PREDICTION OF HEAVY DIESEL T95 USING JUST IN TIME LEARNING MODELS

by

Sevi Zeynep Hasdemir

B.S., Chemical Engineering, Middle East Technical University, 2013

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering
Boğaziçi University

2017

to Taner Yüzgeç

ACKNOWLEDGEMENTS

First, I would like to express my gratitude to my thesis advisor Assoc. Prof. Burak Alakent who guided me with his fund of knowledge. It was honor for me to work with him since he shaped my perspective related to application of statistic for process control and industrial processes.

I am also grateful to my family, especially my mother, Ferhan Yüzgeç. They are always supporting for everything.

I would like to thank to my team leader, Merve Aygün Esastürk since she supported me not only with her deep knowledge but also her encouragement and understanding. I am also grateful to my team members, Mehmet Yağcı, Gamze İş, Gizem Kuşoğlu, Işıl Kabacaoğlu and Yeşim Teke. It was real chance to know that they are always on the side of me.

I would also express my thanks to my colleagues in Research and Development and APC department in Tüpraş since they are always helped me and share their deep knowledge with me.

BAP Project No. 8041 is gratefully acknowledged for the funding.

ABSTRACT

One of the middle distillates of atmospheric distillation column is Heavy Diesel (HAD). T95 is the temperature, at which 95% volume of a sample is boiled, and it is the main controlled variable so accurate T95 predictions are required for a satisfactory performance of the model predictive control (MPC) algorithm, in which online T95 predictions are used to determine control actions. In the current thesis, just in time learning (JITL) methodology is used on historical process data to develop soft sensors for real time predictions of HAD T95. Local models are constructed using samples located in the neighborhood of a query point, and the constructed model is used for prediction of the response variable. Using 47 process variables in the historical data, three main groups of predictive models are constructed for HAD T95. In the first group of models, various subsets of variables, which are assumed to carry the highest information on variation of T95, are included into static and dynamic models. While there is no time lag between the selected input variables and the predicted quality variable in static models, previous day's T95 values (response variable) are included in dynamic models, as in autoregressive exogenous (ARX) input modeling. In the second group of models, least-squares (LS), partial LS (PLS), and subset regression via stepwise regression methods are employed on a predictor set, which consists of seven "most important" process variables. JITL models are evaluated with respect to various reference data selection methods, reference set size, window size and neighborhood size. The best model of this group is found to have predictive root means square error (RMSE) and mean absolute error (MAE) statistics equal to 5.66 and 4.23 ^{o}C, respectively. In the last group of models, interaction and quadratic predictor terms are included in the JITL model, and neighboring samples are selected a different subset of predictors. Using this method, RMSE and MAE of prediction statistics are decreased to 4.77 and 3.82 ^{o}C, respectively. This, to our knowledge, is the first time predictor and neighbor selection predictor subsets are separated from each other in the literature, and this seems a promising method in constructing soft-sensors for industrial applications.

ÖZET

Bu çalışmanın amacı atmosferik distilasyon ürünü olan Ağır Dizel'in hacimce %95'inin kaynama noktasının (T95) gerçek zamanlı olarak tahmin edilmesidir. Ağır Dizel'in T95 kaynama noktası kolon kontrolünde önemli bir kontrol edilen değişken olup, kolon yapısındaki MPC algoritmasının doğru aksiyonlar alabilmesi için bu değerin en doğru tahmin edilmesi operasyonel açıdan önemlidir. Bu çalışmada, Ağır Dizel T95 değerinin gerçek zamanlı tahmini için "anlık öğrenme" (JITL) yöntemi kullanılmıştır. JITL belirli sayıda ve tahmin edilecek veri noktasına en yakın komşuları tespit ederek, bu komşu veriler ile tahmin edilecek numuneye özel yerel modeller kurulmasını sağlayan bir modelleme yöntemidir. 47 süreç değişkeni ile yürütülen modelleme çalışmaları üç ana grupta incelenmiştir. İlk olarak, farklı değişken setleri denenerek Ağır Dizel T95 ile en yüksek ilişkili değişkenler tespit edilmeye çalışılmıştır. Bu çalışma değişkenler arasında zaman farkı olmadığı durum ve bir gün önceki Ağır Dizel T95 değerinin ARX girdisi olarak değişkenlere eklendiği durum olmak üzere iki ayrı şekilde incelenmiştir. İkinci modelleme grubunda ise belirlenen yedi değişken ile en küçük kare regresyonu (LS), kısmi en küçük kare regresyonu, ve model değişkenlerinin adım adım regresyon yöntemiyle belirlendiği yöntemler kullanılmıştır. Bu yöntemlerle kurulan JITL modelleri, referans kümesinin belirlenme yöntemi, referans kümesinin büyüklüğü, komşu sayısı, pencere büyüklüğü gibi parametreleri değiştirilerek değerlendirilmiştir. Bu grup çalışmada kurulan en iyi modelin kare ortalama hatası (RMSE) ve ortalama mutlak hatası (MAE) sırasıyla 5.66 ve 4.23 ^{o}C olarak hesaplanmıştır. Son modelleme grubunda, JITL modelleri kurulurken, en yakın komşu verilerin belirlenmesinde kullanılan değişkenler ve model değişkenleri ayrıştırılmış ve iki aşama için farklı değişkenler göz önünde bulundurulmuştur. Bu grup çalışmada kurulan en iyi modelin kare ortalama hatası ve ortalama mutlak hatası sırasıyla 4.77 ve 3.82 ^{o}C olarak hesaplanmıştır. Bu çalışma ile JITL yönteminin tahmin performansı önemli ölçüde artırılmıştır ve bu yöntemin endüstriyel uygulamalar için göz önünde bulundurulması gereken bir metot olduğu ortaya konmuştur.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

C	Covariance matrix
d_i	Auto scaling parameter
d^2	Euclidean distance based measurements as similarity criterion
e_i	i^{th} error value
E	Error matrix
F	Model error of Y
md^2	Mahalanobis distance based measurements as similarity criterion
m_x	Mean of data
n	Sample number
p_i	i^{th} eigenvector of $X^{\text{T}}X$
P	Loading matrix for X
P^T	Transpose matrix of loading matrix for X
Q	Loading matrix for Y
s	Standard deviation of the data
S	Scaling matrix
s_i	Similarity criterion
s_{MAD}	MAD scaling parameter
T	Outer product of scores matrix
t_i	i^{th} score vector (projection of X onto the eigenvector p_i
X	Variance-covariance matrix of the input data
x_i	Variable trajectory

x_{median}	Median of x
x_q	Query sample
Y	Output space of the data
w_{LS}	Least-Squares weighting function
w_B	Bisquare weighting function
w_G	Gaussian weighting function
w_T	Tricube weighting function
w_I	Inverse distance weighting function
β_i	Coefficients of linear model predictor variables
γ	Weighting parameter

LIST OF ACRONYMS/ABBREVIATIONS

AIC	Akaike information criterion
ANN	Artificial neural network
ARX	Autoregressive exogeneous
BPCD	Barrels per calender day
BPSD	Barrels per stream day
CoJIT	Correlation based just in time learning
CDU	Crude distillation unit
CRD	Constant reference dataset
DD	Data driven models
DVPE	Dry vapor pressure equivalent
FCC	Fluid catalytic cracking
FP	First principle models
HAD	Heavy Diesel
HVGO	Heavy Vacuum Gas Oil
IRD	Incremental reference dataset
JITL	Just in time learning
LPG	Liquified Petroleum Gas
LS	Least-squares
LSRN	Light Straight Run Naphtha
LVGO	Light Vacuum Gas Oil

LWR	Local weighted wegression
LW	Locally weighted model
MAE	Mean absolute error
MAD	Median absolute deviation
MPC	Model predictive control
MJIT	Mahalanobis distance based just in time
MRA	Multiple regression analysis
NCI	Nelson capacity index
PA	Pumparounds
PC	Principle component
PCA	Principal component analysis
PCR	Principle component regression
Phys	Physical model
PLS	Partial least-squares
RMSE	Root mean square error
SVR	Support vector regression
sWRD	Sliding window reference dataset
TBP	True boiling point
T95	Temperature at which 95% volume of a sample is boiled

1. INTRODUCTION

Refineries are used to convert crude oil into various products, ranging from transportation fuels and petrochemical feedstocks to asphalt and coke, and more than $8 billion of refined products take their places in marketplace (Chang *et al.,* 2012). Main products of refining sector are gasoline (aviation and motor gasoline, and light distillates), middle distillates (diesel fuel, jet fuel and home heating oil), fuel oil and other products (fuel gas, lubricants, wax, solvents, and refinery fuels) (Gary *et al.,* 2007).

Refining begins with distillation processes in atmospheric and vacuum distillation columns, used to separate crude oil into fractions and cuts based on molecular size and boiling point ranges. Distillation is followed by cracking, reforming, and other conventional processes, which are, in turn, followed by various treatments to remove the undesired contaminants. Then, blending processes are used to bring the final products at desired qualities (Gary *et al.,* 2007 a). During refinery processes, specifications must be met, and the refinery operation must stay within tight environmental and legal constraints.

Since refining is an integrated, complex, continuous operation, elaborated optimization that link operation variables, product specifications, economic constraints and environmental limitations is required to be performed. Various operation variables and conditions, such as temperature, pressure, residence time, feed quality, cut points, space velocity and catalyst are used to balance feedstock, product and quality. Since each undesired product creates a new waste stream, controlling these variables within their limits is essential to optimize the operation of the refinery (Gary *et al.*, 2007 b).

Advanced monitoring and control systems are installed in process units to comply with important requirements and regulations, such as final product quality, pollution efficiency, process safety and environmental concerns (Sliškovic *et al.,* 2011). Online measurement of quality variables is crucial to gain insight about productivity and economic development. Control of distillation operations is complicated not only due to nonlinear

relations between process variables, but also due to lack of maintenance and regular calibrations of online analyzers, which may yield inaccurate and imprecise measurements (Al-Dunainawi and Abbod, 2016). Furthermore, online analyzers are high cost equipment, generally time consuming, and used at a low frequency, usually once in a laboratory shift or even less (Chatterjee and Saraf, 2004).

Soft sensors are generally developed to monitor product quality in refinery process units. More importantly, online predictions via soft sensors are used in MPC of distillation columns. Soft sensors are mathematical algorithms, which produce real time predictions of unmeasured variables using mechanistic models or historical data. Advantages of soft sensors over hardware sensors are their low costs, facility to work in parallel with hardware sensors, easy implementation on existing hardware, allowance of real time prediction of variables, and overcoming the time delays introduced by slow hardware sensors (Fortuna, Graziani, and Xibilia 2005). Soft sensors are used in exploiting correlations between process variables, which are easy to measure, and quality variables, which are difficult to measure, in forms of mathematical models. These mathematical models may be model-based, or data-based, depending on the modeling approach. The former approach is based on physical and chemical principles, while the latter method consists of building input-output relations based on historical plant data (Waddams, 2013). For most of the chemical processes, physicochemical knowledge is not complete, process relations are generally time varying and nonlinear, and process conditions may change abruptly. For these reasons, data-based methods are preferably used in industrial processes. However, for complex systems, even the empirical model structure may be difficult to develop, and adapting the model to the operation is generally laborsome due to changing process conditions and nonlinear system behavior (Cheng and Chiu, 2004). To overcome these problems, just in time learning (JITL) methodology is developed. This methodology is also known as instance-based learning, locally weighted model (LW) or lazy learning. JITL models are constructed dynamically using samples located in the neighborhood of a query point, and response of the query point is predicted. After the output is predicted, the constructed local model is discarded (Saptoro, 2014).

Tüpraş is Turkey's largest industrial enterprise, consisting of four operational oil refineries in İzmit, İzmir, Kırıkkale, and Batman, with a total 28.1 million tons of annual crude oil processing capacity. Tüpraş owns 59% of the total petroleum product storage capacity in Turkey, composed of 1.7 million tons of crude oil, 1.3 million tons of white product, and 0.9 million tons of black product. Nelson Capacity Index (NCI) is an indicator for technical ratings of refineries, and NCI is equal to 7.25 and 5.95 (in 2016) for Tüpraş and Mediterranean refinery complexity, respectively, showing the quality of Tüpraş refineries.

In İzmit Tüpraş Refinery, there are three main Crude Distillation Units (CDUs), in which crude oil is first processed via atmospheric distillation, and the feed capacities of these units are approximately equal to 13000 m3/d. Naphtha, Liquefied Petroleum Gas (LPG), Kerosene, Diesel are produced by atmospheric distillation. In the CDUs, product specifications are to be met according to instructions of planning department. In these units, product specifications are mainly based on T95, flash point of Kerosene and Dry Vapor Pressure Equivalent (DVPE).

Canadian Fuel Association states that European refineries have surplus of gasoline and shortage of diesel in recent years, due to large scale conversion of domestic vehicles from gasoline to diesel, so maximization of Heavy Diesel yield, as currently achieved in TUPRAŞ, is an important economic criteria (Canadian Fuels Association 2013). Yield in an atmospheric distillation column is mainly dependent on the degree of separation between bottom and heavy diesel products, controlled via heavy diesel T95 point. Heavy diesel product properties are determined using laboratory analyses on intermittent samples. Online or real time estimation of the controlled variables is essential to hold the operation at its optimum point, and to guarantee high efficiency and profitability. In other words, the economic goal for prediction of Heavy Diesel T95 is to obtain maximum amount of product with the highest quality within specifications (Macias-Hernandez, Angelov, and Zhou, 2007).

In Tüpraş, Heavy Diesel T95 is measured in laboratory three times a day in every shift. Operational conditions in crude distillation unit may change instantaneously, number of Heavy Diesel sample analyses in a day is not sufficient to follow process transitions, thus instantaneous changes in the product quality. Moreover, inaccuracies may occur due to experimental and instrumental errors. In İzmit Tüpraş refinery, various kinds of crude oils are blended and fed to the crude oil units from different tanks having different compositions. Since changes in feed composition both effects process conditions, and changes gap overlap and cut points, it is difficult to develop robust soft sensors that can adapt these abrupt changes. Additionally, nonlinear process behavior also makes estimating accurate product qualities difficult. In this study, JITL methodology is employed for building online soft sensors with the purpose of coping with these obstacles, and real time Heavy Diesel T95 values are estimated via local regression techniques.

In this thesis, first, soft sensor applications in process industry are represented. During the modeling construction of soft sensors, both global and local learning algorithms are adopted. In the following sections, global modeling techniques such as principal component analysis (PCA) and partial least squares (PLS), are discussed, and JITL methodology, developed by local modeling approach, is elucidated. Following a general discussion of refinery processes, Tüpraş refinery, from which the data used for statistical models in the current thesis were obtained, is briefly represented. In the Results and Discussion section, various global and local JITL models are constructed. Finally, assessments according to refinery application of JITL, are done and discussed elaborately.

2. SOFT SENSORS DEVELOPMENT

Soft sensors are mathematical algorithms used as quality estimators in industry. In the current section, soft sensors are discussed in detail. General information on soft sensors and their industrial applications, and types of soft sensors are given in Section 2.1 and Section 2.2, respectively. Effect of data imperfections on soft-sensor development is discussed in Section 2.3. Data driven soft sensor development methodology is elucidated in Section 2.4, in which PCA and PLS are also discussed.

2.1. Soft Sensors and Their Applications in Industry

Companies are getting stricter in their product prices and specifications, and these are main competing parameters in industrial market due to economic and environmental regulations. Power and raw material consumption, safety rules, management strategies of abnormal situations are to be carefully considered to reach operational excellence. Thus, most of the companies are in need of reliable and long lasting tools to solve different operational problems, such as measuring system back-up, what-if analysis, real time prediction for plant control, sensor validation and fault diagnostic strategies by increasing performance of their control and monitoring systems (Fortuna *et al.*, 2007).

Process measurements in industry are usually performed by hardware sensors, but these sensors may fail due to different sensor faults, data acquisition system fail, sensor removal in maintenance period (Sliškovic *et al.*, 2011). According to the survey of PSE143 in 2004, while 27% of hardware sensors problems are due to time consuming maintenance, 21% of them are caused by the need for calibration, and 15% occur due to age deterioration (Kano and Fujiwara, 2013). Thus, generally, laboratory measurements are periodically performed, particularly for variables, which give information on final product quality. However, laboratory measurements are performed off-line, rendering them inconvenient

for automatic process control, and they may be too expensive, and less reliable compared to automatic measurements.

Since the beginning of 1990's, various novel methods, which are used to make online predictions of difficult-to-measure variables, and online monitoring, have been studied widely. These methods are developed using the relations between easily measured process variables and difficult-to-measure variables (Sliškovic *et al.*, 2011). In general, these systems are called inferential models, virtual sensors, or soft sensors, and various properties of inferential models may be stated as follows: (i) soft sensors represent low cost alternative to hardware sensors, (ii) it is possible to realize more comprehensive monitoring networks, (iii) it is easy implement on existing hardware, such as microcontrollers, and (iv) soft sensors provide real time estimation of process variables to deal with time delays, improving performance of the control strategies (Fortuna et al. 2007). According to a survey related to the present state of soft sensor technologies in Mitsubishi Chemical Corporation in 2006, 37% of the soft sensors are used for stabilizing quality, 28% are used for reducing feed and utility, and 18% are used for improving reliability, simultaneously employed with online analyzers. Soft sensors may also be used for stabilizing operations, reducing manual analyses, avoiding installation of special analyzers (Kano *et al.*, 2013).

For various applications of soft sensors, there are three important points to consider: (i) changes in process characteristics, (ii) individual differences of equipment, and (iii) reliability of soft sensors. First of all, even if a soft sensor is developed successfully by selecting predictor variables and tuning input parameters carefully, its estimation performance may deteriorate upon perturbations in process characteristics and operation conditions. It is important to construct soft sensors, which adapt to these changes via minimum manual and repeated constructions. Furthermore, when a soft sensor is developed for one device or process, it cannot be applied to another, so it is should be customized. Finally, modeling strategy has significant impact on reliability of the soft sensors; data driven soft sensors, particularly, are prone to deteriorate due to incorrect measurements and sensor fails (Kano *et al.*, 2013). Thus, application purpose of soft-

sensors and methodologies used in their construction are to be carefully designed for long lasting and reliable soft sensors.

There are broad application areas of soft sensors. Overall, soft sensor applications can be divided into three main groups: (i) online prediction, (ii) process monitoring and process fault detection, and (iii) sensor fault detection and reconstruction (Figure 2.1). The most common application area of soft sensors is the online prediction of process variables, which are either sampled at a low rate, or measured by offline analysis. Online predictions are generally performed for variables, which are indicators of output quality, or which give critical assessment related to the process, at a high sampling rate, and low financial cost. As shown in Figure 2.2., easy-to-measure variables, such as pressure and temperature of a process, may be used to develop models to make online concentration predictions, which may be used for process control applications. Soft sensors are also employed for process monitoring and process fault detection applications, in which deviation of process conditions from the normal operating plant conditions is detected, and causes for the disturbances are identified. Generally, these types of soft sensors are developed based on historical data, additionally receiving support from operators. Process monitoring soft sensors are widely based on PCA based methods and statistical process control (SPC) charts (Kadlec, 2009). Since processing plants contains a large number of sensors, it is difficult to detect whether sensor faults are temporary, or permanent. Soft sensors are also widely used for sensor fault detections (Kadlec *et al.*, 2009).

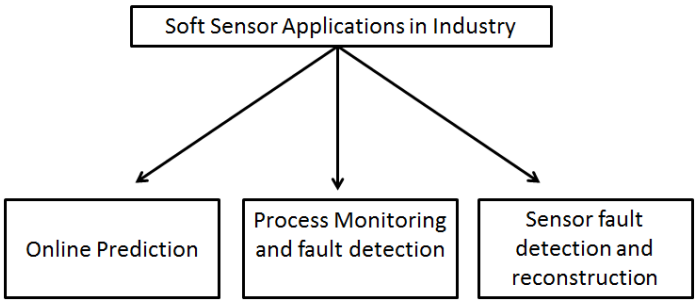

Figure 2.1. Soft sensor applications in process industry.

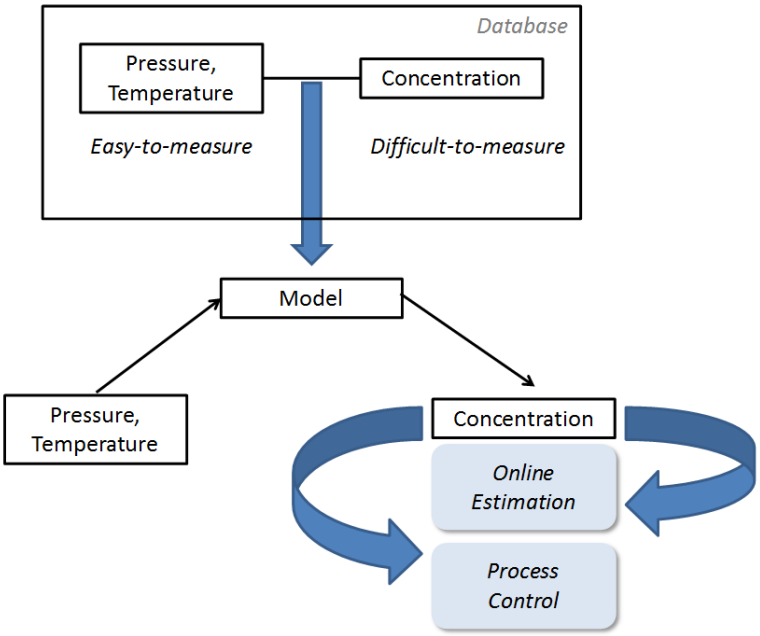

Figure 2.2. The fundamental working principle concept of inferential sensors (Okada *et al.*, 2012).

2.2. Types of Soft Sensors

Availability of priori knowledge in the application area of soft sensors, especially for complex systems, is the main criterion for determining the modeling approach for soft sensors. This knowledge is helpful in determining modeling procedure, type of the model, complexity of the model and applied methodology. There are mainly two modeling approaches: (i) first principle (FP) models, also known as white-box, mathematical model-based models, and (ii) data based, or data-driven (DD), models, also called black box models.

Utilizing process knowledge in modeling the process is the fundamental advantage of FP models, while data driven models cannot be easily used to associate the model parameters with the actual process. FP models generally yield physically meaningful links between process variables. However, for complex industrial processes, parameters of the FP models may be inaccurate, and development of FP models from physical laws are

usually expensive. Process operation data are used to construct DD models. Figure 2.3 shows the model structures for both FP and DD models. In contrast to FP models, DD models is used to provide useful results with little domain of knowledge. At the same time, DD Models are adaptable to specific problem instances, and are used to handle disturbances and errors (Czop *et al.*, 2011; Sliškovic *et al.*, 2011).

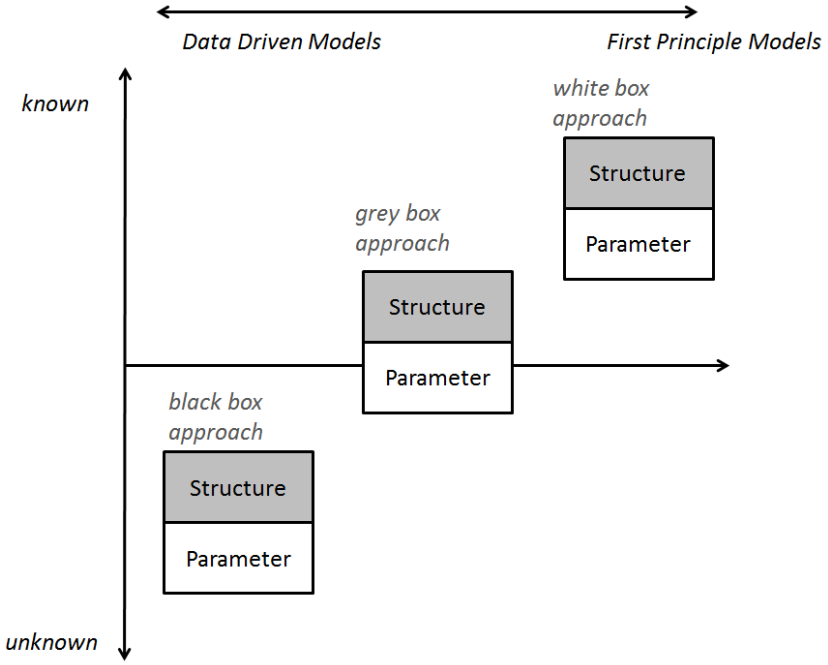

Figure 2.3. FP and DD models based on recognition of their structures and parameters (Czop *et al.*, 2011).

2.3. Effect of Imperfections of the Data in Soft Sensor Development

Data obtained from industrial processes usually have "imperfections", which may prevent construction of reliable soft-sensors. The main imperfections with the process are listed as (a) missing values, (b) outliers, (c) data co-linearity, and (d) measurement noise.

Missing values may consist of single observations, or a group of successive samples, which do not represent the real state of the process variables (Figure 2.4a). While failure

and removal of hardware sensors are common reasons, problems in transmission of the data from the sensors to the database, errors in the database, and problems in accessing the database are other possible causes of for missing values.Outliers are measurement, which deviate from the typical ranges of the measured values (Figure 2.4b) (Kadlec *et al.,* 2009). Outliers can be classified as “obvious” and “non-obvious” outliers. Obvious outliers can be detected from the violation of the physical or technological constraints. However, non-obvious outliers do not exceed any constraints, but do not reflect the true state of the variables.

Drifting data is also another problem for continuous plant operations, and can be classified as process and sensor drifts. Process drifts are due to steady abrasion of mechanical elements during operation of the plant. In addition to mechanical pump abrasion, changing environmental conditions, changes in raw material, and catalyst deactivation are also causes for process drifts. As a result of the shift in the process state, process data show a drift and do not form a stationary time series. Sensor drifts, on the other hand, are the result of changes on the measuring devices, usually for recalibration. These drifts are only observed in the measured data, but do not represent a true change in the state of the process.

Co-linearity is a consequence of high dimensional variable-rich but information-poor datasets (Figure 2.4c) (Kadlec *et al.,* 2009). Measured data in the industry are strongly co-linear, and successive measurements are strongly correlated with each other. Including all the variables directly in the statistical models increases the model complexity, and it has a negative effect on predictive performance of the model.Measurement noise is another artifact in industrial data, and is usually eliminated in the data preprocessing step (Figure 2.4d). Generally, a smoothing filter is employed in this step to filter out the noise (Kadlec *et al.,* 2009).

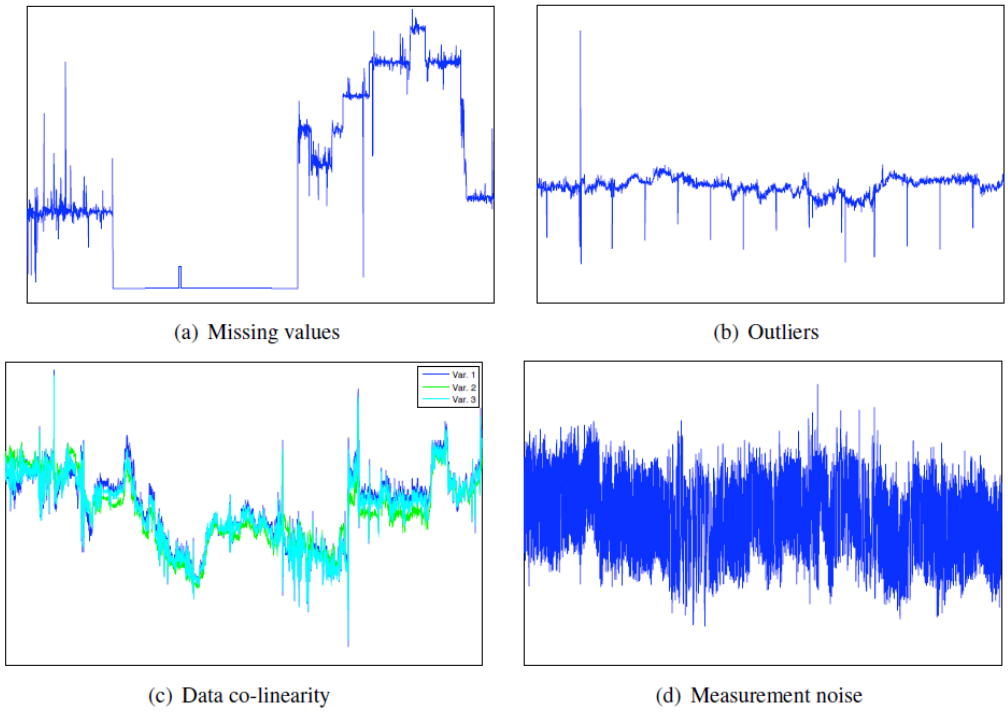

Figure 2.4. Common imperfections in industrial data sets (Kadlec *et al.*, 2009).

2.4. Data Driven Soft Sensor Modeling

Data for soft sensors can be obtained either through specifically designed experiments, or from normal plant operation in complex industrial plants. Industrial plants have a large number of installed sensors, which measure various process variables. These measurements are accumulated in the database and may be utilized during the modeling stage. As can be seen in Figure 2.5, data preprocessing step is required to cope with problems related to dataset. After a soft sensor model is developed and validated, the soft sensor should still be maintained to render sustainability and adaptability to the processes.

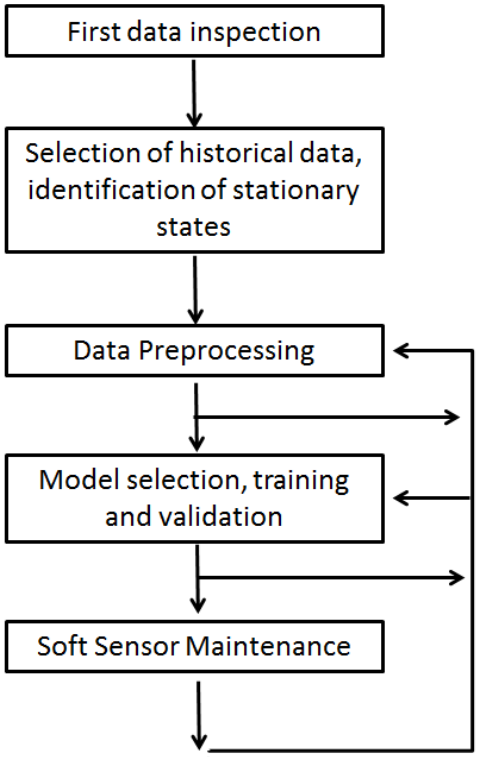

Figure 2.5. Soft sensor development methodology (Kadlec *et al.*, 2009).

2.4.1. Data Preprocessing

Data preprocessing step is usually performed to handle outliers or missing data. Outliers and missing data, which can also be considered as outliers, may particularly deteriorate data driven soft sensor models, such as PCA or PLS. If missing data are occasional, then these values are generally interpolated from the measured neighboring observations. However, if missing data cover large partition within dataset, then these values are usually discarded before constructing the soft sensor (Lin *et al.*, 2007).

Another important step during data preprocessing is scaling. Variables with large absolute values in the dataset are likely to have larger variance, and these variables may cloud variables with smaller variance. On the other hand, it is important for all variables having equal weights, particularly in statistical models, such as PCA. For this purpose, auto-scaling, i.e. mean-centering followed by normalization using standard deviation, is

commonly used on all variables. Auto scaling procedure for a variable trajectory x_i is shown in equation (2.1), in which m_x is the mean, and s is the standard deviation of the data.

$$d_i = \frac{x_i - m_x}{s} \tag{2.1}$$

In addition to auto-scaling, median absolute deviation from median (MAD) scaling may also be used. It is known that MAD scaling is more reliable for scaling data with outliers, while auto-scaling is more effective for data with close-to-normal distribution. Equation (2.2) shows MAD equation, in which x_{median} is the median of x, and the constant 1.4826 is required to make MAD an unbiased estimate of the standard deviation for Gaussian data (Chiang *et al.*, 2003).

$$s_{MAD} = 1.4826 median\{|x_i - x_{median}|\} \tag{2.2}$$

2.4.2. Model Selection

Model selection is a critical step in soft sensor development. Although different strategies are discussed in the literature, it is difficult to determine a general method, which may be applied to all cases. It is clear that process knowledge, degree of nonlinearity in the process, selection of input variables, model order, operational ranges, time delay, and sampling time are to be carefully considered during model construction. Overall, it can be said that model structure should be selected with respect to the capability of models to cope with high dimensionality, collinearity and nonlinearity. In this stage, highly correlated variables are usually eliminated, dimension of variables is reduced using convenient transformations, and nonlinear characteristics may be removed and adapted to model structure (Fortuna *et al.*, 2007). Various approaches are used in the literature in determining the model structure. In the following section, two of the widely used data driven modeling methods, namely PCA and PLS, are discussed.

2.4.2.1. Principal Component Analysis (PCA). PCA is a dimension reduction technique, used to construct latent variables from linear combinations of variables with the highest variance in the input space. Multicollinearity problem may be solved via PCA, since the new constructed variables (principal components) are orthogonal to each other. The resulting principal component (PC) subspace is assumed to contain the most useful information about the process in the smallest dimensions, mostly free from noise. Additionally, it is possible to visualize the high dimensional data via projection of the original data onto PC space (Sanguansat, 2012).

In PCA analysis, eigenvectors of the variance-covariance matrix of the input data *(X)* are ordered in accordance with the descending order of eigenvalues, and each eigenvector (PC) describes a direction in the original variable dimensions, accounting for the largest amount of unexplained variance in the data. Mathematically, these can be described as;

$$Xp_i = t_i \tag{2.3}$$

Here, p_i. is the i^{th} eigenvector of XTX, and ti is the ith score vector, which is the projection of *X* onto the eigenvector p_i. Similarly, *X* can be reconstructed by the outer product of scores matrix *T*, whose columns are the score vectors associated with each PC, and the transpose of the loadings matrix *P*, whose columns are the loadings vectors associated with each PC. When a subset of PC space is used to approximate *X*, the rest of PCs form the residual subspace and the corresponding error matrix *E*.

$$X = TP^T + E \tag{2.4}$$

PCA, while being a powerful and widely used method, has a number of shortcomings, which need to be addressed. Nonlinear relations between the variables cannot be easily modeled using PCA. It is not straightforward to determine the optimal number of principal components. Although PCA may be used to describe relations in the input space, relations between the input and output spaces cannot be modeled using only PCA. A number of variations on PCA, such as nonlinear and adaptive PCA, are shown in

Figure 2.6. PCA be made adaptive using recursive, moving window and time-lagged forms. Additionally, PLS may be used to model the relation between high dimensional input and output spaces (Kadlec *et al.*, 2009).

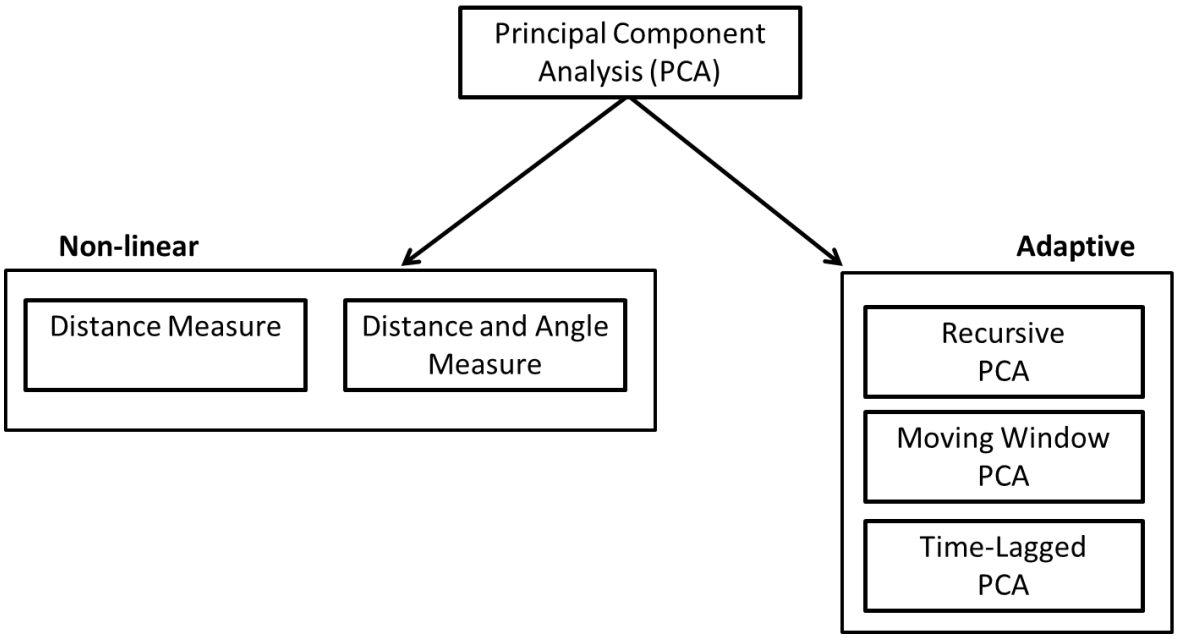

Figure 2.6. PCA and its variations (Kadlec *et al.*, 2009).

2.4.2.2. Partial Least Square (PLS). In PLS method, correlation both within the input variable space and between input and output spaces are taken into account (Sliškovic et al. 2011). In other words, PLS is used to determine the latent variables, which represent the highest correlated subspace in X, mostly correlated with subspace of Y, showing the highest variability. Mathematically, these relations may be written as the following:

$$X = TP^T + E \tag{2.5}$$

$$Y = TQ + F \tag{2.6}$$

Q is the matrix of loadings for *Y*, and *F* is the model error of *Y*. T is the matrix of scores, *P* is the matrix of loadings for *X*, and *E* is the model error associated with *X*. The loadings for PLS are not identical to the eigenvectors found by PCA. In PLS, the objective is not to represent only the measurement space *X*, or the output space of *Y*, but to represent both spaces, simultaneously.

