The Thesis Committee for Courtney Marie Fowler
Certifies that this is the approved version of the following thesis:

Ion Implant Virtual Metrology for Process Monitoring

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Thomas F. Edgar

Dragan Djurdjanovic

Ion Implant Virtual Metrology for Process Monitoring

by

Courtney Marie Fowler, B.S.

Thesis

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science in Engineering

The University of Texas at Austin
December 2009

Dedication

This thesis is dedicated with love to my husband, Michael L. Neel, who kept my spirits up when my determination failed me. Without his encouragement when this thesis seemed endless, I doubt it should ever have been completed.

Acknowledgements

I wish to thank Dr. Edgar and Dr. Qin for their guidance and help during my time in their research groups. I would also like to thank Bala Balakrishnan and Victor Martinez for their support and encouragement while I was at Freescale Semiconductor as well as The University of Texas. I am extremely grateful for the opportunity to pursue a graduate degree while working in the semiconductor industry. Finally, I want to thank all my friends and family that have helped me through this whole experience.

December 2009

Abstract

Ion Implant Virtual Metrology for Process Monitoring

Courtney Marie Fowler, M.S.E.
The University of Texas at Austin, 2009

Supervisor: Thomas F. Edgar

This thesis presents the modeling of tool data produced during ion implantation for the prediction of wafer sheet resistance. In this work, we will use various statistical techniques to address challenges due to the nature of equipment data: high dimensionality, colinearity, parameter interactions, and non-linearities. The emphasis will be data integrity, variable selection, and model building methods. Different variable selection and modeling techniques will be evaluated using an industrial data set. Ion implant processes are fast and depending on the monitoring frequency of the equipment, late detection of a process shift could lead to the loss of a significant amount of product. The main objective of the research presented in this thesis is to identify any ion implant parameters that can be used to formulate a virtual metrology model. The virtual metrology model would then be used for process monitoring to ensure stable processing conditions and consequent yield guarantees.

Table of Contents

List of Tables

List of Figures

Chapter 1: Virtual Metrology in Semiconductor Manufacturing

The purpose of this chapter is to provide an introduction of virtual metrology in semiconductor manufacturing. It begins with a discussion of the motivation for using virtual metrology as a solution for process monitoring. Several techniques are identified for variable selection and modeling that have been successful in practice. Finally, this chapter concludes with a discussion of the objectives and organization of the thesis.

1.1 MOTIVATION FOR USING VIRTUAL METROLOGY FOR PROCESS MONITORING

During wafer fabrication, semiconductor equipment needs to be monitored on a periodic basis to ensure stable processing conditions [1], [2]. Generally, this is performed by measuring specific monitor wafers (off-line), or measuring production wafers as they move through the factory (on-line). Any unexpected equipment drift has the potential to lower device yield, or worse, result in a complete loss of the product. Depending on the monitoring frequency of the equipment, late detection of a process shift could lead to the loss of a significant amount of product.

1.1.1 Off-line Process Monitoring

Off-line process monitoring utilizes non-production wafers to periodically measure both equipment and process performance. During off-line monitoring, production on a particular process tool is halted. A monitor wafer is then processed using a selected program. Before returning the tool to production, the monitor wafer must be measured and pass the quality test. If the monitor wafer does not pass, the equipment remains in an unproductive state until the recovery is complete. Figure 1.1 graphically illustrates the off-line monitoring scenario. This flow of events results in two main disadvantages for using off-line process monitoring: Cycle time increase and monitor wafer cost.

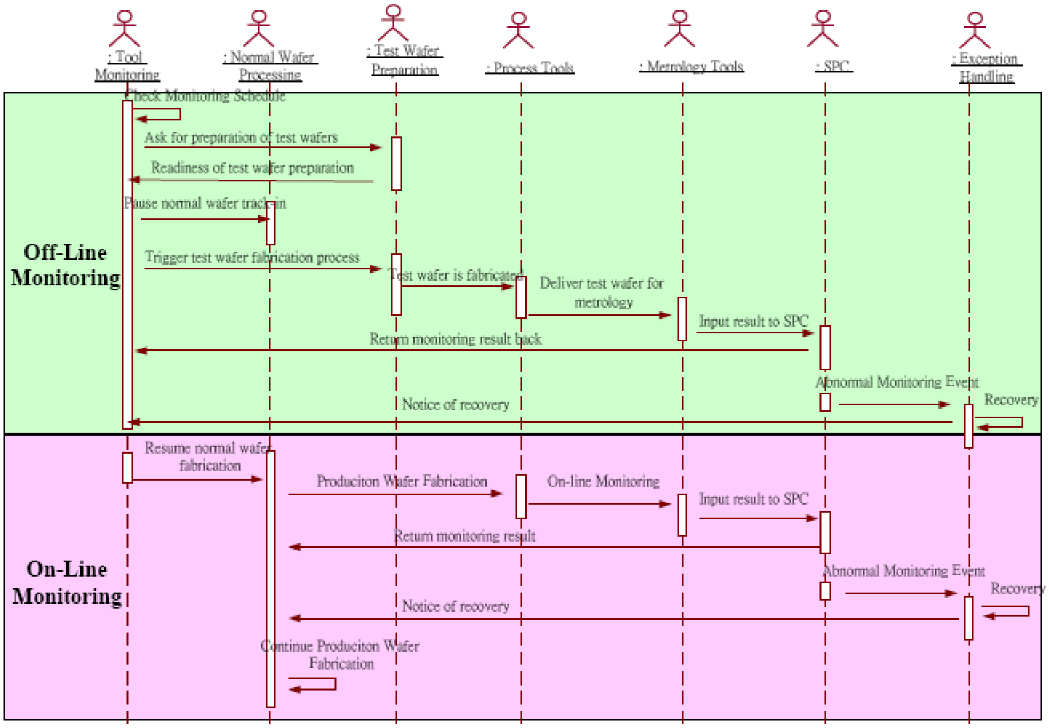


Figure 1.1: Physical Metrology Operation Scenarios [3].

The interruption of the equipment for the processing, measuring, and quality check of the monitor wafers all contribute to lost production and increased cycle times of product in the factory. Additionally, the operator performing the process monitoring is also contributing to lost production by not utilizing their time to run product instead of monitor wafers.

For a high mix manufacturing environment, such as a foundry factory, routine off-line monitoring is estimated to consume fifteen to thirty percent of daily production [3]. Aside from the obvious increases to cycle time, the consumption of these monitor wafers can significantly increase the cost of operation of the factory.

1.1.2 On-line Process Monitoring

Unlike off-line process monitoring, on-line process monitoring utilizes production wafers to periodically measure both equipment and process performance. During on-line monitoring, production on a particular process tool is not halted; rather, it is allowed to continue throughout the measurement cycle of the production wafer. Equipment may be removed from production if the quality test fails; otherwise, there is no interruption of production. Figure 1.1 graphically illustrates the on-line monitoring scenario. Although on-line process monitoring does not consume monitor wafers, the cycle time for the production wafers with on-line monitoring certainly increases. Additionally, the extra operations performed on the product wafers as a result of on-line monitoring may impact their overall quality.

1.1.3 Virtual Metrology as a Solution

To overcome the problems of traditional process monitoring, virtual metrology (VM) has been proposed as a technique to predict the process quality of every wafer using equipment and sensor data. Specifically, VM would allow a quality measurement to be conducted in real time without actually performing any on-line or off-line monitoring. In practice, the wafer quality would be known immediately after process completion as opposed to post physical metrology measurement. Virtual metrology has a number of significant advantages over traditional process monitoring that include:

- Quality assurance in real time
- A reduction in quality tests that utilize monitor wafers
- An overall reduction in metrology operations resulting in a reduction of cost and cycle time

Virtual metrology has been identified as part of the future strategy for overall equipment effectiveness (OEE) [2], [4]. The implementation of VM in a factory has the potential to replace a number of physical metrology operations and significantly improve factory productivity and quality assurance [3], [5].

1.2 VIRTUAL METROLOGY TECHNIQUES IN SEMICONDUCTOR MANUFACTURING

Virtual metrology is often selected to augment or replace off-line or on-line metrology for the purposes of improving run-by-run process control or process monitoring. In general, VM schemes are challenging to implement due to the low number of actual metrology values, equipment drift, and periodic equipment maintenance actions [6]. Investigations that compare various variable selection and model techniques can ensure the best possible prediction accuracy for the data set provided.

Virtual metrology models have been investigated for the following semiconductor processes [5], [7], [8], [9], [10]:

- Plasma Etch
- Chemical Vapor Deposition
- Rapid Thermal Anneal
- Copper Electroplate

1.2.1 Variable Selection Techniques

In order to achieve high VM prediction accuracy, it is necessary to select an optimal set of predictor variables that will represent the production environment. If too many variables are selected, the irrelevant variables can add noise that will affect the prediction accuracy [11], [12]. On the other hand, too few variables may prevent the model from ever reaching the necessary prediction accuracy [13].

Variable selection, or feature selection, is defined as selecting a subset of variables from the proposed set of predictor variables [14]. Ultimately, the goal is to make good predictions with as few variables as possible. Several techniques are employed in practice such as Principal Component Analysis (PCA), ranking of correlation coefficients, multi-regression based stepwise selection (MR-SS), and neural-net based stepwise selection (NN-SS) [6], [15]. Although multi-regression based stepwise selection is the most commonly used method for solving prediction problems, neural-network based stepwise selection has been shown to increase prediction accuracy when coupled with neural-net prediction algorithms [15].

In this thesis, we will compare the PCA, ranking of correlation coefficients, and stepwise selection techniques for variable selection.

1.2.2 Modeling Techniques

After the key variables have been selected using one of the techniques in section 1.2.1, it is time to build a prediction model using the selected key variables. Common modeling techniques employed in VM applications are multiple linear regression (MLR), principal component regression (PCR), partial least squares regression (PLS), and neural-net approaches.

Recently, efforts have been directed towards the neural-net approaches. It has been shown that neural-net based models of semiconductor processes produce better prediction accuracy when compared to the traditional statistical methods [5]. In particular, back propagation neural network (BPNN) and radial basis function neural network (RBFN) algorithms have been effective for metrology prediction [5], [7].

In this thesis, we will generate MLR, PCR, and PLS traditional models and compare them with the neural-net based BPNN and RBFN.

1.3 OBJECTIVES AND ORGANIZATION

The main objective of the research presented in this thesis is to identify any ion implant parameters that can be used to formulate a virtual metrology model. The virtual metrology model would then be used for process monitoring to ensure stable processing conditions and consequent yield guarantees.

This thesis is composed of six chapters. The first chapter is the introduction and overall description of the thesis. It contains a literature review of the variable selection and modeling techniques used for VM in the semiconductor industry. After the introduction, an overall description of the ion implant process and hardware is presented in chapter two.

Chapter three details the collection and pre-processing of the ion implant data.

In chapter four, the variable selection methods PCA, ranking of correlation coefficients and stepwise selection are described and used to extract features from the data set.

Chapter five describes three linear, and two non-linear modeling techniques. The linear techniques presented are MLR, PCR, and PLS. Neural networks are used for non-linear modeling, specifically BPNN and RBFN.

This thesis concludes with a discussion of the results presented.

Chapter 2: Ion Implant in Semiconductor Manufacturing

The purpose of this chapter is to describe the Ion Implant process and equipment. Because of its high throughput and low detection capability, Ion Implant is a prime candidate for a VM application that will predict sheet resistance for improved process monitoring. This chapter begins with a description of the Ion Implant process and concludes with an overview of the standard Ion Implant equipment configuration.

2.1 DESCRIPTION OF THE ION IMPLANT PROCESS

Ion Implant is the primary process step used to create the doped regions that form the semiconducting junctions at the heart of a Complementary Metal Oxide Semiconductor (CMOS) device [16]. Initially, doping was performed by Diffusion, but the higher control achievable with Implant has brought this technique to almost exclusive use.

2.1.1 Basics of Ion Implant

Doping of a semiconductor material involves introducing a small percentage of atoms with either more (N-type) or fewer (P-type) valence electrons than the bulk material. The interface between the two doped regions forms a semiconducting junction. The junction’s conductivity is controlled by the charge applied to it which provides the core functionality of microelectronics. In the case of Silicon, the bulk material is Group IV, (4 valence electrons). P-type doping is achieved with a group III element such as Boron (B), Gallium (Ga), or Indium (In). N-type doping is achieved with a group V element such as Phosphorous (P), Arsenic (As), or Antimony (Sb). The two most common ways of doping a material are Diffusion and Ion Implantation, both illustrated in Figure 2.1. In Diffusion, a gas containing the desired dopant is introduced above the wafer. Areas of the wafer where doping is not desired are masked using a soft mask (photo resist) or a hard mask (silicon nitride). The temperature is raised to increase dopant mobility resulting in the dopant diffusing from the high concentration region in the gas to the low concentration region in the solid. In contrast, Implant generates ions of the desired dopant and accelerates them towards the wafer. The ions are propelled into the crystal structure of the wafer by kinetic energy and reach a depth that is a function of their energy and the number of collisions they encounter on their path. As in Diffusion, a mask is used to define the area to be implanted. Implant is typically characterized by 3 parameters: the dopant, the dose, and the profile. The dopants can be either P-type or N-type. Dose is the amount of dopant introduced into the substrate, described in terms of ions/cm2. Finally, the profile describes the dopant concentration vs spatial dimension, both laterally and vertically.

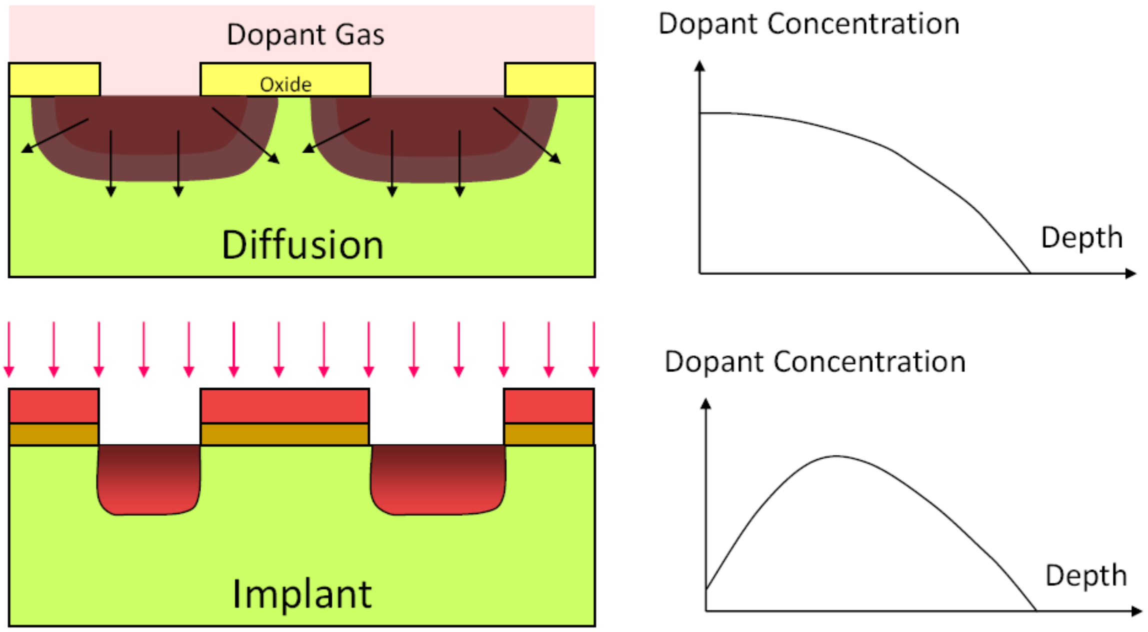


Figure 2.1: Comparison of doped regions formed by Diffusion and Implant.

One important advantage to using Ion Implant instead of Diffusion to change the electrical characteristics of the bulk material is the ability to control the dopant concentration profile. In Diffusion, the slope of concentration the profile is shallow and isotropic causing the maximum concentration to occur at the wafer surface. Because the process is driven by diffusion gradients, the resulting profile cannot be significantly engineered. The portion of the wafer in contact with the dopant gas will always have the highest dopant concentration. Concentration will decrease uniformly in all directions from the surface because the diffusion gradient driving the dopant into the wafer is uniform in all directions

In contrast, the kinetic energy driving dopant into the wafer in Implant is uniform and unidirectional. Only three factors can blur the lateral sharpness of the profile: masking non-uniformity, random collisions, and beam non-uniformity. If regions of the mask can be penetrated by the ion beam, some lateral non-uniformity will result. As the ion tunnels into the bulk Si, some collisions will scatter it laterally, resulting in a less sharp lateral profile. Finally, the beam is never completely collimated, and the approaching ions have some lateral velocity. Even with these imperfections, the lateral profile achieved with implant is orders of magnitude steeper than what is possible with Diffusion. This sharp profile allows the entire device to shrink and has enabled recent advances in critical dimension reduction.

Vertical profile in implant is largely a function of kinetic energy distribution in the beam and distribution in free path between collisions with the bulk material. Roughly speaking, a certain number of collisions are required to dissipate the incoming ion's energy. There will be statistical variation in the distance covered between these collisions, and this will result in a horizontal concentration gradient of dopant, as not all atoms stop at the same depth. Implant vertical profiles are primarily dependent on the energy of the implanted ions and the crystallographic orientation of the silicon with respect to the ion beam.

The greatest advantage of Ion Implant is the ability to tune Rp, (projected range). This parameter is used to quantify the depth of doping. Unlike Diffusion, where peak concentration must occur at the surface, Implant can adjust the vertical location of peak concentration by adjusting the ion energy. There are limits to Rp due both to the ion energies required and resulting damage to the device resulting from large values of Rp.

Unlike Diffusion, where the crystal structure smoothly adjusts to the dopant impurities, implant induces crystal damage that must be repaired by a subsequent anneal. The anneal must be performed quickly enough to maintain the sharp, non-equilibrium dopant concentration profile, but must last long enough to repair the damage. Clearly, there is some level of lattice dislocation that cannot be repaired without other side effects.

2.1.2 Implant Applications in a Modern CMOS Device

The Ion Implant process is highly flexible resulting in its extensive use in CMOS design. Specialized CMOS circuits can use up to 40 Implant steps during fabrication [16]. Species, doses, and energies of the Implant can be quickly varied over several orders of magnitude to tailor the electrical characteristics of the device. Figure 2.2 illustrates the operational space of Ion Implant in CMOS devices. Notice that Ion Implant can provide a five decade range for energy as well as a six decade range for dose.

Figure 2.3 illustrates the doped regions that are required to form a modern CMOS device. Each of the regions requires specific ion beam characteristics. Some of these characteristics can be achieved by recipe modifications of a single tool, while others require a different toolset. The well and basic S-D junctions are good examples of contrasting implant requirements. Each transistor requires a well that is doped in complement to the source-drain region. This well requires a deep (high energy) but lightly doped (low dose) implant. In contrast, the source-drain (S-D) junctions require shallow (low energy) but highly doped (high dose) implants. In addition, modern high-performance devices require delicate shaping of the junction through the use of source-drain extensions and halo implants. These implants require extremely precise control, as device performance requires the generation of a complex 3-dimensional doping profile. Such features as halo implants and anti-punch through can only be achieved via Ion Implantation. As discussed in the previous paragraph, the features require more than different masks, they also require different beam characteristics, and thus different recipes. Since device characteristics are largely controlled by semiconductor junction characteristics, each device essentially needs a different junction design, thus a different implanter recipe. The end result is that a high-mix fab must maintain hundreds if not thousands of different recipes. Both process diagnosis and simple accuracy of the recipe set is a major challenge.

