(t)) \tag{4.9}$$

4.4.1.3 Fuzzy Forecasting and Signal Regeneration

Once the optimal mask has been determined and before it can be integrated into the mixed simulation, its prediction capability must be checked. For this purpose, the values of T_m obtained from the quantitative simulation are compared with the forecast and regenerated values. This process is shown in Figure 4.5.

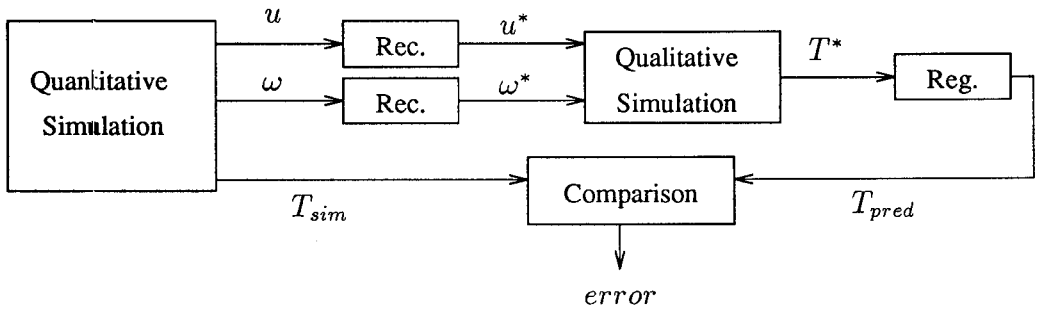


Figure 4.5: Validation of Qualitative Model

Notice that, although Figure 4.5 contains both a quantitative and a qualitative simulation, this does not mandate a mixed simulation approach yet. Since there are no feedback loops in Figure 4.5, it is perfectly feasible to

first complete the quantitative simulation, storing all the data away for later reuse, and *then* perform the qualitative simulation, recalling the previously stored quantitative trajectories.

As mentioned earlier, the first 900 rows of the raw data matrix are used as past history data to compute the optimal mask. Fuzzy forecasting is employed to predict new qualitative triples, T_m^*, for the last 100 rows of the raw data matrix only. From the predicted qualitative triples, quantitative values can then be regenerated that can be compared with the "true" values obtained from the purely quantitative simulation for validation purposes.

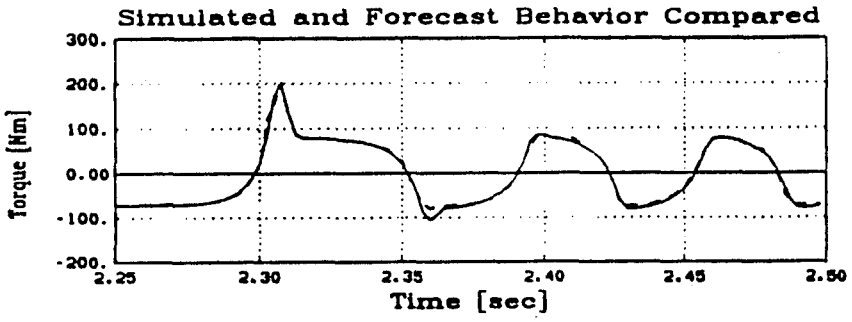


Figure 4.6: Comparison between Simulated and Forecast Signals (Open Loop)

Figure 4.6 compares the true "measured" values of T_m obtained from the purely quantitative simulation (solid line) with the forecast and regenerated values obtained from the qualitative simulation (dashed line).

The results are very encouraging indeed. Quite obviously, the optimal mask contains sufficient information about the behavior of the hydraulic subsystem to be used as a valid replacement of the true quantitative differential equation model, although the chosen recoding scheme was extremely coarse using three levels for each variable only. Notice that the fuzzy inductive reasoning model was constructed solely on the basis of measurement data. No insight into the functioning of the hydraulic subsystem was required other than the knowledge

that the torque, T_m, dynamically depends on the control signal, u, and the angular velocity, ω_m.

4.4.2 Mixed Modeling and Simulation

Once the prediction capability has been demonstrated, the fuzzy inductive reasoning model can be used to replace the former differential equation model of the hydraulic subsystem in a mixed simulation, where the electrical and mechanical subsystems are still modeled using differential equations, whereas the hydraulic subsystem is modeled using a fuzzy optimal mask. The mixed model is shown on Figure 4.7.

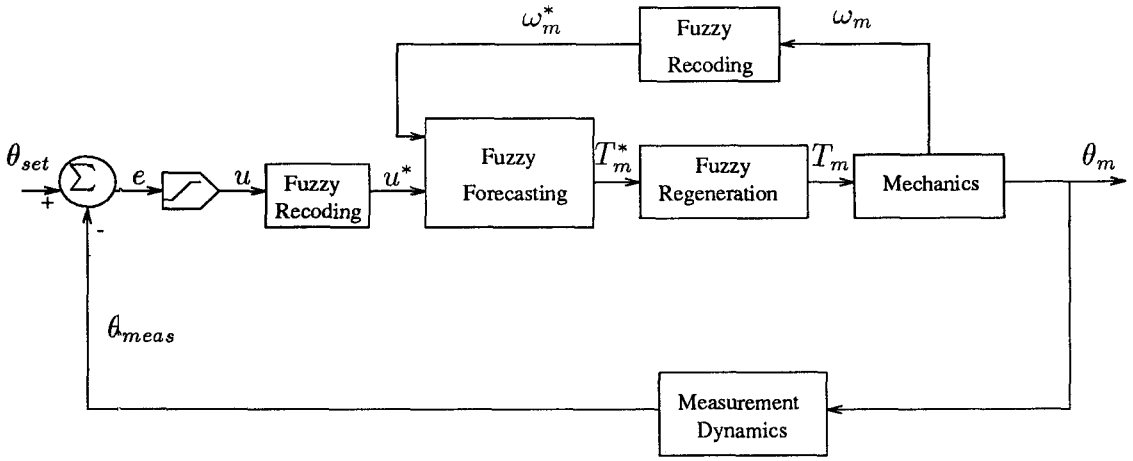


Figure 4.7: Mixed Model of the Hydraulic System

Notice that this model contains feedback loops that involve both quantitative and qualitative submodels. Thus, a mixed simulation capability is indispensable for handling this model. The quantitative control signal, u, is converted to a qualitative triple, u^*, using fuzzy recoding. Also the quantitative angular velocity, ω_m, of the hydraulic motor is converted to a qualitative triple, ω_m^*. From these two qualitative signals, a qualitative triple of the torque of the hydraulic motor, T_m^*, is computed by means of fuzzy forecasting. This qualitative signal is then converted back to a quantitative signal, T_m, using fuzzy signal regeneration.

Forecasting was restricted to the last 100 sampling intervals, i.e., to the time span from 2.25 seconds to 2.5 seconds. Figure 4.8 compares the angular

position, θ_m, of the hydraulic motor from the purely quantitative simulation (solid line) with that of the mixed quantitative and qualitative simulation (dashed line).

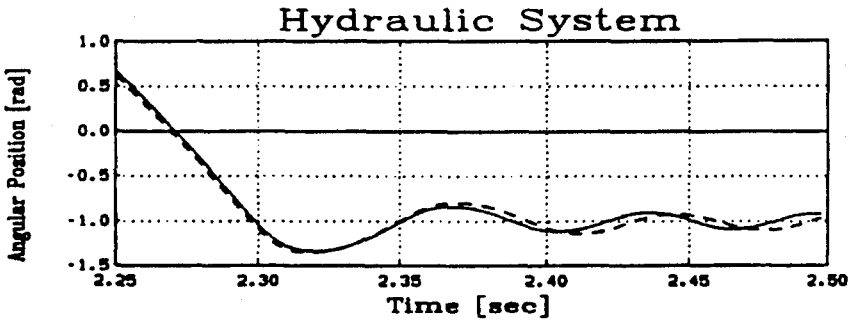


Figure 4.8: Comparison between Simulated and Forecast Signal (Closed Loop)

As was to be expected, the mixed model behaves like a sampled–data control system. The mixed simulation exhibits an oscillation amplitude that is slightly larger and an oscillation frequency that is slightly smaller than those shown by the purely quantitative simulation. Surprisingly, the damping characteristics of the mixed model are slightly better than those of the purely quantitative model.

This example demonstrates the validity of the chosen approach. Mixed simulations are similar in effect to sampled–data system simulations. *Fuzzy recoding* takes the place of analog–to–digital converters, and *fuzzy signal regeneration* takes the place of digital–to–analog converters. However, this is where the similarity ends. Sampled–data systems operate on a fairly accurate representation of the digital signals. Typical converters are 12–bit converters, corresponding to discretized signals with 4096 discrete levels. In contrast, the fuzzy inductive reasoning model employed in the above example recoded all three variables into qualitative variables with the three levels 'small,' 'medium,' and 'large.' The quantitative information is retained in the fuzzy membership functions that accompany the qualitative signals. Due to the

small number of discrete levels, the resulting finite state machine is extremely simple. Fuzzy membership forecasting has been shown to be very effective in inferring quantitative information about the system under investigation in qualitative terms.

4.5 Conclusions

In this chapter, a mixed quantitative and qualitative modeling and simulation methodology has been introduced. Quantitative subsystems are modeled and simulated as differential equation models, whereas qualitative subsystems are described using fuzzy inductive reasoning. Mixed quantitative/qualitative modeling is a highly valuable methodology when dealing with any type of systems composed of well–known and poorly–understood subsystems. This approach allows to encode exactly as much information about the system under study as is available to the modeler. Purely quantitative modeling approaches require detailed information to be provided for *all* subsystems, forcing the modeler to make assumptions, the correctness of which may be hard to verify. Such models will generate *unique* answers when simulated, but it may be an impossibly difficult task to assess the correctness of the simulation results obtained in this fashion. Purely qualitative modeling approaches, on the other hand, prevent the modeler from providing detailed information that he or she may possess, leading to unnecessary ambiguity. These models will certainly produce *valid* answers when simulated, but these answers may often not contain much useful information due to the inherent ambiguity problem. Mixed simulation is thus an indispensable tool for obtaining valid and yet unambiguous results when dealing with systems for which only partial knowledge is available.

Our implementation of the mixed simulation environment employs ACSL as the underlying simulation language. Quantitative subsystems are implemented in a straightforward manner by encoding the corresponding differential and algebraic equations in the *derivative* section of the ACSL program. This is what ACSL had been designed for. The qualitative subsystems, however, are encoded in one or several *discrete* sections of the ACSL program that are executed once per sampling interval. These sections call upon ACSL macros, which in turn make calls to Fortran–encoded SAPS routines for *fuzzy recoding*, *fuzzy forecasting*, and *fuzzy signal regeneration*.

In order to demonstrate the feasibility of mixed simulation using the proposed approach, a hydraulic motor control system has been simulated,

whereby the hydraulic subsystem has been modeled by means of qualitative techniques, whereas the mechanical and electrical parts were modeled using differential equations. Special attention had to be paid to the stability of the sampled feedback control system, satisfying simultaneously the Shannon sampling theorem and the needs of the inductive reasoner. The excellent results obtained from the mixed simulation of this system demonstrate the validity of the chosen approach.

A mixed model has meanwhile also been used in the simulation of a fairly complex biomedical application, describing aspects of the cardiovascular system of the human body. This work is presented in Chapter 6 of this dissertation.

In chapters 3 and 4 of this doctoral thesis, a new and very promising modeling and simulation methodology was presented. The general descriptions provided in these two chapters have laid the ground for an understanding of the biomedical applications that will be described in subsequent chapters. The remainder of this thesis deals with biomedical questions exclusively. The discussion begins in Chapter 5 with a compilation of the differences between technical and biomedical applications to modeling and simulation, with a discussion of the inherent difficulties that need to be overcome when dealing with biomedical systems, and with an explanation of why modeling and simulation, when applied to biomedicine, must still be considered an art rather than a science.

Chapter 5

Difficulties in Biomedical Applications

5.1 Introduction

In previous chapters of this dissertation, a general state–of–the–art review of techniques and tools used in qualitative modeling and simulation of biomedical systems was presented, and one of these techniques, the Fuzzy Inductive Reasoning (FIR) methodology, was described in more detail, together with an extension that allows mixed quantitative (differential equation) and qualitative (FIR) models to be treated.

It is now time to dig a little more deeply into the biomedical engineering field, and to discuss the specific difficulties that have hampered progress in modeling and simulation of biomedical systems in the past, and that make these systems much more difficult to handle than practically all other types of systems met anywhere in science and engineering. This chapter provides first insights into these domain–specific problems. It also forms a basis for the subsequent chapters, where several of the problems presented here will be addressed in order to try to provide a general solution to them, or at least, to alleviate their implications.

As has been mentioned in Chapter 2, qualitative techniques have not been as extensively applied to modeling and temporal reasoning in biomedicine as in other problem domains. The inherent complexity of medicine has contributed to this lack of success. The difficulties encountered when dealing with biomedical applications are manyfold and formidable. They are summarized

in this chapter.

5.2 Problems of Biomedical Applications

Biomedical systems pose particular problems that make it difficult to come up with models capturing their dynamical behavior. Some of those are summarized here:

- **Lack of information from the point of view of quantity**: Biomedical time constants are often long, and the length of experiments from which data can be extracted is limited. For example, a surgical operation has a certain length specific to the operation. If a particular surgery lasts for a duration of one hour, yet the slowest time constants of the patient are of the order of 10 minutes, it is unfortunately not possible to observe the system for more than six of these time constants in order to obtain better data.

  This is of special importance when working with inductive methodologies. All inductive techniques, such as fuzzy inductive reasoners and neural networks, need lots of data to work with. It is not possible to generate meaningful and reliable inductive models without ample and rich data. This poses a serious problem for most health care applications, where data may not be as easily to come by as in other application domains, such as electrical engineering.

- **Lack of information from the point of view of quality**: Data stemming from biomedical experiments are poor in quality for two reasons. On the one hand, there is a strong dependence on the patient from whom the data are taken. Thus, it is difficult to separate system–generic information from patient–specific information. Even more detrimental, however, are the problems that have to do with the limited range of experiments that can be performed on human subjects. It is not possible to expose a patient to experiments just for the purpose of getting data. Even if a heart surgery lasts for several hours, the data that can be extracted from the patient may be (and hopefully are) quite uniform. However, in order to come up with models to predict potential problems, we would need recordings of earlier disasters of the same or a similar kind from which specific symptoms leading to the problem can be extracted. Such data are very difficult to come by. Having rich data is

indispensable for inductive methodologies. Inductively generated models cannot predict system behavior that they have never seen before.

- **Technical difficulties with obtaining information**: The interest of the patient must always come first. Therefore, data from e.g. a surgery can only be gathered if this can be accomplished in a non–intrusive fashion. If the data gathering setup is in any way intrusive and carries the potential for increasing the risk to the patient, then data gathering must not be done in the interest of the patient.

- **Ethical difficulties with obtaining information**: It is excellent for the modeler if he or she can apply random input signals to the system under study in order to exert it equally at all frequencies. However, if the "system under study" is a human patient, this is hardly feasible. Moreover, it is always necessary to obtain the permission of the patient to record any data for whatever purposes, irrespective of whether this puts the patient in danger or not. Many patients are reluctant to grant this permission, since they don't understand the purpose of the measurements and cannot truly judge whether this puts them at extra risk or not.

- **Diversity in patient behavior**: Usually, the aim of modeling is that of knowledge generalization. It is rarely useful to obtain a model that applies to a single patient only. Obtaining data from one single patient in order to identify a model that is specific to him or her for future use with the same patient makes sense only in cases of long–term support such as the determination of Insuline levels for diabetic patients, Levadopa dosage for people suffering from Parkinson's disease, or the administration of anticoagulants for people with chronic heart disorders. In must other situations, it only makes sense to have a model that is generic, i.e., can be used for all patients suffering from a given type of disease. Unfortunately, biomedical systems exhibit a much larger variability in system behavior than those in most other application areas, and it is not easy to separate system–generic patterns from patient–specific patterns in observed data records. Many observations from the same type of application are usually needed in order to be able to filter out patient–specific behavior.

- **Incomplete information**: Biomedical data records are notorious for being incomplete. A patient on a specific monitor is routinely taken off the monitor while he or she is being cleaned by the nurse. A particular instrument may exist in one copy only. Although the instrument is currently in use by one patient, it is temporarily removed to be given to another patient who needs it more urgently. Thus, it is quite common

that data records contain big holes where no data is recorded for one or more variables during some period of time. It is necessary to be able to deal with such kinds of data sets. In other types of systems, e.g. mechanical or electrical, it may be possible to throw out incomplete data sets and repeat the experiment. In medical systems, this is rarely an option. A sufficiently long and rich data stream obtained from a patient is far too valuable an asset to being thrown out on the basis of incomplete data records alone.

Yet, there exists a second and quite different reason why medical data records are notoriously incomplete. Many important variables are not being recorded, either because the monitors in use are not equipped for permanent recording, or because these variables are not recorded at all by any monitors. For example, an anaesthetist may decide to change the amount of an anaesthetic drug to be administered to a patient on the basis of his or her skin coloring, or after looking at lacrimation levels. These variables are easy to be observed manually, but they are difficult to record, and this is therefore never done. Furthermore, the same anaesthetist may decide to reduce the dosage of an anaesthetic agent after being told by the surgeon that the operation will last for another 10 minutes only. This information is not something that *any* type of monitor could possibly pick up.

- **Different sampling rates**: Many biomedical types of measurements cannot be taken continuously, but are sampled at discrete points in time only. The sampling frequencies often differ from one variable to another, and not even an assumption of equidistant sampling holds true in all situations. During critical time periods, samples are usually drawn more frequently than during other time periods. In this situations, it is necessary to at least synchronize if not consolidate the data records of the different variables that have been sampled at different rates and have been recorded by different pieces of equipment, before modeling and simulation can even begin.

- **Diverse time constants**: It is quite common in biomedical domains that time constants of interest differ by one or two orders of magnitude. In such situations, the modeler is dealing with stiff systems, and such systems pose equally severe if not even more severe difficulties to qualitative simulation as to quantitative simulation. Hardly any results have been reported in the artificial intelligence literature about the effects of system stiffness on qualitative modeling and simulation.

- **Retrospectivity**: As soon as a system is to be modeled that contains variables whose values can be influenced by a conscious decision on the part of the object of the modeling effort, we are faced with the problem that the understanding of the object under investigation of what is going on may influence the very results of the effort. This phenomenon is called the retrospectivity of a system. It can pose a serious problem when modeling any system involving humans. During surgery, signals such as the heart beat or breathing rate of the patient are normally indicated by acoustic signals. This is useful since the eyes of the surgeon need to be focused on the patient, and not on a variety of monitors. Thus, any signal that can be picked up by the ears rather than by the eyes relieves the surgeon from having to get visual information from other sources than the patient. Unfortunately, not only the surgeon has ears, but also the patient, and this can create a hazard in situations where the patient is left conscious (e.g. surgery with epidural anaesthesia) or if the anaesthesia is not kept at a sufficiently deep level.

Due to all of the above reasons, data stemming from biological sources are commonly scarce and of poor quality, with large variability from one patient to another, but of boring monotonicity for any one patient during most of the time. Data are notoriously incomplete and frequently quite unreliable. Coming up with decent models for such systems is therefore an impossibly difficult task. *Knowledge–based* approaches to modeling usually fail because of a lack of meta–knowledge. *Pattern–based* approaches to modeling commonly fail due to a lack of data, both in quantity and in quality.

Several of the aforementioned factors are analyzed and discussed in this thesis from the fuzzy inductive reasoning methodology point of view, in the hope to at least alleviate some of the difficulties, possibly even eliminate them once and for all. The results of these efforts are presented in the subsequent chapters of this dissertation.

The previously mentioned difficulties are the most prevalent problems that we have to deal with when modeling and simulating biomedical systems. However, progress in biomedical engineering has been impeded by several other factors as well, factors that are not inherent in the medical sciences *per se*, but that are derived from human limitations and more specifically the difficulties of communication between the engineering researchers and the health care personnel, and between the health care personnel and their end users, the patients. Some of these factors are the following:

- Health care personnel usually offer resistance to the implantation of computer technology in their operating environment. Resistance in the medical community is sometimes attributed to the natural conservatism of physicians, as well as to a lack of understanding on the part of these physicians, what computer technology can to for them. Yet, the most important reason for the reluctance of clinicians to embrace modern computer technology is a quite different one.

  Especially in a country, such as the United States of America, where malpractice suits are a common and accepted way of boosting one's income levels, clinicians live under a constant fear of getting sued. In the early days of qualitative reasoning, this fear impeded the proliferation of artificial intelligence techniques in the medical field. If the clinicians had to take the full responsibility for their decisions, they at least wanted to be in full control of what these decisions were.

  Ironically enough, these very same considerations that originally had impeded introduction of artificial intelligence into the medical field, now make its use almost mandatory. It has happened several times that a physician got sued after a complication, because he or she had not ordered the most expensive and outlandish laboratory experiment to be performed on the patient, an experiment that might possibly have unveiled the problem before it occurred. Thus, physicians are more and more forced to order all kinds of experiments and use the most fancy monitors they can get, even if it hardly makes any sense, just to cover their own back. This is the main reason for the astronomic increase in health care cost in countries such as the United States. The conglomerate of fancy monitoring equipment in use generates a wealth of data that need to be screened in real time, because overlooking an essential piece of information provided by one of the monitors may be even more likely to trigger a malpractice suit than not using the monitor in the first place. So now the surgeons are suffering from human overload, and they badly need intelligent monitors and smart alarms that screen the incoming data for them and alert them if any of the patient's physiological signals look suspicious.

- On the other hand, the engineering researchers still have a frightening lack of understanding of the cognitive processes that underlie medical decision making. The decisions taken by a physician is founded not only in medical knowledge, but also and primarily in his or her gut feelings acquired through years of medical experience, knowledge that is vague and diffuse, and almost impossible to extract and codify in any systematic fashion. This makes it very hard for engineering researchers to

understand the decision making process in a logical way, and therefore, the deduction and synthetization of the relevant pieces of information needed to identify the model becomes an impossibly complicated issue.

In order to overcome these difficulties, a frequent and intense dialogue between the medical doctors and the engineering researchers is of paramount importance. Without such a dialogue, the engineers don't understand what the medical personnel needs, and the physicians don't understand what the engineers have to offer. It has been only in the very recent past that this dialogue has begun to emerge and flourish. This dialogue can lead to a deepened understanding of the other party's perspectives, which in turn is the only way to achieve profound mutual respect between the parties involved. It is this respect finally that allows to form well–functioning teams of cooperating experts that ultimately will be able to address and solve some of the most fascinating problems of this century.

5.3 Conclusions

This chapter has presented the main problems characteristic of biomedical applications. These problems make the modeling and simulation of biomedical systems a very difficult task of frightening complexity. The greater part of these problems cannot truly be solved, because they are inherent in the medical sciences. However, it is of major importance to understand these difficulties, and to be able to offer advice how these difficulties can be minimized. Without such understanding, the modeling and simulation of biomedical systems becomes a hopeless undertaking.

However, other factors beside the practical aspects of biological systems themselves contribute to the difficulties encountered when dealing with these types of systems. These problems are related to the difference in education between the engineering researchers, who are trained to think in terms of non–disputable mathematical logic, and the physicians, who are educated to rely on a wealth of imprecisely formulated clinical experience and draw conclusions by means of qualitative associations. The engineers have a very hard time to extract this clinical experience from the physicians and codify it in terms of their beloved facts and rules. The physicians, on the other hand, have an equally hard time to experience anything that the engineers present them with as being more than just a toy and of having clinical relevance.

It is the aim of the subsequent chapters of this thesis to address several of the aforementioned difficulties, and to come up with a methodology that, although being unable to do away with the system–inherent problems, at least can make use of the available data in an optimal fashion, and can do so in the context of realistic clinical applications and not only when confronted with toy problems.

Chapter 6

Qualitative Control of Biomedical Systems

6.1 Introduction

The aim of this chapter is to demonstrate by means of two different biomedical systems that the Fuzzy Inductive Reasoning (FIR) methodology can indeed be used to qualitatively capture the dynamic behavior of systems stemming from soft sciences.

In the previous chapter, the difficulties that arise when modeling biomedical systems for the purpose of predicting their future behavior had been enumerated. This chapter proposes how to tackle some of these problems, such as *lack of information from the point of view of quantity*, *lack of information from the point of view of quality*, and *diversity of time constants*, using the FIR methodology.

Also, it has been considered essential and valuable to be able to compare the results obtained using the FIR methodology with those obtained using other either quantitative or qualitative methodologies. In this manner, it will be possible to know not only if the FIR technique works, but also how it fares in comparison with these other methodologies. To this end, two biomedical problems are being tackled that had been previously addressed in the literature, one from the field of **Anesthesiology**, the other from the field of **Cardiology**.

Anesthesiology is an area of major interest in the medical domain. Ideally,

the aim of anaesthesia is to allow surgeons to perform a surgical operation on a patient within physiological normality while not causing any response in the autonomic nervous system of the patient due to stimulation. However, it is equally essential that the patient have no memory of the operation to prevent post–traumatic stress disorder (PTSD) syndrome. According to information obtained from an anaesthetist of the Can Ruti Hospital (Barcelona, Spain), the foremost source of incident reports after surgery is that patients have been operated on under a too light anaesthesia leading to PTSD syndrome because, although the surgeon was able to complete the operation without any technical difficulties, the patient remembers the operation (García, 1993). This is very difficult to detect during the surgery, since muscle relaxation, pain suppression, and awareness suppression are three quite separate and independent issues, and patients may be suffering pain or may be aware of what is going on without being able to express any signs of their pain or awareness.

The anesthesiology application presented in this chapter is the *control of an anaesthetic agent*, Isoflurane, during surgical operations, an agent that is predominantly being used for awareness suppression. The control of the depth of anaesthesia is a difficult undertaking, because the level of anaesthesia cannot be measured directly. Progress has been made during recent years by use of different monitoring systems that allow to estimate the depth of anaesthesia in indirect ways, i.e., by measuring other physiological variables and deducing the depth of anaesthesia from these observations. Often, artificial intelligence systems go one step beyond mere estimation of the depth of anaesthesia and directly suggest a safe amount of the anaesthetic agent to be administered to the patient, considering his or her individual conditions (such as weight, age, known heart conditions, etc.). Despite these improvements, anaesthetists still rely heavily on personal experience (gut feeling) when suggesting the anaesthetic dosage to be used during surgery.

A comparison between a neural network approach and the fuzzy inductive reasoning technique is also presented in the first part of this chapter. To this end, an already exiting system, ANNAD (Artificial Neural Network for Anaesthetic Dose determination), is taken as a starting point. This system had been previously developed in the United Kingdom (Linkens and Rehman, 1992a). As an alternative, a new system, FIRAD (Fuzzy Inductive Reasoning model for Anaesthetic Dose determination), has been developed by the author of this thesis (Nebot *et al.*, 1993a). The design of FIRAD is discussed in full detail here. The same data that had been used to drive ANNAD have also been used with FIRAD, in order to be able to objectively compare the results obtained from the two approaches.

Cardiology is an area of medicine that has been studied from many different angles, and with equally many different focus points. In the second part of this chapter, the *cardiovascular control system* is being investigated. The cardiovascular system is composed of the hemodynamical system comprising the heart and the blood vessels, and the central nervous system control that is responsible, among other things, for the beating of the heart.

The hemodynamical system can be modeled as a mechanical (hydraulic) system. The dynamics of this system have been extensively studied, and good differential equation models exist describing the behavior of the hemodynamical system in much detail.

In contrast, the central nervous controller is still not fully understood. It is quite clear which are the input output variables of the controller, and some differential equation models have already been proposed. However, and different from the hemodynamical differential equation models in use, these models are not based on an accurate knowledge of the precise structural relationships among the internal variables governing the controller. In our study, it was decided to generate a pattern–based input/output model of the central nervous controller by means of fuzzy inductive reasoning. The mixed quantitative and qualitative modeling and simulation technique, described in Chapter 4 of this dissertation, is required in order to be able to close the loop between the (quantitative) hemodynamical system model and the (qualitative) central nervous controller model.

A previous dissertation (Vallverdú, 1993) worked with two different quantitative central nervous controller models, a differential equation model and a NARMAX (Nonlinear AutoRegressive Moving Average with eXternal inputs) model. The hemodynamical model used in our own study is exactly the same that had been used by Vallverdú. Consequently, it will be possible to make quantitative comparisons, for a single patient, between our own FIR model and the two models developed by Vallverdú.

In the remainder of the chapter, these two biomedical studies are presented.

6.2 Anesthesiology

6.2.1 About Anesthesiology

Both sleep and general anaesthesia are states of unresponsiveness, which vary in depth. While sleep is healthy, natural, and repeats itself rhythmically once every 24 hours, anaesthesia is an artificial state maintained by the continuing presence of chemical agents in the brain.

Anaesthetic agents affect the respiratory system, the cardiovascular system, the central nervous system, and the muscles. The use of anaesthetic agents can produce severe complications and side effects, which, under extreme conditions, may even cause the death of the patient. It is therefore essential that the dose of anaesthetic agents is limited to the minimum amount necessary for proper anaesthesia thereby reducing undesired side effects and minimizing the risk to the patient.

Monitoring devices can be used to record the values of indicator variables, to reason about the consistency of these values, and suggest to the anaesthetist an appropriate dose of anaesthetic agent. Research results have recently been reported in the area of monitor integration that enhance the clinical robustness of such monitoring devices by improving their reasoning capabilities through the detection of critical events and by means of enhancing their alarm accuracy (Navabi *et al.*, 1991).

Several new results have been reported in the past few years relating to the control of the depth of anaesthesia. Both open–loop and closed–loop techniques have been explored (Linkens, 1992; Linkens and Rehman, 1992b).

One of these studies resulted in the development of a computer–based on–line expert system called RESAC (Real–time Expert System for Advice and Control) (Linkens *et al.*, 1986; Linkens *et al.*, 1990). RESAC comprises a rule–based backward chaining inference engine with about 400 rules and makes use of fuzzy logic and Bayesian reasoning. The rule base was obtained through knowledge acquisition in consultation with expert anaesthetists. RESAC has been tested during real surgeries, and its advice was found to be consistent with clinical necessities. The major problem of the approach underlying RESAC is the formidable size of the employed rule base. Evidently, such a system is of no practical use if it cannot provide advice in real time. As the understanding of the mechanisms underlying anaesthesia and their measurable effects grew, so did the size of the rule base that was integrated into the system, up to a point,

where a much faster and more expensive computer would have to be acquired in order for the system to still be able to function in a real–time setting.

Triggered by the aforementioned difficulties, another study was carried out by the same group that promised to enhance the run–time efficiency of the monitoring system. A new system, ANNAD — an Artificial Neural Network for Anaesthetic Dose determination, was created that was supposed to replicate the capabilities of RESAC (Linkens and Rehman, 1992a, 1992b). ANNAD employs a feedforward neural network trained through back–propagation. This work is being reviewed in the next section.

6.2.2 Background: ANNAD

The artificial neural network approach was chosen due to its inherent ability to learn the input/output behavior of a system in situations where it is possible to specify the inputs and outputs, but where it is difficult to define analytically a relationship between them. This is precisely the situation in biomedical applications, such as anaesthesia, since clinical signals are readily available through measurements, but no precise analytical relationships are known between them, and variations between patients are large. Also, neural networks are inherently parallel in nature, and are therefore well suited for real–time environments.

The clinical variables comprising heart rate (HR), respiration rate (RR), systolic arterial pressure (SAP), gender, age, and weight of the patient were selected as the key clinical indicator signals to be used for suggesting an anaesthetic dose (control signal).

A patient model and a controller model were independently synthesized by means of the neural network methodology. The control loop was then closed as shown in Figure 6.1.

6.2.2.1 Artificial Neural Network Patient Model

An Artificial Neural Network (ANN) patient model was obtained using a back–propagation algorithm applied to a set of data measured on a patient during a surgical operation.

Three separate neural networks were trained, one for each output: HR, SAP, and RR. The inputs for the training networks were the $Dose$, older (delayed)

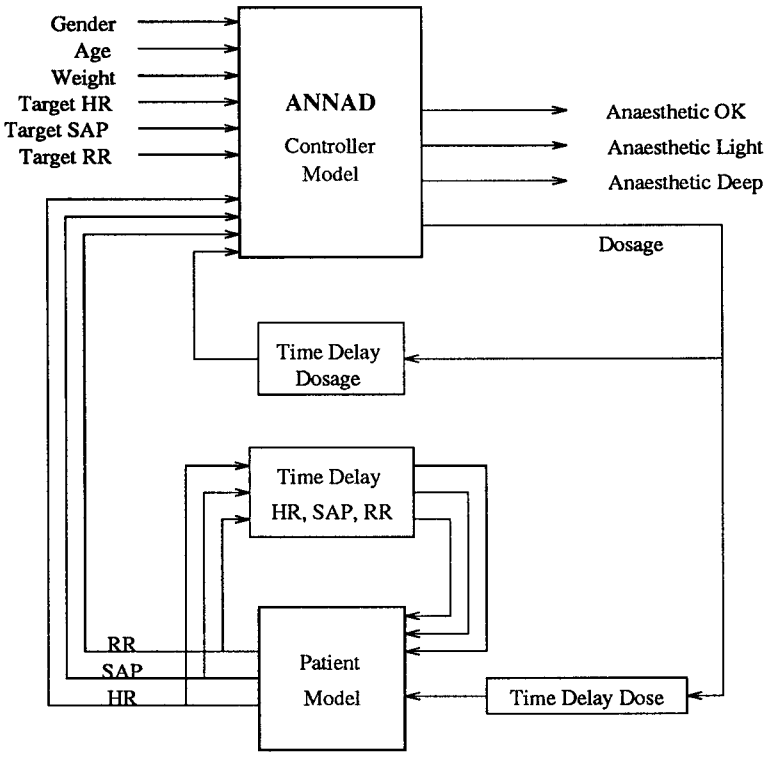


Figure 6.1: Feedback Loop Involving Patient Simulator and Drug Controller

values of the *Dose*, as well as delayed values of *HR*, *SAP*, and *RR*. Each neural network employed two hidden layers.

6.2.2.2 Artificial Neural Network Controller

An ANN controller model (ANNAD) was obtained using a back-propagation algorithm applied to another set of data collected from a second patient during a similar surgical operation as for the patient model. In this case, a neural network with three hidden layers was found to be optimal (showing the smallest deviation from the measured data after training). The inputs for this neural network were gender, age, weight, *RR*, *SAP*, *HR* and the desired values of the latter three variables, while the output was the anaesthetic agent, *Dose*.

The anaesthetic agent used in all these operations was *Isoflurane*, because the elimination of this agent by the body is quicker than that of other anaesthetic agents, such as Enflurane. Obviously, both the patient model

and the controller model depend heavily on the choice of the anaesthetic agent in use, and the process of model identification must therefore be repeated from scratch if another agent is to be used in the process.

6.2.2.3 Closed–loop Control

As shown in Figure 6.1, the control loop was then closed by connecting ANNAD with the ANN patient model. The results of this experiment demonstrate the stability of the control loop. ANNAD was able to replicate satisfactorily the advice that was obtainable from RESAC. ANNAD also produced good control performance when coupled to a patient simulator. Contrary to RESAC, which had actually and successfully been used during surgical operations, ANNAD has not yet undergone real–life testing.

For a deeper insight into this work, the reader is referred to (Linkens and Rehman, 1992b; Rehman *et al.*, 1993).

6.2.3 FIRAD

The motivation for the research described in this section was to investigate how the *fuzzy inductive reasoning* methodology would perform in comparison with the neural network approach when applied to the identification of dynamic processes from the soft sciences. To this end, we first tried to develop a fuzzy inductive reasoning model for the patient, and then to find a fuzzy inductive reasoning model for the controller. The controller model has been named FIRAD (Fuzzy Inductive Reasoning model for Anaesthetic Dose determination). The insights gained during this research effort are detailed in the following subsections.

6.2.3.1 SAPS Patient Model

The patient model should be identified from the qualitative relationship between its single input variable, the administered *Dose*, and its output variables, the clinical signals of the patient (*SAP*, *HR*, and *RR*) that reflect his or her body reaction to the amount of agent applied.

In order to determine the patient model, we worked with the data sets of two different patients. The available measurement data are plotted in Figure 6.2.

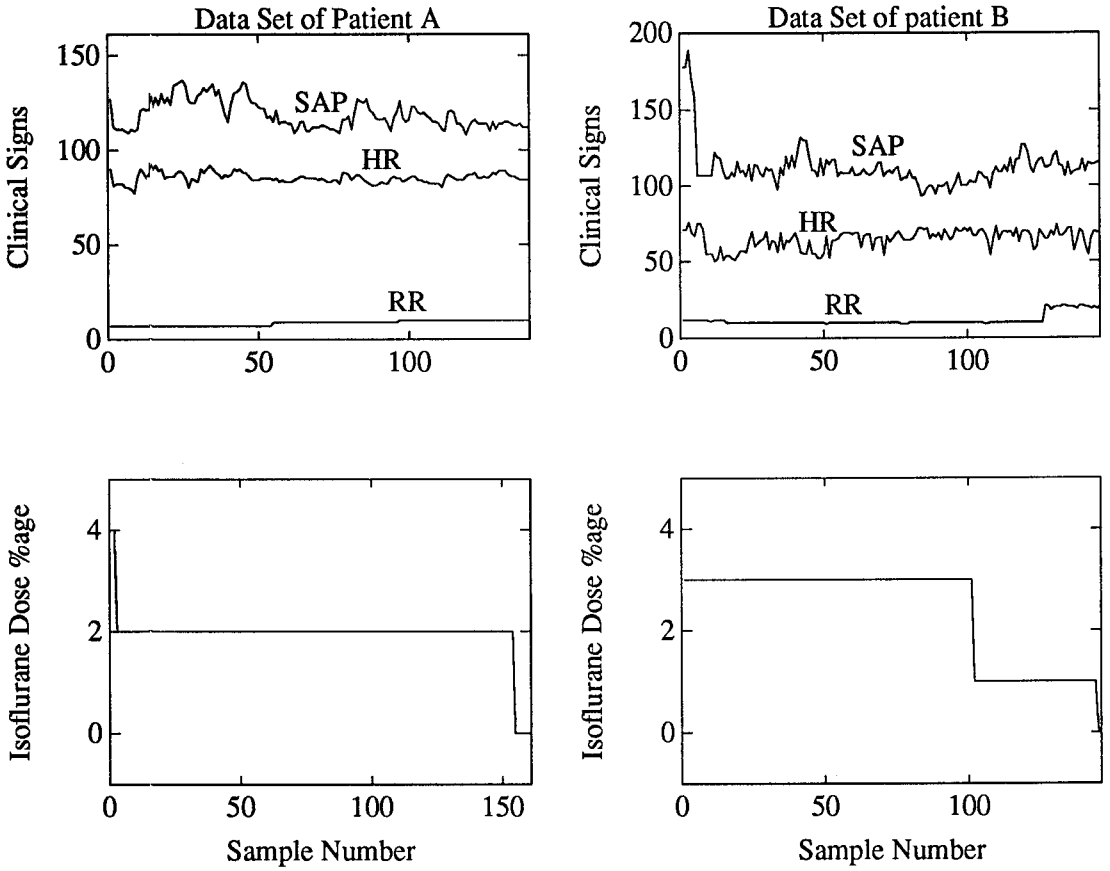


Figure 6.2: Patient Model Measurement Data

The two data sets are real data sets obtained from two patients undergoing surgery at a hospital in Glasgow.[1] These data were provided to a research group in Sheffield, the group that subsequently developed RESAC and ANNAD, from whom we finally obtained the data for use in our own research study. Data set A stems from a male patient, 35 years of age, weighing 50 kilograms. The data have been recorded during a renal transplant operation. Data set B stems from another male patient, 57 years of age, weighing 75 kilograms. This data set has been recorded during an abdominal–perineal resection operation.

These plots reveal that the input variable, *Dose*, varies very little over time. In both cases, it is initially "high," assumes a value of "medium" for most part

[1]The clinical studies were approved by the Western Infirmary Ethics Committee, Glasgow, Scotland, and the patients involved gave informed consent.

of the experiment, and goes "low" only at the very end of the experiment. It is quite clear that, in the meantime, the output variables react in various ways that are obviously not driven by the input directly, since the input does not change at all. The changes in the output variables were caused by other extraneous factors that were not recorded, and therefore, the variations in the output variables look like *noise* to the inductive reasoner. In fact, the recorded data do indeed contain considerable *digitization noise*, since all variables were recorded as integers only.

