Development and Validation of Standard Area Diagrams as Assessment Aids for Estimating the Severity of Citrus Canker on Unripe Oranges

Ricardo Braido and **Aline M. O. Gonçalves-Zuliani,** Núcleo de Pesquisa em Biotecnologia Aplicada, Universidade Estadual de Maringá, Maringá, Paraná, Brazil; **Vanderly Janeiro,** Departamento de Estatistica, Universidade Estadual de Maringá, Maringá, Paraná, Brazil; **Sérgio A. Carvalho,** Centro APTA Citros "Sylvio Moreira", Instituto Agronômico de Campinas, São Paulo, Brazil; **José Belasque Junior,** Departamento de Fitopatologia e Nematologia, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Brazil; **Clive H. Bock,** United States Department of Agriculture–Agricultural Research Service SEFTNRL, 21 Dunbar Road, Byron, GA 31008, USA; and **William M. C. Nunes,** Núcleo de Pesquisa em Biotecnologia Aplicada, Universidade Estadual de Maringá, Paraná, Brazil

Abstract

Braido, R., Gonçalves-Zuliani, A. M. O., Janeiro, V., Carvalho, S. A., Belasque Junior, J., Bock, C. H., and Nunes, W. M. C. 2014. Development and validation of standard area diagrams as assessment aids for estimating the severity of citrus canker on unripe oranges. Plant Dis. 98:1543-1550.

Asiatic citrus canker (ACC) is an important disease of citrus in Brazil and elsewhere in the world. Infection with the causal pathogen, *Xanthomonas citri* subsp. *citri*, can cause severe disease on the fruit. Visual estimation of severity is the usual method used to quantify ACC on diseased fruit. The objective of this research was to construct and validate standard area diagram (SAD) sets as assessment aids for raters to improve the accuracy and reliability of visual estimates of ACC on unripe (green) fruit of sweet orange. Two SAD sets were constructed. A five-diagram SAD set had five severities depicted (0.5, 2.0, 8.0, 27.0, and 40.0%) and a six-diagram SAD set had six severities depicted (0.5, 1.0, 3.0, 9.0, 20.0, and 40.0%). Fifteen raters evaluated 40 images of cankered, unripe fruit. Both the five- and six-diagram SAD sets significantly improved the accuracy and reliability of estimates. Agreement, measured by Lin's concordance correlation coefficient, was 0.220 to 0.913 when not using SADs, 0.814 to 0.955 when using five-diagram SAD sets, and 0.863 to 0.925 when using six-diagram SAD sets. The five-diagram SAD set was significantly more accurate and reliable compared with the six-diagram set. Possible reasons for this are discussed. Based on the results, the five-diagram SAD set is preferable to use. Although the SAD set was developed for sweet orange, it doubtless has applicability to other citrus, including grapefruit. These SAD sets should be useful for research endeavors where accurate and reliable estimates of the severity of ACC are required.

Asiatic citrus canker (ACC), caused by *Xanthomonas citri* subsp. *citri*, originated in Asia but has been spread through human activity and is currently found in many tropical and subtropical citrus-growing regions. It is considered one of the most important bacterial diseases of citrus worldwide (16,19,35). In Brazil, ACC was first reported in 1957, in Paraná state, in the municipality of Lupionópolis (7,23). The disease symptoms are characterized by erumpent, necrotic lesions with yellow halos (on green tissues), and are found on leaves, fruit, and shoots (6,35). Within fields, the disease is spread by rain splash, often associated with wind, and also by transport on contaminated equipment (12,13,27,34).

Subsequent to the first report in Brazil, attempts were made to eradicate the pathogen from Paraná state, an approach still used in other states in Brazil (e.g., São Paulo). However, in the southern states (including Paraná, Santa Catarina, and Rio Grande do Sul), the practice of eradication was abandoned, and ACC is now considered endemic (23). Paraná remains one of the most important citrus-producing states in Brazil (20) but the severity of ACC during the 1960s and 1970s prevented cultivation of citrus orchards in the northwestern and northern regions of the state. The Agronomic Institute of Paraná continues to encourage citrus production throughout the state but ACC remains an issue.

In areas with environmental conditions conducive for the disease, management of ACC is critical. Much of the research activity and some management decisions are based on estimates of disease severity. In plant disease epidemiology, the severity of a disease represents the proportion of area that is diseased on a plant organ (2). There is a need for accurate and reliable quantification of the severity of ACC on fruit, on which to base sound scientific conclusions or management options. Accuracy in measurement science can be defined as the closeness of the estimated values to the true values (26,33), and reliability is the extent to which the same estimate obtained under different conditions yields similar results (26,33).

Several methods are used to measure or estimate plant disease (13,29) but visual rating remains the most widely used. Due to recognized error in plant disease estimation (10,11,13,31), standard area diagrams (SADs) are used in plant pathology to increase accuracy and reliability of disease severity estimates made by raters (2,21,32). The improvement that can be achieved by using SADs is now well documented for different plant diseases (5,17,21,28, 37,38,41), including for foliar symptoms of ACC (5). The use of SADs reduces error of the estimates made by raters. Correct use of SADs requires the rater to estimate as close to the actual value by interpolating between two severities represented in the SADs (1).

Although some of the methods used to estimate disease on fruit have been very sophisticated, such as three-dimensional fruit surface image analysis (14), other methods have relied on two-dimensional SADs to represent a single face of the approximately spherical fruit. Two-dimensional SADs have been developed for fruit diseases (37) and have been validated and demonstrated to improve both accuracy and reliability of rater estimates. ACC on fruit is important because is can lead to premature abscission and reduced marketability (4,9,18), yet no SADs have been developed for the purpose of assessing severity of ACC on citrus fruit.

However, green, unripe fruit of citrus are a contrast to ripe fruit, which are ordinarily yellow to orange in color. On unripe fruit, the typical corky lesions of canker have chlorotic haloes surrounding them but lesions on ripe fruit lack the chlorotic halo, the edge of which ends abruptly at the healthy, ripe yellow-orange-colored

Corresponding author: W. M. C. Nunes, E-mail: william.nunes@pq.cnpq.br

Accepted for publication 28 April 2014.

http://dx.doi.org/10.1094/PDIS-01-14-0090-RE

fruit rind. There is a need to develop specific SADs that can be used as an assessment aid for both unripe and ripe fruit. The objectives of this research were, first, to construct SADs for unripe fruit as an assessment aid for raters; second, to validate the SADs by comparing disease estimates made without and with the use of these SADs; and, third, to compare two different SAD sets, one with an additional severity depicted to ascertain whether there was an advantages to an additional illustrated severity to help estimate ACC severity in the range of most frequently encountered severities.

Materials and Methods

Development of SADs. Green, unripe fruit (n = 100) of sweet orange (*Citrus sinensis*) 'Baia' presenting symptoms of ACC (necrotic, erumpent, corky lesions, the edge of which are surrounded by a chlorotic halo that blends into the green rind of the unripe fruit) were harvested from a field experiment located in Maringá, in the northwest of Paraná state (latitude 23°25′ S, longitude 52°10′ W, elevation 555 m). Fruit were 6 to 8 cm in diameter. The orchard was planted in April 2003 with trees spaced 3.5 by 5.0 m. Crop management practices were carried out as for commercial citrus production areas in Paraná state, except no ACC control was practiced and no copper-based products were used.