2.4.3. Model Validation

Model validation is another important step after model selection phase. In model validation, two statistical concepts should be considered simultaneously; under fitting leads to high bias, while overfitting leads to high variance. Training and validation data set sizes have important impact on determining the best predictive model. Increasing the training sample size can reduce the effect of overfitting and reduce the variance of estimators, but bias is not much affected. Instead, addition of more predictors to a model can reduce the bias.

During model validation, two statistics, root mean square error (RMSE) and mean absolute error (MAE), are used to assess the predictive quality of the constructed models in the current study.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {e_i}^2} \tag{2.7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i| \tag{2.8}$$

As can be seen in equations (2.7) and (2.8), higher weights are assigned to errors with larger absolute values than errors with smaller absolute values in RMSE, while equal weights are given to all errors in MAE. Therefore, RMSE is always greater than MAE, and RMSE tends to increase more, as error distribution moves away from normality. Since

higher weights are assigned to larger errors in RMSE, RMSE is usually regarded better at evaluating the performance of a model (Chai and Draxler, 2014).

After soft sensors are validated, they are ready to be used in the field. However, permanent changes in process conditions usually lead to degradation in soft sensors, unless maintained. Soft sensors should be adaptive to these changes, so various techniques, such as recursive PCA and PLS, neuro-fuzzy-based soft sensors, and moving window methods are used (Kadlec *et al.*, 2009; Chen *et al.*, 2015).

3. GLOBAL AND LOCAL MODELING PARADIGM

Data based modeling can mainly be classified into two groups: (i) global modeling, and (ii) local modeling. While linear models, neural network applications and splines are studied mostly via global modeling approach, lazy learning and nearest neighbor methods are utilized via local modeling (Figure 3.1). Global models are built based on an analytical function that is expected to represent the whole set of operating conditions. These models can be stored in a small memory even for large datasets, and execution time is short. However, there is no guarantee that the suggested functions are able to represent the relations in the future data.

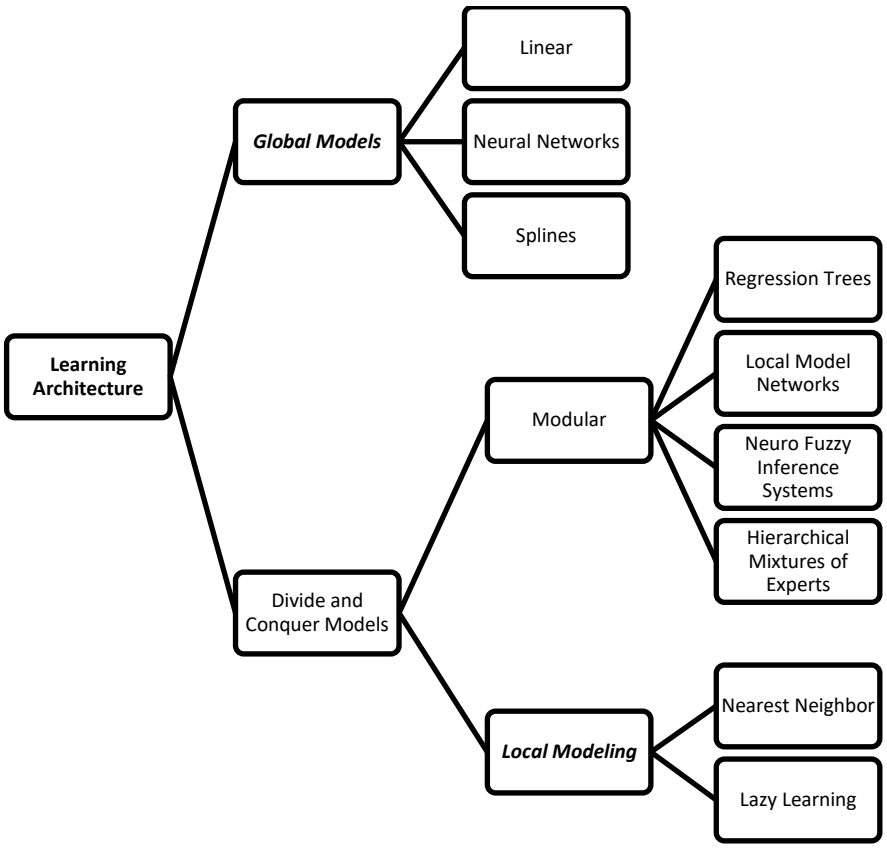

Figure 3.1. General view of learning architecture (Bontempi, Birattari, and Bersini 2001).

As can be seen in Figure 3.2, estimation of the global model parameters using the data forms the induction part, while prediction using the constructed models is performed in the deduction part. Existence of these two successive steps makes it difficult to determine predictions, so local modeling techniques, called transduction, are developed as an alternative approach. In local modeling paradigm, predictive functions are built within the neighborhood of the query point, whose response is to be predicted. Contrary to global models, the reference dataset is always kept in memory.

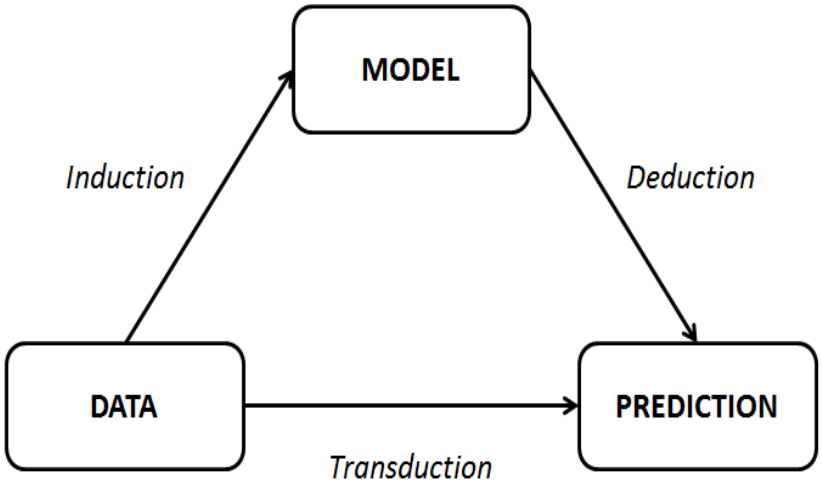

Figure 3.2. Function estimation vs. value estimation (Bontempi *et al.*, 2001).

There are well known global modeling methods in the literature, such as linear regression, nonlinear regressions, splines and neural network modeling. Additionally, data driven models, such as PLS and PCA, are widely used in industry. Although most of these methods are originally developed from a global modeling point of view, local versions of these techniques are developed in the recent years. In the following section, methodology of PCA and PLS will be represented, and adaptive forms of these techniques are also discussed.

3.1. Just in time Learning (JITL)

Data-driven soft sensors are widely used in industry, since they can be developed from already available historical database, and provide new generation of operational excellence at low cost. Moreover, hidden or latent information can be extracted, while analyzing historical database that gives different point of view related to complex and highly integrated processes (Saptoro, 2014). However, after a certain time, a soft sensor will gradually deteriorate due to several reasons, such as: (i) changes in process input materials, (ii) process fouling, (iii) abrasion of mechanic components, such as equipment ageing, fouling clogging, (iv) catalyst activity changes, (v) production of different product quality grades, and (vi) changes in external environment (Kadlec *et al.,* 2011). In addition to these issues, measurement noise, missing values, data outliers, co-linear features and varying sampling rates are also main reasons of soft sensor failures (Kadlec *et al.,* 2009).

Model maintenance may be one of the main solutions to cope with the issues represented above, and maintain accurate predictions. However, models reconstruction prevents the continuity of soft sensors. JITL has been proposed to cope with the continuity issues and process nonlinearity. JITL is inspired by the ideas of database technology and local modelling techniques, and generally known as instance-based learning, LW model, lazy learning or model-on-demand (Saptoro, 2014).

Locally weighted learning is a form of lazy learning approach, in which training data are stored in the memory to determine a relevant set of data points for a query point. In contrast to global models, local models are built considering the neighboring data in a region around a query point. Basically, a local model is developed in two steps: i) a neighborhood is selected around a query point, and ii) a local model is built using the points in the neighborhood. Nearest neighbors are chosen to be the closest points to query data, and locally weighted regression (LWR) is used to fit a surface to predict the response of the query point. For some applications, contributions of neighborhood points to the model are weighted with respect to their distances to the query point. This methods makes it easy to add new training points to the dataset (Englert, 2012).

Figure 3.3 shows the difference between global and local model structures. While computation of global models may be difficult in the presence of large number of data points and model parameters, which are required to be updated, local models are easily updated for new query data. In JITL method, relevant dataset are searched to match the query data with respect to some nearest neighborhood criterion, and a local model is built using the relevant data. Finally, model output is calculated by the local model, and the current query point and the local model are discarded (Cheng and Chiu, 2004).

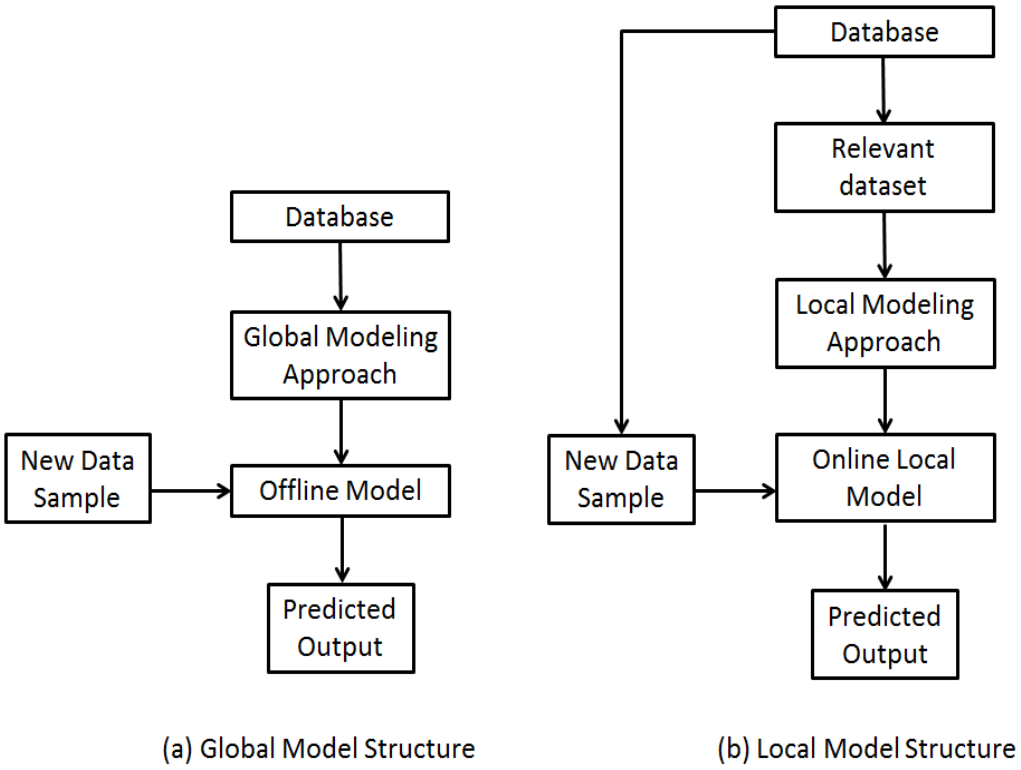

Figure 3.3. Comparison between global model and local model structure (Ge and Song, 2010) .

Figure 3.4 is a schematic representation of LW learning algorithm, and activation region and weighting curves are shown. When a query point is to be predicted, a neighborhood is determined that includes the closest data points to the query point, and weights are assigned to these neighbors. This method increases the contribution of the

closest neighborhoods on predictions and helps in obtaining more accurate results. As can be seen in weight curve, Gaussian weighting curve is used and highest weights are assigned to the closest points to the query sample, and a local linear regression model is constructed (red line in the upper the figure).

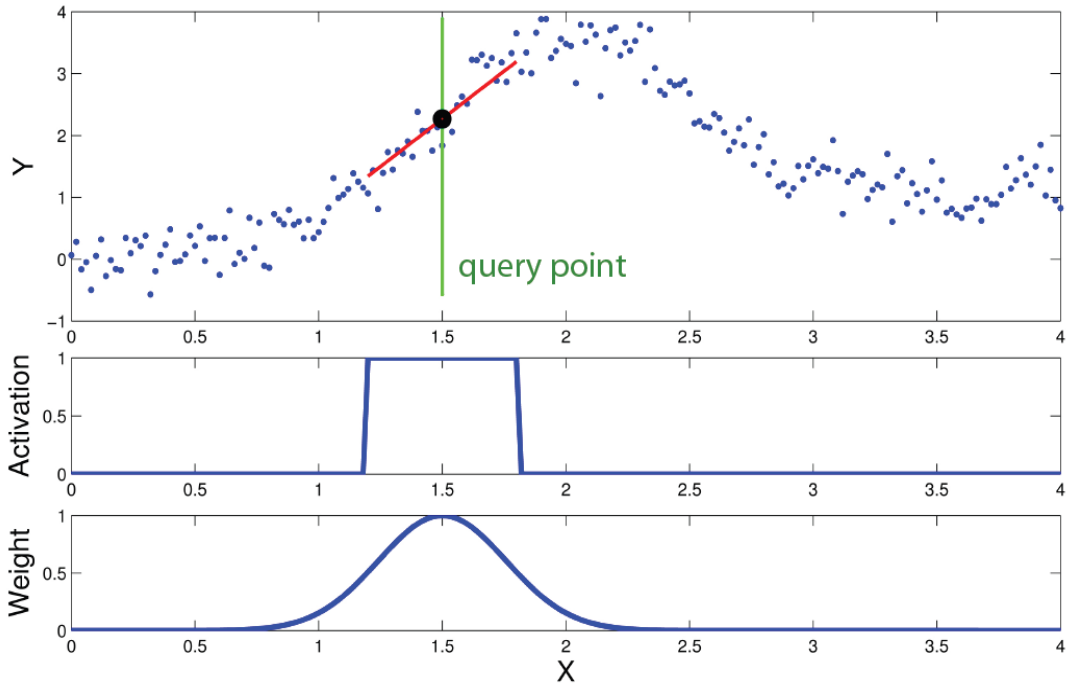

Figure 3.4. An Example of LW learning (Englert, 2012).

Overall, JITL has three main characteristics: (i) Postponing model building until an output for a given query data is requested, (ii) predicting output for the query data by exploiting the stored data in the database, and (iii) discarding results after predicted output is obtained (Saptoro 2014; Cheng *et al.*, 2004).

3.2. Similarity Criterion and Weighting Function in Local Modeling

One of the key stages in JITL modeling is to determine the neighborhood data points. Neighborhood is defined as some portion of the data, having similarity with the query point. Similarity criteria can basically be measured using three different metrics: distance

measurement, combination of distance and angle measurements, and correlation measurement (Figure 3.5).

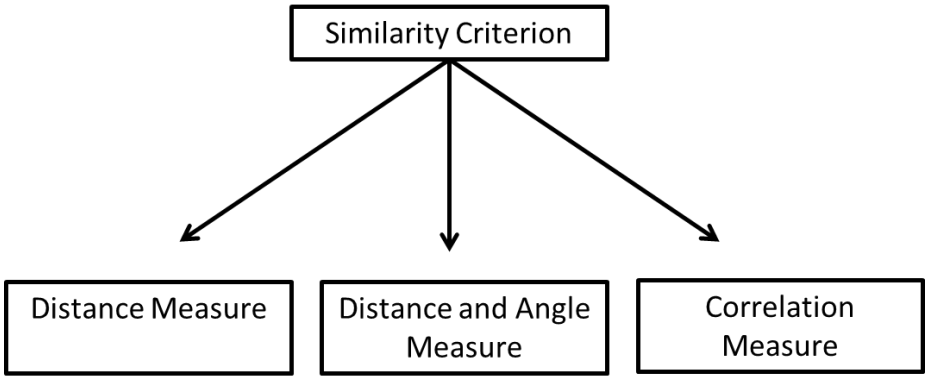

Figure 3.5. Similarity criterion for neighborhood determination in JITL applications (Saptoro, 2014).

Distance based measurements can be determined by calculating Euclidean distance, weighted form of Euclidean distance, or Mahalanobis distance, which are shown in equations (3.1), (3.2) and (3.3) respectively:

$$d^2 = (x_q - x_i)^T (x_q - x_i) \tag{3.1}$$

$$d^2 = (x_q - x_i)^T S (x_q - x_i) \tag{3.2}$$

$$md^2 = (x_q - x_i)^T C^{-1} (x_q - x_i) \tag{3.3}$$

In the equation above, S is scaling matrix, and C is covariance matrix of the neighborhood data points.

As mentioned before, distance and angle based similarity criterion can be used together, as shown in equation (3.4).

While the first term represents distance measure criterion, the second term represents the contribution of the angle measurement.

$$s_i = \gamma\sqrt{e^{-d^2(x_q,x_i)}} + (1-\gamma)cos(\theta_i) \tag{3.4}$$

Here, $d^2(x_q, x_i)$ is the Euclidean distance between x_q and x_i, γ is the weighting parameter, and $\cos(\theta_i)$ is

$$cos(\theta_i) = \frac{\Delta x_q{}^T \Delta x_i}{\|\Delta x_q\|_2 . \|\Delta x_i\|_2} \tag{3.5}$$

$$\Delta x_q = x_q - x_{q-1} \tag{3.6}$$

For Gaussian-distributed data, correlation based similarity criterion are computed as follows:

$$J = \lambda T^2 + (1-\lambda)Q \tag{3.7}$$

Here, $0 \leq \lambda \leq 1$, and T^2 and Q statistics are derived using PCA.

For high dimensional database, most of the training data stay in the far region from the query point, so zero weight is assigned to these distant points, decreasing the computational complexity significantly. Weighted JITL can be applied to systems, which are not too high dimensional, and in which too much data are not accumulated in one particular area of the input space (Schaal and Atkeson, 2002). Different weighting functions used in JITL applications in the literature are shown in Table 3.1.

3.3. Industrial Applications of JITL Modeling

Table 3.2 shows a statistics on soft sensor methodologies employed in Japan chemical industry in 2009 (Kano *et al.,* 2013). Although traditional regression analysis is

generally much more preferred and JITL models cover only ~2% of all statistical models in the industry until 2009, the number of JITL applications in industry has possibly increased in the recent years, as shown by a number of examples in the following paragraphs. In the following paragraphs, recent LWR and JITL applications in cement, petrochemical and refinery processes are discussed.

3.3.1. Blast Furnace Application

Blast furnace operation shows highly complicated and nonlinear behavior. If a large scale database is constructed for modeling, a high calculation load is difficult for online application is to be expected. To avoid these complexities, JITL modeling is employed on the large scale database (Ito *et al.*, 2004). In this study, large number of variables related to physical properties in blast furnace is measured, and used as an input to the JITL local models. Output is the molten pig iron temperature, which is predicted from these high dimensional inputs. Using 8800 data points in a large scale database for online modeling, 200 points were used for verification.

3.3.2. Blast Furnace Application

Blast furnace operation shows highly complicated and nonlinear behavior. If a large scale database is constructed for modeling, a high calculation load is difficult for online application is to be expected. To avoid these complexities, JITL modeling is employed on the large scale database (Ito *et al.*, 2004). In this study, large number of variables related to physical properties in blast furnace is measured, and used as an input to the JITL local models. Output is the molten pig iron temperature, which is predicted from these high dimensional inputs. Using 8800 data points in a large scale database for online modeling, 200 points were used for verification.

Table 3.1. Weighting functions applied in JITL.

Methods	Weighting Functions
Least- Squares function (Huang, Cabral, and Torre 2012)	$w_{LS} = 1$
Bisquare function (Huang *et al.*, 2012)	$w_B = \begin{cases} 1 & for \ \|e\| \le k \\ k/\|e\| & for \ \|e\| > k \end{cases}$
Gaussian function (Ito *et al.*, 2004)	$w_G = e^{-d^2}$
Tricube function (Ito *et al.*, 2004)	$w_T = \begin{cases} (1-d^3)^3 & for \ \|d\| \le 1 \\ 0 & otherwise \end{cases}$
Inverse Distance function (Ito *et al.*, 2004)	$w_I = \frac{1}{1+d^p}$, $p \ is \ a \ positive \ integer$

Table 3.2. Statistics of soft sensor applications in Japan industry (Kano *et al.*, 2013).

Process	Methodology				
	Phys[1]	**MRA[2]**	**PLS**	**ANN**	**JITL**
Distillation	20	256	41	0	5
Reaction	5	32	43	0	5
Polymerization	0	4	8	3	0
Others	0	1	1	0	0
Total	25	293	93	10	10

[1]: Physical model

[2]: Multiple regression analysis

3.3.3. Industrial Splitter Column and Crude Column Application

LWR approach is used on an Industrial Splitter Column, which has high dimensionality, collinearity and nonlinearity problems (Park and Han, 2000). In this study, both the splitter column, which shows a nonlinear behavior, and the crude column with linear characteristics are analyzed. For splitter column, estimation of the composition of toluene at the bottom using 16 process variables are the main objective, while, for crude column, the aim is to estimate the temperature of 90% distilled diesel, using 57 online process measurements. During soft sensor development procedure for both cases, firstly, the original variables are transformed to a small number of new variables globally using linear transformations, such as PCA and PLS, to overcome collinearity. Secondly, a local regression method, such as PLS or OLS, is employed, and predictions are obtained. In local regression, Euclidean distance of new variables and tricube function as a weighting function are used. This study demonstrates that combination of PLS and LWR may be used to simultaneously determine the latent variables and local regression with better prediction results.

3.3.4. Cracked Gasoline Fractionation Application

An application of correlation based JITL (CoJIT) modeling is applied to Cracked Gasoline Fractionation to estimate the aroma concentration accurately (Fujiware, Kano, Hasebe and Takinami, 2009). Eight of the 19 process variables are selected as input variables for the soft sensor, on the basis of physical process knowledge. Recursive PLS, conventional JITL and CoJIT modeling methods are used on the process, recursive PLS has been found not to function well. In addition, recursive PLS and conventional JITL methods are not adaptive to abrupt changes in process characteristics. CoJIT modeling improves RMSE by 28% in comparison with recursive PLS, and it is proposed that CoJIT modeling is convenient for maintaining soft sensors in the real world.

3.3.5. Debutanizer Column Application

Debutanizer column is widely used in refinery operations, in which propane and butane are removed as overhead products from the naphtha stream. Online prediction of bottom butane content is required for minimization purpose (Ge and Song, 2010) . 2000 samples are collected, and half of these samples are used as training. Local forms of artificial neural network (ANN), PLS and support vector regression (SVR) are developed to overcome the difficulties arising from process nonlinearity. The results show that JITL-LSSVR based soft sensor is the most efficient online prediction method.

3.3.6. Methane Steam Reforming Process Application

Methane steam reforming process, which is used to generate hydrogen from natural gas, consists mainly of four steps: humidification of feed gas, a reforming reaction, a shift reaction and a preferential oxidation reaction. Since the calculation accuracy of CO concentration at the outlet of the oxidation reactor is problematic, and reactor has highly nonlinear characteristics, Mahalanobis distance based Just in Time (MJIT) modeling is used to estimate product CO concentration (Nakabayashi *et al.*, 2010). In this study, Principle Component Regression (PCR) is used, two types of online adaptive methods, Kalman Filter and MJIT are compared. According to the results, while both methods are adaptive to feed changes, MJIT is more prone to update the model.

3.3.7. Polymerization Reactor Application

JITL methodology is employed on a polymerization reactor, in which an isothermal, free radical polymerization of methyl methacrylate is carried out using azo-bis-isobutyronitrile as an initiator and toluene as a solvent (Cheng, Hashimoto and Chiu, 2004). While the output variable is taken to be the number average molecular weight, the input variable is taken to be initiator flow rate. During similarity evaluation step of JITL procedure, combination of both distance and angle measure are used, and both adaptive and non-adaptive methods are applied.

4. REFINERY PROCESSES AND UNITS

Refineries are highly complex, integrated and huge industrial establishments used to convert crude oil into different ranges of products. In this section, overview of main refinery processes and relations between process units in a refinery are discussed in section 4.1. Atmospheric distillation is explained in detail in section 4.2. It is known that diesel products have a higher impact on profitability in Turkey, so a brief assessment of demand and supply strategy of Tüpraş is discussed in final section 4.3.

4.1. Overview of Refinery Processes

Petroleum refineries are large industrial complexes consisting of different processing units, utility units and tank fields. Refining begins with distillation of crude oil, i.e. separation of oil into different cuts and fractions. Since all crude oil is exposed to separation process via distillation, refinery capacity is quantified by two measurements. The first of these measurements is barrels per stream day (BPSD), which is defined as the maximum number of barrels of inputs that a distillation facility can process without downtime. The other measurement is barrels per calendar day (BPCD), that is the amount of input that a distillation facility can process under usual operating conditions, allowing for the types and grades of products to be manufactured, environmental constraints, and unscheduled and scheduled downtime due to maintenance, repairs, and shutdown (Gary *et al.,* 2007).

In refineries, crude oil is physically separated using atmospheric and vacuum distillation columns into fractions based on molecular size and boiling point ranges. After physical separation in crude distillation units, each stream is further processed using catalytic and thermal chemical conversions in order to change the size and structure of molecules. To remove undesirable contaminants in the products, to meet the specified

product quality, and to improve the current product quality, products obtained via chemical conversions are subjected to various treatment processes (Table 4.1). It is also important to note that refinery processes vary with respect to the types of crude oil they process, planning decisions, product specification, market needs, environmental and economic considerations; hence process architecture is not identical in various complex plants.

Table 4.1. Main refinery processes (Fahim, Alsahnaf and Elkilani, 2012).

Physical Separation	**Chemical Conversion**	
	Catalytic	**Thermal**
Distillation	Fluid Catalytic Cracking	Delayed Coking
Solvent Deasphalting	Hydrotreating	Flexicoking
Solvent Extraction	Hydrocracking	Visbreaking
Solvent Dewaxing	Catalytic Reforming	
	Alkylation	
	Isomerization	

Figure 4.1 shows a basic refinery flow diagram, in which crude oil enters the distillation units and converted into final product form through formulating and blending operations. In crude oil distillation units, cuts are obtained having specific boiling point ranges and can be classified in the order of decreasing volatility, as gases, light distillates, middle distillates, gas oils and residuum. Residuum can further be distilled in the vacuum towers maintaining vapor liquid flow at reduced pressures. Following the distillation unit, Liquefied Petroleum Gas (LPG) is sent to LPG treating unit, and Fuel Gas and LPG are the lightest products. Light Naphtha is sent to the isomerization unit, and isomerate is produced as the alkylation feedstock. Heavy Naphtha is processed in naphtha hydrodesulfurizer unit, in order to remove sulphur contaminant, and converted to LPG and Reformate product. All the light products are processed in Gasoline blending. Kerosene and Diesel are the main middle distillates of the atmospheric distillation column. Light Vacuum Gas Oil from vacuum distillation column is sent to the hyrodesulfurizer units, followed by distillate blending, to yield Jet, Kerosene and Diesel products. Heavy Vacuum

Gas Oil produced in vacuum units is fed to both fluid catalytic cracking (FCC) and hydrocracker units. While FCC units are the main gasoline producers in refineries, hydrocracker unit also converts heavy Vacuum Gas Oil into Gasoline, Kerosene, Diesel and Fuel Oil in the presence of hydrogen and durable catalysts.

In Figure 4.2, main products produced via refinery processes and their fields of applications are summarized. Among all the operations in oil refinery, atmospheric distillation plays the most important role in setting the desired product specifications, and increasing the yield of the desired product based on planning strategies.

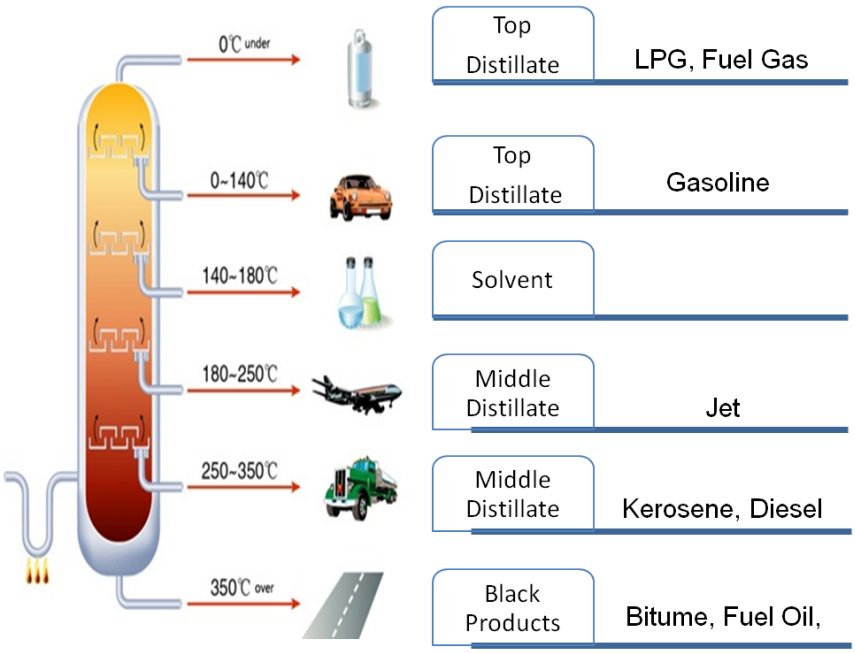

Figure 4.1. Main refinery products.

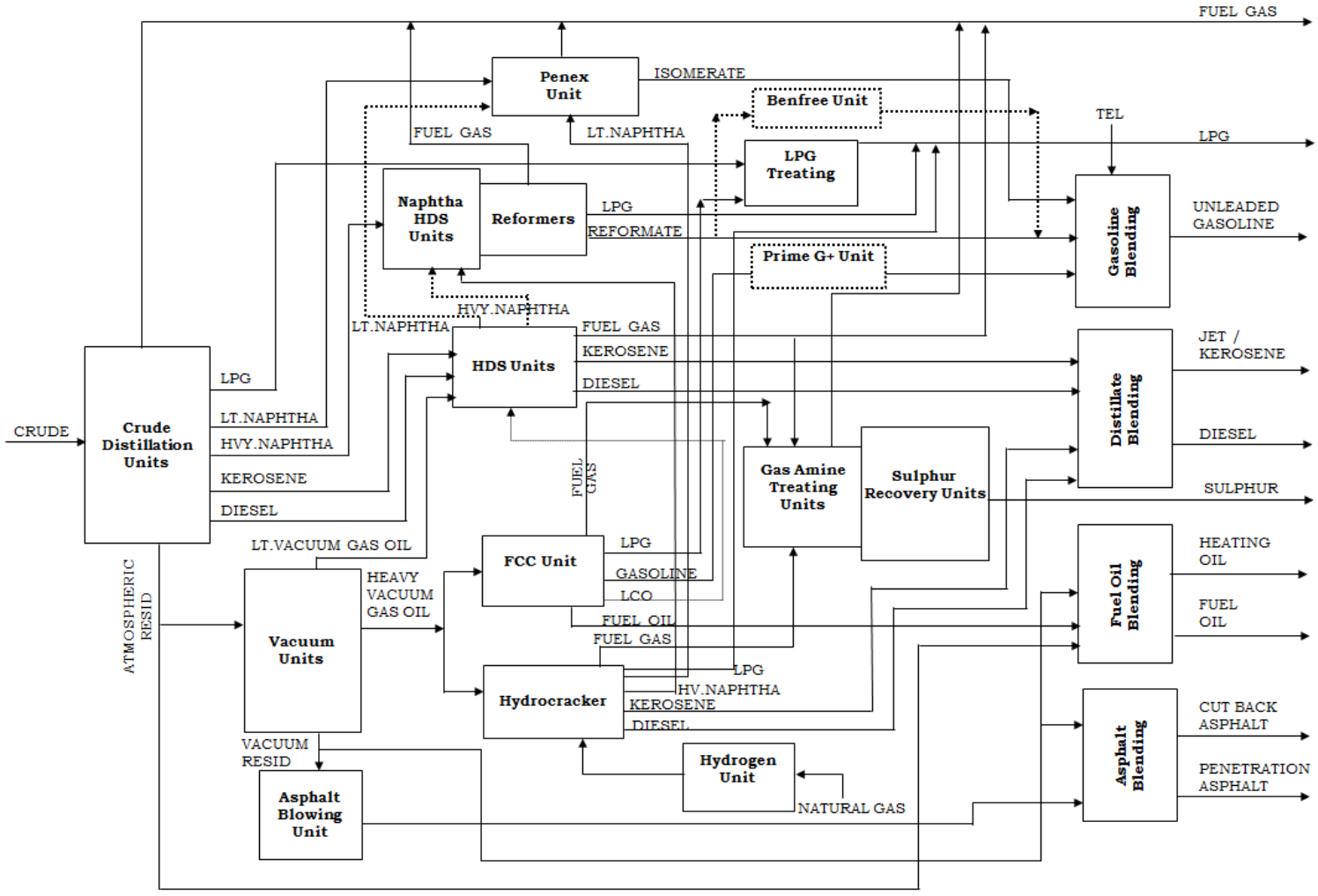

Figure 4.2. A basic refinery flow diagram

4.2. Atmospheric Distillation Units

Atmospheric distillation is the first unit in petroleum refinery, used to separate the crude oil mixture into various distillates such as Naphtha, Kerosene, Diesel, Gas Oil and Residuum as shown in Figure 4.2. Heavy middle distillates are named diesel and gas oil, but it should be noted that gas oil in the scheme refers to distillate having heavy diesel composition. After crude oil is heated in preheating section, it is sent to the desalter, in order to remove the salt, which is harmful to downstream equipment because of fouling and scaling. The desalted crude oil enters the furnace and heated approximately to 340-372 ^{o}C for partial vaporization. The stream leaving the furnace is fed to the distillation column through the feed region, called flash zone (Chang, Pashikanti and Liu, 2012).

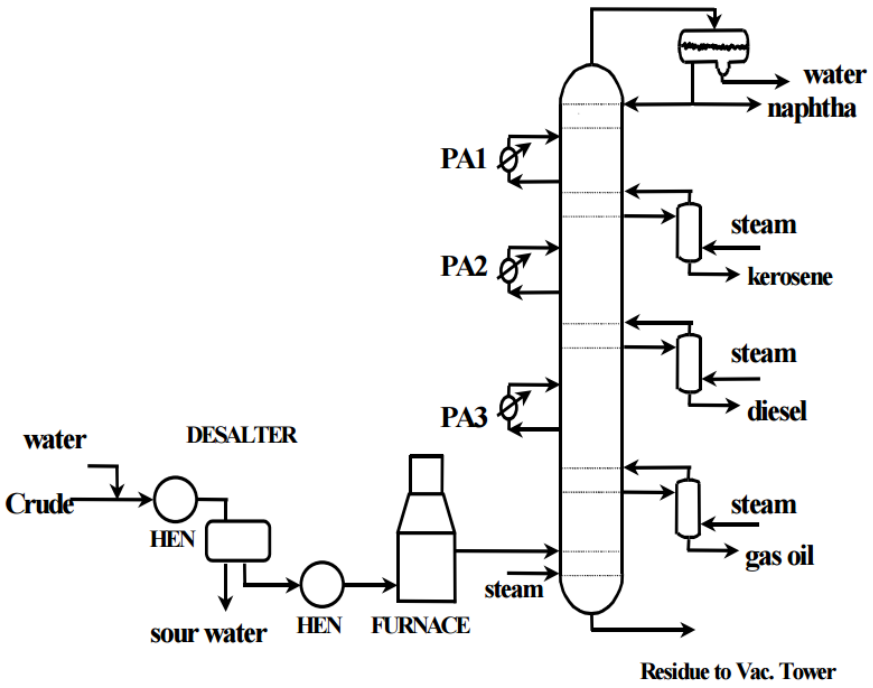

Figure 4.2. Atmospheric distillation unit.

Generally, significant amount of steam is given to the bottom of the atmospheric distillation column to reduce partial pressure of the bottom. Reduction in the partial

pressure renders stripping residue from the bottom and increasing the yield of the distillate. Middle products, on the other hand, are withdrawn from the column, and light components are stripped by superheated steam, shown by three flash units on top of each other, following the distillation column, in Figure 4.3. This operation adjusts the boiling point of 5% volume of products boiled (T5) and flash point of each product. To remove heat effectively, liquid is removed at various points of the column, called pumparounds (PA in Figure 4.3), and cooled and reinjected to the column at a different position. The overhead condenser of the atmospheric column condenses lighter products, and some portion of this condensate is returned to the top of the column, also rendering heat removal from the column.

In addition to the significant effect of pumparounds and top reflux to the heat balance of the column, these streams affect the internal reflux, which does not only contribute to the fractionation efficiency and heat balance of the column, but also play an important role in products yield (Golden, 2009). Heat input to the column resulting from the composition and temperature of feed limits the reflux rates and capability to separate products. Separation can be monitored mainly by measuring gap-overlap and cutpoint. Gap-overlap is a measure of separation between two adjacent products with respect to their exit location from the column, and cutpoint is the temperature of true boiling point (TBP) curve, which reports borders on yields of specific cuts (Sloley, 2014). To be able to monitor these performance indicators and maintain continuous column operation at high efficiency levels, robust quality estimators that give accurate and continuous estimates of product qualities are widely used in crude distillation units.

