2008

Analysis of casting surface anomalies captured through spatial mapping

Scott Alan Von Busch
Iowa State University

Recommended Citation

Von Busch, Scott Alan, "Analysis of casting surface anomalies captured through spatial mapping" (2008). *Graduate Theses and Dissertations*. 11071.
https://lib.dr.iastate.edu/etd/11071

Analysis of casting surface anomalies captured through spatial mapping

by

Scott Von Busch

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Major: Industrial Engineering

Program of Study Committee:
Frank E. Peters, Major Professor
Matt C. Frank
Scott L. Chumbley

Iowa State University
Ames, Iowa
2008

Table of Contents

Abstract

The data needed for quality control in the steel casting process is often difficult to obtain. This is especially true when referring to the surface quality of the part as it undergoes multiple inspections. A typical inspection involves identifying the location of anomalies and marking them for further processing in the cleaning room. Each time an inspector views a casting, information on the part surface quality is conveyed. This information however, is rarely available for analysis since it is recorded directly on the casting. A few foundries have attempted to collect this surface quality data (anomaly type, size, and location) as identified during inspection. Unfortunately, their data format is difficult to manage and has limited analysis opportunities. This paper presents a software program which removes the problems associated with current attempts at data collection. The program provides an easy to use interface for recording anomaly type and location directly on a 3D CAD model. Analysis modules designed for this data include histograms, frequency plots, area calculations, correlations, and principal component analysis (flaw pattern recognition). A case study for collecting and analyzing real data from a steel casting foundry was completed using this program. Some sample results from this study are included in this paper to illustrate benefits achieved from the data collected.

Chapter 1. Introduction

Manufacturing industries rely on data collection to gain understanding and control of their process. Typically, this information is related to part quality at critical stages in its production. For the steel casting industry, this type of data can be obtained, but is often difficult to manage. This is particularly true when referring to the surface quality of the casting as evaluated during inspection.

Inspections identify casting surface anomalies through non-destructive evaluation (NDE) methods. Anomalies are defined as casting features which must be mitigated before the customer will accept the part; common types include cracks, porosity, sand inclusions, and shrinkage. After anomalies are identified, their locations are marked to advise subsequent operators where further processing (grinding/welding) is required (Figure 1). All castings undergo multiple inspections where each time it may be approved or marked for additional work. It is during these inspections where valuable information related to the quality of the casting is identified, including the size, type, and location of anomalies. Given that the information is only recorded on the actual casting, it can rarely be used for analysis.

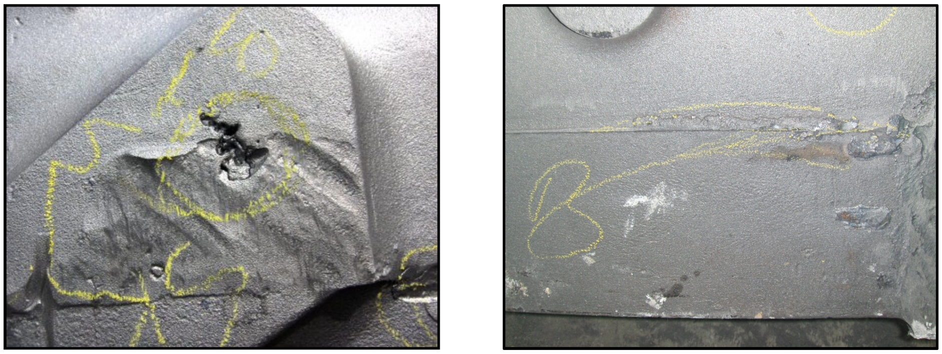


Figure 1: Examples of casting surface anomalies marked for cleaning operations.

The value of this data for achieving better process control is acknowledged by foundries and efforts have been made to capture it. Recent studies show the current best-case practice is to collect inspection details (inspector, part number, detection method) along with the size and location of anomalies on a 2D part drawing. Since this data is collected on paper, storing it in an accessible manner is difficult. Also as difficult is achieving any evaluation of anomaly locations across many parts. While the effort made to collect the data is good, the format is difficult to manage and has limited analysis opportunities. The lack of a method for collecting the data in an effective format prevents the majority of foundries from even attempting.

In 2007, Iowa State University proposed a new software program for collecting this casting surface data (Boonsuk, 2007). The program, titled Surface Anomaly Mapping*, provides a 'point and click' interface for marking the anomaly type and location directly on a 3D CAD model. Additionally, users can specify other types of information to record with regards to the part and the inspection. All data entered is stored in an electronic database for easy accessibility on the foundry's server. The database can be readily linked to other foundry databases housing information on process variables related to the part. What results is a repository of data available for analysis. This program addresses problems associated with past collection attempts by providing foundries with manageable data for use in multiple analysis applications.

The goal of this study was the development of analysis modules specifically designed for use with the software program. Each module was created to fill an immediate need in the steel casting industry as identified through interviews with foundry personnel. The modules range from purely visual to more advanced statistics and provide new analysis options previously unavailable or difficult to attain.

The analysis modules developed are as follows:

- **Histograms**-for describing how often each anomaly type is appearing on a particular part
- **Frequency plots**-for displaying the distribution of anomaly occurrences across a series of parts on the 3D CAD model
- **Anomaly Area**-for calculating the size of the anomaly region
- **Correlations**-for investigating relationships between process variables and anomaly locations
- **Principal Component Analysis**-for classifying parts by their surface flaw pattern

A 10 week case study at a steel casting foundry was completed to trial the software program. During this case study, real data was collected and analysis tools were refined. Sample results from the study are provided as real examples of benefits achieved by utilizing the software program. The benefits from the analysis tools presented in this paper are just the beginning to what foundries could receive in the future.

*Software is copyright protected

Chapter 2. Literature Review

A literature review with regards to casting surface data was conducted with three objectives. They were to confirm the need for the program, to investigate automated data collection methods, and to examine related work. The following paragraphs summarize the information gathered while pursuing each objective.

The first objective was to confirm no data collection system currently existed for this type of data. After a robust review, no academic or commercial substitutes were identified. These findings were validated through additional interviews with Steel Founders' Society of America (SFSA) members and foundry quality control personnel. While some foundries were collecting this data, their data format produced shortcomings in their analysis. Research outside the steel casting industry did discover a comparable method for tracking defect locations. Though the approach is not suitable for casting surface data, it was still closely analyzed for constructive insight.

The automotive case study by Bala et al. (2005) presented a quality management system for tracking the location of car door defects. The system provided users with a software interface for recording data and rapidly analyzing it in a visual manner. Included in their system was the ability to create simple statistical outputs, such as pie charts, Pareto diagrams, and concentration diagrams. The latter analysis was of particular interest because it involved defect locations.

Defect locations were recorded by selecting a cell inside an Excel spreadsheet. The cell represented a specific region of the car door defined by the 2D CAD drawing in the background. The simplicity of this method leads to several limitations regarding the data. First, the data does not capture the size and shape of the defect. Second, the locations are defined in an arbitrary fashion. This prevents the data from being used in other analysis programs and applications. The data collection program was designed for use in the automotive industry and has not been made commercially available.

The second objective was to evaluate whether automated inspection systems were capable of rapidly collecting casting surface data. Multiple studies by Mery (2002-2006) were examined for this purpose. In these studies, casting defects (anomalies) were identified through x-ray or camera images captured from multiple angles. Each image was then analyzed through a set of algorithms for classifying defect regions. The majority of this research pertained to the inspection of aluminum wheels, in which up to 83% were inspected correctly through this system.

Another automated inspection system proposed by Prieto (2002) et al. used a 3D point cloud with the 3D CAD model for identifying defects. The point cloud was created by taking a laser scanned image of the part from all sides. These points were then mapped to the 3D CAD model in a software program. The deviation from the point and the 3D model represented areas of possible defects.

There are many other examples of automated inspection systems, but the two studies presented here are enough to explain why it is not a popular option in the steel casting industry. The problem starts with the amount of environmental control needed to obtain an accurate image of the casting. This includes the orientation and surface condition of the part, as well as lighting and equipment. In the foundry environment, this can be too difficult to achieve. This is regardless of whether the image is captured by camera or through a 3D laser scan. Secondly, the inspection systems are not designed to capture data. They are ideally used for pass or fail conditions and often don’t classify the type of defect, size, or location. This alone prevents it from being an option for rapidly collecting casting surface data. As technology improves in the future, these types of systems should be re-evaluated for use in this manner.

The final objective of the literature review was to examine past research in which surface data was collected. Two particular studies on visual inspection were examined for their approach in identifying and recording anomaly locations. The study by Daricilar (2005) focused on quantifying the amount of variability within and between inspectors. This was achieved by placing stickers directly on the casting where the inspector identified an anomaly. An image of the casting was then taken and sticker locations were automatically identified through a computer algorithm. The results from the study showed the overall effectiveness of the inspector was around 67%.

An independent study conducted by Schorn (2006) also investigated the effectiveness of visual inspection. The objective for this study was less concerned with the locations of defects and placed more focus on whether the part was correctly approved or scrapped. The results agreed with the previous study as visual inspection effectiveness ranged from 69-90%. This implies up to 30% of parts were incorrectly inspected. The cost implications of this ineffectiveness encourage foundries to identify the variability in their own process. Focus can be centered on determining methods for improving their inspection process in the form of better training, lighting, and equipment.

A clean steel project conducted by Carpenter (1996) provides a third example of collecting casting surface data. In this project, experiments were designed to determine the process variable(s) impacting the total amount of dirt on the cope surface. Each experiment involved recording values for each process variable and the size and location of dirt defects on a 2D CAD sketch. The number of variables examined during these experiments exceeded 30, but in the end a gating system change resulted in the largest decrease of dirt.

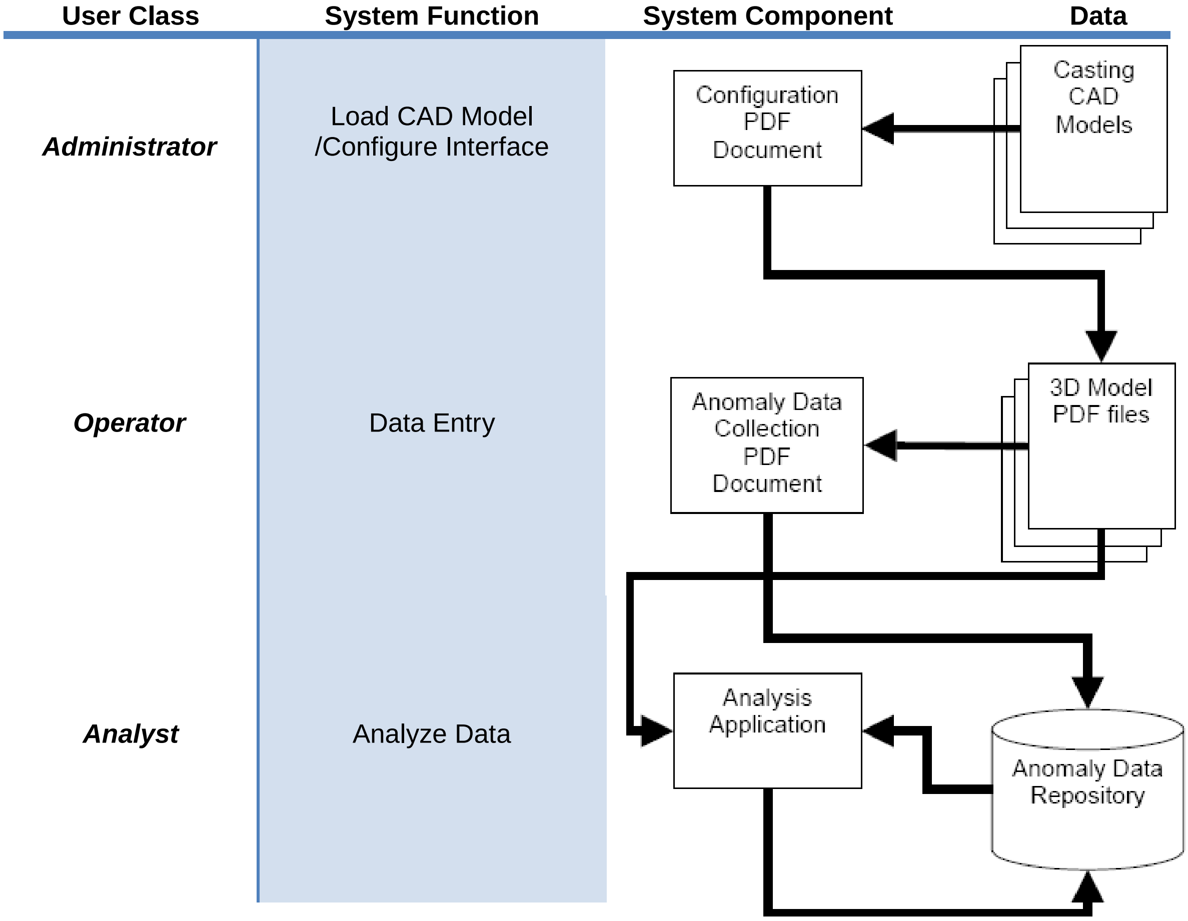
The analysis performed by Daricilar, Schorn, and Carpenter would not be possible without information regarding the surface quality of the casting. The data has proven useful in measuring the variability in the visual inspection process and in identifying root causes in defect formation. This demonstrates need for a system that collects and stores casting surface data in an efficient manner.

This literature review met all three of the intended objectives. First, it confirmed the need for the software program presented in this paper. Second, using automated inspection for rapidly collecting surface quality data was determined infeasible. Thirdly, research examples of where casting surface data provided benefits to foundries were identified. The knowledge gained during this research was helpful in shaping the direction for this project.

Chapter 3. System Architecture

The software program consists of three distinct components for configuration, data input, and analysis. Configuration and data input components were created using the Adobe Acrobat 3D platform and JavaScript programming language. The data analysis component was developed separately as a windows application written in VisualBasic.NET language. Data can also be readily exported to other software programs such as Excel, JMP, and Minitab to take advantage of analysis tools already available in those commercial packages. The three separate components communicate with each other through data stored in the anomaly database. Details on the database are included in Appendix A. A diagram illustrating the system architecture is provided in Table 1.

Table 1: System architecture for Surface Anomaly Mapping program.



3.1 User Classes

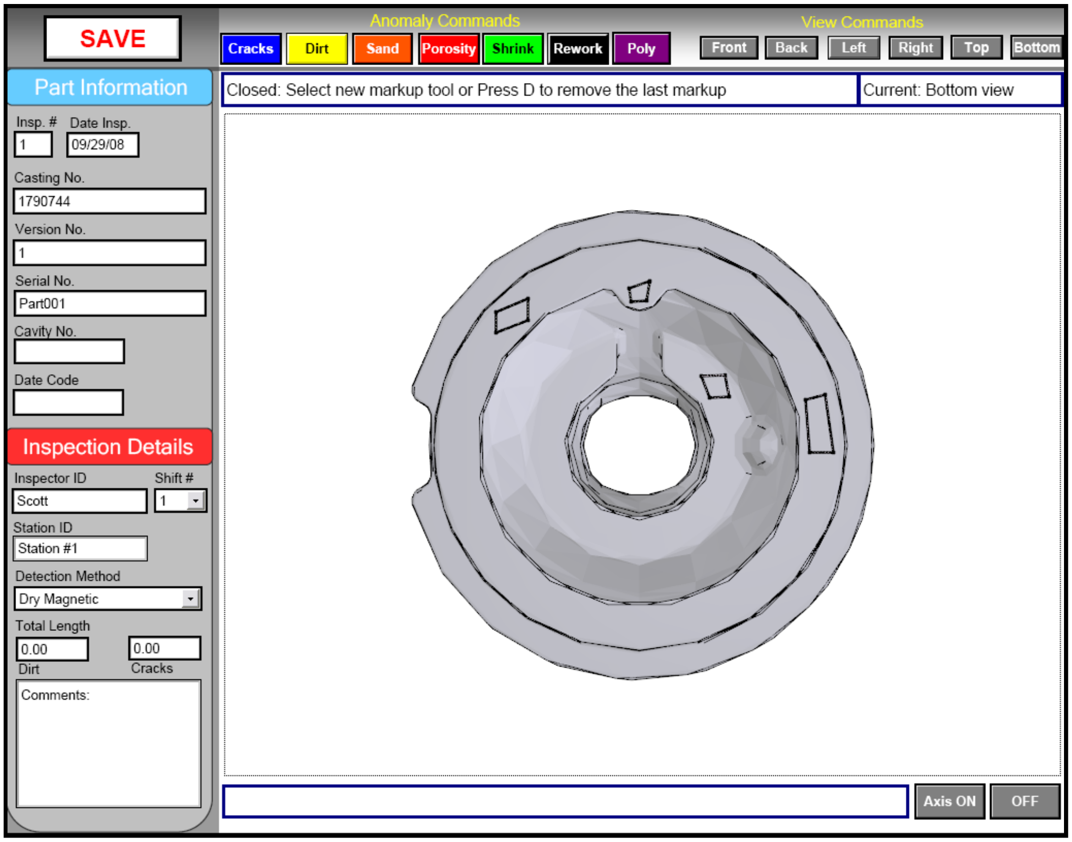
The accessibility of each component is determined through software user classification. Users are classified as *administrators*, *operators*, or *analysts* based on the function they serve. *Administrators* are responsible for loading the 3D CAD file into the program and customizing the data input interface. *Operators* physically enter the data, and *analysts* pull historical data into analysis tools and interpret. The following sections provide more details on the roles of each user class and the components they interact with.