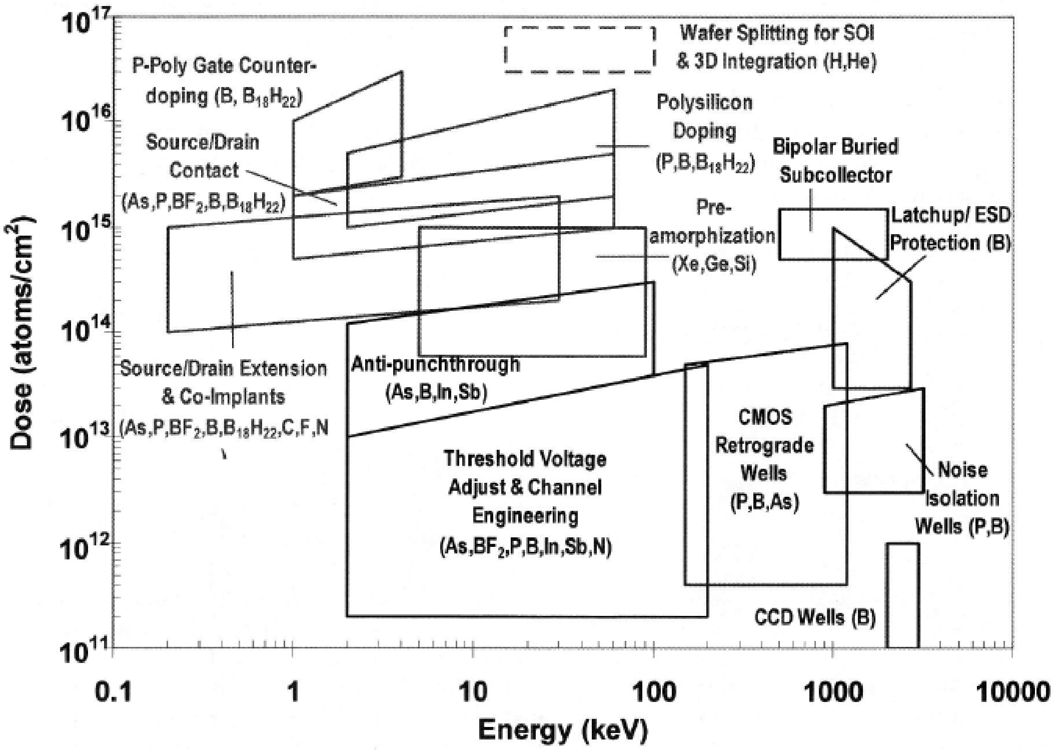


Figure 2.2: CMOS Implant applications in dose-energy space [16].

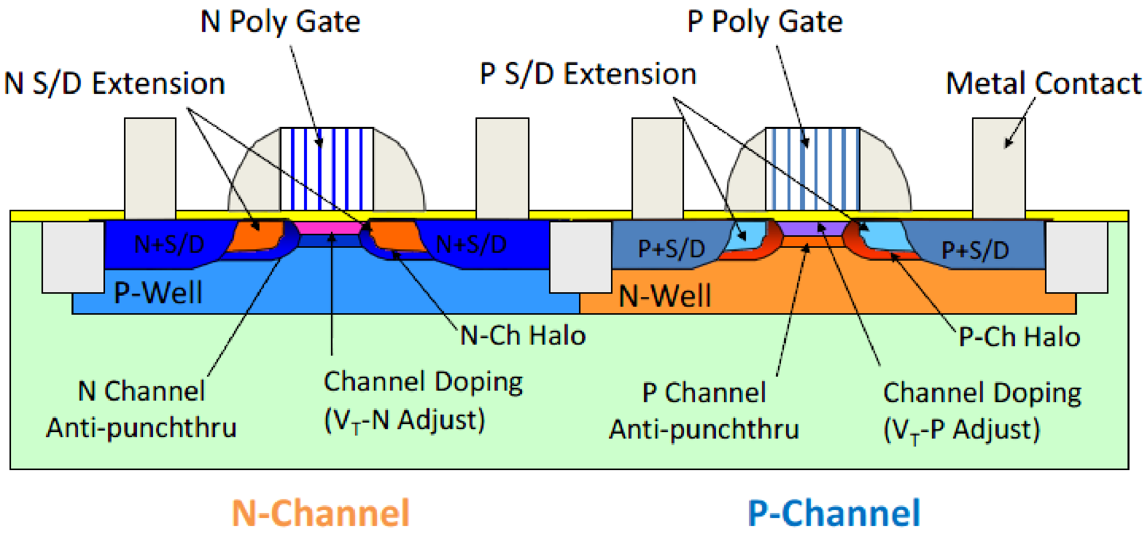


Figure 2.3: Illustration of doped regions in a CMOS device.

2.2 ION IMPLANT EQUIPMENT

It is useful to think of an implanter as a very large mass spectrometer with a wafer in the place of the detector. Figure 2.4 is a schematic of a generic implanter.

A feed gas that includes the desired atom, for example Boron Trifluoride (BF3) is introduced into the ion source. It is ionized into a number of fractions, including a singly charged Boron ion B+. All ions created in the source are extracted with a large negative extraction voltage and accelerated into the magnetic sector mass filter, or analyzer magnet.

The analyzer magnet uses a magnetic field to induce a charged particle to follow a curved path. The path radius is a function of the ion kinetic energy and magnetic field strength. The exit slit of the mass filter acts as both a mass and energy filter, eliminating other mass fragments such as BF2+ as well as B+ ions with significantly different energies. The complexity of the acceleration stage is a function of the maximum beam energy and is a major factor affecting tool design. 'Low Energy' beams are on the order of 1 KeV to 100 KeV and achieve these levels with the use of a simple DC bias. Typically, electrostatic and mechanical (slit) lenses are used prior and post acceleration to collimate the beam as well as tighten the energy distribution. 'High Energy' beams are on the order of 10 keV to 4 MeV and require the use of a linear accelerator. As can be imagined, the linear accelerator is a large and complex component. It consists of multiple RF accelerator and quadrupole filter stages. Achieving the desired energy level, collimation and energy distribution requires complex tuning of this multistage system.

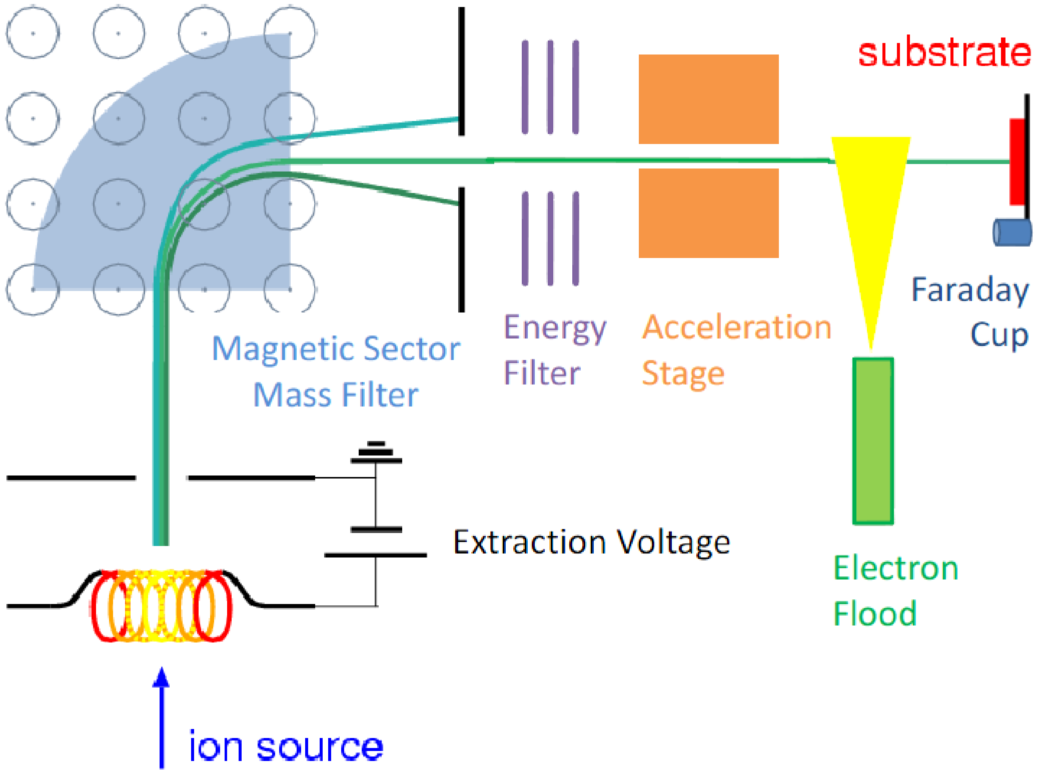


Figure 2.4: Implant Equipment Schematic.

Large parts of the wafer are covered with insulating material, making charging a major concern of ion implant. If uncontrolled, charge can build up and cause discharges that destroy gate oxide structures. Charging is mitigated by the use of an electron flood that neutralizes the positive charge. The electron flood needs to be precisely calibrated, as excess negative charge is even more likely to induce damage than excess positive charge.

An advantage of the ion beam is that the delivered dose can be readily inferred using a Faraday cup. The Faraday cup does not overlap the substrate target, thus the ratio of ions collected by the Faraday cup to ions that hit the target must be constant. This ratio can change significantly as beam geometry changes with different recipe parameters.

The beam of dopant ions of roughly constant energy is then targeted towards the wafer. Typically, multiple wafers are mounted on a rotating platform that can be angled relative to the beam. A single run consists of implanting all the wafers on the disc.

Chapter 3: Implanter Data Collection and Data Preprocessing

The purpose of this chapter is to discuss the methods used to collect and treat raw equipment data prior to creating VM models. This chapter begins with a description of how data was collected from both the implant and metrology equipment. This is followed by a discussion of rejection criteria used to filter outlier data and data with poor quality. The chapter concludes with the methods used to frame and summarize the filtered raw equipment data.

3.1 DATA COLLECTION

Two types of data are commonly used as input for virtual metrology applications: data collected from equipment during process and process measurement data from metrology equipment. Equipment data are collected from integrated sensors using a communication port on the controller module. Although there are a number of different communication protocols available, Implanters are integrated using the Semiconductor Equipment Communication Standard I (SECS-I) or High-Speed SECS Message Service (HSMS). Metrology data are collected from dedicated metrology equipment using proprietary File Transfer Protocol (FTP) transactions that populate the fields in third-party SPC packages.

Equipment data can be collected in a continuous stream, or batched by processing step. For Ion Implant, the batch collection method is used. After each group of wafers is processed, a record is created that contains the time-series data. The process record is then framed, summarized, and normalized for use in unsupervised run-by-run analyses.

Metrology data are collected at some predetermined frequency that depends on the level of risk a factory is willing to take compared to the cost of collecting the data. It can easily be seen that more metrology measurements would certainly lower the risk to product, however, more metrology measurements also means the factory experiences lower productivity. One of the primary metrology types used in Ion Implant is a sheet resistance measurement of a specific low-dose implant into a test wafer with known dopant profile and crystalline structure. This measurement is typically performed once every 48 hours.

3.1.1 Effect of SECS Communication on Data Collection

Because SECS communication is a polling protocol, it requires a primary message and a corresponding secondary, or reply, message to complete a communication transaction. This sequence of ask and reply messages sent between the Equipment and Host can introduce several complications to the data collection process including: determining the optimum polling frequency, determining the internal parameter update frequency of the equipment, and synchronization of data streams. The Implant processes discussed here are of a sufficient duration such that internal parameter update frequencies and synchronization become less of a concern after statistical summarization is performed. Determining the optimum polling frequency, however, is critical to the health of the controller module.

Finding the optimum polling frequency for a specific type of equipment is always a tradeoff between the desired data collection rate and the degradation in controller response from repeated SECS requests. The optimum polling frequency depends on the data collection rate and also the number of parameters requested from the equipment. For example, a 200 mm multi-chambered process equipment is known to crash the controller when more than 30 parameters per chamber are requested at a frequency 1 Hz. Implant equipment have a single process chamber and are able to collect ~100 parameters at a frequency of 0.5 Hz.

After determining the optimum polling frequency for a specific type equipment is at a reasonable rate for data collection, it may be disappointing to discover that the internal parameter update is somewhat slower. The controller module of any process equipment is performing a number of tasks simultaneously with equipment control always the priority. Responding to SECS messages and updating internal parameters are given a lower priority in the processing queue. Figure 3.1 shows time series data from an Implanter where the response to SECS messages is 0.5 Hz, however, the internal parameter update is on the order of 1.5 to 2.0 Hz. The resulting graph is a step function as the Implanter reports the same value multiple times before updating the internal parameter.

For data collection using SECS, synchronization of the time series data streams is automatic if all parameters are contained within the same request message. Each return message containing the parameter values has a time stamp from the equipment that can be applied to each data point in the set. When data streams are collected using multiple protocols, the frequencies may not be identical and a synchronization technique is required. Since the Implant data will be reduced to statistical summaries by batch, synchronization beyond the SECS time stamp is not necessary.

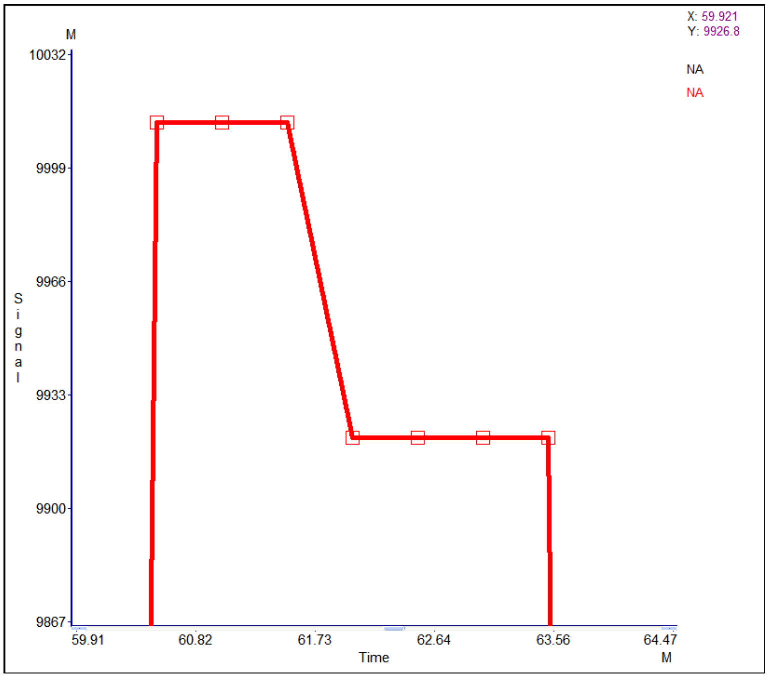


Figure 3.1: Trace data showing the Implanter SECS response at 0.5 Hz with an internal parameter update of 1.5 to 2.0 Hz.

3.1.2 Metrology Data Collection

Because there is no acceptable method for measuring implanted dose on product, a standard low dose recipe is used to implant a test wafer for the periodic qualification of implant equipment. When the implanter is due for qualification, the equipment operator will stop production and run the low dose recipe on a test wafer. After the implant is complete, the operator will then take the test wafer to a metrology measurement station and record the sheet resistance of the test wafer.

The sheet resistance of a silicon wafer is measured using a four-point probe. This technique requires a fixed current to be injected into the wafer surface through two outer probes. The resulting voltage is measured between two inner probes, see Figure 3.2.

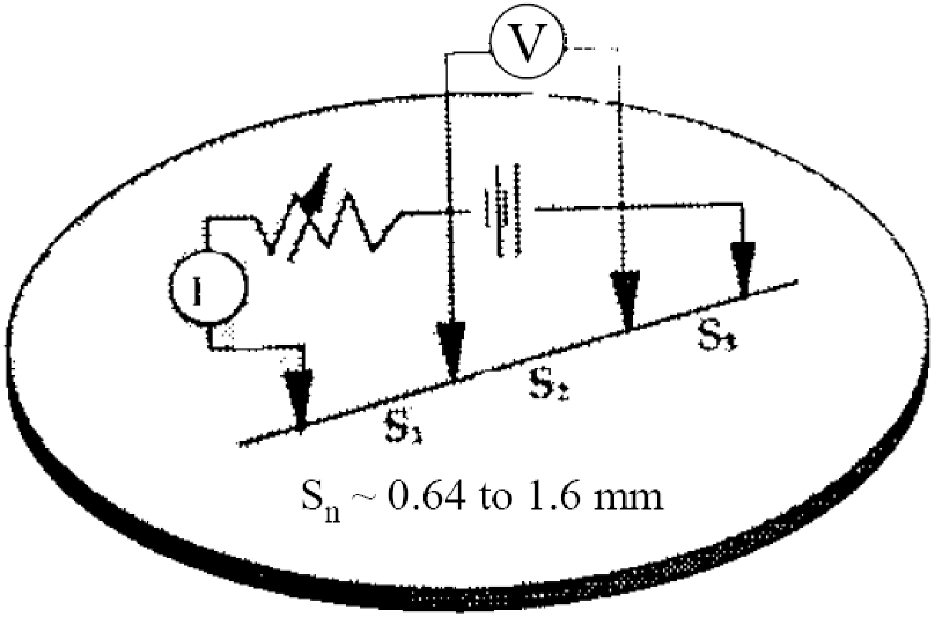


Figure 3.2: Schematic of a four-point probe measurement [19].

A set of 49 sheet resistance measurements are taken in concentric rings across a 200 mm test wafer. These measurements are summarized by calculating the arithmetic mean and standard deviation of the set. Figure 3.3 illustrates the sheet resistance measurement pattern.

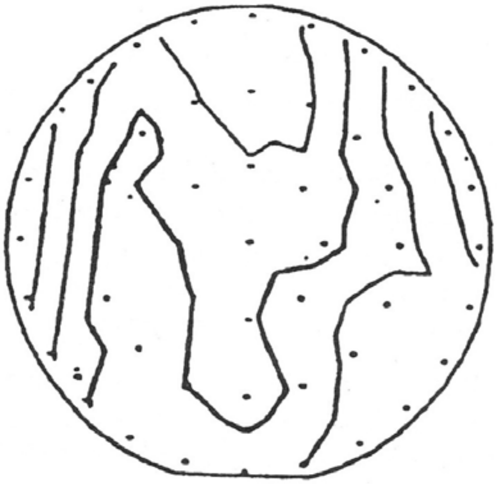


Figure 3.3: Pattern of individual sheet resistance measurements on a 200 mm test wafer [17].

3.2 REJECTION CRITERIA

Data collected from equipment and metrology tools may be inaccurate due to communication or measurement errors. Communication errors generally result in problems with data quality, while measurement errors produce outliers. Missing data can be the result of either communication or measurement errors. It is necessary to screen the data and repair any errors before attempting to build a model [11].

