NOTE TO USERS

The original manuscript received by UMI contains pages with indistinct and slanted print. Pages were microfilmed as received.

This reproduction is the best copy available

UMI

THE UNIVERSITY OF CALGARY

A Hybrid Approach Applied to an Industrial Distillation Column that Compares Physical and Neural Network Modeling Techniques

by

Amish Sabharwal

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

DEPARTMENT OF CHEMICAL ENGINEERING

CALGARY, ALBERTA

OCTOBER, 1997

Abstract

A novel approach to hybrid modeling of an industrial Xylene distillation column that uses physical and neural network based techniques is proposed in this thesis.

The physically based dynamic model is developed using a commercial simulator, HYSYS™ (v.1.1). The development of the physical model involves rigorous thermodynamics, and a first principle representation of the column's equipment. An iterative procedure is used to tune the dynamic model that is initialized using the steady-state model and benchmarked using plant data. The physical model captures the column's dynamic changes and benchmarks and upper limit values, based on a performance index (*I.sim*), are recommended to evaluate the accuracy of the physical model.

Artificial neural network models using Process Insights™ (v.3.2) are developed for the column. The optimum topologies for the ANN models are selected using an iterative procedure that involves a number of quantitative selection criteria. The ANN models are identified using transient plant data (ANN.t models), nominal data (ANN.n models). Performance indices (*I.ann*) that represent the prediction accuracy of an ANN are developed independent of the *I.sim* values developed for the physical model. Comparisons between the physical model and ANN models shows that the ANN.t models predicted the plant dynamic changes more accurately than the physical model and ANN.n models. The lack of excitation in the training data of the ANN.n models proves to be significant in the model's inability to predict the plant's transient response. However, using the benchmarked physical model the plant's transient changes are simulated. A hybrid approach is incorporated that involves re-training the ANN using both simulated and nominal data (ANN.n+s models).

The ANN.n+s models prediction of plant transient data are compared to the ANN.t and ANN.n models predictions and the result verifies that the hybrid strategy proved to be a very accurate modeling strategy. The hybrid modeling technique is applied successfully to the prediction of the column's condenser level.

Acknowledgments

I would like to thank Dr. Bill Svrcek and Dr. Dale Seborg for their guidance and support throughout this project. Without their insight, I would not have accomplished the goals of this research. I would also like to thank Hyprotech Ltd. and Pavilion Technologies for providing the necessary software to help engineer this research. Furthermore, I'd especially like to thank NKD of Japan for their financial support.

To all my peers at both the University of California, Santa Barbara and at the University of Calgary, thanks for all your help and encouragement. I especially would like to acknowledge Tony Swanda for his efforts in collecting the data used in this thesis from the Mizushima Oil Refinery. To all of my colleagues at both Universities, I will always remember all the good times we shared and I wish you all the best of luck in your future endeavors.

To all my friends who kept asking, "when are you going to be done?" Well I am done! I probably would have finished it sooner, but you guys kept nagging me. Special thanks to Christine Coutts for being more than just my best friend, but for being my crutch to help me stand through all the tough times.

Last but definitely not least, to my family. Thanks Mom, Dad, and Aashish for your constant love, support and encouragement.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the chemical process industries, distillation is the most widely used separation technique when high purity components are required. To comply with stringent quality requirements, it is estimated that distillation columns consume 3 % of the total energy usage in the US alone (Humphrey et al., 1991). Distillation control can be onerous because of the following well-known process characteristics:

1. Inherent non-linearity of columns due to production of high purity products
2. Severe control loop interaction for dual composition control
3. Large upsets in feed flow rate and feed composition disturbances causing variability in valued products
4. Infrequent measurement of product quality.

These attributes have sparked an increasing amount of interest in developing suitable control methodologies. One of the most recently studied control strategies is the use of model predictive control (MPC). Model predictive control is a technique, which minimize future output deviations from the desired setpoint, while taking into account for the sequence of control actions required to achieve a cost function objective. This method may seem alluring, but there has not been an overwhelming acceptance by the chemical process industry. The main concern is that its success depends substantially on the quality of the process model used. Traditional approaches in developing a process model assume a linear approximation approach such as step input response models. Algorithms such as Dynamic Matrix Control (DMC), where a linear description of the process is assumed, have shown major cost benefits on many industrial systems (Qin and Bodgwell, 1997). However, for distillation where the non-linear behavior (changes in

input variables do not produce proportional changes in the output variable) is an inherent characteristic and a linear model based control algorithm can lead to unacceptable performance. Obtaining a suitable non-linear dynamic model is often the key part of the effort necessary to implement MPC. Examples of non-linear models are models based on engineering first principles or artificial neural networks (ANN).

First principle modeling is a mechanistic approach that involves representing the physical behavior of an operating process by non-linear equations. Physical modeling methods tend to be computationally demanding and require simplifying assumptions that limit their accuracy. Furthermore, these models need to be validated with plant data, which is not a trivial task. Some of these restrictions are limited to the use of physical models to represent only steady-state behavior of chemical processes. However, recent advances in software and computational speed make it possible to develop dynamic models using the same knowledge base, the underlying thermodynamics and physics of the steady-state process models. An example of a commercial simulator that combines steady-state and dynamic simulation in an integrated environment is known as HYSYS^{TM} from Hyprotech Ltd.

Dynamic simulators are applied to problems such as design, training, and operations (Fisher et al, 1985; Marquardt, 1991; Longwell, 1993; Vogel, 1991 Tyreus and Mahoney (1994) provide a good overview. In general, most research studies focus on using dynamic simulators to facilitate a better understanding of process dynamic behavior (Dolph, 1995). Other researchers employ dynamic simulations to obtain process data that provide insight into the interaction between inputs and outputs of a process for other empirical techniques (Ramchandran and Rhinehart, 1994). Dynamic simulators for physical modeling of distillation columns employ a rigorous stage by stage model that solve component mass and energy balances, liquid flow dynamics, and pressure dynamics on each stage (Luyben, 1992). There are some disadvantages to developing such process models. Physical models that try to account for certain phenomenological details tend to become rigorous and computationally intensive. Hence, few dynamic

simulations of distillation columns validated against industrial data are available in the literature; which is an issue that is addressed in this thesis.

An alternative approach to physical modeling is the use of empirical (based on experimental data) models, which rely on process data to develop a relationship between process inputs and outputs. Artificial Neural Networks (ANN) show great promise for such modeling tasks. In contrast to conventional statistical techniques, ANN's use no explicitly specified "knowledge" or procedure to analyze the process data. They are simply "trained" from a representation set of data. In general ANN's require less prior knowledge of the process than physical modeling methods. ANN's consist of interconnected non-linear activation functions that learn by manipulating its internal parameters to predict a set of outputs based on specified inputs. ANN's have the advantage of high accuracy and the ability to adapt to changing process conditions. The fast execution times of these models make them ideal for real-time process optimization. Their universal approximation capabilities (Hornik et. al., 1989) make ANN good candidates for modeling nonlinear chemical process. The major limitation of these models is their prediction accuracy is only as good as the data used to train the models (Baratti et al., 1995). Therefore, a model developed with process data that do not span a wide operating region may have limited utility.

Bhat et al. (1990) explains the use of ANN's in the chemical process industry and provide three case studies: a steady state reactor, a dynamic pH stirred tank, and interpretation of biosensor data. MacMurray and Himmelblau (1995) provide an industrial application of ANN modeling to a packed distillation tower. For a good overview of applications and methodologies of ANN for process modeling and control of chemical processes refer to Morris et al.(1994); Astrom and McAvoy (1992); Thibault and Grandjean (1991). Bhat and McAvoy (1990) applied ANN's to model the dynamic response of pH in a CSTR. Willis et al. (1990) discusses the applicability of ANN's in process control of chemical processes, specifically, ANN's are successfully applied to distillation columns. ANN's are applied to the identification of a packed distillation

column where the ANN model is used in an MPC structure (MacMurray and Himmelblau, 1995) and for on-line monitoring of industrial columns (Baratti et al. (1995). ANN's are also used to control a distillation column in a multivariable MPC framework using a dynamic simulation model (Willis et al, 1990) and pilot plants (MacMurray and Himmelblau, 1995). ANN model development of both high purity industrial columns and pilot columns (e.g. Willis et al. 1992) is limited to the use of steady state simulation for developing training data sets (Ramchandran and Rhinehart, 1995). However, recently there are cases where dynamic simulators are used (e.g. Basualdo and Ceccatto (1995), Willis et al. (1990), Munsif (1997)). The motivation for studying ANN models for distillation columns stems from the fact that most of the research focuses on identifying models from steady state simulation case studies. There has been little work on identification of dynamic ANN using actual plant operating data of distillation columns; an issue which is addressed in this thesis. Furthermore, there are many readily available software companies that are successfully commercializing the ANN technique such as Process Insights from Pavilion Technologies Inc.

With the increasing number of ANN applications in chemical process industries, engineers have been trying to create a suitable working environment for neural networks and physical models. Inadvertently, the two modeling methods are indirectly competing with one another in a very lucrative market. Therefore, an alternative approach to modeling that combines neural networks and first principle modeling, known as hybrid modeling, would indeed attract the attention of many potential users and vendors such as Hyprotech and Pavilion Technologies. In the literature there are few reported uses of hybrid modeling. For example, artificial neural nets are simply used to empirically model a simulation or mechanistic model, to perform a data analysis to improve physical models, or sometimes to put combinations of models (empirical models in series or in parallel with physical models) together for optimization purposes (Martin and Bhat, 1997). Psichogios and Ungar (1991) use a hybrid model to examine a fed batch bioreactor. Pulley et al. (1996) investigated using a neural network trained on process measurements and data derived from a non-linear mechanistic model as inputs to predict

polymer product quality. Pottemann and Seborg (1995) proposed a strategy for forcing ANN models to agree with know steady-state relationships or data. However, these examples are proprietary and there are very few publications of these types of hybrid techniques.

In general, there is a lack of published hybrid modeling case studies because artificial neural networks and physical models are quite different. Physical models represent engineering knowledge in the form of material and energy balance equations, whereas ANN's do not require *a priori* knowledge of the system and are based on a simple architecture that is able to "generalize" predictions if the ANN is well trained. Hence, ANN's do require knowledge in the form of voluminous excited data whereas physical models are fitted to a design condition and are used to extrapolate beyond regions they are intended to model because they capture the physics of the actual process. However, ANN's knowledge of the process is limited to data and cannot extrapolate to regions beyond the range of data they were intended to model. Hence, the two technologies are indeed different but are viewed as complementary because the physical model can be used to provide knowledge in the form of training data to extrapolate the ANN model beyond normal operating regions.

The published studies mentioned above have one or more of the following limitations:

1. Very little published comparisons of dynamic simulation models to actual plant data.
2. Most of the identification data to build a neural network model of a distillation column is limited to industrial steady state or dynamic simulation data.
3. There are very few comparisons of artificial neural network and physical models validated with actual plant operating data.
4. The reported examples of hybrid modeling approaches consist of using artificial neural networks to make models of only computer simulation data (ANN models of physical models). There is little application of physical modeling to provide

additional stimulated data for the development of ANN models from industrial data.

The focus of this research is to address these issues using an industrial case study involving a distillation column, which is part of a refinery benzene-toluene-xylene (BTX) fractionation train. The column is a xylene splitter that separates C_8 from C_{9+} components and is known as TW252. TW252 is located in the Japan Energy Corporation Refinery in Mizushima, Japan. Operating data indicative of various operating periods are collected from the distributed control system (DCS) for nominal, step response, and feed composition disturbance conditions.

The research objectives are to:

1. Build and validate a first principle model of TW252 using plant step response data.
2. Build and validate an artificial neural network model using plant step response data and nominal operating data.
3. Compare the two modeling approaches using a performance index that determines which technique is more accurate in predicting actual plant data.
4. Combine the simulated data that represents a wider range of operating conditions, such as step response and disturbance data obtained from the physical model, with nominal plant steady-state data to train ANN models to interpolate and extrapolate with greater accuracy beyond normal operating regions.

The remainder of the thesis is organized such that each chapter addresses a specific part of the objectives outlined above. Chapter 2 describes in detail the industrial distillation column studied and actual process operating data collected. The development of a physically based dynamic model based on first principles using a commercial simulation package, HYSYS^{TM}, is described in Chapter 3. In Chapter 4, artificial neural network models are developed for TW252 internal tray temperatures, distillate compositions, levels, and the plant inverse from the actual plant operating data using a

commercial ANN software package, Process Insights™, from Pavilion Technologies. Furthermore a comparison is made to the HYSYS™ model predictions. The performance of these two modeling techniques are compared and analyzed, setting the foundation for the combined modeling strategy introduced in Chapter 5. An application of an integrated model that uses HYSYS™ to generate supplemental simulated data to broaden the range of applicability of a nominally trained ANN model using Process Insights™ is presented in Chapter 5. The thesis concludes with a summary and suggestions for future research in Chapter 6.

Chapter 2

Process Description and Data Acquisition

A basic understanding of the industrial process is required in order to develop reliable high fidelity physical and neural network models that are accurate representations of the process. In particular, physically based models require the underlying knowledge of the inherent dynamic characteristics, the actual equipment specifications and thermo-physical properties to ensure that the chemistry and physics of the process are modeled correctly. On the other hand, ANN's require little understanding of the process but need the information from the process in the form of "excited" data. Excited data usually contains information on the dynamics of the process, therefore the amount of dynamic information in the identification of an ANN is very important. In this chapter, the industrial distillation column and the plant data acquired from various plant tests and nominal operation are described in detail.

2.1 Xylene Splitter

The No. 2 xylene splitter (tower TW252 is the middle distillation column in the benzene-toluene-xylene (BTX) fractionation unit at the Japan Energy Corporation's Mizushima Oil Refinery. The feed to TW252 is the bottom product of an upstream tower (TW251). Therefore, the operation of this column depends on the upstream conditions. The TW252 feed enters as a saturated liquid at a design temperature of around 160 °C and contains four chemical components as indicated in Table A.1 in Appendix A. The xylene splitter essentially separates the feedstock made up of traces of ethyl benzene, meta (m), para (p), ortho (o) xylene, and C_{9+} components into a high purity m-p-o xylene

distillate and a C_{9+} component bottom product. An important product specification is a C_{9+} composition in the distillate of less than the specification of 1 wt %. Hence, the column employs a one-point composition control strategy to monitor the distillate C_{9+} contaminant and maintain it below 1 wt. %. The column, TW252, has a diameter of 3.25 m with 34 trays, a horizontal thermosiphon reboiler, and overall condenser that subcools the reflux. Note, the trays are numbered from the bottom to the top. The equipment specifications for the reboiler and condenser and total column volume were determined from the internal material flows and are described in Chapter 3. The main process variables and the nominal steady-state conditions for TW252 are listed in Table 2.1. The process flow diagram (PFD) is given in Figure 2.1.

Table 2.1 Key process variables for TW252

Tag Name	Tag Description	Units	Nominal 1 N.1	Nominal 2 N.2	Nominal 3 N.3
F284	Feed Flowrate	m^3/h	39.16	34.33	39.73
F256	Reflux Flow Rate	m^3/h	43.79	35.40	40.51
F267	Steam Flow Rate	ton/h	11.37	9.87	10.98
F271	Bottoms Flow Rate	m^3/h	14.90	12.91	11.91
F257	Distillate Flow Rate (rotometer)	m^3/h	24.75	21.91	23.01
Y280	Feed Temperature	°C	161.42	160.83	160.09
Y297	Tray 17 Temperature	°C	136.09	139.47	140.81
Y255	Tray 27 Temperature	°C	131.55	135.77	137.02
Y282	Top tray Temperature	°C	128.65	133.57	134.71
Y256	Exit Condenser Temperature	°C	49.48	40.03	45.53
Y257	Tray 1 Temperature	°C	163.30	166.92	182.83
Y284	Bottoms Temperature	°C	169.01	172.54	173.58
L253	Bottoms Level	%	55.01	60.01	59.97
L254	Condenser Level	%	58.03	55.00	54.98
P252	Condenser Pressure	mmHg	550.55	629.99	650.00
A254	C_{9+} Concentration in Distillate	wt. %	0.78	0.70	0.59
A255	Ethyl Benzene in Distillate	wt. %	15.87	16.05	14.89
A256	M/P-Xylene in Distillate	wt. %	58.75	57.67	58.01
A257	O-Xylene in Distillate	wt. %	25.88	25.46	26.38
A264	C_{9+} Concentration in Feed	wt. %	36.87	34.00	33.94
A265	Ethyl Benzene Concentration Feed	wt. %	10.19	10.30	9.57
A266	M/P-Xylene Concentration Feed	wt. %	38.84	37.87	38.16
A267	O-Xylene Concentration Feed	wt. %	18.36	17.84	18.34

2.2 Control Strategy

The column is controlled by several single input and single output (SISO) temperature and pressure control loops as shown on the simplified process flow diagram (PFD) in Figure 2.1.

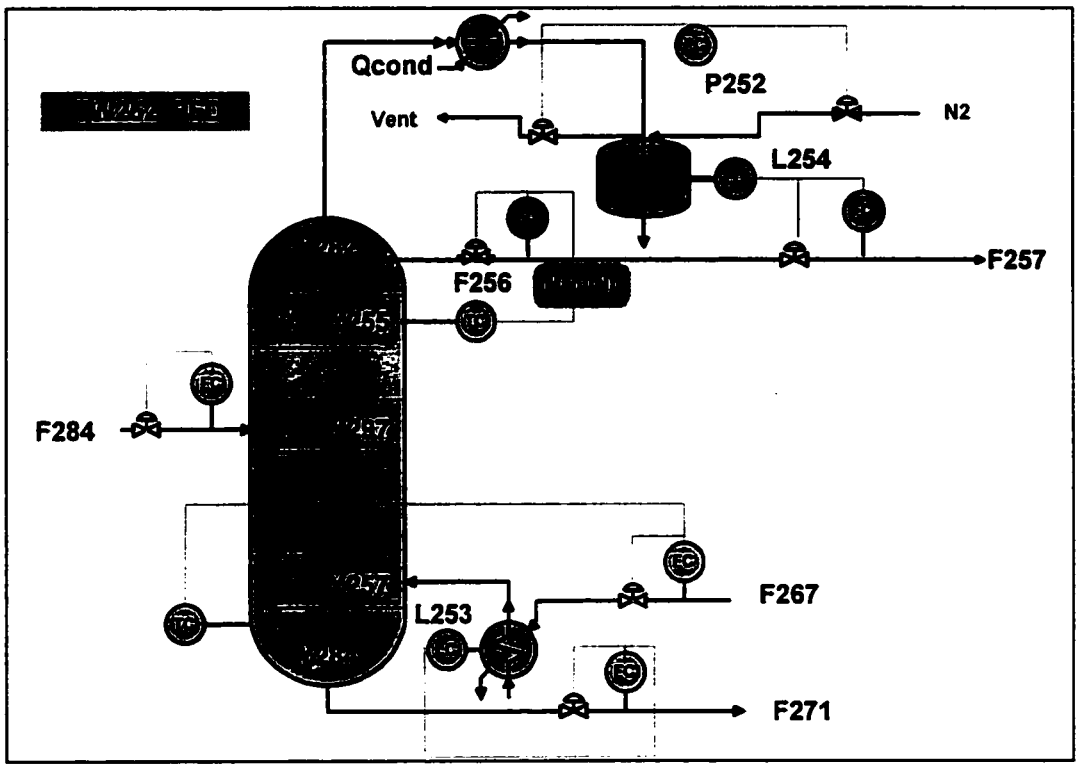


Figure 2.1: TW252 PFD

TW252 contains five degrees of freedom; therefore there are five manipulated flow rates and five controlled variables as shown in Table 2.2:

Table 2.2: TW252 control loops

	Loop 1	**Loop 2**	**Loop 3**	**Loop 4**	**Loop 5**
Manipulated	F256	F267	F257	F271	Vent/N_2
Controlled	Y255	Y284	L254	L253	P252

From the figure and table, the distillate holdup, manipulating the distillate flow rate, F257, controls L254 (condenser liquid level). Similarly, the bottom liquid level, L253, is controlled by manipulating the bottom production rate, F271. The column pressure at the top of tower, P252, is controlled by manipulating the condensate flow rate, Vent, (no measurable tag name associated with this variable) and the low-pressure nitrogen, N_2, flow rate to the condenser. Therefore, the column control scheme reduces to a 2X2-composition control problem. Typically, tray temperatures are usually used as a simple means to infer compositions. Hence, for TW252, the composition both the distillate and bottoms are inferred from two tray temperatures, which are cascaded to the reflux flow rate, F256 and steam flow rate, F267.

The current strategy is to control the temperature at stage 27, Y255, by manipulating a slave reflux flow controller, F256. However, due to unsatisfactory performance the loop is now open and the reflux controller is in the manual mode. Adjusting the steam flowrate to the reboiler (F267) controls the bottoms temperature (Y284). The controller settings are listed in Table A.2 of Appendix A.

2.3 Data Collection

Data collection is the most important step in the development of high fidelity models. The variations in the input/output data and the regions of operation are important factors in determining if the data is a good representation of process behavior over a wide range of conditions. In other words, the data must be "rich" with information. It is critical that the data collected from the process, either through a Yokogawa DCS (Distributed Control System) or OSI Software Inc., PI data historian, be accessed before the data is "compressed". Compressed data or averaged data removes important information about process behavior. Typically, process data must be collected over a reasonably long time, which may extend from several weeks to months, in order to provide information over a wide range of process operations.

The TW252, process data were collected from data stored in an information database known as the "Total System" at the Mizushima Oil Refinery. The Total System acquires data directly from the DCS at three minute sampling intervals without any form of data compression. Three types of process datasets were collected: nominal, step response tests, and disturbance data. A summary outline of the process data is listed in Table 2.3.

The datasets are rich in dynamic information and represent both dynamic and steady state operating conditions. Eight experimental "bump tests" (open and closed loop tests) were performed on TW252; step changes up and down in the setpoints of Y255 (t.1), P252 (t.2), L254 (t.3 and t.7), F256 (t.4), F257 (t.5), Y284 (t.9), and F267 (t.10). The input test signal used for all the open and closed loop experiments is shown in Figure 2.2. However, for some tests, only a step up or a step down is allowed because of possible product specification violations.

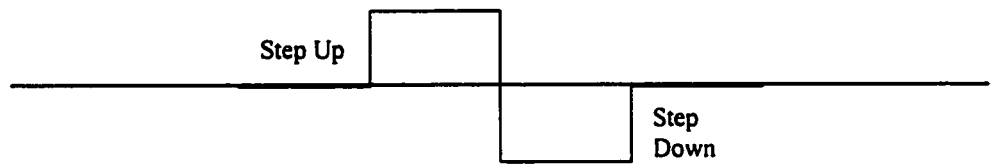


Figure 2.2: Example of a Test Signal

In Figure 2.1, controllers F256, F267, F257, F271 are slave controllers to controllers Y255, Y284, L254, and L253, respectively. Therefore from Table 2.3 datasets t.4, t.5, and t.10 were used to validate the prediction capability of the model during open loop setpoint response behavior. Datasets t.1, t.2, t.3, t.7, t.8 and t.9 are conducted for the purpose of evaluating the predication capability of a model during closed loop setpoint response behavior. Furthermore, dataset t.6 is not included in any model validation due to its lack of input excitation and was neglected. Incidentally, t.3 was repeated to test the new controller tuning parameters for L254 controller and this dataset was t.7. The step response datasets (t.1, t.2, t.3, t.4, t.5, and t.9) were assembled

and labeled as dataset I.t and are plotted in Figures E.1 (a) through (d) with the dataset statistics tabulated in Table E.1, Appendix E.

2.3.1 Additional Data

In addition to the experimental "bump test" data, a feed composition disturbance resulting from an upset condition in an upstream crude oil distillation unit was also collected and labeled as dataset t.8. Dataset t.8 is plotted in Figures E.3 (a) through (d) with dataset statistics tabulated in Table E.3, Appendix E. Furthermore, nominal datasets were collected consisting of three very different normal operating conditions as shown in Table 2.1. Nominal 1 (N.1) dataset consists of six consecutive days of nominal operating conditions recorded during the month of July 1995. From 7/01/95-7/04/95, the column was approximately at steady state. However, on 7/05/95, a large change in feed temperature, Y280, occurred. Nominal 2 (N.2) and 3 (N.3) datasets were deemed to be at steady state and are indicative of May and June operation, respectively. These nominal datasets are assembled and labeled as dataset I.n and are plotted in Figures E.2 (a) through (d) with the dataset statistics tabulated in Table E.2, Appendix E.

2.4 Process Knowledge

It should be noted that seasonal effects have a significant effect on the column's operation. The setpoints for temperature and level controllers stay approximately the same during seasonal changes. However, flow setpoints do differ because lower cooling water temperatures during winter months reduces reflux temperature (Y256). To keep the top and bottom temperatures of the column at setpoint conditions, it is necessary to decrease the reflux flow rate (F256). Furthermore, during winter months the demand for products is lower, therefore, the feed flow rate (F284) is decreased as a direct result of lower production rates from upstream units.

Table 2.3: Summary of the TW252 data collected from the Mizushima Oil Refinery

Test	Date of Test	Type of Test	Perturbed Variable	Process Description	Variable Size up / Size Down	Location
t.1	07/17/95	Setpoint	Y255	Y256 & Y284 OL	0.5 / 0.0 °C	Fig. E.1(a), Samples 237-288
t.2	07/18/95	Setpoint	P252	Y255 & Y256 OL	10 / 10 mm Hg	Fig. E.1(d), Samples 766-846
t.3	07/19/95	Setpoint	L254	Y255 OL	10 / 5 %	Fig. E.1(d), Samples 1258-1388
t.4	07/20/95	Setpoint	F256	Y255, Y256 & Y284 OL	5.0 / 2.5 kl/h	Fig. E.1(b), Samples 1722-1804
t.5	08/03/95	Setpoint	F257	L254 & Y255 OL	1.5 / 1.5 kl/h	Fig. E.1(b), Samples 2155-2210
t.6	08/07/95	Disturbance	Y280	Y255 OL	0 / 3.7 °C	-
t.7	04/23/96	Tuning	L254	All CL	10 / 5 %	Fig. 5.7, Samples 1-57
t.8	01/15-21/96	Disturbance	A264-A257	All CL	-	Fig. E.3, Samples 1-121
t.9	06/25/96	Setpoint	Y284	Y255 OL	2.0/2.0 °C	Fig. E.1(a), Samples 2400-2603
t.10	01/16/97	Setpoint	F267	Y255 & Y284 OL	0.1/0.1 t/h	-
N.1	07/01-06/95	Nominal	-	Y255 OL	-	Fig. E.2, Samples 1-2899
N.2	05/01/96	Nominal	-	Y255 OL	-	Fig. E.2, Samples 2900-3380
N.3	06/14-15/96	Nominal	-	Y255 OL	-	Fig. E.2, Samples 3381-3860

OL = Open Loop
CL = Closed Loop

2.5 Summary

In this chapter, an industrial distillation column, TW252, is described. Various operating data, which included step responses, nominal and disturbance information were collected from the Mizushima Oil Refinery and summarized.

Chapter 3

Physical Dynamic Model Development

In order to investigate the predictive capabilities of a dynamic simulation model for TW252 against plant data, a realistic physically based model is developed. The model is validated using actual process data to ensure both the steady-state values and transient responses are closely matched to the step response data collected from the plant tests outlined in Chapter 2.

3.1 Introduction to HYSYS

Physical models based on first principles and model parameter estimation are the traditional approaches to developing a physical model. In the past, physical modeling was computationally demanding and required simplifying assumptions that limited its fidelity. Furthermore, there were few attempts at rigorously validating the predictive nature of simulation models.

For complex, multivariable control systems such as distillation, steady-state techniques have been widely used to develop a physical model for distillation control evaluation. Steady-state techniques, such as the Relative Gain Array (Bristol, 1966) and steady-state sensitivity analysis (Tolliver and McCune, 1978), are very effective in evaluating different control structures (Fruehauf and Mahoney, 1994). However, such methodologies seem to be incomplete because they only screen out unworkable schemes and do not provide the engineer an assessment of the control strategies chosen. The lack of process understanding is one of the fundamental reasons why many industrial columns still operate in manual mode with very ineffective controls. Hence, there is a need to develop an engineering tool to improve the methodology for distillation control, design,

and operability. Dynamic simulation was deemed to be the solution with the advantages of control system modification, operator training, set-point optimization, and controller reconfiguration, without actually introducing minor upsets to the process that may have the potential for loss of profit.

During the 1970's steady-state process simulation was time consuming and required experts to create meaningful simulations. With the increased computational power and the advent of PC's, by the mid 1980's, formulating these simulations became a less specialized and less time consuming problem. Furthermore, from the 1980's on, simulators were interactive, graphically oriented and entire plants could be built on one flowsheet in a CAD like environment. With these enhancements, the steady-state simulation of chemical processes became a common everyday engineering task (Lawrence, 1996).

In parallel to the steady-state simulation timeline, dynamic simulation modeling was regarded as highly specialized form of modeling that was very labor intensive requiring coding and solving differential equations that accurately represented the dynamics of a process. As a result, this type of simulation was considered amongst the engineering community as "sacred". However, today, engineers are taking advantage of the increased computational power of PC's and workstations coupled with new and powerful programming architectures such as Object-Oriented Programming to create dynamic simulation packages (Fruehauf and Mahoney, 1994).

With today's technology, dynamic physical models can be built on the same knowledge base, underlying thermodynamics and physics of the process, as steady-state models. An example of a commercial software package is HYSYS™ from Hyprotech Ltd. (Calgary, Canada). HYSYS™ is a rigorous interactive simulator that integrates steady-state and dynamic applications into one common model based on the same thermodynamic property package. Figure 3.1 outlines the basic steps to creating a HYSYS™ simulation model.

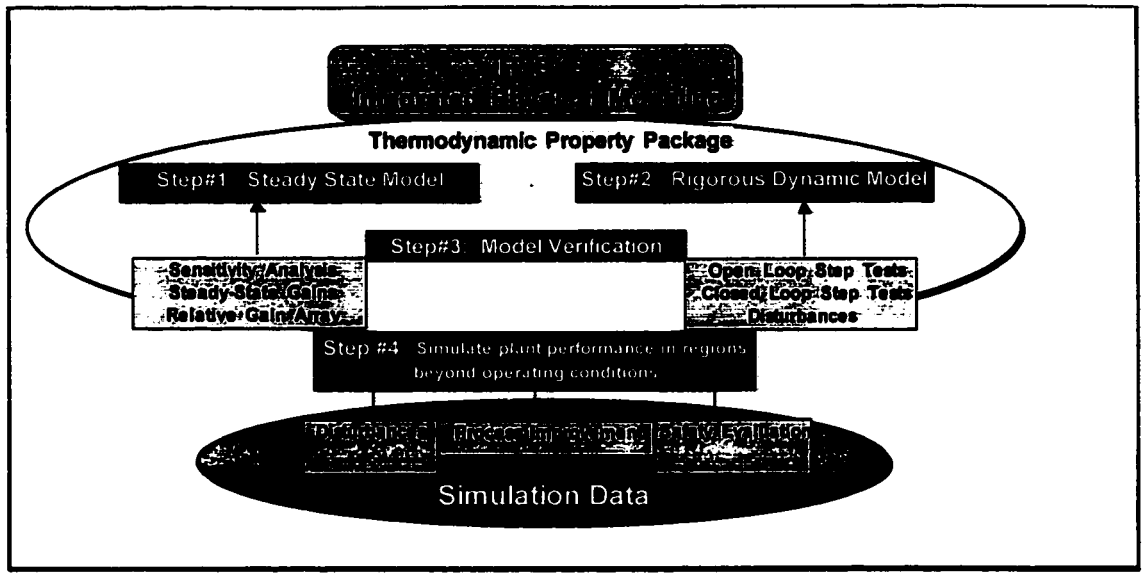


Figure 3.1: Modeling steps in simulation model development

3.2 Modeling of TW252

To develop a physically based process model of TW252 that realistically predicts the plant dynamics, a dynamic distillation model is developed using HYSYS^{TM}. The development of a HYSYS model is based both on engineering process knowledge and judgement. A steady-state model is first developed that solves the differential mass and energy balances in the simulation as algebraic equations and initializes the dynamic model. Next, a dynamic model for TW252 is developed that includes the continuity equations (Appendix C) for mass and energy are written for each tray in the column, condenser, reboiler, and all integrating controllers.

3.2.1 Dynamic Distillation Stage Model

The dynamic model for a single distillation stage is developed and a schematic is shown for the nth tray in Figure 3.2. Liquid enters the through the downcomer of the tray above (L_{n+1}). Vapor enters the tray from the tray below (V_{n-1}). The vapor and liquid completely mix on the tray and the vapor (V_n) leaves, in equilibrium with the tray liquid compositions, and passes through to the tray above. The liquid (L_n) flows over the outlet

weir into the downcomer and to the tray below. The tray may also contain a feed (F), depending on its location in the column.

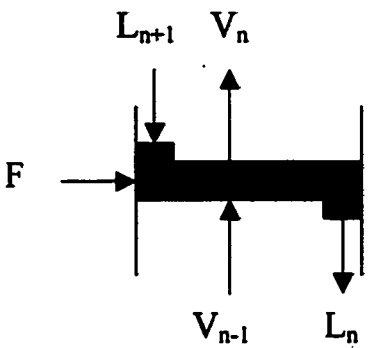


Figure 3.2 Distillation tray model

The dynamic model for a single tray contains N_{c-1} differential material balances, where N_c is the number of components in the system:

$$\frac{dx_i M_n}{dt} = z_{i,f} F + x_{i,n+1} L_{n+1} + y_{i,n-1} V_{n-1} - x_{i,n} L_n - y_{i,n} V_n \qquad \text{Equation 3.1}$$

where: M_n = Material on the n^{th} tray

x_i, y_i = liquid and vapor mole fractions

z_i = feed composition (mole fraction)

One overall material balance for the nth tray:

$$\frac{dM_n}{dt} = F + L_{n+1} + V_{n-1} - L_n - V \qquad \text{Equation 3.2}$$

One overall energy balance for the nth tray:

$$\frac{dh_n^L M_n}{dt} = h_f F + h_{n+1}^L L_{n+1} + h_{n-1}^V V_{n-1} - h_n^L L_n - h_n^V V_n \qquad \text{Equation 3.3}$$

$$\frac{dh_n^L}{dt} \approx 0$$

where: h = specific enthalpy (J/mol)

The hydraulics are accounted for using the Francis Weir Equation:

$$L_n = C_{p,n}^{L}\, WL\, WH^{1.5} \qquad \text{Equation 3.4}$$

where: WL = Weir Length (m)
WH = Weir Height (m)
C_p = Heat capacity (J/mol K)

The general vapor-liquid equilibrium equation is given by the following equation and the generalized model is described in detail in Appendix C:

$$y_{..,n} = K_{..,n} x_{..,n} \qquad \text{Equation 3.5}$$

where: K_i = vapor-liquid equilibrium constant

The following assumptions are required in order to solve the dynamic stage model:

1. Single flow pass tray hydraulics
2. The change in specific enthalpy is very small compared to the total tray enthalpy. Therefore the energy balance equation reduces to an algebraic expression.
3. Vapor mass is less than 30% of the total tray material, hence vapor holdup is negligible.
4. Stage pressure or tray pressure drop is constant and the pressure profile is determined linearly from the condenser pressure.

Luyben (1992) has suggested that the above assumptions are "good enough" to solve 95% of the industrial distillation problems. Given these assumptions and Equations 3.1 and 3.5 the following procedure is used to solve the stage model, which starts from the bottom of the column and proceeds upward.

1. The condenser sets the pressure profile through the column

2. Liquid compositions for each tray, $x_{i,n}$, throughout the column are linearly initialized from the steady-state solution, which is based on a product specification for the distillate.
3. The vapor compositions for each tray, $y_{i,n}$, are calculated using the VLE equation and the temperature for each tray, T_n, is calculated using a bubble point iterative flash calculation (see Appendix C for details).
4. The liquid and vapor enthalpies for each tray, h_n^L and h_n^V are calculated from the EOS (see Appendix C for details).
5. The liquid internal flow rate leaving the tray, L_n is calculated using the Francis Weir Equation (Equation 3.4).
6. The vapor leaving each tray, V_n, is calculated from the energy balance, which is simply an algebraic expression from Equation 3.3.
7. The derivatives for the component and total mass balance, Equations 3.1 and 3.2, are calculated.
8. The equations are then integrated using the Euler Method (which will be described later).

3.2.2 Condenser

The energy dynamics of the condenser are fast relative to column composition dynamics. For TW252, the condenser pressure is actually controlled by regulating a nitrogen valve and vent flow valve as depicted in Figure 2.1. If the condenser pressure increases above the setpoint, the vent valve is allowed to open, however if condenser pressure is below the setpoint nitrogen is injected into the condenser. This control system was deemed to be very complicated to emulate using HYSYS due to the way an inert is modeled in HYSYS. The nitrogen inert would actually condense in the distillate at steady-state causing very different steady-state conditions. In dynamics, nitrogen can be specified as an inert, however it only leaves the liquid after the simulation starts causing the condenser duty and top tray temperatures to change as nitrogen enters the columns vapour phase (Chen, 1997). Therefore, it became impossible to achieve an appropriate steady-state to initialize the dynamics of the column (Chen, 1997).

The condenser is used to condense vapour from the stage model by removing its latent heat with a coolant. The condenser, used in the development of the TW252 physical model, is a modified total condenser. In HYSYS a partial condenser partially condenses a vapor feed into vapor and liquid product streams and the heat removed is determined from utility fluid parameters. The amount of condensation determines the condenser pressure. On the other hand, in a total condenser the entire vapor feed is completely condensed and HYSYS determines the amount of heat required to condense the overhead vapor to its bubble point temperature plus the sensible heat for any subcooling. A modified total condenser is an un-vented (no vapor product) partial condenser with varying pressure and the heat removed is calculated from the utility fluid parameters. The condensate temperature of the liquid leaving the condenser, allowing for subcooling is specified. From the data sets outlined in Chapter 2, four distinct subcooling temperatures exist for TW252 based on the season of operation. For the summer months the reflux temperature, Y256, is as high as 60 °C. However, in the winter Y256 is as low as 30 °C. The utility fluid used to cool the overhead vapor stream is water. The parameters for the cooling water are the heat capacity (C_p) and temperature approach (condensed vapor temperature minus the water outlet temperature) which are defined as 20 kJ/kgmol °C and 10 °C, respectively. Using the condensation temperature and the temperature approach (ΔT_{app}) the outlet temperature of the water ($T_{w,o}$) is calculated. The inlet temperature of the water ($T_{w,i}$) is defaulted in HYSYS to be 30 °C less than the outlet. The enthalpy removed from the vapor feed (Q_{cond}) is calculated from the following formula:

$$Cooling\ Duty(Q_{cond}) = H_V - H_R - H_D \qquad \text{Equation 3.6}$$

where: H_v = enthalpy of vapor stream to condenser (J/mol)

H_R = enthalpy of reflux stream (J/mol)

H_D = enthalpy of distillate stream (J/mol)

The flow rate of the cooling water is then calculated as:

$$F = \frac{Q_{cond}}{C_p \Delta T_w}$$ Equation 3.7

where: ΔT_w = temperature increase of water (30 °C default)

The product of the overall heat transfer coefficient and the heat transfer surface area (UA) between the cooling water tubes and the vapor feed is then calculated from the following expression:

$$UA = \frac{Q_{cond}}{\Delta T_{app}}$$ Equation 3.8

3.2.3 Reboiler

For the reboiler, the duty is calculated by means of enthalpy balances around the column. Again as with the condenser, instead of a direct duty, a "utility fluid" option may be specified. The specified parameters for the utility fluid for the reboiler are: the product of the overall heat transfer coefficient and heat exchange area (UA), holdup, minimum and maximum flow rates, heat capacity, and inlet temperature. For TW252. the utility fluid used to provide heat to the reboiler is steam. Furthermore, the minimum and maximum flow rates, holdup, heat capacity (C_p) of steam, inlet temperature (Tutil_in), were all specified. The reboiler duty (Q) is calculated from the steady-state value. The initialization procedure for the reboiler is as follows:

$$T_{util\_out} = T_{reb} + \Delta T_{app}$$ Equation 3.9

$$UA = \frac{Q}{\frac{T_{util\_in} + T_{util\_out}}{2} - T_{reb}}$$ Equation 3.10

$$W = \frac{Q}{(T_{util\_in} + T_{util\_out})C_p}$$ Equation 3.11

where:

W	=	steam flow rate
Cp	=	steam heat capacity
T_{util_in}	=	steam inlet temperature
T_{util_out}	=	steam outlet temperature
ΔT_{app}	=	temperature approach (10 °C by default)
T_{reb}	=	reboiler temperature
Q	=	reboiler duty

In HYSYS, the steam flow rate, W, can only be calculated in units of kgmole/hr. However, in this form the steam could not be used directly as the manipulated variable in the Y284 cascade control configuration as seen in Figure 2.1, Chapter 2. Hence, it was necessary to calculate the steam flow rate to match the plant data, using a spreadsheet in HYSYS with the following conversion factors:

$$F267 = \frac{W \times M_{steam}}{1000}$$ Equation 3.12

where M_{steam} = Molar Mass of Steam

For validation of the HYSYS model, neither the T_{util_in} nor T_{util_out} were available from the actual plant operating data for any of the step response tests. As a result, matching the steady-state values of F267 to plant data required a great deal of tuning of the parameter T_{util_in}.

From Figure 3.1, before a steady-state or dynamic model is built, the thermodynamic properties, physical dimensions of the process equipment, and modeling assumptions must be specified.

3.2.4 Thermodynamics

In order to define the process, the property package used to model both steady-state and dynamics of TW252 must be specified. The feed to TW252 is considered to be a relatively ideal mixture of C_8's and C_9's. The C_9's are primarily characterized as 1M2-E-Benzene, and traces of 1,3-E-Benzene and Naphthalene. The Peng-Robinson Equation of State (EOS) is used to model the thermodynamics of TW252 for both steady-state and dynamic operations (HYSYS Reference, 1995):

$$P = \frac{RT}{V-b} - \frac{a}{V(V+b)+b(V-b)} \qquad \text{Equation 3.13}$$

The above equation can be expressed in terms of compressibility (Z):

$$Z^3 - (1-B)Z^2 + (A-2B-3B^2)Z - (Ab - B^2 - B^3) = 0$$

where

$$Z = \frac{PV}{RT}$$

$$b = \sum_{i=1}^{n} x_i b_i \qquad \text{Equation 3.14}$$

$$b_i = 0.077796 \frac{RT_{ci}}{P_{ci}}$$

$$a = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j (a_i a_j)^{0.5}(1-k_{ij})$$

$$a_i = a_{ci}\alpha_i$$

$$\alpha_i^{0.5} = 1 + m_i(1 - T_{ri}^{0.5})$$

$$m_i = 0.37646 + 1.54226\varpi_i - 0.26992\varpi_i^{\,2}$$

$$a_{ci} = 0.457235 \frac{(RT_{ci})^2}{P_{ci}}$$

$$A = \frac{aP}{RT^2}$$

$$B = \frac{bP}{RT}$$

For dynamic modeling of TW252, the Peng Robinson Equation of State was found to simulate TW252 faster than real time. When performing the dynamic simulation, HYSYS permits a user selected thermodynamic calculation procedure. The routines selected for TW252 are summarized in Table 3.1 and defined in Appendix C.

Table 3.1: Dynamic Property Model for TW252

Property to be Calculated	Model Used
K Values	Local Model
Vapour Enthalpy	EOS (PR)
Liquid Enthalpy	EOS (PR)
Vapour Entropy	Linear Model
Liquid Entropy	Linear Model

Additionally, the allowable maximum and minimum temperature and maximum pressure over which dynamic properties are calculated and is user defined in HYSYS. For the HYSYS model the default values were selected. Usually the default minimum and maximum temperatures are 10 °C below the lowest value and 10 °C above the highest temperature value in the flowsheet, respectively. The maximum pressure was selected to be 1 atm above the highest pressure in the flowsheet (HYSYS reference, 1995). Additional details of the dynamic model used in the simulation are given in Appendix C.

3.2.5 Equipment Specifications

The HYSYS model is based on physical dimensions of the design specifications received from the refinery. Physical dimensions for TW252 such as, tray sizes (weir heights, weir lengths, and tray volume), tower volume and cooling volume can all be inferred from TW252's column diameter. Furthermore, for dynamic operations controllers are implemented to mimic control strategy of the plant.

Tray Section

The tray section diameter is completely specified by choosing an appropriate residence time, weir height, and internal liquid flow rate as shown in Equation 3.3 below. From equipment specification sheets, TW252 had a column diameter of 3.25 m, therefore the only unknown in Equation 3.15 is the weir height (WH).

$$D = \sqrt{\frac{4qt}{W.H.\pi}} \qquad \text{Equation 3.15}$$

The liquid volume flowrate, q, is taken as the maximum allowable reflux flow rate (F256) (50 m^3/h) into the column. The holdup time, t, on each tray is generally taken to be 30 seconds (HYSYS Reference, 1995). Therefore the weir height for each tray, WH, is determined to be 0.05 m. Furthermore the maximum liquid volume allowable on each tray is determined to be 0.83 m^3 using the following equation:

$$Tray\ Volume = \frac{\pi}{4} D^2 WH \qquad \text{Equation 3.16}$$

Based on the single tray volume the vapor space in the entire column (tower volume) is assumed to be 10 times the single tray volume multiplied by the number of trays in the column (30 theoretical trays). For TW252, the tower volume is 250 m^3 and the cooling volume (The volume around the condenser tubes), which is 10% of the tower volume, is 25 m^3. It is very important that reasonable values for the tower volume and cooling volume be specified because these parameters define the vapor traffic profile in the column during dynamic simulation.

Vessels

The condenser and reboiler are sized based on a liquid holdup, with a holdup set to between 5 to 15 minutes for both vessels. The condenser and reboiler volumes were not available from the equipment specification sheets. Both the vessel volumes are calculated using Equation 3.17.

$$V = \frac{t\,q}{\frac{Levels\,Setpo\,int\,\%}{100\%}}$$ Equation 3.17

TW252's condenser and reboiler are both sized using a holdup time, t, of 10 minutes (HYSYS reference, 1995). For the condenser, the internal liquid flow rate, q, is calculated from the maximum allowable reflux (F256) and distillate (F257) flow rates and is determined to be 24 m^3. However, the required cooling volume needed to vaporize the overhead vapor was 25 m^3 as determined in the previous section. Hence the condenser volume was increased to 30 m^3. For the reboiler, the internal liquid flow rate, q, is calculated from the steady-state simulation results of liquid fed to the reboiler and the volume is determined to be 30 m^3.

Controllers

In order to control TW252, normal PI controllers were required for level control of the reboiler, condenser, as well as temperature, flow, and pressure controllers. All the controller settings and process variable spans for the simulation of TW252's controllers were made available from the DCS at the Mizushima Oil Refinery and are shown in Table A.2, Appendix A. The actual flow through the control valve is a function of the controller output and valve flow characteristics, which are scaled to instrument ranges. Therefore, the control valves were sized based on maintenance equipment specification sheets, which include both the minimum and maximum flow rates and the percentage of valve openings. The control law or characteristic equation for the PI controllers used to control TW252 is given as:

$$OP(t) = OP_{ss} + K_p(P_v(t) - S_p(t)) + \frac{K_p}{T_i}\int_0^t (P_v(t) - S_p(t))dt$$ Equation 3.18

where: $OP(t)$ = controller output at time t

OP_{ss} = steady-state controller output

K_p = proportional gain of the controller

T_I = Integral (reset) time of the controller

$P_v(t)$ = process variable at time t

$S_p(t)$ = setpoint at time t

3.2.6 Integration Methods

To solve the set of differential equations that are in the dynamic stage model, integration is required. The integration procedure must be started with a set of initial conditions for each state variable. In HYSYS the initial conditions used are the steady-state solution, therefore the method of initialization is very stable (Luyben, 1995). There are three different varying step size integration methods available in HYSYS; Euler, Runge-Kutta-Merson and Richards-Lanning-Torrey. For modeling of TW252, the Euler method was employed because it is the simplest form of an explicit method. Explicit in the sense that no information is required at the next time step, t_{n+1}, the local integration error is estimated and is used to change the integration step size. The Euler method is known as a rectangular integration method. Graphically speaking, it uses a step size, h, to measure a straight line with a slope = 0, from t_n to t_{n+1}. The area under the straight line curve is estimated by a rectangle with the dimensions of h and height $f_n(Y_n,U_n)$ (HYSYS Reference, 1995):

$$Y_{t+\Delta t} = Y_t + \frac{dY_t}{dt}\Delta t \qquad \text{Equation 3.19}$$

where: $Y(t=0) = Y_0$

The Euler integration method works well for TW252 because the ordinary differential equations (ODE) used to model TW252 are not considered to be very stiff. A stiff system of ODE's occurs when the ratio of time constants in a set of differential equations is large. For example, in the distillation column simulation the residence time for the vapour is less than one second per tray, while the composition changes on trays

can take several minutes. Therefore, if a large step size is used then the variables such as vapour flow in the column with small constants can be inaccurately calculated (Luyben, 1992). However, for the dynamic simulation of TW252 the vapor hydraulics and pressure dynamics are not considered, therefore the system of differential equations is not stiff.

3.3 Model Verification with Plant Data

The key to development of useful simulation models based on a first principle model is the "validation" with real plant data. Before the dynamic simulation model is representative of TW252 behavior, the model must be verified with plant operated data. A good model must match the plant at both steady-state and accurately track the process during dynamic upsets. A verification of the model includes a check of the overall material balances, steady-state temperature profiles, product flows and a steady-state gain analysis. Furthermore, the model must be able to track dynamic open loop disturbance tests. Finally, good closed loop performance will guarantee that the model is accurate and useable.

The HYSYS model developed for TW252 is validated using plant open and closed loop step response disturbances and nominal plant data as described in Chapter 2. The steady-state behavior of the column is first validated using nominal plant operating data. A quantitative assessment of TW252's non-linearity and interactive nature is investigated using steady-state process gains and a relative gain array (RGA) analysis. Also, a qualitative assessment of the physically based dynamic model's ability to predict transient conditions in the form of open and closed loop step responses, is conducted.

3.3.1 Material Balance

Initially, the biggest difficulty in obtaining a reasonable steady-state prediction arose in the initialization of TW252's feed conditions, internal flow rates, temperatures and compositions to match the nominal plant data. From the available nominal historical data (N.1, N.2, and N.3 as outlined in Table 2.3) an overall mass balance is performed in the following manner:

1. The temperature, mass compositions, and volumetric flow rate of the feed stream, F284, the distillate stream, F257, and the bottoms, F271, of TW252 were retrieved from the nominal historical process data.
2. Using HYSYS, the mass flow rate for the streams was determined. The mass flows of the individual components were then calculated.

Note only distillate and feed compositions were available from the data historian. Therefore, using the HYSYS model the bottom compositions are inferred based on the actual distillate and feed compositions from the plant. This provides an indication of how well the composition balances would have closed. The errors in the overall mass and component balances for TW252 were calculated using Equation 3.20 and are tabulated in Table 3.2.

$$Error(E) = \frac{Out - In}{In} \times 100\% \qquad \text{Equation 3.20}$$

From Table 3.2 the overall balance closure had a satisfactory error (<10%) of closure for all three nominal conditions. Dataset N.3, was remarkable good for both the overall and component balances with an error in the range of 1.5%. However, for datasets N.1 and N.2 closure was not as good. The error in the overall mass balance may be caused by unmeasured disturbances or the oscillating nature of the feed due to poorly tuned level controllers upstream or just simply sensor measurement error.