3.1.1 Administrators

The *administrator* user class is responsible for two primary functions. First, to load the 3D CAD files into the program and second, to customize the data input interface for meeting the foundry's specific needs. Loading models is a simple process due to the many different file formats accepted by Adobe Acrobat 3D. All models created in major CAD packages, as well as standard transfer file formats such as .STL or .IGES are recognized. The *administrator* can also at this time modify the model's color, lighting, and rendering. Once a 3D CAD model is loaded into the program it becomes a unique PDF document for easy accessibility.

The *administrator's* second function is to customize the data input interface. An example of a fully configured interface is shown in Figure 2. In this example, the anomaly types selected include cracks, dirt, sand, porosity, and shrinkage. A color is also associated to each anomaly type for distinguishing its locations on the part model. *Administrators* can configure how many anomaly types to collect, the terminology used (example: gas instead of porosity), and the color associated with it.

In addition to specifying anomaly types, *administrators* determine which CAD views are accessible to the *operator* for marking anomaly locations. This is necessary for multiple reasons. First, allowing free rotation of the 3D CAD model would increase the complexity and the computer knowledge required by the *operator*. Second, by specifying a set of CAD views, data entry is quicker and the data collected can be analyzed in an efficient manner (See Chapter 4). The default setting is to include the six standard CAD views (front, back, top, bottom, left, and right). However more views can easily be added or removed. Additional views could display cross sectional areas for capturing interior geometry, as well as views for marking otherwise inaccessible regions. It is estimated that a typical casting will require 6-10 views for capturing data.



Closed: Select new markup tool or Press D to remove the last markup

Current: Bottom view

Insp. # 1

Date Insp. 09/29/08

Casting No. 1790744

Version No. 1

Serial No. Part001

Cavity No.

Date Code

Inspector ID Scott

Shift # 1

Station ID Station #1

Detection Method Dry Magnetic

Total Length

0.00 Dirt

0.00 Cracks

Comments:

Figure 2: Example of fully configured data input interface for a specific foundry. Configurable items include the ‘Anomaly Commands’, ‘View Commands’, ‘Part Information’, and ‘Inspection Details.’

The fields for recording part and inspection information are the final configurable aspect of the data input interface. It is here where *administrators* specify what part information (pattern number, mold date, heat code, etc.) and inspection details (inspector, inspection station, detection method, etc.) to collect. The only requirement is the part information and inspection details make the data entry unique. Typically, this is accomplished by including a serial number for representing a specific part and an inspection number for denoting when the data was captured. Recording the inspection number also allows greater analysis of the entire cleaning room process. For instance, it can identify what is being marked the 2nd, 3rd, and 4th time inspected. Another good practice is to record the necessary information for linking the specific part entry to data store in other foundry databases. This thereby increases the data set available for performing analysis later.

3.1.2 Operators

Operators are responsible for entering data once the *administrator* has loaded the CAD models and configured the work space. Users in this class require only basic computer knowledge due to the user friendly design of the data input interface. The procedure for data entry begins by recording part information and inspection details in the appropriate fields (See Figure 2). Entering this data is accomplished through typing or selecting items in a drop down list. The following steps in the procedure assume the operator is entering data while the casting is being inspected.

After entering the required information, the next step is to mark anomalies on the CAD model. To accomplish this, the *operator* first selects the CAD view representing the side of the casting being inspected. Selecting the CAD view will automatically rotate the 3D model to the appropriate side. The o*perator* can then pick the anomaly type to be marked from the anomaly command list. Upon choosing one, the *operator* is prompted to define a set of points for creating a polygon. Each point is created by clicking anywhere on the model. Subsequent points are connected with lines for outlining the region containing the anomaly. Polygons can be of any shape and size, but must contain at least three points and have no self intersecting segments.

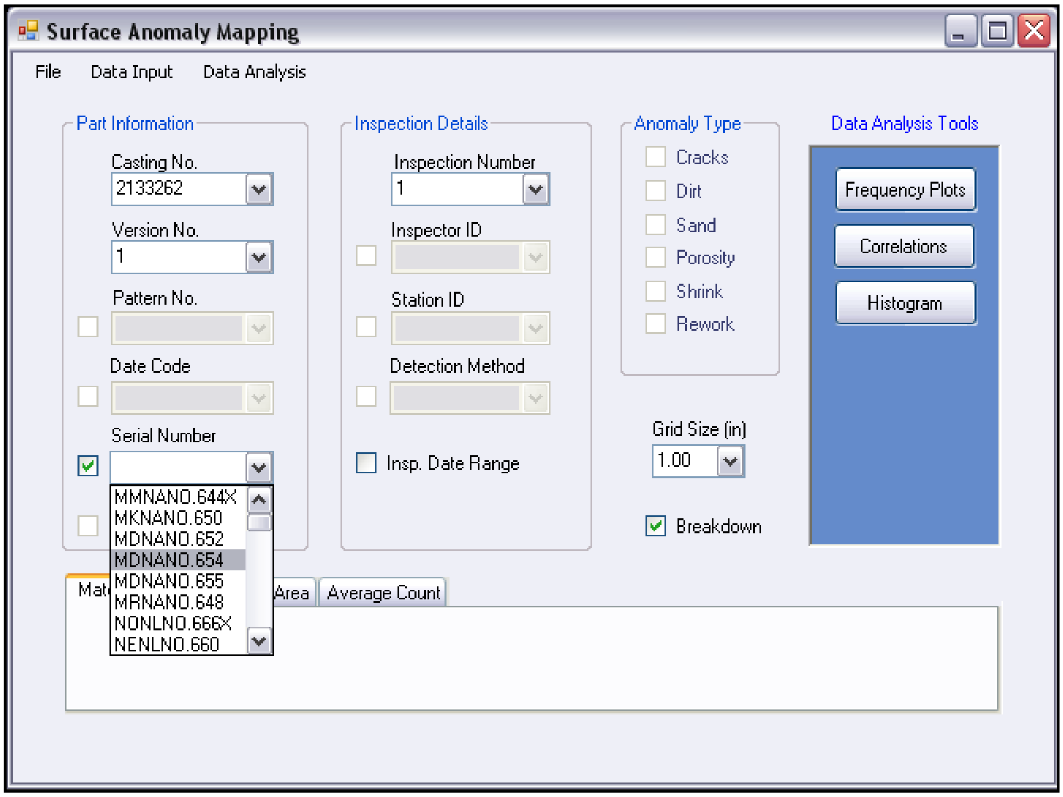
To aid in marking anomalies, *operators* can zoom and pan to specific areas on the model. The distance between the last point and the mouse position is also displayed to assist the *operator* in marking the correct anomaly size. This process of selecting a CAD view, choosing an anomaly type, and creating a polygon repeats for every anomaly identified by the inspector.

The final step for the *operator* is to click the SAVE button for populating the database with the new data. Data transferred to the anomaly database includes all part and inspection details, as well as the *x, y, z* coordinates of each point marked on the model. The associated CAD view used for marking each anomaly polygon is also saved to the database. The relevance of saving the CAD view is explained later when frequency plots are discussed. For understanding where the information is stored in the database, please reference Appendix A.

3.1.3 Analysts

Analysts are responsible for analyzing the historical data located in the anomaly database. The different analysis tools available for them to use are histograms, frequency plots, area calculations, correlations, and principal component analysis. Data can also be readily exported to other software programs for taking advantage of other analysis applications. Due to the complexity of some of these analysis tools, the users in this class would typically include quality control managers or engineering staff. Knowledge of how to use statistical software packages (JMP or Minitab) is also required.

The procedure for conducting any analysis begins by determining what historical data needs retrieved from the anomaly database. *Analysts* accomplish this by defining a criterion which parts must match in order to be included in the data set (Figure 3). This search criterion is created by first selecting a particular casting type (usually denoted by a casting number). At this point, the user can choose to run an analysis that includes all parts matching the casting type. If desired however, *analysts* have the option to constrain the data set even more by choosing additional attributes the part data must match. This could include specifying a serial number, an inspection date range, or an inspector ID. The type and number of attributes available is dependent on how the *administrator* configured the data input interface. Only parts matching the criteria will be included in the dataset. The data can be constrained even further still by specifying a particular anomaly type for analysis. For example, if the *analyst* is only interested in cracks they can ignore data regarding other anomaly types.



Casting No. 2133262

Version No. 1

□ Pattern No.

□ Date Code

☑ Serial Number

MMNANO.644X
MKNANO.650
MDNANO.652
MDNANO.654
MDNANO.655
MRNANO.648
NONLNO.666X
NENLNO.660

Inspection Number 1

□ Inspector ID

□ Station ID

□ Detection Method

□ Insp. Date Range

□ Cracks
□ Dirt
□ Sand
□ Porosity
□ Shrink
□ Rework

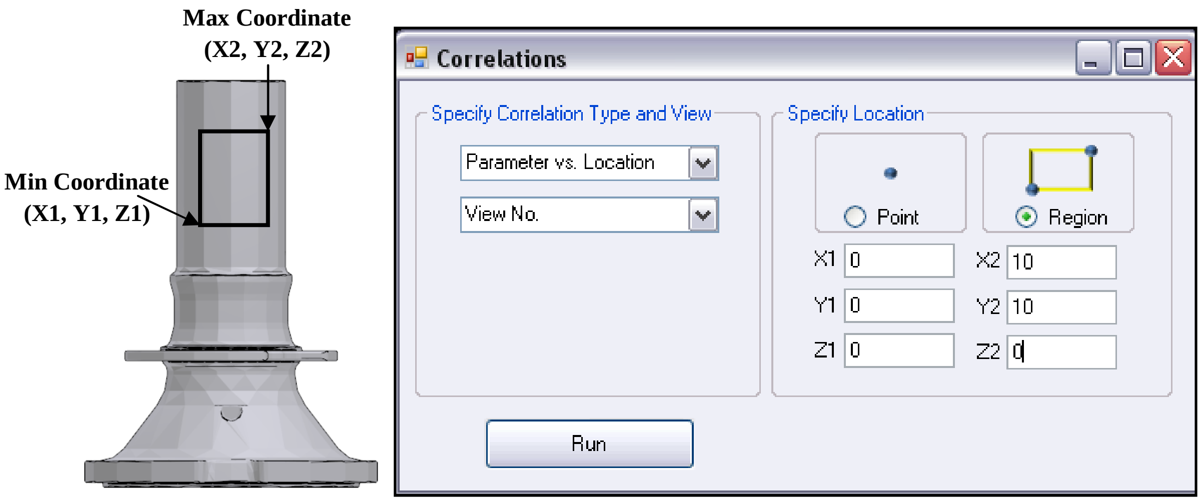
Grid Size (in) 1.00

☑ Breakdown

Figure 3: Windows application for specifying analysis type and for defining search criteria for what parts to include in the data set.

The next step after defining the data set depends on what type of analysis is being performed. Histograms require no additional information and the output will appear inside the window application. For all other modules, the *analyst* must specify a grid size. Aspects to consider when selecting a grid size are saved for later discussion in Chapter 6. Once the grid size is chosen, frequency plots and area calculations can be conducted and outputs appear in Adobe or Excel.

The correlation and principal component modules require additional details since these analyses are with respect to specific locations on the CAD model. *Analysts* define locations by entering the minimum and maximum *x*,*y*,*z* coordinates for the region. These coordinates are obtained from the 3D CAD model in Adobe. The CAD view related to the location must also be entered in the program (Figure 4). The output from these modules is a dataset ready for further analysis inside statistical software packages such as JMP or Excel. More details on the different types of analysis modules are described fully in the next chapter.



Parameter vs. Location

View No.

○ Point ● Region

X1 0
Y1 0
Z1 0
X2 10
Y2 10
Z2 0

Figure 4: Windows application for running correlations and principal component analysis. Regions are defined by their minimum and maximum coordinates and CAD view number.

Chapter 4. Data Analysis Modules

The data analysis modules were designed to assist foundries by evaluating data captured with the software program. Most require minimal human intervention and are easy to interpret. The following sections outline the approach for creating histograms, frequency plots, calculating anomaly area, performing correlations (process and location based), as well as principal component analysis. The analysis modules are designed to either create a data set and perform the analysis or create a data set and export it to a statistical software program. Results then may appear as a PDF, inside Excel, or within a statistical software program (such as JMP or Minitab). For information on how the software collects and stores the data used for analysis, please reference Chapter 3.

4.1 Histograms

Histograms are a visual tool for describing how often an anomaly type appears on a particular casting. A histogram is comprised of a series of columns, a vertical axis for quantifying the frequency, and a horizontal axis for identifying each bin or variable. The frequency for each anomaly type is calculated through a database count query and shown through the height of its corresponding column. Each column is color coded to distinguish anomaly type and to match the color scheme used in the data input interface.

The main benefit of generating histograms is the increased awareness of those anomaly types which produce the majority of additional cleaning room operations. Commonly, histograms are organized to show decreasing frequency levels from left to right. This format is often called a Pareto chart and is used extensively by quality departments for concentrating their efforts. An example of a histogram is shown in Figure 5.

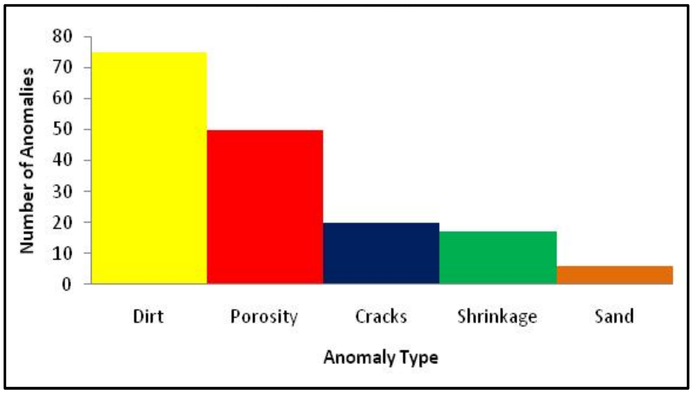


Figure 5: Histogram showing the frequency of anomaly type.

4.2 Frequency Plots

Frequency plots are a graphical tool for displaying the distribution of anomaly occurrences with respect to CAD model locations. The frequency plot appears in Adobe Acrobat 3D as a series of CAD views matching the ones used for data input. Each CAD view is mapped with a range of colors for representing the frequency count at different locations. The output also includes a scale for quantifying the count associated to each of the five colors. An example frequency plot is shown in Figure 6. From this output, users can readily identify high and low anomaly activity by distinguishing between the red and blue regions respectively.

