Fully Automated Rating of Slab Segregation Images for Pipeline Steel on a Continuous Scale

Digital technologies are transforming industry at all levels. Steel has the opportunity to lead all heavy industries as an early adopter of specific digital technologies to improve our sustainability and competitiveness. This column is part of AIST's strategy to become the epicenter for steel's digital transformation, by providing a variety of platforms to showcase and disseminate Industry 4.0 knowledge specific for steel manufacturing, from big-picture concepts to specific processes.

Authors

Ahmad van der Breggen
EVRAZ North America Regina Steel, Regina, Sask., Canada
ahmad.vanderbreggen@evrazna.com

Kendal Dunnett
EVRAZ North America Regina Steel, Regina, Sask., Canada

Laurie Collins
EVRAZ North America Regina Steel, Regina, Sask., Canada

For high-strength, low-alloy (HSLA) steel produced for pipelines, the centerline segregation of continuously cast slabs is an important metric of quality. In the current market, measurement of centerline segregation is done by comparison to reference images, by measurement of the fraction of surface area darkened by etching, or by measuring and counting of darkened dots on etched surfaces. To remove sensitivity in these adjustments while balancing precision and understandability, this paper proposes a fully automated image analysis method based on industrially recommended dot measuring methods. A software prototype using this method has been developed; since this method uses images of samples etched for other methods, the method has been tested on thousands of samples.

Existing Visual Methods

Common methods for visually analyzing centerline segregation can be broken down into three categories: reference image comparison, dark area fraction and individual dark area measurement. Each of these methods can be executed manually by operators or automatically by image analysis, but the complexity of the process and quality of the measurements varies with the method of execution. For example, it is possible to do an area fraction analysis by hand, but it is very tedious and has questionable repeatability across operators, while image analysis software can easily count light and dark pixels once the threshold for darkness is determined. The following outlines a few of the possible issues of each method and some that all methods have in common.

Reference image comparison methods, such as the work described by SMS group,[1] require an operator or image analysis software to look at the sample segregation image, compare it to reference images and pick the closest reference image. The sample then gets the rating of the closest reference image. When executed by different operators, these methods can become subjective if the criteria for analysis are not clearly defined beforehand. For example, reference images tend to be homogeneous in their segregation across the width to be clear on the difference between classes, but centerline quality can vary across the width of the transverse cut and across individual samples. In order to choose the most similar reference image, it should be agreed upon whether to look at the quality of the centerline across the whole sample, to focus the worst section of a predetermined width, or to focus on the worst concentration of indications regardless of the sample area affected. Automated image analysis using image comparison methods have been used in the past, but the measurement of success tends to be based on agreement with the same method executed by one or more experienced operators, so the same subjectivity can be built into the implementation of an automated method.

Dark area fraction methods, like the work of Abraham et al.,[2] require the operator or image analysis software to measure the area of all dark indications and divide that by the total area being measured to obtain a dark area fraction. From this measured area fraction, a rating is determined. This type of method has a different set of issues: a large number of small indications can

have the same influence on the final rating as a small number of large indications, and a larger section size can wash out acute problems (assuming the worst section's rating is used as the final rating for the sample). A trivial example to demonstrate these two issues: 20 indications that are 1 x 1 mm spread out across the sample contribute the same 20 mm^2 to area fraction as a single indication that is 2 x 10 mm. On a 200-mm-thick slab, if the slab is cut into 200-mm-wide sections (40,000 mm^2), the 20 mm^2 contribution would have 20/40,000 = 050% impact on area fraction, while cutting the sample into 400-mm-wide sections (80,000 mm^2) would make that same contribution have 025% impact on the section's area percentage. While this seems obvious, more evenly distributed indications, like the ones seen in reference images, do not typically present this problem.

Individual dark area measurement techniques, like the work of Rapp,[3] and that of Steel Institute VDEh,[4] also called "dot counting" methods, resolve most of the problems above, particularly when executed by automated image analysis software. These methods measure the dimensions of individual dark areas and group the indications by their size. Each method has a set of thresholds for the number of indications allowed in each size category for each final rating classification. The remaining issues are the thresholds between size classes and the thresholds to determine final rating class based on the number of indications in each size class. For example, both methods have four final rating classes. Both use a 10-mm threshold for entry into the worst class, but an unlimited number of 9-mm indications are allowed in the second-worst class. The biggest differences between the two mentioned methods are the dimensions of the dark areas used to classify indications, and treatment of close but disconnected indications and thinly joined indications. The method of Steel Institute VDEh will be referred to several times in this paper as SEP 1611.

One of the issues that is common to all of the methods mentioned is the decision to include or exclude an indication based on its darkness relative to the surroundings. Each analysis method handles this decision differently, having different wording or methods to decide that an indication is dark enough to be counted. While an operator executing two different methods might use a similar threshold for darkness, operator-to-operator differences in interpretation can be large, and attempting to follow the rules strictly or implementing the rules in an automated analysis can make a measurable difference between methods on the number and size of indications detected.

The last issue to discuss here is that of classifying samples into a small number of ratings. There is such a wide range of possibilities in each rating class that the difference between two samples in the same class can be far larger than the difference between samples in adjacent classes. This makes the measurement of improvement efforts based on these ratings very difficult.

Proposed Method

In order to address these issues, the new method incorporates the following:

- The influence of each indication is calculated based on the indication's area and its outside diameter, with larger indications having more impact on final rating than the same area made up of smaller indications. Additional consideration is included for the total length of centerline covered by the sum of equivalent circular diameters.
- The method scans across the samples to find the worst 100-mm-wide window, regardless of section width. This only solves the size problem partially because a section cut through a large indication or batch of indications can still reduce the rating. The worst 100-mm window method is also used in the two referenced methods for individual dark area measuring.
- To improve on the issue of a small change in indication size pushing the indication into another size category and possibly into another rating category, the indication measurements are used to calculate a rating without rounding or categorization.
- To improve the ability to compare between slabs of similar rating, the method uses a continuous rating scale.
- To address the possibility of subjectivity between operators, the method is fully automated once the sample image is scanned.

The remainder of this section describes the rating process with an overview of the sample preparation, followed by details about the image adjustments and analysis done to make up the new rating method. Note that for all segregation images with grids overlaid for size reference, the grid spacing is 5 mm.

Sample Preparation — Full-width, full-height samples are cut from the end of a continuously cast slab by an automated torch. The thickness of the sample (in the casting direction) must be sufficient to allow all torch heat effects to be removed in the machining step. The full-width samples are sent to a machine shop, where they are cut into pieces across the width so that they fit in the automated etching machine and on the scanning bed, and to reduce the weight of the pieces for handling by operators. The pieces are then machined on one of the cut faces to remove all torch effects

and make a smooth surface. Once the samples are returned from the machine shop, they are chemically etched to make the centerline segregation (which is somewhat different in chemistry than the surroundings) stand out.

