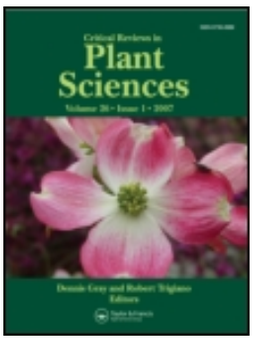

Critical Reviews in Plant Sciences

ISSN: 0735-2689 (Print) 1549-7836 (Online) Journal homepage: http://www.tandfonline.com/loi/bpts20

Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging

C. H. Bock , G. H. Poole , P. E. Parker & T. R. Gottwald

To cite this article: C. H. Bock , G. H. Poole , P. E. Parker & T. R. Gottwald (2010) Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging, Critical Reviews in Plant Sciences, 29:2, 59-107, DOI: 10.1080/07352681003617285

To link to this article: https://doi.org/10.1080/07352681003617285

Critical Reviews in Plant Sciences, 29:59–107, 2010
ISSN: 0735-2689 print / 1549-7836 online
DOI: 10.1080/07352681003617285

Taylor & Francis
Taylor & Francis Group

Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging

C.H. Bock,[1] G.H. Poole,[2] P.E. Parker,[3] and T.R. Gottwald[2]
[1]*USDA-ARS-SEFNTRL, Byron, GA, USA**
[2]*USDA-ARS-USHRL, Ft. Pierce, FL, USA*
[3]*USDA-APHIS-PPQ, Moore Air Base, Edinburg, TX, USA*

Table of Contents

Address correspondence to Clive Bock, USDA-ARS-SEFNTRL, 21 Dunbar Rd., Byron, GA 31008, USA. E-mail: clive.bock@ars.usda.gov
*Dr. Bock was formerly associated with the University of Florida/USDA, Ft. Pierce, FL, USA, at the time this article was written.

Reliable, precise and accurate estimates of disease severity are important for predicting yield loss, monitoring and forecasting epidemics, for assessing crop germplasm for disease resistance, and for understanding fundamental biological processes including coevolution. Disease assessments that are inaccurate and/or imprecise might lead to faulty conclusions being drawn from the data, which in turn can lead to incorrect actions being taken in disease management decisions. Plant disease can be quantified in several different ways. This review considers plant disease severity assessment at the scale of individual plant parts or plants, and describes our current understanding of the sources and causes of assessment error, a better understanding of which is required before improvements can be targeted. The review also considers how these can be identified using various statistical tools. Indeed, great strides have been made in the last thirty years in identifying the sources of assessment error inherent to visual rating, and this review highlights ways that assessment errors can be reduced—particularly by training raters or using assessment aids. Lesion number in relation to area infected is known to influence accuracy and precision of visual estimates—the greater the number of lesions for a given area infected results in more overestimation. Furthermore, there is a widespread tendency to overestimate disease severity at low severities (<10%). Both interrater and intrarater reliability can be variable, particularly if training or rating aids are not used. During the last eighty years acceptable accuracy and precision of visual disease assessments have often been achieved using disease scales, particularly because of the time they allegedly save, and the ease with which they can be learned, but recent work suggests there can be some disadvantages to their use. This review considers new technologies that offer opportunity to assess disease with greater objectivity (reliability, precision, and accuracy). One of these, visible light photography and digital image analysis has been increasingly used over the last thirty years, as software has become more sophisticated and user-friendly. Indeed, some studies have produced very accurate estimates of disease using image analysis. In contrast, hyperspectral imagery is relatively recent and has not been widely applied in plant pathology. Nonetheless, it offers interesting and potentially discerning opportunities to assess disease. As plant disease assessment becomes better understood, it is against the backdrop of concepts of reliability, precision and accuracy (and agreement) in plant pathology and measurement science. This review briefly describes these concepts in relation to plant disease assessment. Various advantages and disadvantages of the different approaches to disease assessment are described. For each assessment method some future research priorities are identified that would be of value in better understanding the theory of disease assessment, as it applies to improving and fully realizing the potential of image analysis and hyperspectral imagery.

Keywords plant disease assessment, variance, error, image analysis, hyperspectral imagery, remote sensing

I. INTRODUCTION

A. Background

There are diverse reasons why we need to estimate or measure disease on plants. Knowledge of the quantity of disease is particularly important to decision-makers in crop situations where disease must be related to yield loss, in plant breeding where various germplasm, varieties and/or cultivars need to be rated, and for disease management decisions, for example, applying pesticides to control disease epidemics, but also for understanding fundamental processes in biology, including coevolution and plant disease epidemiology (James, 1971; Berger, 1980; Kranz, 1988; Gaunt, 1995; Nutter and Gaunt, 1996; Burdon *et al.*, 2006; Cooke, 2006; Nutter *et al.,* 2006).

Quantifying disease on plants by measuring symptoms generally falls under the broad definition of "remote sensing" (there are one or two exceptions). Remote sensing can be defined as obtaining information about an object without having direct physical contact with it (de Jong *et al.*, 2006). In contrast, measurement of pH, soil temperature or moisture (and many other phenomena) is done by direct contact with the object. Thus, both visual estimation of disease and using cameras or other imaging technologies to measures disease can be considered as remote sensing (Nutter, 1990). Furthermore, remote sensing of disease is a passive remote sensing process, rather than an active method that would generate imaging radiation, such as X-rays. Exceptions to measuring disease by remote sensing could be by weighing diseased vs. healthy portions of an actual leaf or photograph to ascertain percent severity (Tucker and Chakraborty, 1997; Nita *et al.*, 2003). Some planimeter measurements, and quantifying the pathogen using molecular or immunological techniques are also direct measurements and do not fall under the broad definition of remote sensing.

Much of the historic background to the development of ways to measure or estimate disease is considered in various reviews (Chester, 1950; Large, 1966; James, 1971; Berger, 1980; Nilsson, 1995; Nutter *et al.*, 2006), and in books (Campbell and Madden, 1990; Cooke, 2006; Madden *et al.*, 2007). Early

estimation of disease severity included visual estimation qualitatively or quantitatively, often with the aid of various assessment scales or keys (Cobb, 1892; Horsfall and Barratt, 1945; Chester, 1950; James, 1971), a process that has become better understood and thus more exacting over the last thirty years (Nutter *et al.*, 1993; Nutter and Schulz, 1995; Madden *et al.*, 2007; Bock *et al.*, 2009b). The value of assessment training and the use of color (or black and white) standard area diagrams to assist raters are now acknowledged (Nutter and Schulz, 1995; Nutter *et al.*, 2006).

The application of technologically advanced remote sensing techniques to locate and measure plant disease started early with aerial photography (Neblette, 1927; Taubenhaus *et al.*, 1929) and has been applied to various pathosystems (Colwell, 1956; Brenchley, 1964; Wallen and Jackson, 1971; Jackson *et al.*, 1978; Lillesand *et al.*, 1981; Gerten and Weise, 1984; Edwards *et al.*, 1985; Lee, 1989). In the last thirty years these and newer methods continue to be explored to quantify plant disease, which include methods of detecting and measuring the pathogen or disease population at various scales, ranging from molecular to regional views (Chester, 1950; Price and Osborne, 1990; Nilsson, 1995, Nutter *et al.*, 2006). Thus, the pathogens can be detected and quantified by immunological and molecular methods (ELISA, PCR etc; Nutter *et al.*, 2006; Jackson *et al.*, 2007) and by microscopy (Hilber and Scheupp, 1992). The disease caused by the pathogen can be detected and quantified as symptoms by laser induced fluorescence, radar, microwave, thermography, nuclear magnetic resonance imaging, multi- and hyper-spectral imagery, digital (or film) cameras and image analysis, and visual assessment (Pederson and Nutter, 1982; Price and Osborne, 1990; Nilsson, 1995). Of the methods available to measure disease, it is important to choose a method that is appropriate and will estimate the correct variable with sufficient quality so that the data represent a realistic estimate of the actual level of disease in a host population, thereby assuring that subsequent conclusions drawn from the data are correct. While achieving these ends it is also necessary to utilize limited resources (equipment, labor and time) as effectively and efficiently as possible (Campbell and Madden, 1990; Nutter and Gaunt, 1996).

B. Aim and Scope of This Review

Disease in plant populations can be quantified in several ways (intensity, prevalence, incidence and severity, defined in the next section). Only estimation and measurement of plant disease severity will be considered in this review. Furthermore, the majority of plant pathologists quantify disease severity at the scale of individual plant organs (leaves, stems, fruit, and roots), plants or in small quadrats, and of the many methods that have been used or investigated for estimating or measuring disease severity, this review will consider only three approaches: visual assessment, digital photography and image analysis, and hyperspectral imagery. Visual assessment is now reasonably well understood (the practice is over 100 years old; Chester, 1950; Cobb, 1892). Digital (and film) photography with image analysis is being increasingly applied as the technology evolves and has found considerable application in the plant sciences including forestry, horticulture and agriculture, often dealing with whole tree and for plant stand issues (defoliation, vegetative cover), which is beyond the scope of this review (Price and Osborne, 1990; Nilsson, 1995). It is of note that digital images are used not only for assessing disease severity, but also for the diagnosis of plant diseases and other disorders (Holmes *et al.*, 2000). Finally, hyperspectral imagery is new to plant disease severity measurement and is still in the early phase of developmental use in plant disease detection and quantification, but offers interesting opportunities for application (Coops *et al.*, 2003; Larsolle and Muhammad, 2007; Huang *et al.*, 2007; Qin *et al.*, 2008). The statistical tools that are used to measure the quality of estimated disease assessments compared to the actual (or true) disease will be described. Visual assessment, digital image analysis, and hyperspectral imagery will be considered from an historic context, including a description of the technologies and their application in plant disease assessment, as well as the potential problems, among the three methods. Finally an outline of future areas of research interest and need will be presented.

II. TERMS AND DEFINITIONS USED IN PLANT DISEASE ESTIMATION AND MEASUREMENT

The terms used to describe concepts and their interpretation are important in plant disease assessment, and are subject to redefinition as a result of advances made in other fields, including measurement science (Madden *et al.*, 2007). The branch of plant pathology dealing with plant disease assessment was christened "plant pathometry" or "phytopathometry" (from the greek *phyto* = plant; *pathos* = disease, *metron* = measure) by Large (1953, 1966); this term was also used by Horsfall and Cowling (1978) and Nutter *et al.*, (2006). Based on Large's (1953, 1966) description, phytopathometry encompasses the mensuration, rules, and history of plant disease symptom assessment. Some preliminary attempts to discuss the importance of plant disease assessment were made by Moore (1943, 1949), but the first definitive review of the literature was by Chester (1950). The need for clear terminology in phytopathometry was recognized by the late 1980s (Nutter *et al.*, 1991), when a subcommittee of the American Phytopathological Society (APS) was formed to provide a consensus on disease assessment terms and concepts.