Table 3.2: Material balance on TW252 (Units = kg/hr)

	N.1			N.2			N.3		
	In	**Out**	**E (%)**	**In**	**Out**	**E (%)**	**In**	**Out**	**E (%)**
Overall	29680	32032	8.0	25736	28305	10.0	29608	30053	1.5
C_{9+}	10943	11309	3.3	8750	10175	16.3	10049	10207	1.6
M/P-Xylene	11528	12256	6.3	9746	10773	10.5	11298	11424	1.1

For example, from Table 3.2, the discrepancies in the component balances could be caused by the inaccuracies of the on-line gas chromatograph composition measurement. For N.1 and N.2 the measured mass fractions of each stream (distillate and feed) did not add up to 1. Furthermore, on average there is insufficient C_{9+} and M/P xylene entering the column to produce the products at the design purities, indicating that some kind of data compression might have been performed prior to data storage.

3.3.2 Steady-state Verification

The steady-state behavior of TW252 was benchmarked against three possible nominal operating regimes of TW252, datasets N.1, N.2 and N.3 (as outlined in Chapter 2). To test the reliability of the model for prediction of the steady-state operation, the model was operated from one steady-state to another to determine the responses of all variables in the model. The responses were then verified using plant data (N.1, N.2 and N.3). Datasets N.1, N.2 and N.3 are steady-state output for TW252 at different operating conditions, which contain changing pressure, feed temperatures, feed compositions and internal flow rate distributions throughout the column. Steady-state comparisons of the HYSYS model to plant data are shown in Tables 3.3 through 3.5. The reflux flowrate (F256), bottoms temperature (Y284), feed flowrate (F284), feed temperature (Y280), and feed compositions (A264, A265, A266, A267) were utilized as specified parameters to match the other simulated variables to plant data. From Tables 3.3 through 3.5 it can be seen that the HYSYS model predictions of the plant at steady-state show good agreement. There is no standard in the process control industry, but for the basis of this work an error

Table 3.3: Comparison of Actual and Simulated Nominal Conditions for Dataset N.1

Statistics for 480 Recorded 3 minute ACS Data Points						
Tag Name	***Min***	***Max***	***Std***	***Mean***	***HYSYS***	***Error (%)***
F284	37.41	41.13	0.706	39.16	40.37	3.10
F256	43.39	44.17	0.140	43.79	43.75	-0.08
F267	11.22	11.54	0.066	11.37	11.39	0.18
F271	13.08	17.10	0.684	14.90	13.80	-7.43
F285	23.76	26.30	0.399	24.75	26.57	7.37
Y280	160.70	162.30	0.329	161.42	161.70	0.17
Y297	135.70	136.80	0.208	139.09	139.40	0.22
Y255	131.30	132.00	0.120	131.55	130.33	-0.92
Y282	128.30	129.20	0.161	128.65	128.30	-0.27
Y256	46.90	52.70	1.506	49.48	51.05	3.17
Y257	162.20	164.70	0.557	163.30	165.80	1.53
Y284	168.40	169.90	0.299	169.01	168.82	-0.11
L253	53.50	56.40	0.444	55.01	55.00	-0.02
L254	56.00	59.70	0.569	58.03	58.00	-0.05
P252	547.60	556.60	1.632	550.55	550.00	-0.10
A254	0.65	0.96	0.074	0.78	0.76	-3.79
A255	15.00	16.39	0.328	15.87	14.99	-5.54
A256	58.20	59.10	0.227	58.75	56.91	-3.12
A257	25.27	26.83	0.419	25.88	27.38	5.80
A264	35.10	39.89	1.268	36.87	34.96	-5.17
A265	9.52	10.62	0.285	10.19	9.80	-3.84
A266	37.65	39.57	0.411	38.84	37.26	-4.08
A267	17.78	18.65	0.188	18.36	17.98	-2.12

Table 3.4: Comparison of Actual and Simulated Nominal Conditions for Dataset N.2

	Statistics for 480 Recorded 3 minute ACS Data Points					
Tag Name	***Min***	***Max***	***Std***	***Mean***	***HYSYS***	***Error (%)***
F284	33.50	36.15	0.401	34.33	34.57	0.71
F256	34.62	35.94	0.222	35.40	35.40	-0.01
F267	9.66	10.05	0.090	9.87	9.93	0.53
F271	11.55	14.81	0.500	12.91	11.65	-9.73
F285	20.81	22.78	0.338	21.91	22.93	4.65
Y280	160.30	161.40	0.212	160.83	160.70	-0.08
Y297	139.10	139.80	0.148	142.47	143.23	0.54
Y255	135.60	136.00	0.074	135.77	135.00	-0.57
Y282	133.20	133.80	0.126	133.57	133.38	-0.14
Y256	36.70	42.90	0.808	40.03	40.87	2.11
Y257	165.00	168.00	0.554	166.92	169.37	1.47
Y284	171.50	173.20	0.321	172.54	172.25	-0.17
L253	57.30	64.70	1.068	60.01	60.00	-0.02
L254	54.20	55.70	0.234	55.00	55.00	0.00
P252	628.40	631.30	0.422	629.99	630.00	0.00
A254	0.58	0.83	0.050	0.70	0.71	1.92
A255	15.60	16.47	0.238	16.05	15.67	-2.41
A256	57.00	58.50	0.303	57.67	56.88	-1.36
A257	24.62	26.24	0.335	25.46	26.74	5.01
A264	33.15	35.36	0.475	34.00	34.12	0.36
A265	9.98	10.66	0.209	10.30	10.36	0.54
A266	36.90	38.34	0.304	37.87	36.65	-3.22
A267	17.70	18.04	0.081	17.84	17.87	0.17

Table 3.5: Comparison of Actual and Simulated Nominal Conditions for Dataset N.3

Statistics for 202 Recorded 3 minute ACS Data Points						
Tag Name	***Min***	***Max***	***Std***	***Mean***	***HYSYS***	***Error (%)***
F284	37.59	42.21	0.950	39.73	40.56	2.10
F256	39.71	41.26	0.364	40.51	40.80	0.71
F267	10.88	11.31	0.053	10.98	11.00	0.23
F271	10.26	13.98	0.725	11.91	14.15	18.78
F285	21.74	24.14	0.459	23.01	26.41	14.76
Y280	159.35	160.80	0.283	160.09	160.00	-0.05
Y297	140.33	141.13	0.173	143.81	143.77	-0.03
Y255	136.77	137.22	0.083	137.02	136.07	-0.69
Y282	134.45	134.98	0.121	134.71	134.24	-0.35
Y256	43.34	50.46	1.737	45.53	46.05	1.14
Y257	180.13	185.00	0.850	182.83	168.15	-8.03
Y284	172.86	174.10	0.248	173.58	171.80	-1.03
L253	56.58	66.50	1.893	59.97	60.00	0.06
L254	54.22	55.67	0.297	54.98	55.00	0.04
P252	648.18	651.52	0.480	650.00	650.00	0.00
A254	0.50	0.69	0.043	0.59	0.62	4.56
A255	14.40	15.69	0.442	14.89	14.63	-1.77
A256	57.69	58.38	0.232	58.01	58.05	0.07
A257	25.62	27.15	0.458	26.38	26.70	1.21
A264	32.61	35.12	0.532	33.94	33.93	-0.02
A265	9.25	10.19	0.289	9.57	9.51	-0.64
A266	37.30	39.05	0.359	38.16	38.07	-0.24
A267	17.82	18.73	0.222	18.34	18.49	0.81

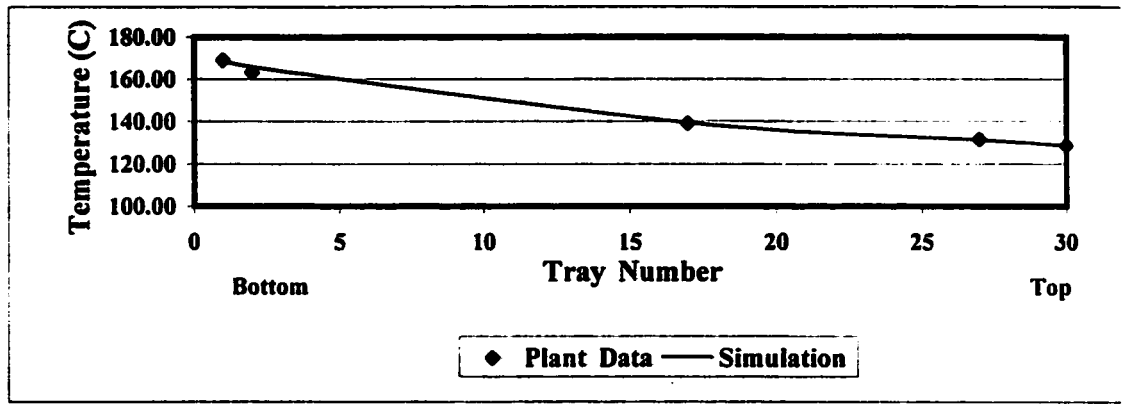
less than 10% is being taken as successful steady-state verification. The shaded areas in Tables 3.3 through 3.5 require further explanations.

Distillate and Bottom Flow Rates

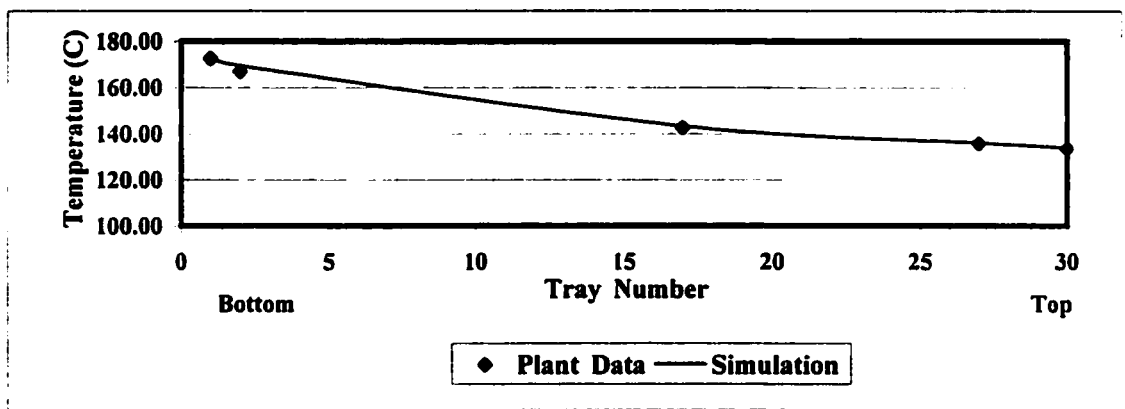
The distillate and bottom flowrates, F257 and F271 for TW252 exhibit a significant discrepancy from the steady-state values in the nominal data sets. For N_1 (Table 3.2), the error is primarily due the oscillatory nature of the feed flow rate F284. Plots of the nominal data collected for selected TW252 variables are shown in Figures 3.2(a) through (d) in Appendix E. From Figure E.2 (b) and the statistics from Table E.2, Appendix E, the oscillatory nature of F257 and F271 is evident. Furthermore, these large fluctuations are mostly likely the direct result of poorly tuned level controllers (L254 and L253), that is investigated in Chapter 5.

Steady-state Temperature Profiles

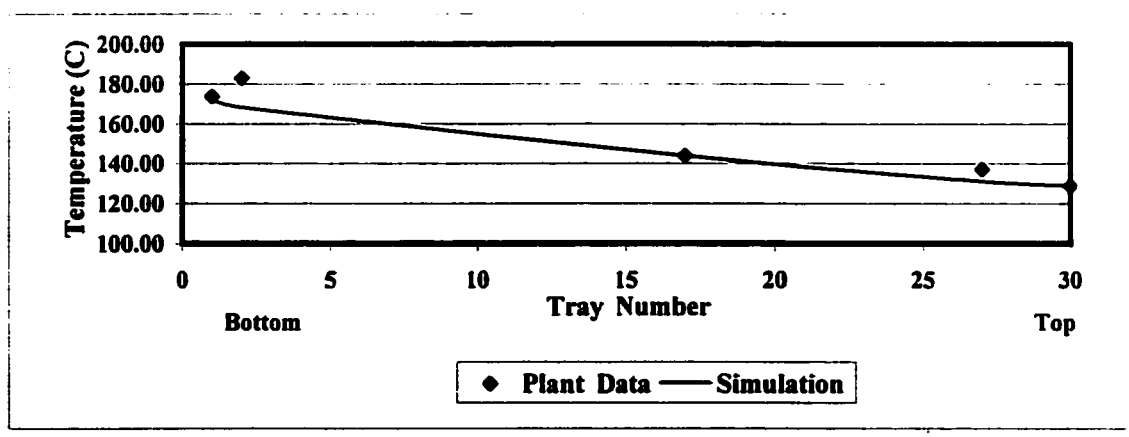
From Tables 3.3 through 3.5, the top tray temperatures, Y255 and Y282 for TW252 are offset from the steady-state value, on average, by 1 and 0.5 °C, respectively. The offset is due to a combination of faulty sensor measurements and changing sensor locations. To compensate for the offset a temperature correction factors of magnitudes 1 and 0.5 C are employed throughout the modeling and verification of the HYSYS model in order to match the plant data more precisely. Figure 3.3 (a) through (c) compares the corrected steady-state temperature profile computed using the HYSYS model with plant data of datasets N.1, N.2 and N.3.



(a)



(b)



(c)

Figure 3.3: TW252 steady-state temperature profile for datasets (a) N.1 (b) N.2 (c) N.3

3.3.3 Sensitivity Analysis

After the steady-state model was validated using various normal operating conditions, the model is used to conduct a steady-state sensitivity analysis to assess quantitatively the non-linearity and interactive nature of TW252. With the level and pressure control loops regulated as shown in Figure 2.1 (Chapter 2), the control problem reduced to a 2x2 composition control problem. To control distillate and bottom compositions, internal tray temperatures are usually used because the frequency of composition measurements is too low for practical control purposes, i.e composition measurements, particularly when done with a GC introduce a significant dead time into the composition control loop.

TW252 employs an L-V control strategy. The current L-V strategy for TW252 is to control top temperature at stage 27, Y255, by manipulating a slave reflux flow controller F256. However, due to unsatisfactory performance the loop is usually open and the reflux flow controller manually controls the distillate purity. In practice, the distillate purity involves control of the light key components (A256, A257), however, for TW252 the reflux rate, F256, is used to control A254, the heavy key component, which is the distillate contaminant. Hence, the focus of the sensitivity analysis will be on C_{9+} components and not m-p-o xylene components. Furthermore at the bottom of the column, the steam flowrate to the reboiler (F267) controls the bottom temperature (Y284). Hence, the L-V control strategy used for TW252 is known as one point composition control, since only A254 is controlled. Skogestad (1992) concluded that the L-V configuration control configuration is a good choice for one-point composition control in most distillation applications.

Steady-State Gains

The steady-state gains were calculated by subtracting the base case temperatures (or compositions) from the final steady-state value and then dividing by the change in the manipulated variable. From the plant test data outlined in Table 2.3 and utilizing the steady state HYSYS model a comparison is done that evaluates the steady-state gains for TW252. For the HYSYS model, step changes in the manipulated variables are

implemented one at a time and the ultimate change in each of the process outputs (controlled variables) are observed. From the plant data, datasets t.4 and t.10 are used to verify the open loop steady-state gains obtained from the HYSYS model. The open loop steady-state gain, k_{ij}, between the *ith* control variable and *jth* manipulated variable are calculated using the following equation:

$$k_{ij} = \frac{\Delta c_i}{\Delta m_j} \qquad \text{Equation 3.21}$$

where: c_i = controlled variable (Y255, Y284)

m_j = manipulated variable (F256, F267)

From Equation 3.21 the open loop steady-state gains for Y255. Y284 (indicative of an L-V configuration) and A254 (one point composition control) can be determined for step changes in F256 and F267 which are represented by the plant data (t.4 and t.10). A comparison between the HYSYS model and plant open loop steady-state gains are shown in Tables 3.6 and 3.7 for datasets t.4 and t.10. Table 3.6, shows the steady-state open loop gains for Y255, Y284, and A254 for a +12% and –6% change in F256 from its nominal value of 41.2 m^3/h as (dataset, t.4).

Table 3.6: Comparison of simulated and plant steady-state gains for Y255, Y284, A254 for step changes in F256

	+12% (41.2-46.2 m^3/h)			-6% (41.2-38.7 m^3/h)		
	Plant	**Simulated**	**$\bar{E}$ (%)**	**Plant**	**Simulated**	**E (%)**
$k_{Y255,F256}$	-0.12	-0.14	-16.67	-0.40	-0.52	-30.00
$k_{Y284,F256}$	-1.92	-1.98	-3.13	-0.84	-0.64	-23.81
$k_{A254,F256}$	-0.097	-0.118	-21.65	-0.489	-0.701	-45.19

From Table 3.6, the signs of the tray temperatures and distillate composition gains are negative for reflux flow changes as would be expected. The open loop steady state gain, $k_{Y255,F256}$, expressed in the units of $C/m^3/h$, exhibits a large non-linear behavior with 4 times more of a gain change for -6% change in F256 compared to a +12 % in the

opposite direction. The same can be said for A254, where $k_{A254,F256}$ (units of wt%/m^3/h) is about 5 times greater for F256 changes in the –6% direction compared to +12% in the opposite direction. However, the non-linear nature of Y284 is not evident from $k_{Y284,F256}$ (C/m^3/h) using the same analogy as before. Therefore, the non-linearity of the column is not as severe at the bottom as is the case at the top of the column. Furthermore, in comparing the simulated open loop gains to the plant open loop gains there is a significant error ((plant-simulated/plant) x 100%). However using the HYSYS model the trends are the same when compared to the plant data. The non-linearity of the column at the top is probably the primary reason the column currently operates using the one-point composition control strategy. The strategy to employ conventional PI controllers to try and regulate either A254 or Y255 with F256 is difficult because of this non-linear relationship. Hence, the Y255 temperature controller operates in open loop and F256 controller is used to manually regulate A254.

Table 3.7 shows similar results for steady-state open loop gains for Y255, Y284, and A254 for a ± 1 % change in F267. The simulated values are compared to the plant data of from t.10, in which ± 0.1 ton/h step changes occurred in F267 from the nominal value of 12.0 ton/h.

Table 3.7: Comparison of simulated and plant steady-state gains for Y255, Y284 and A254 for step changes in F267

	+1% (11.9-12.0 ton/h)			-1% (12.0-11.9ton/h)		
	Plant	**Simulated**	**E (%)**	**Plant**	**Simulated**	**E (%)**
$k_{Y255,F267}$	-3.2	-2.2	31.25	-2.0	-2.3	-15.00
$k_{Y284,F267}$	-8.2	-7.6	7.32	-9.2	-7.5	18.48
$k_{A254,F267}$	-1.03	-1.22	-18.45	-0.93	-0.75	19.35

From the Table 3.7 the $k_{Y255,F267}$ in the positive and negative manipulated variable directions were reasonably consistent indicating that the relationship between Y255 and F267 is fairly linear. The same can be said for the relationship between Y284 and F267, where the gains in the opposite directions are similar. The non-linearity of the $k_{A254,F267}$

is not as severe as $k_{A254,F256}$. The relationship between the Y255, Y284 and A254 to the manipulated changes of F267 is not as non-linear as was seen with changes in F256. Hence, Y284 can be regulated using F267 with fairly good success.

3.3.4 Relative Gain Array Analysis

Once the open loop steady-state gains are determined the degree of interaction for the single loop L-V control configuration (Y255 & Y284 are controlled by manipulating F256 and F267) can be evaluated using the relative gain array (RGA). The RGA can be directly calculated from the steady-state gains. Each RGA element, λ_{ij}, can be determined by performing two experiments. The first experiment determines the open-loop steady-state gain by measuring the response of y_i to input m_j, when all the other loops are opened. In the second experiment, all the other loops are closed and the response of y_i to a change in input m_j is re-determined. Using Equation 3.22, the ratio of these gains gives the desired relative gain elements:

$$\lambda_{ij} = \frac{\left(\frac{\partial y_i}{\partial m_j}\right)_{all\ loops\ \ open}}{\left(\frac{\partial y_i}{\partial m_j}\right)_{\substack{all\ loops\ closed \\ except\ for\ them_j}}} = \left[\frac{open\ \ loop\ \ gain}{closed\ \ loop\ \ gain}\right] = \frac{k_{ij}}{k_{ij}^{*}} \qquad \text{Equation 3.22}$$

where: c_i = control variables, Y255 and Y284

m_j = manipulated variables, F256, F267

The variable k_{ij}^{*} represents the ith loop steady-state gain when all the other loops except the one in question are closed. Whereas k_{ij} represents normal, open loop gain that are calculated as shown in the previous section. Then the RGA can be calculated for two inputs and two outputs from the following equation:

$$RGA = \Lambda = \begin{vmatrix} \lambda_{Y255,F256} & 1-\lambda_{Y255,F256} \\ 1-\lambda_{Y255,F256} & \lambda_{Y255,F256} \end{vmatrix} \qquad \text{Equation 3.23}$$

The results of the RGA analysis using the HYSYS model results for control variables Y255, Y284 are shown in Tables 3.8 and 3.9 for ±1% changes in manipulated variables F256, F267. The simulation model was used to develop the relative gain array elements instead of the plant data because the plant data did not contain any information on k_{ij}^*.

Table 3.8: Relative gain array for TW252

Y255	-0.03	1.03
Y284	1.03	-0.03

Y255	1.03	-0.03
Y284	-0.03	1.03

Table 3.9: Relative gain array for TW252 using opposite gains

Y255	12.04	-11.04
Y284	-11.04	12.04

The calculated RGA values for Tables 3.8 and 3.9 indicate different degrees of interaction and surprisingly different suggested controller pairings depending on the direction of change of the manipulated variables. The suggested controller pairings for various changes are circled in the tables. For negative changes in the manipulated variables the RGA analysis suggests that Y255 (Tray 27th Temperature) should be paired

with reflux flow rate and Y257 (Tray 1 Temperature) with reboiler duty. However for positive changes, the correct pairing is the exact opposite.

Also, if both temperatures increase (maybe a feed composition disturbance) causing a setpoint change in the reflux flow rate to increase and the reboiler duty setpoint to decrease. Therefore, the RGA is calculated using the gains corresponding to changes in opposing direction to the manipulated variables. The result Y255 should be paired with F256 and Y284 with F267. Furthermore, if a disturbance were to occur there would be sever interactions between the two control loops because $\lambda_{11} = 12.04$. Note that McAvoy (1983), indicated that large λ_{ij} for the RGA matrix are typical for L-V configuration schemes like TW252's control strategy. A $\lambda_{ij} > 1$ indicates that the open loop gain between y_i and m_j is larger than the closed loop gain. Therefore, the loops interact, and the retaliatory effect from the other loops act in opposition to the main effect of m_j on y_i (thus reducing the loop gain when the other loops are closed). However, the main effect is still dominant, otherwise λ_{ij} will be negative. For large values of λ_{ij}, the controller gain for loop *i* will have to be chosen to have a larger open loop gain than the other loops because loop *i* could become unstable if the other loops are open. The instability explains why Y255 is currently in open loop. One recommendation is not to pair m_j (F256) with y_i (Y255) without a decoupler.

From the RGA analysis it can be concluded that the interaction effects between Y255 and Y284 are significant and that other control techniques that involve decoupling should be evaluated. One example may the use of a non-linear process model built using neural networks that could be used to provide the necessary decoupling control action. Furthermore McAvoy noted that dynamic interactions tend to be more important for a 2 x 2 process when $\lambda > 1$. Hence, the dynamic responses of TW252 will be the next focus.

3.4 Dynamic Model Validation

The steady-state model was refined to include the effect of dynamics on the actual process. Using the steady-state results the dynamic simulation of TW252 was initialized and then validated using plant setpoint response and disturbance tests, as indicated in Table 2.2 (Chapter 2). The initial conditions of the dynamic model for all the step response tests can be found in Tables D.1-D.7, Appendix D.

Three-minute historical data from the DCS as described in Chapter 2 for all the process variables listed in Table 2.1 are used to verify the dynamic simulation model by qualitatively comparing the dynamic response of actual plant data to the HYSYS model predictions for the selected process variables. Figure 3.4 shows the process variables (designated by a "*") that must be verified by the HYSYS model.

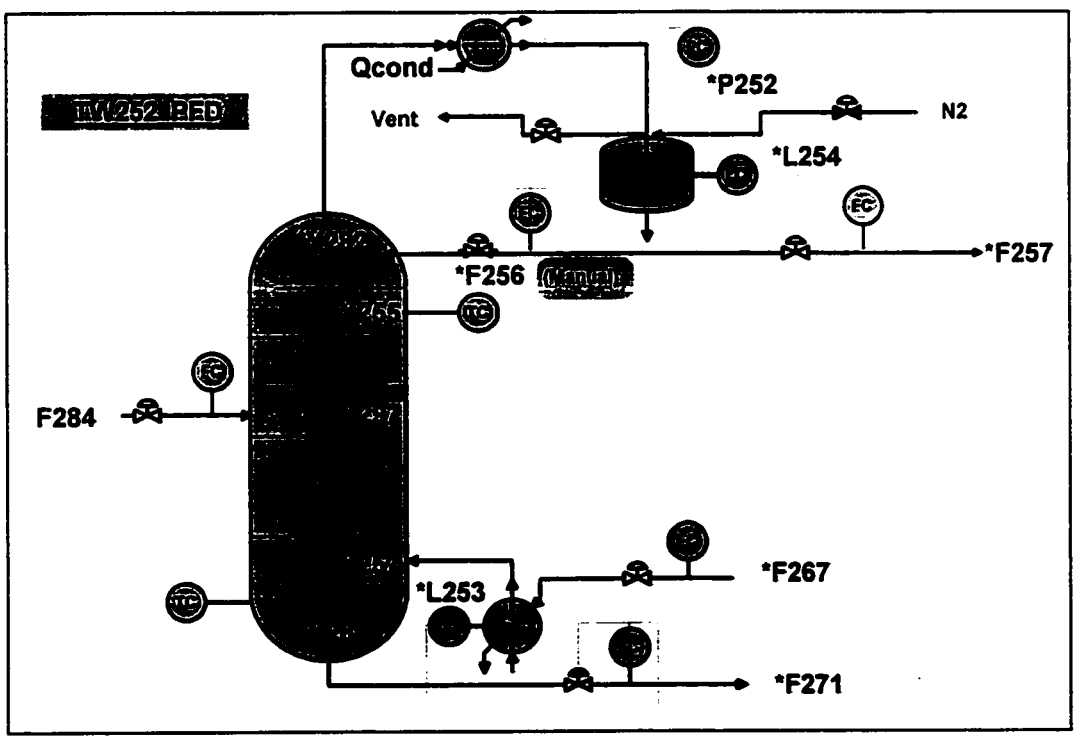


Figure 3.4: The selected process variables to be verified by the HYSYS model

To verify the HYSYS model's predictions of TW252 process variable, the first step was to verify the predictions using open loop dynamic setpoint tests; Test 4 (t.4), Test 5 (t.5), and Test 10 (t.10). The next test performed was to ensure the correctness of the HYSYS model process response to closed loop conditions using setpoint response data from Tests 1 (t.1), 2 (t.2), 3 (t.3), and 9 (t.9). The final test was to validate the HYSYS model response for feed composition disturbances as shown in Test 8 (t.8).

The quality of the HYSYS model predictions is measured using a performance index, I, shown below:

$$I.sim = \frac{\sum_{k=1}^{K}(Y_k - S_k)^2}{\sum_{k=1}^{K}(Y_k - \overline{Y})^2} \times 100\% \qquad \text{Equation 3.23}$$

where: Y_k = actual output at the k sample

S_k = predicted output at the k sample

$\overline{Y}$ = mean of the output for all the patterns (K)

For the HYSYS model predictions of the plant data the extension "sim" is used to distinguish between the performance indices utilized in the subsequent chapters. The *I.sim* is simply the normalized sum of squared errors between the predicted HYSYS value and the actual plant value for the entire dataset. The denominator in equation 3.23 is used to act as a normalization factor to take into account the diverse variability of the datasets used to validate the HYSYS model. Therefore, a relatively unbiased comparison of the HYSYS model predictions of the various degrees of "excitable" plant data can be done. In industry, there is no value that defines a good performance index. Since the index is an "invented" statistical measure it has no definition of what is "good" or "bad". The index is merely used as a mechanism to compare the HYSYS model's predictions of various datasets.

3.4.1 Open Loop Validation

The open loop setpoint response tests used to verify the HYSYS model are shown below. For a detailed description of the tests, refer to Table 2.2 and for a visual representation, see Figure E.1 (a) through (d), Appendix E.

1. Test #4 - Reflux flow rate (F256) setpoint changes - t.4
2. Test #5 - Distillate flow rate (F257) setpoint changes - t.5

The response behavior of the HYSYS model compared to the plant data, using *I.sim*, is only shown for certain variables in dataset t.4 and t.5. For t.4, the responses of selected TW252 process variables to a series of setpoint changes to the manipulated variable, F256, are shown in Figure 3.5 (a) through (d) (+ 5.0 m^3/h at sample number 2 and -2.5 m^3/h at sample number 34 from the nominal value of 41.2 m^3/h). The figures show that the HYSYS model predictions match the response of most of TW252 variables with very satisfactory results. However, for some process variables, certain response characteristics could not be matched. These include oscillations in the feed, F284, the feed compositions, A264 through A267 which are caused by upstream changes in crude feedstock occurring at the made crude fractionator. Other variables that deserve special attention are A254 (C_{9+} hydrocarbons in the distillate), F284 (distillate flowrate), F271 (bottoms flowrate), L253 (reboiler level), and L254 (condenser level).

A254

In Figure 3.5 (c), the HYSYS model was able to predict A254 at the on-line composition analyzer measurement points, which occurred every 50 minutes (16 samples), as indicated by the "-x-" except for one measurement at about the 73 sample point. The low sampling frequency of the on-line analyzer measurement of the distillate compositions seem to miss the true process dynamics between samples 56 and 76 for A254. Note that the allowable A254 measurement range was only 0-2 wt. % as ascertained from DCS specification sheets. Hence, from the actual process data there is

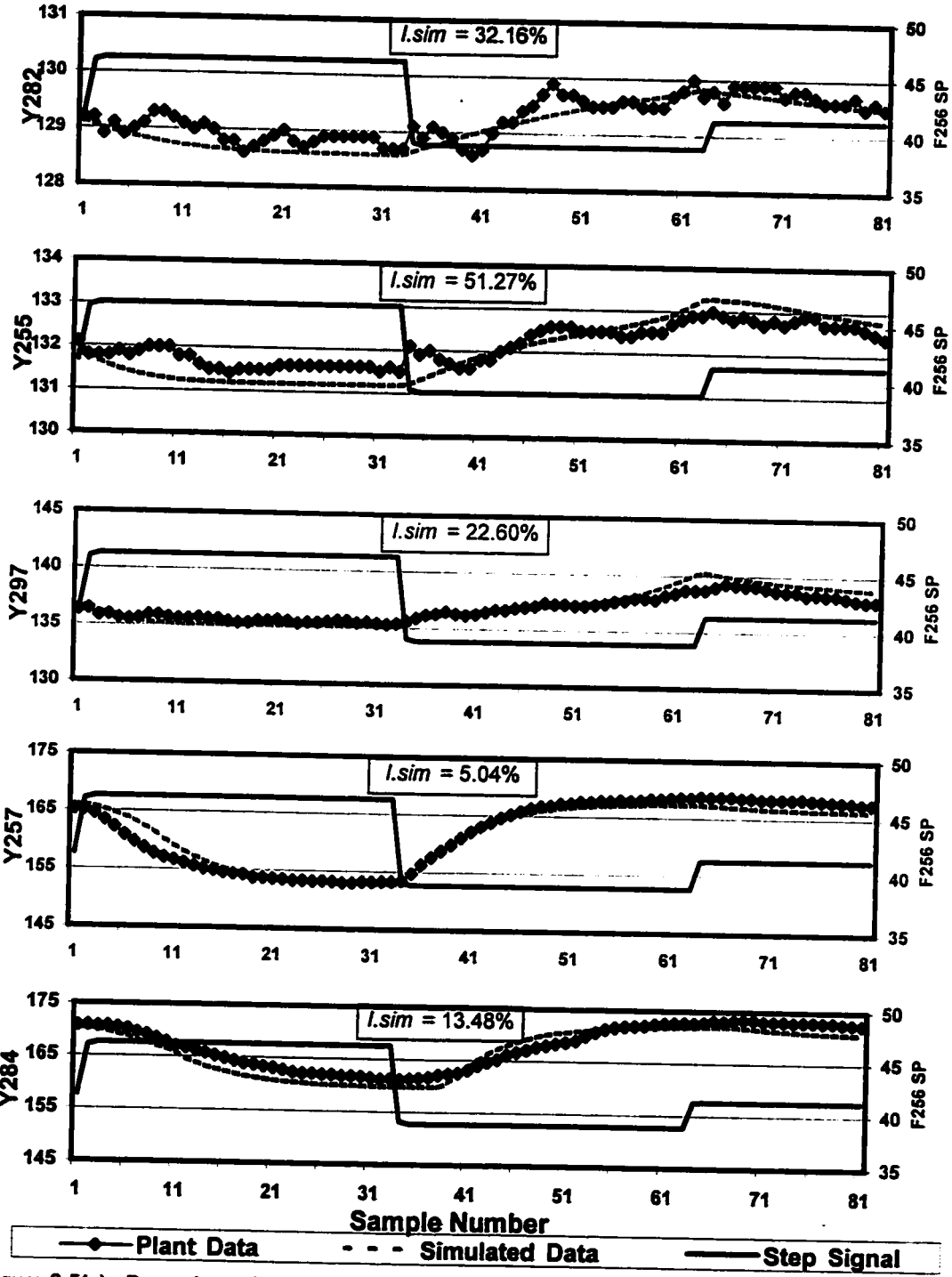


Figure 3.5(a): Dynamic verification of the HYSYS model for TW252 temperatures using plant open loop setpoint changes as seen in dataset t.4

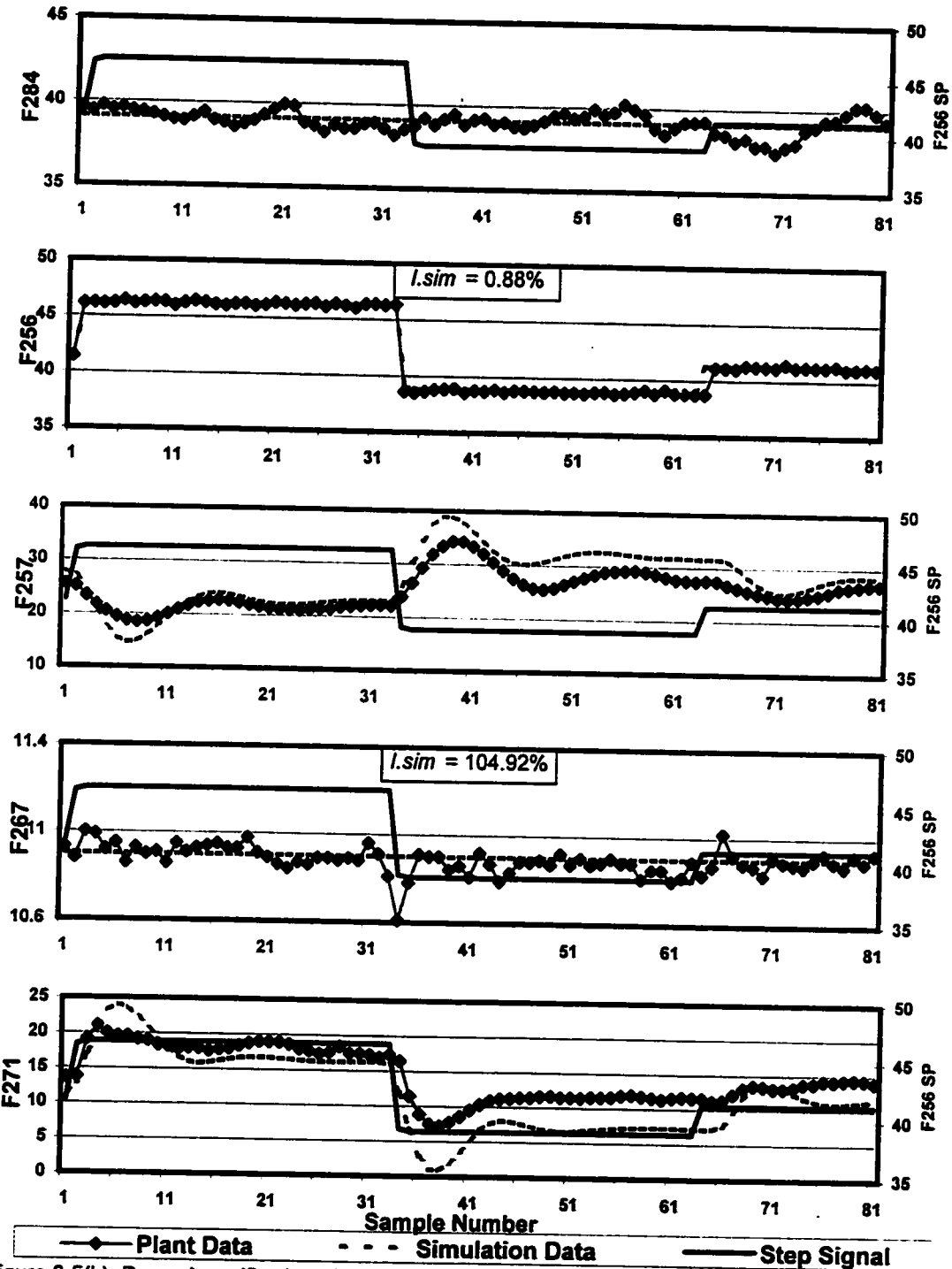


Figure 3.5(b): Dynamic verification of the HYSYS model for TW252 flowrates using a plant open loop setpoint changes as seen in dataset t.4

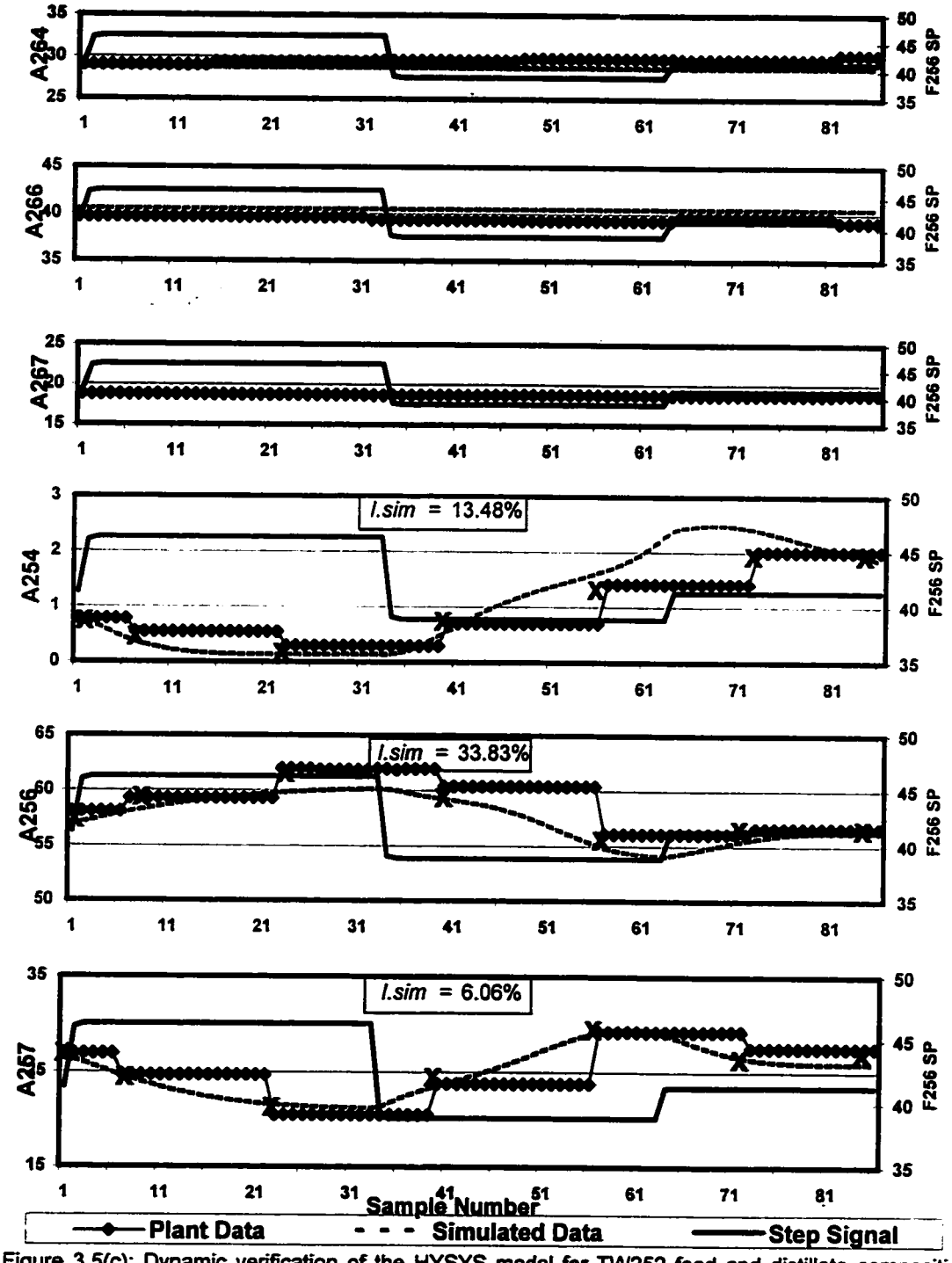


Figure 3.5(c): Dynamic verification of the HYSYS model for TW252 feed and distillate compositions using a plant open loop setpoint change as seen in dataset t.4

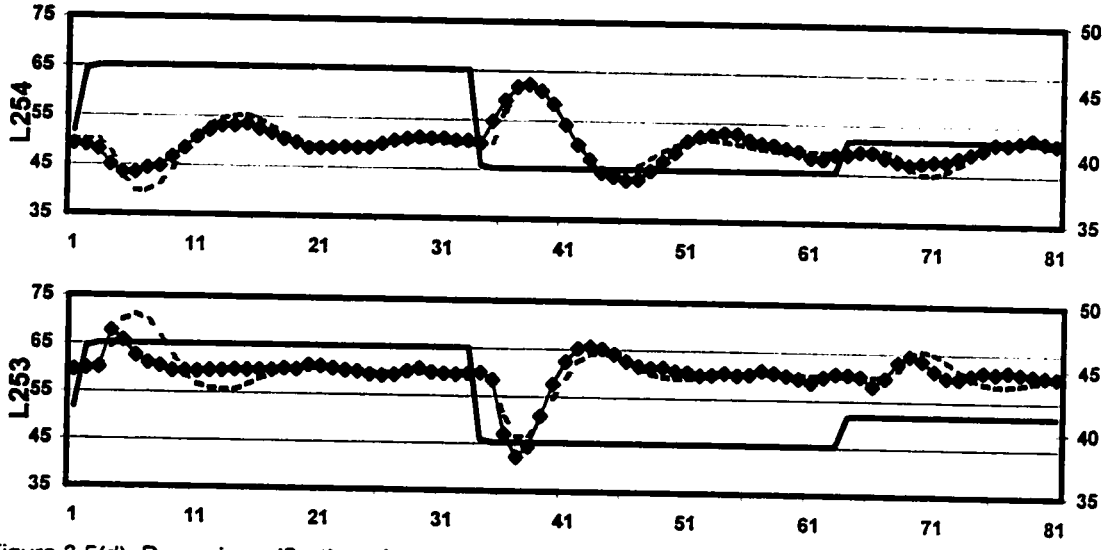


Figure 3.5(d): Dynamic verification of the HYSYS model for TW252 condenser and reboiler levels using a plant open loop setpoint change as seen dataset t.4

no accurate indication of A254 above 2 wt. % since the measurement limit is 2 wt %. The HYSYS model predicted that the response of A254, to a F256 step change down, clearly should be above the on-line analyzer limit of 2 wt. %. Hence, the on-line analyzer low frequency sampling did not detect the apparent dynamic behavior of A254 in dataset t.4.

F257 and F271

The response of variables F257 and F271 to the F256 step change was a little bit aggressive as seen by the overshoot in Figure 3.5 (b). The discrepancies between plant and simulated data for variables F257 and F271 were directly related to the allowable holdup time used to model the reboiler and condenser. The holdup time in the condenser and reboiler was modified from the default value of 10 to 11 minutes and the size of the condenser was changed from 24 m^3 to 30 m^3 and that of the reboiler from 30 m^3 to 34 m^3. A separate plant open loop test was used to model the response characteristics of F257. L254, F271 and L253, known as t.5. Dataset t.5 consists of a series of setpoint changes to the manipulated variable, F257 (+2.5 m^3/h at sample number 4 and -2.5 m^3/h at sample number 24 from the nominal value of 25 m^3/h). Figure 3.6 is a comparison of plant data (dataset t.5) and HYSYS simulated response data for these variables. Using the new vessel sizes, the HYSYS model was able to accurately predict the plant response in dataset t.5 for variables F257, L254, F271 and L253 with very little steady-state error.

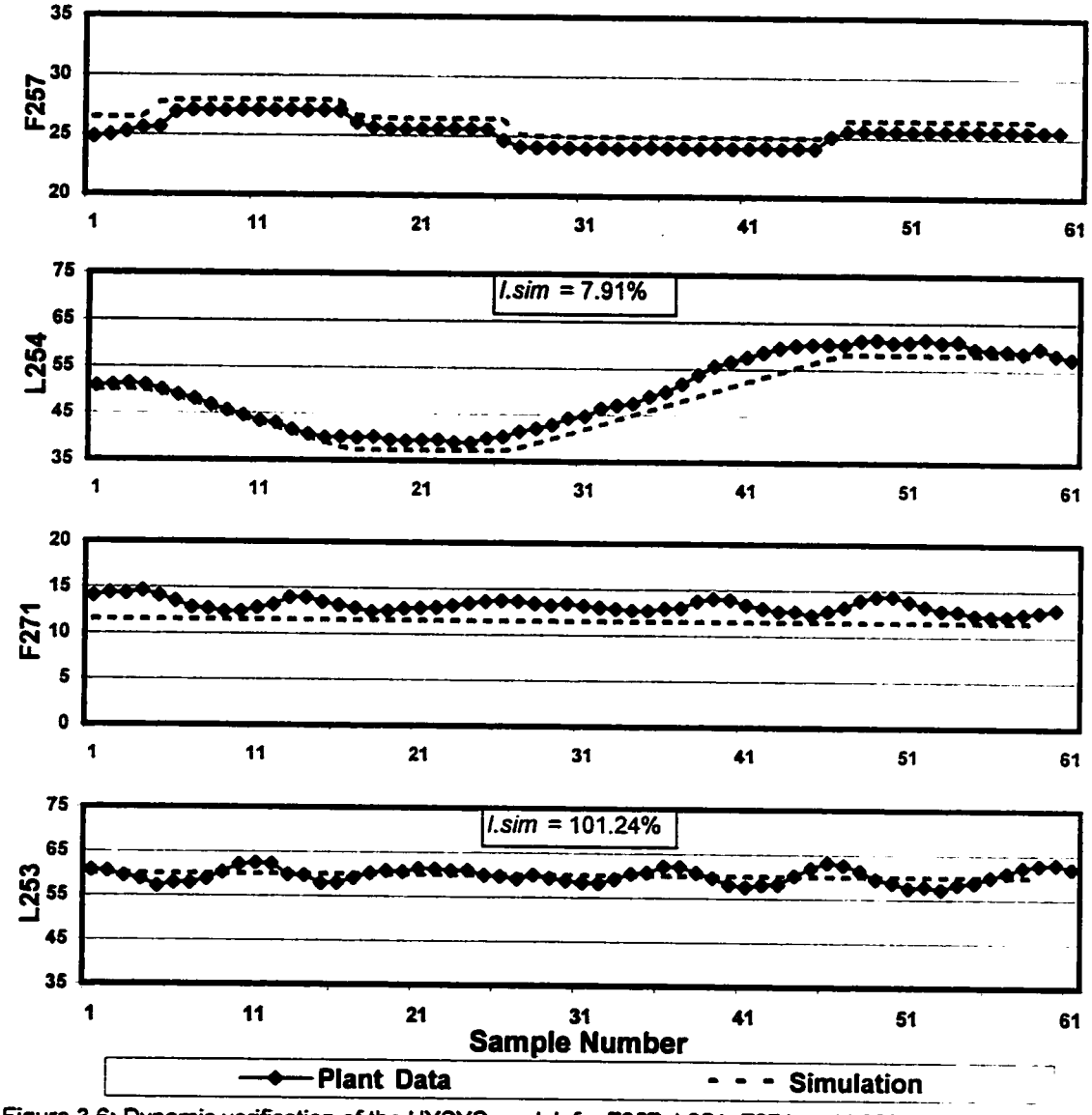


Figure 3.6: Dynamic verification of the HYSYS model for F257, L254, F271 and L253 using plant open loop setpoint changes as seen in dataset t.5

3.4.2 Closed Loop Validation

The second validation step, for the HYSYS model, was to ensure that the simulated process response matched the closed loop response. The validation of the HYSYS model during closed loop setpoint changes was conducted using four plant tests. For a detailed description of the tests refer to Table 2.2 and for a visual representation consult Figure E.1 (a) through (d), Appendix E.

1. Test #1 - Stage 27 Temperature Setpoint Change (Y255) - t.1
2. Test #2 - Condenser Pressure Setpoint Change (P252) - t.2
3. Test #3 - Reflux Drum Level Setpoint Change (L254) - t.3
4. Test #9 - Bottoms Temperature Setpoint Change (Y284) - t.9

To avoid redundancy, one example of the response behavior of the HYSYS model compared to the plant data, using *I.sim*, is shown for selected variables for the above tests. For Test 1 (t.1) the HYSYS model prediction of TW252 process variables to a Y255 closed loop setpoint change of +0.5 C from the its nominal value of 131.2 C at sample number 11 is shown in Figure 3.7(a) through (c). Note that the data shown is only a snapshot of dataset t.1 consisting of 51 samples of the 480 samples in the dataset. Furthermore, the setpoint step signal is plotted on the secondary Y-axis for all the figures.

The HYSYS model accurately predicts the response of most of TW252 variables for dataset t.1. The same variable characteristics exist in dataset t.1, as was mentioned earlier for dataset t.4. For example, A254 from Figure 3.7(c), the HYSYS model predicted that a maximum occurred at 0.9 wt. % at approximately sample 30, which was very close to violating the product specification (A254 must be below 1.0 wt. %). The apparent near violation was not detected by the infrequent sampling of the on-line analyzer for TW252, which is also evident in dataset t.4.