To measure the symptomatic area on the surface of the unripe fruit, the face of each fruit with the most severe symptoms of ACC was photographed with a digital camera (Sony CyberShot 5.1MP). In practice, raters would most often need to assess both faces of a fruit, taking care to ensure that each side was fully represented (i.e., no overlap); however, for the purposes of this study and the validation process, only a single side was assessed (where both sides are assessed, total severity would be divided by 2, so as to obtain the overall severity per fruit). For the photography, fruit were illuminated using a 40-W light bulb placed 15 cm from the fruit to ensure uniform light conditions. The resulting images were analyzed for diseased area (necrotic + chlorotic) using the image analysis program Quant V1.0.2 (39). The diseased area in relation to the total surface area of the fruit was used to calculate the percent fruit surface area with ACC. The minimum and maximum percent severity obtained from the 100 images of the unripe fruit were 0.5 and 39.6%, respectively. The frequency of the severities in different ranges (Fig. 1) shows that approximately 75% of the fruit had an ACC severity ≤24%. The range and maximum disease severity for various plant diseases has been previously characterized (22), and this distribution is common (if ACC is very severe on immature fruit, it often results in premature fruit abscission; 4).

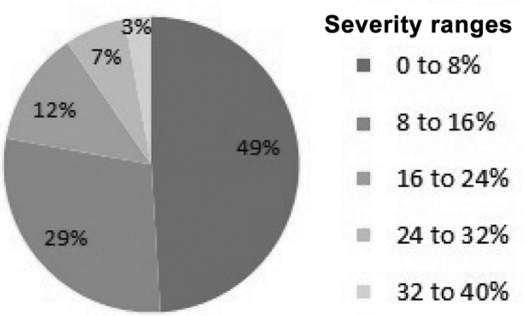

Fig. 1. Proportion of Baia sweet orange fruit (total n = 100 fruit) with different severities (percent area diseased) of Asiatic citrus canker as measured using image analysis. Fruit were collected from an orchard in Parana state, Brazil. Symptoms area measured included the chlorotic halo and the necrotic area.

Based on the range and frequency of severity, two separate SAD sets were constructed. The first had five diagrams and the second had six diagrams. The two SAD sets had slightly different structure; with the six-diagram SAD set, an extra diagram was included in the range of more frequently encountered severity (≤20%). Upper and lower limits were based on the image analysis-measured minimum and maximum ACC severity in the sample of 100 fruit. Thus, the severity values illustrated for the five-diagram SAD set were 0.5, 2.0, 8.0, 27.0, and 40.0% ACC (Fig. 2A), and for the six-diagram SAD set were 0.5, 1.0, 3.0, 9.0, 20.0, and 40.0% ACC (Fig. 2B). The images in both the five- and six-diagram SAD sets were constructed from images of fruit harvested from the same orange trees in the experiment described above. The SAD images were prepared using the image analysis program Quant. V1.0.2. The two different sets were developed to explore whether an additional diagram in the range where most frequent severity was encountered might increase accuracy and reliability, without dramatically increasing the complexity of the SAD set.

Validation of SAD set. To validate and compare these SAD sets, 15 raters estimated the severity of a subset of 40 images of the cankered fruit using a PowerPoint slide presentation, showing each fruit image at random on a computer screen. The raters had a range of experience with disease assessment and familiarity with ACC symptoms. Prior to the first assessment, all raters received the same instructions describing the symptoms of ACC and instructions in use of the SAD set. Initially, each rater estimated the severity of ACC without the aid of the SAD set. After a 30-min break, each rater reestimated severity of ACC on the same fruit images shown

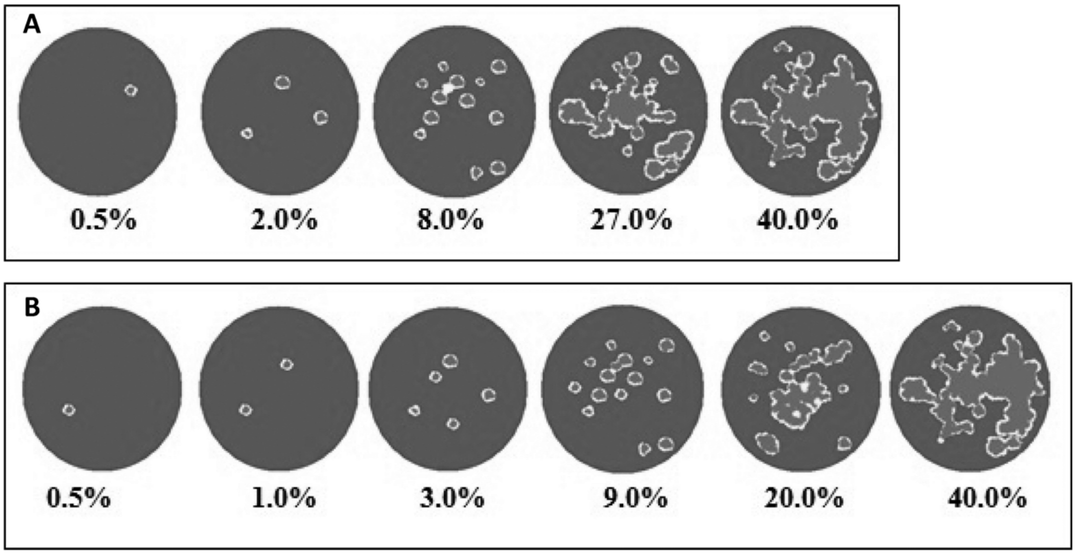

Fig. 2. Standard area diagrams (SADs) used with the same range but with five **A,** or six **B,** severities depicted. The SADs were used by 15 raters as aids to estimate citrus canker on 40 images of unripe (green) Baia sweet orange displaying a range of disease severity.

at random but with the aid of the five-diagram SAD set as an assessment aid; and. finally, after a further 30-min break, the raters assessed the images shown at random a third time using the six-diagram SAD set as an assessment aid.

Data analysis. Analyses were performed using either SAS (V9.3; SAS Systems) or MS Excel (Microsoft Corp.). Data were analyzed using a general linear model (GLM) to explore the main effects of number of SAD set, fruit number, rater, and all two-way interactions: $Y_{ijklm} = \vartheta + \alpha_i + \beta_j + \gamma_j + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk}$, where ϑ is the intercept term; α_i is the effect of the ith SAD set; β_j is the effect of the jth fruit; γ_k is the effect of the kth rater; $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, and $(\beta\gamma)_{jk}$ are the interaction terms; and ε_{ijk} is the residual error term.

Table 1. General linear model (GLM) analysis of the effect of standard area diagrams (SADs) (zero or SADs with five or six diagrams), raters ($n = 15$), and actual severity ($n = 40$) on estimates of Asiatic citrus canker on unripe (green) Baia sweet orange[z]

Effect	DF	F value	P value
Number of SADs	2	256.46	<0.0001
Fruit number	39	142.14	<0.0001
Rater	14	3.26	<0.0001
Number of SADs × Fruit number	78	4.87	<0.0001
Number SADs × rater	28	11.69	<0.0001
Fruit number × rater	546	1.04	0.3

[z] GLM model F value (P value) = 10.4 (<0.0001). Error degrees of freedom = 1,092.

