METHODOLOGY FOR ASSESSING MEASUREMENT ERROR FOR CASTING SURFACE INSPECTION

G. Daricilar and F. Peters
Iowa State University, Ames, IA

Abstract

Visual assessment of objects is critical to many fields including metalcasting. While the output for such a task is often a simple attribute, the problem studied here is for inspection tasks requiring an output that defines shape, size and location of anomalous areas, which are random and are not defined a priori. This paper defines a methodology to quantify the amount of repeatability and reproducibility variation. The application of the methodology for the visual surface assessment of steel castings reveals significant repeatability and reproducibility error. The work presented here is the impetus for current efforts that are defining the capabilities of the visual inspection process and ways to improve it through the selection, training and retraining of operators and through better control of the process.

Keywords: *metalcasting, subjective evaluations, surface inspection*

Introduction

This paper develops a methodology for determining the measurement error for the visual assessment process used to determine the presence and size of surface anomalies. Specifically, the method was developed for the inspection of steel casting surfaces, but could be extended to many other visual assessment tasks. During the processing of the castings, the surface is typically inspected several times to determine if it meets customer requirements. The inspections identify unacceptable casting surface defects such as inclusions, porosity, burnt-on sand, and flash. The parts that are marked for surface defects are then taken through a series of cleaning operations, where the marked defects are mitigated by welding and/or grinding.

The most common method for communicating the surface requirements and determining acceptable surfaces is to use surface comparator plates or photographs. These methods are only qualitative and based on human judgment as there is not a method of specifically classifying casting surface conditions. For instance, employing a surface prolifometer as is done for machined surfaces cannot capture the longer range of the surface indications found on a casting. One commonly used standard, ASTM A802, relies on the comparison of the subject surface with comparator plates from the Steel Castings Research and Trade Association.[1,2] Photographs of some sample plates are shown in Figure 1. Another standard, MSS SP-55, relies on photographs of casting surfaces, such as shown in Figure 2, for the inspector to make comparisons with the subject surface.[3] The result is that the description of the casting surfaces is uncertain at best and impossible at worst. The lack of consistent surface evaluations potentially results in unnecessary costs. Undetected surface defects result in unacceptable quality standards and returns from the customer; marking minor surface imperfections as defects will result in excessive processing.

NOTE: The comparators for definition of casting surface quality, referred to in this paper can be purchased from Castings Technology International (England).

As with all assessment procedures, measurement error will inherently exist. Measurement is more complicated when concerned with freeform areas; likewise, the assessment of measurement error in these cases is also more difficult. The objective of this study is to develop a methodology to quantify the amount of variation in terms of repeatability (variation within the same operator) and reproducibility (variation between different operators) for visual assessment tasks.

There are many literature sources on measurement error for a variety of measurement systems. Much of this work is for measurements in which the output is a continuous variable such as a length, viscosity, or temperature, which does not apply to the problem at hand with visual inspections. The Automotive Industry Action Group (AIAG) published a reference manual for analysis of such measurement systems.[4]

A significant group of literature sources include those regarding attribute measurements in which the output is binary (e.g. pass/fail) or from a discrete list of options. Lyu et al.[5] developed a method for evaluating gage R&R for attribute data collected from inspection tasks such as inclusions in micrographs of steel or individual pixels of a monitor. Specifically, their method takes into special consideration the distribution of the defects within the inspection field. Burke studied errors associated with industrial inspection tasks, but only considered variable measurements and binary decision inspection tasks.[6] Das[7] conducted a case study on the measurement error associated with a subjective length decision.

The medical field also has needs for subjective evaluation of images; these are not immune to measurement error.[6-7] Lee et al.[8] investigated the subjective evaluation of X-ray CT head images by radiologists. This work showed that the agreement among six readers was statistically significant, but there were large variations among readers as well as within a reader. This method did not include the variation of size of the region of interest. Arendts et al.[9] investigated misinterpretation rate of CT scans of hospital emergency department staff. None of these, however, provides a method for determining measurement error when the inspection output is the size, shape, and location of a region.

The most relevant work in the metalcasting field is that by Schorn [10-11] who reported the difficulty in achieving consistency in the inspection of aluminum automotive wheel castings. His approach also did not take into account the size or the shape of the defective region.

While there are many sources available on measurement error, the review of literature reveals very little regarding inspection tasks pertinent to this specific situation (when there could be several types of anomalies, and the measurement output is their shape, size and location.) Because of this need, the methodology presented in the next sections was developed.