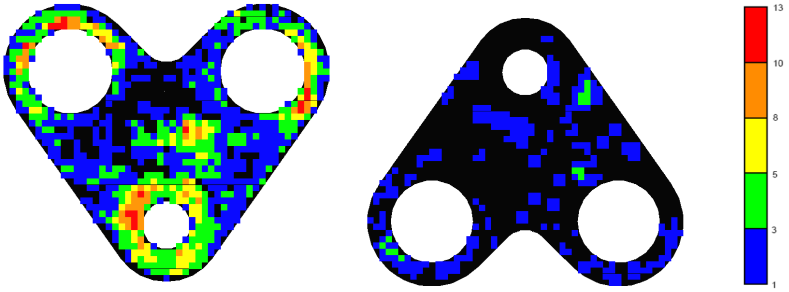


Figure 6: Frequency plot example showing two CAD views.

As is the case with all analysis modules, frequency plots are created based on data derived from a series of parts. The parts included in the data set are at the discretion of the user (See Section 3.1.3). The following sections provide detail on the three main operations for generating frequency plots. In order, the operations are as follows: defining locations, checking locations for anomaly activity (point in polygon testing), and coloring locations based on their final anomaly count.

4.2.1 Defining Locations

The first step in creating a frequency plot is producing a set of locations to check for anomalies. The original approach aimed at defining these locations directly on the 3D CAD model. Unfortunately, this proved infeasible due to limitations inside the current Adobe Acrobat platform. As an alternative approach, the CAD bounding box is utilized to create 2D representations of actual locations on the 3D model.

A bounding box contains all points for a 3D CAD model, where each face represents a specific side or view. The primary use of bounding boxes is to rotate 3D models to individual views inside CAD packages. Likewise, the bounding box allows users to change views during data entry for this program. The bounding box created for defining locations is identical to the one used for data entry. Bounding boxes are typically axis aligned, meaning each side is parallel to one of the major axes. When defining locations, points on each side of the bounding box are used to represent points in the associated view of the 3D model. Even though these locations are not defined on the model itself, this method comes closest to achieving the original intent.

Calculating the bounding box requires the minimum and maximum coordinates of the 3D CAD model. Once these are acquired, each bounding box side is defined through a combination of those coordinates. The example provided in Figure 7 shows how the front face is created. The next function is to create a set of points on the six surfaces of the bounding box. To accomplish this, all sides are partitioned into a grid of equally sized cells. The midpoint of each cell denotes a site on the bounding box surface for representing an actual point on the 3D model (Figure 7). The final step in defining locations is storing grid cell midpoints in a data table. These points, stored as 2D coordinates, are later used in a point-in-polygon algorithm for determining how many times they fall inside an anomaly.

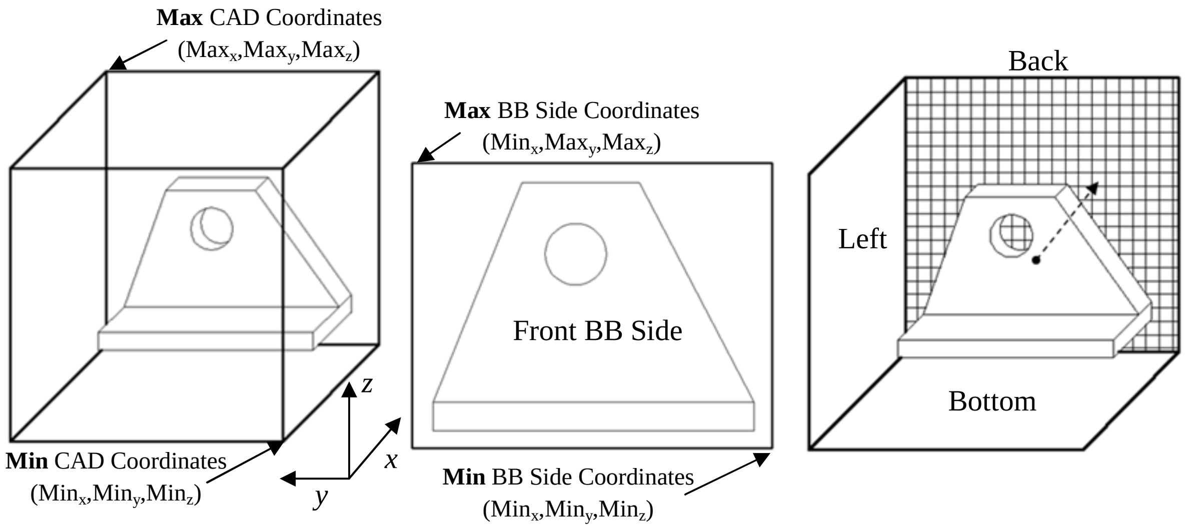


Figure 7: Steps for defining locations by utilizing the CAD bounding box (BB). Left: Bounding box created by maximum and minimum coordinates of the CAD model. Middle: Bounding box sides defined from different combinations of CAD coordinates. Right: Bounding box side partitioned into grid for defining locations.

4.2.2 Point-in-Polygon Test

A point-in-polygon test determines if grid cell midpoints are located inside of anomaly regions. Recall that anomalies are defined by polygons on the surface of the 3D CAD model. These polygons are stored as a set of 3D coordinates and must be projected to the appropriate bounding box face in order to lie in the same plane as the points (locations) being checked. This provides reasoning for why the CAD view, used when marking the polygon, is stored in the database. It ensures polygon vertices are projected to the correct bounding box surface.

Projecting the polygon vertices is a matter of performing an orthogonal transformation for every point. An orthogonal transformation removes the coordinate whose axis is parallel to the normal vector of the bounding box face. For instance if the face has a normal vector of [0, 0, 1], the *z* coordinate is removed. This transformation must occur for all polygon vertices prior to use in the point-in-polygon algorithm. If the bounding box is not axis aligned, a more complicated transformation must occur.

The point-in-polygon algorithm selected for this study is commonly coined the “number of crossings” or “even/odd” method (O’Rourke, 1998). The “even/odd” method works by counting the number of time a ray, originating at the cell midpoint, crosses each line segment of the anomaly polygon. If the ray crosses an even number of times, the point is outside the polygon. Odd number of crossings signifies a point within. This method was chosen because it works for all simple polygon shapes (concave or convex) and is calculated quickly.

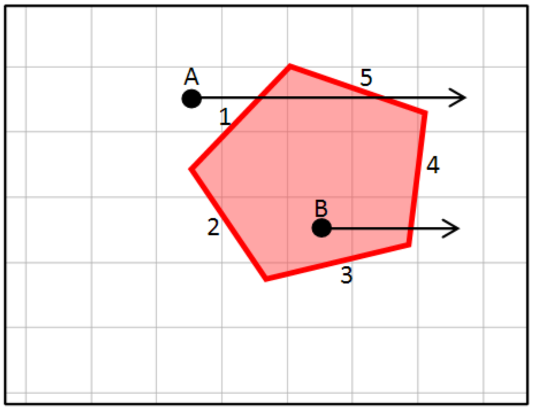


Figure 8: Point-in-polygon test example. Point A crosses twice and is outside. Point B crosses once and is inside.

The logic for the test is quite simple in nature and is best described through the visual shown in Figure 8. In this example, the polygon lies in the *xy* plane and Points A and B are examined. The test begins by classifying what side of the line segment the point sets on. This is done to ensure the rays, created from the points, are crossing all line segments from the same side. The importance of this can be seen in the Figure 8 above. If Point B had no constraint on the direction of the ray, it would intersect line segments 2 and 4, causing the test to incorrectly categorize the point as outside the polygon. The equation shown in Figure 9 is a combination of the point coordinate with the equation for a line segment. Points generating negative values are on one side of the line, positives on the other. By only continuing the test with points generating negative values, the rays will cross the line segments from the same side.

$$(P_y\text{-}Y_1) – [(Y_2\text{-}Y_1) / (X_2\text{-}X_1)] * (P_x\text{-}X_1) < 0$$

$(P_{x,}\ Py_{)}$ = coordinate of point

$(X_1$, Y1) & $(X_2$, Y2) = endpoints of line segment

Figure 9: Equation for ensuring rays all intersect from the same side of each line segment.

Points on the correct side of the line will then be tested to determine if their ray intersects the line segment. This is accomplished by comparing the *y* coordinate of the point with the *y* coordinates of the line segment's vertices. If the point lies between them, the ray intersects. This test repeats for every line segment of the polygon. The final number of intersections is then used to classify the point. In the given example, Point A crosses twice and is outside and Point B is inside by crossing only once. The set of inequalities used for checking the point with each line segment is shown in Figure 10. Only points satisfying the equation are used in the test.

$$Y_1 < P_y \text{ AND } P_y <= Y_2 \text{ OR } P_y <= Y_1 \text{ AND } Y_2<P_y$$

Figure 10: Inequality for testing whether the ray intersects the line segment. In order for the ray to intersect, the point coordinate must satisfy these conditions.

The inequality also is designed to handle the situation when a ray intersects a polygon vertex. A problem could occur in this situation since the ray technically crosses both line segments sharing the same end point. An example of this is shown in Figure 11, where the ray intersects line segments 1 and 2. The number of crossings for this point would falsely classify it as being inside the polygon.

To handle this situation, the inequality only allows one of these line segments to be counted. This ensures the output from the test remains accurate. When a ray intersects a polygon vertex, only line segments located below the point are tallied. The complete logic for the point-in-polygon test is provided in Appendix B.

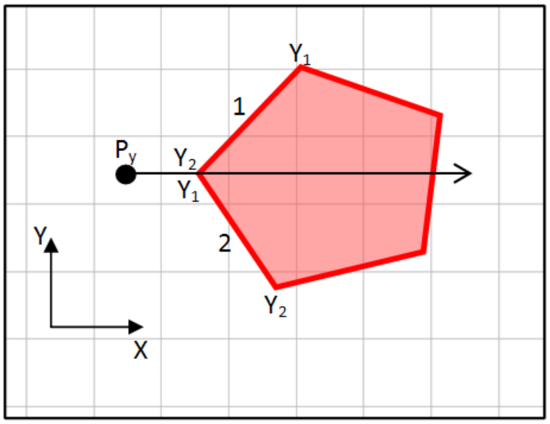


Figure 11: Example where ray intersects a polygon vertex. The inequality for checking intersections prevents both line segments from being counted. In this case, line segment 2 is counted and segment 1 is ignored.

4.2.3 Color Mapping

Color is used to identify how many times anomalies appeared at specific locations on the part. The color mapped to each location is determined by the final count associated with each grid cell midpoint. Cell midpoints which have a final count of ‘0’ receive no color, while all other midpoints are painted blue, green, orange, yellow, or red.

Each of the colors represents a specific range of frequency count. This range is determined by subtracting the lowest frequency count from the highest and dividing the value equally into the five color categories. For example, if the maximum count is 60 and the minimum is 10, each color would include a range of 10 values. Figure 12 shows the scale created in this situation. Colors were chosen to match similar outputs found in finite element analysis (FEA) and solidification software.

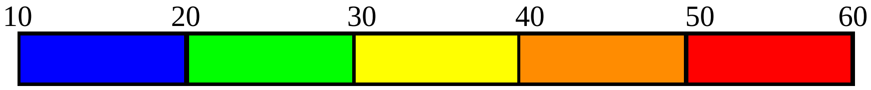


Figure 12: Frequency plot color scale based on example values.

The frequency plots developed here provides many benefits to foundries. Visually, they describe the location, severity, and type of anomalies on a particular part. This information can be used in multiple ways. Foundries can identify variances within their inspection process. This could include differences between operators, between stations (lighting), or equipment (dry or wet mag). Secondly, the output can be utilized for analyzing the effect of process changes. As an example, consider the case where a foundry wants to determine the impact of pouring temperature. This type of analysis would require two separate frequency plots. One, with a data set from parts poured at a 'high temperature range' and the other from parts poured at a 'cold temperature range'. The foundry could then compare these two plots for trends in anomaly location, type, and severity. Later in this paper a more quantitative method for measuring the effect of process parameters is discussed.

4.3 Anomaly Area

The size of an anomaly is an important quality measure calculated from frequency plot results. It is computed by counting how many grid cells are located inside an anomaly polygon. This frequency number is then multiplied by the grid cell size to determine the area. For instance, if the grid cell is one square inch in size, the area would equal the count (Figure 13). Area calculations are all 2D estimates of a 3D polygon.

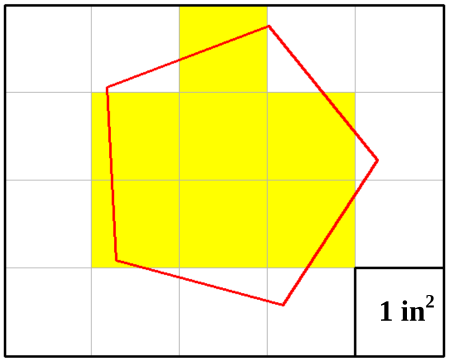


Figure 13: Anomaly area is based on the number of grid cells inside the polygon region. In the example above, the number of cells is equal to 7. Assuming the grid cell size is 1 in^2, the anomaly area would equal 7 in^2.

The total anomaly area can also be calculated for a group of parts and displayed as a histogram. This type of output is called a 'Total Anomaly Area Breakdown' and is used to identify parts which have significantly lower or higher values with respect to the average. An example of this is shown in Figure 14. By identifying the outliers, quality personnel can focus on discovering differences in the processes used to make those parts. Upper and lower control limits can also be placed on the graph for easier identification of parts with abnormal values.

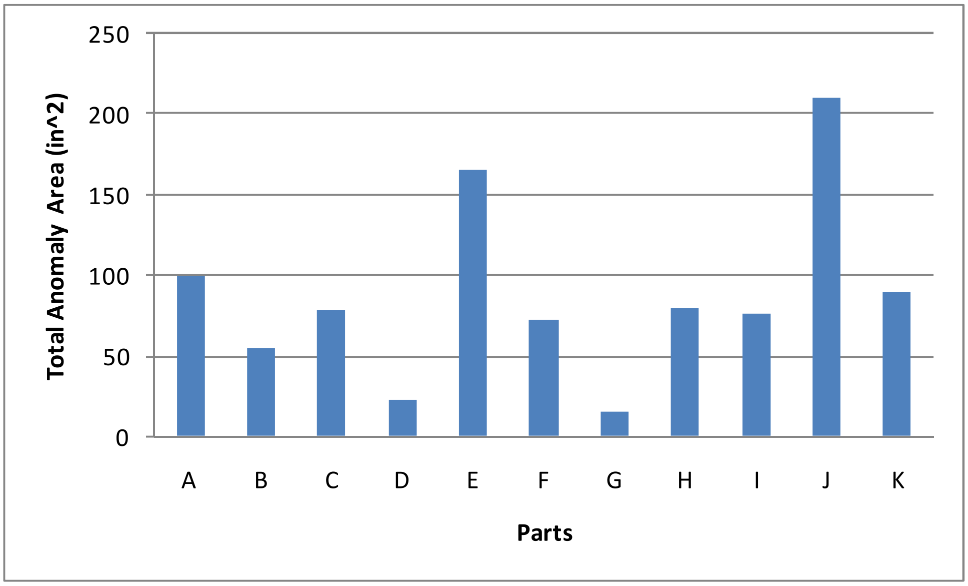


Figure 14: Total anomaly area breakdown histogram.

4.4 Correlations

Correlations are used to quantify the strength of relationship between different variables. For this study, three different types of correlations are shown to illustrate this analysis tool:

- *Process Parameter vs. Anomaly Area*
- *Process Parameter vs. Anomaly Location*
- *Anomaly Location vs. Anomaly Location*

The purpose of the correlation module is to generate data sets for analysis in regression models. All data sets created with the program are easily exported for evaluation inside commercial statistical software. The focus of the following sections will discuss how these data sets are created for each correlation type.