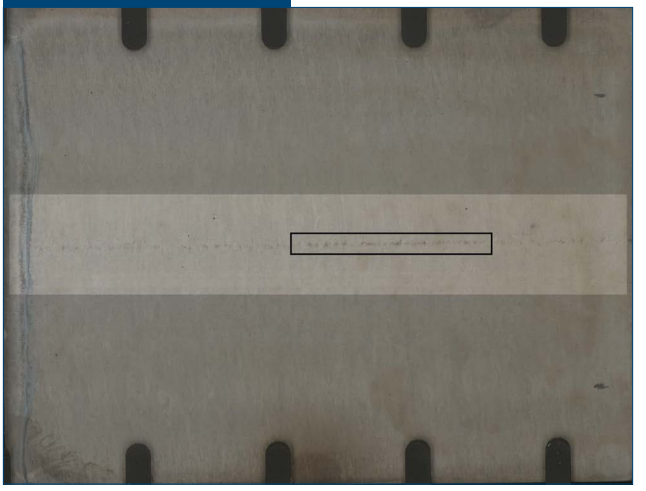
Once the etching is complete, the sample is rinsed and then scanned, either by a flatbed scanner or a wand scanner. Images could also be acquired by a well-placed camera with well-controlled lighting, but it has been found that scanners give more consistent results because they contain their own well-controlled lighting and they reduce the possibility of geometric distortions that can arise from different camera setups (either due to the shape of the optics or placement relative to the sample). A very helpful feature of using a scanner to acquire images is that the dimensions of the sample and the centerline indications can be measured on the image with acceptable accuracy by dividing the pixel dimensions by the pixel density (dots per unit length). To get accurate dimensions from a camera setup, measuring scales should be photographed on the sample.

Image Adjustment and Analysis — Most methods of visual centerline segregation analysis can be broken down into three main steps: detection of dark regions, measuring detected regions and calculating a rating. Each step should be analyzed very carefully before committing to the use of a given method. This section discusses the details of the proposed method as it stands today, and a subsequent section outlines a sensitivity analysis done as a part of this study.

Fig. 1 shows the original image from a 341-mm-wide sample of a 254-mm-thick slab; the dark "fingers" on the top and bottom are the brackets that hold the sample above the scanner's glass. The lightened rectangle across the center of the image is the 50-mm-high region normally considered for analysis. This height was chosen to ensure that the centerline was within the considered region if the sample was shifted up or down on the scanner bed. The width of the considered region is 98% of the sample width, which was chosen because there are often machining, etching or scanning anomalies outside this width on historical images. The smaller black rectangle of centerline, on the right side of the image, is a 100 x 10 mm portion being shown through the detection process for this document.

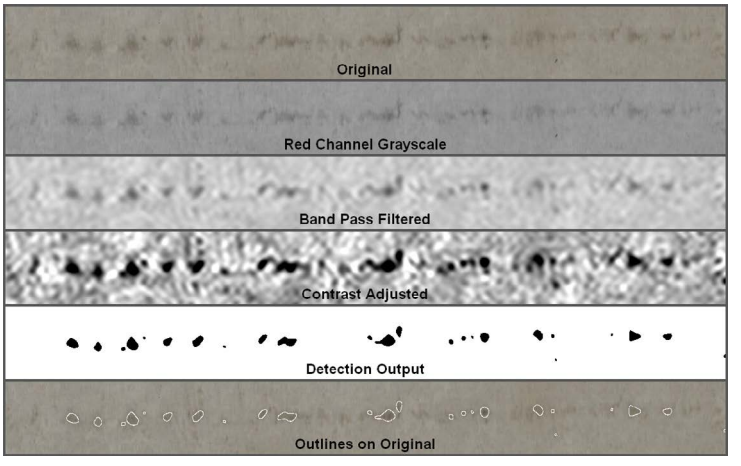
Detection of Dark Regions — The detection of dark regions in the proposed method is accomplished using a combination of filtering, contrast adjustment and thresholding. This process is shown in Fig. 2. First, the red channel of the RGB (red, green and blue) scanned image is used because it is less sensitive to rust staining that can result if the sample is not dried soon enough after etching and rinsing. The red channel, now a grayscale image, is put through a band pass filter to remove noise and very small indications on the high-frequency side and to remove etching variation across the sample on the low-frequency side. The frequency cutoffs of the band pass filter are adjusted based on the image's pixel density (dots per millimeter), which helps to reduce the differences that can arise from different scanning settings. The filtered

Figure 1



Original sample image used to demonstrate proposed method, with the 50-mm-high considered region lightened and a 10 x 100 mm region for further analysis shown with a black rectangle.

Figure 2



Progress from original image to detection output on 100 x 10 mm portion of sample image.

image is then adjusted so that the median brightness is at 75%, and the darkness relative to the median is scaled based on the difference between the 40th and 60th percentile brightness. This makes the areas that are much darker than the background stand out and provides the ability to threshold the image based on a fixed number.

Measuring Detected Regions — The measurement of detected dark regions is done on the detection output. The measurements available include bounding rectangle, bounding circle, bounding ellipse and equivalent circular area. Fig. 3 shows the rectangular and circular measurements of the detection output for comparison.

Calculating a Rating — Once the indications in the considered region are measured, some strange indications are removed; for example, thin vertical indications near the left or right edge of the image are considered etching artifacts. These detected artifacts are drawn on the output image as purple rectangles and saved with the detection details. With the remaining indications that have equivalent diameter of at least 1 mm, Eq. 1 is used to calculate the rating for the indications in a 100-mm window:

$$\left(a_a \sum (\pi r_e r_b)^{c_a} + a_p \left(\sum \frac{d_e}{w} \right)^{c_p} \right)^{c_o} + a_o \qquad \text{(Eq. 1)}$$

where

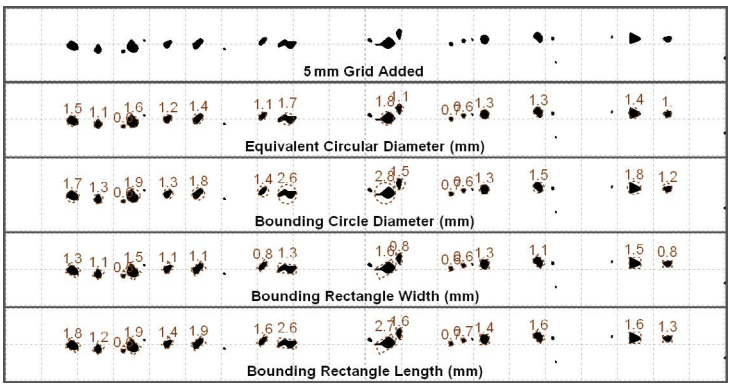
r_e and d_e = the radius and diameter of the equivalent area circle (in mm),
r_b = the radius of the bounding circle (in mm) and
w = the window width (in mm).