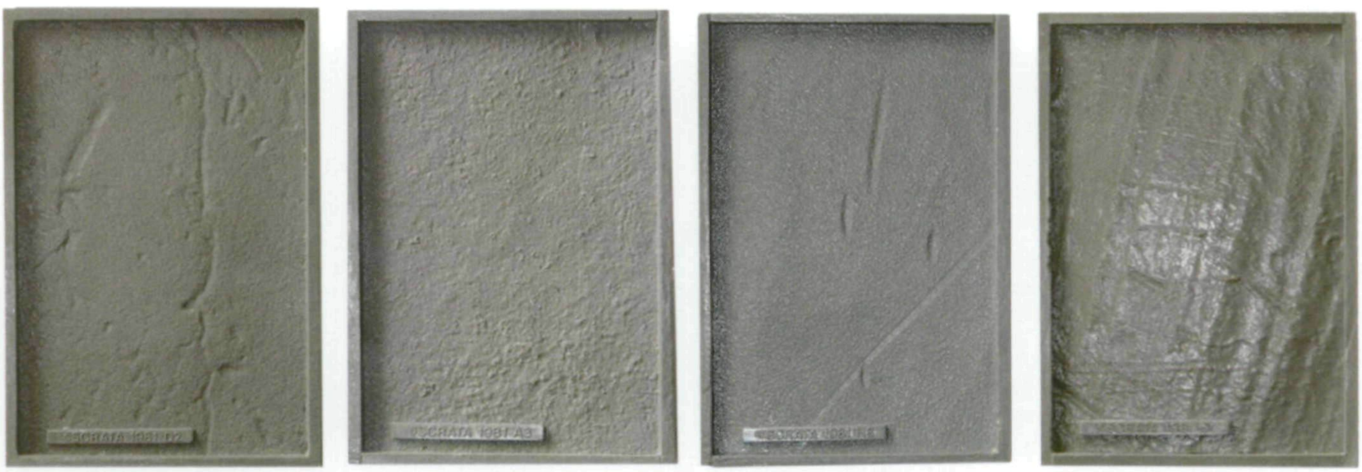

Figure 1. Photographs of some representative comparator plates from the SCRATA Comparator Plate set. [2]

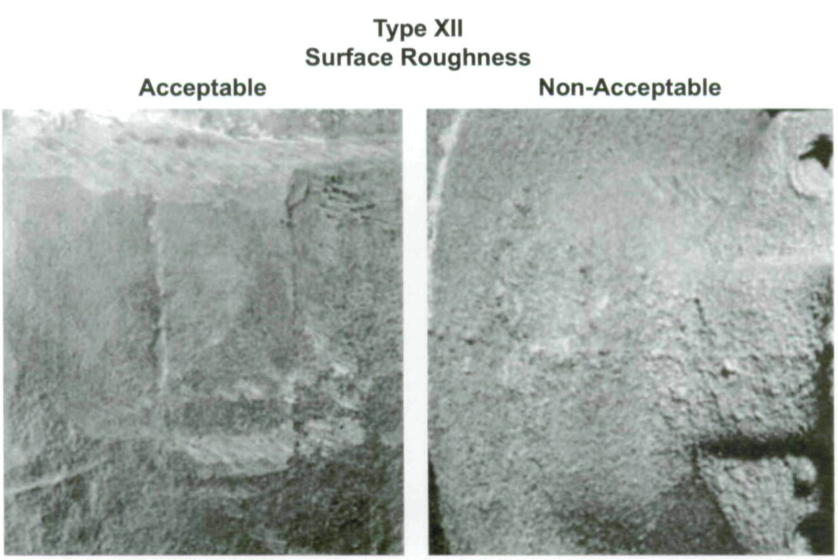

Figure 2. Representative photographs that are included in the MSS SP-55 standard. [3]

Defining Measurement Error for Visual Inspection Applications

The methodology presented here is to determine the measurement error when the objective of an inspection task is to identify freeform areas. This methodology was included in the work by the author.[12] The generic description of the methodology is presented first, followed by an application to the metalcasting surface inspection task. The first step is for regular inspectors to identify the areas that are anomalous or otherwise of interest. Multiple, equally-sized circles, which are smaller than the area of interest, are used to approximate the size and location of the identified area. A series of tightly packed circles are placed until the area of interest is covered. When completed, a boundary line drawn around the circles would encapsulate the defective area identified by the inspector. In this paper, these circles are defined as *elements*. Figure 3 shows an example of how the method is applied to a part marked by an operator; the defective area is quantified by circles, or elements. For each inspection trial, an *element map* that contains the location of each element is constructed; each element map has the same coordinate system with respect to the part. Figures 4 and 5 show examples of castings at two different foundries which were inspected twice by two inspectors each, and then the inspection marks were identified with circular labels.

Defining Master Clusters

To assess repeatability and reproducibility error, each part must be assessed at least twice by at least two operators. Each of these separate inspections is defined here as a *trial*. This leads to an interesting situation, since the defect regions will likely be defined differently for each trial. For instance, a single defect region on one trial may be defined as two separate defect regions for a subsequent trial on the same part. This leads to the need to define a *master cluster*. A master cluster is a composite of all of the elements that would have contact with each other if the element maps from all trials for that part were superimposed. To determine the elements that comprise each master cluster, a search is conducted of all element maps for elements that would be adjacent or coincident.

The search for adjacent elements is done across all the element maps (one for each trial) for the part. The master cluster search zone is circular and centered on a particular element's center. During the master cluster determination process, if the center position of another element from any of the element maps of the same part falls within this zone, then the two elements are assigned to the same master cluster region. Since the goal is to find overlapping and adjacent

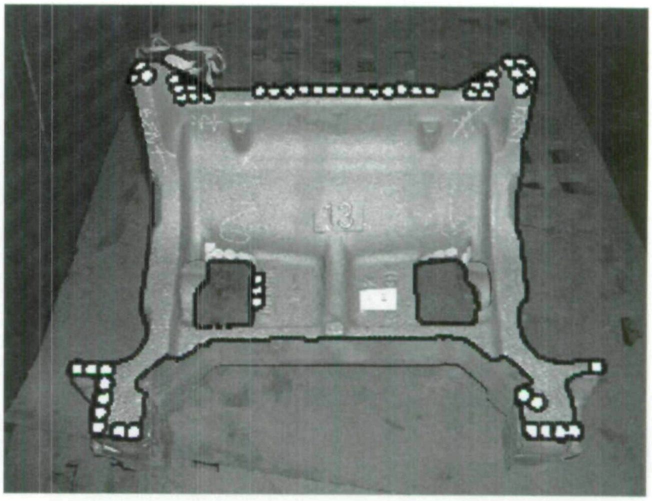

Figure 3. A photograph of a casting with white stickers applied to identify the locations of the inspector's marks. These stickers then become the elements used in the methodology defined here.