4.4.1 Process Parameter vs. Anomaly Area

The first correlation example attempts to identify process variables impacting the total anomaly area of the part. To generate the appropriate data set, the total anomaly area must first be calculated. The procedure for calculating this area was discussed previously in Section 4.3. Process variables being investigated must be pulled from existing foundry databases. A hypothetical dataset is shown in Table 2. Notice each row represents a specific part with a pouring temperature and total anomaly area value. From here, the data set is readily exported to a statistics software program for undergoing regression analysis. The example data set only contains one process variable, but additional ones could be added for multiple regression analysis. Since there are many process variables influencing the quality of the part, it may be best to include more than one.

Table 2: Hypothetical data set for Process Parameter vs. Total Anomaly Area correlation.

Casting	Pouring Temperature	Total Anomaly Area
#100	2800	20
#200	2820	35
#300	2810	30
#400	2890	50
#500	2880	45

A unique aspect of this correlation is the ability to use data from multiple types of castings. This is done by dividing the total anomaly area by the total surface area for the casting. The percentage which results can be used for identifying process variables impacting total anomaly area on more than one casting type. For example, the time of year may affect the permeability of the mold. This could cause more porosity to appear across all parts.

4.4.2 Process Parameter vs. Anomaly Locations

The second type of correlation aims at discovering process conditions causing the formation of anomalies at specific regions on the casting. In contrast to the macro approach of the first correlation, this type allows users to narrow in on locations of interest. This in turn increases the possibility of discovering why certain anomalies occur at specific locations.

To conduct this analysis, users must first determine what location to pursue. This is accomplished by creating a frequency plot for a casting of choice and examining the output. The output of the frequency plot visually shows which regions have high, medium, and low severity through color. From this information, users select a location of interest. For example, if the user wants to determine what causes anomalies to randomly appear, they can select a location with low to medium severity. Regions with unexplained high severity also make good candidates.

Figure 15 shows an example of a frequency plot for a particular casting. From this image, locations having different levels of severity are easily identified. Region #1 represents a location with randomly appearing anomalies, while Region #2 is a location with high concentration of anomaly activity.

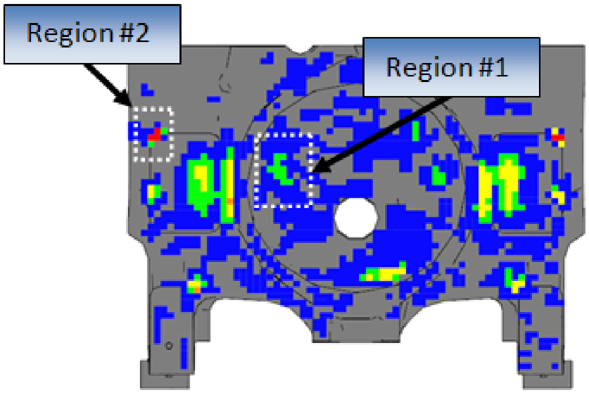
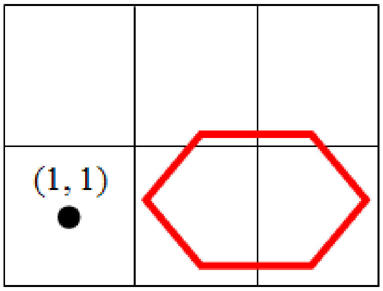


Figure 15: Frequency plot output for identifying locations of interest.

A location of interest is then chosen and inputted into the software program. Details on how the location is entered can be found in Section 3.1.3. The ensuing computations then determine how many times an anomaly was present in the particular location. How these computations work is discussed during the frequency plot section of this paper (4.2). The output however, is slightly different then frequency plots. Instead of summing the data together, details on each individual part are preserved. This includes a binary response for every point tested within the location. A point receives a ‘1’ if it falls inside an anomaly, ‘0’ otherwise. An example data row for one individual casting is shown in Figure 16. The particular part had one anomaly in the location chosen, resulting in two points receiving a value of “1”. In reality, the data set for this correlation would contain many more rows for including more parts.



Part No.	(1,1)	(1,2)	(2,1)	(2,2)	(3,1)	(3,2)
#200	0	0	1	0	1	0

Figure 16: Left: Location divided into six cells with red polygon representing an anomaly region. Right: Sample data set created from this situation.

There are two options for continuing with this data set. The first option is to leave the data in the current format and perform a logistic regression analysis. This type of analysis provides information on the probability an anomaly will occur anywhere in the region given a change in the process variable. Again, this is easily calculated in a statistical software program. The other option is to use traditional linear regression models for noticing relationships. Before this can occur however, the data set must be adjusted.

In order to utilize a linear regression model, the binary data is converted to a continuous variable. If left in the current binary format, the results from the regression model would not be accurate. Linear regression models require data which is continuous and unbounded. Binary variables violate both of these conditions. A fast and efficient method for manipulating the data from binary to continuous is to sum the values for each part. The sum of these binary variables is then multiplied by the grid cell size for representing the total anomaly area for the region. Table 3 below shows a sample data set before and after the binary values are converted. This new data set no longer violates the use of a linear regression model.

Table 3: Data set with binary values summed to total anomaly area.

Part No.	(1,1)	(1,2)	(2,1)	(2,2)	(3,1)	(3,2)	Anomaly Area
#200	0	0	1	0	1	0	2
#300	1	0	0	1	1	1	4
#400	0	0	0	0	1	0	1

4.4.3 Anomaly Location vs. Anomaly Location

Type three of the correlations investigates if a relationship exists between anomalies appearing at two or more distinct locations on a casting. The procedure for selecting the locations involves analyzing the frequency plot for regions with similar levels of severity. The program then creates a data set in the same fashion as the previous type. Once the data set is generated, the binary data must be converted to total anomaly area as before. An example data set is shown in Table 4.

Table 4: Hypothetical data set for Anomaly Location vs. Anomaly Location correlation.

Part No.	Location #1 (Anomaly Area)	Location #2 (Anomaly Area)
#100	14	9
#200	7	11
#300	0	17
#400	12	10
#500	0	0

Parts are then grouped based on the presence of anomalies in the different locations. These three categories are:

- Parts with anomalies at both locations
- Parts with anomalies at one location
- Parts with no anomalies at either location

The number of parts in each group is then used for calculating percentages. These percentages represent the probability an anomaly will appear at one location, given an anomaly was found at the other. Table 5 shows the final data set for calculating these percentages. According to the data, there is a 75% probability an anomaly will appear in both locations if it appears in one. It can also be estimated that the anomaly appearing in the second location will be approximately 9.4 in^2. The purpose for doing this type of analysis is identifying anomalies which share the same root cause, making the appropriate process change, and eliminating both.

Table 5: Data set with parts categorized.

Part No.	Location #1	Location #2	Both	One	None
#100	14	9	1	0	0
#200	7	11	1	0	0
#300	0	17	0	1	0
#400	12	10	1	0	0
#500	0	0	0	0	1
		Total:	3	1	1

After completing this analysis, users may want to pull in process parameters used to make the casting. This provides the data set necessary for performing a Process Parameter vs. Anomaly Location correlation. By knowing the relationship between two separate locations, the result from one correlation may identify the process variable influencing both. Therefore, a change to eliminate one anomaly from occurring may in fact eliminate two or more. The example described here only looks at two locations, but there is no limit to how many locations can be analyzed. The same procedure can be applied.

4.5 Principal Component Analysis

Prior to this section, the majority of analysis tools were designed to be quick, simple, and efficient measures. This section describes a more advanced application available for use with the data set. The statistical tool described in this section is called Principal Component Analysis (PCA). It is commonly used for reducing multi-dimensional data sets into lower dimensions for analysis purposes. In other words, it is used to simplify data sets into as few variables as possible while still maintaining the data's meaning. The intended purpose for applying PCA is to characterize and group parts by their anomaly flaw pattern.

This type of application is well suited for the data set generated by the program. Each casting is characterized by a vast amount of points depending on the size. For example, if a casting is 4' x 4' x 4', the number of specific locations would approximately be 45,000 (assuming cell size is 1x1"). This means the data set would have 45,000 coordinates with binary responses for characterizing one part flaw pattern. PCA tries to describe the same data in significantly lesser amount of variables. As was the case with correlations, the main purpose of the module is to present the data in the appropriate format for analysis. This data is readily exported to statistical packages for the actual principal component analysis. A sample data set is shown below in Table 6. This simplified set contains only 3 variables (locations) for 10 parts. In a real application, this number is significantly greater. Two principal component values were generated from this data set. Note how the values are identical for parts with identical flaw patterns.

Table 6: Data set for principal component analysis.

Part No.	Point 1	Point 2	Point 3	PC 1	PC 2
#100	1	1	0	1.40	0.72
#200	1	1	0	1.40	0.72
#300	0	1	0	0.51	-0.99
#400	0	1	0	0.51	-0.99
#500	0	0	1	-1.91	0.26
#600	0	0	1	-1.91	0.26

In brief, PCA works by obtaining principal component values for each part in the data set. A data set with 'n' number of parts has the potential of creating 'n' number of principal components. The goal however is to use the least number of principal components for representing the original data set. If all possible principal components are used there is no real advantage. For data from this program, principal components will replace the large set of part coordinates.

The procedure for transforming the original data into a set of principal component values involves first normalizing the data set. Once the data is normalized, a covariance matrix is created. Covariance matrices describe the relationship between all points with respect to each other. A set of ‘n’ points, results in an ‘n’ x ‘n’ matrix, where the diagonal values are the variance for each individual point. Eigen vectors are then calculated from this covariance matrix. These vectors are unique, because when multiplied by the covariance matrix the direction of the vector does not change. In fact, the vector is either scaled up or down. The amount the vector changes size is given the term eigen value. The higher the eigen value, the more variance it accounts for in the data set. The eigen vectors for the covariance matrix are placed inside their own matrix. The original data set is then taken in linear combination with this matrix for determining principal components. For example, five eigen vectors results in the creation of five principal components.

Principal components account for a specific percentage of the total variance in the original data. The first principal component contains the majority of variance since it is created by combining the original data with the eigen vector having the largest eigen value. For instance, the first value may describe 25% of the total variance in the data set. Subsequent components account for greatest variance not correlated to previous principal components (Jolliffe, 2002). The effect is cumulative. Two component values by themselves may contain 24% and 8% respectively, but together they represent 32% of the variance in the data.

Identifying parts with irregular flaw patterns or groups of parts with similar ones is then accomplished by sorting the data by a particular principal component. Irregular flaw patterns appear in parts which have abnormally high or low principal component values. To determine their flaw pattern, a frequency plot for the specific part must be created and compared to the frequency plot for all parts. Parts with similar types of flaw patterns will contain principal component values relatively close to each other. Typically, only a few principal component values need to be analyzed to discover these types of trends. A good measure for knowing when to stop analyzing is when successive principal component values identify the same parts as being outliers. Principal component values can also be used in correlations with process variables for identifying process conditions which result in a particular flaw pattern. An example of a PCA is included in the Chapter 5 case study. It is important to note that PCA was not designed for use with binary variables, but has proven to work remarkably well in this application.

Chapter 5. Case Study

During the summer of 2008, a case study at Harrison Steel Casting Company in Attica, Indiana was completed. The goal was to implement the software program in a steel casting foundry and begin collecting and analyzing data. Harrison Steel was selected in part based on their established practice of recording surface quality information during inspections. The products produced by this foundry range from 300 to 12,500 lbs and are made from carbon and low alloy steel.

The foundry's quality department is in charge of collecting data on casting surface quality as marked by inspectors. To collect this data, operators are positioned at inspection stations for recording anomaly locations on a 2D drawing. Operators denote the anomaly type by the shape of the marking and classify its location (interior or exterior) through color. Other valuable information such as the serial number, inspector, and detection method are written down as well. An example of one of these data collection sheets is shown in Figure 17. Data is collected under these conditions for every first time part inspection, as needs arise, and on a continual basis for some castings. All data sheets are filed away and kept for later reference upon request.

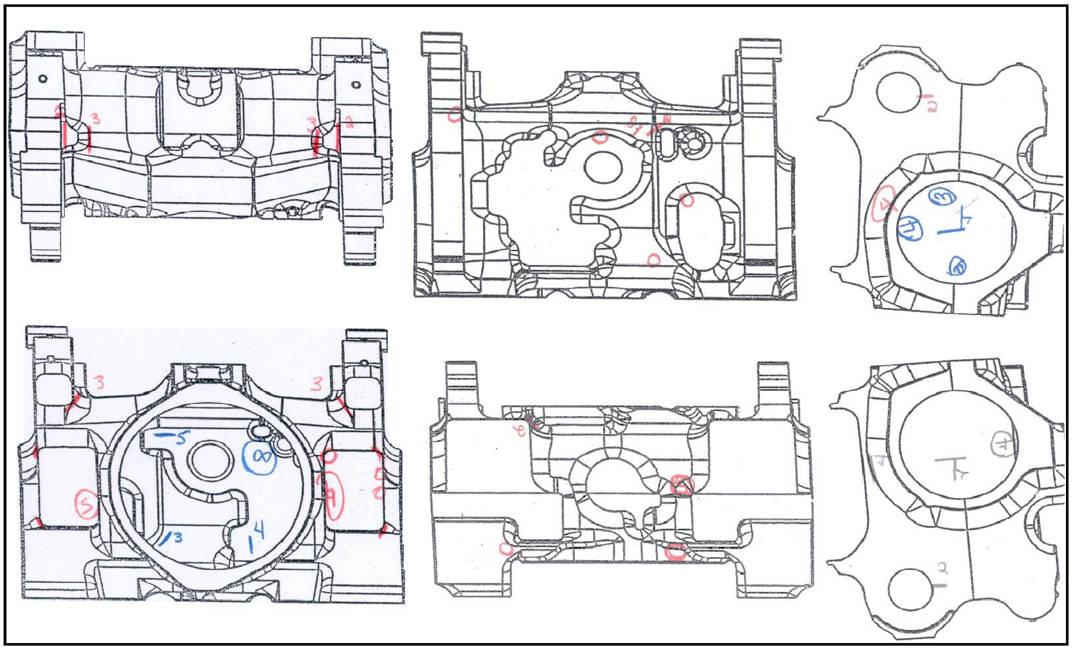


Figure 17: Data sheet for casting surface quality information at Harrison Steel.

5.1 Methodology

Five castings were chosen based on interest from the foundry in obtaining a data set for making process improvements. These castings ranged from 2000 to 10,000 lbs and were manufactured on two separate production lines. The castings chosen were regularly produced.

Surface quality data was collected by transferring the information recorded on data sheets (See Figure 17) to the software program. The time involved transferring the data varies on the number of anomalies present. Each anomaly takes approximately 10 seconds to enter into the program. Average anomaly size recorded was around two square inches. Examples of how the data appears before and after being entered into the program are shown in Appendix B.

The data captured at the foundry was stored in a Microsoft Access 2007 database located on a single computer. At the end of the 10 week study, over 300 parts were entered with a minimum of 30 for each of the five castings. Analysis of this data was completed using information from the anomaly database and from current part databases already on the foundry's SQL server.

5.2 Results

Analysis of the data collected during the case study was conducted in the subsequent months after leaving the foundry. Sample results are used to illustrate the types of benefits foundries can achieve. The subject matter selected for analysis was chosen by the author and is by no means a complete list of what could be investigated using the data. In most cases, the examples require a larger data set with further analysis for making broad conclusions. Some results were modified to protect the confidentiality of the data, without altering its meaning.

5.2.1 Variation in Inspection

The first example shows how frequency plots are used to identify valuable information on the inspection process. The focus of this study was to determine if there was any noticeable difference in parts inspected by one operator compared to another. To accomplish this objective, two frequency plots were created. One using parts inspected by Operator A, the other with parts by Operator B. Both outputs were then visually contrasted to identify differences in anomaly cluster locations, frequency, and area.