3.2.1 Data Quality

Data quality is a general term that describes the accuracy, correctness, completeness, and relevance of data. The use of the term data quality in this paper is restricted to the relevance of the data as it compares to equipment communication. In other words, we would like to know how well the equipment is sending data compared to the actual processing of wafers.

Data quality, in this respect, begins with a measurement of data latency. Remember that SECS communication is accomplished by polling the equipment. The basic sequence of events is as follows: (1) Ask the equipment for data, T_{poll}, (2) The equipment responds with data values, T_{arrive}, (3) The resulting data is posted to the process record, T_{run}. The duration between (2) and (3) is generally constant and set by the data collection system. The duration between (1) and (2), however, can be problematic depending on the processing load of the equipment control computer. The data latency can be computed at each time period for a set of variables by

$$\text{Data Latency} = \text{maximum}(T_{run} - T_{arrive} - T_{poll}). \qquad (3.1)$$

The data quality attributes can be collected by the system and used to determine the relevance of the data in the process record. If a piece of equipment consistently communicates poorly during a process record, it is highly likely that the accuracy of that record is suspect and it should be excluded from any analysis and or model creation. Specifying a threshold for latency allows the fraction of data points that meet the quality criteria to be calculated per process record. Records can be excluded based on the data quality fraction.

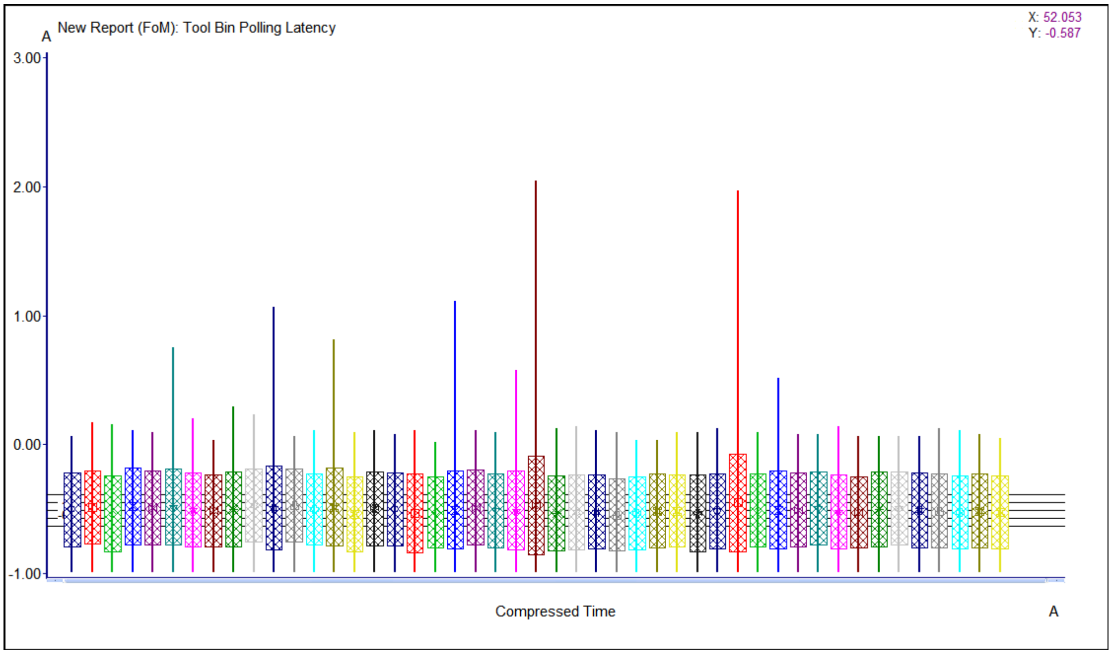


Figure 3.4: Data latency measurements for Ion Implant low dose qualification process records.

In practice, data latency values should not exceed two seconds when the equipment is responding properly to SECS requests. Figure 3.4 shows the minimum, maximum, average, and standard deviation of the low dose qualification process records used for this analysis. Notice that the process record averages are all well below the two second limit, however, there are a number of records whose maximum values exceed the limit. For these records, it was found that a single data point with high data latency was at fault. Figure 3.5 shows the data latency values by time for one of these records. The SECS communication of the equipment is well behaved for these high dose qualification records and no records were removed from the set.

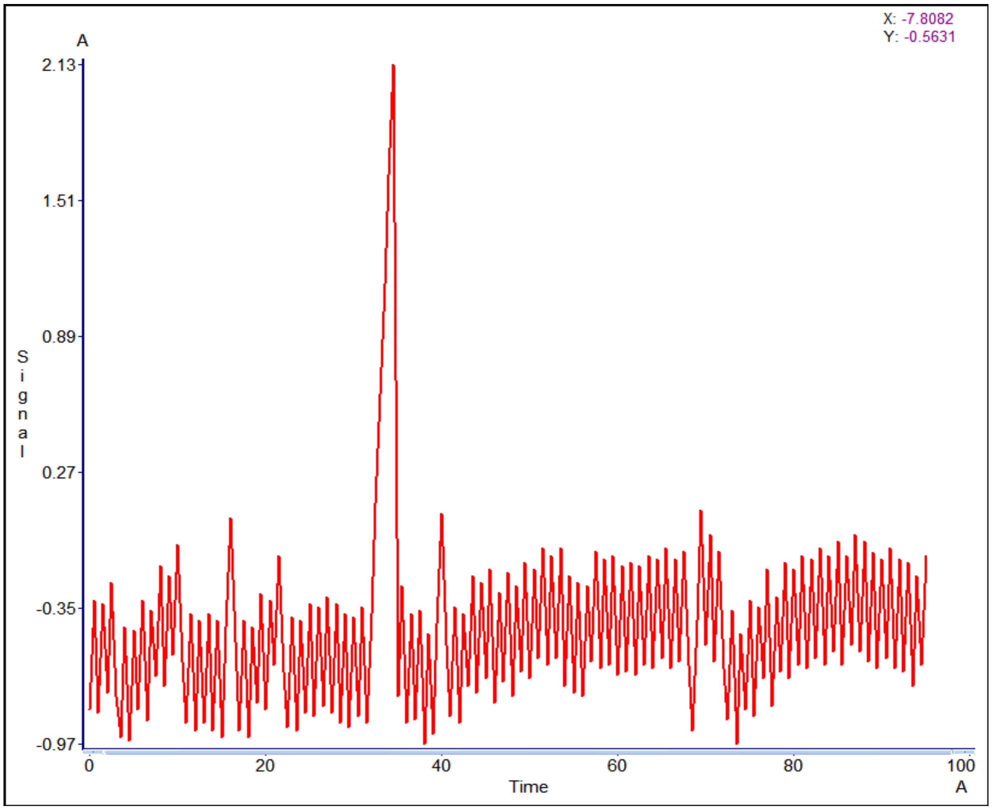


Figure 3.5: Data latency measurements by time for one Ion Implant low dose qualification process record.

3.2.2 Outlier Rejection

To further improve the accuracy and robustness of the data, outliers will need to be detected and either corrected or removed prior to building the model. For metrology data, this task requires a simple three sigma univariate control chart to compare the sample averages and standard deviations. Data from the equipment, however, requires a multivariate technique for efficient outlier rejection.

Within wafer metrology data is not available for analysis, only the sample average and standard deviation are reported. This makes the task of outlier rejection simple. Using a univariate control chart, the data can be quickly screened for outliers. The low dose qualification average sheet resistance and standard deviation values are plotted along with three and six sigma levels in Figures 3.6 and 3.7. Although a few data points exceed the three sigma level, no data points exceed the six sigma level and no data will be removed from the analysis set.

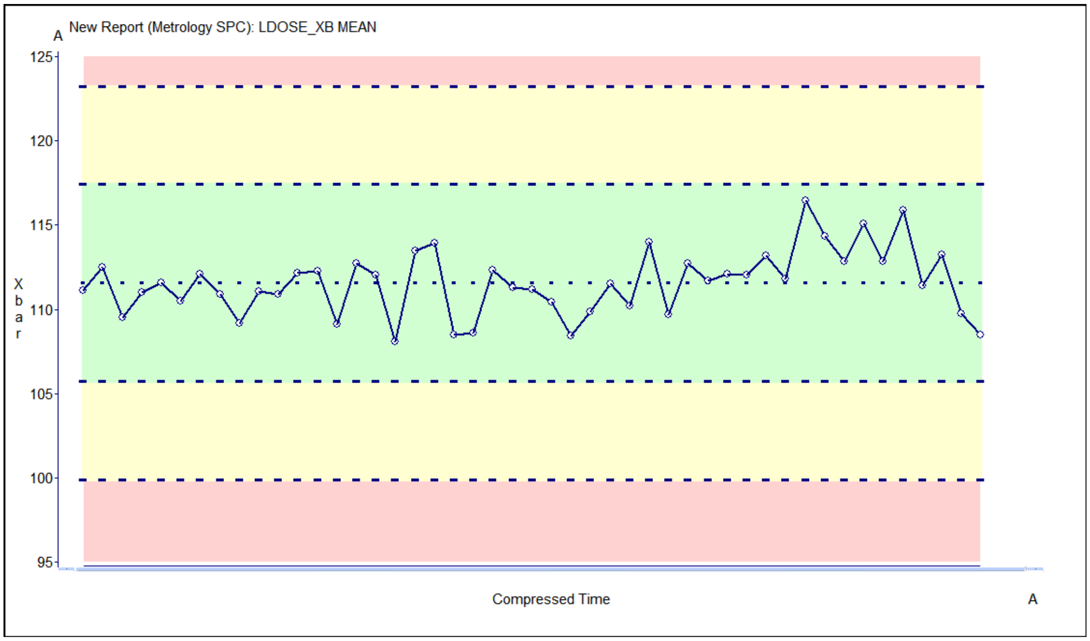


Figure 3.6: Low dose qualification average sheet resistance values.

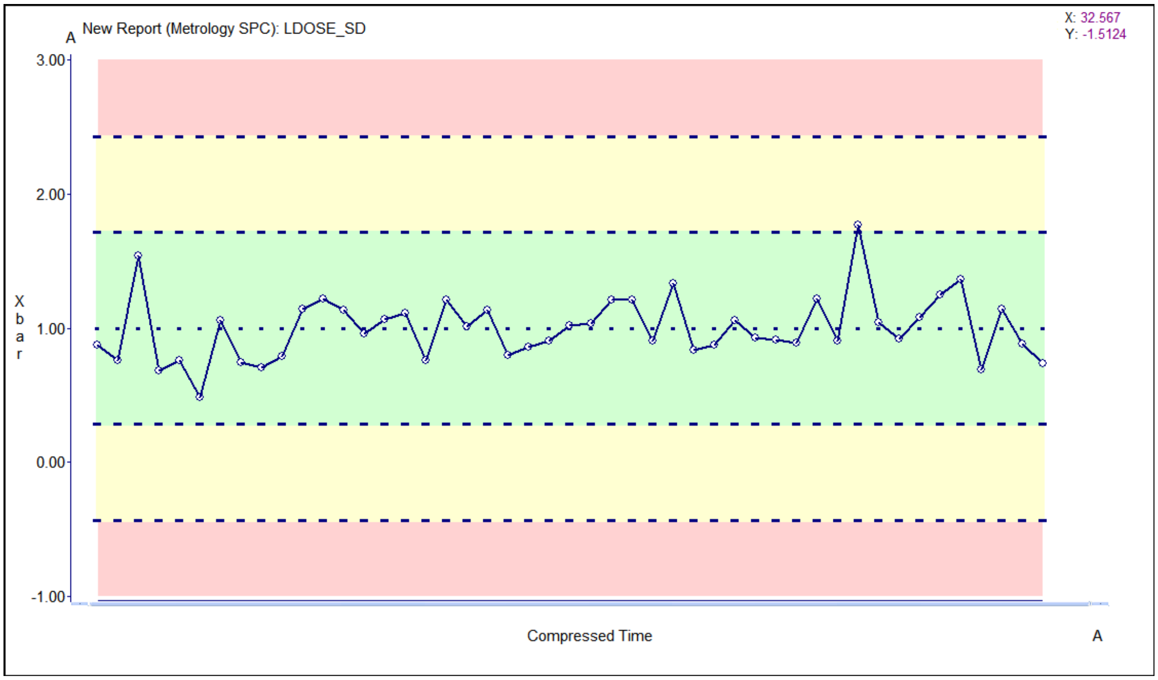


Figure 3.7: Low dose qualification standard deviations of sheet resistance values.

We will use Principal Component Analysis (PCA) as a dimensionality reduction technique to explore the variability of the trace-data set and identify outliers. In particular, we will examine the PCA scores as well as the T^2 and Q statistics. The PCA methodology is described in Section 4.1.

PCA scores were calculated from the 47 predictor variables of 48 process records. A scores plot was generated from the projections of the data onto the first to PCA loading vectors. Figure 3.8 shows the resulting scores plot with a 95% confidence band. One of the low dose qualification data falls outside of this confidence band, however the T^2 and Q statistics detailed below indicate that it should not be removed from the set.

Hotelling's T^2 statistic is used to measure the systematic variation of the process. T^2 statistics were computed and plotted, see Figure 3.9. The T^2 plot indicates that there is not a significant amount of systematic variation in these process records.

The Q statistic is used similarly to T^2, however, Q is a measure of the random variation in the process. Q statistics were computed and plotted, see Figure 3.10. The Q plot indicates that there is not a significant amount of random variation in these process records.

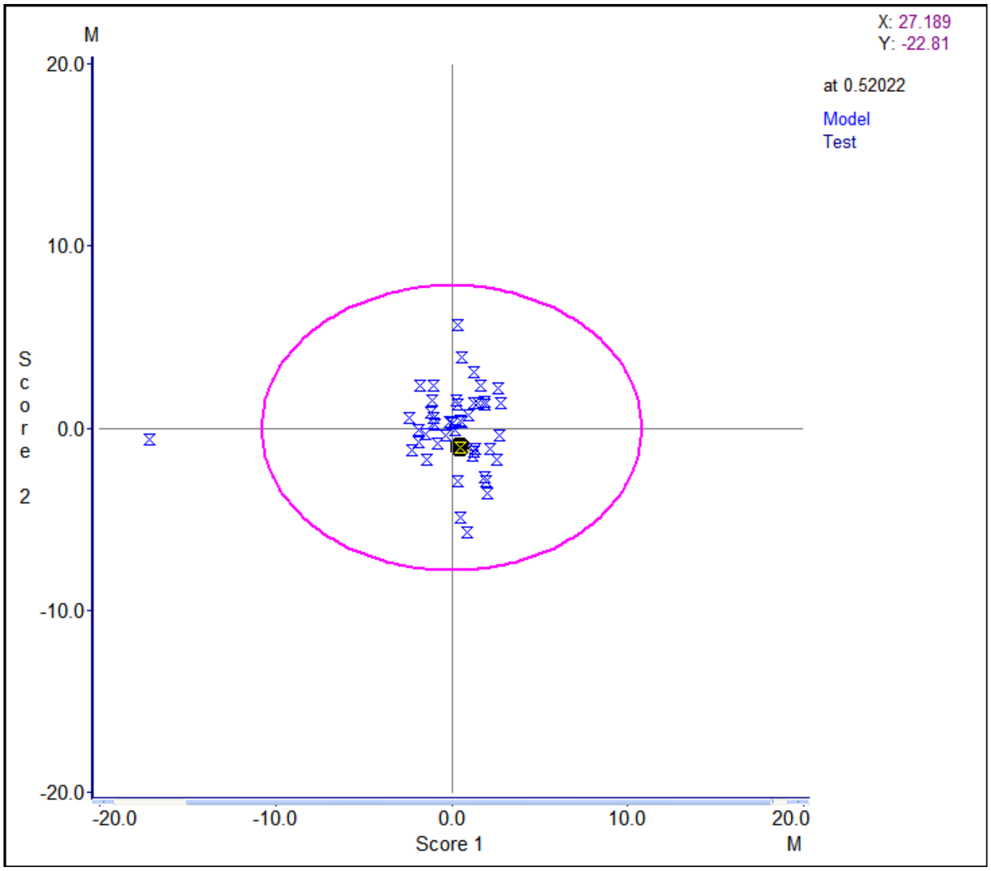


Figure 3.8: PCA scores plot of low dose qualification data with 95% confidence band.

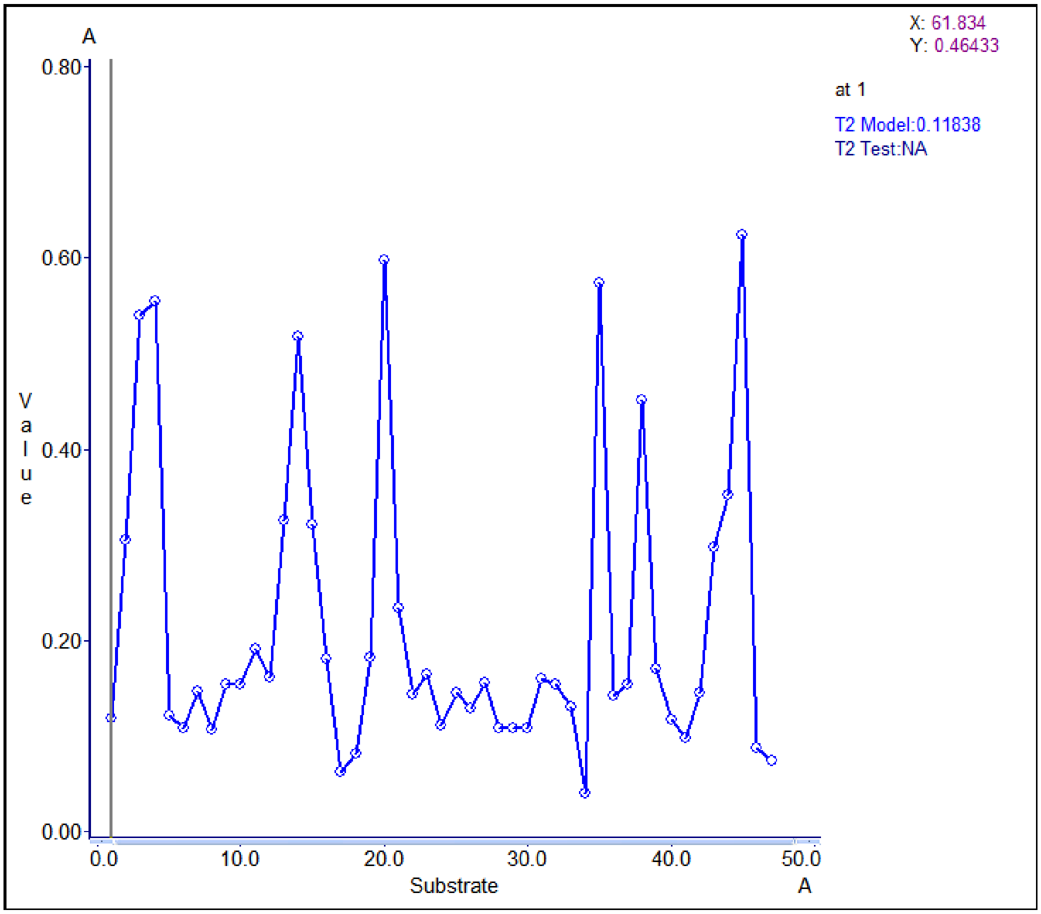


Figure 3.9: Hotelling's T^2 plot of low dose qualification data.