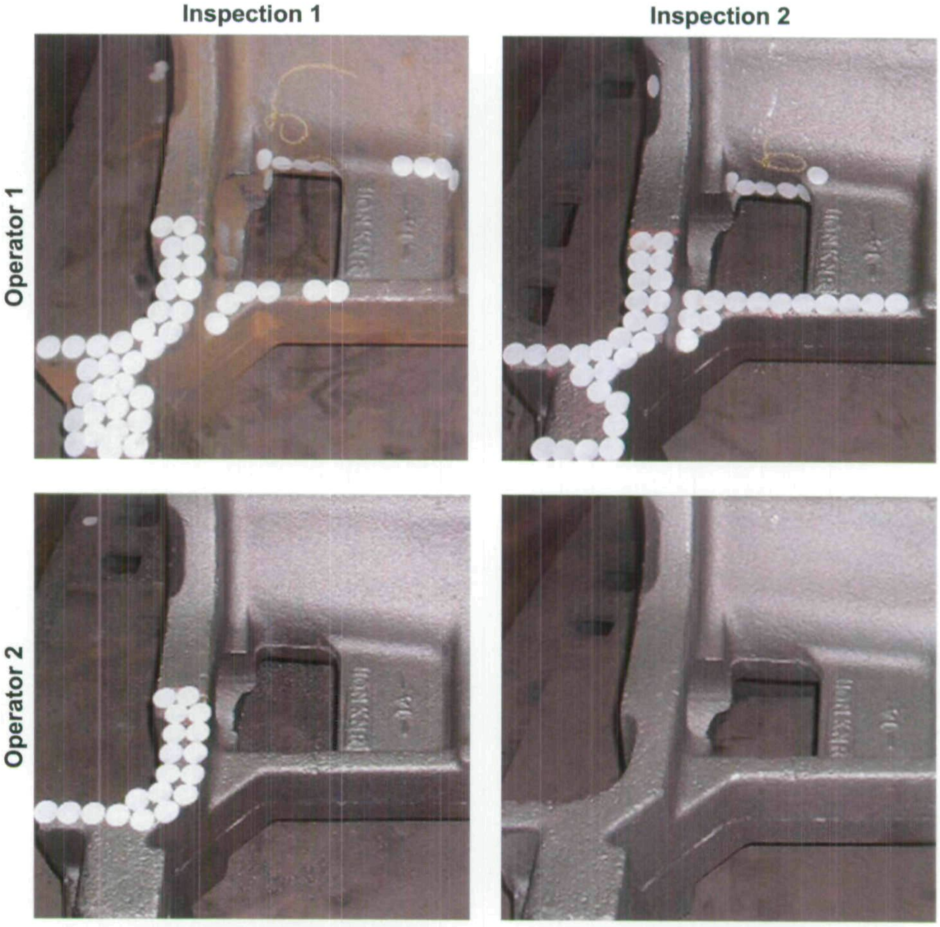

Figure 4. Results obtained from the same casting inspected twice by each of two operators at Foundry 2.

elements, the radius of the search zone is typically 2.5 times the radius of the element. This is illustrated in Figure 6. Theoretically, a radius of 2.0 will find adjacent elements, but the larger value is used to account for minor errors in element placement. Discussion on the sensitivity of the search zone size follows in the case study. This search method is repeated for each element.

A visual representation of the master cluster concept is displayed in Figure 7. In this case, operator 1's trial 1 contains seven elements for three defect regions, and trial 2 contains four elements for three defect regions for the same part. For this same part, operator 2's trial 1 contains five elements for three defect regions, and trial 2 contains four elements for three defect regions. When the four element maps collected for this part are superimposed, the result is four master cluster regions. Note that one of the master clusters is only comprised as one element since this region was only identified by operator 2 on trial 1, and no other elements were within the master cluster search zone of this element.

Defining Repeatability

Repeatability error is the amount of variation among inspections by the same operator on the same part. To assess repeatability, each operator therefore needs to inspect each part at least twice. In determining repeatability, the objective is to find those elements that are in the same location as elements from other trials. This element matching process is similar to that done to determine the master clusters, but differs because we are looking for elements that are in identical locations, not adjacent locations. During the element matching process, if the center position of another element from another trial of the same part and operator falls within the repeatability search zone, then the two elements are considered to be matched (see Figure 8). The nominal repeatability search zone radius is 1.5 times the radius of the element to accommodate minor error in element placement.