4.3. An Assessment of Refinery Products Demands and Supply in Turkey

TÜPRAŞ is the only refinery company of Turkey with 610 kb/d crude oil capacity, and processed 13 types of crude oil from nine countries in 2011. Almost 65% of crude oil processed in Turkey was medium and heavy sour crude in 2011, followed by heavy sour (28%) and light sweet (7%) (International Energy Agency 2014). However, fuel demand in the transportation sector directly affects Tüpraş's production strategies and future

investments. From 2006 to 2013, while gasoline consumption decreased in half, diesel consumption increased by 5%, and LPG demand stayed almost constant (Figure 4.4). Demand and supply fuels in 2012 shows that there is approximately 52% deficit in domestic demand of diesel (Figure 4.5), while there is high surplus in gasoline production (Atalay and Kumbaroğlu, 2014). It is previously estimated that diesel demand is expected to rise to 14 m, 15.8 m, and 18 m tones by 2010, 2015 and 2020, respectively. Tüpraş's plan to boost diesel output at the expense of standard fuel will help shift this balance in Turkey's favor (International Energy Agency, 2014).

Figure 4.3. Comparison of fuel consumption in Turkey in 2006 and 2013 (International Energy Agency, 2014).

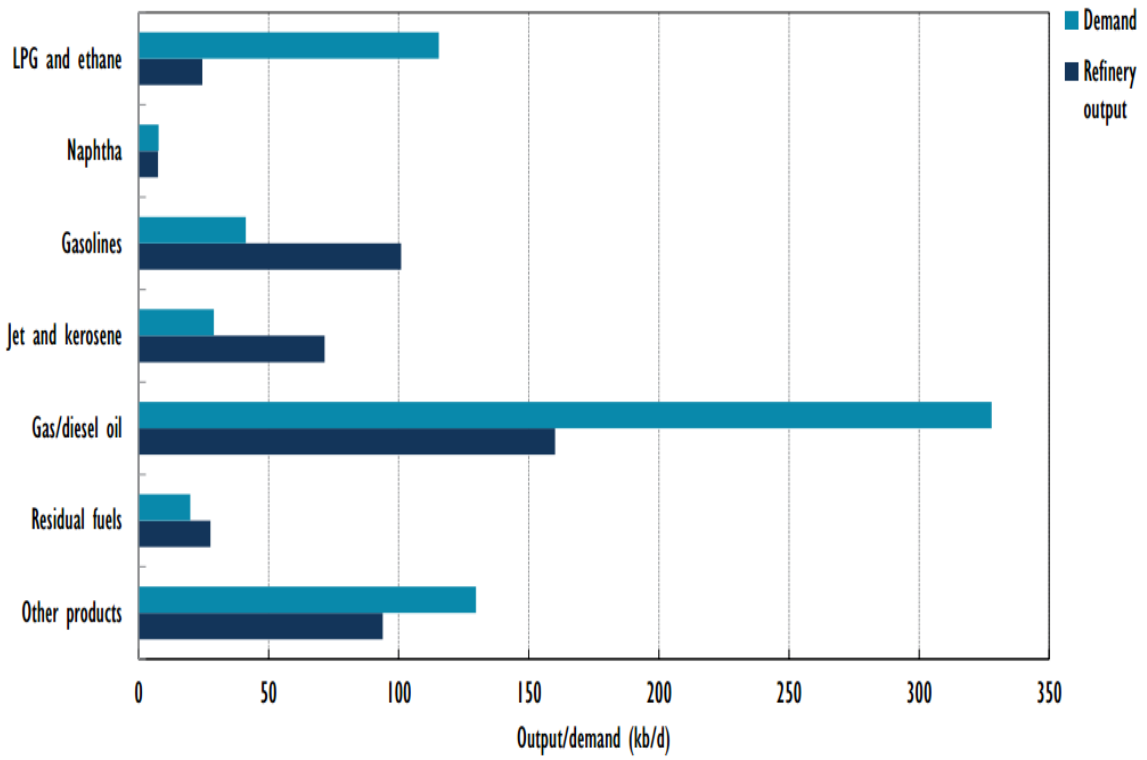

Figure 4.4. TÜPRAŞ output vs. demand in 2012 (Atalay and Kumbaroğlu, 2014).

5. EXPERIMENTAL STUDY IN TÜPRAŞ REFINERY

İzmit Tüpraş refinery has three different crude distillation units. Each of these units has different capacities and subtly differs with respect to their subunits. In Section 5.1, the distillation unit, in which variables used in JITL modeling are collected, is discussed. Historical data set collection methodology and selected parameters for modeling are elucidated in Section 5.2.

5.1. Process Unit Description of Experiment

Plant-5 is one of the main crude distillation units in İzmit Tüpraş refinery, and has maximum 13000 m3/d crude oil processing capacity. Main products of this unit is Light Vacuum Gas Oil (LVGO), Heavy Vacuum Gas Oil (HVGO), Liquified Petroleum Gas (LPG), Light Straight Run Naphtha (LSRN), Heavy Straight Run Naphtha (HSRN), (KERO), Light Diesel (LAD) and Heavy Diesel (HAD).

After heating crude oil in the heat exchanger network and fired heater, feed is sent to the atmospheric distillation column (Figure 5.1). Main products of the distillation unit are KERO, LAD and HAD, which are sent to the desulphurizer units. The top product of atmospheric distillation column is sent to the naphtha splitter columnand the debutanizer column, which yield HSRN, LPG and LSRN. While naphta products are sent directly to storage tank, LPG is processed in LPG treating unit. Atmospheric residue (the stream which leaves the bottom of the distillation column in Figure 5.1 is processed in vacuum column that is operated at negative pressure generated by steam ejectors at the top side of the column. While HVGO is sent to hydrocracker unit, LVGO is directly contributed to diesel blending.

For this complicated system, producing maximum amount of products with the desired qualities with respect to product specifications is main target for process control

structure of crude distillation units. For this purpose, quality estimators providing accurate online predictions of product qualities is crucial for process controllers to take correct actions, while trying to sustain operation at optimum conditions. However, it is expected that these quality estimators must be adaptive to changes in operational conditions, some faults in operations and unexpected and unmeasurable disturbances. With this purpose, JITL models are constructed to predict T95 of HAD product.

5.2. Selected Parameters for Historical Dataset

Historical data from the process unit is extracted from database for ~1 year. During the operation, HAD samples had been taken every day at ~06:00 a.m. and sent to the laboratory. To take the time lags in the process into consideration and to filter out the noisy components in the process measurements, process data between 02:00 and 10:00 a.m. is averaged for each day. Process measurements of temperatures, pressures and flow rates of atmospheric distillation column are shown in Table 5.1. Inclusion of these variables in JITL models have been mainly decided with respect to their significance in distillation, as dictated by field experience.

Table 5.1. Selected process variables for JITL models.

Variables	Column Parameters	Units
1	Crude Feed Flow Rate	m^3/d
2	Desalter Pressure	kg/cm^2
3	2nd group Heat Exchangers Entrance Temperature	^oC
4	HADPA Reflux Flow Rate	m^3/d
5	HADPA Reflux Temperature	^oC
6	Column Top Pressure	kg/cm^2
7	Column Top Temperature	^oC
8	Condenser Temperature	^oC

Table 5.1. Selected process variables for JITL models (cont.)

Variables	Column Parameters	Units
9	Top Reflux Temperature	^oC
11	Superheated Steam Temperature	^oC
12	Column Bottom Flow Rate,1	m^3/d
13	Column Bottom Flow Rate,2	m^3/d
14	Column Bottom Total Flow Rate	m^3/d
15	Furnace Transfer Temperature, 1	^oC
16	Furnace Transfer Temperature, 1	^oC
17	HADPA Reflux Outlet Temperature	^oC
18	Kerosene Column Outlet Temperature	^oC
19	Light Diesel Column Outlet Temperature	^oC
20	HAD Column Outlet Temperature	^oC
21	Light Diesel Flow Rate	m^3/d
22	HAD Flow Rate	m^3/d
23	Top Reflux Flow Rate	m^3/d
24	Bottom Temperature,1	^oC
25	Flash Zone Pressure	kg/cm^2
26	Condenser Pressure	kg/cm^2
27	Column Bottom Temperature,2	^oC
28	Flash Zone Temperature	^oC
29	Column Top Temperature, 2	^oC
30	Superheated Steam Flow Rate	m^3/d
31	Kerosene Top Temperature	^oC
32	Bottom Temperature,2	^oC
33	Superheated Steam Flow Rate/ Crude Feed Flow Rate	-
34	HADPA Reflux Flow Rate/ Crude Feed Flow Rate	-
35	HADPA Reflux Flow Rate/ Crude Feed Flow Rate- Column Bottom Total Flow Rate	-
36	HADPA Reflux Flow Rate/Top Reflux Flow Rate	-

Table 5.1. Selected process variables for JITL models (cont.)

Variables	Column Parameters	Units
37	Flash Zone Temperature-Column Bottom Temperature, 1	^{o}C
38	Column Bottom Temperature, 1-HAD Column Outlet Temperature	^{o}C
39	Column Bottom Temperature, 1-Kerosene Column Outlet Temperature	^{o}C
40	Bottom Temperature- Light Diesel Column Outlet Temperature	^{o}C
41	HAD Column Outlet Temperature- HADPA Reflux Temperature	^{o}C
42	Bottom Temperature, 1- Column Top Temperature	^{o}C
43	Bottom Temperature, 1- Bottom Temperature, 2	^{o}C
44	Flash Zone Temperature- Kerosene Top Temperature	^{o}C
45	HAD T95 of the day before	^{o}C
46	Logarithmic Function of Top Reflux Temperature	-
47	Logarithmic Function of Bottom Temperature,1	-

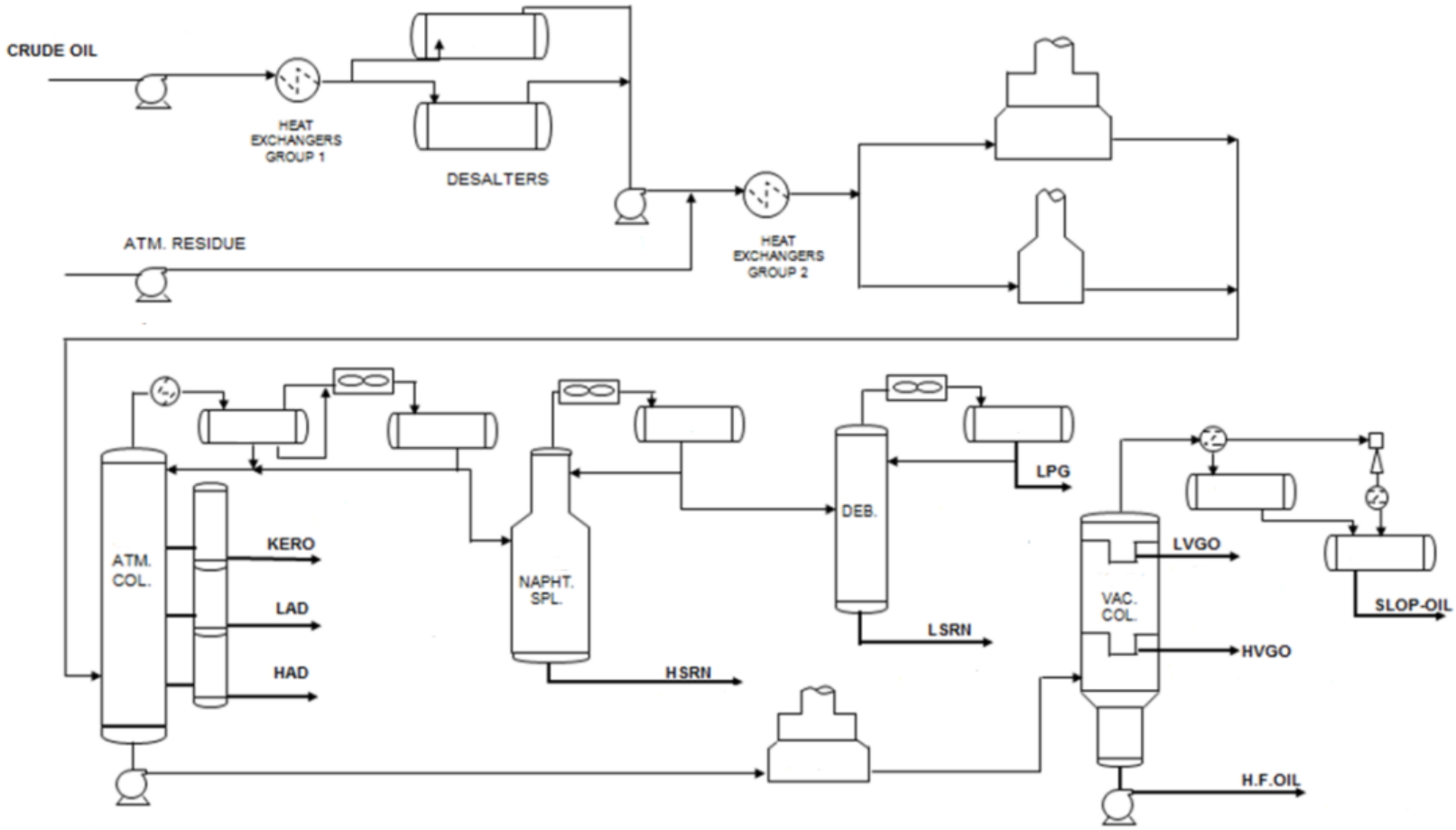

Figure 4.6. Process flow scheme of Plant 5 Crude Unit

6. RESULTS AND DISCUSSION

Predictions of HAD T95 values via JITL methods are shown and discussed in this section. JITL studies are conducted mainly in four groups. First, 47 variables, defined in the previous section, are analyzed to identify the most correlated variables with HAD T95. Second, different modeling methods are integrated to JITL methodology, and prediction accuracy is enhanced. Third, various parameters in JITL models such as reference set size, neighborhood size and sliding window size are changed to see their effects on prediction quality. Finally, besides being included of interaction and quadratic predictor terms, for the first time in the literature, neighborhood variables and predictor variable spaces are separated in the construction of JITL models.

6.1. A Preliminary Inspection of the Effects of Variable Selection Methods in Quality Prediction

Out of the 47 process variables introduced in Section 0, identification of a subset variables, which carries the highest information for determining HAD T95 values, and including these variables into predictive models are the first crucial steps for constructing reliable soft sensor models, especially in JITL methodology.

For this purpose, static and autoregressive exogenous (ARX) local models are developed in the following two sections, respectively, via selecting different variables. In static models, there is no time lag between selected input variables, and the previous value of the predicted quality variable is not included in the predictor vector. Dynamic models are ARX models, which are developed by including the previous HAD T95 measurements in the predictor vector. Here, training and validation sets consist of 100 and 264 observations, respectively, and neighborhood size is taken to be equal to 70. Similarity criterion is based on distance between observations, while tricube weighting function is used.

6.1.1. Predictions of Static Models

Static models are developed by initially including variables, which are deemed to affect product quality in the regression model, and including/excluding predictors in a stepwise fashion, depending on the RMSE and MAE statistics of the models.

Three of the obtained models with the lowest RMSE and MAE statistics are shown in Table 6.1. The predictor variables common in all three models are top fan outlet temperature (X10), column bottom flow rate (X12), HAD flow Rate (X22), bottom temperature (X24) and the difference between flash zone temperature and column bottom temperature (X37).

Table 6.1. Results of JITL application of static models.

Model	Variables	RMSE	MAE
1	X4,X10,X12,X22,X24,X37	6.79	5.23
2	X9,X10,X12,X22,X24,X37	6.83	5.26
3	X28,X10,X12,X22,X24,X37	6.91	5.24

In the current study, in addition to measures, which are used to state the performance of the predictors for the whole validation set, such as RMSE and MAE, validation set is divided into various intervals, and correlation of the laboratory measurements with the predictions and predictive bias are computed for each of these intervals. The motivation here is to see whether the predictive performance of JITL models changes with respect to sampling time in the validation set.

HAD T95 predictions of Model 1 and 2 are shown in Figure 6.1. Furthermore, two different intervals in the validation set, the first (interval 1) between samples 40 and 150, and the second (interval 2) between 150 and 264 (the index of the final sample) are analyzed for both static model 1 and static model 2. In general, correlation between laboratory measurements and predictions is higher in interval 2 compared to interval 1 (Table 6.2, Table 6.3).

Table 6.2. Correlation and bias between laboratory measurements and predictions via static model 1 and static model 2 for interval 1.

	Static Model 1	**Static Model 2**
Correlation	0.23	0.18
Bias	-0.56	-1.56

Table 6.3. Correlation and bias between laboratory measurements and predictions via static model 1 and static model 2 for interval 2

	Static Model 1	**Static Model 2**
Correlation	0.36	0.24
Bias	-0.83	-0.59

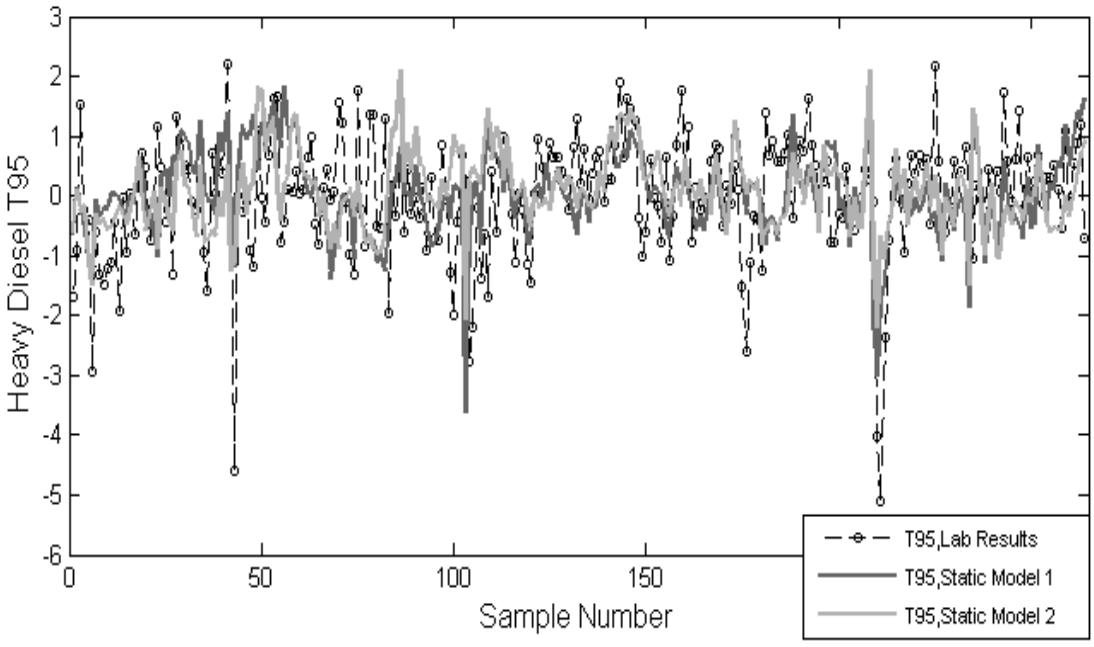

Figure 6.1. HAD T95 predictions vs. sample number of static model 1 and static model 2.

6.1.2. Prediction of ARX Models

ARX Models are developed by including previous HAD T95 measurements to the inputs. In these models, predictor variables are identical to those used in Section 6.1.1, and stepwise regression is performed to select the statistically significant regressors. P-values of 0.15 and 0.20 are taken as the threshold values to include and remove variables in the predictor subset, while variables, which are highly correlated with output variable, are "forced" to be included in the regression model. Table 6.4 shows the dynamic models with the lowest RMSE and MAE values. In addition to predictors used in the static models, light diesel flow rate (X21), top reflux flow rate (X23) and as the day before HAD T95 values (X45) are also included in the development of the dynamic models.

Table 6.4. Results of JITL ARX models variables selected by stepwise regression.

Model	**Variables**	**Obligatory Variables**	**RMSE**	**MAE**
1	X10,X12,X21,X22,X23,X24,X28,X37,X45	X22,X45	6.08	4.56
2	X12,X16,X21,X22,X23,X24,X28,X42,X45	X22,X45	6.17	4.74
3	X14,X16,X21,X22,X23,X24,X28,X35,X36 X42,X45	X22,X45	6.14	4.71

HAD T95 predictions of ARX models 1 and 3 are shown in Figure 6.2 to 6.4. Prediction performances of the estimators on the validation data is evaluated on the same two intervals used in the previous section. By addition of AR character, correlations are increased for both data regions (Table 6.5 and Table 6.6).Although predictive bias for interval 1 decreases significantly, compared to that of static model 1, bias increases slightly for interval 2. In summary, one may say that adding a dynamic element to the JITL model and choosing the regressors in a stepwise fashion decrease the prediction errors significantly.

Table 6.5. Correlation and bias between laboratory measurements and predictions of ARX model 3 and Static model 1 for interval 1

	ARX Model 3	**Static Model 1**
Correlation	0.27	0.23
Bias	0.06	-0.56

Table 6.6. Correlation and bias between real samples and predictions of ARX model 3 and static model 1 for interval 2.

	ARX Model 3	**Static Model 1**
Correlation	0.52	0.36
Bias	0.97	0.83

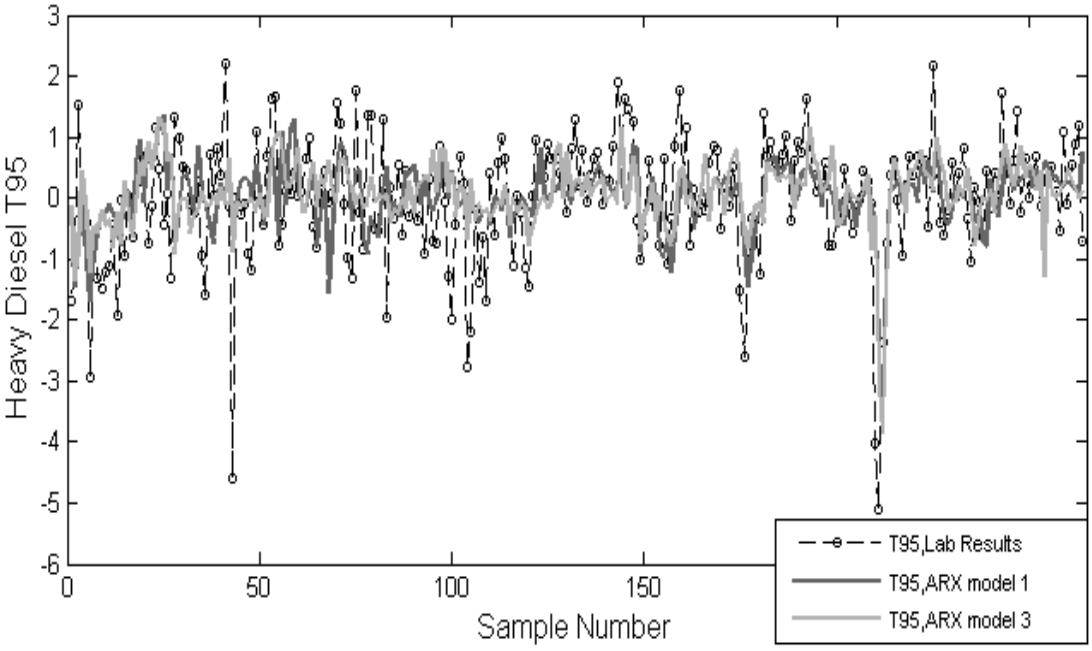

Figure 6.2. HAD T95 predictions vs. sample number of ARX model 1 and model 3.

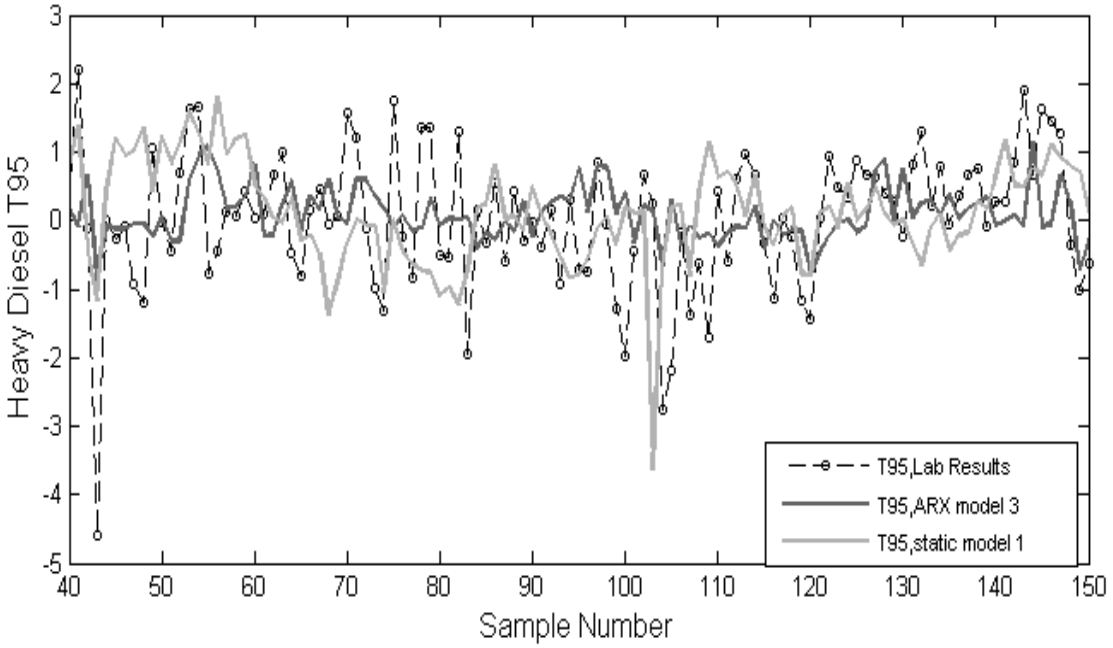

Figure 6.3. HAD T95 predictions vs. sample number between 40 and 150 predictions of ARX model 3 and static model 1.

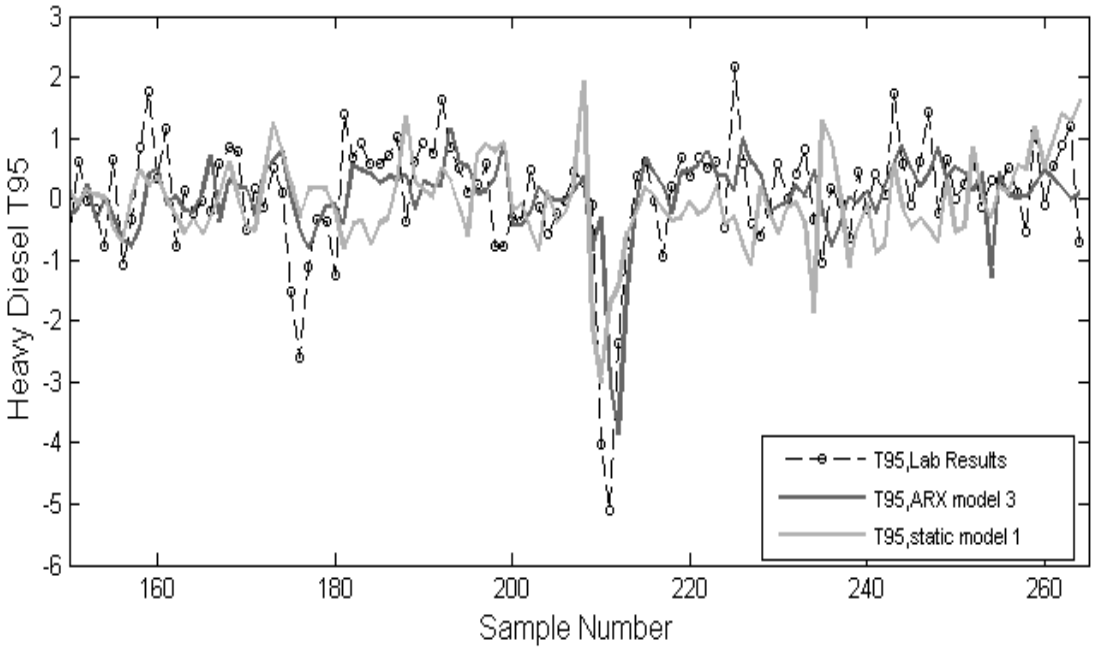

Figure 6.4. HAD T95 predictions vs. sample number between 150 and 265 of ARX model 3 and static model 1.

6.2. Effect of Modeling Methods on Prediction Quality

In the previous section, various predictor subsets from 47 process variables were constructed, and static and ARX local models were developed and assessed with respect to their RMSE and MAE values. In the current section, a constant subset of seven variables is chosen with respect to process experience, and different modeling methods are applied on this subset of crude feed flow rate (X1), inlet temperature of the 2nd group heat exchangers (X3), HAD flow rate (X22) and the ratio of HADPA reflux flow rate to top reflux flow rate (X36) are assumed to be the most important process variables, which affect the variation in HAD quality. Due to process engineers practice, crude oil feed and HAD flow rate are related to HAD yield, which may provide insight to HAD product quality. In practice, temperature difference between the inlet and outlet of the 2nd group heat exchangers is known to be correlated with feed type changes. Refluxes of atmospheric distillation column are used to remove heat from the distillation column, so they are known to have significant effect on column temperature profile, and compositions. In addition to the four predictors discussed above, logarithmic functions of process variables, top reflux temperature (X46) and bottom temperature (X47), which have the highest difference, are included in the predictor subset. ARX character of the model is rendered via adding the previous HAD T95 measurement in the predictor subset.

All models are developed by taking reference and validation set sizes as 100 and 264 data points, respectively, and using a sliding window, i.e. reference set is shifted by one observation when a new query point is predicted, with a constant size of 100 observations. Similarity criterion is based on distance between observations. Two neighborhood sizes with 70 and 99 observations, and uniform and tricube weighting functions are used in all models.

6.2.1. Prediction of Global Models

In this section, LS and PLS analysis are performed globally using three different selection methods of reference sets: Constant reference dataset (CRD), incremental reference dataset (IRD), sliding window reference dataset (sWRD) . Reference set is taken

to be equal to the first 100 data points in the historical dataset for CRD regression; reference set is increased by one sample including the previous query point data in IRD; and reference set is shifted by one observation upon the arrival of a new query point in sWRD.

6.2.1.1. LS Analysis. RMSE and MAE values of validation set show that LS models constructed using CRD and IRD perform equally well (Table 6.7). It is seen that T95 predictions obtained both via CRD and IRD describe the trends in the laboratory measurements well (Figure 6.5).

Table 6.7. Results of Global LS Models with respect to different selection methods of reference data.

Model	Selection method of reference data	RMSE	MAE
1	CRD	6.09	4.63
2	IRD	6.17	4.63
3	sWRD	6.41	4.80

6.2.1.2. PLS Analysis. RMSE and MAE values of validation set shows that PLS models constructed using CRD and IRD perform equally well (Table 6.7 and Table 6.8), and slightly lower than those obtained from LS models. It is seen that T95 predictions obtained both via CRD and IRD characterize the trends in the laboratory measurements well (Figure 6.6).

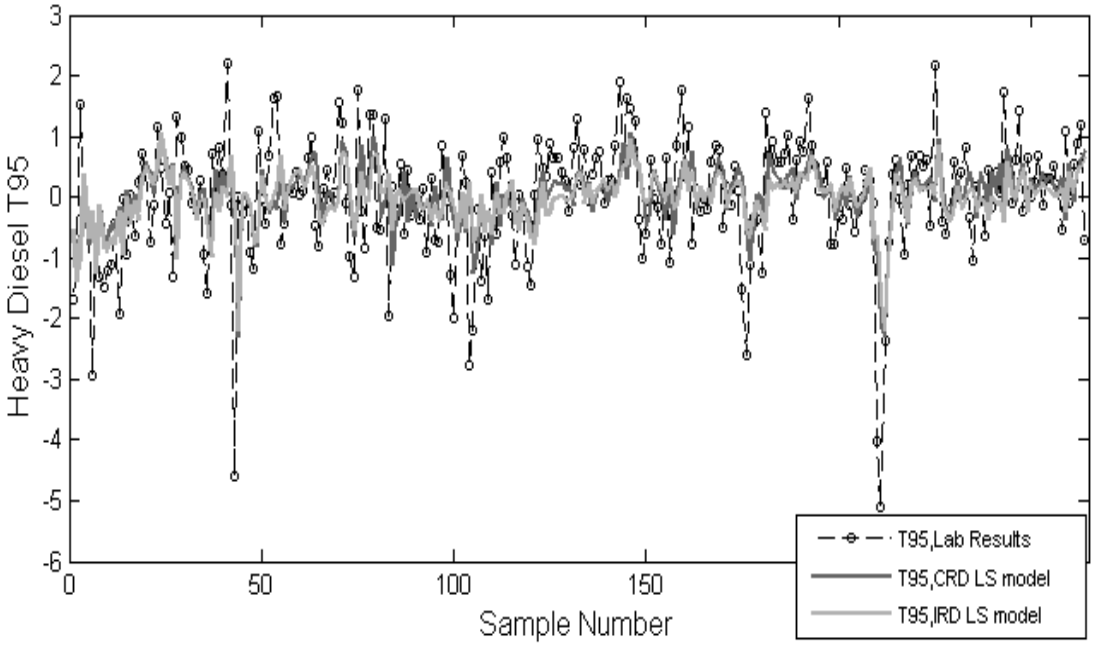

Figure 6.5. HAD T95 predictions vs. sample number of CRD LS and IRD LS models.

Table 6.8. Results of Global PLS Models.

Model	Selection method of reference data	RMSE	MAE
1	CRD	6.04	4.55
2	IRD	5.97	4.51
3	sWRD	6.15	4.74

It is interesting that local ARX models (Section 6.1.2 and Table 6.4) gives better prediction results than global models applied on sWRD. In addition, global model results developed using seven "intelligently" chosen variables are close to the results obtained from local models with variables selected by stepwise regression (Section 6.1.2).

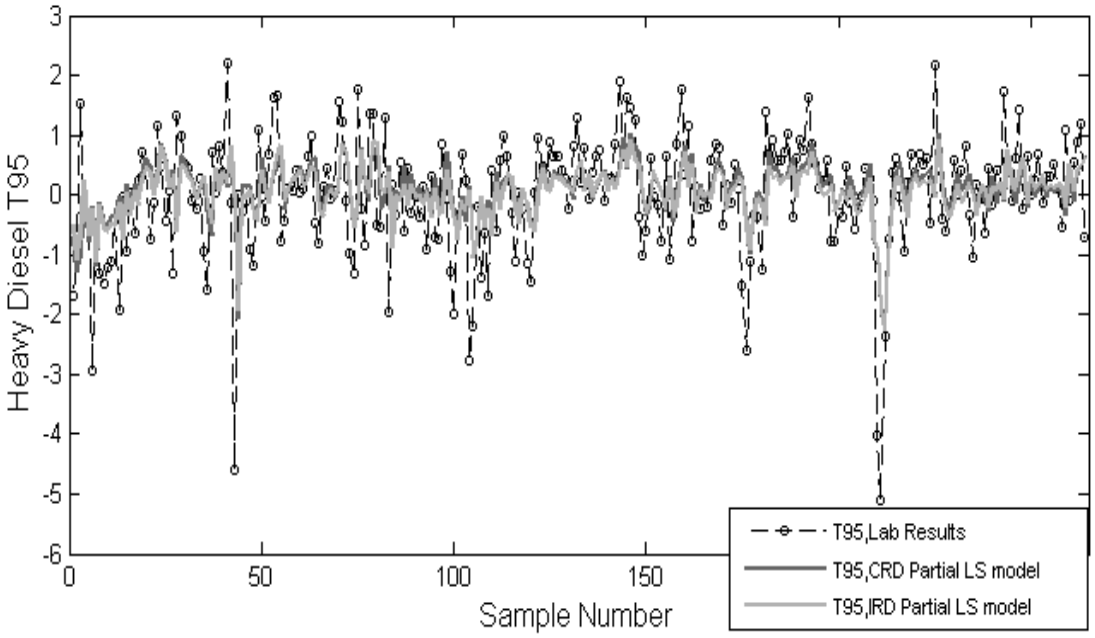

Figure 6.6. HAD T95 predictions vs. sample number of CRD PLS and IRD PLS models.

6.2.2. Prediction by JITL Models and Effect of Neighborhood size in JITL Applications

In this section, all models are constructed using JITL applied on sWRD using 100 observations, since this dataset selection method gave the best results for JITL models (Section 6.1.2). Here, local LS and PLS models are employed, using various subsets of predictors, determined via stepwise regression. Various predictor variables, assumed to be highly related with T95 values, are obligatory included in the LS and PLS models, while rest of the predictors are selected based on their Akaike Information Criterion (AIC), and p-values in stepwise regression. During stepwise regression, predictors are included in the model when individual p-values are smaller than 0.15, and removed from the model when individual p-values exceed 0.20.

All seven predictors are included in Local LS and PLS models constructed at two neighborhood sizes equal to 70 and 99 observations using uniform and tricube weighting functions. Errors are found to be lower for uniform weighting functions (for 99 closest neigbors are closer to each other than those obtained for 70 closest neighbors) (Table 6.9).

This may be caused by the small reference set and neighborhood size preventing to determine adequate number of closest points, since RMSE values obtained via uniform and tricube weighting functions for 99 closest neighbors are closer to each other than those obtained for 70 closest neighbors.

Table 6.9. Results of JITL LS models.

