Eur J Plant Pathol (2015) 141:327–337
DOI 10.1007/s10658-014-0544-0

A standard area diagram set to aid estimation of the severity of Asiatic citrus canker on ripe sweet orange fruit

Ricardo Braido · Aline M. O. Gonçalves-Zuliani · Paula T. R. Nocchi · José Belasque Junior · Vanderly Janeiro · Clive H. Bock · William M. C. Nunes

Accepted: 2 October 2014 / Published online: 11 October 2014

Abstract Asiatic citrus canker (ACC) is a major disease of citrus in many tropical and subtropical citrus-growing regions. Severe disease can develop on the fruit, making it unmarketable. Research activities need accurate and reliable methods to estimate the severity of citrus canker, most often done visually by raters. The objective was to validate and compare two standard area diagram sets (SADs) as aids for raters to improve the accuracy and reliability of visual estimates of canker severity on ripe fruit of sweet orange (unripe fruit are green, and have slightly different symptoms compared to ripe fruit, warranting different SADs). A 5-diagram SAD set (0.7, 2.0, 7.0, 21.0 and 39.0 %) and a 6-diagram SAD set (0.7, 2.0, 4.0, 10.0, 21.0 and 39.0 %) were compared; the 6-diagram set had an additional diagram in the most common severity range. Fifteen raters evaluated 40 images of cankered, unripe fruit. Both the 5- and 6-diagram SADs improved the accuracy and reliability of rater estimates. Agreement, measured by Lin's concordance correlation coefficient (ρ_c) was 0.735 to 0.906 when not using SADs, 0.931 to 0.985 when using 5-diagram SADs, and 0.879 to 0.979 when using 6-diagram SADs. Mean inter-rater reliability (R^2) was 0.638 when not using SADs, 0.889 when using 5-diagram SADs, and 0.834 when using 6-diagram SADs. The 5-diagram SADs resulted in significantly more accurate and reliable estimates compared with the 6-diagram set, and is thus recommended for use. The SADs should have application for estimating ACC severity on ripe fruit of other citrus, including grapefruit.

Keywords Phytopathometry · Diagrammatic scales · Field keys · Disease assessment

R. Braido · A. M. O. Gonçalves-Zuliani · P. T. R. Nocchi · W. M. C. Nunes (✉)
Núcleo de Pesquisa em Biotecnologia Aplicada, Universidade Estadual de Maringá, 87020-900 Maringá, Paraná, Brazil
e-mail: wmcnunes@uem.br

J. B. Junior
Departamento de Fitopatologia e Nematologia, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, São Paulo, Brazil
e-mail: belasque@usp.br

V. Janeiro
Departamento de Estatística, Universidade Estadual de Maringá, Maringá, Paraná, Brazil
e-mail: vjaneiro@uem.br

C. H. Bock
USDA-ARS-SEFTNRL, 21 Dunbar Road, Byron, GA 31008, USA
e-mail: clive.bock@ars.usda.gov

Introduction

The plant-pathogenic bacterium, *Xanthomonas citri* subsp. *citri* causes Asiatic citrus canker (ACC), which is one of the world's most important diseases of citrus (Das 2003; Gottwald et al. 2002; Schubert et al. 2001). Although the disease originated in Asia, it was first reported in Brazil in 1957 (Bitancourt 1957). Symptoms are typified by erumpent, necrotic lesions with variable yellow halos and are found on leaves, fruits and shoots

(Belasque et al. 2008; Schubert et al. 2001). Older lesions, particularly on fruit, lack the chlorotic halo.

In those regions where ACC causes epidemics, research into ACC management to minimize the impact of the disease requires estimation of disease severity (Behlau et al. 2009; Gottwald et al. 2009). Severity of a disease is often represented as a percentage of the area that is diseased on a plant organ (Amorim and Bergamin Filho 1996; Nutter et al. 1991). For many research activities aimed at studying ACC, accurate and reliable estimates of the severity of the disease are required. Accuracy in measurement science is defined as the closeness of the estimated values to the true values (Madden et al. 2007; Nutter et al. 1991), whereas reliability is the extent to which the same estimate obtained under different conditions yields similar results (Madden et al. 2007; Nutter et al. 1991).

Visual estimates are widely used to assess disease severity (Bock et al. 2010). To help minimize error in estimation (Bock et al. 2008a, b, 2010; Nutter et al. 1993; Parker et al. 1995), standard area diagrams (SADs) have been developed as an assessment aid (Amorim and Bergamin Filho 1996; James 1971; Nutter et al. 1998). The effectiveness of these aids is well documented on many crops (Belasque et al. 2005; Godoy et al. 2006; James 1971; Michereff et al. 2006; Spolti et al. 2011; Spósito et al. 2004; Yadav et al. 2013), including for both foliar symptoms of ACC (Belasque et al. 2005) and as an aid for assessing ACC severity on green, unripe fruit of sweet orange that show chlorotic halos (Braido et al. 2014). Ripe fruit of sweet orange (or other citrus) with canker lesions do not show a chlorotic zone as part of the symptom complex, but it should be included in estimation of severity on unripe fruit (Braido et al. 2014). Thus a SAD set specifically for ripe fruit of sweet orange is desirable.

