Automated Image Analysis of the Severity of Foliar Citrus Canker Symptoms


ABSTRACT

Citrus canker (caused by Xanthomonas citri subsp. citri) is a destructive disease, reducing yield and rendering fruit unfit for fresh sale. Accurate assessment of citrus canker severity and other diseases is needed for several purposes, including monitoring epidemics and evaluation of germplasm. We compared measurements of citrus canker severity (percent area infected) from automated image analysis to visual estimates by raters and true values using images from five leaf samples (65, 123, 50, 50, and 200 leaves; disease severity from 0 to 60%). Severity on leaves was measured by automated image analysis by (i) basing threshold values on a presample of leaves, or (ii) replacing healthy leaf color on a leaf-by-leaf basis before automating image analysis. Samples 1 to 4 were assessed by three trained plant pathologists, and sample 5 was assessed by an additional 25 raters. Healthy leaf area color replacement gave the most consistent agreement with the true severity data. Using color replacement agreement, with true values based on Lin’s concordance correlation coefficient (ρ) was 0.93, 0.79, 0.71, 0.85, and 0.89 for each of the samples, respectively. The range and consistency of agreement was generally less good for automated thresholds based on a presample (ρ = 0.35–0.90) or visual raters (ρ = 0.30–0.94). The constituents of agreement (precision and accuracy) showed similar trends. No one rater or method was best for every leaf sample, but replacing healthy color in each leaf with a standard color before automation of image analysis improved agreement, and was relatively quick (20 s per image). The accuracy and precision of automated image analysis of citrus canker severity can be comparable to unaided, direct visual estimation by many raters.

Corresponding author: Clive Bock
E-mail: clive.bock@ars.usda.gov

MATERIALS AND METHODS
Leaves, leaf images, and rater assessments. Five samples of detached leaves used in this study came from three sources of infected grapefruit (Table 1): (i) Naturally infected, container-grown plants in Dade County, Miami, FL. Leaves were chosen arbitrarily to reflect the frequency and range of citrus canker severity at the time. Two sets of images (65 and 123 leaves) were taken from these trees on two occasions (leaves photographed under ambient light). (ii) Leaves from container-grown grapefruit seedlings inoculated with a suspension of citrus canker. Two sets of images (50 leaves each) taken at the same time came from this source (leaves photographed under 4 × 250W incandescent blue photoflood bulbs). (iii) A sample of 200 leaves from several naturally infected grapefruit trees in south Florida, collected by survey crews of the

Table 1. Numbers of raters, leaves, and true disease severity range and the image analysis threshold settings used for automated image analysis for measuring citrus canker severity in each sample of canker-infected grapefruit leaves

<table>
<thead>
<tr>
<th>Leaf sample</th>
<th>Number of raters</th>
<th>Number of leaves</th>
<th>True severity range (%)</th>
<th>Threshold setting for automated estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>65</td>
<td>0-59</td>
<td>103-111, 100-109, 102-107</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>123</td>
<td>0-34</td>
<td>103-111, 100-109, 102-107</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>50</td>
<td>0-33</td>
<td>105-117, 106-113, 108-114</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>50</td>
<td>0-60</td>
<td>105-117, 106-113, 108-114</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>200</td>
<td>0-37</td>
<td>96-103, 94-104, 95-102</td>
</tr>
</tbody>
</table>

doi:10.1094/PDIS-93-6-0660
This article is in the public domain and not copyrightable. It may be freely reprinted with customary creditining of the source. The American Phytopathological Society, 2009.
Citrus Canker Eradication Program (CCEP) in 2005. The leaves were sampled arbitrarily from survey bags and imaged under the same conditions (leaves photographed under ambient light).

For all leaf samples, leaf miner-infested, incomplete leaves, or leaves affected by symptoms of other diseases were not used. Some aspects of the reliability, precision, and accuracy of the rater estimates and sources of error associated with the estimates for some of these image samples have been previously described (3–5).