Visual comparison of the two frequency plots identified a casting region with a significant discrepancy in frequency count.   Operator A marked anomalies in the region over 80% of the time, while Operator B, only found anomalies 17% of the time. Figure 18 displays the frequency plots derived from each operator. The region of question is circled on each image for reference.

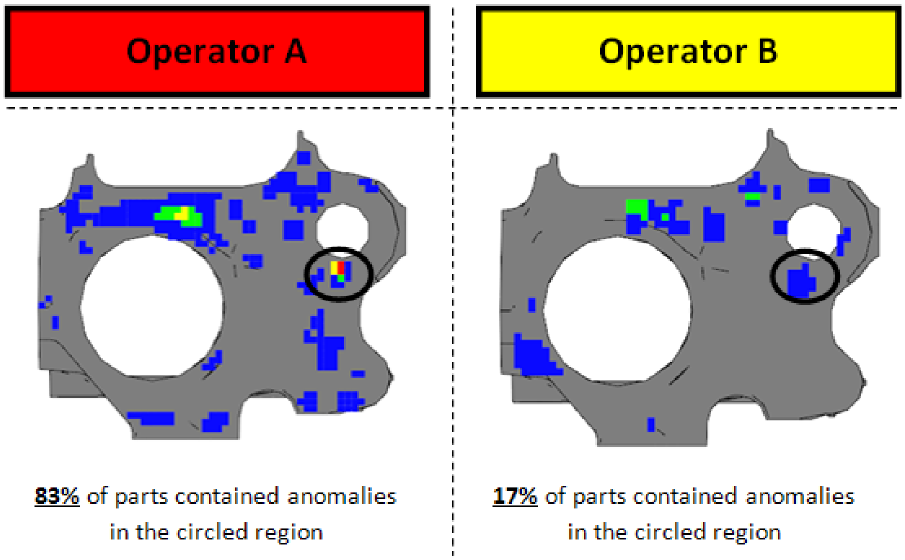


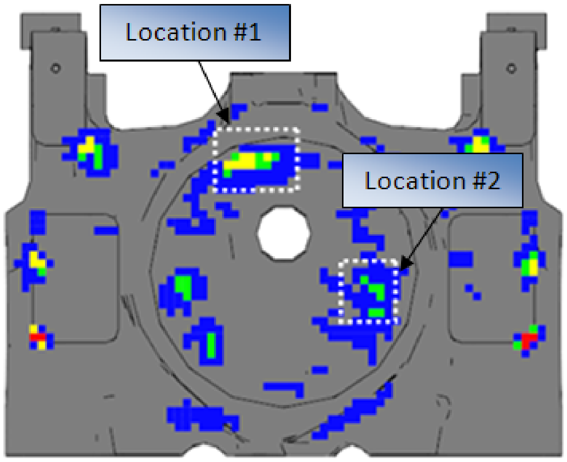
Figure 18: Frequency plot showing casting region containing large discrepancy between operators.

The following steps concentrated on what may have caused this variability. This included comparing the process variables used to create the parts inspected by each operator. When no clear trend presented itself, investigation switched to the process used by each operator for inspection. Items such as inspection location, inspection equipment, experience level, and work shift were targets for comparison. The inspection location and equipment were identical, but Operator A had far greater experience inspecting this particular casting. The question then becomes is Operator B insufficiently trained or is Operator A over inspecting. Time did not permit during the case study to collect more data and analyze this further. Operator measurement error was also not a specific focus of this project. It does however provide a powerful example of what can be identified through frequency plots.

5.2.2 Anomaly Correlations

The next example describes how the correlation module was used to identify the relationship between cracks appearing at two distinct locations. Figure 19 below highlights the two locations of interest on this casting. These locations were selected based on their similar levels of severity.

The results from the study are also displayed in Figure 19. From these results, there is a 60% probability of cracks appearing in both locations. This provides motivation to determine the cause for cracks in one location with the hope it may be the reasoning for both. The stronger the relationship between anomalies at separate locations, the greater the likelihood they share the same root causation.



Both	One	None
29	17	2
60.4%	35.4%	4.2%

Figure 19: Left: Frequency plot with two locations of interest highlighted. Right: Results from correlation study.

The process variables used to make the parts along with their anomaly area for the two locations was then analyzed. The analysis was capable of determining which parameters had no influence, but was unable to identify ones having significant impact. It is likely then, the cause of the cracks may be related to the design of the casting. This could include anything from the casting’s geometry to the gating system. Experiments would be necessary to validate this assumption. While the module provides foundries with information on potential relationships, it cannot prove causation.

5.2.3 Casting Classification

The final example shows how principal component analysis (PCA) is used for classifying parts by their surface flaw pattern. PCA was conducted on a data set including 80 parts of the same casting. To improve performance, the data set was first reduced. This was accomplished by limiting the analysis to only one side of the casting. Points on the side which never contained anomalies were also removed to further reduce the data set. Even with this reduction, there was still over 1500 points for describing the flaw pattern for one part.

The final data set used for analysis was created by the software program and contained 80 rows and 1500 columns. Details on how this set is generated can be found in Section 4.5. The frequency plot output from all 80 parts is shown in Figure 20. From the output, the typical flaw pattern for this casting can be identified as regions colored in red, orange, and yellow, representing high anomaly frequency. Atypical flaw patterns consist of regions with low frequency count appearing blue.

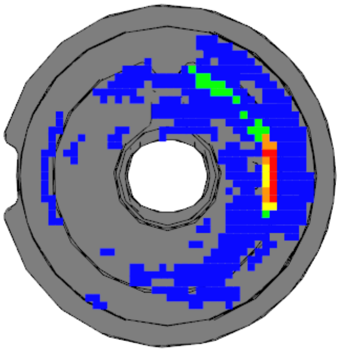


Figure 20: Frequency plot output from 80 parts.

The PCA produces a set of principal component (PC) values for all 80 parts. The number of PC values is equivalent to the number of points representing the flaw pattern. However, it is not necessary or valuable to analyze them all. Each PC value accounts for a certain percentage of variance from the original data set. The PC values used for analysis are those which contain the highest percentage level. In this example, only a handful is needed for grouping the parts effectively. A single PC value for a particular part contains no real meaning. It only becomes practical for analysis when it is compared with others.

PC values are used as criteria for sorting the parts. The parts are initially sorted by the first PC value. This naturally groups parts by their surface flaw pattern as those sharing similar PC values and identifies outliers in the data set. These outliers are of particular interest because they represent parts with unique patterns. Table 7 below provides a list of interesting parts produced from this initial sort.

In this case only two parts contained extreme values. Serial numbers 100 and 531. The first contained a unique flaw pattern and large amount of anomaly area. The second, 135, exhibited a unique flaw pattern when compared to the norm. Parts which received identical PC values were identified as having zero anomalies on the surface for this view. All other parts had PC values ranging from 2 to -2, with a mean value of -1.24. The majority of these parts were noted as having common location of anomaly activity. These common locations are identical to the regions on the overall frequency plot (Figure 20) which are red. The frequency plots from Table 7 are displayed in Figure 21 for illustrating flaw patterns.

Table 7: Results from sorting parts by the first PC value.

Serial Number	PC 1 Value	Comments
#100	77.88	Large Anomaly Area
#531	18.08	Unique Flaw Pattern
#022	-0.93	No Anomalies
#200	-0.93	No Anomalies
#175	-0.93	No Anomalies
#010	-0.93	No Anomalies
#084	-1.24	Typical Flaw Pattern

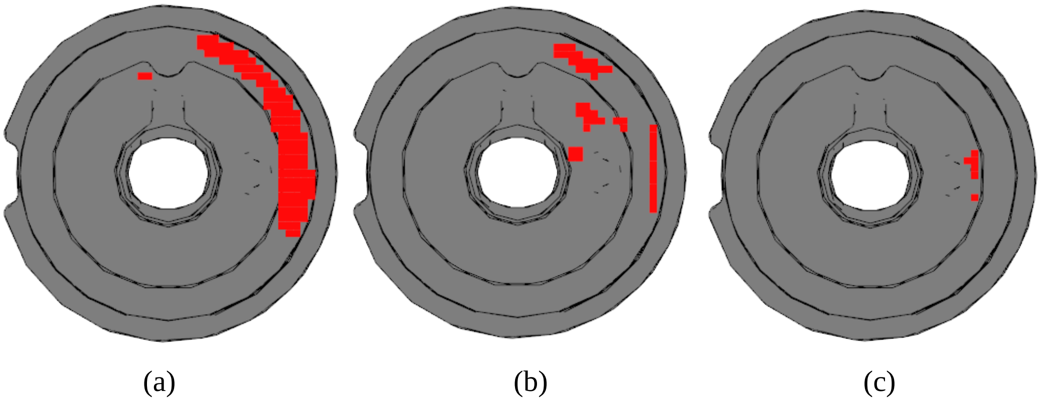


Figure 21: Frequency plot output from parts identified through first principal component value. Images show outputs from # 100 (a), #531 (b), and #084 (c).

The same process of sorting is used for identifying casting flaw patterns through the second PC. Table 8 contains the results from this sort. Again, parts with extreme PC values had unique patterns and parts receiving PC values near the mean shared typical flaw patterns. Figure 22 displays the frequency plots of these parts. The process of sorting by PC values stopped here since subsequent values produced similar results. This shows the benefit of using PCA. Instead of having over 1500 points to analyze for classifying castings, only two variables are needed. This is a 99.87% reduction in data.

Table 8: Results from sorting parts by the second PC value.

Serial Number	PC 2 Value	Comments
#122	-39.29	Unique Flaw Pattern
#150	14.85	Flaw Pattern
#148	14.00	Flaw Pattern
#075	0.05	Typical Flaw Pattern

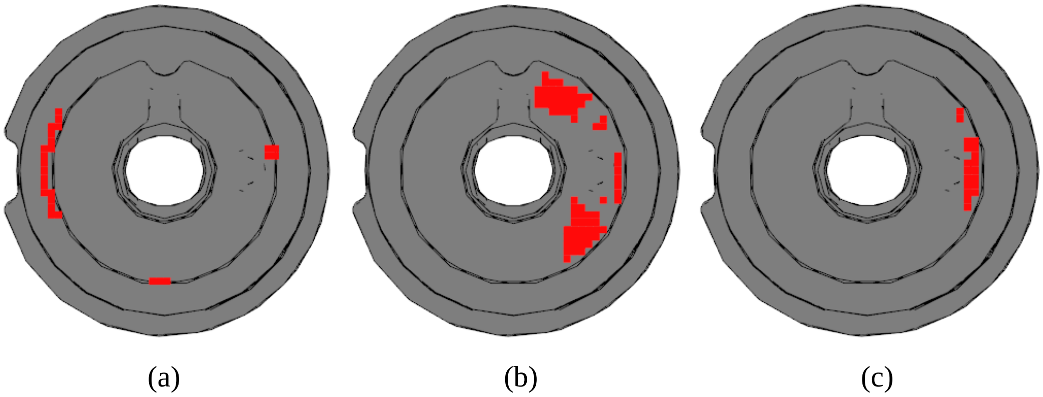


Figure 22: Frequency plot output from parts identified through second principal component value. Images show outputs from # 122 (a), #150/148 (b), and #075 (c).

Up to this point, PCA has been used to group parts by their surface flaw pattern. The focus now switches to identifying what process variables cause each flaw pattern type to occur. Recall, that the PC value contains no real meaning for data analysis. It cannot be used as a continuous variable in a regression analysis for instance. Instead, it is best to give parts containing similar PC values a category name. The process variables used to make the parts in each category can then be compared through ANOVA methods. This would again require a much larger data set before any robust conclusions are made.

Chapter 6. Discussion

This discussion provides an opportunity to examine some of the more important issues related to the study. The first discussion topic explores the impact of grid cell size selection on analysis. Next, the error associated with data entry during the case study is measured and analyzed. Finally, valuable insight gained from implementing the program at a steel casting foundry is presented.

6.1 Selecting Grid Cell Size

Prior to employing many of the analysis modules, users must first specify a grid cell size. Recall from Section 4.2 that grid cells are used to represent specific locations with regards to the casting. The size is entered as a single value which defines the length and width of the cell. For example, a cell size value of 2.00 results in a 2.00 x 2.00 cell with an area of 4.00 in^2. Careful consideration must be made when selecting a size due the implications it has on the output.

First, grid cell size affects the capability of capturing different sized anomalies for analysis. A smaller cell increases the chances of capturing the location of smaller anomalies. Figure 23 provides an example of this. Notice when the grid cell size is large, one anomaly is missed. This is because no grid cell midpoint is located inside the anomaly polygon. When the grid cell size is decreased however, both anomalies can be identified. For identifying anomaly locations, the grid cell size should be closely associated to the smallest range of anomalies marked.

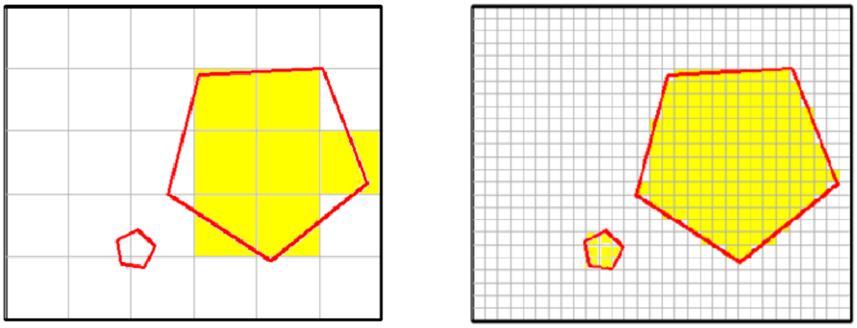


Figure 23: Shaded cells are located inside the anomaly. Left: Smaller anomaly is missed and anomaly shape is crudely defined. Right: Smaller cell size results in both anomalies identified, plus better shape definition.

The size of the grid cell also affects the area estimate for an anomaly. Recall that area is approximated by counting the number of cells inside an anomaly and multiplying by the cell size. A smaller cell will always come closer to the polygon's theoretical area. Proof of this is shown through the images provided in Figure 23. The large polygon in both images has an area of 6.88 in^2. This is the nominal area since it was calculated using the polygon's dimensions. The estimated area of the same polygon using the large cells was 7.00 in^2, resulting in a 1.7% approximation error. The smaller cells estimated the area at 6.84 in^2 and contained a 0.05% error. It appears that either cell size does a good job of approximating the actual area of the polygon with both containing less than 2% error. The use of smaller cells does again result in a better output.

Finally, the cell size determines how well the anomaly shape is represented on the frequency plot output. Grid cells can be compared to pixels used for images. More pixels results in better resolution. Figure 23 visually shows the effect grid cell size has on the anomaly shape. The plot using large cells provides a very crude representation of the polygon, while the plot using smaller cells defines the shape quite well. To produce the best frequency plot images, grid cell size should be set small.

The tradeoff when selecting a small cell size is the increase in computation time. The graph in Figure 24 shows an example of computation time vs. grid cell size. As expected, the graph shows an inverse relationship between computation time and cell size. If time is not a factor when running an analysis, cell size can be set extremely small. A larger size should be chosen in cases where the user wants a quicker response.

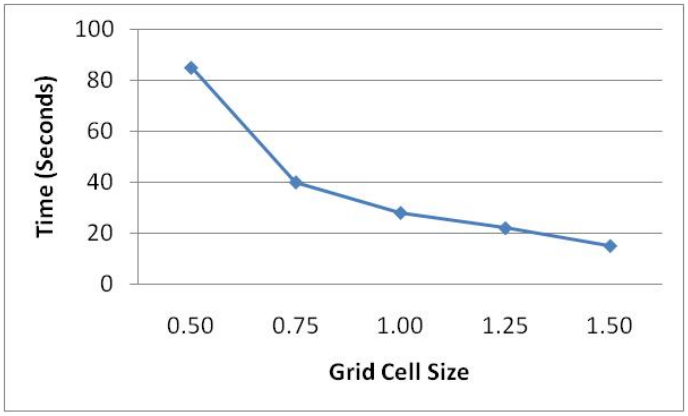


Figure 24: Graph showing computation time vs. grid cell size.