The Braido et al. (2014) study validated the effectiveness of SADs for unripe fruit in improving estimates of ACC severity by different raters, with significant improvements demonstrated. In that study a comparison was made of two different SAD structures, 1 with 5 diagrams, and the second with 6 diagrams. They had the same range of severity (0.5 to 40 %), but the inter-SAD ranges were different in the 2 sets (and the 1 with 6 diagrams had an additional diagram in the most common severity range); interestingly, the 5-diagram SAD set was slightly better at aiding more accurate and reliable estimates compared to the 6-diagram SAD set. The reason for this is unknown, but might be due to the structure of the two SADs, or the disease severity range experienced on those particular fruit. A previous study has suggested that the number of SADs affects rater accuracy and reliability, at least when used for categorizing disease severity (Corrêa et al. 2009). Considering the importance of accurate estimates of disease severity, knowing and understanding how the number of SADs affects rater ability estimating disease to the nearest percent (as opposed to placing in categories) is valuable for improving accuracy, but is very likely sensitive to the range of actual disease severity being estimated, and that depicted in a particular SAD set. Thus, these observations on effects of SAD number also warrant further research to establish whether a 5- or 6-diagram SAD set is a consistently better aid for estimating severity of ACC.

The objectives of this research were to: 1) develop and validate SAD sets as assessment aids for rating ACC severity on yellow, ripe fruit of sweet orange, and 2) to compare two different SADs (with 5 and 6 diagrams, respectively) to confirm and further explore any effect of, or advantage to having an additional diagram in the range of the most frequently encountered severities.

Material and methods

Development of SADs

One hundred ripe fruit (6 to 8 cm in diameter) of sweet orange (*C. sinensis*) cultivar 'Baia' showing symptoms of ACC were collected arbitrarily from trees in a citrus orchard located at Iguatemi Experimental Farm (State University of Maringá, Paraná State, Brazil; latitude 23°25' S, longitude 52°10 W, elevation 555 m). The orchard was planted in April 2003 with trees spaced 3.5×5.0 m. Crop management practices were as for commercial citrus production in Paraná state, except canker control was not practiced and no copper-based products were applied.

The symptoms that developed were typical of ACC on mature sweet orange fruit, with necrotic, erumpent, corky lesions, the edge of which ends abruptly at the healthy, ripe yellow-orange fruit rind interface (in contrast to symptoms on green, unripe citrus (Braido et al. 2014)). The surface of one side of each of the 100 fruit was photographed (Sony CyberShot® 5,1MP), and the actual symptomatic area on that surface of the fruit in the image was analyzed using the image analysis software Quant V1.0.2 (Vale et al. 2003). The side of each fruit selected had the most severe symptoms of ACC, and was illuminated with a 40 W lamp placed at 15 cm distance to ensure uniform light conditions. The percentage diseased area was subsequently calculated ([area with canker symptoms ÷total fruit area]×100). The frequency of severities measured in different severity ranges shows that the majority were in the range 0-8 % area cankered (Fig. 1).

Based on the range of disease severity (Kranz 1977) two separate SADs were constructed: i) a SAD set with five diagrams, and ii) a SAD set with six diagrams. 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 (Braido et al. 2014). The 6-diagram SADs had an extra diagram included in the lower severity range (in the range of severity observed for the majority of the fruit, with the 7 % severity diagram (Fig. 2a) in the 5-diagram SAD set being switched for a 4 and a 10 % severity diagramin the 6-diagram set (Fig. 2b)). For both sets, the upper and lower limits (0.7 and 39 %) were based on the results of the actual minimum and maximum severity on fruit collected from the field. Thus the values for the 5-diagram SAD set were 0.7, 2.0, 7.0, 21.0 and 39.0 % canker severity (Fig. 2a), and the for 6-diagram SAD set were 0.7, 2.0, 4.0, 10.0, 21.0 and 39.0 % canker severity (Fig. 2b). The diagrams were constructed from images of fruit harvested from the same sweet orange trees as described above, and were prepared using Quant. V1.0.2 (Vale et al. 2003).

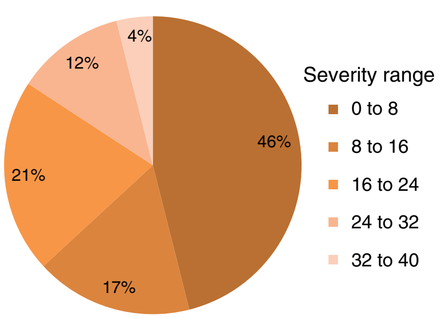

Fig. 1 The proportion of ripe sweet orange fruit (total 100 fruit) with different severities (% area diseased) of Asiatic citrus canker as measured using image analysis. Fruit were collected from an orchard in Paraná state, Brazil

Validation of SADs

As described previously (Braido et al. 2014), the SADs were validated by 15 raters (who had a range of experience of disease assessment and of familiarity with ACC symptoms), who estimated the canker severity on a subset of 40 images using a PowerPoint slide presentation, with each fruit image shown at random. All raters received the same instructions. The symptoms of ACC were described so they were readily recognizable, and the use of the assessment aids, the SADs, was described. First, the raters assessed the canker severity on each fruit at random without the SADs, repeating the process a second time with the aid of the 5-diagram SADs after a 30-min break, and finally a third time with the aid of the 6-diagram SADs after a further 30-min break. In practice when assessing disease in actual experiments, raters would assess both faces of a fruit, taking care to ensure that each side was fully represented (i.e., no overlap). 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 (Braido et al. 2014)).