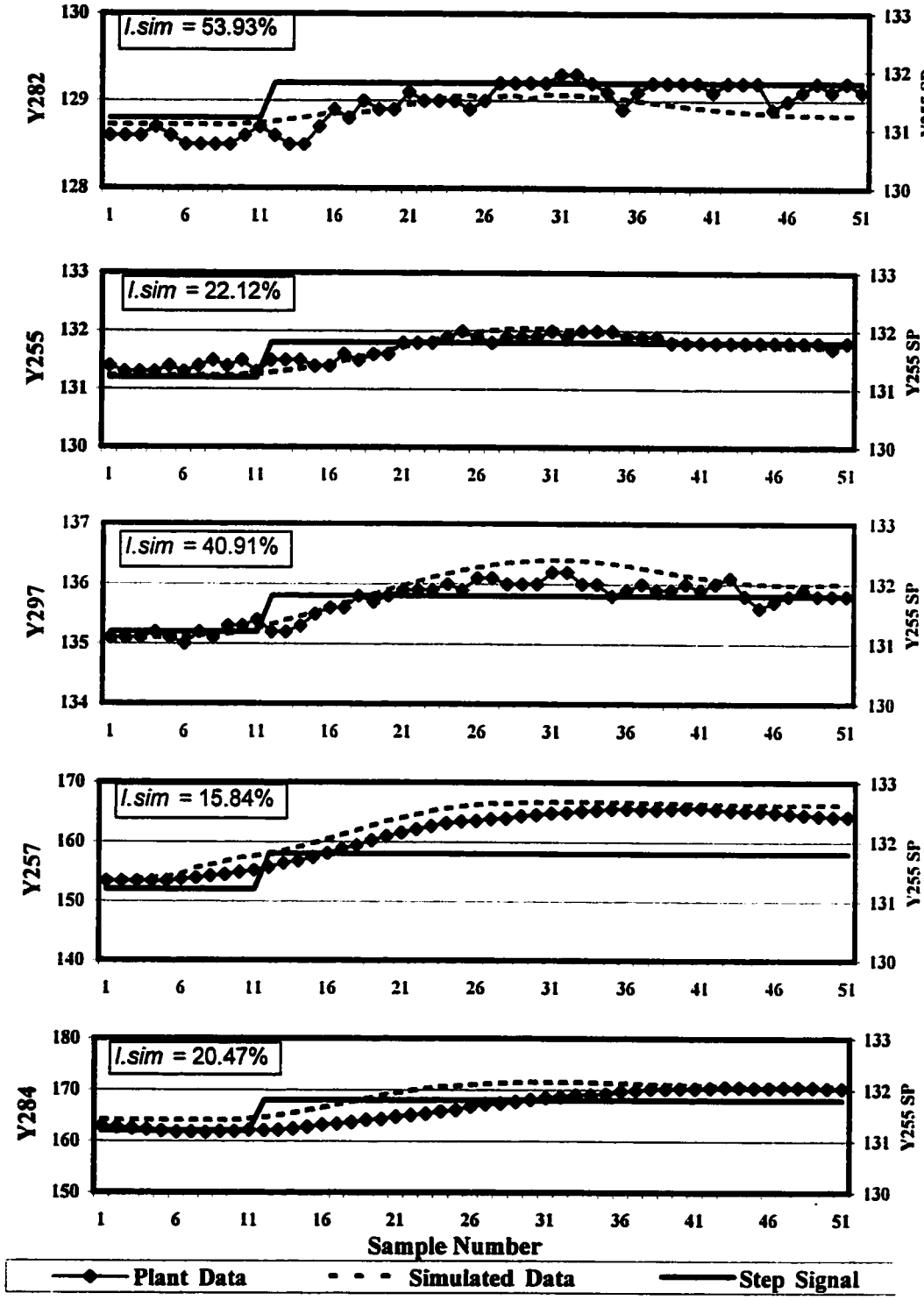


Figure 3.7 (a): Dynamic verifiction of the HYSYS model for TW252 temperatures using plant closed loop setpoint changes as seen in data set t.1

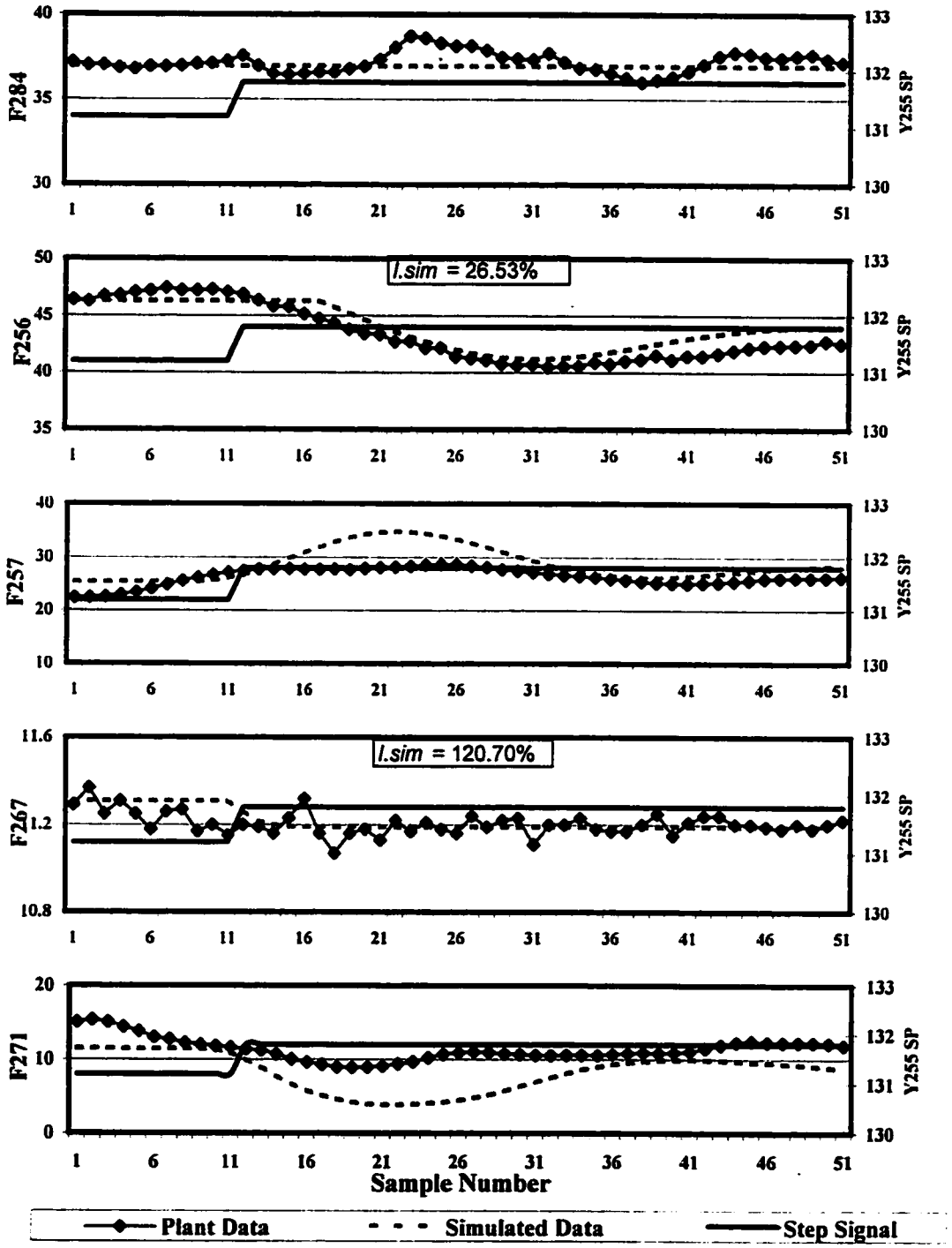


Figure 3.7 (b): Dynamic verifiction of the HYSYS model for TW252 flowrates using plant closed loop setpoint changes as seen in dataset t.1

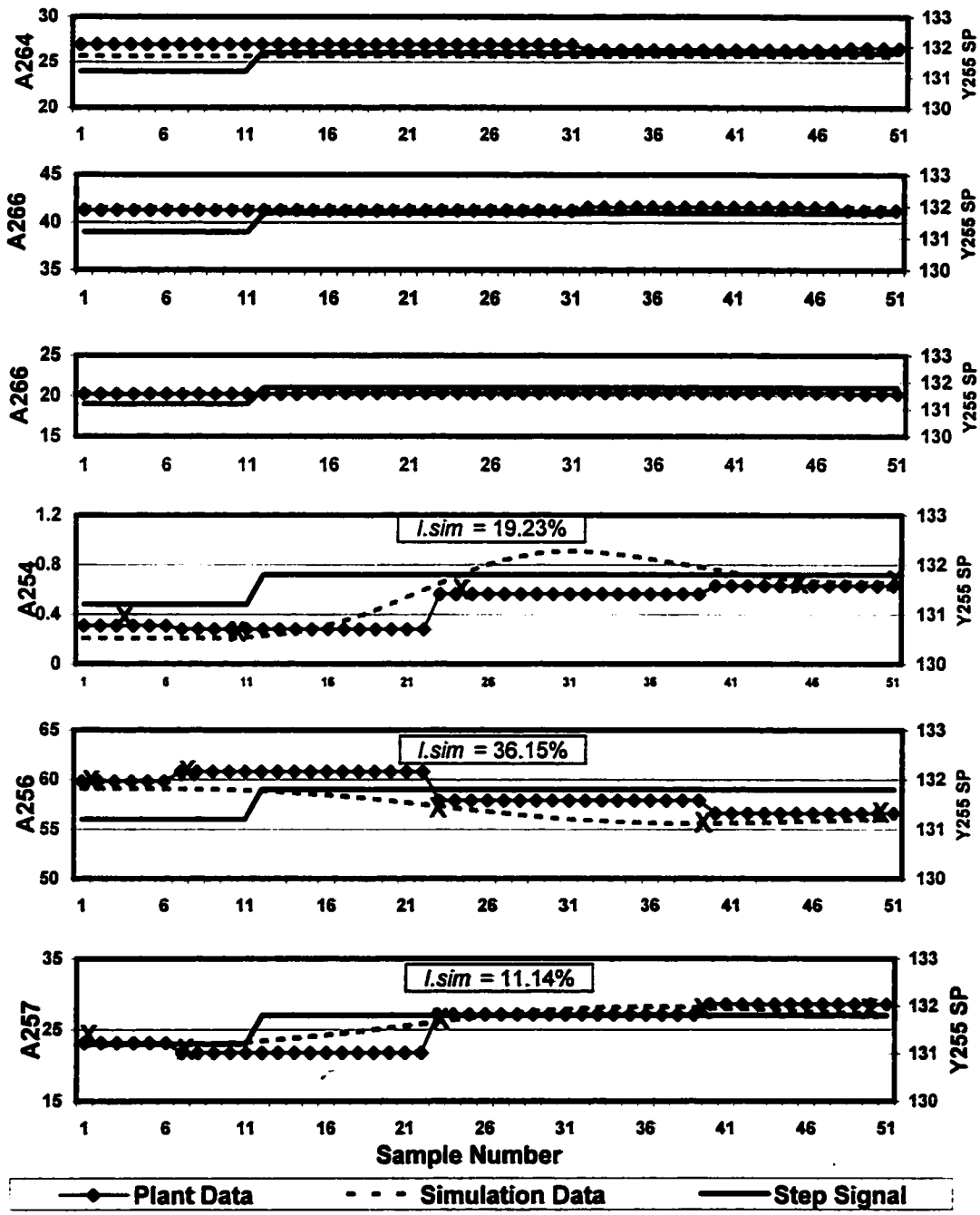


Figure 3.7 (c): Dynamic verifiction of the HYSYS model for TW252 feed and distillate compositions using plant closed loop setpoint changes as seen in dataset t.1

The simulated closed loop response of TW252 variables F257, L254, F271 and L253 are compared using plant data from Test 3 (t.3) and are shown in Figure 3.8. Figure 3.8 compares the HYSYS model prediction of F257, L254, F271 and L253 to a series of setpoint changes in the controlled variable, L254 (+ 10 % at sample number 3 and – 5% at sample number 42 from the nominal value of 50 %). Note that in the figure, the data shown is only a snapshot of dataset t.3 consisting of 121 samples of the entire dataset of 480 samples. Furthermore, the setpoint step signal is plotted on the secondary Y-axis for all the figures.

The HYSYS model predicted the closed loop setpoint changes in the controlled variable L254 very accurately, *I.sim* = 1.74 %. On the other hand, *I.sim* for L253 was about 100%. Furthermore, the manipulated variables, F257 and F271, predicted response was also fairly accurate, *I.sim* 26.53% and 120.70%, respectively. There was a slight difference in the prediction of F257 and F271, which are directly related to the inaccuracies of the material balances in the column as was previously discussed (section 3.3.1)

In comparing *I.sim* values for closed loop tests (t.1 and t.3) and open loop tests (t.4 and t.5) for selected variables, on average, the HYSYS model predictions were able to match the response of certain variables to the open loop test better than the closed loop test. The results are summarized in Table 3.10. For example, the HYSYS model predicted all the distillate compositions and internal temperatures, that were not involved in any of the tests directly (i.e. they were not "perturbed" in any way), better for the open loop tests than the closed loop tests. However, for Y255 and L254, the simulated response was closer to the plant closed loop response than the plant open loop response because these variables were directly "perturbed". The opposite is true for F256 as would be expected.

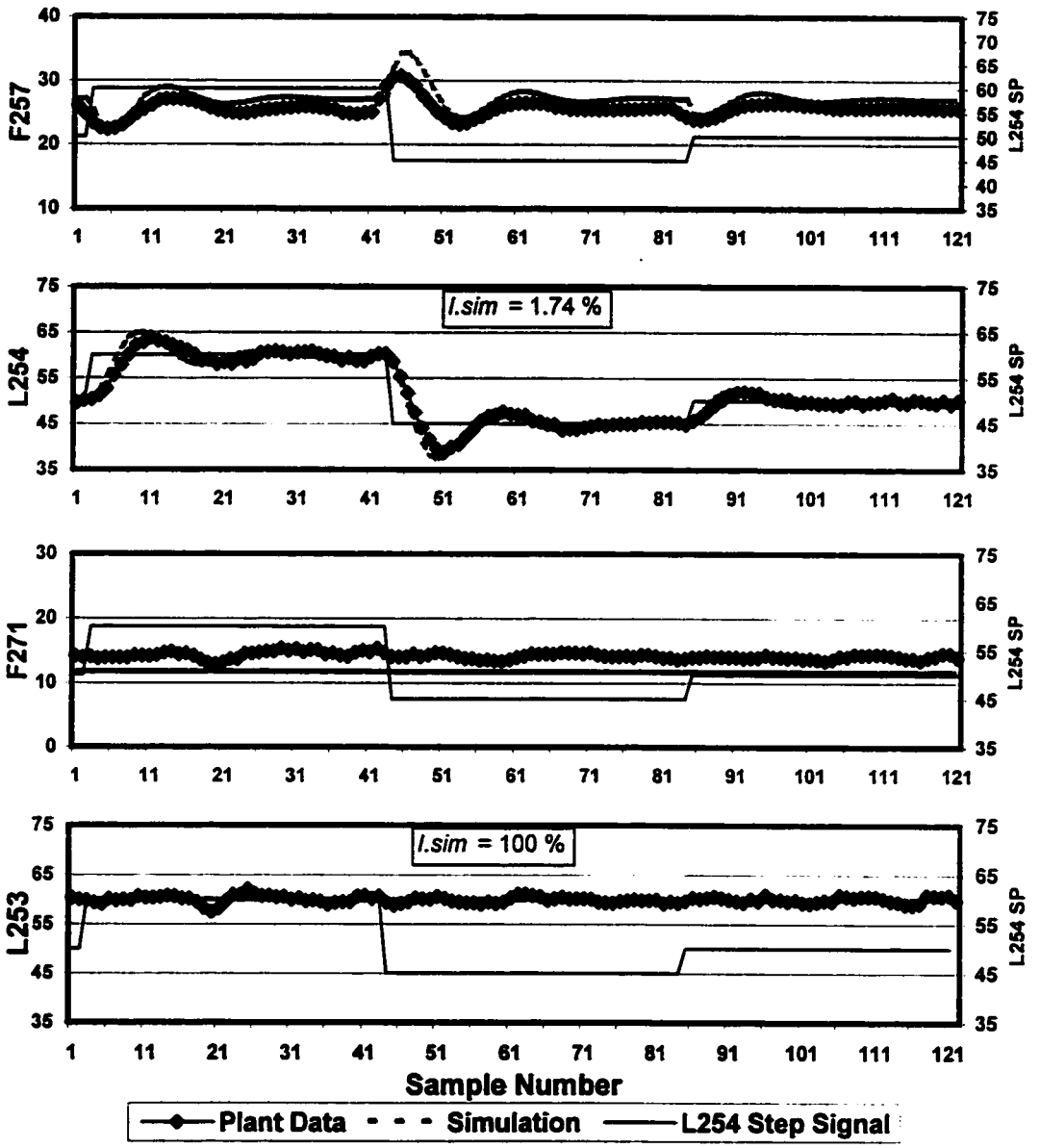


Figure 3.8: Dynamic verification of the HYSYS model for F257, L254, F271 and L253 using plant closed loop setpoint changes as seen in dataset t.3

Table 3.10: Comparison of HYSYS model predictions of closed loop and open loop responses using *I.sim* for selected TW252 process variables.

Variable	Closed Loop	Open Loop
	I.sim (%)	*I.sim (%)*
Y282	53.93	32.16
Y255	22.12	51.27
Y297	40.91	22.60
Y257	15.84	5.04
Y284	20.47	13.48
A254	19.23	33.83
A256	36.15	29.30
A257	11.14	6.06
F256	26.53	0.88
F267	120.70	104.92
L254	1.74	7.91
L253	100.00	101.24

3.5 Performance Benchmarks

Using the validation results as a guideline, proposed benchmark values based on the performance index, *I.sim*, are suggested for the physically based HYSYS model for TW252. The indices are developed for the four major categories, which represent the entire column: internal tray temperatures, distillate compositions, internal flow rates and levels. The reason for the categorization will become evident in the next chapter. The proposed benchmark performance indices for the selected categories are based on the open loop and closed loop verification results of the HYSYS model and are tabulated in Table 3.11.

Table 3.11: Proposed benchmark and recommended upper limit values for HYSYS physical model prediction of plant data

Category	TW252 Variables	Benchmark	Upper Limit
		I.sim (%)	*I.sim (%)*
Internal Tray Temperatures	Y282		
	Y255		
	Y297	28	42
	Y257		
	Y284		
Distillate Compositions	A254		
	A256	24	37
	A257		
Internal Flowrates	F256	50	75
	F267		
Levels	L254	52	78
	L253		

From Table 3.11, the optimum *I.sim* values for TW252 were chosen using a very straightforward procedure. For example, for the measured internal tray temperatures (Y282, Y255, Y297, Y257 and Y284), each individual variable's *I.sim* value for datasets t.4 and t.1 are averaged. Then the five averaged *I.sim* values for each variable are again averaged together to determine the benchmark value. The recommended upper limit values are 1.5 times the benchmark values for each category, which is somewhat of an arbitrary specification. The upper limit values for *I.sim* could be used to verify if the HYSYS model prediction results are satisfactory. For example if the HYSYS model predicted Y282 for a particular dataset with an *I.sim* value > 42%, then the prediction results of the HYSYS model are deemed to be inaccurate for that particular category. The procedure is repeated for the other categories and the results are shown in Table 3.11. Note that performance indices may be very different for various systems and that the suggested benchmarks and limits are only recommended for the simulation modeling of TW252.

3.6 Other Closed Loop Validation Using Plant Data

The HYSYS model was further validated using plant data from Test 2 (t.2) and is shown in Figure 3.9 (a) and (b) for selected TW252 variables. In verifying the HYSYS model using the other datasets, the condenser pressure (P252) was assumed to be constant. Test 2 was specifically performed on the column to verify the HYSYS model ability to predict changing column pressure. Dataset t.2 consists of a series of setpoint changes in the controlled variable, P252 (+10 mmHg at sample number 5 and –10 mmHg at sample number 20 from its nominal value of 555 mmHg). Note that in the figures, the data shown are only a snapshot of dataset t.2 consisting of 81 samples of the entire dataset of 480 samples. Furthermore, the P252 setpoint step signal is plotted on the secondary Y-axis for all the figures. The distillate compositions and the internal tray temperatures were plotted since these were the only measured variables that the test had any effect on. The HYSYS model was able to match the plant data for some of the variables quite well. Based on the performance benchmarks in Table 3.11, the only *I.sim* values that met the criteria were for only the top tray temperatures (Y255 and Y282) and the A254 distillate composition. Since the *I.sim* values for the other tray temperatures were well above the upper limit of 42 %, the predictions are inaccurate. However, the HYSYS model detected the general trend. Using the internal tray temperature benchmark for P252, the *I.sim* value of 13.89% is less than 28%, hence indicates a good prediction by the HYSYS model. Note that the responses of all the internal tray temperatures took a similar shape to that of P252 because the pressure profile in the column must first be specified and then the tray temperatures are calculated. The distillate compositions do not follow the same shape as P252 because the composition dynamics are much slower. Furthermore, from Figure 3.9 (b), the HYSYS model accurately predicted the plant data at the on-line analyzer points (as indicated by the –x-) except for at sample number 48 because there may have been a change in feed composition that was not included in the model. However, the HYSYS model accurately predicted the general shape of the distillate compositions.

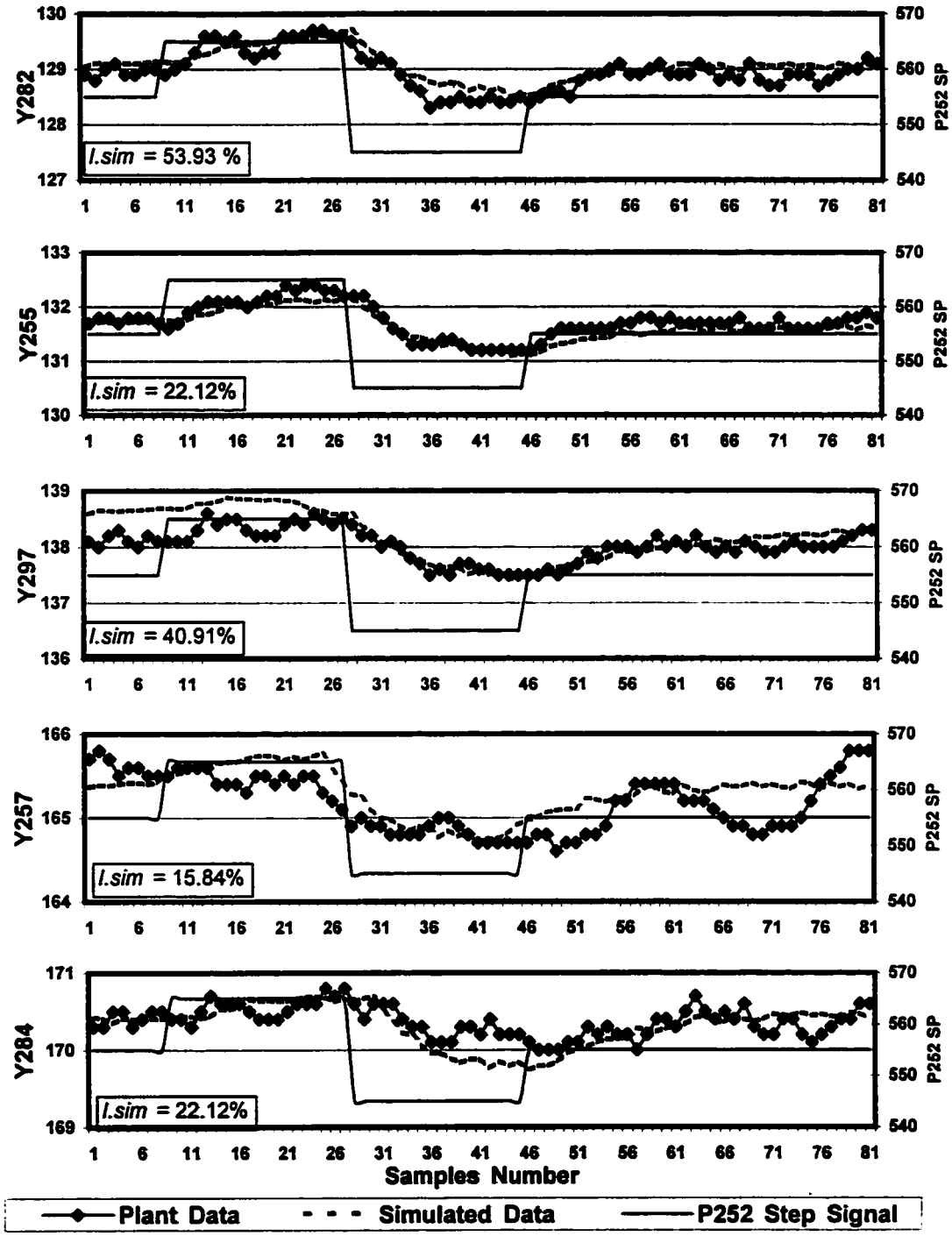


Figure 3.9 (a): Dynamic verification of the HYSYS model for TW252 temperatures using plant setpoint closed loop changes as seen in dataset t.2

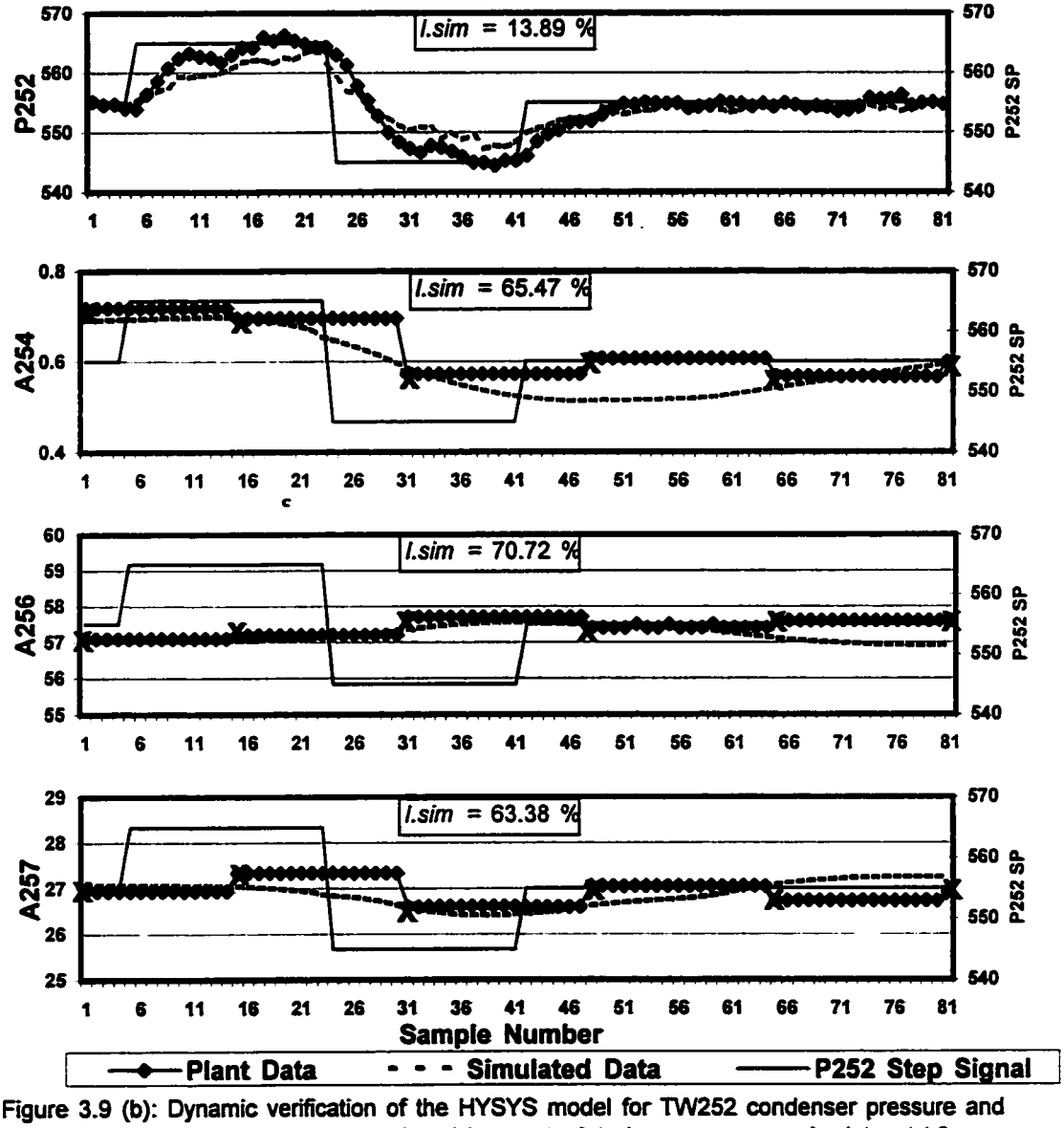


Figure 3.9 (b): Dynamic verification of the HYSYS model for TW252 condenser pressure and distillate compositions using a plant closed loop setpoint changes as seen in dataset t.2

3.7 Disturbance Validation

The final test used to verify the HYSYS model is with a dynamic disturbance test as seen in plant Test 8 (t.8). Dataset t.8 consists of a feed composition disturbance in process variables A264, A266, and A267 at approximately sample number 25 as shown in Figure 3.10 (b). During the disturbance, the operators made many manipulations to compensate for these disturbances such as ramping P252 from 555 mmHg to 540 mmHg as shown in Figure 3.10 (a). Note that in the figures, the data shown is only a snapshot of dataset t.8 consisting of 106 samples of the entire dataset of 480 samples. The internal tray temperatures and distillate compositions are greatly affected by the pressure profile in the column and feed compositions entering the column. Dataset t.8 exhibits some interesting disturbance dynamics that the HYSYS model can be used to predict. From Figure 3.10 (a) and (b) the HYSYS model was able to predict the affect of the feed composition disturbances for selected TW252's process variables. The *I.sim* values were within the upper limits for variables Y282 and Y255; however, the prediction of Y257 and Y284 were not as good because the *I.sim* values violated the upper limit of 42 %. The *I.sim* values for A254 and A257 show that the HYSYS model was not able to predict the distillate compositions within the upper limit of 37 % for dataset t.8.

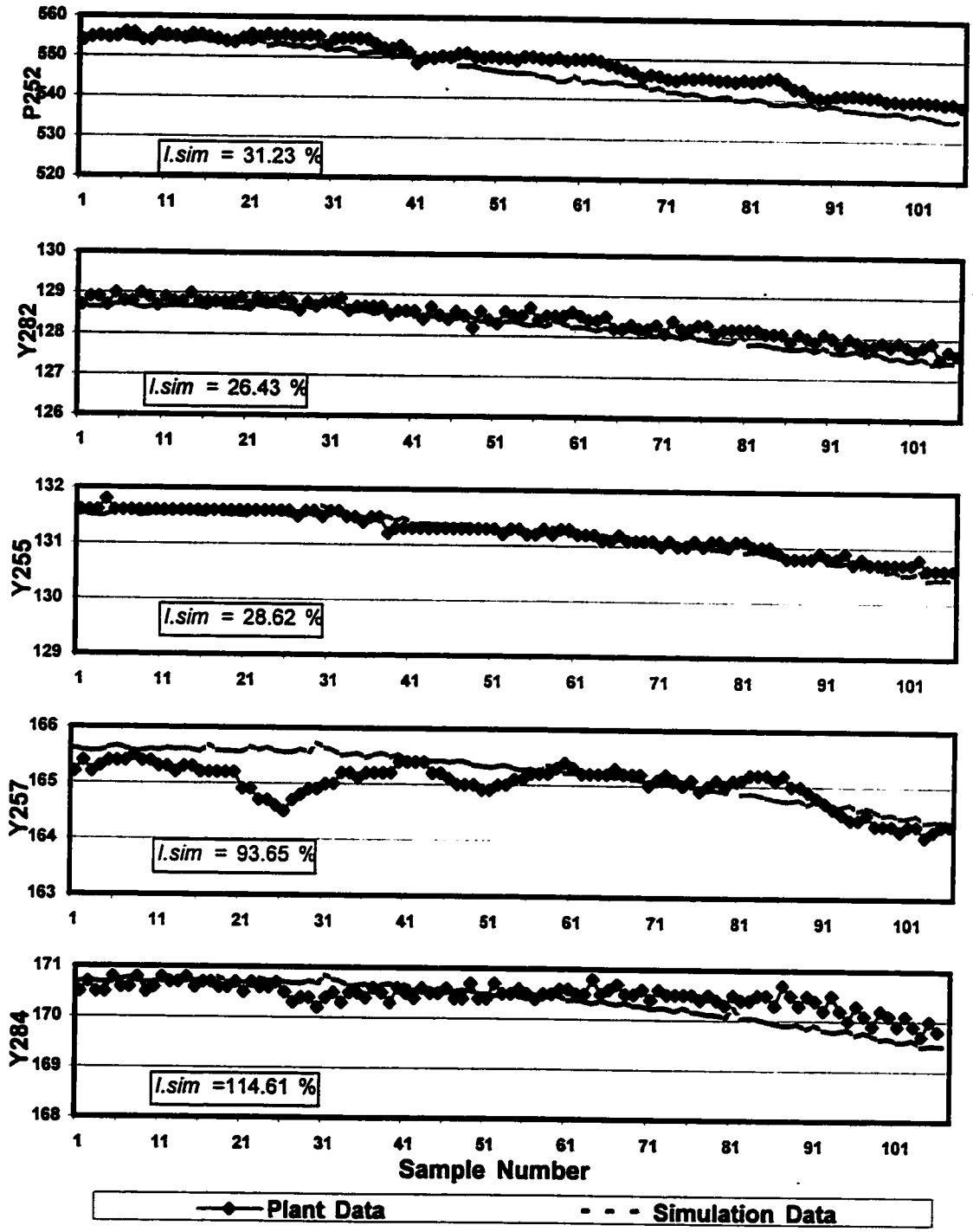


Figure 3.10 (a): Dynamic verification of the HYSYS model for TW252 condenser pressure and temperat using disturbance data as seen in dataset t.8

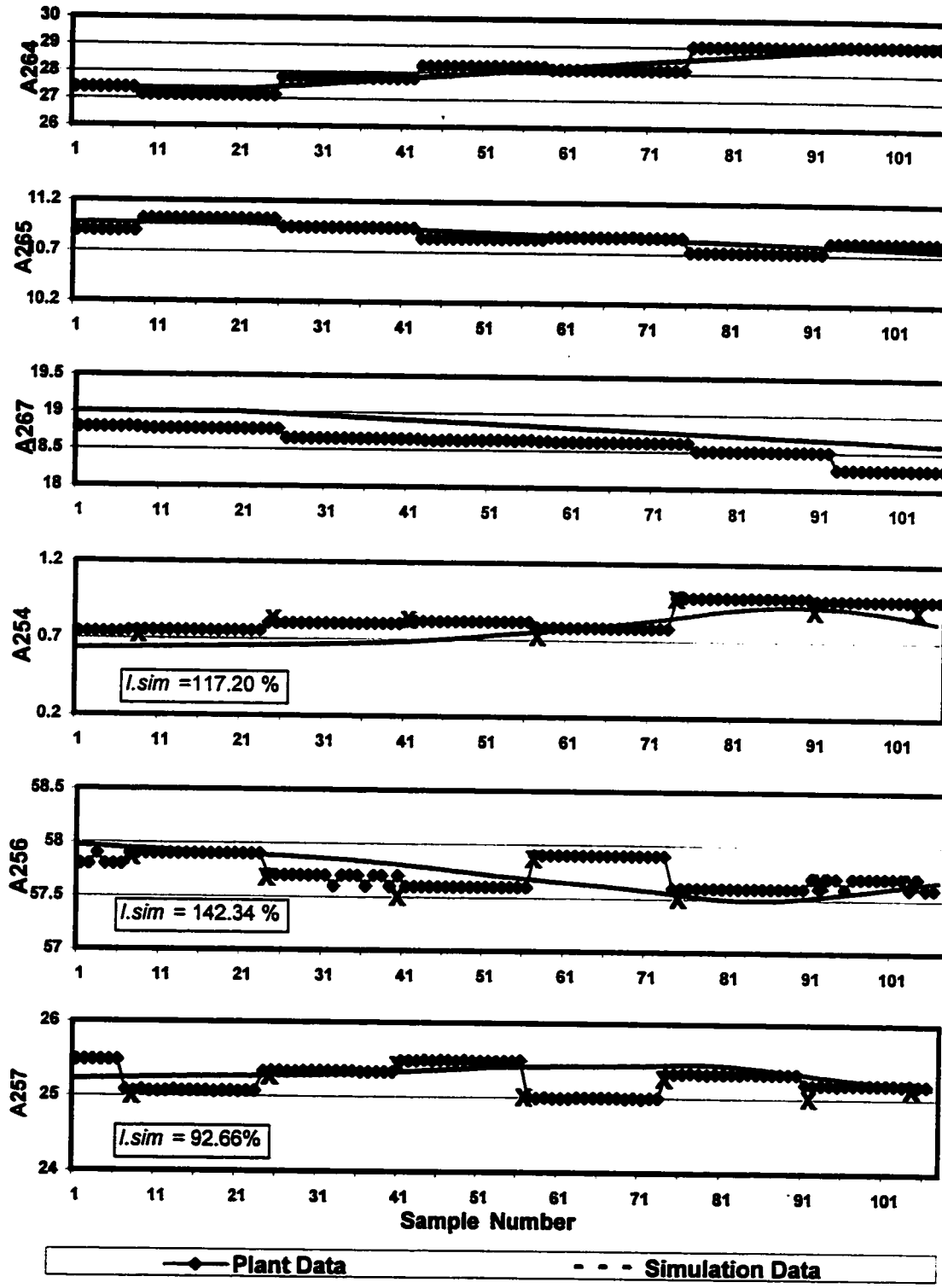


Figure 3.10 (b): Dynamic verification of the HYSYS model for TW252 feed and distillate compositions using plant disturbance data as seen in dataset t.8

3.8 Summary

In this chapter, a physically based model for TW252 is developed using HYSYS for both the steady-state and dynamic behavior. The HYSYS model is validated at steady state for various nominal conditions and used to quantitatively assess the non-linearity and interaction effects of TW252 using a steady state gain and RGA analysis, respectively. A dynamic validation of the HYSYS model is performed using plant step response tests, which included open loop, closed loop and random disturbance data. Discrepancies in the physically based model prediction of these tests were also discussed in detail. The most interesting observation in this chapter is the apparent ability of the physical model to capture dynamic changes in the distillate compositions not detected by infrequent sampling of the GC-based overhead composition analyzer. Furthermore, from the validation of the HYSYS model, benchmarks and upper limit values, based on the performance index (*I.sim*), are recommended for the prediction of TW252 variables using the developed HYSYS model.

Chapter 4

Artificial Neural Network Model Development

This chapter introduces artificial neural network (ANN) modeling as a process-modeling tool. In particular, its applications in the field of chemical process engineering modeling and more specifically distillation column modeling are briefly reviewed. A novel application of dynamic ANN modeling of an industrial distillation column is presented. The ANN models were generated using a non-linear ANN model predictive control tool known as Process Insights™ from Pavilion Technologies Inc. (Austin, Texas) to predict the following TW252 variables:

- The Distillate Compositions
- The Internal Column Tray Temperatures
- The Condenser and Bottom Levels
- The Internal Flow Rates

4.1 Artificial Neural Networks

An alternative approach to physical modeling is to identify a model directly from input/output data collected from a plant known as empirical modeling. Empirical modeling techniques can be linear or non-linear depending on their structure. Linear system identification has been extensively, and successfully, used in advanced control systems. However, most physical systems are highly non-linear, therefore control systems based on linear identification techniques may not provide the best results.

In recent years, researchers have focused their attention on the applicability of ANN technology to a vast range of complex and demanding real-world problems. Much

of the interest lies in the potential of ANN to solve problems that have been difficult, or even impossible to solve with more traditional approaches. In particular, neural nets have the ability to learn new relationships or patterns from various types of input/output data. ANN models involve specifying interconnected non-linear basis functions to represent the behavior of the system. The inputs and outputs can be quite general (e.g. patterns or attributes, rather than physical variables). Some successful applications of ANN are (Hunt et al, 1992):

1. Optical character recognition (e.g. handwriting, Chinese characters)
2. Speech recognition
3. Medical diagnosis
4. Analysis of sonar signals
5. Process Control

The key advantage of ANN is their ability to model any non-linear process by non-linear regression. ANN learn to recognize relationships between inputs and outputs by changing its internal structure of the net and its parameters. The internal structure is simple allowing for fast computational performance at each time step. ANN are an ideal complement or alternative to traditional first principle based process modeling.

In industry, there has been a tremendous amount of interest and speculation about applying ANN to advanced process control of chemical processes (Bhat and McAvoy, 1990). ANN's are used in a variety of control structures and applications as process models and/or controllers. ANN have been used to detect and diagnose faults (Hoskins and Himmelblau, 1988; Ungar et al, 1990), for control system design (Birky and McAvoy, 1989), to solve nonlinear optimization problems (Kennedy and Chua, 1988), to perform statistical quality control, recognize and forecast disturbances (Ungar et al, 1995), and to validate sensors (Keeler, 1993). Furthermore, recent studies (Bhat et al, 1990; Willis et al, 1990; Psichogios and Ungar, 1991) have demonstrated the flexibility of ANN technology for modeling chemical processes.

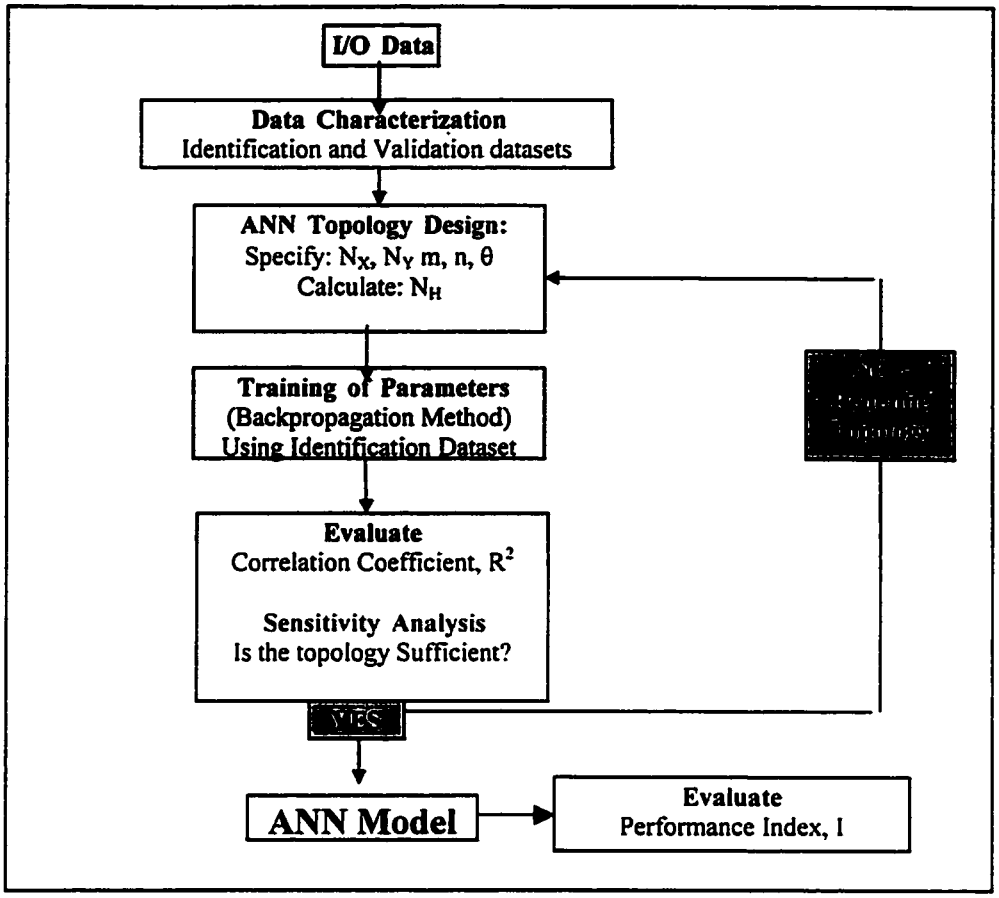


Figure 4.1: ANN model building algorithm

The optimum ANN models are determined using an iterative procedure that involve various steps as shown in Figure 4.1. The iterative algorithm presented in Figure 4.1 involves specifying the network topology or architecture, defining node (neuron) characteristics and deciding on an adequate training or learning method (Lippman, 1987). Next, the predictive capability of the model is evaluated on the identification and verification datasets to determine if the specified architecture is valid. A sensitivity analysis is performed to initially eliminate the unnecessary inputs that have no relationship to the selected outputs in order to reduce the number of parameters in the model. If the model is satisfactory, then the ANN model development is halted. If the model is unsatisfactory, then the ANN model topology is redesigned and the procedure is repeated. The redesigning step involves reducing or increasing the number of inputs to

the model and specifying the correct number of past inputs and outputs. Before an ANN model is trained and verified, adequate data that contains all relevant information about the dynamics of the process being modeled, must be preprocessed.

4.1.1 Data Preprocessing and Classification

Raw process data typically needs to be "preprocessed" before it can be used for modeling. Preprocessing involves the removal of outliers, elimination of data corresponding to process down-time (or other unusual operations such as maintenance periods, switch-over, etc.) or undesired outliers. Such manipulation of data may be done graphically or using advanced DSP (Digital Signal Processing) methods (Sabharwal et al, 1997). Another important aspect of data preprocessing is obtaining the correct distribution of data. Raw process data typically contains a large amount of data-points in a few operating regimes and very few data points in other regions of operation. In order to develop a good empirical process model that performs well over a wide range of operating conditions, it is necessary to have uniformly distributed data.

Data classification involves subdividing the input and output (I/O) data for identification and verification of a high fidelity ANN model. The identification process or learning involves partitioning the dataset into training and testing data patterns. A data pattern is a complete set of inputs and outputs. If there are time delays between the variables, the data pattern will contain values gathered from different rows in the dataset. During the training patterns, the ANN modifies the internal structure of its internal parameters based on the error between the actual output value from the dataset and the predicted output from the model (described in section 4.1.3). During the testing patterns, the ANN does not learn and only compares its output to the target output. The training and testing cycle is referred to as an epoch. A test set is randomly chosen to be 15 % of the original identification dataset. Therefore, the mean and standard deviation of the identification and test sets should be similar. Furthermore the test set must not contain points that that are outside the range of the training data. In addition to the required sets of testing and training patterns, a verification dataset is used for validation. The

verification dataset are patterns that are completely independent of the identification patterns. Note that, in general, the verification patterns usually do not contain points that are outside the range of the identification patterns as ANN models do not extrapolate, in general, at all!

4.1.2 ANN Topology Design

In an extensive literature review, Hecht-Neilson (1987b) reported that at least 50 different types of neural networks have been reviewed (Thibault and Grandjean, 1991). Below is a list of some of the main architectures mentioned in his paper with the appropriate references:

- Perceptron (Rosenblatt, 1958)
- Adaline (Widrow and Hoff, 1960)
- Hopfield Nets (Hopfield 1982, 1984)
- Feedforward or Backpropagation networks (Parker, 1982; Rummelhart et al. 1986; Werbos, 1974)
- Adaptative Resonance Theory (Carpenter and Grossberg, 1990)
- Counter-propagation networks (Hecht-Nielsen, 1987)
- Cognitron and Neowcognitron (Fukushima, 1990)
- Self Organizing Map (Kohonen, 1990)

In general, ANN consists of a large number of neurons arranged in layers. Each node is connected to other neurons by means of connection links, each with an associated weight. Neurons in the same layer behave in a similar manner. Within each layer a typical neuron takes in a set of inputs, sums them together, applies it to a non-linear function and passes the output signal forward through to another weighted connection to other neurons in the next layer. The neuron is a function of a nonlinear combination of predictor variables. The connection weights serve as adjustable parameters, which are determined by the training method. The arrangement of these neurons into layers and the connection patterns within the layers are known as the *ANN topology*. The neurons that

represent the inputs to the neural network form the input layer. The neurons that represent the outputs of the neural network form the output layer. The layers, which are independent of the input/output dataset, are known as the hidden layers because they are transparent to the user and problem specific. The convention used to name an ANN topology was developed by Bremmerman and Anderson (1989) that specify the number of layers and the number of neurons used in each layer. For example, an ANN architecture with three layers, representing; one input layer with three input neurons, one hidden layer with five neurons and an output layer with two neurons is referred to a **3-5-2** network.

In the literature, many different architectures are used, typically with hundreds or thousands of parameters (Ungar et al. 1990). The most widely used architectures are: the multilayer feedforward network more commonly known as the backpropagation network (Rummelhart et al, 1986) and external recurrent backpropagation networks. These two architectures are the focus of this section.

Feedforward Artificial Neural Nets (FANN)

Feedforward artificial neural network (FANN) are static ANN because the outputs of the network are not included as inputs to any of the networks nodes in the input layer. Figure 4.2 shows a FANN topology, that consists of nodes organized in layers of neurons (processing units) where neuron connections occur only between adjacent layers for a multiple input multiple output (MIMO) case.

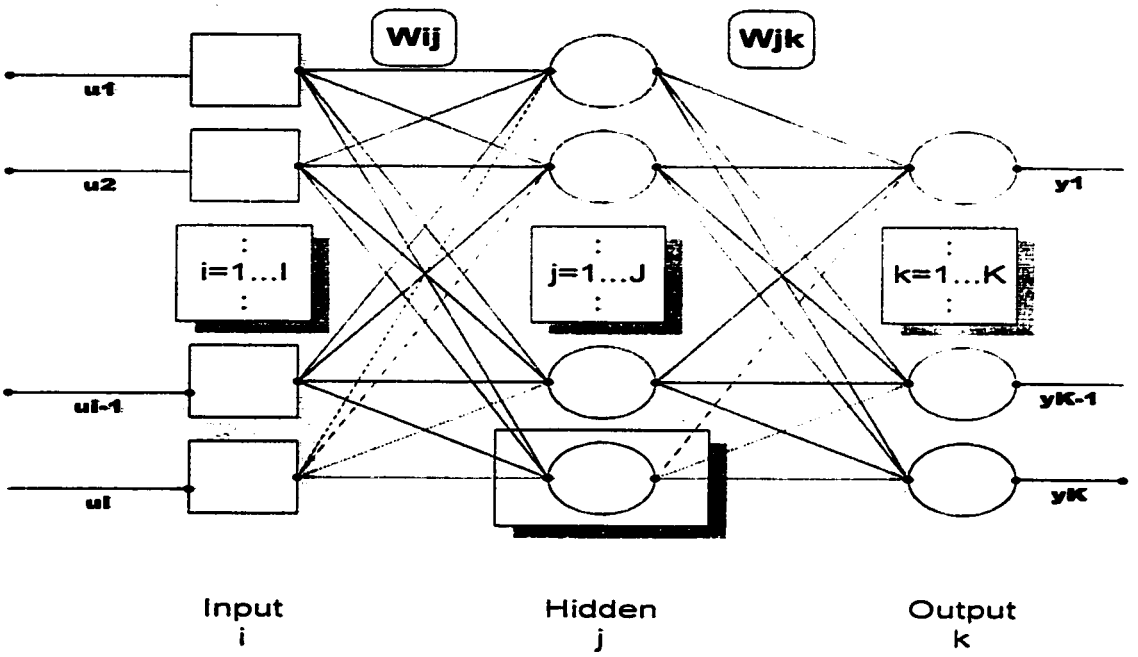


Figure 4.2: Flow through a static three layer FANN

The squares represented in the input layer are non-processing neurons indicating that each input is distributed to the neurons of the hidden layer by a connection weight (W_{ij}). The second and third layers of processing neurons (represented by circles) consist of two parts, a weight summation element and a non-linear transfer function as seen in Figure 4.3.