Whether a cell is small or large is all relative to the size of the casting. A starting point for approximating the appropriate cell size is to make it 2% of the smallest casting dimensions. For example, if a casting has dimensions of 60” x 48” x 36” the resulting cell size would be around 0.75” (0.02 x 36”).

6.2 Data Entry Error

A reasonable concern regarding the case study is how much error occurred during data entry. The error in question is the result of transferring the data to the software program. Types of error would include incorrectly marking the size, shape, and type of anomaly, as well as missing anomalies all together.

An experiment was setup to test the error associated with data entry. The setup included a single operator entering the same data sheet three times into the system. The operator chosen for this experiment had previous experience entering data into the program. To remove bias, the operator was unaware of the test and the duplicated data sheets did not appear in sequential order. The type of error measured in this experiment is commonly called repeatability error. Repeatability error is the result of a single operator performing the same task multiple times.

The experiment lasted approximately 50 minutes, allowing the operator to enter data from 6 sheets, 3 of which were identical. All data sheets entered contained the same casting model. This was the largest casting and featured many more anomalies when compared to others studied. It was chosen to represent the worst case scenario for data entry error. The large casting size creates more difficulty in consistently locating small anomalies. While the greater number of anomalies, increases the likelihood one will be missed.

Repeatability error was first measured by calculating the number of anomalies marked in each entry. Each data entry should contain 33 anomalies as recorded on the duplicated data sheet. The operator achieved this value on two of three data entries. On the third entry, the operator only recorded 32 anomalies. The operator correctly marked 98 out of 99 or 96.97% of anomalies. The repeatability error associated with marking the correct number of anomalies is 3%. All anomaly types were correctly identified.

Next, the repeatability error associate with marking the correct anomaly size was measured. This error was calculated by comparing the total anomaly area for each data entry. The results are shown in Figure 25. The standard deviation was 17.21 with a mean of 192 in^2. A 7% average variation in total anomaly area between data entries implies some amount of repeatability error. This however may have been inflated since Entry 3 did not contain all 33 anomalies.

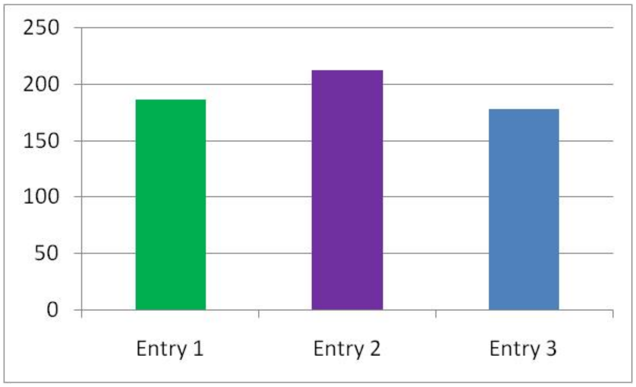


Figure 25: Histogram showing the total anomaly area for each data entry.

The third aspect investigated was the repeatability error linked to anomaly locations. This measured how well the operator marked anomalies in identical locations each time. Frequency plot dated was exported to Excel for measuring this error. In Excel, cell midpoints for representing casting locations were categorized by their frequency count (1, 2, or 3). The total number in each category was then used to determine a percent match. A 100% match refers to cell midpoints which received a frequency count of 3, meaning they were inside an anomaly for all three data entries. A 67% match refers to cell midpoints receiving a frequency count 2. Finally, a 0% match is regarding midpoints inside an anomaly in only one data entry. Ideally, the test would show a 100% match for all midpoints.

Results from this analysis are shown in Figure 26. The operator had a 100% match on 37.41% of the total midpoints. A 67% match on 27.78% of the points and a 0% match on 34.81%. This would suggest there was a significant amount of repeatability error when marking locations. However, the high percentage of midpoints with a frequency count of 1 may be tied to the error associated with marking anomaly size and less with regards to marking the correct location. The grid cell size also has a large impact on the repeatability error with regards to location and size. Smaller cells will be more sensitive to minor changes in anomaly location and size between data entries.

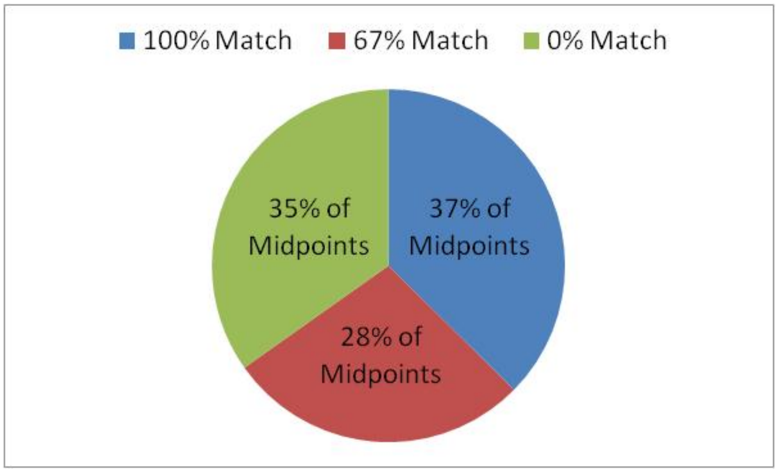


Figure 26: Pie chart showing percentage of midpoints and % Match.

A more important measure is how well the frequency plot compared to the original data sheet. This was evaluated in two ways. First, a visual comparison between the two data formats was completed. The comparison resulted in good agreement between anomaly locations. Next, the number of anomaly clusters in the frequency plot was counted. Anomaly clusters represent regions on the model where anomalies were marked multiple times during multiple data entries (Figure 27). Preferably, there should be one anomaly cluster for every one of the 33 anomalies. If the number of anomaly clusters matches the number of anomalies, then the operator did a good job of defining them in similar regions each time. In this test, the operator performed well by creating exactly 33 clusters.

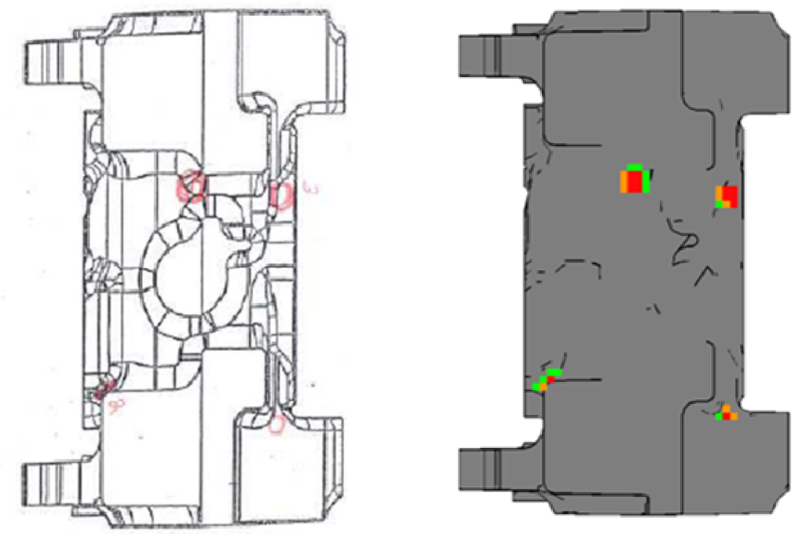


Figure 27: Visual comparison between original data sheet and frequency plot. Note the number of anomaly clusters matches the number of anomalies. For comparing the other five views, reference Appendix B.

The result of this experiment provided two conclusions with regards to data entry. First, it showed that marking anomaly size and location the same way each time is challenging. This was proven by the low 100% match and by the discrepancy between total anomaly areas. In many ways these results were expected since the casting was very large and the anomalies on average were less than 2 in^2. The grid cell size was also very sensitive to minor changes in position and size.

More importantly, the study proved that data entered into the program was capable of representing the original data sheet well. This can be seen in the comparison provided in Figure 27. While there was some repeatability error associated with the location and size of anomaly markings, the operator only created 33 anomaly clusters. This means that each anomaly was marked in a similar region each time.

6.3 Case Study Insights

This section takes a closer look at the first implementation of the program in the industry. It will discuss potential challenges and provide suggestions for future implementations in other facilities. Key insights from case study results will also be discussed.

6.3.1 Case Study Challenges

The majority of challenges faced when using the software program was limited due to the small scale implementation. The software was setup on a single computer and only that computer had access to the database. The only significant concern with this setup was the location of the data. Storing it on a single computer increases the likelihood of a computer failure losing all the data. This was prevented by consistently backing up the data on an external hard drive.

The other challenge was determining when and where to collect the data. There were two options to consider. The first option was to collect data on site at each of the inspection stations. This would require setting up the computer at each location and entering data during the inspection. There were multiple concerns with using this method. First, since the program was installed on a single computer the amount of data entered would be limited. The foundry had six inspection stations spread throughout the facility. Data could only be captured one station at a time. Second, there were concerns with the durability of a laptop computer unprotected in the foundry environment. Dust, vibration, and impact could cause complete failure of the computer.

The second option was to bring the data to the computer in the form of data sheets. These sheets were the current method for collecting surface quality data by the foundry. Multiple operators were trained to collect information in this way at many inspection stations simultaneously. This option was chosen over the previous since more data could be entered and the computer was protected. It does not represent the best method for a full implementation of the program. If more computers were available and could be protected inside the foundry, data would have been collected on site. Entering the data on paper and then transferring it to the software program creates non value added waste. It also increases the likelihood for error.

6.3.2 Suggestions for Full Scale Implementation

The lessons learned during this small scale implementation can be applied to a full scale implementation in the future. A full scale implementation would require a computer at each inspection station with networking capability. The computers would also need a protective case for lasting in the foundry environment. A server for storing the data collected inside the program would also be required.

Training for operators entering the data is also necessary. This was not required during the case study since the data was entered by the author. Basic computer knowledge is all that is necessary for using this program. Therefore, the training should be minimal.

While it would be preferred to collect data on every part produced, this is not feasible in the majority of foundries. Castings with good quality records (based on customer complaints) should have data collected on an audit basis. This will provide data for analysis if something changes with regards to the casting's quality. The amount of data collected on other castings should be based on the quality problems they are experiencing. If a casting is causing lots of problems, more data should be collected. It is also recommended that data be collected on the same part during multiple inspections. This will provide data for foundries to use when analyzing their cleaning room process. They can identify what anomalies are marked on the 2nd or 3rd inspection, as well as how much rework is occurring in their system.

6.3.3 Implication of Results

The results from the case study are very reflective of the foundry chosen. Harrison Steel is one of the top foundries in the industry for data collection and process control. They collect a significant amount of data on every casting produced that is used for problem solving through root causation.

Providing them with a software program for collecting and storing the location of anomalies just adds to their impressive data set available for analysis. Harrison's past dedication to quality prevented the software program from identifying quick fixes to their process.

Analysis of this data set did produce some very interesting results. First, it showed there was a significant amount of variation between two operators inspecting the same casting. Secondly, it suggested cracks appearing at two locations on the casting were related. Thirdly, PCA was proven effective at grouping castings by their surface flaw pattern. Unfortunately, no broad conclusions could be made. This was true for other analyses not presented in this paper.

There are many reasons for why conclusions could not be made at this time. The lack of a sufficient data set on anomaly locations was the first reason. Collecting data on only 50 parts does not provide a very large set for analysis. It does supply enough data for creating accurate frequency plots, but not enough for identifying strong relationships in correlations. Secondly, when it came time to analyze the data, no resources were readily available at the foundry for conducting experiments. Experiments would be necessary in many cases for validating the relationships identified during analysis.

Casting quality issues take a significant amount of data and time to identify root causation. Sometimes it requires data not currently collected. Even with good data collection methods and process control it may still take weeks or months to make any broad conclusions. The time to pick a single casting and fully analyze it using the software program was not available inside this particular case study.   Instead, data was collected on a variety of castings for illustrating the types of analysis possible from the new software program.

Chapter 7. Conclusions

The software program presented here, along with the new set of analysis modules, provides foundries with a much needed tool for controlling their manufacturing process. Process parameters effect on anomaly type, location, and severity can all be calculated. Relationships between variables can now be quantified and used for making appropriate design changes to improve initial part quality. Foundries can also gain a greater understanding into their inspection process. This includes identifying variance between inspectors, equipment, and inspection locations.

The development of new analysis tools provides motivation for collecting it on more parts and during multiple stages in the cleaning room process. This promotes a more proactive approach to investigating and eliminating surface quality problems. Collecting information prior to a new anomaly occurring also helps with route cause analysis, which may not be achieved otherwise. For example, comparisons can be made to see the effects of the season, pouring temperature, pattern number and many other process parameters. The data is also valuable for validating and improving defect simulation software.

Not only will this software program benefit foundries but their customers as well. Foundries can provide customers with information on where the part was welded (weld maps), crack locations, and other details. By supplying management with a new tool for collecting and analyzing inspection data, better insight can be gained into the entire steel casting process. Excessive “rework loops” can be analyzed and the true production cost of a casting can be estimated better.

The steel casting industry was the focus for this project; however, the application of the software program is not limited to this industry. Any industry interested in improving quality through tracking surface quality data will benefit from the program. The data input interface and analysis modules are capable of collecting and analyzing data related to all types of parts.

Chapter 8. Future Work

Future work on this project will focus on two areas of the software program. The first focus will be on the data input interface. This could include modifying the interface functions to increase the speed of data collection. Potential work may also involve investigating automated data collection methods to reduce the amount human interaction and data entry error.

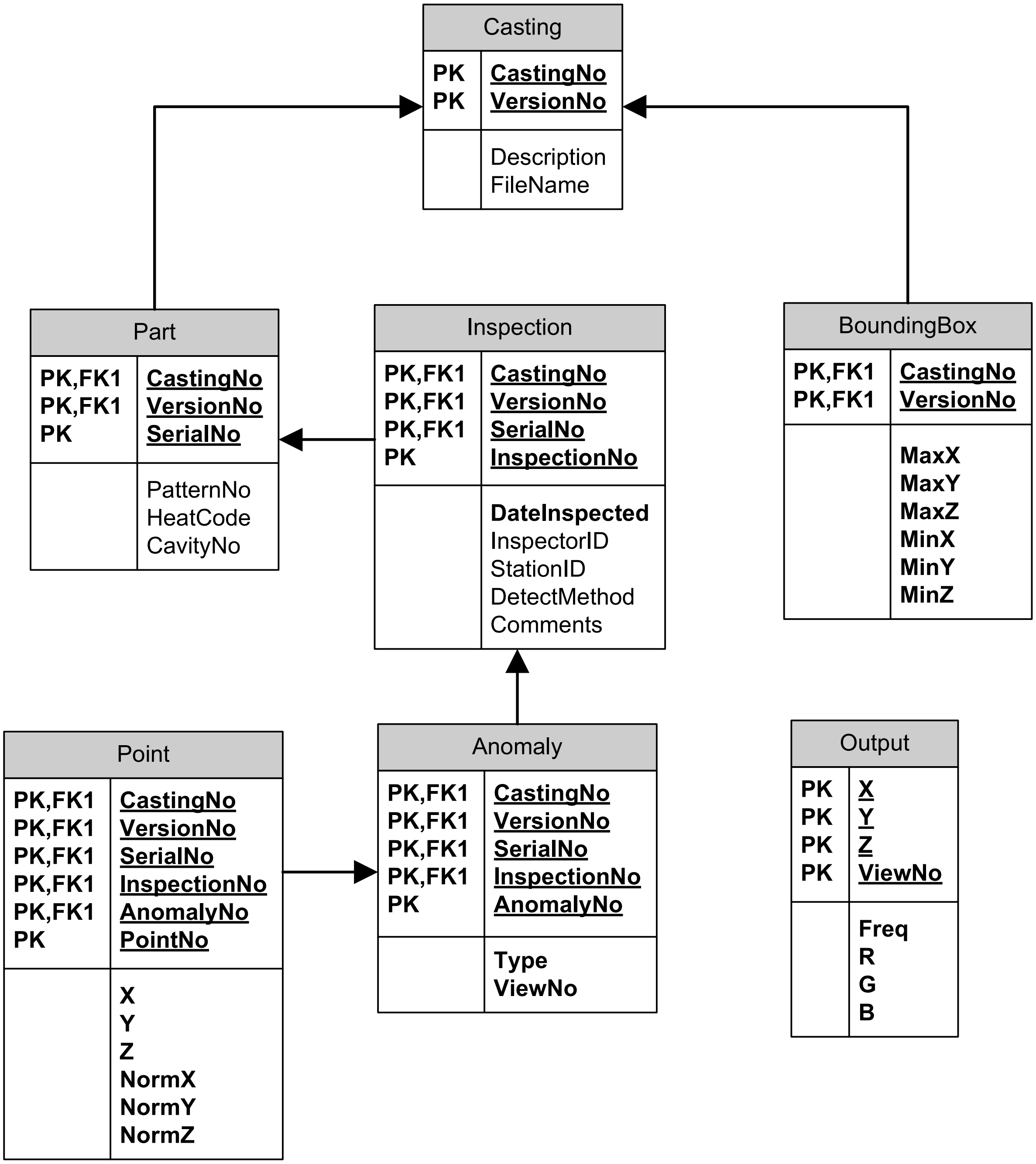
The second focus is on the development of new data analysis tools and improvements on the current set. The current set of analysis tools are forced to convert the 3D data collected into other formats due to limitations inside Adobe. If these limitations were removed, the analysis tools could be modified to directly use the 3D data. The analysis modules considered for future program versions include:

- **Gage R&R:** The Gage R&R module would provide foundries with the ability to quantify the performance of their inspection process. It would contain an easy to use interface for setting up Gage R&R experiments and calculating the variation.