This equation can be broken into two main parts: the portion from a_a to c_a is the contribution from individual circles (equivalent to multiplying the area of each indication by its ratio of outside diameter to equivalent diameter), and the portion from a_p to c_p is the contribution from the sum of equivalent diameters divided by the width of the sample (which is the percentage of width covered by equivalent circles laid end to end). The tuning coefficients a_a, c_a, a_p, c_p, a_t and c_t are used to shape the output to meet the goals of the project. Some details on the goals and on how this equation reacts to different inputs are in the next section.

Finding the Worst 100 mm — A 100-mm window is shifted from left to right across the image in 1-mm increments. For each 1-mm step, the indications whose centroids are within the 100-mm window are used to calculate the rating for that step using Eq. 1. Once all of the steps are complete, the software has ratings in 1-mm increments, centered from 50 mm from the left to 50 mm from the right of the considered region. From these ratings, the worst is selected and the rating for this region becomes the rating for the sample.

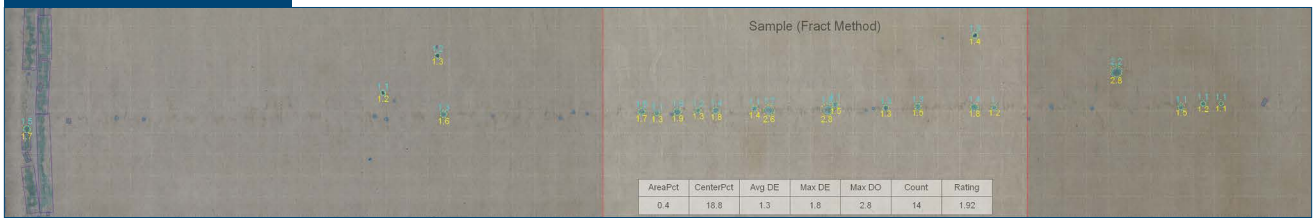
Marking Up the Image — Finally, after measurement is complete, an output image is created showing the measured indications, with the worst 100-mm portion highlighted, and some metrics and a final rating for

Figure 3



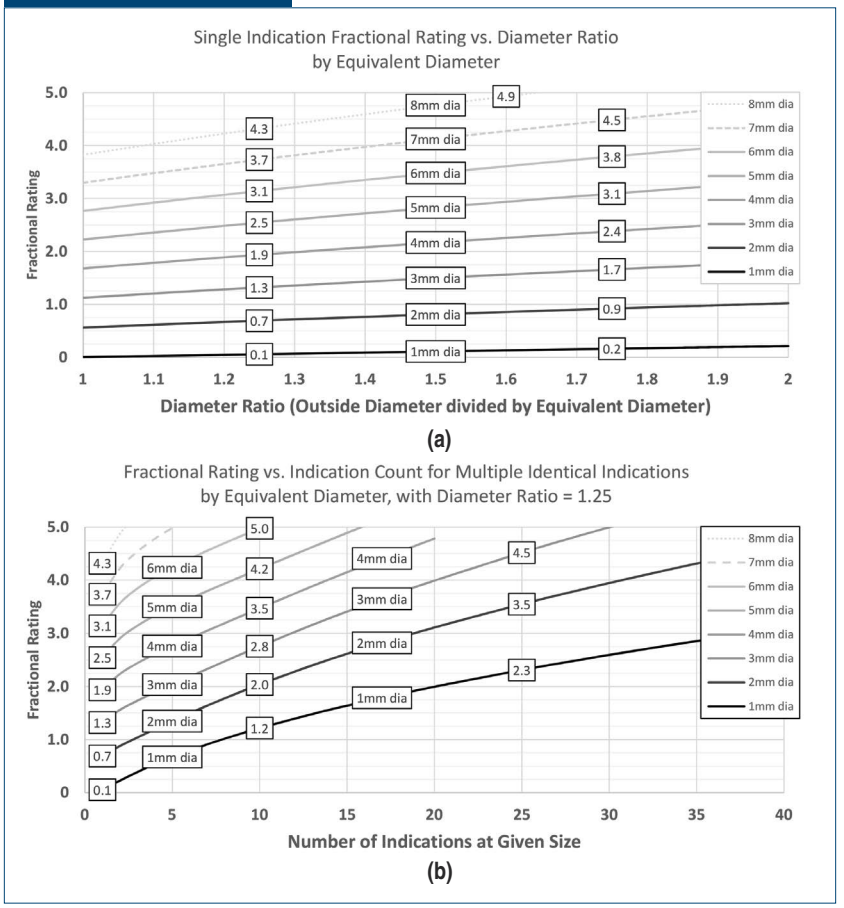
Measurements for calculating rating based on size of indications, applied to the detection output from Fig. 2. Grid with 5-mm spacing added for reference.

Figure 4



Marked-up image for the considered region of Fig. 1, showing the rating for the worst 100-mm area.

Figure 5



Fractional rating vs. diameter ratio for single indications and for multiple indications (a) vs. number of indications for increasing equivalent diameter (up to 8-mm equivalent diameter shown) (b).

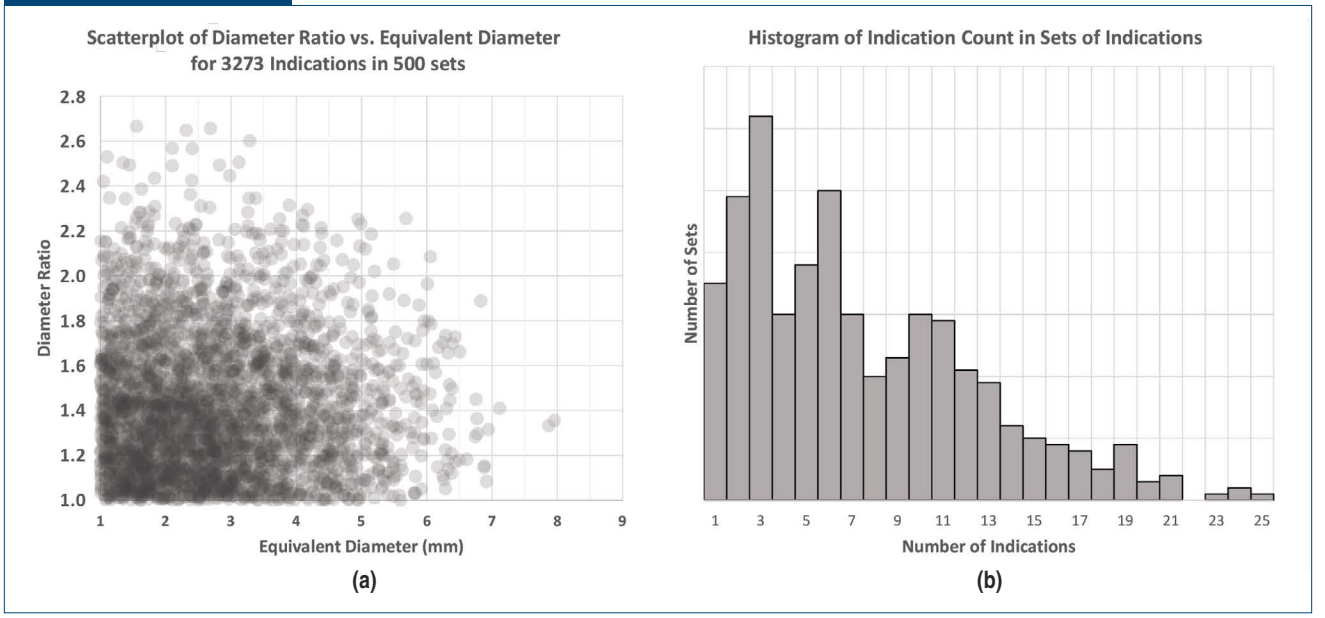
that 100-mm portion. The number above each indication's bounding circle is the equivalent diameter, and the number below each indication's bounding circle is the bounding circle diameter, both in millimeters. This image is shown in Fig. 4.