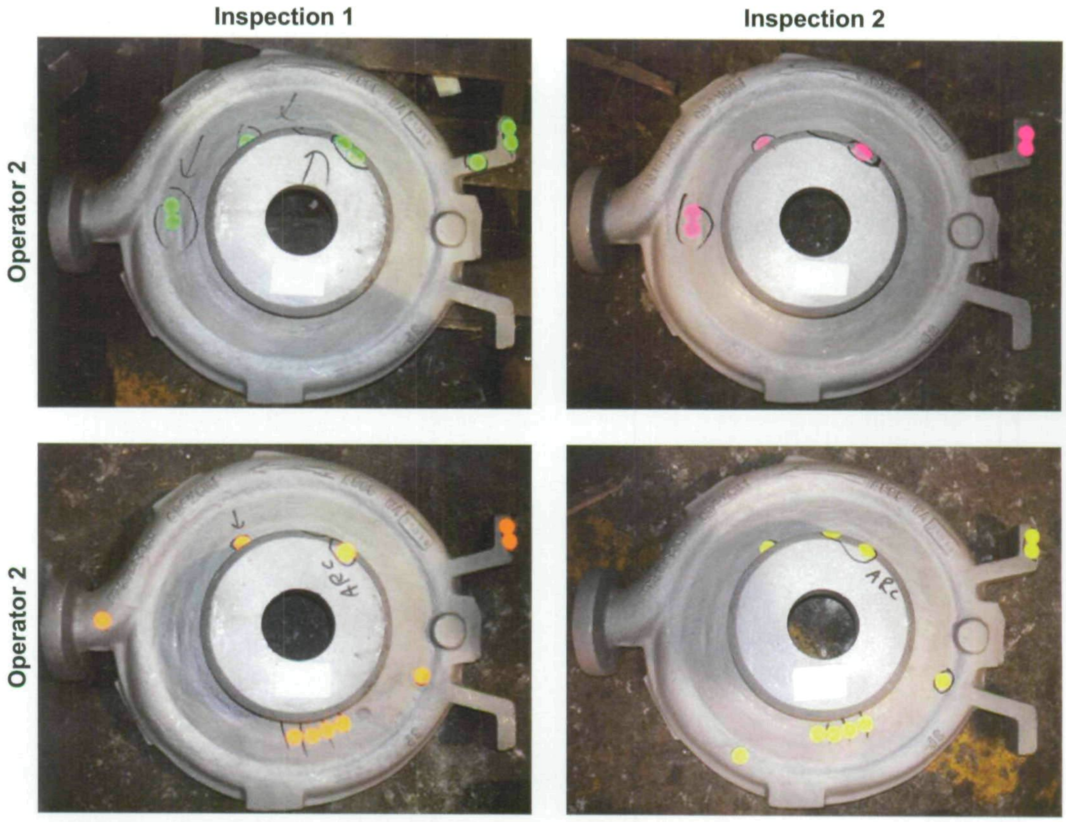

Figure 5. Results obtained from the same casting inspected twice by each of two operators at Foundry 3.

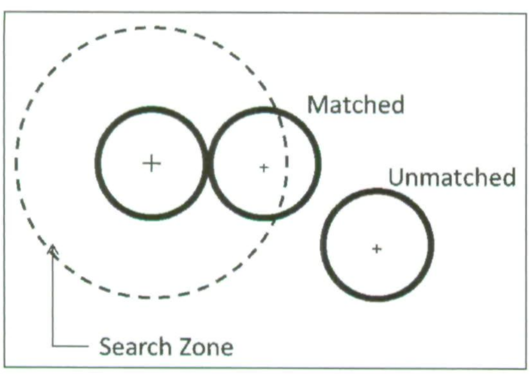

Figure 6. Search zone coefficient example with default values of 2.5 for master cluster.

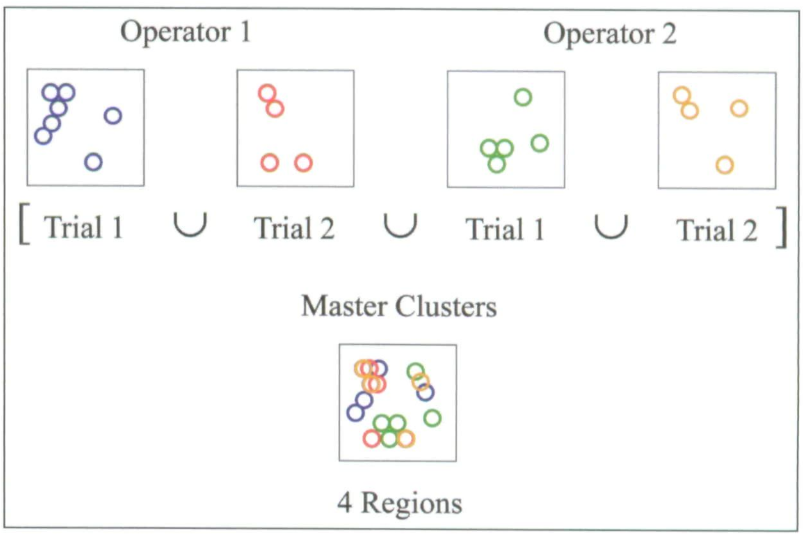

Figure 7. Defining master clusters with the four combined inspection trials of the same part. The two elements at the bottom of the part from trial 2 of operator 1 both fall in the same master cluster region, since they are connected by elements from other trials.

In this study, the repeatability of an operator is reported as two different values. The first percentage value indicates how well the operator performs in identifying the same defect regions for the inspection trials of the same part, reported as master cluster match percentage. For this measure, a match between two trials (from the same operator and part) is established if any element of the master cluster from the trial is matched with any element of the other trial. The total potential cluster matches, MRT_{MC} for each combination of operator and part is found as follows,

$$MRT_{MC} = C_j*(T-1)! \quad \text{Equation 1}$$

where C_j is the number of master clusters that operator j finds on all trials and T is the number of trials for the combination of operator and part.