Model	Weighting Function	Neighborhood Size	RMSE	MAE
1	Uniform	70	6.54	4.94
2	Tricube	70	7.10	5.16
3	Uniform	99	6.45	4.84
4	Tricube	99	6.50	4.91

For the rest of the section, all predictions of T95 trajectories are plotted for uniform weighted models. When predictions from models 1 and 3 in Table 6.9 are compared, they represent similar trends (Figure 6.7).

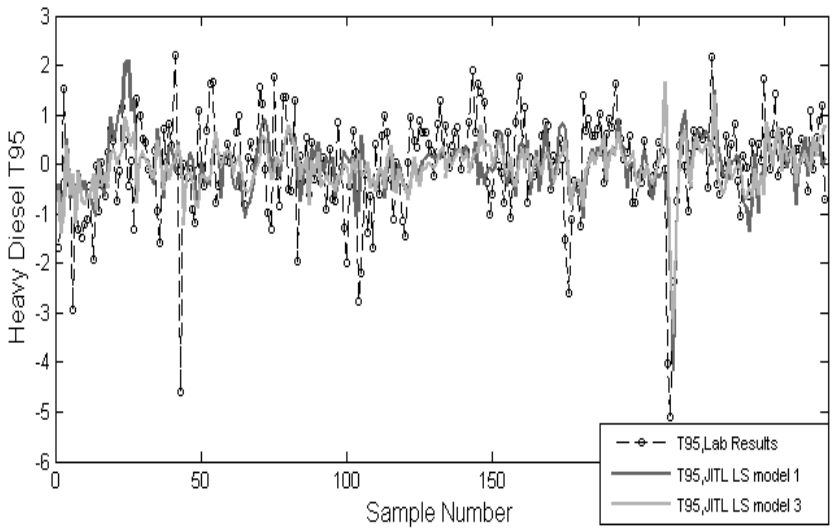

Figure 6.7. HAD T95 predictions vs. sample number of JITL LS model 1 and model 3.

Table 6.10. Results of JITL PLS Models.

Model	Neighborhood Size	RMSE	MAE
1	70	6.40	4.84
2	99	6.33	4.75

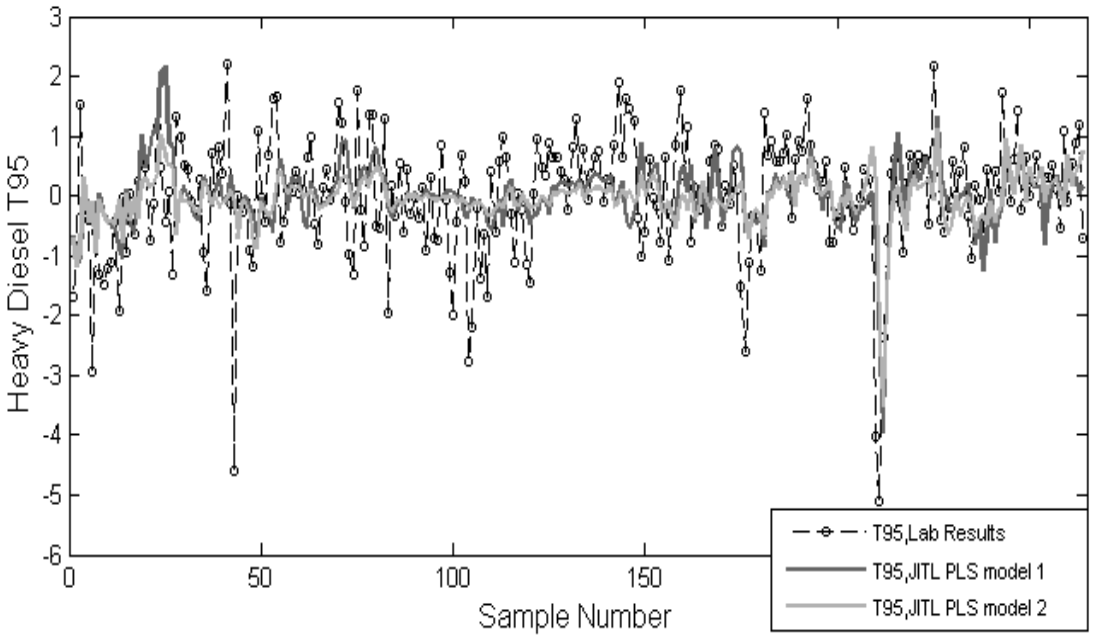

Figure 6.8. HAD T95 predictions vs. sample number of JITL PLS models with different neighborhood sizes.

Stepwise regression is employed on the subset of predictors containing seven process variables. Three types of variable selection methods are employed on stepwise regression. The first selection method uses either AIC, which is a measure of the predictive quality of the model, or p-values to include predictors into the model, while a number of the variables is compulsory included in the model, irrespective of their p-values, and the rest of the predictors are included in the models using their p-values in stepwise regression. Results of the stepwise regression method using the first variable selection method gave unsatisfactory results with RMSE values higher than 6.50.

X3, 2nd group heat exchangers entrance temperature, and X45, HAD T95 measurement of the day before, are directly included to predictive variables for all samples. RMSE and MAE values on validation set using stepwise regression with the second variable selection method are shown in Table 6.11. Model 1 and 2, for which uniform weighting functions with neighborhood sizes 70 and 99 are used, respectively, give the lowest RMSE and MAE values in Figure 6.9

Table 6.11. Results of JITL SR with obligatory variables.

Model	**Weighting Function**	**Obligatory Variables**	**Neighborhood Size**	**RMSE**	**MAE**
1	Uniform	X3, X45	70	6.32	4.80
3	Tricube	X3, X45	70	6.88	5.06
2	Uniform	X3, X45	99	6.22	4.66
4	Tricube	X3, X45	99	6.52	4.88

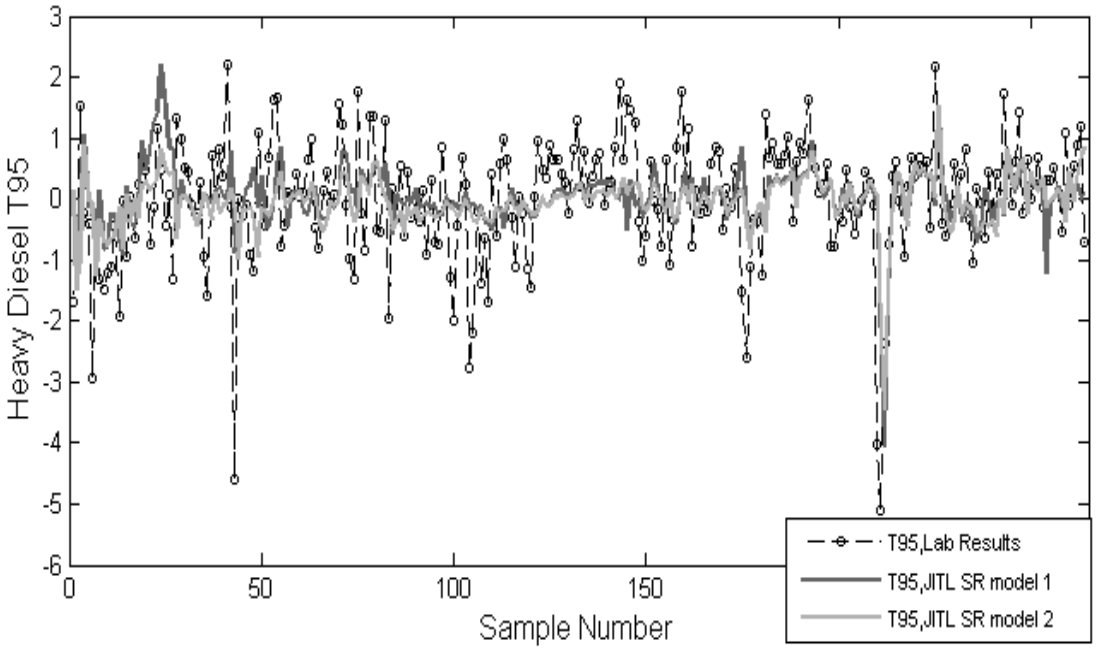

Figure 6.9. HAD T95 predictions vs. sample number of JITL SR model 1 and model 2.

In the following analysis, JITL SR model 2, JITL LS model 1 and CRD Global LS model, which have the lowest RMSE and MAE values, are compared, using the two intervals defined in Section 6.1. CRD Global LS model have highest correlation and lowest bias for both intervals (Table 6.13). Correlation of JITL SR model 2 predictions with the laboratory measurements is close to that of CRD Global LS model, although bias is higher than other two models for interval 1. Correlations of these three models are nearly same for samples in the interval 2, and bias is highest for JITL LS model 1 (Table 6.12). This shows that a stepwise predictor selection method may yield a JITL model with a higher predictive power compared to a JITL model with a pre-determined subset of predictors. Figure 6.10 to Figure 6.13 show that, except a number of intermittent time segments, JITL SR model 2 and CRD Global LS model predictions exhibit similar trends for both regions. Results in this section show that JITL model performance may change with respect to time and operation conditions.

Table 6.12. Correlation and bias between laboratory measurements and predictions of JITL SR model 2, JITL LS model 1 and CRD global LS model for interval 1.

	JITL SR Model 2	**JITL LS Model 1**	**CRD Global LS**
Correlation	0.24	0.11	0.26
Bias	0.70	0.59	0.32

Table 6.13. Correlation and bias between laboratory measurements and predictions of JITL SR model 2, JITL LS model 1 and CRD global LS model for interval 2.

	JITL SR Model 2	**JITL LS Model 1**	**CRD Global LS**
Correlation	0.45	0.49	0.53
Bias	0.44	0.79	-0.04

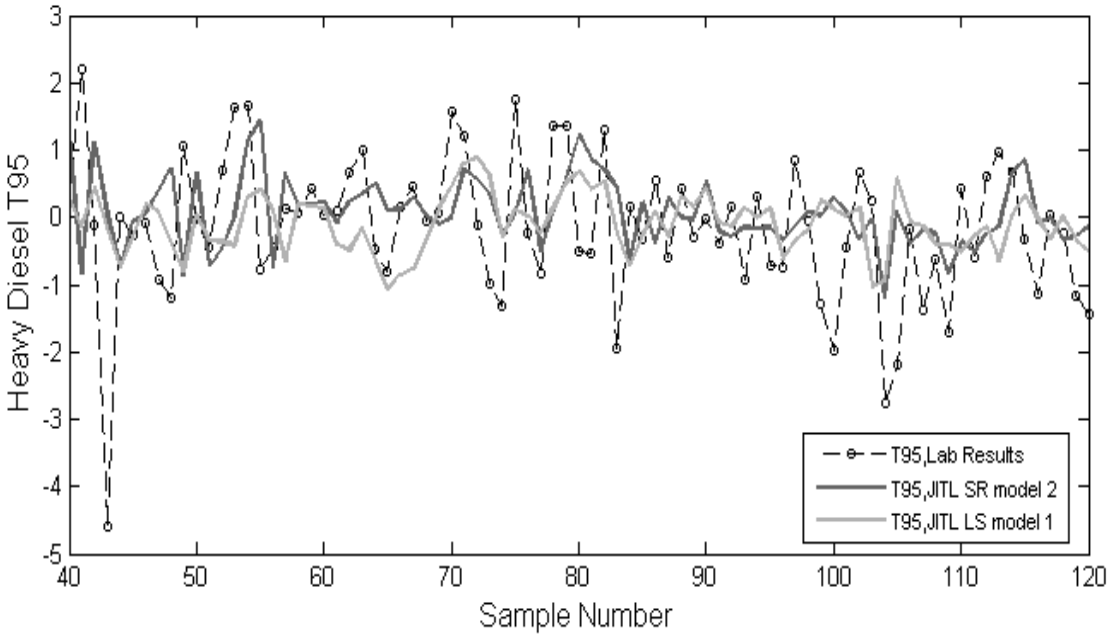

Figure 6.10. HAD T95 predictions vs. sample number of JITL SR model 2 and JITL LS model 1 for interval 1.

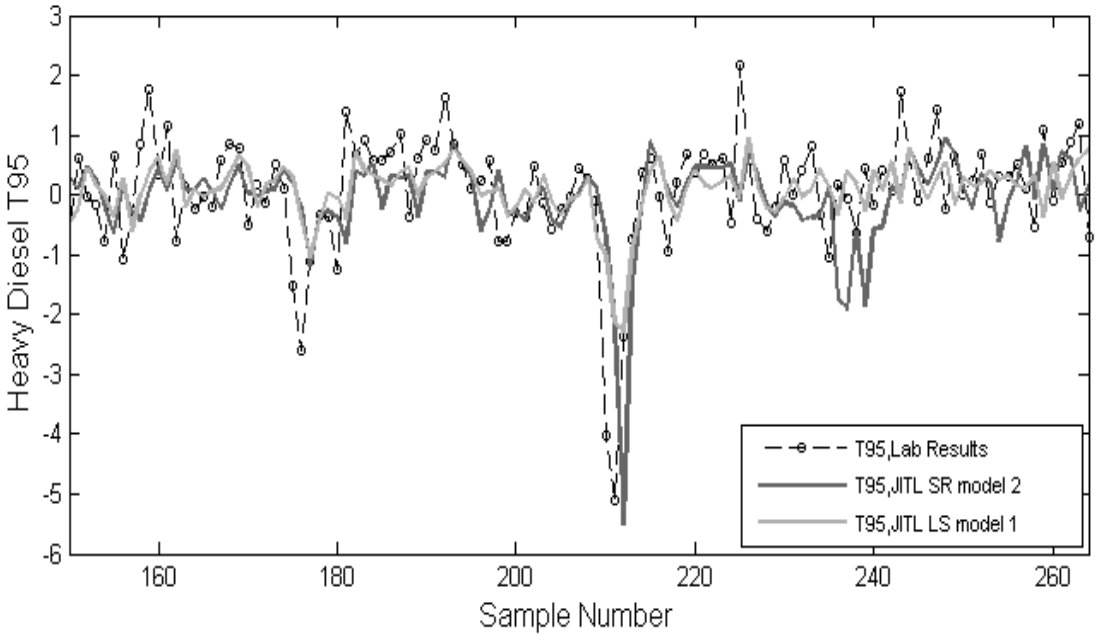

Figure 6.11. HAD T95 predictions vs. sample number of JITL SR model 2 and JITL LS model 1 for interval 2.

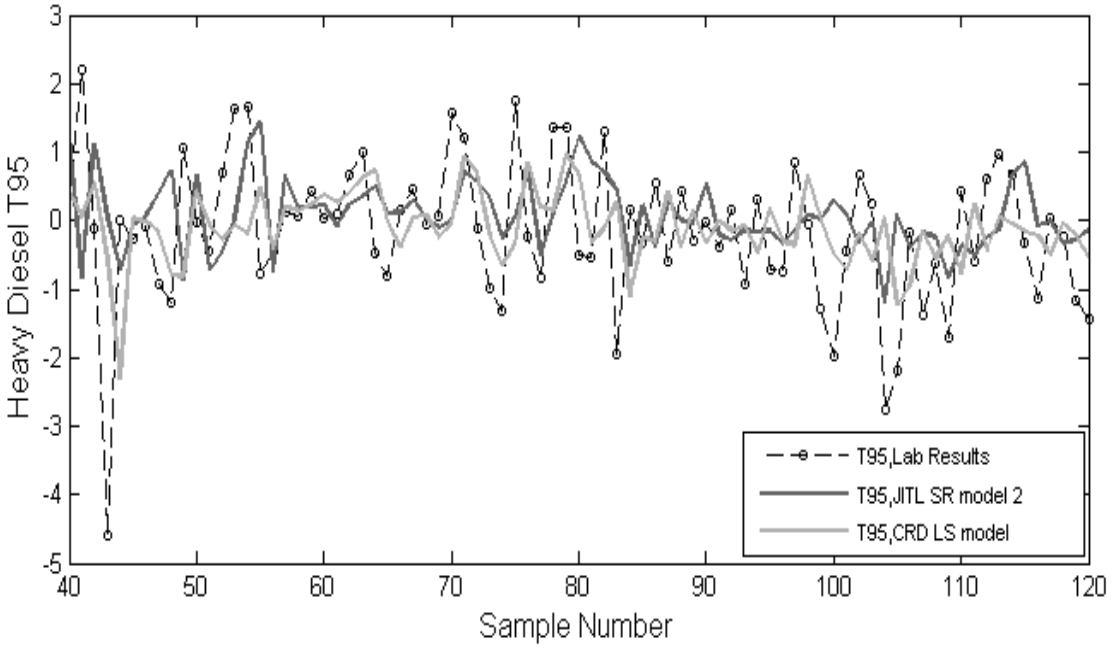

Figure 6.12. HAD T95 predictions vs. sample number of JITL SR model 2 and global LS constant model for interval 1.

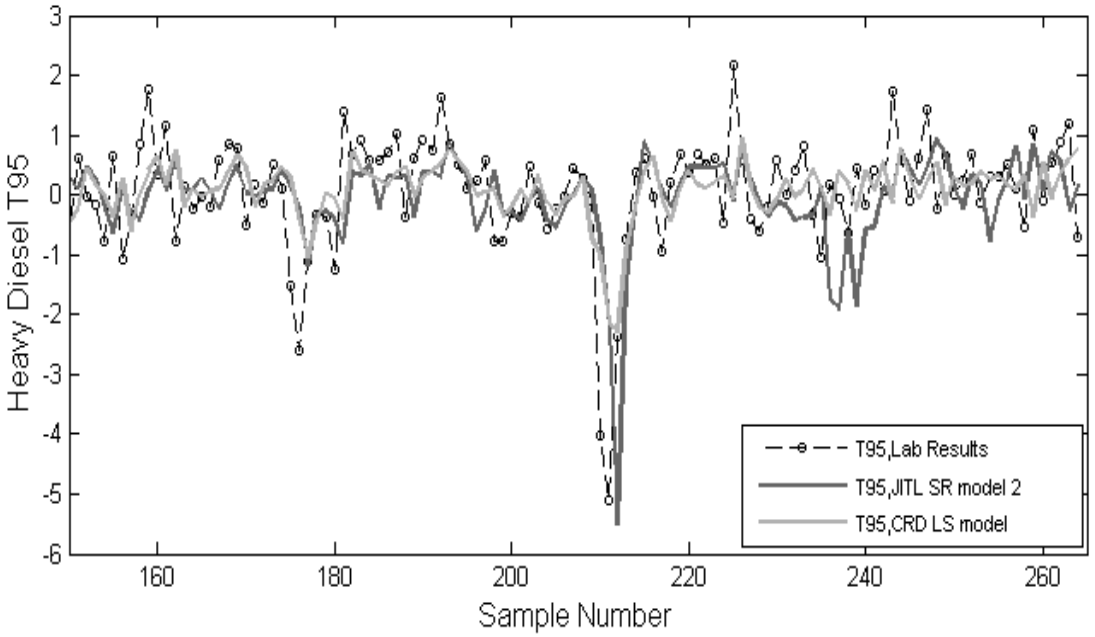

Figure 6.13. HAD T95 predictions vs. sample number of JITL SR model 2 and CRD global LS model for interval 2.

6.3. Effect of Reference Set, Neighborhood and Window Sizes on Prediction Quality

Various modeling techniques based on global and local models of LS and PLS analyses are conducted by considering constant subset of seven variables, as performed in the previous section. In the previous analyses, reference set and test data sizes were taken to be equal to 100 and 264, respectively, and a sliding window of 100 observations was used in constructing sWRD models. In the current analysis, reference set size is increased to 200 observations, and developed models are validated by 164 samples. A sliding window with 200 observations is used in the sWRD model. Two neighborhood sizes with 120 and 185 observations, and uniform and tricube weighting functions are used.

6.3.1. Predictions of Global Models using Increased Reference Set and Window Size

In this section, LS and PLS analyses are performed globally using three different selection methods of reference sets: CRD, IRD and sWRD. These models are built by using 200 reference data, and validated by 164 test data.

6.3.1.1. LS Analysis for Increased Reference Set Size. Table 6.14 shows the prediction RMSE and MAE values of the LS models developed by CRD, IRD and sWRD methods on validation set, on 164 test data. In the table, the value on the left and right of the slash represent the prediction error statistics using the initial 200 observations, as performed in the current section, and initial 100 observations, as performed in the previous section on the same 164 test data.

It should be noted that model predictions via the IRD method, which consists of accumulated reference observations, are identical using both sets, so a single prediction error value is shown in the table. It is rather surprising that the global model using a constant reference dataset of 100 observations is still the best predictive model among all tested models. CRD global LS model prediction trajectories for 200 and 100 reference set sizes are shown in Figure 6.14.

Table 6.14. Results of global LS models for increased reference set size.

Model	Selection method of reference data	RMSE	MAE
1	CRD	5.68/5.42	4.31/4.18
2	IRD	5.59	4.23
3	sWRD	5.75/5.97	4.36/4.52

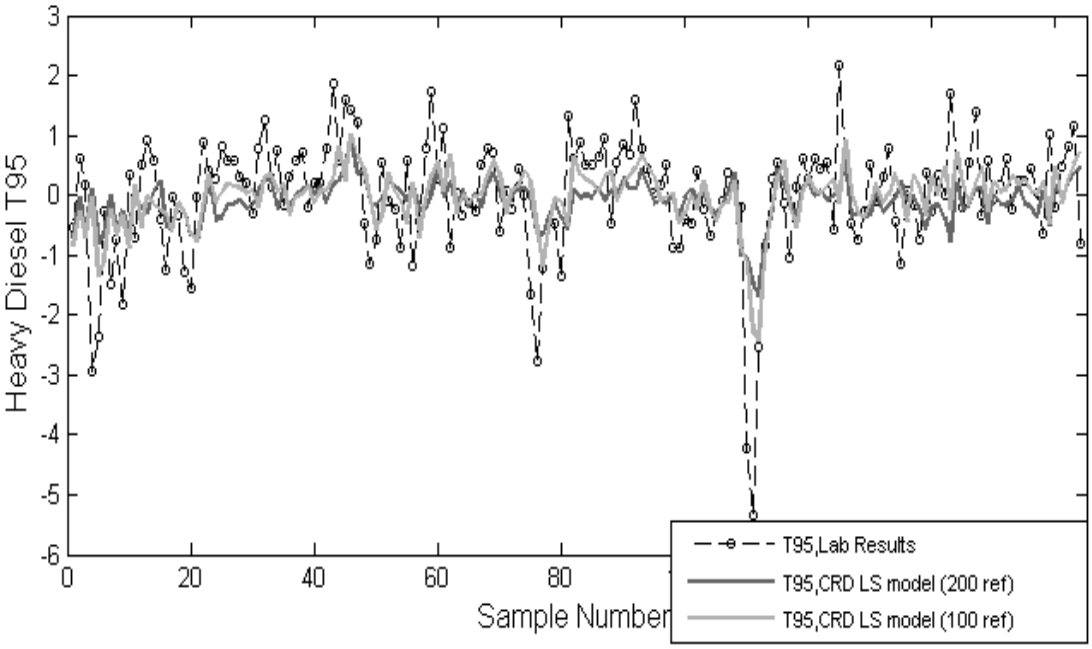

Figure 6.14. HAD T95 predictions vs. sample number of CRD global LS models for 200 reference set size and 100 reference set size.

6.3.1.2. PLS Analysis for Increased Reference Set Size. Table 6.15 shows the prediction RMSE and MAE values of the PLS models developed by CRD, IRD and sWRD methods on validation set, on 164 test data. Similar to that seen for LS model results, global models using a window size of 100 observations have higher predictive power compared to global models using a window size of 200 observations. sWRD global PLS model prediction trajectories for 200 and 100 reference set sizes are shown in Figure 6.15.

Table 6.15. Results of global PLS models for increased reference set size.

Model	Selection method of reference data	RMSE	MAE
1	CRD	5.87/5.37	4.34/4.04
2	IRD	5.35	4.06
3	sWRD	5.30/5.23	4.10/4.14

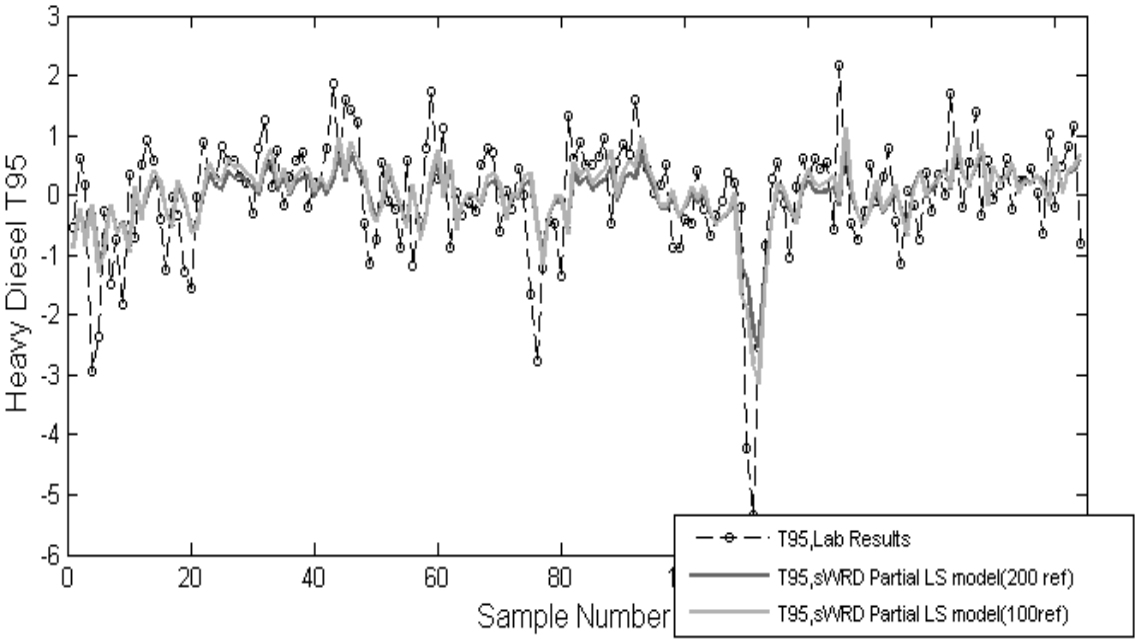

Figure 6.15. HAD T95 predictions vs. sample number of sWRD global PLS models for 200 reference set size and 100 reference set size.

6.3.2. Predictions of JITL Models using Increased Reference Set and Window Size

In this section, all models are constructed using JITL model applied only on sWRD. Here, local LS and PLS models are employed, using various predictors determined via stepwise regression. A number of predictor variables, which are assumed to be highly related with T95 values, are obligatory included in the LS and PLS models, while the rest of the predictors are selected based on their Akaike Information Criterion (AIC), and p-values in stepwise regression. During stepwise regression, predictors are included in the

model when individual p-values are smaller than 0.15, and removed from the model when individual p-values exceed 0.20.

Seven predictors are included in the construction of JITL LS models using 200 reference data set with 120 and 185 neighborhood data points, and using uniform and tricube weighting functions. In contrast to what is seen in global models, JITL LS models developed by 100 data points in the reference set have higher RMSE than the those constructed from 200 data points in the reference set (Table 6.16). This shows that having a larger reference set from which the neighboring points are chosen is more convenient for the performance of a JITL model. In addition, as in Section 6.2.2, models have uniform weighting function give better results than the ones with tricube weighting. However, when neighborhood size increases from 120 to 185, RMSE values from the models with different weighting function (model 3 and model 4) get close to each other, showing that weighting function may play a more important role in the predictive quality for smaller neighborhood size.

Prediction errors of local PLS models using uniform weighting on validation data are shown in Table 6.17. RMSE and MAE values are improved compared to JITL LS models, and prediction errors for local PLS models at different neighborhood sizes are similar.

Then, stepwise regression is employed on the subset of predictors containing seven process variables by increasing reference set size to 200. This procedure is performed for two different neighborhood sizes: 120 and 185. Two types of variable selection methods are employed on stepwise regression. The first selection method uses either AIC or p-values, to include predictors into the model, while a number of the variables, irrespective of their p-values, is compulsory included in the model. Results of the stepwise regression method using AIC selection methods gave unsatisfactory results, so the results obtained by p-value screening are shown in the following analysis.

Table 6.18 shows model results for 200 reference data set, 120 and 185 neighborhoods. Sliding window size is regarded as 185. JITL SR analysis for 100 reference data set, 70 and 99 neighborhoods are also conducted by validating model with 164 data

are shown in in Table 6.18. Sliding window size is regarded as 100 for this part. JITL SR models are net affected too much by reference set size in contrast to global models so these models can handle with nonlinear data behavior and wide ranging data sets.

Table 6.16. Results of JITL LS models for increased reference set size.

Model	**Weighting Function**	**Reference Set size**	**Neighborhood Size**	**RMSE**	**MAE**
1	Uniform	200	120	5.87	4.47
2	Tricube	200	120	6.31	4.54
3	Uniform	200	185	5.74	4.38
4	Tricube	200	185	5.73	4.37
1*	Uniform	100	70	5.93	4.53
3*	Uniform	100	99	6.02	4.57

*: Models 1 and 3 with reference set size equal to 100 refer to the previous section.

Table 6.17. Results of JITL PLS models for increased reference set size.

Model	**Neighborhood Size**	**RMSE**	**MAE**
1	120	5.94	4.42
2	185	5.81	4.35
1*	70	5.80	4.43
2*	99	5.92	4.48

*: Models 1 and 2 with reference set sizes equal to 70 and 99 refer to the previous section.

Table 6.18. Results of JITL application via stepwise regression with obligatory variables for increased reference set size and window size.

Model	Weighting Function	Obligatory Variables	Neighborhood Size	RMSE	MAE
1	Uniform	X3, X45	120	5.66	4.23
2	Tricube	X3, X45	120	6.21	4.41
3	Uniform	X3, X45	185	5.85	4,38
4	Tricube	X3, X45	185	5.74	4.32
1*	Uniform	X3, X45	70	5.82	4.36
2*	Tricube	X3, X45	70	6.36	4.53
3*	Uniform	X3, X45	99	5.72	4.30
4*	Tricube	X3, X45	99	6.05	4.49

6.4. Addition of Higher order terms and changing the neighborhood variables in the JITL models

In the previous section, none JITL models show better performance than global models so in this section, besides being chosen different variables, interaction and quadratic predictor terms are also included in the JITL models firstly. Crude feed flow rate (X1), HADPA reflux flow rate (X4), column top temperature (X7), furnace transfer temperature (X16), HADPA reflux outlet temperature (X17), Light Diesel flow rate (X21), bottom temperature (X24) and ARX character that is previous HAD T95 measurements (X45) are included to model predictors. Quadratic term of HAD flow rate (X222), interaction between furnace transfer temperature and HAD flow rate (X16×X22) and interaction between Light Diesel flow rate and HAD flow rate (X21×X22) are also included to model.

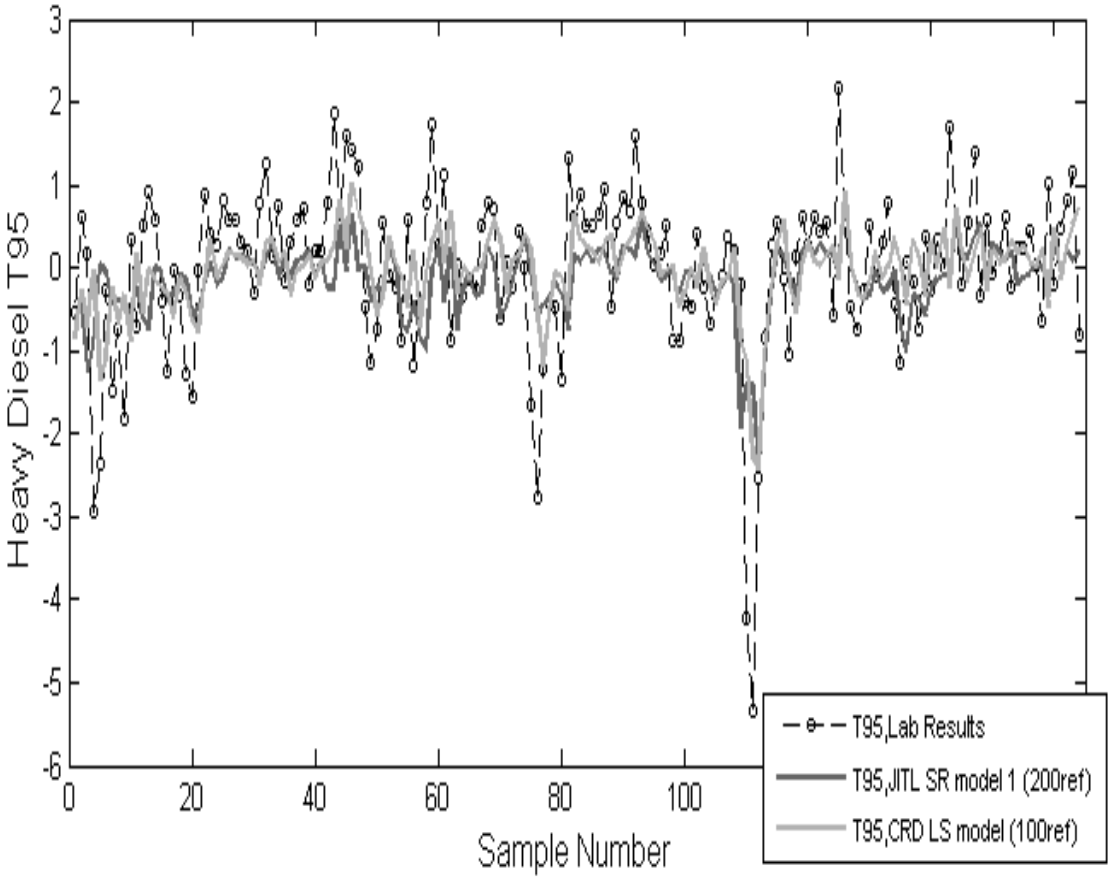

Figure 6.16. HAD T95 predictions vs. sample number of JITL SR model 1 for increased reference set size and global LS constant model for 100 reference set size.

$$T95 = \beta_0 + \beta_1 X1 + \beta_2 X4 + \beta_3 X7 + \beta_4 X16 + \beta_5 X17 + \beta_6 X21 + \beta_7 X22 + \beta_8 X24 + \beta_9 X45 + \beta_{27} X16 \times X16 + \beta_{67} X21 \times X22 + \beta_{66} X22^2 \quad (6.2)$$

Second, for the first time in the literature, neighborhood variables and predictor variable spaces are separated in the construction of JITL models. As in Section 6.3, reference set size is taken as 200 observations and developed models are validated by 164 samples. Both CRD, IRD and sWRD methods of reference data selection are considered. Neighborhood size is taken as 150 and uniform weighting function is used in the current section.

6.4.1. Predictions of Global LS Models with Higher Order Terms using Increased Reference Set and Window Size

In this section, LS and PLS analyses are performed globally using three different selection methods of reference sets: CRD, IRD and sWRD. These models are built by using 200 reference data, and validated by 164 test data.

6.4.1.1. LS Analysis for Increased Reference Set Size and Addition of Higher Order Terms. Table 6.19 shows the prediction RMSE and MAE values of the LS models developed by CRD, IRD and sWRD methods on validation set, on 164 test data. Models developed by IRD method has higher predictive models compared to other global models, especially sWRD model. sWRD global LS model performance using a window size of 200 observations deteriorates with these new predictor variables. CRD and IRD global LS model prediction trajectories for 200 reference set size are shown in Figure 6.17.

Table 6.19. Results of global LS models with higher order terms for increased reference set size.

Model	Selection method of reference data	RMSE	MAE
1	CRD	5.59	4.36
2	IRD	5.46	4.19
3	sWRD	7.05	4.53

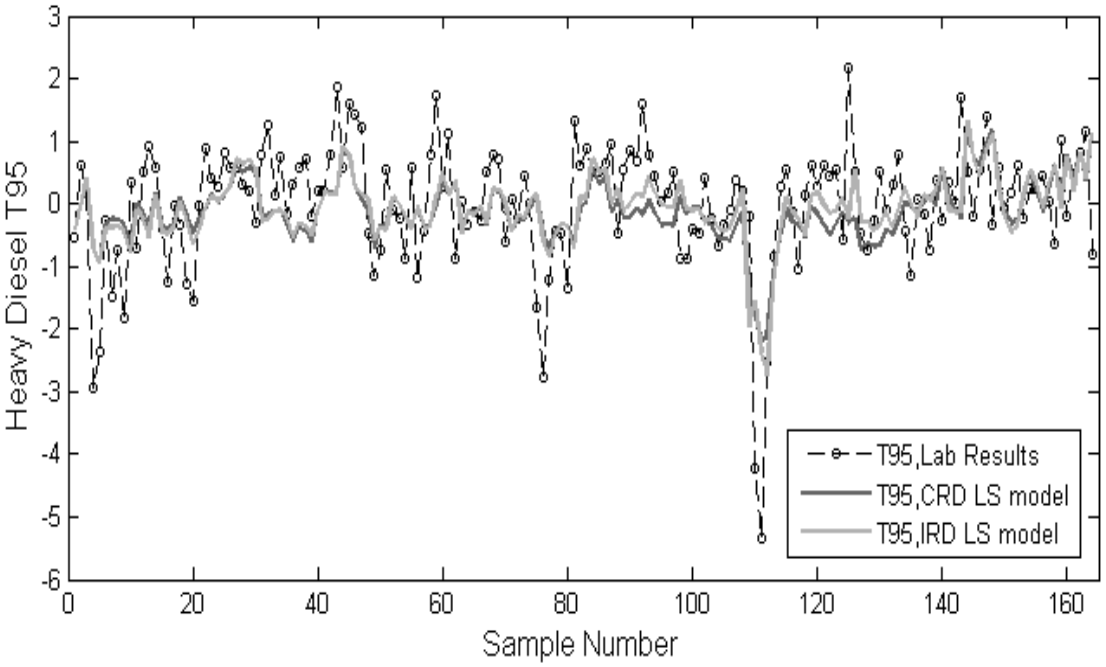

Figure 6.17. HAD T95 predictions vs. sample number of CRD and IRD global LS models with higher order terms for 200 reference set size.