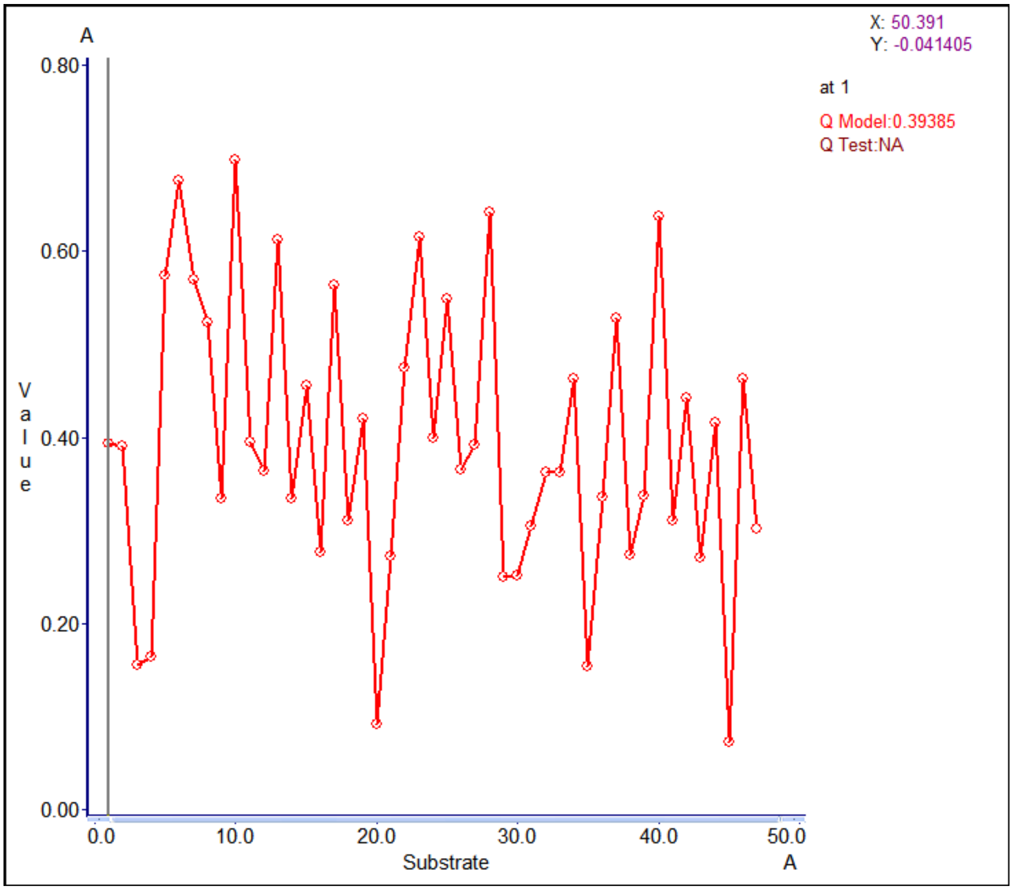


Figure 3.10: Q plot of low dose qualification data.

No low dose qualification records were removed based on the PCA results.

3.2.3 Missing Values

Missing values in time-series data and metrology sample data can affect the statistical summarization. For time-series data, missing values are the result of communication failures with the equipment. Missing values from a metrology sample are generally caused by faulty measurement equipment.

Missing values in time-series data result in high data latency values and are easily detected. When they occur, missing values affect approximately one out of one hundred data points in a process record. Since Ion Implant processes tend to be stable during single processing step, we will adopt the method of replacing the missing value with the average from the process.

Since the metrology equipment reports only summarized data, there is no detection for missing values from a metrology sample. We will rely on the outlier detection methods described above to filter any non-standard measurements.

3.3 DATA PREPROCESSING

Now that the data set has been filtered for errors and outliers, the data set must be preprocessed before using it as input or output for model creation. The three main preprocessing types are framing, normalization, and summarization of the time-series data. Framing is accomplished by filtering the data to select only the portions that are valid for analysis. Normalization scales the data so that it can be properly compared despite scale differences. Finally, time-series data needs to be statistically summarized for many model applications.

3.3.1 Framing

Framing is used as another method of removing noise from process data. Data collection systems frequently begin and end their data collection using equipment states that are useful for capturing the entire process and not necessarily the portion of interest for analysis. For example, an Ion Implant process may consist of several equipment steps: (1) Wheel spin up, (2) Beam tune, (3) Implant, and (4) Wheel spin down. For our purposes, only the Implant step is useful for modeling. Framing can be used to impose logic on the time-series data to eliminate the data from steps 1, 2, and 4. Figure 3.11 shows the Beam Current data for an entire process record with the Implant Status overlaid for comparison. Knowing that an Implant State of 60 means that the equipment is actively implanting the wafers, we can frame the data accordingly. Figure 3.12 shows the results of applying framing to the Beam Current signal. The primary benefit of framing is increased accuracy of the statistical summarization performed on the time series data. All low dose qualification data is framed by the Implant State for this analysis.

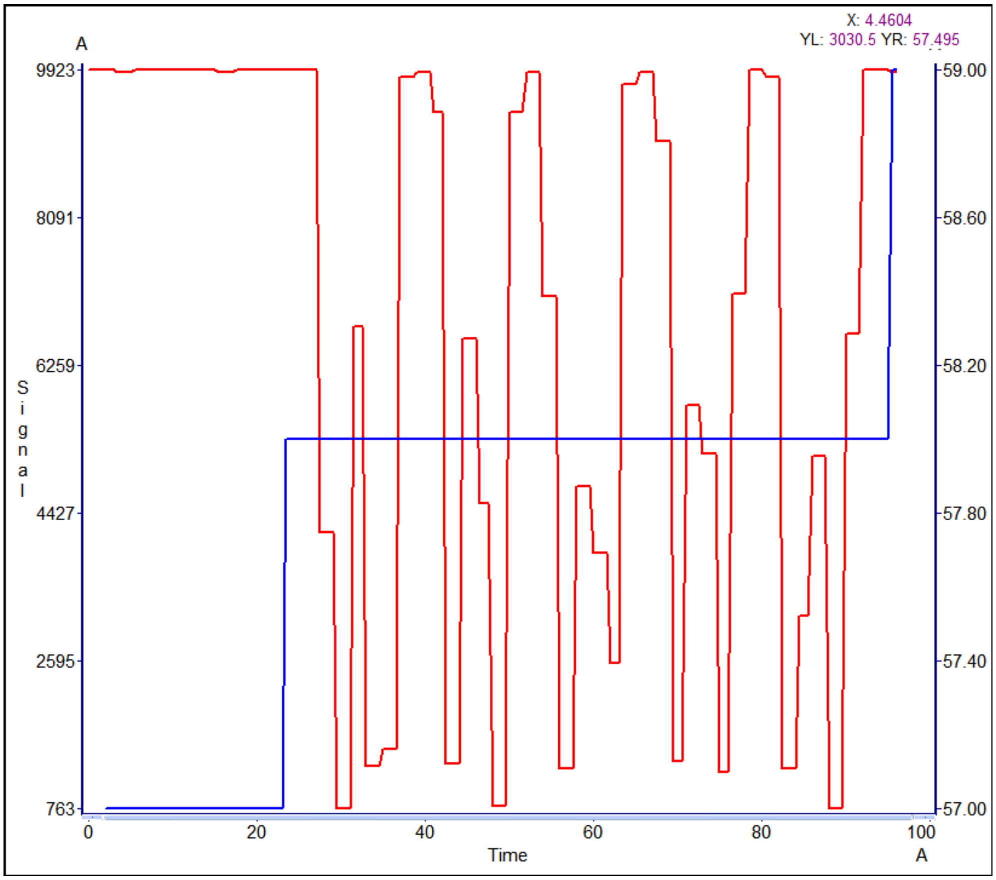


Figure 3.11: Implant Beam Current (red) and Implant Status (blue).

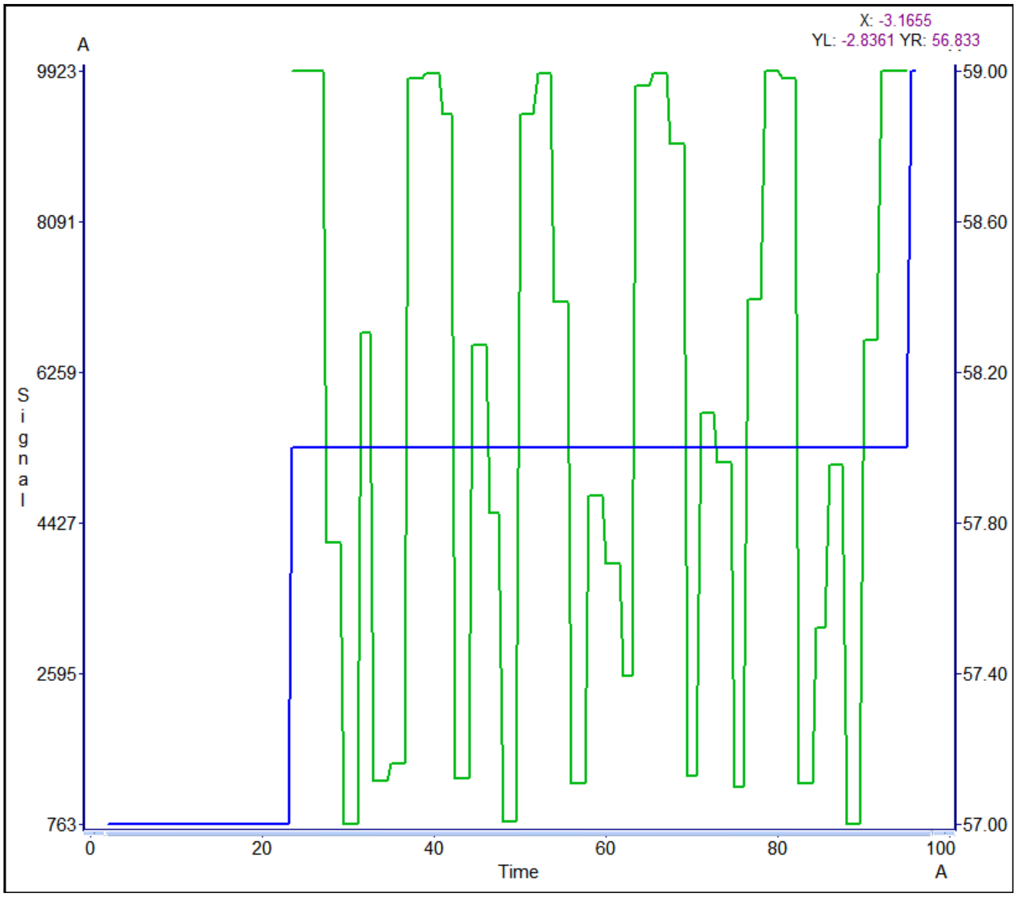


Figure 3.12: Framed Implant Beam Current (green) and Implant Status (blue).

3.3.2 Summarization

The standard arithmetic average is applied to each time-series data variable after it has been framed by Implant State.

3.3.3 Normalization

Data normalization is an important preprocessing procedure. The purpose of normalization is to scale the time-series data averages so that data of different scales can be effectively compared. A standard autoscaling method is used on this data set, see equation 3.2 [18]. First, subtract each variable by its sample mean. Second, divide the resulting variable by its sample standard deviation. This will result in data that has been autoscaled to zero mean and unit variance.

$$x_i' = \frac{x_i - \bar{x}}{sd} \qquad (3.2)$$

Chapter 4: Variable Selection

During Ion Implant processing, a significant amount of data can be collected from the equipment. This generally results in a surplus of available data for each processed wafer. Deciding which variables should be used to predict physical metrology during model build is challenging, particularly when model building is known to be extremely dependent on the variables selected [11], [12], [13]. There are several modeling techniques that are capable of handling co-linearity among the variables such as Principal Components Regression (PCR), Partial Least Squares Regression (PLSR), and neural net, however, elimination of uncorrelated variables would improve the prediction accuracy of the resulting model. This chapter examines three different methods for variable selection: Principal Components Analysis (PCA), Correlation Methodology, and Stepwise Regression.

4.1 PRINCIPAL COMPONENT ANALYSIS

4.1.1 Overview of PCA

Principal Component Analysis [20] is a mathematical method used to transform a set of correlated variables into new uncorrelated variables known as principal components (PCs). Each PC is a linear combination of the original variables. They are then arranged in order of importance based on the amount of variance of the original data set that is captured by each PC.

Mathematically, PCA is an eigenvector decomposition of the covariance or correlation matrix of the process variables. For a data matrix **X** with *m* sample rows and *n* variable columns, the covariance matrix of **X** is

$$\mathrm{cov}(\mathrm{X}) = \frac{\mathrm{X}^T \mathrm{X}}{m-1} \tag{4.1}$$

when the variable columns of **X** have been autoscaled, i.e., adjusted to zero mean and unit variance. PCA decomposes the data matrix **X** into the outer products of two vectors, scores $\mathbf{t}_\mathrm{i}$ and loadings $\mathbf{p}_\mathrm{i}$, and a residual matrix **E**

$$\mathrm{X} = t_1 p_1^T + t_2 p_2^T + \ldots + t_l p_l^T + \mathrm{E} \tag{4.2}$$

where *l* is less than or equal to the smallest dimension of the data matrix **X**. The $\mathbf{p}_\mathrm{i}$ loading vectors are the eigenvectors of the covariance matrix.

$$\mathrm{cov}(\mathrm{X})\, p_i = \lambda_i \, p_i \tag{4.3}$$

If we define $\mathbf{t}_\mathrm{i}$ to be the i$^{\mathrm{th}}$ column of T in the training set, then the following properties can be shown [18]

1. $t_i^T t_k = 0;$ for $\mathrm{i} \neq \mathrm{k}$.
2. $p_i^T p_k = 0;$ for $\mathrm{i} \neq \mathrm{k}$.
3. $p_i^T p_k = 1;$ for $\mathrm{i} = \mathrm{k}$.

The score vector $\mathbf{t}_\mathrm{i}$ is the linear combination of the original data matrix **X** variables defined by $\mathbf{p}_\mathrm{i}$

$$\mathrm{X} p_i = t_i\,. \tag{4.4}$$

Overall, the loadings represent how the original variables are combined to make the principal components. The scores represent the original data projected onto the new uncorrelated variables, while **E** represents the data that is not represented in the model.

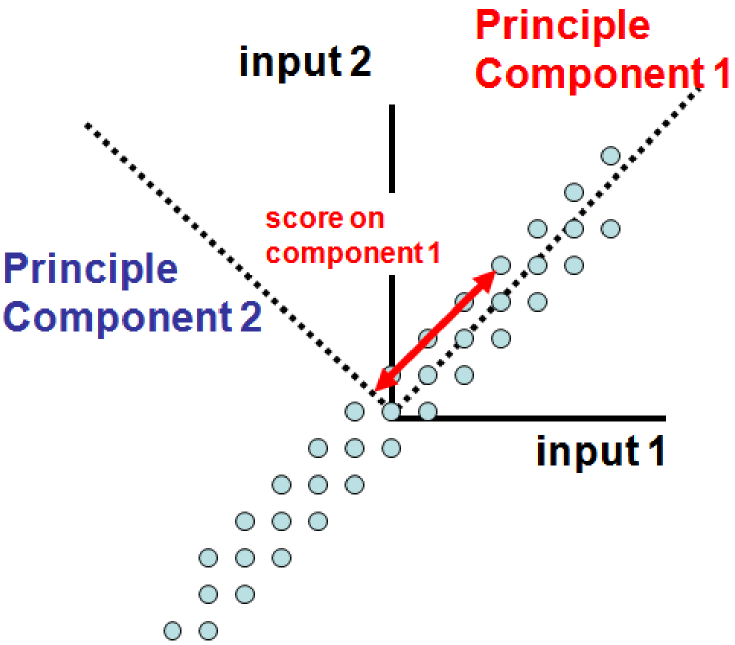


Figure 4.1: Plot of input 2 vs. input 1 with the first two principal components overlaid. Notice that PC 1 describes the direction of the greatest variation in the data set.

4.1.2 PCA for Variable Selection

Although PCA is generally used as a dimensionality reduction method, it can also be used as a variable selection technique by examining the loading vectors for the first few principal components [6], [14]. Traditionally, PCA is interpreted such that a high correlation between PC1 and a variable indicates that the variable is associated with the direction of the maximum amount of variation in the data set. Alternatively, if a variable does not correspond to any of the principal PC axes, this suggests that the variable has little effect on the distribution of the data set. Therefore, Principle Component Analysis can indicate which variables in a data set are important and which ones may be eliminated to simplify the overall analysis.

4.1.3 Application of PCA for the LDM Data Set

When a Principal Components Analysis was performed on the LDM data set, the first PC explained 99.9226% of the variance in the data matrix **X**, Figure 4.2 and Table 4.1. This indicates that at least some of the variables are highly correlated.

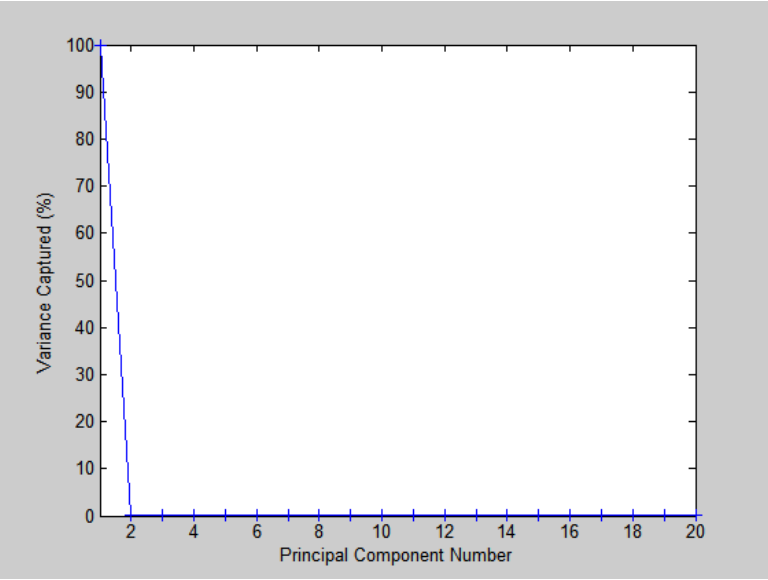


Figure 4.2: Plot of percentage of variance captured by Principal Component Number for the LDM data set.

PC	**Variance Captured (%)**
1	99.9226
2	0.059362
3	0.006958
4	0.005015

Table 4.1: Percentage of variance captured by Principal Component Number for the LDM data set.

The loadings of PC 1 indicate that a significant number of variables can be eliminated. Choosing a cutoff of 0.02 for the loading value reduces our variable count from 47 down to 14, Figure 4.3 and Table 4.2.