Using the visual estimates and comparing them to the actual values measured by image analysis, Lin's concordance correlation (LCC; 24,26,30) analysis was applied to evaluate the degree to which the pairs of observations for each fruit fell on the concordance line of 45° (slope = 1, intercept = 0). If there is perfect concordance between the estimates and the actual values, then the LCC statistics of systematic bias, $\upsilon = 1$; location bias, $\mu = 0$; accuracy, $C_b = 1$; precision, $r = 1$; and agreement, $\rho_c = 1$. Any deviation from these values indicates bias or loss of accuracy, precision, or agreement. LCC was calculated for estimates made by each rater without an SAD set, and for estimates made using the five- or six-diagram SAD set.

Inter-rater reliability was measured using the coefficient of determinations (R^2) for each pairwise combination of rater based on estimates without SAD set and for the estimates made using the

Table 2. Bias, accuracy, precision, and agreement measures using Lin's concordance correlation (LCC) for estimates of severity of Asiatic citrus canker on unripe (green) Baia sweet orange by 15 raters without use of standard area diagram (SAD) sets as assessment aides (None), or using SADs with five or six diagrams depicting different severities of disease

LCC	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6	Rater 7	Rater 8	Rater 9	Rater 10	Rater 11	Rater 12	Rater 13	Rater 14	Rater 15
None															
υ[v]	0.649	0.276	0.419	0.560	0.451	0.822	1.100	0.667	0.697	0.791	0.968	0.836	0.861	0.610	0.717
μ[w]	–0.840	–1.989	–1.561	–1.076	–1.274	–0.464	–0.213	–0.651	–0.295	–0.450	0.012	–0.559	–0.578	–0.661	–0.674
C_b[x]	0.691	0.255	0.382	0.571	0.466	0.887	0.974	0.772	0.901	0.886	0.999	0.853	0.849	0.744	0.779
r[y]	0.891	0.864	0.783	0.729	0.889	0.936	0.806	0.869	0.712	0.818	0.914	0.895	0.743	0.538	0.522
ρ_c[z]	0.615	0.220	0.299	0.416	0.414	0.830	0.785	0.671	0.642	0.725	0.913	0.764	0.630	0.400	0.407
Five															
υ[v]	1.197	1.252	0.931	0.990	1.059	0.908	0.931	0.856	0.942	0.918	0.937	0.900	1.032	0.975	0.945
μ[w]	0.103	0.281	–0.039	0.156	0.127	–0.153	–0.112	–0.232	–0.355	–0.191	–0.042	–0.112	–0.072	–0.120	–0.099
C_b[x]	0.979	0.939	0.997	0.988	0.990	0.984	0.991	0.962	0.939	0.979	0.997	0.988	0.997	0.992	0.994
r[y]	0.861	0.867	0.836	0.850	0.890	0.914	0.958	0.957	0.876	0.901	0.905	0.887	0.958	0.960	0.950
ρ_c[z]	0.843	0.814	0.833	0.839	0.882	0.899	0.950	0.921	0.823	0.882	0.903	0.876	0.955	0.953	0.944
Six															
υ[v]	0.930	0.876	0.947	0.980	0.856	0.825	0.888	0.831	0.915	0.998	0.836	0.865	0.849	0.892	0.887
μ[w]	–0.040	–0.098	–0.012	–0.010	–0.051	–0.176	–0.157	–0.256	–0.282	–0.053	–0.154	–0.138	–0.111	–0.169	–0.089
C_b[x]	0.997	0.987	0.998	1.000	0.987	0.967	0.981	0.952	0.958	0.999	0.973	0.980	0.981	0.980	0.989
r[y]	0.852	0.828	0.852	0.836	0.823	0.866	0.943	0.940	0.891	0.886	0.825	0.901	0.907	0.932	0.901
ρ_c[z]	0.849	0.817	0.851	0.835	0.812	0.837	0.925	0.896	0.854	0.884	0.802	0.883	0.889	0.913	0.891

[v] Scale bias, or slope shift (υ, 1 = no bias relative to the concordance line).

[w] Location bias, or height shift (μ, 0 = no bias relative to the concordance line).

[x] Correction factor (C_b) measures how far the best-fit line deviates from 45° and, thus, is a measure of accuracy.

[y] Correlation coefficient (r) measures precision.

[z] Lin's concordance correlation coefficient (ρ_c) combines both measures of precision (r) and accuracy (C_b) to measure the degree of agreement with the true value.

Table 3. Mean concordance statistics (bias, accuracy, precision, and agreement) with bootstrap analysis of the differences between means when estimating severity of Asiatic citrus canker on unripe (green) Baia sweet orange without or with a five-diagram standard area diagram (SAD) set[r]

	Mean			
LCC[s]	No SAD	Five-SAD set	Mean difference[t]	95% CIs (upper and lower)[u]
υ[v]	0.318	0.087	**–0.230**	**–0.319 to –0.144**
μ[w]	0.753	0.146	**–0.604**	**–0.871 to –0.358**
C_b[x]	0.734	0.981	**–0.246**	**–0.364 to –0.148**
r[y]	0.794	0.905	**–0.111**	**–0.195 to –0.044**
ρ_c[z]	0.582	0.888	**–0.305**	**–0.400 to –0.206**

[r] Bold text indicates a significant difference.

[s] Lin's concordance correlation (LCC) statistic.

[t] Mean of the difference between each rating.

[u] Confidence intervals (CIs) were based on 2,000 bootstrap samples. If the CIs embrace zero, the difference is not significant ($\alpha = 0.05$).

[v] Scale bias (systematic bias or slope shift, υ, 1 = no bias relative to the concordance line) can be ≤1≥; therefore, it was necessary to obtain standardized (as $1 - \upsilon$) absolute data prior to calculating the mean difference.

[w] Location bias (constant bias or height shift, μ, 0 = no bias relative to the concordance line) can be ≤0≥; therefore, it was necessary to obtain absolute data prior to calculating the mean difference.

[x] Correction factor (C_b) measures how far the best-fit line deviates from 45° and, thus, is a measure of accuracy.

[y] Correlation coefficient (r) measures precision.

[z] LCC coefficient (ρ_c) combines both measures of precision (r) and accuracy (C_b) to measure the degree of agreement with the true value.

five- or six-diagram SAD set. The R^2 is the proportion of the variation explained by the association between two sets of measurements, and indicates how closely one measurement predicts the other. A different measure of inter-rater reliability, the intraclass correlation coefficient (ICC, ρ) was also calculated. The ICC is a comparison of between-subject and within-subject variance and, thus, accounts for chance correspondence of the variance between the two measurements. The ICC and its confidence limits was calculated for the estimates of ACC severity made without an SAD set and using the five- or six-diagram SAD set using an SAS macro (25).