All leaves were photographed using either a 6.5 megapixel Canon EOS 10D (24 to 85 mm AF lens) or a 1.6 megapixel Sony FD Mavica (used on leaf samples 1 and 2 only). Image size was standardized and printed (3.25″ × 2.25″) to approximately reflect the size of the leaf for a total of six images per page. All leaf images were assessed once for percent area of the combined necrotic plus chlorotic symptoms by three plant pathologists. In addition, two plant pathologists and 23 inexperienced raters assessed the leaf images from sample 5 (3–5). All image analysis measurements were taken by an individual trained in the use of the image analysis software and familiar with citrus canker symptoms, but not involved in the visual estimates of disease.

Image analysis measurement of disease. The image analysis software used to measure the diseased area was Assess V1.0: Image Analysis Software for plant disease quantification (American Phytopathological Society, St. Paul, MN, 12). The true values of disease were measured using image analysis on a leaf-by-leaf basis (3,5). Images like the ones used in this study are comprised of pixels with particular values for each of the primary colors red, green, and blue, which result in the color of the perceived image (27). Each pixel has characteristic hue, saturation, and intensity (HSI). The hue is the pure color of the pixel, while the saturation of a pixel is the amount of color (pure hue, which is very specific, to white, which contains a wider range), and the intensity of the pixel relates to its brightness (from the pure hue to completely black). Hue is the most useful color plane for separation of foliar disease on color images (12,31). The HSI option in Assess V1.0 was applied to identify and separate the diseased area from the healthy tissue by using the threshold function. Initially, each image was sized to approximately 1,000 × 600 pixels at a resolution of 300 pixels per inch, allowing a view of the entire leaf surface. Studies have shown that relatively low image resolution is sufficient to provide accurate measurement of disease (31). Images were opened individually and the leaf separated from the background by applying an overlay which provided a clear view of the selected area. Assess V1.0 generates a histogram that reveals the distribution of color in the pixels within the selected area of the image and allows choice of the correct settings to threshold healthy leaf and diseased areas of the image. With the overlay set to outline, the thumbnails (program controls which determine the color range) were adjusted until the best fit of the diseased area was selected. Once selected, the program calculates the area in pixels and retains that value to calculate the percent area diseased. Thumbnail settings for the leaf and lesion were retained, providing a guideline for the subsequent image measurements. To measure true values accurately, image enhancement was necessary to optimize measurements because of inconsistency in lighting among images, slight differences in the color of healthy leaf tissue, and variation in necrosis or chlorosis color. Assess V1.0 has filters, contrast, and color saturation functions, as well as color balancers that were applied on a leaf-by-leaf basis and were used to consistently enhance the entire area of interest to maintain measurement accuracy recording the true values. Repeatability of this system has been demonstrated (3).

Subsequent to measuring the true values, the images were automatically batch processed in Assess V1.0 using different thresholds. The batch processing macro automates image analysis of leaf samples, and can be paused so the threshold on each image can be reset. However, this is relatively time-consuming for large numbers of images and requires that an operator be present during the measurements. Thus, various thresholds based on presamples, and image pretreatment were tested to determine whether complete automation could be achieved with minimal loss in quality of measurement of canker severity.

(i) The effect of setting thresholds at predetermined levels for each sample was

**Fig. 1.** Correlation coefficients of various raters and automated image analysis methods for the estimate of citrus canker severity (% area infected) against the true severity (measured by image analysis) as a measure of precision of the estimate. Data shown for leaf samples 1 to 5 are A to E, respectively.
examine by taking different subsets of 10 images from the samples and identifying “average” threshold settings for the symptoms. Based on these tests, threshold settings of 103-111, 100-109, and 102-107 were applied to images in samples 1 and 2, of 105-117, 106-113, and 108-114 in samples 3 and 4, and of 96-103, 94-104, and 95-102 in sample 5. The samples of leaf images were batch processed without pausing as described in Assess V1.0 at these thresholds to compare the effect of fixed threshold choices on estimates of severity.