Data analysis

Data analysis was as described previously (Braido et al. 2014), and all analyses were calculated in SAS V9.3 (SAS Systems, Cary, NC, USA) or MS Excel (Microsoft Corp, Redmond, WA, USA). A general linear model (GLM) was used to explore main effects of number of SADs used, fruit and rater and all two-way interactions: $Y_{ijklm} = \vartheta + \alpha_i + \beta_j + \gamma_k + (a\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk}$, where *nnn* is the intercept term, α_i is the effect of the i^{th} SAD set, β_j is the effect of the j^{th} fruit, γ_k is the effect of the k^{th} rater, $(a\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$ are the interaction terms, and ε_{ijk} is the residual error term. Subsequently the accuracy (Madden et al. 2007) and inter-rater reliability (Madden et al. 2007) were compared without the use of SADs and with the use of 5- or 6-diagram SADs. Lin's concordance correlation (LCC) (Lin 1989; Madden et al. 2007; Nita et al. 2003) was used to evaluate agreement between the pairs of observations (visual estimates and the actual image analysis-measured values). If the estimates and actual values are identical, the LCC statistics of systematic bias, υ =1 (also called scale shift or slope shift), constant bias, μ =0 (also called location bias or height shift), accuracy, C_b =1, precision, r =1, and agreement, ρ_c =1. Any deviation from these

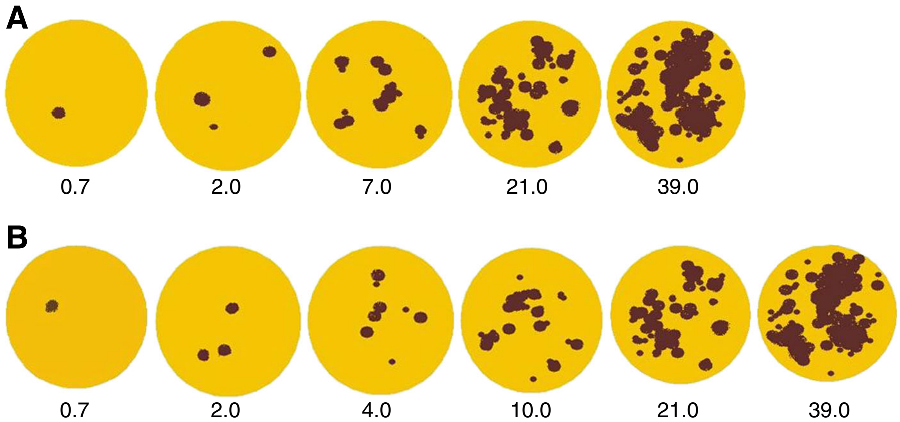

Fig. 2 The standard area diagrams with the same severity range but five **a** or six **b** severities (area showing symptoms in percent) depicted and used by 15 raters as aids to estimate Asiatic citrus canker severity

values indicates loss of accuracy, agreement or precision.

Inter-rater reliability was measured using the coefficient of determinations (R^2) for each pairwise combination of rater based on estimates without SADs, and for the estimates made using the 5- or 6-diagram SADs. Inter-rater reliability was also measured using the intra-class correlation coefficient (ICC, ρ) without SADs and using the 5- or 6-diagram SADs using a SAS macro (Lu and Shara 2007).

A test of equivalence (difference between the means) between treatments (no SADs, 5 SADs and 6 SADs) for each statistic (υ, μ, C_b, r, ρ_c and R^2) was calculated (Bardsley and Ngugi 2012; Yadav et al. 2013; Yi et al. 2008). The 95 % confidence intervals (CIs) were calculated based on 2,000 bootstraps using the percentile method. If the CIs of the mean difference spanned zero there was no significant difference between them.

To explore the gain (or loss) in rater ability when using five or six SADs, the difference between the two assessments (2nd assessment – 1st assessment) was regressed against the 1st assessment statistics. This calculation was made for all LCC statistics (υ, μ, C_b, r, ρ_c) and for 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-υ) 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 2nd assessment – 1st assessment. The absolute error (visual estimate made with or without SADs – actual disease severity) was calculated for all estimates.

Results

There were significant main effects of number of SADs, fruit, rater, and all interactions (Table 1). The mean severity for actual data was 7.92 % (st dev =8.25). A Tukey's means separation (α =0.05) of the effect of SADs showed significant overestimation without SADs (9.64 %), compared with the mean severity estimated using 5 (7.61 %) or 6 (7.29 %) SADs. Individual raters had a range in bias (υ, μ), accuracy (C_b), precision (r) and agreement (ρ_c) without SADs, or when using 5 or 6-diagram SADs. Overall, the range in agreement was 0.735 to 0.906 when not using SADs, 0.931 to 0.985 when using 5-diagram SADs, and 0.879 to 0.979 when using 6-diagram SADs (measures of bias, accuracy and precision followed a similar

Table 1 General linear model (GLM[a]) analysis of the effect of standard area diagrams SADs (0, or SADs with 5 or 6 diagrams), raters, fruit, and interactions on estimates of the severity of Asiatic citrus canker on ripe fruit of sweet orange

Effect	DF	*F*-value	*P*-value
Number of SADs	2	131.4	<0.0001
Fruit	39	322.6	<0.0001
Rater	14	12.4	<0.0001
No. SADs × Fruit	78	2.4	<0.0001
No. SADs × Rater	28	9.4	<0.0001
Fruit × Rater	546	1.2	0.003

[a] GLM model F-value (*P*-value) =20.0 (<0.0001). Error degrees of freedom =1092

pattern showing improvement when using 5 or 6-diagram SADs compared to without SADs).

Compared to estimates of severity without SADs, those done using a 5-diagram SAD set showed that, with the exception of systematic bias (υ), all other parameters of LCC (constant bias (μ), accuracy (C_b), precision (r) and agreement (ρ_c)) improved significantly (Table 2a). Similarly, all parameters (with the exception of systematic bias (υ)) showed improvement with the 6-diagram SADs (Table 2b). The comparison of the 5 and 6-diagram SADs (Table 2c) showed neither systematic bias (υ) nor constant bias (μ) were affected, but accuracy (C_b), precision (r) and agreement (ρ_c) were all significantly better using the 5 SAD set.