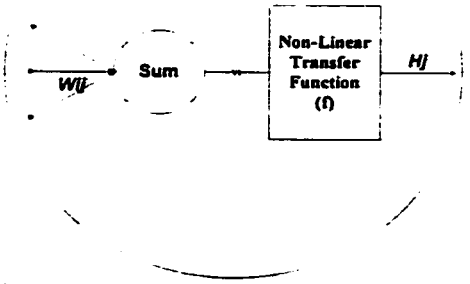


Figure 4.3: Basic model of a typical neuron

The processing neuron takes a weighted sum of all the inputs, then applies a transfer function, $f1$ (described later) to this sum:

Hidden Layer: $$H_j = f1\left[\sum_{i=1}^{N_X} W_{ij} U_i + U_o\right]$$ Equation 4.1

where:
- H_j = the j^{th} hidden node in the hidden layer
- U_i = the i^{th} input node in the input layer
- U_o = 0 or 1
- W_{ij} = the weight associated with the i^{th} input in the input layer connected to the j^{th} hidden node in the hidden layer.
- N_X = total number of inputs used in the topology
- $f1$ = non-linear transfer function

Expanding, we get:

$$H = f1\left[\begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1J} \\ W_{21} & W_{22} & \cdots & W_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ W_{I1} & W_{I2} & \cdots & W_{IJ} \end{bmatrix} x \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_I \end{bmatrix} + U_o\right]$$ Equation 4.2

The function ($f1$) can be linear or non-linear (see section 4.1.3). The outputs of the neurons in the hidden layer (Hj) are distributed to the neurons of the next layer (output layer) by another weight (W_{jk}). The weighted outputs are summed and another transfer function, $f2$, is applied:

Output Layer: $$S_k = f2\left[\sum_{j=1}^{N_H} W_{jk} H_j\right]$$ Equation 4.3

where:
- S_k = the k^{th} predicted output

W_{jk}	=	the weight associated with the j^{th} hidden node in the hidden layer connected to the k^{th} output in the hidden layer.
N_H	=	the total hidden nodes used in the topology
$f2$	=	non-linear transfer function

The values received by neurons S_k are different than H_j because each signal is scaled by its appropriate weight (W_{jk}). Figure 4.2 shows a single hidden layer but several hidden layers can be used. A neural network with a single hidden layer is sufficient to adequately represent the dynamics of a process (Cybenko, 1989; Hornik et al, 1989). Furthermore, the number of neurons in the hidden layer is problem dependent and must be determined in an iterative manner.

Dynamic vs. Static FANN

The backpropagation algorithm can be used to model dynamic systems by simply presenting past outputs of the network as inputs. The dynamic FANN, originally introduced by Rumelhart et al., (1986), is called "backpropagation in time" or known as "external recurrent backpropagation" (Hecht-Neilson, 1984). Past values are important because they include the transient nature of the data in the ANN topology, whereas a static model is only concerned with steady state data. The process model for a single input and single output (SISO) in its most general form is given as:

$$y(k) = f[y(k-1),\ y(k-2),\ldots,y(k-n),\ u(k-1-\theta),\ u(k-2-\theta),\ldots,\ u(k-m-\theta)]$$

Equation 4.4

where:	y(k)	=	is the process output at the k^{th} sample pattern
	u(k)	=	is the process input at k^{th} sample pattern
	m	=	is the number of past inputs
	n	=	is the number of past outputs
	θ	=	is the sample time delay between y and u
	f	=	specified functional relationship

A simple example of backpropagation in time is shown in Figure 4.4 for a SISO case. In this example, to predict the 6^{th} sample in a dataset with n = 3, m = 3, and $\theta = 2$, the above equation becomes:

$$y(6) = f[y(5), y(4), y(3), u(3), u(2), u(1)] \qquad \text{Equation 4.5}$$

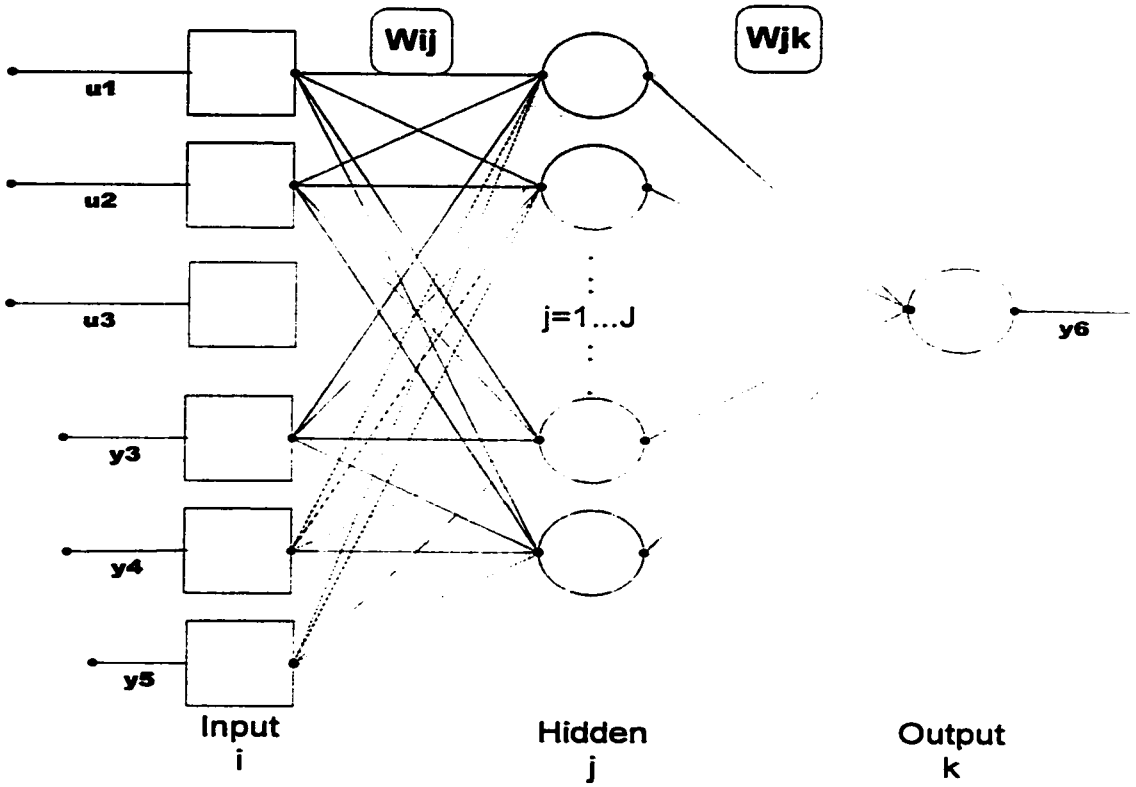


Figure 4.4: Three-layer external recurrent ANN for SISO case

Note that y(1), y(2), u(4), u(5) and u(6) are not used in the prediction of y(6). The methodology of formulating the input vector is known as the moving window method (Bhat et al., 1990). For more information on recurrent ANN the reader is referred to Levin and Narendra (1995) and Delgado et al. (1995).

Determining the ANN Topology Design Parameters

Once the system to be modeled is selected and the outputs determined the following ANN topology design parameters must be evaluated in order to define the optimum ANN model:

1. The optimum number of inputs used to accurately derive the output values is determined from a Sensitivity Analysis (which is described later).

2. The optimum number of past values for each input (m) and output (n) is determined using an iterative procedure that involves varying m and n and comparing the predictive power of many different topologies.

3. The apparent time delay (θ) between each input and output is determined from communication with the operators of the plant and analyzing the data collected.

Initially, a static topology is chosen with as many inputs as possible that may describe the system's output(s). The ANN is trained and a sensitivity analysis is used to filter out the unnecessary inputs. Then the optimum number of past inputs and outputs and apparent time delay are included in the ANN topology and re-trained using the identification dataset.

4.1.3 Training of ANN Parameters

From Figure 4.1, the next step in the building of an ANN model is the training or calculation of the internal parameters. An ANN must be trained using successive presentations of input-output data pairs. In essence, ANN is a parameter estimator that can simply be viewed as a large dimensional regression model which can be used to correlate a vector **U** of input variables to a vector **Y** of output target variables that can be approximated by a vector **S** (Thibault and Grandjean, 1991). Given input-output vectors **U-Y** (training data), the objective of ANN mapping is to obtain accurate values of the weights that connect the layers of the ANN together, know as connectionist weights. The connectionist weights represent the underlying non-linear relations between the **U-Y** pair.

Therefore, the backbone of ANN model is the learning or training process to obtain fitted parameters or connectionist weights.

The number of model parameters or connectionist weights in a feedforward neural network with one single hidden layer and a bias neuron, is given by:

$$N = N_X N_H + N_H N_Y + N_H \qquad \text{Equation 4.6}$$

where: N_X = number of inputs in the input layer

N_Y = number of outputs in the output layer

N_H = number of hidden neurons in the hidden layer

Therefore to develop an ANN process model, the topology of the network must be specified and the parameters must be evaluated and minimized for computational speed requirements.

Back Propagation Algorithm

One of the most popular learning algorithms is back propagation (Rumelhart et al., 1986). This algorithm is designed to minimize the squared error between the calculated vector **S** and expected vector **Y** of the outputs of the network.

$$E = \frac{1}{2}\sum_{k=1}^{K}(Y_k - S_k)^2 \qquad \text{Equation 4.7}$$

where: Y_k = actual output at k^{th} sample pattern

S_k = predicted output at k^{th} sample pattern

The initial weight values are usually set to small random values between [-1, 1] and the internal transfer functions are sometimes sinusoidal transfer functions. The training inputs are mean centered to clear the effects of variable units in the input layer. Then the inputs are propagated forward through the network and then the outputs are computed and the sum of the square error (E) is calculated using Equation 4.7.

After the error is calculated the connectionist weights must be modified. Using the same nomenclature as in Figure 4.1, the weights from the hidden to the output layer are computed and updated using the following backpropagation equation:

$$W_{jk}^{m} = W_{jk}^{m-1} - \varepsilon^{m}\left[\frac{\partial E}{\partial W_{jk}}\right] + \beta^{m}\left[W_{jk}^{m-1} - W_{jk}^{m-2}\right] \qquad \text{Equation 4.8}$$

Momentum Term
Forces the change of weight to proceed in the same direction as the previous change

Error Gradient Term (δ)
Fraction of error gradient that is backpropagated through the network

where:

m	=	iteration counter
ε^m	=	Tuning parameter - provides the step size during gradient descent (learning rate).
β^m	=	Tuning parameter - controls the convergence of the algorithm

Each output unit (Y_k) receives a target pattern corresponding to the input training pattern and computes its error gradient, δ, by taking the derivative of Equation 4.7 with respect to W_{jk} and substituting Equation 4.3 for S_k:

$$\frac{\partial E}{\partial W_{jk}} = \delta_k = \left[f'\left(\sum_{j=1}^{N_H} W_{jk} H_j\right)(S_k - Y_k)\right] H_j \qquad \text{Equation 4.9}$$

where: f' = first derivative of the linear transfer function $f1$

During training, each output unit computed, S_k, is compared to its target value Y_k to determine the associated error for that pattern. Based on this error, the factor δ_k is computed using Equation 4.9. The calculation of the updated weights from the input to the hidden layer is given as follows:

$$W_{ij}^{m} = W_{ij}^{m-1} - \varepsilon^{m}\left[\frac{\partial(E)}{\partial W_{ij}}\right] + \beta^{m}\left[W_{ij}^{m-1} - W_{ij}^{m-2}\right] \qquad \text{Equation 4.10}$$

where:

$$\left[\frac{\partial(E)}{\partial W_{ij}}\right] = f'\left(\sum_{i=1}^{N_I} W_{ij} Ui\right)\left[\sum_{k=1}^{N_H} f'\left(\sum_{j=1}^{N_J} W_{jk} H_j\right)(S_k - Y_k) W_{jk}^{m-1}\right] U_i$$

Equation 4.11

δ_k is used to distribute the error at output unit S_k back to all the units in the previous layer (hidden layer). Hence, the name, "backpropagation method". The above expression is simplified in a similar manner as before with the factor δ_j and is computed for each hidden unit:

$$\delta_j = f'\left(\sum_{i=1}^{N_I} W_{ij} U_i\right)\left[\sum_{k=1}^{N_J} \delta_k W_{jk}^{m-1}\right] \qquad \text{Equation 4.12}$$

Momentum is added to the back propagation learning by setting weight changes that are functions of the last weight changed and new change suggested by the back propagation rule. The magnitude of the effect that a last weight change is allowed to have is given by β^m. The momentum term accelerates the weight adjustments as long as the corrections are in the same general direction for several iterations. A smaller learning rate ε^m is used at the same time to prevent a large response to the error from any one training pattern. When momentum is used, the reduction in error proceeds in the

direction that involves a combination of the current gradient and the previous direction of weight corrections. Limiting cases are:

1. $\beta^m \rightarrow 0$; weight change is solely based on the gradient
2. $\beta^m \rightarrow 1$; weight change is heavily based on last weight change (momentum)

After all of the δ factors have been determined, the weights for all layers are adjusted simultaneously. The adjustment to the weight W_{jk} (from hidden unit H_j to output unit S_k) is based on the factor δ_k, the activation function (which will be described later) of the hidden unit H_j and the momentum term. The adjustment to the weight W_{ij} (from input unit U_i to hidden unit H_j) is based on the factor δ_j, the activation function of the input unit U_i and the momentum term again.

Activation Transfer functions

Activation transfer functions for a backpropagation nets are generally non-linear and must have several important characteristics: continuous; differentiable; monotonically non-decreasing. Linear functions are sometimes used in combination with non-linear transfer functions. The most common transfer function used in the ANN topology is the multi-layer Perceptron (Lippmann 1987). Typically the hidden layer consists non-linear transfer functions whereas the output layer uses linear transfer functions. Often the transfer function of choice for the hidden layer is a sigmoid or an S-shaped curve. Mathematically, this function is convenient because its derivatives are easy to calculate as shown in Equation 4.13.

$$f(z) = \frac{1}{1+e^{-z}}; \qquad f'(z) = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) \qquad \text{Equation 4.13}$$

4.1.4 Evaluation of ANN Performance

As described in the data classification in section 4.1.1, the data collected is classified as identification and verification datasets. The verification data are not used during the learning process. The identification data are used to identify the ANN model using a training/testing cycle known as an epoch. During the training pass of the epoch, the ANN modifies the internal parameters based on the error (Equation 4.7) between the actual output value and predicted output. During the testing pass of the epoch, the ANN is not allowed to learn but only compares its output (Y_k) to the target output (S_k).

Initially during training the discrepancy between the ANN model's predicted values and the datasets actual values is relatively high and the model can not derive output values based on the inputs. As training progresses, the ANN modifies its internal parameters minimizing the error (Equation 4.8) of the identification dataset for both the training and testing cycle until the relative error reaches its lowest value. The Relative Error (rel_err) for a single output in an entire dataset pattern is defined as:

$$rel\_err = \sqrt{\frac{\sum_{k=1}^{P}(Y_k - S_k)^2}{P \times \sigma^2_{out}}} \qquad \text{Equation 4.14}$$

where: k = is the sample index in the dataset pattern

P = is the total number of patterns in the dataset

σ^2_{out} = the standard deviation of the actual values of the output from the total number of patterns in the dataset

Note that Relative Error (rel_err) is not as commonly used as the statistical measure, R^2. R^2 is known as the correlation coefficient and its relationship to the relative error is:

$$R^2 = (1 - rel\_err^2)$$ Equation 4.15

In general, $R^2 = 1$ is considered to be optimal however there really is no definition of what a reasonable R^2 value is, for a specific ANN topology. However,. As the identification continues the ANN memorizes or "over-trains" the training data and the Relative Error will continue to decrease but also reduces the model's ability to generalize new data. For example, the Relative Error for the test data will tend to increase if the model tends to be over-trained. Hence, the highest R^2 values tend to be at the point where the test Relative Error reaches a minimum. Once the highest R^2 is obtained for a certain topology, the procedure is repeated for various ANN topology structures. The optimum ANN model for a process, in compliance with the algorithm in Figure 4.1, is determined by comparing R^2 values of the various topologies. The selected ANN topology for the model of the process is based on the following criterion:

- R^2 must be as close to 1.0 as possible, but no less than 0.850 by utilizing the minimum number of internal parameters, N (Equation 4.6).

The success of training not only depends on minimizing the error (Equation 4.7) of the training data set, but also on the validation of the ANN model with an independent verification dataset. The criterion above is extended to include verification datasets.

Sensitivity versus Rank analysis calculates the sensitivity of output variables to input variables (that is the effect or influence the inputs have on the outputs) over the patterns in the Dataset, and, for each output ranks the inputs in order of sensitivity. The sensitivity analysis is incorporated in the iterative procedure, shown in Figure 4.1, to determine the least amount of inputs required to generalize the outputs. Sensitivity vs. Rank analysis eliminates inputs that have negligible sensitivity from the model. Hence, the number of training parameters will decrease, the computational load will decrease and increases in the predictive speed of the ANN model will be occur. The three types of sensitivity measures used are:

- Average Absolute Sensitivity - average sum of the absolute values of the partial derivatives of the input-output pairs (Process Insights Reference Manual, 1996).

$$AverageAbsolute = \frac{\left|\sum_{k=1}^{P} \frac{\partial Y_{k,:}}{\partial X_{k,:}}\right|}{P} \qquad \text{Equation 4.16}$$

where: P = the number of patterns in the Dataset

$Y_{k,i}$ = the i^{th} output for the k^{th} sample pattern

$X_{k,i}$ = the i^{th} input for the k^{th} sample pattern

- Average Sensitivity - actual average of the partial derivatives in Equation 4.16.
- Peak Sensitivity - the maximum of the partial derivatives over all the patterns in the Dataset.

Another coefficient used to evaluate the performance of ANN model is known as the performance index (which was introduced in Chapter 3), *I*, and is given by the following equation:

$$I = \frac{\sum_{k=1}^{P}(Y_{k,i} - S_{k,i})^2}{\sum_{k=1}^{P}(Y_{k,i} - \overline{Y}_i)^2} \times 100\% \qquad \text{Equation 4.17}$$

where: $Y_{k,i}$ = actual i^{th} output at k^{th} sample pattern

$S_{k,i}$ = predicted i^{th} output at k^{th} sample pattern

$\overline{Y}_i$ = mean of the i^{th} output for all the patterns (P) in the dataset

There is no general minimum threshold used to determine if the performance index (I) is satisfactory. I is only used as a relative comparison to gauge the predicted power of an ANN relative to other ANN's and in particular to the HYSYS model predictions in Chapter 3. However, based on the work by Bomberger and Seborg (1997), who utilize the same performance index to evaluate radial basis function network models, it is inferred that I values of less than 10% are good values.

4.2 TW252 Application

In this section, the following ANN models are developed for TW252 with a commercially available software package know as Process Insights™, from Pavilion Technologies:

- Internal Tray Temperature Model (Temp Model)
- Distillate Composition Model (Comp Model)
- Internal Flow Model (Rev Model)
- Level Model (Lev Model)

Process Insights™, like many other commercially available applications, incorporates a similar algorithm to the one introduced in section 4.1, which is summarized in Figure 4.5.

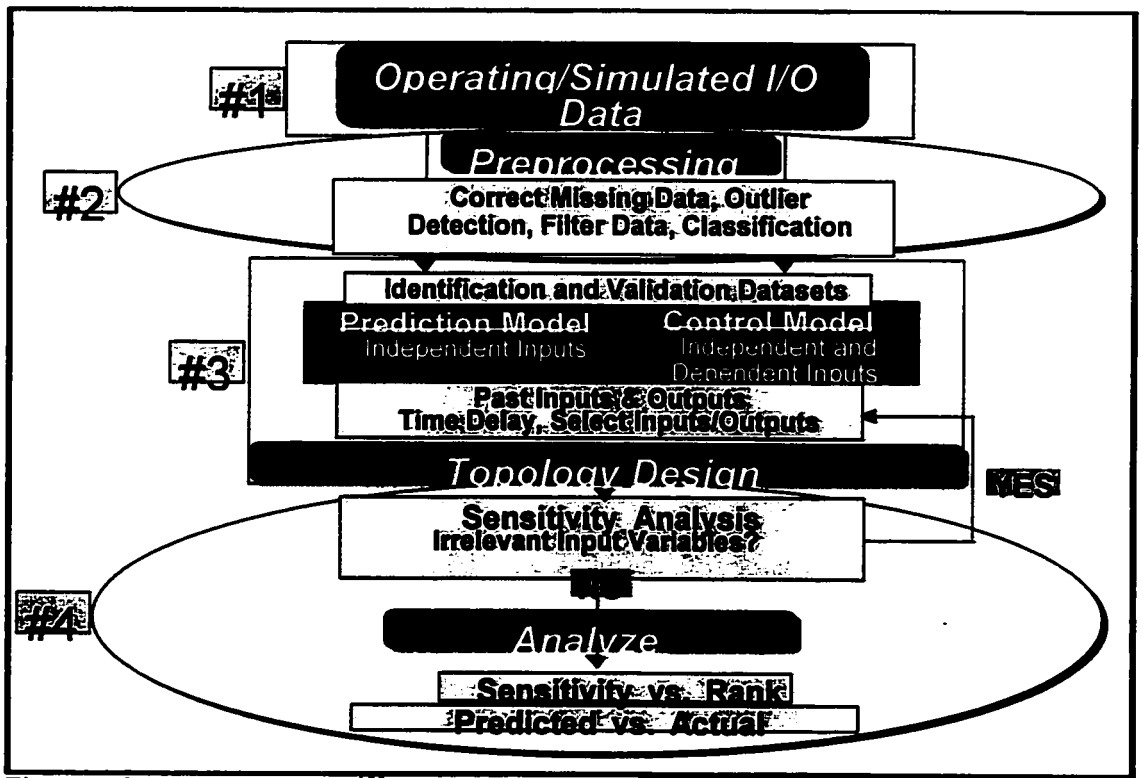


Figure 4.5: Process Insights™ model building steps

4.2.1 Data Classification for TW252

As shown in Chapter 2 the plant data collected for TW252 contained open loop, closed loop setpoint tests and disturbance tests. To train an ANN, the training signal should contain all relevant information about the dynamics of the controlled system; however, no general specification is available in the literature for selecting an optimal training signal. Furthermore, care must be taken with the type of data to be used for identification and validation. The plant data collected in Chapter 2 and tabulated in Table 2.2 is classified using the following nomenclature, where:

t	=	All available setpoint datasets
tOL	=	Open loop setpoint datasets: Tests 4 (t.4), 5 (t.5), 10 (t.10)
tCL	=	Closed loop setpoint datasets: Tests 1 (t.1). 2 (t.2). 3 (t.3). 9 (t.9)
d	=	All disturbance datasets
d8	=	Varying feed composition dataset: Test 8 (t.8)
n	=	Nominal process data (N.1, N.2, N.3)

Using these classifications (x) the data can be subcategorized into identification (I.x) and validation (V.x) datasets.

4.2.2 Topology Design Parameters

Before the ANN technology can be used to model TW252, the topology of the models must be defined. The design of the four distinct ANN topologies that represent TW252 must minimize the number inputs needed to define a relationship to the outputs of the models. In turn, optimizing the inputs used to define the models will decrease the computational load of the ANN by minimizing the number of weights needed between the inputs and the outputs. Furthermore, for dynamic modeling the number of past values used in the recurrent ANN structure must be defined.

Optimum number of inputs

To illustrate the method used to determine the optimum number of inputs, consider the internal tray temperature model (Temp model). The objective is to predict five outputs: Y282, Y255, Y297, Y257, and Y284. From the 23 measurable tags (Table 2.1) available for TW252, nine possible inputs that could be used to map a relationship to the outputs were chosen to initialize the Temp model. These inputs are shown in Table 4.1 and consist of flowrates, feed compositions, and pressure tags. Using the iterative procedure outlined in Figure 4.1, a sensitivity analysis was used to filter the inputs. The Temp model was identified using the nominal dataset (I.n) and the average absolute sensitivities of each output to a change input were evaluated and ranked and shown in Table 4.1.

Table 4.1: Sensitivity analysis ranking of the inputs based on the average absolute sensitivity of each input to the outputs

Rank	Input	Avg. Abs	Avg	Peak
Output: Y282				
1	P252	0.144	0.144	0.289
2	Y280	0.050	0.05	0.097
3	F256	0.019	-0.019	0.052
4	F267	0.018	-0.018	0.051
5	F284	0.016	0.016	0.023
6	A266	0.010	-0.01	0.029
7	A264	0.008	-0.008	0.019
8	A265	0.002	0	0.006
9	A267	0	0	0.001
Output: Y255				
1	P252	0.162	0.162	0.274
2	Y280	0.06	0.06	0.099
3	F284	0.021	0.021	0.027
4	F267	0.012	-0.011	0.031
5	F256	0.012	-0.011	0.03
6	A266	0.008	-0.008	0.021
7	A264	0.007	-0.007	0.014
8	A265	0.001	0	0.005
9	A267	0	0	0.002
Output: Y297				
1	P252	0.163	0.163	0.249
2	Y280	0.074	0.074	0.112
3	F284	0.03	0.03	0.043
4	F256	0.012	-0.006	0.016
5	A266	0.011	-0.011	0.02
6	F267	0.011	-0.004	0.019
7	A265	0.003	0.001	0.005
8	A264	0.002	0.002	0.005
9	A267	0.001	0.001	0.002
Output: Y257				
1	P252	0.102	0.102	0.149
2	Y280	0.052	0.052	0.078
3	F267	0.045	0.045	0.116
4	F284	0.041	0.041	0.084
5	F256	0.034	0.034	0.11
6	A266	0.016	0.016	0.022
7	A265	0.01	-0.01	0.031
8	A267	0.004	0.003	0.008
9	A264	0.003	0.001	0.008
Output: Y284				
1	P252	0.068	0.068	0.097
2	F256	0.042	-0.039	0.071
3	Y280	0.034	0.034	0.055
4	F267	0.026	-0.022	0.048
5	F284	0.019	0.019	0.028
6	A265	0.009	0.008	0.014
7	A264	0.003	-0.002	0.009
8	A266	0.002	0.001	0.007
9	A267	0.001	0	0.003

The number of inputs used to represent the five outputs is evaluated according to how many times the inputs are ranked in the top five. Based on the rankings the inputs effect on each output are evaluated using a point system as follows:

- 10 points for each input ranked number 1
- 7 points for each input ranked number 2
- 5 points for each input ranged number 3
- 3 points for each input ranked number 4
- 1 point for each input ranked number 5

Using the above system the inputs are eliminated and the top five variables used to develop the Temp model for the internal tray temperatures as shown in Table 4.2.

Table 4.2: The inputs selected to predict the temp model outputs

Rank	Variable	Points
1	P252	50
2	Y280	33
3	F267	20
4	F256	17
5	F284	15

The procedure to determine the optimum number of inputs is repeated for the Comp, Rev and Lev models using dataset I.n. The selected inputs with the corresponding outputs are summarized in Table 4.3.

Table 4.3: Summary of the selected inputs and outputs for the ANN models

Model	Possible Initial Inputs	Selected Inputs Based on Sensitivity Analysis	Selected Inputs	Selected Outputs
Temp	P252, Y280, F284, F256, F267, A267, A266, A265, A264,	P252, Y280, F267, F256, F284	P252, Y280, F267, F256, F284	Y282, Y255, Y297, Y257, Y284
Comp	P252, Y282, Y255, Y257, Y284, Y280, F284, F256, F267, A264, A266, A267	P252, A266, Y257, F267, A264, Y255, F256	P252, F256, F267, A264, A265, A267	A254, A256, A257
Rev	F284, P252, Y282, Y297, Y255, Y257,Y284,A264, A265, A267, L253, L254	Y257, Y284, F284, Y282, P252, Y255	Y284, Y255, P252	F256, F267
Lev	P252, Y282, Y255, Y257, Y284, Y280, F284, F256, F267, F271, F285, A264, A266, A267	F271, F285, F284	F271, F285, F284	L254, L253

The Comp, Rev and Lev model sensitivity results are shown in Appendix B. Note for the Comp and Rev models a different set of inputs were chosen then recommended by the absolute sensitivity results from Process Insights®. For example, the Comp model the inputs 27th and 1st tray temperatures (i.e. Y255 and Y257) are highly correlated with the other selected inputs, reflux flow rate (F256) and steam flow rate (F267). Hence, to avoid redundant inputs and to eliminate the temperature dependencies, it was decided that only the manipulated variables (F256, F267), the feed compositions (A264, A266, and A267) and the condenser pressure (P252) were chosen to represent the outputs (A254, A256, and A257) of the Comp model. The Comp model's topology is similar to other published architectures (Rhinehart, 1997). For the Rev model, the five inputs selected by Process Insights were further filtered to three inputs. The objective of the Rev model was to build an inverse relationship between the controlled variables (i.e. the tray

temperatures) and the two main manipulated variables (F267 and F256). Process Insights selected four tray temperatures (Y282, Y255, Y257, and Y284) as inputs to the model. Since, only two temperatures are used to control TW252's temperature profile by manipulating two variables (F256, F267), only two temperatures were selected. Y255 was the obvious choice for the top because it is the only temperature that was sometimes controlled with F256 during normal operation. For the bottom of the column there were two possible temperatures Y257 and Y284. It was decided that since Y284 was the current controlled variable that is regulated by the steam flow rate and hence determines the bottoms temperature profile, that only Y284 would be used in the model. Furthermore, F284 was not used any of the models.

Optimum number of past values

In conjunction with determining the optimum number of inputs, the number of past values used to represent the dynamic nature of the recurrent ANN was also determined. Again to illustrate the procedure used to determine the number of optimum past values, the Temp model is considered. To keep the iterative nature of this procedure simple, the number of past values for both the inputs (m) and outputs (n) were kept the same (m=n). Therefore, the number of past inputs and past outputs to be included as additional inputs to the dynamic Temp model were evaluated using six case studies as shown in Table 4.4.

Table 4.4: Various ANN topologies tested for ANN development of the Temp model

Case	Model	m	n	N_h	N	θ	Topology
1	Static	0	0	5	55	0	5-5-5
2	Dynamic	1	1	10	210	0	15-10-5
3	Dynamic	3	3	17	697	0	35-17-5
4	Dynamic	5	5	20	1220	0	55-20-5
5	Dynamic	7	7	22	1782	0	75-22-5
6	Dynamic	9	9	24	2424	0	95-24-5

From Table 4.4, cases 2 to 6 are dynamic ANN architectures and case 1 is for the static situation. For example, case 2, the input vector (U) contains one past value of the process outputs ($Y282_{t-1}$, $Y297_{t-1}$, $Y255_{t-1}$, $Y257_{t-1}$, $Y284_{t-1}$), manipulated variables ($F256_{t-1}$, $F267_{t-1}$), feed disturbances ($F284_{t-1}$, $Y280_{t-1}$), controlled variable ($P252_{t-1}$) and the current value of all the input variables ($P252_t$, $Y280_t$, $F267_t$, $F256_t$, $F284_t$) totaling 15 inputs to the neural network. The output vector (Y) of the Temp ANN model had five outputs ($Y282_t$, $Y255_t$, $Y297_t$, $Y257_t$, $Y284_t$). From these input/output configurations, the number of hidden nodes was internally determined using a Process Insights™ proprietary method, yielding 10 hidden nodes and a 15-10-5 network. Thus from Equation 4.7 there were 210 adjustable parameters (weights) that were trained using the back propagation method. The apparent time delays (θ) allows the ANN to model a process with significant time delays between the process variables. The time delay adjusts the temporal relationship of the process variable with respect to the other variables. A positive θ advanced the variable in time; a negative θ value moves it back in time. One time delay unit represents the time interval between the rows in the dataset (i.e. sampling period is three minutes for all the plant data collected for TW252). From discussions with plant operators, the inputs and outputs chosen for the Temp model had no significant time delays between the process variables as shown in Table 4.4 for all the cases. The different topologies for the Temp model, represented by the cases, are identified using a combination of Tests 1 and 4 (I.t1t4). Also, to aid in the selection, the test cases are validated using other independent datasets. The results are tabulated in Table 4.5.

Table 4.5: Selection of the appropriate ANN topology based on the sum of the squared errors for the identification dataset and for various validation datasets

Case	I.t1t4 R^2	V.t2 R^2	V.t3 R^2	V.t5 R^2	V.t8 R^2
1	0.545	0.382	0.264	0.000	0.000
2	0.886	0.892	0.910	0.889	0.833
3	0.890	0.891	0.911	0.887	0.854
4	0.894	0.887	0.910	0.886	0.801
5	0.895	0.889	0.910	0.902	0.790
6	0.895	0.886	0.909	0.885	0.775

In theory as the number of adjustable parameters increase the correlation coefficients, R^2, should also increase exponentially till a maximum is reached (Pavilion Reference Manual, 1997). From Table 4.5, the maximum R^2 obtained for dataset I.t1t4 was case 5, which corresponded to having 1782 calculated parameters as shown in Table 4.4. However, when the cases were tested with various validation datasets the architecture with a R^2 value consistently above the 0.850 criteria only occurred for case 3, with 697 adjustable parameters. For the subsequent cases, introducing more adjustable parameters did not enhance the prediction capabilities of the Temp model on the identification dataset drastically and for some of the cases the R^2 for the validation datasets decreased. Hence, a dynamic ANN topology of m=3 and n=3 was selected as the most appropriate ANN representation of the internal tray temperatures of TW252. In the literature, the use of 3 past values is consistent with the work of Bhat et al. (1990) and Cheng et al. (1995).

A similar analysis was done for the other ANN models and the selected architectures are shown in Table 4.6.

Table 4.6: The selected topologies for the other TW252 ANN models

Model	Model	m	n	N_h	Inputs	θ	Outputs	Topology
Temp	Dynamic	3	3	17	P252	0	Y282	35-17-5
					Y280	0	Y255	
					F256	0	Y297	
					F267	0	Y257	
					F284	0	Y284	
Comp	Dynamic	2	2	13	P252	-20	A254	24-13-3
					F256	-20	A256	
					F267	-20	A257	
					A264	-40		
					A265	-40		
					A266	-40		
Rev	Dynamic	3	3	10	P252	0	F256	18-10-2
					Y255	0	F267	
					Y284	0		
Level	Dynamic	3	3	10	F257	0	L254	18-10-2
					F271	0	L253	
					F284	0		

Note, that for the Comp Model in Table 4.6, the apparent time delays between the inputs and the three outputs were approximately 60 minutes or 20 sampling intervals for P252, F256, F267 and 40 sampling intervals for the feed compositions. The large apparent time delay is due to the sampling location of the composition analyzer. The same analyzer is used to measure the feed and distillate compositions, hence the time delay associated between the inputs and outputs incorporates both inherent and processing lags. For the Rev and Lev models, the apparent time delay between the each of the inputs and all of the outputs were considered to be negligible.

4.2.3 Analyze

The optimum Temp, Comp, Rev and Lev ANN topologies of TW252 were identified using dataset I.t (Temp.t, Comp.t, Rev.t and Lev.t) and validated using plant setpoint step tests; V.tOL, V.tCL, (subsets of I.t), V.n, and V.t8. The results are investigated using the correlation coefficient R^2 and tabulated in Table 4.7.

Table 4.7: Comparison of ANN models identified on setpoint response data using the correlation coefficient

Model	ANN Architecture	Output	Set I.t	Set V.tOL	Set V.tCL	Set V.n	Set V.t8
			R^2	R^2	R^2	R^2	R^2
Temp.t	35-17-5	Y282	0.910	0.994	0.976	0.891	0.951
		Y255	0.908	0.997	0.972	0.885	0.933
		Y297	0.915	0.982	0.969	0.914	0.943
		Y257	0.994	0.985	0.990	0.932	0.989
		Y284	0.997	0.986	0.985	0.923	0.985
Comp.t	24-13-3	A254	0.955	0.950	0.913	0.978	0.935
		A256	0.963	0.944	0.873	0.967	0.705
		A257	0.962	0.941	0.827	0.972	0.819
Rev.t	18-10-2	F256	0.945	0.991	0.904	0.231	0.449
		F267	0.979	0.958	0.976	0.723	0.922
Lev.t	18-10-2	L254	0.992	0.994	0.965	-	-
		L253	0.966	0.975	0.943	-	-

From Table 4.7 the selected optimum topologies for the four ANN models were able to predict the outputs quite accurately based on R^2, for identification dataset I.t. The ANN models trained on I.t will be generalized as ANN.t models (Temp.t, Comp.t, Rev.t, Lev.t). A plot of dataset I.t for all the measured TW252 variables in the dataset along with their statistics can be found in Figure E.1(a) through (d) and Table E.1, Appendix F.

In general, the verification results of the ANN.t models validated using datasets, V.tOL, V.tCL (subsets of I.t), V.n, and V.t8, were good for most of the models because the correlation coefficients were well above our criterion of 0.850. The only noticeable difficulty occurred for the Rev.t model where the R^2 values for the prediction of the outputs (F256 and F267) were very low for validation dataset V.n. The low values indicate that the Rev.t model was not able to represent the input/output relationship of dataset V.n for those particular inputs and outputs. The result is similar for the Rev.t model prediction of V.t8, where again the R^2 value is quite low for F256. On the other hand the Rev.t model was able predict F267 quite well. In examining the Temp.t model prediction behavior of the verification datasets, some generalizations can be made. First, the Temp.t model was able to capture the input/output relationship of the verification datasets quite accurately. For the validation dataset, V.tOL, the R^2 values of Table 4.7 show that the Temp.t model predicted the top tray temperatures (Y282, Y255) slightly better than the bottom tray temperatures (Y297, Y257 and Y284). The R^2 values of Table 4.7 for dataset V.tCL show that the Temp.t model accurately predicts the tray temperatures near the bottom of the column, Y257 and Y294 this time around. However, near the feed stage (Y297) and at the top of the column, Y282 and Y255, the accuracy of the model diminishes slightly. The Temp.t model is expected to give better prediction of dataset V.OL than V.CL for outputs Y282 and Y255 because the variances are larger due to the larger gains in the open loop responses of dataset V.OL. The opposite can be said for the bottom temperatures where the Temp.t model predicted better R^2 values due to the closed loop gains for Y257 and Y284 are larger for V.CL than V.OL. For the Comp.t model, the prediction of the independent verification datasets tended to decrease for dataset V.n., however, the R^2 values for dataset V.t8 were very reasonable.

4.3 Comparison of ANN and HYSYS Model Predictions

The ANN.t models and the HYSYS model prediction of plant open loop, closed loop and disturbance data are compared using performance index, *I*. The performance index of the HYSYS model is donated by *I.sim*, whereas, the performance index of the ANN.t models will be donated by *I.ann*. In general, the performance indices represent a model's ability to predict a particular plant output. The index contains a normalization factor so that a relative comparison can be made amongst all the predicted outputs. However, the magnitude of an output variable's deviation from the mean of the output in the dataset is reflected in the nature of the index. The index may get excessively large if the variance of the output variable from the mean of the dataset is small.

4.3.1 Comparison using Open Loop Plant Data

Figures 4.6 (a) and (b) show the comparison of the Temp.t, Comp.t, and Rev.t models to the HYSYS model predictions of a snapshot of the 480 available samples of plant dataset, t.4. To recap, dataset t.4 has a series of manipulated variable setpoint changes in F256 (+5.0 m^3/h and -2.5 m^3/h) from its nominal value of 41.2 m^3/h. Both Y255 and Y284 control loops were opened allowing for identification of the open loop behavior of all the internal tray temperatures in TW252. From Figure 4.7 (a), the *I.ann* and *I.sim* values show that the Temp.t model predicted the internal tray temperatures of the plant better than the HYSYS model for dataset t.4. However, the HYSYS model predictions of the internal tray temperatures were well within the benchmark value of 35%. The HYSYS model prediction of Y255 was the only variable near the *I.sim* upper limit value of 53% for the internal tray temperatures. In fact the HYSYS model prediction of the temperatures near the bottom of the column (Y257 and Y284) were similar to the Temp.t model predictions. A similar trend is shown in Figure 4.6 (b), where much lower performance indices (*I.ann*) occurred for the Comp.t (1.07 – 1.43%) and Rev.t models than the HYSYS model. Note, during the identification of the HYSYS model, only a few initial design condition of dataset t.4 were used to initialize the model,

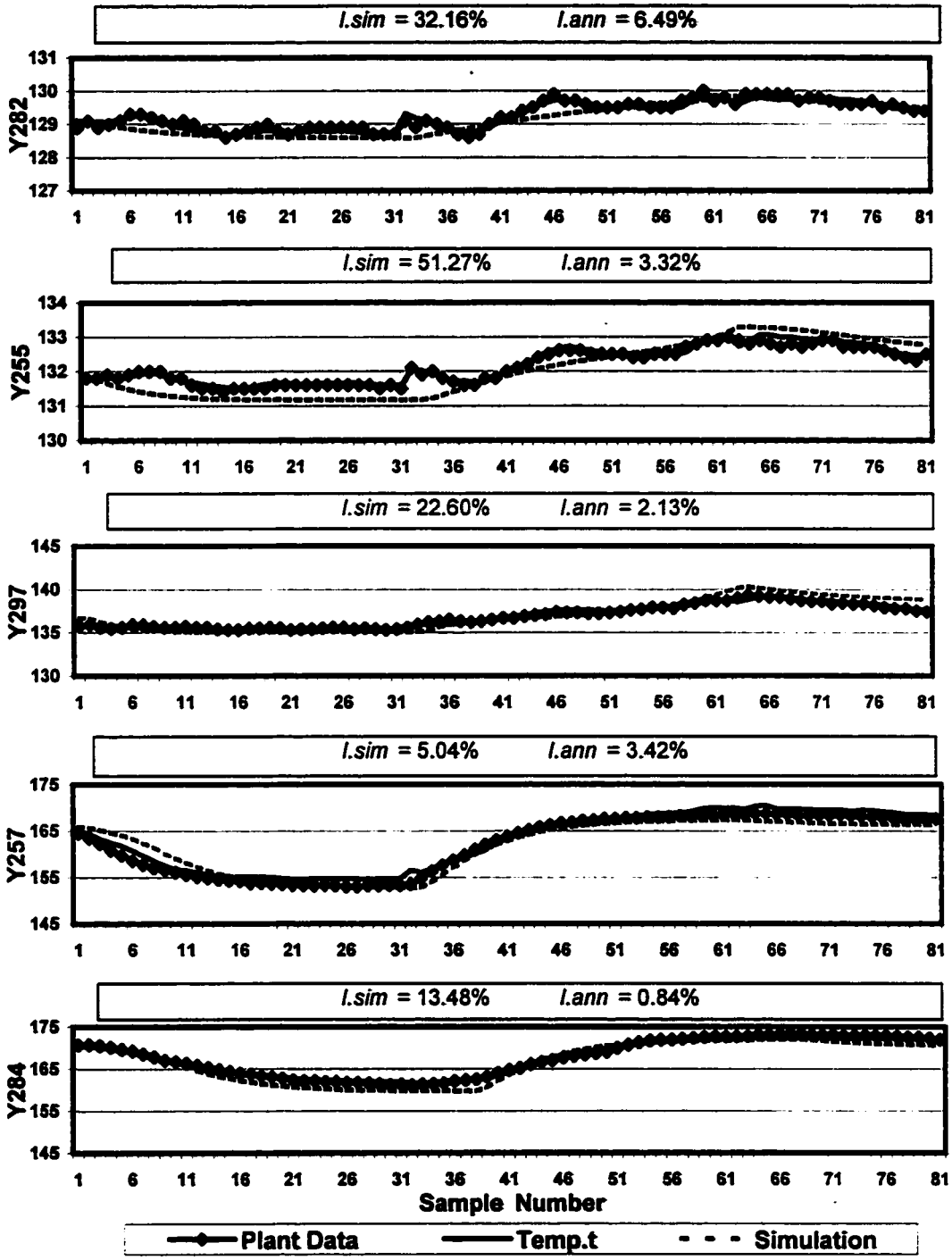


Figure 4.6(a): Comparison of the Temp.t and the HYSYS model prediction of TW252 internal temperatures for open loop dataset, t.4

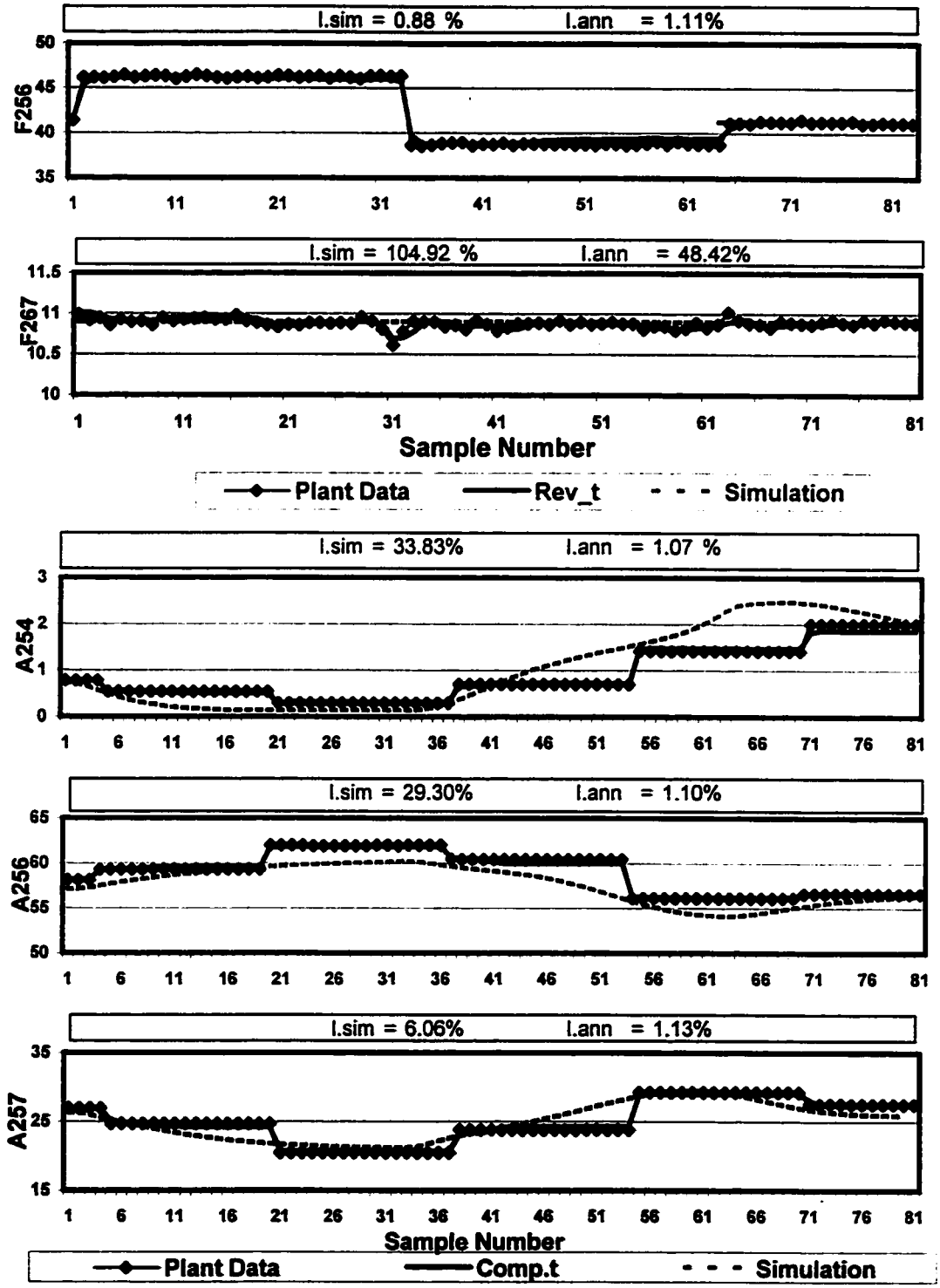


Figure 4.6(b): Comparison of the Comp.t, Rev.t and the HYSYS model prediction of TW252 distillate compositions and internal flowrates for the open loop dataset, t.4

hence, the HYSYS model actually predicted the response of t.4 without being identified on the plant data. In general, Figures 4.6 (a) and (b) show that the HYSYS model captured the dynamic response of TW252 for dataset t.4 quite effectively. Furthermore, the basic "trends" of the plant data were predicted quite well based on the satisfactory *I.sim* values for most of the variables, which were well within the recommended benchmark and upper limit values in Chapter 3. In contrast, the ANN.t models were identified using the step response plant data, therefore based on *I.ann* values, the ANN.t models should be able to predict dataset t.4 much more accurately than the HYSYS model.

Based just on the performance indices (*I.sim and I.ann*), "more accurately" or "better", sometimes does not guarantee that the prediction is right. In Chapter 3, it was concluded that the on-line analyzer's low frequency sampling rate did not capture the dynamic behavior of the distillate compositions. Hence, the recorded plant data for dataset t.4 did not record the distillate compositions actual dynamic response as was shown in Chapter 3. From Figure 4.6 (b) the Comp.t model accurately predicted, for example, A254 with an *I.ann* = 1.07 % in comparison to the HYSYS model prediction of *I.sim* = 33.83% at the composition analyzer points. The accuracy of the ANN to predict the plant data cannot be denied; however, the true nature of A254 has definitely been neglected. The ANN is only as good as the data used to train it, therefore its limited knowledge of the process (i.e. limitation of the DCS measurement device) restricts its ability to extrapolate. Hence, the low frequency sampling of the data hinders the ability of the ANN.t models to predict beyond regions that the model was intended to predict. On the other hand, the HYSYS model accurately captured the true dynamic behavior of A254 at a much higher frequency (i.e., in this case at the sampling interval of 3 minutes) for dataset t.4. The Rev.t model predicted F256 very accurately with an *I.ann* = 1.07%, however, predicted F267 less accurately with an *I.ann* = 48.42%. Furthermore, the *I.sim* values for the HYSYS model were similar to the *I.ann* values with values for F256, which were well within the benchmark of 50 % (*I.sim* = 0.88%), but very poor for F267 (*I.sim* = 104.92 %). The higher performance indices for F267 for both the Rev.t model

and HYSYS model are due to the very small response of F267 to a series of reflux flow rate step changes. The F267 deviation from its mean of dataset t.4 is small, hence, large values for both *I.sim* and *I.ann* are expected.

Figure 4.6 (c) shows the validation of the Lev.t model using dataset t.5. To summarize, dataset t.5 contains a series of manipulated variable setpoint changes in the distillate flow rate, F257, (±1.5 m^3/h). The L254 control loop was opened allowing for the open loop response to be identified by the Lev.t model. The performance indices, in Figure 4.6 (c), show that the Lev.t model accurately predicted the L254 open loop response with *I.ann* values of 1.55% for L254 and 14.03% for L253. Comparatively, the HYSYS model predicted L254 with *I.sim* value = 7.91% which is well below the benchmark value of 52% for the levels.

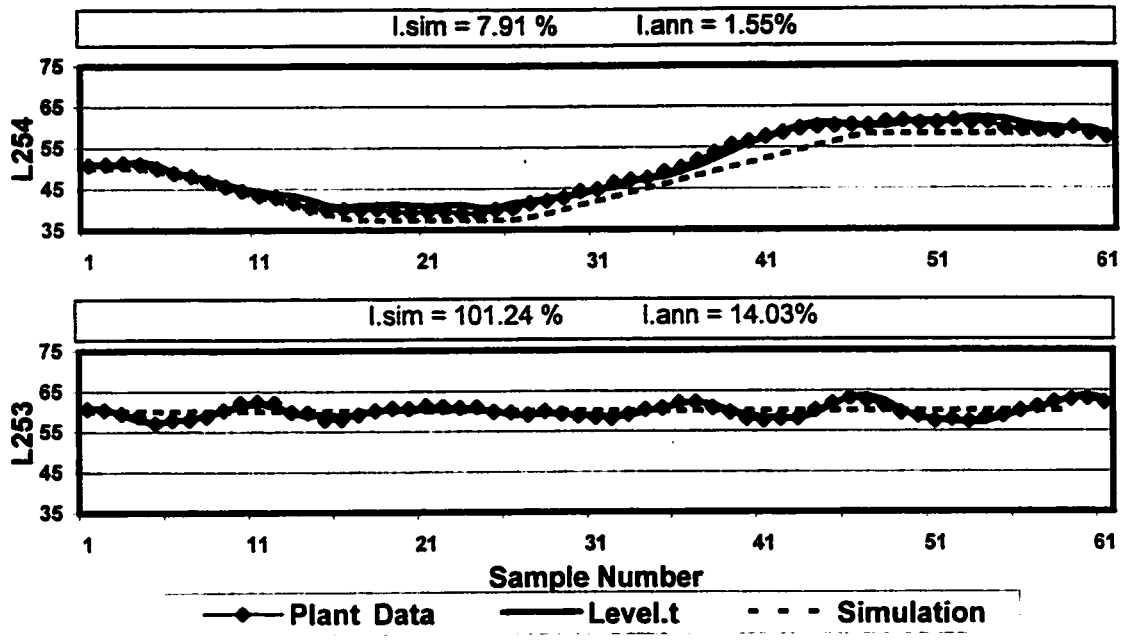


Figure 4.6(c): Comparison of the Lev.t and the HYSYS model predictions of TW252 condenser and reboiler levels for open loop dataset, t.5.