According to (García, 1993) and fortunately for the two patients, both surgeries went by boringly uneventful. Had the surgeon cut into any major blood vessel during the operation, the heart rate would have immediately jumped up to much higher levels, the blood pressure would have risen accordingly, and the anaesthetist would have had to react by increasing the amount of Isoflurane administered to the patient. Due to the monotonicity of the patient behavior (the anaesthetist considers such small variations in the patient signals basically as noise), he or she did the right thing: start out with a high level of Isoflurane to raise the amount of the drug in the blood to the desired level, then reduce the administered dose to a maintenance level for the duration of the surgery, and finally, when the surgeon told him or her that the operation was about to end, switch off the supply of anaesthetic agent and leave. Since nothing major happened during the surgery, the anaesthetist had no reason at all to intervene.

Since in neither of the two cases there exists a causal relationship between the administered *Dose* (the "input") and the physiological variables of the patient (the "outputs"), our fuzzy inductive reasoner could not find masks that would model the patient system in any meaningful way. The best masks found during the two searches, did not forecast correctly, and because the obtained input/output behaviors were not deterministic at all, the confidence in the forecasts dropped immediately to unacceptably low levels, and forecasting stopped almost at once. Thus, the built–in self–assessment capability of the FIR methodology came to play, and no forecast could be produced at all except through overriding the self–assessment feature by declaring that *any* forecast, irrespective of how unlikely it may be, would be considered acceptable.

This was not the case when using the neural network methodology. At least for one of the two data streams, Rehman reported that he had found a neural network that gave reasonable responses for that patient model (Rehman *et al.*, 1993). Evidently, since there really doesn't exist a causal relationship within the system, the match is purely coincidental. We consider it a dubious quality if a methodology produces accidental hits, especially in an application area where

little meta–knowledge is available, and where it is therefore difficult after the fact to distinguish an accidental hit from true success, and in an application area where few of the end users of the tool (i.e., the medical personnel) are educated to make this distinction. We therefore consider the self–assessment capability of the FIR methodology, which prevents accidental hits from being reported as success stories, one of the foremost assets of the FIR methodology.

As for all inductive techniques, inductive reasoners need a lot of data to work with. It is not possible to generate meaningful and reliable inductive models without ample and rich data. This is equally true for the neural network approach (another inductive modeling technique). However, while the neural network will always predict something, the inductive reasoner will not predict anything that cannot be validated on the basis of the available data. SAPS, our inductive reasoner, simply declines to predict anything when confronted with the patient model data, since no prediction can truly be justified given the available facts.

Here, we observe one of the *strengths* of the fuzzy inductive reasoning methodology. It will not generate models that are not justifiable from the given data. The neural network methodology generates models for *any* data, irrespective of whether they are justifiable or not. While SAPS contains an inherent model validation mechanism inside the simulation methodology, the neural network approach does not. The fact that the neural network was able to produce a reasonable response for one of the data sets does not mean that the model is validated. The fact that it was unable to produce a reasonable response for the other data set indicates just the opposite. Since inductive models necessarily lack physical insight, we believe it to be absolutely *essential* for any inductive modeler to contain an intrinsic model validation mechanism that is inseparable from the modeling/simulation tool itself. Our fuzzy inductive reasoner, SAPS–II, offers such a mechanism.

6.2.3.2 SAPS Controller Model

The controller model is determined by the qualitative relationship between its three "input" variables: *SAP*, *HR*, and *RR*; and its single "output" variable: the administered *Dose*.

We could, of course, have worked with the data sets shown in Figure 6.2 for the identification of the controller model, after simply swapping inputs and outputs. However, as it was explained earlier, also this system does not contain sufficient causality between its inputs and outputs. The only causality that

exists is the fact that nothing major happened to the three input variables, thus no drastic measures were indicated, and so, the "lazy" anaesthetist didn't react at all to the variations in the input variables. The experiment had supplied us with *ample* data, but not with *rich* data.

Thus, already the Sheffield group decided to go a different route. Since the purpose of ANNAD was to replicate the behavior of the, already field–tested, RESAC system, the physiological variables of the two patients, preprocessed using a filter with low–pass characteristics, were fed to RESAC, and RESAC's reaction was recorded. The two data sets obtained in this fashion have been plotted in Figure 6.3.

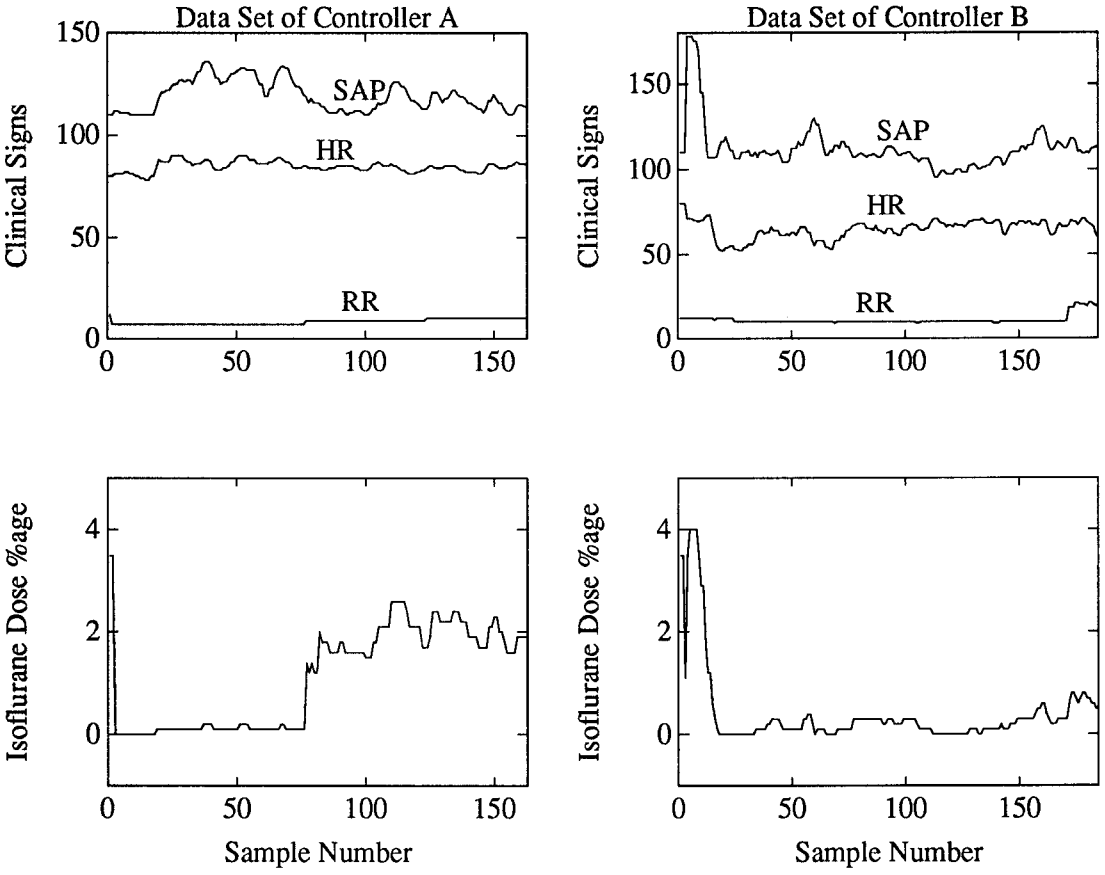


Figure 6.3: Controller Model Measurement Data

The "input" variables are not exactly the same as the "output" variables of the previous experiment. The data had been filtered prior to feeding them to RESAC in order to reduce the noise level inherent in these data.

Looking at the plots, we can see that the new output variable, *Dose*, varies here considerably more than in the clinical data sets that had been used for the patient model. The human anaesthetist didn't find the variations in the biological variables (*HR*, *RR*, and *SAP*) alarming, and therefore, reacted very little during the entire operation. Fortunately for us, RESAC is more "industrious" than a human anaesthetist would ever be, and reacts to small variations in the biological variables by recommending a slightly modified *Dose* of the anaesthetic drug. Therefore, there now exists a direct causal relationship between the observed biological data and the recommended *Dose*, and it should, therefore, be possible to correlate the administered *Dose* with the biological variables, and come up with a causal inductive model that can be used to replace the anaesthetist (or RESAC) in his or her (its) decision making process.

Notice that the *Dose* recommended by RESAC is quite different from that recommended by the anaesthetist. To an engineer, who is used to much more precise and unambiguous information, the two signals look even qualitatively quite distinct, and he or she would therefore be inclined to conclude that RESAC has failed to make a meaningful suggestion for the appropriate amount of Isoflurane to be administered. However, medical personnel is much more generous than engineers are in their interpretation of what is similar or acceptable. RESAC *has* already been field–tested, i.e., anaesthetists have observed RESAC at work during several real surgeries in a real operating theater and have concluded that the recommendations made by RESAC were clinically acceptable in all cases. Consequently, the chosen approach makes sense.

The two data sets contain 163 and 185 records, respectively. They were sampled once per minute. According to information obtained from the Sheffield group (Linkens, 1993), the slowest time constant of interest in our system is on the order of 10 minutes, and the fastest time constant of importance is on the order of one minute.

In accordance with Shannon's sampling theorem we should therefore use a sampling rate of approximately one sample every 0.5 minutes. For this reason, before starting to identify an optimal model, we decided to use a Spline interpolation to find one new data record per interval, located exactly in the middle between the two neighboring measurement data records. Thereby, the length of the data records was enhanced to 325 and 369 records, respectively. We can assume that not too much information was lost by sampling once instead of twice per minute, because the plots reveal that the data vary sufficiently slowly.

It was decided to recode (discretize) the variables *SAP*, *HR*, and *Dose* into three qualitative levels (classes), whereas *RR* was recoded into two qualitative levels only. This decision has been taken on the basis that, in any class analysis, each legal discrete state should be recorded at least five times, see Equation 6.6 (Law and Kelton, 1990).

Due to the fact that there are available less than 400 data points, it is not possible to recode each variable in three classes. A maximum of three of the four system variables can be recoded into three classes, whereas the fourth one has to be recoded into two classes. It was decided that the respiration rate would be recoded into two classes because its behavior contains less variation than that of the other variables.

Due to the difference between the slowest and the largest time constants of importance, it was necessary to use a mask candidate matrix of depth 21 with nine zero rows in between rows that contain potential inputs.

$$
\begin{array}{l|cccc}
{}_t\backslash^x & SAP & HR & RR & DOSE \\
t-20\delta t & -1 & -1 & -1 & -1 \\
t-19\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-11\delta t & 0 & 0 & 0 & 0 \\
t-10\delta t & -1 & -1 & -1 & -1 \\
t-9\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 & 0 & 0 \\
t & -1 & -1 & -1 & +1
\end{array}
\tag{6.1}
$$

In this way, one new forecast is produced every 0.5 minutes, satisfying Shannon's sampling theorem, and yet, the inductive reasoner looks at input values 5 minutes and 10 minutes back to capture the slowest time constant. This technique, that permits to operate on data at different sampling rates, has proven successful in the past (Cellier *et al.*, 1992).

The first 270 (320) rows of the data matrix were used as past history data to compute the optimal mask. Fuzzy forecasting is used to predict new qualitative class and fuzzy membership values for *Dose* for the last 55 (49) rows of the raw data matrix, respectively.

For the first data set, the optimal mask obtained was the following:

$$
\begin{array}{c|cccc}
{}_t\backslash^x & SAP & HR & RR & DOSE \\
t-20\delta t & 0 & 0 & 0 & 0 \\
t-19\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-11\delta t & 0 & 0 & 0 & 0 \\
t-10\delta t & 0 & 0 & 0 & -1 \\
t-9\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 & 0 & 0 \\
t & -2 & 0 & 0 & +1
\end{array}
\tag{6.2}
$$

This mask denotes the qualitative relationship:

$$Dose(t) = \tilde{\mathrm{f}}(Dose(t-10\delta t), SAP(t)) \tag{6.3}$$

For the second data set, the optimal mask obtained was:

$$
\begin{array}{c|cccc}
{}_t\backslash^x & SAP & HR & RR & DOSE \\
t-20\delta t & 0 & 0 & 0 & 0 \\
t-19\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-11\delta t & 0 & 0 & 0 & 0 \\
t-10\delta t & -1 & 0 & 0 & -2 \\
t-9\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 & 0 & 0 \\
t & 0 & -3 & 0 & +1
\end{array}
\tag{6.4}
$$

This masks denotes the qualitative relationship:

$$Dose(t) = \tilde{\mathrm{f}}(SAP(t-10\delta t), Dose(t-10\delta t), HR(t)) \tag{6.5}$$

It turns out that the two masks obtained are *different*. Although RESAC used the same causal reasoning, SAPS decided that, by proposing a different causal relationship in the two cases, the quality of the forecast can be improved. The proposed controller is thus different for each of the two patients.

Since our gauge is the decision making process of RESAC, it makes sense to use the best possible mask, i.e., the mask that produces results that are

as consistent as possible with those obtained from RESAC. However, both optimal masks will produce answers that a human anaesthetist would consider clinically plausible, and so, for practical purposes, it doesn't really matter which of them we use.

In reality, human anaesthetists use imprecise patient models (gut feeling) in their decision making. Thus, different optimal masks correspond to slightly inconsistent decision making, a fact that we are well prepared to accept since human decision making is never fully consistent. Different anaesthetists may decide differently when exposed to the same data, and the same anaesthetist may decide differently depending on e.g. the number of hours he or she slept the night before, or depending on whether he or she had gotten into an argument with his or her spouse the evening before, or simply, depending on the current mood he or she is in. There is no such thing as a "correct" *vs.* an "incorrect" decision. We can only talk about "clinically acceptable" *vs.* "clinically unacceptable" decisions. Within the range of the clinically acceptable decisions, it doesn't matter too much how the decision is drawn. Thus, SAPS is simply being realistic in its assessment.

One fact that is common to both optimal masks is that the output of the controller model depends on the amount of previously administered anaesthetic agent. This is clinically plausible since the chemical substance accumulates in the patient for some time.

The forecast results for the two data sets are shown in Figure 6.4.

The results are quite good. The optimal masks contain sufficient information about the behavior of RESAC to be used as a valid controller of the dosage of Isoflurane given to the patient. In contrast, Rehman reported that the neural network gave good responses for the controller model only for one of the two data files.

From these results, we can conclude that the SAPS methodology is fairly *robust*, i.e., it consistently generates a decent inductive model whenever the data allow it to, and it categorically declines to generate a model if the available data do not permit to validate an inductive model.

The neural network approach is different in this respect, since it uses a gradient technique (back–propagation) for optimization in the original (i.e., continuous) search space, whereas SAPS uses an exhaustive search in a reduced (discrete) search space. Therefore, it is perfectly feasible that the neural network does not converge (as it happened with one of the data records), whereas SAPS will come up with the "best possible" model (within the

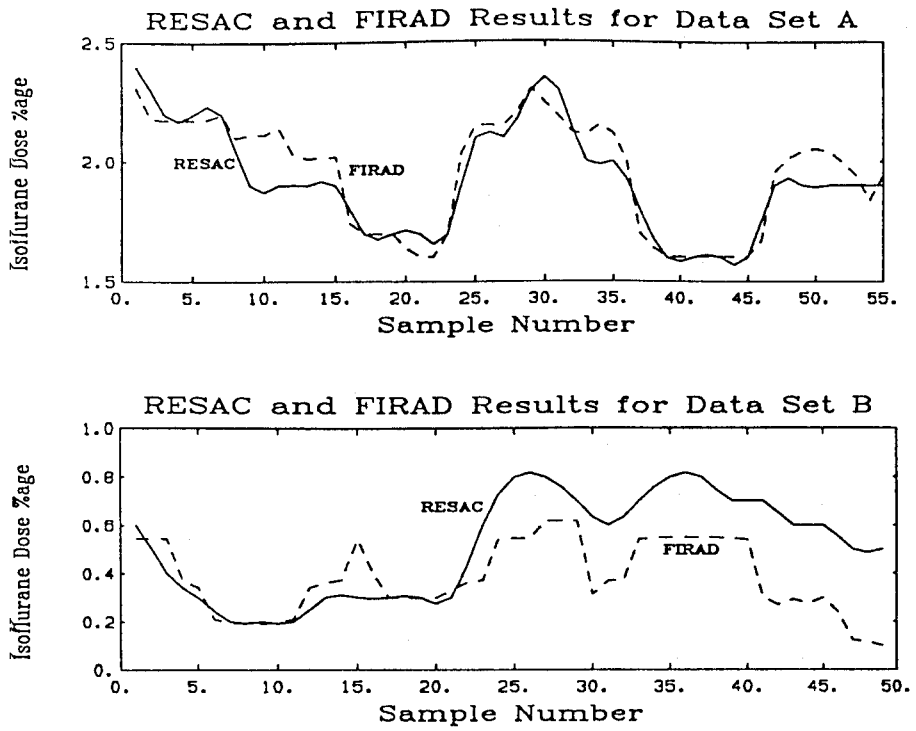


Figure 6.4: Comparison RESAC/FIRAD

framework of the discrete search space) whenever the data justifies a model.

The attentive reader may object that SAPS meticulously has identified the "noise" while ignoring the real "signal," i.e., that SAPS has generated a model of the small, and medically insignificant, changes rather than a model of the big, medically important, changes. Such an objection would be justified. SAPS could not identify a model of dramatic changes, since no dramatic changes ever took place during either of the two operations. However, this is not a weakness of the FIR methodology, only a shortcoming of the available data sets. SAPS simply exploits all the (linear and nonlinear) correlation that exists between inputs and outputs, and comes up with a dynamic model that makes use of this information in an optimal manner. The reader ought to remember that the main emphasis of this thesis is to demonstrate the applicability of the chosen modeling and simulation methodology when dealing with soft science systems, rather than to discuss in great detail a biomedical problem taking into consideration every facet of the application at hand.

Would it have been possible then to simply swap inputs and outputs of the data sets shown in Figure 6.3, i.e., use the RESAC–generated data to obtain also a patient model? The answer to this question is no. Causal modeling is an extension to the concept of *uni–valued functions*. Given the function $y = \sin(x)$, it is always possible to find a unique value of y for any given value of x, because $\sin(x)$ is a uni–valued function. On the other hand, $x = \sin^{-1}(y)$ is multi–valued, and therefore, it is not possible to conclude a unique value of x given a value of y. It therefore makes sense to call a uni–valued function a *strictly causal function*, whereas a multi–valued function is not a strictly causal function.

Causal modeling is an extension of this concept. A causal model is a uni–valued (deterministic) function that accepts, as inputs, not only current values of its input variables, but also past values of all its inputs as well as its output. The controller model is obviously a fairly causal model, since the anaesthetist (or RESAC) bases his or her (its) decision making in a semi–deterministic (fully–deterministic) fashion on the available inputs. The reverse, however, is not true. It is not evident that it is possible to conclude current and future values of the biological parameters in a unique fashion from measurements of their own past, and from current as well as past *Dose* values, and SAPS indeed concludes that this is not a meaningful proposition.

6.2.4 Comparison of Results from the two Modeling Methodologies

Before comparing the results obtained from ANNAD and FIRAD, we wish to make a comment about the SAPS methodology. The original idea was that FIRAD should forecast the Dose during 63 minutes in order to obtain the same plot length as was shown in the previously published ANNAD report (Linkens and Rehman, 1992a). This was not possible because SAPS needed more data points (training data) than the neural network methodology to identify a model.

Previous investigations involving SAPS, presented in Chapter 3 of this thesis, have led to a recommendation with respect to the minimum number of data records to be used in the identification of an inductive model. This rule is based on statistical considerations, and states that, in any class analysis, each (discrete) state should be recorded at least five times (Law and Kelton, 1990). Thus:

$$n_{rec} \geq 5 \cdot \prod_{\forall i} k_i \tag{6.6}$$

where n_{rec} denotes the total number of recordings, i.e., the total number of observed states, i is an index that loops over all variables, and k_i denotes the number of levels (i.e., discrete class values) of the variable i.

In the given application, the number of suggested records is:

$$n_{rec} \geq 5 \cdot (3 \cdot 3 \cdot 3 \cdot 2) = 270 \tag{6.7}$$

Consequently, the first 270 data records should be used for model identification, which, for the first data stream, leaves us with only 55 records, or 27 minutes worth of measurement data for forecasting.

To improve the situation, tests were made to find the minimum number of records needed to identify the same controller model that was found using the set of 270 records. It was determined that, if at least 240 records are used for identification, the same controller model can still be found. This then allows us to forecast the system over the last 43 minutes of the recorded data. This forecast may be compared with the forecast obtained from ANNAD and with the original *Dose* recommendations made by RESAC. The comparative results are presented in Figure 6.5.

As can be seen from this plot, FIRAD forecasts the *Dose* quite well, even a good deal better than ANNAD. Thus, the fuzzy inductive reasoning methodology has been shown to be able to synthesize, at least for this example, an inductive biomedical model that works equally well or better than a neural network would.

6.2.5 Summary of the Relevant Issues

The results shown in this section confirm the ability of the FIR methodology to produce good control performance of the anaesthetic agent delivery system. For the two patients used in the experiment, the FIRAD software not only replicated the advice from RESAC correctly, but it performed even more accurately, more reliably, and more consistently than the ANNAD software.

Evidently, the selected approach will never permit us to obtain results that are better than those obtained by RESAC, since RESAC was used as the

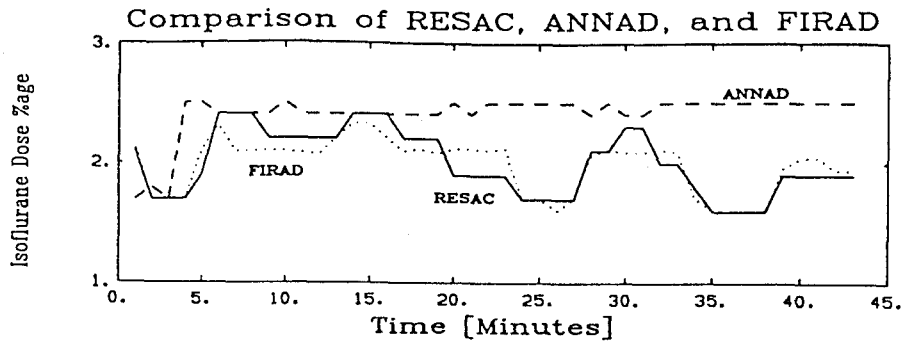


Figure 6.5: Comparison RESAC/ANNAD/FIRAD

"system" from which FIRAD tried to learn. Hence a comparison of FIRAD and ANNAD allows us to relate their abilities at capturing the dynamic behavior of the anaesthetic agent delivery system to each other, but FIRAD and RESAC cannot be compared with each other for the purpose of judging their relative abilities. In order to be able to compare the latter two approaches to each other, the FIR model would have to be based on real measurement data, data that are sufficiently ample in quantity and rich in quality to permit a FIR model to be developed from them. Such data were, unfortunately, not at our disposal.

Contrary to the neural network approach, the inductive reasoner contains information about the likelihood of any particular state transition. This is important for model validation purposes. If the accumulated likelihood of a particular state drops below a level to be specified by the user, forecasting will come to an immediate halt. In this manner, the technique guarantees that the model will not forecast behavior beyond a time for which the available data are insufficient to substantiate the prediction. Also and similar to some of the knowledge–based approaches to qualitative modeling/simulation, the FIR technique is able to enumerate *all possible* system behaviors that are consistent with the available knowledge (data), and it can furthermore assign a measure of likelihood of occurrence to each of them.

This is why the SAPS methodology would not generate a patient model that is not justifiable from the given data. We consider this intrinsic model validation mechanism a *distinct and significant advantage* in comparison with the neural network methodology, especially in the context of soft sciences.

For the controller model, two different masks have been found from the two different data sets available. Although RESAC used the same causal reasoning, SAPS decided that, by proposing a different causal relationship in the two cases, the quality of the forecast can be improved.

However, from a medical point of view, it would be much more useful to have available a unique model for a class of similar patients undergoing similar operations. In that manner, it would not be necessary to wait until the first operation of a given patient is over in order to have data available that would allow the identification of a controller model to be used for the same patient during a later surgery of the same or a similar kind. If we are able to obtain a decent model that can be used for all similar patients undergoing similar operations, it would be possible to use this model for a new patient of this set, i.e., for patient who never underwent surgery before. This is an interesting and important issue that will be discussed in more detail in Chapter 9 of this dissertation.

Was this a success story? From an artificial intelligence perspective, it most certainly was. Not only was the FIR methodology (FIRAD) shown to be able to replicate the behavior obtained by a neural network (ANNAD), but it outperformed the latter by leaps and bounds. The quality of the forecast of FIRAD is considerably better than that of ANNAD, and FIRAD is much more reliable. It *always* works when it is supposed to work, and it avoids the pitfalls of accidental hits. Also, the training of FIRAD is considerably faster than that of ANNAD. We were not able to provide quantitative information about this fact since the report by Rehman does not provide any numbers to this effect, but a SAPS model can usually be trained within a few minutes of CPU time, whereas a back–propagation neural network usually takes hours if not days of CPU time to train (Korn, 1991). Consequently, it is much more feasible using the FIR approach to try out different alternatives, e.g., look at different suboptimal masks, to see which one offers the best forecast. Retraining a feedforward neural network is so painfully slow that most researchers would shun the expense.

From a clinical perspective, the success is less evident. We cannot claim that FIRAD has learned everything that RESAC knows. An inductive reasoner can only replicate behaviors it has seen before. Yet, we know that all that was

shown to FIRAD are two boringly uneventful trajectories. Would FIRAD know how to react in the case of a true emergency? It most certainly wouldn't. In our opinion, it would have been better to forget the clinical data altogether and to drive RESAC by streams of random noise of clinically meaningful magnitude to exert the software "at all frequencies" (whatever that may mean in the context of an expert system). In this way, it would have been possible (and even cheap) to obtain arbitrarily long sequences of data streams that exhibit every possible mode RESAC con operate under, from which both a neural network and a FIR model could have been deduced that would then truly replicate the behavior of RESAC in a clinically meaningful and significant way. However, since we had no access to either RESAC or ANNAD, we had to live with the limited data sets we had been provided with. In addition, this approach would have allowed us to also make use of the additional input variables that RESAC is working with (gender, age, and weight), which we were unable to use in the FIR model described in the previous sections of this chapter.

6.3 Cardiovascular System

6.3.1 About the Cardiovascular System

Understanding of the central nervous system is essential because it is of paramount importance to the functioning of the human body. It comprises among others the signals that are transmitted from the brain to the heart and to the blood vessels, controlling the hemodynamical system. Failure of the central nervous system to deliver the triggering impulses to the heart results in an almost immediate death of the person.

In this section, a model of the Central Nervous System (CNS) control of the cardiovascular system is developed. The CNS control and the hemodynamical system compose the cardiovascular system. A simplified diagram of the cardiovascular system is presented in Figure 6.6.

In order to validate the cardiovascular system model, it is of high utility to have experimental data from a patient carrying out the *Valsalva maneuver* due to the fact that, under this scenario, all control mechanisms that belong to the central nervous system operate in a significant way. The validation of the cardiovascular system for one particular patient is described in Section 6.3.4 of this chapter.

The Valsalva maneuver has been described by *Antonio M. de Valsalva*

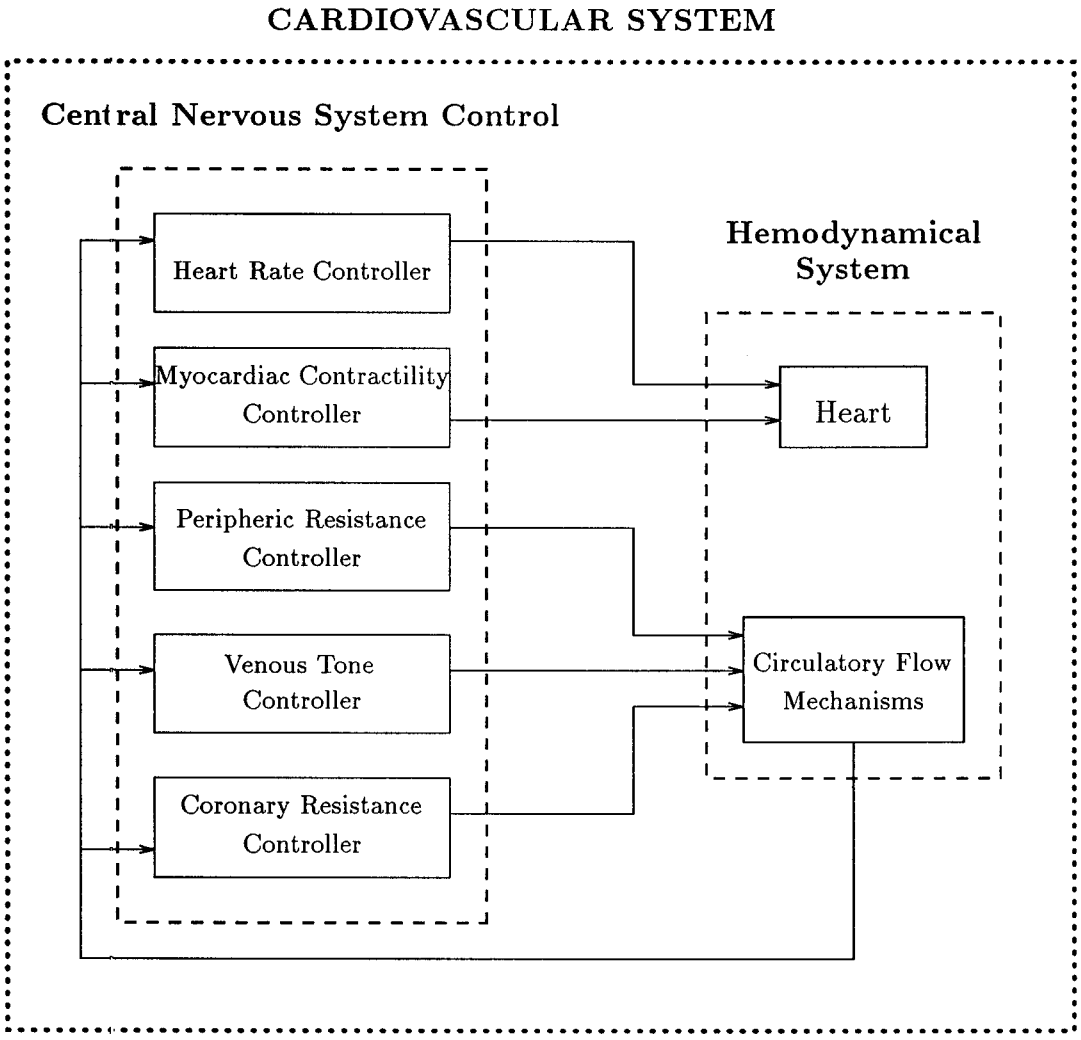


Figure 6.6: Simplified Diagram of the Cardiovascular System (Adapted from (Vallverdú, 1993))

(1704). It was originally used to study the ear by means of increasing the pressure in the Eustachian tube thereby mobilizing the tympanic membrane. The Valsalva maneuver consists in attempting a brisk exhalation with nose and mouth closed, blocking the passage of air, and thereby provoking the increase of intra–thoracic and intra–abdominal pressures.

This maneuver has been, and still is, a useful test in cardiology, because it causes important hemodynamical changes in a short time span, can be carried out easily and painlessly by any patient, and does not have any undesirable side effects. It is used not only for cardiac diagnosis, but also for the evaluation of CNS control performance.

A Valsalva maneuver is presented in Figure 6.7, with the temporal evolution of the right auricular pressure $P_{AD}(t)$, the aortic pressure $P_A(t)$, and the coronary flow $F_C(t)$.

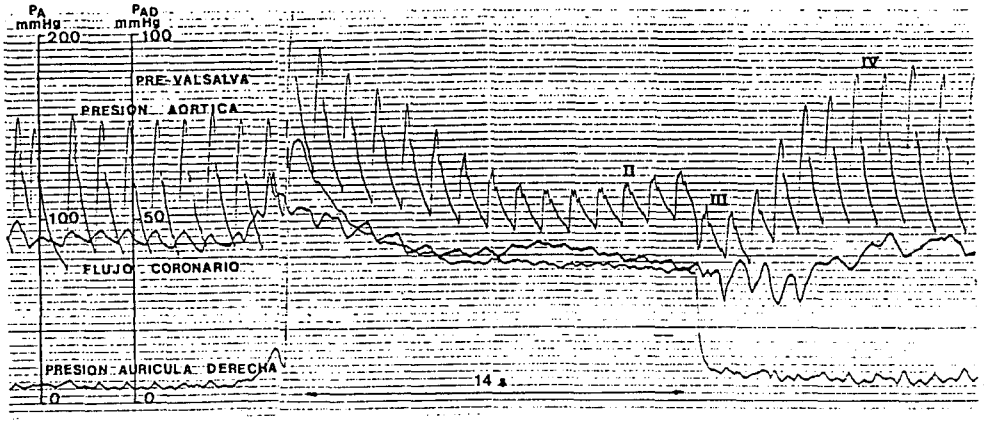


Figure 6.7: Example of Valsalva's Maneuver (Provided by M. Vallverdú and obtained from the Hemodynamical Division of the *Hospital de la Santa Creu i Sant Pau* in Barcelona)

In Figure 6.7, the four phases that comprise Valsalva's Maneuver are presented as phases I, II, III, and IV. Each phase corresponds to different

cardiovascular repercussions. Phase I occurs just after the onset of the maneuver, phase II takes place just before the effort is concluded, phase III corresponds to the ending of the maneuver, and finally, phase IV occurs after the attempt has been aborted.

6.3.2 Hemodynamical System

Over the years, the mathematical models describing the hemodynamical system have grown in size and complexity, proportional to the computational capacity of the available computers and progress made in cardiovascular system clinical research. Elaborate models of the arterial, vein, and cardiac systems that together form the hemodynamical system have been developed by researchers such as Beneken, Snyder, Rideout, and Sagawa (Beneken and Rideout, 1968; Snyder and Rideout, 1969; Sagawa, 1973). They are in compliance with the laws of fluid mechanics.

A recent and very detailed hemodynamical model was developed in (Vallverdú, 1993). In this model, the heart is composed of four chambers, modeled from the relations between pressure, volume, and elasticity variables. This model is influenced by the publications of Leaning (Leaning *et al.*, 1983) and Suga (Suga *et al.*, 1974) as well as those of a few other authors. The hemodynamical system has been widely studied, and its mechanisms are quite well understood. They are not fundamentally different from those of a hydro–mechanical pump. There only exist much larger parameter variations from one specimen to the next than would be the case among hydro–mechanical pumps.

It does not make much sense to use a qualitative methodology to identify a hemodynamical model, since no new knowledge can be acquired in this way. Consequently, the quantitative hemodynamical model presented in (Vallverdú, 1993), described through a set of differential equations, has been adopted in our study.

6.3.3 Central Nervous System Control

The central nervous system is composed of the spinal cord, the brain, and the hypothalamus. One of the systems that the central nervous system controls is the hemodynamical system by generating the regulation signals for the blood vessels and the heart. These signals are transmitted through the bundles of sympathetic and parasympathetic nerves, producing stimuli in

the corresponding organs and other body parts. Sympathetic stimuli cause an increase in the cardiac frequency and the contraction of the heart, as well as blood flow resistance in the peripherical vascular system and the venous tone. Parasympathetic stimuli, on the other hand, reduce the pumping of the heart.

The functioning of the central nervous system is of high complexity and not yet fully understood. This is the reason why many of the cardiovascular system models developed so far have been designed without taking into account the effects of its CNS control.

Although the central nervous system control is, at present, still not completely known, individual differential equation models for each of the hypothesized control mechanisms have been postulated by various authors on the basis of measurements obtained from experiments with dogs. Yet, these models are less than perfect, and in many respects, they fall short of providing a true understanding of the structural mechanisms responsible for the control actions inside the physiological system. The use of inductive modeling techniques with their reduced explanatory power but enhanced flexibility for properly reflecting the input/output behavior of a system may offer an attractive alternative to these differential equation models. It is the aim of this section to apply the FIR methodology to find a qualitative model of the CNS control that accurately represents the input/output behavioral patterns of the CNS control that are available from observations for a particular patient.

To this end, the previous research efforts of Vallverdú are taken as a starting point (Vallverdú, 1993). She developed two separate CNS control models, a differential equation model and a NARMAX (Nonlinear AutoRegressive Moving Average with eXternal inputs) model. The differential equation model developed by Vallverdú is an enhancement of many individual previous research efforts described by various authors, and certainly represents one of the most complete deductive CNS control descriptions currently available. The NARMAX model is an inductive model that shares many of the advantages and shortcomings of neural network (NN) models. Just like the NN models, the NARMAX model is basically a quantitative model with slow training capabilities but easy adaptation possibilities. Both NN and NARMAX models may predict anything when provided with inputs outside the range for which the model had been trained. The NARMAX model developed by Vallverdú contains parameter values that are specific for every analyzed patient, yet a common equation structure that was optimized using the data from five different patients.

The CNS control is composed of five separate controllers: the *heart rate controller*, the *peripheric resistance controller*, the *myocardiac contractility controller*, the *venous tone controller*, and the *coronary resistance controller*.

Some of the controller signals cannot be measured directly from human patients, and dog experiments are not viable for this purpose. In qualitative modeling, it would be possible to hypothesize causal relationships between variables that are measurable, and probably, this would work quite well. However, in the interest of being able to compare our results with those previously obtained by Valverdú, it was decided not to deviate from the approach chosen by her. The signals obtained from the simulation of the CNS control modeled with differential equations were used by Vallverdú as initial data for the identification of the NARMAX models common, in structure, to a set of different patients, and were also used as initial data for the identification of the FIR models for a single patient. This may be an acceptable solution since the differential equation model, from which the data were generated, had been previously validated.

Of course, the question must be raised whether it makes any sense to identify an inductive model if a deductive model is needed first. Evidently, we cannot expect to obtain in this way models that are better than the differential equation model that was used as the basis for their construction.

It is always a virtue to work with the *simplest models possible* that explain the available data in order to make potentially necessary patient–specific adjustments to the models as quickly and painlessly as possible. It serves no purpose whatsoever to carry in the models highly sophisticated parameters that are conceptually satisfying, the values of which are however impossible to determine either through direct measurements or through indirect parameter identification techniques. The NARMAX models are extremely simple in their internal structure, and therefore satisfy the requirement for simple models. The FIR models are less simple, but they are non–parametric anyway, and setting up a new FIR model can be done easily and quickly. Both types of inductive models are much more manageable than the differential equation model.

Although in Vallverdú's work it was deemed necessary to obtain a structurally unique model, it was not required that this model also contain the same parameter values for all the available patients. However, for other purposes, the parametric dependence of a model can constitute a major difficulty. A discussion of what could be done in the context of the FIR methodology to alleviate this problem will be presented later in this thesis.

As had been explained earlier, it would have been perfectly feasible to generate a qualitative FIR model on the basis of true measurement signals alone. This approach might have made more sense in some ways, but it would have had drawbacks as well. Most of the variables at the interface between the (qualitative) CNS control model and the (quantitative) hemodynamical model are not directly measurable. Thus, a model derived from measurements alone would not have allowed us to gain such a clear–cut separation between these two models.

After these methodological explanations, let us now return to the task at hand. Five controllers have to be identified. All of them are of the single–input single–output (SISO) type. In each case, the input signal is the *carotid sinus pressure*. The outputs of the five controllers are respectively: the *heart rate control signal*, the *peripheric resistance control signal*, the *myocardiac contractility control signal*, the *venous tone control signal*, and the *coronary resistance control signal*.

The input and output signals of the system had originally been recorded with a sampling rate of 0.12 seconds. In Figures 6.8 and 6.9, the input and output signals of a specific patient suffering from an at least 70% coronary arterial obstruction are presented.