For all LCC statistics (υ, μ, C_b, r, and ρ_c) and inter-rater reliability (R^2), the difference between the statistics for each rater from each pair of assessments was calculated and an equivalence test (3,41,42) was used to calculate 95% confidence intervals (CIs) by bootstrapping using the percentile method. With an equivalence test, the null hypothesis is the converse of H_0 (i.e., the null hypothesis is nonequivalence). In total, 2,000 balanced bootstrap samples were taken: if the CIs of the mean difference spanned zero then there was no significant difference between them.

To investigate the relationship between the change in rater ability from estimates made without SAD sets and those made using the five- or six-diagram SAD sets, and the change in rater ability from estimates made with the five-diagram SAD sets and those made using the six-diagram SAD sets, the difference between the two assessments (second assessment – first assessment) was regressed against the first assessment statistics. This calculation was made for all LCC statistics (υ, μ, C_b, r, and ρ_c) and inter-rater reliability (R^2). Because values of υ and μ can be ≤1≥ and ≤0≥, respectively, it was necessary to standardize these data. Systematic bias was standardized (as $1 - \upsilon$) using absolute data prior to calculating the mean difference, and constant bias was converted to absolute values prior to calculating the mean difference of the second assessment – first assessment. The absolute error (visual estimate

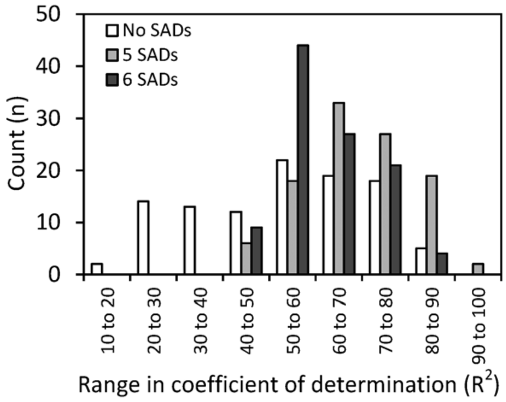

Fig. 3. Frequency of the inter-rater reliability of 15 raters measured by the coefficient of determination (R^2) without and with use of a standard area diagram set with five or six diagrams as an aid for assessment of Asiatic citrus canker on 40 images of unripe (green) Baia sweet orange.

Table 4. Mean concordance statistics (bias, accuracy, precision, and agreement) with bootstrap analysis of the differences between means when estimating severity of Asiatic citrus canker on unripe (green) Baia sweet orange without or with a six-diagram standard area diagram (SAD) set[r]

	Mean			
LCC[s]	**No SAD**	**Six-SAD set**	**Mean difference**[t]	**95% CIs (upper and lower)**[u]
υ[v]	0.318	0.108	**–0.208**	**–0.321 to –0.107**
μ[w]	0.753	0.120	**–0.630**	**–0.931 to –0.360**
C_b[x]	0.734	0.982	**–0.246**	**–0.369 to –0.144**
r[y]	0.794	0.879	**–0.086**	**–0.167 to –0.017**
ρ_c[z]	0.582	0.863	**–0.280**	**–0.378 to –0.177**

[r] Bold text indicates a significant difference.

[s] Lin's concordance correlation (LCC) statistic.

[t] Mean of the difference between each rating.

[u] Confidence intervals (CIs) were based on 2,000 bootstrap samples. If the CIs embrace zero, the difference is not significant ($\alpha = 0.05$).

[v] Scale bias (systematic bias or slope shift, υ, 1 = no bias relative to the concordance line) can be ≤1≥; therefore, it was necessary to obtain standardized (as $1 - \upsilon$) absolute data prior to calculating the mean difference.

[w] Location bias (constant bias or height shift, μ, 0 = no bias relative to the concordance line) can be ≤0≥; therefore, it was necessary to obtain absolute data prior to calculating the mean difference.

[x] Correction factor (C_b) measures how far the best-fit line deviates from 45° and, thus, is a measure of accuracy.

[y] Correlation coefficient (r) measures precision.

[z] LCC coefficient (ρ_c) combines both measures of precision (r) and accuracy (C_b) to measure the degree of agreement with the true value.

Table 5. Mean concordance statistics (bias, accuracy, precision and agreement) with bootstrap analysis of the differences between means when estimating severity of Asiatic citrus canker on unripe (green) Baia sweet orange using a five- or six-diagram standard area diagram (SAD) set[r]

	Mean			
LCC[s]	**Five-SAD set**	**Six-SAD set**	**Mean difference**[t]	**95% CIs (upper and lower)**[u]
υ[v]	0.087	0.108	0.022	–0.019 to 0.057
μ[w]	0.146	0.120	–0.026	–0.067 to 0.015
C_b[x]	0.981	0.982	–0.001	–0.010 to 0.008
r[y]	0.905	0.879	**0.026**	**0.012 to 0.040**
ρ_c[z]	0.888	0.863	**0.025**	**0.007 to 0.044**

[r] Bold text indicates a significant difference.

[s] Lin's concordance correlation (LCC) statistic.

[t] Mean of the difference between each rating.

[u] Confidence intervals (CIs) were based on 2,000 bootstrap samples. If the CIs embrace zero, the difference is not significant ($\alpha = 0.05$).

[v] Scale bias (systematic bias or slope shift, υ, 1 = no bias relative to the concordance line) can be ≤1≥; therefore, it was necessary to obtain standardized (as $1 - \upsilon$) absolute data prior to calculating the mean difference.

[w] Location bias (constant bias or height shift, μ, 0 = no bias relative to the concordance line) can be ≤0≥; therefore, it was necessary to obtain absolute data prior to calculating the mean difference.

[x] Correction factor (C_b) measures how far the best-fit line deviates from 45° and, thus, is a measure of accuracy.

[y] Correlation coefficient (r) measures precision.

[z] LCC coefficient (ρ_c) combines both measures of precision (r) and accuracy (C_b) to measure the degree of agreement with the true value.

made with or without SAD sets – actual disease severity) was calculated for all estimates.

Results

The GLM analysis showed that there were significant main effects of number of SAD sets, fruit number, rater, and interactions of number of SAD sets–fruit number, and number of SAD sets–rater (Table 1). LCC analysis of these data demonstrated a range in bias (υ and μ), accuracy (C_b), precision (r), and agreement (ρ_c) among raters whether rating without SAD sets or when using five- or six-diagram SAD sets (Table 2). Overall, the range in agreement was 0.220 to 0.913 when not using SAD sets, 0.814 to 0.955 when using five-diagram SAD sets, and 0.863 to 0.925 when using six-diagram SAD sets (measures of bias, accuracy, and precision followed a similar pattern, suggesting an improvement with use of five- or six-diagram SAD sets).