The second repeatability value indicates how consistent the operator is in defining the specific shape and size of the defect regions of the same part across multiple trials, reported as the element match percentage. For each trial, the elements are checked for a match to an element on each of the other trials. The potential number of matches, $MRT_{El,}$ for operator j inspecting a particular part is found by Equation 2.

$$MRT_{El} = E_j*(T-1) \quad \text{Equation 2}$$

E_j is the number of elements across all trials for operator j on this part.

A visual representation for the repeatability is presented in Figure 9. In this example, operator 1's trial 1 contains seven elements for three defect regions, and trial 2 contains four elements for three defect regions for the same part. Operator 1's two inspection trials for the same part match two of three master clusters and seven of eleven elements. For this hypothetical example, operator 1's repeatability is 67% for master cluster matching and 64% for element matching.

Defining Reproducibility

Reproducibility error is the variation in inspections between different operators on the same part. Similar to repeatability, reproducibility is also defined by two aspects in this paper: how well the operators perform in identifying the same defect regions and the specific size and shape of the defect regions between the inspection trials of the same part. For both measures, higher values indicate better reproducibility or agreement.

The first step in calculating reproducibility is to create a *composite element map* for all of the trials conducted by each operator for each part. This is shown in Figure 10 as the union of all trials (two in this example) for each operator. For each part, element matching is conducted between each operator's composite element maps. This matching is done in the same manner as used for repeatability, and once again the goal is to determine those elements that are in (nearly)

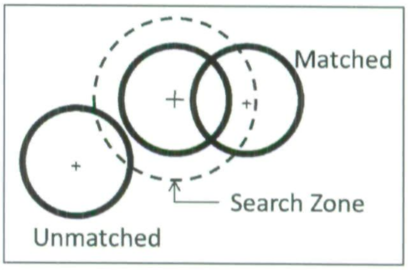

Figure 8. Search zone coefficient example of 1.5 for repeatability and reproducibility operations.

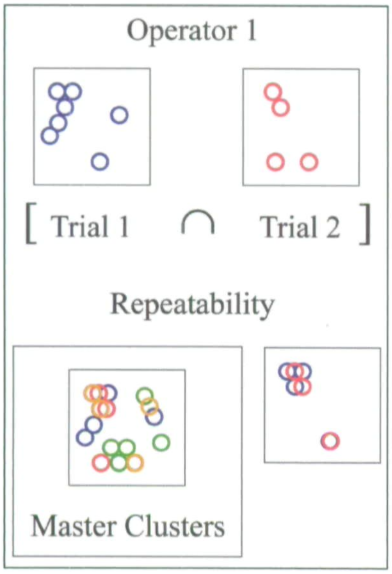

Figure 9. Defining repeatability of operator 1 for the two trials on one part. For this example, 2 of 3 master clusters match and 7 of 11 elements.

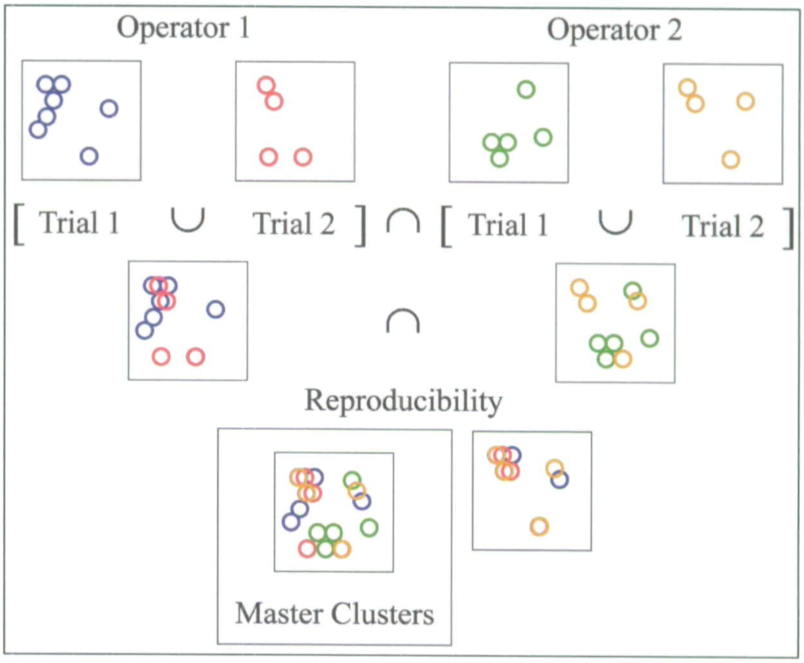

Figure 10. Reproducibility of two operators with the same part. For this example, 3 of 4 master clusters match and 12 of 20 elements.

identical locations as any element from the other operator's composite element map. The reproducibility search zone would theoretically have a radius equal to that of the elements, but 1.5 times this value is again used to account for any minor discrepancies in element placement.