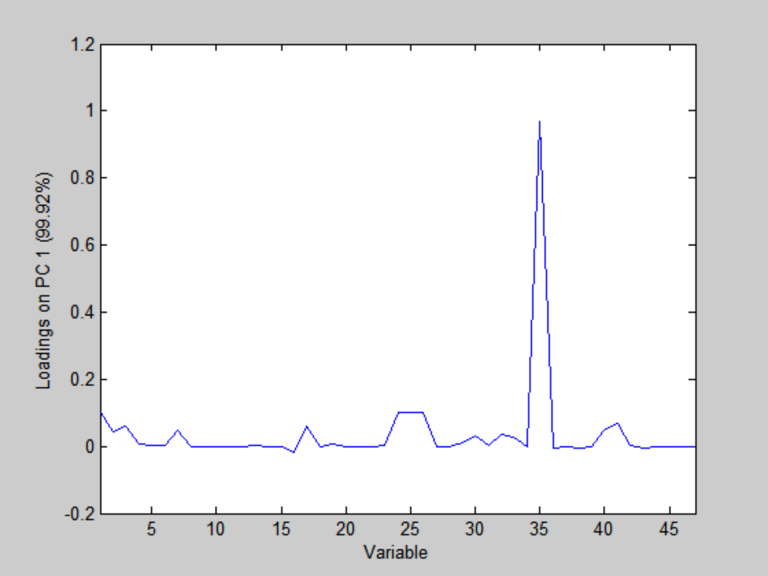


Figure 4.3: Plot of the loadings on PC 1 by Variable for the LDM data set.

Variable	Loadings on PC 1
Spin Speed During Process AVERAGE	0.96835
AMag Current During Process AVERAGE	0.10215
Phase C Voltage During Process AVERAGE	0.10054
Phase B Voltage During Process AVERAGE	0.1005
Phase A Voltage During Process AVERAGE	0.099994
TEM Actual Energy Feedback During Process AVERAGE	0.067895
AMU During Process AVERAGE	0.059393
Implant Time During Process MAXIMUM	0.058889
TEM Actual During Process AVERAGE	0.047353
Extraction (PreA) Volts Actual During Process AVERAGE	0.047251
AMag Voltage During Process AVERAGE	0.042422
Source Arc Current During Process AVERAGE	0.037166
Pre Accel Voltage During Process AVERAGE	0.031754

Variable	Loadings on PC 1
Source Arc Voltage During Process AVERAGE	0.02791
Pre Accel Current During Process AVERAGE	0.010649
Lateral Actual During Process AVERAGE	0.0079084
Beam Current During Process AVERAGE	0.0049459
Scan Speed During Process AVERAGE	0.0041577
Number of Scans During Process MAXIMUM	0.0034295
C of G During Process AVERAGE	0.0030762
Tetrode Current Actual During Process AVERAGE	0.0024075
Extraction (PreA) Current During Process AVERAGE	0.0017655
FG Min Charge V During Process AVERAGE	0.0010081
Transfer Ratio During Process AVERAGE	0.00054513
FG Tube Current During Process AVERAGE	0.00053437
Plasma Gun Emissions During Process AVERAGE	0.00051372
Suppression Current During Process AVERAGE	0.00050544
MRS Position During Process AVERAGE	0.00036963
MRS Gap Demand During Process AVERAGE	0.00033408
Dose Ratio Average	0.00025999
FG Filament Current During Process AVERAGE	0.00012006
FG Arc Voltage During Process AVERAGE	3.19E-05
Wheel Current During Process AVERAGE	2.54E-05
Inert Flow During Process AVERAGE	9.49E-06
FG Arc Current During Process AVERAGE	4.53E-06
FG Filament Voltage During Process AVERAGE	2.42E-06
Source Pressure During Process AVERAGE	8.35E-10
Magnet Pressure During Process AVERAGE	7.85E-10
Target Pressure During Process AVERAGE	5.81E-10
Post A Decel During Process AVERAGE	-1.97E-07
FG Tube Voltage During Process AVERAGE	-2.10E-05
Tilt Angle During Process AVERAGE	-0.0017148
FG Max Charge V During Process AVERAGE	-0.0028507
Tetrode Voltage Actual During Process AVERAGE	-0.0053249
Src Suppression Voltage During Process AVERAGE	-0.0058061
Suppression Voltage Actual During Process AVERAGE	-0.0059438
Focus Voltage During Process AVERAGE	-0.015797

Table 4.2: Loadings on PC 1 by variable for the LDM data set.

4.2 VARIABLE RANKING WITH CORRELATION COEFFICIENTS

4.2.1 Overview of Correlation Coefficients

In statistics, correlation is a measure of the linear dependence between two variables X and Y. In particular, Pearson's correlation coefficient, ρ, is defined as the covariance of two variables divided by the product of their standard deviations

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y}, \tag{4.5}$$

where μ_X and μ_Y are the mean values of the variables.

A correlation coefficient can range from -1 to 1 depending on the strength of the linear relationship between X and Y. A coefficient of 1 indicates a perfect linear relationship between X and Y where Y increases as X increases. A coefficient of -1 also indicates a perfect linear relationship, however, Y decreases as X increases. A coefficient of 0 indicates that there is no linear relationship between X and Y.

4.2.2 Correlation Coefficients for Variable Selection

Correlation coefficients can be used for variable selection by evaluating the linear correlations between the data variables and the metrology results [6]. The coefficients are calculated for each data variable-metrology pair and are ranked in order of increasing linear relationship. The data variables that are most correlated with the metrology results are used as inputs to VM models.

4.2.3 Application of Correlation Coefficients to the LDM Data Set

Correlation coefficients were calculated between the implant variables and the LDM metrology results, Figure 4.4 and Table 4.3. The magnitude of the largest coefficient was 0.5402, indicating weak linear relationships between the data variables and the LDM metrology results at best. Choosing a cutoff of ±0.3 for the correlation coefficient reduces our variable count from 47 down to 9.

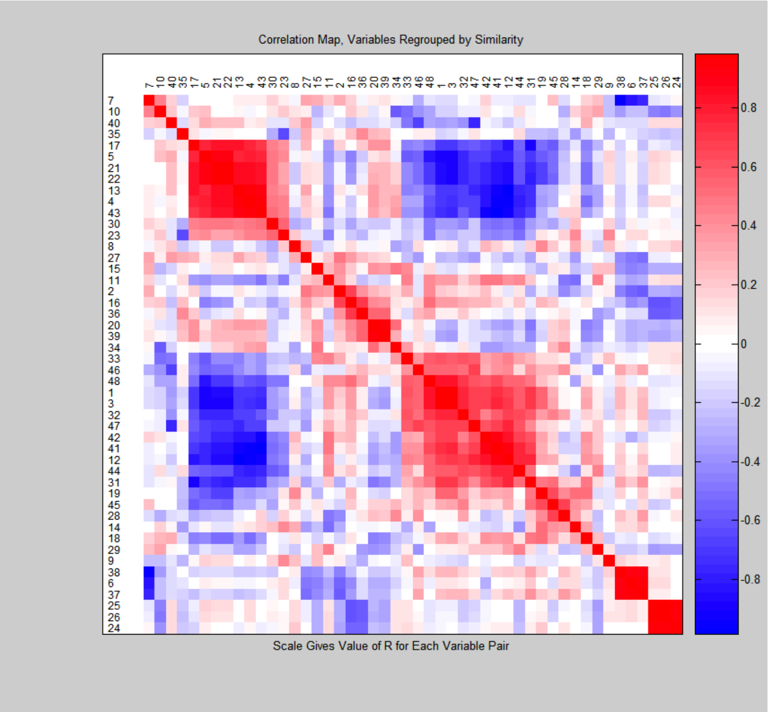


Figure 4.4: Correlation map for the LDM data set, variable 48 is the metrology output data.

Variable	Correlation to LDM Metrology data
AMag Current During Process AVERAGE	0.5402
AMU During Process AVERAGE	0.5327
Source Arc Current During Process AVERAGE	0.3922
Focus Voltage During Process AVERAGE	0.3917
TEM Actual Energy Feedback During Process AVERAGE	0.3177
AMag Voltage During Process AVERAGE	0.2875
Wheel Current During Process AVERAGE	0.2693
Src Suppression Voltage During Process AVERAGE	0.2606
Source Arc Voltage During Process AVERAGE	0.2349
Tilt Angle During Process AVERAGE	0.211

Variable	**Correlation to LDM Metrology data**
Scan Speed During Process AVERAGE	0.2109
Tetrode Current Actual During Process AVERAGE	0.208
Dose Ratio Average	0.1984
FG Filament Voltage During Process AVERAGE	0.196
Source Pressure During Process AVERAGE	0.1907
FG Tube Voltage During Process AVERAGE	0.1701
FG Max Charge V During Process AVERAGE	0.1247
Suppression Voltage Actual During Process AVERAGE	0.0669
Plasma Gun Emissions During Process AVERAGE	0.065
FG Arc Voltage During Process AVERAGE	0.0624
FG Arc Current During Process AVERAGE	0.048
Magnet Pressure During Process AVERAGE	0.0203
Phase C Voltage During Process AVERAGE	0.0093
Post A Decel During Process AVERAGE	0.0078
Phase B Voltage During Process AVERAGE	-0.0024
Extraction (PreA) Current During Process AVERAGE	-0.0106
Spin Speed During Process AVERAGE	-0.0174
Suppression Current During Process AVERAGE	-0.02
Phase A Voltage During Process AVERAGE	-0.0299
Extraction (PreA) Volts Actual During Process AVERAGE	-0.0352
Inert Flow During Process AVERAGE	-0.0426
Transfer Ratio During Process AVERAGE	-0.0673
Target Pressure During Process AVERAGE	-0.0692
Lateral Actual During Process AVERAGE	-0.0808
FG Min Charge V During Process AVERAGE	-0.1189
Pre Accel Current During Process AVERAGE	-0.1227
TEM Actual During Process AVERAGE	-0.1358
Implant Time During Process MAXIMUM	-0.16
Pre Accel Voltage During Process AVERAGE	-0.1893
FG Filament Current During Process AVERAGE	-0.2233
Beam Current During Process AVERAGE	-0.245
Number of Scans During Process MAXIMUM	-0.245
FG Tube Current During Process AVERAGE	-0.274
Tetrode Voltage Actual During Process AVERAGE	-0.3166
MRS Position During Process AVERAGE	-0.37
MRS Gap Demand During Process AVERAGE	-0.4142
C of G During Process AVERAGE	-0.5226

Table 4.3: Correlation between the data variables and the LDM metrology measurements for the LDM data set.

4.3 STEPWISE REGRESSION

4.3.1 Overview of Stepwise Regression

Stepwise regression, Efroymson's algorithm, is a modified forward-selection technique where a linear regression model is produced from input variables that are chosen automatically by some criteria [14]. Stepwise regression begins by using the forward selection method for variable addition to the model. It is then followed by a backward elimination to see if any of the previously selected variables can be removed from the model.

4.3.2 Forward Selection

The forward selection method begins with an empty model, i.e. no variables. Variables are then added to the model one at a time based on how correlated they are with the output variable. If we let RSS_p denote the residual sum of squares with p variables in the model and a constant, and RSS_{p+1} denote the residual sum of squares that is obtained by adding the next variable to the model, then the ratio

$$F_{RSS} = \frac{RSS_p - RSS_{p+1}}{RSS_{p+1} / (n - p - 2)}, \tag{4.6}$$

can be compared to a threshold F_s (F to select) value. If F_{RSS} is greater than F_s, the new variable is added to the model. Note that the F_{RSS} ratio for both forward selection and backward elimination is approximated by the F-distribution [14].

4.3.3 Backward Elimination

With p variables and a constant in the model, let RSS_{p-1} denote the residual sum of squares that is obtained after deleting any variable from the model. The ratio

$$F_{RSS} = \frac{RSS_{p-1} - RSS_p}{RSS_p / (n - p - 1)}, \tag{4.7}$$

can be compared to a threshold F_e (F to eliminate) value. If F_{RSS} is less than F_e, the variable is removed from the model.

4.3.4 Stepwise Regression for Variable Selection

The data variables are added individually to a model, and the residual sum of squares is calculated. When it is determined that a variable should be added to the model, it is included, and all other variables in the model are retested against the new result. Current variables are then subject to removal based on the backwards elimination criteria.

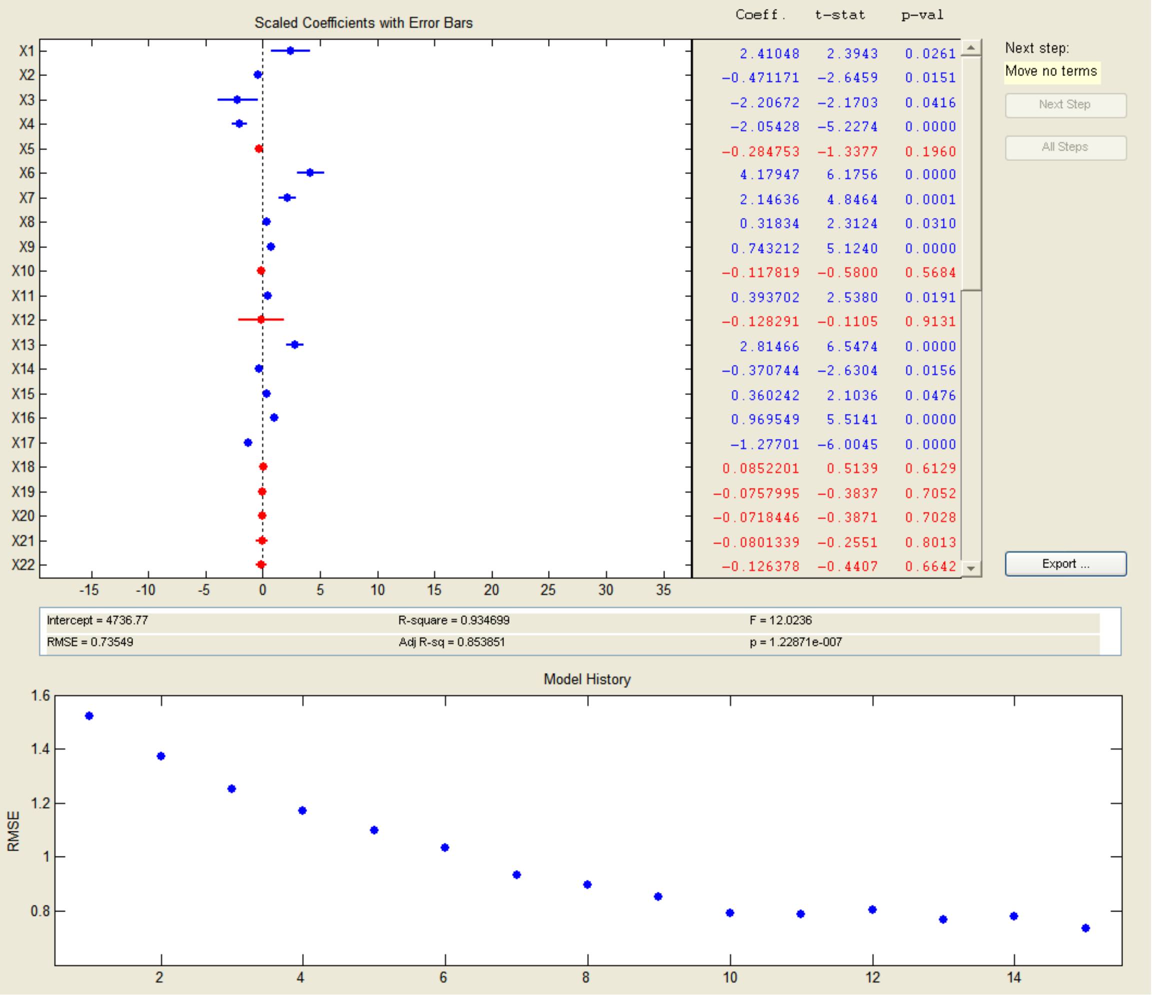
The stepwise regression will continue as long as the new RSS computation, denoted by RSS^*, meets the following criteria

$$RSS_p^* \le RSS_p \frac{1 + F_e / (n - p - 2)}{1 + F_s / (n - p - 2)}. \tag{4.8}$$

The stepwise regression will stop when variables cannot be added or removed based on the criteria.

4.3.5 Application of Stepwise Regression to the LDM Data Set

Stepwise regression was performed on the LDM data set, Figure 4.5 and Table 4.4. The resulting regression metrics are as follows: RMSE = 0.73549, R-square = 0.934699, Adj R-square = 0.853851. According to these results, approximately 85-93% of the variation in the LDM metrology values can be explained by the selected data variables. P-values of 0.10 were selected for both the forward and backward portions of the stepwise regression. This resulted in a reduction in variable count from 47 down to 25.



```
   Coeff.      t-stat     p-val
   2.41048     2.3943    0.0261
 -0.471171    -2.6459    0.0151
  -2.20672    -2.1703    0.0416
  -2.05428    -5.2274    0.0000
 -0.284753    -1.3377    0.1960
   4.17947     6.1756    0.0000
   2.14636     4.8464    0.0001
   0.31834     2.3124    0.0310
  0.743212     5.1240    0.0000
 -0.117819    -0.5800    0.5684
  0.393702     2.5380    0.0191
 -0.128291    -0.1105    0.9131
   2.81466     6.5474    0.0000
 -0.370744    -2.6304    0.0156
  0.360242     2.1036    0.0476
  0.969549     5.5141    0.0000
  -1.27701    -6.0045    0.0000
 0.0852201     0.5139    0.6129
-0.0757995    -0.3837    0.7052
-0.0718446    -0.3871    0.7028
-0.0801339    -0.2551    0.8013
 -0.126378    -0.4407    0.6642
```

```
Intercept = 4736.77          R-square = 0.934699          F = 12.0236
RMSE = 0.73549               Adj R-sq = 0.853851          p = 1.22871e-007
```

Figure 4.5: MATLAB® workspace showing the results of the stepwise selection method.

Variable	**Matrix Column Assignment**
AMag Current During Process AVERAGE	1
AMag Voltage During Process AVERAGE	2
AMU During Process AVERAGE	3
Beam Current During Process AVERAGE	4
Extraction (PreA) Current During Process AVERAGE	6
Extraction (PreA) Volts Actual During Process AVERAGE	7

FG Arc Current During Process AVERAGE	8
FG Arc Voltage During Process AVERAGE	9
FG Filament Voltage During Process AVERAGE	11
FG Min Charge V During Process AVERAGE	13
FG Tube Current During Process AVERAGE	14
FG Tube Voltage During Process AVERAGE	15
Focus Voltage During Process AVERAGE	16
Implant Time During Process MAXIMUM	17
Number of Scans During Process MAXIMUM	23
Phase B Voltage During Process AVERAGE	25
Post A Decel During Process AVERAGE	28
Pre Accel Current During Process AVERAGE	29
Source Arc Current During Process AVERAGE	32
Source Pressure During Process AVERAGE	34
Suppression Current During Process AVERAGE	37
Suppression Voltage Actual During Process AVERAGE	38
TEM Actual During Process AVERAGE	40
TEM Actual Energy Feedback During Process AVERAGE	41
Tilt Angle During Process AVERAGE	44

Table 4.4: List of data variables retained in the model using stepwise selection for the LDM data set.