6.4.2. Predictions of JITL LS Models with higher order terms and changing neighborhood variables using Increased Reference Set Size

In this section, all models are constructed using JITL LS model applied on both CRD, IRD and sWRD. Here, constant nine predictors and higher order terms of some predictors are all included in the construction of JITL LS models while different variables are taken into consideration during neighborhood selection step. Top fan outlet temperature (X10), Light Diesel column outlet temperature (X19) and HAD column outlet temperature (X20) are included to variables considered in neighborhood selection. During these studies, reference set size is taken as 200 observations and developed models are validated by 164 samples. Neighborhood size is taken as 150 and uniform weighting function is used in the current section. Window size is selected as 200 for sWRD reference date selection method.

Three different groups of variables are considered during neighborhood selection step in this study and these are all applied on both CRD, IRD and sWRD. In contrast to

what is seen in global LS models, CRD method has the highest predictive performance and has lowest RMSE among all models while sWRD has the highest RMSE for all constructed models. The best model in Section 6.3 was JITL SR model 1 and RMSE of this model was equal to 5.66. When different model variables are chosen including interaction and quadratic terms to the model and all variables are used in neighborhood selection, RMSE value is equal 5.05 for CRD method as shown in CRD JITL LS model 0 (Table 6.20). CRD JITL LS model 1 constructed via different variables and considering different group of variables during neighborhood selection phase, RMSE value are decreased to 4.77. Both JITL SR models adopted sWRD method in the previous section and JITL LS models applied on sWRD in the current section have high RMSE values ranging between approximately 5.6 to 5.8 and predictive performance are lower than JITL LS models applied on CRD and IRD. JITL SR model 1 and CRD JITL LS model 1 prediction trajectories for 200 reference set size are shown in Figure 6.18.

6.5. Comparison of the Prediction Performance of the Constructed Models

In models construction, it is seen that some crucial steps have significant effect on model performances. First, predictor variables selection is highly significant in developing accurate soft sensor for refinery applications. While RMSE values of the static models are approximately equal to 6.8 ^{o}C, RMSE is decreased significantly, when lagged T95 measurements are included in the model. Two different constant predictor subsets are considered in Section 6.2, 6.3 and 6.4. In both predictor subset groups, variables representing the column temperature profile, such as furnace outlet temperature, column top and bottom temperatures, product draw temperatures, side reflux temperatures and flow rates are found to be significant in predicting the HAD T95. Second, LS and PLS models are used in local and global regression analysis. Three types of JITL models, i.e. local linear regression (JITL LS), local PLS analysis (JITL PLS), and local linear regression parameters selected by stepwise regression (JITL SR), are developed in Section 6.2 and 6.3. RMSE of the CRD LS model, which is constructed from the first 100 observations in the training set, and found to be the global model with the smallest RMSE, is 5.42 ^{o}C, while that of JITL SR model 1, which is repeatedly constructed using the most recent 200 observations, and found to be the local model with the smallest RMSE, is 5.66 ^{o}C. Hence, JITL models developed in Section 6.2 and 6.3 could not outperform global

models. In Section 6.4, different variables including interaction and quadratic predictor terms are considered and variables used in neighborhood selection and predictor spaces are separated while constructing CRD JITL models. Reconstructing the local model using different process variables including interaction and quadratic terms, RMSE value is decreased to 5.05 ^{o}C (CRD JITL LS model 0). Using the local regression model with different neighbor selection variables (CRD JITL LS model 1), RMSE value is further decreased to 4.77 ^{o}C. A summary of the predictive performance of various models is listed in Table 6.21. It should furthermore be noted that an error margin of ± 3 ^{o}C is considered to be acceptable for the laboratory HAD T95 measurements in TÜPRAŞ refinery, so the accuracy and precision of the predictive algorithm suggested in the current thesis seem to be satisfactory for operational purposes.

Table 6.20. Results of JITL LS models for increased reference set size.

Model	**Selection method of reference data**	**Variables used in neighborhood selection**	**RMSE**	**MAE**
0	CRD	X1, X4, X7, X16,X17, X21, X22, X24,X45	5.05	4.00
	IRD		5.50	4.16
	sWRD		5.72	4.49
1	CRD	X1, X7, X10, X20, X21, X22, X23	4.77	3.82
	IRD		5.01	4.03
	sWRD		5.73	4.49
2	CRD	X1, X10, X20, X21, X22, X23	4.78	3.84
	IRD		4.87	3.95
	sWRD		5.71	4.49
3	CRD	X1, X10, X19, X21, X22, X23	4.80	3.81
	IRD		5.02	4.11
	sWRD		5.72	4.48

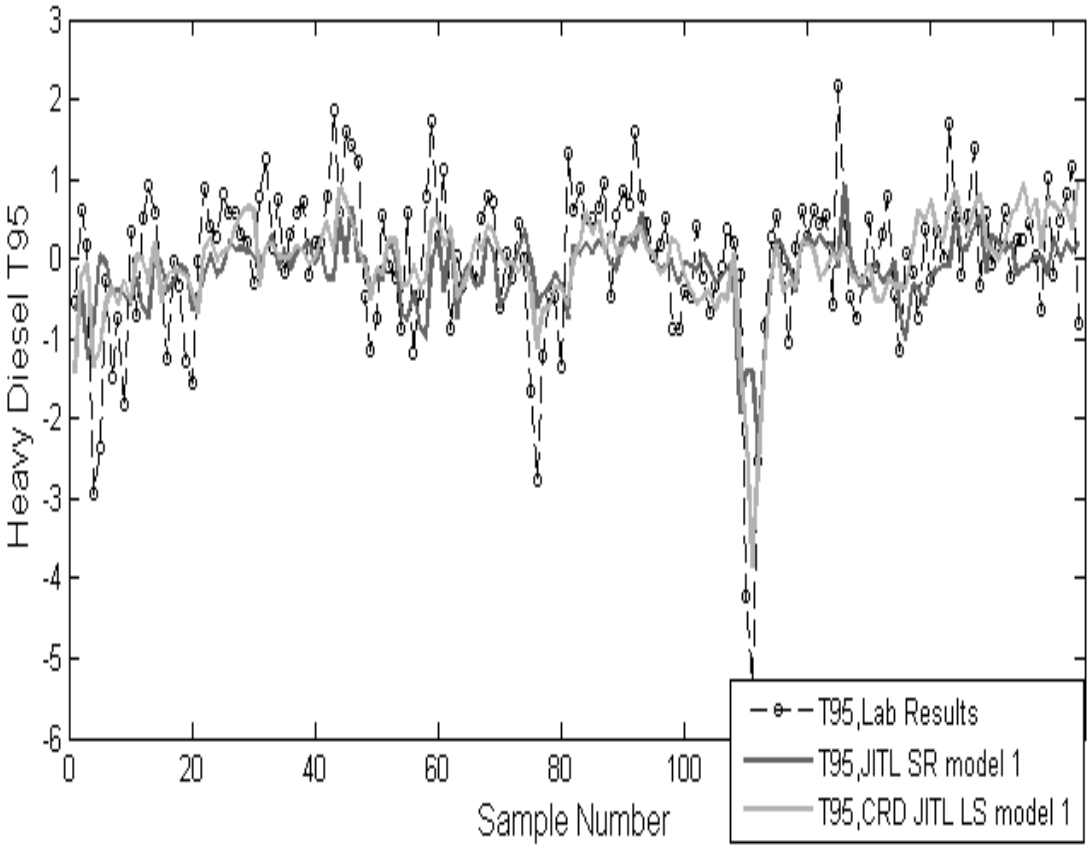

Figure 6.18. HAD T95 predictions vs. sample number of JITL SR model 1 and CRD JITL LS model 1 for 200 reference set size

Table 6.21. Selected JITL models having different properties

Model	**Predictor Variables**	**Selection Method of reference data**	**Variables used in neighborhood selection**	**Neighborhood Size**	**RMSE**	**MAE**
Static model 1	X4,X10,X12,X22,X24 ,X37	CRD	X4,X10,X12,X22,X24, X37	70	6,79	5.23
ARX model 1	X10,X12,X21,X22,X2 3,X24,X28,X37,X45	sWRD	X10,X12,X21,X22,X23, X24,X28,X37,X45	70	6.08	4.56
JITL SR model 1 (100ref)	X1,X3,X22,X36,X45, X46,X47	sWRD	X1,X3,X22,X36,X45, X46,X47	99	5.66	4.23
JITL SR model 1 (200ref)	X1,X3,X22,X36,X45, X46,X47	sWRD	X1,X3,X22,X36,X45, X46,X47	120	5.72	4.30
CRD JITL LS model 0	X1, X4, X7, X16,X17, X21, X22, X24,X45	CRD	X1, X4, X7, X16,X17, X21, X22, X24,X45	150	5.05	4.00
CRD JITL LS model 1	X1, X4, X7, X16,X17, X21, X22, X24,X45	CRD	X1, X7, X10, X20, X21, X22, X23	150	4.77	3.82

7. CONCLUSION AND RECOMMENDATIONS

Soft sensors are mathematical algorithms, which produce real time predictions of unmeasured variables using mechanistic models or historical data. In Izmit Tüpraş refinery, prediction of HAD T95 is a challenge because of the instantaneous changes in operational conditions, feed compositions, instrumental and experimental errors. In the current thesis, just in time learning (JITL) methodology, also known as locally weighted, instance-based, or lazy learning models, is used on historical process data to develop soft sensors for real time predictions of HAD T95.

Three main groups of predictive JITL models are constructed for HAD. In the first group, various subsets of variables, which are assumed to carry the highest information on variation of HAD T95, are included into static models and dynamic models. In the second group, seven process variables are selected and LS, PLS and subset regression via SR are employed on the predictor set. JITL models are evaluated with respect to reference data selection methods, reference set size, window size and neighborhood size. Here, it is shown that including previous HAD T95 measurement in the predictor vector increases prediction performance of JITL models significantly. Furthermore JITL model performances are found to vary with operational conditions. Bias and correlation of the JITL predictions with the laboratory measurements are found to be significantly different for two non-overlapping intervals of the historical data (see Section 6.2). Last but not the least, RMSE values of JITL model are decreased as neighborhood size is increased. The best predictive model of this group is found to be JITL SR model 1, using sliding window reference dataset (sWRD), and yielding RMSE and MAE values equal to 5.66 and 4.23 ^{o}C, respectively. In the third group, a different set of process variables with interaction and quadratic terms are included in the predictor set, and neighborhood variable set is separately treated. Performances of JITL models are highly increased. RMSE and MAE values of CRD JITL LS model 1 are found to be equal to 4.77 and 3.82 respectively. This shows that the predictor and neighborhood selection predictor subspace do not need to be identical, and a more convenient neighborhood selection predictor subspace may be chosen which yield a

smaller RMSE. To our knowledge, this is the first time in the literature, predictor and neighbor selection variable subsets are separated from each other; and the current study shows that using a high number of predictor variables in the regression model and a lesser number of variables in determining the neighbors to a query point is likely to increase the prediction accuracy.

Similarity criterion of all models in this thesis is distance based criterion so for future studies, angle based similarity criterion or hybrid form that is combination of distance and angle based criterion may be studied. In addition to that, although results considering tricube weighting function are not satisfactory enough, different weighting functions may give better results for different data set, or for data sets with a larger set size. Another future suggestion to increase JITL model prediction performance is elimination of outliers during the model construction step. Lastly, if robust and accurate estimations of hard-to-measure controlled variables can be done, JITL models can be also used in Model Predictive Control.

REFERENCES

1. Al-Dunainawi, Yousif and Maysam F. Abbod. 2016. “Hybrid Intelligent Approach for Predicting Product Composition of Distillation Column.”, *International Journal of Advanced Research in Artificial Intelligence* 5(4), pp. 28–34.

2. Atalay, By Özlem and Gürkan Kumbaroğlu. 2014. “A Regional Demand Forecasting Study for Transportation Fuels in Turkey.”, *Third Quarter*, pp.50-51.

3. Bontempi, Gianluca, Mauro Birattari, and Hugues Bersini. 2001. “Lazy Learning : a Local Method for Supervised Learning Introduction.”, *In New Learning Paradigms in Soft Computing*, pp. *97–137.*

4. Chai, T. and R. R. Draxler. 2014. “Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)-Arguments against Avoiding RMSE in the Literature.”, *Geoscientific Model Development* 7(3), pp. 1247–12550.

5. Chang, A., Pashikanti, K., and Liu, Y. A. 2012. “Refinery engineering: Integrated process modeling and optimization ”, *Weinheim: Wiley-VCH.*

6. Chatterjee, Tirtha and Deoki N. Saraf. 2004. “On-Line Estimation of Product Properties for Crude Distillation Units.” , *Journal of Process Control* 14(1), pp. 61–77.

7. Chen, K., I. Castillo, L. H. Chiang, and J. Yu. 2015. “Soft Sensor Model Maintenance: A Case Study in Industrial Processes.” , *IFAC Proceedings Volumes (IFAC-PapersOnline)* 48(8): pp. 427–32.

8. Cheng, Cheng and Min Sen Chiu. 2004. “A New Data-Based Methodology for Nonlinear Process Modeling.”, *Chemical Engineering Science* 59(13), pp. 2801–2810.

9. Cheng, Cheng, Yoshihiro Hashimoto, and Min-sen Chiu. 2004. “An Enhanced Just-in-Time Learning Methodology for Process Modeling.”, *Environmental Engineering* (1), pp. 1–6.

10. Chiang, Leo H., Randy J. Pell, and Mary Beth Seasholtz. 2003. “Exploring Process Data with the Use of Robust Outlier Detection Algorithms.”, *Journal of Process Control* 13(5), pp. 437–49.

11. Czop, P., G. Kost, D. Stawik, and G. Wszotek. 2011. “Formulation and Identification of First- Principle Data-Driven Models.” , *Journal of Achievements in Materials and Manufacturing Engineering* 44(2), pp.179–86., http://www.journalamme.org/papers_vol44_2/4427.pdf., accessed at January 2017.

12. Englert, P., 2012. “Locally Weighted Learning.” *Seminar Class on Autonomous Learning Systems* (1), pp. 1–9. http://www.ias.informatik.tudarmstadt.de., accessed at January 2017.

13. Fahim, M. A., Alsahhaf, T. A., and Elkilani, A. S. 2010., “Fundamentals of petroleum refining”. *Elsevier*, chapter 2.

14. Ferreira, Luciane S. and Jorge O. Trierweiler. 2009. “Modeling and Simulation of the Polymeric Nanocapsule Formation Process.”, *IFAC Proceedings Volumes (IFAC-PapersOnline)* 7. pp. 405–10.

15. Fortuna, L., Grazini, S., Risso, A., and Xibilia, M. G. 2007. “Soft Sensors for Monitoring and Control of an Industrial Processes. ” , *Springer*, pp. 1-10.

16. Fortuna, L., S. Graziani, and M. G. Xibilia. 2005. “Soft Sensors for Product Quality Monitoring in Debutanizer Distillation Columns.” , *Control Engineering Practice* 13(4), pp. 499–508.

17. Fujiwara, K., Kano, M., Hasebe, S., and Takinami, A. 2009. “Soft-sensor development using correlation-based just-in-time modeling. ” , *AIChE Journal AIChE J.*, 55(7), pp. 1754-1765.

18. Gary, J. H., Handwerk, G. E., and Kaises, M. J. 2007. “Petroleum Refining, Technology and Economics.”, *CRC Press*, pp. 32-37.

19. Ge, Zhiqiang and Zhihuan Song. 2010. “A Comparative Study of Just-in-Time-Learning Based Methods for Online Soft Sensor Modeling.” , *Chemometrics and Intelligent Laboratory Systems* 104(2), pp. 306–17.

20. Golden, Scott W. 2009. “Maximising Diesel Recovery from Crude Important to Increasing Diesel Yield Are Discussed in Detail.”, *Petroleum Technology Quarterly*,pp. 60-65

21. Huang, Dong, Cabral, and Fernando Torre. 2012. “Robust Regression.

22. International Energy Agency. 2014. “ENERGY SUPPLY SECURITY 2014 Part 3.” , *Energy Supply Security: The Emergency Response of IEA Countries - 2014 Edition*, pp. 1–105.

23. Ito, M. *et al*. 2004. “Large Scale Database Online Modeling for Blast Furnace.” , *Proceedings of the 2004 IEEE International Conference on Control Applications*, pp. 906–11.

24. Kadlec, Petr, Gabrys, and Strandt. 2009. “Data-Driven Soft Sensors in the Process Industry.” , *Computers and Chemical Engineering* 33(4), pp. 795–814.

25. Kadlec. (2009). “On robust and adaptive soft sensors.” , MS Dissertation, School of Design, Engineering & Computing Bournemouth University.

26. Kadlec, Petr, Ratko Grbić, and Bogdan Gabrys. 2011. “Review of Adaptation Mechanisms for Data-Driven Soft Sensors.”, *Computers and Chemical Engineering* 35(1), pp. 1–24.

27. Kano, Manabu and Koichi Fujiwara. 2013. “Virtual Sensing Technology in Process Industries: Trends and Challenges Revealed by Recent Industrial Applications.” , *Journal of Chemical Engineering of Japan* 46(1), pp. 1–17.

28. Kosanovich, K. A. and M. J. Piovoso. n.d. “Process Data Analysis Using Multivariate Statistical Methods.” *E. i. Du Pont De Nemours and Company inc.*, pp. 721–24.

29. Lin, Bao, Bodil Recke, Jørgen K. H. Knudsen, and Sten Bay Jørgensen. 2007. “A Systematic Approach for Soft Sensor Development.” *Computers and Chemical Engineering* 31(5–6), pp. 419–25.

30. Macias-Hernandez, J. J., Plamen Angelov, and Xiaowei Zhou. 2007. “Soft Sensor for Predicting Crude Oil Distillation Side Streams Using Takagi Sugeno Evolving Fuzzy Models.” 44(1524), *IEEE*, pp. 3305–10.

31. Nakabayashi, Akio et al. 2010. “A Process Simulator Based on Hybrid Model of Physical Model and Just-In-Time Model.” *Proceedings of the SICE Annual Conference*, pp. 1497–1501

32. Okada, T., Kaneko, H., and Funatsu, K., Original. 2012. “Development of a Model Selection Method Based on the Reliability of a Soft Sensor Model.”, *Songklanakarin Journal of Science and Technology*, 34(2), pp. 217–221.

33. Park, S. and C. Han. 2000. “Anonlinear Soft Sensor Based on Multivariate Smoothing Procedure for Quality Estimation in Distillation Columns.”, *Computers and Chemical Engineering* 24, pp. 871–77.

34. Sanguansat P. 2012. “Principal Component Analysis-Engineering Applications.”, *InTech,* chapter 1 .

35. Saptoro, Agus. 2014. “State of the Art in the Development of Adaptive Soft Sensors Based on Just-in-Time Models.” , *Procedia Chemistry* 9, pp. 226–34.

36. Schaal, Stefan and Christopher G. Atkeson. 2002. “Scalable Techniques from Nonparametric Statistics.” , *Applied Intelligence* 16(1), pp. 49–60.

37. Slikovic, Drazen, Ratko Grbic, and Emmanuel K. Nyarko. 2009. “Data Preprocessing in Data Based Process Modeling.”, *IFAC Proceedings Volumes (IFAC-PapersOnline)* 2, pp. 559–65.

38. Sliškovic, Dražen, Ratko Grbic, and Željko Hocenski. 2011. “Methods for Plant Data-Based Process Modeling in Soft-Sensor Development.”, *Automatika* 52(4), pp. 306–18.

39. Sloley, Andrew W. 2014. “Atmospheric Distillation Process: Fundamental Concepts.”, *Fuels and Petrochemicals Division 2014 - Core Programming Area at the 2014 AIChE Spring Meeting and 10th Global Congress on Process Safety* 1, pp. 59–63.

40. The Economics of Petroleum Refining-Understanding the business of processing crude oil into fuels and other value added products. 2013. *Canadian Fuels Association*., http://www.canadianfuels.ca/, accessed at January 2017.

41. Turkey’s refining industry: Competition set to heat up for an increasingly lucrative market. 2015., http://www.bcct.org.tr/, accessed at January 2017.