The reproducibility is reported as two different percentage values. The first value indicates how well operators perform in identifying the same defect regions for the same part, reported as master cluster match percentage. For a master cluster to be considered a match, an element from an operator's composite element map must match at least one element from another operator's composite element map. For each operator and part, the total potential matches, MRD_{MC}, is found by Equation 3.

$$MRD_{MC} = C*(P-1)! \qquad \text{Equation 3}$$

Where C is the number of master clusters defined for a part and P is the number of operators. It is important to note that the composite element maps for each part are the union of all of the element maps for each operator. Therefore, an element from one trial of the first operator only requires the same location as one element from any trial of any other operator for a master cluster to be considered a match.

Since the size and shape of the areas identified by operators could be grossly different in size, it is also important to have an indication of how the size and shape correspond. The second percentage value indicates how well the operators perform in defining the same size and shape of the defect regions. These comparisons are done between each operator and for each part and reported as element match percentage. Each element for a particular operator is checked to see if it matches an element in the composite element map for each other operator. The total potential matches, MRD_{EI}, for each operator and part combination are found by the following.

$$MRD_{EI} = Ej*(P-1) \qquad \text{Equation 4}$$

Methodology Implementation

The experimental data were collected at three steel casting producers, which collectively represent a broad spectrum of the North American steel casting industry. This experiment had three stages. The first was the inspection of the castings and the collection of images of the castings. The next stage was the extraction of the coordinates of the elements from the images collected. The final stage was the analysis of the coordinate data. At each company, a particular casting design was chosen for the study, and six to ten parts of the chosen design were then picked to be used in this experiment. Two operators at each of the three locations participated in the experiment. The overall size of the casting at foundry 1 was 350 mm (14 in.) and those at the other two foundries were 600 mm (24 in.). Some of the results of the case study without the methodology were previously reported by the authors.[13-14]

Data Collection and Analysis

The parts were marked for defects by the operators employing the same method that they typically use to inspect castings, except that only one side of the castings was inspected for this study. This study wanted to measure the actual variability that existed in the inspection operations, while minimizing the influence of the study on the current process. At the three companies, each of two operators inspected each part twice. To reduce bias, the inspections were done on different days and the order of the parts was randomly selected by the experiment moderator. After the parts were marked by the inspectors with industrial crayons per their normal protocol, round stickers of 19 mm (¾ in.) diameter were used to cover the markings, as previously shown in Figure 3. These stickers became the elements that were used in the measurement error analysis. Careful attention was paid to carry out the sticker placement process as consistently as possible. Digital pictures of the sticker-covered parts were then taken. Between trials, the parts were shot blasted to remove all markings. There were initial concerns that the shot blasting between trials could affect the results, but this was ruled out through analysis.

The digital pictures were cleaned using photo software to eliminate noise and to delineate the element (sticker) boundaries. This process did not change the location of the elements. A cleaned image with elements identified is shown in Figure 11. These images were then input into DVT FrameWork software to determine the center coordinates of each element (sticker) which were recorded in a spreadsheet. (Other software would have worked equally as well)

The algorithms presented in the previous sections were implemented via macros in the spreadsheet. The master cluster operation was used to take the element coordinates and group the elements into master clusters. The nominal search zone coefficient for the master cluster search was set at 2.5 times the element radius. However, for Companies 2 and 3, a search zone of 3.0 times the element radius was used because of lower image quality and a high percentage of elements that were at an angle to the image view.

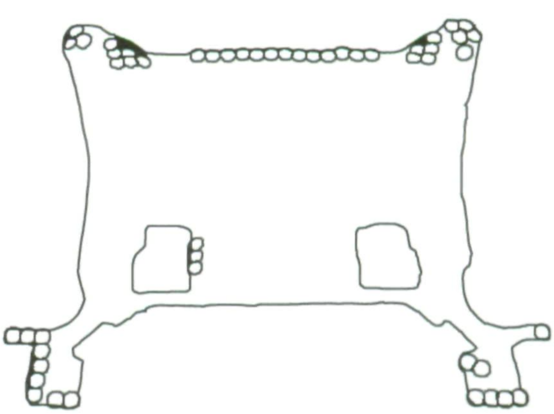

Figure 11. Cleaned image created from the steel casting shown in Figure 3.

The percentage matches for repeatability and reproducibility was also carried out via macros programmed in the spreadsheet. The radius of the search zone was increased from 1.5 to 1.8 times the element radius because of the image quality.

The results showed that there is a significant amount of repeatability and reproducibility error in the visual casting surface inspection process. The repeatability and reproducibility values are shown in Figure 12 and 13, respectively. On average, the repeatability was better than the reproducibility measurements, and the master cluster match percentages were lower than the element match percentages for the same operator.

Sensitivity Analysis for Search Zone Coefficients

A sensitivity analysis was conducted to examine the effects of varying the size of the search zone coefficients on the total number of master clusters as well as repeatability and reproducibility percentages using the data from Company 1. For the search zones used for establishing the master clusters, the number of master clusters approached a minimum as the search zone coefficient was increased (see Figure 14.) As the search zone approached zero, the number of master clusters approached the number of elements, as expected.