The bootstrap analysis comparing assessments done with no SAD sets to those done using the five-diagram SAD set showed that all of the parameters of LCC (υ, μ, C_b, r, and ρ_c) were significantly improved with the use of the five-diagram SAD set (Table 3). Similarly, the bootstrap analysis comparing assessments done with no SAD sets with those done using the six-diagram SAD set also showed that all of the parameters of LCC were significantly improved with the use of the six-diagram SAD set as an aid for estimating ACC severity (Table 4). The comparison of the five- and six-diagram SAD sets as aids for assessment showed that the six-

Table 6. Inter-rater reliability of the estimates of Asiatic citrus canker severity made by 15 raters of 40 images of unripe (green) Baia sweet orange displaying a range in severity of disease either without or with use of a standard area diagram (SAD) sets with five or six diagrams as an aid for assessment[u]

			***F* value (*P* value)**				
Variable tested	**Means (R^2)[v]**	**Mean difference[w]**	**Fruit number**	**Rater number**	**CV[x]**	**ICC[y]**	**95% CIs[z]**
R^2							
No SADs × five SADs	0.54, 0.69	**–0.152**	…	…	…	…	**–0.202 to –0.107**
No SADs × six SADs	0.54, 0.62	**–0.083**	…	…	…	…	**–0.128 to –0.041**
Five SADs × six SADs	0.69, 0.62	**0.069**	…	…	…	…	**0.046 to 0.091**
ICC (ρ)							
No SADs	…	…	29.3 (<0.0001)	18.6 (<0.0001)	51.9	0.93 a	0.92 to 0.94
Five SADs	…	…	66.4 (<0.0001)	6.1 (<0.0001)	32.2	0.96 b	0.95 to 0.97
Six SADs	…	…	54.5 (<0.0001)	1.3 (0.2)	33.0	0.95 ab	0.94 to 0.96

[u] Inter-rater reliability was measured using the coefficient of determination (R^2) and the intraclass correlation coefficient (ICC; ρ). Numbers with the same letters are not significantly different at $P = 0.05$.

[v] R^2 is the proportion of the variation explained by the association between two sets of measurements.

[w] Mean of the difference between each rating.

[x] Coefficient of variation.

[y] ICC compares the between-subject variance with the within-subject variance and is the relative amount of variation from the combined mean of the two test sessions explained by differences between the subjects. The F values (P values) indicate the significance of the effect.

[z] Confidence intervals (CIs) were based on 2,000 bootstrap samples. If the CIs embrace zero, the difference is not significant ($\alpha = 0.05$).

Table 7. Regression solutions for the relationship between bias, precision, and agreement without the use of standard area diagrams (xSADs, where $x = 0$ or 5) assessment aides and the difference (xSADs – ySADs, where $y = 5$ or 6) for estimates of Asiatic citrus canker on unripe (green) Baia sweet orange

LCC, relationship[s]	**Intercept**	**Slope**	***F* value (*P* value)**	**CV[t]**	**R^2[u]**
υ[v]					
Five SADs – no SADs	0.95	1.10	72.1 (<0.0001)	12.6	0.85
Six SADs – no SADs	0.90	0.95	102.1 (<0.0001)	10.8	0.89
Six SADs – five SADs	1.00	–0.78	5.8 (0.03)	9.7	0.31
μ[w]					
Five SADs – no SADs	–0.15	0.99	441.5 (<0.0001)	–12.4	0.97
Six SADs – no SADs	–0.17	0.91	862.3 (<0.0001)	–8.9	0.99
Six SADs – five SADs	–0.08	–0.89	3.5 (0.09)	–264.1	0.21
C_b[x]					
Five SADs – no SADs	0.98	–1.01	1735.1 (<0.0001)	2.7	0.99
Six SADs – no SADs	0.97	–0.97	4077.8 (<0.0001)	1.8	0.99
Six SADs – five SADs	0.98	–0.73	13.6 (0.003)	1.4	0.51
r[y]					
Five SADs – no SADs	0.88	–0.83	161.3 (<0.0001)	4.6	0.93
Six SADs – no SADs	0.86	–0.82	184.9 (<0.0001)	4.3	0.93
Six SADs – five SADs	0.89	–0.55	2.2 (0.2)	4.6	0.15
ρ_c[z]					
Five SADs – no SADs	0.90	–1.02	220.1 (<0.0001)	8.8	0.94
Six SADs – no SADs	0.86	–0.99	384.6 (<0.0001)	6.7	0.97
Six SADs – five SADs	0.87	–0.85	9.9 (0.008)	4.4	0.43
R^2					
Five SADs – no SADs	0.64	0.66	511.7 (<0.0001)	13.8	0.83
Six SADs – no SADs	0.60	0.72	458.8 (<0.0001)	14.5	0.82
Six SADs – five SADs	0.65	0.60	68.8 (<0.0001)	12.8	0.40

[s] Lin's concordance correlation (LCC) statistic.

[t] Coefficient of variation (CV) is a unitless measure of variation, and is calculated as [(mean square error/mean) × 100].

[u] Coefficient of determination (R^2) is the proportion of the variation explained by the association between two sets of measurements.

[v] Scale bias (systematic bias or slope shift, υ, 1 = no bias relative to the concordance line) can be ≤1≥; therefore, it was necessary to obtain standardized (as 1 – υ) absolute data prior to calculating the mean difference.

[w] Location bias (constant bias or height shift, μ, 0 = no bias relative to the concordance line) can be ≤0≥; therefore, it was necessary to obtain absolute data prior to calculating the mean difference.

[x] Correction factor (C_b) measures how far the best-fit line deviates from 45° and, thus, is a measure of accuracy.

[y] Correlation coefficient (r) measures precision.

[z] LCC coefficient (ρ_c) combines both measures of precision (r) and accuracy (C_b) to measure the degree of agreement with the true value.

diagram SAD set provided less precision (r) or agreement (ρ_c) compared with the five-diagram SAD sets but there was no effect on systematic or constant bias (υ or μ, respectively) or accuracy (C_b) (Table 5).

The frequency of inter-rater reliability (R^2) improved with use of both five- and six-diagram SAD sets as aids for assessment of severity of ACC on unripe sweet orange (Fig. 3). The bootstrap analysis showed that use of both five- and six-diagram SAD sets resulted in a significant improvement in the inter-rater reliability (Table 6). However, as with the LCC measures of agreement, although there was a significant difference between the reliability of estimates made using five- or six-diagram SAD sets, it was the five-diagram SAD sets that provided estimates that were more reliable. The ICC indicated a significant difference in inter-rater reliability between estimates made without SAD sets using the five-diagram SAD sets, but the six-diagram SAD sets apparently provided no significant improvement in inter-rater reliability compared with using no SAD sets or using the five-diagram SAD sets (Table 6B).