- **Weld Maps:** Creating weld maps for a customer is currently available with the program, but has not been integrated yet. Weld maps describe the location on the part where welding occurred. Many customers are beginning to require them and this program would provide a great method for meeting the customer's request.

- **Cost Calculations:** The cost calculation module would take the anomaly area and multiply it by a "dollar per inch" rule. This "dollar per inch" rule is the cost associated for cleaning a particular type of anomaly of a certain size. Other information such as the amount of weld wire could also be stored to connect the anomaly type/size to operation cost. This is a new cost metric for management to use in determining project priority and process change results.

Appendix A. Database Diagram



Appendix B. Data Entry Examples

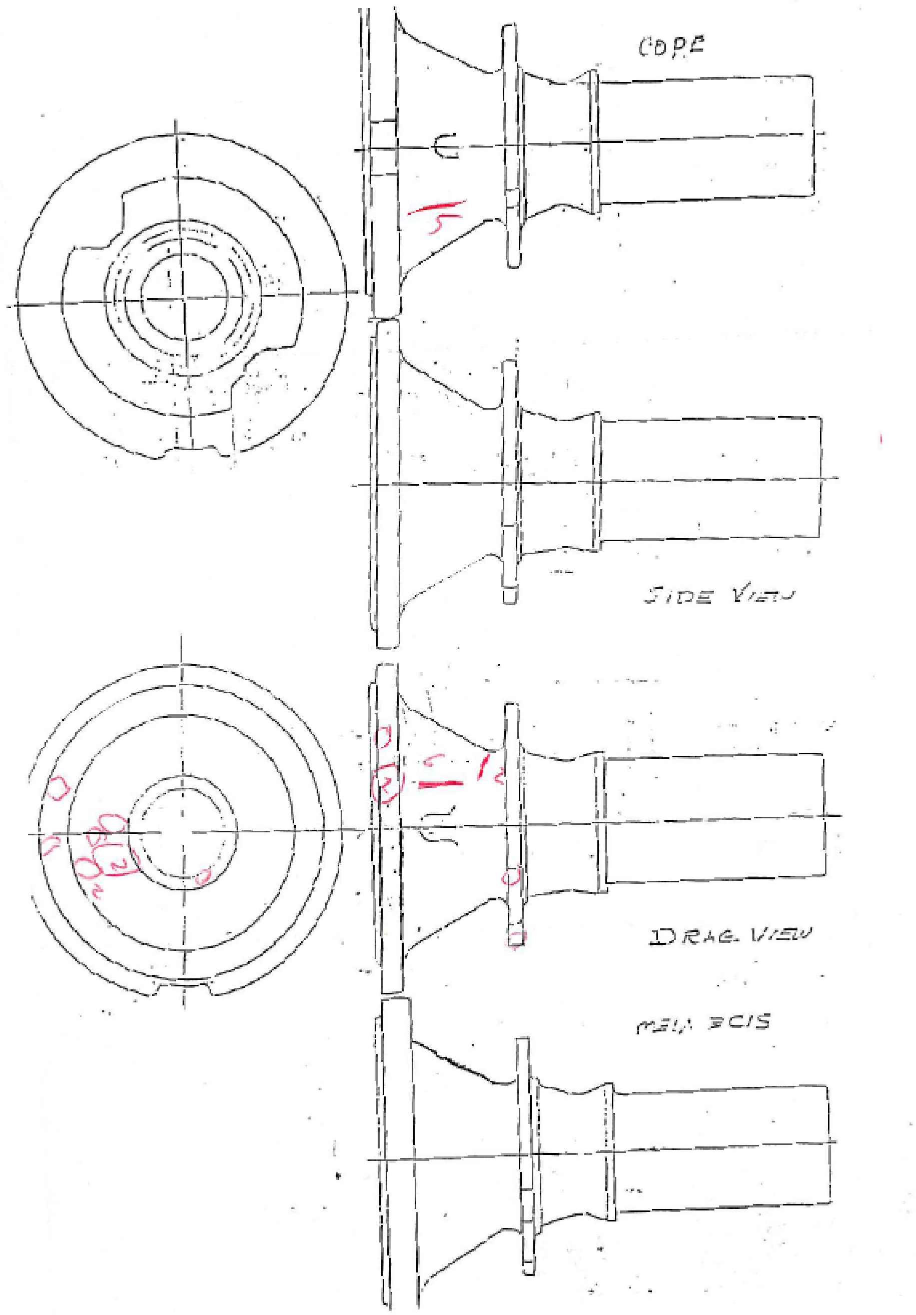


Figure 1: Data collection sheet for casting at Harrison Steel Casting Company.

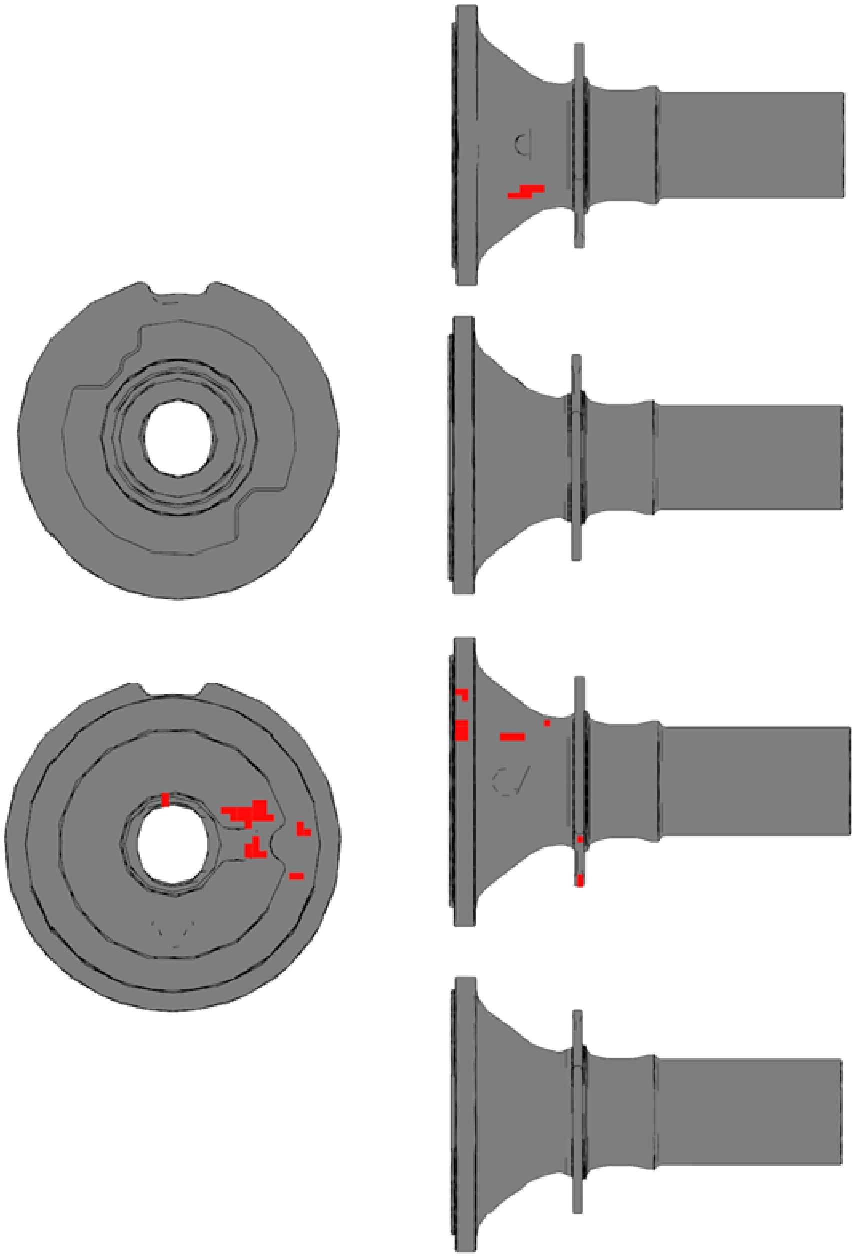


Figure 2: Frequency plot showing casting surface data transferred from data collection sheet shown in Figure 1.

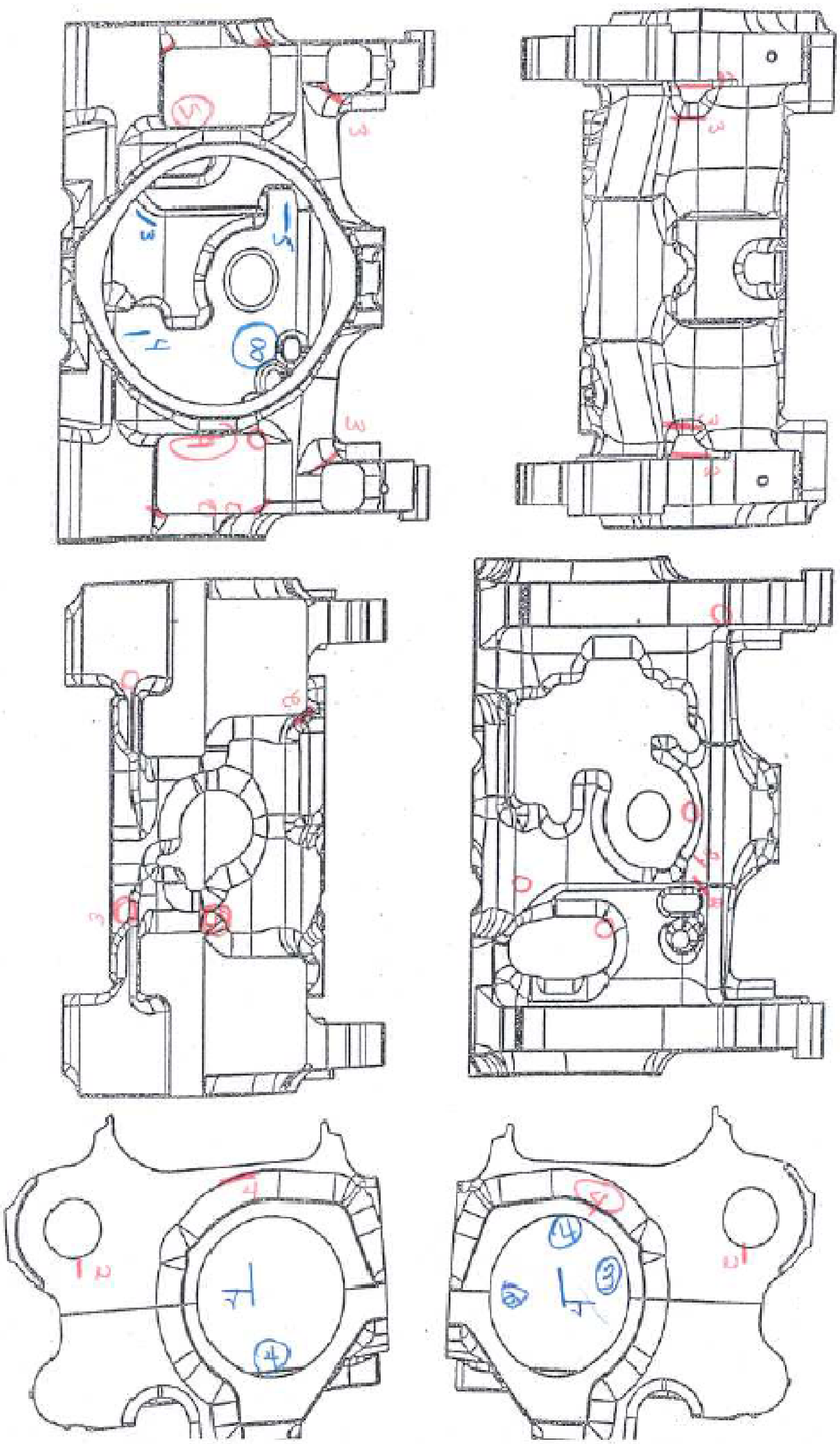


Figure 3: Data collection sheet duplicated for data entry error study.