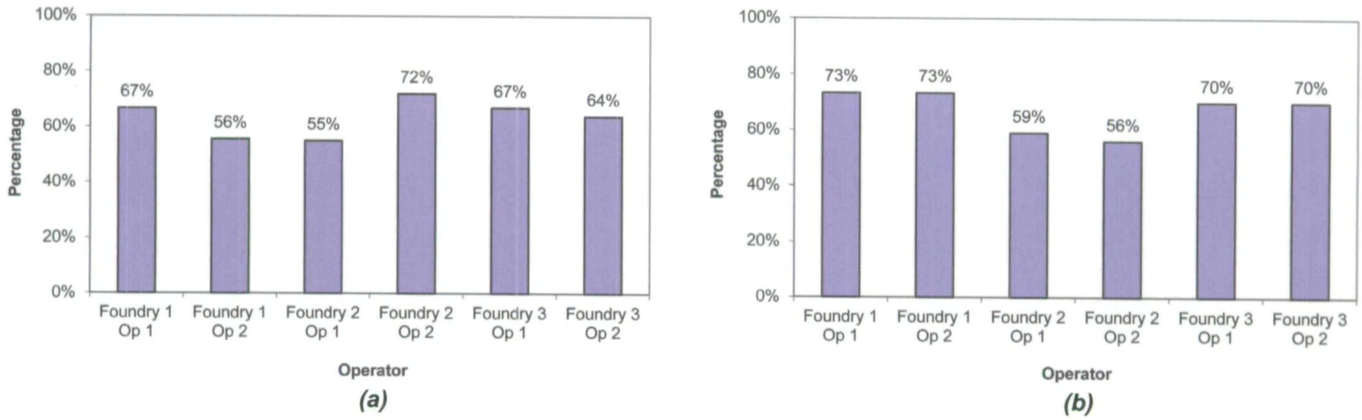

Figure 12. Repeatability results for percent master cluster match (a) and percent element match (b).

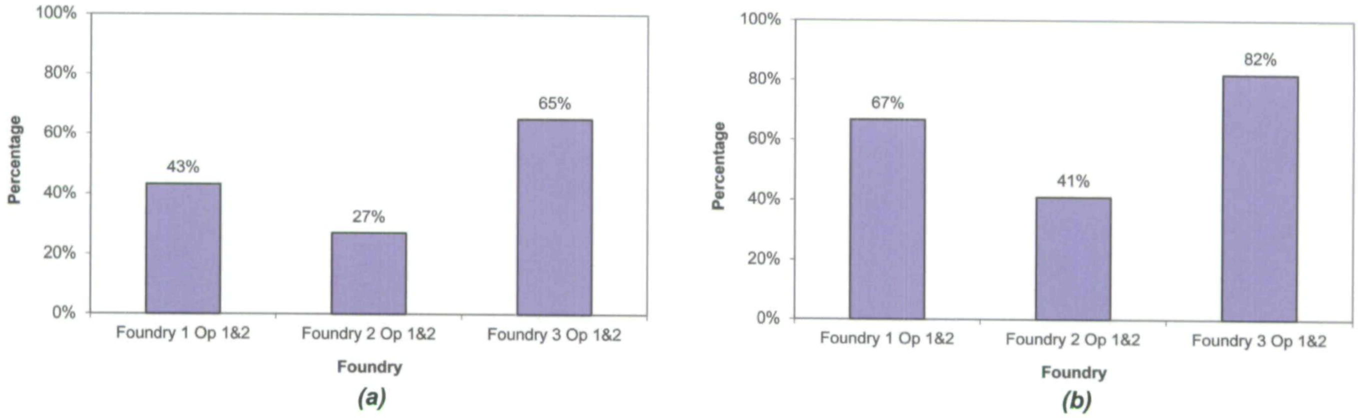

Figure 13. Reproducibility results for percent master cluster match (a) and percent element match (b).

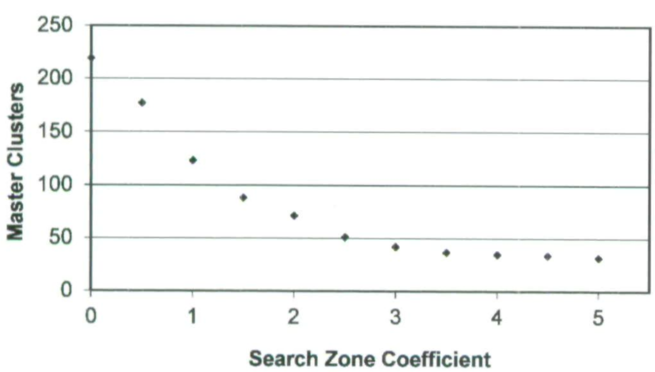

Figure 14. Effects of search zone coefficients on the total number of master clusters.

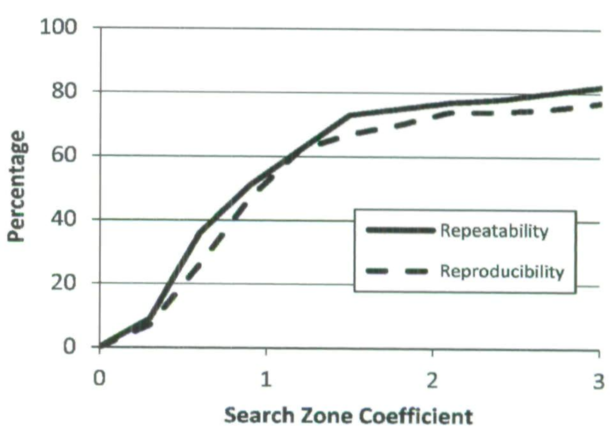

Figure 15. Effects of search zone coefficients on the repeatability of one operator and reproducibility.

For the search zones used to determine repeatability and reproducibility, the percentage of elements matched reached a maximum as the search zone coefficient increased (see Figure 15.) Based on these observations, the search zone sizes were considered appropriate.