In the modeling process, the normalized mean square error (in percentage) between the simulated output, $\hat{y}(t)$, and the system output, $y(t)$, is used to determine the validity of the models. This error is described in Equation 6.8.

$$MSE = \frac{E[(y(t) - \hat{y}(t))^2]}{y_{\text{var}}} \cdot 100\% \tag{6.8}$$

where y_{var} is the variance defined as:

$$y_{\text{var}} = E[y^2(t)] - \{E[y(t)]\}^2 \tag{6.9}$$

This error measure will also be used to compare the quality of the models for a single patient obtained by the NARMAX *vs.* FIR methodologies.

The signals used for identifying the NARMAX models for all the controllers for a single patient are plotted in Figures 6.8 and 6.9. It is desirable to use the same data to identify the FIR models for the same patient. Each of these signals contains 7279 data points. As the currently used version of SAPS is limited in size to raw data models containing no more than 5000 data records, it

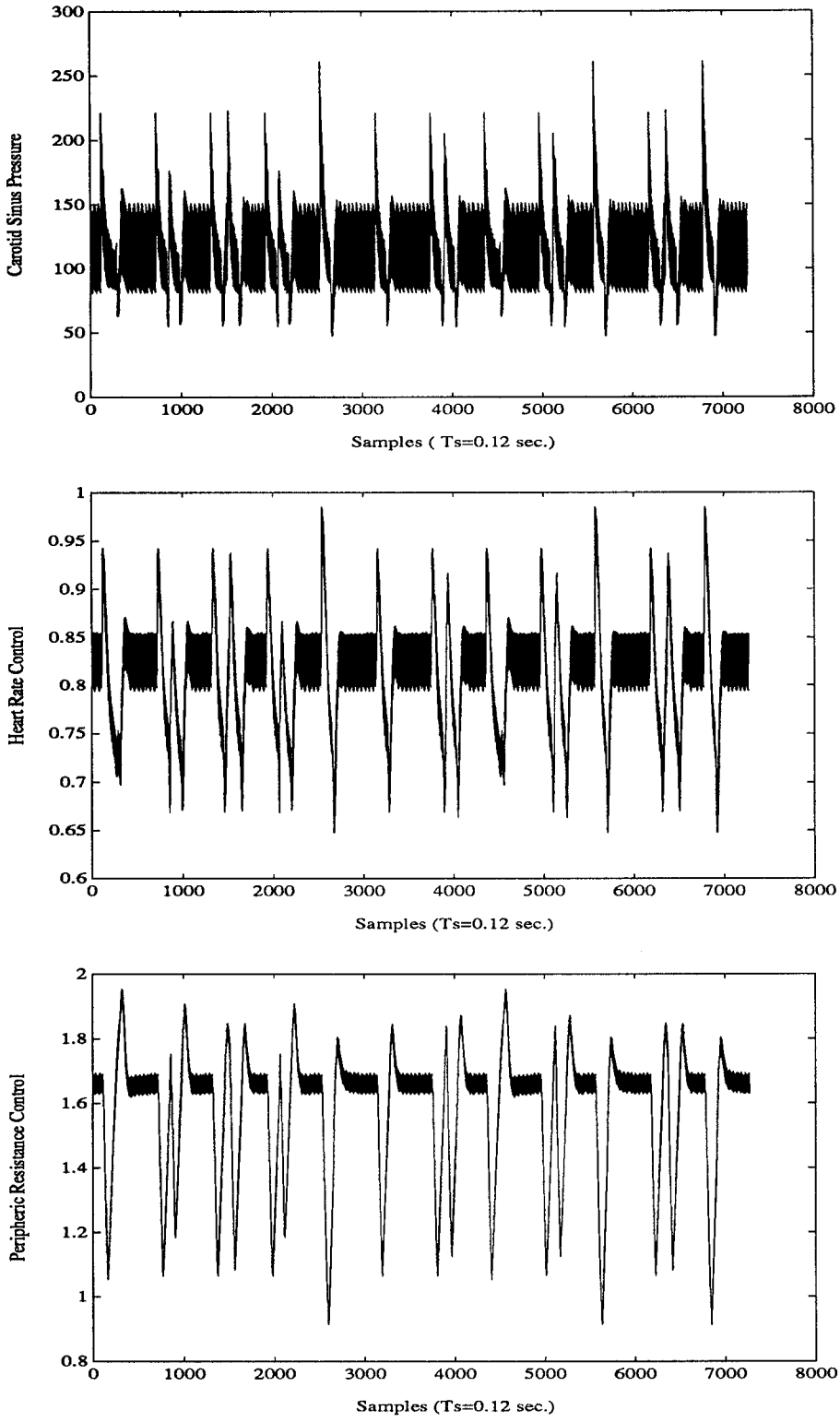


Figure 6.8: Carotid Sinus Pressure, Heart Rate and Peripheric Resistance Control Signals

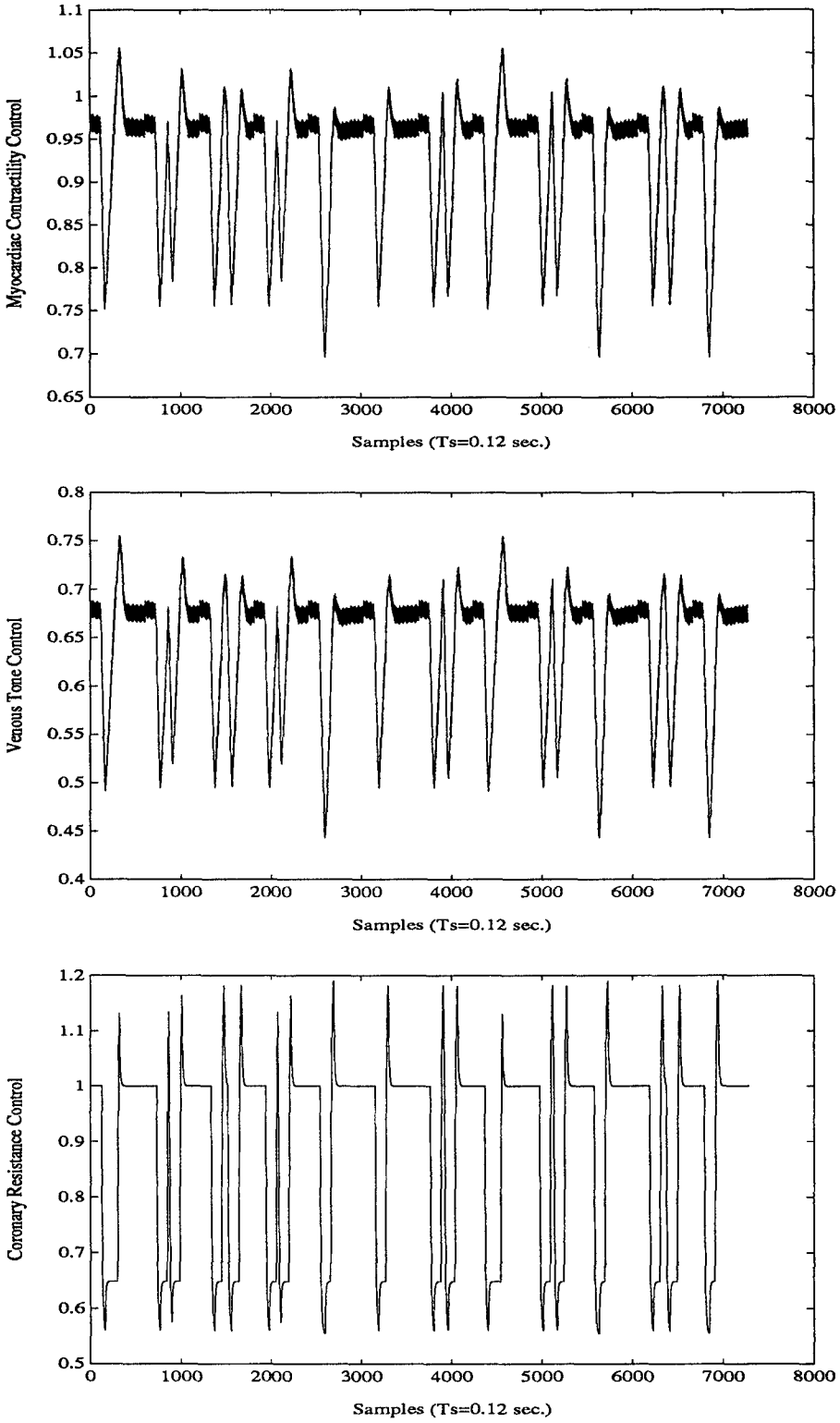


Figure 6.9: Myocardiac Contractility, Venous Tone, and Coronary Resistance Control Signals

was decided to reduce the amount of data used to identify the model. However if a portion of the initial data is thrown away, potentially useful patterns may be lost. Therefore, it was decided to keep the entire length of the trajectories, while decreasing the sampling frequency. A sampling rate of 0.24 seconds was used for identification. This could be achieved easily by eliminating every second data record from the original data set. This process is called *decimation*.

The plots of the reduced data set look exactly the same as the ones shown in Figures 6.8 and 6.9. Yet, in order to be absolutely sure, a *spectral power density* analysis was performed. The spectral plots obtained from the reduced data set look identical to the ones obtained from the original data set. Hence it can be concluded that no important system frequencies have been filtered out in the process of decimation. Therefore from now on, we will say that the data used for the identification of the FIR models are those plotted in Figures 6.8 and 6.9, although being precise, we did not use the entire set of 7279 points, but only half of them, throwing away every other point. Due to the fact that the data containing about 3600 points look exactly the same as the data plotted in those figures, it does not make sense to show here the same identical plots once more using a different abscissa.

6.3.3.1 Heart Rate Controller

The qualitative heart rate controller model has to be determined from the relationship between its input, the *carotid sinus pressure*, and its output, the *heart rate control signal*. The signals used for the heart rate controller identification process are plotted in the first two graphs of Figure 6.8.

The input and output variables were recoded into three qualitative levels each. Three classes are sufficient to obtain a good qualitative model for the system, and consequently, it was not necessary to work with more complex models.

For the heart rate controller, the optimal mask obtained was the following:

$$\begin{array}{c} {}_t\backslash^x \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{cc} CSP & HR \end{array} \\ \begin{pmatrix} 0 & 0 \\ -1 & -2 \\ -3 & +1 \end{pmatrix} \end{array} \tag{6.10}$$

This mask denotes the qualitative relationship:

$$HR(t) = \tilde{\mathrm{f}}(CSP(t-\delta t), HR(t-\delta t), CSP(t)) \tag{6.11}$$

Once the model has been identified, it must be validated. The data used in the identification process (Figure 6.8) constituted only a subset of the data available from the studied patient. Model validation is done by forecasting six data sets that were not used in the identification process, i.e., data that the model has never seen before. Each one of these six data sets, with a size of about 300 points, contains signals with specific morphologies, allowing the validation of the model for different system behaviors. Data set $^{\#}1$ represents two consecutive Valsalva maneuvers of 10 seconds duration separated by a two second break, data set $^{\#}2$ shows two consecutive Valsalva maneuvers of 10 seconds duration separated by a four second break, and data set $^{\#}3$ exhibits two consecutive Valsalva maneuvers of 10 seconds duration separated by an eight second break. Data set $^{\#}4$ shows a single Valsalva maneuver of 10 seconds duration with an intensity (pressure) increase of 50% relative to the previous three data sets. Data set $^{\#}5$ describes a single Valsalva maneuver of 20 seconds duration with nominal pressure. Finally, data set $^{\#}6$ characterizes a single Valsalva maneuver of 10 seconds duration with nominal pressure. Data set $^{\#}6$ is called the *reference data set*, since it represents a standardized Valsalva maneuver, from which all the other variants are derived by modifying a single parameter.

A comparison of the results obtained by simulation with the quantitative differential equation model and the qualitative FIR model is presented in Figures 6.10 and 6.11. The solid lines show the quantitative simulation results, whereas the dashed lines represent the qualitative simulation results.

The results obtained are quite good, and consequently, the identified model can be accepted. The normalized mean square errors (in percentage), MSE, of the heart rate controller model have been computed for each of the data sets individually, and also for all data sets together. These results are given in Table 6.1.

As can be seen from this table, the average error is quite low, and a comparison of this result with the one obtained using a NARMAX model of four terms for the same patient will confirm this assessment. This comparison is presented in Section 6.3.3.6. The FIR heart rate model requires a delay of one sample only in order to capture the dynamics of the system.

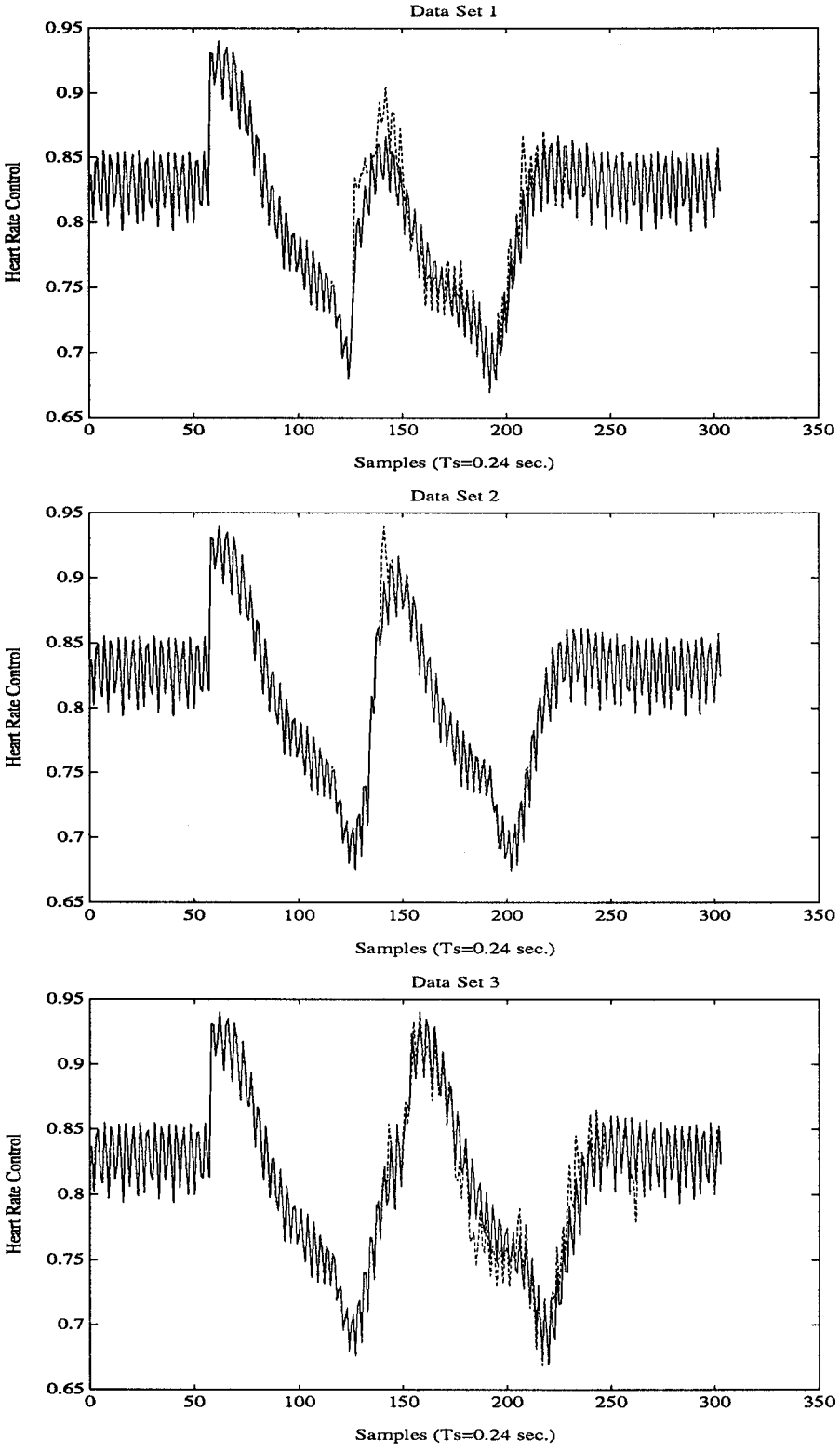


Figure 6.10: HR Controller: Forecast Results of the First Three Data Sets

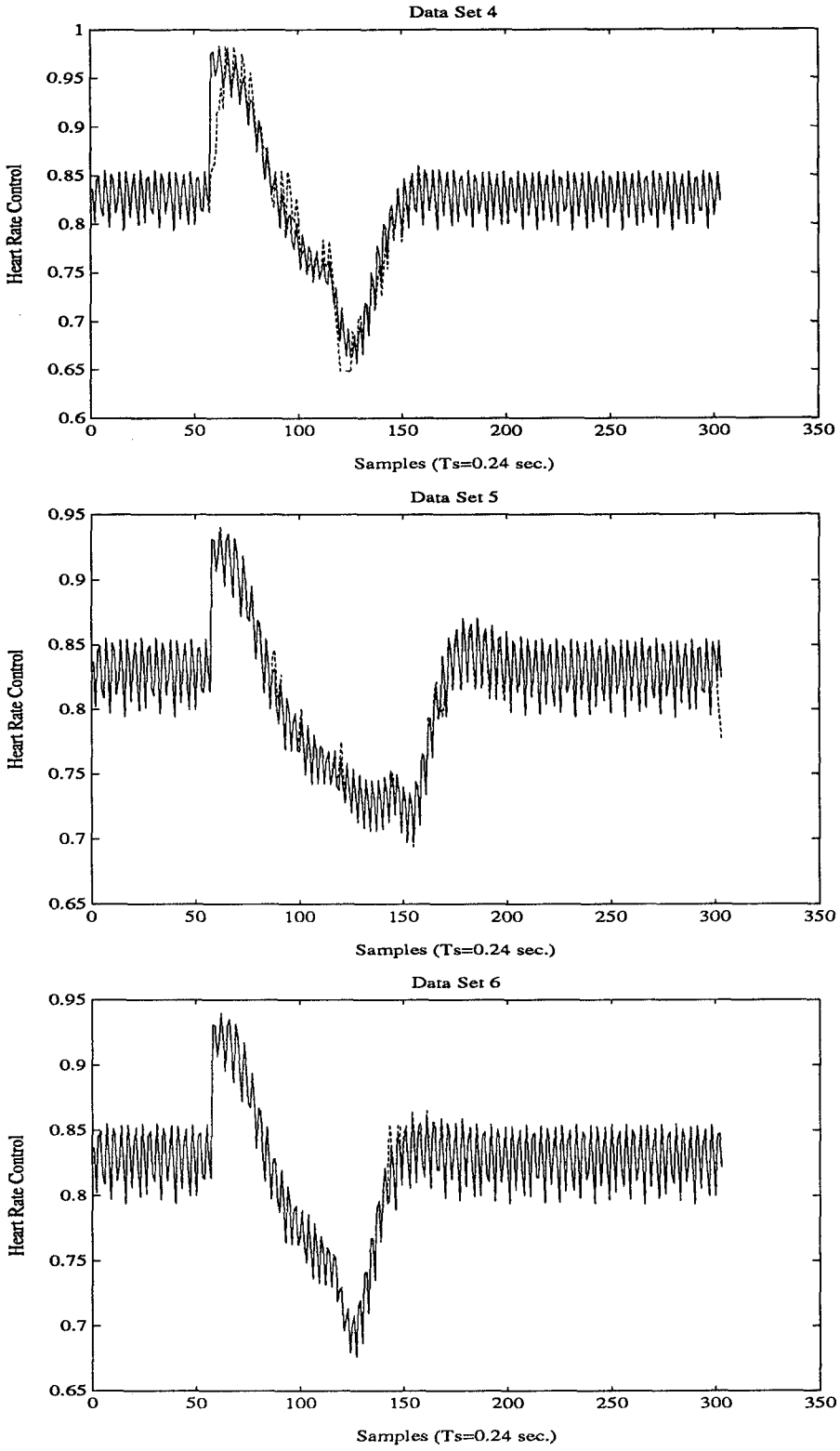


Figure 6.11: HR Controller: Forecast Results of the Last Three Data Sets

	MSE
Data Set 1	5.74 %
Data Set 2	0.97 %
Data Set 3	4.94 %
Data Set 4	8.64 %
Data Set 5	0.84 %
Data Set 6	0.82 %
Average Error	3.65 %

Table 6.1: MSE Errors of the Heart Rate Controller Model

6.3.3.2 Peripheric Resistance Controller

The qualitative peripheric resistance controller model is determined from the relationship between its input variable, the *carotid sinus pressure*, and its output variable, the *peripheric resistance control signal*. The signals used for the identification process are plotted in Figure 6.8.

For this model, the output variable has been recoded into four qualitative levels whereas the input variable has been recoded into three levels. In accordance with Equation 6.6, the number of classes for both variables could be made quite large, but four output classes and three input classes are sufficient to warrant a good model.

For the peripheric resistance controller, the optimal mask obtained was the following:

$$\begin{array}{l} {}_t\backslash^x \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{cc} CSP & PR \end{array} \\ \begin{pmatrix} 0 & -1 \\ -2 & -3 \\ 0 & +1 \end{pmatrix} \end{array} \tag{6.12}$$

This mask denotes the qualitative relationship:

$$PR(t) = \tilde{\mathrm{f}}(PR(t-2\delta t), CSP(t-\delta t), PR(t-\delta t)) \tag{6.13}$$

The same procedure was used to validate the model that had been employed to validate the heart rate controller model. A comparison of the quantitative

and qualitative simulation results is presented in Figures 6.12 and 6.13. As before, the solid line shows the quantitative simulation results, whereas the dashed line depicts the qualitative simulation results.

The results are for the most part as good as those obtained for the heart rate model. The normalized mean square errors for each data set individually as well as for all six data sets together have been computed. The results are shown in Table 6.2. It turns out that the results from data set #4 are considerably poorer than those from all other data sets. Evidently, this controller is quite sensitive to the pressure used in the Valsalva maneuver. More data varying the pressure would need to be used in the identification data set in order for this error to be reduced to a smaller value.

	MSE
Data Set 1	2.08 %
Data Set 2	0.03 %
Data Set 3	1.31 %
Data Set 4	10.51 %
Data Set 5	0.62 %
Data Set 6	0.12 %
Average Error	2.44 %

Table 6.2: MSE Errors of the Peripheric Resistance Controller Model

This model is the simplest FIR model that was found to represent the peripheric resistance controller at a good quality. The FIR peripheric resistance model only requires a delay of two samples in order to capture the dynamics of the system. A comparison between the results obtained with this FIR model and a NARMAX peripheric resistance controller model of two terms for the same patient is provided in Section 6.3.3.6.

6.3.3.3 Myocardiac Contractility Controller

The third controller to be determined is the myocardiac contractility controller. Its qualitative model is determined from the relationship between the *carotid sinus pressure*, its input, and the *myocardiac contractility control signal*, its output. The signals used for the identification process are plotted in Figures 6.8 and 6.9.

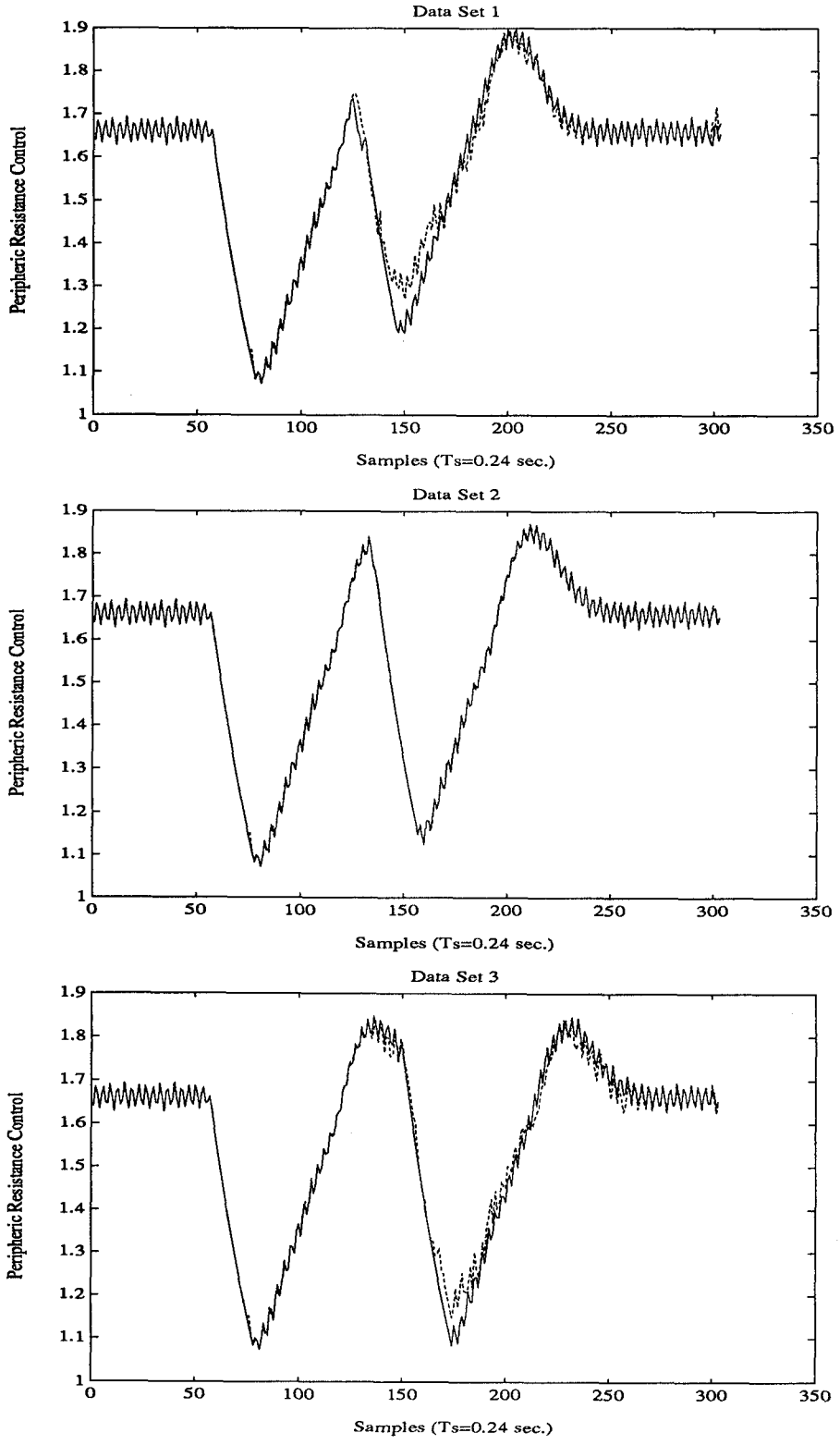


Figure 6.12: PR Controller: Forecast Results of the First Three Data Sets

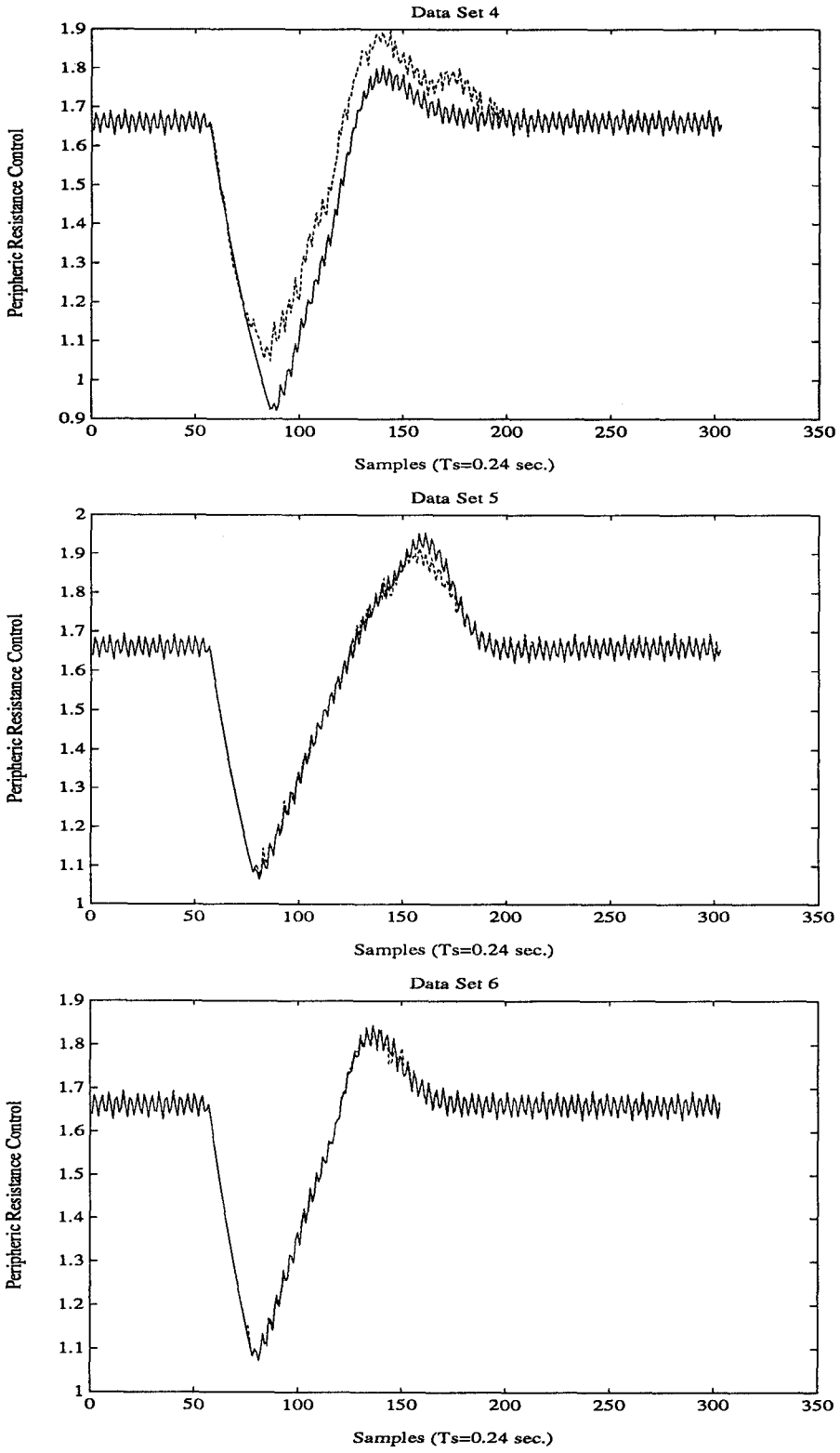


Figure 6.13: PR Controller: Forecast Results of the Last Three Data Sets

For this model, the input variable has been recoded into three qualitative levels whereas the output variable has been recoded into five levels.

For the myocardiac contractility controller, the optimal mask obtained was the following:

$$\begin{array}{c} {}_t\backslash^x \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{cc} CSP & MC \end{array} \\ \begin{pmatrix} -1 & 0 \\ 0 & -2 \\ -3 & +1 \end{pmatrix} \end{array} \tag{6.14}$$

This mask denotes the qualitative relationship:

$$MC(t) = \tilde{\mathrm{f}}(CSP(t-2\delta t), MC(t-\delta t), CSP(t)) \tag{6.15}$$

The model has been validated using the same approach as before. The comparative results for this controller are shown in Figures 6.14 and 6.15.

On the average, the results obtained are fairly good. However, data sets $^{\#}1$ and $^{\#}4$ are not truly optimal. An identification data set that characterizes the controller a little better could cure this problem easily. However, in order to follow as closely as possible the design process proposed in the dissertation by Vallverdú, we decided to stick with the same identification experiment that was used before.

The MSE errors for this controller are summarized in Table 6.3.

	MSE
Data Set 1	15.00 %
Data Set 2	0.10 %
Data Set 3	3.77 %
Data Set 4	8.31 %
Data Set 5	1.36 %
Data Set 6	0.61 %
Average Error	4.86 %

Table 6.3: MSE Errors of the Myocardiac Contractility Controller Model

As shown in Table 6.3, the average error obtained for the myocardiac contractility controller model is the largest obtained for all five controllers.

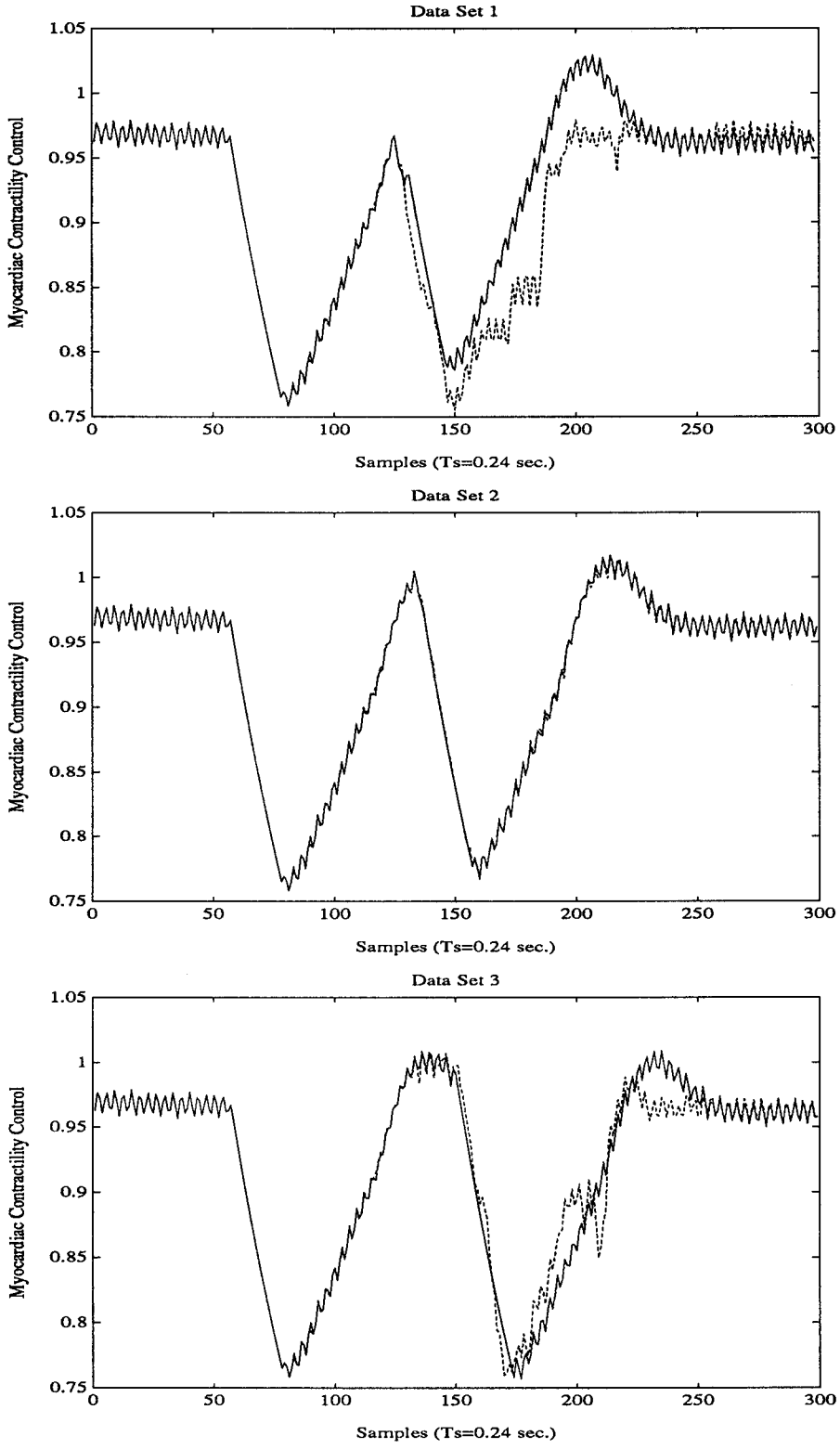


Figure 6.14: MC Controller: Forecast Results of the First Three Data Sets

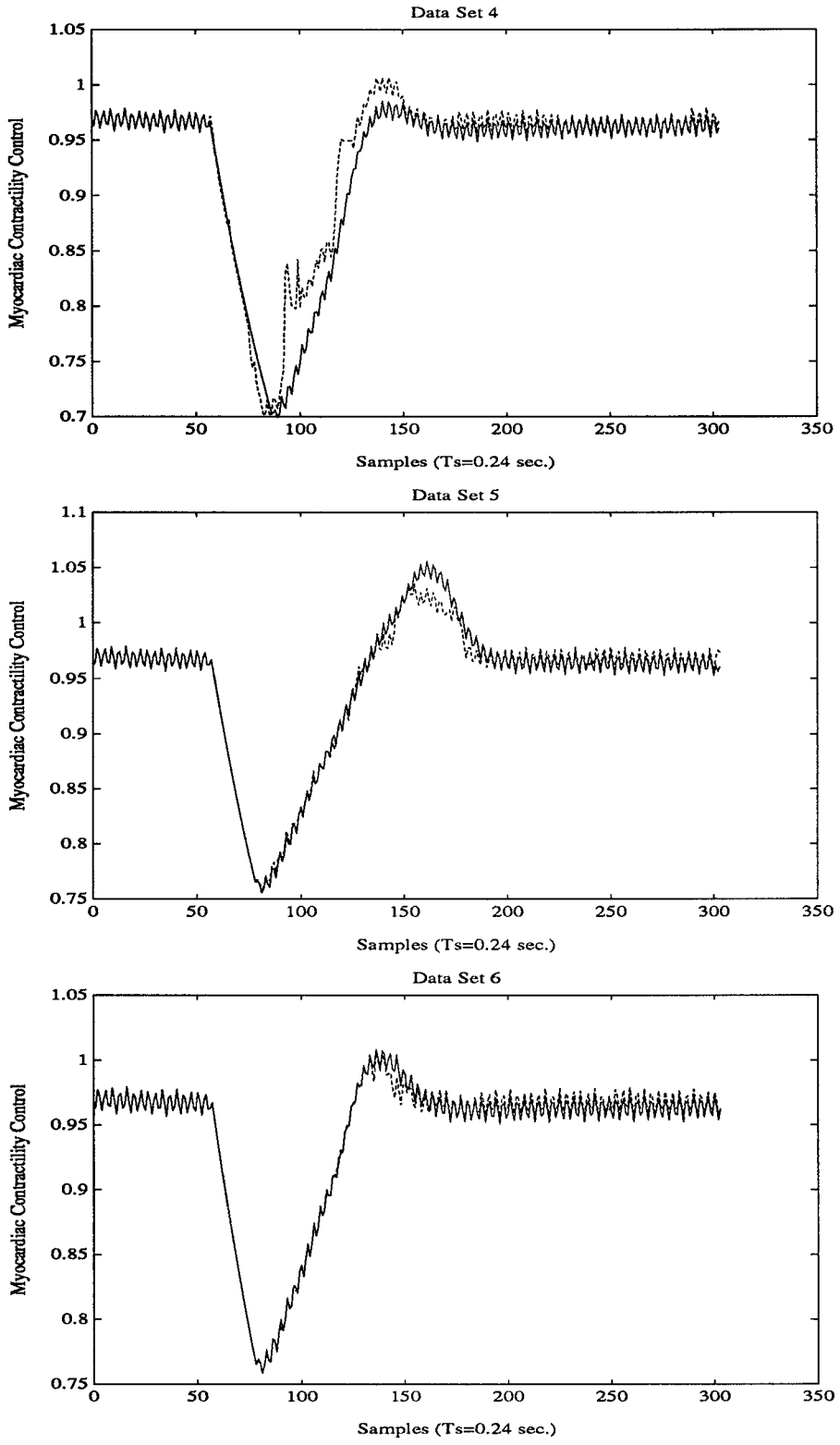


Figure 6.15: MC Controller: Forecast Results of the Last Three Data Sets

However, the error is still acceptably small, and consequently, the FIR model can be used as a valid representation of the myocardiac contractility controller. This will be confirmed in Section 6.3.3.6 where the results obtained by this FIR model are compared with the ones obtained with a NARMAX model of two terms for the same patient.

6.3.3.4 Venous Tone Controller

The fourth controller is the venous tone controller that uses the *carotid sinus pressure* as its input variable, and the *venous tone control signal* as its output variable. The signals used for the identification process are plotted in Figures 6.8 and 6.9.

Here, the input variable has been recoded into three qualitative levels whereas the output variable has been recoded into four classes.