In this review the term plant disease severity "estimate" is used specifically for those assessments made visually, and the term "measurement" for those assessments by image analysis or hyperspectral equipment. Plant disease symptoms can be estimated or measured in various ways that quantify the

intensity, prevalence, *incidence*, or *severity* of disease. The following relate to the way in which disease symptoms are quantified:

i) *Disease intensity* is a general term used to describe the amount of disease present in a population (Nutter *et al.*, 1991).
ii) *Disease prevalence* is the proportion (or percent) of fields, counties, states, etc. where the disease is detected, and reveals disease at a grander scale than incidence (Nutter *et al.*, 1991).
iii) *Disease incidence* is the proportion (or percent) of plants (or plant units, leaves, branches, etc.) diseased out of a total number assessed (Nutter *et al.*, 1991; Madden *et al.*, 2007).
iv) *Disease severity* is the area (relative or absolute) of the sampling unit (leaf, fruit, etc.) showing symptoms of disease. It is most often expressed as a percentage or proportion (Nutter *et al.*, 1991). Disease severity is the only measure considered in this review.

Thus there are several ways to quantify disease, and understanding the differences among these methods is fundamental to disease assessment. Furthermore, it is useful to be aware of the relationships between disease incidence and severity (Seem, 1984), which is varied and dependent on the pathosystem and time of sampling, although in some cases incidence and severity are equivalent. When choosing a variable to assess, forethought should be given to the kind of data required. Various studies and reviews have considered the subject of disease severity and disease incidence in detail (Seem, 1984; McRoberts *et al.*, 2003).

A further consideration is that traditional concepts of plant disease as symptoms on a leaf or other organ have been challenged by the advent of immunological and molecular methods and the ability to measure the quantity of the pathogen, rather than the symptoms, which can then be used to relate to yield, disease progress, or host resistance (Nutter *et al.*, 2006; Jackson *et al.*, 2007). This is an important and valuable development, but has reinforced the need for clear terminology to avoid confusion. Pathogen intensity, prevalence, incidence, and severity are parallel terms that measure the quantity of the pathogen in a population (Nutter *et al.*, 2006). It is important to make the observation that disease estimated by symptoms or by pathogen quantity can have different results which has ramifications, for example, for disease versus pathogen progress estimation (Nutter, 1997b; 2001).

Disease symptoms can be incorrectly assessed by raters or disease measurement devices which results in error (Nutter *et al.*, 1993; Nita *et al.*, 2003; Nutter *et al.*, 2006). This error requires understanding to form the basis for improving the accuracy and precision of assessments. The terms and concepts of plant disease assessment that describe its quality, thus, need to be understood (Nutter *et al.*, 1991; Madden *et al.*, 2007). Many terms have been used concerning the quality of plant disease assessments including precision, reliability, reproducibility, repeatability, accuracy and agreement; all suggest a gauge of quality, but bear fuller examination to understand the relationships among them. Comparisons can be made between assessment data, and these fall primarily into two groups: those that relate to comparisons among raters or disease measurement devices (methods), and those in which ability of the rater or disease measurement device are being compared to the "actual" value of disease (the "true" value or "golden standard"). As the following description of concepts progresses, the reason for these divisions should become evident.

Various aspects of rater and assessment device error have to be measured and understood. Conceptualizations and the terms used in measurement science suggest that the measurement of plant disease severity should readdress some of these terms, their equivalence, and their usage (Madden *et al.*, 2007). For a recent and comprehensive review of terms and concepts used for assessing quality of estimates in measurement science, the reader is directed to Barnhart *et al.*, (2007). The concepts and definitions that follow adhere to those of Madden *et al.* (2007), reflecting the most recent developments in plant pathology:

i) *Reliability of estimates:* Reliability can be defined as "the extent to which the same measurement of individuals obtained under different conditions yield similar results" (Everitt, 1998). In studies of plant disease assessment there are two aspects of reliability that need to be considered. The first is intrarater reliability, or how similar repeat measurements taken by the same rater are to one another and is also known as repeatability (Nutter *et al.*, 1991) and equates to the test-retest procedure often referred to in medical studies (Deen *et al.*, 2000; Warke *et al.*, 2001). The second is interrater reliability, which is how similar measurements of the same specimens are between or among raters or rating methods at the same time. Interrater reliability is also called reproducibility (Madden *et al.*, 2007). The term "precision" is widely used in statistics to denote the amount of variability, with less precision resulting in greater variability (= less similarity). Madden *et al.* (2007) points out that although reliability equates to precision (variability) statistically, in measurement science they are not entirely synonymous. Also, highly precise estimates are not necessarily close to an actual value (Figure 1). Furthermore, as is presently described, precision is a component of accuracy and agreement, and to avoid confusion the term reliability provides a measure of variability in situations where actual values are not being considered (i.e., for inter and intrarater comparisons).
ii) *Accuracy and agreement of estimates*: According to Barnhard *et al.* (2007) "accuracy," *sensu strictissimo*, measures only systematic bias, while the term "precision" has been

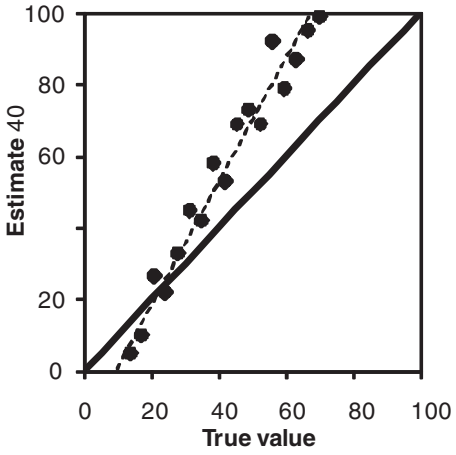

FIG. 1. Example illustrating good precision of estimates of X that are not necessarily close to the actual values which lie along the line of concordance (the solid line at 45^o), where the estimates would lie if there was a one-to-one relation between the estimate and the actual value, i.e., if they were accurate. Overestimates are above the concordance line and underestimates lie below the line of concordance. There is a transition point from under- to over-estimate at an actual value of approx. 20%.

used to measure random error. However, accuracy is often used in a somewhat broader sense and has included usage in relation to precision as well. This makes it particularly important to explicitly define what is meant by accuracy. In plant pathology, a definition of accuracy that has been used is "the degree of closeness of measured values to some recognized standard, true or actual values" (Everitt, 1998; Nutter, 2001; Madden *et al.*, 2007). Using this general definition, and the premise that perfect accuracy is possible only if there is no variability (imprecision) and no bias, Madden *et al.* (2007) reasoned, that in the case of estimates of actual values, accuracy was the product of both bias and precision. Thus measurement of accuracy is meaningful when the comparisons are made to actual or accepted standard values, and when estimate of disease are close to the actual value, they are considered accurate. Conceptually, agreement measures the closeness between readings; it is a broader term than accuracy, and by definition encompasses both accuracy and precision (Barnhart *et al.*, 2007). In comparing estimates to actual values agreement is the product of coefficients of precision and accuracy (Lin, 1989). Based on the aforementioned description of the concepts and the general definition of accuracy and agreement in plant pathology, it is apparent that when estimates are being compared to actual values, then accuracy and agreement are equivalent terms (Figure 2; Lin, 1989; Barnhart *et al.*, 2007; Madden *et al.*, 2007). In plant pathology the term reliability, already described in the previous section, is used to describe the agreement between measurements when actual values are not involved (i.e., inter- and intrarater estimates; Madden *et al.*, 2007).

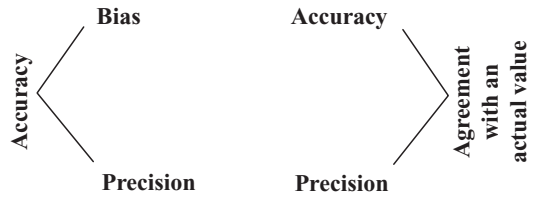

FIG. 2. The concepts that make accuracy and agreement equivalent in relation to estimates of "actual" or "true" value based on bias, accuracy and precision, and the definitions of these terms described in Section III (after Madden *et al.*, 2007).

III. APPROACHES TO IDENTIFY ERROR, TEST RELIABILITY, ACCURACY, AND AGREEMENT OF DISEASE SEVERITY ESTIMATES AND MEASURMENTS

A. The Need for Actual Values

Without an estimate of the actual disease value it is not possible to assess accuracy or agreement of the measurement or estimate (Madden *et al.*, 2007). However, a measure of interrater (or method) reliability or intrarater (or method) reliability can be gained for raters and methods of severity measurement based on comparisons of estimates or measurements on two occasions by the same rater or method (intrarater or intra-method reliability) and between one rater or method and another (interrater or inter-method reliability) at the same time. Using rater values as a standard inevitably runs the risk of adding error inherent to that rater ability, although raters have been used to provide a "true" or "standard" value as well (O'Brein and van Bruggen, 1992).

There are several ways to measure the actual disease severity. An accurate and reliable method is to make prints of the leaves and cut out healthy and diseased parts and weigh them (Sherwood *et al.*, 1983; Nita *et al.*, 2003). In various reports using this system repeatability has been excellent (Sherwood *et al.*, 1983; Nutter *et al.*, 1993; Nita *et al.*, 2003). However, any system will be subject to operator error, including defining exactly where diseased areas end and healthy tissue begins. Image analysis can be used to estimate actual values of disease on a leaf-by-leaf basis, and is highly repeatable. Bock *et al.* (2008a) showed good intra-method reliability (coefficient of determination, $r^2 = 0.95$) for estimates of area infected with citrus canker (with an r^2 of 0.99 for counts of lesion number). Reliable and accurate estimates of disease severity using image analysis have been reported previously (Lindow and Webb, 1983; Martin and Rybicki, 1998). Painting infected areas prior to actual value image analysis measurements has also been used (Olmstead *et al.*, 2001). Nutter and Gaunt (1996) noted a subjective step in image processing where the operator assigns pixels to represent healthy versus diseased tissue – however, a certain level of subjectivity exists when marking a diseased area with a pen, projecting leaf

images onto acetates and drawing in diseased areas with a black marker pen and then performing image analysis, or cutting and weighing portions of leaves. Done with care, these are probably all reasonably accurate ways of obtaining an actual value. Various planimeters have also been used to obtain actual values of disease (Lindow and Webb, 1983; Price *et al.*, 1993; Tucker and Chakraborty, 1997).

Without actual values, the absolute ability of raters, or new disease assessment methods cannot be evaluated. Various statistical methods exist to compare a new technique to actual values, and assess agreement and investigate reliability (Nutter *et al.*, 1993; Nutter and Schulz, 1995; Madden *et al.*, 2007). Several of these methods are standard statistical tools used in many aspects of plant pathology. However, other methods are more recent, or have not been used as much in this discipline. One example is Lin's concordance correlation coefficient, which has only recently been applied in plant pathology (Nita *et al.*, 2003; Madden *et al.*, 2007; Bock *et al.*, 2008a), but has been increasingly used in many areas of measurement science, including in medicine. However, before analysis the quality of the data should be checked to ensure that it does not violate any of the assumptions of a proposed analysis (normality, heteroscedasticity, etc.) and if it does, data transformation should be considered.