Calculation Analysis

Eq. 1 was developed with three goals: (i) to make a rating that increases with indication area and with outside diameter so that a small number of large or long indications gives a high rating; (ii) to make the rating increase with centerline coverage percentage so a large number of small indications also gives a high rating; and (iii) to scale the final rating to ensure the reference images from SEP 1611 rated at least their reference level of 1, 2, 3 or 4, to make the resulting rating comparable to other methods of segregation image analysis.

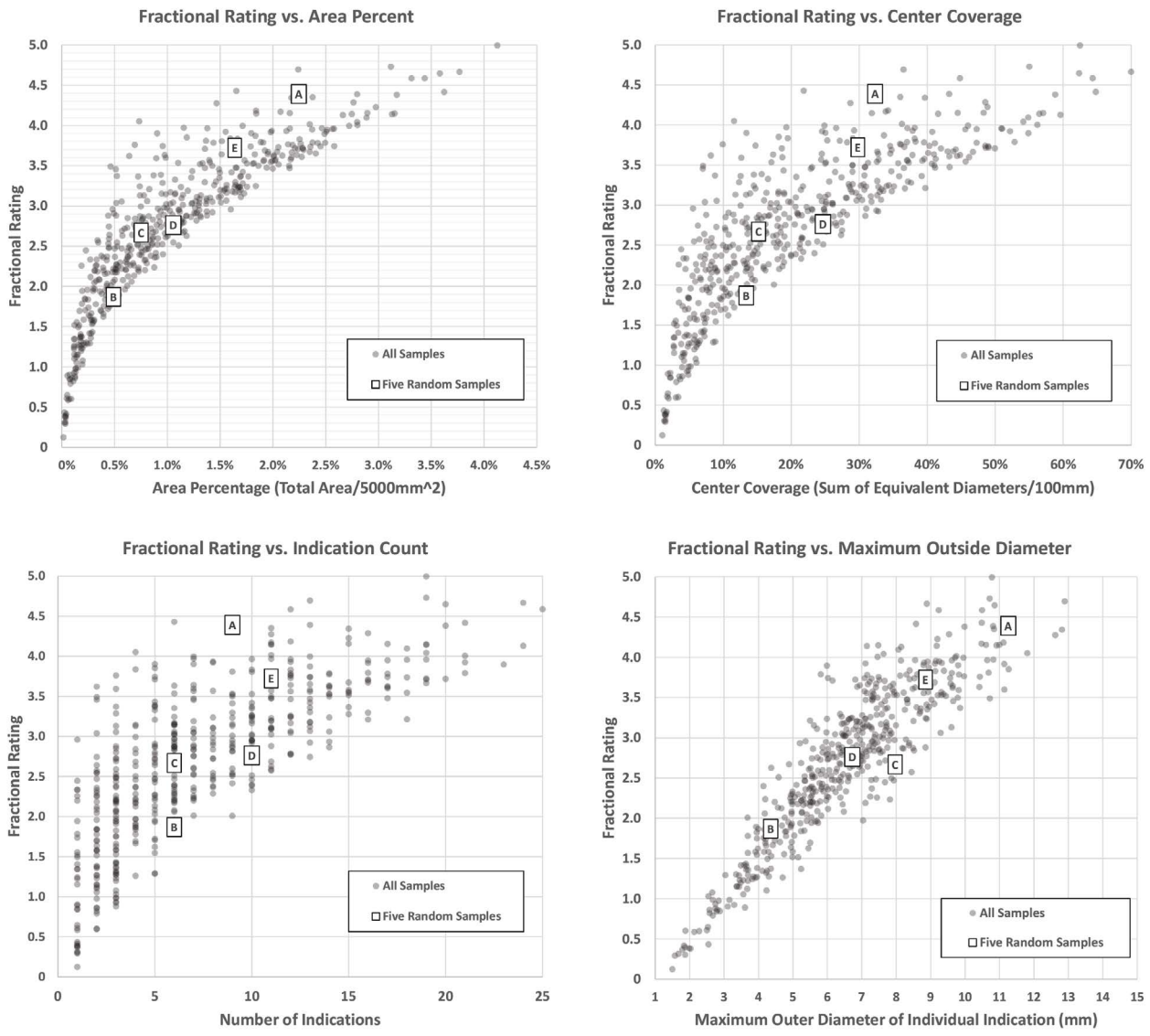
To show how the equation meets goals (i) and (ii), Fig. 5a illustrates the relationship between the diameter ratio and fractional rating for a single indication of different sizes.

Figure 6



Charts showing the indication sizes (a) and counts for randomly generated sets used to test the rating equation (b).

Figure 7



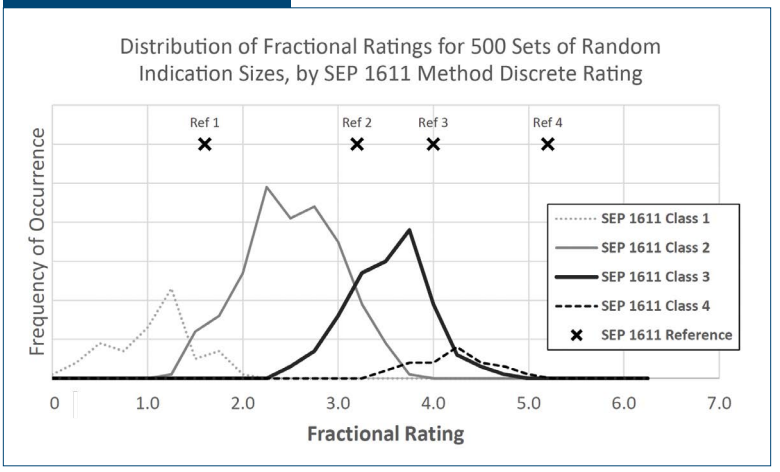
Charts showing the relationship between fractional rating and other metrics for randomly generated sets of randomly sized indications.