For the venous tone controller, the optimal mask obtained was the following:

$$\begin{array}{c} {}_t\backslash^x \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{cc} CSP & VT \end{array} \\ \begin{pmatrix} 0 & -1 \\ -2 & -3 \\ 0 & +1 \end{pmatrix} \end{array} \tag{6.16}$$

This mask denotes the qualitative relationship:

$$VT(t) = \tilde{\mathrm{f}}(VT(t-2\delta t), CSP(t-\delta t), VT(t-\delta t)) \tag{6.17}$$

Once again using the same approach as before, the venous tone controller model has been validated. The comparative results are plotted in Figures 6.16 and 6.17.

The results obtained using the FIR venous tone controller model were generally quite good. The normalized mean square errors are tabulated in Table 6.4.

This model is the simplest FIR model that was found to represent the venous tone controller at a good quality level. The FIR venous tone controller requires a delay of two samples in order to capture the dynamics of the system. A comparison between the results obtained with this FIR model and a NARMAX venous tone controller model of two terms for the same patient is provided in Section 6.3.3.6.

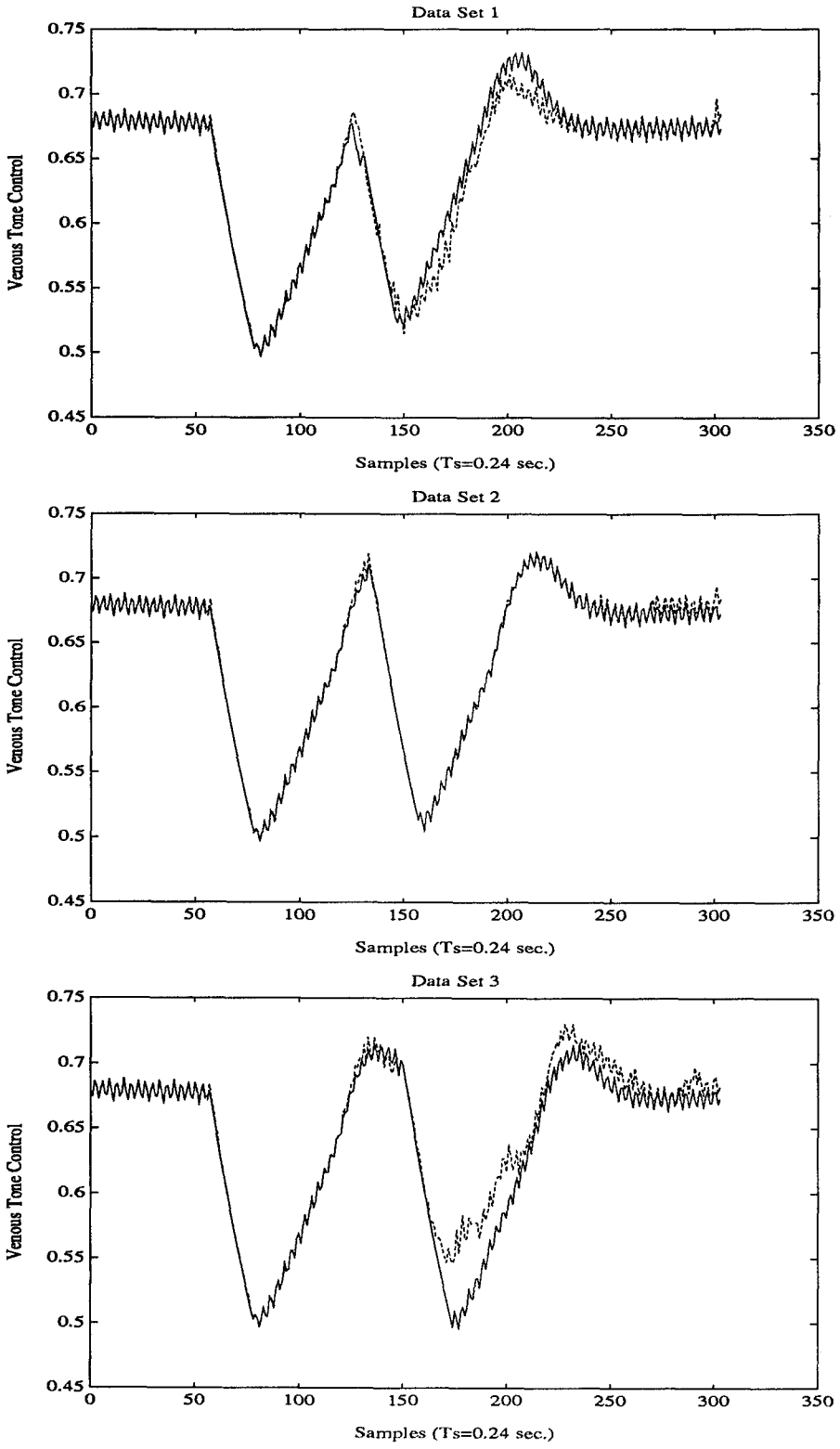


Figure 6.16: VT Controller: Forecast Results of the First Three Data Sets

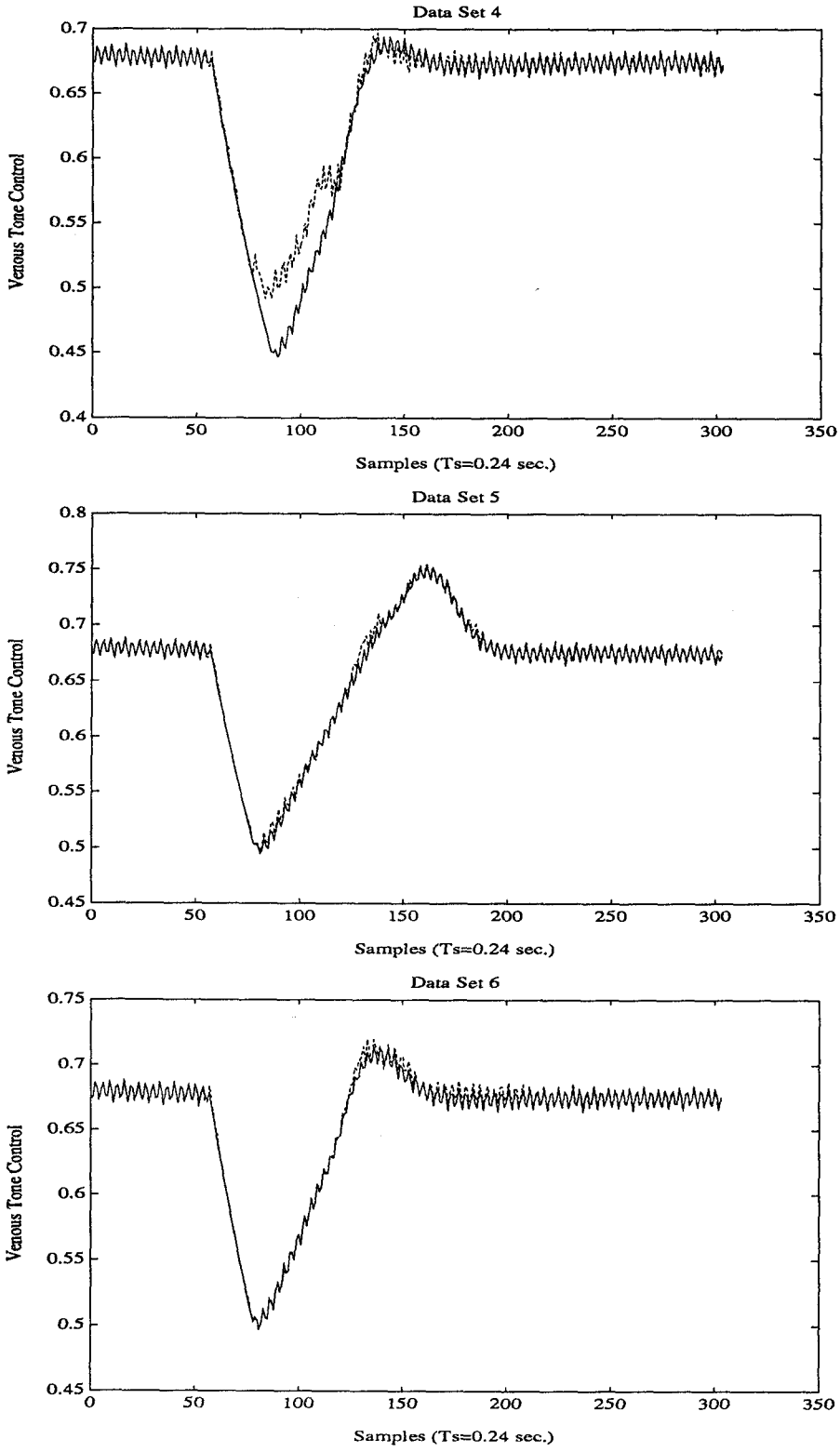


Figure 6.17: VT Controller: Forecast Results of the Last Three Data Sets

	MSE
Data Set 1	2.59 %
Data Set 2	0.28 %
Data Set 3	7.20 %
Data Set 4	6.18 %
Data Set 5	0.42 %
Data Set 6	0.42 %
Average Error	2.85 %

Table 6.4: MSE Errors of the Venous Tone Controller Model

6.3.3.5 Coronary Resistance Controller

The fifth and last qualitative CNS controller is the coronary resistance controller. It describes the qualitative relationship between the *carotid sinus pressure* and the *coronary resistance control signal*. The signals used for the identification process are plotted in Figures 6.8 and 6.9.

For this model, the input variable has been recoded into three qualitative levels whereas the output variable has been recoded into two classes only.

The optimal mask obtained for the coronary resistance controller was the following:

$$\begin{array}{c} {}_t\backslash^x \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{cc} CSP & CR \end{array} \\ \begin{pmatrix} -1 & 0 \\ -2 & -3 \\ -4 & +1 \end{pmatrix} \end{array} \tag{6.18}$$

This mask denotes the qualitative relationship:

$$CR(t) = \tilde{\mathrm{f}}(CSP(t-2\delta t), CSP(t-\delta t), CR(t-\delta t), CSP(t)) \tag{6.19}$$

The validation results for this model are shown in Figures 6.18 and 6.19.

The results obtained using this model are quite excellent. It turns out that the coronary resistance controller model obtained using the FIR methodology

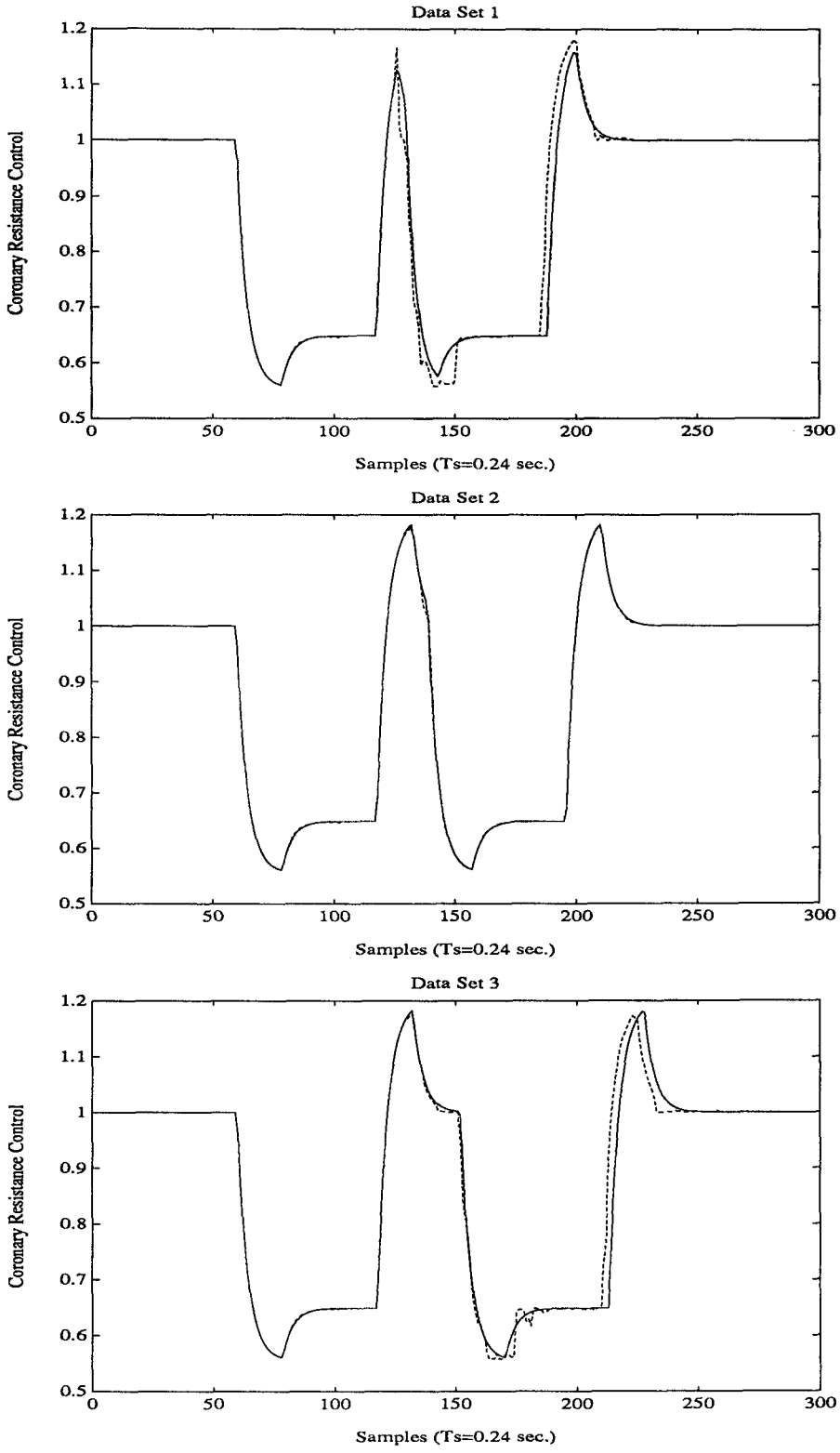


Figure 6.18: CR Controller: Forecast Results of the First Three Data Sets

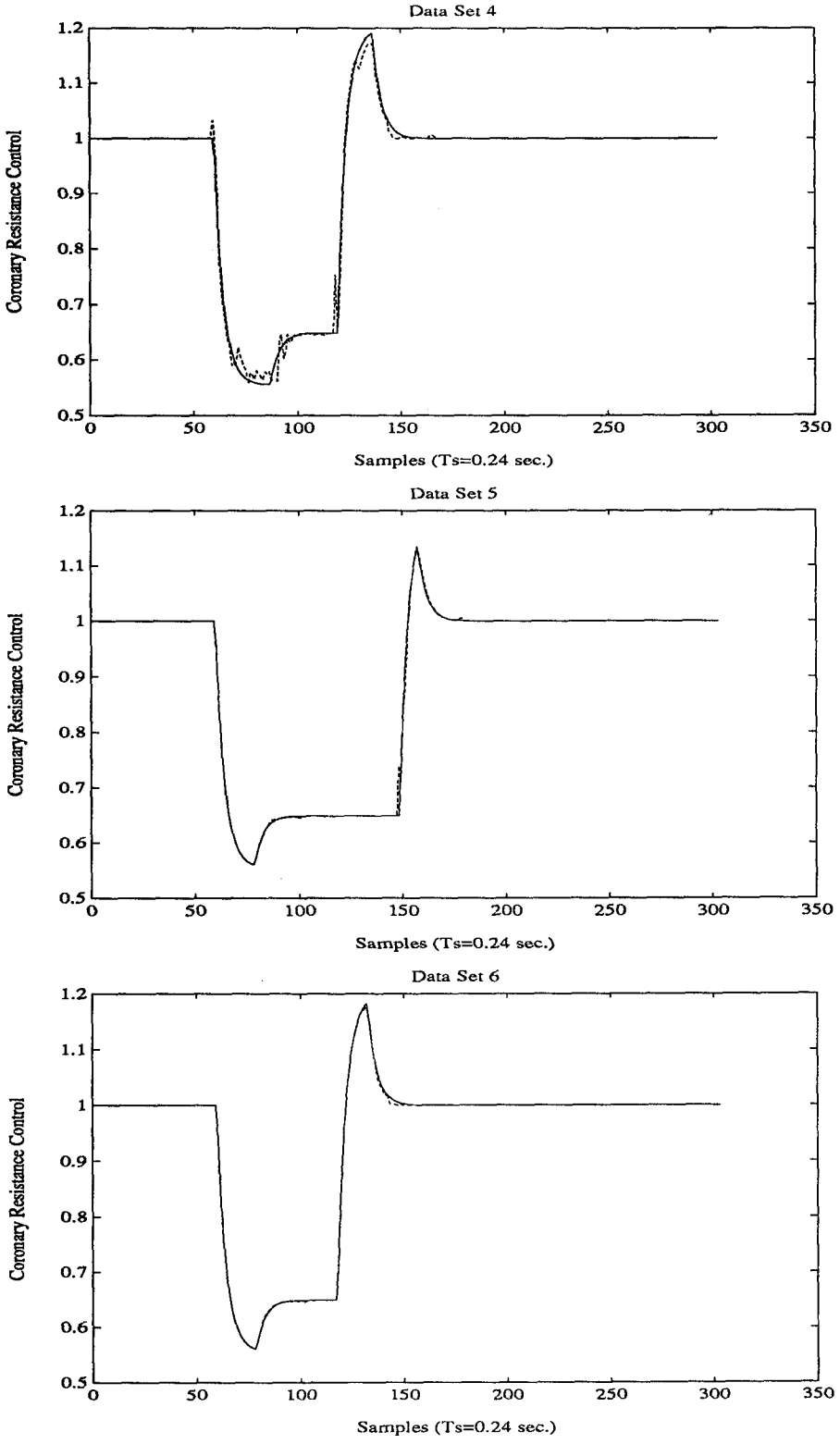


Figure 6.19: CR Controller: Forecast Results of the Last Three Data Sets

	MSE
Data Set 1	3.07 %
Data Set 2	0.02 %
Data Set 3	2.62 %
Data Set 4	0.56 %
Data Set 5	0.15 %
Data Set 6	0.11 %
Average Error	1.09 %

Table 6.5: MSE Errors of the Coronary Resistance Controller Model

exhibits the smallest MSE error of all FIR controller models. The normalized mean square errors are tabulated in Table 6.5.

This model, as the previous ones, is the simplest FIR model that was found to represent the coronary resistance controller with a high degree of quality. A comparison between the results obtained with this FIR model and a NARMAX coronary resistance controller model of five terms for the same patient is provided in Section 6.3.3.6.

6.3.3.6 Comparisons of NARMAX vs. FIR Controller Models

In this section, a comparison of the five controller models obtained for a given patient using two different methodologies is presented. These methodologies are the NARMAX quantitative inductive modeling technique *vs.* the FIR qualitative inductive modeling technique.

There is something that must be clarified before the comparison. Both methodologies have used the same data sets presented in Figures 6.8 and 6.9 for model identification. In the validation process, the same six data subsets described in Section 6.3.3.1 were used by both the NARMAX and FIR methodologies. However, while in the FIR validation process, the six subsets of data were predicted separately, in the NARMAX validation process these data subsets have been concatenated one behind the other, and the prediction of the overall set was done at once. It is necessary to bear this in mind since this difference may influence the obtained errors somewhat.

Heart Rate Controller Model

The NARMAX model of four terms found in Vallverdú's thesis that represents best the heart rate controller relative to the signals of the patient shown in Figure 6.8 is described in Equation 6.20.

$$\begin{aligned} y(t) = 0.0346 + 7.9151 \cdot 10^{-4} * x(t) \\ +0.7612 * y(t-1) + 0.1133 * y(t-7) \\ -1.5930 \cdot 10^{-6} * x(t) * x(t-3) \end{aligned} \tag{6.20}$$

where $y(t)$ is the output, i.e., the heart rate control signal, and $x(t)$ is the input, i.e., the carotid sinus pressure. This model is composed of a constant term, three linear terms, and one bi–linear term.

In order to validate this model of five terms, the six data sets given for validation purposes were simulated, yielding an overall MSE error of **9.3%**, which is larger than the MSE error of **3.65%** obtained with the FIR model. It can thus be concluded that the FIR model of the heart rate controller describes the system somewhat more accurately than a NARMAX model of five terms.

However, the reader should bear in mind that the NARMAX model is a *much simpler* model than the FIR model. Including more terms in the NARMAX model would certainly reduce the error of the NARMAX model further at the cost of having to carry along more parameters. The best model is not necessarily the one that carries the smallest error, but the simplest (easiest to identify) model that still explains the available facts. In favor of the FIR model, it should be mentioned that, although the FIR model is much more complex than the NARMAX model, it is as easily identifiable as the NARMAX model, and the identification process consumes very little time.

Peripheric Resistance Control Model

The NARMAX model of two terms found in Vallverdú's thesis that represents best the peripheric resistance controller relative to the signals of the patient shown in Figure 6.8 is described in Equation 6.21.

$$y(t) = 0.0489 + 0.9851 * y(t-1) \\ -2.0074 \cdot 10^{-6} * x(t-1) * x(t-7) \tag{6.21}$$

where $y(t)$ is the output, i.e., the peripheric resistance control signal, and $x(t)$ denotes the input, i.e., the carotid sinus pressure. The model is composed of a constant term, a linear term, and a bi–linear term.

To validate this model of three terms, the same six data sets were used, yielding an MSE error of **18.5%**, clearly much larger than the **2.44%** MSE error obtained with the FIR model. Hence it can be concluded that the FIR model of the peripheric resistance controller describes the system considerably more accurately than a NARMAX model of three terms.

Myocardiac Contractility Control Model

The NARMAX model of two terms found in Vallverdú's thesis that represents best the myocardiac contractility controller relative to the signals of the patient shown in Figures 6.8 and 6.9 is described in Equation 6.22.

$$y(t) = 0.0177 + 0.9897 * y(t-1) \\ -6.5093 \cdot 10^{-7} * x(t-1) * x(t-7) \tag{6.22}$$

where $y(t)$ is the myocardiac contractility control signal, i.e., the output, and $x(t)$ is the carotid sinus pressure, i.e., the input. This model, just like the one used for the peripheric resistance controller, is composed of one constant term, one linear term, and one bi–linear term.

The same six data sets were used to validate the NARMAX model of three terms. This experiment generated an MSE error of **22%**, clearly much larger than the **4.86%** MSE error obtained with the FIR model.

Venous Tone Control Model

The NARMAX model of two terms found in Vallverdú's thesis that represents best the venous tone controller relative to the signals of the patient shown in Figures 6.8 and 6.9 is described in Equation 6.23.

$$y(t) = 0.01374 + 0.9897 * y(t-1) \\ -5.6402 \cdot 10^{-7} * x(t-1) * x(t-7) \quad (6.23)$$

where $y(t)$ is the venous tone control signal, and $x(t)$ is the carotid sinus pressure. Also this model is composed of a constant term, a linear term, and a bi–linear term.

Use of this model of three terms with the same six data sets leads to an MSE error of **22%**, much larger than the average **2.85%** MSE error obtained with the FIR methodology.

Coronary Resistance Control Model

The NARMAX model of six terms found in Vallverdú's thesis that represents best the coronary resistance controller relative to the signals of the patient shown in Figures 6.8 and 6.9 is described in Equation 6.24.

$$y(t) = -0.0245 - 7.6672 \cdot 10^{-5} * x(t-10) \\ +0.9914 * y(t-1) - 9.0794 \cdot 10^{-7} * x(t-6) * x(t-8) \\ +3.9906 \cdot 10^{-7} * x(t-10) * x(t-10) \\ -9.4522 \cdot 10^{-5} * x(t-10) * y(t-1) \quad (6.24)$$

where $y(t)$ is the coronary resistance control signal, and $x(t)$ is the carotid sinus pressure. The model is composed of a constant term, two linear terms, and three bi–linear terms.

The average MSE error for this model is **25.5%**, substantially larger than the average MSE error of the FIR model of **1.09%**. It is interesting to notice that the NARMAX methodology had most difficulties with this controller model, whereas the FIR methodology had the least problems with it.

After these comparisons, it can be concluded that the FIR methodology is indeed able to find five controller models that represent, for a given patient, the controllers with surprising accuracy, *much* better so than NARMAX models.

6.3.4 Cardiovascular Closed–Loop System

The aim of this section is to close the loop between the hemodynamical system, modeled by means of differential equations, and the central nervous control system, modeled in terms of the fuzzy inductive reasoning methodology.

It has been demonstrated in previous sections of this chapter that the FIR methodology is indeed able to model every one of the five CNS controllers individually. Now it is time to study the behavior of the cardiovascular system as a whole.

Real physiological data obtained from cardiac catheterization are used for this study. These data stem from the hemodynamical division of the *Hospital de la Santa Creu i de Sant Pau* in Barcelona and have been supplied by Vallverdú. The data stem from a patient with coronary arterial obstruction of at least 70%. The physiological variables of interest are: the *right auricular pressure*, $P_{AD}(t)$, the *aortic pressure*, $P_A(t)$, the *coronary blood flow*, $F_C(t)$, and the *heart rate*, $HR(t)$. The physiological variables were recorded during the following phases of the Valsalva maneuver (Figure 6.7):

- Pre–Valsalva phase.
- Valsalva Phase I, immediately after the onset of the maneuver.
- Valsalva Phase II, just before the end of the maneuver.
- Valsalva Phase III, immediately after the maneuver has ended.
- Valsalva Phase IV, after the maneuver has come to an end.

From the trajectories of the right auricular pressure, the aortic pressure, the coronary blood flow, and the heart rate, mean values are computed for each of the five phases of the maneuver. P_{ADM} denotes the average right auricular pressure during a phase, P_{AM} stands for the mean aortic pressure, F_{CM} is the average coronary blood flow, and HR_M signifies the average heart rate during any one of the phases.

The measurement results obtained through cardiac catheterization for the studied patient are summarized in Table 6.6. Only the mean values computed for the pre–Valsalva phase, the Valsalva phase II, and the Valsalva phase IV are shown in the table, because these are the most significant phases.

P_{ADM}	Pre–V	4
	II	38
	IV	5
P_{AM}	Pre–V	107
	II	99
	IV	119
F_{CM}	Pre–V	123
	II	106
	IV	118
HR_M	Pre–V	77
	II	82
	IV	70

Table 6.6: Measurement Results Obtained from Catheterization (From Vallverdú's Thesis)

The mean values presented in Table 6.6 are obtained from real measurements. They will consequently be used as reference values in the model validation process. In order for a model to pass the acceptance test, none of the four key variables, i.e., the *average right auricular pressure*, the *mean aortic pressure*, the *mean coronary blood flow*, and the *average heart rate* must deviate from the reference values by more than $\pm 10\%$ during any of the three key phases of the Valsalva maneuver, i.e., the pre–Valsalva phase, the Valsalva phase II, and the Valsalva phase IV.

In Vallverdú's dissertation (Vallverdú, 1993), two different cardiovascular system models were studied: a model described solely by means of differential equations, and another model whose hemodynamical subsystem was modeled using differential equations and whose central nervous system control was modeled embracing the NARMAX methodology.

The simulation results obtained from either of the two cardiovascular system models was found to lie inside the $\pm 10\%$ error margin permitted, and therefore, both models were considered to be valid for the task at hand. The results obtained with the purely deductive differential equation model are presented in Table 6.7.

The largest negative relative deviation from the measurement values is -5.19%, and the largest positive relative deviation is $+5.93\%$. Thus, all the indicators are clearly within the requested $\pm 10\%$ margin. The average relative

P_{ADM}	Pre–V	4
	II	38
	IV	5
P_{AM}	Pre–V	111
	II	100
	IV	118
F_{CM}	Pre–V	119
	II	105
	IV	125
HR_M	Pre–V	73
	II	79
	IV	74

Table 6.7: Results Obtained from Purely Deductive Differential Equation Model (From Vallverdú's Thesis)

deviation from the measurement values is 2.52%.

The results obtained from the mixed differential equation and NARMAX model are summarized in Table 6.8. It is necessary to point out that these results have been obtained using models of each of the five controllers with a structure common to all patients rather than the individual models shown in Section 6.3.3.6. However, the parameters of the NARMAX models were post–optimized for each patient individually in closed loop, comparing the results obtained from simulating the mixed model with the actual measurement data.

Here, the largest negative relative deviation from the measurement values is −4.06%, and the largest positive relative deviation is +6.09%. Thus, all the indicators are again within the requested ±10% margin. The average relative deviation from the measurement values is 1.48%. Since the average deviation is a little smaller than in the case of the differential equation model, the mixed differential equation and NARMAX model can be considered to be of higher quality than the pure differential equation model. These excellent results were possible as a consequence of the post–optimization.

At this point, the question to be raised is whether a mixed model of the cardiovascular system, whereby the hemodynamical subsystem is described by means of differential equations and the CNS control is described using a FIR model also generates results inside the ±10% error margin permitted and can therefore also be considered a valid model for the task at hand.

P_{ADM}	Pre–V	4
	II	38
	IV	5
P_{AM}	Pre–V	108
	II	99
	IV	117
F_{CM}	Pre–V	128
	II	102
	IV	118
HR_M	Pre–V	76
	II	77
	IV	70

Table 6.8: Results Obtained from mixed differential equation and NARMAX Model (From Vallverdú's Thesis)

To answer this question, the mixed quantitative and qualitative modeling and simulation methodology described in detail in Chapter 4 of this dissertation have been applied.

A graph of the entire cardiovascular system showing the interactions between the hemodynamical system and the central nervous system control is presented in Figure 6.20.

In this figure, the solid boxes represent physiological components of the cardiovascular system, such as the heart and the arteries and the veins, and the solid arrows represent the blood flow passing between these body components. These are controlled by the CNS. The conceptual controllers within the CNS control are drawn as dashed boxes, and the information flow from the body to these controllers and from the controllers back to the body are indicated through dashed arrows. The heart, the thorax, and the abdomen are the principal areas where the control takes place. The baroreceptors are the sensorial organs that record the pressure changes, and thereby, are able to determine the carotid sinus pressure, the input variable to all of the CNS controllers. Figure 6.20 clarifies the overall structure of the cardiovascular system model and its implementation.

The differential equation model of the hemodynamical system is implemented using the Advanced Continuous Simulation Language (ACSL), whereas the qualitative central nervous system control is realized using SAPS.

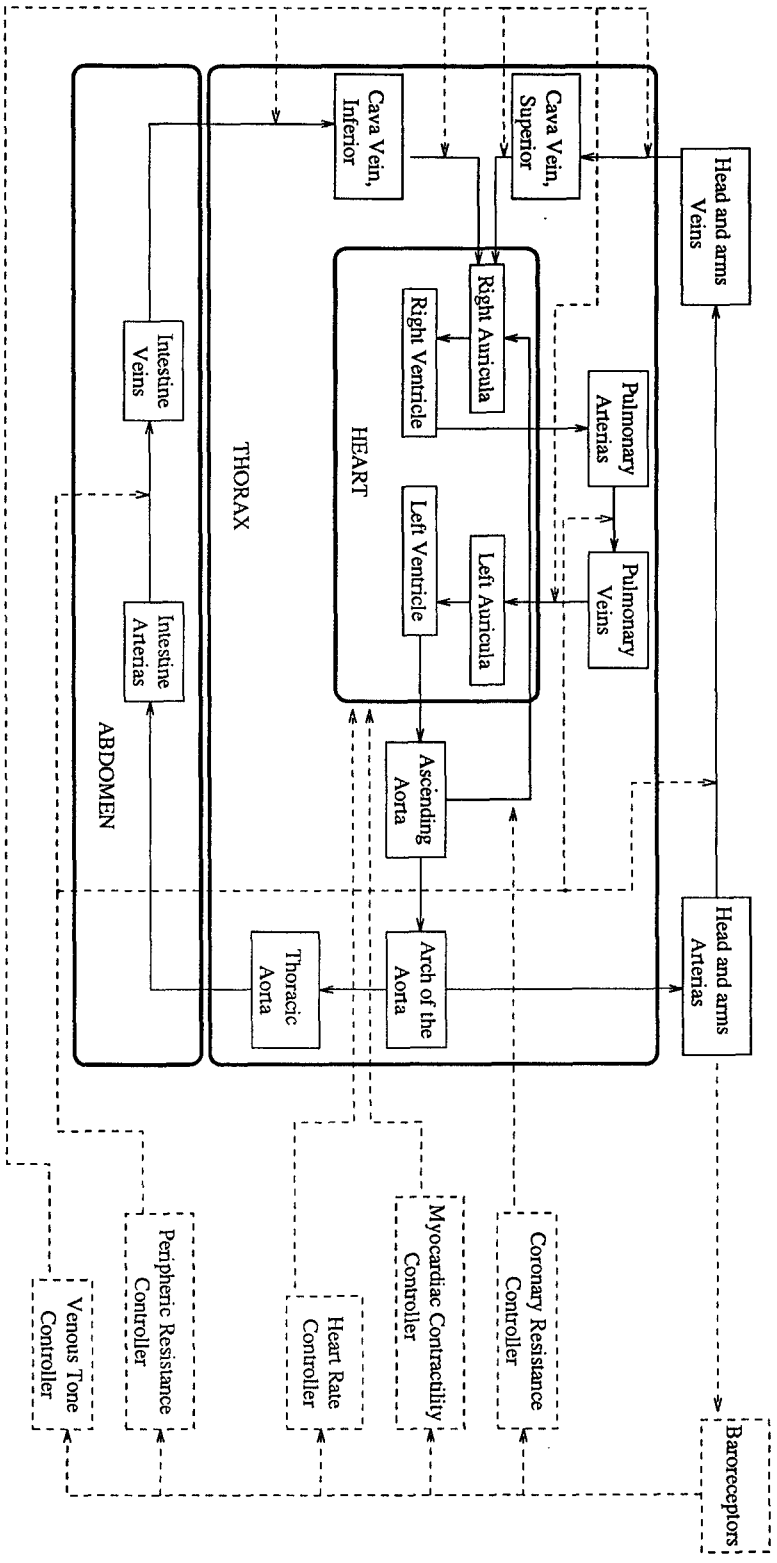


Figure 6.20: Schematic of the Cardiovascular System with Interconnections (Adapted from (Vallverdú, 1993))

Therefore, an interface between ACSL and SAPS is needed. A simplified scheme of the simulation structure is shown in Figure 6.21.

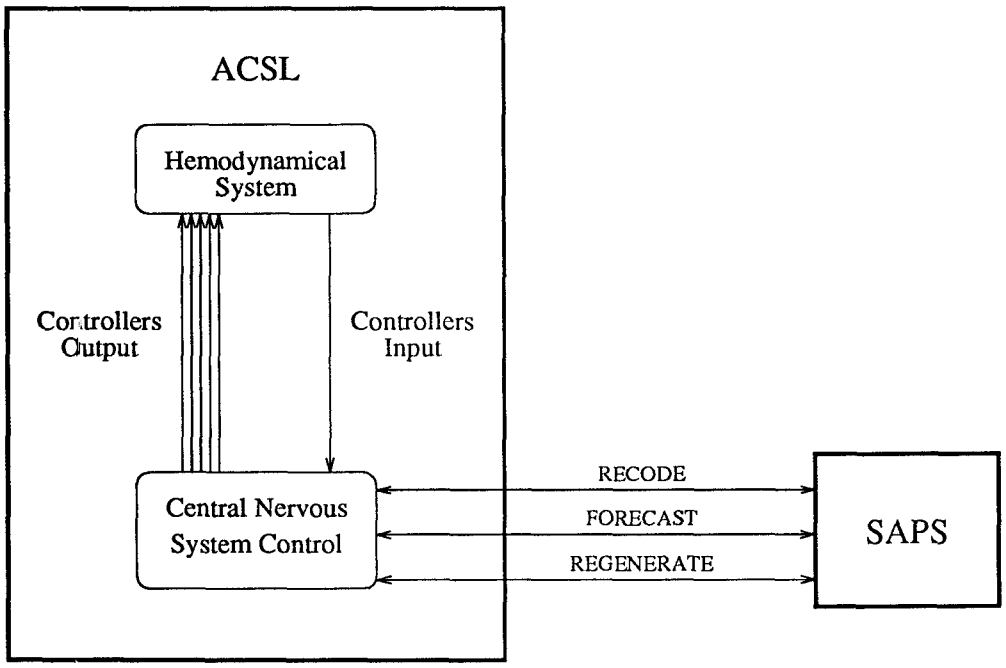


Figure 6.21: Simulation Structure of the Cardiovascular System

The hemodynamical system, modeled and simulated in a strictly quantitative fashion, is implemented in full within ACSL. Its differential equations are implemented as a continuous process within ACSL to be integrated across time using one of the standard integration algorithms offered by ACSL. The CNS control, on the other hand, is implemented inside the ACSL program as a discrete process to be executed once every 0.24 seconds.

The simulation process operates in the following way. The hemodynamical system generates a continuous-time trajectory representing the *carotid sinus pressure.* This variable is sampled by the discrete process once every 0.24 seconds and is immediately being recoded into three qualitative classes using the same landmarks and the same fuzzy membership functions that had been used for the training data set, and using the SAPS fuzzy recoding function that is coupled to ACSL through an interface routine. The discrete process then calls five times upon the SAPS fuzzy forecasting routine to predict qualitative values of the five controller outputs. These five qualitative triples are then regenerated into quantitative (real-valued) controller outputs using

the SAPS fuzzy regeneration function. The regenerated signals are finally made available to the hemodynamical system for use within the differential equation model. The overall effect of the qualitative CNS control model is that of a single–input multi–output (SIMO) digital controller with sample–and–hold (ZOH) circuitry at each of the five controller outputs. The qualitative processes, *recode*, *forecast*, and *regenerate*, are run inside SAPS–II, reducing the ACSL implementation of the CNS control to a mere interface.

The simulation results obtained from the mixed quantitative and qualitative cardiovascular system model using fuzzy inductive reasoning for the description of the CNS control are summarized in Table 6.9.

P_{ADM}	Pre–V	4
	II	38
	IV	5
P_{AM}	Pre–V	110
	II	100
	IV	117
F_{CM}	Pre–V	118
	II	105
	IV	125
HR_M	Pre–V	71
	II	78
	IV	73

Table 6.9: Results Obtained from mixed differential equation and FIR Model

As can be seen from this table, the largest negative relative deviation from the measurement values is -5.093%, and the largest positive relative deviation is $+7.79\%$. Thus, all the indicators are again within the requested $\pm 10\%$ margin, and in accordance with the requirements, also this model is to be accepted as a valid representation of reality for the task at hand. The average relative deviation from the measurement values is 2.78%. Consequently, the average deviation from the measurement data is here a little larger than in the other two models.

Notice that it is not fair to compare the mixed differential equation model and FIR model with the mixed differential equation model and NARMAX model, due to the fact that, in the latter case, the CNS control models are common, in structure, to all patients and its parameters were post–optimized

for the single patient studied, whereas the CNS control models of the mixed differential equations and FIR models were obtained for the single patient studied, without any kind of generalization.

It should also be remembered how the FIR model was created, namely as a replica of the purely quantitative differential equation model and not as a replica of the measurement data. Comparing the FIR model with the differential equation model, we find that the largest negative relative deviation between the two models is 0.0%, and the largest positive relative deviation is +2.73%. The average relative deviation between the two models is 0.66%.

This is an impressive similarity. The FIR model replicates exceptionally well what it was told to be the "reality," i.e., the differential equation model.

6.3.5 General Comments About NARMAX and FIR Methodologies

NARMAX models are *parametric models*. Once their structure has been determined, the five NARMAX models form a set of algebraic equations containing a bunch of parameters. Thus, NARMAX models can be easily optimized by any off–the–shelf curve–fitting algorithm. Training a NARMAX model consists of two separate steps: (i) determining the optimal equation structure, and (ii) optimizing the parameters of the selected structure. Most of the computational effort is spent on the optimization. Thus, designing a NARMAX model is predominantly an optimization problem.

FIR models are *non–parametric models*. Training a FIR model also consists of two steps: (i) determining the qualitative equation structure, i.e., the optimal mask, and (ii) composing a historical data base for holding the previous experience, i.e., the previously observed input/output patterns. Designing a FIR model is a synthesis procedure, not an optimization problem. Hence, although it is fairly simple and fast to set up a FIR model, the methodology does not offer an easy means for post–optimizing it.