APPENDIX A: TRAJECTORIES OF HISTORICAL DATASET OF SIGNIFICANT COLUMN PARAMETERS

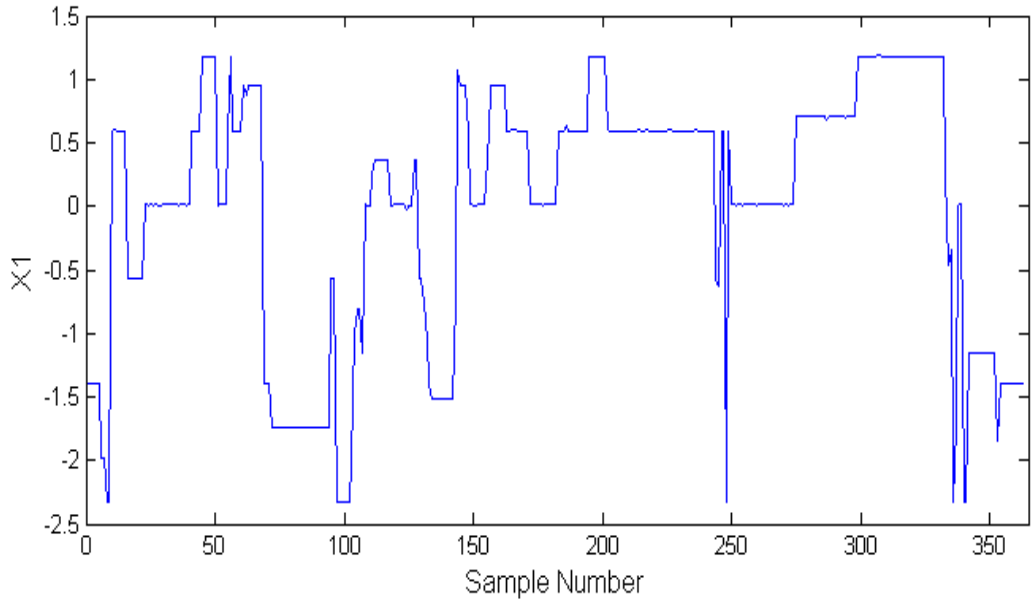

Figure A.1.Trajectories of historical dataset of X1

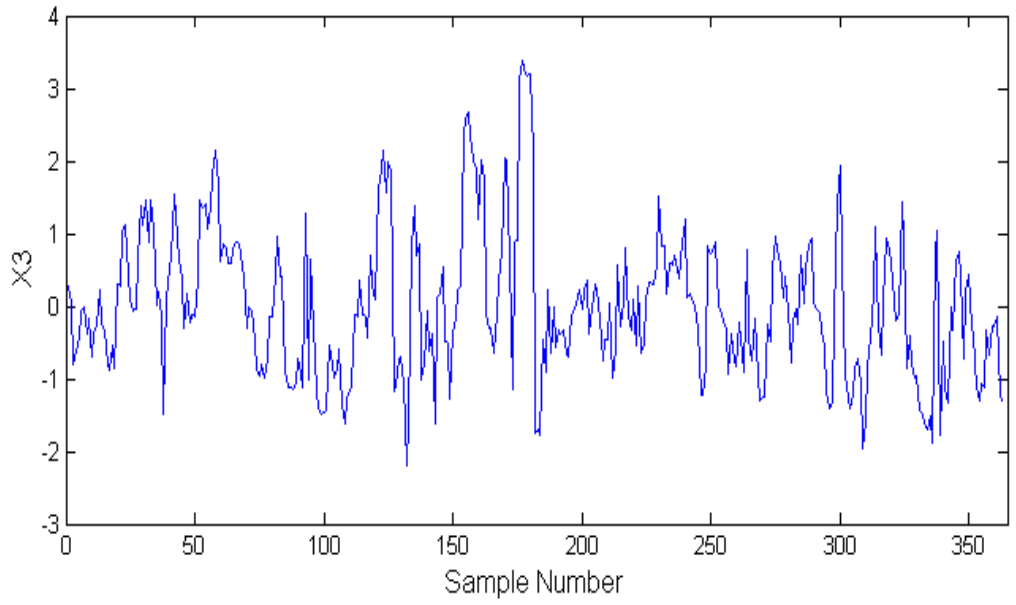

Figure A.2.Trajectories of historical dataset of X3

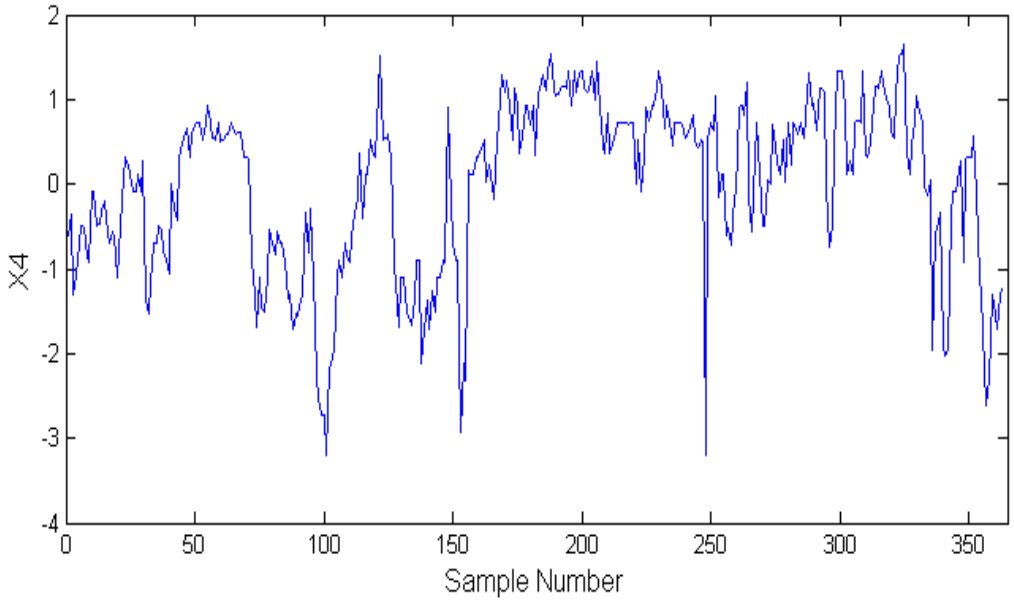

Figure A.3.Trajectories of historical dataset of X4

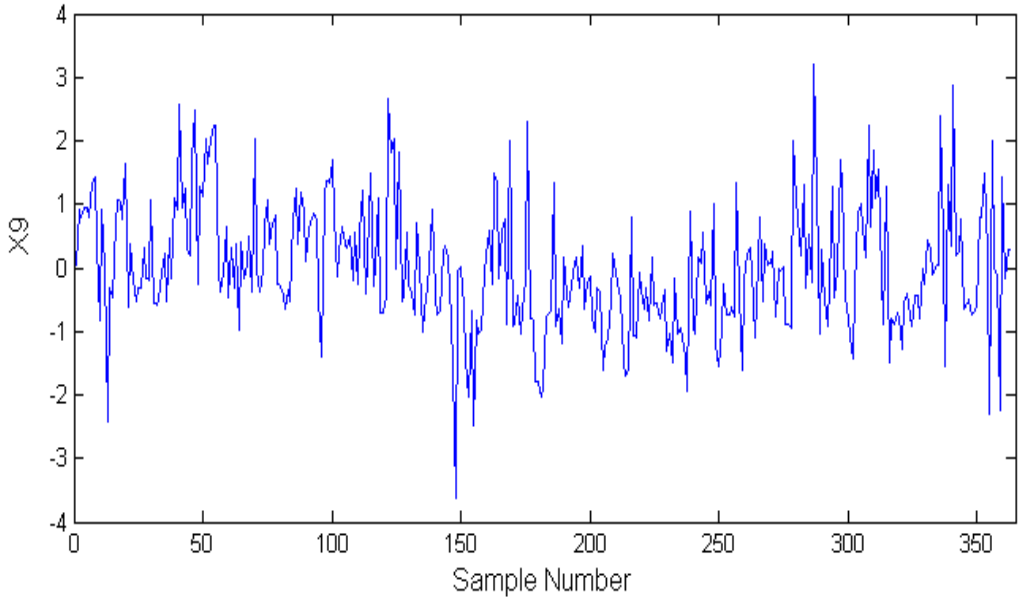

Figure A.4.Trajectories of historical dataset of X9

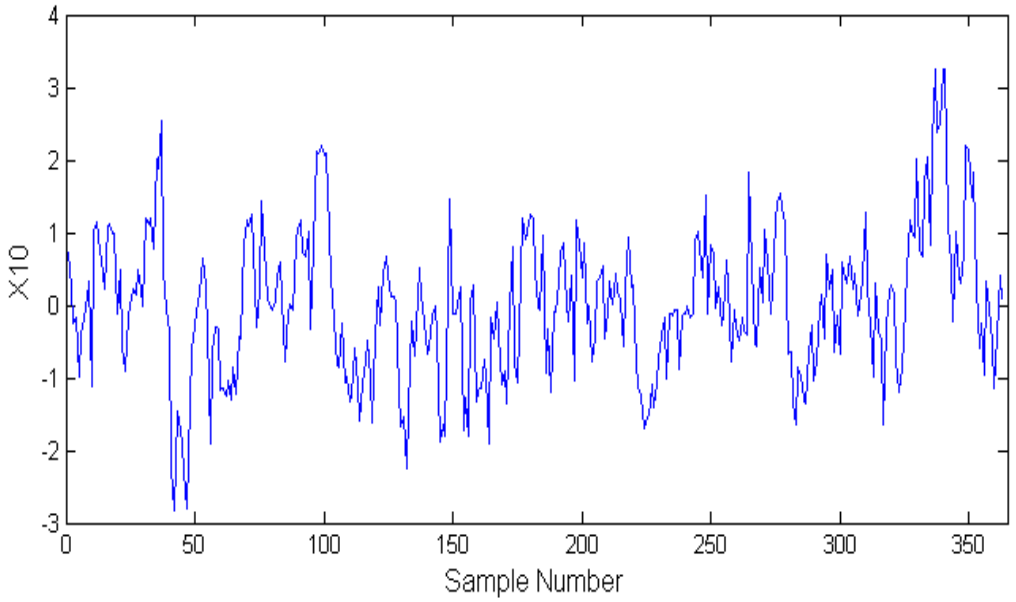

Figure A.5.Trajectories of historical dataset of X10

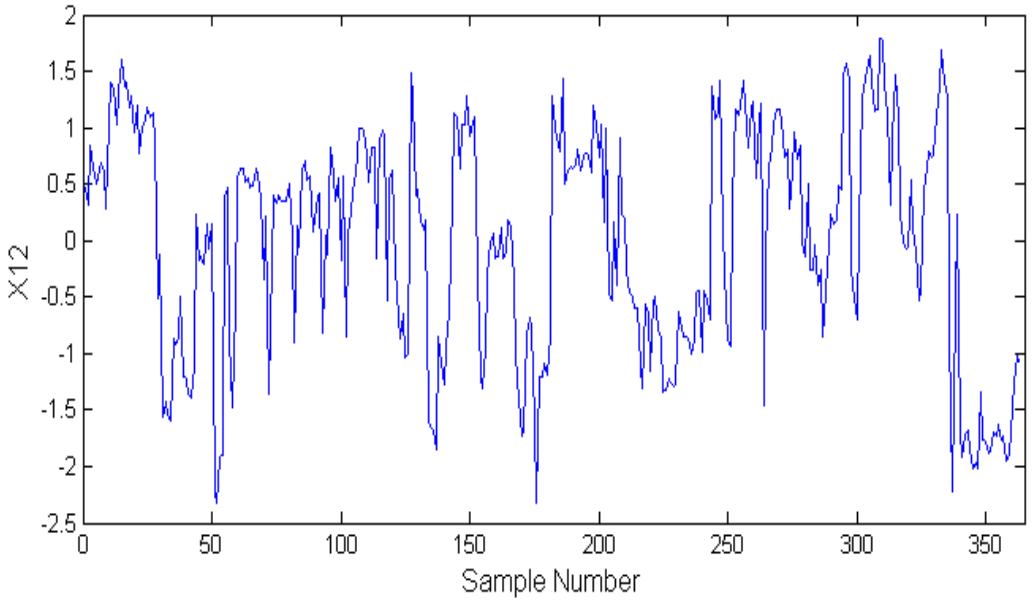

Figure A.6.Trajectories of historical dataset of X12

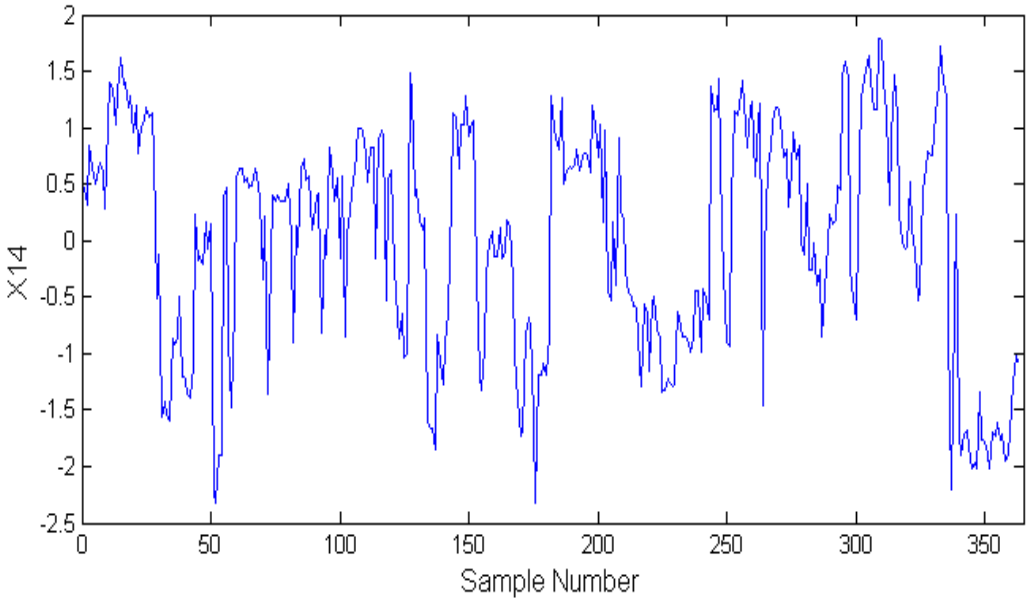

Figure A.7.Trajectories of historical dataset of X14

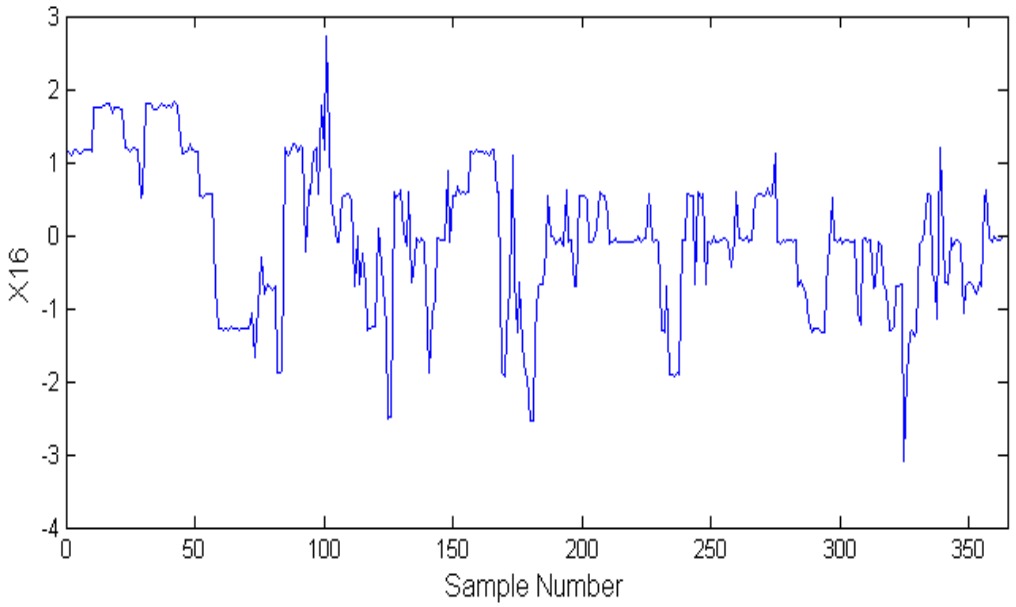

Figure A.8.Trajectories of historical dataset of X16

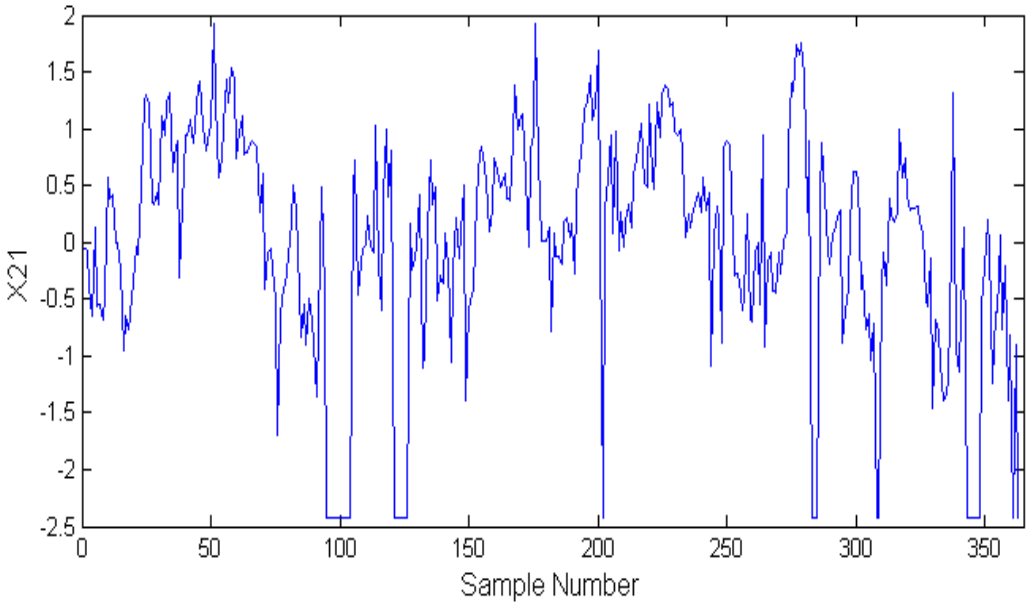

Figure A.9.Trajectories of historical dataset of X21

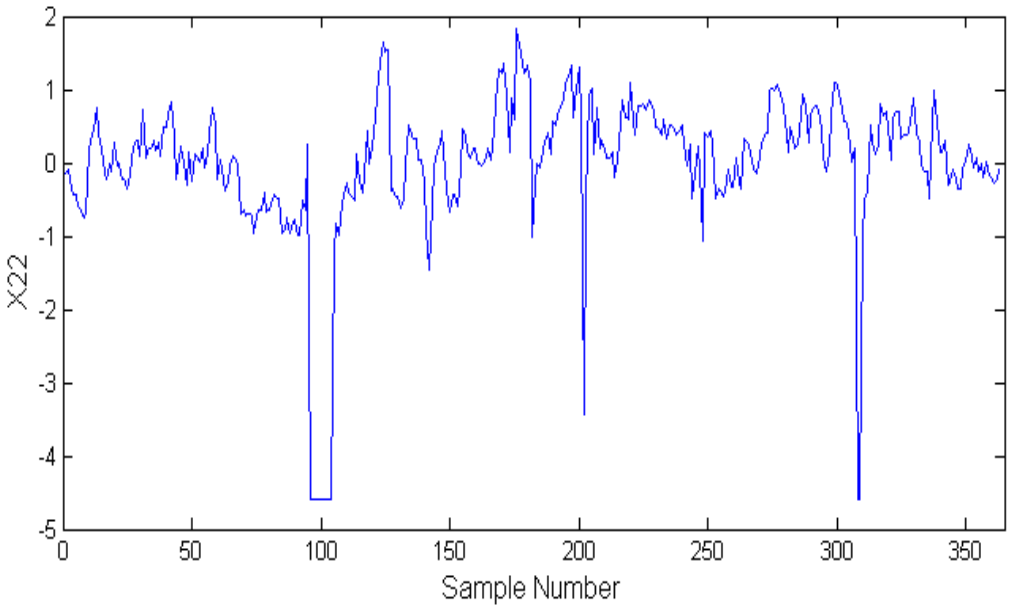

Figure A.10.Trajectories of historical dataset of X22

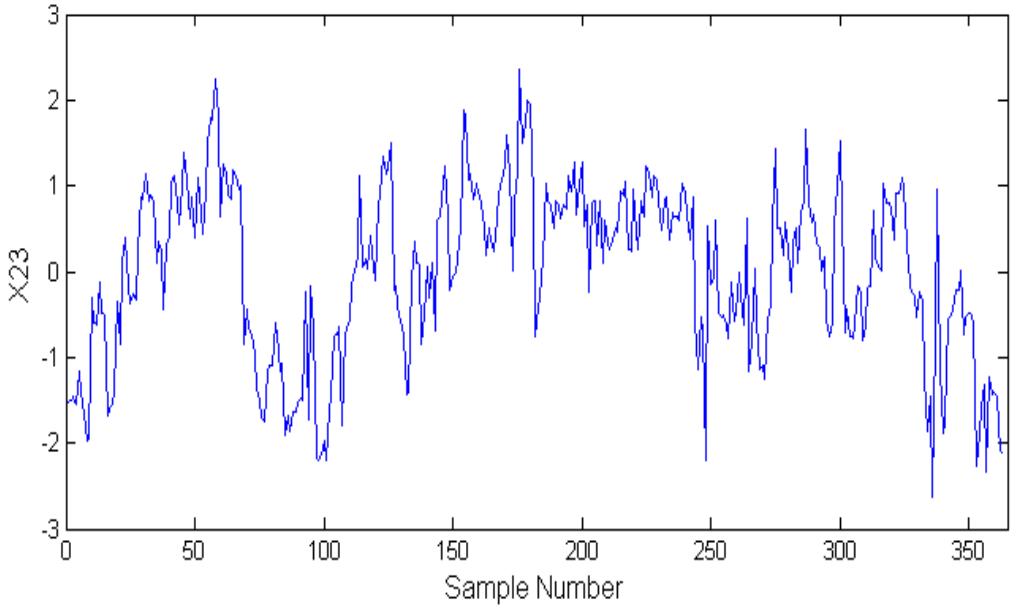

Figure A.11.Trajectories of historical dataset of X23

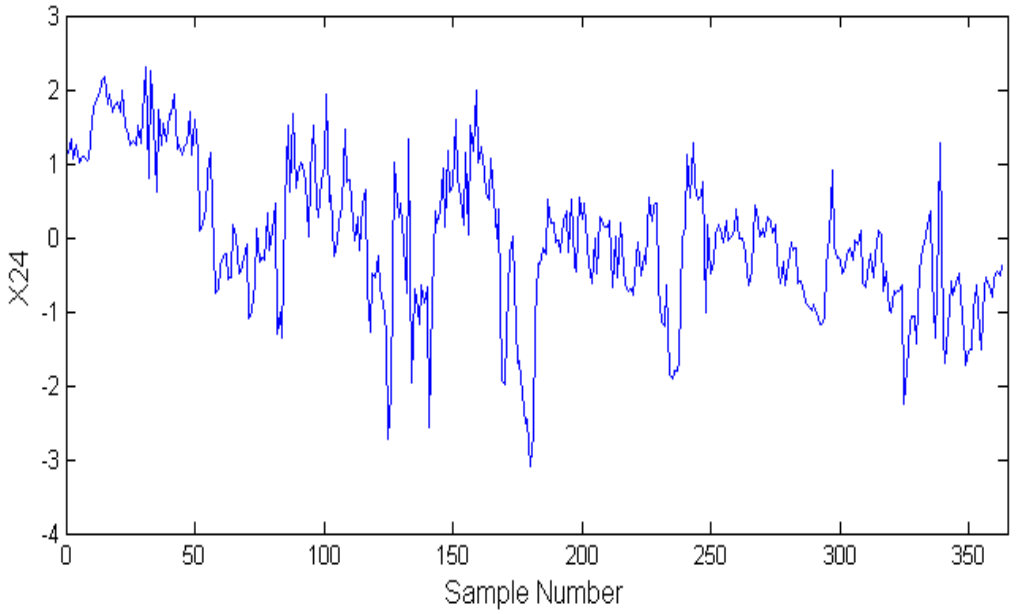

Figure A.12.Trajectories of historical dataset of X24

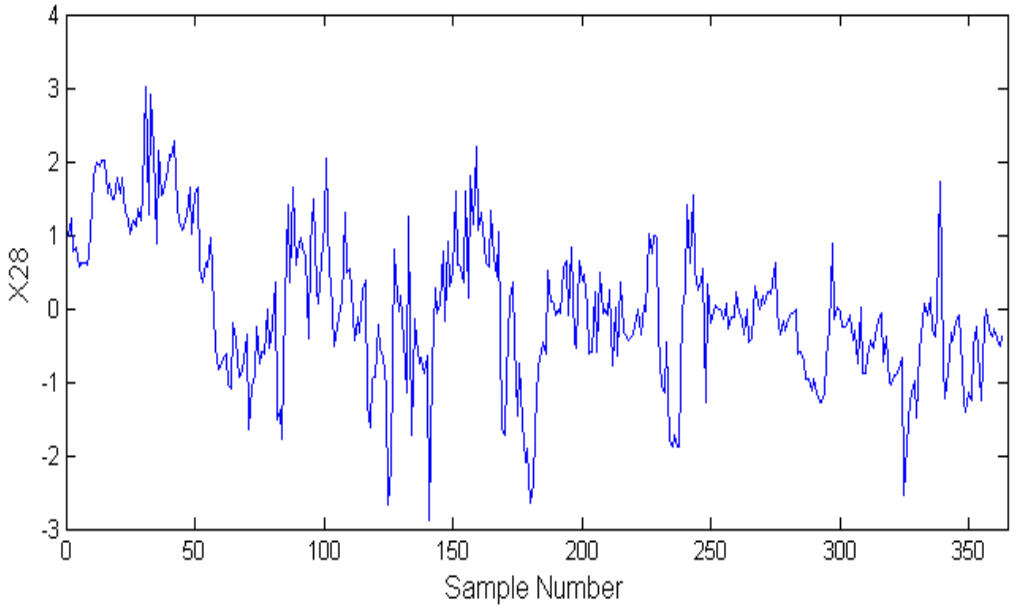

Figure A.13.Trajectories of historical dataset of X28

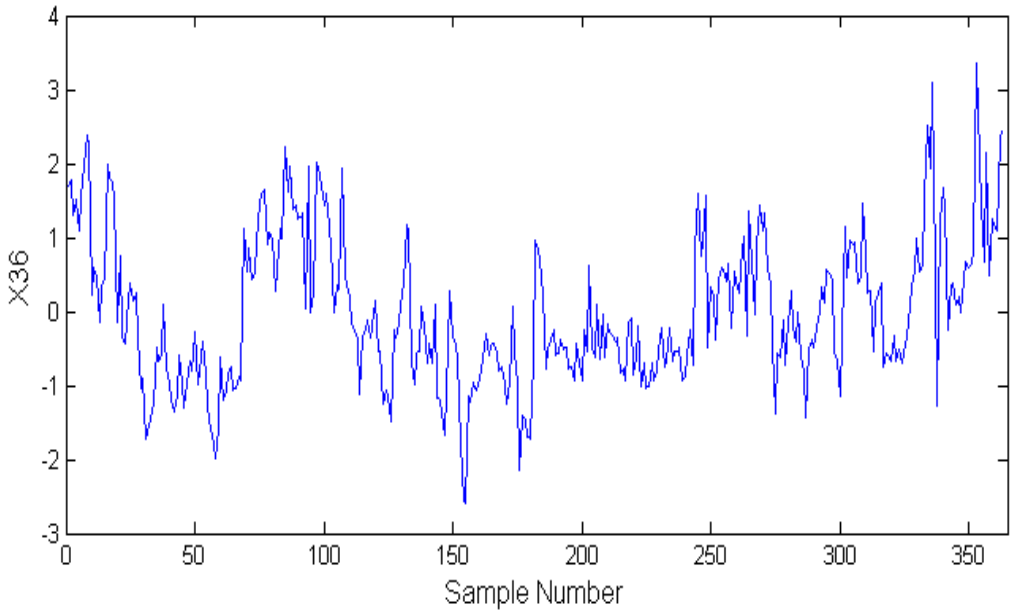

Figure A.14.Trajectories of historical dataset of X36

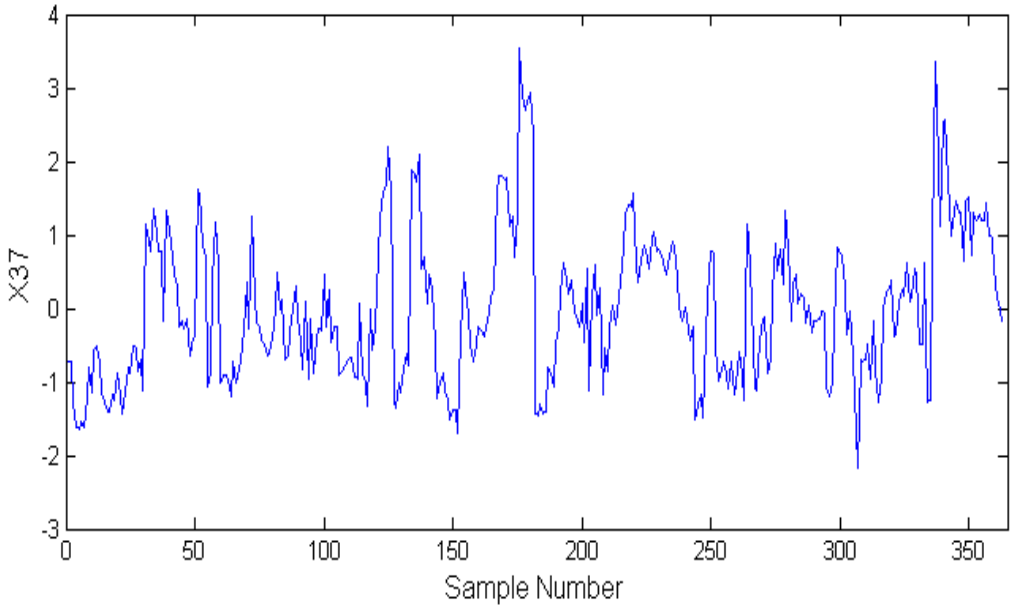

Figure A.15.Trajectories of historical dataset of X37

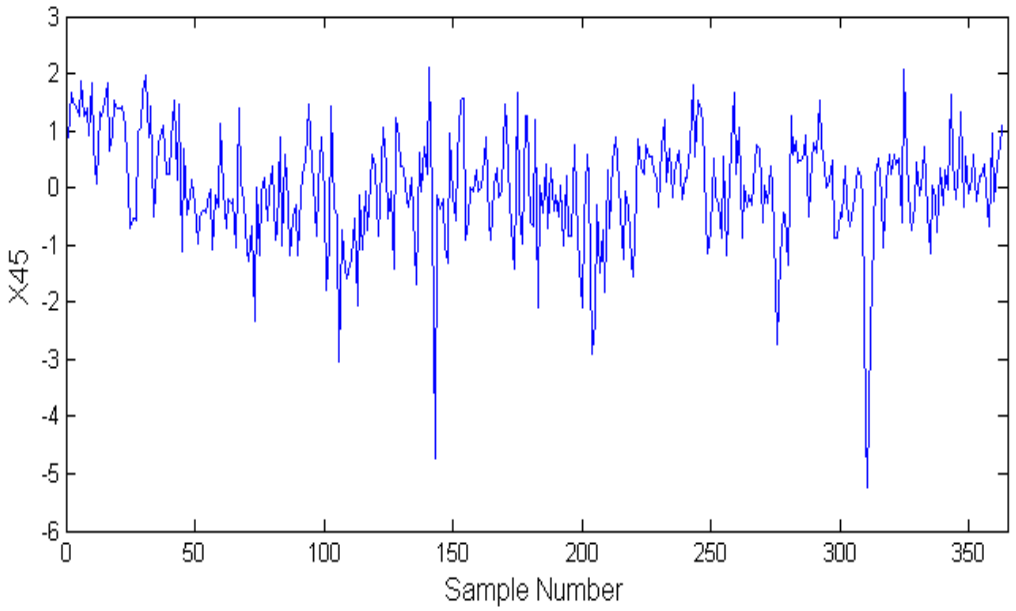

Figure A.16.Trajectories of historical dataset of X45

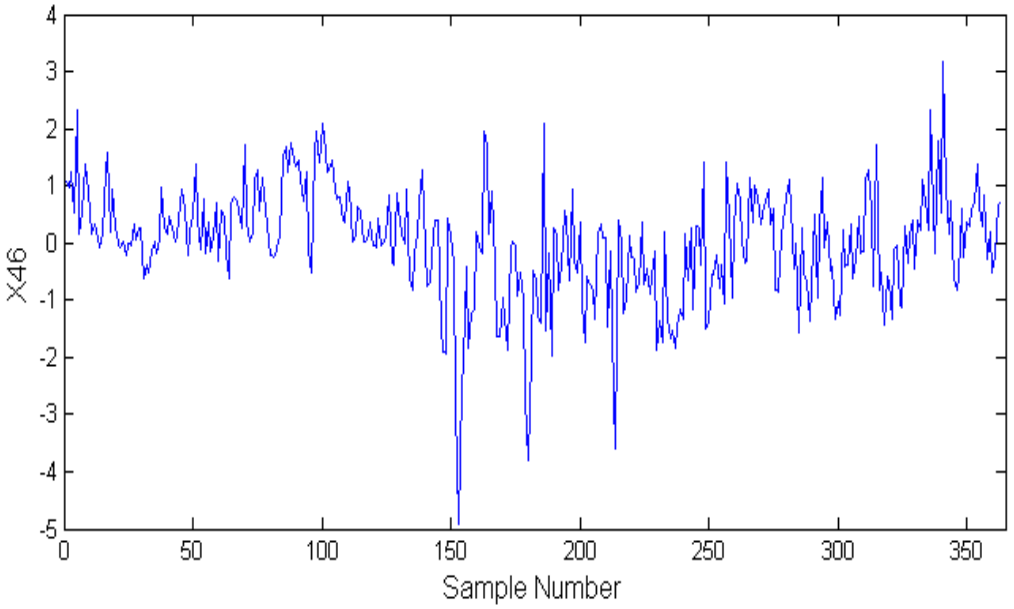

Figure A.17.Trajectories of historical dataset of X46

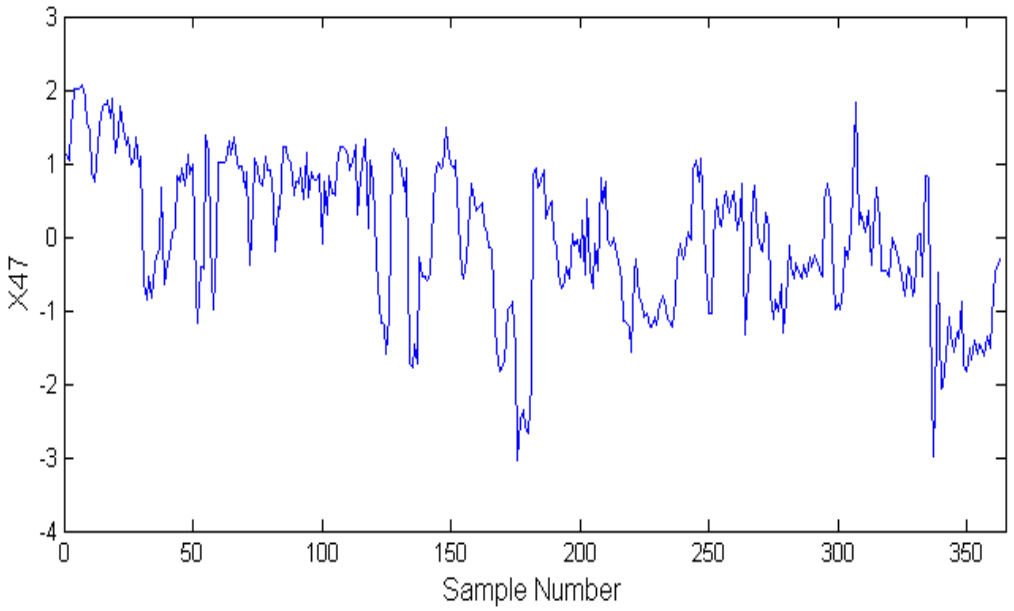

Figure A.18.Trajectories of historical dataset of X47

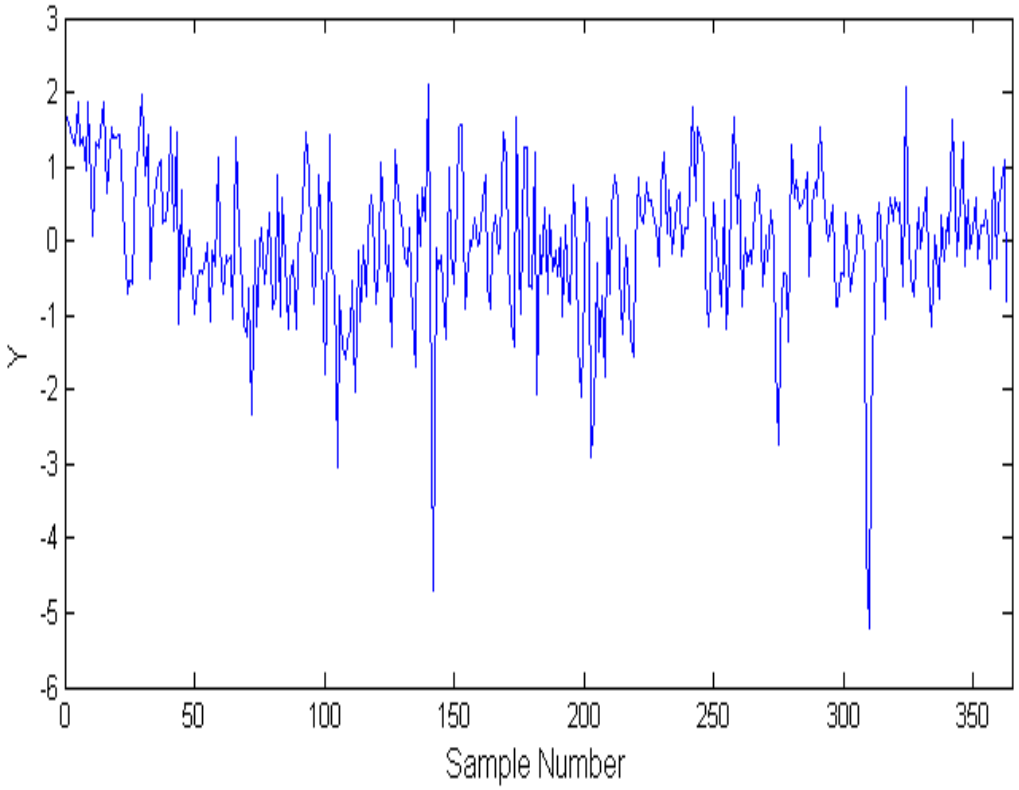

Figure A.19. Trajectories of historical dataset of Y

APPENDIX B: PREDICTION TRAJECTORIES OF STATIC MODELS

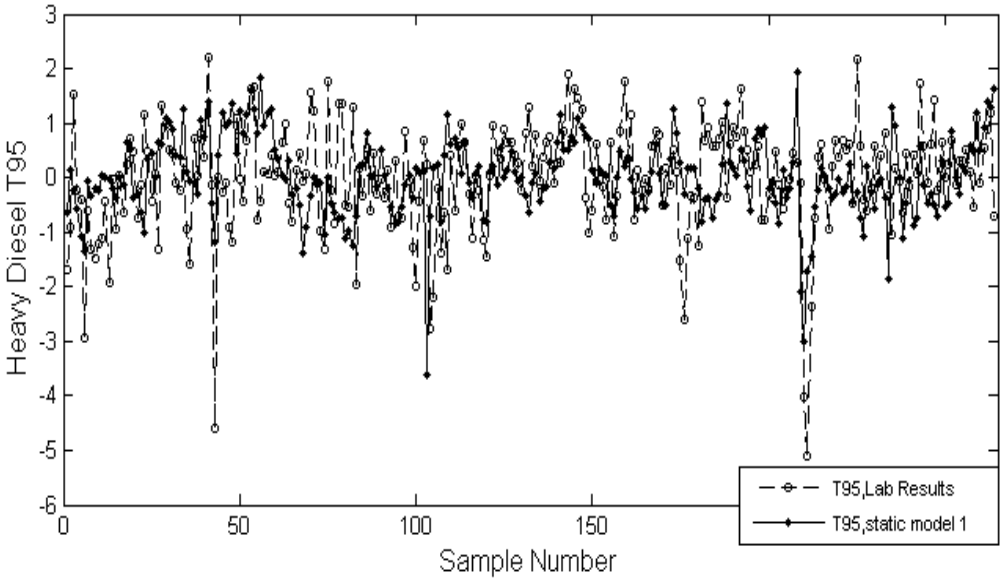

Figure B.1. HAD T95 predictions vs. sample number of static model 1

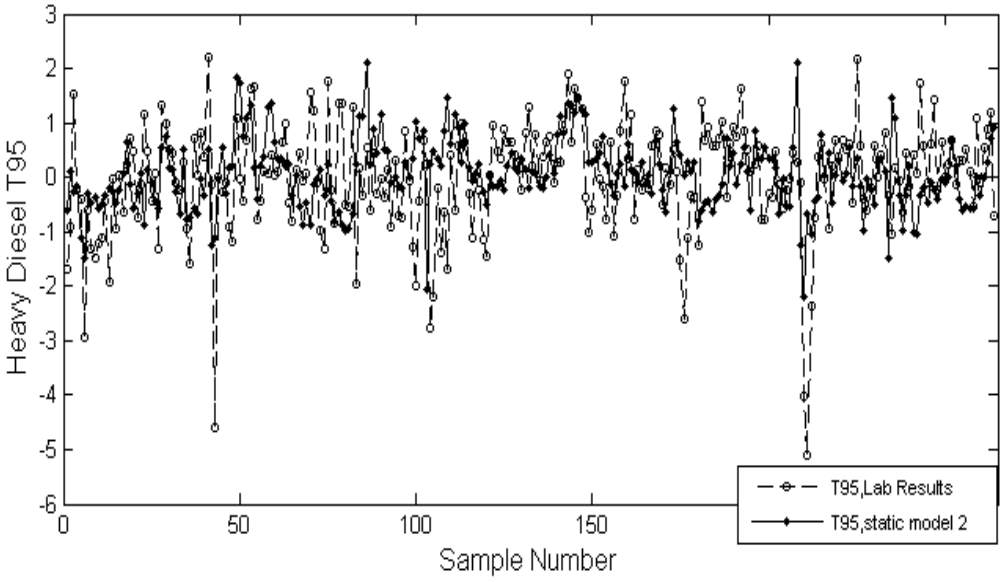

Figure B.2. HAD T95 predictions vs. sample number of static model 2

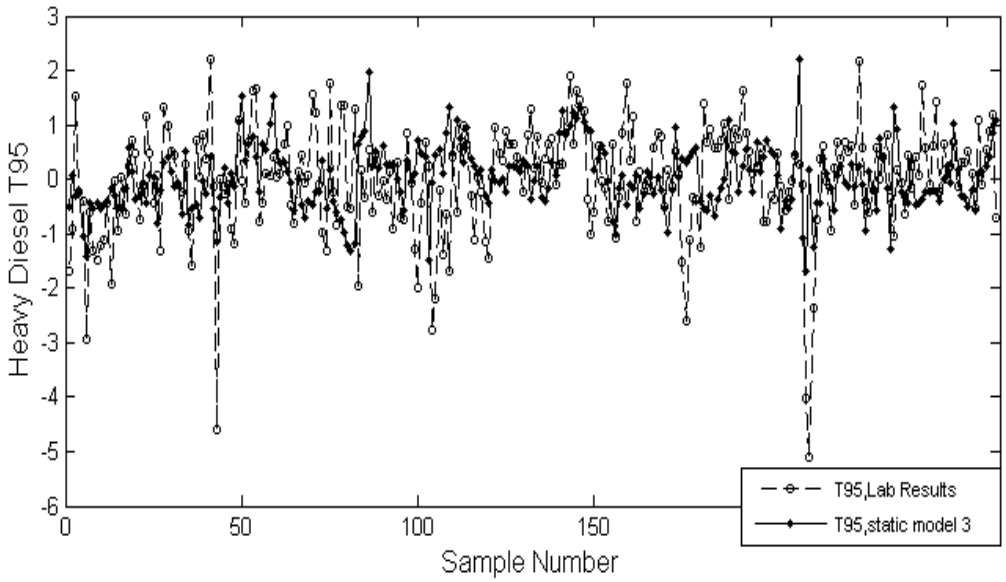

Figure B.3. HAD T95 predictions vs. sample number of static model 3

APPENDIX C: PREDICTION TRAJECTORIES OF DYNAMIC MODELS VARIABLES SELECTED BY STEPWISE REGRESSION

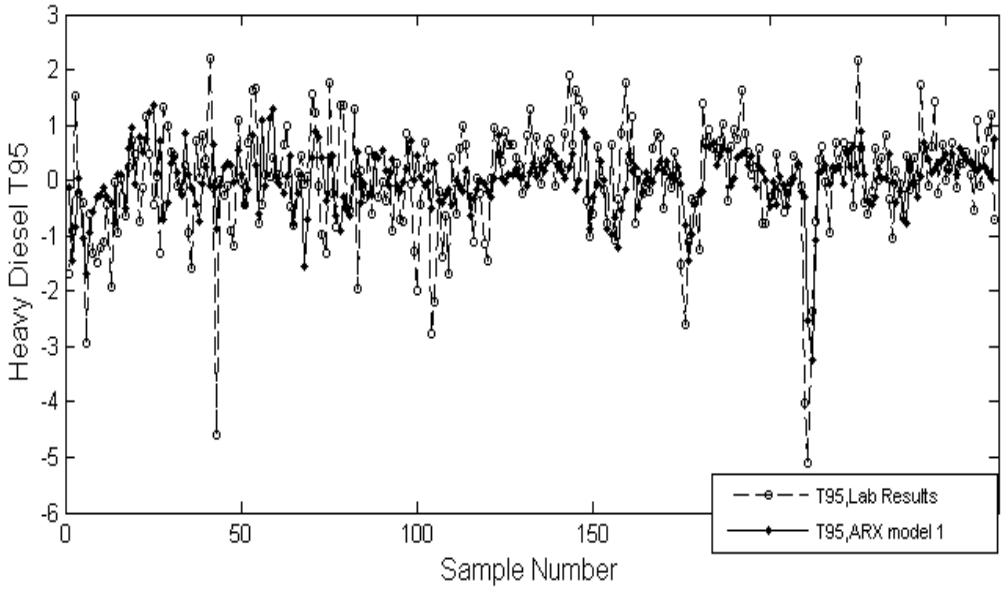

Figure C.1. HAD T95 predictions vs. sample number of ARX model 1

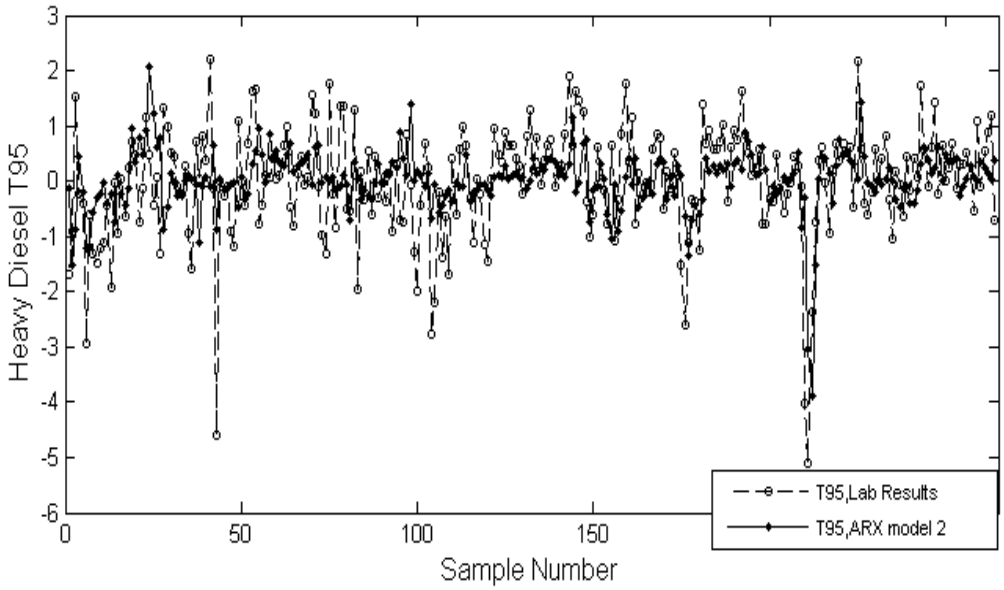

Figure C.2. HAD T95 predictions vs. sample number of ARX model 2

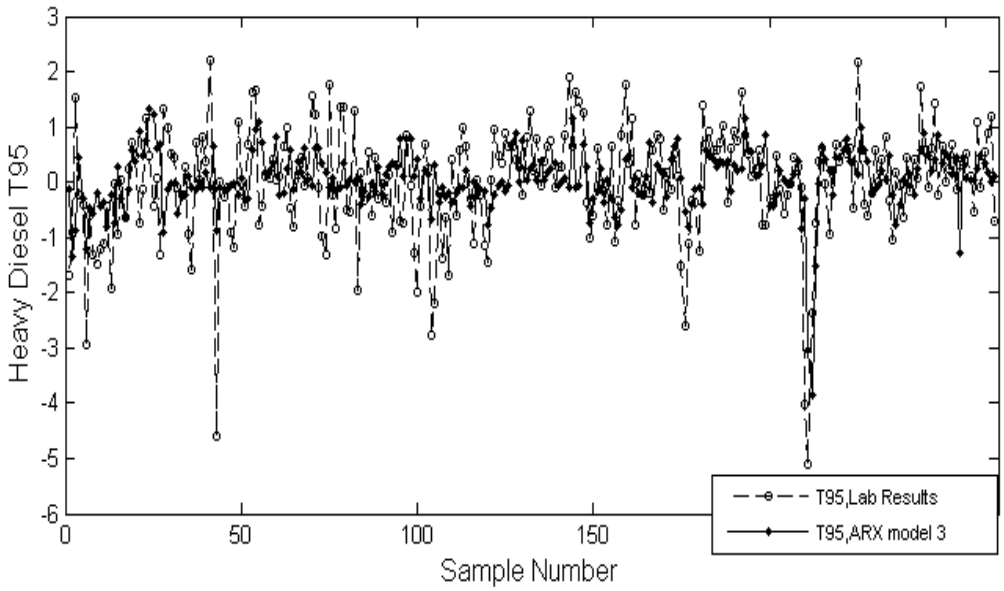

Figure C.3. HAD T95 predictions vs. sample number of ARX model 3

APPENDIX D: GLOBAL LS PREDICTION TRAJECTORIES

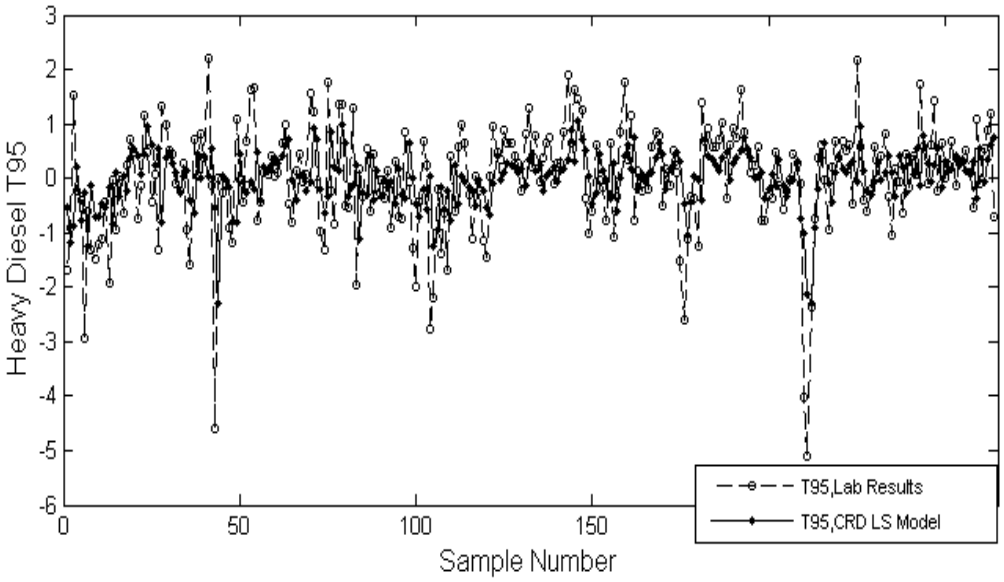

Figure D.1. HAD T95 predictions vs. sample number of CRD global LS model

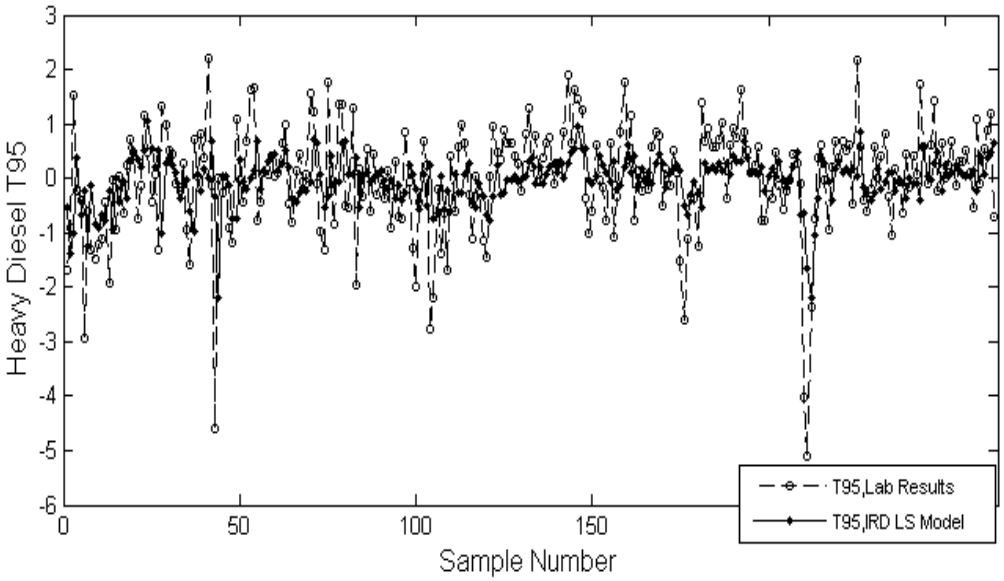

Figure D.2. HAD T95 predictions vs. sample number of IRD global LS model

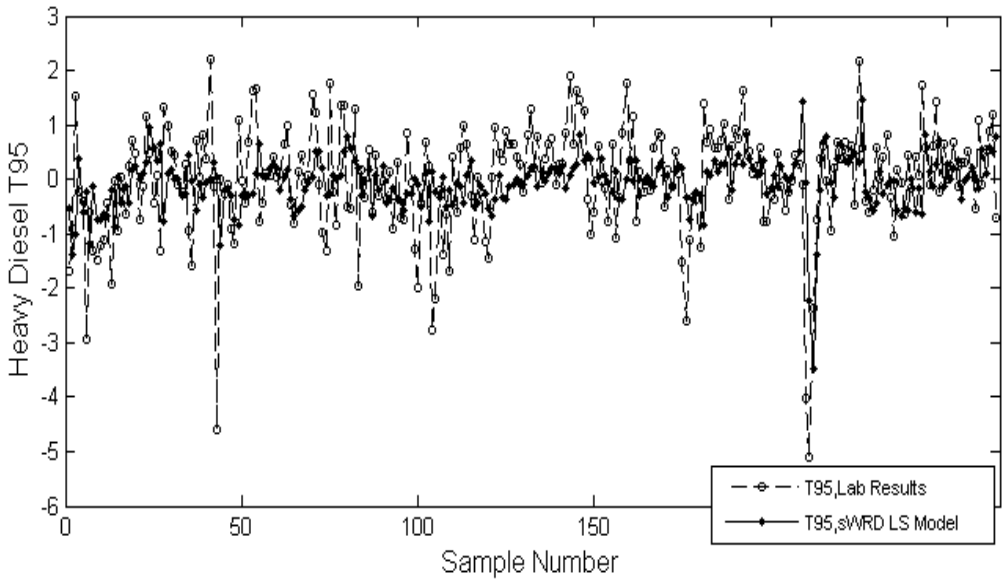

Figure D.3. HAD T95 predictions vs. sample number of sWRD global LS model

APPENDIX E: GLOBAL PLS PREDICTION TRAJECTORIES

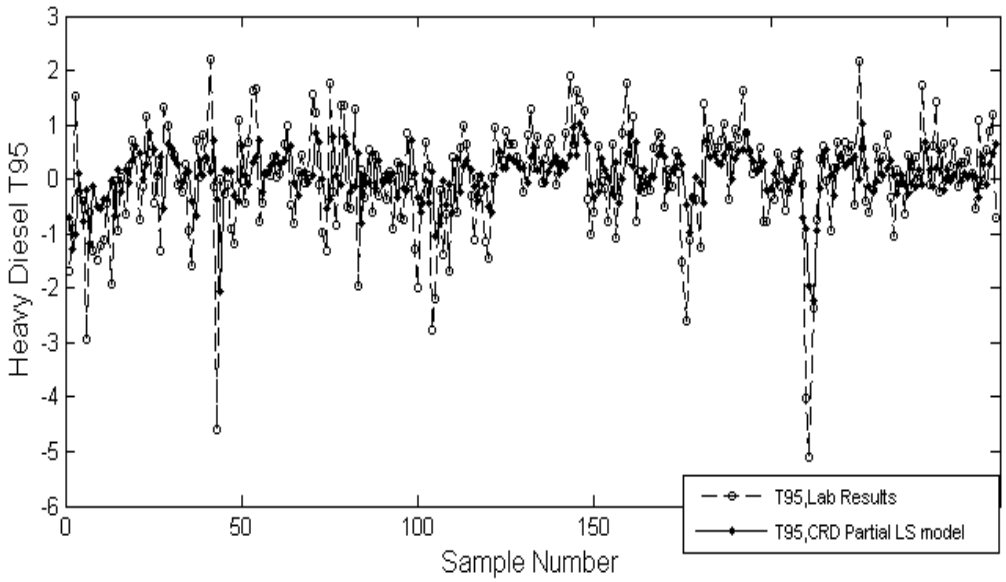

Figure E.1. HAD T95 predictions vs. sample number of CRD global PLS constant model