There was little relationship between prior ability and magnitude of gain or loss in systematic bias (υ) using either 5- or 6-diagram SADs (Fig. 3a and Table 3). However, with constant bias (μ), prior ability and magnitude of gain or loss were described by a linear function using either 5- or 6-diagram SADs (Fig. 3b, markers to the *left* of the vertical dashed line – negative values - indicate a reduction in constant bias), but there was no relationship between 5- and 6-diagram SADs (i.e., raters who had less bias with 5-diagram SADs were not relatively or consistently more or less biased with 6-diagram

Table 2 Mean concordance statistics (bias, accuracy, precision and agreement) with bootstrap analysis of the differences between means when estimating severity of Asiatic citrus canker on ripe fruit of sweet orange **a** without or with a 5-diagram SAD set, **b** without or with a 6-diagram SAD set, and **c** using a 5- or a 6-diagram SAD set

Comparison	LCC statistic	Mean		Mean diff[a]	95 % CIs[b] (upper and lower)
		No SAD	5-SAD set		
A. Without or with a 5-diagram SAD set	υ^c	0.093	0.071	−0.021	−0.067 to 0.019
	μ^d	0.313	0.061	−0.251[h]	−0.315 to −0.185
	$C_b{}^e$	0.939	0.993	−0.054	−0.069 to −0.039
	r^f	0.873	0.968	−0.095	−0.119 to −0.074
	$\rho_c{}^g$	0.819	0.961	−0.142	−0.163 to −0.122
B. Without or with a 6-diagram SAD set					
	υ	0.093	0.101	0.009	−0.031 to 0.052
	μ	0.313	0.096	−0.216	−0.280 to −0.152
	C_b	0.939	0.984	−0.044	−0.061 to 0.029
	r	0.873	0.954	−0.081	−0.105 to −0.061
	ρ_c	0.819	0.938	−0.119	−0.142 to −0.097
C. Using a 5- or a 6-diagram SAD set.	υ	0.071	0.101	0.031	−0.005 to 0.065
	μ	0.061	0.096	0.035	−0.007 to 0.083
	C_b	0.993	0.984	0.009	0.001 to 0.019
	r	0.968	0.954	0.014	0.007 to 0.022
	ρ_c	0.961	0.938	0.023	0.011 to 0.036

[a] Mean of the difference between each rating

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

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

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

[e] The correction factor (C_b) measures how far the best-fit line deviates from 45° and is thus a measure of accuracy