This shows that a single indication of the smallest considered size (1-mm equivalent diameter with diameter ratio of 1.0) gives a fractional rating 0, and also shows that a single indication can take the fractional rating well above 4.0 as the combination of equivalent diameter and diameter ratio increases. Fig. 5a shows the increase in fractional rating with the number of indications of a given size (assuming identically sized indications for simplicity). This chart was generated with a diameter ratio of 1.25, so the first column of labels (for a single indication of the given size) matches the column of labels at diameter ratio 1.25 in Fig. 5a.

To demonstrate the ratings that arise from Eq. 1 on a wider range of inputs, 500 sets of indications were generated, each with between 1 and 25 indications, with each indication having random equivalent diameter between 1 and 9 mm and random outside to equivalent diameter ratio between 1.0 and 2.7. Fig. 6 shows a scatterplot of diameter ratio versus equivalent diameter for the individual indications, and a histogram of the number of indications in each set. These sets of indications were counted with the proposed method as well as the counting method outlined in SEP 1611.

Fig. 7 shows the relationships between fractional rating and four other metrics. Squares labeled A through E are the ratings for five randomly selected sets of indications that can be used to compare what makes a rating between the charts; for example, D has only slightly higher rating than C, even though

Figure 8



The distribution of fractional ratings for randomly generated sets of randomly sized indications, grouped by the four discrete rating classes of SEP 1611.

Figure 9

Worst 20 x 100-mm portions of 12 samples with increasing fractional rating.

D has 10% more centerline coverage and a higher number of indications. This is explained by C having a higher maximum outside diameter.

Finally, to illustrate how the equation meets goal (iii), Fig. 8 shows the distribution of fractional ratings for the same 500 sets of random indications used for Fig. 6 and 7, but grouped by the discrete rating found using the counting method of SEP 1611. Near the top of the chart, there are four X markers labeled with "Ref 1" through "Ref 4" showing where the reference images from SEP 1611 were measured on the fractional scale using the proposed method. When the same measured indications were counted using SEP 1611, the images fell into the correct classes of 1, 2, 3, and 4, while using the proposed equation put their fractional rating at 1.62, 3.16, 4.03 and 5.19, respectively.

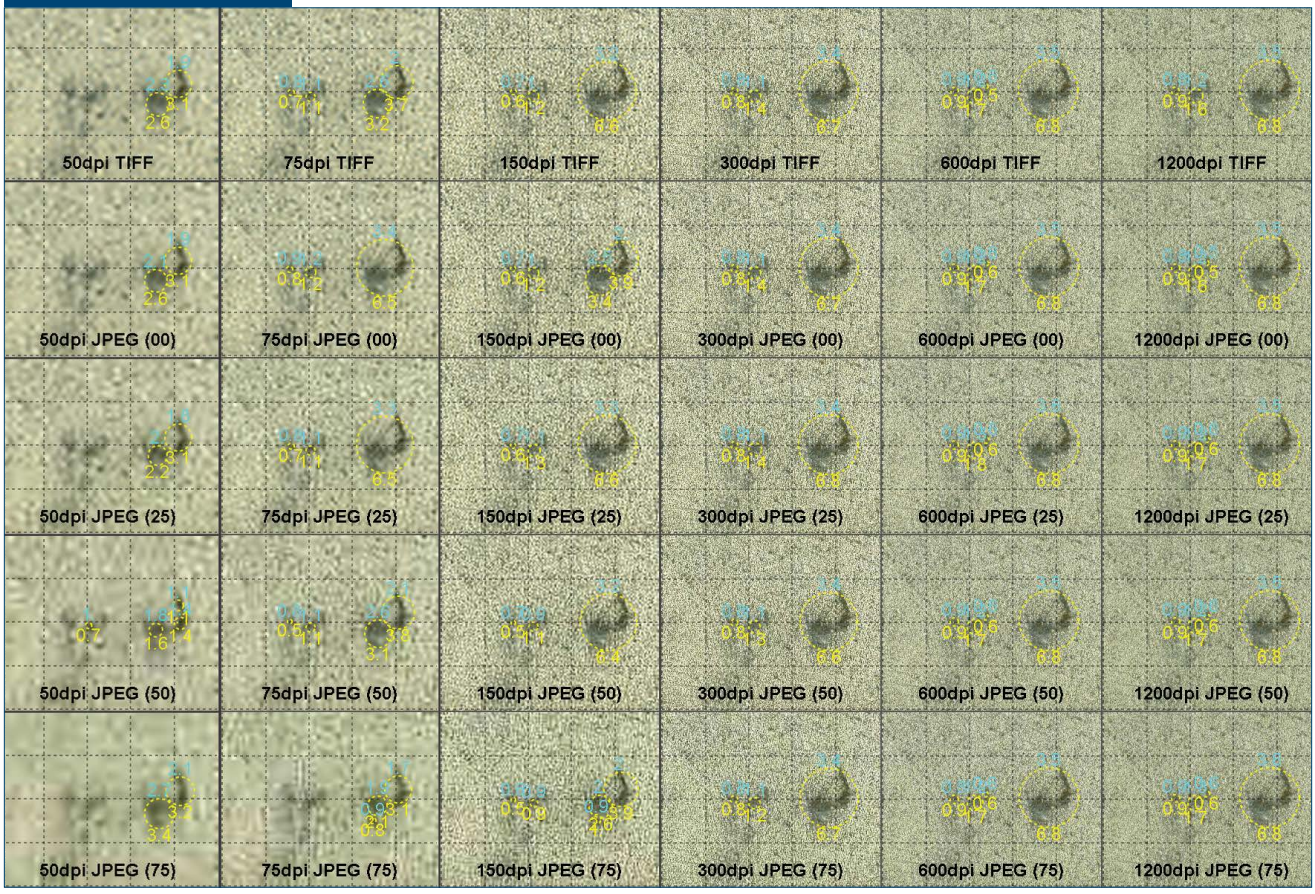
Example Results

To demonstrate varying levels of detected segregation, spanning the fractional range from 0.5 to over 3.0, 12 images were selected with fractional rating increasing by approximately 0.25 per image, and the worst 100-mm-wide area, cropped to 20 mm tall for each image is shown in Fig. 9, with the original image on the left and the measurements overlaid on the filtered image on the right.