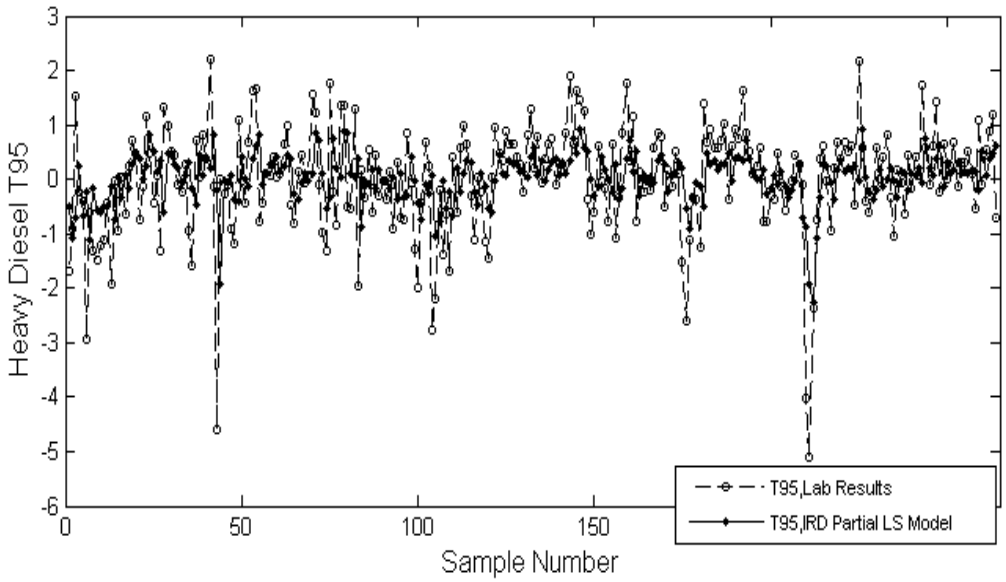

Figure E.2. HAD T95 predictions vs. sample number of IRD global PLS model

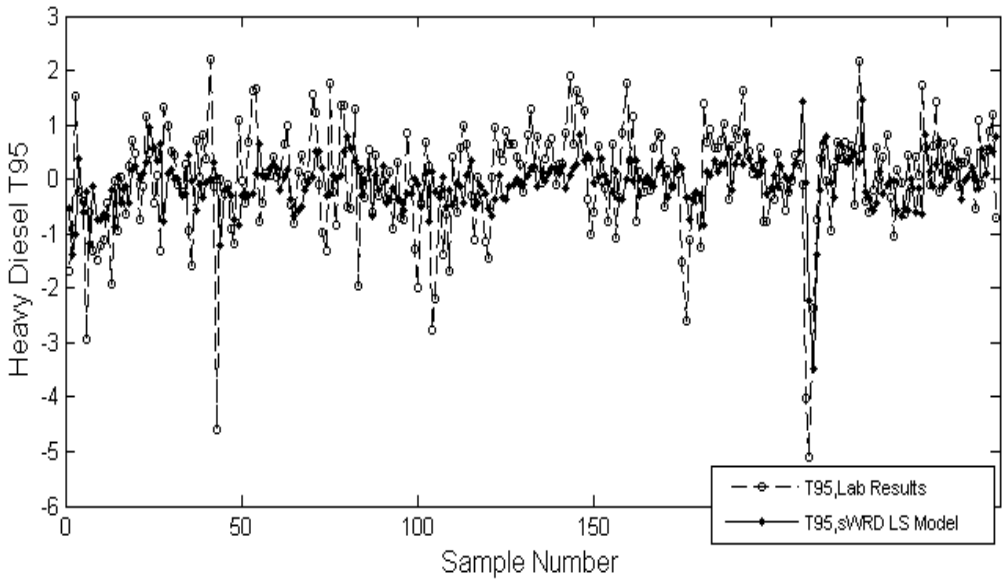

Figure E.3. HAD T95 predictions vs. sample number of sWRD global PLS model

APPENDIX F: JITL LS PREDICTION TRAJECTORIES

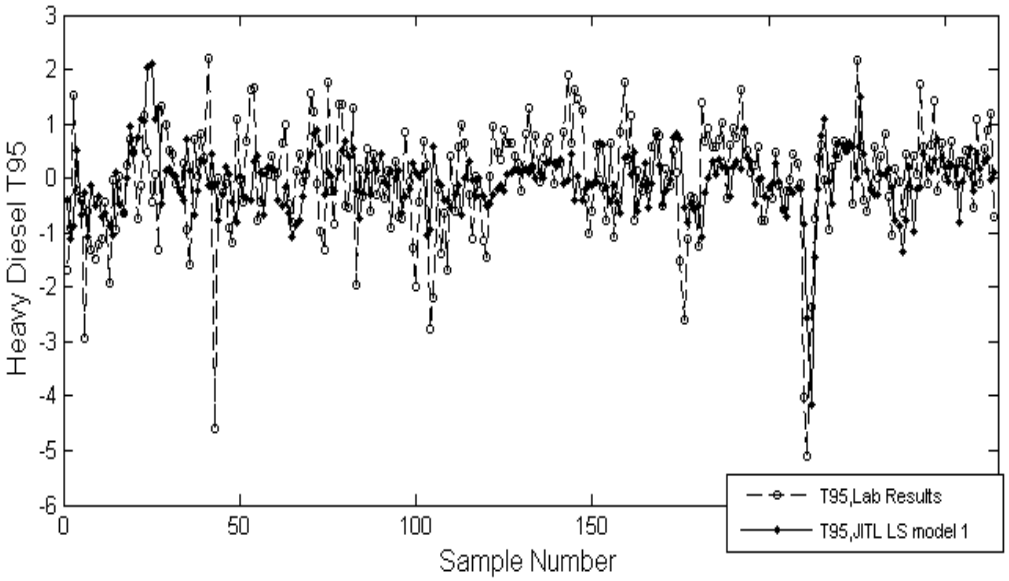

Figure F.1. HAD T95 vs. sample number of JITL LS analysis model 1

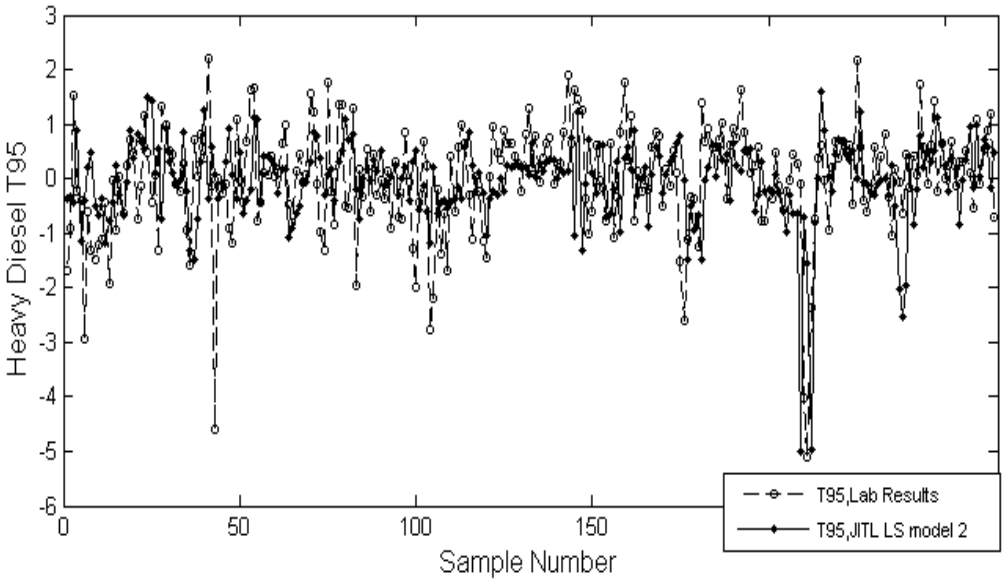

Figure F.2. HAD T95 vs. sample number of JITL LS analysis model 2

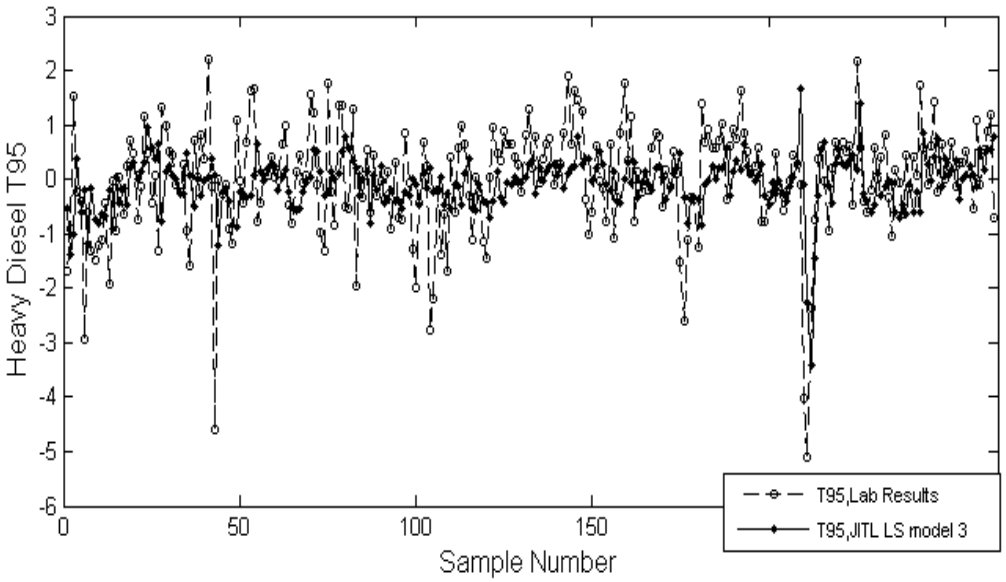

Figure F.3. HAD T95 vs. sample number of JITL LS analysis model 3

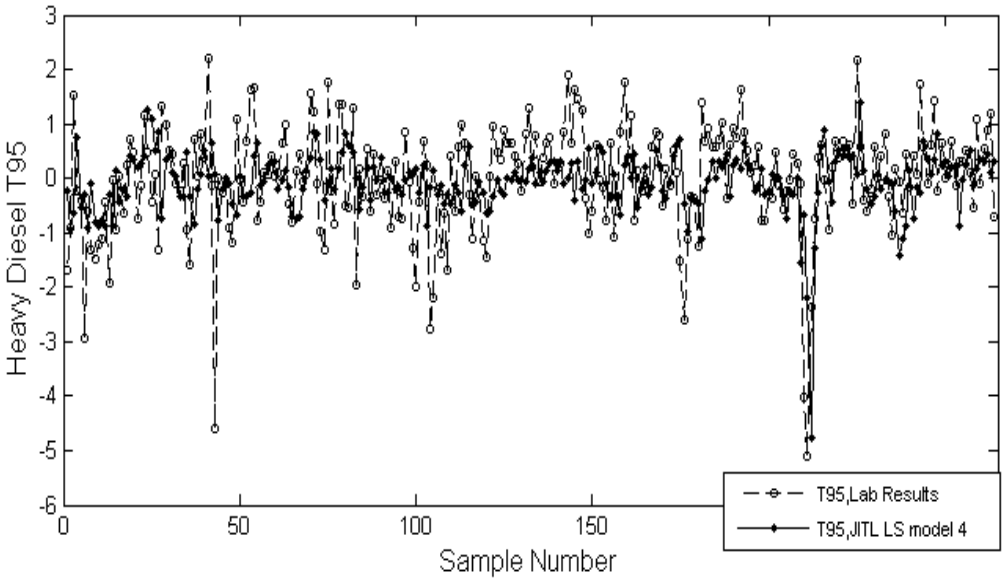

Figure F.4. HAD T95 vs. sample number of JITL LS analysis model

APPENDIX G: JITL PLS PREDICTION TRAJECTORIES

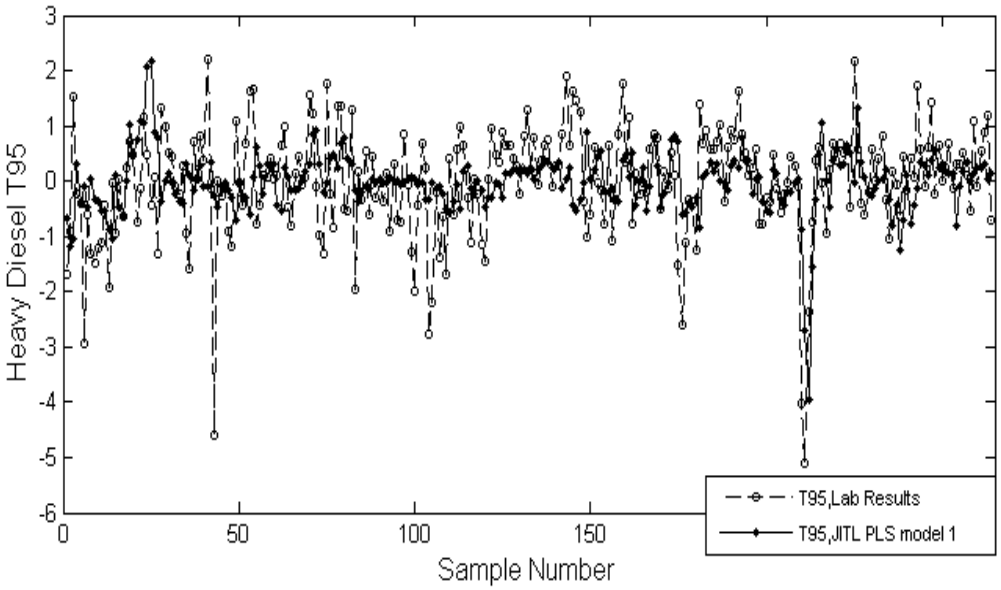

Figure G.1. HAD T95 vs. sample number of JITL PLS model 1

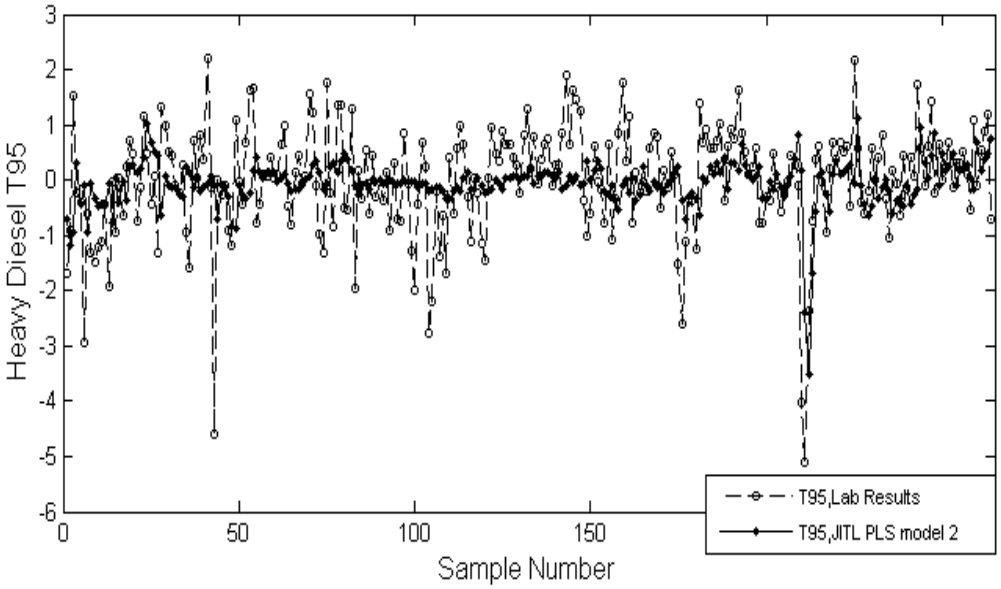

Figure G.2. HAD T95 vs. sample number of JITL PLS model 2

APPENDIX H: JITL SR PREDICTION TRAJECTORIES

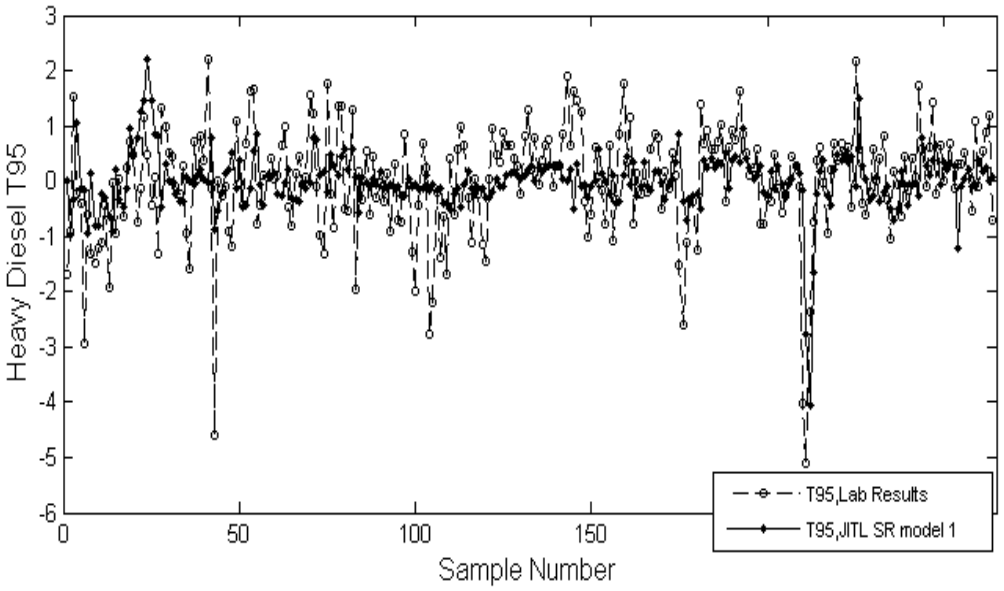

Figure H.1. HAD T95 vs. sample number of JITL SR model 1

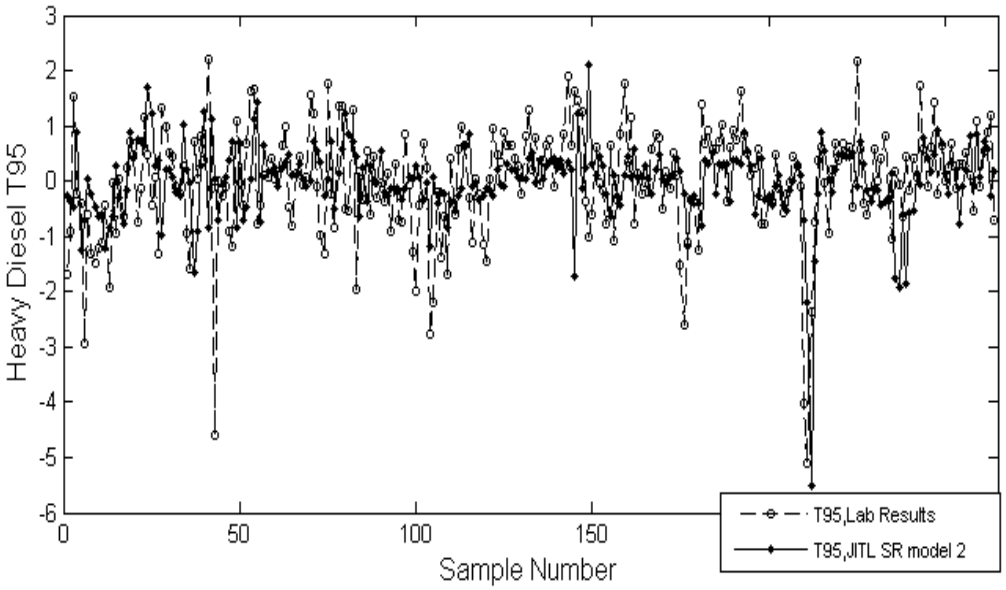

Figure H.2. HAD T95 vs. sample number of JITL SR model 2

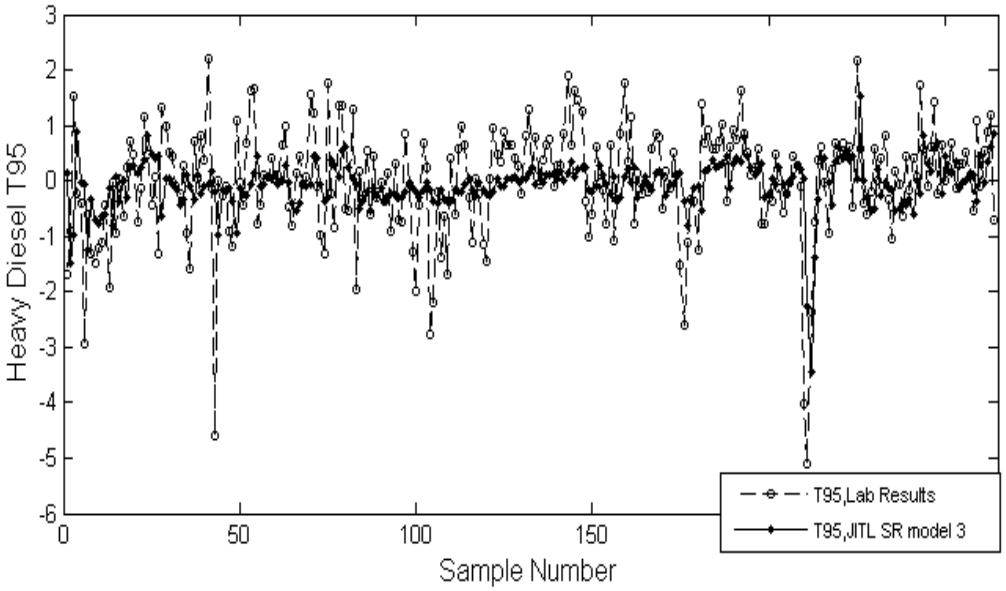

Figure H.3. HAD T95 vs. sample number of JITL SR model 3

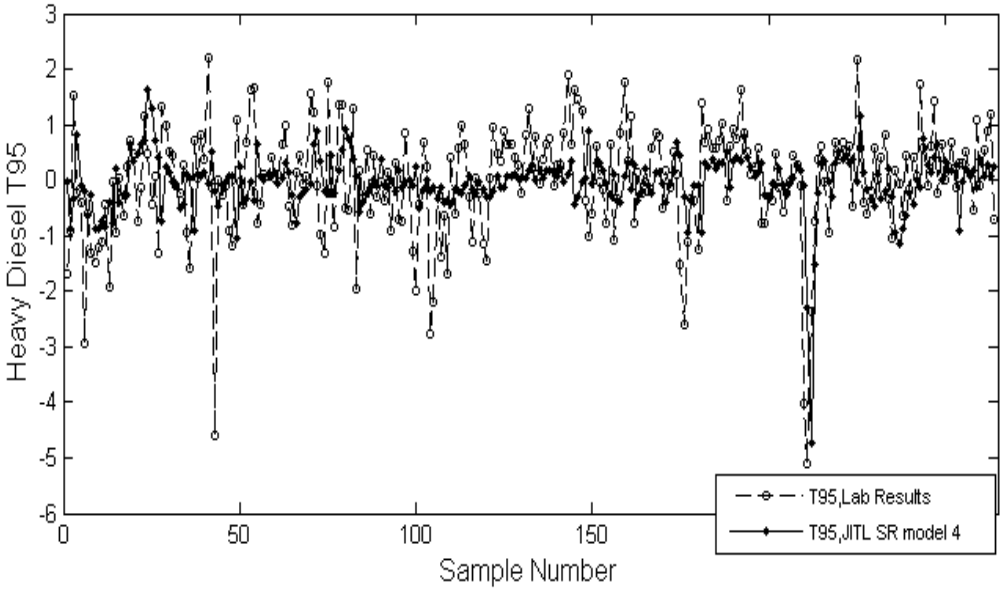

Figure H.4. HAD T95 vs. sample number of JITL SR model 4

APPENDIX I: GLOBAL LS PREDICTION TRAJECTORIES FOR INCREASED REFERENCE SET SIZE

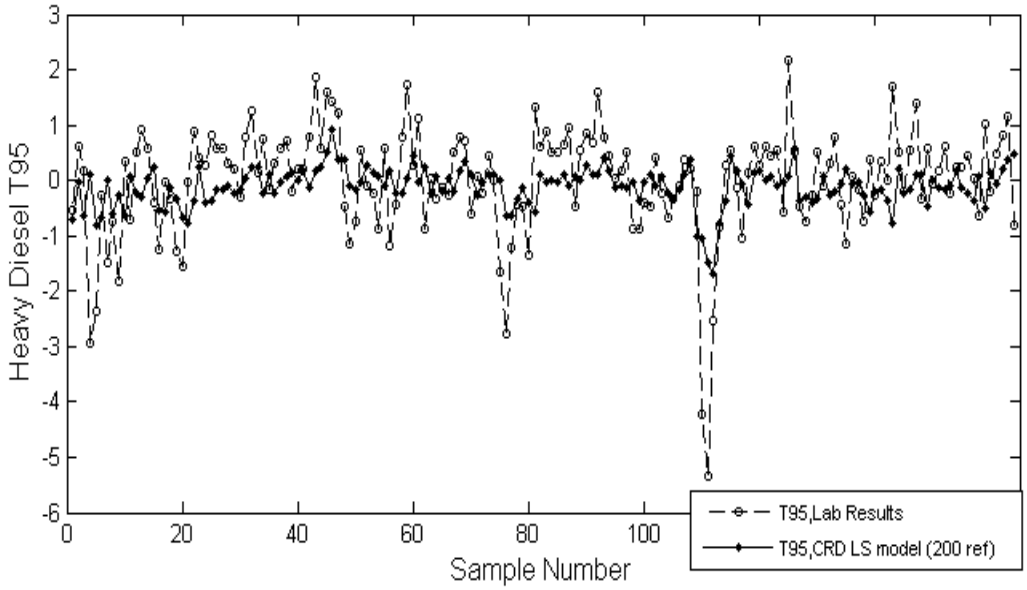

Figure I.1. HAD T95 vs. sample number of CRD global LS model for 200 reference set size

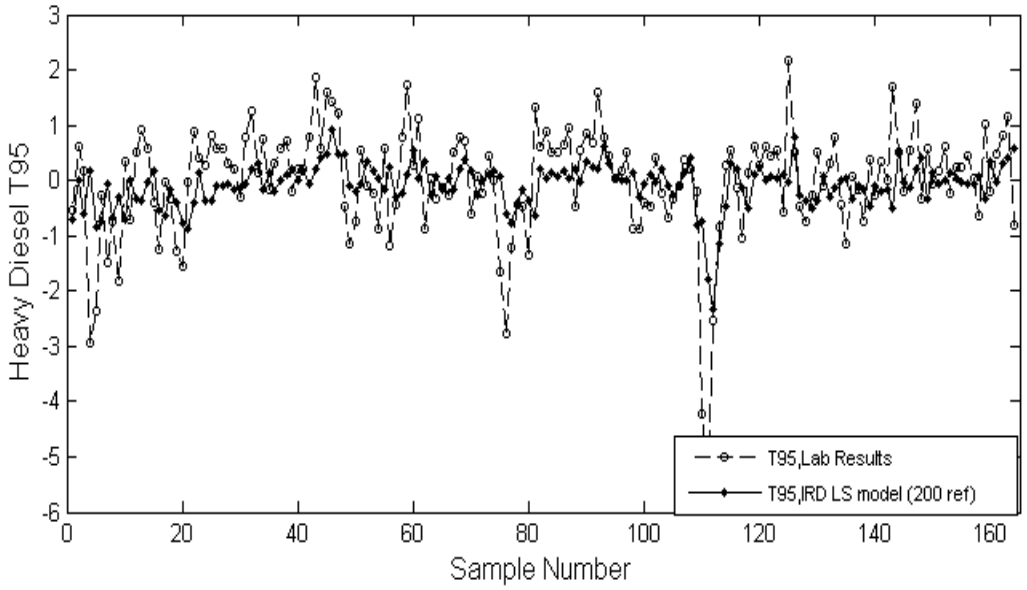

Figure I.2. HAD T95 vs. sample number of IRD global LS model for 200 reference set size

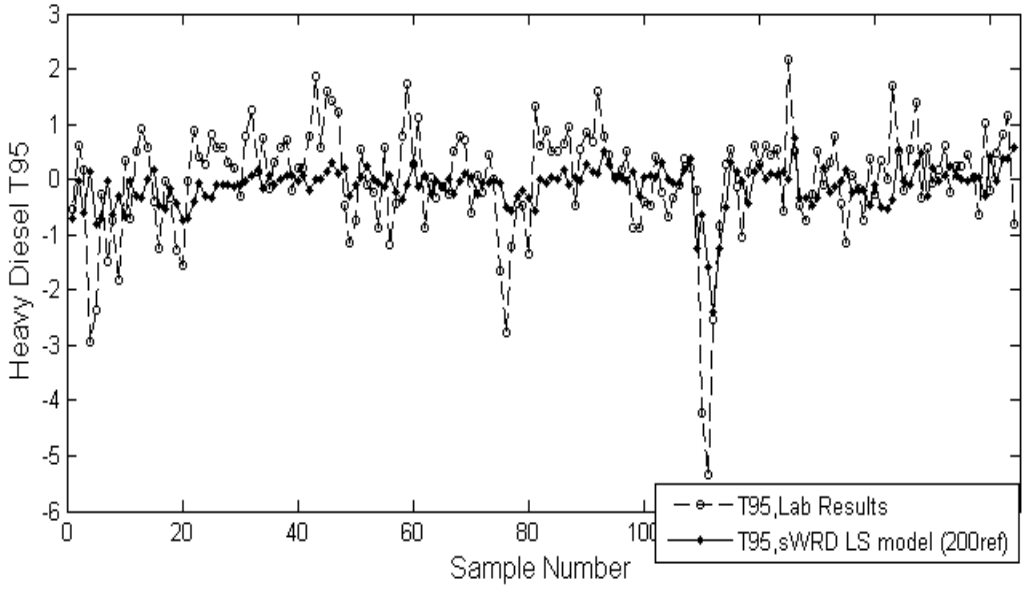

Figure I.3. HAD T95 vs. sample number of sWRD global LS recursive 2 model for 200 reference set size

APPENDIX J: GLOBAL PLS PREDICTION TRAJECTORIES FOR INCREASED REFERENCE SET SIZE

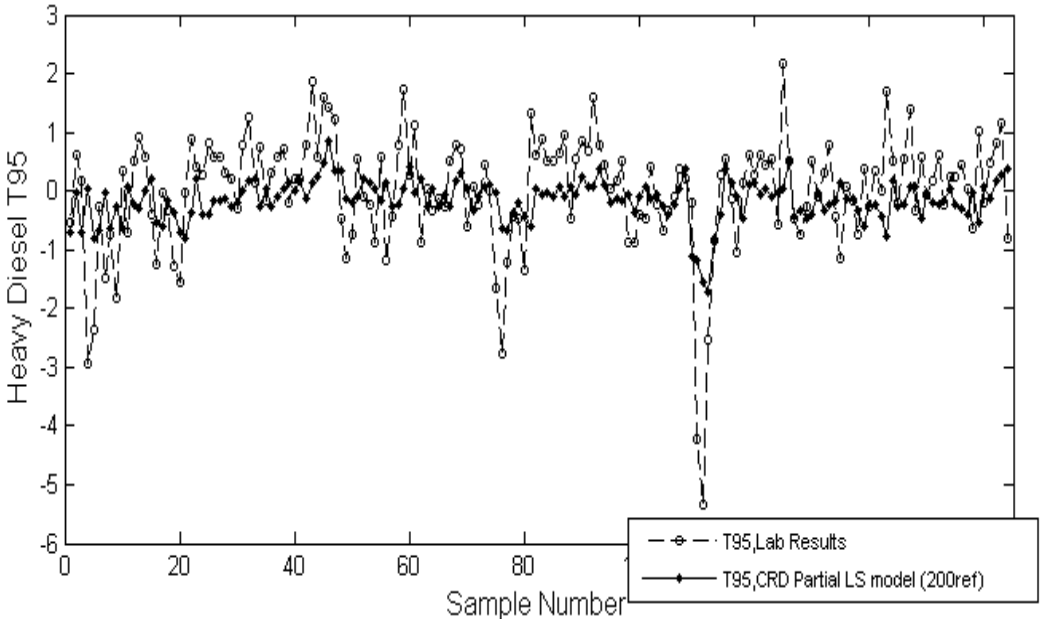

Figure J.1. HAD T95 vs. sample number of CRD global PLS model for 200 reference set size