B. Statistical Tests to Identify Error and Quantify the Quality Of Disease Severity Estimates and Measurements

1. *Analysis of Variance (ANOVA) and General Linear Modeling (GLM)*

Analysis of variance is a parametric test that partitions variance in the sample into predefined explanatory variables or effects. In the analysis, the sums of squares are partitioned relating to the effects in the model, and an F-test is used to test for significant sources of variation. Several studies have used ANOVA, or GLM, to investigate sources (and magnitude) of variation or error in estimates or measurements. An early report using ANOVA to identify factors affecting assessment was that of Sherwood *et al.* (1983) who analyzed disease assessments of *Stagonospora* leaf spot on orchardgrass. Two- and three-way ANOVA were used to assign variance to the factors that influenced disease assessment, and the results showed that apart from the actual disease, there were significant effects of lesion number, leaf size, individual rater and rater experience. Using a similar approach with GLM, similar factors were found to influence assessment of citrus canker (*Xanthomonas citri* subsp. *citri*, *Xcc*) on grapefruit leaves (Bock *et al.*, 2008b). In another study comparing application of image analysis to estimates on a category scale, a nested ANOVA was used to test various components of variance in image analysis estimation (including image frame, individual specimen, and numbers of specimens per frame and specimen side and orientation). The results showed that none of these factors accounted for $> 4\%$ of the error in the measurement of infected area, and the image analysis data were more objective and relatively more precise (reliable) measurements compared to visual estimates on a category scale (Kokko *et al.*, 1993 and 2000). Application of ANOVA has shown how error in assessor and plant organ assessed affects disease estimation on simulated images of infected plant parts (Forbes and Jeger, 1987; Table 1). Precision was measured with the standard deviation of an individual's repeated measurements (the reliability of measurement), and accuracy was measured as average deviation between the actual and estimated values (closeness to the actual value). The analysis showed that low disease severity ($<25\%$ disease) was more reliably and accurately estimated, and that plant structures were a significant factor affecting quality of assessment.

TABLE 1
Various sources of error identified in a study with effects of assessor, order of assessment, plant structure, and severity. ANOVA was performed on the absolute error, proportional error and log-transformed error (after Forbes and Jeger, 1987).

		Mean square error values		
Source of variability	Df[1]	Absolute error[2]	Proportional error	Log-transformed error
Assessor (A)	11	3723**[3]	28**	38**
Order (O)	11	203	2*	8
Plant structure (PS)	11	1961**	10*	36**
A x O x PS	110	209	1	10
Intensity (I)	5	3876**	13*	172**
PS x I	55	1020**	3*	29**
A x I	55	277**	2**	13**
Residual	605	95	6	0.4

[1]Degrees of freedom.

[2]Absolute error = estimated percent area diseased – actual percent area diseased; Proportional error = actual percent area diseased; and log-tranformed error = *ln*true diseased area – *ln*proportional error.

[3]*, **represents 0.05 and 0.01 levels of significance, respectively. Assessor, order and plant structure were tested with A x O x PS mean square. Severity, PS x I and A x I were tested with the residual mean square.

For testing reliability and general agreement of rater estimates and/or method measurements, ANOVA has been used to calculate the intra-class correlation coefficient (ρ). The intra-class correlation coefficient can be defined as the ratio of actual variance to total variance (Shokes *et al.*, 1987; Nita *et al.*, 2003; Madden *et al.*, 2007; Bock *et al.*, 2008a). It is calculated from the variance components of the ANOVA, for example, in a case where only rater effects are being tested, there are the following sources of variance (σ), σ^2_{true}, $\sigma^2_{\text{observer}}$, and σ^2_{error}, the variance can be portioned into components by a two-way random effects ANOVA where $\rho = \sigma^2_{\text{true}}/(\sigma^2_{\text{true}} + \sigma^2_{\text{rater}} + \sigma^2_{\text{error}})$. Shokes *et al.* (1987) used a two-way ANOVA for testing reliability and agreement of various assessment procedures and assessors on late leafspot of peanut and showed that training improved interrater reliability, and certain symptoms were consistently more reliably estimated (for example, necrotic area) compared to other symptoms. The intra-class correlation was used to demonstrate that reliability of direct estimation of *Phomopsis* blight of strawberry was greater compared to using the Horsfall and Barratt (H-B) scale (Nita *et al.*, 2003) and also to compare reliability of different assessment measures (lesion numbers and severity of citrus canker) by Bock *et al.* (2008a). The results showed that lesion number was the most reliably estimated compared to percent area necrotic or percent area necrotic+chlorotic. Madden *et al.* (2007) considers ρ a better indicator of reliability and general agreement than the correlation coefficient (r) as it is generally smaller than r, and has the advantage that it can be calculated for multiple raters, methods and assessment times and incorporates a measure of general agreement.

Stonehouse (1994) used ANOVA to show differences among raters in the level of agreement with an image analysis standard, and other studies have used ANOVA and means separation to directly compare visual estimates of disease and image analysis measurement for comparing cultivars (Todd and Kommedahl, 1994; Niemira *et al.*, 1999). Todd and Kommedahl (1994) applied Duncan's multiple range test to show that image analysis data provided superior means separations of isolate and variety compared to visual analysis. These techniques have also been used to compare image analysis measurements and molecular quantification of pathogens (Jackson *et al.*, 2007). Finally, it should be mentioned that where data are not normally distributed, or for nonparametric data sets (ordinal scales and some interval scales) both image analysis measurements, rater assessments and actual values have been compared using non-parametric methods such as the Kruskal-Wallis analysis of variance (Olmstead *et al.*, 2001). In this case the data suggested that visual raters were more precise than image analysis.

2. *The Correlation Coefficient*

Correlation analysis measures the strength and nature (positive or negative) of the association between estimated disease and the actual values or between individual rater or method estimates, but it gives no insight into the relationship between the two variables. The statistic, the correlation coefficient, r, ranges from -1.0 to +1.0 and is calculated:

$$r = \left(\frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \right)$$

Where x is the actual value, $\bar{x}$ is the mean actual value, y is the estimate and $\bar{y}$ is the mean estimate. There are different correlation methods, the most commonly used being Pearson's product-moment correlation coefficient, which is a parametric statistic, and thus usually applied to continuous, normally distributed data and is generally pertinent to use for measures of disease severity. Correlation has been used in assessing inter- and intrarater (method) reliability (Shokes, *et al.*, 1987; Nita *et al.* 2003; Bock *et al.*, 2008a, 2009b). As noted above, statistically speaking, correlation analysis defining variability can be considered as a measure of precision. However, in measurement science where there is a need to accommodate the concepts of agreement it is not entirely synonymous and it is preferential to use the term reliability when comparing raters or methods in the absence of actual values (Madden *et al.*, 2007). Furthermore, the correlation coefficient does not measure closeness to an actual value, the nature of the relationship, bias or error (Nutter and Schulz, 1995). A high correlation coefficient (+/−1, or close to +/−1) does not mean that the estimates are close to the actual values, as on average they can be much greater or less, and still return a high r value (see Figure 1), and in this respect the intra-class correlation coefficient described above is more powerful for discerning reliability of measurement.

Shokes *et al.* (1987) used correlation analysis to show improvement in repeat assessments, and to identify the more reliably estimated indicators of disease severity. Some raters were more reliable in repeat estimates of disease, and estimates of some symptoms (defoliation) were less reliable compared to disease severity estimates. Nita *et al.* (2003) showed direct disease estimation ($r = 0.81$-0.97) was more reliable than estimates based on the Horsfall-Barratt (H-B) interval scale (Horsfall and Barratt, 1945, $r = 0.72$-0.94). Similarly, direct estimation of the severity of citrus canker was more reliable than the H-B scale-based estimates (Bock *et al.*, 2009b).

3. *Regression Analysis*

Regression analysis is a well-known tool that has been used to investigate various aspects of error in disease assessment (Sherwood *et al.*, 1983; Bock *et al.*, 2008b) and reliability, precision and accuracy of measured or estimated data (Amanat, 1976; Nutter *et al.*, 1993; Nutter and Schulz, 1995; Parker *et al.*, 1995a and b; Guan and Nutter, 2003; Nutter and Esker, 2006; Nutter *et al.*, 2006). However, it should be applied with awareness that erroneous conclusions might be drawn from the analysis under certain circumstances (Lin, 1989; Madden *et al.*, 2007). Nonetheless, it does remain an invaluable tool for exploring the relationships between variables in plant disease assessment.

Linear regression compares a series of paired observations (estimated disease and actual disease values) which are designated as X and Y. The dependent variable Y is assumed to depend on X, the independent variable. Estimated disease is typically the Y variable in regression analysis. Linear regression analysis draws a line of best fit that minimizes the sum of the squared deviations of the estimated from the actual disease. Various different statistics are used to assess the relationship between X and Y. These have been described in some detail in studies of disease assessment, and are briefly reiterated here (Nutter and Schulz, 1995; Nutter, 2002; Bock *et al.*, 2008a). The slope, and intercept, and associated standard errors can be calculated for the regression solution and are indicative of the quality (accuracy and/or precision) of the disease severity estimates depending on whether the disease severity estimates are being regressed against actual values or against other disease severity estimates. If slope = 1 and intercept = 0, and coefficient of determination, $r^2 = 1.00$, the assessments are perfectly accurate when estimates are being compared to actual values (if there is deviation of the slope and/or intercept from expected values then there is bias, and thus loss in accuracy of the estimates compared to the actual values). If the data being compared do not include actual values, the slope and intercept of the regression can only be used to explore inter- or intrarater reliability (precision). The coefficient of determination (r^2), like the correlation coefficient, is a measure of reliability. Statistically, it describes the proportion of variability accounted for by the regression model and provides a measure of the proportion of overlapping variance (how much X explains Y, Nutter and Schulz, 1995). The coefficient of variation (*CV*) of the regression can also be calculated and provides a dimensionless, standardized estimate of the amount of error associated with the estimates in the regression model (Nutter and Schulz, 1995), and is calculated:

$$CV = \left(\sqrt{MSE}/\bar{x}\right) x100$$

Where *MSE* is the mean square error and $\bar{x}$ is the mean for the sample. A high *CV* is indicative of poor precision (Nutter *et al.*, 1993), and can be used in interrater or method comparisons of reliability. The standard error of the Y-estimate can also be calculated to indicate precision (Nutter and Schulz, 1995). The standard error of the Y-estimate offers additional information beyond the *CV* as it indicates the error involved in the actual prediction, compared to the overall error indicated by the *CV*. The standard error of the estimate (σ_{ypred}) is calculated:

$$\sigma_{ypred} = \sigma_y\sqrt{1 - r^2}$$

Where σ_y is the standard deviation of the dependent variable Y, and r is the correlation coefficient between X and Y. The residual plots of the regression analysis (the vertical distance of each Y point from the regression line vs actual disease-estimated) should always be checked as they provide insight into the applicability of, and assumptions underlying the regression analysis (Gomes *et al.*, 2004; Andrade *et al.*, 2005; Belasque *et al.*, 2005). If there is evidence of heteroscedasticity (non-constant variance with magnitude of X) some of the assumptions of the regression might be violated, although if using robust standard errors heteroscedasticity is less critical (SAS, 2004). There are various tests for heteroscedasticity, including White's and Breush-Pagan's tests (Breusch and Pagan, 1979; White, 1980; SAS, 2004; Bock *et al.*, 2008b). Both tests are based on the residuals of the fitted regression model, but White's test is more general and makes no assumptions about the form of the heteroscedasticity.