Conclusions and Future Work

This study introduces an approach to assess measurement error for visual inspection processes. This method determines the amount of measurement error when the location of the defect regions are not defined a priori. This method also handles instances where two or more defects from one trial could overlap only one larger defect of another trial. The applications of this methodology, however, are not limited to the steel casting industry. If utilized, it could prove useful in other fields where subjective evaluations are used, such as medical image analyses, nondestructive evaluation techniques, and meat carcass inspections.

The results of the application showed that there was significant variation in both repeatability and reproducibility measurements of casting surface inspections. Although the repeatability measurements were somewhat consistent within the companies, the reproducibility measurements exhibited considerably more variation.

Propelled by the need shown here, current efforts are striving to improve the visual inspection process. Better control over the process in terms of the environmental conditions as well as limiting the expected inspection speed based on the quantity of surface information that a human operator is able to accomplish is being investigated. In addition, better training tools for the initial as well as ongoing training periods is being researched.

Acknowledgements

We would like to thank the Department of Energy (DOE) Office of Industrial Technology, which provided funding via grant No. DE-FC07-02ID14228. However, any opinions, findings, conclusions, or recommendations expressed herein are those of the author and do not necessarily reflect the views of the DOE.

REFERENCES

1. ASTM, *A 802 Standard Practice for Steel Castings, Surface Acceptance Standards, Visual Examination*, ASTM International, West Conshohocken, PA (2006).
2. SCRATA, Comparator Plates, The Castings Development Centre, Sheffield, UK.
3. MSS, *SP-55 Quality Standards for Steel Castings for Valves, Flanges and Fittings and other Piping Components*, Manufacturers Standardization Society, New York, NY (1961).
4. AIAG, *Measurement System Analysis,* Automotive Industry Action Group, Detroit, MI (2002).
5. Lyu, J. and M.-N. Chen, "Gauge Capability Studies for Attribute Data," *Quality and Reliability Engineering International*, vol. 24, pp. 71-82 (2008).
6. Burke, R.J., Davis, R.D., Kaminsky, F.C., and Roberts, A.E.P., "The Effect of Inspector Errors on the True Fraction Nonconforming: An Industrial Experiment," *Quality Engineering*, vol. 7, pp. 543-543 (1995).
7. Das, N., "A Study on Estimating Measurement Error in the Measurement of Earing of Pilfer-Proof Caps," *Quality Engineering*, vol.16, pp. 265 - 268 (2003).
8. Lee, H., Rowberg, A.H., Frank, M.S., Choi, H.S. and Kim, Y., "Subjective evaluation of compressed image quality," *Medical Imaging VI: Image Capture, Formatting, and Display*, Newport Beach, CA (1992).
9. Arendts, G., A. Manovel, and A. Chai, "Cranial CT Interpretation by Senior Emergency Department Staff," *Australasian Radiology*, vol. 47, pp. 368-74 (2003).
10. Schorn, T., "Visual Inspection Errors: Measurement and Impact," *AFS Transactions,* vol. 114, pp. 17-23 (2006).
11. Schorn, T., "Factors That Influence Visual Inspection Effectiveness," *AFS Transactions*, vol. 115, pp. 67-75 (2007).
12. Daricilar, G., "Measurement Error of Visual Casting Surface Inspection," M.S. Thesis, Iowa State University (2005).
13. Daricilar, G., F. Peters, and M. Blair, "Visual Assessment of Casting Surface Quality," *Materials Science and Technology 2005 Conference*, Pittsburgh, PA (2005).
14. Daricilar, G. and F. Peters, "Measurement Error of Visual Casting Surface Inspections," *Steel Founders Society of America Technical & Operating Conference*, Chicago, IL (2005).

Methodology for Assessing Measurement Error for Casting Surface Inspection

G. Daricilar, F. Peters, Iowa State University, Ames, IA

Reviewer: This is a subject in continued demand and a problem between casting manufacturers and casting users. Visual inspection at best is about 85% reliable, so this also needs to be discussed.

Authors: *The authors have heard about the notion of an upper limit on the effectiveness of visual inspection. However, the source for this number, nor the scientific evidence to support it has not been found. The authors do not want to propagate this premise without being able to support it. Current efforts may help quantify this number for casting operations, but several conditions will still be required because of the variability of operators and inspection requirements.*

Reviewer: For some metalcasters, the solution is to fire inspectors when a bad part is shipped to a customer. I am not sure this solves the problem.

Authors: *This study showed that significant variability exists in the visual inspection process. Firing workers because a bad casting is received by the customer is not going to solve the problem. A comprehensive improvement of the system needs to be implemented including control of environmental conditions that degrade operator performance, selection of operators based on their skill set not on their seniority or other factors, providing adequate training and tools to make comparisons, and providing ongoing retraining to 're-calibrate' the human measurement system. After this is accomplished, then it may be most appropriate to reassign an operator that allows castings to be shipped to the customer.*

Reviewer: The best foundries do not inspect quality into their casting. They establish the right process controls to assure things are made correctly without defects.

Authors: *For many castings, especially larger and shorter run castings, it is not always feasible (or economically feasible) to have them acceptable as they come out of the mold. We believe that visual inspectors still perform a vital need. This goal of this experiment was to capture the variability of the current measurement methods, using their standard inspection techniques and requirements. The authors are focused on providing the metalcasting industry with tools that can lead to improvements.*