The NARMAX approach has the advantage of being naturally adaptive, i.e., it lends itself to post–optimization. This does not hold true for the FIR model. However, since setting up a new FIR model is usually a simple and fast process, post–optimization is not truly needed. When a FIR model needs to be modified, it is acceptable to simply identify a new model, since this procedure doesn't require much time. Also, some adaptation would be possible in the FIR approach as well, not by optimizing parameters, but by updating the

experience data base on the fly.

The NARMAX model is much simpler to implement and does not require an experience data base as is the case for the FIR model. Thus, the NARMAX model needs much less memory, and also the simulation is somewhat faster than using the FIR model. However, the additional implementational effort of the FIR methodology goes hand in hand with a much increased flexibility and capability of replicating arbitrarily nonlinear system behavior.

6.3.6 Conclusions

In this chapter, the fuzzy inductive reasoning methodology has been applied to two quite different applications from two very different biomedical areas: *anesthesiology* and *cardiology*. In the first example, the system to be modeled was the decision making process of the anaesthetist. Hence we were fighting problems of limited causality between input variables and output variables. The system could just as easily have been classified as a *psychological system* instead of being called a *biomedical system*. In the second example, the system to be modeled was truly a biomedical system. Here, we were modeling a portion of the human central nervous system control, namely the portion that is responsible for the functioning of the heart, and, more generally, for the blood transport through the body. In this example, we were fighting problems of dealing with model variables for which no measurement data were available, variables that possibly do not even represent physiological quantities.

The main objective of this chapter was to address some of the problems that arise when working with this kind of systems, namely the problems of poor data availability and/or poor data quality, and diversity of time constants. Rather than tackling these problems in a theoretical manner (which may prove to be impossibly difficult due to the complexity of the issues involved), the problems were dealt with by means of two practical examples.

It is also the purpose of this chapter to offer comparisons between the FIR methodology and other methodologies, such as the ANN and NARMAX methodologies. This is valuable for validating the FIR methodology *per se*, i.e., for getting a feeling for what it can accomplish, what is clearly outside its capabilities, and how it fares in comparison with other methodologies that promise to solve similar problems. Since the FIR methodology is still fairly new and has not been used widely by the research community at large, such comparisons should be of great interest.

To this end, the control of an anaesthetic agent during surgical operation has been presented. A previous publication that made use of an ANN for the same purpose has been taken as a starting point, in order to be able to compare the FIR results with those obtainable by an already established modeling technique. The same identification data were used for finding the ANN and FIR models. In that manner, a fair comparison between the two methodologies could be achieved.

As Figure 6.5 shows, the FIR model captures the system behavior *much* better than the ANN model. Moreover, it could be concluded from the results that the FIR methodology is fairly robust, because it consistently generates a decent inductive model whenever the data allow it to, and it categorically declines to generate a model if the available data do not permit to validate an inductive model.

The second part of the chapter has presented an effort to model and simulate a part of the central nervous system control of the human body, namely the part responsible for the blood delivery system throughout the body. Five controller models, for a unique patient, for different control actions have been identified separately using the FIR methodology. In a previous research effort due to Vallverdú, different NARMAX models for the same five CNS controllers where obtained for the same patient. The first NARMAX controller had five terms, the next three controllers contained three terms each, and the last NARMAX controller model was characterized by six terms.

The average MSE errors obtained using the two methodologies are summarized in Table 6.10 for each of the five controllers.

Controller	**NARMAX**[2]	**FIR**
Heart Rate	9.30 %	3.65 %
Peripheric Res.	18.5 %	2.44 %
Myocardiac Cont.	22.0 %	4.86 %
Venous Tone	22.0 %	2.85 %
Coronary Res.	25.5 %	1.09 %

Table 6.10: MSE Errors of NARMAX and FIR Controller Models

The FIR methodology proved much more capable than the NARMAX

[2]Notice that the number of terms of each NARMAX model differs as explained in the text.

methodology to accurately reproduce the input/output behavior of an arbitrarily nonlinear unknown dynamic system.

To summarize, it has been demonstrated that the qualitative non–parametric FIR model synthesis technique is a powerful tool for the identification of inductive models of the CNS controllers. It compares favorably with the quantitative parametric NARMAX model optimization technique, when used for such purpose.

The FIR CNS control models had only been developed for a single patient. However, in later chapters of this dissertation, a technique will be presented that could allow us to find, for each of the five CNS controllers, a single FIR model, a model that is independent of the individual patient characteristics. However, this has not actually been tried yet for the CNS control system.

The two applications presented in this chapter demonstrate unequivocally the validity of the FIR approach to qualitative modeling and simulation of biomedical systems, and show that the FIR approach is at least as powerful as if not better suited than other well–established modeling techniques, such as the ANN and NARMAX methodologies, for these kinds of tasks.

Yet, although the results presented in this chapter are remarkable as they stand and raise high expectations for the future, there are still lots of known limitations to the FIR methodology. The next chapter tries to explain some of these limitations.

Chapter 7

Limitations to Predictability of Behavior Using FIR

7.1 Introduction

In Chapter 6, the validity of the FIR methodology for use in biomedicine has been demonstrated. It was also shown that this new methodology compares quite favorably with more established inductive modeling techniques, such as the ANN and NARMAX approaches. Two real biomedical applications have been discussed relating to modeling and control problems in anesthesiology and in cardiology.

The good results obtained make fuzzy inductive reasoning a promising qualitative modeling and simulation methodology, offering good perspectives when used in the biomedical domain. Yet, and in spite of the good results obtained in the previous biomedical applications, some limitations to predictability of behavior exist when using fuzzy inductive reasoning. These limitations have also been studied in this doctoral thesis and are explained in some detail in the following sections.

The reader may already have noticed that the forecasts obtained in the cardiological problem were considerably better than those obtained in the anaesthesia example. This has probably to do with the fact that the signals in cardiology are more repetitive than those in anaesthesia, beside from the fact that we had more past data available in the cardiology case. Somehow, the fact that the relationship between inputs and output in the anaesthesia example was not totally *causal*, hurt our prediction capabilities badly. The question

now is, can this –until now quite vague– notion of *causality* in input/output behavior be quantified? Can we look at a data set and determine beforehand how causal the data are, how likely it is that good predictions can be obtained relating the observed "inputs" to the perceived "output"?

The concept of a **causality horizon** will be introduced. The causality horizon helps determine the likelihood of success of a qualitative prediction. This research is a consequence of personal experiences that have demonstrated limitations to predictability of behavior in biomedical applications, whereas predictions in technical areas had mostly been accurate far beyond our original expectations.

A word of warning! Whereas the last chapter showed the FIR methodology in all its glory, and may have raised high expectations in the reader, this chapter is a chapter of doom. It points out an important problem, and the answer is rather depressing for the most part. Yet, this is an important chapter. While it may by pleasant to know your friends, it is more essential to know your enemies. Some battles can be won by taking proper precautions. Others cannot, and this is very useful knowledge also. Battles that cannot be won should best be avoided.

7.2 The Causality Horizon

This section focuses on limitations to predictability of system behavior through induction. These limitations are demonstrated by two types of systems: a *linear state–space model*, and observations of input/output behavior of a *biomedical system*. The causality horizon is introduced, a conceptual barrier limiting the predictability of future states of the system under investigation.

Experiences with technical and biomedical applications have shown that the quality of predictions is not always the same. In particular, it is much more difficult to obtain even half–way decent predictions for many of the biomedical applications, whereas the predictions in technical applications are often accurate far beyond expectations. It is the purpose of this section to illuminate and explain this discrepancy.

7.2.1 Determination of the Mask Depth

Chapter 3 offers a complete description of the FIR methodology. In that chapter, the determination of the mask depth is explained in detail. It has been claimed that the mask should cover the largest time constant of interest, t_l, and that the sampling rate, δt, should be no larger than one half of the shortest time constant, t_s, of the system to be captured by the model, thus:

$$\Delta t \geq t_l \quad ; \quad \delta t \leq \frac{t_s}{2} \tag{7.1}$$

where Δt, is the time span to be covered by the mask.

The *depth* of the mask should then be computed as follows:

$$depth = \text{round}(\frac{\Delta t}{\delta t}) + 1 \tag{7.2}$$

In this section, it is shown that the mask depth is not only dictated by the two time constants mentioned earlier, but is limited also by yet another factor that has been coined the *causality horizon*, H_C. The causality horizon can be defined as an *upper limit of the depth of the optimal mask beyond which no good forecast quality can be expected.*

Up to this point, no measure of the causality between inputs and outputs was taken into account. Seemingly, any two signals can be declared as "input" and "output" of a "system," and a system response can be predicted between them. Obviously, this cannot be done. A *measure of causality* should be introduced that allows to determine the likelihood of success of a qualitative prediction. Such a measure is the *correlation function.*

For the following mask:

$$\begin{array}{c|ccccc} {}_t\backslash^x & u_1 & u_2 & u_3 & y_1 & y_2 \\ t-2\delta t & 0 & 0 & -1 & -2 & -3 \\ t-\delta t & 0 & -4 & 0 & 0 & 0 \\ t & -5 & 0 & 0 & +1 & 0 \end{array} \tag{7.3}$$

representing the qualitative equation:

$$y_1(t) = \tilde{\text{f}}(u_3(t-2\delta t), y_1(t-2\delta t), y_2(t-2\delta t), u_2(t-\delta t), u_1(t)) \tag{7.4}$$

the output y_1 depends on current and past values of the inputs u_i, and on past values of the outputs y_i. The autocorrelation function for y_1, as well as the cross–correlation functions between u_i and y_1 and the cross–correlation between y_2 and y_1 can be computed. All these functions decay for sufficiently large values of the time lapse Δt. The correlation functions can be viewed as *measures of causality*. Once a sufficiently long time span Δt has elapsed, the output, $y_1(t)$, is no longer causally related to any of the inputs, $u_i(t - \Delta t)$, the other output, $y_2(t - \Delta t)$, or its own past, $y_1(t - \Delta t)$, since the corresponding correlation functions for this value of Δt are small.

By making Δt larger and larger, the inductive reasoner is told to predict the future from old data values that are no longer causally related to the current time. This obviously can't work. The effect is that, even if the best possible mask spanning Δt time units is used, recurrences of the same input patterns lead to all legal output values with approximately equal probability. This is just another way of saying that the output does not causally depend on these inputs. The forecasting algorithm within the FIR methodology is therefore uncertain which value to predict and chooses one of the values arbitrarily, assigning to its forecast a low confidence value. In those cases, the forecast is poor and looks like noise.

If Δt is chosen smaller than the shortest time constant to be captured, the FIR forecast basically consists of a constant value (in lack of better knowledge, tomorrow's weather is predicted to be the same as today's). If Δt is increased to cover the fast time constants but not the slow ones, the forecast exhibits local maxima and minima where the real data show them, but the forecast won't follow the general trend, i.e., it cannot follow the slow time constants. If Δt can be chosen sufficiently large for all time constants to be covered but not larger than the causality horizon, the forecast will be the best that can be obtained. All these types of behaviors can be seen in the following examples.

7.2.2 Linear System

The linear system used in this example is described by the following equations:

$$\dot{\mathbf{x}} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -3 & -4 \end{pmatrix} \cdot \mathbf{x} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \cdot u \tag{7.5}$$

$$y = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \cdot \mathbf{x}$$

It is demonstrated with this example, how the selection of δt and Δt influence the quality of the forecast. To this end, Δt is varied, and an optimal mask is computed in each case using a subset of the available data. These optimal masks are then used to forecast the remainder of the data stream.

For the different tests presented in this section, each of the two variables, u and y, was recoded into three qualitative classes.

Three major types of behavior can be observed in this experiment:

$\underline{\delta t < t_s, t_l < \Delta t < H_C}$

In this experiment, δt and Δt were calculated using the characteristics of the linear system. This system exhibits two time constants, a slow (large) one of $t_l = 2.7$ seconds, and a fast (small) one of $t_s = 0.3$ seconds. Consequently, $\delta t = 0.15$ seconds and $\Delta t = 2.7$ seconds were used, which yields a mask depth of $depth = 19$, as can be seen from Equation 7.6:

$$depth = \text{round}(\frac{2.7}{0.15}) + 1 = 19 \tag{7.6}$$

With these values for Δt and δt, the reasoner operates in a region where both time constants are captured and, as shown in Figure 7.1, the cross–correlation function between input and output is comfortably large. The autocorrelation function of the output variable has also been computed. It exhibits the same wide correlation as the cross–correlation function between the input and the output. Therefore, the predictions of future output values are expected to be good, as can be verified in Figure 7.2.

The optimal mask that represents this system and that was used to obtain the good results presented in Figure 7.2 is shown in Equation 7.7.

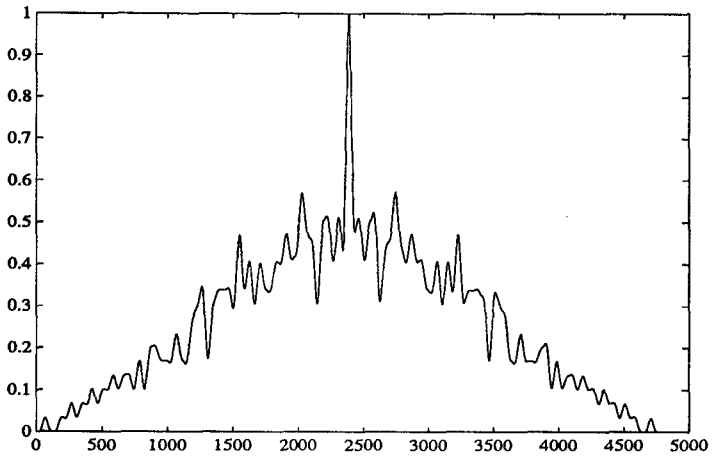


Figure 7.1: Cross–Correlation Function

$$
\begin{array}{l|cc}
{}_t\backslash^x & x & y \\
t-18\delta t & -1 & -2 \\
t-17\delta t & 0 & 0 \\
\vdots & \vdots & \vdots \\
t-10\delta t & 0 & 0 \\
t-9\delta t & -3 & -4 \\
t-8\delta t & 0 & 0 \\
\vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 \\
t & -5 & +1
\end{array}
\tag{7.7}
$$

$\underline{\delta t < t_s, t_l, H_C < \Delta t}$

In the second test, the concept of the causality horizon is being illustrated. The mask depth is repetitively increased, leading to a progressive deterioration of the prediction quality. This happens because the mask now stretches beyond the limit imposed by the causality horizon. When the mask depth is increased to a value of 101 rows, corresponding to $\Delta t = 15$ seconds (see Equation 7.8):

$$depth = \text{round}(\frac{15.0}{0.15}) + 1 = 101 \tag{7.8}$$

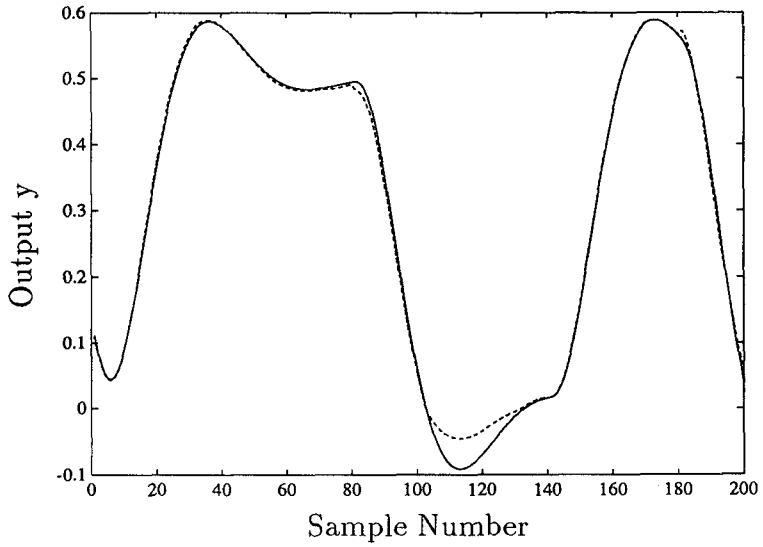


Figure 7.2: Linear System: Real *vs.* Predicted Behavior (Mask Depth 19)

the optimal mask obtained in that case is the following:

$$
\begin{array}{l\backslash x} & x & y \\
t-100\delta t & -1 & -2 \\
t-99\delta t & 0 & 0 \\
\vdots & \vdots & \vdots \\
t-51\delta t & 0 & 0 \\
t-50\delta t & -3 & -4 \\
t-49\delta t & 0 & 0 \\
\vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 \\
t & 0 & +1
\end{array}
\tag{7.9}
$$

The future output values predicted by this mask look like noise, as shown in Figure 7.3. The cross–correlation and autocorrelation functions have decayed significantly, so that the causal relationship between input and the output is no longer sufficiently strong.

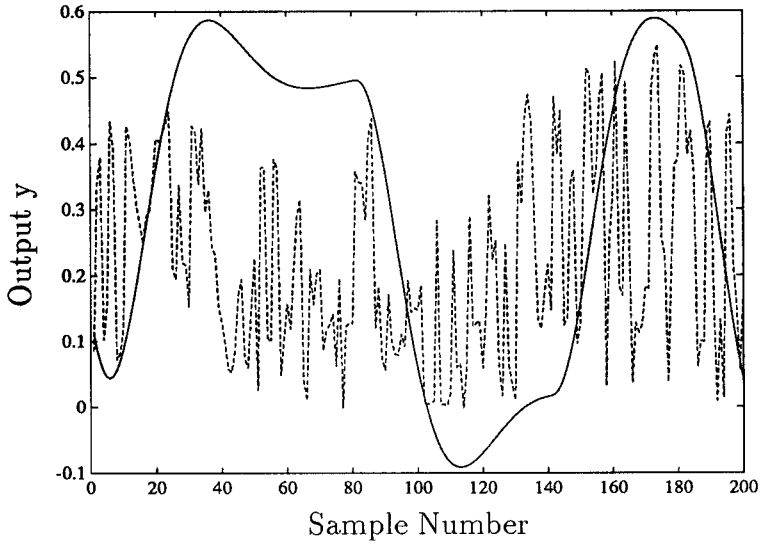


Figure 7.3: Linear System: Real *vs.* Predicted Behavior (Mask Depth 101)

$\underline{\delta t, \Delta t < t_s, t_l, H_C}$

In a third experiment, the time span covered by the mask is made smaller than the shortest time constant to be captured by the model. A δt of 0.003 seconds and a mask depth of $depth = 3$ are used in this test, therefore, Δt covers only 0.006 seconds, a value smaller than the faster of the two time constants. This relation is shown in Equation 7.10.

$$depth = \text{round}(\frac{0.006}{0.003}) + 1 = 3 \tag{7.10}$$

The optimal mask obtained in this case is presented in Equation 7.11.

$$\begin{array}{c} {}_t\backslash^x \\ t - 2\delta t \\ t - \delta t \\ t \end{array} \begin{array}{c} \begin{array}{cc} x & y \end{array} \\ \begin{pmatrix} -1 & -2 \\ -3 & -4 \\ 0 & +1 \end{pmatrix} \end{array} \tag{7.11}$$

As was to be expected, the reasoner predicts a constant value for the output. This is shown in Figure 7.4.

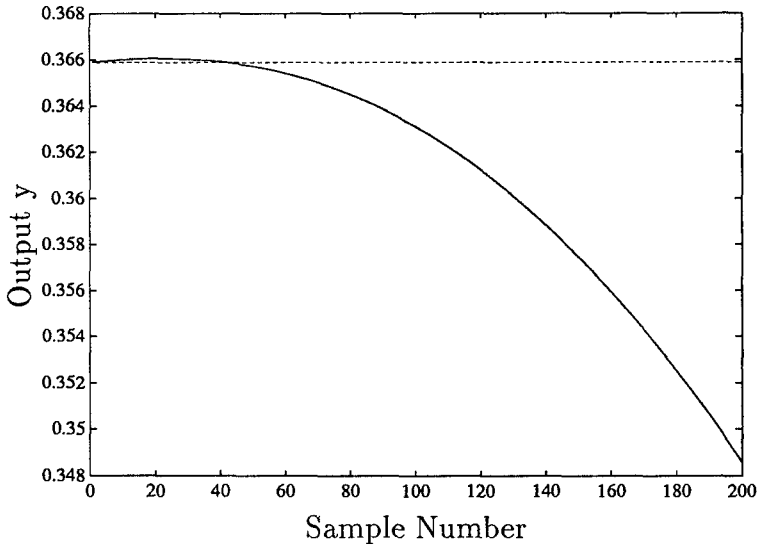


Figure 7.4: Linear System: Real *vs.* Predicted Behavior (Mask Depth 3)

This example shows clearly that the upper limit of the predictability of system behavior is defined by the causality horizon. The correlation functions provide a causality measure that is essential for understanding why the quality of predictions is not always the same. Figure 7.5 shows the forecast quality plotted across the mask depth. The *forecast quality* [1] is defined as:

$$Q = 1 - \frac{\varepsilon_{\text{RMS}}}{\bar{y}} \tag{7.12}$$

where ε_{RMS} denotes the square root of the mean value of the squared error over the prediction period:

$$\varepsilon_{\text{RMS}} = \sqrt{\frac{\sum_{i=1}^{n_{pts}} (\hat{y}_i - y_i)^2}{n_{pts}}} \tag{7.13}$$

[1]The reader may notice the difference in definitions of the term "forecast quality" here and in Chapter 3 of this thesis. In that chapter, we were talking about a *measure* of forecast quality, i.e., an estimate of the forecast quality that was obtained from the qualitative analysis alone by modeling the process of making errors, whereas here, we are talking about the true forecast quality that is obtained by comparing the true and the forecast output values.

and $\bar{y}$ denotes the mean value of the predicted output over the same period. $\hat{y}_i$ denotes the predicted value at point i, and n_{pts} is the total number points of the prediction period.

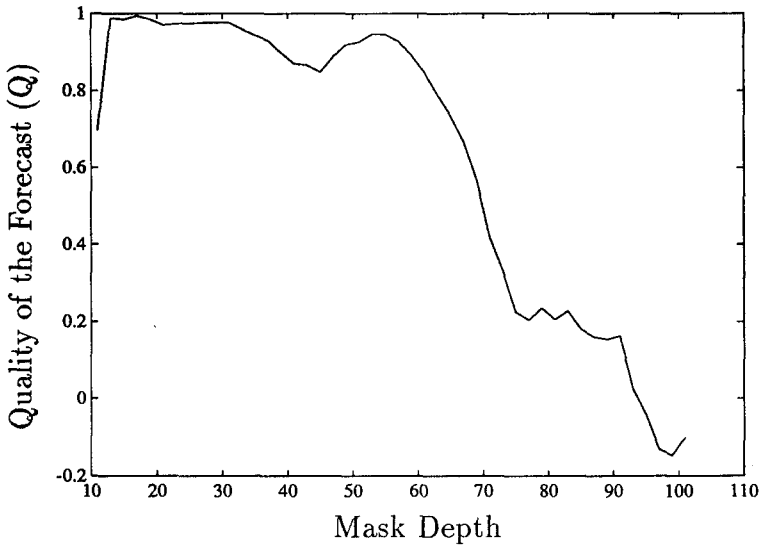


Figure 7.5: Quality of the Forecast *vs.* Mask Depth

It can be noticed that the forecast quality grows up to a mask depth of 13, and then stays at a fairly constant value up to a mask depth of roughly 60. Thereafter, the forecast quality decays rapidly until it becomes totally useless, as shown in Figure 7.3. Thus, the causality horizon in this example is about $H_C = 9$ seconds. From Figure 7.1, it can be seen that, for $\Delta t = 9$ seconds, the cross–correlation has decayed to 55% of its maximum value. Similarly the autocorrelation of the output has decayed to 60% of its maximum value.

Which of these correlation functions is responsible for the decay of predictability? The answer to this question is not straightforward. Let us assume for a moment that we are dealing with a five–variable system with the following optimal mask:

$$
\begin{array}{c} t\backslash^{x} \\ t-2\delta t \\ t-\delta t \\ t \end{array}
\begin{array}{c}
\begin{array}{ccccc} u_1 & u_2 & y_1 & y_2 & y_3 \end{array} \\
\begin{pmatrix} 0 & 0 & 0 & 0 & -1 \\ 0 & -2 & -3 & 0 & 0 \\ -4 & 0 & +1 & 0 & 0 \end{pmatrix}
\end{array}
\tag{7.14}
$$

Evidently, the cross–correlation between y_2 and y_1 is completely harmless in this example, since y_2 is not used at all to predict values of y_1. Let us now assume that the sampling frequency is decreased (δt is increased) until such a point, where the cross–correlation between y_3 and y_1 is not wide enough any more to cover the entire mask depth of $2\delta t$. In this case, the relationship between y_3 and y_1 will no longer be deterministic, and the optimal mask analysis will find a new optimal mask in which y_3 is excluded from the list of input variables. The new optimal mask obtained in this situation has been:

$$\begin{array}{l} {}_t\backslash^x \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{ccccc} u_1 & u_2 & y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -2 & -3 & 0 & 0 \\ -4 & 0 & +1 & 0 & 0 \end{pmatrix} \end{array} \tag{7.15}$$

What will be the effect of this mask change? We shall probably notice a decrease in forecasting power, since y_3 can no longer be considered in the evaluation of y_1. How large this decrease is, depends on the importance of y_3 in the determination of y_1. If y_3 is *very* important, the forecast may have become useless, and the causality horizon has already been reached. On the other hand, if y_3 only added confidence to a prediction that could also have been reached without looking at y_3, the forecast may still be acceptable.

Thus, we need to know how *critical* a variable is in the prediction process. This is not always easy to assess. What we can say is the following: Whenever the cross–correlation of the first critical variable has decayed to a value of 60%, the causality horizon has been reached. In the linear system analyzed in this section, only two variables were considered, and both are critical to the forecasting capability. This explains the causality horizon of $H_C = 9$ seconds found for the above example.

The same study has been realized with a biomedical system, namely the anaesthesia example that had already been used in Chapter 6 of this thesis. This work is presented in the next section.

7.2.3 Biomedical System

As has been mentioned previously, it is much more difficult to obtain good predictions for biomedical than for technical applications. This is due to the qualitative shape of the correlation functions in the two cases. Whereas the technical systems usually offer wide correlation functions, biomedical

correlation functions are often quite narrow.

The biomedical system, presented in Chapter 6, for predicting the right value of an anaesthetic agent to be applied to patients during surgery is being used in this study. The clinical variables comprising the heart rate (HR), the respiration rate (RR), and the systolic arterial pressure (SAP), are the key clinical indicator signals to be used for suggesting an anaesthetic *Dose* (the control signal).

For all the tests presented in this section, the variables SAP, HR, and *Dose* were discretized into three qualitative classes, whereas RR was discretized into two qualitative classes only.

According to information obtained from anaesthetists, the slowest time constant of interest in this system is on the order of 10 minutes, and the fastest time constant of importance is on the order of one minute (Nebot *et al.*, 1993a).

As in the case of the linear system, the variation in forecast quality as a function of the time span covered by the mask, Δt, will be shown.

$\underline{\delta t < t_s, t_l < \Delta t < H_C}$

In accordance with the previously made recommendations, values of $\delta t = 0.5$ seconds and $\Delta t = 10$ seconds were chosen. Consequently, the mask depth is 21, computed as shown in Equation 7.16.

$$depth = \text{round}(\frac{10.0}{0.5}) + 1 = 21 \tag{7.16}$$

The optimal mask obtained for this depth is the one shown in Equation 7.17.

$$
\begin{array}{l|cccc}
{}_t\backslash^x & SAP & HR & RR & DOSE \\
t-20\delta t & 0 & 0 & 0 & 0 \\
t-19\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-11\delta t & 0 & 0 & 0 & 0 \\
t-10\delta t & -1 & 0 & 0 & -2 \\
t-9\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 & 0 & 0 \\
t & 0 & -3 & 0 & +1
\end{array}
\tag{7.17}
$$

With this choice, the forecast exhibits indeed the best results that can be obtained for this system. These results are shown in Figure 7.6.

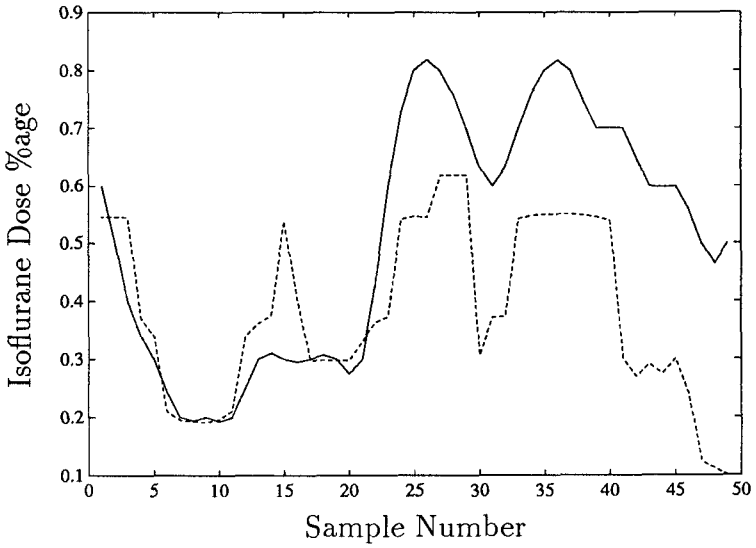


Figure 7.6: Biomedical System: Real *vs.* Predicted Behavior (Mask Depth 21)

It turns out that, in this example, the causality horizon is just about as large as the largest time constant to be modeled. Thus, the forecast is never as good as in the case of the linear system since the effects of causality degradation set in before the slowest time constants are truly and fully covered by the mask. The slowest time constant is just at the limit of the causality horizon. The maxima and minima are predicted correctly, but the forecast

is not bias free. There is a tendency for drifting away. Figure 7.7 shows the autocorrelation function of the output, *Dose*. The autocorrelation function presented in Figure 7.7 is only one of four correlation functions that have been computed in this study. However, only this correlation is presented here, because it is the function that exhibits the most narrow correlation, and since it turns out to be a critical variable, it is the one that limits the forecasting power.

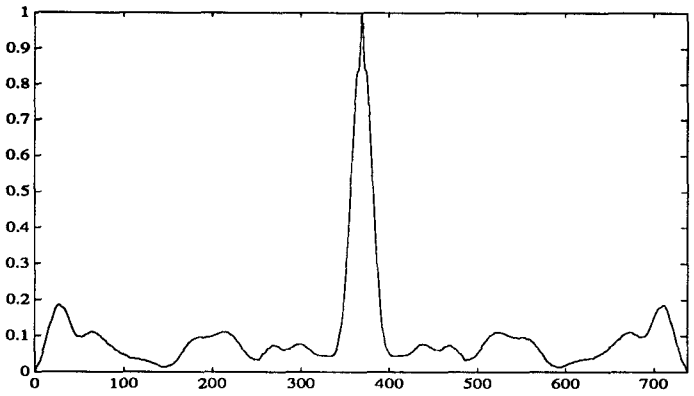


Figure 7.7: Autocorrelation Function (DOSE)

$\underline{\delta t < t_s, t_l, H_C < \Delta t}$

When Δt is increased, the quality of the forecasts deteriorates rapidly, and the forecasts look like noise. In the second test of this series, Δt is chosen bigger than both time constants and the causality horizon. It was decided to choose $\Delta t = 17$ minutes.

Accordingly, the depth of the mask is:

$$depth = \text{round}(\frac{17.0}{0.5}) + 1 = 35 \tag{7.18}$$

Figure 7.8 shows the results obtained with a mask depth of 35. However, test with different mask depths have been computed. In a range from 25 to 51, ε_{RMS} grows proportionally with the mask depth.

The optimal mask obtained with a depth of 35 is different from the mask obtained with a depth of 21. The optimal mask obtained with a depth of 35

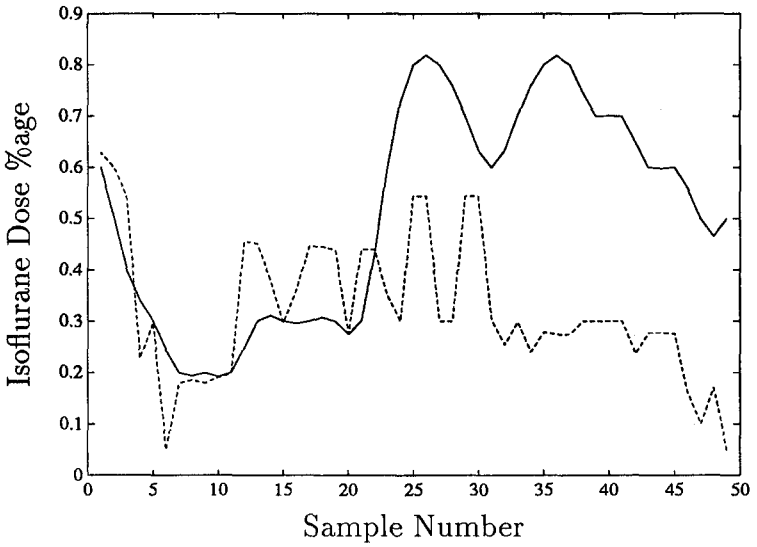


Figure 7.8: Biomedical System: Real *vs.* Predicted Behavior (Mask Depth 35)

is shown in Equation 7.19.

$$
\begin{array}{l\backslash x} & SAP & HR & RR & DOSE \\
t-34\delta t & 0 & -1 & 0 & 0 \\
t-33\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-18\delta t & 0 & 0 & 0 & 0 \\
t-17\delta t & 0 & 0 & 0 & 0 \\
t-16\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 & 0 & 0 \\
t & -2 & -3 & 0 & +1
\end{array}
\tag{7.19}
$$

With this mask depth, the results aren't complete garbage yet, but they are indeed considerably poorer than those obtained with a mask depth of 21.

$\underline{\delta t, t_s < \Delta t < t_l, H_C}$

In a third experiment, Δt will be chosen somewhere between the fastest and the

slowest time constants. The sampling rate, δt is kept at a value of 0.5 minutes, but the mask depth is reduced to 11, such that $\Delta t = 5$ minutes.

$$depth = \text{round}(\frac{5.0}{0.5}) + 1 = 11 \tag{7.20}$$

The optimal mask obtained in this case was:

$$\begin{array}{l|cccc} {}_t\backslash^x & SAP & HR & RR & DOSE \\ t-10\delta t & -1 & 0 & 0 & -2 \\ t-9\delta t & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t-6\delta t & 0 & 0 & 0 & 0 \\ t-5\delta t & 0 & 0 & 0 & 0 \\ t-4\delta t & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t-\delta t & 0 & 0 & 0 & 0 \\ t & 0 & -3 & 0 & +1 \end{array} \tag{7.21}$$

In this situation, the forecast exhibits local maxima and minima where the real data show them, but the forecast does not follow the general trend any longer, because it cannot follow the slowest time constant. This result is illustrated in Figure 7.9.

$\underline{\delta t, \Delta t < t_s, t_l, H_C}$

In the fourth and last experiment of this series, both the sampling rate and the total coverage of the mask were reduced until the mask covered a time interval smaller than the fastest time constant of 0.5 seconds. To this end, a sampling rate of $\delta t = 0.2$ minutes and a mask coverage of $\Delta t = 0.2$ minutes were chosen, leading to a mask depth of $depth = 2$ in accordance with Equation 7.22.

$$depth = \text{round}(\frac{0.2}{0.2}) + 1 = 2 \tag{7.22}$$

As was to be expected, the fuzzy inductive reasoner now predicts that the amount of anaesthetic agent to be administered to the patient should always be the same as during the previous sampling period. This becomes evident already by looking at the optimal mask found in this case:

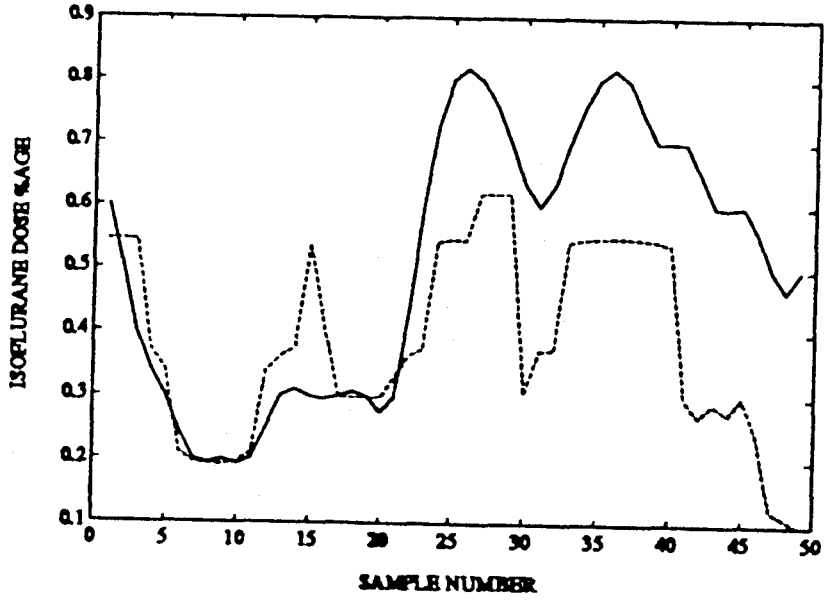


Figure 7.9: Biomedical System: Real *vs.* Predicted Behavior (Mask Depth 11)

$$
\begin{array}{c} {}_t\backslash^x \\ t-\delta t \\ t \end{array}
\begin{array}{c} \begin{array}{cccc} SAP & HR & RR & DOSE \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & +1 \end{pmatrix} \end{array} \tag{7.23}
$$

Consequently, the FIR model here forecasts a straight line. This result is illustrated in Figure 7.10.

7.2.4 Summary of the Relevant Issues

In this section, it was demonstrated by means of two examples that the *causality horizon* is an important factor that influences the forecast quality. It was also shown that the correlation functions are good indicators to estimate the causality horizon.

The first example used was a linear system. In this situation, the causality horizon is comfortably large, much larger than the slowest time constant of interest in the system, thereby guaranteeing a good prediction, if Δt covers

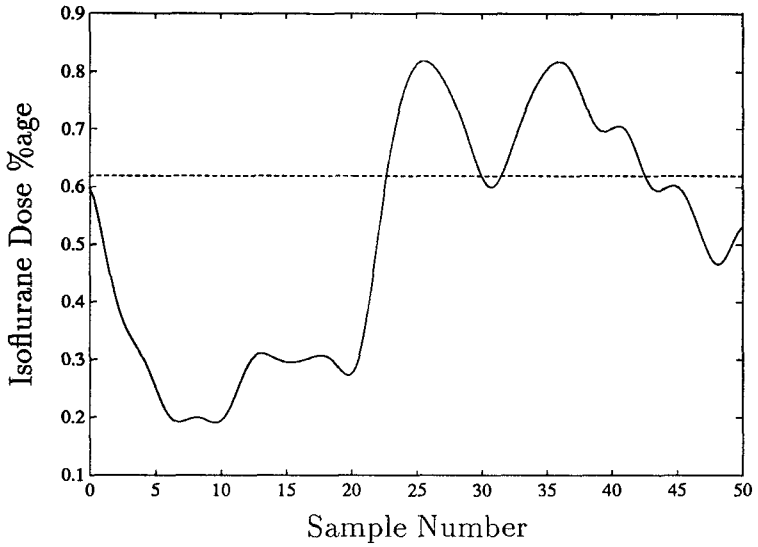


Figure 7.10: Biomedical System: Real *vs.* Predicted Behavior (Mask Depth 2)

both time constants and is not chosen unreasonably large. Several tests were conducted varying Δt, which showed the influence on forecast quality. A proper mathematical definition of the forecast quality was proposed. The forecast quality was computed and plotted across the mask depth in Figure 7.5.

For the second example, a soft science system was selected. Such systems don't lend themselves as readily to obtaining good predictions as technical applications do. This is primarily due to the inherent reduction of their causality horizons. Some biomedical applications exhibit a causality horizon *smaller* than the largest time constant of interest in the system. In such a situation, no decent predictions can be expected under *any* circumstances.

The worst situations are those where human decision making forms part of the model, as this was the case in the anaesthesia example. Such cases severly limit the input/output causality of the system under study and make proper forecasting virtually impossible. One should not blame the FIR methodology for not being able to predict the outcome of throwing a dice. If the so–called input/output behavior of a system under study is not properly correlated, i.e., is not properly characterized by a causality relationship, no methodology in the world can save our neck!