4.3.2 Comparison using Closed Loop Plant Data

Figures 4.7 (a) through (c) shows the ANN.t models predictions for a closed loop setpoint change of +0.5 °C for Y255 from its nominal value of 131.2 °C compared to the HYSYS model prediction from Chapter 3. For high purity columns, the temperature at the top of the column is very sensitive, therefore, only minimal allowable setpoint changes at the top of the column (i.e. Y255 setpoint change) can cause very large gains in the temperatures at the bottom of the column. The *I.ann* values of Figure 4.7 (a) show that the Temp.t model accurately predicts the tray temperatures near the bottom of TW252, Y257 and Y284. However, near the feed stage (Y297) and at the top of the column (Y282 and Y255), the prediction accuracy tends to decrease because the Y255 setpoint changes had very small gains compared to temperatures at the bottom of the column (Y257 and Y284). A similar prediction pattern for the Temp.t model at the bottom of the column was also seen in the correlation coefficients (R^2) values shown in Table 4.7 for the same model validated on dataset V.tCL. In comparison, the performance indices of the HYSYS model are larger than the Temp.t model indices except for Y255 where the indices are similar because, for dataset t.1, the HYSYS model was tuned to fit Y255. Furthermore, again, the *I.sim* values are well within the upper limit values indicating that the HYSYS model captured the closed loop behavior of the internal tray temperatures quite successfully as was concluded in Chapter 3.

From Figure 4.7 (b), the *I.ann* values of 0.39 – 1.53% indicate that the Comp.t model was very successful in capturing the variable gains of the distillate compositions (A254, A256, and A257) for dataset t.1. In comparing the HYSYS model and Comp.t model predictions of the distillate compositions, the difference between the performance indices for the HYSYS model are much more greater than what we saw before with the Temp.t model. For example, the Comp.t model accurately predicted the A254 distillate composition, with an *I.ann* = 1.53 %. However the HYSYS model predicted an *I.sim* = 19.23% which is well within the benchmark value of 24%. Hence the HYSYS model was able to predict A254 at the on-line composition analyzer measurement points, as

indicated by "-x-" on the Figure 4.7 (b), quite well. From Chapter 3, the low sampling frequencies of the distillate composition on-line analyzer seem to miss the true A254 process dynamics and that the HYSYS model representation is likely more accurate. Hence, the Comp.t model identified on plant data was not able to capture the distillate compositions true dynamics because of the ANN's knowledge of the dynamic response of the distillate compositions was limited to only the plant data.

Also, from Figure 4.7 (b), the Rev.t ANN model seemed to predict F256 and F267 fairly accurately. For F256 the standard deviation from the mean in the dataset is quite large and little signal noise is present, therefore the Rev.t model was successful in capturing the dynamic behavior of F256 to a step change in Y255. In comparison the HYSYS model accurately predicted the F256 response with an *I.sim* = 26.53%. However, for output F267 the standard deviation of dataset t.1 is very small. Hence. signal noise is very evident, hence. the performance index is as expected to be an order of magnitudes higher for both the Rev.t and HYSYS model predictions.

Figure 4.7 (c) shows the closed loop response of the condenser level, L254, and reboiler level, L253 to ±10% L254 setpoint changes from its nominal process value of 50% of dataset t.3. From the Figure, the Lev.t model accurately predicted L254 with an *I.ann* = 1.74%, however was less accurate in predicting L253, with an *I.ann* = 52.32%. The large *I.ann* value is directly related to the lack of dynamic information as indicative by the small standard deviation of L253 in the dataset. Hence, noise again becomes the driving force behind the large *I.ann* values.

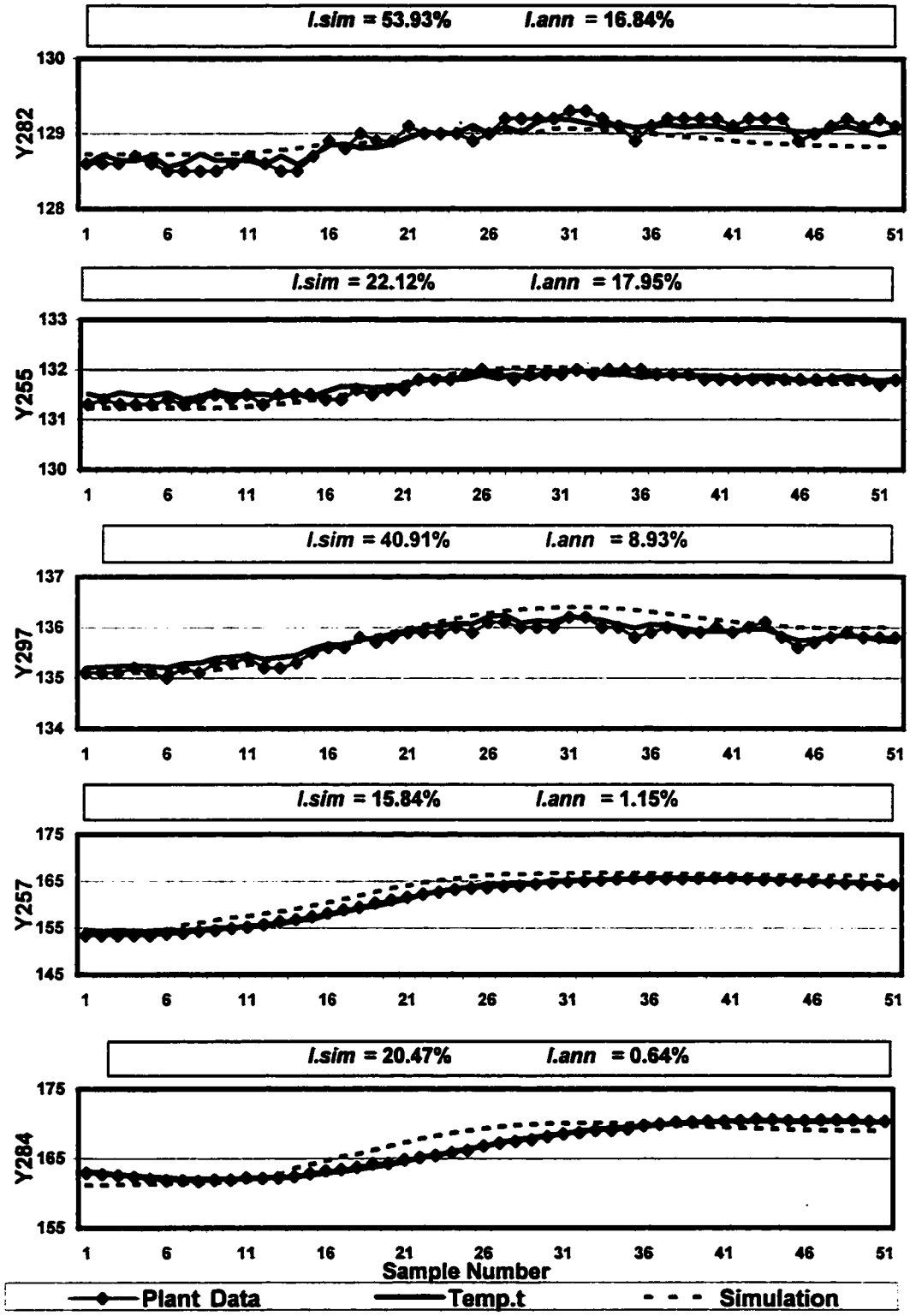


Figure 4.7(a): Comparison of the Temp.t and the HYSYS model predictions of TW252 internal temperatures for the closed loop dataset, t.1

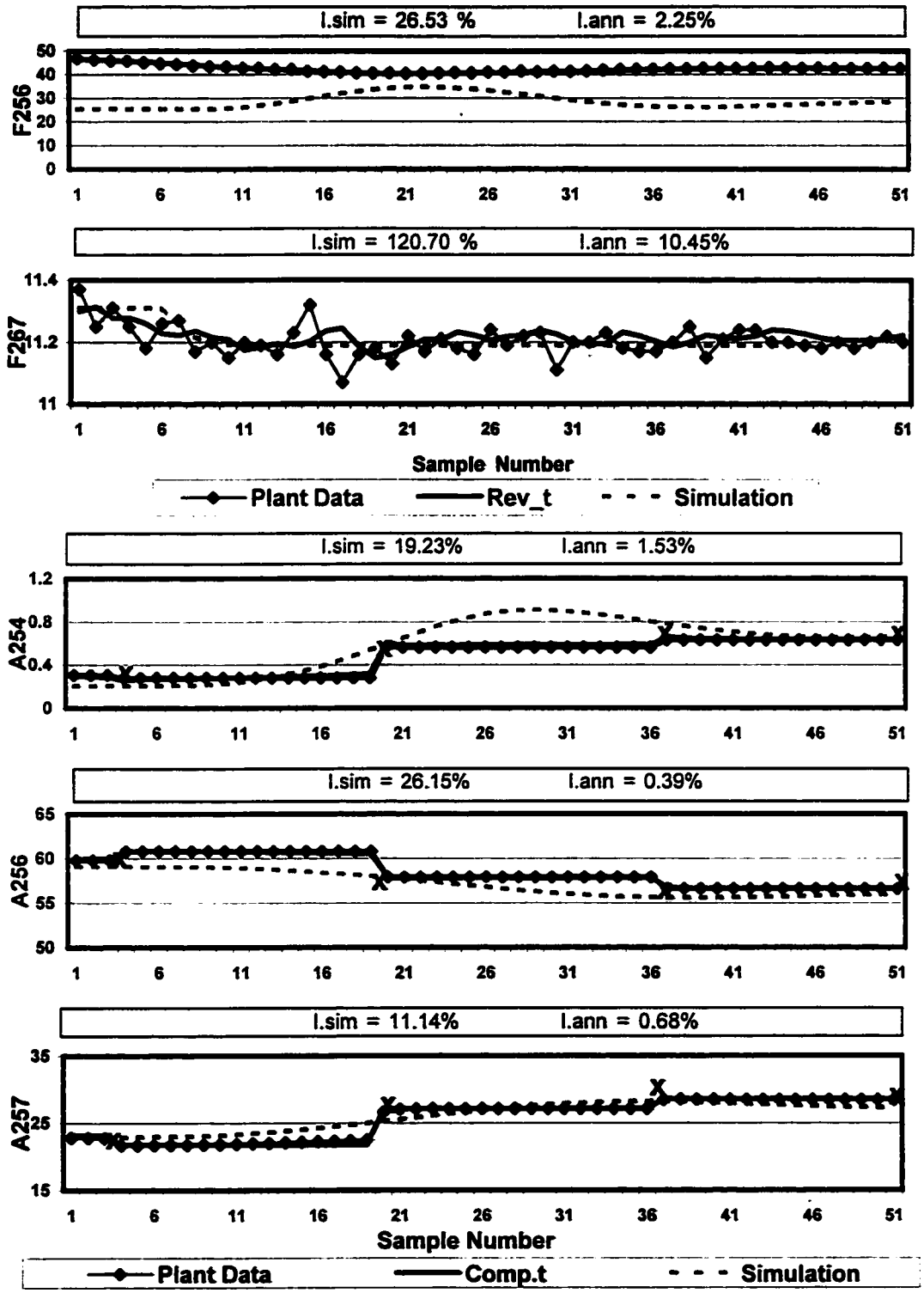


Figure 4.7(b): Comparison of the Comp.t, Rev.t and the HYSYS model predictions of TW252 distillate compositions and internal flow rates for the closed loop dataset, t.1

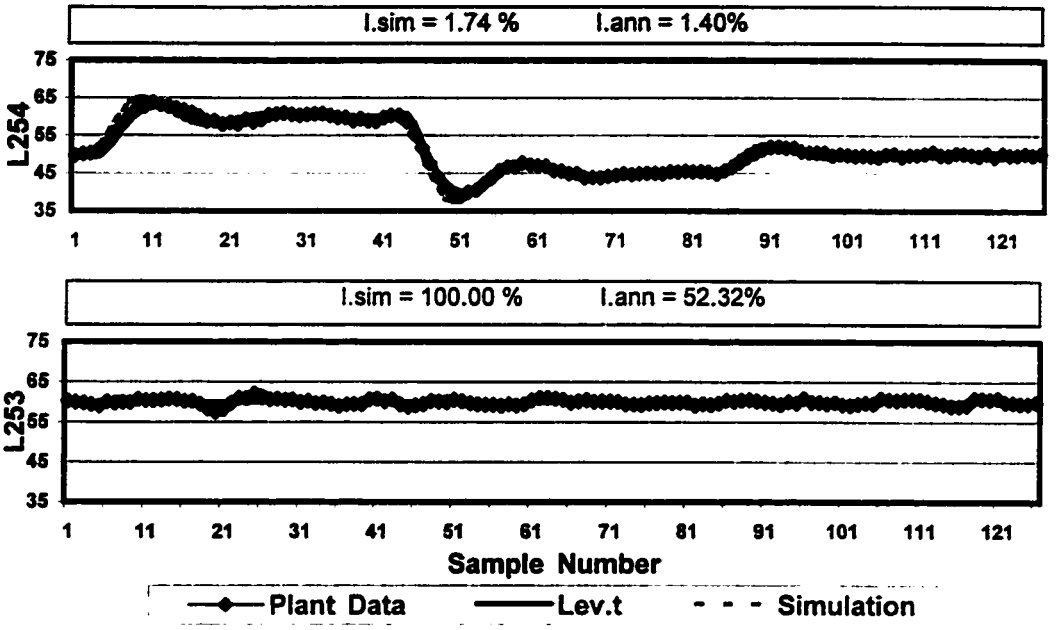


Figure 4.7(c): Comparison of the Level.t and the HYSYS model predictions of TW252 condenser and reboiler levels for closed loop dataset, t.3

4.3.3 Performance Index Benchmarks for the ANN models

The ANN.t model prediction results from the open loop and closed loop plant data presented in the previous sections are used as a guideline to determine benchmark and upper limit values for the performance indices, *I.ann*. The indices are developed for the four basic ANN models, which represent the entire column; internal tray temperatures (Temp), distillate compositions (Comp), internal flow rates (Rev) and levels (Lev) and are intended to be used to benchmark any ANN model prediction independent of the dataset the models are trained on. The benchmarks are tabulated in Table 4.8.

Table 4.8: Proposed benchmark and recommended upper limit values for ANN models prediction of plant data

Category	Output	*Benchmark*	*Upper Limit*
		I.ann (%)	*I.ann (%)*
Temp	Y282 Y255 Y297 Y257 Y284	6.5	10
Comp	A254 A256 A257	1.1	1.6
Rev	F256 F267	15	23
Lev	L254 L253	17	25

From Table 4.8, the optimum performance indices for the ANN models were chosen using the same procedure as outlined in Chapter 3 for the *I.sim* values for the HYSYS model. Furthermore the upper limit for I.ann can be used to determine the accuracy of an ANN prediction of a certain dataset. For example, if the Temp.t model predicted Y282 for a particular dataset with an *I.ann* value > 10%, then the prediction results of the Temp.t model are not accurate for a particular variable. Note that the

benchmark and upper limit values for the Rev and Lev type models are quite a bit larger that the Temp and Comp type models because of large *I.ann* values for F267 and L253.

Using these benchmark values, Table 4.9, compares the prediction of datasets t.1 and t.4 for models Temp.t, Comp.t and Rev.t and for datasets t.3 and t.5 for model Lev.t with the HYSYS model predictions. The prediction of the outputs that violated the recommended upper limit values for both the *I.sim* and *I.ann* values are highlighted in grey in Table 4.9.

Table 4.9: A comparison of performance indices (*I.sim and I.ann*) values for the ANN.t and HYSYS model prediction results

Model	Output	Closed Loop (t.1)		Open Loop (t.4)	
		I.sim (%)	*I.ann.* (%)	*I.sim* (%)	*I.ann* (%)
$Temp_t$	Y282	[illegible]	[illegible]	***32.16***	6.49
	Y255	22.12	[illegible]	[illegible]	3.32
	Y297	***40.91***	8.93	22.60	2.13
	Y257	15.84	1.15	5.04	3.51
	Y284	20.47	0.64	13.48	0.84
$Comp_t$	A254	19.23	***1.53***	***33.83***	1.07
	A256	***36.15***	0.39	***29.30***	1.10
	A257	11.14	0.68	6.06	***1.43***
Rev.t	F256	26.53	2.55	0.88	1.11
	F267	[illegible]	10.45	[illegible]	[illegible]
		Closed Loop (t.3)		**Open Loop (t.5)**	
Lev.t	L254	1.74	1.40	7.91	1.55
	L253	[illegible]	[illegible]	[illegible]	14.03

The results show that for the closed loop datasets the HYSYS model violated the recommended upper limit *I.sim* values the same number of times as the ANN.t models did for *I.ann* upper limit values. However for the open loop case the violations for the HYSYS model comparatively were 3 to 1 compared to the ANN.t models. Furthermore, the number of *I.sim* and *I.ann* values above their respected benchmark values are italicized and bolded in Table 4.9. These values show that for the HYSYS model

predicted output values above the benchmark values but below the upper limit values twice as many times compared to the ANN.t models for the closed loop dataset and three times more for the open loop dataset.

4.4 Comparison Using Other Closed Loop Tests

The HYSYS and ANN.t model predictions are further compared using other closed loop datasets t.2 and t.9. The performance indices for both the HYSYS model and ANN.t models are tabulated in Table 4.10 for comparison.

Table 4.10: Performance indices of the HYSYS and ANN.t model predictions of closed loop setpoint response datasets t.2 and t.9

Model	Output	t.2		t.9	
		I.sim (%)	*I.ann (%)*	*I.sim (%)*	*I.ann (%)*
Temp.t	Y282	27.74	***9.88***		
	Y255	21.36	5.99		
	Y297				
	Y257				
	Y284			***34.16***	
Comp.t	A254		1.00		***1.15***
	A256				***1.59***
	A257			23.16	***1.14***
Rev.t	F256				
	F267				8.61

Examining the results from Table 4.10, it can be seen that in most instances both the HYSYS model and Temp.t model had poor prediction accuracy for most of the internal tray temperatures for datasets t.2 and t.9. The result is a bit surprising for dataset t.2, however, it was expected for t.9 since there were undetected process disturbances and a few faulty temperature sensors (especially Y297). The HYSYS and Rev.t models see a similar pattern for the prediction of the internal flowrates. The HYSYS model and the Comp.t model predictions of the distillate compositions showed that the ANN was able to meet the upper limit values (< 1.6%) quite well for dataset t.9, and only for variable A254

for dataset t.2. On the other hand, the HYSYS model prediction of dataset t.2 was inaccurate (*I.sim* > 37%) for all the output variables but within the benchmark value for only variable A257 for dataset t.9.

4.5 Comparison Using Disturbance Plant Data

The HYSYS model and the ANN.t models were further compared using an independent disturbance dataset, t.8. Dataset t.8 is an independent dataset because neither model used the dataset during its identification process. In summary, dataset t.8 is comprised of a series of feed composition disturbances in A264, A266, and A267. Using the recommended benchmarks and upper limits, the prediction accuracy of the HYSYS and Temp.t and Comp.t models are compared and shown in Table 4.11.

Table 4.11: Comparison of the HYSYS model and ANN.t model prediction of selected TW252 variables for the independent disturbance dataset, t.8

Model	Output	t.8	
		I.sim (%)	*I.ann (%)*
Temp.t	Y282	26.43	[illegible]
	Y255	***28.62***	[illegible]
	Y257	[illegible]	[illegible]
	Y284	[illegible]	[illegible]
Comp.t	A254	[illegible]	[illegible]
	A257	[illegible]	[illegible]

In examining the results of Table 4.11 it is evident that the Temp.t and Comp.t models gave poor predictions for selected internal tray temperatures and distillate compositions as indicated by the high performance indices that are well above the recommended upper limit values. Similar prediction results occurred for the HYSYS model prediction of Y257 and Y284, however adequate *I.sim* values for Y255 (< 42%) and within benchmark values for Y255 (< 28 %) were obtained. In general, the HYSYS model *I.sim* values, relative to the *I.ann* values, were diligent in comparison and the basic trends of the variables were detected as concluded in Chapter 3. Hence, it can be

concluded that the Temp.t and Comp.t models trained only on dataset I.t could not capture the dynamic behavior of dataset t.8 and prediction results of the HYSYS model were far from satisfactory, however, better than the ANN.

4.6 Comparison of ANN.t Model Predictions of Simulated and Plant Data

The Temp.t, Comp.t, Rev.t, and Lev.t models were used to predict the generated simulated response of the plant data as another check to determine if the ANN can predict HYSYS model generated data that is similar to the plant data it was identified on. The *I.ann* values of the ANN.t models prediction of the simulation data should be similar to the *I.ann* values obtained for the ANN.t models prediction of the plant data and this will assure that both models are correct. Table 4.12 shows the Temp.t, Comp.t and Rev.t models prediction results using the performance index for the simulated data compared to the prediction results of the plant data. The simulated data is given the suffix "sim". For example the simulated representation of plant dataset t.1 is labeled as t.1.sim. A similar nomenclature is used to represent the other plant datasets.

Table 4.12: Comparison of the Temp.t, Comp.t and Rev.t model predictions of simulated and plant data using performance index, *I.ann*

Model	Output	Dataset t.1	Dataset t.1.sim	Dataset t.4	Dataset t.4.sim
		I.ann (%)	*I.ann (%)*	*I.ann (%)*	*I.ann (%)*
Temp.t	Y282	[illegible]	[illegible]	6.49	9.09
	Y255	[illegible]	[illegible]	3.32	8.89
	Y297	8.93	3.51	2.13	2.25
	Y257	1.15	0.44	3.51	2.94
	Y284	0.64	2.10	0.84	2.53
Comp.t	A254	1.53	[illegible]	1.07	[illegible]
	A256	0.39	1.38	1.10	[illegible]
	A257	0.68	1.45	1.43	[illegible]
Rev.t	F256	2.55	2.36	1.11	1.34
	F267	10.45	23.44	[illegible]	[illegible]

From the Table 4.12, a comparison can be made between the ANN.t models ability to predict plant data and simulated data. For example for dataset t.1 and t.1.sim, the Temp.t model predicted both datasets with similar performance indices. The *I.ann* values for both datasets were similar for tray temperatures at the bottom of the column (Y284, Y257, and Y297), however for both datasets the Temp.t model prediction of Y255 and Y282 exceeded the upper limit of 10%. The Temp.t model had similar performance indices for the prediction of datasets t.4 and t.4.sim; however, in this case, all the predicted outputs had *I.ann* values below the upper limit.

The Comp.t model's performance indices for the prediction of the outputs (A254, A256, A257) for both dataset t.1 and t.1.sim were within the upper limit of 1.6 % except for the prediction of A254 in dataset t.1.sim. As stated in Chapter 3, the actual dynamic response of A254 to the Y255 setpoint change was not available from the plant data due to the low frequency sampling of the on-line analyzer. Furthermore, the HYSYS model predicted the A254 response at the composition analyzer points quite accurately with *I.sim* = 19.23 %, which was below the benchmark *I.sim* value (24%) indicating that the prediction was very accurate. The Comp.t model (which was identified using dataset I.t) did not contain any information of A254 true dynamic nature during the identification stage, hence, the model was not able to extrapolate to the dynamic conditions presented in dataset t.1.sim for A254 and is reflected in the *I.ann* value of 5.19%. The lack of the ANN model's ability to extrapolate is further recognized in the Comp.t models prediction of the outputs for dataset t.4.sim, where not only was A254 *I.ann* values well above the upper limit, so was A256 and A257.

Table 4.12 also shows the Rev.t model gave similarly good *I.ann* values for both datasets t.1 and t.1.sim for outputs, F256 and F267. Moreover, accurate and resembling *I.ann* values are obtained for F256 for datasets t.4 and t.4.sim. The Rev.t model's prediction of F267 for both dataset t.4 and t.4.sim are equally unsatisfactory.

The *I.ann* values for the Lev.t model prediction of the simulation data and plant data are compared in Table 4.13.

Table 4.13: Comparison of the Lev.t model predictions of simulated and plant data using performance index, *I.ann*

Model	Output	Dataset t.3	Dataset t.3.sim	Dataset t.5	Dataset t.5.sim
		I.ann (%)	*I.ann (%)*	*I.ann (%)*	*I.ann (%)*
Lev.t	L254	1.74	1.31	1.55	2.27
	L253	[redacted]	[redacted]	14.03	[redacted]

Table 4.13 shows that the Lev.t model predicted both dataset t.3 and t.3.sim very similarly with *I.ann* values exceeding the upper limit (>25%) for L253 and within benchmark values for L254 (<17%). For datasets t.5 and t.5.sim (open loop response) the Lev.t model predicted L254 as indicated by the low *I.ann* values. For L253 the Lev.t model predicted *I.ann* = 14.03% for dataset t.5, which is within the benchmark value, however for t.5.sim the *I.ann* values was 127.83%, which is well above the upper limit value. The discrepancy between the plant data and the Lev.t model prediction of L253 for dataset t.5 is due to the excessive amount oscillation in the L253 signal in dataset t.5, causing large *I.ann* values. Whereas dataset t.5.sim contained little or no oscillations, therefore, the Lev.t model easily predicted L253 within benchmark values.

4.7 Summary

Four dynamic ANN models (Temp Comp, Rev and Lev) were discussed and developed for TW252 using plant step response data. The optimum ANN architectures were chosen using a rigorous iterative procedure that involved various steps. The first step was selecting the number of inputs to represent the selected outputs for each model based on the correlation coefficients. Once the inputs and outputs are chosen the second iterative procedure involved using the correlation coefficient again to optimize the

number of past values (determines the dynamic nature of the models) and the number of internal parameters. The optimum ANN architectures were identified using dataset I.t and given the classification of ANN.t models (Temp.t Comp.t, Rev.t and Lev.t). The ANN.t models were validated using various independent datasets and then the predictive power of the models were compared to the HYSYS model predictions from Chapter 3 using a new performance index, *I.ann*. From comparisons using datasets t.1 and t.4, it was obvious that the performance benchmarks for the HYSYS model in Chapter 3 were not applicable to the ANN.t models. Hence, benchmarks and upper limit values for *I.ann* values were developed so that the models could be compared correctly. In general, based on a comparison of performance indices, the ANN.t model predicted the plant data much more accurately than the physical model as was expected because the ANN.t models were directly identified from the compared data. However, for disturbance dataset t.8. which was completely independent of the training data, the ANN.t models were inferior to the HYSYS model indicating that the ANN.t models lacked the ability to extrapolate beyond the data they were trained on. The inability to extrapolate was further exhibited in the ANN.t model prediction of simulated data for the distillate compositions for datasets t.1.sim and t.4.sim were the lack information from the trained data clearly did not help the ANN.t model predict the true simulated response.

Chapter 5

Integration of the First Principle and ANN Modeling Techniques

For industrial processes, data that is rich with dynamic information such as setpoint step response data are often difficult to obtain. It is usually not feasible to deliberately upset a column by introducing large setpoint or load changes. There are some commercial applications such as DMCplus from Aspen Technology that are successful at upsetting a process to obtain models that can be used for control of the process. The modeling however requires many perturbations of all the manipulated variables to obtain linear, dynamic step response models of the controlled variables response. Developing these step response models and determining which variables to step is time consuming and could involve at least two weeks of testing and then countless days of trying to re-stabilize the unit after the testing.

Another alternative is to create models from available data to identify an ANN. The data to build ANN models are present in a data historian. However, the data in a historian is usually compressed or averaged and consists mostly of steady state information (nominal conditions) with very little dynamic information. Therefore, an ANN trained on this type of data may not represent the entire range of operation desired to obtain a good model of the process. However, data generated from a physical model can be used to simulate process information beyond normal operating regions, such as changing large setpoint changes or feed composition disturbances. The data from the physical model can be merged with historical plant nominal operating data to increase the

range of data on which an ANN model is identified. In essence, the resulting ANN is identified from hybrid data that represents the nominal plant data and simulated data outside the nominal regions.

5.1 Recommended Integrated Modeling Strategy

The following is an overall strategy for the integration of the physical model and the ANN models for TW252:

1. Develop a physically-based dynamic model bench-marked with available plant operating data.
2. Simulate transient response data such as known disturbances or setpoint changes that may occur in the process using the physical model
3. Using the optimum ANN topologies, the ANN models (Temp, Comp. Rev, and Lev) are identified using historical plant data and validated using a collection of the step response data.
4. The physical model is then used to provide process knowledge to the ANN models in the form of data to extrapolate the intended prediction region of the ANN models.
5. Then a comparison may be performed using the performance benchmarks and upper limit values of *I.ann* for the ANN models trained using plant nominal and simulated process data (ANN.n+s models) to ANN models trained only on plant nominal data (ANN.n models).

From the modeling strategy above, the development of the physical and ANN models have been presented in Chapter 3 and 4, respectively. In this chapter using the optimum ANN topologies selected in Chapter 4, the ANN models are identified using nominal operating data, and then validated using the step response data. Furthermore, an integrated model that combines simulated step response data and nominal operating data to aid in the development of a robust Hybrid model for TW252 is investigated.

5.2 ANN Models of TW252 Identified with Nominal Data

For most industrial applications such as TW252 the available data to identify an ANN model is usually in the form of nominal data available from a data historian. In most instances for an industrial application, step response data is a luxury. Hence, the ANN model topologies selected in Chapter 4 are re-trained on nominal data and used to predict step response or disturbance datasets.

The optimum ANN model (Temp, Comp, Rev and Lev) topologies are re-identified using three regions of nominal historical data, which are labeled as dataset I.n. The new identification dataset I.n and its statistics can be found in Table E.2 and plotted in Figures E.2, Appendix E for selected TW252 variables. Dataset I.n contains 3860 sample data points for each of the measured TW252 variables and represented three months of different steady-state operating regimes. From Appendix E, in comparing the statistics of dataset I.t and I.n, dataset I.n is considered to represent closed loop and nominal operation with most of the points near the three steady state operating regimes. Whereas dataset I.t is completely made up of transient data representing a wide range of open and closed loop operation.

Similar to the ANN models identified on dataset I.t (in Chapter 4), the ANN models identified on dataset I.n are given an extension "n". For example the Temp, Comp, Rev and Lev models identified on nominal data are represented as Temp.n, Comp.n, Rev.n, and Lev.n and are generalized as ANN.n models. Table 5.1 shows the correlation coefficient results for the ANN.n model prediction of both identification dataset I.n and various independent datasets; V.tOL, V.tCL (subsets of I.t), and V.t8

Table 5.1: Correlation coefficient results for ANN models identified on nominal data and validated using other closed loop datasets

Model	ANN Topology	Output	Set I.n	Set V.tOL	Set V.tCL	Set V.t8
			R^2	R^2	R^2	R^2
Temp.n	35-17-5	Y282	0.997	0.993	0.992	0.997
		Y255	0.998	0.994	0.994	0.996
		Y297	0.992	0.990	0.960	0.993
		Y257	0.994	0.988	0.988	0.996
		Y284	0.995	0.987	0.991	0.995
Comp.n	24-13-3	A254	0.971	0	0.776	0.978
		A256	0.990	0.946	0.945	0.996
		A257	0.983	0.888	0.853	0.986
Rev.n	18-10-2	F256	0.996	0.983	0.983	0.996
		F267	0.991	0.978	0.988	0.992
Lev.n	18-10-2	L254	0.950	0.765	0.821	-
		L253	0.990	0	0	-

From Table 5.1 the selected optimum topologies for the four ANN.n models were able to predict the outputs very accurately, based on R^2 values for identification dataset I.n. The very high R^2 values for the prediction of dataset I.n indicate that the output variables data in dataset I.n consist mostly of groups of samples around a steady state value. Therefore, it is easier for the ANN.n models to approximate dataset I.n than the ANN.t models in approximating dataset I.t, because the information in I.n only represented a few regions of possible operation. This is in contrast to dataset I.t, which is completely made up of transient data representing a wide range of operation.

In general, the verification results of the Temp.n and Rev.n models validated using datasets, V.tOL, V.tCL (subsets of I.t), and V.t8, were notably good for most of the models because the correlation coefficients were well above the criterion of 0.850. On the other hand the Comp.n and Lev.n models had noticeable difficulty in predicting the outputs for the validation detests. For example, for validation dataset V.tOL, the Comp.n model predicted the output, A254, with an R^2 value = 0 and for dataset V.CL, $R^2 < 0.850$, indicating that the inputs chosen were not able predict the output at all. However, for

dataset V.t8 the results were much better. On the other hand, the Comp.n model predicted A256 and A257 successfully for all the validation datasets. For the Lev.n model the negligible R^2 value was also obtained for the prediction of L253 for dataset V.tOL and V.tCL. Furthermore, correlation coefficients results for L254 were also below the 0.850 criterion for both datasets V.tOL and V.tCL. The low values indicate that the Lev.t model was not able to represent the input/output relationship of datasets for those particular outputs. The Lev.n model was not used to predict the outputs in dataset V.t8.

5.3 Comparison of ANN.t and ANN.n Model Predictions

The ANN.n models identified in the previous section will be used to predict both open loop and closed loop data. As was clearly evident in Chapter 4, the most important consideration in the reliability of a model is the ability of the model to predict independent datasets. The ANN.n models (Temp.n, Comp.n, Rev.n, and Lev.n) predictions are compared to the ANN.t models developed in Chapter 4.

5.3.1 Comparison using Open Loop Data

Figure 5.1 (a) and (b) shows the comparison of the Temp.n, Comp.n, and Rev.n models and the Temp.t, Comp.t, and Rev.t predictions of dataset t.4. To review, dataset t.4 consists of manipulated variable setpoint changes (5.0 m^3/h step up and 2.5 m^3/h step down) in the reflux flow rate, F256, from the nominal value of 41.2 m^3/h. Both Y255 and Y284 control loops were in manual mode (a full description of Test 4 can be found in Chapter 2). Figure 5.1 (a) is a display of all the internal tray temperature responses to this open loop step test. The Temp.n model predicted the open loop behavior of the internal tray temperatures less accurately (i.e. Y282, *I.ann* = 224.10%) than the Temp.t model (i.e. for Y282, *I.ann* = 6.49 %) indicated by the higher performance indices. For most of the outputs (Y282, Y255, and Y257) the Temp.n model predicted dataset t.4 with *I.ann* values well above the upper limit value for the Temp model of 10% indicating that the

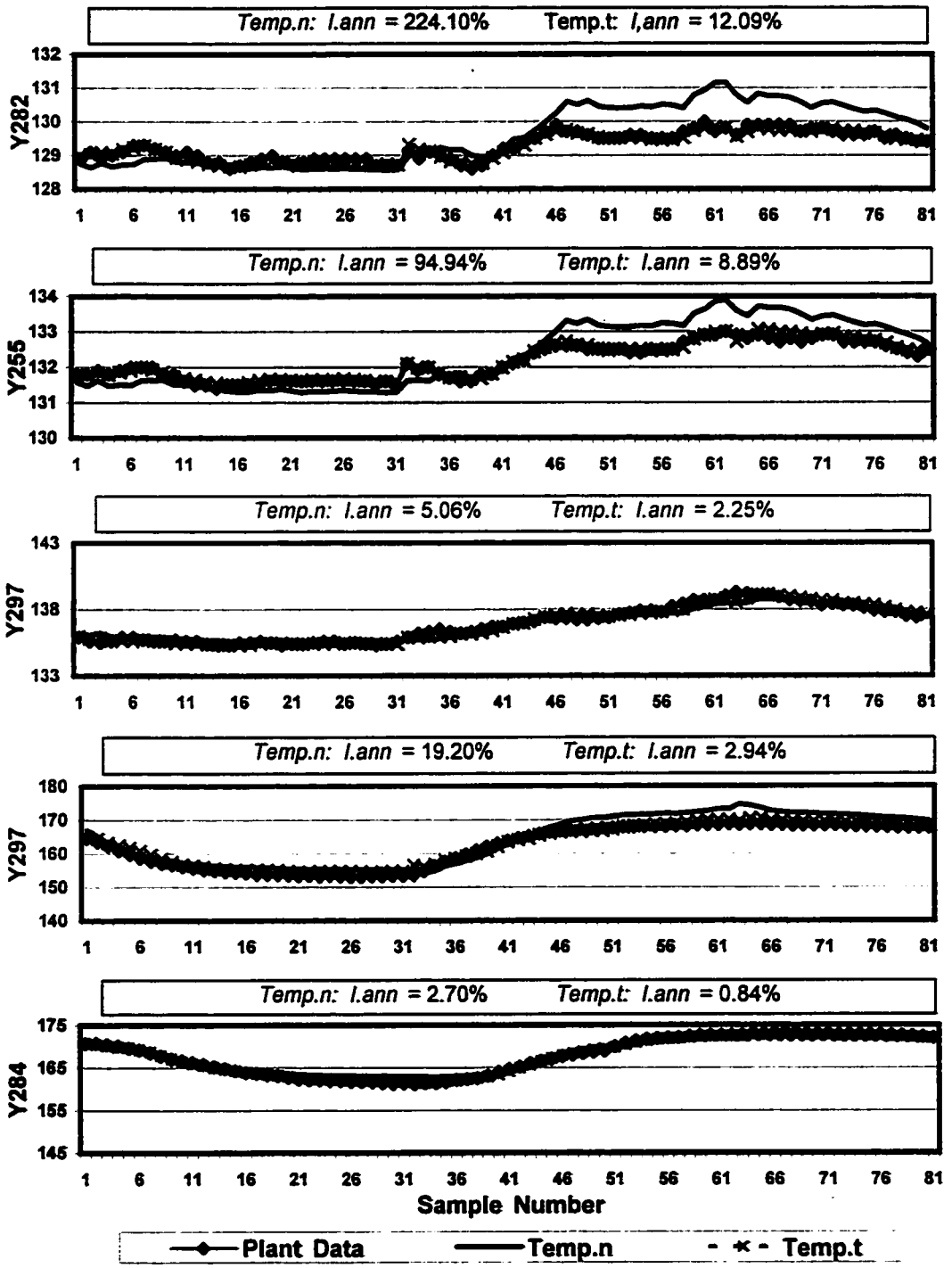


Figure 5.1 (a): Comparison of the Temp.t and Temp.n models prediction of TW252 internal temperatures for the open loop behavior of dataset, t.4.

Temp.n model was not able to interpolate the dynamics of dataset t.4 for some of these outputs. However for outputs, Y297 and Y284 Temp.n model prediction was very accurate indicated by the very low *I.ann* values that were within the benchmark value of 6.5 % for the Temp model.

Figure 5.1 (b) illustrates the ability of the Comp.n and Rev.n models identified using dataset I.n to predict dataset t.4. The Comp.n model predicted A256, and A257 with an *I.ann* = 4.70% and 6.67%, respectively. These performance indices were also higher than the Comp.t model prediction of t.4 and are also well above the *I.ann* upper limit value for the Comp model of 1.6 % for Comp models. The Comp.n model predicts the distillate composition. A254 of dataset V_{t4}, even less accurately with *I.ann* = 52.53 %, which is well above the upper limit value of 1.6%. On the other hand the Comp.t model predicted the distillate compositions with indices well within the upper limit values. The poor performance of the Comp.n model is directly related to the model's inability to extrapolate beyond the conditions in the nominal training data. Similarly the Rev.n model greatly under estimated F267 (*I.ann* = 357.42%). However, the model predicted F256 very accurately with an *I.ann* = 5.43 %, which was well within the benchmark value of 15% for the Rev model.

Figure 5.1 (c) shows a comparison of the Lev.t and Lev.n model predictions of dataset t.5. It clearly demonstrates that the Lev.n model can not approximate the open loop response of L254 based the nominal data used to train the model. The *I.ann* value was well above the upper limit of 25% for Lev models. The poor result of the Lev.n model is further echoed in its prediction of L253 where *I.ann* was 35%.

The most important observations that can be inferred from Figures 5.1(a) through (c) is the ANN.n models were not able to capture the dynamic open loop nature of datasets t.4 and t.5 very effectively. The ANN.n models trained on dataset I.n are not reliable enough to be able to infer open loop conditions beyond the training dataset.

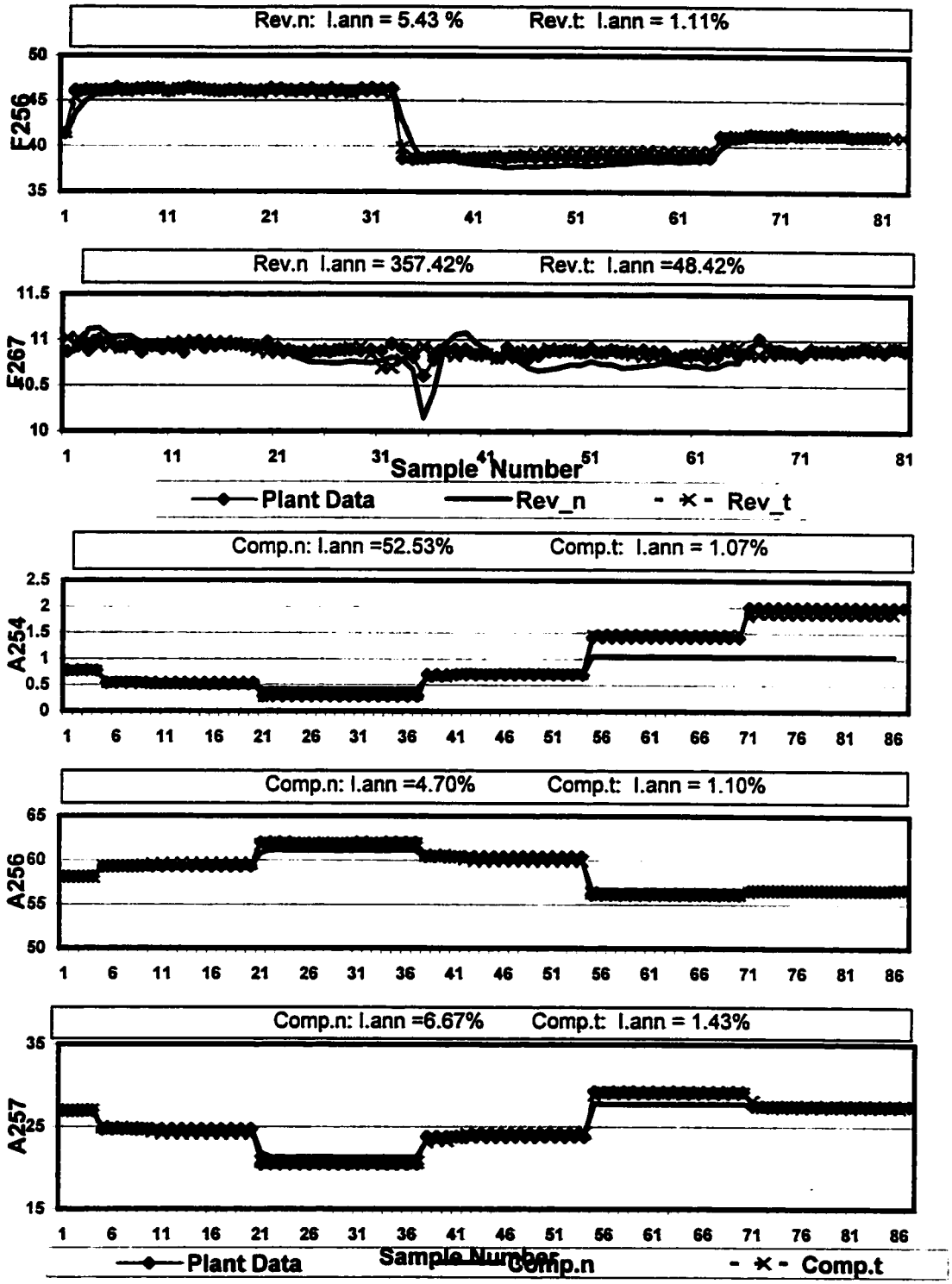


Figure 5.1(b): Comparison of the Comp.t, and Rev.t models to Comp.n and Rev.n models prediction of TW252 distillate compositions and internal flowrates for the open loop dataset, t.4

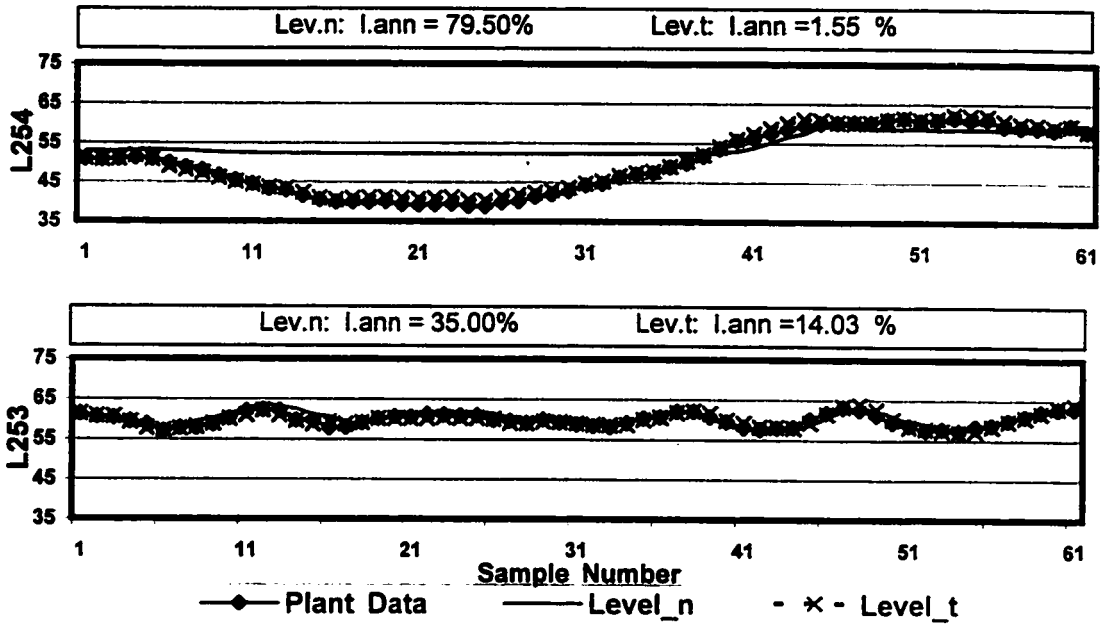


Figure 5.1(c): Comparison of the Lev.t and Lev.n model predictions of TW252 condenser and reboiler levels for open loop dataset, t.5

5.3.2 Comparison using Closed Loop Data

Figures 5.2 (a) through (c) show the Temp.n, Comp.n, Rev.n, and Lev.n model predictions of closed loop data. From Figure 5.2 (a), the Temp.n model had better prediction results for closed loop dataset t.1 than the for the open loop dataset t.4. The results are as expected because the closed loop test more closely mimics nominal operation. However, the Temp.n model still predicted poor *I.ann* values that were well above the upper limit value of 10% for outputs Y282, Y255 and Y297, which again was expected because Y255 rarely operates in closed loop (hence, the nominal data would not have this information). On the other hand the Temp.n model predicted very low *I.ann* values for outputs Y257 and Y284 (bottom of the column), which is not surprising considering that Y284 nominally operates in closed loop. The Temp.n model is expected provide poorer prediction for dataset t.4 than t.1 because t.4 has more transient excitation than t.1. Therefore the Temp.n model identified on nominal data (I.n) would be less effective because the model lacks knowledge of the transient nature of dataset t.4.

Figure 5.2 (b) shows the Comp.n and Rev.n predictions of the closed loop dataset t.3. The *I.ann* values for the Comp.n are much higher than the Comp.t models. As was evident for dataset t.4, the Comp.n model was not able predict the distillate compositions with the performance index upper limit of 1.6% indicating that the Comp.n model again has poor extrapolation characteristics. For the internal flowrates (F256 and F267) the Rev.n model seemed to capture the trend of the dataset, however the *I.ann* value (27.08%) was above the allowable upper limit of 23%.

Figure 5.2 (c) shows a comparison of the Lev.n and Lev.t predictions of the condenser level, L254, and reboiler level, L253, for a series of closed loop setpoint changes as represented in dataset t.3. The Level model was also identified using dataset I_n ($Level_n$) and validated using the dataset V_{t3}. Again, it is evident that the Lev.n model identified on nominal operating data was not able to capture the dynamic nature of

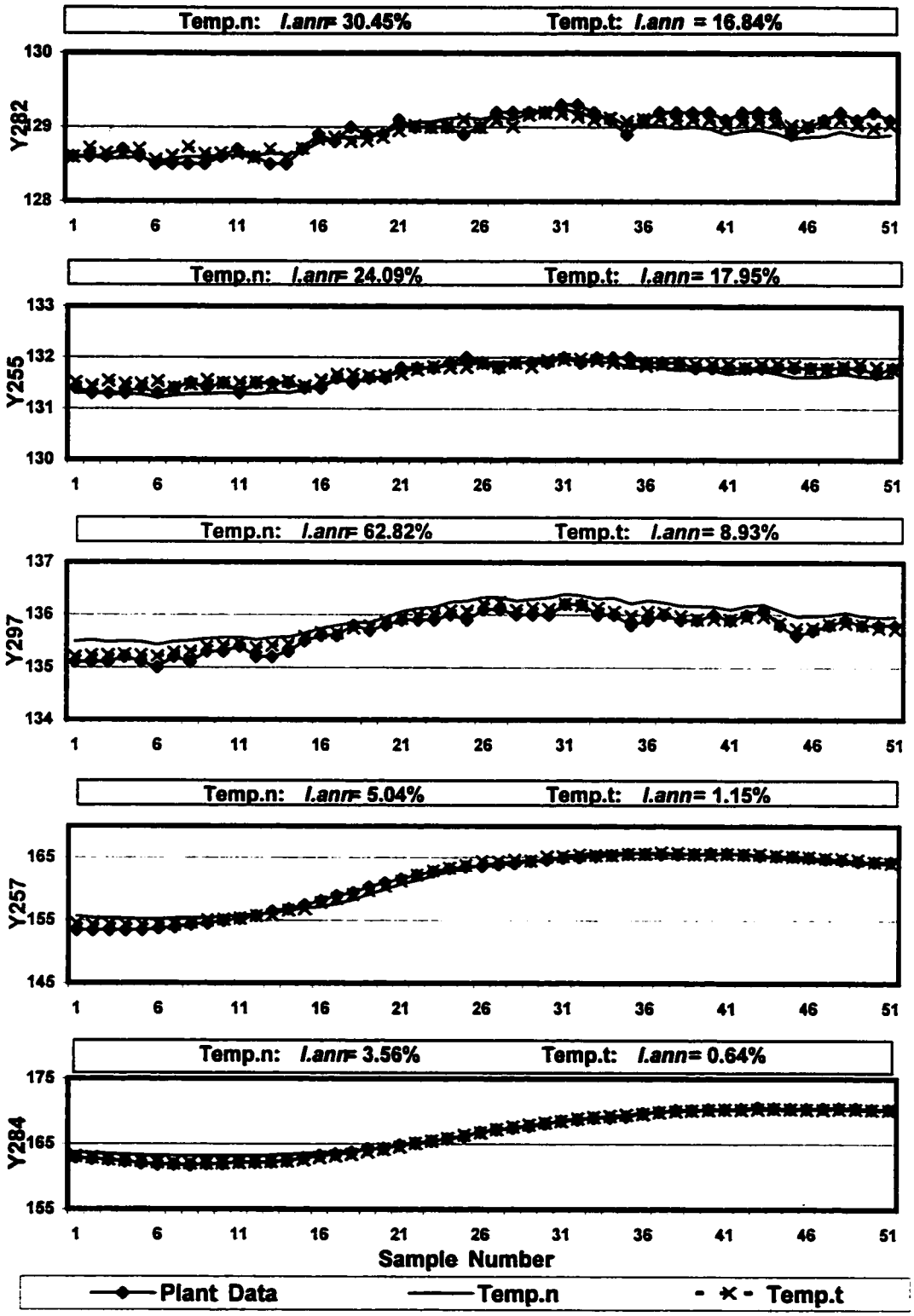


Figure 5.2(a): Comparison of the Temp.t and Temp.n models prediction of TW252 internal tray temperatures for the closed loop dataset, t.1.