Chapter 5: Modeling Methods

There are numerous empirical modeling methods available for VM applications. The purpose of this chapter is to compare several key methods, MLR, PCR, PLS, BPNN, and RBFN. Each model has advantages and disadvantages based on model/computational complexity and the nature of the data used for model generation. MLR, for instance, is not suitable for highly dimensional or collinear datasets. PCR can handle these issues, however, only linear relationships can be modeled and output variables are not even considered. While still a linear model, PLS has the advantage of tilting the latent variables towards the output. The Neural Networks BPNN and RBFN are non-linear models, however they have problems with over fitting if the training sample size is too small. Additionally, BPNN in particular is complex to train. This chapter begins with a discussion of the linear models, MLR, PCR, and PLS, and concludes with a presentation of the non-linear models, BPNN and RBFN.

5.1 MULTIPLE LINEAR REGRESSION

5.1.1 Multiple Linear Regression Overview

Multiple Linear Regression (MLR) [21] is a linear method that attempts to model the relationship between two or more explanatory variables to a single response variable by fitting a linear equation to the observed data. MLR extends the idea of simple linear regression in that the response variable y_i is a straight-line function of multiple explanatory variables x_i instead of just a single explanatory variable. This relationship can be written for k explanatory variables x_i as

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_k x_{i,k} + \varepsilon_i , \tag{5.1}$$

where the β_is are the regression parameters, y_is are the measured outputs, and the ε_is are random errors. Alternatively, the equation can be written in matrix notation as

$$y = X\beta + \varepsilon \, . \tag{5.2}$$

The solution to the problem is a vector ***b*** which estimates the unknown vector of parameters, **β**. The least squares solution is

$$b = \hat{\beta} = (X^T X)^{-1} X^T y \tag{5.3}$$

Multiple Linear Regression is computationally quick to train, however there are two distinct issues that will often cause the approach to fail: 1) the collinearity of the data matrix **X** can lead to unstable matrix inversions, and 2) **X** may contain fewer samples than variables leading to an underdetermined situation.

5.1.2 MLR Model Build Using the Training Data Set

Multiple Linear Regression was used to build three models based on the variables selected by each of the three techniques: PCA, Pearson Correlation ranking, and Stepwise Selection. These are referred to as PCA-MLR, PCorr-MLR, and Step-MLR respectively. Figure 5.1 shows the model residuals from each of the 35 Runs in the training data set by variable selection technique. It is interesting to note that each model identified Run number 10 as an outlier based on the residual confidence intervals. Based on the R^2 metric, it can be seen that the Step-MLR model explains over 95% of the variability in the observations and clearly outperforms the PCA-MLR and PCorr-MLR models with R^2 values of 0.4639 and 0.5211 respectively.

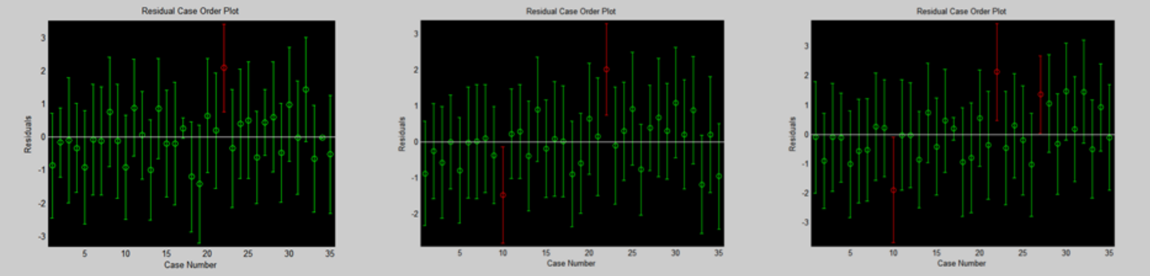


Figure 5.1: Plot of residuals with 95% confidence intervals for PCA-MLR, PCorr-MLR, and Step-MLR for the LDM data set. Run numbers 10, 22, and 27 are identified as outliers.

Variable Selection Technique	**R^2 Metric**
PCA	0.4639
Pearson Correlation	0.5211
Stepwise Selection	0.9558

Table 5.1: MLR modeling results for different variable selection techniques for the LDM data set.

5.2 PRINCIPAL COMPONENTS REGRESSION

5.2.1 Principal Components Regression Overview

In Principal Components Regression (PCR) [21], a principal component analysis is performed on the data matrix **X** and the resulting PCs are used as the input to the regression model. The orthogonality of the PCs and dimensionality reduction of the data set **X** solve the two major computational issues with MLS; the collinearity of the data matrix **X** and the underdetermined data matrix **X**. Compared to MLR, PCR is much more stable, primarily due to the orthogonality of the PCs. The underdetermined data matrix **X** issue is resolved by the fact that the maximum number of PCs is equal to the lesser of the number of response variables and the number of data samples.

An important part of PCR is determining the optimal number of PCs to retain in the model as they relate to prediction accuracy. Typically, the optimal number of PCs to retain is determined by cross-validation, where the data matrix **X** is subdivided into training and test sets. The prediction residual error (PRESS) on the test set is then computed as a function of the number of PCs retained in the model. The procedure is repeated a number of times, using different test sets for each iteration so that each sample in the data set is part of a test set at least once. The composite prediction error is computed as a function of the number of PCs over all test sets and is used to determine the optimal number of PCs to include in the model. Note that if all of the PCs are retained in the model, the result is identical to that for MLR when there are more samples than variables. The convergence of the PCR model to the MLR model under these conditions provides a good way to quickly verify both computational code and model results.

5.2.2 Principal Components Regression Algorithm

1. Perform a PCA on the data matrix **X**.

$$X = TP^T \tag{5.4}$$

2. Perform a Multiple Linear Regression of **Y** on the A major principal components using the principal components $\mathbf{T}_{[A]}$. The solution to the problem is a vector ***b*** which estimates the unknown vector of parameters, $\boldsymbol{\beta}$. The least squares solution is

$$b = \hat{\beta} = P_{[A]}(T_{[A]}^T T_{[A]})^{-1} T_{[A]}^T y \,. \tag{5.5}$$

5.2.3 PCR Model Build Using the Training Data Set

Principal Components Regression was used to build three models based on the variables selected by each of the three techniques used. These are referred to as PCA-PCR, PCorr-PCR, and Step-PCR respectively. Figure 5.2 shows plots of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the PCA-PCR model of the LDM data set. The optimal number of PCs to retain in the model is a tradeoff between captured x-variance and the error introduced by retaining too many PCs. Nine PCs captured 95.01% of the variance in the data matrix **X** while minimizing the RMSE.

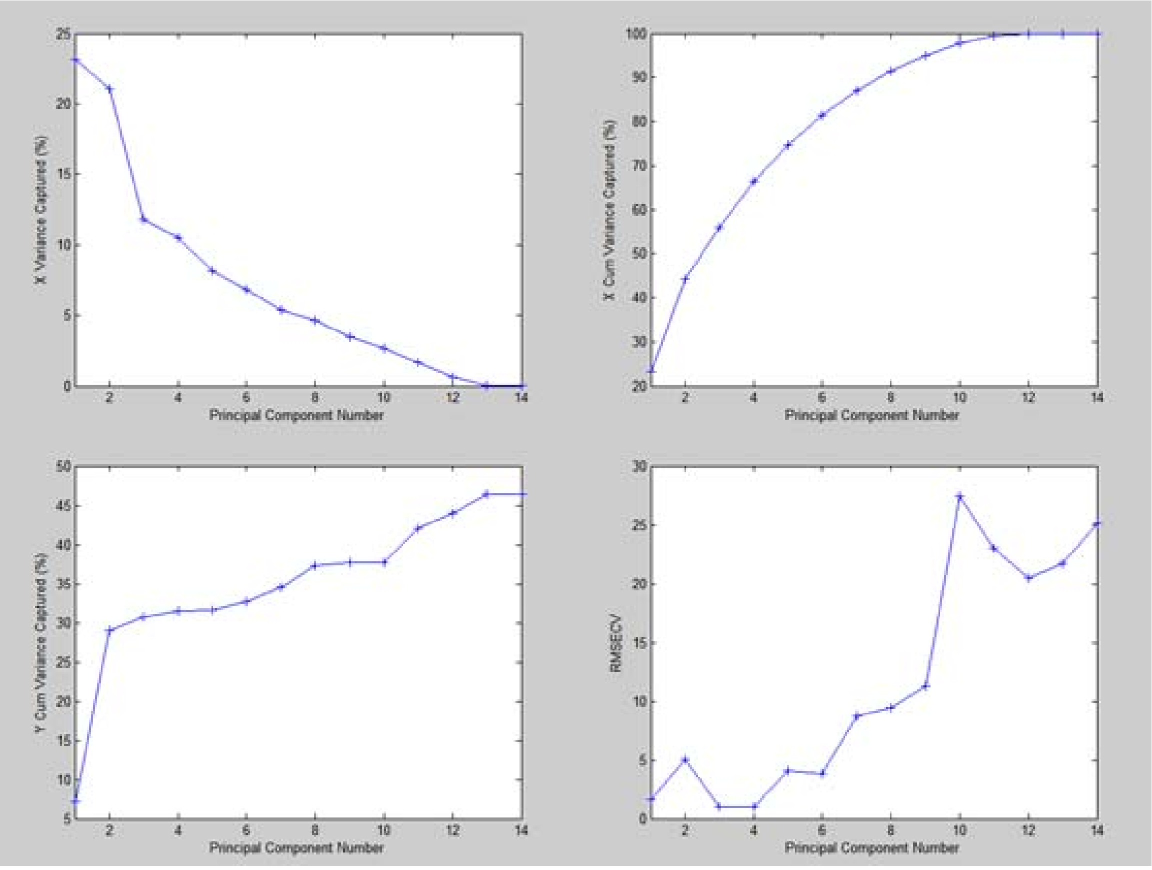


Figure 5.2: Plot of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the PCA-PCR model of the LDM data set. The optimal PC number based on RMSECV is 9.

Figure 5.3 shows plots of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the PCorr-PCR model of the LDM data set. Six PCs captured 95.98% of the variance in the data matrix **X** while minimizing the RMSE.

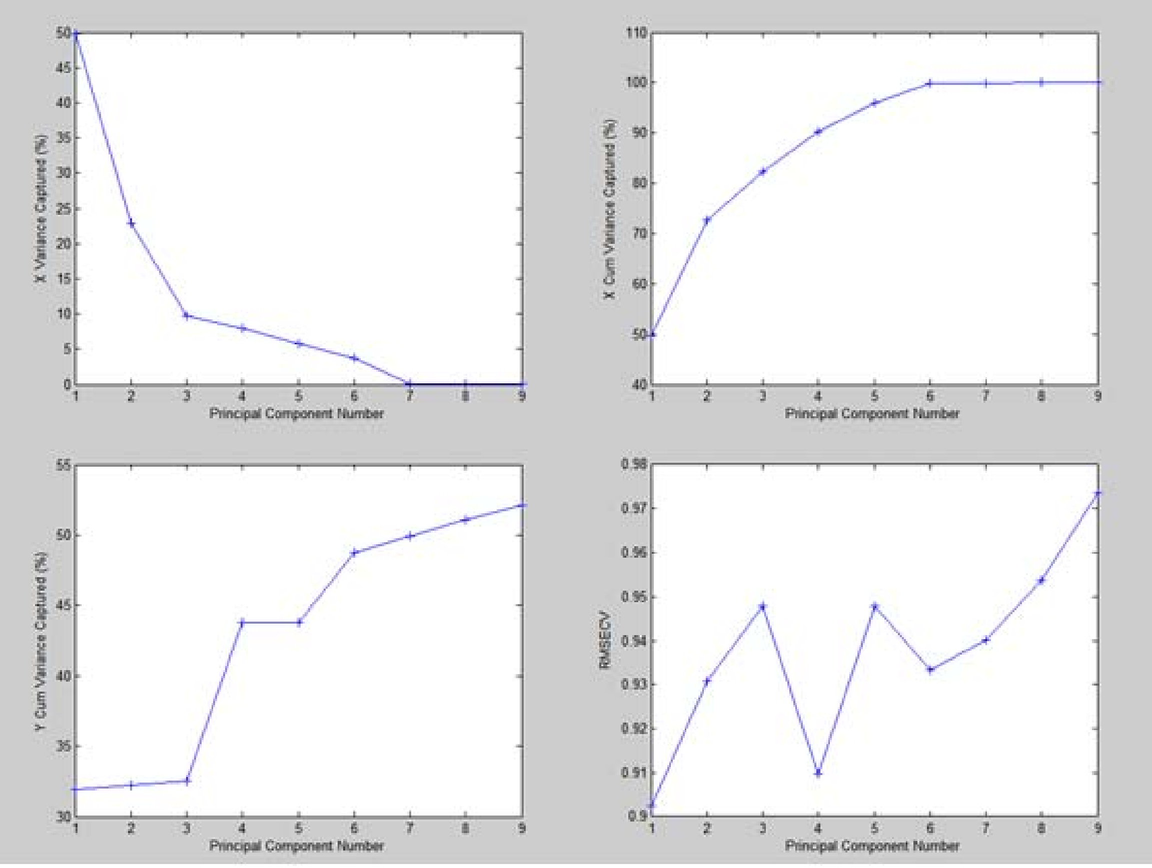


Figure 5.3: Plot of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the PCorr-PCR model of the LDM data set. The optimal PC number based on RMSECV is 6.

Figure 5.4 shows plots of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the Step-PCR model of the LDM data set. Sixteen PCs captured 95.68% of the variance in the data matrix **X** while minimizing the RMSE. From the RMSECV plot, a selection of 12 PCs would have better minimized the error, but at a sacrifice of captured x-variance.

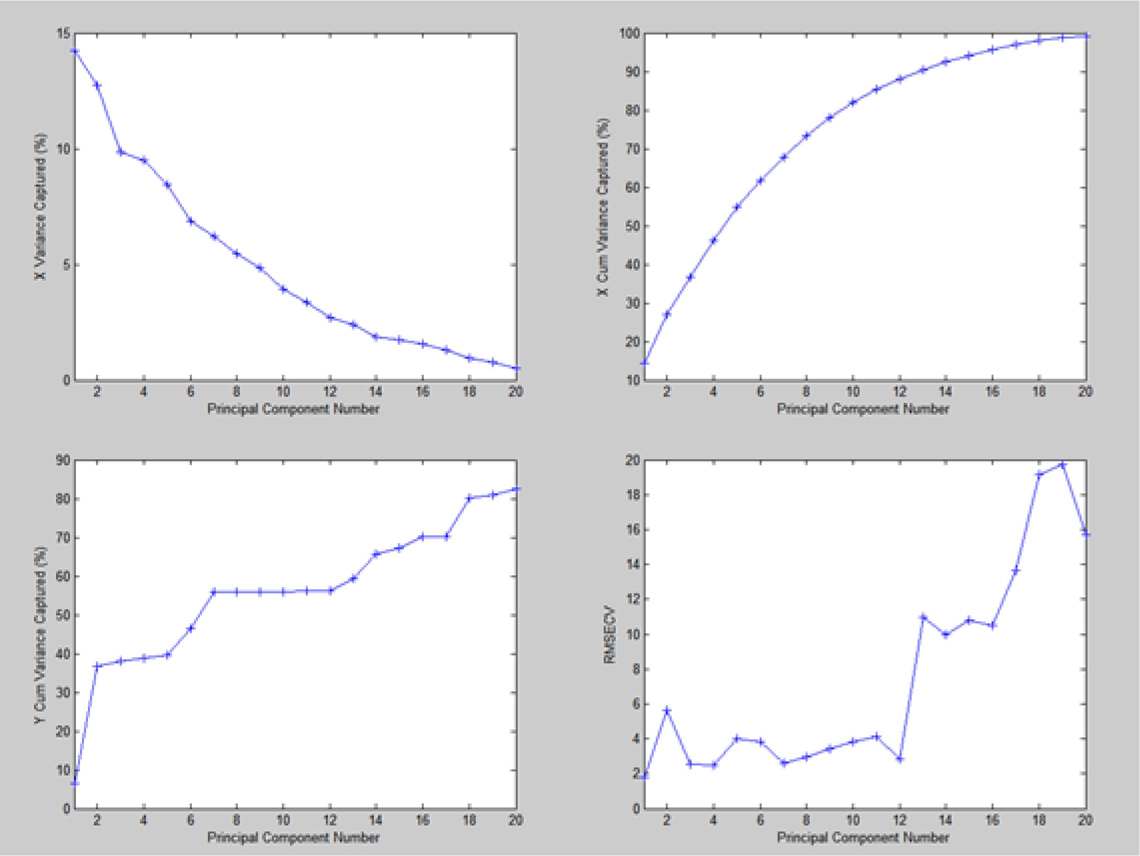


Figure 5.4: Plot of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the Step-PCR model of the LDM data set. The optimal PC number based on RMSECV and R^2 is 16.

Based on the R^2 metric, Table 5.2, it can be seen that the Step-PCR model explains over 70% of the variability in the observations and moderately outperforms the PCA-PCR and PCorr-PCR models with R^2 values of 0.3765 and 0.4381 respectively.

Variable Selection Technique	**R^2 Metric (all PCs)**	**R^2 Metric (optimal PCs)**
PCA	0.4639 (14)	0.3765 (9)
Pearson Correlation	0.5211 (9)	0.4381 (6)
Stepwise Selection	0.9558 (25)	0.7007 (16)

Table 5.2: PCR modeling results for different variable selection techniques for the LDM data set. The R^2 metric is reported for the optimal number of PCs retained in the model, as well as for all PCs retained in the model.

5.3 PARTIAL LEAST SQUARES REGRESSION

5.3.1 Partial Least Squares Regression Overview

Much like PCA, Partial Least Squares (PLS) attempts to find factors, or latent variables, that capture the maximum amount of variance in a data set. PLS, however, also includes the objective of achieving correlation, the result of which is maximizing the covariance between the data matrix **X** and the output matrix **Y**. One major difference between PCA and PLS is that PLS reduces the dimensions of both **X** and **Y** matrices simultaneously to find the most highly correlated factors for **X** and **Y**.

There are a number of ways to calculate the PLS model parameters, however, non-iterative Partial Least Squares (NIPALS) and statistically inspired modification of Partial Least Squares (SIMPLS) [22] are commonly used [21]. The SIMPLS regression algorithm is used here primarily because of its computational speed compared to the NIPALS algorithm. For univariate Y, SIMPLS produces the exact same result as NIPALS. Additionally, the SIMPLS algorithm actually maximizes the covariance criterion, while NIPALS has been shown to perform less than optimally in this respect

[21]. Regardless of the method of calculation, there are two relationships considered in PLSR: 1) Outer: the linear relationships between data variables X and the latent variables, 2) Inner: the set of univariate regressions between each latent variable pair, up to the full rank of the model.

As in PCR, the optimal number of latent variables to retain in the model can be determined by performing cross-validation and computing the prediction residual error (PRESS).