[f] The correlation coefficient (r) measures precision

[g] 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

[h] Bold text indicates a significant difference

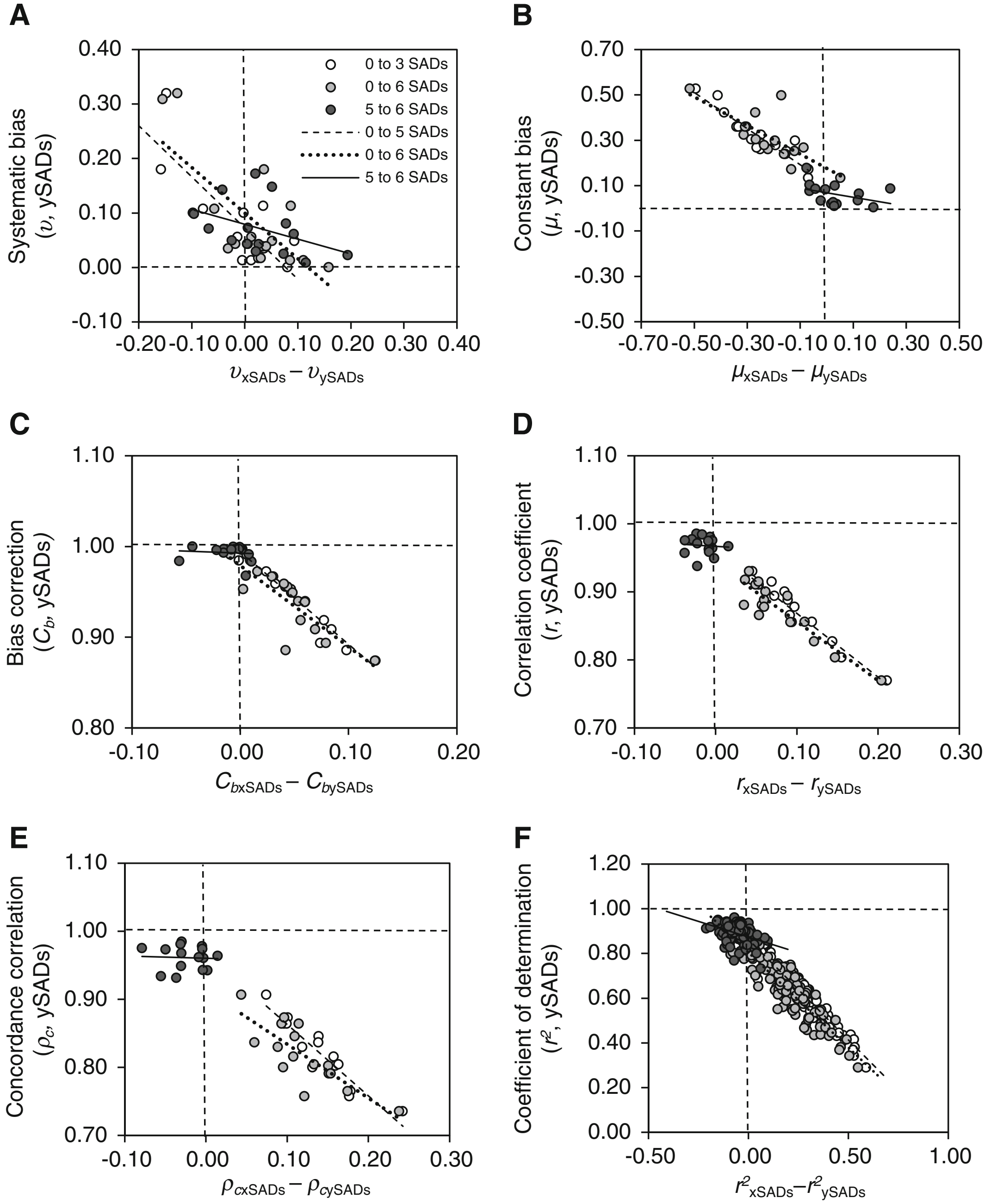

Fig. 3 The relationship between bias, precision and agreement without the use of standard area diagram (xSADs, where x =0 or 5, see axis titles) assessment aides and the difference (xSADs – ySADs, where y =5 or 6, see axis titles) demonstrating rater gain or loss for **a** systematic bias, **b** constant bias, **c** the bias correction factor, **d** precision (correlation coefficient, **e** agreement (Lin's concordance correlation coefficient), and **f** inter-rater reliability. Disease was assessed on a set of 40 images of ripe sweet orange fruit with Asiatic citrus canker by 15 different raters. Regression solutions are given in Table 3

SADs). The regression analysis also indicated that accuracy (C_b), precision (r) and agreement (ρ_c) were all improved estimating severity using 5- or 6-diagram SADs relative to no SADs, and those raters who were less accurate or precise improved the most (Fig. 3c, d and e, respectively, markers to the *right* of the vertical dashed line – positive values - indicate improvement), but when using 6- compared with 5-diagram SADs,

Table 3 The regression solutions for the relationship between bias, precision and agreement without the use of standard area diagrams (x SADs, where x =0 or 5) assessment aides and the difference (x SADs – y SADs, where y =5 or 6). See Fig. 3

LCC statistic	Relationship	Intercept	Slope	F-value (P-value)	CV[a]	R^{2b}
υ^c	5 SADs - no SADs	1.01	0.66	3 (0.09)	13.0	0.20
	6 SADs – no SADs	0.99	0.82	5 (0.05)	12.4	0.27
	6 SADs – 5 SADs	0.93	0.19	1 (0.4)	6.7	0.05
μ^d	5 SADs - no SADs	−0.15	−1.38	9 (0.01)	112.8	0.41
	6 SADs – no SADs	0.06	−0.62	1 (0.3)	138.6	0.09
	6 SADs – 5 SADs	−0.05	0.10	1 (0.6)	−166.4	0.02
C_b^e	5 SADs - no SADs	0.99	−1.02	207 (<0.0001)	1.0	0.94
	6 SADs – no SADs	0.98	−0.90	30 (0.0001)	2.2	0.69
	6 SADs – 5 SADs	0.99	−0.04	<1 (0.8)	0.9	0.01
r^f	5 SADs - no SADs	0.96	−0.94	143 (<0.0001)	1.5	0.92
	6 SADs – no SADs	0.94	−0.86	69 (<0.0001)	2.1	0.84
	6 SADs – 5 SADs	0.97	−0.07	<1 (0.8)	1.4	0.01
ρ_c^g	5 SADs - no SADs	0.97	−1.05	82 (<0.0001)	2.2	0.87
	6 SADs – no SADs	0.91	−0.79	23 (0.0003)	3.6	0.64
	6 SADs – 5 SADs	0.96	−0.03	<1 (0.9)	1.9	0.01
R^2	5 SADs - no SADs	0.86	−0.90	927 (<0.0001)	6.2	0.90
	6 SADs – no SADs	0.81	−0.85	475 (<0.0001)	8.3	0.82
	6 SADs – 5 SADs	0.87	−0.28	18 (<0.0001)	4.3	0.15

[a] The coefficient of variation (*CV*) is a unit-less measure of variation, and is calculated as [(Mean Square Error/Mean) × 100]

[b] The coefficient of determination (R^2) is the proportion of the variation explained by the association between two sets of measurements

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

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

[e] The correction factor (C_b) measures how far the best-fit line deviates from 45° and is thus a measure of accuracy

[f] The correlation coefficient (r) measures precision

[g] 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

there was no relationship between prior ability and magnitude of gain or loss.

Using either the 5 or 6-diagram SADs improved inter-rater reliability (Fig. 4 and Table 4a), but the 5-diagram SADs provided the more reliable estimates. The 95 % CIs of the ICC indicated inter-rater reliability was highest using the 5-diagram SADs (Table 4b). Although inter-rater reliability (R^2) was improved estimating severity using 5- or 6-diagram SADs relative to no SADs, and those raters who were less accurate or precise improved the most (Fig. 3f and Table 3, markers to the *right* of the vertical dashed line – positive values - indicate improvement), when using 6-diagram SADs compared to 5-diagram SADs there was no relationship between prior ability and magnitude of gain or loss.

Absolute error was reduced using either 5- or 6-diagram SADs compared to error for estimates made without SADs. There was a >10 % discrepancy for 15 estimates when using no SADs, but all estimates made using 5-diagram SADs ≤10 % discrepant from the actual value (Fig. 5). Using 6-diagram SADs, all but 3 estimates were within 10 % of the actual value. Absolute error reflected the tendency of most of the raters to overestimate the severity of ACC on ripe fruit. Use of either SAD set resulted in the estimates being a better balance of over- and under-estimates, and had fewer extreme values.