(ii) A method of rapidly standardizing healthy leaf tissue prior to automated image analysis in Assess V1.0 was developed. Healthy tissue color was replaced in each image individually with R: 7 G: 228 B: 194 in Photoshop CS (2003 Adobe Systems Inc.). This was achieved for each leaf image by taking a sample of healthy tissue color on that image using the “color picker” tool and replacing it with the standardized healthy tissue replacement color, a process that took approximately 20 s per image, but could not be done in Assess V1.0. Up to 40 images could be opened at once in Photoshop CS. The new image was saved as a copy of the original image for automatic processing in Assess V1.0 at a set threshold (94-114, which separated disease from the standardized leaf color). The replacement color was chosen to maintain contrast of infected areas with healthy tissue and to contrast with the image background (royal blue: R: 51 G: 5 B: 248). Thus, using Photoshop for recoloring was an alternative approach that allowed full automation in Assess V1.0.

Data analysis. The data were analyzed for agreement with the assumed true values using Lin’s concordance correlation coefficient (LCC, 3,13,15,19). LCC is a widely used tool in measurement science and gauges agreement between estimated or measured values (dependent variable, Y) and an assumed true value (independent variable, X) (15). It involves components of accuracy and precision of the estimate and provides a quantifiable method to evaluate the characteristics of agreement. The coefficient (ρc) is the product of elements of precision (correlation coefficient, r, and accuracy, Cb). Cb is the bias correction factor that measures accuracy (generalized bias; 15) and comprises slope (scale) shifts (coefficient υ, with a perfect relation of 1 between X and Y) and elevation (location) shifts (coefficient μ, where 0 is a perfect relation between X and Y). To further compare the consistency of the accuracy of the estimates, the range of discrepancy from the true value was calculated for each rater and automated assessment method. All data were analyzed using SAS (SAS Systems, Cary, NC).

RESULTS

Agreement of visual estimates and image analysis measurements compared to true values. The disease severity range in each sample of leaf images varied, as did the number of leaves (Table 1). In all samples, the leaf images were in focus, although leaf and symptom color and lighting conditions varied among samples. Lighting was constant within samples. There were 50 leaves in the smallest groups (samples 3 and 4) and 200 in the largest (sample 5). Among individual leaves, the lower value for thresholds ranged from 90 to 113, and the upper values for the thresholds ranged from 97 to 121.

Precision. Among these five samples, estimates and measurement suffered some imprecision by both raters and automated image analysis (Fig. 1A to E). Based on the correlation coefficient, precision of rater estimates ranged from 0.80 to 0.95. Precision of measurements by manual-CR was similar (r = 0.81 to 0.94). The ranges of precision among automated measurements based on threshold settings from leaf presamples depended on the setting and the sample on which they were based. For samples 1 and 2, threshold setting precision, r, ranged from 0.51 to 0.94, and for samples 3 and 4, from 0.53 to 0.91. Sample 5 threshold-based measurements were all imprecise (r = 0.69 to 0.71).

Accuracy. Accuracy was measured by the bias correction factor (Cb), which incorporates elements of slope and elevation bias. All methods of estimating or measuring diseased area affected both the slope and elevation (Tables 2 and 3). Evidence of proportional error (error related to magnitude of the true values) was evident among raters.

Table 3. Lin’s concordance correlation coefficient analysis showing slope and elevation shifts of rater estimates and different automated image analysis measurements of the percent area showing symptoms of citrus canker on a sample of canker-infected grapefruit leaves.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>Sample 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slope (υ)</td>
<td>Elevation (μ)</td>
<td>Slope (υ)</td>
<td>Elevation (μ)</td>
</tr>
<tr>
<td>Rater 1</td>
<td>1.34</td>
<td>0.24</td>
<td>1.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Rater 2</td>
<td>0.87</td>
<td>0.08</td>
<td>0.81</td>
<td>-0.08</td>
</tr>
<tr>
<td>Rater 3</td>
<td>1.22</td>
<td>0.19</td>
<td>1.07</td>
<td>0.11</td>
</tr>
<tr>
<td>Manual-CR</td>
<td>1.07</td>
<td>0.14</td>
<td>0.83</td>
<td>-0.08</td>
</tr>
<tr>
<td>1</td>
<td>1.90</td>
<td>0.70</td>
<td>1.36</td>
<td>0.13</td>
</tr>
<tr>
<td>2</td>
<td>1.57</td>
<td>0.22</td>
<td>0.89</td>
<td>-0.17</td>
</tr>
<tr>
<td>3</td>
<td>1.19</td>
<td>0.70</td>
<td>0.62</td>
<td>-0.49</td>
</tr>
</tbody>
</table>

a True percent area diseased measured with image analysis on a leaf-by-leaf basis, υ = slope shift (1 = no bias relative to the concordance line, μ = elevation shift (0 = no bias) relative to the concordance line.