The relative gain or loss among raters using the five- or six-diagram SAD set (regression of method 1 against method 2 – method 1) for each of the LCC statistics and for inter-rater reliability had a linear relationship in most cases (Fig. 4, with regression solutions shown in Table 7). The linear relationship suggests that those raters with the least good score improved the most, whereas those raters who already showed little bias (υ and μ) or had good accuracy (C_b), precision (r), agreement (ρ_c), or reliability (R^2) did not improve much (or, in a few cases, were less good) when using either five- or six-diagram SAD sets compared with using no SAD sets. Use of either SAD set resulted in reduction in systematic bias (υ; Fig. 4A; Table 7) and constant bias (μ; Fig. 4B; Table 7). For both υ and μ, markers to the left of the vertical dashed line (negative values) indicate reduction in bias; based on the R^2, there was no relationship between five- and six-diagram SAD sets (i.e., raters who had less bias with five-diagram SAD sets were not relatively or consistently more or less biased with six-diagram SAD sets). The regression analysis showed that accuracy (C_b), precision (r),

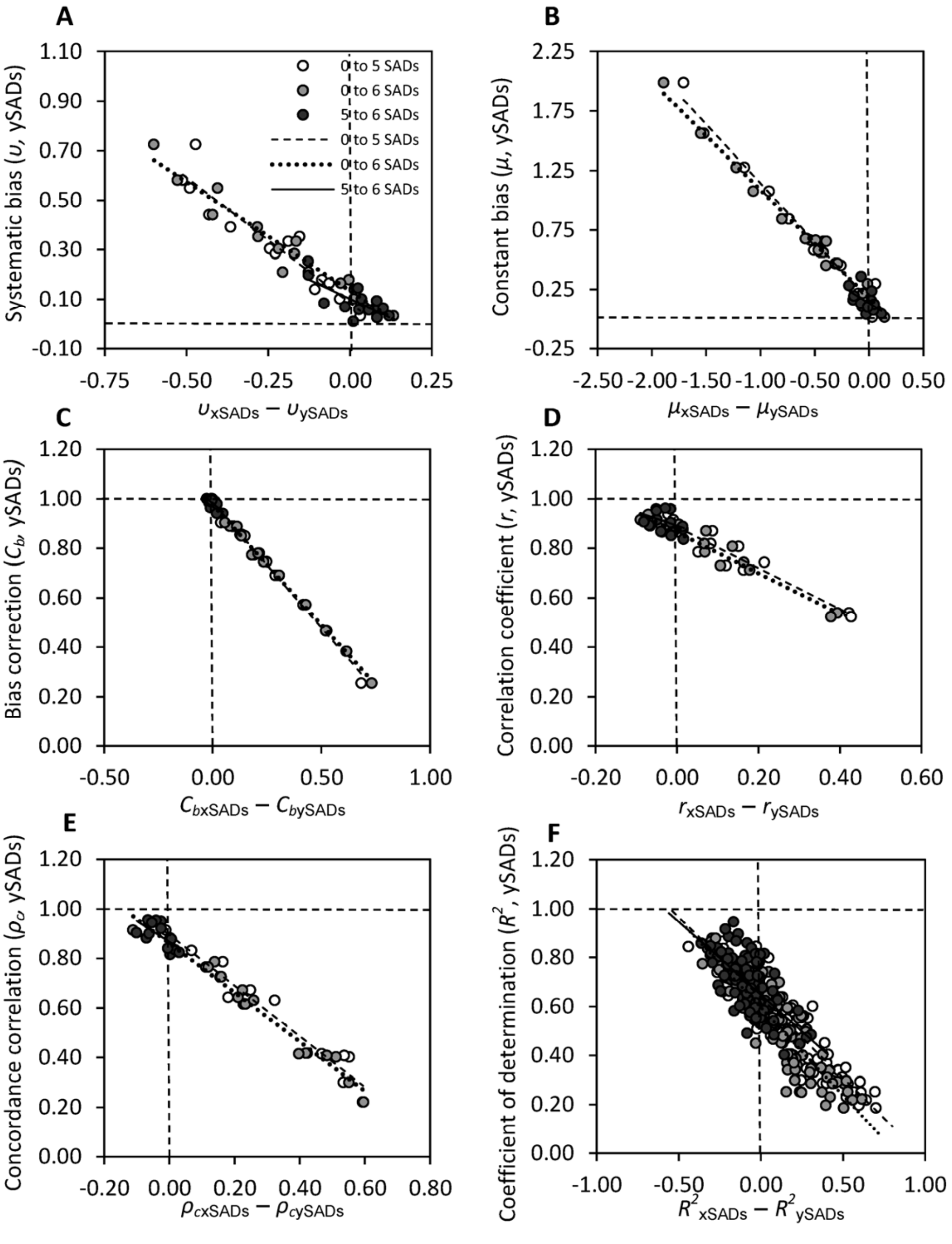

Fig. 4. Relationship between bias, precision, and agreement without the use of standard area diagram (xSAD, where x = 0 or 5) assessment aides and the difference (xSADs – ySADs, where y = 5 or 6), demonstrating that raters with the least good scores benefitted the most for all variables. **A,** Systematic bias; **B,** constant bias; **C,** bias correction factor; **D,** correlation coefficient; **E,** Lin's concordance correlation coefficient; and **F,** Inter-rater reliability. Disease was assessed on a set of 40 images of unripe (green) Baia sweet orange with symptoms of Asiatic citrus canker by 15 different raters. Regression solutions are given in Table 7.

agreement (ρ_c), and inter-rater reliability all improved when estimates of ACC severity were made using a five- or six-diagram SAD set relative to no SAD set, and the worst raters tended to improve the most (Fig. 4C, D, E, and F, respectively; Table 7); markers to the right of the vertical dashed line (positive values) indicate improvement. However, when using six- compared with five-diagram SAD sets, there was generally a weaker relationship and, most often, a slight loss in individual rater accuracy and reliability.

Absolute error was reduced using either five- or six-diagram SAD sets compared with error for estimates made without SAD sets, and tended to be more centered around zero (Fig. 5), with a greater proportion of estimates falling within ±10%. Without use of SAD sets, absolute error reflected the tendency of most of these raters to underestimate the severity of ACC on unripe orange fruit. However, use of either SAD set resulted in the estimates being more evenly balanced between over- and underestimates, with fewer extreme values.

Discussion

The results confirm that SAD sets can greatly improve the agreement of visual estimates of the actual severity of ACC on unripe, green fruit of sweet orange. Both the five- and six-diagram SAD sets improved the estimates compared with estimates made without using SAD sets. These results also confirm several previous studies that have demonstrated the utility of SAD sets in improving the accuracy and reliability of visual estimates of disease severity (2,5,17,28,32,37,38,41).

Estimates using the five-diagram SAD sets were slightly better than those using the six-diagram SAD sets. Although this may seem counterintuitive, there are possible explanations for the apparent paradox of less agreement despite using an SAD set comprising more diagrams. The difference in SAD structure (the individual severities presented) might have made the six-diagram SAD sets less useful for guiding raters who were interpolating estimated disease severities over the range experienced in this sample. Alternatively, having an extra diagram in the low severity range (≤9%) might have caused raters to overanalyze their estimate, resulting in an estimate with greater bias. Whatever the cause, the number of diagrams depicted in an SAD set might affect agreement (15) except, in that particular study, the raters did not interpolate severity but rated the disease based on the severity depicted in the SAD, limiting its comparison with estimating severity to the nearest percent.

The distribution of ACC on fruit is not uniform (8). However, sweet orange is an almost spherical fruit and, when viewed from the side, 50% of the surface area is visible. They are unlike apple fruit, which tend to have an upper and a lower surface which are less visible from a side view and, thus, require additional SAD sets to ensure that the severity of Sooty blotch and flyspeck is assessed appropriately (37). With ACC, although there is a tendency for more disease on the upper half of the fruit (8), a side view still provides very good coverage of 50% of the fruit surface. Thus, using a two-dimensional SAD as an aid to assess the two sides of a fruit can provide a more accurate assessment of ACC severity on the whole fruit compared with unaided assessment.