Regression analysis has been widely used to demonstrate interrater (method) reliability (Smith *et al.*, 1969; Nutter *et al.*, 1993; Newton and Hackett, 1994; Nutter and Schulz, 1995; Parker *et al.*, 1995a and b; Guan and Nutter, 2003; Bock *et al.*, 2008a), intrarater (method) reliability (Newton and Hackett, 1994; Guan and Nutter, 2003; Bock *et al.* 2008a), accuracy or agreement (Lindow and Webb, 1983; Nutter *et al.*, 1993; Nutter and Schulz, 1995; Martin and Rybicki, 1998; Olmstead *et al.*, 2001; Bock *et al.*, 2008a) and also applied to show the benefits of computer training (Nutter and Schulz, 1995; Parker *et al.*, 1995b) and the use of standard area diagrams, or SADs (Hock *et al.*, 1992; Parker *et al.*, 1995b; Godoy *et al.*, 1997; Nutter *et al.*, 1998; Leite and Amorin, 2002; Gomes *et al.*, 2004; Belasque *et al.*, 2005; Godoy *et al.*, 2006). In almost all cases the parameters of the regression solution improved subsequent to training. More than any other method, regression analysis has been used to gauge the quality of assessments.

Regression analysis has also found useful application for investigating the many factors that influence rater estimates of disease. There are illusions in disease assessment and regression analysis has been used to demonstrate the role of lesion number in estimating diseased area, and the tendency to overestimate disease at low severity (Sherwood *et al.*, 1983; Forbes and Jeger, 1987; Nita *et al.*, 2003; Bock *et al.*, 2008a and 2009a). Studies using regression analysis have confirmed a linear relationship exists between estimated and actual disease (Forbes and Jeger, 1987; Nutter and Schulz, 1995; Nita *et al.*, 2003; Bock *et al.*, 2008a). It was also used to investigate the relationship between variance of the mean estimate of 28 raters and magnitude of actual severity of citrus canker, showing variance of the mean increased up to 35% disease, the maximum in that study (Bock *et al.*, 2008b), indicating that estimates of more severe disease were less precise among raters.

An F-test is a useful analysis to compare regression lines (also known as a "test of parallelism" or a "test for heterogeneity of regression") and ascertains whether two or more dependent variables have different slopes and intercepts, or not. Using this analysis, Parker *et al.* (1995a) were able to demonstrate that individual rater's were significantly different in their ability to rate disease, and Bock *et al.* (2009a) showed differences between

experienced and inexperienced raters assessing citrus canker. The regression slopes are tested to see if there are differences in the dependent variable (dv) due to the independent variable of group (g) when compared to the continuous independent variable, the regressor (r). A test for interaction between regressor and group ($dv = r, g, r \times g$) tests for the presence of group differences which are indicated in separate slopes for the two groups. If the interaction is not different a further test for differences in the elevation of the regression lines is required, which is achieved by testing the reduced model (no interaction term, $dv = r, g$). If there is a significant group (g) effect the two groups have different elevations but the same slope.

Stepwise linear regression has been used to assess the factors that significantly influence raters assessing *Phomopsis* on strawberry leaves (Nita *et al.*, 2003). Stepwise linear regression analysis showed that lesion number and size significantly contributed to the error, but these sources were minor. Non-linear regression analysis was also used to investigate the functional relationship between estimated and actual disease severity in the strawberry-*Phomopsis* pathosystem, but the non-linear functions were not significant, and only improved the co-efficient of determination by one percentage point, confirming the applicability of the linear functions for this relationship.

4. *Lin's Concordance Correlation Coefficient*

Lin described the reasons for evaluating measurements: "In an assay validation or an instrument validation process, the reproducibility of the measurements from trial to trial is of interest. Also, when a new assay or instrument is developed, it is of interest to evaluate whether the new assay can produce the results based on a traditional gold standard assay" (Lin, 1989). Lin's concordance correlation coefficient has been widely used in various disciplines to quantify and compare accuracy and agreement (Nita *et al.*, 2003; Critchley *et al.*, 2005; Trager *et al.*, 2006; Bock *et al.*, 2008a), reproducibility (Richard *et al.*, 2000; Lin *et al.*, 2004) and test-retest characteristics (Deen *et al.*, 2000; Klassen *et al.*, 2001; Warke *et al.*, 2001; Bock *et al.*, 2008a) of estimates and measurements.

The disadvantage of using regression analysis in some situations to quantify accuracy or agreement was previously addressed (Lin, 1989; Madden *et al.*, 2007). The claim was that regression analysis did not detect departure from intercept 0 and slope 1 if data are very scattered (the less precise the data, the less likely the hypothesis will be rejected), and conversely a highly reproducible system could be rejected due to very small error. This observation prompted Lin (1989) to develop a new concordance correlation coefficient. The coefficient has been called "Lin's concordance correlation coefficient"; it provides an unbiased and quantifiable method to test accuracy or agreement. The analysis calculates and evaluates the degree to which pairs of observations fall on the concordance line of 45° (slope = 1, intercept = 0), and the concordance correlation coefficient, ρ_c, combines the measures of accuracy and precision to assess the relational fit to the line of concordance (45°):

$$\rho_c = rC_b,$$

Where C_b is a bias correction factor that measures how far the best-fit line deviates from 45° and is thus a measure of bias or accuracy, and r, the previously described correlation coefficient between X and Y, which measures in this case, the precision of the best fit line. C_b the bias correction factor is derived from:

$$C_b = \left[\left(\upsilon + 1/\upsilon + u^2\right)/2\right]^{-1},$$

where $\upsilon = \sigma_x/\sigma_y$, where σ is the variance of X and Y, respectively; and $u = (\mu_x - \mu_y)/\sqrt{\sigma_x}\sigma_y$, where μ is the mean value of X and Y, respectively, and σ is defined as above.

The coefficient υ defines the scale, or slope, shift (1 = perfect relation between X and Y,and deviation from 1 occurs because bias in Y depends on the magnitude of X), and μ is the location, or height, shift relative to the perfect relation (0 = perfect relation between X and Y,and deviation from 0 occurs when the means of Y and X are not the same). Other symbols are defined as already described. Madden *et al.* (2007) considered C_b to be a *generalized* bias correction factor as it is based on more than just the difference between the means, including measuring the effect of magnitude of X on the estimate (the scale shift) – or where the variance of the estimate varies with X.

Perfect accuracy (yet some loss in precision) is found when the fitted line falls on the line of concordance, but data points are scattered (or imprecise, Figure 3A). A location shift occurs when there is a constant bias in Y across the range of magnitude of X (Figure 3B), and a scale shift occur when there is a systematic bias in the magnitude of Y that is influenced by the magnitude of X (Figure 3C). A perfect relationship is shown when the fitted line and data points fall on the line of concordance (Figure 3D). There are an infinite range of possibilities between these scenarios.

Lin's concordance correlation coefficient has only relatively recently been used to analyze how well disease assessment data relate to actual values or relate to repeat estimates (Nita *et al.*, 2003; Bock *et al.*, 2008a, 2009a, 2009b). Relating estimated to actual disease severity, Nita *et al.* (2003) found that a group of six raters returned ρ_c values of 0.82-0.93 when assessing *Phomopsis* severity on strawberry, showing reasonable to good agreement (Figure 4). There were losses in precision or accuracy due to bias for each of the raters. Similar results were found with citrus canker estimates (Bock *et al.*, 2008a). In this study, Lin's concordance correlation coefficient was also used to characterize repeat estimates by raters and image analysis estimating citrus canker. It demonstrated that repeat measurements of all disease symptoms (lesion number or percent area diseased) were more consistent using image analysis compared to visual estimates by raters. Not only were the image analysis

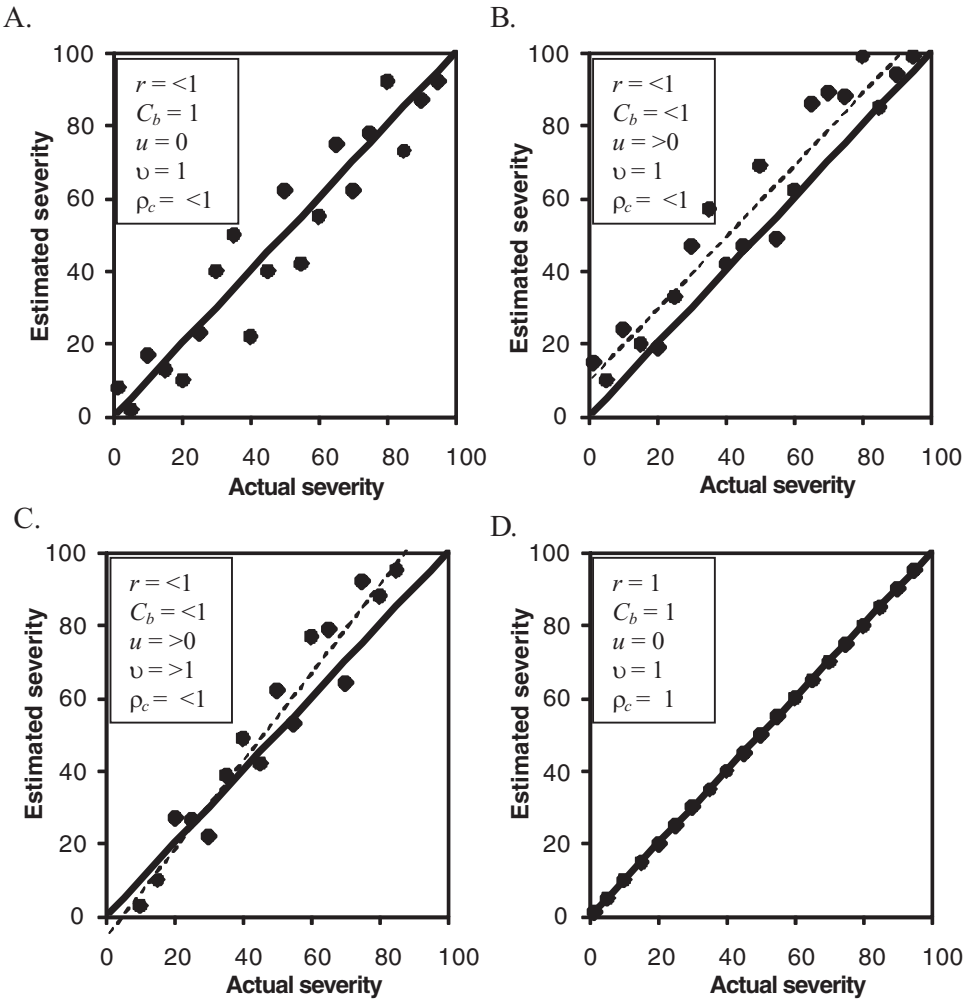

FIG. 3. Lin's concordance correlation coefficient for agreement and the terms used to measure accuracy and precision in relation to an actual value, or for comparing methods, or raters for elucidating characteristics of reproducibility (interrater) and test-retest data (intrarater). A. The fitted line falls perfectly on the line of concordance showing no bias in estimates ($C_b = 1$) due to scale or location shifts, but the measurements are imprecise ($r < 1$); B. The fitted line shows a consistent difference in elevation compared to the concordance line – a location shift ($u > 0$) and a loss in precision ($r < 1$); C. The fitted line shows a scale shift ($\upsilon > 0$), bias (variance) is not consistent with magnitude of X, a location shift ($u > 0$), and data are imprecise (r < 1); D. Shows perfect precision and complete accuracy, and thus complete agreement between X and Y, and no bias due to scale ($\upsilon = 1$) or location shift ($u = 0$) (after Nita *et al.*, 2003 and Madden *et al.*, 2007).

repeat estimates more reliable, they also showed much less bias.