b Threshold settings used for leaf samples 1 and 2 were 1 = 103-111, 2 = 100-109, and 3 = 102-107, and for leaf samples 3 and 4 were 1 = 105-117, 2 = 106-113, and 3 = 108-114.

c All image analysis measurements were done by an individual familiar with citrus canker symptoms but not involved in the visual estimates in this study.
estimates by raters with slopes (υ) ranging from 0.57 to 2.45. The slope for measurement by manual-CR was more consistent and ranged from 0.79 to 1.07. With automated measurements using threshold values from presamples, slope was variable and dependent on the sample and threshold setting. For samples 1 and 2, slopes of threshold setting ranged from 0.62 to 1.90. For sample 4, some threshold settings (106-113, 108-114) and samples had no slope bias (υ = 1.00), but the overall range for sample 3 and 4 was from 0.45 to 1.16. Sample 5 also showed slope bias (1.15 to 1.47) with measurements based on presample threshold values.

Among raters, elevation (over- or under-estimation) of the line (μ) ranged from -0.41 to 0.83, reflecting a systematic error in some raters estimates resulting in bias (Tables 2 and 3). The elevation of the line for measurements by manual-CR ranged from -0.34 to 0.14, and for automated measurements based on threshold settings from presamples, the elevation was variable and dependent on sample and setting. Thus, elevation of samples 1 and 2 ranged from 0.13 to 1.07, and for samples 3 and 4 from -0.49 to 0.13. For sample 5, the threshold settings also showed some elevation bias (0.03 to 0.16).

The product of slope and elevation bias, the bias correction factor (Cb), showed that all methods suffered some generalized bias compared to true values (Fig. 2A to E). Among raters, bias ranged widely from 0.38 to 1.00 (no bias), but was consistently good for all samples measured by manual-CR (Cb = 0.92 to 0.99). For other automated measurements based on presample threshold settings, the magnitude of the bias correction factor was highly dependent on the sample and setting, although certain settings did give unbiased measurements, indicated by values close to 1.00. Samples 1 and 2 had a Cb of 0.56 to 0.98, and samples 3 and 4 from 0.69 to 0.99. For sample 5, the threshold settings based on leaf presamples also resulted in some inaccuracies (Cb = 0.92 to 0.99), although at least one setting (95-102) was virtually unbiased (0.99).

Agreement. The overall measure of agreement (LCC), ρc, between the estimated or measured values and the true values reflected the measures of precision and accuracy (Fig. 3A to E). Among raters, agreement ranged widely from 0.30 to 0.94, but was more consistent for measurements by manual-CR (ρc = 0.71 to 0.93). For automated measurements based on presample thresholds, the magnitude of agreement was dependent on the sample and setting. For samples 1 and 2, ρc ranged from 0.35 to 0.90, and for samples 3 and 4 from 0.50 to 0.90. For sample 5, the three threshold settings used resulted in mediocre agreement (ρc = 0.61 to 0.70).

The range of estimate or measurement discrepancy from the true value of disease severity was less for manual-CR and tended to be evenly represented by over- and underestimation and was similar in magnitude to the rater estimates discrepancies (Fig. 4A to E). Automated measurements based on presample thresholds had greater ranges and were more biased to over- or underestimation.

**Time to assess images.** By far the most rapid method was direct estimation by a rater of the area diseased, which took, on average, approximately 7 s per leaf. Time to photograph each leaf took approximately 10 s. To analyze each leaf image individually using image analysis (but most accurately and precisely) took 7 min (7:10 min total time including photography) per leaf (including all enhancements to ensure accuracy). Manual-CR took 50 s per image (1:00 min total time including 10 s for photography, 20 s for recoloring, and 30 s for automated measurement of the diseased area), eight to nine times longer per leaf to record the area infected compared to raters, but provided similar accuracy and precision compared to many raters. Automated measurements based on presample thresholds took on average 30 s per image (40 s total time including photography), but were inconsistent.