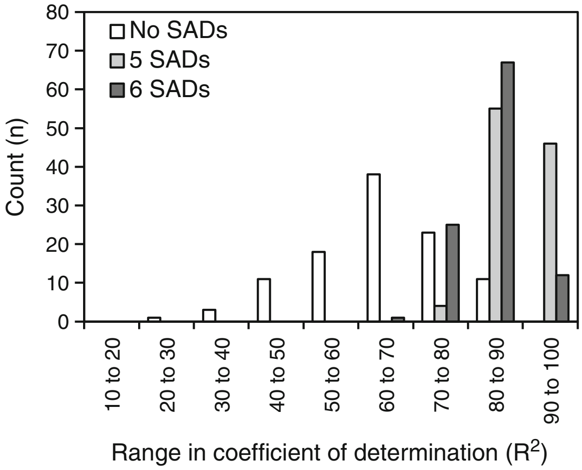

Fig. 4 The frequency of the inter-rater reliability of 15 raters measured by the coefficient of determination (R^2) without and with use of a standard area diagram set (SAD) with 5 or 6 diagrams as an aid for assessment of 40 images of ripe sweet orange fruit displaying a range in severity of Asiatic citrus canker

Discussion

The results corroborate many previous studies showing that SADs improve accuracy and reliability of visual estimates of disease severity (Amorim and Bergamin Filho 1996; Belasque et al. 2005; Braido et al. 2014; Godoy et al. 2006; Michereff et al. 2006; Nutter et al. 1998; Spolti et al. 2011; Spósito et al. 2004; Yadav et al. 2013). However, few studies have compared number of diagrams in a SAD set (Braido et al. 2014; Corrêa et al. 2009). This study agrees with a previous study that either 5- or 6-diagram SADs with the same range of severity (0.7 to 39.0 %) but slightly different severities illustrated improved estimates of ACC severity on unripe fruit of sweet orange (Braido et al. 2014). It is worth noting that the range of severity used in these SADs is representative of that observed in the field on ripe citrus fruit with symptoms of ACC. Rarely, if ever, are fruit observed with ACC symptoms >40 %; if infection is more severe fruit are aborted (Braido et al., personal observation).

ACC symptoms are estimated visually on both ripe and unripe fruit for various reasons, including monitoring epidemic progress and comparing treatments for purposes of hypothesis testing. Symptoms on unripe and ripe fruit differ considerably. Typically, on unripe fruit there may be a chlorotic halo associated with the canker lesions that must be included in disease severity estimation. On ripe fruit, which are yellow-orange, there is no halo, and the corky, erumpent lesion ends abruptly

Table 4 The inter-rater reliability of the estimates of Asiatic citrus canker severity made by 15 raters of 40 images of ripe fruit of sweet orange displaying a range in severity of symptoms either without or with use of a standard area diagram set (SADs) with 5- or 6-diagrams as an aid for assessment. Inter-rater reliability was measured using **a**. the coefficient of determination (R^2)[a] and **b**. the intra-correlation coefficient (ρ)[b]

A. Coefficient of determination					
Variable relationship tested	Means (R^2)	Mean diff[c]			95 % CIs[d]
No SADs x 5 SADs	0.638, 0.889	0.252			0.227 to 0.275
No SADs x 6 SADs	0.638, 0.834	0.197			0.173 to 0.221
5 SADs x 6 SADs	0.889, 0.834	−0.054			−0.065 to −0.045
B. Intra-class correlation coefficient (ICC, ρ)					
Variable tested	F-value (P-value) Fruit no.	Rater no	CV	ICC[e]	95 % CIs
No SADs	54.2 (<0.0001)	13.1 (<0.0001)	40.3	0.95a	0.941 to 0.958
5 SADs	230.3 (<0.0001)	2.9 (0.0003)	25.3	0.99c	0.988 to 0.992
6 SADs	148.0 (<0.0001)	4.4 (<0.0001)	30.1	0.98b	0.976 to 0.983

[a] The coefficient of determination (R^2) is the proportion of the variation explained by the association between two sets of measurements