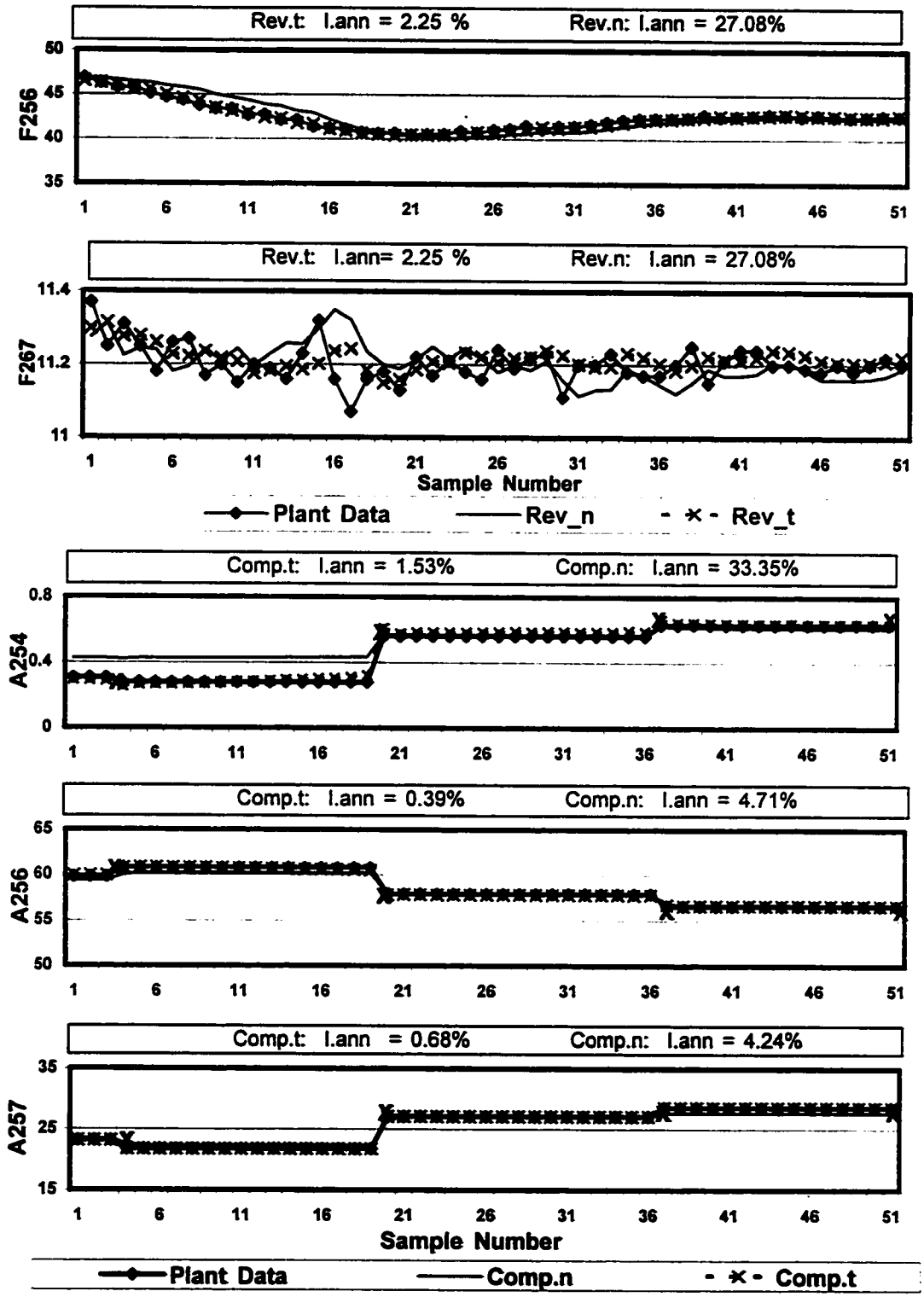


Figure 5.2(b): Comparison of Comp.t and Rev.t models to Comp.n and Rev.n models prediction of TW252 distillate compositions and internal flowrates for the closed loop dataset, t.1

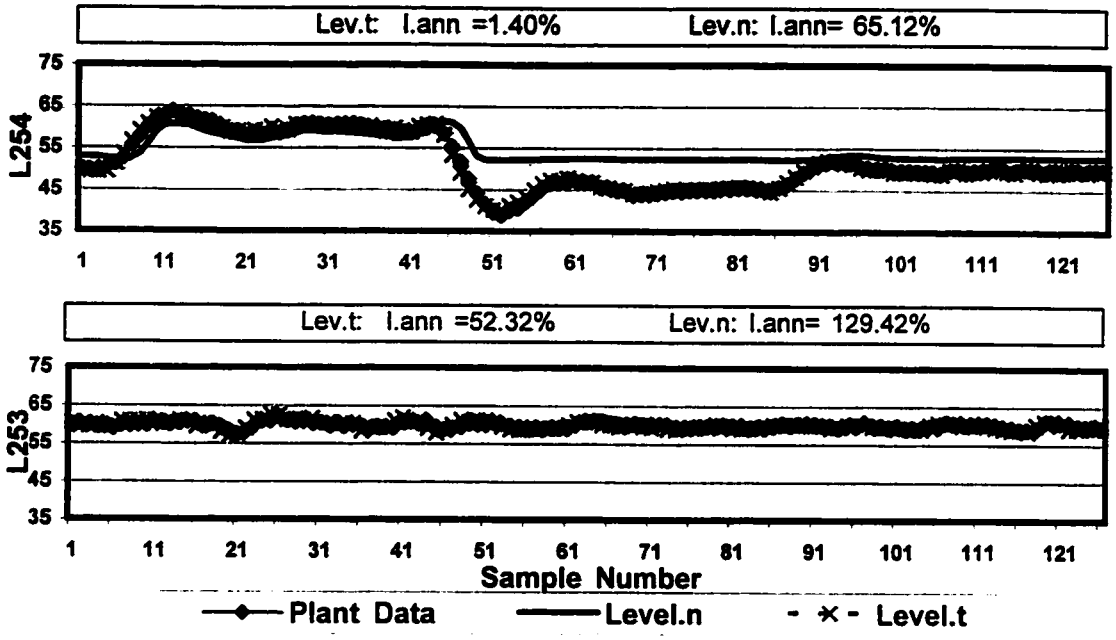


Figure 5.2 (c): Comparison of the Lev.t and Lev.n model predictions of TW252 condenser and reboiler levels for the closed loop dataset, t.3.

dataset t.3. The Lev.n model predicted L254 with $I.ann = 65.12\%$, which was well above the recommended upper limit value of 25%.

The most important observations that can be inferred from Figure 5.2 (a) through (c) are that the ANN.n models were not able to capture the outputs as effectively as the ANN.t models for dataset t.1.

5.4 Gaining Process Knowledge Using the Physical Model

In Chapter 3, the HYSYS model was benchmarked against plant data. The prediction of the plant data for step response tests (t.1, t.2, t.3, t.4, and t.5) were deemed to be accurate and the data were recorded in a new dataset labeled with the suffix "sim". From the integration strategy presented in section 5.1, it is suggested to employ the HYSYS model to provide supplemental identification data to aid the ANN.n model prediction of operating regions that represents operations beyond nominal conditions such as the closed and open loop step response tests.

The HYSYS simulation of the step response tests as outlined in Chapter 2 provides an abundant source of transient data that contains extensive knowledge of the column's dynamics that can be used for the identification of robust ANN models. The dynamic information may expose non-linear input/output relationships that were transparent in the nominal operating data. Hence, the ANN.n models are re-initialized with an augmented dataset labeled I.n+s. Dataset I.n+s represents simulations of Tests 1, 2, 3, 4, 5, 9, and 10 along with the three regions of historical nominal plant operating data. The ANN models (Temp, Comp, Rev, and Lev) identified on dataset I.n+s are given the extension "n+s" and are generalized as ANN.n+s models. A plot of dataset I.n+s is included in Appendix E. Since the ANN.n+s models are trained partly on simulation data. The ANN should be able to learn the partial derivatives of the inputs with respect to the outputs; hence, an increase in the reliability of the ANN.n+s models to predict the outputs for plant step response data should occur.

5.4.1 Comparison of ANN.n+s and ANN.n Model Predictions of Open Loop Plant Data

The performance indices, *I.ann* values, for ANN.t, ANN.n and ANN.n+s model prediction of dataset t.4 are compared in Table 5.2. The table shows that the ANN.n+s models had a much lower number of *I.ann* violations of the upper limit values, as indicated by the smaller amount of highlighted values, than the ANN.n models.

Table 5.2: Comparison of ANN models identified on I.t, I.n and I.n+s predictions of dataset t.4 using *I.ann*.

Category	Output	Identification Dataset		
		I.t	I.n	I.n+s
		I.ann (%)	*I.ann (%)*	*I.ann (%)*
Temp	Y282	6.49	[illegible]	[illegible]
	Y255	3.32	[illegible]	[illegible]
	Y297	2.13	5.06	2.47
	Y257	3.51	[illegible]	2.76
	Y284	0.84	2.70	1.53
Comp	A254	1.07	[illegible]	0.60
	A256	1.10	[illegible]	1.58
	A257	1.43	[illegible]	1.53
Rev	F256	1.11	5.43	0.74
	F267	[illegible]	[illegible]	[illegible]

The Temp.n+s model predicted all the internal tray temperatures with relatively better *I.ann* values than Temp.n but not quite as accurate as the Temp.t model. The Temp.n+s had *I.ann* violations for outputs Y282 and Y255, however for Y257 the Temp.n+s had a slightly better *I.ann* value than the Temp.t model. The Temp.n model greatly overestimates the open loop gain for variables Y282, Y255 and Y257. However, with the inclusion of the simulation data the Temp.n+s model clearly eliminated the overshoot as seen in Figure 5.3. Also, there was a definite improvement in the prediction of the distillate compositions using the Comp.n+s model compared to the Comp.n for

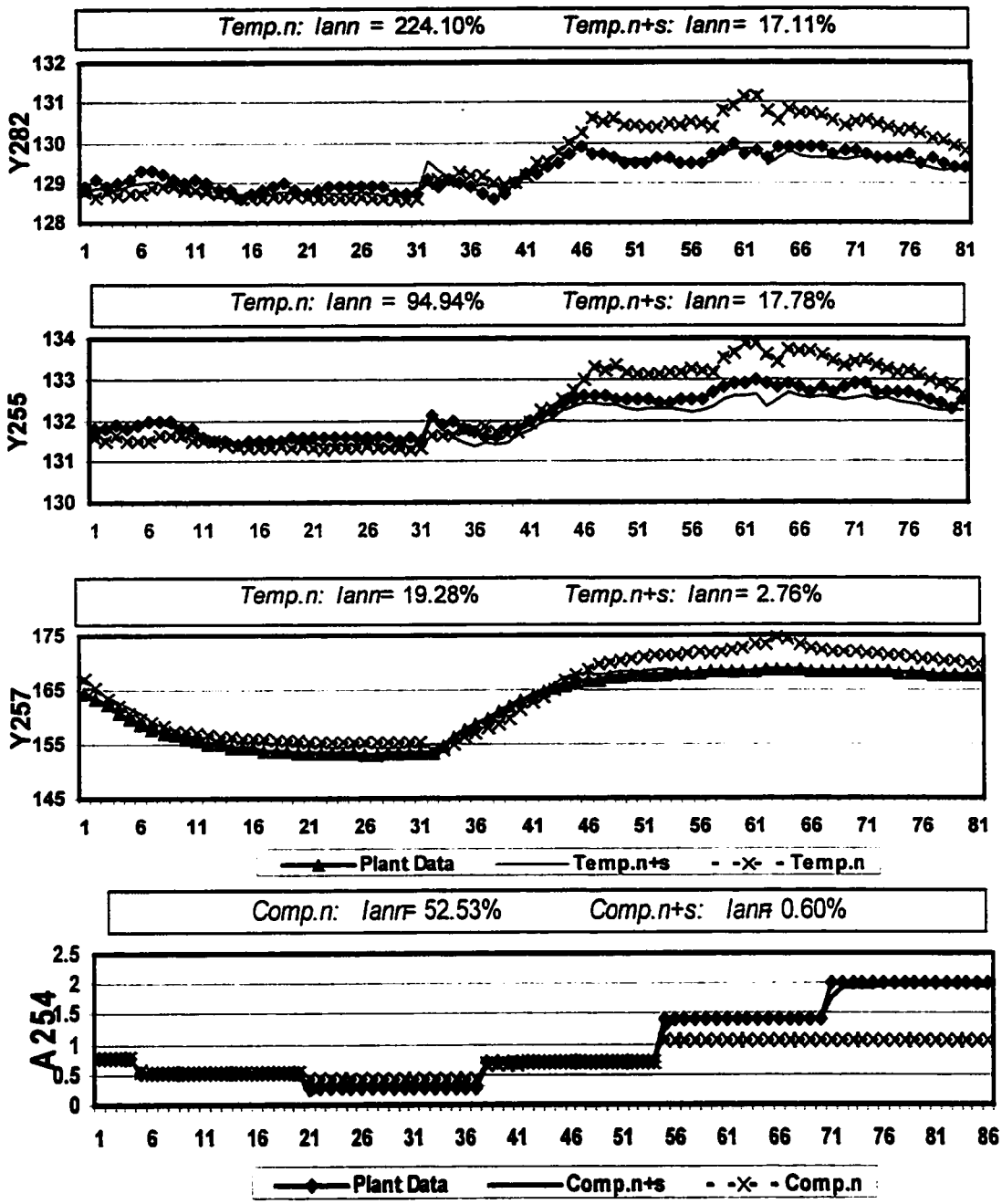


Figure 5.3: The addition of simulated data increases the *I.ann* values of certain variables.

dataset t.4. The Comp.n+s model outperformed the Comp.n model and was more accurate in predicting A254 than the Comp.t model. The introduction of the simulated data extended the range of the nominally trained model, which is evident in the Comp.n+s model, prediction of A254, also seen in Figure 5.3. From Table 5.2 a similar trend is apparent for F256, where in this case the Rev.n+s model provided better predictions than the Rev.t model. The inclusion of simulated data for the Rev.n+s model prediction of F267 did not achieve the same results as F256, however it did manage to increase the performance considerably.

A similar comparison between Lev.t, Lev.n and Lev.n+s models prediction of open loop dataset t.5 is shown in Table 5.3. The Lev.n+s model outperformed both Lev.n considerably for the prediction L254 and moderately achieved better *I.ann* values than the Lev.t model. Furthermore, for L253 the Lev.n+s was able to reduce the *I.ann* value to well below the benchmark value of 17% and again was moderately better than the Lev.t prediction.

Table 5.3: Comparison of Lev models identified on I.t, I.n and I.n+s predictions of dataset t.5 using *I.ann*.

Category	Output	Identification Dataset		
		I.t	I.n	I.n+s
		I.ann (%)	*I.ann (%)*	*I.ann (%)*
Lev	L254	1.55	[redacted]	0.50
	L253	14.03	[redacted]	10.73

5.4.2 Comparison using Closed Loop Plant Data

The prediction results of the ANN models (Temp, Comp, and Rev) trained on I.n+s (ANN.n+s), I.n (ANN.n) and I.t (ANN.t) datasets, for closed loop dataset t.1 are compared in Table 5.4. In general the ANN.n+s models again decreased the number of *I.ann* violations of the upper limit values, as indicated by the smaller amount of highlighted values.

Table 5.4: Comparison of ANN models identified on I.t, I.n and I.n+s predictions of dataset t.1 using *I.ann*.

Category	Output	Identification Dataset		
		I.t	I.n	I.n+s
		I.ann (%)	*I.ann (%)*	*I.ann (%)*
Temp	Y282	16.84	30.45	47.28
	Y255	17.95	24.09	40.25
	Y297	8.93	62.82	9.96
	Y257	1.15	5.04	1.50
	Y284	0.64	3.56	0.33
Comp	A254	1.53	33.55	3.86
	A256	0.39	4.84	0.46
	A257	0.68	4.24	0.52
Rev	F256	2.55	27.08	5.57
	F267	10.45	195.77	13.60

The Temp.n+s model had considerably better prediction results than the Temp.n models. For outputs Y297, Y257 and Y284, the Temp.n+s predictions were as reliable as Temp.t model; however, the Temp.n+s model appeared to underestimate Y255 and Y282 with *I.ann* values greater than the Temp.n model. The reason for the wide discrepancy in predicting the outputs is that the Temp.n+s model was trained on the simulated data, which is an emulation of the plant data, hence the actual noise associated with signals Y282 and Y255 are neglected causing the high *I.ann* values. For the Comp.n+s model prediction of dataset t.1, there is a definite improvement compared to Comp.n predictions. For outputs A256 and A257, the introduction of simulated data to the nominal data during the identification step reduced the *I.ann* value well below the

benchmark value of 1.0%. For A254 there was also a significant improvement in the prediction, however, not quite enough to bring the *I.ann* value within the upper value limit of 1.6%. Furthermore, the prediction of the Comp.n+s model is comparable to even to the Comp.t model prediction indicating the significance of adding simulation data to the training step. The Rev.n+s model also experienced considerable improvement in the prediction of variables F267 and F256 compared to the Rev.n model.

From Table 5.5 the prediction results of the Lev models, for closed loop dataset t.3, are compared. Lev.n+s model outperformed both Lev.n and Lev.t models in predicting dataset t.3 with better *I.ann* values for both L254 and L253. For L254 the introduction of the simulation of Test 3 to the training decreased the *I.ann* value well within the benchmark indicating that the model prediction is excellent.

Table 5.5: Comparison of Lev models, identified on datasets I.t, I.n and I.n+s, predictions of dataset t.3 using *I.ann*.

Category	Output	Identification Dataset		
		I.t	I.n	I.n+s
		I.ann (%)	*I.ann (%)*	*I.ann (%)*
Lev	L254	1.74	[illegible]	1.36
	L253	[illegible]	[illegible]	[illegible]

In summary the ANN models identified on nominal plant data and simulated step response data (ANN.n+s models) were able to capture the transient nature of the plant closed loop step test, t.1, with approximately the same reliability as if they were actually identified on t.1 (ANN.t models). Furthermore, there was a significant improvement in the ANN.n+s models ability to predict t.1 compared to the ANN models that were trained on nominal data only (ANN.n models).

5.4.3 Comparison of the ANN Model Prediction of other Datasets

Using dataset t.2, the ANN.n+s models are compared to ANN.n and ANN.t in Table 5.6. From the table, the Temp.n+s model provided very little improvement in predicting most of the outputs compared to the Temp.n model. However, for output Y255 a significant improvement in the *I.ann* values occurred and comparable to the Temp.t prediction of dataset t.2. Furthermore, the Comp.n+s model accurately predicted the distillate compositions in dataset t.2 with *I.ann* values below the benchmark for A254 and A256, which is a significant improvement from the Comp.t predictions. The Rev.n+s model also predicted the outputs (F267 and F256) for dataset t.2 with better accuracy than the other two models.

Table 5.6: Comparison of ANN models, identified on datasets I.t, I.n and I.n+s, predictions of dataset t.2 using *I.ann*.

Category	Output	Identification Dataset		
		I.t	I.n	I.n+s
		I.ann (%)	*I.ann (%)*	*I.ann (%)*
Temp	Y282	9.88	[illegible]	[illegible]
	Y255	5.99	[illegible]	5.96
	Y297	[illegible]	[illegible]	[illegible]
	Y257	[illegible]	[illegible]	[illegible]
	Y284	[illegible]	[illegible]	[illegible]
Comp	A254	1.00	[illegible]	0.92
	A256	[illegible]	[illegible]	0.67
	A257	[illegible]	[illegible]	[illegible]
Rev	F256	[illegible]	[illegible]	22.69
	F267	[illegible]	[illegible]	8.48

5.4.4 Comparison of the ANN Models using Disturbance Data

In Chapter 4, it was concluded that the ANN.t model's prediction of the independent validation dataset t.8 was very poor. Hence, using the recommended integrated modeling strategy; the Temp and Comp models were reinitialized and trained using an augmented dataset (I.t+s) consisting of the plant step response tests (I.t) and a simulation of t.8. These resulting models are labeled as Temp.t+s and Comp.t+s. The predictions of t.8 using the Temp.t+s and Comp.t+s models for various TW252 variables are shown in Table 5.7. Based on the *I.ann* values there are significant improvements in the prediction of the internal tray temperatures and distillate compositions. In fact, for outputs Y255 and A254 the Temp.t+s and Comp.t+s models respectively, predicted dataset t.8 within the *I.ann* upper limit values indicating a reliable prediction. Due to problems in the data acquisition of Test 8, Y297 and A256 could not be predicted.

Table 5.7: Comparison of ANN models, identified on datasets I.t and I.t+s, predictions of dataset t.8 using *I.ann*

Category	Outputs	Identification Dataset	
		I.t	I.t+s
		I.ann (%)	*I.ann* (%)
Temp	Y282	[illegible]	[illegible]
	Y255	[illegible]	9.99
	Y257	[illegible]	[illegible]
	Y284	[illegible]	[illegible]
Comp	A254	[illegible]	1.36
	A257	[illegible]	[illegible]

5.5 Application of the Recommended Modeling Strategy of TW252

While this research was underway, a problem was observed with the L254 controller. An analysis of the closed loop response of the reflux drum level, L254, to +10% and −10% setpoint changes from the nominal steady state value indicated that the control performance of the L254 (condenser liquid level) controller was poor. A study was conducted using the HYSYS model and the developed Lev (Lev.t, Lev.n, Lev.n+s) models to quantitatively assess the L254 control loop performance. The characteristics used to evaluate the loop were as follows:

1. Stable operation with little oscillation
2. No steady state error
3. Fast and smooth setpoint responses
4. Little or no overshoot of setpoint changes

As was discussed in Chapter 3, the HYSYS model was able to emulate the plant setpoint response data of Test 3 (dataset t.3) fairly accurately for variable L254. The results of the HYSYS prediction are reexamined in Figure 5.4, which represents the prediction of the closed loop response of dataset t.3 to only a +10% setpoint change in L254.

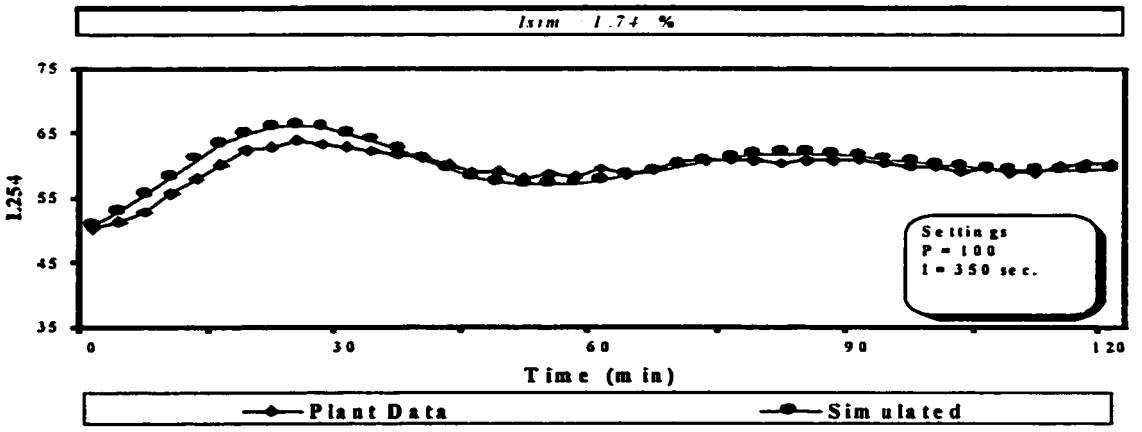


Figure 5.4: Comparison of the HYSYS model prediction of an actual setpoint response for L254 using existing controller settings

From the figure, it is evident that the HYSYS model successfully captured the L254 response characteristics successfully. Furthermore, a control loop performance assessment was performed (Swanda and Seborg 1997) and it was determined that the response of L254 to a setpoint change exhibited an overshoot of 38%. In control theory, a 38% overshoot indicates that the response is very aggressive and controller tuning may be the source of the problem.

Using the benchmarked HYSYS model a simulation was performed to determine new "well tuned" controller settings. In collaboration with University of California, Santa Barbara (UCSB), Anthony Swanda developed new PID controller settings for the L254 controller. The effectiveness of the tuning parameters for L254 was first simulated using the developed HYSYS model. A similar experiment to dataset t.3 was performed using the HYSYS model, which consisted of a +10 % L254 setpoint change. The response of L254 is plotted in Figure 5.5.

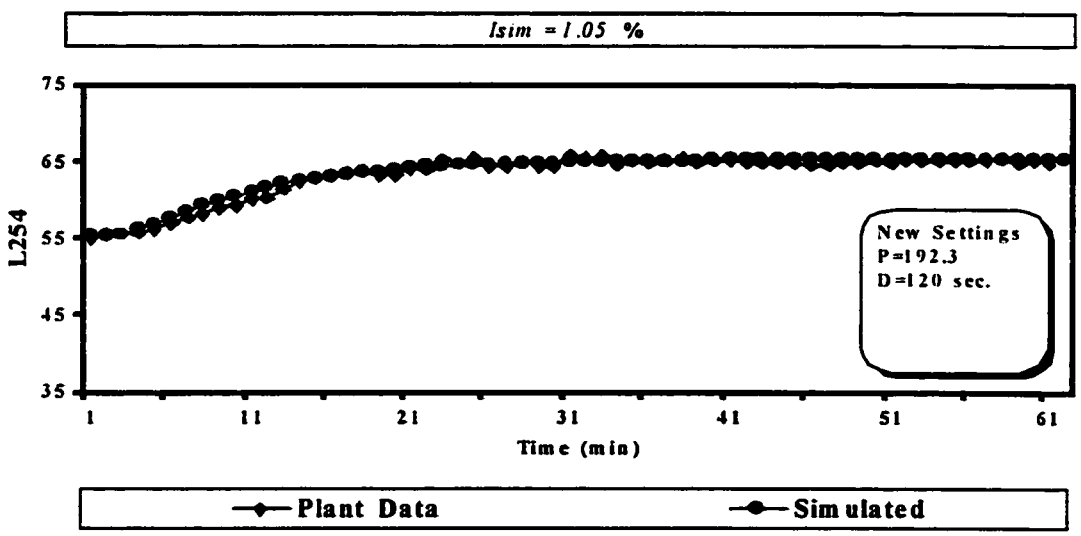


Figure 5.5: Simulated vs. actual L254 setpoint step response using new control settings

From the figure, the response characteristics of the new PI controller for L254 setting simulated in HYSYS were very good. The overshoot was negligible. After simulating the response using the HYSYS model the new tuning parameters were

implemented on the L254 controller for TW252 at the Mizushima Oil Refinery. The same +10% setpoint change in L254 was conducted on the actual column and the data was collected and labeled as Test 7 (t.7) and also plotted on Figure 5.5. For a brief description of Test 7, see Chapter 2. As shown on Figure 5.5, the HYSYS model successfully predicted the response of L254 with an *I.sim* = 1.05 %, which is well within the benchmark value.

For the sake of comparison, the ANN models (Lev.t, Lev.n, Lev.n+s) for the prediction of output variable L254 identified in the previous sections were also tested for their predictive powers on dataset t.7 and plotted in Figure 5.6.

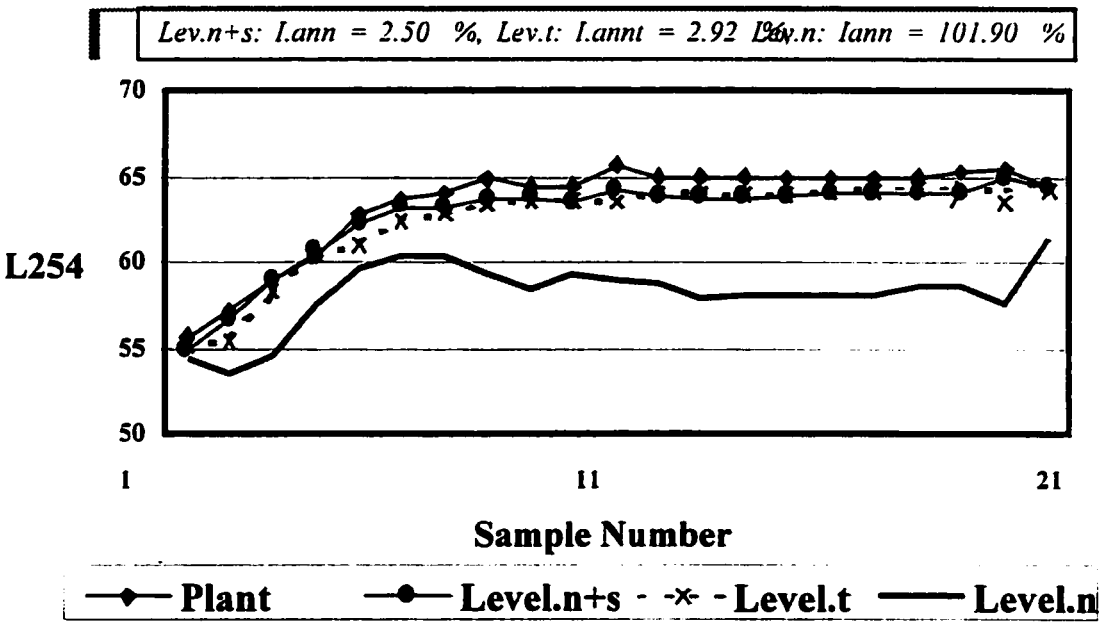


Figure 5.6: Comparison of Lev.t, Lev.n and Lev.n+s model predictions of TW252 condenser level (L254) for dataset t.7.

The Lev.n model predicted L254 for dataset t.7 with an *I.ann* = 101.90%. The Lev.t model predicted L254 with an *I.ann* = 2.92 %. The Lev.n+s model predicted L254 had an *I.ann* = 2.50 %.

5.6 Summary

In this chapter, the concept of integrating a physical model with ANN models was explored. ANN model topologies selected in Chapter 4 were re-identified using nominal data for the internal temperatures (Temp.n), distillate compositions (Comp.n), internal flow rates (Rev.n), and TW252 levels (Lev.n). The prediction accuracy of these models were compared to the ANN.t models, identified on step response data, using a benchmark and upper limit values developed in Chapter 4. In general, the ANN.n models did not predict the excited nature of the step response datasets very accurately because the ANN.n models were identified primarily on steady state operating data. Furthermore, the ANN.n models violated the *I.ann* upper limit boundaries more often than the ANN.t models. On the other hand, once the ANN.n models were re-initialized using simulated transient step response data in conjunction with the nominal data (an augmented dataset, I.n+s). The resulting ANN.n+s models (Temp.n+s, Comp.n+s, Rev.n+s, and Lev.n+s) had much better prediction results on the step response data. Furthermore, on average, the ANN.n+s models were not only better than the ANN.n models, but in some circumstances outperformed ANN.t models and the I.ann violations decreased substantially. An explanation for this phenomenon is that networks trained on "perfect" data (i.e. simulation data) essentially can learn the derivatives of the physical model. If a neural network can learn the derivatives, it can simply look up the derivatives if presented with data similar to the simulation data.

An industrial application of the recommended integrated modeling procedure, that combined neural networks and HYSYS modeling, was shown for the condenser level, L254, where a control loop assessment was done. The HYSYS model was used to predict the response using new tuning parameters for L254. The simulation data generated from this test was augmented with nominal operating data and the Lev.n model was retrained. The results indicated that not only was the HYSYS model able to capture the true dynamic response of L254 to new tuning settings in dataset t.7, but the Lev.n+s model, trained on the augmented dataset, was also able to predict dataset t.7 more accurately than the Lev.t model.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

Using historical nominal and step response data collected from the Japan Energy Corporation, Mizushima Oil Refinery in Japan, a physical model and an ANN model were successfully developed for TW252 (No.2 Xylene splitter). The goal of this research was to compare these two modeling techniques and to suggest and develop an integrated hybrid modeling strategy for TW252.

The physical model was developed using a commercial dynamic and steady-state simulator, HYSYS (v.1.1). A steady-state model was used to evaluate the column's non-linear behavior and the control loop interaction present in TW252's current control structure. The development of the column dynamic model involved using rigorous thermodynamics, and first principle representations of TW252's equipment. An iterative procedure was used to tune the dynamic model. That is the dynamic model, initialized using the steady-state model, was bench-marked using open loop, closed loop, and disturbance data. Characteristics in modeling and the discrepancies in prediction of these tests were discussed in detail. The ability of the HYSYS dynamic model to capture the plant dynamic changes was evidenced by the prediction of the distillate compositions. Furthermore, from the verification of the HYSYS model, benchmarks and upper limit values, based on a performance index (*I.sim*), are recommended for the prediction of TW252 variables.

Artificial neural network models using Process Insights™ were also developed for TW252 for four distinct general ANN models; Temp, Comp, Rev and Lev. It was shown that the designs of the ANN model topologies were very sensitive to the number of inputs and the number of past values incorporated. The optimum topologies for the ANN models were selected using an iterative algorithm that involved varying the number of inputs and then using the correlation coefficient (R^2) to evaluate the ANN models predictive power using independent datasets. After the inputs were selected, the number of past values were investigated using a similar a trial and error procedure that also involved the R^2 coefficient. Using the optimum topologies, the ANN models were identified using setpoint open and closed loop response data (ANN.t models). The ANN.t models were used to predict various open and closed loop changes. Based on the results, new performance indices (*I.ann*) that represented the prediction accuracy of an ANN model were developed independent of the *I.sim* values developed for the HYSYS model. Comparisons between the HYSYS model and ANN.t models, using performance indices *I.sim* and *I.ann*, showed that the ANN.t models predicted the setpoint step response data more accurately than the HYSYS model.

For cases where the collection of step response data that represents the dynamic nature of the process is not possible, the predictive power of ANN models identified on nominal process data were investigated (ANN.n models). The lack of excitation in the training data of the ANN.n models proved to be fatal for prediction of setpoint step responses. It was recommended that the benchmarked HYSYS model be used to aid in the re-identification of the ANN.n models by supplying simulated data representing the step response tests that the ANN.n models were trying to predict. Using an augmented dataset that included nominal and simulated step response tests, the prediction accuracy of the ANN.n+s models are compared to the ANN.t and ANN.n models. The ANN.n+s models were found to enhance the predictive power of the ANN.n models and in some cases made predictions that were comparable to the ANN.t model. Hence, the recommended hybrid modeling strategy verifies that physical models are indeed complimentary to neural network models. The technique of hybrid modeling was applied

to a control loop assessment of TW252's condenser level, L254. The results showed that the Lev.n+s model identified on an augmented dataset extrapolated the model's ability to predict beyond the nominal operating data.

6.2 Future Work

The trust of this thesis was to determine an effective strategy to integrate physical and artificial neural network techniques in modeling an industrial distillation column. Physical models can now aid in the identification of ANN models to represent operating regions beyond nominal conditions. These ANN models can be used to optimize process inputs to obtain desirable outputs, or used to infer quality, or used in a non-linear model predictive control strategy. In the past commercial model predictive control techniques such as Aspen Technologies' DMCplus have successfully applied linear based dynamic models to control large scale multiple input and multiple output (MIMO) systems. With the successful application of the hybrid methodology developed in this thesis, one may infer that some linear models will one day be replaced with non-linear ANN models built using plant and simulated data. Some potential applications and alternative control strategies are considered and described below.

6.2.1 ANN Applications

Recently there has been concern in the petrochemical industry that the frequency of on-line measurement of parameters and/or variables that quantify process behavior is too low. For example, the frequency of sampling of the distillate compositions for TW252 was too low for practical control purposes (the on-line sample rate was every 60 minutes). An alternative to such problems is to use an ANN model that can infer difficult to measure product qualities from secondary process measurements. Hence, if an ANN model could capture the relationship between quality measurements and on-line process variables, the model could be used within a control scheme to infer quality variable(s) at a higher frequency than conventional measurement methods. The concept of inferring a quality variable from other variables is known as "virtual sensing" or "softsensors". The

models developed in this research can be applied as softsensors. For example, the possibility of closed loop inferential control of the distillate quality may be realized if fast and accurate estimates can be obtained from an ANN model. An inferential control strategy uses the inferred estimates of the controlled output ($\hat{y}$) from process measurement (y) to be used directly for feedback control as depicted in Figure 6.1. An interesting continuation of this project would be to evaluate the effectiveness of the ANN models developed via an on-line implementation to predict the distillate quality at a higher sampling frequency than conventional techniques.

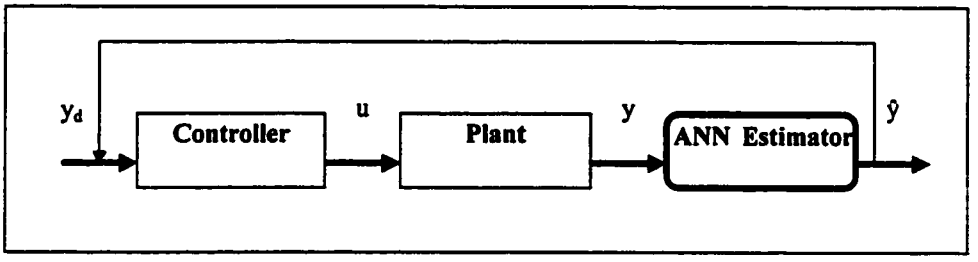


Figure 6.1: An example of an inferential estimator used as a "virtual sensor"

An artificial neural network model could also be used to replace mechanistic or ARMA (Autoregressive moving average with erogenous input) models with the same strengths and weaknesses with very little modification of the control scheme. The popular Internal Model Control (IMC) framework (Garcia and Morari, 1982) provides the typical example of how an ANN model can be used for control purposes. Traditionally, the IMC model of the plant is linear and is inverted to determine the control structure. Psichogios and Ungar (1991) studied the use of an ANN as the controller for direct control or as a plant model for indirect control. For indirect control Psichogios and Ungar (1991) also studied the use of an ANN as the process model within the MPC framework. In short the MPC architecture uses an optimizer to pick a sequence of control actions, u, to minimize the difference between a target, y^{sp}, and the actual value, y, over the next N time steps. In this situation the ANN model could be trained to minimize the multi-step ahead prediction error (N time steps ahead) rather than the one-step ahead prediction error.

6.2.2 Other Empirical Techniques

The use of other empirical methods such as Principle Component Analysis (PCA) or Projection to Latent Structures (PLS) should also be studied as a comparison to ANN modeling. PLS and PCA modeling is a multivariate statistical empirical method that basically finds linear combination of variables that describes major trends in the data. In general, there is a great deal of correlated or redundant information in laboratory and process measurements. PLS and PCA methods extract this information by determining how variables change with respect to one another, i.e. how they co-vary. The co-linearity of the variables are captured into a number of smaller principal components that describe the large portion of the variance in the data. There have been a number of studies on the use of linear PCA models to describe chemical processes (Chen, 1997). For a good overview of the use of these methods, consult papers by Kourti *et al* (1996) and Wise *et al* (1995). Since TW252 contains a large number of process variables that may or may not be dependent on the quality variable, the selection of the appropriate number of inputs to represent the behavior of the output is purely arbitrary. In this thesis, a sensitivity analysis that was proprietary to Process Insights™ was used to help select the necessary input variables to represent the various ANN models developed for TW252. However, using PCA/PLS methods, it may be possible to reduce the relationship between the process variables and quality output variables to a few latent variables that could adequately describe all the important variations and trends in the process. Since TW252 was proven to be a highly non-linear process, future studies could focus on the use of reducing the process data (inputs and outputs) to a fewer latent variables that may be used in the identification process. The reduction of inputs and outputs would eventually exhibit a decrease in the time required to train the ANN.

References

Arkun Y., and Hernandez E., "Study of Control-Relevant Properties of BackPropagation Neural Network Models of Nonlinear Dynamical Systems," *Comp. & Chem. Engg*, Vol. 16, pp. 227 (1992).

Åström, K. J., and McAvoy, T. J., "Towards Intelligent PID Control," *Automatica*, Vol. 28, pp. 1 (1992).

Baratti R., Vacca, G., and Servida, A., "Neural Network Modeling of Distillation Columns", in *Hydrocarbon Processing*, Vol. 74, pp. 6 (1995).

Basualdo, M. S., and Ceccatto, H. A., "Practical Control Methods For Distillation Columns Using Neural Networks," *IFAC Symposium on Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes (DYCORD'95)*, (1995).

Bhat N. and McAvoy T. J., "Use of Neural Nets For Dynamic Modeling And Control of Chemical Process Systems", *Comp. & Chem. Engg.*, Vol. 14, pp. 573 (1990).

Bhat N., Saint-Donat J., McAvoy T. J., "Neural Net Based Model Predictive Control," *Int. J. Control*, Vol. 54, pp. 1453 (1991).

Bhat, N., McAvoy, T. J., Minderman, P. A., and Wang, N. S., "Modeling Chemical Process Systems via Neural Computation," *IEEE Control System Magazine*, Vol. 1, No. 24 (1990).

Birky G. and McAvoy, T., "A neural net to learn the design of distillation controls," *IFAC DYCORD Symp.*, Maastricht, Netherlands (1989).

Bomberger, J.D., and Seborg, D. E., "On-line Updating of Radial Basis Function Network Models," *Proc. IFAC-NOLCOS '95*, Tahoe City, CA, 3-8 (June 1995).

Boozarjomehry, R. B., "Application of Artificial Intelligence In Feedback Linearization," Ph. D. Dissertation, University of Calgary, (1997).

Bremmermann, H. J. and Anderson, R. W., "An Alternative to Backpropagation," *Department of Mathematics*, University of California, Berkeley, (1989).

Bristol, E. H., "On a New Measure of Interactions for Multivariable Process Control," *IEEE Trans. Autom. Cont.*, AC-11, pp.133, (1966).

Chang C., Tsai C., "Dynamic Process Diagnosis via Integrated Neural Networks", *Computers Chem. Eng.* Vol.19, pp. 747 (1995).

Chen S., Billings S. A., and Grant P. M., "Non-linear System Identification Using Neural Networks", *Int. J. Control,* Vol. 51, pp. 1191 (1990).

Chen, C., "Process Monitoring of Two Industrial Distillation Columns Using a Physical Model and Principal Component Analysis," M.S. Thesis. University of California. Santa Barbara (1997).

Chu S. R., and Shoureshi, R., "Neural-Based Identification of Continuous Nonlinear Systems," *Proc. American Control Conference*, San Francisco, CA (1993).

Costanza, V., et al., "An Adaptive Control Strategy for Nonlinear Processes," *Computers Chem. Eng.* Vol. 50, pp. 2041 (1995).

Cybenko, G., "Continuous Valued Networks with Two Hidden Layers are Sufficient," *Math. Cont. Sign. Syst.*, Vol. 2, pp. 303 (1989).

Delgado A., "Dynamic Recurrent Neural Network for System Identification and Control," *IEE Proc.-Control Theory Appl.*, Vol. 45, pp. 307 (1995).

Deshpande P. B., and Ramasamy S., "Consider Neural Networks for Process Identification", *Hydrocarbon Processing*, Vol. 1, pp. 59 (1995).

Dolph, G. A., "Dynamic Simulation for Emergency Control Strategies," *HTI Quarterly*, pp.51 (1995).

Fisher, W. R., and Douglas, J. M., "Analysis of Process Operability at the Preliminary Design Stage," *Comp. Chem. Eng.*, Vol.9, No.5, pp. 499 (1985).

Fruehauf, P. S., and Mahoney, D. P., "Improve Distillation Column Control Design," *Chemical Engineering Progress,* Vol. 90, No. 3, pp. 73 (1994).

GoldFarb S., and Bradley T., "Process Modeling for Design, Commisioning and Operational Analysis," *Comp. & Chem. Engg.*, Vol. 19, pp. 433 (1995).

Guimaaes P. R. B., and McGreavy C., "Flow of information Through An Artificial Neural Network," *Comp. & Chem. Engg.*, Vol. 19, pp. 741 (1995).

Hajare R. P., and Paxton G. E., "Consider Neural Networks for Process Identification," *Hydrocarbon Processing*, Vol. 1, pp. 59 (1995).

Havener, J. and Terhune, K., "Implementation of a Software CEM™ for Nox Emissions Monitoring," *AFRC International Symposium*, Oct (1993).

Hecht-Nielson, R., "Conterpropagation Networks," *Proc. IEEE Int. Conf. on Neural Networks* (1987b).

Hecht-Nielson, R. and Lambert, J., "Application of feedforward and recurrent neural networks to chemical predictive modeling," *I&EC Research*, Vol. 34, No. 4 (1997).

Helget A., Heiderpriem J., and Gariglio D., "Identification and Control of a Simulated Distillation Plant Using Connectionist and Evolutionary Techniques," *Simulation*, Vol. 63, pp. 393, (1994).

Himmelblau, D. M. & MacMurray, J. C., "Modeling and Control of a Packed Distillation Column using Artificial Neural Networks," *Computers Chem. Eng.*, Vol.19, pp. 1077 (1995).

Hornik, K. J., Stinchcombe, D. and White, H., "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, Vol. 2, pp. 359 (1989).

Hoskins, J. C. and Himmelblau, D. M, "Process Control Via Artificial Neural Network and Reinforcement Learning," *Computers Chem. Eng.* Vol. 16, pp. 241 (1992).

Hunt, K. J., and Sbarbaro, R., Zbikowski, R. and Gawthrop, P. J., "Neural networks for Control Systems – A Survey," *Automatica*, Vol.1, pp. 28 (1992).

HYSYS (v1.1) Reference Manual, *Hyprotech Ltd.*, Calgary, Canada, (1995).

Ishida M., and Zhan J., "Neural network Control for a MIMO Process With Time Delay," *Journal of Chem. Eng. of Japan*, Vol. 26, pp. 337, (1993).

Kambhampati C., and Warwick K., "Dynamic Recurrent Neural Network for System Identification and Control," *IEE Proc.-Control Theory Appl.*, Vol. 42, pp. 307, (1995).

Keeler, J. D., "Prediction and Control of Chaotic Chemical Reactions Via Neural Network Models," *Artificial Intelligence in Petroleum Exploration and Production*, Plano, TX (1993).

Keeler, J. D., Hartman E. J., Martin G. D., "Process Modeling and Control Using Neural networks," *Intelligent Control Systems*, Snowmass, CA (1995).

Kourti, T., and MacGregor, J. F., "Multivariate SPC Methods and Process and Product Monitoring," *Journal of Quality Technology*, Vol. 28, pp. 409 (1996).

Kourti, T., Lee, J., and MacGregor, J. F., "Experiences with Industrial Applications of Projection Methods for Multivariable Statistical Process Control," *Computers Chem. Engg.*, Vol. 20, pp. S745 (1996).

Kukarni, S. A., Simulation and Control of a Binary Vacuum Distillation Column, M.S. Thesis, Texas Tech University, (1995).

Lawrence, T., "Process Simulation for Operator Training," M.S. Thesis, University of Calgary, (1996).

Levin A. U., "Recursive Identification Using Feedforward Neural networks," *Int J. Control*, Vol. 61, pp. 533 (1995).

Levin A. U., and Narendra K. S., "Identification Using FeedForward Networks, " *Neural Computation*, Vol. 7, pp. 349 (1995).

Lippmann, R. P., "An Introduction to Computing with Neural Nets", *IEEE ASSP Magazine*, Vol. 1, No. 4, pp. 53 (1987).

Longwell, E.J., "Dynamic Modeling for Process Control and Operability," *Proceedings of the 1993 ISA International Conference*, ISA Research Triangle Park, NC (1993).

Luyben, W. L., *Practical Distillation Control*, 1st Edition, Van Nostrand Reinhold, New York (1992).

Luyben, W.L., *Process Modeling, Simulation, and Control for Chemical Engineers*, 2nd Edition, McGraw Hill (1990).

MacMurray J. C. and Himmelblau D. M., "Modeling and Control of a Packed Distillation Column Using Artificial Neural networks," *Computers Chem. Eng.*, Vol. 19, pp. 1077 (1995).

Marquardt, W., "Dynamic Process Simulation – Recent Progress and Future Challenges," *Proceedings of CPCIV, AIChE*, New York, pp. 131 (1991).

Martin G., and Bhat N., "How Hybrid Modeling Approaches Between Neural Networks and Physical Models can Benefit Customers," *Hyprotech User Conference*, San Antonio, TX (1996).

McAvoy T. J., and Su. H. T., "Integration of Multilayer Perceptron Networks and Linear Dynamic Models: A Hammerstein Modeling Approach," *Ind. Eng. Chem. Res.*, Vol. 32, pp. 1927 (1996).

McAvoy, T. J., "Interaction Analysis Theory and Application," *Instrum. Soc. of America.* Research Triangle Park, NC (1983).

McAvoy, T. J., and Qin, S. J., "Nonlinear PLS Modeling Using Neural Networks". Chem. Engng, Vol. 16, No. 4, pp. 379 (1992).

McGreavy, C., and Guimariaes, P.R.B., "Flow of Information Through an Artificial Neural Network," Computers Chem. Eng. Vol. 19, pp. s741 (1995).

Mogili P., Sunol B.O., Hall L.O., and Camurdan M.C., "Chemical Plant Fault Diagnosis Through A Hybrid Symbolic Connectionist Approach and Comparison with Neural Networks," *Comp. & Chem. Engng.*, Vol. 19, pp. 753 (1995).

Montague G. A., Massimo C. D., and Tham, M. T., "Artificial Neural Network Based Predictive Control," *IFAC Conf. on Adv. Control of Chem. Processes*, Toulouse, France (1991).

Morari, M., and Holcomb, T. R., "PLS/Neural Networks", *Computers Chem. Eng.*, Vol. 16, No. 4, pp.393 (1992).

Morris A. J., Willis M. J., and Montague G. A., "Artificial Neural Networks: Studies in Process Modeling and Control," *Trans. IChemE , Part A*, Vol. 72, No.3, pp. 1 (1994).

Morris A. J., Willis M. J., Tham M. T., Montague G. A., and Massimo C. Di, "Artificial Neural Networks in Process Estimation and Control," *Automatica*, Vol. 28, pp. 1181 (1992).

Morris A.J., Willis M. J., Montague G. A., "Artificial Neural Network Model Based Control," *American Control Conference (ACC94)*, Baltimore, MA (1994).