The essential point to be made, however, is that the FIR methodology

indeed does exploit whatever causality exists among the possible input/output relationships of a multivariable system, and makes the best of the situation, better so than most other inductive modeling techniques such as ANNs and NARMAX models.

7.3 Conclusions

This chapter has focused on limitations to predictability of system behavior using the fuzzy inductive reasoning methodology. The limitation is related to the concept of a *causality horizon*. The causality horizon helps to determine the likelihood of success of a qualitative prediction, providing an upper limit on the mask depth beyond which no good predictions can be obtained any longer.

Two systems, a linear state–space model and a biomedical system, have served to demonstrate the concept. These different types of systems were chosen in order to illustrate the differences in predictability between applications from technical domains and those from soft sciences areas. It is much more difficult to obtain decent predictions for soft science systems than to obtain accurate predictions for systems from the hard sciences. In this chapter, this discrepancy has been explained.

Thus, we now know that no good qualitative predictions of system behavior can be obtained unless there exist strong, either positive or negative, correlations between the inputs and the outputs of that system. Thus, no causality is possible without correlation. The question now is: can this statement also be negated, i.e., is it possible to conclude that a large correlation between inputs and outputs of a system *implies* causality?

This is a dangerous proposition if accepted thoughtlessly. As explained in (Cellier, 1991a), there exists a strong positive correlation between the stork populations in Switzerland and the birth rate of human infants in the same country during the twentieth century. Indeed, it is perfectly feasible to predict human birth rates from the stork populations (or vice–versa). Yet, most researchers would shun away from the biomedical implications of assuming a direct causality between these two correlated facts.

Consequently, a high correlation implies *predictability* . . . and it also implies a *latent causality*, yet, it does not imply an *open causality*. What is a latent causality? It is correct that higher educated women in a well–functioning

modern society give, on the average, less births than women with a lower degree of education or women living under primitive conditions. In order to support an advanced social environment and provide for the educational infrastructure needed, more industry is required, which, in turn, eats up the living space of the storks. So indeed, the two facts have a common cause, and this is what we define as *latent causality*. A latent causality between two phenomena indicates that both phenomena are caused by yet another phenomena that is not contained in the data set, rather than one being caused by the other. In our experience, a high correlation is *never* coincidental. It just doesn't happen that way. Thus, at least a latent causality can indeed be implied from high correlation.

The results obtained in this chapter may sound somewhat depressive. Well, if biomedical system modeling and simulation were that easy, all problems would have been solved long ago, and this doctoral thesis would not have been needed. Yet, it is the knowledge of the limitations inherent in a technology that allows us to exploit its strengths by by–passing them, and the very positive results reported in the previous chapter of this thesis and in the chapters yet to come are by no means curtailed by carefully exploring these limitations.

Chapter 8

Dealing With Incomplete Data Records

8.1 Introduction

As has been explained in Chapter 5, one of the major problems in biomedical qualitative modeling is the *lack of information*. Inductive, pattern–based modeling techniques are extremely data hungry. It is therefore essential for behavioral qualitative methodologies to have available a large amount of rich data to work with. Unfortunately, in biomedical applications, this is hardly ever the case.

The lack of information may have several different causes, all of them related to acquisition difficulties. The problems are further amplified when the data records obtained from medical experiments are incomplete. *Incomplete information* is another crucial problem of biomedical applications that makes the information deficiency even worse.

Biomedical data records are notorious for being incomplete. A patient on a heart monitor is routinely taken off the monitor while being cleaned by the nurse. A particular instrument may exist only in one copy. Although the instrument is in use by one patient, it is temporarily removed in order to give it to another patient who needs it more urgently. Signal detectors that are taped to the patients' body often fall off during the night. The recording device (e.g. a tape cassette) is full and is not replaced for a while. There are dozens of circumstances that can produce gaps of information for one, several, or all of the parameters. Qualitative methodologies that cannot deal with missing data

values are therefore quite useless for dealing with biomedical applications.

In other types of systems, e.g. mechanical or electrical, it may be possible to throw out incomplete data sets and repeat the experiment. In medical systems, this is rarely an option. A sufficiently long and rich data stream obtained from a patient is far too valuable an asset to being thrown out on the basis of incomplete data records alone.

The purpose of this chapter is to tackle the incomplete information problem by means of the fuzzy inductive reasoning methodology. To this end, a technique called *missing data option* is proposed that allows to work with incomplete medical data records. This technique represents an enhancement to the FIR methodology. The first part of this chapter describes how the missing data option has been implemented.

A practical study of the limits to predictability of behavior when incomplete data records are present in the training data set is discussed in the second part of this chapter. Two different applications have been used to show those limits: a (generic) linear state–space model, and observations of input/output behavior stemming from a biomedical system.

8.2 Missing Data Option

Even when sufficient and sufficiently rich data are available, incomplete data sets can still pose difficult problems to the modeler, and this situation is unfortunately all too common in the biomedical domain.

The missing data feature enables the researcher to work with incomplete data records and extract as much information from them as they contain. The feature makes it possible to convert incomplete quantitative data sets to reduced qualitative data sets in order to derive the best possible qualitative model for prediction of future system behavior. The feature has been designed for use with all the FIR modules implemented in the SAPS–II tool kit.

As described in detail in Chapter 3 of this dissertation, the fuzzy inductive reasoning methodology is composed of four basic functions: *fuzzification* (fuzzy recoding), *qualitative modeling* (fuzzy optimization), *qualitative simulation* (fuzzy forecasting), and *defuzzification* (fuzzy regeneration). In order to explain how the missing data option has been implemented within the SAPS–II software environment, a description of how this feature affects these four primary FIR modules is provided next.

Medical data bases usually denote missing data by a physiologically impossible value. For example, a value of -999 is a rather uncommon value do denote the systolic arterial pressure of a patient. Thus, when a blood pressure value is missing in the data base for whatever reason, the medical personnel may decide to mark this value down as -999.

It is very convenient if medical data records can be read into the programs that use them without first having to modify them. Thus, it was decided to mark missing values also in SAPS–II by a number that the user can freely allocate. It was also decided that the missing data option should be made as transparent to the user as possible. Thus, rather than modifying all the SAPS–II functions and adding the missing data value to them as an additional argument, a global variable, *miss_data*, has been introduced that influences the behavior of many of the SAPS–II modules when set.

$\gg$ *global miss_data*;
$\gg$ $miss\_data = -999$;

By default, *miss_data* assumes a value of 0, which indicates that the missing data option is disabled, thus 0 is an illegal value for marking missing values.

The *fuzzy recoding* function computes a qualitative triple for each quantitative data entry in the raw data matrix. When a missing data value is encountered, the class value of the corresponding qualitative triple is not computed, keeping the missing data marker as the class value. The membership value is set to one, and the side value is set to zero. For example, if the global variable *miss_data* is set to 500, a quantitative value of 500 is recoded into the qualitative triple $< 500, 1, 0 >$. In this manner, the location of the missing data elements is preserved within the qualitative data model. This is important, since data entries in the qualitative data model are time–stamped, yet the time stamp is implied, i.e., it is not written into the data record itself.

The call of the *recode* module looks exactly the same as presented in Chapter 3 of this thesis. It thus makes no sense to repeat the description of the syntax of calls to the *recode* module in this chapter. Only the semantics of the recoding operation have been slightly modified by introducing the global variable *miss_data*.

Whereas SAPS–II can take care of missing data values correctly within its intrinsic functions, the user is responsible for properly handling missing

values when manipulating data records manually. For example, SAPS–II does not currently offer a function to compute landmarks. This operation is so simple that, until now, landmarks were always computed in a short Matlab (or CTRL–C) code segment, as indicated in Chapter 3. In the following example, landmarks are computed in the presence of missing data values. Evidently, the approach advocated in Chapter 3 will fail, since missing data values are usually marked through large either positive or negative numbers outside the legal domain of the variable, numbers that, without special care, would be treated as highly visible and prominent landmark values.

$\gg$ $[pointer, m]$ =sortm$(meas)$;
$\gg$ $[row, column]$ =size$(meas)$;
$\gg$ **for** $i = 1 : colum$,
$\gg$ $p1$ =round($pointer(i)/3$);
$\gg$ $p2$ =round($2 * pointer(i)/3$);
$\gg$ $lm = [\ m(1, i)$
$\gg$ $0.5 * (m(p1, i) + m(p1 + 1, i))$
$\gg$ $0.5 * (m(p2, i) + m(p2 + 1, i))$
$\gg$ $m(pointer(i), i)\ ]$;
$\gg$ **end**

In the above code segment, *sortm* is an auxiliary SAPS–II function that knows about the significance of missing values, places them at the end of each of the sorted trajectories, represented by column vectors of m, and sets the corresponding pointer variable, *pointer(i)*, to the last significant data value within the i^{th} sorted data vector, $m(:, i)$. Notice that a single call to *sortm* sorts the entire raw data model by looping over its columns (trajectories).

In order to identify the model that best represents the system, the *fuzzy optimal mask* function is used. It performs an optimization that finds the best model (mask) among all legal models (masks) through a mechanism of exhaustive search. Each of the candidate masks is compared to the others with respect to its potential merit. The optimality of the mask is evaluated with respect to the maximization of its forecasting power using an uncertainty measure and a complexity measure (cf. Chapter 3).

In order to evaluate the quality of a mask, it is necessary to have available the input/output matrix. As was mentioned in Chapter 3 of this doctoral thesis, the input/output matrix is obtained by shifting the mask over the episodical behavior, picking out the selected inputs and outputs, and writing them together in one row. Hence it can happen that the input/output matrix

contains missing elements that might get used during the quality evaluation. It is therefore essential to eliminate from the input/output matrix all data records that are contaminated by missing values. The SAPS function that computes the optimal mask has been modified to eliminate all contaminated records from the input/output matrix.

As with the *recode* function, the syntax of calls to the *foptmask* function is unchanged from that described in Chapter 3. Only the semantics of such a call are slightly modified. The same holds for the *fforecast* and *regenerate* functions.

Once the optimal model that describes the system has been found, it can be used to forecast future outputs of the system. When the missing data option is activated, the forecasting function (just like before the optimal mask function) goes through the input/output matrix, deleting all rows that are contaminated by missing values. This corresponds to the elimination of illegal rules from a rule base. Once the input/output matrix is free of contaminated data records, the behavior matrix is computed from it, and the class, membership, and side values of the desired output can be forecast.

It is important to distinguish between two types of past data: (i) the *history data* that is being used to recognize similar behavioral patterns in the past, and (ii) the *immediate past data* that is used by the recursion of the finite state machine. Whereas missing values in the history data base are comparatively harmless since contaminated records can simply be eliminated, missing data values among the immediate past data are much more critical. If the current forecast requires an immediate past value that is missing, the routine needs to backtrack to first come up with a prediction of that value.

The modifications needed to upgrade the *regenerate function* were trivial. Contaminated qualitative triples are simply converted to missing quantitative values. The quantitative regenerated value of a missing qualitative triple (e.g. $< 500, 1, 0 >$), is simply the class value of that triple (i.e., 500).

The idea behind this kind of an implementation is to remove the user as much as possible from the details of the internal codification. In this manner, by pre–setting the *miss_data* variable to zero at the beginning of the program, a user who does not need the missing data option can ignore this feature altogether. If the same user later on encounters a data file that is contaminated by missing data values, all he or she needs to do is to reset the global variable *miss_data* to a new value. Most of the previously written (Matlab) M–files should still be operating correctly. Exceptions would be

functions that manipulate the data manually, as was the case in the earlier presented example of computing landmarks.

8.3 Limits to Predictability

How many missing data values can be tolerated before the forecasting power of the fuzzy inductive reasoner deteriorates? This question is difficult to answer in a precise quantitative fashion. The missing data feature is implemented in such a way that missing data do not affect the forecasting power of the model *per se*. It is the lack of training data that affects the forecasting power. Thus, as long as the training data set is sufficiently rich, missing data will not affect the forecasting very much, i.e., missing data can be compensated for by redundancy in data records. A significant degradation of the forecasting power will be experienced when the number of missing data records is so large that the richness of the training data set no longer suffices to compensate for the loss, or if the missing data are in some way systematic. For example, if the missing data always occur when one of the variables is at its peak value, then obviously, the training data set no longer contains any information as to how the model should behave when that variable is at its peak value, and the model will be unable to forecast appropriately in that situation. After all, forecasting in fuzzy inductive reasoning is only a smart way of remembering (associating) similar past behavior with the current situation. Two examples are used in this chapter to clarify these statements.

8.3.1 Biomedical Application

The biomedical system presented in this example is a subsystem of the central nervous control system of the human heart. More specifically, it concerns the *venous tone controller* that had already been introduced in Chapter 6 of this thesis. The most important characteristics of this system can be captured by a SISO model where the input signal is the *carotid pressure* and the output signal represents the *control of the venous tone*, i.e., a signal that influences the compliance of the vein that finally dictates the blood pressure in the vein itself.

Figure 8.1 shows the input and output signals that are used to identify the qualitative model. In this data set, there are no missing values. 1500 data records are used as identification data to obtain the model, whereas the final 300 data records are used for validation. Clearly, there is a large amount of

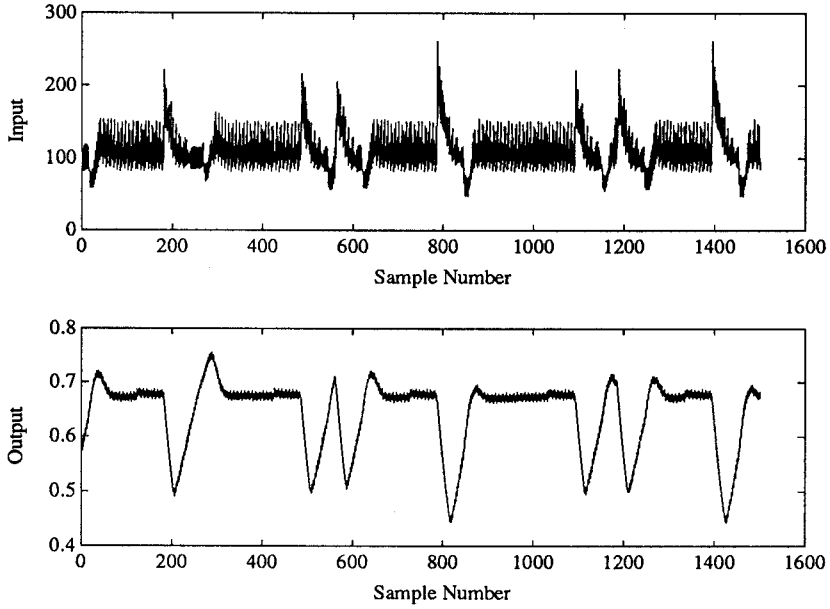


Figure 8.1: Input/Output Data for Model Identification

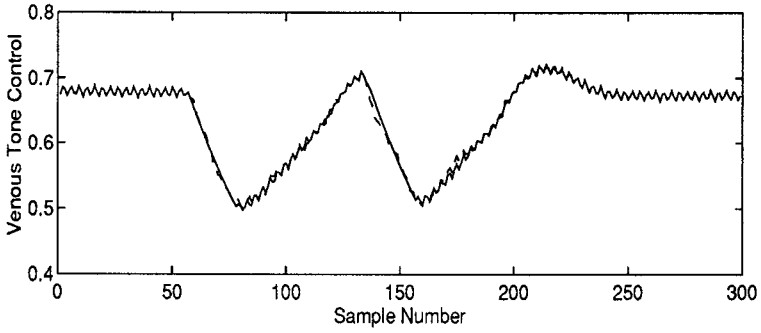


Figure 8.2: Forecast of the Biomedical System without Missing Data

rich data available for the identification of the model.

The input and output signals were first recoded in such a way that the best model representing the dynamics of this system could be obtained. It was decided to recode the two signals into three levels each. In this example, three levels were considered sufficient for obtaining good results. The best qualitative model (the optimal mask) for representing the venous tone controller is the following:

$$\begin{array}{c|cc} {}_t\backslash^x & u & y \\ t-2\delta t & -1 & 0 \\ t-\delta t & 0 & -2 \\ t & -3 & 1 \end{array} \tag{8.1}$$

This mask indicates that the control output at the current time depends on its own past one time step back, and on the carotid pressure at the current time as well as two time steps back.

Now that the best model has been found, we are ready to forecast future behavior. To this end, a new set of data that has not been used in the identification process will be forecast. The measured output signal of the new data set can be used for validation purposes. It will be compared with the forecast output signal. The forecasting results obtained from this model are presented in Figure 8.2. The solid line represents the measurement data, whereas the dashed line denotes the forecast. As can be seen, they are very close. The fast time constant is captured with high accuracy, and also the slow time constant is captured quite well.

It is evident that the identified qualitative model represents the system rather accurately. Note that the data used in the model identification process did not contain any missing values. A series of tests have then been carried out modifying the number and position of missing data entries. A few of them are presented in this chapter.

8.3.1.1 Adjacent Missing Data

In this test, varying amounts of missing values have been inserted in the identification data set adjacent to each other.

In a first test, 10% of the identification data were declared missing, namely those from sampling points 501 to 650. Figure 8.3 shows the identification

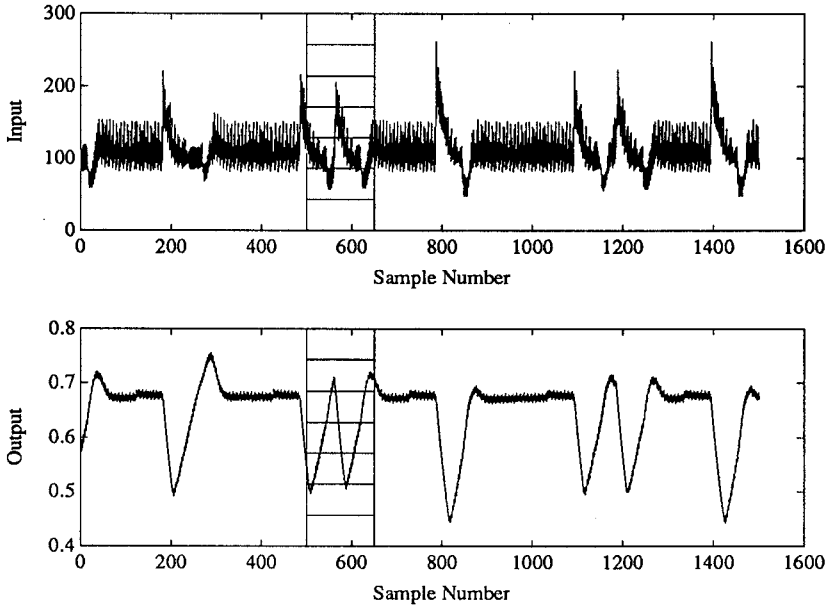


Figure 8.3: Input/Output Data for Model Identification (10% Adjacent Missing Data)

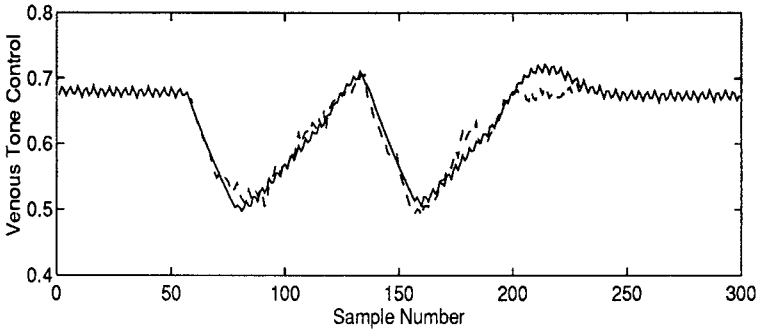


Figure 8.4: Forecast of the Biomedical System with 10% of Adjacent Data Missing

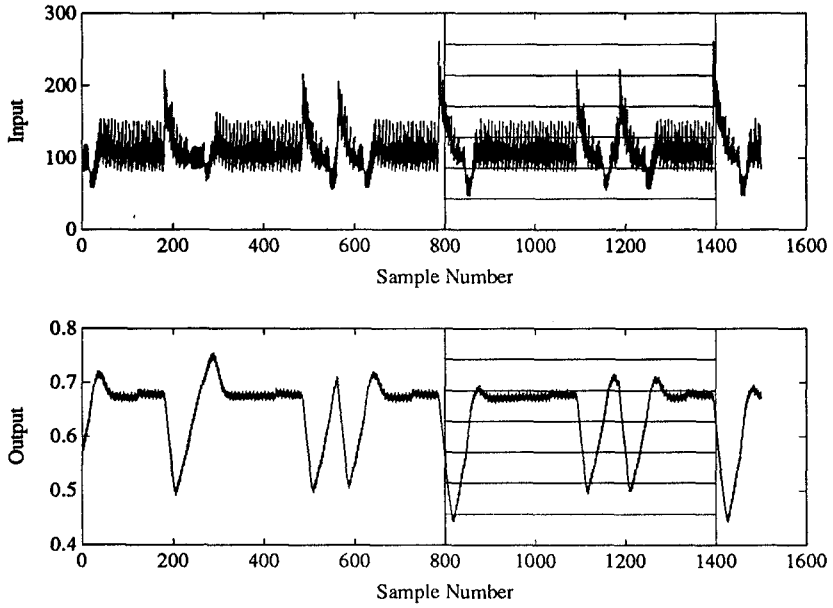


Figure 8.5: Input/Output Data for Model Identification (40% Adjacent Missing Data)

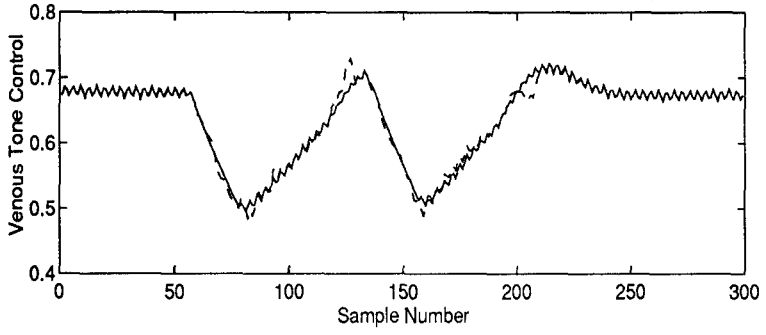


Figure 8.6: Forecast of the Biomedical System with 40% of Adjacent Data Missing

data used to obtain the best mask. The portion of the curves that are marked by a ladder represent missing data, and consequently, unknown behavioral knowledge of the system. In this case, the optimal mask analysis found, as best model, the same mask obtained when no missing values were encountered within the identification data set. Notice that, although the mask is the same, the behavior matrix is different.

As can be seen from Figure 8.3, this portion of the data is the one that is most similar to the curve to be forecast. Therefore, although the amount of missing data is not very high, its loss is significant for the prediction. The forecast results obtained in this test can be seen in Figure 8.4. The prediction still works, but its quality is definitely reduced with respect to the results obtained when the identification data set does not contain any missing data (cf. Figure 8.2).

In a second test, the identification data set contains a gap of 40% missing values from sampling points 800 to 1400. Figure 8.5 shows this training set.

Here, a different mask is obtained as the best model to represent the system:

$$\begin{array}{c} {}_t\backslash^x \\ t-2\delta t \\ t-\delta t \\ t \end{array} \begin{array}{c} \begin{array}{cc} u & y \end{array} \\ \begin{pmatrix} -1 & 0 \\ -2 & -3 \\ -4 & 1 \end{pmatrix} \end{array} \tag{8.2}$$

The set of missing data chosen in this test is not similar to the curve to be predicted (cf. Figure 8.5), and consequently, it is not essential for the model to capture them. Hence the prediction is quite good and definitely better than in the previous case in spite of the much larger chunk of missing data, as shown in Figure 8.6.

8.3.1.2 Scattered Missing Data

In this test, 40% data values are missing like in the previous one. However, they are scattered throughout the data file. Groups of 30 to 50 missing data records are distributed arbitrarily along the identification data toggling between input stream and output stream: from samples 51 to 100, 91–140, 232–281, 262–311, 423–472, 451–500, 623–672, 731–780, 821–840, 961–1010, 1061–1110, 1310–1354, and 1410–1459. The identification data set, including the missing data groups, is shown in Figure 8.7.

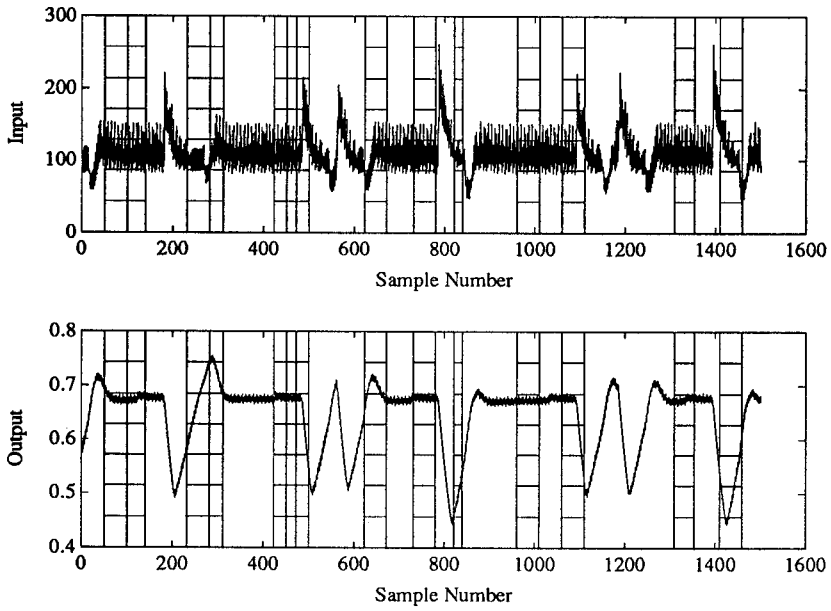


Figure 8.7: Input/Output Data for Model Identification (40% Scattered Missing Data)

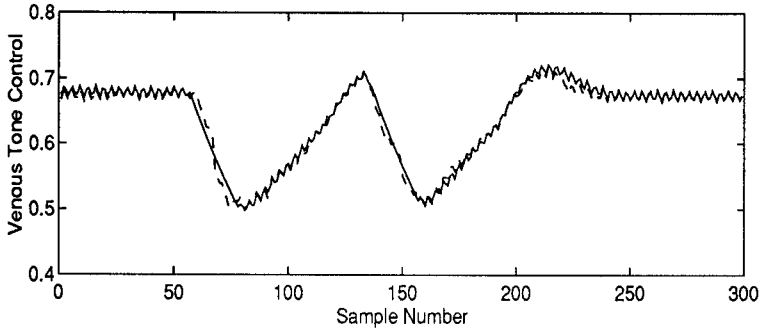


Figure 8.8: Forecast of the Biomedical System with 40% of Scattered Data Missing

In this test, the original model obtained with the full data set is still valid. The results of the forecast are shown in Figure 8.8.

As can be seen from Figure 8.8, the forecasting power is not significantly reduced by the introduction of the missing data values. This is due to the fact that the relevant information contained in the identification data set is not lost.

8.3.2 Linear Model Application

The linear system presented in this example is a position servo mechanism. The input of the system is a square wave signal, and the output is the angular position of the servo measured in radians. The input and output signals used to identify the model are shown in Figure 8.9. The data don't contain any missing value. 1150 data points are used as identification data to obtain the model, whereas 500 values are used for validation.

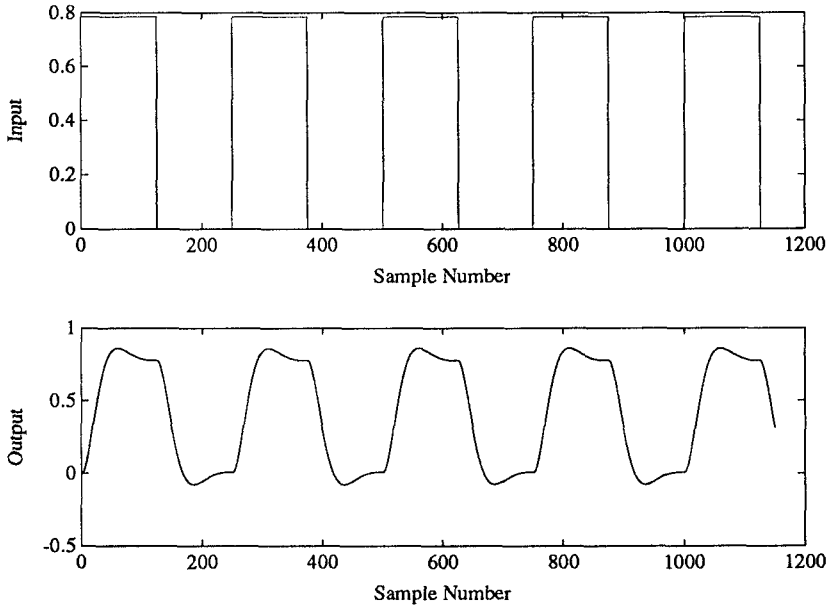


Figure 8.9: Input/Output Data for Model Identification

In this example, the input is recoded into 2 levels, because it is a square wave signal, whereas the output signal is recoded into 3 levels. The best model found representing the position servo mechanism is the following:

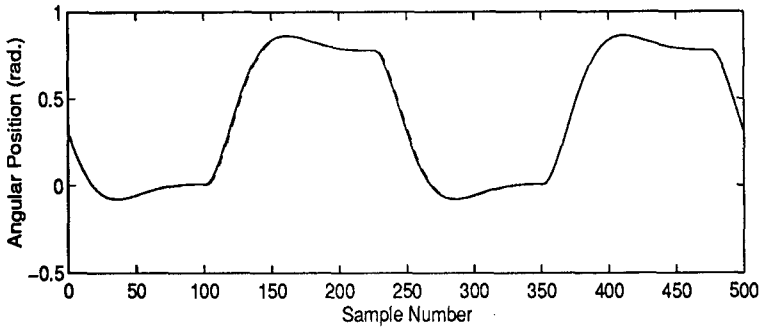


Figure 8.10: Forecast of the Linear System without Missing Data

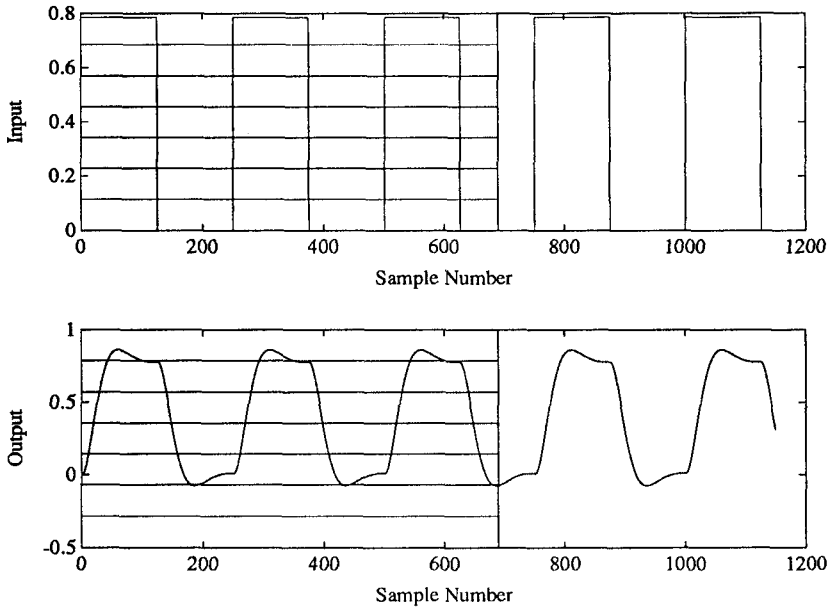


Figure 8.11: Input/Output Data for Model Identification (60% Adjacent Missing Data)

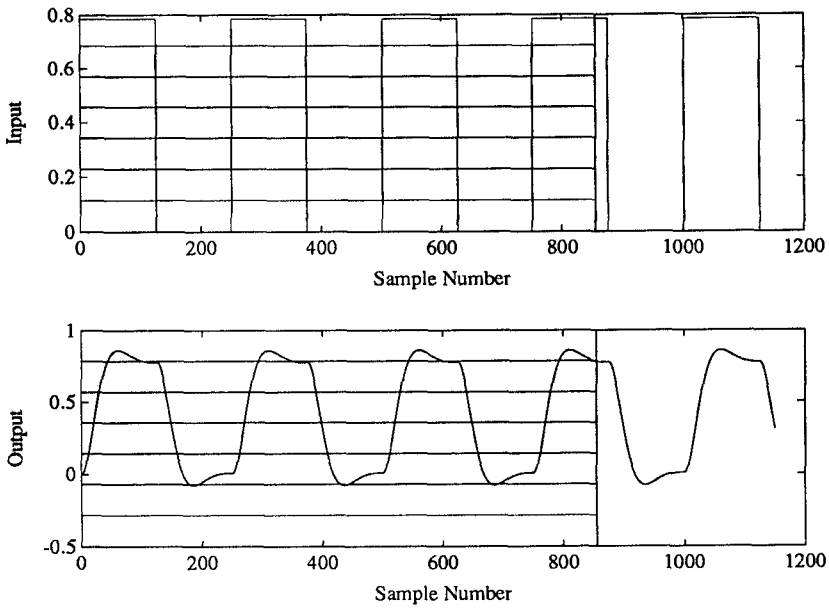


Figure 8.12: Input/Output Data for Model Identification (75% Adjacent Missing Data)

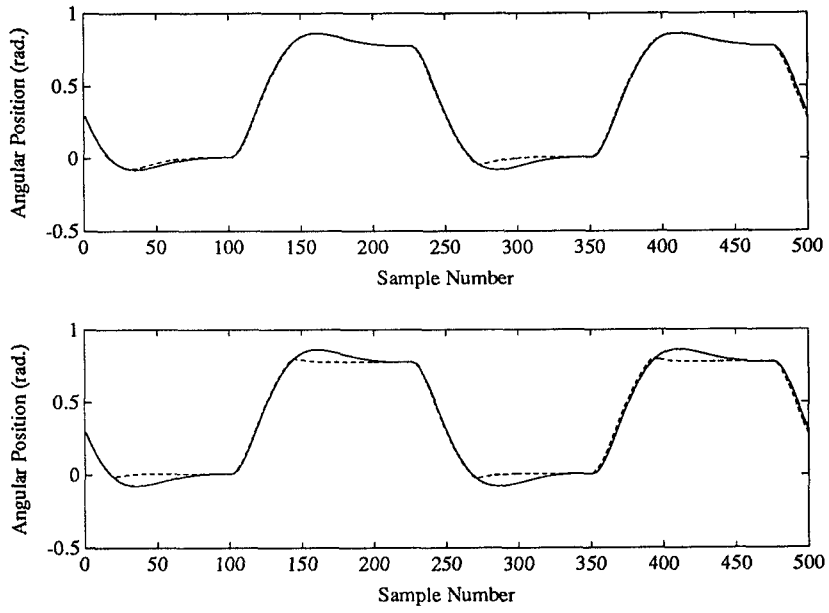


Figure 8.13: Forecast of the Linear System with 60% and 75% of Adjacent Data Missing

$$\begin{array}{c|cc} {}_t\backslash^x & u & y \\ t-2\delta t & -1 & -2 \\ t-\delta t & -3 & -4 \\ t & -5 & 1 \end{array} \tag{8.3}$$

The results of the prediction using this model are presented in Figure 8.10.

As can be seen, the results obtained are excellent, which is not further surprising thanks to the regularity of the output pattern.

8.3.2.1 Adjacent Missing Data

In a first test, different amounts of missing data values are inserted in the identification data set adjacently. As can be seen in Figure 8.9, the redundancy on the original data is quite large, therefore the forecast results obtained with 50% missing data included in the raw data are as good as the results shown in Figure 8.10.

The amount of missing data has then been increased to a level of 60% (from samples one to 690) on the identification data set, as can be seen in Figure 8.11. In this example, the best mask found during the optimal mask analysis was the same one encountered when no missing data is included in the identification data set.

In this case, the output forecast is affected by the reduction of the available data, as can be seen in the first plot of Figure 8.13. If the amount of missing data is further increased to 75% (from samples one to 855), as shown in Figure 8.12, the results get worse (second plot of Figure 8.13). Here, identification is performed with only one period of the signal, and therefore, no more redundancy exists in the training data.

8.3.2.2 Scattered Missing Data

A different test has been realized in this example. Here, the missing data groups are not distributed arbitrarily as in the biomedical example. The missing groups coincide always with the maximum of the servo position.

If the missing data are always located at the maximum of the output variable, the training data set no longer contains any information indicating

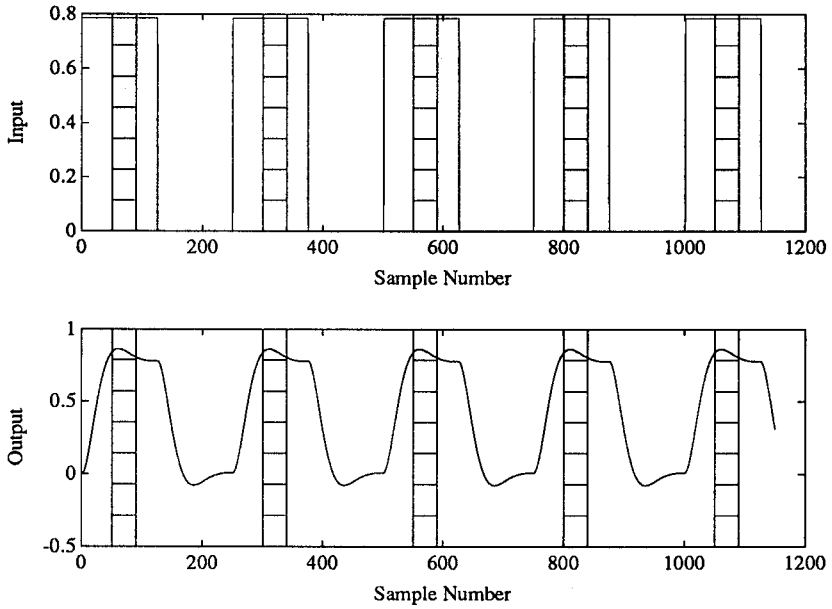


Figure 8.14: Input/Output Data for Model Identification (17% Scattered Missing Data)

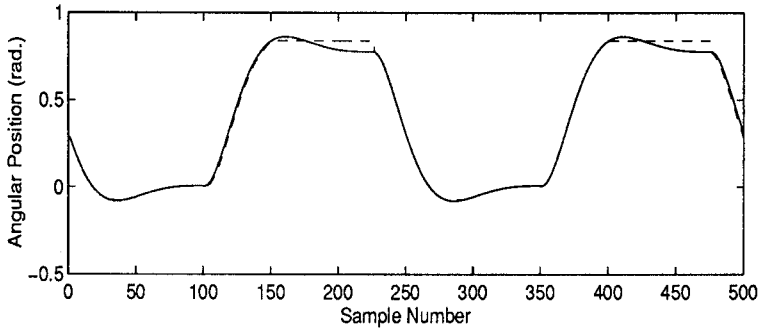


Figure 8.15: Forecast of the Linear System with 17% Scattered Data Missing

how the model should behave when the position of the servo mechanism is at its maximum. In this test, the missing data gaps were inserted at samples 51 to 90, 301–340, 551–590, 801–840, and 1051–1090 of the output variable, corresponding to a data loss of 17%. The identification data set is presented in Figure 8.14. The best model found in this example by means of the optimal mask analysis is the same one as in the previous tests. Using this model, the forecast has been computed and the results are shown in Figure 8.15.

Evidently, a severe degradation of the forecasting power at the peak takes place due to the missing behavioral information in the identification data set.

8.4 Conclusions

This chapter contains a detailed explanation of how the missing data feature for use in fuzzy inductive reasoning has been implemented. It is described how the four pillars of the FIR methodology: *fuzzification*, *qualitative modeling*, *qualitative simulation*, and *defuzzification* are affected by missing data, and how the corresponding modules within SAPS–II have been enhanced to work reliably with incomplete data streams.

A practical study of the limits to predictability with respect to the amount of missing data encountered in the input/output data of the system under investigation is also presented in the chapter. Two different applications were used to show those limits: a (generic) linear state–space model, and observations of input/output behavior stemming from a biomedical system.

The tests done with both systems show that the limits to predictability are difficult to quantify in a precise fashion. The degradation of the forecasting power depends on the richness and the redundancy of the data records in the data history. It is the lack of training data (previous behavioral experience) that affects the forecasting power and not the presence or absence of data gaps.