[b] The ICC (ρ) 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

[c] Mean of the difference between each rating

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

[e] Numbers with the same letters are not significantly different at α =0.05

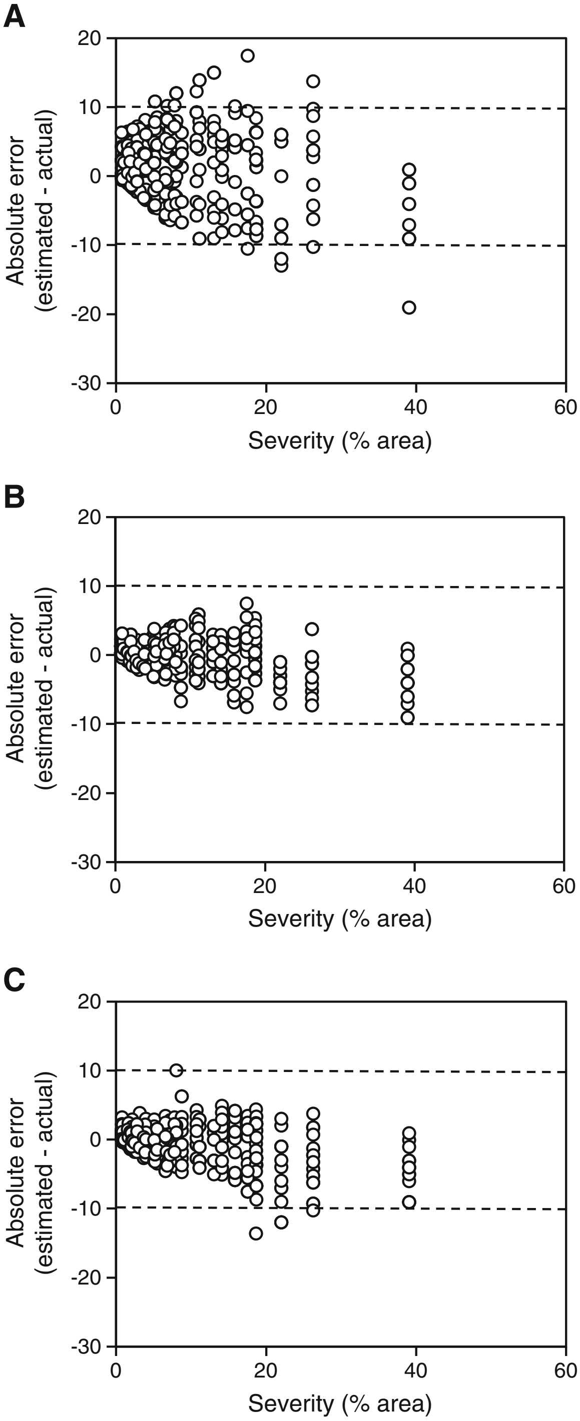

Fig. 5 The absolute error (estimate minus true disease) of assessments of a set of 40 images of cankered, ripe sweet orange fruit by 15 raters without use of standard area diagram sets (SADs) as assessment aides **a**, or using a SADs with 5 **b** or 6 **c** diagrams depicting different severities of Asiatic citrus canker. The dashed lines indicate ±10 % absolute error limits

where the healthy, ripe fruit tissue starts. This difference in symptoms warrants a specific SAD set for ripe fruit. A previous report described assessment on unripe fruit using 5- or 6-diagram SAD sets (Braido et al. 2014), and the results of that study were similar to this one – the 5-diagram SAD set resulted in more accurate, reliable estimates compared with the 6-diagram SAD set.

Why use of the 5-diagram SADs resulted in slightly more accurate results with both unripe (Braido et al. 2014) and ripe (this study) fruit is unclear. As discussed in the previous article on unripe fruit, the difference in SAD structure (the individual severities presented) might result in the 6-diagram SADs being less useful for guiding raters when interpolating estimates of disease severity. Or, the extra diagram at low severity (≤10 %) might cause raters to over-analyze their estimate, resulting in less accurate or reliable data. The number of diagrams depicted in a SAD set can affect agreement (Braido et al. 2014; Corrêa et al. 2009), although in the Corrêa et al. study, the SADs were used as a key to categorize disease, rather than interpolate an estimate.

In this study there was a tendency for unaided raters to overestimate, particularly at disease severity <20 % (Fig. 5). Although raters also overestimated severity on unripe fruit when severity was <20 % (Braido et al. 2014), raters in that study also tended to underestimate disease on unripe fruit at severity >20 %. Use of SADs helped reduce these tendencies. Underestimation on fruit has been noted before using SADs to aid assessment of apple fruit disease (Spolti et al. 2011), which contrasts with estimates of foliar disease, where there is most often a tendency to overestimate (Bock et al. 2008b; Sherwood et al. 1983). Fruit shape, disease ditribution and lesion size might affect the tendency to over or under-estimate (Bock et al. 2011; Braido et al. 2014; Sherwood et al. 1983; Spolti et al. 2011).

With the exception of systematic bias (υ), all aspects of accuracy and reliability were consistently improved when raters used SADs as an aid to estimate severity. Improvement in estimation using SADs is well-established (Braido et al. 2014; Spolti et al. 2011; Yadav et al. 2013), and this study confirms this point. Greater accuracy (agreement with the actual values) results in better inter-rater reliability, although where multiple raters are used to assess experiments they should be deployed in a manner to ensure that rater variability is spread across treatments.

As with the unripe sweet orange fruit (Braido et al. 2014), both the 5- and 6-diagram SADs improved the accuracy and reliability of visual estimates of severity of ACC, although use of the 5-diagram SADs appeared to provide slightly better agreement and reliability compared to the 6-diagram SADs, so they are recommended in preference to the 6-diagram SAD. These SADs developed for ripe sweet orange fruit doubtless have applicability to other citrus, and will be useful for research where visual estimates of severity of ACC are required.

Acknowledgments We acknowledge support of The National Council for Scientific and Technological Development – CNPq, for financial aid provided and granted to the seventh author, and to Coordination of Improvement of Higher Education Personnel – CAPES, for financial assistance granted to the second and third author for a scholarship.

References

Amorim, L., & Bergamin Filho, A. (1996). *Doenças de Plantas Tropicais: Epidemiologia e Controle Econômico*. Editora Ceres: São Paulo.

Bardsley, S. J., & Ngugi, H. K. (2012). Reliability and accuracy of visual methods used to quantify severity of foliar bacterial spot symptoms on peach and nectarine. *Plant Pathology, 62*, 460–474.

Behlau, F., Belasque, J., Jr., Graham, J. H., & 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 Protection, 29*, 300–305.

Belasque, J., Jr., Bassanezi, R. B., Spósito, M. B., Ribeiro, L. M., Jesus, W. C., Jr., & Amorim, L. (2005). Escalas diagramáticas para a avaliação da severidade do cancro cítrico. *Fitopatologia Brasileira, 30*, 387–393.