Where data is not on a continuous ratio or interval scale, and comprises ordinal category values agreement can be determined using Cohen's weighted kappa statistic, which is a non-parametric analogue to Lin's concordance correlation coefficient (Madden *et al.*, 2007).

5. *Other Methods to Explore the Quality of the Estimate or Measurement*

There are various other methods that have been used in measuring quality of disease severity assessments. These are very briefly considered.

Bland-Altman plots (Bland and Altman, 1986, 1999) can be used to assess the characteristics of the reproducibility and bias within and among rating methods and provide insight into inter- and intrarater reliability and agreement. They have been widely used in clinical assays to assess accuracy and precision of medical and pharmaceutical tests (Bland and Altman, 1999). The plots show the difference ($estimated - actual$) against the mean value. The plot can readily detect absolute systematic error, proportional error or heterogeneity of variance. The reliability can be tested statistically with the coefficient of repeatability (CR) where:

$$CR = 1.96\sqrt{\frac{\sum (d_2 - d_1)^2}{n - 1}}$$

d_2 and d_1 are the sample data for the two assessments, respectively, and n is the number of leaves assessed. Agreement

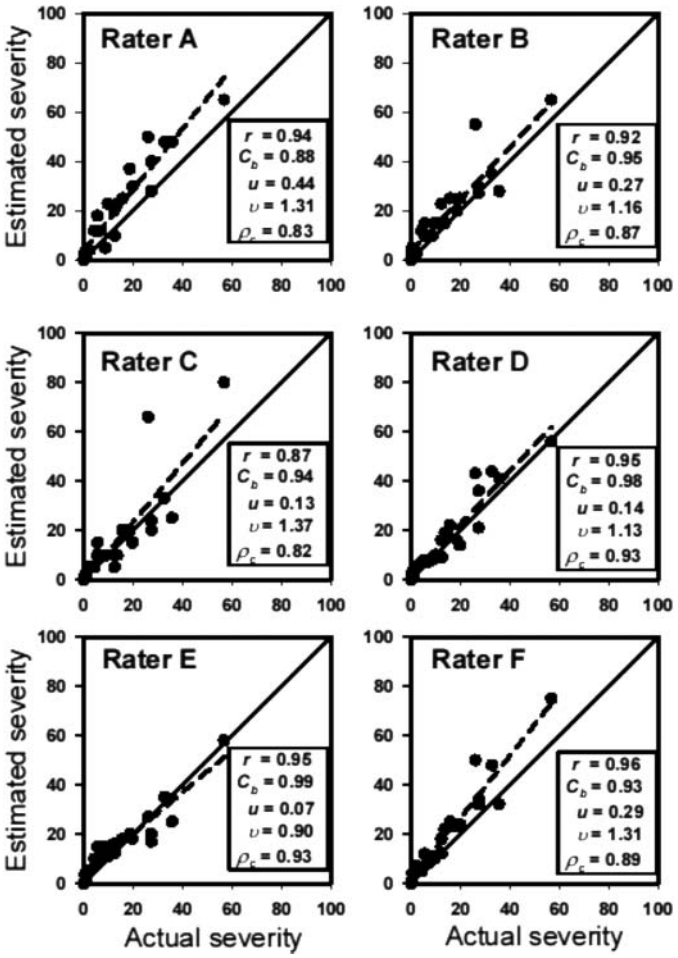

FIG. 4. Lin's concordance correlation analysis of estimated disease severity of *Phomopsis* leaf blight of strawberry, based on direct visual estimation, versus actual severity for six raters. The solid line is the concordance line, representing perfect agreement between actual and estimated severity. The broken line is the best fitting line. Accuracy is determined with Lin's (Lin *et al.*, 1989) concordance correlation coefficient (ρ_c), calculated as the product of the correlation coefficient (r) and the bias correction factor (C_b). C_b is a function of location shift (u) and scale shift (υ), indicating changes in line height and slope, respectively (after Nita *et al.*, 2003).

with actual data can be tested based on the 95% limits of agreement (Bland and Altman, 1999), by taking the mean difference d, and the standard deviation of the difference σ_d, such that the 95% agreement limit is equal to d +/−1.96 × σ_d. Bland-Altman plots have been used to demonstrate trends in rater agreement and reliability in studies of citrus canker assessment (Bock *et al.*, 2008b).

The absolute error (estimate minus actual disease) and relative error (absolute error ÷ actual severity x 100) can be calculated for the mean estimate at each actual severity to describe how actual disease severity is related to inaccuracy and imprecision of the estimate over the range of disease assessed. Hock *et al.* (1992) used absolute error to demonstrate characteristics of individual raters in assessing tar spot of maize, and also showed how mean relative error was influenced by lesion size. Bock *et al.* (2008b) showed that the mean absolute error increased with greater disease severity. Relative error was greatest at low disease severity (up to nearly 600% overestimation).

Frequency plots of disease severity estimates can be made and provide an insight where data are biased or imprecise. They can be useful for exploring trends and showing differences among methods or symptoms (Bock *et al.*, 2008b). The relationships between frequency of rater estimates and actual disease clearly demonstrated the preference raters had for specific values (resulting in clusters of estimates known as "knots"; Koch and Hau, 1980; Hau *et al.*, 1989; Bock *et al.*, 2008b, 2009b). Frequency of discrepancy of the estimate from the actual value has been fit to a normal probability density function for different disease categories and provided a basis for comparing estimation error over different disease ranges (Bock *et al.*, 2008b). The resulting parameters suggested that the difference between the actual value and the estimate increased with actual disease severity.

The methods described above have been used to explore the characteristics, sources of error, accuracy, precision and agreement of disease estimates and measurements using visual rating of various types, image analysis and hyperspectral imagery.

IV. VISUAL ASSESSMENT

A. How the Eye Works

The eye acts as a remote sensing device, and combined with the brain acts as an image analysis system (Nilsson, 1995; de Jong *et al.*, 2006). The way the eye functions and creates an image compares with the way a camera operates. The focusing section is the cornea, which takes widely diverging rays of light and bends them through the pupil (which is surrounded by the iris). The iris and the pupil perform the function of the aperture of a camera. The inbound light then encounters the lens, which helps focus light on the retina at the back of the eye. The retina is coated in photoreceptor nerve cells which convert light rays into electrical impulses and sends them through the optic nerve to the brain, where they are translated and perceived as an image. The sharp vision, as when concentrating on an individual leaf, or reading a book, as now, takes place in the center 10% of the retina, which is called the macula, the rest of the retina being responsible for peripheral vision. The eye and the relevant parts of the brain work together extremely rapidly to acquire the image, analyze and interpret it (Hubel, 1995). This process allows individuals to estimate the type and quantity of disease tissue on a particular leaf or plant. The response of visual raters is not the same every time, and is dependent on several factors, some of which are known. There are individual differences in how light and color are perceived among individuals, and thus in estimating disease – quite apart from the cognitive ability to estimate disease that might vary with individual and with assessment occasion. Conditions such as color blindness have

also been demonstrated to influence ability to estimate disease (Nilsson, 1995).

There have been great strides made in our understanding of factors that contribute to the quality of a visual disease assessment or measurement over the last thirty years and this information has been used to develop ways of improving rater ability, a process which continues (Horsfall, 1945; Chester, 1950; Large, 1966; James, 1971; Horsfall and Cowling, 1978; Berger, 1980; Nutter *et al.*, 2006).

B. Rater Error and Its Ramifications

Large differences can exist between raters assessing the same leaves (or other plant organs) for disease severity, showing rater reliability and agreement with actual values can be variable (Nutter and Schulz, 1995; Parker *et al.*, 1995a and b; Bock *et al.*, 2008a). The diversity and frequent lack of inter- and intrarater reliability and agreement with actual values is caused by several factors that can be identified (Sherwood *et al.*, 1983; Nutter and Schulz, 1995; Bock *et al.*, 2008a, b). The ramifications of inconsistent and mediocre disease assessments can be important. Nonetheless, gross differences in disease severity will most often be detected even by poor estimates, but in certain situations lack of reliability and/or agreement might lead to different outcomes of a statistical analysis. For example, in a fungicide trial the null hypothesis (H_o) would be that there was no effect of fungicide treatment compared to the control in reducing disease severity. If rater disease assessments failed to show an effect of fungicide when there actually is a reduction in disease, then a Type II error would be committed (failure to reject H_0 when H_0 is false, Everitt, 1998). Visual assessment can fail to discern differences among treatments where more objective measures do (Parker *et al.*, 1995a and b). In other words the experiment failed to prove the primary hypothesis (H_1) and demonstrated that visual estimates did not discern significant effects among treatments, whereas the actual disease severities demonstrated significant differences. Furthermore, methods of assessing the disease (or pathogen) can influence epidemiological studies where disease progress might be modeled (Nutter, 1997b; Nutter, 2001), or for discerning sources of disease resistance among germplasm (Jackson *et al.*, 2007).

C. Sources of Error

Visual disease severity estimation is error prone. Since the early years of plant pathology this has been accepted, but it is only recently the understanding of these sources of error have been placed on an empirical footing. Indeed, the first study that quantified error associated with a visual assessment, and accommodated for it was that by Smith *et al.* (1969), who described error in visual estimates of percent area compared to the actual severity on tomato infected with *Cladosporium* leaf mold in the UK. Since that time many sources of error in rater assessment have been identified and quantified in many pathosystems. The many sources and causes of error include:

TABLE 2
Intra- and interrater reliability and accuracy of visual assessments of dollar spot severity on bentgrass. Percent area showing symptoms in $1m^2$ quadrats were estimated. Regression analysis[1] was used to compare estimates among assessments, raters and against actual values (after Nutter *et al.*, 1993).