The next chapter presents an important application of the missing data feature. The missing data technique is used as a tool to aid the modeler in the elimination of patient–specific behavior. The diversity in patient behavior is another important problem inherent to most biomedical applications.

Chapter 9

Elimination of Patient–Specific Behavior

9.1 Introduction

In Chapter 8, we focused our attention on one of the principal problems of biomedical modeling, *incomplete information*. In Chapter 9, another important difficulty of biomedical modeling and simulation is analyzed, namely the *diversity in patient behavior*.

Is knowledge acquired from and about one patient at all applicable to another, and if so, to what extent? Can an inductive qualitative model acquired by analyzing data retrieved from one patient be used to predict the behavior of another? The purpose of this chapter is to discuss these questions in the context of the fuzzy inductive reasoning methodology.

To this end, a technique based on combining knowledge obtained from different patients is presented in this chapter that makes it possible to derive a single model characterizing a specific class of similar patients undergoing similar operations. Data streams stemming from different patients are grouped together, separated by segments of "missing data" in order to prevent the creation of fake causal relationships at the seam between neighboring data streams in the concatenated data set.

A medical application relating to the control of a specific anaesthetic agent administered to patients undergoing surgery is used to demonstrate the feasibility of this method. Two data streams stemming from two different

patients undergoing different operations were used to obtain a single model identifying a similar patient/operation class.

It will be shown that the predictions obtained by this common model are not as good as those obtained from each patient alone using patient–specific models. This is reasonable and characteristic of all knowledge generalization schemes. However, the results obtained are still significant and useful for medical advice in the operating theater.

9.2 Knowledge Combination Technique

This chapter focuses on the preconditioning of biomedical data to eliminate patient–specific behavior. In most biomedical applications, such as, for instance, the control of depth of anaesthesia of a patient undergoing surgery, it is extremely useful to have available a model that identifies not only the behavior of one concrete patient on a specific day during a specific operation, but one that is able to capture the behavior of a class of similar patients undergoing similar operations. The definition of what makes different patients or operations similar is left to the medical experts.

It does not make practical sense, from a medical point of view, to first have to identify a model for a given patient during surgery to be able to predict his or her behavior at some later time. A reliable model must be ready for use before surgery begins. It is therefore important to be able to synthesize a generic model that is valid for a specific type of patient undergoing a given kind of surgery.

This section presents a knowledge combination technique that allows to merge the knowledge stemming from different patients in order to obtain a general knowledge base. This knowledge base can then be used for the prediction of future states of a new patient with characteristics similar to those of the patients used for obtaining the knowledge base.

In order to be able to merge two or more patient data sets, the *missing data option* presented in the previous chapter has been used. This option makes it possible to merge data streams stemming from different patients for the purpose of desensitizing the derived qualitative model to patient–specific characteristics of the observed data.

Since the research focuses on models of dynamic behavior, the advocated methodology searches for causal relationships between variables measured at

different points in time. Therefore, if the data set stemming from one patient were placed immediately adjacent to the data set stemming from another patient, fake causal relationships would be created at the seam of the two data streams, as can be seen in Figure 9.1.

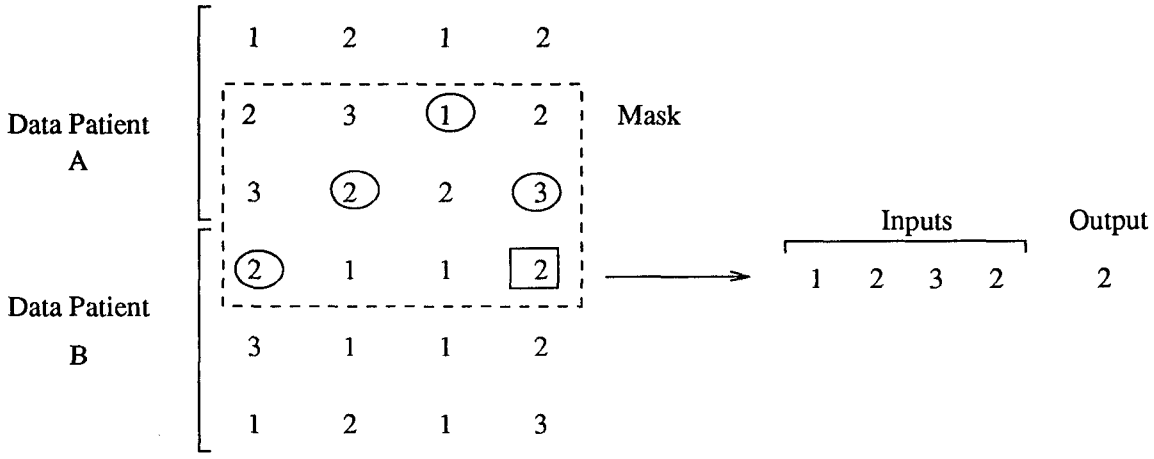


Figure 9.1: Creation of Fake Causal Relationships

The top three rows of recoded data in Figure 9.1 belong to *patient A*, whereas the bottom three rows were obtained from *patient B*. If these data are concatenated, data pertaining to different patients may be combined into a single rule as the mask passes down the qualitative data model converting dynamic relations into static ones. Such rules don't make sense at all. They are thus called *fake relations.*

The creation of fake relations must be avoided under all circumstances, since fake relationships can cause a severe degradation of the forecasting power of the derived qualitative model. The solution is to add gaps of "missing data" between neighboring data streams stemming from different patients, thereby preventing the methodology from extracting, from the combined data set, contaminated data records containing mixed information from different data sources. The gaps must be chosen at least as wide as the mask is deep.

The same physiological signal stemming from different patients may exhibit different dc values, whereas the higher frequency components of the signal may be quite similar. For example, the average blood pressure may vary drastically from one patient to another, yet both patients would react similarly to the application of anaesthetic agents in terms of relative blood pressure increase/decrease patterns.

In the newly proposed technique, data from different patients are merged into a single data stream using a single set of landmarks. In this case, it is necessary to use a linear prefilter to get rid of the dc values of the individual patients' data streams *before* merging them. This simple normalization procedure has been applied to the data of each patient separately, prior to concatenating the two data records.

In order to improve the quality of the prediction and reduce the risk of coming up with entirely incorrect forecasting values, a *voting procedure* is adopted. Instead of working with a single optimal mask, as was done in all the earlier examples, three high–quality masks are determined, and three different state transition matrices are obtained. In the forecasting process, three separate forecasts are computed using the three state transition matrices.

Let M_a, M_b, and M_c be the three selected masks. Each of these masks leads to a different forecast. Let them be called F_a, F_b, and F_c. Three distance measures are computed in the following way:

$$D_a = \text{abs}(F_a - F_b) + \text{abs}(F_a - F_c) \tag{9.1}$$
$$D_b = \text{abs}(F_b - F_a) + \text{abs}(F_b - F_c) \tag{9.2}$$
$$D_c = \text{abs}(F_c - F_a) + \text{abs}(F_c - F_b) \tag{9.3}$$

Once the distance measures have been computed, the predicted value with the largest distance measure is refused. The new forecast value will be the mean value of the two predicted points obtained with the two remaining masks. For instance, if $D_b > D_a$ and $D_b > D_c$, then forecast F_b is rejected, and the new forecast is computed as:

$$F = \frac{Fa + Fc}{2} \tag{9.4}$$

This technique offers a systematic way to compute predictions for all patients in the patient/operation class. Evidently, we could have applied the same technique already earlier, i.e., when dealing with individual patient models. However, good forecasts could be obtained in that case without going through the additional computational effort of implementing the voting algorithm. In the new context, it is more difficult to come up with good forecasts, and therefore, the additional expense is justified.

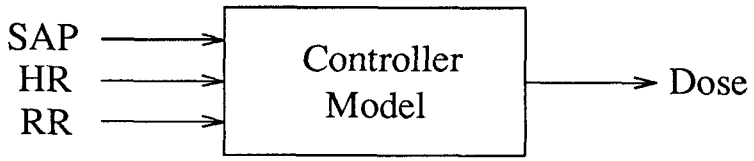
In the next section, the voting method is used in a biomedical application.

9.3 Anesthesiology Application

The efficiency of the advocated technique is being demonstrated by means of an application from anesthesiology. The technique is applied to a biomedical system for predicting the right value of an anaesthetic agent to be administered to patients during surgery. This is the same system that had already been introduced in Chapter 6 of this dissertation. As has been explained earlier, the control of the depth of anaesthesia is a difficult undertaking. The use of anaesthetic agents can produce severe complications and side effects, which, under extreme conditions, may even cause the death of the patient. It is therefore essential that the dose of anaesthetic agents be limited to the minimum amount necessary for proper anaesthesia, thereby reducing undesired side effects and minimizing the risk to the patient.

The same two data streams stemming from two different patients undergoing different types of surgery that had been used in Chapter 6 are analyzed in the current chapter.

Three physiological variables, the *heart rate* (HR), the *respiration rate* (RR), and the *systolic arterial pressure* (SAP) of the patient were selected as the key clinical indicator signals to be used for suggesting the proper amount of anaesthetic agent, *Dose*, to be administered. The controller model is determined by the qualitative relationship between its three input variables, SAP, HR, and RR, and its single output variable, *Dose*.



Separate qualitative models for the two patients had been found earlier. This research effort was described in detail in the first part of Chapter 6 of this dissertation.

For the first patient, the model (optimal mask) obtained was the following:

$$
\begin{array}{l\backslash^{x}_{t}} & SAP & HR & RR & Dose \\
t-20\delta t & 0 & 0 & 0 & 0 \\
t-19\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-11\delta t & 0 & 0 & 0 & 0 \\
t-10\delta t & 0 & 0 & 0 & -1 \\
t-9\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 & 0 & 0 \\
t & -2 & 0 & 0 & +1
\end{array}
\tag{9.5}
$$

denoting the qualitative relationship:

$$Dose(t) = \tilde{\mathrm{f}}(Dose(t-10\delta t), SAP(t)) \tag{9.6}$$

For the second patient, the optimal mask obtained had been:

$$
\begin{array}{l\backslash^{x}_{t}} & SAP & HR & RR & Dose \\
t-20\delta t & 0 & 0 & 0 & 0 \\
t-19\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-11\delta t & 0 & 0 & 0 & 0 \\
t-10\delta t & -1 & 0 & 0 & -2 \\
t-9\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 & 0 & 0 \\
t & 0 & -3 & 0 & +1
\end{array}
\tag{9.7}
$$

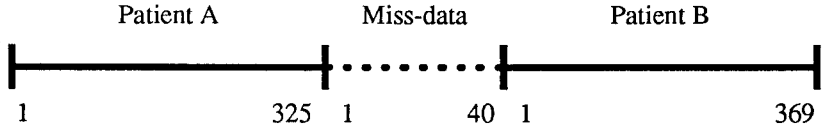
which denotes the qualitative relationship:

$$Dose(t) = \tilde{\mathrm{f}}(SAP(t-10\delta t), Dose(t-10\delta t), HR(t)) \tag{9.8}$$

As can be seen, the obtained qualitative models were distinct. It had not been possible at that time to apply either of the two qualitative models to the other patient and obtain meaningful predictions of that patient's future behavior.

Making use of the new approach for dealing with multiple patient models, it was decided to merge the data from the two patients in order to extract a

set of models common to *both* patients that would hopefully be able to offer acceptable predictions in both cases. The two individual data sets contain 325 and 369 records, respectively. A gap of 40 "missing values" were inserted connecting the two data sets. Consequently, a single data set of 734 values resulted for use in the identification of a set of three suboptimal masks.

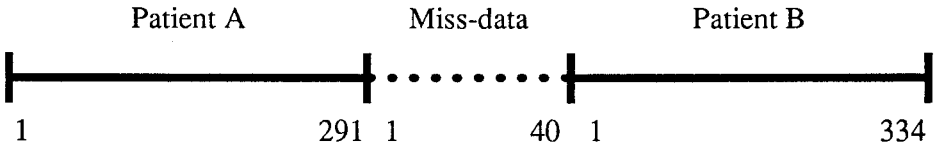


As was to be expected, the dc values of the data sets from the two patients were different, thus, it was decided to normalize the data in the manner previously described.

At this point, the data is ready to start the model identification process. To this end, it is required to recode the data into qualitative variables. It was decided to recode the variables SAP, HR, and $Dose$ into three qualitative levels each, whereas RR was recoded into two qualitative levels only.

The first 291 rows of patient A combined with the gap of "missing values" together with the first 335 rows of patient B were used as past history data to compute the qualitative model.

Past History Data



The optimal mask obtained for the combined data set was the following:

$$
\begin{array}{c}
{}_t\backslash^x \\
t-20\delta t \\
t-19\delta t \\
\vdots \\
t-11\delta t \\
t-10\delta t \\
t-9\delta t \\
\vdots \\
t-\delta t \\
t
\end{array}
\begin{array}{c}
\begin{array}{cccc} SAP & HR & RR & Dose \end{array} \\
\left(\begin{array}{cccc}
0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & -2 \\
0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 \\
-3 & 0 & 0 & +1
\end{array}\right)
\end{array}
\tag{9.9}
$$

This mask denotes the qualitative relationship:

$$Dose(t) = \tilde{f}(HR(t - 20\delta t), Dose(t - 10\delta t), SAP(t)) \tag{9.10}$$

Fuzzy forecasting was then used to predict new qualitative triples for variable *Dose* for the last 34 rows of each patient.

It turns out that the prediction obtained using this optimal mask alone is not good enough, and therefore, the previously described voting method had to be used.

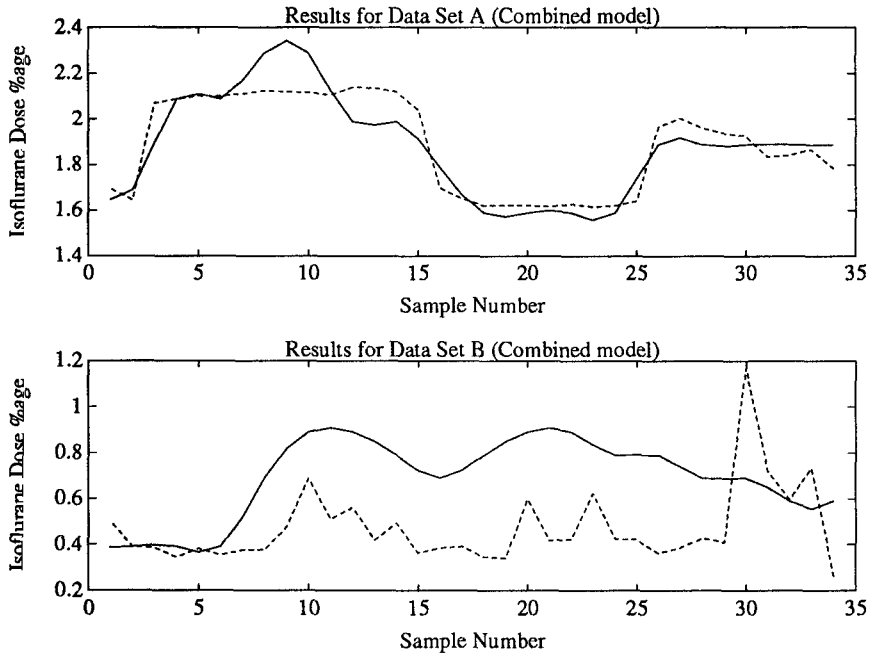


Figure 9.2: Prediction Results Using Combined Models

To this end, a set of three suboptimal masks had to be found. One of the three masks is evidently the optimal mask obtained for the combined data set. The other two masks could be chosen as suboptimal masks from the mask history. However, a different approach was taken. The second mask was obtained using "common sense." We already know that the two optimal masks obtained for the two patient data sets separately are different. Whereas one of them leads to the best forecast for patient A, the other is optimal when used for patient B. However, neither of them gives acceptable results when applied on the other patient. It seems reasonable to assume that if the input patterns

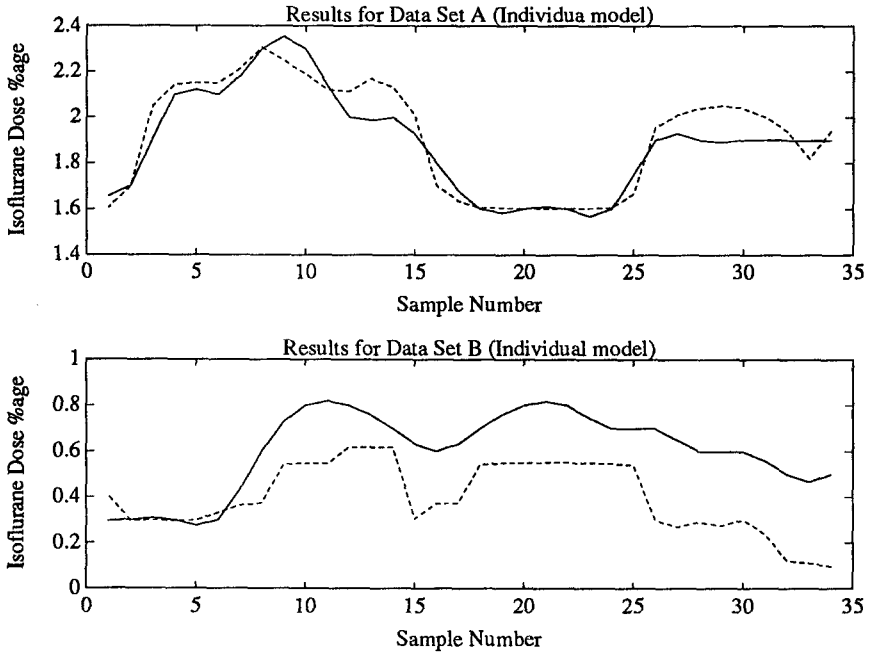


Figure 9.3: Prediction Results Using Individual Models

of the two masks are combined, a decent forecast for both patients could be encountered. Following this reasoning, the second mask was constructed as:

$$
\begin{array}{l|cccc}
{}_t\backslash^x & SAP & HR & RR & Dose \\
t-20\delta t & 0 & 0 & 0 & 0 \\
t-19\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-11\delta t & 0 & 0 & 0 & 0 \\
t-10\delta t & -1 & 0 & 0 & -2 \\
t-9\delta t & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t-\delta t & 0 & 0 & 0 & 0 \\
t & -3 & -4 & 0 & +1
\end{array}
\tag{9.11}
$$

denoting the qualitative relationship:

$$
Dose(t) = \tilde{\mathrm{f}}(SAP(t-10\delta t), Dose(t-10\delta t), SAP(t), HR(t)) \tag{9.12}
$$

This mask, when used alone, gives worse results than the optimal mask, but

in concert with the other two voting masks, it turns out to be acceptable.

Finally, the optimal mask obtained for the second patient when the two data sets were treated separately was chosen as the third mask in the voting set (cf. Equation 9.7).

The forecasting results for the two data sets using the voting scheme are shown in Figure 9.2.

As can be seen, the prediction follows the real curve in an acceptable way. The predictions are not as good as those obtained from the individual models, especially as far as patient B is concerned. However, the results obtained using the combined model are still better than those obtained with the ANNAD system, a feedforward neural network trained for individual patients (Rehman *et al.*, 1993). The results obtained by ANNAD had been judged by medical personnel as "clinically meaningful." Consequently, the results obtained from our combined model are equally valid.

Computing the least square error of the predictions for the two patients, the following results are obtained:

- The predicted *Dose* to be administered to patient A when using the individual model exhibits an error of 0.4886, whereas the error is 0.5449 when the combined model is in use.
- The predicted *Dose* to be given to patient B according to the individual model shows an error of 1.4156, whereas an error of 1.8224 results when using the combined model.

It is quite evident that the predictive power of the combined model *is* poorer than that of the individual models (as had to be expected), but the reduction in predictive power is not unacceptably large.

9.4 Conclusions

This chapter presents a systematic way for *generalizing knowledge* about patient behavior, reaching beyond individual patient–specific behavioral patterns contained in data files reflecting individual patient measurements. The approach chosen is extremely simple, yet very powerful. Individual data streams stemming from different patients are concatenated to each other, while

gaps of missing data are introduced in between each pair of adjacent data streams to prevent a contamination of the generated rule base with fake causal relationships. The dc values are subtracted from each of the data streams before merging them, and, if necessary, the ranges can also be normalized.

It is then left to the faculties of the FIR methodology to *generalize the available knowledge*. The FIR methodology is able to exploit (in an indirect fashion) available correlations between the signals contained in the model, recognize similar behavioral patterns observed in the past, interpolate between these previously seen behavioral patterns, and extrapolate behavior over time.

A representative biomedical example demonstrated how medical information obtained from different patients undergoing similar operations can be processed in a systematic though still not fully automated manner. The measured data streams stemming from different patient/operation pairs are combined in such a way as to allow the synthesis of a common model that can be used for an entire class of similar patients undergoing similar types of surgery.

Using the fuzzy inductive reasoning methodology augmented by the missing data option, it does now seem feasible to generate a single qualitative model that can be used to predict the future behavior of patients within an entire class of similar patient/operation pairs. The predictions are not as good as those obtained from individual models, but they are still clinically meaningful.

A quantification of the degradation of the predictive power of the combined model relative to the individual models was also presented at the end of this chapter. This quantification demonstrates that the prediction errors made by the combined model are not drastically increased in comparison with the prediction errors obtained when using the individual models.

Has the anesthesiology example demonstrated the capability of the FIR methodology to generalize knowledge? Unfortunately, it has not. In order to prove the generalization capability, we would have needed a third data stream stemming from yet another patient, a stream not used in the combined past history data and show that the combined model is capable of also predicting the future behavior of *that* patient.

It is not clear that this might have worked. Although the approach *is* valid –this is evident from the way in which it was designed– it is not clear that two individual patient data streams are enough to lead to a generalization of knowledge that would suffice for predicting the behavior of an arbitrary patient. In all likelihood, considerably more patient data would have to be included in the combined data stream.

Luckily, there are 10 different data streams available for the cardiology example presented in Chapter 6 of this dissertation. In that chapter, only one of these data streams had actually been used. It turns out that these trajectories exhibit an impressive (or depressive) variability one from the other, and thus, true knowledge generalization that would enable us to predict the behavior of one patient given his or her own immediate past data and a history data base comprised of the past behaviors of some other patients would be truly remarkable. It is planned to work on this problem in the near future. We hope to be able to use a subset of the 10 data streams available, and show that the behavior of the other data streams, that have not been used for constructing the history data base, can indeed be predicted using the combined model.

Chapter 10

Summary and Future Research

Biomedical engineering is a discipline that addresses medical and biological problems through the use of theories borrowed from the physical sciences, and technologies inherited from engineering.

As a consequence of the imprecision of knowledge available in biomedicine in general, one would expect that qualitative reasoning techniques, as they have been developed during recent decades by researchers working in the area of artificial intelligence, would be ideally suited to tackle problems stemming from biomedical fields. Yet, the application of artificial intelligence to biomedical sciences has not advanced rapidly in the past. Artificial intelligence techniques have been proliferated much more rapidly to and within other areas of science and technology.

Several difficulties inherent to the biomedical fields have restrained the progress in modeling and simulation of this type of systems in the past, problems that make these systems much more difficult to tackle than practically all other types of systems met anywhere in science and engineering.

The aim of the work developed in this dissertation was to address some of these difficulties, of soft sciences in general and of biomedical engineering in particular, and to come up with a methodology that would make optimal use of the limited knowledge available to the modeler, a modeling methodology that would not get confused by the inevitable incompleteness and even inconsistency of information generally available for these types of systems.

To this end, a qualitative modeling and simulation methodology called *Fuzzy Inductive Reasoning (FIR)* has been employed, a modeling technique of fairly recent vintage that looked promising for the task at hand, and for which a

prototypical implementation was available to us. The methodology has been refined, and a second generation of FIR software was implemented that would allow us to work with biomedical and other soft science systems.

The FIR modeling technique can be classified as *pattern–based*, because it works with behavioral knowledge about the external relationships between inputs and outputs of the system under study. Therefore, the FIR methodology is well suited for dealing with biomedical engineering systems, where the structure of the systems to be modeled is usually either totally or at least partially unknown.

Fuzzy qualitative modeling and simulation are, in the context of biomedical research, topics of a rather recent vintage. For this reason, we decided to show and describe not only the positive results that we obtained during this thesis research, but also to explain the difficulties encountered and to delineate the disappointments that we met on our way. For these reasons will the reader find both types of accounts described in here: the very promising results that were discovered in this research effort, the results that truly made this effort rewarding and exciting, but also some brick walls that we encountered and that we were unable to tear down or at least circumvent.

Also negative results *are* results indeed, and they are often even more valuable that the positive ones, since they may prevent future researchers from falling into the same traps that we ventured into, wasting weeks if not months of their precious time, and because they are all too often bashfully concealed by the research community.

10.1 Summary of Results Obtained

The results obtained in this doctoral thesis address several of the problems that are characteristic of biomedical system modeling and simulation. In a first step, a mixed quantitative and qualitative modeling and simulation methodology was developed. Mixed modeling and simulation is needed for all types of systems that are composed of well–known and poorly–understood subsystems. This characteristic is common to a large variety of different applications.

In biomedical engineering, it is frequently necessary to deal with systems with partial structural knowledge. It was therefore essential to be able to have available a mixed quantitative and qualitative modeling and simulation methodology that could be used to analyze such systems.

The validity of the mixed qualitative/quantitative modeling technique has been demonstrated by means of two examples. First, an example from the technical domain, a position control system involving a hydraulic motor with a four–way servo valve, explains the process of mixed quantitative and qualitative simulation using fuzzy inductive reasoning in detail.

Second, a realistic and meaningful biomedical application has been used to demonstrate the validity of this methodology inside the soft science area. This biomedical application represents aspects of the cardiovascular system of the human body. The cardiovascular control system is composed of the hemodynamical system comprising the heart and the blood vessels, and the central nervous control system that is responsible, among other things, for the beating of the heart.

While the functioning of the hemodynamical system is fairly well known, it operates similarly to a mechanical (hydraulic) system and can therefore be modeled quite accurately by means of differential equations, the central nervous controller is still not fully understood, and it was therefore decided to use the FIR methodology to model it in qualitative terms. The results obtained in both examples were excellent, consolidating the mixed methodology as a valuable technique for dealing with systems with partial structural knowledge.

Some of the problems that arise when working with biomedical systems, namely the problems of *poor data availability* and/or *poor data quality*, have been addressed later on in this dissertation. Rather than tackling these problems in a theoretical manner (which may prove to be impossibly difficult due to the complexity of the issues involved), the problems were dealt with by means of a practical example. The system to be modeled was the control of an anaesthetic agent during surgical operation, where a model of the patient and a model of the controller were to be developed. An earlier publication that made use of an artificial neural network for the same purpose was employed as a starting point in order to be able to compare our results against previously established state–of–the–art technology.

Following the approach taken in the earlier publication, we first tried to develop a fuzzy inductive reasoning model for the patient, and then to find a fuzzy inductive reasoning model for the controller.

The FIR methodology was not able to generate a decent model for the patient due to a lack of information from the point of view of data quality. The available data was extremely poor not only in quantity but, even worse, in quality. As goes for all inductive techniques, inductive reasoners need a lot

of data to work with. It is not possible to generate meaningful and reliable inductive models without ample and rich data. This holds equally true for the neural network approach (another inductive modeling technique). However, while a neural network will always predict something, the inductive reasoner will not predict anything that cannot be validated on the basis of the available data. SAPS–II, our inductive reasoner, simply declined to predict anything when confronted with the patient model data, since no prediction can truly be justified given the available facts.

On the other hand, SAPS–II was able to generate a qualitative model with decent predictive power for the anaesthetic controller. For this model, the data quality was sufficient to allow a meaningful prediction.

This is a definite strength of the FIR modeling methodology, since the approach consistently generates a decent inductive model whenever the available data allow it to, and it categorically declines to generate a model if the data at hand do not permit to validate an inductive model. Generating a seemingly good model that is not justifiable on the basis of the available data (as it happened in the case of the neural network approach with the patient model of one of the two available data streams) is maybe even more dangerous than failing to generate a model if the data at hand would support such a model (as it happened in the case of the neural network approach with the controller model of the other data stream), since the modeler will rely on the not properly validated model and trust its predictions in another situation where it may predict pure nonsense.

Along this dissertation, results obtained using neural network and NARMAX approaches for two specific biomedical systems under study have been compared with those obtained using the FIR methodology for the same systems. At least in the examples analyzed in this thesis, the FIR methodology turned out to be *consistently* far superior to both of the other modeling techniques in terms of its capability to capture a maximum of linear and non–linear correlations in input/output behavior. It is considerably more flexible than the NARMAX approach, and does not call for multiple layers as the neural network approach does. The model identification process is very fast, and alternative models (masks) can be investigated quickly to determine their predictive power.

A study of the *limits to predictability* of system behavior using the fuzzy inductive reasoning methodology has also been presented in the thesis. This research was a consequence of personal experiences that have demonstrated limitations to predictability of behavior in biomedical applications, whereas

predictions in technical areas had mostly been accurate far beyond our original expectations.

The concept of a *causality horizon* was introduced. The causality horizon helps determine the likelihood of success of a qualitative prediction. It has been found that the most appropriate mask depth is not only dictated by the slowest and fastest time constants of interest in the system, but is limited also by the causality contained in the input/output behavior, a concept closely related to the cross–correlations used in statistics. The causality horizon was defined as a measure of causality in input/output behavior allowing to assess beforehand the likelihood of success of a qualitative prediction. This idea has been exemplified by means of two systems: a (generic) linear system and a biomedical system.

No meaningful qualitative predictions of system behavior can be obtained unless there exist strong, either positive or negative, correlations between the inputs and the output of that system. Thus, no causality is possible without correlation. Although a strong correlation between signals does not necessarily imply true causality between these signals, it does imply predictability of behavior. In this sense, the FIR modeling methodology (and the same goes for neural networks) is not fundamentally different from statistical modeling techniques. However, both of these inductive approaches to modeling are much easier to apply than statistical approaches to modeling, and both of them exploit non–linear correlations considerably better than at least the classical statistical techniques.

Another inherent problem of biomedical systems is *incomplete information*. This issue has also been studied in this doctoral thesis. The lack of information problem, that has already been treated before, may have several different causes, all of them related to acquisition difficulties. The problems are further amplified when the data records obtained from medical experiments are incomplete. There are lots of circumstances that can produce gaps of information for one, several, or all of the parameters. Qualitative methodologies that cannot deal with missing data values are therefore quite useless for dealing with biomedical applications.

In this dissertation, the FIR methodology has been enhanced in order to be able to tackle incomplete information. To this end, a technique called *missing data option* has been developed that allows to work with incomplete medical data records.

A practical study of the limits to predictability of behavior when incomplete

data records are present in the training data set is discussed in this dissertation. Two different applications have been used to show those limits: a (generic) linear state–space model, and observations of input/output behavior stemming from a biomedical system.

The tests done with both systems show that the limits to predictability are difficult to quantify in a precise fashion. The degradation of the forecasting power depends on the richness and the redundancy of the data records in the data history. It is the lack of training data (previous behavioral experience) that affects the forecasting power and not the presence or absence of data gaps.

Another important problem in biomedical modeling is the inevitable variability in system behavior from one patient to another. This doctoral thesis thus has also focused on the preconditioning of biomedical data to *eliminate patient–specific behavior*. This is a crucial issue that has received some attention in this dissertation. In most biomedical applications, such as, for instance, the control of depth of anaesthesia of a patient undergoing surgery, it is extremely useful to have available a model that identifies not only the behavior of one concrete patient on a specific day during a specific operation, but one that is able to capture the behavior of a class of similar patients undergoing similar operations.

This study presents a knowledge combination technique that allows to merge the knowledge stemming from different patients in order to obtain a general knowledge base. This knowledge base can then be used for the prediction of future states of a new patient with characteristics similar to those of the patients used for obtaining the knowledge base.

The combination technique makes use of the missing data option that allows to merge different patient data streams avoiding the creation of fake causal relationships at the seam between adjacent streams. This technique offers a systematic way to compute predictions for all patients in the patient/operation class.

The efficiency of the advocated technique has been demonstrated by means of the anesthesiology application presented earlier. It became feasible to generate a single qualitative model that could be used to predict the future behavior of two different patients. The predictions were not as good as those obtained from individual models generated separately for each of the two patients, but they were still clinically meaningful.

A quantification of the degradation of the predictive power of the combined model relative to the individual models was also presented. This quantification

demonstrates that the prediction errors made by the combined model were not drastically increased in comparison with those obtained when using the individual models.

It can be remarked that the FIR methodology has two major restrictions relative to the types of systems that can be handled by it. These restrictions are the following:

- It is necessary to have available a sufficiently large amount of rich behavioral data of the system to be modeled. Data–poor systems are not amenable to be modeled using the FIR methodology. This limitation is shared by *all* inductive modeling approaches and is not specific to the FIR methodology.
- The number of variables that the SAPS–II software can reason with simultaneously is limited. This limitation is partly caused by the current implementation of the toolbox (Fortran does not handle variable arrays very well), but it is also dictated by the computational complexity of the algorithms used within the SAPS–II implementation of the FIR methodology. This restriction will be relaxed in a few months, when another Ph.D. dissertation based on the FIR methodology will be completed. That dissertation deals explicitly with the problems inherent in large–scale system modeling, and the modified algorithms presented in that thesis along with an upgrade of the SAPS–II software will allow the user to work with complex system with a fairly large number of variables.

Summarizing, the major contributions of this doctoral thesis are the following:

- A methodology for modeling and simulation of mixed quantitative and qualitative biomedical systems exemplified by means of the cardiovascular system.
- A methodology for qualitative replication of human decision making processes exemplified by means of the anaesthetic agent delivery system.
- A statistical analysis tool for assessing the forecasting power of qualitative models.
- A methodology for dealing with incomplete measurement data sets in qualitative modeling of biomedical (or other) systems.
- A methodology for filtering out patient–specific behavior from separate measurement data sets stemming from different patients.

10.2 Future Research

As has been described in Chapter 5 of this dissertation, biomedical systems pose particular problems that make it difficult to come up with models capturing their dynamical behavior. Several of those difficulties have been analyzed in this dissertation, from the fuzzy inductive reasoning methodology perspective, in the hope to eliminate at least some of them once and for all, while alleviating the consequences of others. However, some problems have either not yet been tackled at all, or they have been discussed but have not been solved in a conclusive fashion. These remain open to future research.

This is the case, for example, of the problem of *diversity in patient behavior*. In this dissertation, a first step has been realized tackling this difficult problem. A data stream merging technique has been developed that is systematic and that, at least conceptually, allows to obtain a model for a given patient/operation class. However, this technique has only been applied to two patient/operation sets. Therefore, the anesthesiology example has not yet demonstrated the capability of the FIR methodology to generalize knowledge. In order to prove the knowledge generalization capability, we would have needed a third data stream stemming from yet another patient, a stream not used in the combined past history data and show that the combined model derived for the first two patients is capable of also predicting the future behavior of the third patient.

It is not clear that this might have worked. It is doubtful that two individual patient data streams are enough to lead to a generalization of knowledge that would suffice for predicting the behavior of an arbitrary patient. In all likelihood, considerably more patient data would have to be included in the combined data stream for this purpose.

Therefore, a deeper study using a biomedical system with much more data stemming from many different patients needs to be performed to truly demonstrate the knowledge generalization capability of the data merging approach. The cardiology example presented in Chapter 6 of this dissertation would lend itself to such an investigation. For this application, there are already available data streams from 16 different patients. In that chapter, only one of these data streams has actually been used. It turns out that these trajectories exhibit a formidable and frightening variability one from the other, and thus, true knowledge generalization that would enable us to predict the behavior of one patient given his or her own immediate past data and a history data base comprised of the merged past behaviors of some other

patients would constitute a sensational breakthrough in machine learning. It is planned to work on this problem in the near future. We hope to be able to use a subset of the 16 data streams available, and show that the behavior of the other data streams, that have not been used for constructing the history data base, can indeed be predicted using the combined model.

Another interesting topic for additional research in the future is the incomplete information issue. In this dissertation, our interest was focused on the ability to work with sets of data where missing values might be present basically due to recording problems. To this end, the idea was to extract as much information as possible from the available data sets, eliminating data records that are contaminated by missing values. In medical systems, such a feature is extremely useful since missing values are utterly common in data records obtained from this type of systems and since it is not acceptable to throw out incomplete data sets and repeat the experiment, as this would be the case in most engineering systems. A sufficiently long and rich data stream obtained from a human patient is far too valuable an asset to being thrown out on the basis of incomplete data records alone.

However, it could also be of interest to focus new research on the idea of reconstructing missing data encountered in the identification data set. Some research has already been developed along these lines (Albridge *et al.*, 1988) that can provide a starting point to center a new research effort using the fuzzy inductive reasoning methodology for extending the available knowledge.

Different data reconstruction strategies have been studied during past years, such as *interpolation*, *extrapolation*, *patient–specific mean value*, and *patient–specific linear regression over time*. General conclusions from these previous studies can be reached. First, several strategies exist for estimating missing data values, yet none of these strategies works uniformly best in all situations. Second, the quality of the strategy as a predictor depends on the characteristics of the variable that is contaminated by the missing value and the goals of the study being conducted. Third, serious limitations exist in the current literature in that a significant portion of this literature assumes data to be "randomly" missing. This is not a meaningful proposition in the context of medical data. If a sensor breaks down, usually more than one data value will be missing in a row. On the basis of these previous conclusions, a deep study of the prediction of missing values in biomedical system applications can be made. Such a study can provide a rigorous comparison of the already existing methods when applied to biomedical systems as well as propose a new method to tackle this problem.

This new method could be based on first extracting the relevant information contained in the identification data set that is contaminated by missing values. This can be accomplished using the missing data option developed in this dissertation, obtaining a reduced history data set. Then, this historical information could be used to try to forecast the missing values that exist in the training data set, replacing the missing elements by the predicted values.

We have seen that predictions of future behavior of a biomedical system are possible whenever there exists sufficient correlation in the input/output behavior of the system. The FIR methodology will exploit this behavior in a quasi optimal fashion. However, whereas correlation implies predictability, it does not necessarily imply causality as well. It will often be the case that there is a large conceptual distance between the inputs and the output of a qualitative model. In such a case, it will not be easy to reduce the qualitative knowledge available in the FIR model, which is perfectly sufficient for purposes of prediction of behavior, to explanations of causal behavior that a practitioner would be able to interpret in terms of variables that he or she can understand.

To this end, it would be useful to subdivide complex systems into simpler subsystems, although the goal of prediction does not require such a subdivision, in order to limit the conceptual gap between the inputs and the output of each subsystem for the purpose of enabling the methodology to provide meaningful explanations of the reasoning process. In this context, a merge between knowledge–based approaches to modeling (decomposition) and pattern–based approaches to modeling (identification) may prove to be fruitful.

The attentive reader may have noticed that the development of FIR models still contains a fair amount of heuristics. Into how many classes should a variable be discretized? In the cardiology example, some of the controller outputs were recoded into four levels, others into five levels, yet others into two levels only. Why was this done? The answer is simple. We tried a number of different alternatives, and selected the one that provided the best compromise between accuracy of forecast and model complexity in each case.

Where precisely are the landmarks to be placed that separate neighboring classes? A quite heuristic algorithm for the selection of landmarks was proposed in Chapter 3 of this dissertation. This algorithm usually works quite well, but sometimes, it didn't lead to good forecasts, and in those cases, intuition was used to come up with a more appropriate set of landmarks. There is a wealth of literature available that deals with the problem of clustering. Evidently, the selection of landmarks is closely related to the problem of clustering as discussed in the artificial intelligence literature.

Although the necessity of take heuristic decisions turned out to be fairly harmless in the examples described in this dissertation, it must be feared that, in examples of much larger size, this may no longer be the case. More research should be made to reduce the number of heuristic decisions to be made by the user of the tool to a minimum.

As this section shows, many interesting problems are still open to be tackled in future research efforts, and we are eager to continue with this line of research ourselves. We are excited about the very promising possibilities of the FIR methodology and about the results that are documented in this thesis. They were, in many cases, far better than we had expected them to be. After we had designed the central nervous system controllers of the cardiovascular system individually, we were quite sceptical when we put everything together whether the mixed simulation would work at all. After all, this was the very first time that anyone had tried to put five separate qualitative FIR controllers into a single system and make them work together. To our own amazement, the closed–loop control worked on the first try.