5.3.2 SIMPLS Regression Algorithm

1. Autoscale and decompose the **X** and **Y** matrices. **T** and **U** become the scores for **X** and **Y** respectively, and are used to show the inner relationship. Similarly, **P** and **Q** become the loadings for **X** and **Y** respectively. The matrices **E** and **F** contain the residuals.

$$X = TP^T + E = \sum t_h p_h^T + E \tag{5.6}$$

$$Y = UQ^T + F = \sum u_h q_h^T + F \tag{5.7}$$

2. For each h=1,…,c

$$A_0 = X^T Y, \tag{5.8}$$

$$M_0 = X^T X, \tag{5.9}$$

$$C_0 = I. \tag{5.10}$$

3. Calculate q_h, the dominant eigenvector of $A_h^T A_h$.
4. Calculate the following and store the resulting w_h into **W** as a column. The **W** matrix is an additional set of vectors known as weights.

$$w_h = A_h q_h \tag{5.11}$$

$$c_h = w_h^T M_h w_h \tag{5.12}$$

$$w_h = \frac{w_h}{\sqrt{c_h}} \tag{5.13}$$

5. Calculate the following and store the resulting p_h into **P** as a column.

$$p_h = M_h w_h \tag{5.14}$$

6. Calculate the following and store the resulting q_h into **Q** as a column.

$$q_h = A_h^T w_h \tag{5.15}$$

7. Calculate the following

$$v_h = C_h p_h , \tag{5.16}$$

$$v_h = \frac{v_h}{\|v_h\|} . \tag{5.17}$$

8. Finally, calculate the matrices for the next iteration

$$C_{h+1} = C_h - v_h v_h^T , \tag{5.18}$$

$$M_{h+1} = M_h - p_h p_h^T , \tag{5.19}$$

$$A_{h+1} = C_h A_h . \tag{5.20}$$

9. The SIMPLS results are computed as

$$T = XW , \tag{5.21}$$

$$B = WQ^T . \tag{5.22}$$

5.3.3 PLS Model Build Using the Training Data Set

Partial Least Squares Regression was used to build three models based on the variables selected by each of the three techniques used. These are referred to as PCA-PLS, PCorr-PLS, and Step-PLS respectively. Figure 5.5 shows plots of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the PCA-PLS model of the LDM data set. Eleven PCs captured 95.11% of the variance in the data matrix **X** while minimizing the RMSE.

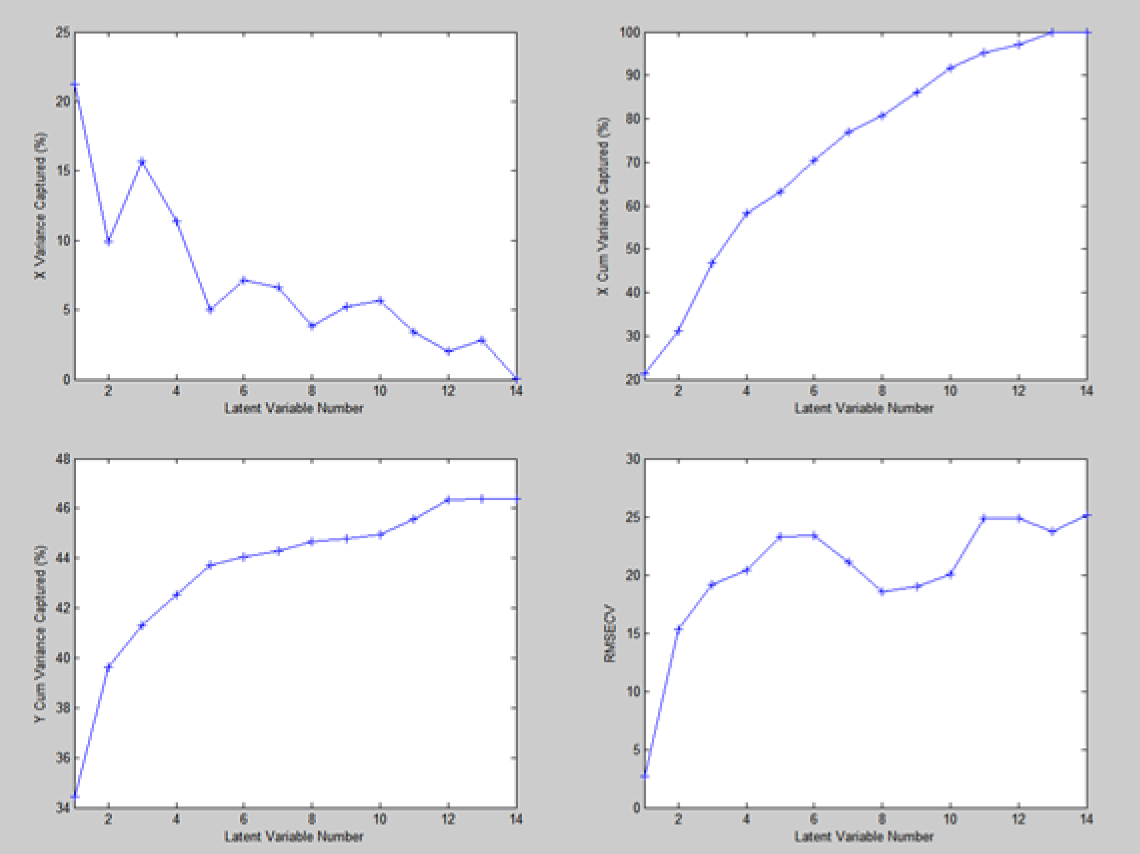


Figure 5.5: Plot of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the PCA-PLS model of the LDM data set. The optimal PC number based on RMSECV is 11.

Figure 5.6 shows plots of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the PCorr-PLS model of the LDM data set. Five PCs captured 93.54% of the variance in the data matrix **X** while minimizing the RMSE.

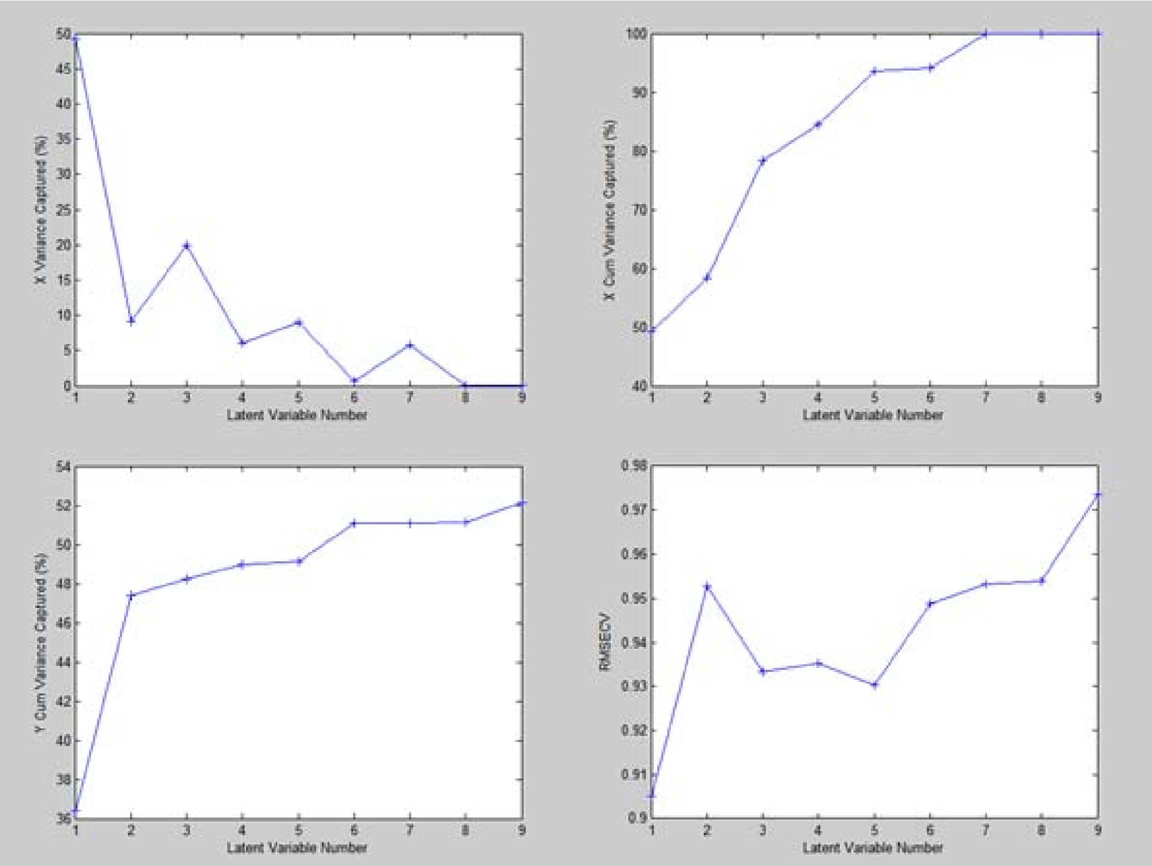


Figure 5.6: Plot of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the PCorr-PLS model of the LDM data set. The optimal PC number based on RMSECV is 5.

Figure 5.7 shows plots of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the PCorr-PLS model of the LDM data set. Eight PCs captured only 56.58% of the variance in the data matrix **X** while minimizing the RMSE.

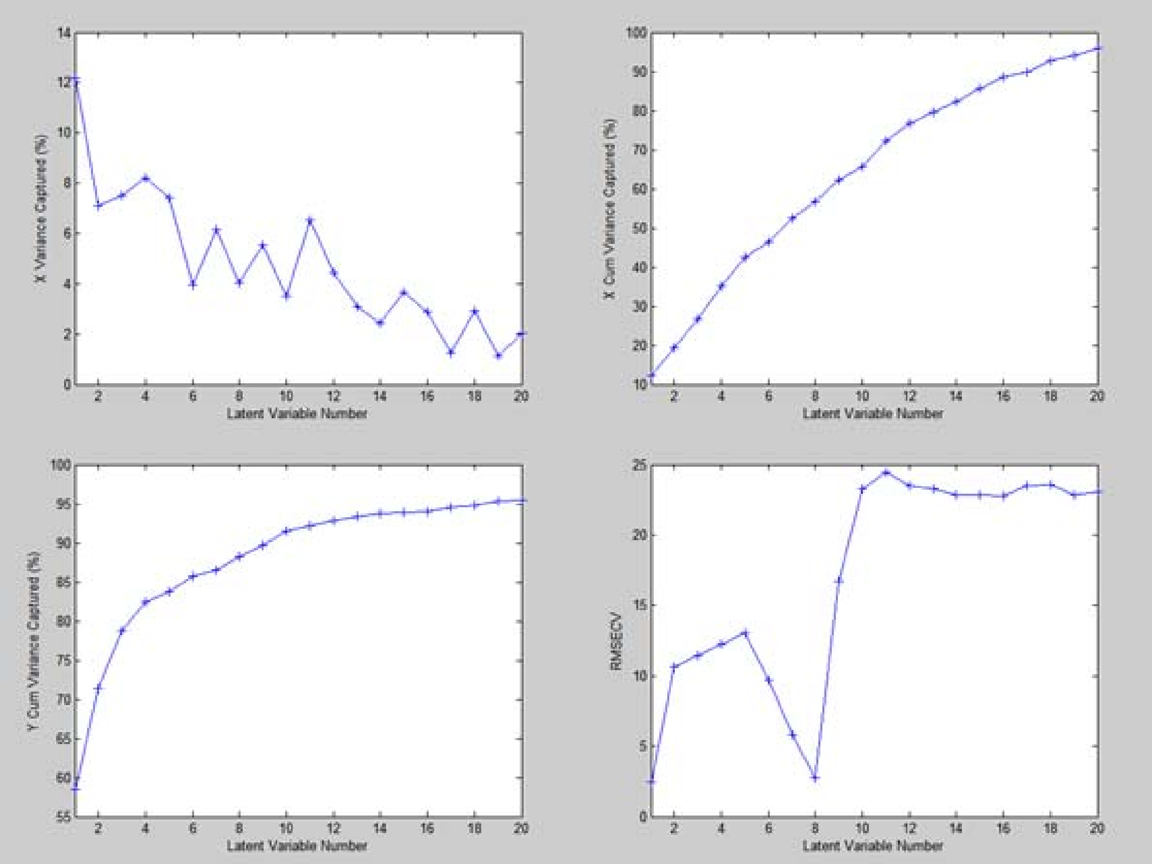


Figure 5.7: Plot of x-variance captured by PC, cumulative x and y-variance by PC, and the Root Mean Squared Error from Cross-Validation (RMSECV) for the Step-PLS model of the LDM data set. The optimal PC number based on RMSECV is 8.

Based on the R^2 metric, Table 5.3, it can be seen that the Step-PCR model explains over 88% of the variability in the observations and moderately outperforms the PCA-PCR and PCorr-PCR models with R^2 values of 0.4555 and 0.4914 respectively.

Variable Selection Technique	R^2 Metric (all PCs)	R^2 Metric (optimal PCs)
PCA	0.4639 (14)	0.4555 (11)
Pearson Correlation	0.5211 (9)	0.4914 (5)
Stepwise Selection	0.9558 (25)	0.8823 (8)

Table 5.3: PLS modeling results for different variable selection techniques for the LDM data set. The R^2 metric is reported for the optimal number of PCs retained in the model, as well as for all PCs retained in the model.

5.4 ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN) [21] can be considered supervised learning tools that link input information to output measurements. Neural networks have been applied to regression problems where the standard linear techniques, PCR, MLR, and PLSR, have not been successful. Neural networks that are used for regression are trained so that a particular input leads to a specific target output, Figure 5.8. The network is adjusted based on how the prediction compares to the target until the prediction matches the target.

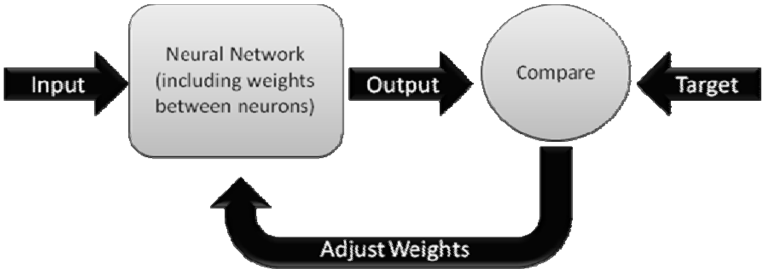


Figure 5.8: The supervised learning flow of a Neural Network where prediction and target information is used to adjust weights in the model.

The basic unit of a neural network is the neuron, which is an abstraction of the biological neuron. The artificial neuron accepts an incoming signal which is then passed to the neuron body to be weighed, summed, and transformed by a transfer function into the outgoing signal, Figure 5.9. A neuron can be characterized by its number of inputs and its transfer function. The most commonly used transfer functions are: linear, sigmoid, and tansigmoid functions. Each neuron represents a node in the network with the nodes forming a defined structure distributed across several layers.

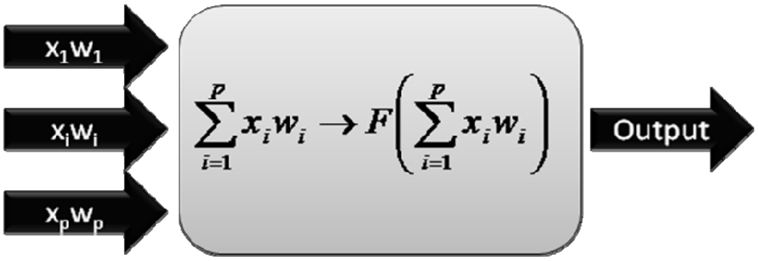


Figure 5.9: An artificial neuron with incoming signals x_i, weighting factors w_i, and a transfer function F.

There are three types of layers in neural networks: the input layer, the hidden layers, and the output layer. Each network contains exactly one input and output layer, with any number of hidden layers. Only the hidden and output layers contain transfer functions.

5.5 BACK PROPAGATION NEURAL NET

5.5.1 Back Propagation Neural Net Overview

Back Propagation Neural Net (BPNN) is a type of Multilayer Feed Forward (MLF) network that utilizes the sigmoidal transfer function to incorporate non-linear transformations of the inputs. In BPNN, all nodes in one layer of the network are connected to all nodes of the subsequent layer. As with all neural nets, the input signals are received through the input layer. The information moves through the hidden layers to the output layer which produces the final result.

The majority of BPNN type networks are created with just a single hidden layer [21]. The number of nodes in the input layer is determined by the number of variables in the data matrix **X**. Similarly, the number of nodes in the output layer is determined by the number of variables in the measurement matrix **Y**. The number of nodes in the hidden layer must be determined empirically by comparing the error in prediction to the number of hidden nodes. In general, too many hidden nodes cause the network to over train, while too few nodes will result in poor predictive capability.

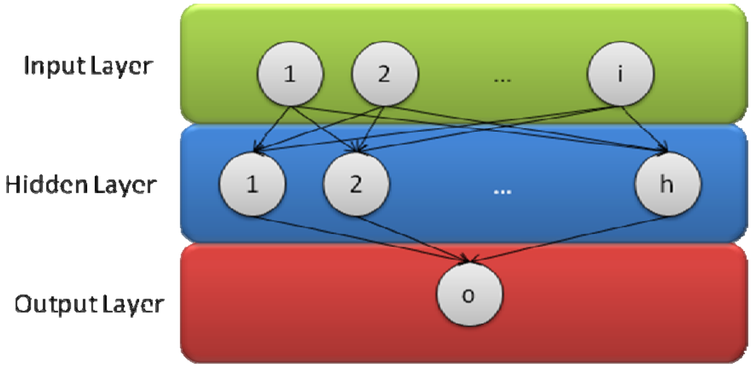


Figure 5.10: General structure of the BPNN network showing the input layer, hidden layer, and output layer used for the LDM data set.

For the hidden layer, a sigmoid transfer function is selected. The BPNN learning rule requires a transfer function for which a derivative exists over the entire domain. The transfer function performs two roles in the signal propagation process. It scales the node output value between 0 and 1, and is also responsible for the useful non-linear properties of the model. The sigmoid transfer function is

$$F(x) = \frac{1}{1+e^{-x}} . \tag{5.23}$$

For the output layer, there are two choices for the transfer function: 1) the sigmoid function and 2) the linear function. For regression modeling, the linear transfer function is preferred for the output layer.

5.5.2 Back Propagation Learning Rule

1. Initialize all weights with small random values in the range (-0.3, 0.3).
2. For each output unit, calculate the output value with the current weight settings and error based on the difference between this value and the target value.
3. Carry out the weight adaptations of the output neurons. Note that in BPNN, the weight adaptation is made in the direction that minimizes error which is essentially a gradient based optimization. Equation 5.24 is used for the weight adaptation. The w_{hj} term is the weight between the h^{th} hidden node and the j^{th} output node, η is the learning rate in the range (0, 1), o_h is the output of the h^{th} hidden node, and δ_j is a term based on the error.

$$\Delta w_{hj} = \eta \delta_j o_h \tag{5.24}$$

4. Carry out the weight adaptations of the hidden neurons. NET_j in Equation 5.25 is the net input of the hidden neuron.

$$\delta_j = \frac{d(sf(NET_j)}{d(NET_j)} \sum_{k=1}^{q} \delta_k w_{jk} \tag{5.25}$$

5. Repeat the process for all inputs. One iteration or epoch is defined as one weight correction for all examples of the training set.