		Regression parameters		
Rater variable	Scorer	r^2	*Slope*	*y-intercept*
Intrarater reliability	Rater 1	0.93	0.93	0.74
	Rater 2	0.83	0.80	−0.54
	Rater 3	0.84	0.88	0.76
	Rater 4	0.88	0.93	−2.66
Interrater reliability	Rater 1/Rater 2	0.77	0.82	6.95
	Rater 1/Rater 3	0.77	0.74	2.33
	Rater 1/Rater 4	0.89	0.74	2.98
	Rater 2/Rater 3	0.70	0.76	−0.42
	Rater 2/Rater 4	0.80	0.77	0.60
	Rater 3/Rater 4	0.86	0.88	6.01
Accuracy	Rater 1	0.83	1.4	4.7
	Rater 2	0.79	1.0	19.6
	Rater 3	0.96	1.1	1.3
	Rater 4	0.97	0.9	12.9

[1]Linear regression analysis ($y = a + bx$) was used to compare estimates to the actual values. The coefficient of determination (r^2) indicates the precision of the estimates, and the slope and *y*-intercept indicate bias.

1. *Individuals Vary in Their Intrinsic Ability*

Various studies with different pathosystems have shown that individuals vary in ability. Measures of agreement, inter- and intrarater reliability all point to differences among individuals, and individual variability between assessments (Amanat, 1976, 1977; Sherwood *et al.*, 1983; Hau *et al.*, 1989; Nutter *et al.*, 1993; Newton and Hackett, 1994; Guan and Nutter, 2003; Nita *et al.*, 2003; Bock *et al.*, 2008a and 2009b). Based on regression analysis, Nutter *et al.* (1993) assessed dollar spot (caused by *Sclerotinia homoeocarpa*) on bent grass *Agrostis palustris* and showed both inter- and intra rater reliability was variable between and among raters, and agreement with actual values varied among individuals (Table 2). Sherwood *et al.* (1983) found that individuals varied among and between both experienced and inexperienced groups of raters in ability to rate symptoms caused by *Stagonospora arenaria* on orchard grass. Newton and Hackett (1994) also found inexperienced and experienced raters varied in ability to assess *Erysiphe* on barley; experienced raters tended to be better at estimating disease and responded less to training using disease assessment training programs. There are some common patterns in error reported. Nutter and Schulz (1995) showed agreement of 80 raters with actual values was variable using computer images of diseased leaves, and the raters

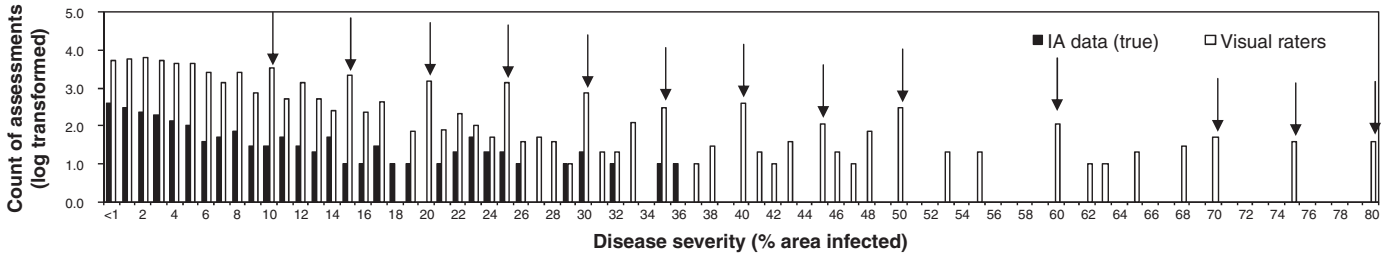

FIG. 5. The frequency of "knots" or preferred values with twenty-eight raters assessing 200 leaves for the percent area infected (area necrotic+chlorotic) by citrus canker. Arrows mark knots where raters appear to prefer specific values when estimating area at most severities greater than 10% (after Bock *et al.*, 2008b).

fell into three identifiable categories – among these raters the most common perception was a V shaped pattern, with absolute error greater and more negative between 25 and 35% severity, another group underestimated disease more with increasing actual disease severity, and a third group showed absolute error at its highest at 50% disease. Other patterns exist – for example, a tendency to overestimate disease increasingly at higher disease severity (Newton and Hackett, 1994; Bock *et al.*, 2008a and 2009b). Bock *et al.* (2008a) found reliability varied between first and second assessments, and individual raters were not constant in the direction or magnitude of their error. Guan and Nutter (2003) found that reliability of intrarater assessment estimates for four raters was poor (as measured by the coefficient of determination, $r^2 = 0.16\text{-}0.95$), and Bock *et al.* (2008a) found intrarater reliability of estimates by three plant pathologists was somewhat variable (correlation coefficient, $r = 0.86\text{-}0.89$). The raters showed reasonable agreement with actual values measured with Lin's concordance correlation coefficient ($\rho_c = 0.82\text{-}0.90$).

2. *Value Preferences by Raters*

Koch and Hau (1980) and Hau *et al.*, (1989) provided data that showed individuals appeared to prefer specific values when estimating disease – they more frequently chose these values compared to what would be expected if estimation was random around the actual disease. These "knots" as they are known, have been observed in other studies, and were observed in direct estimation of citrus canker severity (Bock *et al.*, 2008b, Figure 5), and in estimates of *Erisyphe* infection on barley where there was a tendency to assess severity with values close to that of standard area diagrams (SADs) being used (Parker *et al.*, 1995b). Thus there is some evidence of raters choosing "preferred values," but this tendency has not been fully explored.

3. *Lesion Number and Size Relative to Area Infected*

Amanat (1976) found that areas under large lesions were estimated with less error than small lesions of the same area, a result also reported by Hau *et al.* (1989). Working with the *Stagonospora*/orchard grass pathosystem, Sherwood *et al.* (1983) demonstrated error due to lesion numbers – overestimation was inversely proportional to the natural logarithm of the

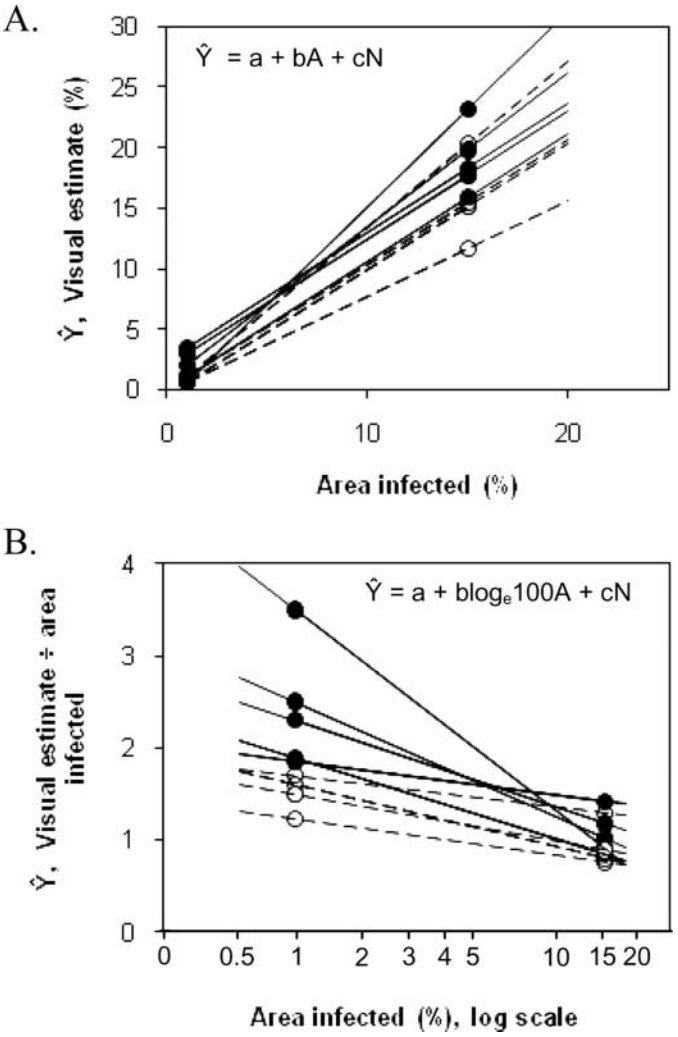

FIG. 6. A. The relationship between rater (Y) estimates of area and actual diseased area (A), lines drawn showed that number of lesions (N) was a significant factor for most raters. B. The relationship between estimated area:actual area symptomatic (Y) vs. natural log of the actual area infected (A) and show an effect of lesion number (N) for many raters. Raters from two different groups of experienced scorers, one routinely used SADs (solid lines) and the second used various other systems (dashed lines). Points drawn assuming number of spots (N) = 15 at 1% A and N =75 at 15% A (after Sherwood *et al.*, 1983).

disease area, and for most raters directly proportional to the number of lesions (Figure 6A and B). A similar relationship exists with citrus canker on grapefruit (Bock *et al.*, 2008b) and with *Phomopsis* on strawberry (Nita *et al.*, 2003). Forbes and Jeger (1987) found that estimation of severity due to fewer larger lesions were less error prone compared to small, random or uniformly distributed lesions, and Godoy *et al.* (1997) observed rater overestimation of bean rust due to large numbers of small lesions, confirming these illusions in visual assessment. Hock *et al.* (1992) also found the greatest error among raters occurred when assessing lower disease severity with lots of small lesions with tarspot complex on maize. In developing SADs to assess soybean rust (*Phakopsora pachyrhizi*), Godoy *et al.* (2006) found such overestimation was common among raters. The overestimation of disease when there are many small lesions, particularly at low severities, has tremendous ramifications for epidemiological studies when projecting yield loss, where disease progress is based on estimates of disease severity; moreover, rater bias that overestimates actual disease severity may adversely affect genetic advance in plant breeding programs as well (Sherwood *et al.*, 1983). Why do raters overestimate at low disease severity? Relative error is greatest in this range (Bock *et al.*, 2008b) and it is a common disease range for many assessments (<10% area). It is not possible to estimate disease less than zero, and thus at low disease severity (at least up to 5-10%) there is an invisible "barrier" (i.e., 0%) to underestimation of disease, yet no "barrier" to overestimation. The cause of overestimation does not appear to have been fully explored, although is now recognized (Amanat, 1976; Sherwood *et al.*, 1983; Beresford and Royle, 1991; Bock *et al.*, 2008b and 2009a).