Most unaided raters underestimated ACC severity on unripe sweet orange fruit in this study. The study by Spolti et al. (37) using SAD sets to aid assessment of apple fruit disease also found that raters had a marked tendency to underestimate disease severity. This is in contrast to most visual estimates of foliar disease, where there is most often a pronounced tendency to overestimate, particularly at severities of 0 to 10% (11,36). The tendency to underestimate disease was resolved by use of both the five- and six-diagram SAD sets. We do not know what the cause of the consistent underestimation might have been but it could have something to do with estimating disease severity on a curved, three-dimensional surface like a citrus fruit.

The results showed that estimates using SAD sets improved the most for raters who had the least good accuracy or reliability without using SAD sets. This has been observed before (3,41), demonstrating the point that using SAD sets helps to ensure that estimates by different raters are all in closer agreement with the actual values and, therefore, with each other. Nonetheless, if assessments are made by multiple raters in an experiment, it is strongly recommended that the raters assess based on replicate to prevent any bias being attributed to treatment.

It is important to reassert that these results relate specifically to estimation of ACC on green (unripe) fruit of citrus. Unripe (often smaller, and always green) citrus fruit are more susceptible to ACC, and become increasingly resistant as they mature (40). Thus, it is important to have an aid to quantify ACC that develops prior to fruit maturity, which is the intended function for the SAD sets described in this study. Furthermore, the difference in symptoms on unripe and ripe fruit is a compelling reason to develop SAD sets to use for ripe citrus that are yellow to orange in color, and on which the canker lesions no longer have an apparent chlorotic halo.

In conclusion, use of either the five- or six-diagram SAD sets improves the accuracy and reliability of visual estimates of the

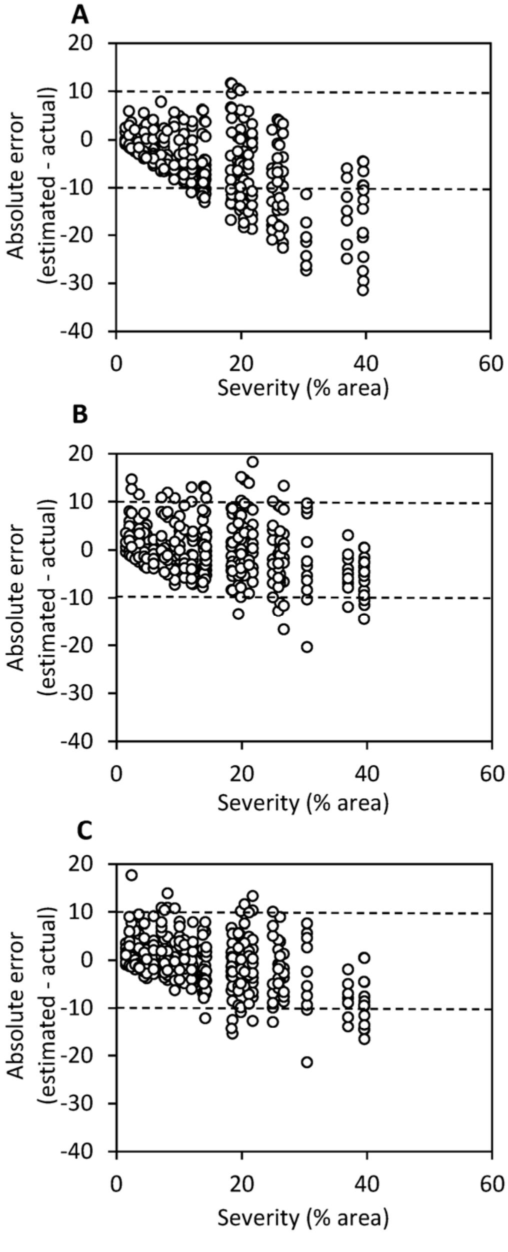

Fig. 5. Absolute error (estimate minus true disease) of assessments of Asiatic citrus canker on a set of 40 images of unripe (green) Baia sweet orange by 15 raters **A,** without use of standard area diagram (SAD) sets as assessment aides or using a SAD set with **B,** five or **C,** six diagrams depicting different severities of citrus canker. Dashed lines indicate ±10% absolute error limits.

severity of ACC on unripe citrus fruit, although the five-diagram SAD sets appeared to provide slightly better agreement and reliability compared with the six-diagram SAD sets. Based on this result, the five-diagram SAD sets is preferred. Although the SAD set was developed for sweet orange, it doubtless has applicability to other citrus, including grapefruit. These SAD sets should be useful for research and disease management, where estimates of the severity of ACC is needed on citrus fruit, including for field surveys, disease progress studies, spread of the disease, citrus genotype rating, and studies of disease control.

Acknowledgments

We acknowledge support of The National Council for Scientific and Technological Development (CNPq) for financial aid provided, and Coordination of Improvement of Higher Education Personnel (CAPES) for financial assistance granted to R. Braido for a scholarship.