Belasque, J., Jr., Jaciani, F. J., Marin, D. R., & Barbosa, J. C. (2008). Tamanho da amostra para quantificação do tamanho de lesões do cancro cítrico. *Tropical Plant Pathology, 33*, 317–322.

Bitancourt, A. A. (1957). O cancro citrico. *O Biologico, 23*, 101–111.

Bock, C. H., Parker, P. E., Cook, A. Z., & Gottwald, T. R. (2008a). Visual assessment and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. *Plant Disease, 92*, 530–541.

Bock, C. H., Parker, P. E., Cook, A. Z., & Gottwald, T. R. (2008b). Characteristics of the perception of different severity measures of citrus canker and the relations between the various symptom types. *Plant Disease, 92*, 927–939.

Bock, C. H., Poole, G., Parker, P. E., & Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical Reviews in Plant Sciences, 29*, 59–107.

Bock, C. H., Gottwald, T. R., & Parker, P. E. (2011). Distribution of canker lesions on the surface of diseased grapefruit. *Plant Pathology, 60*, 986–991.

Braido, R., Gonçalves-Zuliani, A.M.O., Carvalho, S.A., Belasque Junior, J., Bock, C.H., & 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 Disease, 98,* Accepted.

Corrêa, F. M., Bueno Filho, J. S. S., & Carmo, M. G. F. (2009). Comparison of three diagrammatic keys for the quantification of late blight in tomato leaves. *Plant Pathology, 58*, 1128–1133.

Das, A. K. (2003). Citrus canker—a review. *Journal of Applied Horticulture, 5*, 52–60.

Godoy, C. V., Koga, L. J., & Canteri, M. G. (2006). Diagrammatic scale for assessment of soybean rust severity. *Fitopatologia Brasileira, 31*, 63–68.

Gottwald, T.R., Graham, J.H. & Schubert, T.S. (2002). Citrus canker: The pathogen and its impact. Plant Health Progress doi:10.1094/PHP-2002-0812-01-RV. http://www.plantmanagementnet work.org/pub/php/review/citruscanker/

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., & 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 Protection, 28*, 508–524.

James, W. C. (1971). An illustrated series of assessment keys for plant diseases, their preparation and usage. *Canadian Plant Disease Survey, 51*, 39–65.

Kranz, J. (1977). A study on maximum severity in plant disease. *Travaux de'die's a' G. Viennot-Bourgin, 16*, 9–73.

Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45*, 255–68.

Lu, L., & 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.

Madden, L. V., Hughes, G., & van den Bosch, F. (2007). *The Study of Plant Disease Epidemics*. St. Paul, MN: APS Press.

Michereff, S. J., Noronha, M. A., Andrade, D. E. G. T., Oliveira, E. P., Xavier Filha, M. S., & Moreira, P. A. A. (2006). Development and validation of a diagrammatic key for *Cercospora* leaf spot of sweet pepper. *Summa Phytopathologica, 32*, 260–266.

Nita, M., Ellis, M. A., & Madden, L. V. (2003). Reliability and accuracy of visual estimation of Phomopsis leaf blight of strawberry. *Phytopathology, 93*, 995–1005.

Nutter, F.W. Jr., Miller, D.L., & Wegulo, S.N. (1998). *Do standard diagrams improve the accuracy and precision of disease assessment?* In: 7th International Congress of Plant Pathology, Edinburgh. Volume 2, p.2.1.16

Nutter, F. W., Jr., Teng, P. S., & Shokes, F. M. (1991). Disease assessment terms and concepts. *Plant Disease, 75*, 1187–1188.

Nutter, F. W., Jr., Gleason, M. L., Jenco, J. H., & Christians, N. L. (1993). Accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. *Phytopathology, 83*, 806–12.

Parker, S. R., Shaw, M. W., & Royle, D. J. (1995). Reliable measurement of disease severity. *Aspects of Applied Biology, 43*, 205–214.

Schubert, T., Shabbir, J., Rizvi, A., Sun, X., Gottwald, T. R., Graham, J. H., & Dixon, W. N. (2001). Meeting the challenge of eradicating citrus canker in Florida - again. *Plant Disease, 85*, 340–356.

Sherwood, R. T., Berg, C. C., Hoover, M. R., & Zeiders, K. E. (1983). Illusions in visual assessment of *Stagonospora* Leaf Spot of orchardgrass. *Phytopathology, 73*, 173–177.

Spolti, P., Schneider, L., Sanhueza, R. M. V., Batzer, J. C., Gleason, M. L., & Medeiros Del Ponte, E. (2011). Improving sooty blotch and flyspeck severity estimation on apple fruit with the aid of standard area diagrams. *European Journal of Plant Pathology, 129*, 21–9.

Spósito, M. B., Amorim, L., Belasque, J., Jr., Bassanezi, R. B., & Aquino, R. (2004). Elaboração e validação de escala diagramática para avaliação da severidade da mancha preta em frutos cítricos. *Fitopatologia Brasileira, 29*, 81–85.

Vale, F.X.R., Fernandes Filho, E.I., & Liberato J.R. (2003). *QUANT – A software for plant disease severity assessment*. 8th International Congress of Plant Pathology. Christchurch, New Zealand, 2–7. Pp 105, abstract 8.18.

Yadav, N. V. S., de Vos, S. M., Bock, C. H., & Wood, B. W. (2013). Development and validation of standard area diagrams to aide assessment of pecan scab symptoms on pecan fruit. *Plant Pathology, 62*, 325–335.

Yi, Q., Wang, P. P., & He, Y. (2008). Reliability analysis for continuous measurements: Equivalence test for agreement. *Statistics in Medicine, 27*, 2816–2815.