Sensitivity Analysis

When dealing with sample images taken over several years, there can be image quality variation due to equipment and practice changes. In order to test the effect of image quality parameters on the analysis results, four experiments were done: three were done on the level of individual indication measurements to determine the effects of image resolution and compression, height above the scanner platen, and common image adjustments; the fourth experiment determined the effects of common image adjustments on the rating of several samples. Note that for the images of measured segregation in this section, the equivalent diameter circle is not drawn so that the

Figure 10



Effects of image resolution and compression ratio on detection of indications. The numbers in parentheses are the JPEG compression levels in percent.

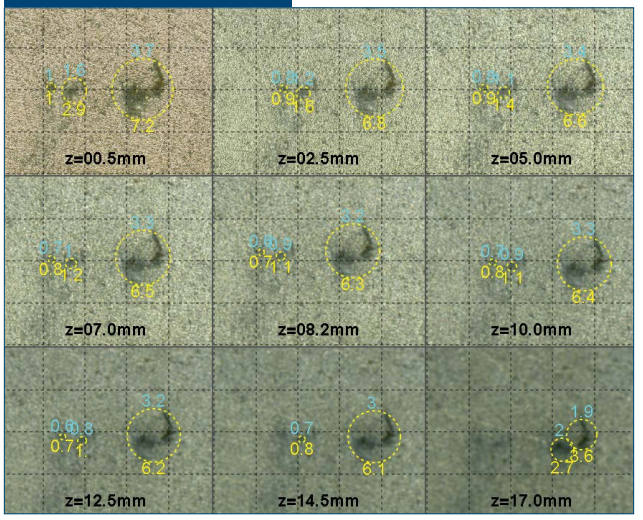
effects of adjustments on indication representation can be examined. As with the previous output images, the number above the indication's outer circle is its equivalent circular diameter, and the number below the indication's outer circle is its bounding diameter.

Effects of Image Resolution and Compression — In order to test the effect of changing the image resolution (dots per unit length of sample) and software image compression on the analysis results, a 20 x 25-mm portion of a segregation sample with some noticeable indications was scanned multiple times on a non-production flatbed photo scanner with pixel density between 50 and 1,200 dots per inch (dpi) (between 2 and 47 dots per mm), and saved in an uncompressed TIFF file. These files were each converted to multiple JPEG-compressed images with increasing compression levels from 0% to 75%. Each image was then processed through the detection and measuring algorithm. After measuring, but before annotating, the images were resized up or down to 300 dpi so that they could be shown together in a grid with the same font size.

Fig. 10 shows that the effects of changing resolution and image compression on detection output is quite small for resolutions at or above 300 dpi (11.81 dots per mm). At lower resolutions, the big detected indication on the right was split into two for some or all compression settings. Looking at the shape of that indication and these results, a decision should be made whether to split thinly connected indications or merge close indications, as has been done in the other referenced dot counting methods.

Effects of Height Above Scanner — Because flatbed scanners used have a glass surface through which the sample is scanned, metal spacers are often used to keep the steel sample from touching the surface and damaging the glass. As the distance from the glass surface increases, the scanned image is affected in two ways: the light becomes more diffused and the focus of the scanner is reduced. Because the analysis is not looking for microscopic indications, both of these effects in small quantities can make it easier to distinguish dark areas from the noise around them, but the effects on detection should be quantified. Using metal shims of increasing thickness, the height above the glass

Figure 11

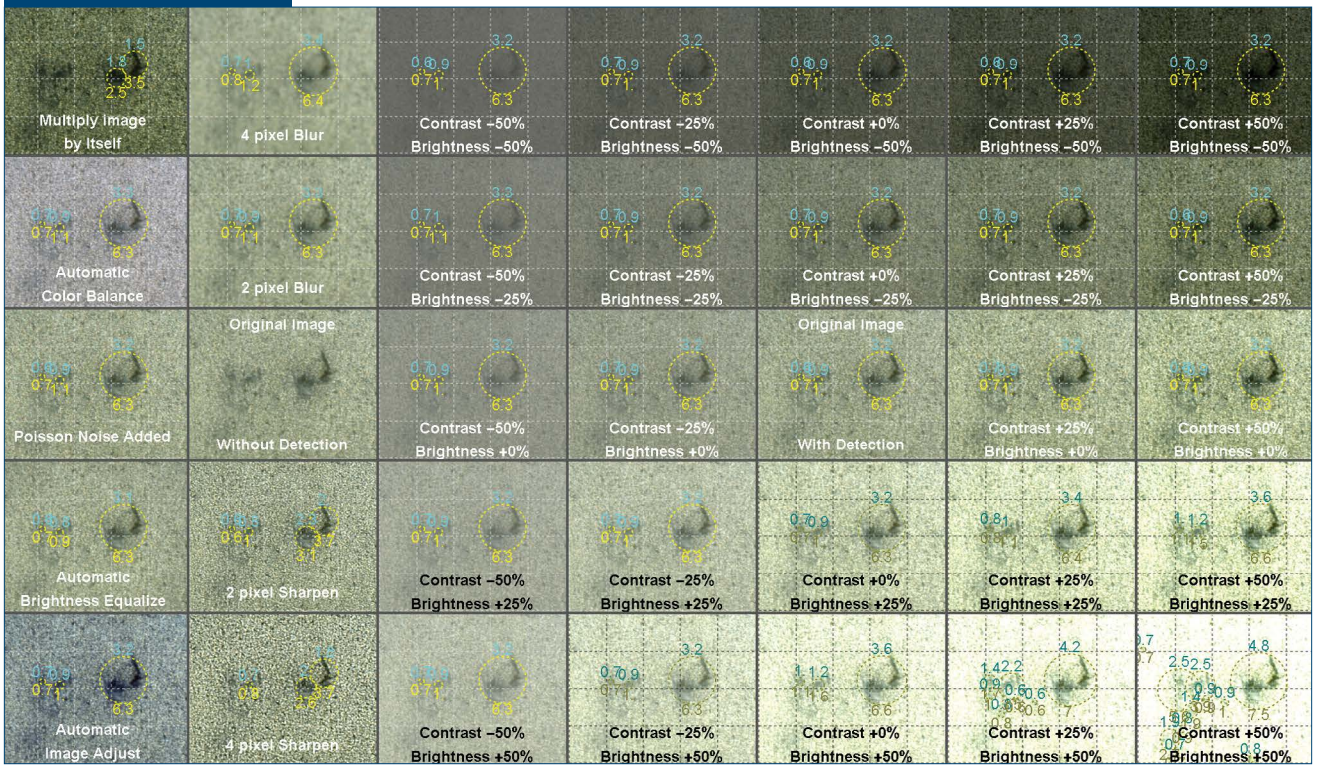


Effects of sample height above scanner glass surface (z dimension) on the detection of indications.

surface was increased and a similar area of indications measured in Fig. 11 was scanned into a 1,200 dpi TIFF file. The results of running these images through the detection and measuring algorithm are shown in Fig. 11. The images scanned for Fig. 10 were scanned at 2.5 mm above the glass because that was the distance when the sample was bridged across the scanner's frame. To get down to 0.5 mm for Fig. 11, very careful handling was necessary to preserve the scanner's glass. Fig. 11 shows that the effects of height above the scanner on measurements are small until the distance exceeds 12.5 mm, when the small indications start to disappear and the large indication is detected as two smaller indications.