Literature Cited

1. Amorin, L. 1995. Avaliação de doenças. Pages 647-671 in: Manual de Fitopatologia, v.1. A. Bergamin Filho, H. Kimati, and L. Amorim, eds. Editora Ceres, São Paulo, Brazil.
2. Amorin, L., and Bergamin Filho, A. 1996. Doenças de Plantas Tropicais: Epidemiologia e Controle Econômico. Editora Ceres, São Paulo, Brazil.
3. Bardsley, S. J., and Ngugi, H. K. 2012. Reliability and accuracy of visual methods used to quantify severity of foliar bacterial spot symptoms on peach and nectarine. Plant Pathol. 62:460-474.
4. Behlau, F., Belasque, J., Jr., Graham, J. H., and Leite, R. P., Jr. 2009. Effect of frequency of copper applications on control of citrus canker and the yield of young bearing sweet orange trees. Crop Prot. 29:300-305.
5. Belasque, J., Jr., Bassanezi, R. B., Spósito, M. B., Ribeiro, L. M., Jesus, W. C., Jr., and Amorim, L. 2005. Escalas diagramáticas para a avaliação da severidade do cancro cítrico. Fitopatol. Brasil. 30:387-393.
6. Belasque, J., Jr., Jaciani, F. J., Marin, D. R., and Barbosa, J. C. 2008. Tamanho da amostra para quantificação do tamanho de lesões do cancro cítrico. Trop. Plant Pathol. 33:317-322.
7. Bitancourt, A. A. 1957. O cancro citrico. Biologico 23:101-111.
8. Bock, C. H., Gottwald, T. R., and Parker, P. E. 2011.Distribution of canker lesions on the surface of diseased grapefruit. Plant Pathol. 60:986-991.
9. Bock, C. H., Graham, J. H., Gottwald, T. R., Cook, A. Z., and Parker, P. E. 2010. Wind speed and wind associated injury affect severity of citrus canker on Swingle citrumelo. Eur. J. Plant Pathol. 128:21-38.
10. Bock, C. H., Parker, P. E., Cook, A. Z., and Gottwald, T. R. 2008. Visual assessment and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. Plant Dis. 92:530-541.
11. Bock, C. H., Parker, P. E., Cook, A. Z., and Gottwald, T. R. 2008. Characteristics of the perception of different severity measures of citrus canker and the relations between the various symptom types. Plant Dis. 92:927-939.
12. Bock, C. H., Parker, P. E., and Gottwald, T. R. 2005. The effect of simulated wind-driven rain on duration and distance of dispersal of *Xanthomonas axonopodis* pv. *citri* from canker infected citrus trees. Plant Dis. 89:71-80.
13. Bock, C. H., Poole, G., Parker, P. E., and Gottwald, T. R. 2010. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. Crit. Rev. Plant Sci. 29:59-107.
14. Corkidi, G., Balderas-Ruíz, K. A., Taboada, B., Serrano-Carreón, L., and Galindo, E. 2006. Assessing mango anthracnose using a new three-dimensional image-analysis technique to quantify lesions on fruit. Plant Pathol. 55:250-257.
15. Corrêa, F. M., Bueno Filho, J. S. S., and Carmo, M. G. F. 2009. Comparison of three diagrammatic keys for the quantification of late blight in tomato leaves. Plant Pathol. 58:1128-1133.
16. Das, A. K. 2003. Citrus canker—a review. J. Appl. Hortic. 5:52-60.
17. Godoy, C. V., Koga, L. J., and Canteri, M. G. 2006. Diagrammatic scale for assessment of soybean rust severity. Fitopatol. Bras. 31:63-68.
18. Gottwald, T. R., Graham, J. H., Bock, C. H., Bonn, G., Civerolo, E., Irey, M., Leite, R., Lopez, M., McCollum, G., Parker, P., Ramallo, J., Riley, T., Schubert, T., Stein, B., and Taylor, E. 2009. The epidemiological significance of post-packinghouse survival of *Xanthomonas citri* ssp. *citri* for dissemination of Asiatic citrus canker via infected fruit. Crop Prot. 28:508-524.
19. Gottwald, T. R., Graham, J. H., and Schubert, T. S. 2002. Citrus Canker: The pathogen and its impact. Plant Health Progress. Online publication. doi:10.1094/PHP-2002-0812-01-RV. http://www.plantmanagementnetwork.org/pub/php/review/citruscanker/
20. IBGE. Fundação Instituto Brasileiro de Geografia e Estatística. Indicadores IBGE. Estatística da produção agrícola. Setembro de 2013. Disponível em: http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/lspa/estProdAgr_201309.pdf
21. James, W. C. 1971. An illustrated series of assessment keys for plant diseases, their preparation and usage. Can. Plant Dis. Surv. 51:39-65.
22. Kranz, J. 1977. A study on maximum severity in plant disease. Travaux de´die´s a` G. Viennot-Bourgin 16:9-73.
23. Leite, R. P., Jr., and Mohan, S. K. 1990. Integrated management of the citrus bacterial canker disease caused by *Xanthomonas campestris* pv. *citri* in the state of Paraná, Brazil. Crop Prot. 9:3-7.
24. Lin, L. I. 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45:255-268.
25. Lu, L., and Shara, N. 2007. Reliability analysis: calculate and compare intra-class correlation coefficients (ICC) in SAS. Statistics and Data Analysis. NeSug 2007. 3 pp. http://www.nesug.org/Proceedings/nesug07/sa/sa13.pdf
26. Madden, L. V., Hughes, G., and van den Bosch, F. 2007. The Study of Plant Disease Epidemics. American Phytopathological Society, St. Paul, MN.
27. Massari, C. A., and Belasque, J., Jr. 2006. A campanha de erradicação do cancro cítrico no Estado de São Paulo—Situação atual e contaminação em viveiros. Laranja 27:41-55.
28. Michereff, S. J., Noronha, M. A., Andrade, D. E. G. T., Oliveira, E. P., Xavier Filha, M. S., and Moreira, P. A. A. 2006. Development and validation of a diagrammatic key for *Cercospora* leaf spot of sweet pepper. Summa Phytopathol. 32:260-266.
29. Nilsson, H.-E. 1995. Remote sensing and image analysis in plant pathology. Annu. Rev. Phytopathol. 15:489-527.
30. Nita, M., Ellis, M. A., and Madden, L. V. 2003. Reliability and accuracy of visual estimation of Phomopsis leaf blight of strawberry. Phytopathology 93:995-1005.
31. Nutter, F. W., Jr., Gleason, M. L., Jenco, J. H., and Christians, N. L. 1993. Accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. Phytopathology 83:806-812.
32. Nutter, F. W., Jr., Miller, D. L., and Wegulo, S. N. 1998. Do standard diagrams improve the accuracy and precision of disease assessment? Page 2.1.16 in: 7th Int. Cong. Plant Pathol. Vol, 2. Edinburgh.
33. Nutter, F. W., Jr., Teng, P. S., and Shokes, F. M. 1991. Disease assessment terms and concepts. Plant Dis. 75:1187-1188.
34. Pruvost, O., Boher, B., Brocherieux, C., Nicole, M., and Chiroleu, F. 2002. Survival of *Xanthomonas axonopodis* pv. *citri* in leaf lesions under tropical environmental conditions and simulated splash dispersal of inoculum. Phytopathology 92:336-346.
35. Schubert, T., Shabbir, J., Rizvi, A., Sun, X., Gottwald, T. R., Graham, J. H., and Dixon W. N. 2001. Meeting the challenge of eradicating citrus canker in Florida—again. Plant Dis. 85:340-356.
36. Sherwood, R. T., Berg, C. C., Hoover, M. R., and Zeiders, K. E. 1983. Illusions in visual assessment of *Stagonospora* Leaf Spot of orchardgrass. Phytopathology 73:173-177.
37. Spolti, P., Schneider, L., Sanhueza, R. M. V., Batzer, J. C., Gleason, M. L., and Medeiros Del Ponte, E. 2011. Improving sooty blotch and flyspeck severity estimation on apple fruit with the aid of standard area diagrams. Eur. J. Plant Pathol. 129:21-29.
38. Spósito, M. B., Amorim, L., Belasque, J., Jr., Bassanezi, R. B., and Aquino, R. 2004. Elaboração e validação de escala diagramática para avaliação da severidade da mancha preta em frutos cítricos. Fitopatol. Brasil. 29:81-85.
39. Vale, F. X. R., Fernandes Filho, E. I., and Liberato J. R. 2003. QUANT—A software for plant disease severity assessment. Page 105 in: (Abstr. 8.18) 8th Int. Cong. Plant Pathol. Christchurch, New Zealand.
40. Verniere, C. J., Gottwald, T. R., and Pruvost, O. 2003. Disease development and symptom expression of *Xanthomonas axonopodis* pv. *citri* in various citrus plant tissues. Phytopathology 93:822-843.
41. Yadav, N. V. S., de Vos, S. M., Bock, C. H., and Wood, B. W. 2013. Development and validation of standard area diagrams to aide assessment of pecan scab symptoms on pecan fruit. Plant Pathol. 62:325-335.
42. Yi, Q., Wang, P. P., and He, Y. 2008. Reliability analysis for continuous measurements: Equivalence test for agreement. Stat. Med. 27:2816-2815.