Munsif, H. P., "Applications of Neural Networks for Distillation Control," Ph.D. Dissertation, Texas Tech University (1995).

Narrendra K. S., Levin A. U., "Control of Nonlinear Dynamical Systems Using Neural Networks- Part II Observability, Identification, and Control," *IEEE Trans. on Neural Networks*, Vol. 7, No. 30 pp. 1 (1996).

Nguyen, D., and Widrow, B., "Improving the Learning Speed of Two-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights." *International Joint Conference on Neural Networks*, San Diego, CA, (1990).

Park S., Song J. J., "Neural Model Predictive Control For NonLinear Chemical Processes," *Journal of Chem. Eng. of Japan*, Vol. 26, pp. 347 (1993).

Parthasarathy K., Narendra K. S., "Neural networks and Dynamical systems." *International Journal of Approximate Reasoning, Vol.* 6, pp. 109 (1992).

Pollard J. F., "Process Identification Using Neural networks," *Computers Chem. Eng.*, Vol. 16, pp. 253 (1992).

Pottman, M. and Seborg, D. E., "A Radial Basis Function Control Strategy and Its Application to a pH Neutralization Process," *presented at European Control Conference* (1993).

Pottman, M. And Seborg, D. E., "Identification of Non-linear Processes using Reciprocal Multiquadric Functions," *J. Proc. Cont.*, Vol. 2, No.4, pp. 189 (1992).

Pottman, M., and Seborg, D. E., "Identification of Nonlinear Process Incorporating a priori Knowledge," *presented at AICHE Annual Meeting,* Miami, FL (1992).

Process Insights (v.3.2) Reference Manual, *Pavilion Technologies Inc.*, Austin, TX (1996).

Psichogios D. C., and Ungar L. H., "Direct and Indirect Model Based Control Using Artificial Neural Networks," *Ind. Eng. Chem. Res.*, Vol. 30, pp. 2564 (1991).

Pulley, R. A., Wainwright, C. E. A., Wilson, J. A. and Jones, S. R., "Combined Neural-Network First Principle Model in Quality Control of Performance Chemicals," *ICHEME Symposium Series*, No. 133, pp. 399 (1996).

Ramasamy, S., Deshphande, P. B., Paxton, and G. E., Hajare, R. P., "Consider Neural Networks for Process Identification" *Hydrocarbon Processing*, pp.59 (June 1995).

Ramchandran, S., and Rhinehart, R. R., "An Introduction to Computing with Neural Nets," *Journal of Process Control*, Vol. 5, No. 115 (1995).

Sabharwal A. "Empirical Modeling using Process Insights®: Optimization and Control of a Xylene Distillation Unit," *Proc. at Pavilion International User's Meeting*, Tokyo, Japan (Nov 1995).

Sabharwal A. "Strategy to Integrate Empirical Modeling and Physical in an Oil Refinery Application," *Proc. at SICE Meeting* Okayama, Japan (Oct 1995).

Sabharwal A., "Dynamic Simulation Modeling and Performance Analysis Using HYSYS: An Oil Refinery Application," *Proc. at Hyprotech User Conference*, San Antonio, TX, (Oct 1996).

Sabharwal A., "Dynamic Simulation Modeling of a Distillation Column Using HYSYS," *Proc. at Hyprotech International User Conference*, Tokyo, Japan (Dec.1995).

Sabharwal A., Wada, T., and Bhat, N. "Benefits of Integrating Empirical Modeling and Physical in an Oil Refinery Application," *Hydrocarbon Processing*, pp. 105 (Oct 1997).

San K. Y., Broussard M. R., and Garrison D. B., "Process Identification Using Neural networks," *Computers Chem. Eng.*, Vol. 16, pp. 253, (1992).

Seborg, D. E., Edgar, T. F. and Mellichamp, D. A., *Process Dynamics and Control*, Wiley, New York (1989).

Sokgestad, S., "Dynamics and Control of Distillation Columns – A Critical Survey", *IFAC Symposium on Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes*, (DYCORD+), College Park, MA (1992).

Stevanovic J. S., "Neural Networks for Process Analysis and Optimization: Modeling and Applications," *Comp. & Chem. Engg.*, Vol. 18, pp. 1149 (1994).

Su, H., Bhat, N., Minderman, P. A., and McAvoy, T. J., "Integrating Neural Networks with First Principle Models for Dynamic Modeling," *IFAC Symposium on Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes*, (DYCORD+), College Park, MA (1992).

Swada, A. and Seborg D., "Evaluation of PID-type Feedback Control Loop Performance using Normalized Setpoint Response Characteristics", IFAC ADCHEM 97 Symposium, Banff, Canada, (1997).

Thibault, J., and Grandjean, B. P. A., "Neural Networks in Process Control – A Survey," Advanced Control of Chemical Processes, *IFAC Symposium Series No. 8*, pp. 251 (1992).

Thompson M. L. and Kramer M., "Modeling Chemical Process Using Prior Knowledge and Neural networks," *AIChE J.*, Vol. 40, pp. 1328, (1994).

Tolliver, T. L., and McCune, L. C., "Distillation Column Control Design Based on Steady State Simulation," *ISA Tran.*, Vol. 17, No. 3, pp.3 (1978).

Trotta A., Barolo M., "Nonlinear Model Based Control of a Binary Distillation Column," *Comp. & Chem. Engng.*, Vol. 19, pp. 519 (1995).

Tsai C., and Chang C., "Dynamic Process Diagnosis Via Integrated Neural Networks," *Comp. & Chem. Engng.*, Vol. 19, pp. 747 (1995).

Tyreus, B. D., and Mahoney, D. P., "Application of Dynamic Simulation," *Proceedings of the Chemical Engineering Chemputers II Conference*, (March 1994).

Ungar L. H., Powell B. A., and Kamens S. N., "Networks for Fault Diagnosis and Process Control, Control," *Comp. & Chem. Engng.*, Vol. 14, pp. 561 (1990)

Vogel, E. F., "An Industrial Perspective on Dynamic Flowsheet Simulation," *Proceedings of CPCIV, AIChE*, New York, pp181 (1991).

Willis, M. J., Montague, G. A. Di Massimo, C., Tham, M. T., and Morris, A. J., "On the Applicability of Neural Networks in Chemical Process Control," *Paper Presented at the AIChE Annual Meeting*, Chicago, IL, Paper 16d (1990).

Willis, M. J., Montague, G. A. Di Massimo, C., Tham, M. T., and Morris, A. J., "Advanced Control of Chemical Processes," *IFAC Symposium Series No. 8*, pp. 261 (1992).

Wise, B. M., N. B. Gallagher, and MacGregor, J.F., "The Process Chemometrics Approach to Process Monitoring and Fault Detection" *IFAC Workshop On-Line Fault Detection and Supervision in the Chemical Process Industries*, June 12-13, Newcastle-Upon Tyne, England (1995).

Xiujuan L., "The Model Validation of Neural networks Describing Non-Linear Dynamical Systems," *2nd Asia Pacific Conf. on Control & Measurement*, China (1995).

Ye X., and Loh N. K., "Dynamic System Identification Using Recurrent Radial Basis Function Network," *American Control Conference*, San Francisco, CA (1993).

Appendix A

Column Specifications

Column Specifications

Table A.1: Design and Operating Conditions for TW252

Design Conditions	TW252
No. of Stages	30
Feed Stage (from Reboiler)	16
Feed quality	Saturated liquid at 160 °C
A264 (C_{9+} Concentration in Feed)	29.50 wt%
A265 (Ethyl Benzene Concentration in Feed)	10.75 wt%
A266 (M/P-Xylene Concentration in Feed)	41.19 wt%
A267 (O-Xylene Concentration in Feed)	18.54 wt%
Internal Reflux Ratio	0.75
Condenser Pressure	550 mmHg
Distillate quality	Less than 1 wt. % of A254
Murphree Stage Efficiency	88 %
Base Case Values used for Steady State Gain Analysis	
Y255 (Tray 27th Temperature)	131.90 °C
Y284 (Bottoms Temperature)	170.20 °C
F256 (Reflux Flowrate)	41.30 m^3/ h
F267 (Steam Flowrate)	10.90 t/h
C_{9+} in Bottoms (no tag)	98.89 wt%
A254(C_{9+} Concentration in the Distillate)	0.79 wt%

Table A.2: DCS Controller Settings for TW252

Controller	CV	MV	MV - span	K_c	τ_I (min)	Mode
L254_LC	L254	F257_FC	-	1.00	5.80	Auto
F257_FC	F257	F257	0-35 m^3/h	0.67	0.85	Cascade
L253_LC	L253	F271_FC	-	1.00	5.80	Auto
F271_FC	F271	F271	0-25 m^3/h	0.67	0.85	Cascade
Y255_TC	Y255	F256_FC	-	1.00	5.80	off
F256_FC	F256	F256	0-50 m^3/h	0.50	0.25	Auto
Y284_TC	Y284	F267_FC	-	1.85	10.00	Auto
F267_FC	F267	F267	0-18 t/h	0.50	0.25	Cascade
P252_PC	P252	Q_{cond}	-	1.25	13.00	Auto

Appendix B

Sensitivity Analysis Results

The following pages contain the sensitivity analysis results from Process Insights™ for the Temp, Comp, Rev, and Lev models.

Table B.1(a): Sensitivity Analysis Ranking for the Temp Models Based on the Average Absolute Sensitivity of Each Input

Rank	Input	Avg. Abs	Avg	Peak
	Output: Y282			
1	P252	0.144	0.144	0.289
2	Y280	0.05	0.05	0.097
3	F256	0.019	-0.019	0.052
4	F267	0.018	-0.018	0.051
5	F284	0.016	0.016	0.023
6	A266	0.01	-0.01	0.029
7	A264	0.008	-0.008	0.019
8	A265	0.002	0	0.006
9	A267	0	0	0.001
	Output: Y255			
1	P252	0.162	0.162	0.274
2	Y280	0.06	0.06	0.099
3	F284	0.021	0.021	0.027
4	F267	0.012	-0.011	0.031
5	F256	0.012	-0.011	0.03
6	A266	0.008	-0.008	0.021
7	A264	0.007	-0.007	0.014
8	A265	0.001	0	0.005
9	A267	0	0	0.002
	Output: Y297			
1	P252	0.163	0.163	0.249
2	Y280	0.074	0.074	0.112
3	F284	0.03	0.03	0.043
4	F256	0.012	-0.006	0.016
5	A266	0.011	-0.011	0.02
6	F267	0.011	-0.004	0.019
7	A265	0.003	0.001	0.005
8	A264	0.002	0.002	0.005
9	A267	0.001	0.001	0.002
	Output: Y257			
1	P252	0.102	0.102	0.149
2	Y280	0.052	0.052	0.078
3	F267	0.045	0.045	0.116
4	F284	0.041	0.041	0.084
5	F256	0.034	0.034	0.11
6	A266	0.016	0.016	0.022
7	A265	0.01	-0.01	0.031
8	A267	0.004	0.003	0.008
9	A264	0.003	0.001	0.008
	Output: Y284			
1	P252	0.068	0.068	0.097
2	F256	0.042	-0.039	0.071
3	Y280	0.034	0.034	0.055
4	F267	0.026	-0.022	0.048
5	F284	0.019	0.019	0.028
6	A265	0.009	0.008	0.014
7	A264	0.003	-0.002	0.009
8	A266	0.002	0.001	0.007
9	A267	0.001	0	0.003

Table B.1(b): The Inputs Selected by Process Insights for the Temp Models

Variable	Points
P252	50
Y280	33
F267	20
F256	17
F284	15

Table B.2(a): Sensitivity Analysis Ranking for the Comp Models Based on the Average Absolute Sensitivity of Each Input

Rank	Input	Avg. Abs	Avg	Peak
	Output: A254			
1	P252	0.039	-0.039	0.047
2	Y255	0.037	0.037	0.048
3	F256	0.035	-0.035	0.046
4	A266	0.029	-0.029	0.039
5	Y257	0.028	0.028	0.037
6	Y284	0.028	-0.028	0.035
7	F284	0.017	0.017	0.024
8	Y280	0.016	0.016	0.021
9	F267	0.01	0.01	0.017
	Output: A256			
1	A266	0.037	-0.037	0.056
2	A264	0.035	-0.035	0.046
3	F267	0.033	-0.033	0.054
4	P252	0.021	-0.021	0.031
5	Y284	0.018	0.018	0.027
6	F256	0.01	-0.01	0.013
7	Y255	0.008	0.008	0.016
8	Y257	0.006	0.002	0.01
9	F284	0.006	0.006	0.008
	Output: A257			
1	Y257	0.088	0.088	0.113
2	P252	0.056	-0.056	0.072
3	F267	0.031	-0.031	0.043
4	Y284	0.012	-0.012	0.017
5	F256	0.012	-0.012	0.022
6	Y282	0.012	-0.012	0.016
7	A266	0.012	-0.011	0.02
8	F284	0.006	-0.004	0.011
9	Y255	0.005	0.004	0.017

Table B.2(b): The Inputs Selected by Process Insights for the Comp Models

Variable	Points
P252	20
A266	13
Y257	11
F267	10
A264	7
Y255	7
F256	6
Y284	4

Table B.3(a): Sensitivity Analysis Ranking for the Rev Models Based on the Average Absolute Sensitivity of Each Input

Rank	Input	Avg. Abs	Avg	Peak
	Output: F256			
1	Y257	0.074	0.074	0.122
2	Y284	0.068	-0.068	0.106
3	P252	0.061	-0.061	0.092
4	Y282	0.054	-0.054	0.083
5	F284	0.054	0.054	0.091
6	A266	0.043	-0.043	0.069
	Output: F267			
1	Y257	0.097	0.097	0.149
2	F284	0.085	0.085	0.136
3	Y284	0.077	-0.077	0.111
4	Y282	0.067	-0.067	0.095
5	Y255	0.056	-0.056	0.078
6	P252	0.055	-0.055	0.077

Table B.3(b): The Inputs Selected by Process Insights for the Rev Models

Variable	Points
Y257	20
Y284	13
F284	8
Y282	6
P252	5
Y255	1

Table B.4(a): Sensitivity Analysis Ranking for the Lev Models Based on the Average Absolute Sensitivity of Each Input

Rank	Input	Avg. Abs	Avg	Peak
	Output: L253			
1	F271	0.393	0.393	0.491
2	F284	0.136	0.136	0.174
3	F257	0.102	-0.102	0.163
4	Y280	0.06	-0.06	0.077
5	A266	0.016	0.016	0.023
6	Y257	0.027	-0.027	0.035
	Output: L254			
1	F285	0.367	0.367	0.439
2	F271	0.053	-0.053	0.097
3	P252	0.05	-0.05	0.06
4	F256	0.042	-0.042	0.054
5	Y257	0.03	-0.03	0.038
6	F284	0.025	0.025	0.039

Table B.4(b): The Inputs Selected by Process Insights for the Comp Model

Variable	Points
F271	17
F285	15
F284	7
P252	5
F256	3

Appendix C

Details of Distillation Dynamic Model Fundamentals

Details of Distillation Dynamic Model Fundamentals

The formulation of dynamic models in HYSYS™ uses the "lumped" parameter approach for all the unit operations. For TW252, HYSYS assumes that on a single tray there are no gradients in temperature, pressure or composition for each phase. (HYSYS Reference, 1995).

The following basic relation can summarize the general mass and energy balance around a system:

Rate of Accumulation of Mass (Energy) = Mass (Energy) flow in – Mass (Energy) Flow out + Mass (Energy) generated

The component balance for an individual tray is given as follows:

$$\frac{dx_i M}{dt} = z_i^{in} F^{in} - x_i F_{out} + M_i^{gen} - M_i^{con} \qquad \text{Equation C.1}$$

For TW252, no chemical components are generated or consumed therefore, M_i^{gen} and M_i^{con} are zero.

The specific potential energy (PE), specific kinetic energy (KE) and pressure-volume work (PV) are negligible, therefore the total energy is equivalent to the specific enthalpy, H. Furthermore, heat can be added or removed in distillation from unit operations such as reboiler and condensers, therefore, W = 0. Hence, the general energy reduces to the following:

$$\frac{d(H^{out} M)}{dt} = H^{in} F^{in} - H^{out} F_t^{out} + Q - W \qquad \text{Equation C.2}$$

Thermodynamic Models

It is essential to have an accurate model or correlation for the vapor-liquid equilibrium (VLE) of all components in the TW252 model. There are many useful sources for obtaining literature data and correlating equations to represent any system (Walas, 1985). For some systems the phase equilibrium is essentially ideal and is governed by Raoult's law:

$$y_i = K_i x_i = \frac{P_i^s}{P}$$ Equation C.3

where: y_i, x_i = vapor, liquid compositions (mole fractions)
P = total pressure (kPa)
P_i^s = vapor pressure (kPa)
K_i = vapor-liquid composition ratio

For chemical systems that have significant non-idealities in the liquid phase a more rigorous VLE model is required that involve *fugacities*. The *fugacities* of vapor and liquid phases are equal at equilibrium and the expression above is written as follows:

$$f_i^V = f_i^L \rightarrow \phi_i^V y_i P = \phi_i^L x_i$$ Equation C.4

where: f_i^V, f_i^L = vapor, liquid fugacity
ϕ_i^V, ϕ_i^L = vapor, liquid fugacity coefficients

If an equation of state (EOS) such as the Peng-Robinson EOS is available that accurately represents the vapor and liquid phases then the equation above is used as the VLE model.

For cases where the liquid phase cannot be modeled by an equation of state the following VLE relationship is applied:

$$\phi_i^V y_i P = \gamma_i x_i P_i^s \exp\left(\frac{v_i^c (P - P_i^s)}{RT}\right) \qquad \text{Equation C.5}$$

where: γ_i = liquid activity coefficient

v_i^c = specific critical volume (m^3/kgmol)

R = gas constant (m^3 kPa/kgmol K)

T = temperature (K)

Equation C.5, includes a liquid phase activity coefficient, that models the non-ideality of the liquid phase and the exponential term is known as the Poynting factor that corrects the liquid fugacity for high pressure.

VLE used for TW252

For TW252 the method used to determine the K-value is different than the model introduced above. The VLE model selected for TW252 is based on the local model, which is an expression developed by Hyprotech Ltd. that sacrifices the details of the above models for speed and is developed from the following equation:

$$K_i P = \frac{y_i}{x_i} P = P_i^S \gamma_i \frac{\phi_i^L}{\phi_i^V} E_i \qquad \text{Equation C.6}$$

The above equation C.6 is modified and expressed in logarithmic form:

$$\ln(K_i) = \ln(P_i^S) + \ln(\gamma_i) + \ln\left(\frac{\phi_i^L}{\phi_i^V}\right) + \ln\left(\frac{E_i}{P}\right) \qquad \text{Equation C.7}$$

The following assumptions are made:

$$\frac{\phi_i}{\phi_i^S} = 1$$

Equation C.8

$$\ln(E_i) = c(\ln P)$$

$$\ln(P^S) = a + \frac{b}{T}$$

$$\ln(\gamma_i) = d(1 - x_i)^2$$

Using these approximations a Local Model, with four parameters, is formulated (HYSYS reference, 1995) as:

$$\ln(K_i) = a_i + \frac{b_i}{T} + c_i \ln(P) + d_i(1 - x_i)^2$$

Equation C.9

The parameter's a_i, b_i, c_i, and d_i are regressed using vigorous VLE predictions.

The Local models, unlike any other models, perform a rigorous update every 600 simulated seconds or if the principle liquid composition changes by 0.05, the pressure changes by more than 50 kPa or temperature changes by more than 5 °C. The model is updated according to a Recursive Least Square Model (HYSYS Reference, 1995).

Enthalpy

The enthalpies of vapor and liquid streams are calculated as a function of temperature, pressure, and composition. TW252's enthalpies are modeled using the Peng-Robinson EOS. In this model the vapour and liquid enthalpies are rigorously determined from the equation of state using the following equation:

$$\Delta H = Pv - RT - \int_{\infty}^{V}\left[P - T\left(\frac{\partial P}{\partial T}\right)_{v}\right]dv \qquad \text{Equation C.10}$$

Substituting the Peng-Robinson equation of state (equation 3.13) the above equation becomes:

$$\frac{H - H^{id}}{RT} = Z - 1 - \frac{1}{2^{1.5}bRT}\left[a - T\frac{da}{dt}\right]\ln\left(\frac{V + (2^{0.5}+1)b}{V + (2^{0.5}-1)b}\right) \qquad \text{Equation C.11}$$

Entropy

For TW252, the liquid and vapor entropy is calculated dynamically using a linear model:

For Vapour:

$$S^{V} = \sum_{i=0}^{n} y_i (r_i^{V} + s_i^{V} T) \qquad \text{Equation C.12}$$

For Liquid:

$$S^{L} = \sum_{i=0}^{n} x_i (r_i^{L} + s_i^{L} T) \qquad \text{Equation C.13}$$

Note, that the for the linear model, HYSYS fits the parameters such as r_i^{L} and s_i^{L} in the above equations to match data predicted by a rigorous VLE calculation that are regressed at the beginning of each dynamic simulation.

Convergence Methods

Simulation of a distillation tower such as TW252 requires solving algebraic equations. HYSYS employs many intuitive methods to solve implicit algebraic expressions such as the Newton-Raphson method. This method requires that an objective

function be written in the form of f(x) = 0. The first derivative of this function is analytically determined which improves the rate of convergence.

$$x_2 = x_1 - \frac{f_1}{f_1'}$$ Equation C.14

The equation above requires an initial estimate of the solution (x_1). An analytical expression gives the derivative of the objective function at the estimate. The next estimate is then computed. The Newton-Raphson is a very powerful method used for thermodynamic calculations, such as the bubble points and dew points.

Equilibrium Bubble Point Calculation

To complete the development of the equilibrium stage model the calculation of the vapor composition and temperature in equilibrium with a known liquid composition and pressure must be performed. That is a bubble point calculation. The Newton-Raphson method is used to converge temperature and compute the equilibrium composition:

$$T_{k+1} = T_k + f \Big/ \left(\sum_i \frac{\partial y_i}{\partial T} \right)$$ Equation C.15

The function (f) in the above expression is unity minus the sum of the calculated vapor compositions:

$$f = 1 - \sum_i y_i$$ Equation C.16

The derivative of this function with respect to temperature using a simplified form of equation C.5 for y_i that neglects the Poynting factor and assumes the vapor phase is ideal (ϕ_i^V=1), can be given as:

$$\frac{\partial f}{\partial T} = -\sum_i \frac{\partial y_i}{\partial T} = -\sum \frac{\gamma_i x_i}{P} \frac{\partial P_i^s}{\partial T} \qquad \text{Equation C.17}$$

The Antoine equation is used to compute the vapor pressure (P_i^S):

$$P_i^s = \exp\left(C_{1,i} + \frac{C_{2,i}}{(T + C_{3,i})} \right) \qquad \text{Equation C.18}$$

The liquid activity coefficient in equation C.17 is a weak function of temperature and is excluded from the analytical derivative. This method when used to compute the tray bubble point temperature and equilibrium vapor composition has been shown to be very fast and robust (HYSYS Reference, 1995).

Appendix D

Initial Steady-State Conditions for Step Response Tests

Steady state initial conditions for all the step response tests outlined in Table 2.3 used to initiate the HYSYS model is summarized in this appendix. Furthermore, the predicted steady state values of the HYSYS model for these step tests are also included.

Table D.1: Comparison of Actual and Simulated Initial Conditions for Dataset t.1

	Statistics for 28 Recorded 3 minute ACS Data Points					
Tag Name	***Min***	***Max***	***Std***	***Mean***	***HYSYS***	***Error (%)***
F284	35.83	37.35	0.398	36.80	36.89	0.23
F256	41.11	41.59	0.107	41.33	41.33	-0.01
F267	11.14	11.44	0.067	11.34	11.35	0.01
F271	8.90	13.29	1.099	10.81	10.56	-2.34
F285	25.77	27.74	0.644	26.73	26.71	-0.08
Y280	156.60	157.50	0.283	157.02	157.00	-0.01
Y281	138.20	138.80	0.198	138.48	138.50	0.01
Y297	135.80	136.40	0.148	136.06	136.05	-0.01
Y255	131.60	131.90	0.093	131.74	131.80	0.05
Y282	128.80	129.20	0.101	129.03	129.00	-0.02
Y256	50.00	53.20	0.868	51.24	51.10	-0.28
Y257	165.00	167.00	0.677	166.36	166.70	0.20
Y284	170.00	171.30	0.463	170.74	170.70	-0.02
L253	56.80	62.20	1.718	59.73	60.00	0.45
L254	48.70	50.80	0.652	49.86	49.80	-0.12
P252	553.60	556.30	0.780	555.09	554.90	-0.03
A254	0.66	0.75	0.045	0.71	0.71	-0.06
A255	16.21	16.27	0.027	16.24	16.24	0.00
A256	56.90	57.10	0.055	57.03	57.00	-0.05
A257	27.79	27.81	0.005	27.80	27.80	0.01
A264	25.81	26.95	0.364	26.53	26.53	-0.02
A265	11.51	11.89	0.116	11.61	11.55	-0.54
A266	41.24	41.98	0.241	41.45	41.28	-0.41
A267	20.22	20.38	0.059	20.30	20.31	0.06

Table D.2: Comparison of Actual and Simulated Initial Conditions for Dataset t.2

Statistics for 51 Recorded 3 minute ACS Data Points						
Tag Name	***Min***	***Max***	***Std***	***Mean***	***HYSYS***	***Error (%)***
F284	35.72	38.22	0.668	36.89	36.32	-1.54
F256	42.11	42.66	0.117	42.40	42.38	-0.05
F267	11.15	11.39	0.051	11.25	11.27	0.21
F271	9.98	12.29	0.615	11.16	9.31	-16.60
F285	25.57	26.58	0.259	26.13	26.92	3.01
Y280	157.60	158.10	0.114	157.86	157.75	-0.07
Y297	138.10	138.60	0.127	138.37	138.66	0.21
Y255	131.50	131.90	0.092	131.68	130.56	-0.85
Y282	128.80	129.20	0.103	129.05	128.64	-0.32
Y256	52.40	55.40	0.622	53.92	54.36	0.82
Y257	165.60	166.60	0.282	166.17	166.00	-0.10
Y284	170.20	170.90	0.161	170.63	170.44	-0.11
L253	57.10	64.50	1.808	60.00	60.00	0.00
L254	48.80	50.80	0.460	49.99	50.00	0.03
P252	553.50	556.80	0.822	555.03	554.96	-0.01
A254	0.74	0.86	0.039	0.82	0.80	-2.28
A255	16.23	16.71	0.137	16.43	15.81	-3.76
A256	56.70	56.80	0.048	56.77	56.42	-0.61
A257	26.71	27.54	0.285	27.36	27.06	-1.09
A264	26.73	27.27	0.073	27.22	26.15	-3.94
A265	11.55	11.70	0.060	11.62	11.70	0.68
A266	41.15	41.39	0.096	41.26	41.68	1.03
A267	19.96	20.14	0.033	20.00	20.00	0.00

Table D.3: Comparison of Actual and Simulated Initial Conditions for Dataset t.3

	Statistics for 26 Recorded 3 minute ACS Data Points					
Tag Name	***Min***	***Max***	***Std***	***Mean***	***HYSYS***	***Error (%)***
F284	38.97	39.98	0.278	39.46	38.90	-1.41
F256	42.05	42.63	0.122	42.30	42.30	0.00
F267	11.05	11.18	0.033	11.10	11.08	-0.17
F271	13.67	14.65	0.265	14.06	11.79	-16.15
F285	22.53	26.29	0.866	25.74	27.12	5.36
Y280	159.40	159.80	0.137	159.60	159.73	0.08
Y297	138.60	139.00	0.099	138.75	138.92	0.12
Y255	132.00	132.20	0.070	132.14	130.90	-0.94
Y282	129.10	129.40	0.070	129.26	129.11	-0.11
Y256	53.10	54.10	0.238	53.74	53.70	-0.07
Y257	163.80	164.50	0.208	164.20	166.06	1.13
Y284	169.70	170.30	0.158	170.05	170.16	0.06
L253	59.10	61.30	0.528	59.94	60.00	0.11
L254	49.30	52.70	0.696	50.19	50.00	-0.38
P252	559.10	561.30	0.530	560.13	560.00	-0.02
A254	0.68	0.72	0.016	0.69	0.69	0.36
A255	16.15	16.40	0.102	16.32	15.63	-4.25
A256	58.10	58.40	0.117	58.32	56.31	-3.44
A257	25.89	26.34	0.184	26.02	27.36	5.15
A264	30.00	30.61	0.298	30.35	30.05	-0.99
A265	10.53	10.73	0.094	10.62	10.87	2.37
A266	38.96	39.31	0.171	39.11	39.20	0.22
A267	18.55	18.68	0.061	18.60	19.21	3.26

Table D.4: Comparison of Actual and Simulated Initial Conditions for Dataset t.4

	Statistics for 25 Recorded 3 minute ACS Data Points					
Tag Name	***Min***	***Max***	***Std***	***Mean***	***HYSYS***	***Error (%)***
F284	36.81	42.10	0.864	39.25	39.24	-0.03
F256	40.86	41.63	0.117	41.29	41.30	0.01
F267	10.87	11.06	0.040	10.96	11.00	0.40
F271	10.39	16.04	0.939	13.41	13.05	-2.70
F285	25.54	27.17	0.354	26.27	26.20	-0.26
Y280	159.00	159.90	0.162	159.44	157.80	-1.03
Y297	135.50	136.60	0.167	136.26	138.20	1.43
Y255	131.80	132.30	0.091	132.06	130.80	-0.95
Y282	128.90	129.60	0.113	129.24	128.90	-0.26
Y256	52.30	57.50	1.316	54.50	54.60	0.19
Y257	163.80	165.80	0.491	164.77	165.00	0.14
Y284	169.90	171.20	0.277	170.52	168.30	-1.30
L253	53.00	65.70	1.910	59.90	60.00	0.16
L254	48.40	51.80	0.620	50.00	50.00	-0.01
P252	555.00	563.00	1.076	559.97	560.00	0.01
A254	0.67	0.82	0.047	0.74	0.75	1.78
A255	15.82	16.32	0.144	16.04	15.37	-4.15
A256	57.80	58.60	0.241	58.20	57.70	-0.85
A257	25.85	26.93	0.340	26.37	26.30	-0.26
A264	28.15	29.36	0.292	28.74	29.05	1.06
A265	10.73	11.11	0.102	10.93	10.90	-0.24
A266	39.58	40.35	0.212	40.01	41.01	2.50
A267	18.60	18.86	0.075	18.72	19.02	1.58

Table D.5: Comparison of Actual and Simulated Initial Conditions for Dataset t.5

	Statistics for 50 Recorded 3 minute ACS Data Points					
Tag Name	***Min***	***Max***	***Std***	***Mean***	***HYSYS***	***Error (%)***
F284	37.13	39.14	0.496	38.31	37.95	-0.94
F256	43.21	43.88	0.153	43.50	43.26	-0.53
F267	10.86	11.06	0.047	10.95	11.10	1.41
F271	11.97	14.73	0.630	13.65	11.51	-15.64
F285	24.09	26.04	0.496	24.95	26.44	5.99
Y280	160.40	161.00	0.153	160.72	159.73	-0.62
Y297	136.10	136.90	0.230	136.51	140.23	2.73
Y255	132.00	132.50	0.102	132.20	131.20	-0.75
Y282	129.00	129.60	0.151	129.33	129.28	-0.04
Y256	57.50	62.60	1.000	60.26	57.76	-4.14
Y257	163.90	166.70	0.960	165.56	166.37	0.49
Y284	169.60	171.10	0.437	170.46	170.70	0.14
L253	57.40	62.50	1.219	60.07	60.00	-0.12
L254	47.00	51.70	0.862	49.85	50.00	0.30
P252	554.40	563.90	1.924	560.16	560.00	-0.03
A254	0.82	1.01	0.070	0.95	0.99	4.79
A255	15.66	15.95	0.109	15.73	15.48	-1.58
A256	57.10	57.70	0.215	57.31	56.37	-1.64
A257	26.94	27.66	0.271	27.47	27.16	-1.13
A264	30.66	31.24	0.147	30.85	31.22	1.21
A265	10.57	10.72	0.067	10.67	10.75	0.76
A266	38.87	39.28	0.160	39.08	39.14	0.14
A267	18.80	18.89	0.034	18.84	18.89	0.26

Table D.6: Comparison of Actual and Simulated Initial Conditions for Dataset t.9

	Statistics for 20 Recorded 3 minute ACS Data Points					
Tag Name	***Min***	***Max***	***Std***	***Mean***	***HYSYS***	***Error (%)***
F284	-	-	0.537	37.40	37.10	-0.80
F256	39.92	40.42	0.134	40.15	40.20	0.13
F267	10.67	10.86	0.041	10.75	10.73	-0.19
F271	12.52	14.61	0.704	13.53	13.48	-0.39
F285	23.54	24.36	0.227	23.99	23.62	-1.55
Y280	160.95	161.41	0.120	161.19	160.70	-0.31
Y297	143.39	143.98	0.136	143.69	145.42	1.21
Y255	136.93	137.21	0.070	137.06	136.20	-0.63
Y282	134.59	134.92	0.091	134.73	134.29	-0.33
Y256	49.48	53.47	1.374	51.31	52.53	2.38
Y257	178.22	180.56	0.539	179.18	170.93	-4.60
Y284	172.61	173.32	0.191	172.97	173.24	0.16
L253	59.63	66.79	2.315	63.17	60.00	-5.02
L254	54.21	55.35	0.289	54.96	55.00	0.08
P252	649.27	651.47	0.567	650.34	650.00	-0.05
A254	0.76	0.87	0.052	0.80	0.88	9.52
A255	15.45	15.53	0.039	15.50	15.39	-0.69
A256	57.37	57.67	0.138	57.55	57.25	-0.52
A257	25.93	26.19	0.130	26.03	26.40	1.41
A264	36.20	36.56	0.146	36.28	36.31	0.09
A265	9.70	9.71	0.002	9.71	9.75	0.44
A266	36.47	36.78	0.122	36.71	36.31	-1.09
A267	17.28	17.35	0.027	17.33	16.82	-2.96

Table D.7: Comparison of Actual and Simulated Initial Conditions for Dataset t.10

	Statistics for 30 Recorded 3 minute ACS Data Points					
Tag Name	***Min***	***Max***	***Std***	***Mean***	***HYSYS***	***Error (%)***
F284	39.24	39.82	0.173	39.50	38.85	-1.65
F256	41.73	42.19	0.110	42.00	42.00	0.01
F267	11.94	12.16	0.043	12.07	12.08	0.11
F271	11.43	12.18	0.174	11.85	11.61	-2.06
F285	26.42	27.19	0.197	26.74	27.24	1.87
Y280	162.36	162.90	0.105	162.58	163.73	0.71
Y297	142.69	143.13	0.098	142.90	143.84	0.66
Y255	136.72	136.86	0.028	136.78	136.02	-0.55
Y282	134.28	134.71	0.103	134.47	134.57	0.07
Y256	33.82	35.84	0.571	34.67	35.52	2.45
Y257	200.01	202.53	0.605	201.19	170.75	-15.13
Y284	171.37	171.81	0.123	171.58	174.35	1.61
L253	54.16	55.75	0.451	54.94	55.80	1.57
L254	64.08	65.85	0.417	64.98	64.75	-0.36
P252	649.04	651.47	0.575	649.94	650.00	0.01
A254	0.60	0.64	0.017	0.62	0.58	-5.98
A255	15.76	15.88	0.042	15.81	14.85	-6.10
A256	58.94	59.00	0.019	58.99	56.76	-3.78
A257	24.32	24.50	0.065	24.42	27.80	13.85
A264	30.19	30.36	0.080	30.27	31.23	3.17
A265	10.62	10.73	0.052	10.68	10.11	-5.35
A266	40.05	40.18	0.060	40.13	38.73	-3.48
A267	18.86	18.93	0.034	18.89	19.23	1.81

Appendix E

TW252 Datasets Collected

For all of TW252 variables as listed in Table 2.1, the step response, nominal and disturbance tests as described in Table 2.3 are plotted in this appendix. Furthermore, the plots are arranged by identification data sets, I.f, I.n, I.n+s and I.t+s and their corresponding statistics are also tabulated.

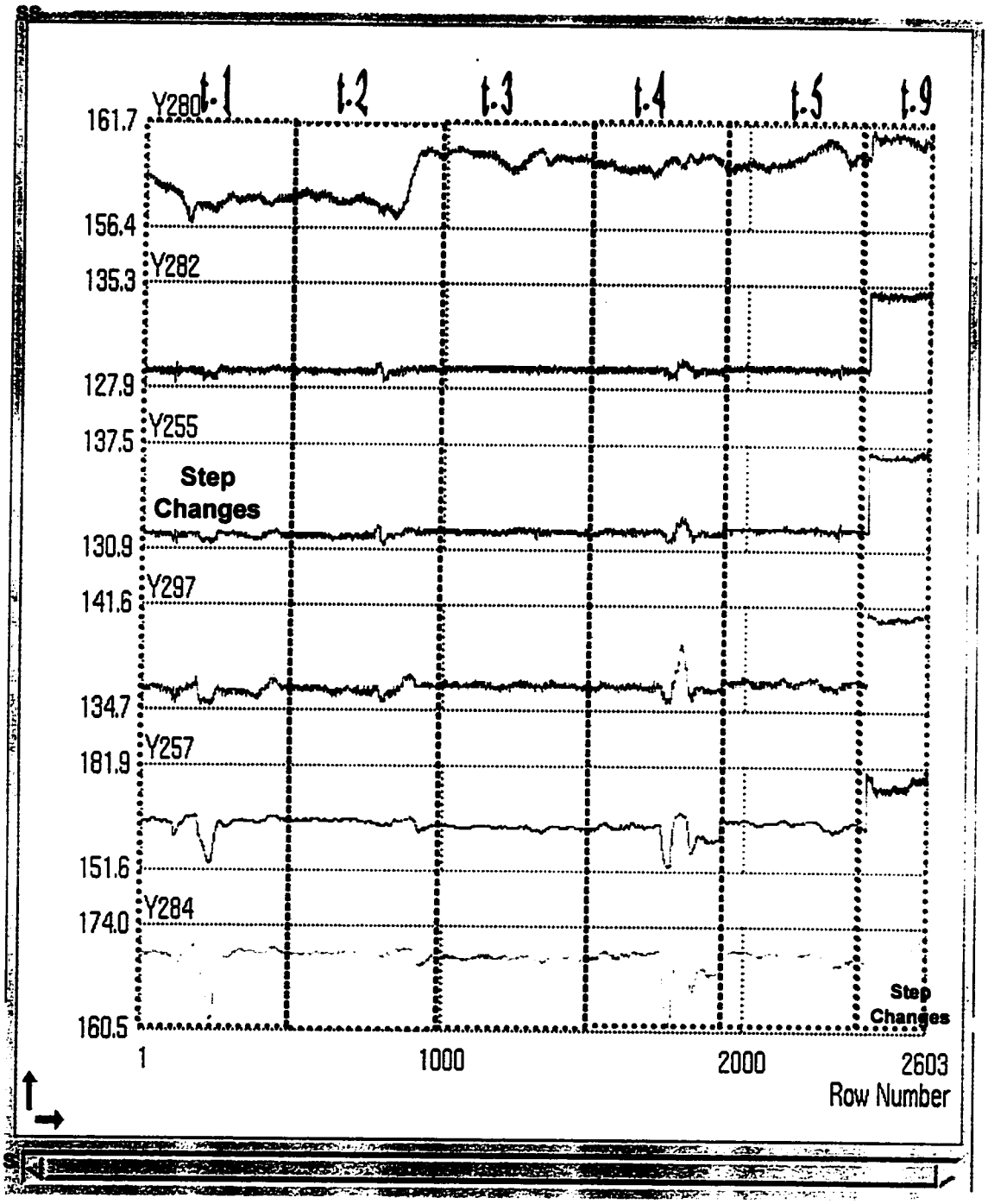


Figure E.1(a): TW252 Temperatures for Identification dataset I.t

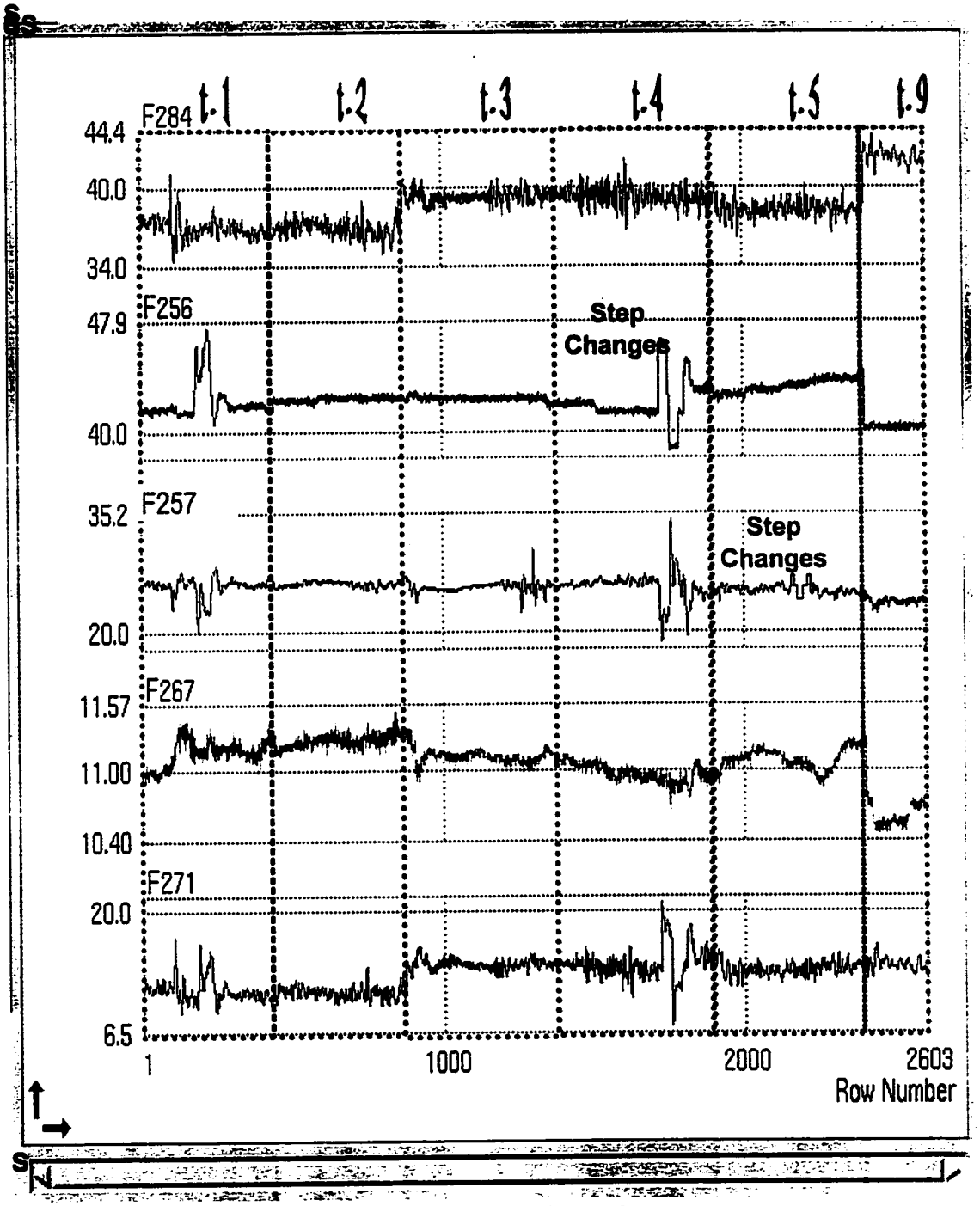


Figure E.1(b): TW252 flow rates for Identification dataset I.t

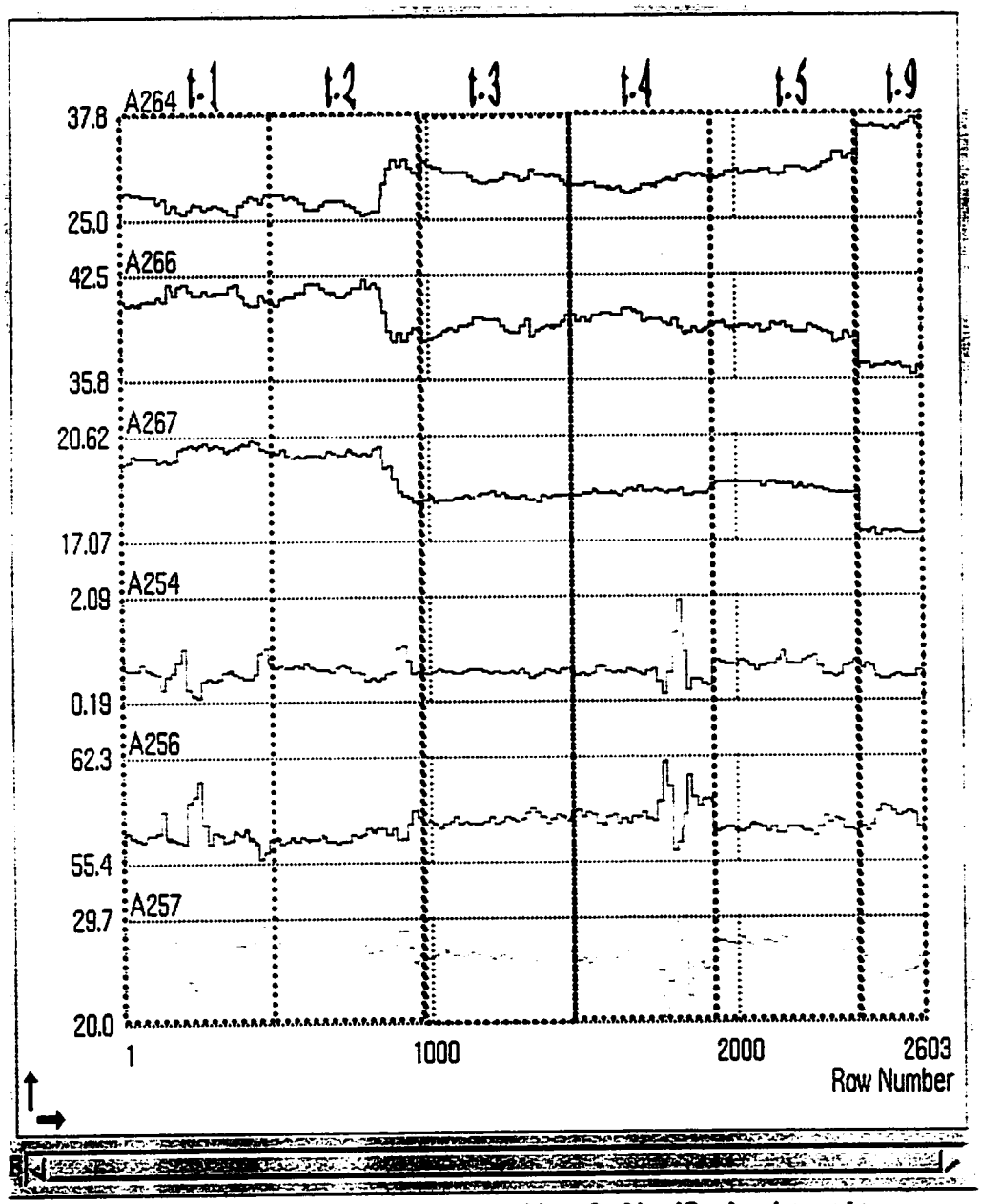


Figure E.1(c): TW252 distillate and feed compositions for Identification dataset I.t

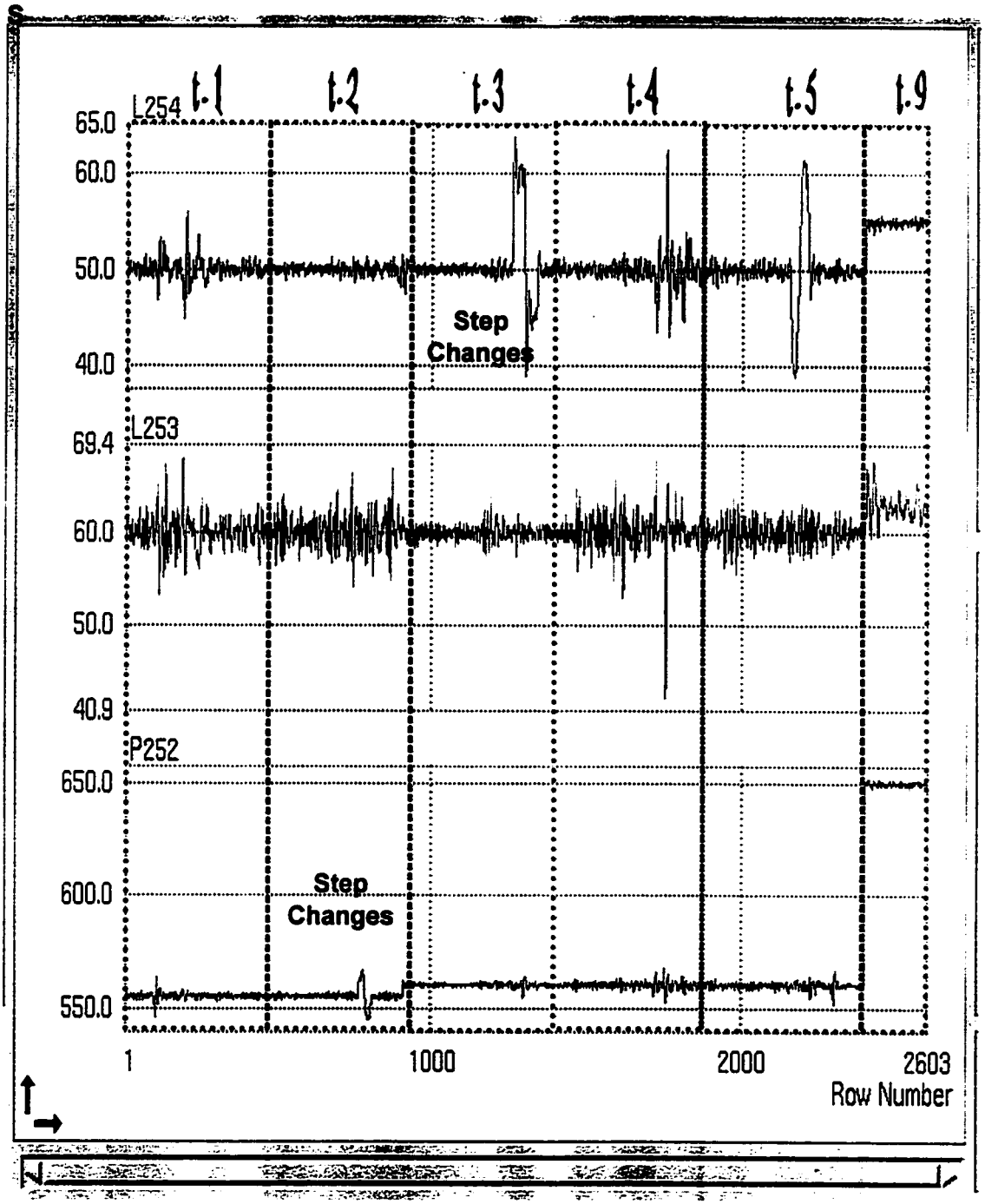


Figure E.1(d): TW252's pressure, condenser and reboiler level for Identification dataset I.t