Effects of Image Adjustments on Detection Output — In order to test the effects of image adjustments on the detection of individual indications, 33 different adjustments were made on detected indication size for two small images: one taken from the previous analysis step where height above glass was 8.2 mm (resized to 300 dpi), and the other taken from a historical sample image that had some noticeable indications. The first column has five images made with some available adjustments of image quality: multiplying the image by itself (each pixel's red, green and blue values on a scale from zero to one are squared), adding noise to the image, and three different types of automatic adjustments. Two levels of blurring and two levels of sharpening adjustments are arranged in the second column on either side of the original image. The 24 remaining adjustments (columns 3 through 7) are brightness and contrast adjustments arranged so

Figure 12



Effects of image adjustments on detection of indications for an image from previous analysis step.

going from left to right shows the contrast adjustments and top to bottom shows the brightness adjustments.

Fig. 12, which displays the adjustments on the image used for the center of Fig. 11, shows that the largest effects of adjustments come from large amounts contrast and brightness together, where the background pixels start to be detected. Some of the other adjustments including sharpening and adding noise caused the same splitting of the large indication that was seen in the previous analysis steps. Fig. 13, which is from a historical sample scanned on EVRAZ Regina's production scanner setup with production settings, is less sensitive to the same set of adjustments.

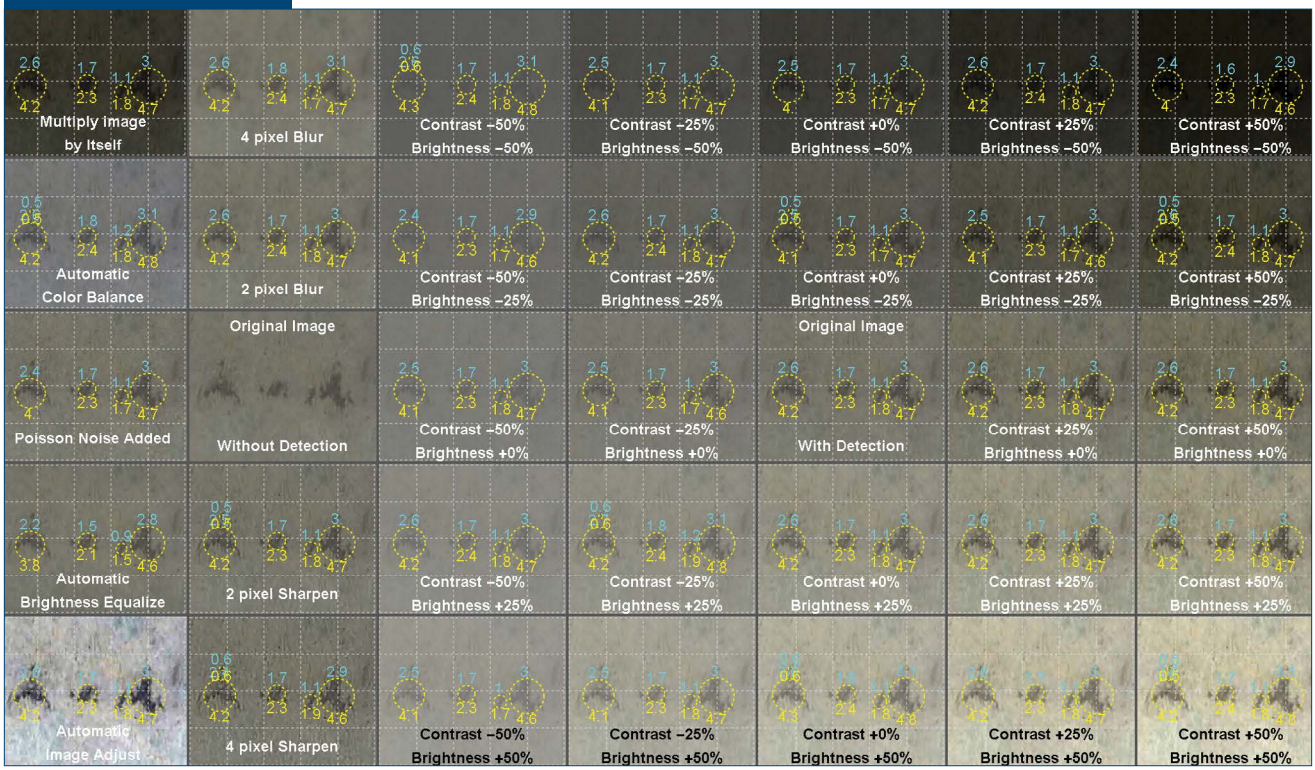
Effects of Image Adjustments on Ratings — In order to test the effect of image adjustments on the proposed method including rating, 21 samples with fractional rating between 0 and 4.0 were put through the proposed method to extract the worst 50-mm-high by 100-mm-wide area. Each 50 x 100 mm image portion was adjusted for all possible combinations of brightness, contrast and blur/sharpen used in the previous analysis step plus the five other adjustments, giving a total of 129 adjustments. The 2,730 resulting images (including one per sample with no adjustments where all three adjustment levels were zero) were run through the analysis to measure their fractional rating. Fig. 14 shows the change in fractional rating vs. the rating for the unadjusted sample, the change in fractional rating vs. the amount of blur/sharpen, brightness or contrast adjustment, with separate markers for that change alone, and the histogram of change in fractional rating.

Table 1 shows the percentage of samples within a given rating amount of the unadjusted image. Reviewing the examples where the rating changed by more than 1.0 (35 of 2,709 adjustments): 32 of them had very high brightness adjustments (+50% or –50%), two were from the automatic brightness adjustment, and the last one was from multiplying the image by itself. The adjustments where the rating increased dramatically tended to be where the background noise started to be detected as segregation. As shown in the charts on the right of Fig. 14, the biggest changes in rating came from combinations of +50% contrast, +50% brightness, with varying amounts of blur/sharpen. This is the same effect seen at the bottom right of Fig. 12.