4. Actual Disease Severity

On average, do raters have closer agreement to the actual value at low disease severity compared to mid-range disease severity? The last thirty years has provided much data, but this issue remains unresolved, partly because the manner in which different studies are done (direct estimates, using SADs, individual estimates or averaging estimates) could influence estimation error. Experiments have been done that show non-constancy of error with actual disease severity (Kranz, 1970; Forbes and Jeger, 1987; Hau *et al.*, 1989; Forbes and Korva, 1994; Bock *et al.*, 2008b and 2009b), but there are also studies that indicate more constant error (Nita *et al.*, 2003; Nutter and Esker, 2006). Forbes and Jeger (1987) found disease intensity was a significant factor, with greater disease severity being more error prone, with a maximum error around 25%. Koch and Hau (1980) found that estimates of low disease severity had a much narrower range compared to estimates of high severity, and Hau *et al.* (1989) similarly reported the standard deviation of the mean of estimates for 200 scores assessing leaf models of known disease and showed that it was less at low and high levels of severity compared to the mid range. Bock *et al.* (2008b and 2009a) found that the variance of multirater direct, unaided estimates of the actual disease increased with actual disease severity of citrus canker on leaves of grapefruit – at least over the disease severity range assessed (0-35%) and Andrade *et al.*, (2005) found that error associated with estimation of leaf spot (*Quambalaria eucalypti*) of Eucalyptus was greatest from 15-30%. However, Nita *et al.* (2003) did not find estimation error was strongly associated with actual disease severity when comparing means of six replicate estimates of *Phomopsis* on strawberry assessed on five occasions. Comparing the estimation of 25, 37.5 and 50% disease, Nutter and Esker (2006) used the method of comparison stimuli to estimate the "just noticeable difference" (jnd) between a reference stimulus (known) and a comparison, which is rated as the same or different to the reference. In this way they showed the jnd divided by the reference stimulus was a constant fraction (Weber's law) and ability to discern differences in disease severity was better than suggested by the Horsfall-Barratt scale (Horsfall and Barratt, 1945). The importance of actual disease severity to magnitude of error remains incompletely characterized and doubtless related to ability of the assessor, training and the use of SADs to assess disease (Newton and Hackett, 1994; Nutter and Schulz, 1995; Nutter and Esker, 2001; Nita *et al.* 2003; Bock *et al.*, 2008b). The characteristics of error estimation can depend on whether the actual disease is compared with an individual's estimates, raters as a group, or the mean values. Thus studies where multiple individual rater estimates were used demonstrate increase in error with actual disease severity (Koch and Hau, 1980; Forbes and Jeger, 1987; Hau *et al.*, 1989; Bock et al, 2008b, 2009b). Some data based on well-trained or guided individuals, or on mean values might be less likely to show increased estimation error or heterogeneity of variance due to actual disease severity (Nita *et al.*, 2003). It is important to characterize all these approaches as assessments might be made by multiple assessors, and averaging data might affect the error. Furthermore, if the mean, or variance of the mean estimates deviates from the actual values, a t-test or $F-$test could result in a Type II error (Parker *et al.*, 1995a and b; Bock *et al.*, 2009c).

Relative error, as opposed to absolute error, tends to decrease with increasing disease (Kranz,1970; Hock *et al.*, 1992; Bock *et al.*, 2009a). Beresford and Royle (1991) found that wheat rust severity estimates at low levels were less accurate, and that overestimation was up to 8.7 times higher than disease severity based on uredinium spore counts. Furthermore, Kranz (1970, 1988) and Parker *et al.* (1995a) found greatest relative error at low severity. These studies suggest non-constant error and that the nature of the relationship has not been fully characterized, although it is unlikely to be the logarithmic (y) – linear (x) advanced by Horsfall and Barratt (Horsfall, 1945; Horsfall and Barratt, 1945; Forbes and Korva, 1994; Nita *et al.*, 2003; Nutter and Esker, 2006; Bock *et al.*, 2008a and 2009b).

5. Plant Structure and Size

The plant structure being assessed can influence the quality of the estimate of severity. This is somewhat peripheral as this review is more concerned with the process of assessment of diseased area, rather than comparing sampling units on plants.

Nonetheless, it bears mentioning as a potential source of error among units (Townsend and Heuberger, 1943; Forbes and Jeger, 1987; Shokes *et al.*, 1987; Christ, 1991; Vereijssen *et al.*, 2003; Danielsen and Munk, 2004;). Forbes and Jeger (1987) compared assessments and found significant effects of plant structure with mean absolute error ranging from 0.3 to 15.9% among the various plant structures assessed for disease (compound leaves, simple leaves, stems, panicles, pods, tubers, rosette heads and roots). Also, illustrating the importance of organ assessed in relation to disease severity, Tinline *et al.* (1994) used a disease severity scale to assess common root rot (caused by *Cochliobolus sativus*) on varieties of wheat, and found that disease rating of lesions on the subcrown internodes was better able to differentiate cultivar response compared to the assessment of lesions on the crowns. Furthermore, how the disease is manifested on plant parts and the ease of assessment might influences rater error differentially, but has not been tested across all plant part types. The size of the unit (e.g., leaf size) on which the estimate has been made has rarely been addressed. Nita et al (2003) investigated the role of leaf size and found only a small, but significant effect of leaflet size on severity estimation o*f Phomopsis* on strawberry.

6. Time Taken to Assess Disease

Parker *et al.* (1995b) found that the time taken by assessors to rate disease affected precision - although fast assessments were not always less precise, they were more often less precise compared to assessments performed slowly. There is little other literature on the amount of time taken to assess disease severity, although on a leaf-by-leaf basis it seems that direct visual estimates of severity take on average approximately 7 secs. (Martin and Rybicki, 1998; Bock *et al.*, 2009a). Nutter *et al.* (1993) found that it took raters approximately 32 min. to assess severity of *Stagonospora* on 80 x 1 m^2 quadrats of bentgrass.

7. Color Blindness

Individuals who suffer from color blindness are impaired in their ability to assess disease. The more severe the red/green colorblindness, the poorer the estimation of disease severity on golf greens (Nilsson, 1995).

8. Complexity of Symptoms and Timing

Apart from lesion size and number (see section IV.C.3), deciding where the symptoms start and end can be a source of subjectivity. Estimates of the necrotic areas of citrus canker might be estimated less well than the total symptomatic area comprising both chorotic and necrotic areas (Bock *et al.*, 2008a, 2009b). Parker *et al.* (1995a) commented that in the *Septoria tritic*/wheat pathosystem, discerning between disease necrosis and natural senescence might be the source of some of the error, although there is evidence those pathogens that appear to be well defined are not necessarily any better assessed (Parker *et al.*, 1995a). The timing of the disease assessment might influence the result, demonstrating the importance of multiple disease assessments during the season. Timing of assessments needs to be made sensitive to the likely rate of development of symptoms. Thus knowledge of the latent periods and timing of infection in relation to crop growth are useful. Other diseases are difficult to rate for severity due to the amorphous way in which symptoms develop, including certain virus diseases which can present a challenge for how best to assess them (Madden *et al.*, 2007). The many ways of rating these diseases are considered in a later section. Some diseases, like huanglongbing of citrus (caused by the bacterium *Candidus liberibacter*) (Gottwald *et al.*, 2007) do not develop symptoms for up to several months after infection, and although infection generally results in death of the plant, rating the progress of the disease requires a more qualitative approach to assessment. Sometimes molecular approaches can be used to check for, or quantify the pathogen in a host population if symptoms is unreliable. The assessment of pathogen "severity" is becoming increasingly feasible (Nutter *et al.*, 2006). With other diseases (particularly some that are systemic), severity can equate to incidence (Seem, 1984; Bock and Jeger, 1996).

9. Interactions among Multiple Factors

Various studies have indicated interactions among multiple factors, but these have rarely been explored in any detail. In an analysis of variance among various factors, Sherwood *et al.* (1983) and Bock *et al.* (2008b) found interactions among lesion number, actual area, rater group and individual rater. Nita *et al.* (2003) also found multiple factors influencing estimates of disease including actual disease, lesion number and leaflet size, although the effects of the latter two were minor. Finally, it should be mentioned that the sampling design and sample size are also critical to obtaining an unbiased estimate of disease (Steddom *et al.*, 2005a; Madden *et al.*, 2007).

D. Methods Used to Visually Estimate Disease Severity

Chester (1950) summarizes the development of disease assessment techniques up to 1950, and the understanding and development of methodology involved in the nascent field of plant disease severity assessment. Large (1966), James (1974), Cooke (2006) Nutter *et al.* (2006), and Madden et al (2007) provide information on various aspects of visual disease assessment since that time. The importance of accurate and precise disease assessment was recognized early. The American Phytopathological Society (APS) established a committee in 1917 (Anon, 1917), which met for a number of years. There were few advances at that time (Chester, 1950), beyond modification to the Cobb scale (Cobb, 1892) for cereal disease assessment. Some years later (1933) the British Mycological Society (BMS) established a similar committee to investigate ways to record and measure disease prevalence and intensity (Beaumont *et al.*, 1933; Moore, 1943, Large, 1955). The BMS committee developed some new key-based scales for various crops, which were deemed useful to plant disease assessment, particularly survey work, most notably a key to assess late blight of potato (Anon., 1947),

which bore some superficial similarity to the Horsfall-Barratt scale (Horsfall and Cowling, 1978). In 1967 the FAO held a conference on crop loss assessment methods (FAO, 1971), and described sampling procedures and assessment methods for various crops. However, there was no empirical test performed to quantify the visual assessment methods for reliability or agreement compared to the actual disease severity in the field, or the ramifications of using scale-based data in analyses. Since the 1970s our understanding of visual disease assessment has taken on a more empirical footing, and the pros and cons of different assessment methods are becoming better elucidated.

A consideration is warranted of the various methods used to visually assess disease severity, both from an historic perspective and to gauge their effectiveness, application and scientific justification today. There are different types of rating scales that comprise continuous or discrete variables (Sheskin, 1997) to measure disease severity. Some strive to be generic, and others are specific to individual host-pathogen systems (Chester, 1950; Campbell and Madden, 1990; Nutter and Esker, 2006; Madden *et al.*, 2007). The various kinds of scales used in visual plant disease assessment include:

- Nominal or descriptive scales.
- Ordinal rating scale.
- Interval (category) scales (with or without standard area diagrams, SADs) and field keys.
- Ratio scales (with or without SADs).

1. Nominal or Descriptive Scales

These are the simplest and probably most subjective of the scales for grading disease severity (Chester, 1950; Nutter and Esker, 2006). Disease is graded into a number of classes with descriptive terms such as "slight," "moderate," or "severe." Due to the subjectivity involved and lack of quantitative definition, these scales have very limited value beyond the individual performing the rating in a particular season and location.

2. Ordinal Scales

These are descriptive disease scales, but they grade disease severity into arbitrary classes that represent increasing severity of symptoms. Newell and Tysdal (1945) used such a scale to measure alfalfa disease severity based on the subjective scale of 1 = very little disease, 5 = medium disease and 9 = very much disease. Very simple ordinal scales of this type are subjective and not particularly transferrable between raters, locations or seasons (Chester, 1950). Ordinal scales can be based on, or accompanied by, diagrams or descriptions indicating the intensity of symptoms. One such example is the ordinal scale developed to assess symptoms of southern corn leaf blight (*Helminthosporium turcicum*) on corn that was developed in the 1940s and related directly to a set of diagrams to aid estimation of the descriptive severity (Ullstrup *et al.*, 1945). Another example is the nine-point ordinal scale developed to measure severity of fungal pathogens (*Puccinia arachidis* and *Cercosporidium personatum*) on peanut (Subrahmanyam *et al.*, 1982). The scale relied on descriptions of the symptoms and was used to assess genotypes for resistance in the breeding program at the International Crops Research Institute for the Semi-Arid Tropics.

TABLE 3
A 0-5 scale used to assess the severity of symptoms of zucchini yellow mosaic virus and watermelon mosaic virus on water melon (after Xu *et al.*, 2004).

0	no symptoms;
1	slightly mosaic on leaves;
2	mosaic patches and/or necrotic spots on leaves;
3	leaves near apical meristem deformed slightly, yellow, and reduced in size;
4	apical meristem with mosaic and deformation; and
5	extensive mosaic and serious deformation of leaves, (or plant dead).