Table E.1: TW252 Variable Statistics for Dataset I.t

Statistics for 2603 Recorded 3 minute ACS Data Points				
Tag Name	***Min***	***Max***	***Mean***	***Std***
F284	34.45	43.93	38.53	1.59
F256	38.54	47.44	42.21	1.14
F267	10.61	11.52	11.11	0.14
F271	7.21	21.08	13.01	1.66
F257	18.52	34.38	25.52	1.27
Y280	156.60	161.40	159.26	1.15
Y297	135.00	141.30	136.58	1.30
Y255	131.20	137.20	132.34	1.36
Y282	128.20	134.90	129.59	1.48
Y256	131.20	137.20	132.34	1.36
Y257	153.00	180.60	165.66	3.92
Y284	161.10	173.40	170.13	1.49
L253	42.20	68.10	60.24	1.81
L254	38.80	63.80	50.47	2.52
P252	544.40	652.10	565.26	24.81
A254	0.28	2.00	0.75	0.19
A255	14.76	18.23	16.08	0.48
A256	55.70	62.00	57.84	0.91
A257	20.47	29.24	26.58	1.27
A264	25.57	37.19	29.74	2.65
A265	9.39	12.12	10.90	0.62
A266	36.14	42.23	39.72	1.40
A267	17.23	20.46	19.05	0.82

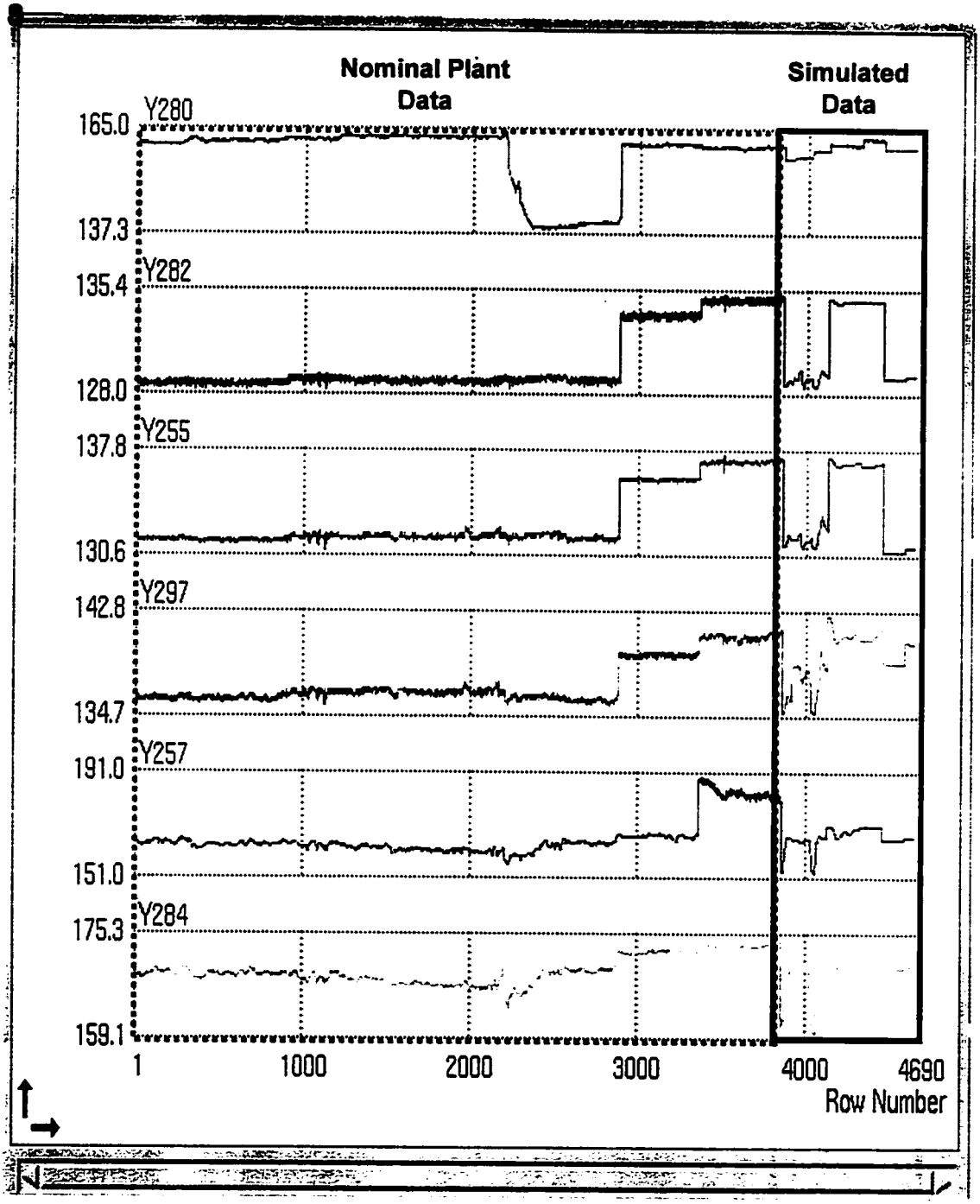


Figure E.3(a): TW252 temperatures for Identification dataset I.n+s

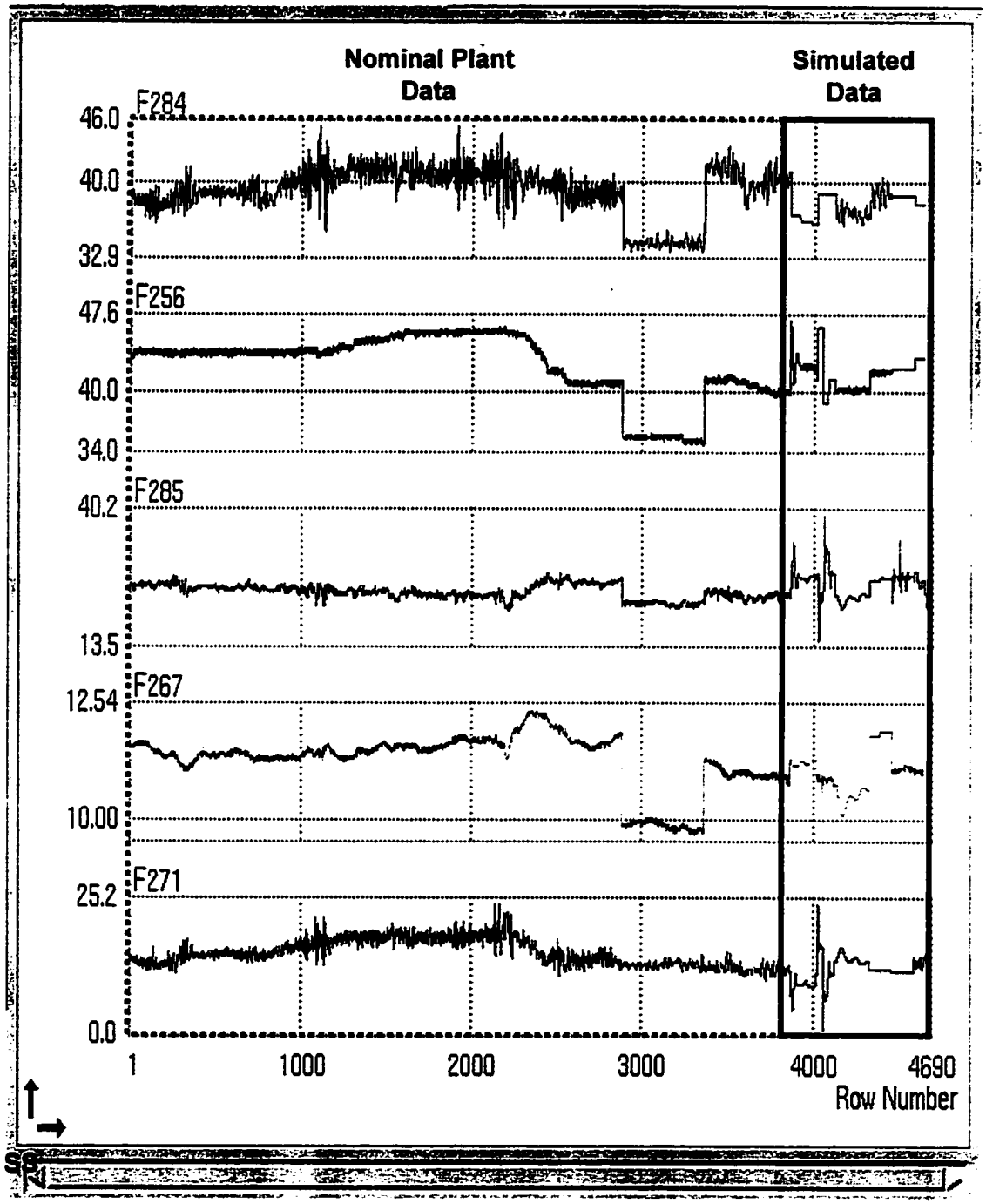


Figure E.3(b): TW252 flowrates for Identification dataset I.n+s

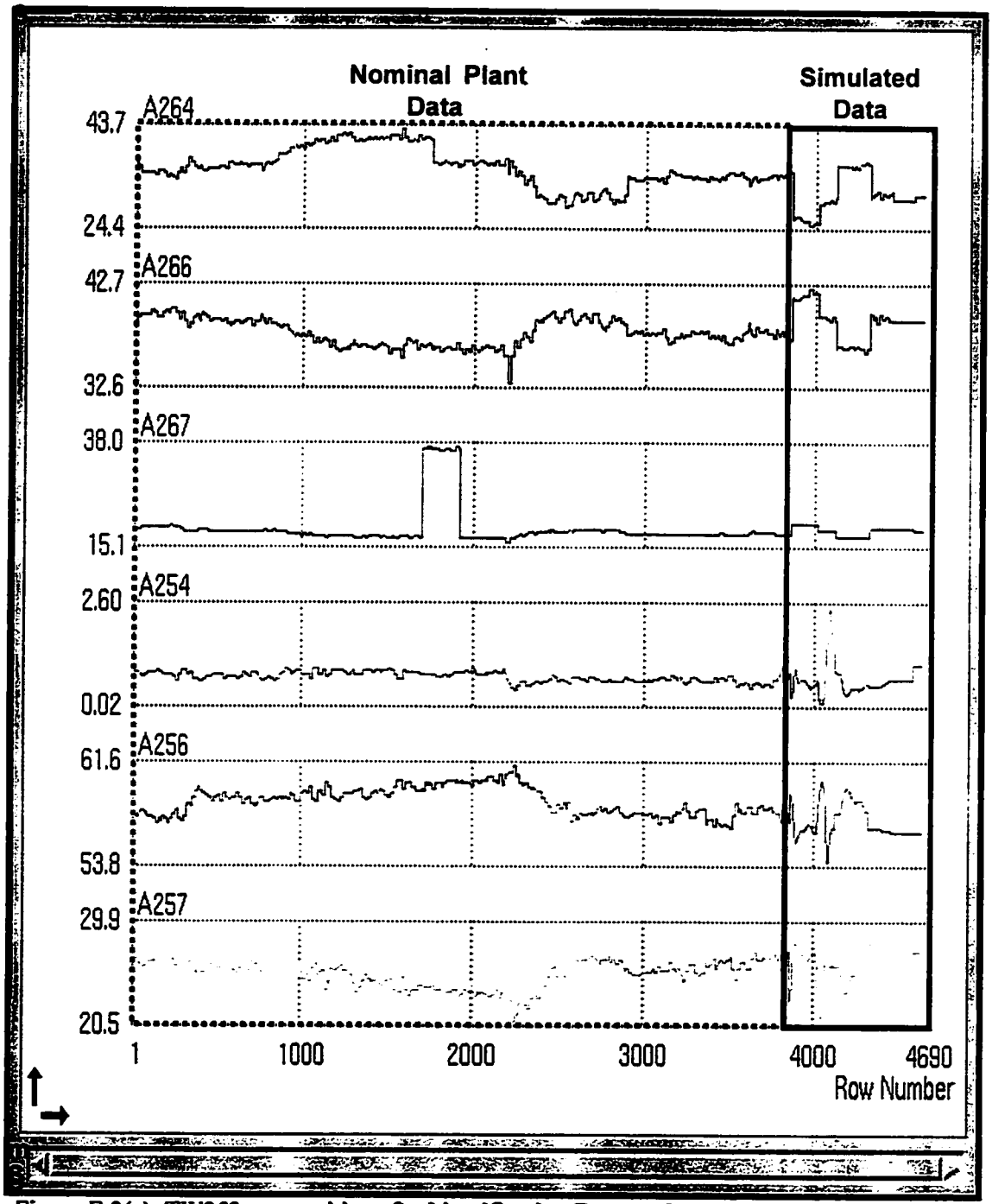


Figure E.3(c): TW252 compositions for Identification Dataset, I.n+s

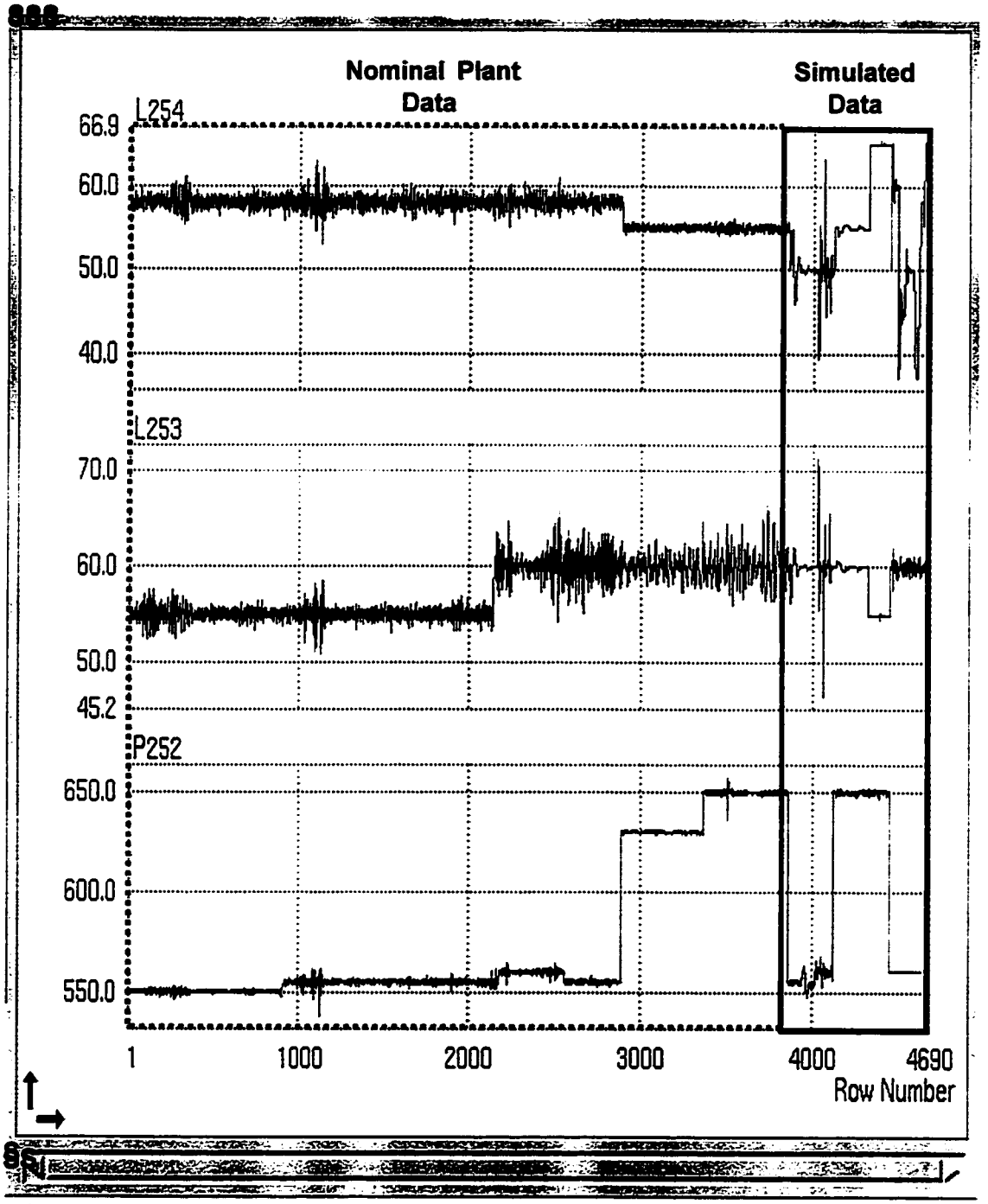


Figure E.3(d): TW252 pressure condenser and reboiler levels for Identification dataset I.n+s

Table E.2: TW252 Variable Statistics for Dataset I.n

Statistics for 3858 Recorded 3 minute ACS Data Points				
Tag Name	***Min***	***Max***	***Mean***	***Std***
F284	33.50	45.43	39.30	2.28
F256	34.62	46.41	42.54	3.21
F267	9.66	12.40	11.29	0.61
F271	10.26	24.07	15.08	2.39
F257	20.60	27.88	24.08	1.34
Y280	138.60	163.70	157.92	7.94
Y297	135.50	141.60	137.26	1.77
Y255	130.90	137.40	132.92	2.07
Y282	128.30	135.10	130.17	2.34
Y256	36.70	57.80	47.29	3.52
Y257	155.90	189.20	165.73	7.23
Y284	163.60	174.60	169.61	2.27
L253	51.00	67.10	57.23	2.72
L254	53.10	63.00	57.24	1.46
P252	537.50	658.00	575.91	37.93
A254	0.45	1.00	0.76	0.11
A255	14.40	18.46	16.27	0.68
A256	56.60	61.20	58.59	1.00
A257	20.94	27.20	25.24	1.22
A264	28.30	42.82	35.72	3.43
A265	8.79	11.81	10.27	0.61
A266	33.08	40.19	37.96	1.24
A267	16.16	36.96	19.10	4.34

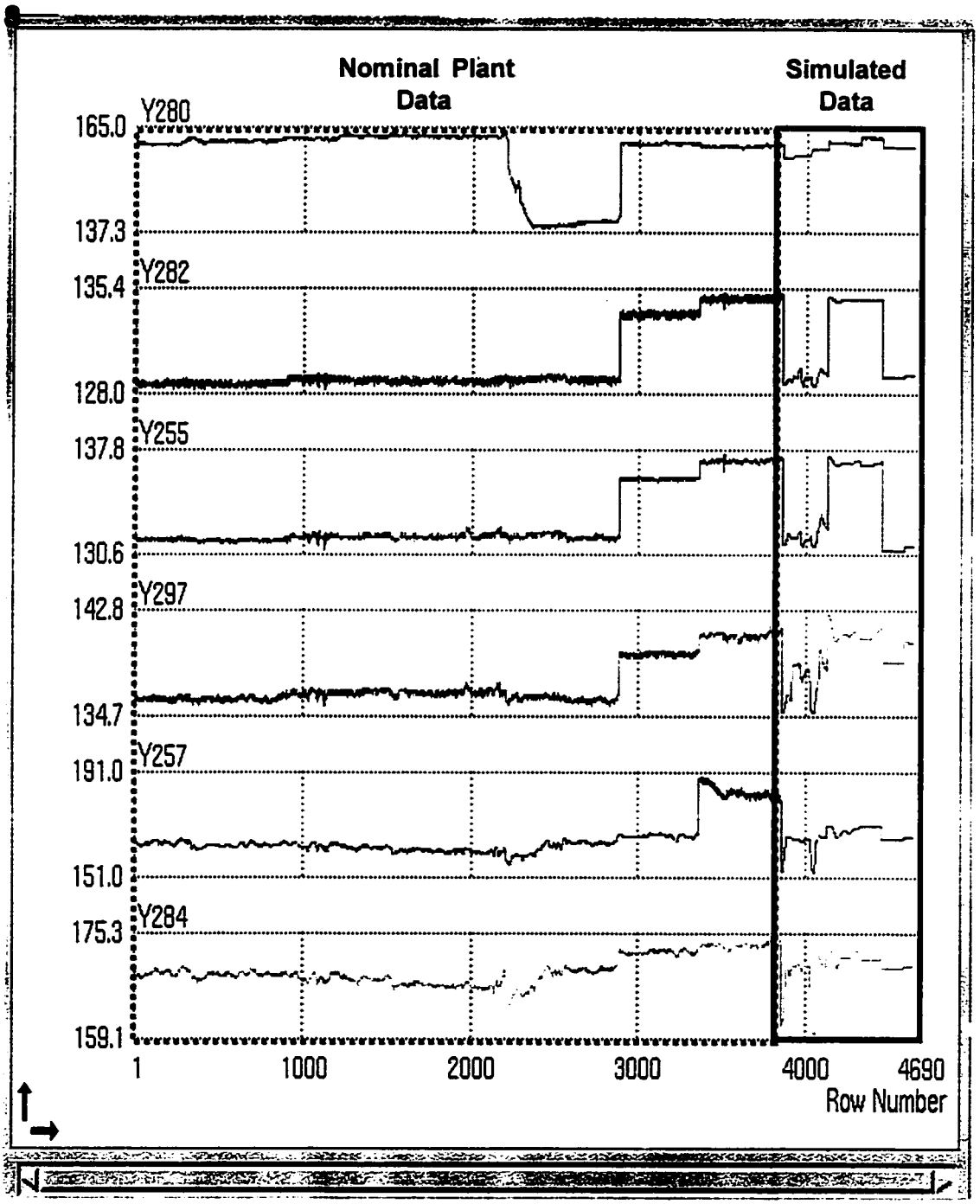


Figure E.3(a): TW252 temperatures for Identification dataset I.n+s

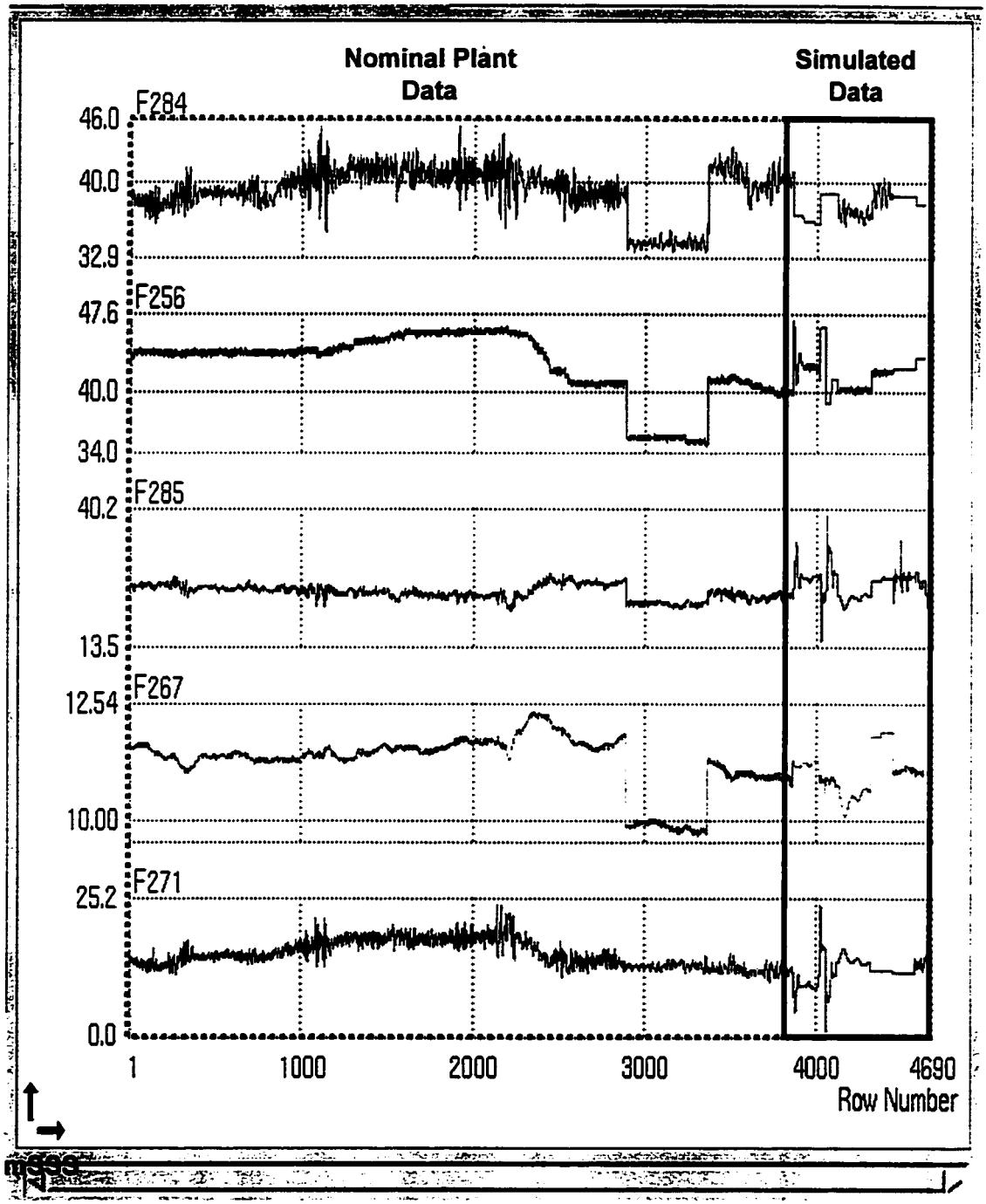


Figure E.3(b): TW252 flowrates for Identification dataset I.n+s

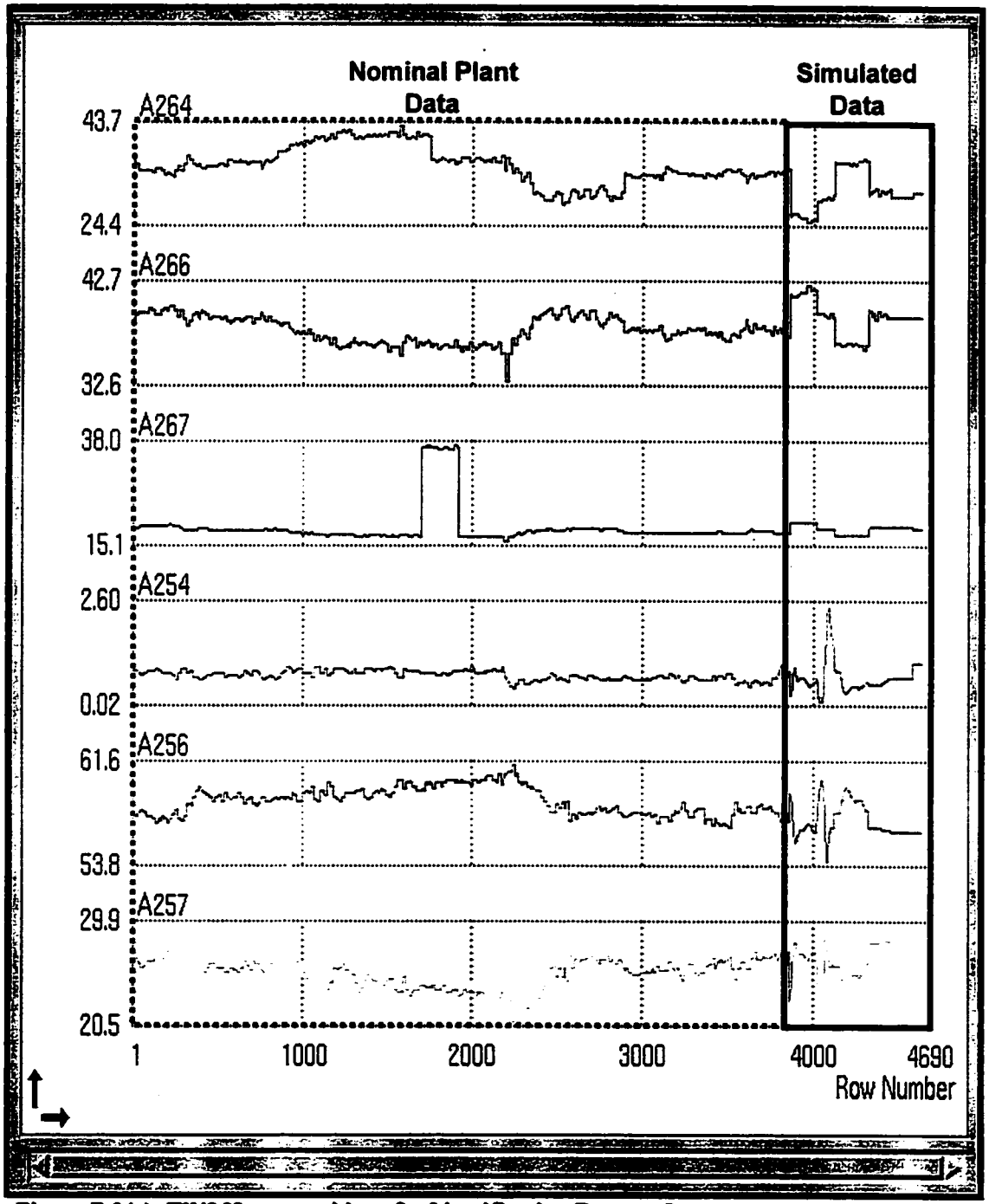


Figure E.3(c): TW252 compositions for Identification Dataset, I.n+s

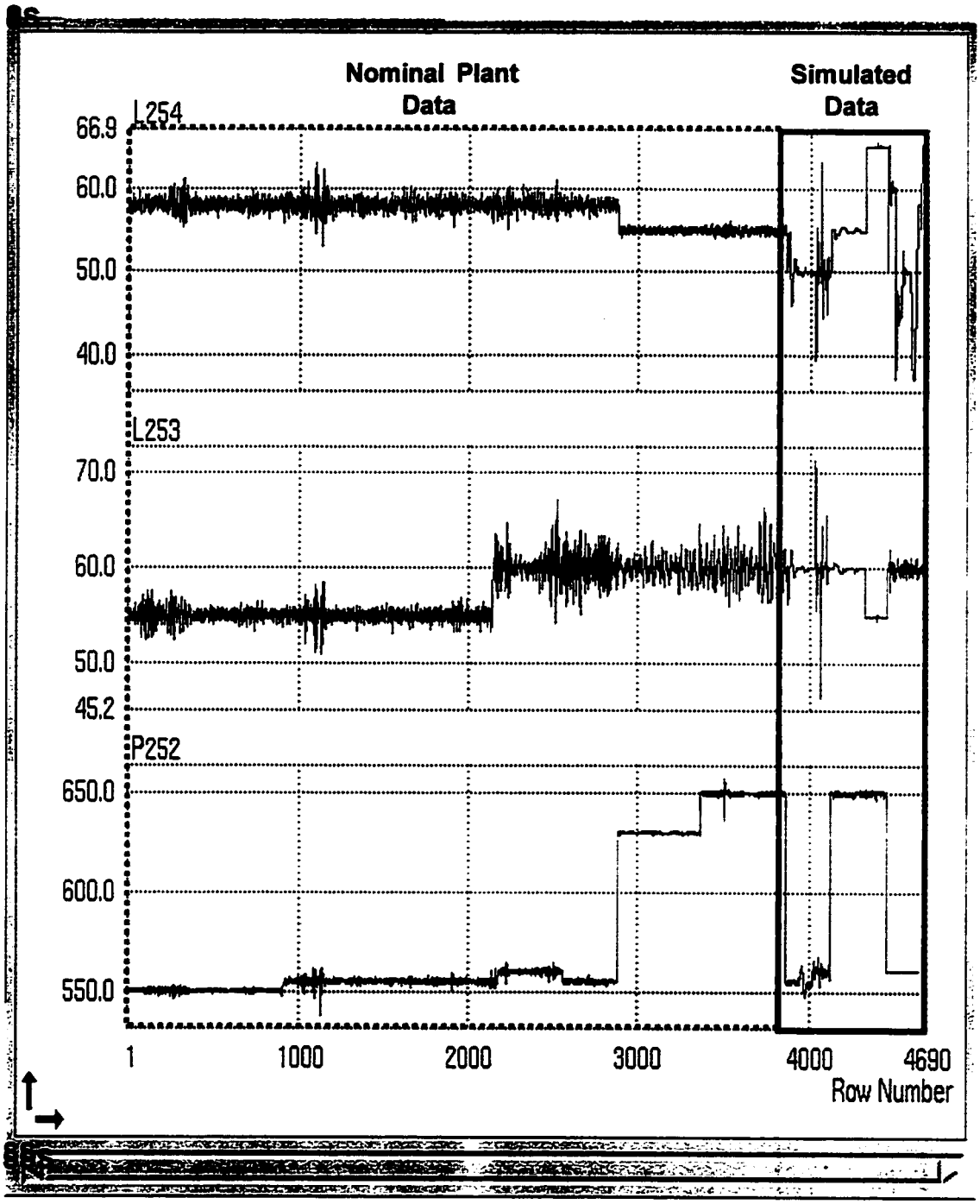


Figure E.3(d): TW252 pressure condenser and reboiler levels for Identification dataset I.n+s

Table E.3: TW252 Variable Statistics for Dataset I.n+s

Statistics for 4626 Recorded 3 minute ACS Data Points				
Tag Name	***Min***	***Max***	***Mean***	***Std***
F284	33.50	45.43	39.08	2.19
F256	34.62	47.00	42.42	3.02
F267	9.66	12.40	11.25	0.59
F271	0.84	24.09	14.57	2.66
F257	14.68	39.01	24.39	1.82
Y280	138.60	163.70	158.27	7.32
Y297	135.10	142.40	137.62	1.96
Y255	130.90	137.40	133.06	2.22
Y282	128.30	135.10	130.40	2.46
Y256	36.70	59.80	47.61	3.86
Y257	152.80	189.20	165.92	6.78
Y284	159.80	174.60	169.72	2.33
L253	46.40	71.40	57.56	2.75
L254	37.20	65.50	56.72	3.29
P252	537.50	658.00	579.55	40.13
A254	0.13	2.49	0.75	0.19
A255	14.25	18.49	16.17	0.73
A256	54.10	61.20	58.36	1.17
A257	20.94	29.48	25.44	1.37
A264	25.27	42.82	34.95	3.86
A265	8.79	11.86	10.32	0.64
A266	33.08	42.28	38.15	1.41
A267	16.16	36.96	19.06	3.98

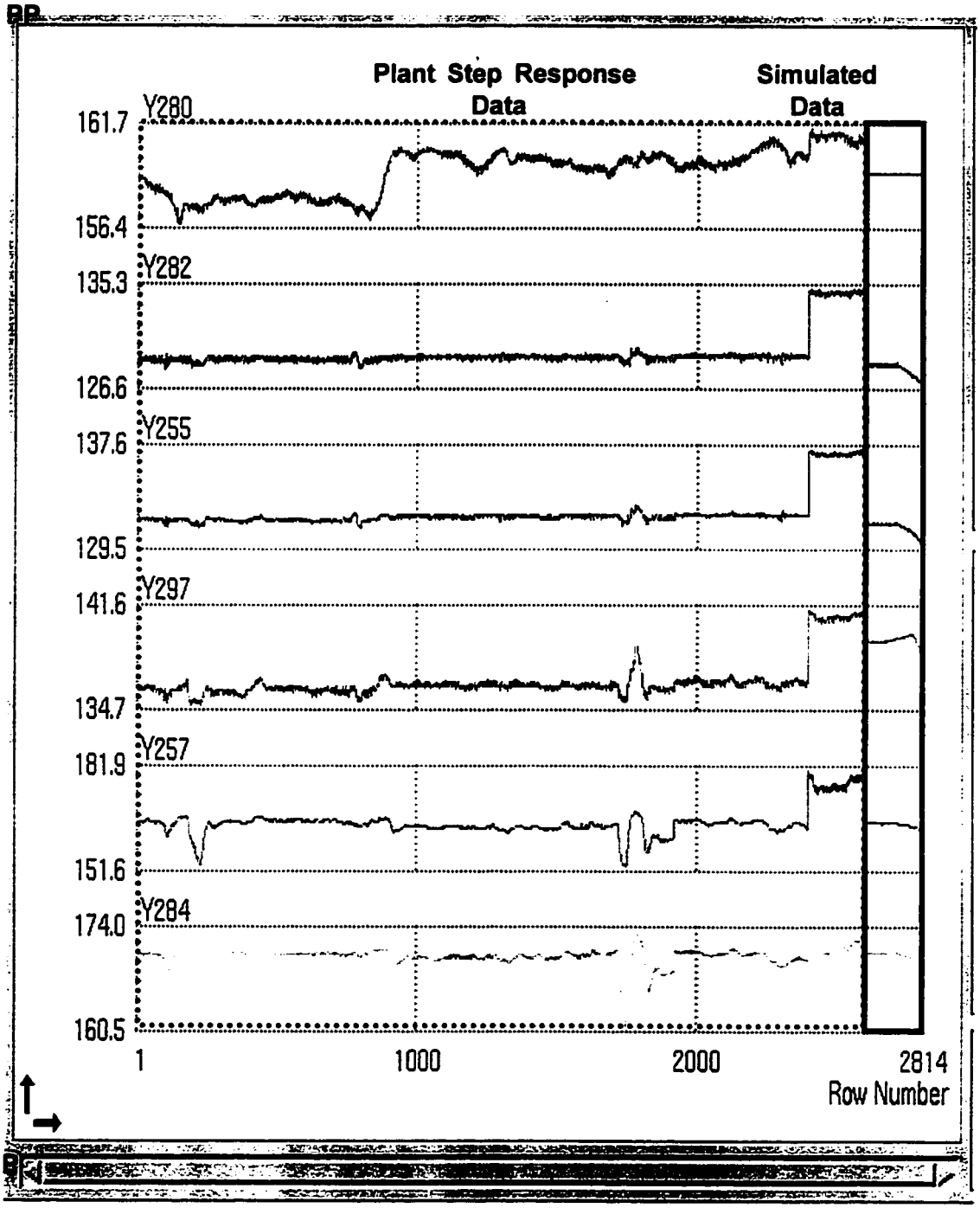


Figure E.4(a): TW252 Temperatures for Identification dataset I.t+s

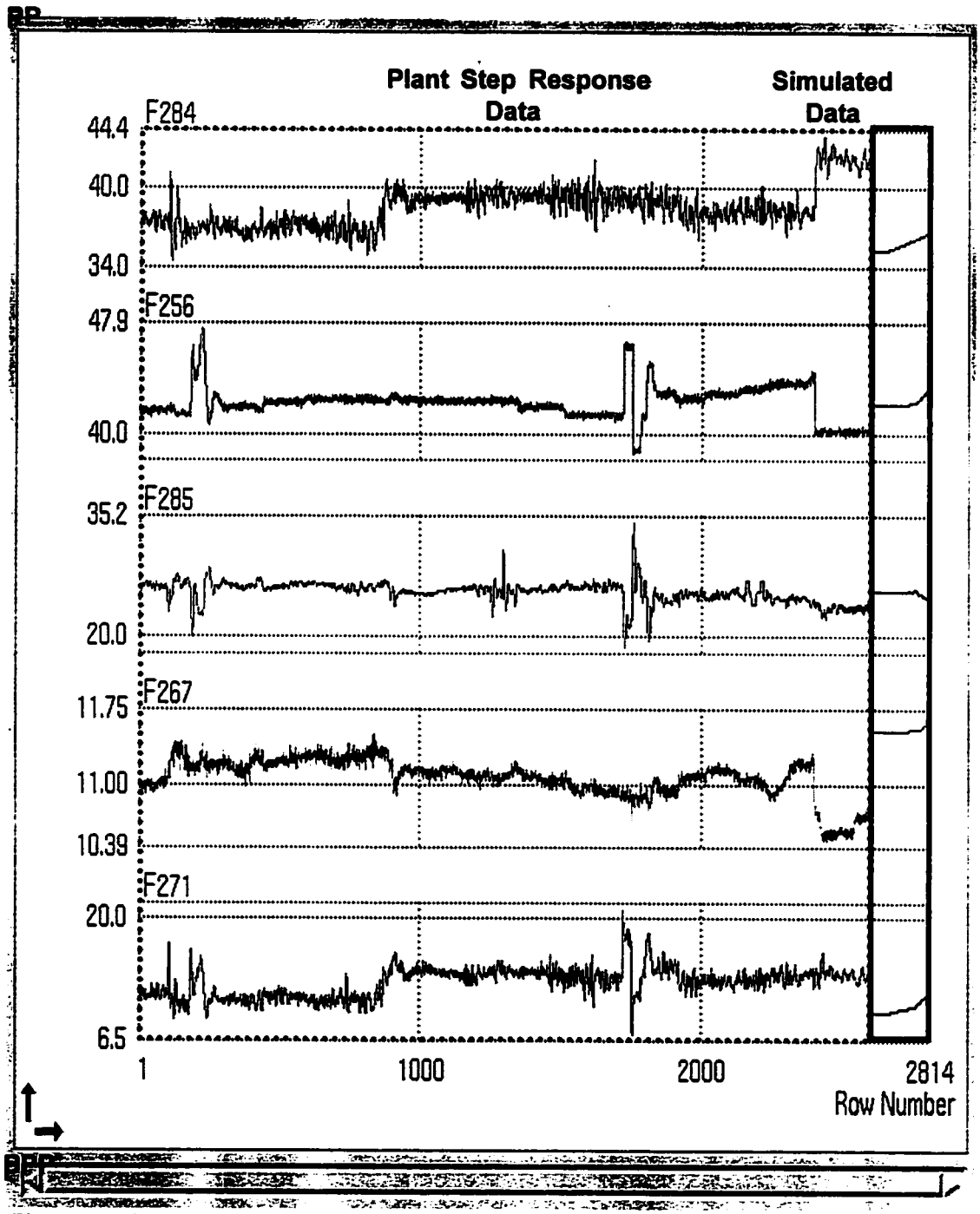


Figure E.4(b): TW252 flowrates for Identification dataset I.t+s

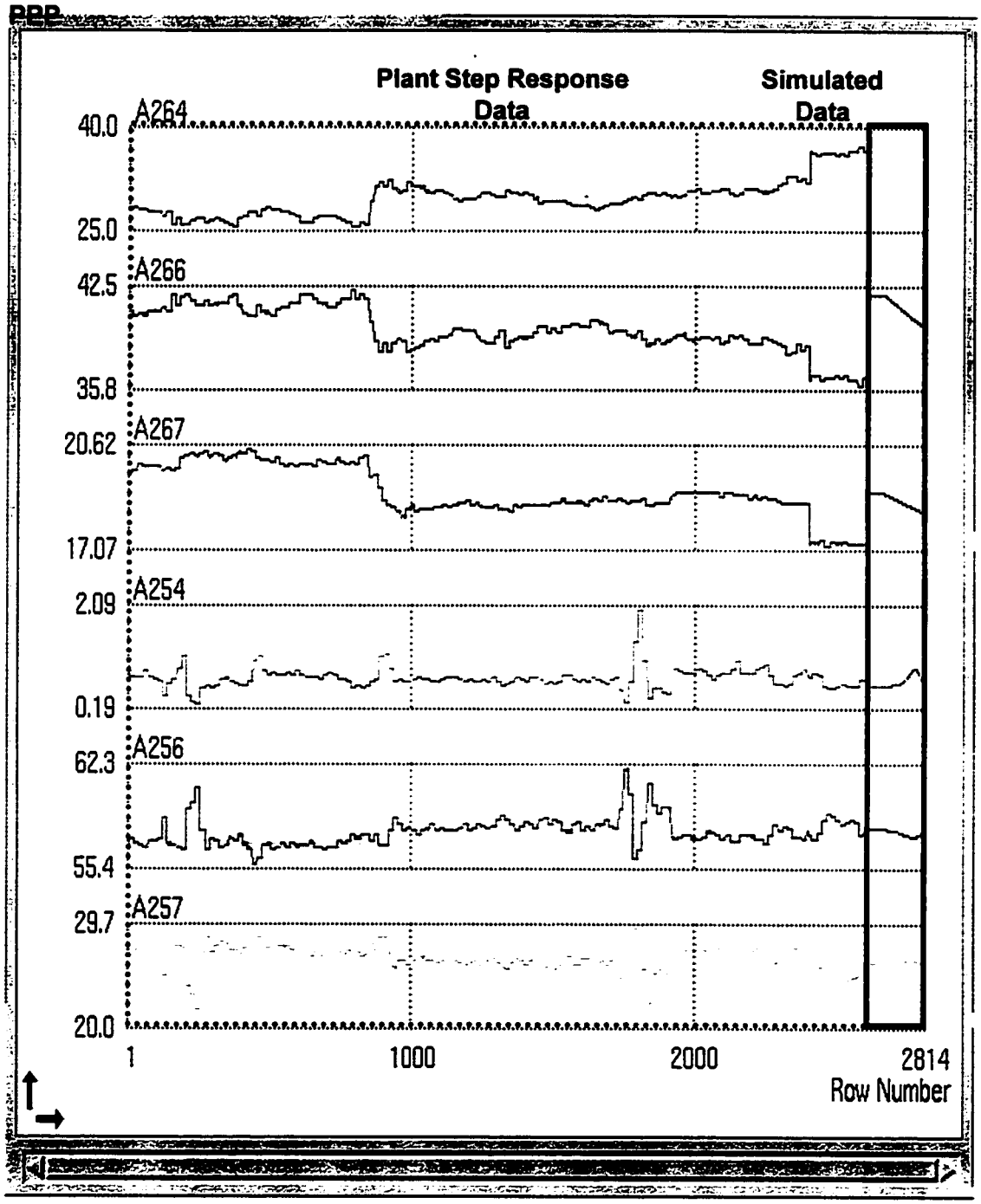


Figure E.4(c): TW252 compositions for Identification dataset I.t+s

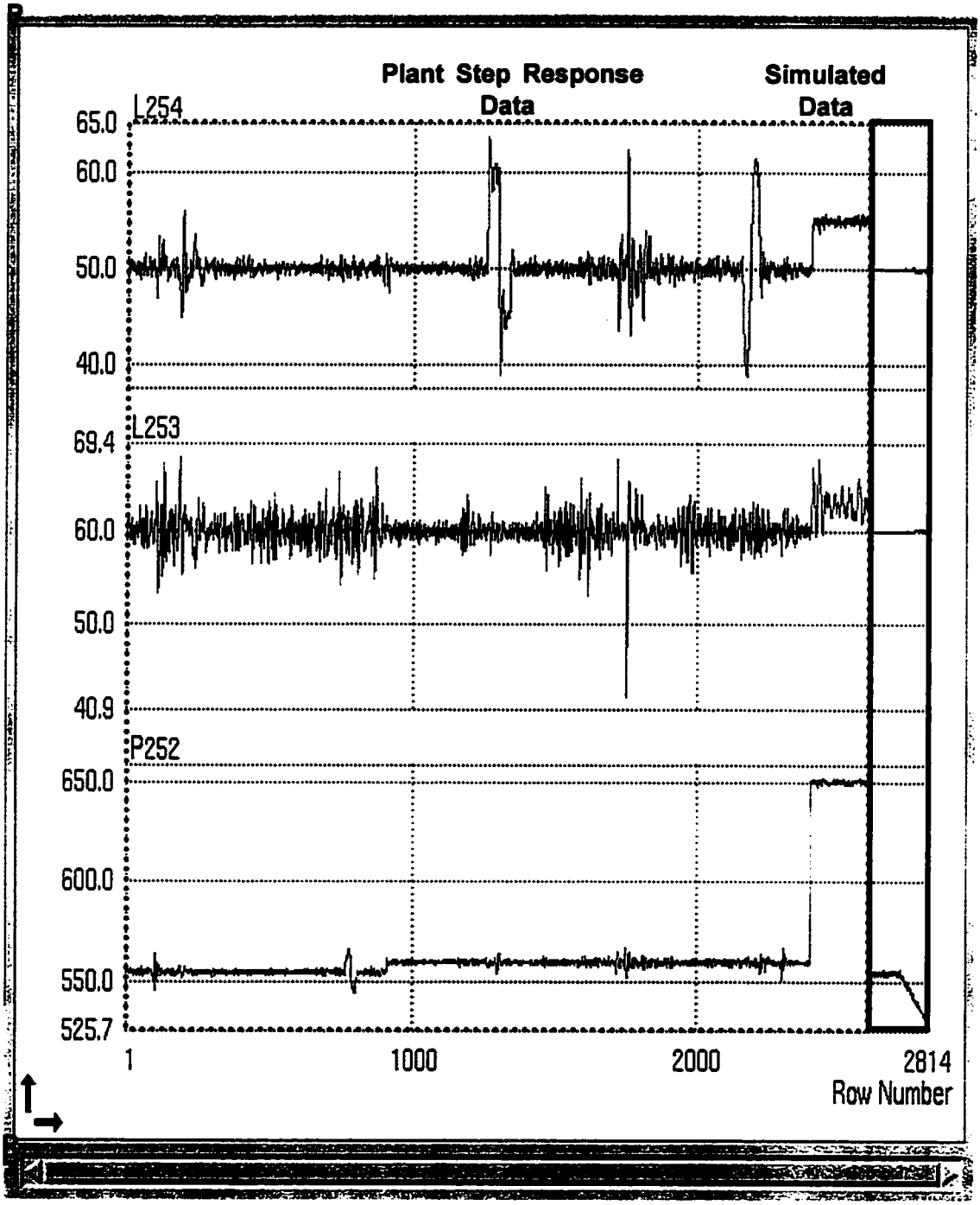
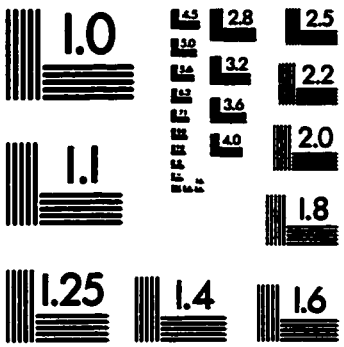


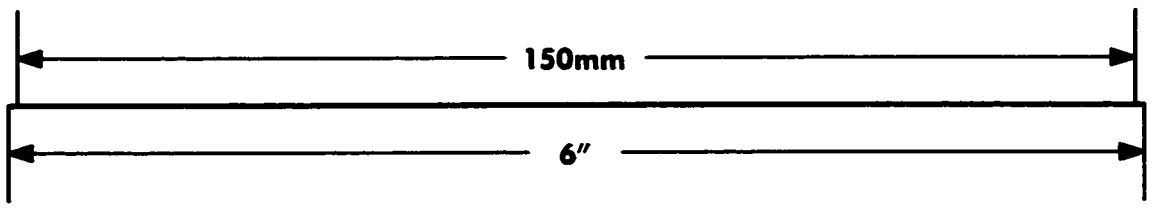
Figure E.4(d): TW252 pressure, condenser and reboiler levels for Identification dataset I.t+s

Table E.4: TW252 Variable Statistics for Dataset I.t+s

Statistics for 2814 Recorded 3 minute ACS Data Points				
Tag Name	***Min***	***Max***	***Mean***	***Std***
F284	34.45	43.93	38.33	1.69
F256	38.54	47.44	42.21	1.10
F267	10.45	11.69	11.11	0.22
F271	7.21	21.08	12.79	1.77
F285	18.52	34.38	25.52	1.22
Y280	156.60	161.40	159.25	1.11
Y297	135.00	141.30	136.79	1.44
Y255	129.90	137.20	132.26	1.34
Y282	127.00	134.90	129.50	1.47
Y256	38.10	62.60	53.27	4.44
Y257	153.00	180.60	165.63	3.78
Y284	161.10	173.40	170.15	1.44
L253	42.20	68.10	60.22	1.74
L254	38.80	63.80	50.43	2.43
P252	531.70	652.10	564.07	24.31
A254	0.28	2.00	0.75	0.18
A255	14.76	18.23	16.01	0.51
A256	55.70	62.00	57.84	0.88
A257	20.47	29.24	26.56	1.22
A264	25.57	37.19	29.74	2.65
A265	9.39	12.12	10.89	0.60
A266	36.14	42.23	39.83	1.42
A267	17.23	20.46	19.03	0.80

TEST TARGET (QA-3)





APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989