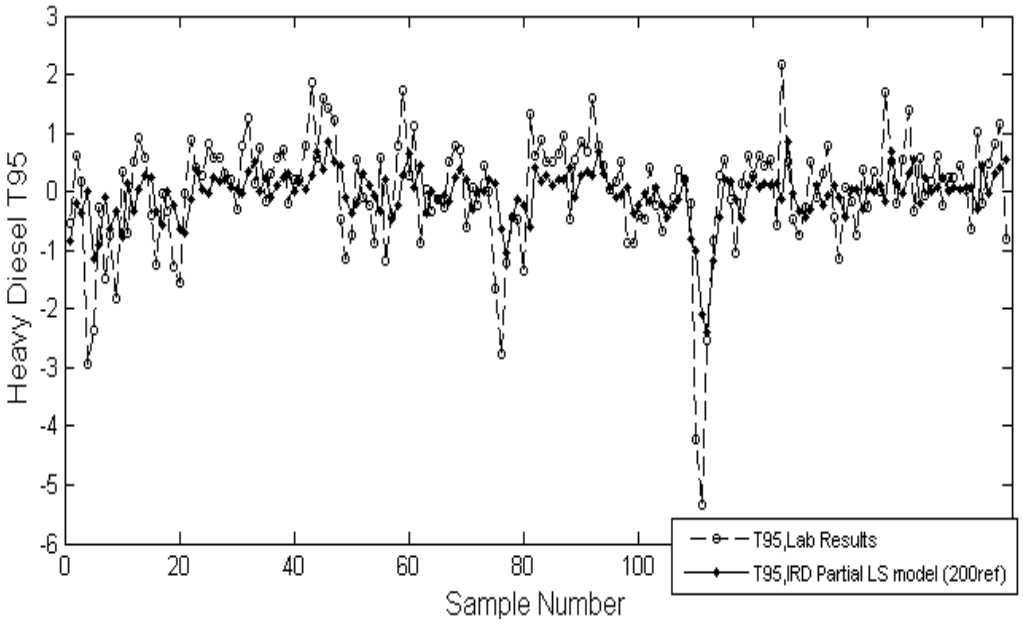

Figure J.2. HAD T95 vs. sample number of IRD global PLS recursive 1 model for 200 reference set size

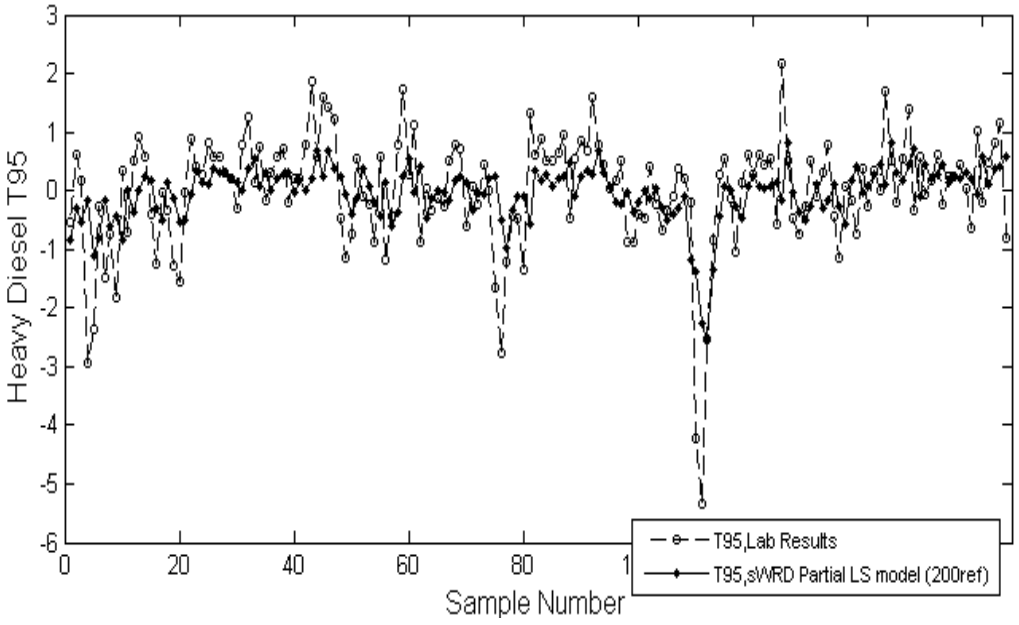

Figure J.3. HAD T95 vs. sample number of sWRD global PLS recursive 2 model for 200 reference set size

APPENDIX K: JITL LS PREDICTION TRAJECTORIES FOR INCREASED REFERENCE SET SIZE

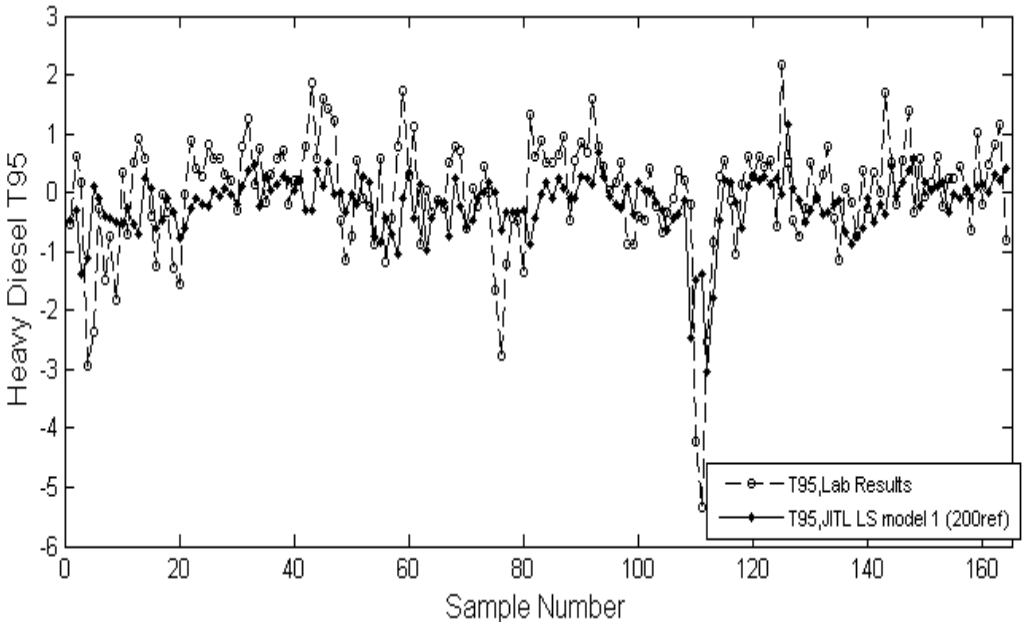

Figure K.1. HAD T95 vs. sample number of JITL LS model 1 for 200 reference set size

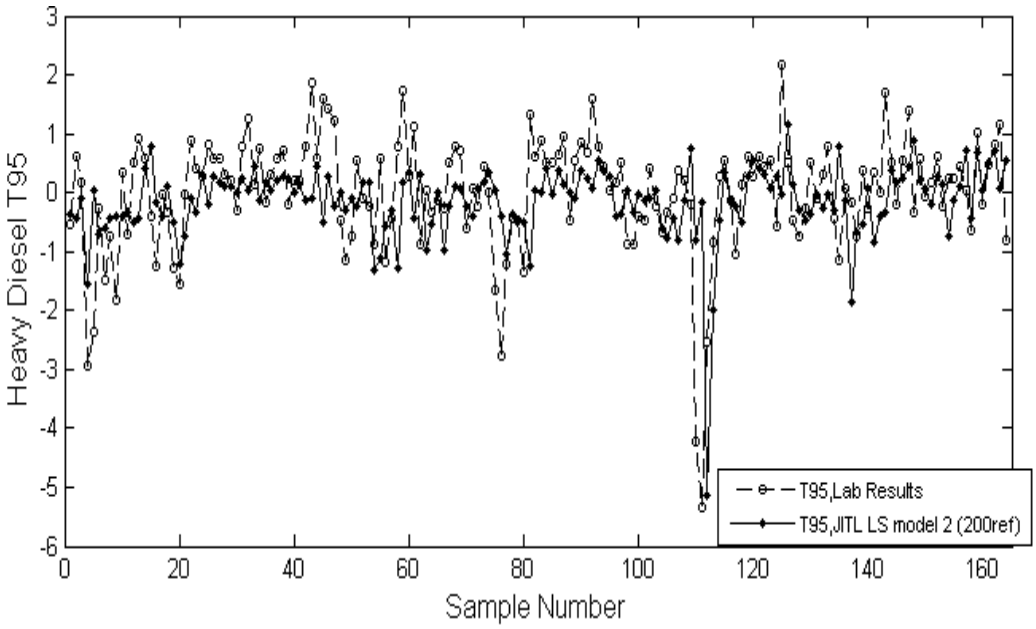

Figure K.2. HAD T95 vs. sample number of JITL LS model 2 for 200 reference set size

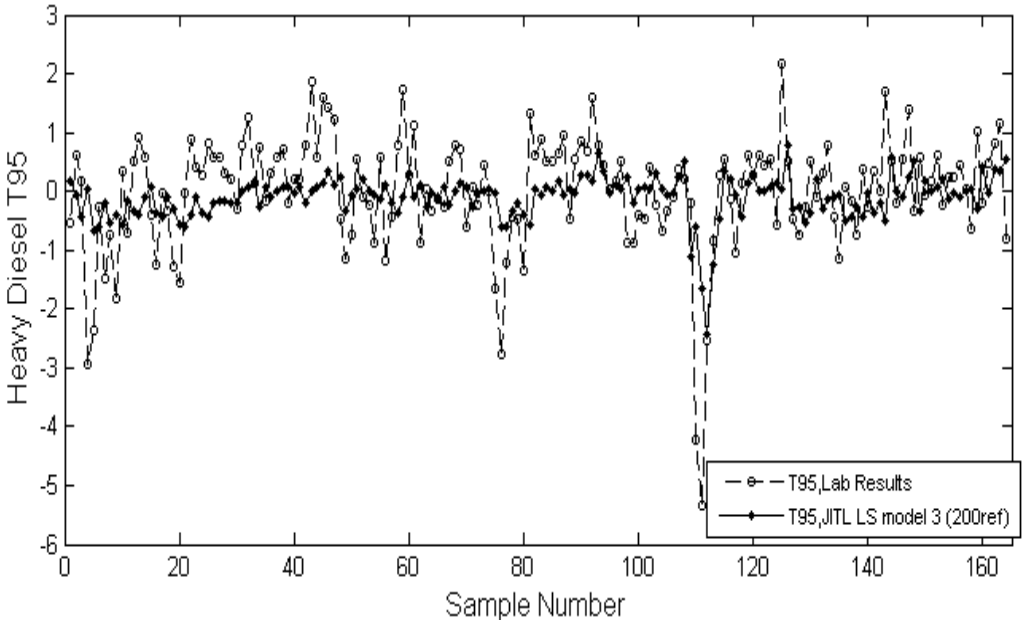

Figure K.3. HAD T95 vs. sample number of JITL LS model 3 for 200 reference set size

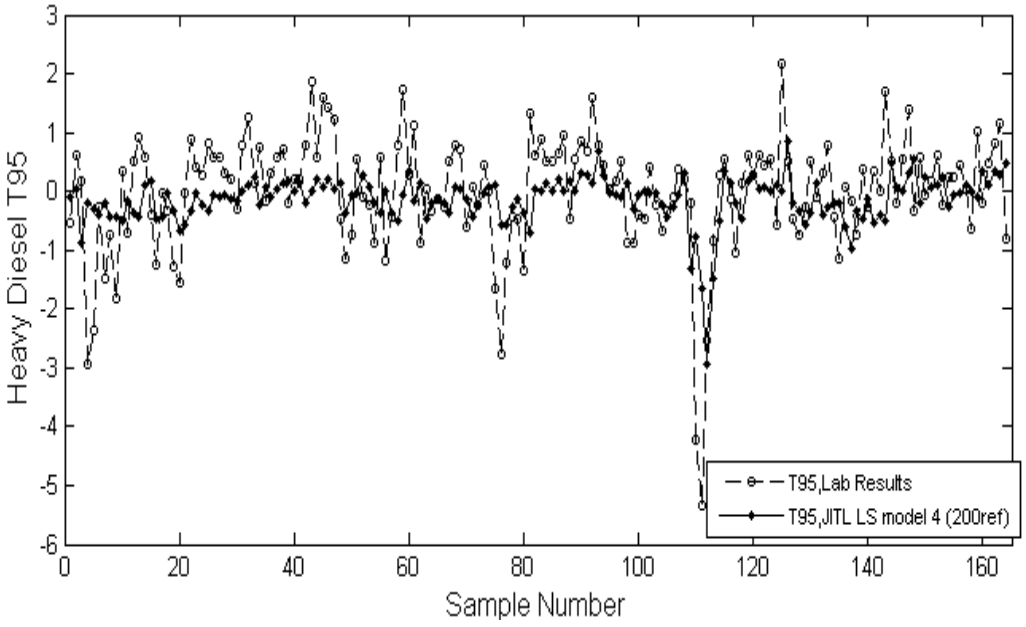

Figure K.4. HAD T95 vs. sample number of JITL LS Model 4 for 200 reference set size

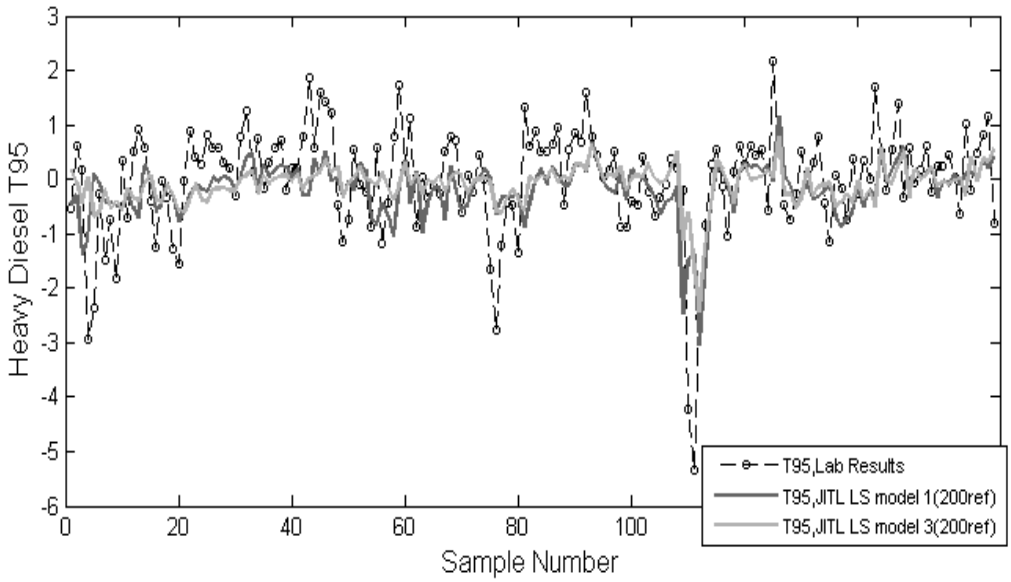

Figure K.5. HAD T95 predictions vs. sample number of JITL LS models 1 and model 3 for 200 reference set size

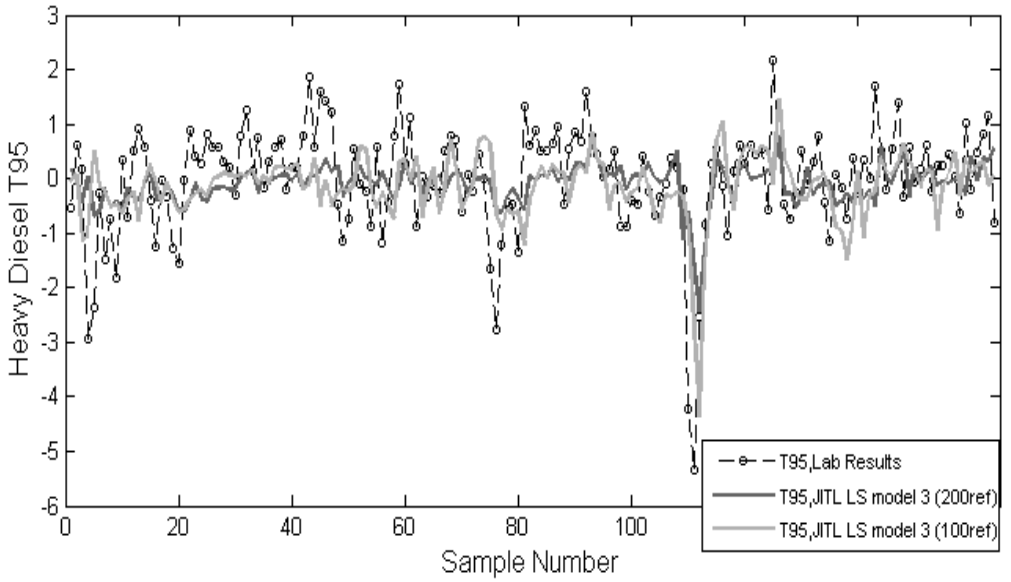

Figure K.6. HAD T95 predictions vs. sample number of JITL LS models 3 for 200 and 100 reference set size

APPENDIX L: JITL PLS PREDICTION TRAJECTORIES FOR INCREASED REFERENCE SET SIZE

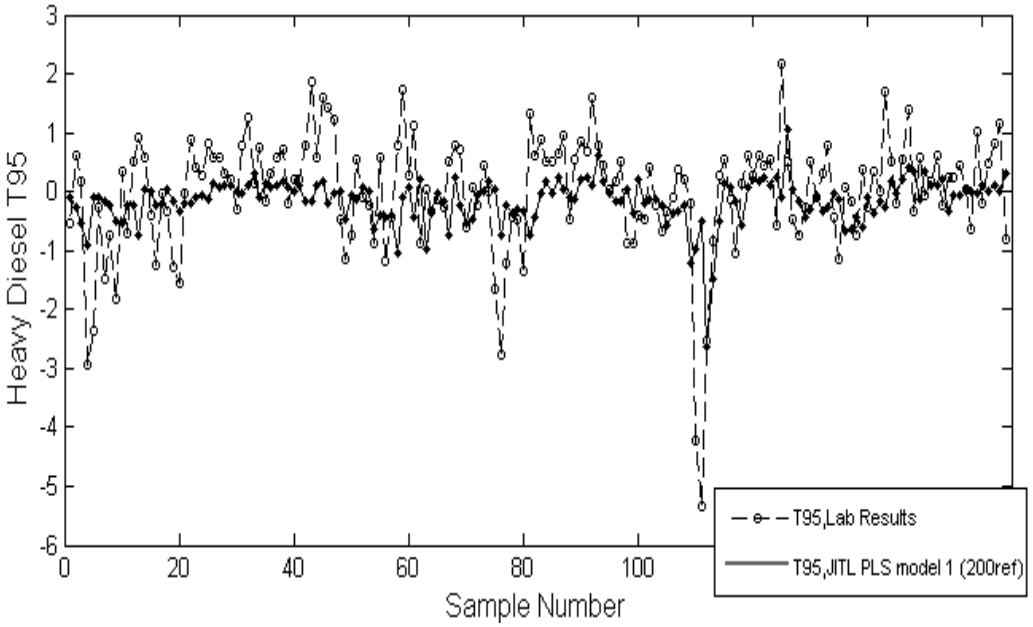

Figure L.1. HAD T95 vs. sample number of JITL PLS constant model 1 for 200 reference set size

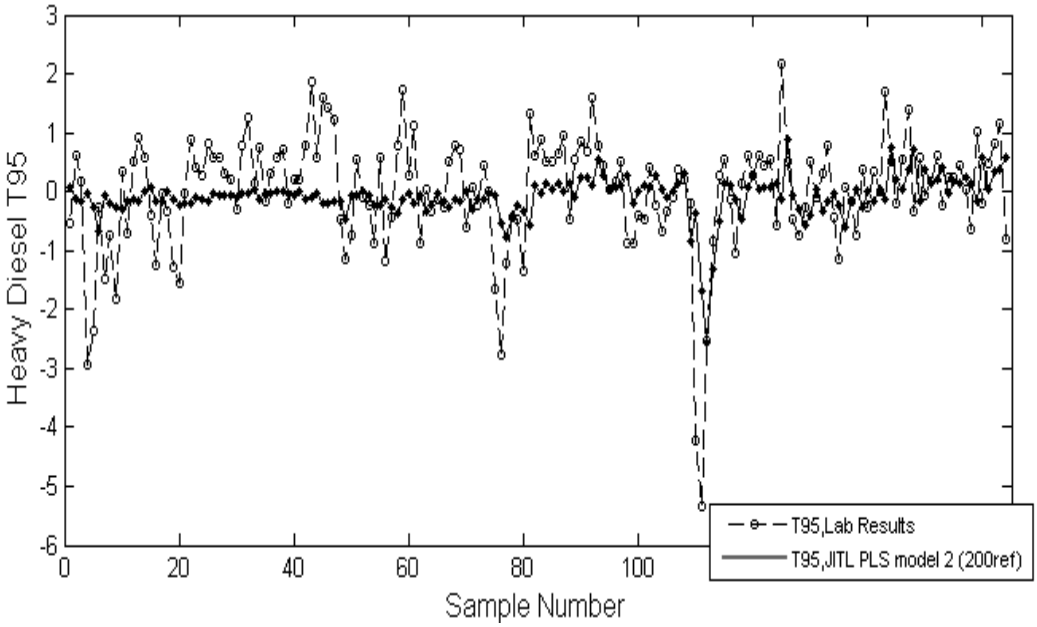

Figure L.2. HAD T95 vs. sample number of JITL PLS constant model 2 for 200 reference set size

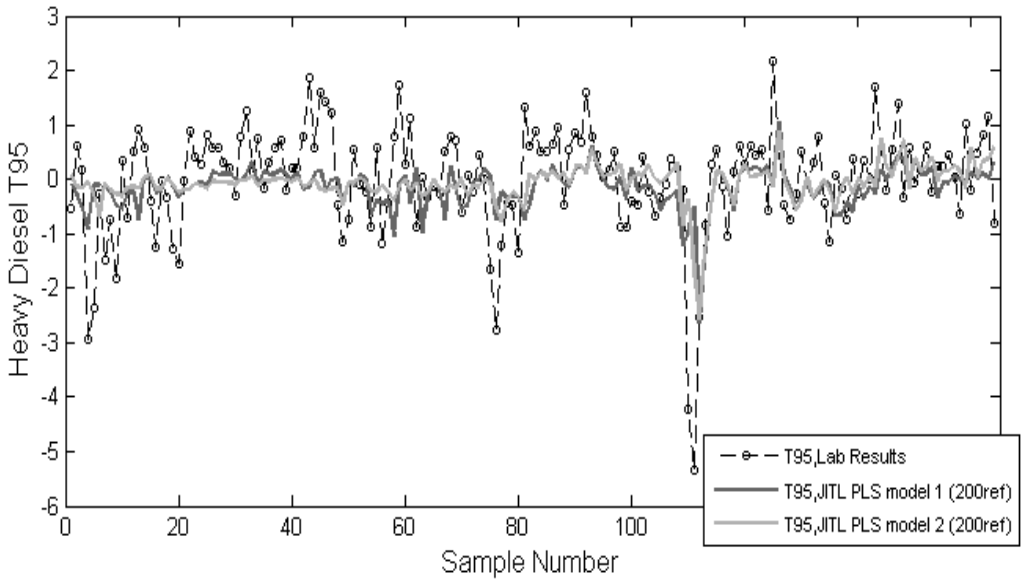

Figure L.3. HAD T95 predictions vs. sample number of JITL PLS models 1 and model 2 for 200 reference set size

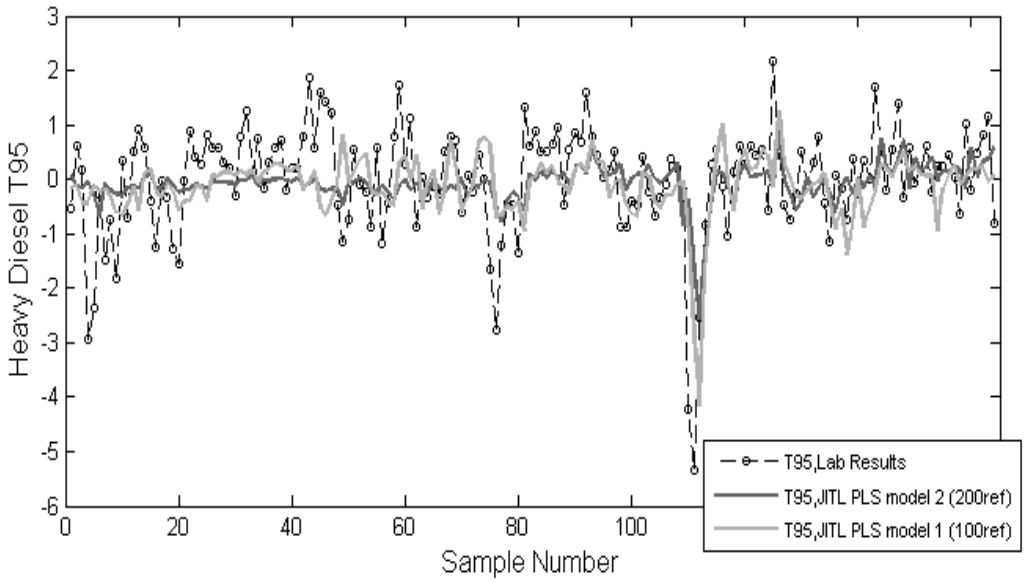

Figure L.4. HAD T95 predictions vs. sample number of JITL PLS models 2 for 200 reference set size and JITL PLS model 1 for 100 reference set size

APPENDIX M: JITL SR PREDICTION TRAJECTORIES FOR INCREASED REFERENCE SET SIZE

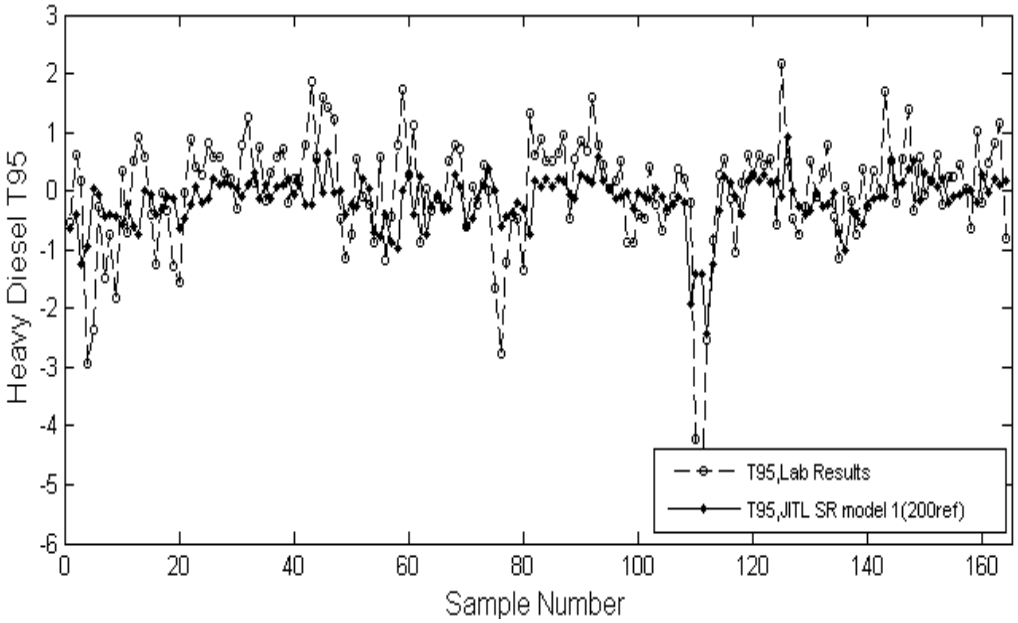

Figure M.1. HAD T95 vs. sample number of JITL SR model 1 for 200 reference set size

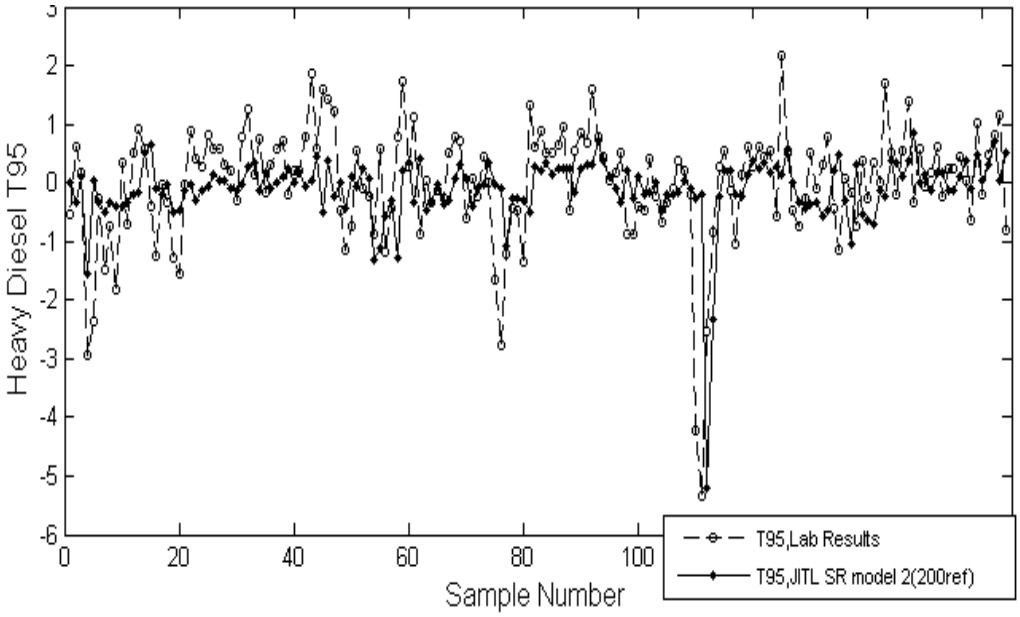

Figure M.2. HAD T95 vs. sample number of JITL SR model 2 for 200 reference set size

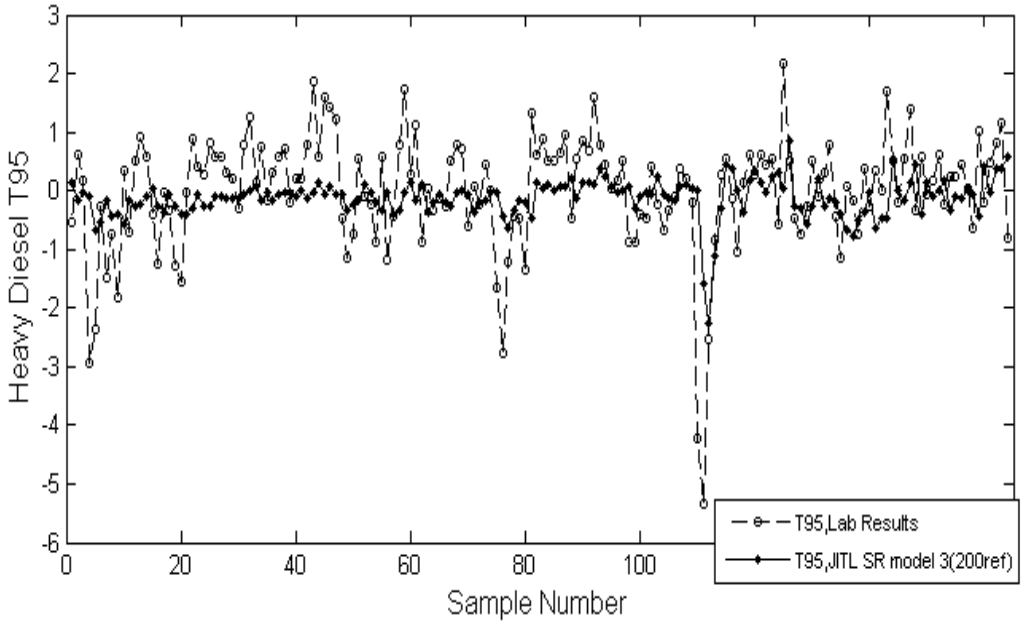

Figure M.3. HAD T95 vs. sample number of JITL SR model 3 for 200 reference set size

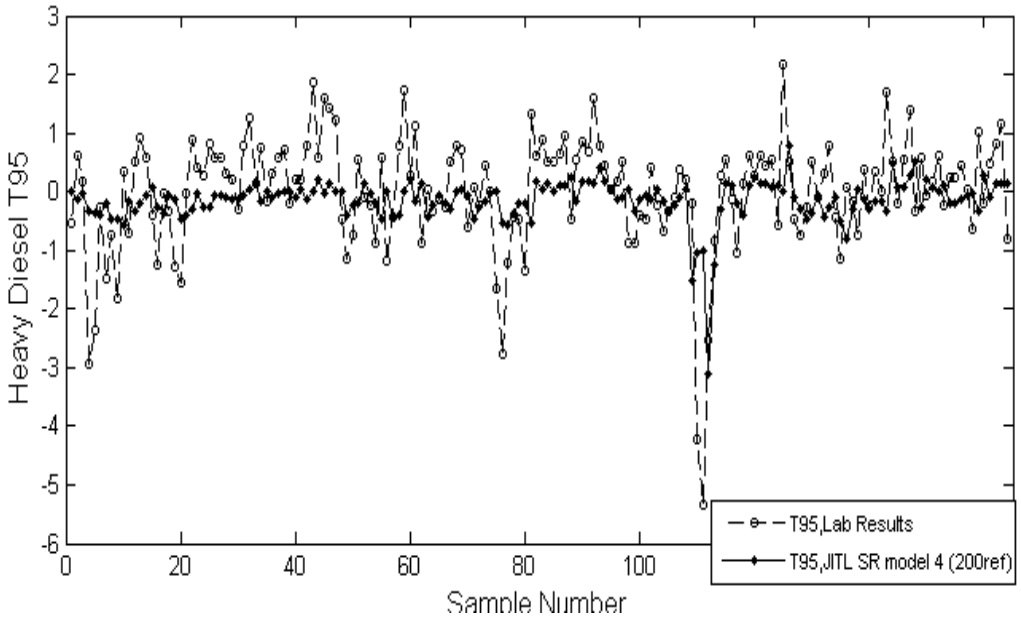

Figure M.4. HAD T95 vs. sample number of JITL SR model 4 for 200 reference set size

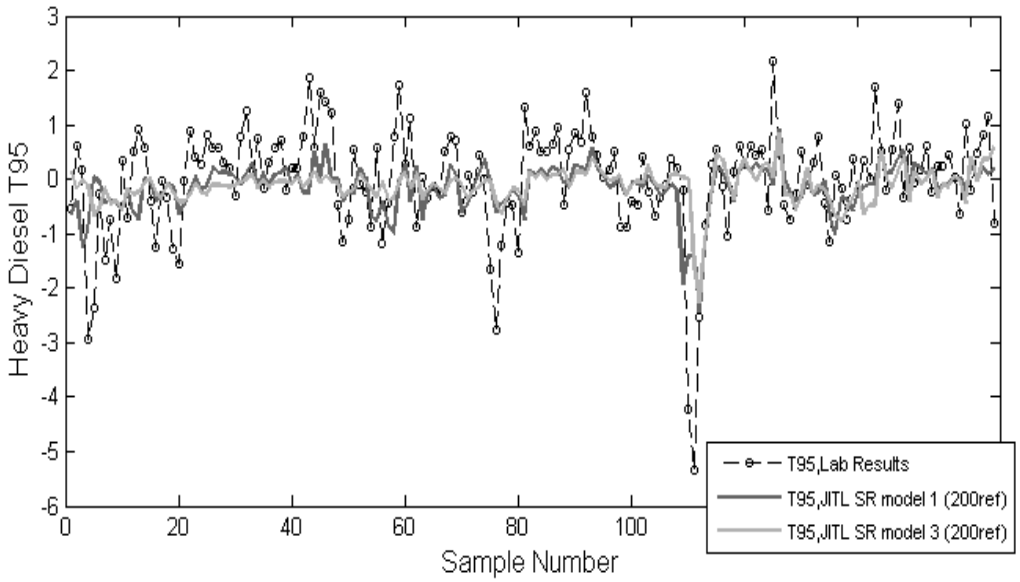

Figure M.5. HAD T95 predictions vs. sample number of JITL SR model 1 and model 3 for 200 reference set size

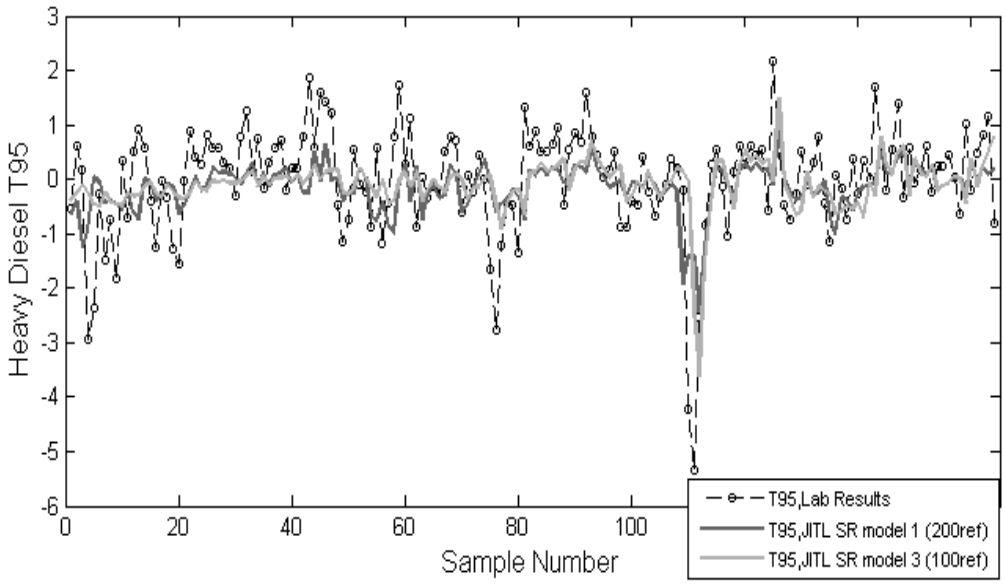

Figure M. 6. HAD T95 predictions vs. sample number of JITL SR model 1 for 200 reference set size and JITL SR model 3 for 100 reference set size