5.5.3 BPNN Model Build Using the Training Data Set

Back propagation neural net was used to build several models based on the variables selected by each of the three techniques used. These are referred to as PCA-BPNN, PCorr-BPNN, and Step-BPNN respectively. Each combination of variable selection technique required a number of model builds to empirically determine the optimal number of nodes in the hidden layer. For PCA-BPNN and PCorr-BPNN, the number of variables plus 1 gave the best modeling result based on RMSE, Figures 5.11, and 5.12. For Step-BPNN, the optimal number of nodes in the hidden layer based on RMSE was less than the number of variables, Figure 5.13.

None of the models created using BPNN captured a significant amount of variance in the observations. The maximum R^2 for any model was 0.2706.

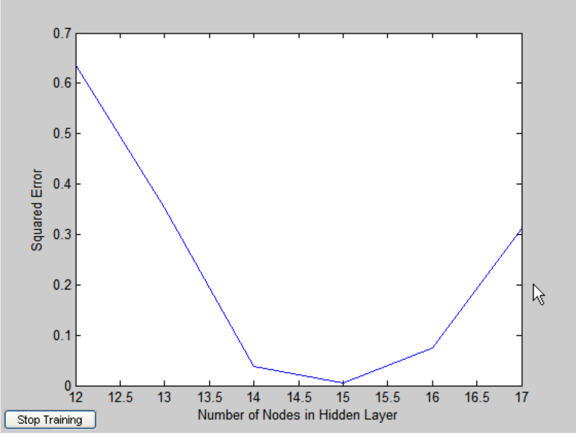


Figure 5.11: RMSE vs. the number of nodes in the hidden layer for PCA-BPNN using the LDM data set. The optimal number of nodes in the hidden layer is 15 based on RMSE.

Number of Hidden Nodes	**RMSE**	**R^2**
12	0.6357	0.1952
13	0.3524	0.2706
14	0.0374	0.0805
15	0.0052	0.0921
16	0.0740	0.0000
17	0.3126	0.0037

Table 5.4: PCA-BPNN modeling results for different numbers of hidden nodes in the network for the LDM data set.

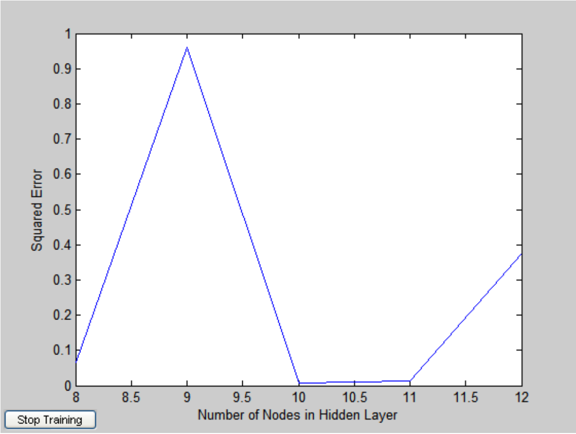


Figure 5.12: RMSE vs. the number of nodes in the hidden layer for PCorr-BPNN using the LDM data set. The optimal number of nodes in the hidden layer is 10 based on RMSE.

Number of Hidden Nodes	RMSE	R^2
8	0.0663	0.0022
9	0.9604	0.1527
10	0.0070	0.2541
11	0.0146	0.1950
12	0.3751	0.0307

Table 5.5: PCorr-BPNN modeling results for different numbers of hidden nodes in the network for the LDM data set.

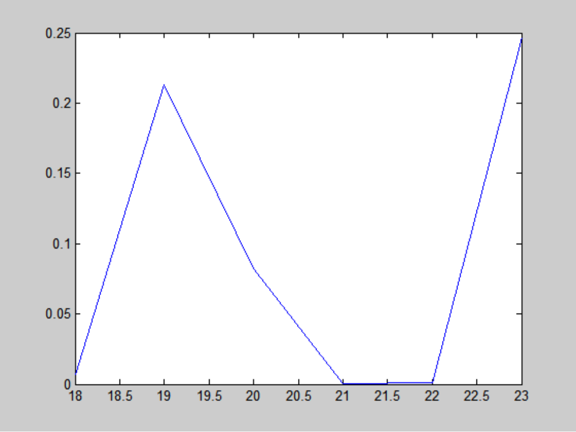


Figure 5.13: RMSE vs. the number of nodes in the hidden layer for Step-BPNN using the LDM data set. The optimal number of nodes in the hidden layer is 21 based on RMSE.

Number of Hidden Nodes	**RMSE**	**R^2**
18	0.0072	0.0268
19	0.2133	0.2208
20	0.0824	0.0428
21	0.0006	0.0409
22	0.0009	0.0026
23	0.2470	0.1415

Table 5.6: Step-BPNN modeling results for different numbers of hidden nodes in the network for the LDM data set.

5.6 RADIAL BASIS FUNCTION NEURAL NET

5.6.1 Radial Basis Function Neural Net Overview

Radial Basis Function Neural Net (RBFN) is a variant of three-layer MLF networks that utilize a gaussian transfer function to define an ellipsoid in the input space. In RBFN, the transfer function is referred to as a kernel or basis function. Like BPNN, all nodes in one layer of the network are connected to all nodes of the subsequent layer. As with all neural nets, the input signals are received through the input layer. The information moves through the hidden layers to the output layer which produces the final result. Radial Basis Function networks usually require more nodes than BPNN networks, however, they can be implemented in a fraction of the time it takes to build a standard feed forward network.

RBFN networks are created with just a single hidden layer [21]. The number of nodes in the input layer is determined by the number of variables in the data matrix **X**. Similarly, the number of nodes in the output layer is determined by the number of variables in the measurement matrix **Y**. Unlike BPNN where the number of nodes in the hidden layer is empirically determined, the number of nodes in the hidden layer of RBRN is set to be equal to the number of training samples.

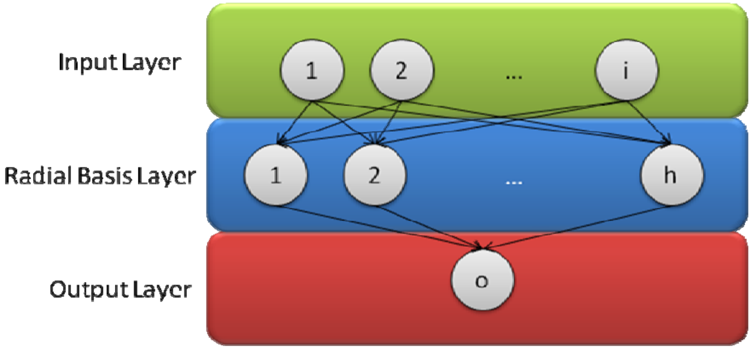


Figure 5.14: General structure of the RBFN network showing the input layer, radial basis layer, and output layer used for the LDM data set.

For the hidden layer, a Gaussian kernel function is selected. The RBFN learning rule uses the back propagation feature of the BPNN networks, and requires a kernel function for which a derivative exists over the entire domain. The Gaussian kernel function is continuously differentiable and satisfies this requirement. The Gaussian kernel function is defined as

$$F(x) = e^{-\left(|x-c_h|/b_{hj}\right)^2}, \tag{5.26}$$

where $|x-c_h|$ is the Euclidean distance between the input vector **x** and $\mathbf{c}_h$ is the centroid of the function. The output layer of RBFN always contains a linear kernel function.

5.6.2 Radial Basis Function Training

1. A number of different methodologies are used to determine the distribution of the position of centers such as:

- Random distribution within a range.
- Random selection of input patterns.
- Maximal coverage of the range of interest.
- Selection of representative input patterns.

2. The widths, *b*, of the kernel functions are determined by error back propagation.
3. The weights, w_{ij}, are determined by error back propagation.

5.6.3 RBFN Model Build Using the Training Data Set

Radial Basis Function neural net was used to build three models based on the variables selected by each of the three techniques used. These are referred to as PCA-RFBN, PCorr-RBFN, and Step-RBFN respectively. PCA-RBFN provided the best linear fit between actual and predicted values, Figure 5.15, however, the R^2 value was 0.5639 indicating that only 56% of the variation in the observations was explained. The PCorr-RBFN and Step-RBFN models performed significantly worse with R^2 values of 0.0683 and 0.1134 respectively, Table 5.7.

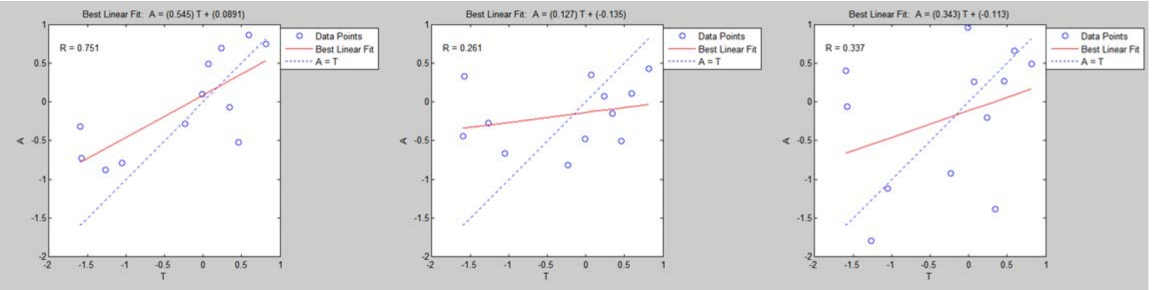


Figure 5.15: Plot of linear fits for the PCA-RBFN, PCorr-RBFN, and Step-RBFN models of the LDM data set.

Variable Selection Technique	**R^2 Metric**
PCA	0.5639
Pearson Correlation	0.0683
Stepwise Selection	0.1134

Table 5.7: RBFN modeling results by variable selection technique for the LDM data set.

Chapter 6: Results and Conclusion

6.1 VARIABLE SELECTION RESULTS

For this thesis, data was collected from an industrial ion implant machine over a period of three months. The 47 independent Runs are obtained from a single ion implanter, and the data set spans approximately three maintenance cycles. The LDM process has a single implant step with static set points. A total of 85 variables are collected and from these, 38 were removed because they were invariant across the data set. Only predictor variables that represent the mean of a parameter were introduced into the models; the variance of a parameter was not used because the sample size for the data set is not large enough. There are 47 predictors and 1 metrology value for each of the 47 implant Runs available for use in variable selection.

In general, variable selection should be performed to decrease the variance in the predicted values. In the particular case of the LDM data set, the small number of samples indicates that variable selection is necessary. Three techniques were used to reduce the number of predictor variables, PCA, Pearson Coefficient ranking, and Stepwise Selection. Stepwise Selection retained the highest number of variables, 25, PCA retained 14, while Correlation Coefficient Ranking only retained 9 of the 47 predictors, Table 6.1. All methods used retained the variables *AMag Current During Process*, *AMU During Process*, *Source Arc Current During Process*, and *TEM Actual Energy Feedback During Process*.

Variable Selection Method	Number of Variables Selected
PCA	14
Correlation Coefficient Ranking	9
Stepwise Selection	25

Table 6.1: Summary of variable selection methods for the LDM data set.

6.2 MODELING METHODS RESULTS

The 47 implant process Runs were divided into a training group of 35 Runs, and a validation group of 12 Runs. The performance ratings used for the application of the training data set are Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE), defined by

$$MAPE(\%) = \frac{\sum_{i-1}^{N} \frac{|(\hat{y}_i - y_i)|}{y_i}}{N} \times 100\,, \tag{6.1}$$

$$RMSE = \sqrt{\frac{\sum_{i-1}^{N} (\hat{y}_i - y_i)^2}{N}}\,, \tag{6.2}$$

where $\hat{y}$ is the predicted value, y is the actual value, and N is the number of samples available.

The results of each variable selection/model combination as applied to the training set are shown in Table 6.2. It can be immediately noticed that the correlation coefficient ranking method for variable selection performs the best out of the three methods examined. With regards to PCA, the main disadvantage of this method as a variable selection technique is that there is a high probability of choosing predictor variables that are correlated with one another. These extra variables do not add any extra information to the prediction accuracy, just additional variance. Although the Stepwise selection method has the advantage that the predictors are unlikely to be highly correlated, it did choose a rather large set containing 25 predictors. In the absence of a significant number of samples, 25 predictors most likely detract from the predictive capability of any model that uses the Stepwise selection method.

The best performing model over all of the techniques investigated is PCA selection coupled with RBFN. The Neural Net based models were expected to perform better than the linear models, and while RBFN did perform significantly better, the BPNN model did not. RBFN is more straightforward to implement compared to BPNN, particularly in the determination of the number of nodes to include in the hidden layer.

Model	**MAPE(%)**	**RMSE**
PCA-MLR	2.43	3.56
PCorr-MLR	1.35	2.19
Step-MLR	1.71	2.50
PCA-PCR	1.60	2.52
PCorr-PCR	1.40	1.94
Step-PCR	2.80	2.93
PCA-PLS	1.86	2.84
PCorr-PLS	1.24	1.67
Step-PLS	1.78	2.56
PCA-BPNN	1.94	2.96
PCorr-BPNN	2.01	2.66
Step-BPNN	1.96	2.56
PCA-RBFN	0.826	1.15
PCorr-RBFN	1.22	1.62
Step-RBFN	1.28	1.89

Table 6.2: Summary of variable selection/model combinations for the LDM data set.

6.3 CONCLUSIONS

This thesis has detailed an examination of various variable selection and modeling techniques in an effort to relate LDM sheet resistance measurements to Ion Implant tool processing parameters. The best variable selection/modeling combination is PCA selection coupled with RBFN, which gives an accuracy of 0.826% MAPE. Further accuracy may be achieved by coupling the tool data with advanced sensor data, such as a Residual Gas Analyzer (RGA), to appropriately determine the process conditions. The models developed in this thesis are certainly capable of following the mean trends in the LDM sheet resistance data as indicated by the MAPE and RMSE. The LDM data set, however, is not entirely suitable for modeling by all of the approaches that are described. This is primarily due to the ratio of samples to available predictor variables. BPNN and RBFN can easily achieve over fitting with just a few samples. Based on the results, a VM model constructed entirely out of tool data might not be able to replace real metrology without a significant effort to obtain an appropriate number of samples.

Bibliography

[1] A.C. Diebold, "Overview of metrology requirements based on the 1994 national technology roadmap for semiconductors," *Proceedings IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop*, (Nov. 1995), 50-60.

[2] International Technology Roadmap for Semiconductors (ITRS), Dec. 2003, [Online]. Available: http://public.itrs.net/.

[3] J.C. Yung-Cheng and F.T. Cheng, "Application Development of Virtual Metrology in Semiconductor Industry," *Industrial Electronics Society, IECON 2005, 31st Annual Conference of IEEE*, (Nov. 2005), 124-129

[4] F.T. Cheng, "Researching Strategy and Development Proposal of e-Manufacturing", *Automation Division of National Science Council*, Taiwan, R.O.C, (Oct. 2004).

[5] M.H. Hung, T.H. Lin, F.T. Cheng, and R.C. Lin, "A Novel Virtual Metrology Scheme for Predicting CVD Thickness in Semiconductor Manufacturing," *Mechatronics, IEEE/ASME Transactions on*, Vol. 12, Issue 3, (June 2007), 308-316.

[6] S. Lynn, J. Ringwood, E. Ragnoli, S. McLoone, and N. MacGearailt, "Virtual Metrology for Plasma Etch using Tool Variables," *Advanced Semiconductor Manufacturing Conference, IEEE/SEMI*, (May 2009), 143-148.

[7] T.H. Lin, M.H. Hung, R.C. Lin, and F.T. Cheng, "A Virtual Metrology Scheme for predicting CVD Thickness in Semiconductor Manufacturing," *Robotics and Automation, 2006 IEEE International Conference on*, (May 2006), 1054-1059.

[8] S.I. Imai, M. Kitabata, and T. Tanaka, "Interconnection Failure Caused by Bath Degradation in Copper Electroplating and its VM-FDC Using Mathematical Model," *Advanced Semiconductor Manufacturing Conference, IEEE/SEMI*, (May 2009), 254-258.

[9] S.I. Imai, "Virtual Metrology for Plasma Particle in Plasma Etching Equipment," *Semiconductor Manufacturing, International Symposium on*, (Oct. 2007), 1-4.

[10] V. Vitale, W. Aderhold, A. Hunter, I. Iliopoulos, N. Kroupnova, A. Yanovich, and N. Merry, "Use of Virtual Metrology for in-situ Visualization of Thermal Uniformity and Handoff Adjustment in RTP Critical Anneals," *Advanced Semiconductor Manufacturing Conference, IEEE/SEMI*, (May 2008), 349-353.

[11] Y.T. Huang, F.T. Cheng, and Y.T. Chen, "Importance of data quality in virtual metrology," *Proc. 32nd Annual Conference IEEE Industrial Electronics Soc.*, (Nov. 2006), 3237-3732.

[12] L. Xu and W. Zhang, "Comparison of different methods for variable selection," *Analytica Chimica Acta*, Vol. 446, (2001), 477-483.

[13] T. Matsuura, "An application of neural network for selecting feature variables in machinery diagnosis," *J. Mater. Processing Technology*, Vol. 157-158, (2004), 203-203.

[14] A. Miller, *Subset Selection in Regression, Second Edition*, Chapman & Hall/CRC, Boca Raton, FL (2002).

[15] T.H. Lin, F.T. Cheng, A.J. Ye, W.M. Wu, and M.H. Hung, "A Novel Key variable Sifting algorithm for Virtual Metrology," *Robotics and Automation, 2008 IEEE International Conference on*, (May 2008), 3636-3641.

[16] J.F. Ziegler, *Ion Implantation – Science and Technology*, Ion Implantation Technology Co., Chester, MD, (2008).

[17] J.F. Ziegler, *Ion Implantation – Science and Technology*, Ion Implantation Technology Co., Yorktown, NY, (1996).

[18] L.H. Chiang, E.L. Russell, and R.D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*, Springer-Verlag London Limited, Great Britain, (2001).

[19] R. Jaeger, *Introduction to Microelectronic Fabrication*, Addison-Wesley, Reading, MA, (1993).

[20] J.E. Jackson, "Principal Components and Factor Analysis: Part 1-Principal Components," *J. Qual. Tech.*, Vol. 13, Issue 1, (1981).

[21] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier Science B.V., Amsterdam, The Netherlands, (1998).

[22] S. de Jong, "SIMPLS: an alternative approach to partial least squares regression," *Chemom. Intell. Lab. Syst.*, Vol. 18, (1993), 251-263.

Vita

Courtney Marie Fowler was born in San Antonio, Texas, on September 27, 1975, the daughter of Donna Marie Fowler. She was raised in Texas, California, and Alaska by her mother, and her stepfather, John L. Nichols. Courtney has one brother, Levi S. Nichols, and one sister, Tiffanie A. Nichols. She received a Bachelor of Science degree in Mathematics from the University of Texas at San Antonio in 2003. From November, 1997, to December, 2007, Courtney worked as a Process Technician and then a Process Engineer for Freescale Semiconductor, formerly Motorola Inc. In January of 2006, she started her graduate studies in the Department of Chemical Engineering at the University of Texas at Austin. In January of 2008, Courtney accepted a position as an FDC Engineer for INFICON Inc.

Permanent address: 14 Oakland Ave., Newton, MA 02466

This thesis was typed by the author.