**DISCUSSION**

When performed on a leaf-by-leaf basis, image analysis can provide repeatable measurements in close agreement with true values (3,5). However, where large numbers of samples need to be assessed, full automation is particularly advantageous to conserve time, while ensuring accuracy and precision of measurements. Rater estimates are often not in agreement with true values (3,5,16,19,21), while auto-

![Fig. 2. Bias correction factor (Cb) of various raters and automated image analysis methods for the estimate of citrus canker severity (% area infected) against the true severity (measured by image analysis) as a measure of accuracy (generalized bias) of the estimate. Data shown for leaf samples 1 to 5 are A to E, respectively.](Image)
mated image analysis can show good agreement (14,16). An objective of this study was to use a readily available image analysis system to test automated image analysis of citrus canker severity on leaf images.

Using a fixed threshold based on a presample was not a reliable way of automating measurement of disease severity. The range of upper and lower threshold values was variable within and between samples, which precludes choosing a setting for the image analysis process, even though a particular threshold setting might provide measurements in close agreement to the true value on a single occasion (Fig. 3A, threshold setting 102-107, sample 1). It was not possible to know this in advance, and if the threshold was based on a different presample it could result in poor measurement agreement due to imprecision and/or inaccuracy, and was often not as good as rater estimates. These results reflect the fact that there are within- and between-sample variations in the characteristics of leaves, and between-sample variation in photographic conditions. Leaf images vary due to difference in lighting, surface reflection, saturation, hue, symptom, or healthy leaf area color which may be due to many factors including leaf age, symptom age and progress, and nutritional status of the plant, or quality of the photography (9,25,31,32). Standardizing imaging of nondestructive samples in the field is especially challenging, particularly if large numbers of leaves need to be photographed. Nonetheless, acceptable agreement at least comparable to that of direct visual rating was achieved with automated image analysis of citrus canker severity by healthy tissue color replacement on a leaf-by-leaf basis (manual-CR). The color replacement was fast (20 s per leaf), and once completed allowed full automation at a set threshold. Replacing healthy leaf color on a leaf-by-leaf basis removed some error due to variation in the color of healthy leaf areas. Further error might be reduced by more standardization of photographic conditions including using various polarizing filters that can help reduce reflection. These results show that automated image analysis of citrus canker symptoms can give measurements of disease severity better than some previously reported (21), and often better than direct estimates by many of the raters.

Some commercial and custom image analysis systems have automated functions that have built-in standards and algorithms to account for variation in leaf and symptom color and characteristics, which enhance agreement of the measured value with the true disease severity (14,16). However, these systems have not been widely applied in measuring plant disease severity, perhaps due to their complexity and/or expense. The measurements we made were done using Assess V1.0. A new version of Assess (V2.0) has just been released and is now being tested. It incorporates an algorithm to provide an automatic threshold feature for each image (Lakhdar Lamari, personal communication). This should improve agreement and the speed of measurement of disease severity. In this study, raters were consistently the fastest at assessing images (7 s per image), a time that agrees with a previous report (16). A potential limitation to automating image analysis of plant disease severity is the ability to differentiate multiple diseases. Symptoms of more than one disease can occur on a leaf, and humans can use their experience, intelligence, and logic to differentiate and quantify these. Nonetheless, where individual diseases are present, automated image analysis can be very useful (16), but in some field applications where symptoms might be associated with multiple diseases, measurement will be error prone without further development of artificial intelligence.

Various characteristics and sources of rater error have already been identified in visual assessment of citrus canker (1,3–5). Advantages of image analysis include its immunity to some of these sources of error, including the illusion of greater area infected in relation to lesion size and number (3,4,30). Image analysis is tireless, particularly advantageous where large numbers of assessments are required. Images can be kept as a permanent record and referred to later if needed, and sampling need not be destructive. Despite the advantages of image analysis, visual assessment will remain a very important method, making training or the use of assessment aids an important facet of canker assessment (1,20), and as a full understanding of sources of error by raters is established, the development and targeting of assessment training can be enhanced.
LITERATURE CITED