Next Steps

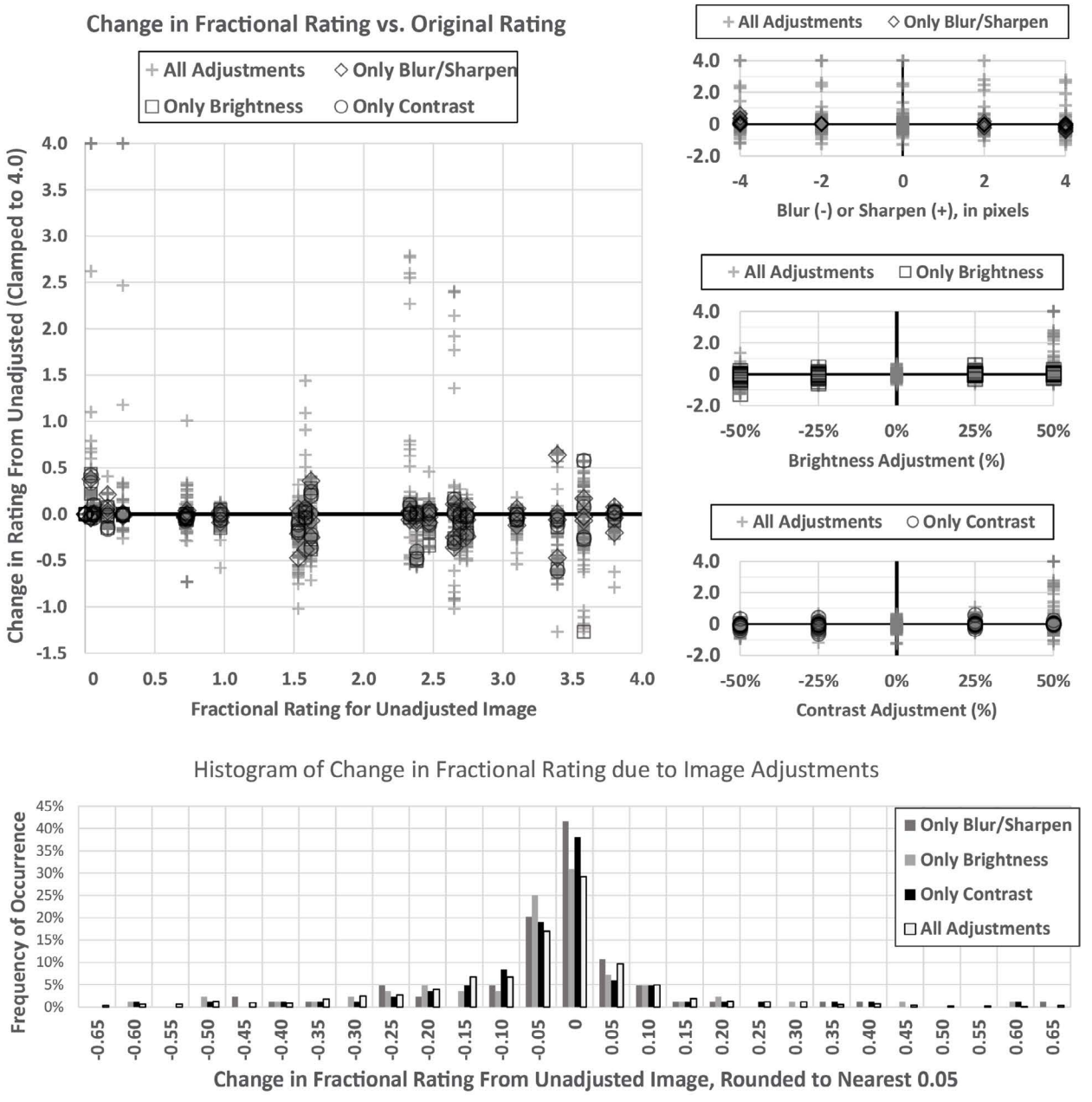
The prototype implementation of this method was developed using the Wolfram Language and runs in

Figure 13



Effects of image adjustments on detection of indications on an image from a historical sample.

Figure 14



Effects of image adjustments on fractional rating for 21 images.

Table 1

Percentage of Image Adjustment Effects Within Increasing Amounts of the Original Rating

Adjusted rating	All adjustments	Only blur/sharpen	Only brightness	Only contrast
Number of samples	2,709	84	84	84
Within 0.1	62.6%	79.8%	67.9%	70.2%
Within 0.2	79.0%	84.5%	81.0%	83.3%
Within 0.5	95.1%	98.8%	96.4%	97.6%
Within 1.0	98.7%	100%	98.8%	100%

Wolfram Mathematica.[5] In order to allow integration with third-party automation systems, the production version of this method will use a wrapper application, and that application will call the Wolfram Language algorithm to run the analysis.

Due to some remaining effect on ratings found during the sensitivity analysis, there should be some work done toward standardizing some aspects of preparation and etching of samples and the acquisition of image in order to make this or any other analysis method comparable across steel producers of similar products. A decision on what to do with thinly connected and very close indications should be made for the proposed method. Detection of some types of adjustments should also be possible using automated image analysis. These will be the topic of a future paper.

Another future topic should be a comparison between ratings generated by as many automated methods as possible on anonymized samples from as many continuous casters as possible. The analysis should include some sensitivity analysis to effects like image adjustment and etching parameters. This will require some collaboration between steel manufacturers, but it will give consumers some confidence in switching between methods once they understand the differences in results.

Conclusions

An innovative method has been developed using image analysis software to address many of the problems presented when quantifying centerline segregation from scanned images of etched steel samples. The method removes subjectivity from the process by automating the detection and measuring once the scanned image is presented and it is robust with respect to some levels of image adjustments and variations in image quality. The method measures on a continuous scale to improve process understanding and to allow smaller improvements to be quantified. The output of the method has been scaled such that the ratings may be considered on a comparable scale to other methods being used in the steel industry. Some further analysis is required before this method can be used across the industry.

Acknowledgments

The authors wish to thank Brad Forster and the EVRAZ Regina Management team for the opportunity to prepare the analysis method and this paper about the method. The EVRAZ Regina Technical Services Team is also acknowledged for their input into the quality and accuracy of this method and for their knowledge of other analysis methods. Thanks also go to Josef Watzinger from Primetals Technologies Austria GmbH for introducing the authors to SEP 1611.

References

1. SMS group, “Classification of Defects in Materials – Standard Charts and Sample Guide,” SN960, October 2011.
2. S. Abraham, J. Cottrell, J. Raines, Y. Wang, R. Bodnar, S. Wilder, J. Thomas and J. Peters, “Development of an Image Analysis Technique for Quantitative Evaluation of Centerline Segregation in As-Cast Products,” *AISTech 2016 Conference Proceedings,* 2016.
3. S. Rapp, “Requirements of the MAOP Rule and Its Implications to Pipe Procurement,” *INGAA Foundation Best Practices in Line Pipe Procurement and Manufacturing Workshop,* Houston, Texas, USA, June 2010.
4. “Evaluation of Centerline Segregation of Continuously Cast Slabs,” SEP 1611, Steel Institute VDEh, Düsseldorf, Germany, October 2018.
5. Wolfram Research Inc., Mathematica, Version 12.0, Champaign, Ill., USA, 2019. ✦