We hope that, with this thesis, we have provided a significant contribution to the field of biomedical engineering, and that our results will prove to be useful for many other researchers dealing with time–dependent bioengineering systems.

References and Bibliography

- **(Albridge et al., 1988)**; K.M. Albridge, J. Standish, and J.F. Fries (1988), "Hierarchical Time–Oriented Approaches to Missing Data Inference", *Computers and Biomedical Research*, **21**, pp. 349–366.

- **(Adlassnig and Scheithauer, 1989)**; Adlassnig, K., Scheithauer, W. (1989), "Performance Evaluation of Medical Expert Systems Using ROC Curves", *Computers and Biomedical Research*, **22**, pp. 297–313.

- **(Aliev et al., 1992)**; Aliev, R. A., F. T. Aliev, and M. D. Babaev (1992), "The Synthesis of a Fuzzy Coordinate–Parametric Automatic Control System for an Oil–Refinery Unit," *Fuzzy Sets and Systems*, **47**, pp. 157–162.

- **(Babbie, 1989)**; E. Babbie (1989), *The Practice of Social Research - Fifth Edition*, Wadsworth Publishing Company, Belmont, CA.

- **(Bacchus et al., 1991)**; Bacchus, F.; J. Tenenberg and J.A. Koomen (1991), "A Non-Reified Temporal Logic," *Artificial Intelligence*, **52**, pp. 87-108.

- **(Ballard, 1990)**; Ballard, D. (1990), "Conjoint Computing: Integrating Numeric, Symbolic and Neural Computing," TH0308-7/90/0000/0194/$01.00 1990 IEEE, pp. 194-201.

- **(Barnett et al., 1987)**; Barnett, G.O., Cimino, J.J., Hupp, J.A., and Hoffer (1987), "DXplain: an evolving diagnostic decision support system, JAMA, 258(1), July 3, pp. 67-73.

- **(Barletta, 1991)**; Barletta, R. (1991), "An Introduction to Case-Based Reasoning," *AI Expert*, August 1991, pp. 43-49.

- **(Barro et al., 1990a)**; S. Barro, R. Ruiz, and J. Mira (1990), "Fuzzy Beat Labeling for Intelligent Arrhytmia Monitoring", *Computers and Biomedical Research*, **23**, pp. 240–258.

- **(Barro et al., 1990b)**; Barro, S.; R. Ruiz; J. Presedo; D. Cabello and J. Mira (1990), "Arrhytmia Monitoring Using Fuzzy Reasoning," *Proceedings of the IEEE Annual International Conference on Engineering in Medicine and Biology Society*, **12**(2), pp. 823-824.

- **(Beneken and Gravenstein, 1987)**; Beneken, J.E.W., and Gravenstein, J.S. (1987), "Sophisticated Alarms in Patient Monitoring: A Methodology Based on Systems Engineering Concepts", Chapter 20 of the book "The Automated Anesthesia Record and Alarm Systems", Edited by J.S. Gravenstein, et al., Butterworths.

- **(Beneken and Rideout, 1968)**; Beneken, J.E., and Rideout, V.C. (1968), "The Use of Multiple Models in Cardiovascular System Studies: Transport and Perturbation Methods", *IEEE Trans. Biomed. Eng.*, Vol. BME-15, no. 4, pp. 281–289.

- **(Beneken and Thevenin, 1993)**; Beneken J.E.W., and Thevenin V. Editors (1993), "Advances in Biomedical Engineering", IOS Press.

- **(Beneken and Van der Aa, 1989)**; Beneken, J.E.W., and Van der Aa, J.J. (1989), "Alarms and their Limits in Monitoring", *Journal of Clinical Monitoring*, Vol. 5, No. 3, July, pp. 205-210.

- **(Bhaskar and Nigam, 1990)**; Bhaskar R., Nigam A. (1990), "Qualitative Physics Using Dimensional Analysis", *Artificial Intelligence*, **45**, 0004-3702/90, pp. 73-111.

- **(Binaghi et al., 1993)**; E. Binaghi, O. De Giorgi, G. Maggi, T. Motta, and A. Rampini (1993), "Computer-Assisted Diagnosis of Postmenopausal Osteoperosis Using a Fuzzy Expert System Shell", *Computers and Biomedical Research*, **26**, pp. 498–516.

- **(Binik et al., 1988)**; Y.M. Binik, D. Servan-Schreiber, S. Freiwald, and K.S.K. Hall (1988), "Intelligent Computer-based Assessment and Psychoterapy: An Expert System for Sexual Dysfunction," *The Journal of Nervous and Mental Disease,* Vol. 176, No. 7, July 1988, pp.387-400.

- **(Bobrow, 1984)**; Bobrow, D.G., ed. (1984), Special Volume on Qualitative Reasoning About Physical Systems, *Artificial Intelligence*, **24**; also as: *Reasoning about Physical Systems*, MIT Press, Cambridge, MA.

- **(Bonissone and Valavanis, 1985)**; Bonnisone, P.P. and K.P. Valavanis, (1985), "A Comparative Study of Different Approaches to Qualitative Physics Theories", CH2215-2/85/0000/0236$01.00 1985 IEEE, pp. 236-243.

- **(Bratko, 1988)**; Bratko, I. (1988), "Qualitative Modelling for Medical Diagnosis," *Selected Papers from the IFAC Symposium on Modelling and Control in Biomedical Systems*, (C. Cobelli and L. Mariani, eds.), Venice, Italy, April 6-8, Pergamon Press, pp. 539-545.

- **(Bratko et al., 1989)**; Bratko, I.; I. Mozetič and N. Lavrač (1991), *KARDIO: A Study in Deep and Qualitative Knowledge for Expert Systems*, M.I.T. Press, Cambridge, MA.

- **(Bylander et al., 1988)**; Bylander, T., Smith, J.W.,Jr., and Svirbely, J.R. (1988), "Qualitative representation of behavior in the medical domain," *Comput. Biomed. Res.*, 21, pp. 367.

- **(Carson and Deutsch, 1992)**; Carson, E.R. and T. Deutsch (1992), "A Spectrum of Approaches for Controlling Diabetes," *IEEE Control Systems*, **12**(6), pp. 25-31.

- **(Cellier and Mugica, 1992)**; Cellier, F.E. and F. Mugica (1992), "Systematic Design of Fuzzy Controllers Using Inductive Reasoning," *Proceedings ISIC-92, IEEE International Symposium on Intelligent Control*, Glasgow, Scotland, U.K., August 11-13, pp. 198-203.

- **(Cellier and Yandell, 1987)**; Cellier, F.E. and D.W. Yandell (1987), "SAPS-II: A New Implementation of the Systems Approach Problem Solver," *International J. of General Systems*, **13**(4), pp. 307-322.

- **(Cellier, 1987)**; Cellier, F.E. (1987), "Qualitative Simulation of Technical Systems Using the General Systems Problem Solving Framework," *International J. of General Systems*, **13**(4), pp. 333-344.

- **(Cellier, 1991a)**; Cellier, F.E. (1991), *Continuous System Modeling*, Springer-Verlag, New York, 755p.

- **(Cellier, 1991b)**; Cellier, F.E. (1991), "Qualitative Modeling and Simulation – Promise or Illusion," *Proceedings of the SCS 1991 Winter Simulation Conference*, Phoenix, AZ, USA, pp. 1086-1090.

- **(Cellier, 1991c)**; Cellier, F.E. (1991), "General System Problem Solving Paradigm for Qualitative Modeling," *Qualitative Simulation*

Modeling and Analysis, (P.A. Fishwick and P.A. Luker, eds.), Springer-Verlag, New York, pp. 51-71.

- **(Cellier et al., 1992)**; Cellier, F.E.; A. Nebot; F. Mugica and A. de Albornoz (1992), "Combined Qualitative/Quantitative Simulation Models of Continuous-Time Processes Using Fuzzy Inductive Reasoning Techniques," *Proceedings SICICA'92, IFAC Symposium on Intelligent Components and Instruments for Control Applications*, Málaga, Spain, May 22-24, pp. 589-593.

- **(Cellier et al., 1994)**; Cellier, F.E.; Nebot, A.; Mugica, F.; and De Albornoz, A. (1994), "Combined Qualitative/Quantitative Simulation Models of Continuous–Time Processes Using Fuzzy Inductive Reasoning Techniques." Article accepted for publication in *International Journal of General Systems* (to be published during 1994).

- **(Cheung and Hull, 1989)**; Cheung, J.Y. and S.S. Hull (1989), "Detection of Abnormal Electrodiagrams Using a Neural Network Approach," *Proceedings of the IEEE Engineering in Medicine and Biology Society, Eleventh Annual Conference*, November 9-12, pp. 2015-2016.

- **(Ciarrocca, 1989)**; Chiarrocca, S. (1989), "Using a Translation-Invariant Neural Network to Diagnose Heart Arrhytmia," *Proceedings of the IEEE Engineering in Medicine and Biology Society, Eleventh Annual Conference*, November 9-12, pp. 2025-2026.

- **(Cios et al., 1991)**; Cios, K.J.; I. Shin and L.S. Goodenday (1991), "Using Fuzzy Sets to Diagnose Coronary Artery Stenosis," *IEEE Computer*, March, 0018-9162/91/0300-0057$01.00, pp. 57-63.

- **(Clancey and Letsinger, 1984)**; Clancey, W.J. and Letsinger, R. (1984), "NEOMYCIN: reconfiguring a rule-based expert system for application to teaching," *Readings in Medical Artificial Intelligence: The first Decade,* Clancey, W.J. and Shortliffe, E.H., Eds., Addison-Wesley, Reading, MA, pp. 361.

- **(Cochran and Paul, 1990)**; Cochran J.K., Paul B.K. (1990), "QUAL: A Microcomputer System for Qualitative Simulation," *Simulation*, November, pp. 300-308.

- **(Coiera, 1990)**; Coiera, E.W. (1990), "Monitoring diseases with empirical and model-generated histories", *Artificial Intelligence in Medicine,* vol. 2, pp. 135-147.

- **(Cox, 1992)**; Cox E., (1992), "Solving Problems with Fuzzy Logic," *AI Expert*, March 1992, pp. 28-37.

- **(Czogala and Rawlik, 1989)**; Czogala, E. and T. Rawlik (1992), "Modelling of a Fuzzy Controller With Application to the Control of Biological Processes," *Fuzzy Sets and Systems*, **31**(1), pp. 13-22.

- **(de Albornoz and Cellier, 1993a)**; de Albornoz, A., and F. E. Cellier (1993), "Variable Selection and Sensor Fusion in Automated Hierarchical Fault Monitoring of Large Scale Systems," *Proceedings QUARDET'93, Qualitative Reasoning and Decision Technologies*, Barcelona, Spain, pp. 722–734, 1993.

- **(de Albornoz and Cellier, 1993b)**; de Albornoz, A., and F. E. Cellier (1993), "Qualitative Simulation Applied to Reason Inductively About the Behavior of a Quantitatively Simulated Aircraft Model," *Proceedings QUARDET'93, Qualitative Reasoning and Decision Technologies*, Barcelona, Spain, pp. 711–721, 1993.

- **(de Albornoz, 1995)**; de Albornoz, A., *Inductive Reasoning and Reconstruction Analysis as a Tool for Fault Monitoring and Decision Support in Large–Scale Systems*, Ph.D. Thesis (in preparation).

- **(de Kleer and Brown, 1984)**; de Kleer, J. and J.S. Brown (1984), "A Qualitative Physics Based on Confluences," *Artificial Intelligence*, **24**.

- **(Dordan, 1992)**; Dordan, O. (1992), "Mathematical Problems Arising in Qualitative Simulation of a Differential Equation," *Artificial Intelligence*, **55**, pp. 61-86.

- **(Dubois and Prade, 1980)**; Dubois, D. and H. Prade (1980), "Fuzzy Sets and Systems: Theory and Applications," *Mathematics in Science and Engineering*, **144**, Academic Press.

- **(Dvorak et al., 1990)**; Dvorak, D.L., D.T. Dalle Molle, B.J. Kuipers, and T.F. Edgar (1990), "Qualitative simulation for expert systems", *Automatic Control Wordls Congress 1990. 'In the sevice of Mankind'. Proceedings of the 11th Triennial World Congress of the International Federation of Automatic Control,* Jaaksoo, V. and Utkin, V.I. editors, Pergamon, Oxford, Vol. 4, pp. 217-22.

- **(Feng et al., 1991)**; Feng, J.; W.-C. Lin and C.-T. Chen (1991), "Epicardial Boundary Detection Using Fuzzy Reasoning," *IEEE Trans. Medical Imaging*, **10**(2), pp. 187-199.

- **(First et al., 1982)**; First, M.B., Weimer, B.J., McLinden, S., and Miller, R.A. (1982), "LOCALIZE: Computer-assisted localization of peripheral nervous system lesions," *Computers and Biomedical Research,* **15**, pp. 525-543.

- **(Fishwick and Zeigler, 1990)**; Fishwick, P.A. and B.P. Zeigler (1990), "Qualitative Physics: Towards the Automation of Systems Problem Solving," TH0308-7/90/0000/0118$0.100 IEEE, pp.118-134.

- **(Fishwick, 1988a)**; Fishwick, P.A. (1988), "Qualitative Simulation: Fundamental Concepts and Issues," *Proceedings Artificial Intelligence and Simulation 88*, The Society for Computer Simulation, San Diego, CA, pp. 25-31.

- **(Forbus, 1984)**; Forbus, K.D. (1984), "Qualitative Process Theory," *Artificial Intelligence*, **24**, pp. 85-168.

- **(Forsythe et al., 1992)**; D.E. Forsythe, B.G. Buchanan, J.A. Osheroff, and R.A. Miller (1992), "Expanding the Concept of Medical Information: An Observational Study of Physicians' information Needs", *Computers and Biomedical Research*, **25**, pp. 181–200.

- **(Fox and Frost, 1985)**; Fox, J. and D. Frost (1985), "Artificial Intelligence in Primary Care," *Artificial Intelligence in Medicine*, pp. 137-153.

- **(García, 1991)**; García, A. (1991), "Aplicaciones actuales de la lógica borrosa," *Automática e Instrumentación*, Septiembre, No.216, pp. 113-119.

- **(García, 1993)**; García, R. (1993), Personal Communication.

- **(Genesereth and Nilsson, 1987)**; Michael R. Genesereth, and Nils J. Nilsson (1987), *Logical Foundations of Artificial Intelligence,* Morgan Kaufmann Publishers, Palo Alto, Calif.

- **(Hayes, 1979)**; Hayes, P.J. (1979), "The Naive Physics Manifesto," *Expert Systems in the Micro-Electronic Age*, (D. Mitchie, ed.), Edinburgh University Press, Edinburgh, Scotland, UK.

- **(Hayes-Roth et al., 1992)**; Hayes-Roth, B., Washington, R., Ash, D., Hewett, R., Collinot, A., Vina, A., and Seiver, A. (1992), "Guardian: a prototype intelligent agent for intesive-care monitoring," *Artificial Intelligence in Medicine*, vol. 4, pp. 165-185.

- **(Hill and Dolan, 1982)**; Hill, D.W. and A.M. Dolan (1982), *Intensive Care Instrumentation*, Academic Press, London, 423p.

- **(Hobbs and Moore, 1985)**; Hobbs, J.R. and R.C. Moore, eds. (1985), *Formal Theories of the Commonsense World*, Ablex Publishing Corporation, Norwood, New Jersey.

- **(Hunter et al., 1989)**; Hunter, J.; Gotts, N.; Hamlet, I.; Kirby, I. (1989), "Qualitative Spatial and Temporal Reasoning in Cardiac Electrophysiology," *AIME-89: Proc. Second Europ. Conf. on Artificial Intelligence in Medicine*, (Springer-Verlag, Berlin, 1989).

- **(Hunter et al., 1991)**; Hunter, J.; I. Kirby and N. Gotts (1991), "Using Quantitative and Qualitative Constraints in Models of Cardiac Electrophysiology," *Artificial Intelligence in Medicine*, **3**(1), pp. 41-61.

- **(Isaka and Sebald, 1992)**; Isaka, S. and A.V. Sebald (1992), "An Optimization Approach for Fuzzy Controller Design," *IEEE Trans. Systems, Man, and Cybernetics*, **22**(6), pp. 1469-1473.

- **(Kahn et al., 1985)**; Kahn M.G., Fergusson J.C., Shortliffe E.H. and Fagan L.M. (1985), "Representation and use of temporal information in ONCOCIN", *Proc. of the Ninth Annual Symp. on Computer Applications in Medical Care,* Ackerman M.J., ed., IEEE Comput. Soc., Baltimore, MD, pp. 172-176, November 10-13.

- **(Kalagnanam et al., 1991)**; Kalagnanam, J.; Simon, H.A.; Iwasaki, Y. (1991), "The Mathematical Bases for Qualitative Reasoning," *IEEE Expert*, 0885/9000/91/0400-0011$1.00, April 1991, pp. 11-19.

- **(Kandel, 1986)**; Kandel, A. (1986), *Fuzzy Mathematical Techniques With Applications*, Addison Wesley.

- **(Kinney et al., 1988)**; Kinney, E.L., Brafman, D., and Wright II, R.J. (1988), "An Expert System on the Diagnosis of Ascites", *Computers and Biomedical Research*, **21**, pp. 169–173.

- **(Kitzmiller, 1988)**; Kitzmiller, C.T. (1988), "Simulation and AI: Coupling Symbolic and Numeric Computing", *AI and Simulation*, ISBN 0-911801-30-8, pp. 3-7.

- **(Klir and Folger, 1988)**; Klir, G.J. and T.A. Folger (1988), *Fuzzy Sets, Uncertainty and Information*, Prentice-Hall, Englewood Cliffs, N.J.

- **(Klir, 1985)**; Klir, G.J. (1985), *Architecture of Systems Problem Solving*, Plenum Press, New York.

- **(Klir, 1989)**; Klir, G.J. (1989), "Inductive Systems Modelling: An Overview." *Modelling and Simulation Methodology: Knowledge Systems' Paradigms*, (M. S. Elzas, T. I. Ören, and B. P. Zeigler eds.), Elsevier Science Publishers B.V. (North–Holland), Amsterdam, The Netherlands, pp. 55–75.

- **(Korn, 1991)**; G. Korn (1991), *Neural–Network Experiments on Personal Computers*, M.I.T. Press, Cambridge, MA.

- **(Kosko, 1992)**; Kosko, B. (1992), *Neural Networks and Fuzzy Systems – A Dynamical Systems Approach to Machine Intelligence*, Prentice-Hall, Englewood Cliffs, N.J.

- **(Kuipers and Kassirer, 1985)**; Kuipers, B.J. and J.P. Kassirer (1985), *Qualitative Simulation in Medical Physiology: A Progress Report*, Memorandum MIT/LCS/TM-280, Laboratory for Computer Science, M.I.T., Cambridge, MA.

- **(Kuipers, 1986)**; Kuipers, B.J. (1986), "Qualitative Simulation," *Artificial Intelligence*, **29**, pp. 289-338.

- **(Kuipers, 1988)**; Kuipers, B.J. (1988), "Abstraction by Time-Scale in Qualitative Simulation for Biomedical Modeling," *Selected Papers from the IFAC Symposium on Modelling and Control in Biomedical Systems*, (C. Cobelli and L. Mariani, eds.), Venice, Italy, April 6-8, 1988, Pergamon Press, pp. 547-551.

- **(Kuipers, 1989)**; Kuipers, B.J. (1989), "Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge," *Automatica*, **25**(4), Pergamon Press, UK, pp. 571-585.

- **(Kulikowski and Weiss, 1982)**; Kulikowski, C. and S.M. Weiss (1982), "Representation of Expert Knowledge for Consultation: The CASNET and EXPERT Projects," in: *Artificial Intelligence in Medicine*, (P. Scolovits, ed.), Westview Press, Boulder, CO, p. 21-55.

- **(Kuncheva, 1990)**; Kuncheva, L.I. (1990), "Fuzzy Multi-level Classifier for Medical Applications", in *Comput. Biol. Med.*, Vol.20, No. 6, pp. 421-431.

- **(Kunz, 1983)**; Kunz, J.C. (1983), "Analysis of physiological behavior using causal model based on first principles," *Proceedings of National Conference on Artificial Intelligence,* Morgan Kaufmann, San Mateo, CA, 1983,225.

- **(Langlotz et al., 1987)**; Langlotz, C.P., Fagan, L.M., Tu, S.W., Sikic, B.I., and Shortliffe, E.H. (1987), "A therapy planning architecture that combines decision theory and artificial intelligence techniques", *Computers and Biomedical Research*, **20**, pp. 279-303.

- **(Lanzola et al., 1990)**; G. Lanzola, M. Stefanelli, G. Barosi, and L. Magnani (1990), "NEOANEMIA: A Knowledge-BAsed System Emulating Diagnostic Reasoning", *Computers and Biomedical Research*, **23**, pp. 560–582.

- **(Law and Kelton, 1990)**; Law A., and D. Kelton (1990), *Simulation Modeling and Analysis*, 2nd Edition, McGraw–Hill, New York.

- **(Leaning et al., 1983)**; M.S. Leaning, H.E. Pullen, E.R. Carson, and L. Finkelstein (1983),"Modelling a complex biological system: the human cardiovascular system", *Trans. Inst. Meas. Control,* Vol. 5, pp. 71-86.

- **(Leung and Lam, 1988)**; Leung, K.S. and Lam, W. (1988), "Fuzzy Concepts in Expert Syatems," *IEEE Computer*, September, pp. 43-56.

- **(Levy et al., 1989)**; M. Levy, P. Ferrand, and V. Chirat (1989), "SESAM-DIABETE, an expert system for insulin-requiring diabetic patient education," *Comput. Biomed. Res.*, vol. 22, No. 5, pp. 442-453.

- **(Li and Cellier, 1990)**; Li, D. and F.E. Cellier (1990), "Fuzzy Measures in Inductive Reasoning," *Proceedings 1990 Winter Simulation Conference*, New Orleans, LA, pp. 527-538.

- **(Linkens and Hasnain, 1991)**; D.A. Linkens, and S.B. Hasnain (1991), "Self-organising fuzzy logic control and application to muscle relaxant anaesthesia," in *IEE Proceedings-D*, Vol. 138, No. 3, May, pp. 274-284.

- **(Linkens and Mahfouf, 1988)**; D.A. Linkens, and M. Mahfouf (1988), "Fuzzy logic knowledge-based control for muscle relaxant anaesthesia," in *Proc. IFAC Symp. Modelling & Control in Medicine,* Venice, Italy, pp. 123-128.

- **(Linkens and Rehman, 1992a)**; Linkens, D.A. and H.U. Rehman (1992), "Neural Network Controller for Depth of Anaesthesia," *Colloquium on Applications of Neural Networks to Modelling and Control*, Liverpool, U.K.

- **(Linkens and Rehman, 1992b)**; Linkens, D.A. and H.U. Rehman (1992), "Nonlinear Control for Anaesthetic Depth Using Neural Networks and Regression," *Proceedings 1992 IEEE International Symposium on Intelligent Control*, Glasgow, U.K.

- **(Linkens, 1992)**; Linkens, D.A. (1992), "Adaptive and Intelligent Control in Anaesthesia," *IEEE Control Systems*, **12**(6), pp. 6-11.

- **(Linkens, 1993)**; Linkens, D.A. (1993), Personal Communication.

- **(Linkens et al., 1986)**; Linkens, D.A.; D.A. Greenhow and A.J. Asbury (1986), "An Expert System for the Control of Depth of Anaesthesia," *Biomed. Meas. Infor. Contr.*, **1** (4), pp. 223-228.

- **(Linkens et al., 1990)**; Linkens, D.A.; D.A. Greenhow and A.J. Asbury (1990), "Clinical Trials with the Anaesthetic Expert Advisor RESAC," *Proceedings Expert Systems in Medicine*, 6th Annual Meeting, London, pp. 11-18.

- **(Maiers and Sherif, 1985)**; Maiers, J. and Y. S. Sherif (1985), "Applications of Fuzzy Set Theory," *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-15**, pp. 175–186.

- **(Marín and Mira, 1991)**; Marín, R. and J. Mira (1991), "On Knowledge-Based Fuzzy Classifiers: A Medical Case Study," *Fuzzy Sets and Systems*, **44**(3), pp. 421-430.

- **(Marín et al., 1990)**; R. Marín, M.J. Taboada, G. Mouriño, J. Mira, A. Barreiro, R.P. Otero, and M. Pereira (1990), "Representation of Temporal Knowledge in Obstetrics", *Fourth International Symposium on Knowledge Engineering Technical Sessions*, 7-11 May, Barcelona, Spain, pp. 367-372.

- **(McCulloch and Pitts, 1943)**; W.S. McCulloch, and W. Pitts (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, **5**, pp. 115–133.

- **(Meier et al., 1992)**; Meier, R.; J. Nieuwland; A.M. Zbinden and S.S. Hacısalihzade (1992), "Fuzzy Logic Control of Blood Pressure During Anaesthesia," *IEEE Control Systems*, **12**(6), pp. 12-17.

- **(MGA, 1986)**; MGA (1986), *ACSL: Advanced Continuous Simulation Language - User Guide and Reference Manual*, Mitchell & Gauthier Assoc., Concord, MA.

- **(Miller et al., 1986)**; Miller, R.A.; M.A. McNeil; S.M. Chalinor; F.E. Masarie, Jr. and J.D. Myers (1986), "The INTERNIST-1 / QUICK MEDICAL REFERENCE Project - Status Report," *The Western Journal of Medicine*, **145**(6), December, pp. 816-822.

- **(Mira et al., 1991)**; Mira, J.; A. Yáñez; A. Barreiro; A.E. Delgado and J.M. Couselo (1991), "Fuzzy Specification in Data Capture and Knowledge Representation for an Expert System in Oncology," *Fuzzy Sets and Systems*, **44**(3), pp. 431-448.

- **(Mozetič, 1990)**; Mozetič, I. (1990), "Diagnostic Efficiency of Deep and Surface Knowledge in KARDIO," *Artificial Intelligence in Medicine*, **2**(2), pp. 67-83.

- **(Mugica and Cellier, 1993)**; F. Mugica, and F.E. Cellier (1993), "A New Fuzzy Inferencing Method for Inductive Reasoning," *Sixth International Symposium on Artificial Intelligence, Intelligent Systems in Industry and Business*, Monterrey, México.

- **(Mugica, 1995)**; F. Mugica (1995), *Diseño Sistemático de Controles Difusos Mediante Razonamiento Inductivo Difuso*, Ph.D. Thesis (in preparation).

- **(Navabi et al., 1991)**; Navabi, M.J.; R.C. Watt; S.R. Hameroff and K.C. Mylrea (1991), "Integrated Monitoring Can Detect Critical Events and Improve Alarm Accuracy," *J. of Clinical Engineering*, **16**(4), pp. 295-306.

- **(Nebot and Cellier, 1994a)**; Nebot, A.; and Cellier, F.E. (1994), "Preconditioning of Measurement Data for the Elimination of Patient-Specific Behavior in Qualitative Modeling of Medical Systems," in: *Proceedings CISS'94, First Joint Conference of International Simulation Societies*, Zurich, Switzerland, August 22–25, pp. 584–588.

- **(Nebot and Cellier, 1994b)**; Nebot, A.; and Cellier, F.E. (1994), " Dealing With Incomplete Data Records in Qualitative Modeling and Simulation of Biomedical Systems," in: *Proceedings CISS'94, First Joint Conference of International Simulation Societies*, Zurich, Switzerland, August 22–25, pp. 605–610.

- **(Nebot, 1994)**; Nebot, A. (1994), **IC-DT-9404**; Informe para la Suficiéncia Investigadora. Departamento de LSI.

- **(Nebot et al., 1993a)**; Nebot, A.; Cellier, F.E.; and Linkens, D.D. (1993), "Controlling an Anaesthetic Agent by Means of Fuzzy Inductive Reasoning," in: *Proceedings QUARDET'93, IMACS International Workshop on Qualitative Reasoning and Decision Technologies*, Barcelona, Spain, June 16–18, pp. 345–356.

- **(Nebot et al., 1993b)**; Nebot, A.; Cellier, F.E.; and Linkens, D.D. (1993), "Controlling an Anaesthetic Agent by Means of Fuzzy Inductive Reasoning." Article to be submitted shortly to *Artificial Intelligence in Medicine.*

- **(Nebot et al., 1994)**; Nebot, A.; Medina. S.; and Cellier, F.E. (1994), "The Causality Horizon: Limitations to Predictability of Behavior Using Fuzzy Inductive Reasoning," in: *Proceedings ESM'94, European Simulation Multiconference*, Barcelona, Spain, June 1–3, pp. 492–496.

- **(Orr and Westenskow, 1989)**; Orr, J.A. and Westenskow, D.R. (1989), "A breathing circuit alarm system based on neural networks," *Anesthesiology,* 71:A339.

- **(Patil, 1982)**, Patil, R.S., Szolovits, P., and Schwartz, W.B. (1982), "Modeling knowledge of the patient in acid-base and electrolyte disorders," *Artificial Intelligence in Medicine*, Szolovits, P., Ed., Westview Press, Boulder, CO, pp. 191-226.

- **(Pedrycz, 1993)**; Pedrycz, W. (1993), *Fuzzy Control and Fuzzy Systems*, Research Studies Press LTD., England, 350p.

- **(Pedrycz et al. 1991)**; Pedrycz, W.; G. Bortolan and R. Degani (1991), "Classification of Electrocardiographic Signals: A Fuzzy Pattern Matching Approach," *Artificial Intelligence in Medicine*, **3**(4), pp. 211-226.

- **(Peng et al. 1991)**; Peng, X.T.; Kandel A. and Wang P. (1991), "Concepts, Rules, and Fuzzy Reasoning: A Factor Space Approach," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.21, No.1, January/February 1991, pp. 194-204.

- **(Poblet, 1988)**; Poblet, J.M., ed. (1988), *Introducción a la Bioengeniería*, Marcombo, Barcelona, 316p.

- **(Pople, 1982)**; Pople, H.E., Jr. (1982), "Heuristic methods for imposing structure on ill-structured problems: the structuring of medical diagnosis," *Artificial Intelligence in Medicine*, Szolovits, P., Ed., Westview Press, Boulder CO, pp. 119-190.

- **(Prokosch et al., 1989)**; Prokosch, H.U., Seuchter, S.A., Thompson, E.A., and Skolnick, M.H. (1989), "Applying Expert System Techniques to Human Genetics", *Computers and Biomedical Research*, **22**, pp. 234–247.

- **(Rehman et al., 1993)**; Rehman, H.U.; D.A. Linkens and A.J. Asbury (1993), "Dynamic Nonlinear Modelling and Control of Depth of Anaesthesia for Spontaneously Breathing and Ventilated Patients," *Computer Methods and Techniques in Biomedicine*, U.K.

- **(Rucker et al., 1990)**; Rucker, D.W., Maron, D.J., and Shortliffe E.H. (1990), "Temporal Representation of Clinical Algorithms Using Expert-System and Database Tools", *Computers and Biomedical Research*, **23**, pp. 222–239.

- **(Rumelhart et al., 1986)**; D.E. Rumelhart, G.E. Hinton, and R.J. Williams (1986), "Learning Internal Representations by Error Propagation", in: *Parallel Distributed Processing: Explorations in the Microstructure of Cognitions, Vol.1: Foundations,* (D.E. Rumelhart, and J.L. McClelland, eds.), M.I.T. Press, Cambridge, MA, pp. 318–362.

- **(Rushmer, 1972)**; Rushmer, R.F. (1972), *Medical Engineering: Projections for Health Care Delivery*, Academic Press, New York, pp. 391.

- **(Rutledge et al., 1989)**; Rutledge, G., Thomsen, G., Beinlich, I., Farr, B., Sheiner, L., and Fagan, L.M. (1989), "Combining qualitative and quantitative computation in a ventilator therapy planner," in *Proc. 13th Annual Symposium on Computer APPlications in Medical Care*, IEEE Computer Society Press, Washington, D.C., November 5-8, pp. 145-151.

- **(Sacks, 1990)**; Sacks, E. (1990), "A Dynamic Systems Perspective on Qualitative Simulation," *Artificial Intelligence*, **42**, 0004-3702/90/$3.50, pp. 349-362

- **(Sagawa, 1973)**; Sagawa, K. (1973), "Advances in Biomedical Engineering", vol. 3, ed. J.H.V. Brown, J.F. Dickson, A.P. Academic.

- **(Sanchez, 1986)**; Sanchez, E. (1986), "Medical Applications with Fuzzy Sets", in *Fuzzy Sets Theory and Applications*, A. Jones et al. (eds.), pp. 331-347.

- **(Sarjoughian, 1994)**; Sarjoughian, H. (1994), *An Approach to Inductive Modeling Based on Systems Theory and Non-monotonic Reasoning*, Ph.D. Thesis.

- **(Schaltenbrand et al., 1993)**; N. Schaltenbrand, R. Lengelle, and J. Macher (1993), "Neural Network Model: Application to Automatic Analysis of Human Sleep", *Computers and Biomedical Research*, **26**, pp. 157–171.

- **(Schwartz, 1991)**; Schwartz, T. (1991), "Fuzzy Tools for Expert Systems," *AI Expert*, February, pp. 34-41.

- **(SCT, 1985)**; SCT (1985), *CTRL-C, A Language for the Computer-Aided Design of Multivariable Control Systems, User's Guide*, Systems Control Technology, Inc., Palo Alto, Calif.

- **(Shafer, 1976)**; Shafer, G. (1976), *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N.J.

- **(Shannon, 1948)**; Shannon, C.E. (1948), "The Mathematical Theory of Communication," *The Bell Syst. Tech. J.*, **27**, pp. 379-423 and pp. 623-656.

- **(Shortliffe, 1976)**; E. H. Shortliffe (1976), *Computer-Based Medical Consultations: MYCIN*, Elsevier, New York.

- **(Sierra, 1989)**; Sierra Garcia, C.A. (1989), "MILORD: Arquitectura multi-nivell per a sistemes experts en classificació", Tesis Doctoral UPC, Dept. LSI, Director: López de Mántaras.

- **(Snyder and Rideout, 1969)**; M.F. Snyder, and V.C. Rideout (1969), "Computer Simulation Studies of the Venous Circulation," *IEEE Trans. Biomed. Eng.*, Vol. 17, pp. 207–211.

- **(Soo et al., 1985)**; Soo, V.-W.; C.A. Kulikowski and D. Garfinkel (1985), "Qualitative Modeling and Clinical Justification for Experimental Design," *Artificial Intelligence in Medicine*, pp. 21-36.

- **(Stefanelli et al., 1988)**; Stefanelli, M.; G. Lanzola; G. Barosi and L. Magnani (1988), "Modelling of Diagnostic Reasoning," *Selected Papers from the IFAC Symposium on Modelling and Control in Biomedical*

Systems, (C. Cobelli and L. Mariani, eds.), Venice, Italy, April 6-8, Pergamon Press, pp. 553-564.

- **(Suga et al., 1974)**; H. Suga, K. Sagawa (1974), "Instantaneous pressure-volume relationships and their ratio in the excised, supported canine left ventricle", *Cir. Res.*, Vol. 53, pp. 117-126.

- **(Szolovits)**; Szolovits, P., "Types of Knowledge as Bases for Reasoning in Medical AI Programs," p. 209.

- **(Todd, 1988)**; Todd, S. (1988), "Qualitative Measurement Interpretation Using Quantity Spaces," *Proceedings Fourth Conference on Artificial Intelligence Applications*, IEEE Computer Society Press, Washington, D.C., pp. 379.

- **(Tu and Guerriere, 1993)**; Tu, J.V., and Guerriere, M.R.J. (1993), "Use of a Neural Network as a Predictive Instrument for Length of Stay in the Intensive Care Unit Following Cardiac Surgery", *Computers and Biomedical Research*, **26**, pp. 220–229.

- **(Uckun and Dawant, 1992)**; Uckun, S. and B.M. Dawant (1992), "Qualitative Modeling as a Paradigm for Diagnosis and Prediction in Critical Care Environments," *Artificial Intelligence in Medicine*, **4**(2), pp. 127-144.

- **(Uckun, 1992)**; Uckun, S. (1992), "Model-Based Reasoning in Biomedicine," *Critical Reviews in Biomedical Engineering*, **19**(4), pp. 261-292.

- **(Uckun et al., 1993)**; Uckun, S.; B.M. Dawant and D.P. Lindstrom (1993), "Model-Based Diagnosis in Intensive Care Monitoring: The YAQ Approach," *Artificial Intelligence in Medicine*, **5**(1), pp. 31-48.

- **(Ursino et al., 1992)**; Ursino, M.; G. Avanzolini and P. Barbini (1992), "Qualitative Simulation of Dynamic Physiological Models Using the KEE Environment," *Artificial Intelligence in Medicine*, **4**(1), pp. 53-73.

- **(Uyttenhove, 1979)**; Uyttenhove, H.J. (1979), *SAPS - System Approach Problem Solver*, Ph.D. Dissertation, (G.J. Klir, Adv.), SUNY Binghamton, USA.

- **(Vallverdú ,1993)**; Vallverdú, M. (1993), *Modelado y Simulación del Sistema de Control Cardiovascular en Pacientes con Lesiones Coronarias*, Doctoral Thesis, Universidad Politécnica de Cataluña, August 1993.

- **(Van Der Rhee et al., 1990)**; Van Der Rhee, F.; Van Nauta Lemke, H.R.; Dijkman, J.G. (1990), "Knowledge Based Fuzzy Control of Systems," *IEEE Transactions on Automatic Control*, Vol.35, No.2, February 1990, 0018-9286/90/0200-0148$01.00, pp. 148-155.

- **(Verdaguer et al., 1992)**; Verdaguer, A., Patak, A., Sancho, J.J., Sierra, C., and Sanz, F. (1992), "Validation of the Medical Expert System PNEUMON-IA", *Computers and Biomedical Research*, **25**, pp. 511–526.

- **(Vesanterä and Cellier, 1989)**; Vesanterä, P.J. and F.E. Cellier (1989), "Building Intelligence into an Autopilot – Using Qualitative Simulation to Support Global Decision Making," *Simulation*, **52**(3), pp. 111-121.

- **(Wang and Klir, 1992)**; Wang, Z., and G.J. Klir (1992), *Fuzzy Measure Theory*, Plenum Publishing Corporation, New York.

- **(Weinberg et al., 1989)**; Weinberg, J.B.; G. Biswas and L.A. Weinberg (1989), "Adventures in Qualitative Modeling – A Qualitative Heart Model," *Proceedings IEEE International Conference on Systems, Man and Cybernetics*, Cambridge, MA, November 14-17, Vol.3, pp. 1003-1008.

- **(Weintraub and Bylander, 1989)**; Weintraub, M.A. and Bylander, T. (1989), "QUAWDS: a composite diagnostic system for gait analysis," *Proc. 13th Annual Symposium on Computer Applications in Medical Care*, IEEE Computer Science Press, Washington, D.C., November 5-8, pp. 145-151.

- **(Ying et al., 1990)**; Ying, H.; D.M. Tucker; M. McEachern; D. Eddleman; M. Perl; A.W.H. Stanley and L.C. Sheppard (1990), "Nonlinear Fuzzy Control of Mean Arterial Pressure in Critically Ill Patients," *Proceedings of the IEEE Annual International Conference on Engineering in Medicine and Biology Society*, **12**(3), pp. 936-937.

- **(Zadeh, 1986)**; Zadeh, L.A. (1986), "A Simple View of the Dempster-Shafer Theory of Evidence and its Implication for the Rule of Combination," *The AI Magazine*, Summer Issue, pp. 85-90.

- **(Zadeh, 1992)**; Zadeh, L.A. (1992), "The Calculus of Fuzzy If/Then Rules," *AI Expert*, March, pp. 23-27.

- **(Zeigler, 1976)**; Zeigler, B.P. (1976), *Theory of Modelling and Simulation*, John Wiley & Sons, New York.

- **(Zimmermann et al., 1984)**; Zimmermann, H.J.; L.A. Zadeh and B.R. Gaines (1984), *Fuzzy Sets and Decision Analysis*, North-Holland, Amsterdam, the Netherlands.