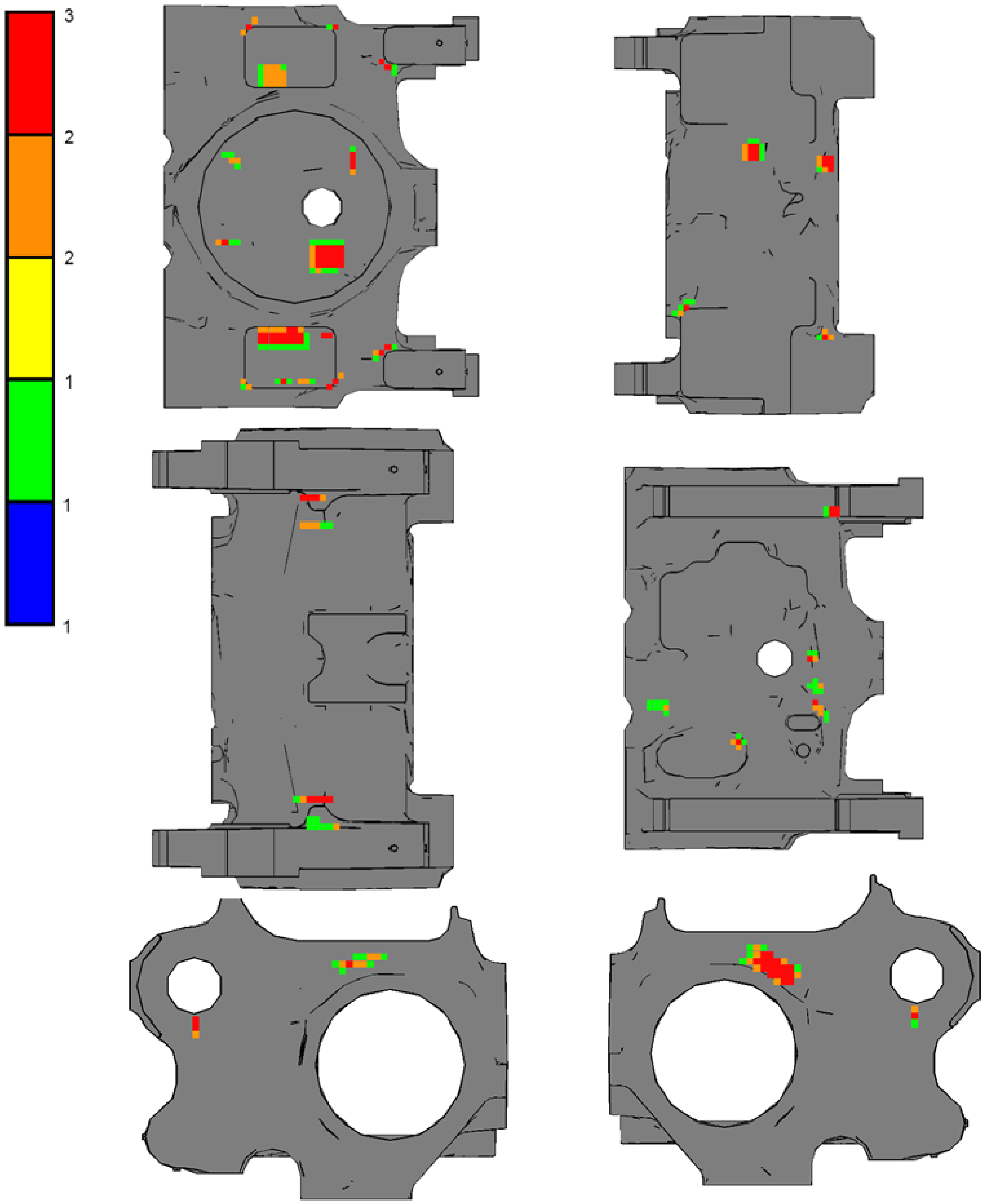


Figure 4: Frequency plot output from data entry study. Frequency plot comprised of the three duplicate data sheets.

Appendix C. VB.NET Code

VB.NET Code for Histogram Output

```
Private Sub Histogram ()

Dim CastingNo As String
Dim VersionNo As String
Dim i As Integer
Dim Histo As String
Dim Value As DataTable
Dim Value2 As Integer


CastingNo = Form1.txtCastingNo.Text
VersionNo = Form1.txtVersionNo.Text

Chart.ColumnCount = 7
Chart.RowCount = 1

         For i = 1 To 7

            Chart.Column = i
            Chart.Row = 1
            Chart.RowLabel = "Anomaly Type"

            If i = 2 Then

                Chart.ColumnLabel = "Dirt"

            End if

            If i = 3 Then

                Chart.ColumnLabel = "Sand"

            End If

            If i = 4 Then

                Chart.ColumnLabel = "Gas"

            End If

            If i = 5 Then

                Chart.ColumnLabel = "Shrink"

            End If

            If i = 6 Then

                Chart.ColumnLabel = "Rework"
            End If



            If i = 7 Then

                Chart.ColumnLabel = "Cracks"

            End If


            Histo = "Select Count(AnomalyNo)As cValue From Anomaly Where
            CastingNo='" & CastingNo & "' And VersionNo='" & VersionNo &
            "' And Type='" & i & "'"

            Value = aDB.Query(Histo)

            Value2 = Value.Rows(0)("cValue")

            Chart.Data = Value2

        Next i

Form2.Show()

End Sub
```

VB.NET Code for Frequency Plot Calculations

```
Public Sub Run()
‘Program functions by evaluating each view of the casting separately.

For View = 1 To 6

      Call CreateGrid()

      Call PointPop()

      If Serial.Rows.Count = 0 Then

            MessageBox.Show("No parts match the search criteria!",
            "Error", MessageBoxButtons.OK, MessageBoxIcon.Exclamation)

            Exit Sub

            Else

            Call Store()

            End If

Next View

End Sub

Public Sub CreateGrid()

‘Create grid on bounding box surface’
Dim rows As Integer
Dim columns As Integer
Dim BB As DataTable
Dim nRow As DataRow
Dim BBQuery As String

        Grid = New DataTable("Grid Midpoints")
        Grid.Columns.Add("X", Type.GetType("System.Double"))
        Grid.Columns.Add("Y", Type.GetType("System.Double"))
        Grid.Columns.Add("Z", Type.GetType("System.Double"))
        Grid.Columns.Add("Output", Type.GetType("System.Double"))
        Grid.Columns.Add("R", Type.GetType("System.Double"))
        Grid.Columns.Add("G", Type.GetType("System.Double"))
        Grid.Columns.Add("B", Type.GetType("System.Double"))
        Grid.Columns.Add("ViewNo", Type.GetType("System.Double"))

        BBQuery = "Select MaxX,MaxY,MaxZ,MinX,MinY,MinZ From BoundingBox
        Where " & "CastingNo='" & CastingNo & "' And VersionNo='" &
        VersionNo &"'"

        BB = aDB.Query(BBQuery)

        If BB.Rows.Count = 0 Then
```

```
            Exit Sub

        End If

'The bounding box is partitioned into a grid based on what view you want.

        If View = 1 Or View = 2 Then

        Dim Width As Double = Abs(BB.Rows(0)("MaxX") - BB.Rows(0)("MinX"))
        Dim Height As Double = Abs(BB.Rows(0)("MaxZ")- BB.Rows(0)("MinZ"))

        rows = Height / GridSize
        columns = Width / GridSize

            For j = 0 To rows - 1

                For i = 0 To columns - 1

                    x1 = BB.Rows(0)("MinX") + (i* (Width / columns))
                    x2 = BB.Rows(0)("MinX") + ((i + 1)* (Width / columns))

                    z1 = BB.Rows(0)("MinZ") + (j * (Height / rows))
                    z2 = BB.Rows(0)("MinZ") + ((j + 1) * (Height / rows))


                    Dim midpointX As Double = ((x2 - x1) / 2) + x1
                    Dim midpointZ As Double = ((z2 - z1) / 2) + z1


                    nRow = Grid.NewRow
                    nRow("X") = midpointX

                    If View = 1 Then

                        nRow("Y") = BB.Rows(0)("MinY")

                    Else

                        nRow("Y") = BB.Rows(0)("MaxY")

                    End If

                    nRow("Z") = midpointZ
                    nRow("ViewNo") = View
                    Grid.Rows.Add(nRow)

                Next i

            Next j

        End If

    End Sub
```

```
‘Similar code is used for views 3-6

Private Sub PointPop()

SerialQuery = "Select Distinct SerialNo From Inspection Where CastingNo='"
& CastingNo & "' And VersionNo='" & VersionNo & "'And InspectionNo=" &
InspectionNo & ""

Serial = aDB.Query(SerialQuery)

   If Serial.Rows.Count = 0 Then

      Exit Sub

   End If

'For loop which cycles through each serial number in the data table

      For s = 0 To Serial.Rows.Count - 1

      QueryMax = "Select * From Anomaly Where CastingNo='" & CastingNo &
      "' And VersionNo='" & VersionNo & "' And " & "InspectionNo=" &
      InspectionNo & " And SerialNo='" & Serial.Rows(s)("SerialNo") & "'
      And ViewNo=" & View

'Stores max anomaly number for specific serial number into datatable

            MaxAnomaly = aDB.Query(QueryMax)

If MaxAnomaly.Rows.Count > 0 Then

      For k = 0 To MaxAnomaly.Rows.Count - 1


            PointQuery = "Select Point.X,Point.Y,Point.Z " & " From Point
            " & " Where CastingNo='" & CastingNo & "' And VersionNo='" &
            VersionNo & "' And AnomalyNo=" MaxAnomaly.Rows(k)("AnomalyNo")
            & " And " & " InspectionNo=" & InspectionNo & " And
            SerialNo='" & Serial.Rows(s)("SerialNo") & "'"



            PolyPts = aDB.Query(PointQuery)

                    If View = 1 Or View = 2 Then

                        For j = 0 To Grid.Rows.Count - 1

                            value = False

                            For i = 0 To PolyPts.Rows.Count - 1

                                x1 = PolyPts.Rows(0 + i)("X")
                                z1 = PolyPts.Rows(0 + i)("Z")
                                If i < PolyPts.Rows.Count - 1 Then

                                    x2 = PolyPts.Rows(i + 1)("X")
                                    z2 = PolyPts.Rows(i + 1)("Z")

                                Else

                                    x2 = PolyPts.Rows(0)("X")
                                    z2 = PolyPts.Rows(0)("Z")

                                End If

                                If (z1 < Grid.Rows(0 + j)("Z") And
                                Grid.Rows(0 + j)("Z") <= z2) Or
                                (Grid.Rows(0 + j)("Z") <= z1 And z2 <
                                Grid.Rows(0 + j)("Z")) Then
     (O’Rourke,1998)
                                    If ((Grid.Rows(0 + j)("X") - x1) -
                                    (Grid.Rows(0 + j)("Z") - z1) * ((x2 -
                                    x1) / (z2 - z1)) < 0) Then

                                    value = Not value


                                    End If

                                End If

                            Next i

                            If value = True Then

                                cell = Grid.Rows(j)("Output")

                                cell2 = cell + 1

                                Grid.Rows(j)("Output") = cell2

                            End If

                        Next j

                    End If

            Next k

      End If

 Next s

End Sub

‘Similar code is used for views 3-6
```

```
Private Sub Store()

Dim Insert As String

        For i = 0 To Grid.Rows.Count - 1

            If Grid.Rows(i)("Output") > 0 Then

            Insert = "Insert INTO Grid(X,Y,Z,Freq,ViewNo)Values(" &
            Grid.Rows(i)("X") & ", " & Grid.Rows(i)("Y") & "," &
            Grid.Rows(i)("Z") & "," & Grid.Rows(i)("Output") & "," &
            Grid.Rows(i)("ViewNo") & ")"

                aDB.Insert(Insert)

            End If

        Next i

End Sub


Public Sub Color()

Dim MaxOutput As DataTable

MaxOutput = aDB.Query("Select Max(Freq)As cMax, Min(Freq) As cMin From
Grid")

      aDB.Update("Update Grid Set R = 0,G=0,B=1 Where Freq >= " &
      MaxOutput.Rows(0)("cMin") & " And  Freq< " & 0.2 *
      MaxOutput.Rows(0)("cMax"))

      aDB.Update("Update Grid Set R = 0,G=1,B=0 Where Freq >= " & 0.2 *
      MaxOutput.Rows(0)("cMax") & " And Freq< " & 0.4 *
      MaxOutput.Rows(0)("cMax"))

      aDB.Update("Update Grid Set R = 1,G=1,B=0 Where  Freq >= " & 0.4 *
      MaxOutput.Rows(0)("cMax") & " And Freq< " & 0.6 *
      MaxOutput.Rows(0)("cMax"))

      aDB.Update("Update Grid Set R = 1,G=0.554,B=0 Where  Freq >= " & 0.6
      * MaxOutput.Rows(0)("cMax") & " And Freq< " & 0.8 *
      MaxOutput.Rows(0)("cMax"))

      aDB.Update("Update Grid Set R = 1,G=0,B=0 Where Freq >= " & 0.8 *
      MaxOutput.Rows(0)("cMax") & " And Freq<= " &
      MaxOutput.Rows(0)("cMax"))

End Sub
```

```
Public Sub averagenum()

Dim Count As Integer
Dim temp As DataTable
Dim temp2 As DataTable
Dim total As Integer
Dim total2 As Integer

        temp2 = New DataTable("Anomaly Count")
        temp2.Columns.Add("AnomalyCount", Type.GetType("System.Double"))


            For i = 0 To Serial.Rows.Count – 1

            Dim MaxAnomaly As String = "Select Count(AnomalyNo)As cValue
            From Anomaly Where CastingNo='" & CastingNo & "' And
            VersionNo='" & VersionNo & "' And " & "InspectionNo=" &
            InspectionNo & " AND SerialNo='" & Serial.Rows(i)("SerialNo") &
            "'"

                temp = aDB.Query(MaxAnomaly)

                total = temp.Rows(0)("cValue")
                Dim newRow As DataRow
                newRow = temp2.NewRow
                newRow("AnomalyCount") = total
                temp2.Rows.Add(newRow)

            Next i


Dim Value As Integer

        total = 0
        total2 = 0

        For i = 0 To temp2.Rows.Count – 1

            total2 = temp2.Rows(i)("AnomalyCount")
            Value = total + total2
            total = Value

        Next i

        Count = Serial.Rows.Count

        Dim result As Integer

        result = Value / Count

        TextBox2.Text="The Average Number of Anomalies per Part ="& result

End Sub
```

```
Public Sub averagearea()

Dim Count As Double
Dim Sum As DataTable

Count = Serial.Rows.Count

Sum = aDB.Query("Select Sum(Freq) As cValue From Grid ")

Dim Sum2 As Double = Sum.Rows(0)("cValue")

Dim AverageLengthin As Integer = (Sum2 * 0.00155 * (GridSize ^ 2))/(Count)

TextBox1.Text = "The Average Anomaly Area per Part = " & AverageLengthin &
" in^2 "


End Sub
```

*The correlation modules all involve variations of the code provided here.

References Cited

Bala, D., Batson, R.G., Moynihan, G.P., Ray, P.S. (2004), "A visually-oriented quality improvement data management system," *International Journal of Automotive Technology and Management*, vol. 4, no. 1, pp. 91-108

Boonsuk, W., Von Busch, S., Jackman, J., Peters, P. (2007), "Surface Anomaly Mapping: What has your data done for you lately?" *SFSA: Technical and Operating Conference*, pp. 1-6

Carpenter, J. (1996), "Clean Steel Project-Identification of Variables that Affect Cope Oxide Inclusions in Steel Castings," *SFSA: Technical and Operating Conference*, pp. 1-39

Daricilar, G. (2005), "Measurement error of visual casting surface inspection," *Thesis. Iowa State University*, pp. 13-35

Jolliffe, I.T. (2002), "Principal Component Analysis," New York: Springer

O'Rourke, J. (1998), "Computational Geometry in C," New York: Cambridge UP, pp. 239-245

Prieto, F., Redarce, T., Lepage, R., Boulanger, P. (2002) "An Automated Inspection System," *The International Journal of Advanced Manufacturing Technology*, vol. 19, no. 12, pp. 917-925

Mery, D., Filbert, D. (2002), "Classification of Potential Defects in the Automatic Inspection of Aluminum Castings using Statistical Pattern Recognition," *European Conference of Non-Destructive Testing*, pp. 1-13

Mery, D., Acuna, G. (2004-2006), "Automated Multiple View Inspection of Castings using Un-calibrated Image Sequences," *Fondecyt Cooperatcion International*, pp. 1-2

Schorn, T.J. (2006), "Visual Inspection Errors: Measurement and Impact," *American Foundry Society Transactions*, pp. 1-6

Acknowledgements

I would like to begin by expressing my gratitude for my major professor, Dr. Frank Peters. His encouragement and guidance throughout this research project was greatly appreciated. I would also like to thank my other graduate committee members, Dr. Matt Frank and Dr. Scott Chumbley for their time and effort. Special thanks to Dr. John Jackman for his support on database design and programming as well as Dr. Maxwell Morris for the direction he provided regarding data analysis applications.

At this time, I would also like to thank fellow graduate students who supported this project. This includes Wutthigrai Boonsuk for his advice during coding and Greg Saveraid and Mike Rickers for their encouragement and feedback.

Special thanks to the Steel Founders’ Society of America for their involvement with this project and to Harrison Steel Casting Company for allowing me to implement the software program at their foundry. This project benefitted greatly from their participation.

Lastly, I would like to thank my fiancé Lia and my family for all of their support and encouragement throughout my academic career at Iowa State University.

Research was sponsored by the U.S. Army Benet Laboratories and was accomplished under Cooperative Agreement Number W15QKN-07-2-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of U.S. Army Benet Laboratories or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation heron