Ordinal scales are still quite widely used for specific diseases, particularly for rating some virus diseases where symptoms are not easy to measure quantitatively (Madden *et al.*, 2007). Allocating a number to a symptom description allows the stage of development of the disease to be numerically assessed or denoted. For example, zucchini yellow mosaic virus and watermelon mosaic virus on water melon have been rated on a 0–5 scale (Xu *et al.*, 2004, Table 3). There is inevitably still a risk of subjectivity in these scales, and there are some situations where more quantitative alternatives are difficult to devise. There are some advantages to ordinal rating scales particularly that they often describe the disease development as the symptoms become increasingly severe and thus can be interpreted by the rater. They are also easy to learn and use, albeit somewhat subjectively, and provide a rapid way for assessing large numbers of plants, as might sometimes be the need in plant breeding programs. Standard sets of illustrations that show the symptom characteristics of each category allow more objective assessment, but because the ratings values are arbitrary they have less interrater, inter-site or temporal consistency. Because of the nature of ordinal data it should be analyzed using non-parametric tests, which are becoming more sophisticated and powerful for these types of data (Shah and Madden, 2004; Madden *et al.*, 2007), although some other options are sometimes applied to allow parametric testing, including converting the ordinal scale to a disease severity index (Chaube and Singh, 1991). The calculation of the disease severity index is described in section IV.G.

3. Interval or Category Scales

A disease interval (or category) scale comprises a number of categories where the numeric values are known – in the case of plant disease this is generally the percent area with symptoms. The first interval scale developed and used was the Cobb scale (Cobb, 1892) to assess severity of rusts on wheat in Australia.

The scale included a standard area diagram (SAD) set that had five levels of rust 1–5 (representing 1, 5, 10, 20 and 50% disease), and the rater placed the sample leaf in the most appropriate category. This diagrammatic scale was modified over the years to improve accuracy (Melchers and Parker, 1922; Peterson *et al.*, 1948). Like the Cobb scale, other disease assessment interval scales are often accompanied by SADs (Chester, 1950; Large, 1966). The development of these scales was spurred on by the perception that descriptive scales were less-than-ideal, indeed McKinney (1923) and Horsfall and Heuberger (1942) developed early interval scales to assess disease severity in a more quantitative way (Table 4). Disease scales were, and still are widely perceived as a way to save time in disease assessment, assumed to be quick to learn and thus allot a score, and in some cases control perceived error in disease assessment (Horsfall and Barratt, 1945). Some interval scales can be analyzed directly, but they are often converted to a disease index, providing a continuous variable which is amenable to parametric statistical analysis (Marsh *et al.*, 1937; Horsfall and Heuberger, 1942; Horsfall and Barratt, 1945; Chester, 1950; Chaube and Singh, 1991; Shah and Madden, 2004). Interval scales are still widely used (Bruton *et al.*, 2000; Harveson and Rush, 2002). For example, Tinline *et al*. (1979; 1994) developed a scale to ascribe disease severity on wheat and barley. Plants were assigned a severity rating from a 4-category scale where 0 = healthy, 1-25% = slight, 26-50% moderate and >50% severe, with values of 0, 1, 2 and 5 being applied to wheat and 0, 2, 5 and 10 being assigned to barley for the four categories.

The intervals of some disease assessment scales are unequal and bear similarity to the well-known Horsfall and Barratt (H-B) scale (Horsfall and Barratt, 1945). Horsfall and Barratt (1945), Horsfall (1945) and Horsfall and Cowling (1978) noted that many rating scales resulted in high mean scores, especially where actual disease severity might be at the lower end of a category, and because grades were so broad it was difficult to show differences, particularly at these low or high levels of disease. They also believed it did not reflect innate human ability to distinguish disease severity based on their "discovery" of the so-called Weber-Fechner Law which was described thus ". . . the Weber-Fechner law which states that visual acuity depends on the logarithm of the intensity of the stimulus." (Horsfall, 1945).

TABLE 4
Horsfall and Heuberger (1942) developed early interval scales to assess disease severity in a more quantitative way.

Category	Severity
0	apparently infection-free;
1	trace-25% leaf area infected;
2	26-50% leaf area infected;
3	51-75% leaf area infected;
4	>75% leaf area infected.

To this end Horsfall and Barratt (1945) developed a new category scale that they believed addressed these issues. This scale became known as the H-B scale. It is a widely applied scale and has been used in several disciplines beyond plant pathology and cited at least 571 times (Web of Science, 2009). Indeed, the H-B scale, or modified versions of the H-B scale are probably the most widely used rating systems over the last 70 years, and it was formulated on a very limited empirical base. Because the wide usage of the H-B scale, it is worth spending time examining it in some detail. It has 12 categories of unequal, logarithmic structure (0–100% disease) symmetrical around 50%. The symmetry around 50% was explained by Horsfall and Barratt's (1945) realization that the eye reads diseased tissue below 50% severity, and healthy tissue above 50% severity (Table 5). Horsfall and Barratt (1945) and Horsfall (1945) stipulated the need for a large sample in disease estimation using this method (preferably a sample size of at least 20). Analysis of data from these unequal-interval scales had to be approached in a unique way. Initially a chart graphed on semi-log paper was used to relate mean disease by category on the X-axis to mean estimated % on the Y-axis (Horsfall, 1945; Horsfall and Barratt, 1945). However, it was realized that recalibration after taking the arithmetic mean of the grades lead to unacceptable bias (Redman *et al.*, 1968), and recalibration of each allotted category to the interval mid-point was required prior to taking the mean, and analysis; eventually a set of tables was developed specifically for this purpose (Redman *et al.*, 1968). Mid-point recalibration provided data acceptable for analysis using parametric approaches. But it also appears to reinstate variance the scale sought to remove in the first place (Bock *et al.*, 2009d). Furthermore, the H-B scale was first described with categories of 1–12 (Horsfall and Barratt, 1945), but was shown on a recalibration chart to

TABLE 5
The Horsfall and Barratt interval scale showing the disease severity ranges, midpoints and interval sizes (Horsfall and Barratt, 1945).

H-B category	Disease severity range	Midpoint	Interval size
1	0	0	0
2	0^{+}–3	2.34	3
3	3^{+}–6	4.69	3
4	6^{+}–12	9.38	6
5	12^{+}–25	18.75	13
6	25^{+}–50	37.50	25
7	50^{+}–75	62.50	25
8	75^{+}–87	81.25	13
9	87^{+}–94	90.62	6
10	94^{+}–97	95.31	3
11	97^{+}–100	97.66	3
12	100	100	0

be a 0-11 scale (Horsfall, 1945; Horsfall and Cowling, 1978). Although the calibration curve is defunct, it is important to be clear which version of the scale is being used as it could affect the calibration (choice of mid-point value).

Since the development of the H-B scale it was revealed the Weber-Fechner law *per se* never existed. It comprises two laws, Weber's law and Fechner's law (Baird and Noma, 1978; Nutter and Esker, 2006). Unfortunately, Horsfall and Barratt were neither explicit about where they obtained information on this law (no citation was ever given), nor what their understanding and interpretation of it really was. Weber's law actually states that the physical size of a just noticeable difference is a constant proportion of the value of the standard for a given dimension; and Fechner's law, which assumes Weber's law, adds that the subjective value is a logarithmic function of the physical value (Baird and Norma, 1978; Birnbaum, 1994; Nutter and Esker, 2006). Thus, the H-B scale was presumably designed to take into account a perceived (but untested at the time) hypothesis that estimated disease (perceived disease) was in some way logarithmically related to actual disease.

Disease severity estimation falls within the realm of psychophysics, which is the study of the relationships between physical stimuli and their subjective correlates, or percepts. The logarithmic assumptions described by the Fechner law have been questioned, and this has cast the rationale for the H-B scale in doubt (Stevens, 1961; Hebert, 1982, Nutter and Schulz, 1995; Nutter and Esker, 2006). There are several ways in which estimates and the actual magnitude of a stimulus relate to each other and include linear, power function or perhaps logarithmic relationships for some kinds of stimuli (Stevens and Galanter, 1957; Baird and Norma 1978; Birnbaum, 1994). Indeed, there remains some controversy regarding the understanding and applicability of different functions to various stimuli, and the possibility of a unifying psychophysical law that incorporates various relationships (including Fechner's) has been postulated (Kreuger, 1989; Norwich and Wong, 1997). Various methods exist to test perception of stimuli (Stevens and Galanter, 1957; Baird and Norma 1978; Birnbaum, 1994; Ehrenstein and Ehrenstein, 1999; Nutter and Esker, 2006) and all evidence to date shows a linear relationship between visually estimated and actual disease severity (Sherwood *et al.*, 1983; Forbes and Korva, 1994; Nita *et al.*, 2003; Nutter and Esker, 2006; Bock *et al.*, 2008a).

However, the nature of the relationship between the error (or variance) of the estimate and the actual disease has not been completely established. As already noted, individuals vary in the characteristics of their ability to estimate disease, and this can include varying degrees of under and/or overestimation and is affected by the use of SADs, training and the innate ability of the rater, and the symptoms themselves (Sherwood *et al.*, 1983; Hock *et al.*, 1992; Nutter and Schulz, 1995; Bock *et al.*, 2008a). Some results indicate non-constancy of variance associated with increasing magnitude of actual disease, at least up to 25% severity (Redman and Brown, 1964; Smith *et al.*, 1969; Koch and Hau, 1980; Forbes and Jeger, 1987; Hau *et al.*, 1989; Bock *et al.*, 2008b and 2009a). It should be stressed that this relationship is not necessarily logarithmic, or over the complete range of disease severity. Others have found little effect of magnitude of actual disease on estimation error (Nita *et al.*, 2003; Nutter and Esker, 2006). Apart from contrasting results, differences in how the studies were conducted, and the data processed and analyzed might affect conclusions (including use of individual, multiple rater, or averaged data). Thus the relationship between estimation error (both relative and absolute error) and actual disease remains to be fully explored and explained under different conditions.

At least four studies have indicated that there is nothing to be gained in accuracy or precision by applying the H-B scale compared to direct estimation of disease severity, and in some cases it might be detrimental (Forbes and Korva, 1994; Nita *et al.*, 2003, Nutter and Esker, 2006; Bock *et al.*, 2009c). Although use of the scale itself might standardize variance across a range of actual disease, it has to be reconverted to percent midpoints for analysis and thus transforming the raw direct estimate of percent area could achieve the same end (Forbes and Korva, 1994). And it remains to be demonstrated whether applying a 12-point scale with unequal categories like the H-B scale to estimate disease severity of individual leaves or plants is really faster than direct percent estimation – especially considering the unequal categories in the H-B scale, of which assessors must be aware, and the possibility that assessors might unconsciously linearize the intervals when using it directly in the field (Forbes and Korva, 1994). Considering how fundamental disease assessment is to plant pathology, it is surprising these relationships between assessment methods are not more fully understood. Furthermore, in recent studies, (Bock *et al.* 2009d) simulated the disease assessment process and hypothesis testing using data from both direct estimation to the nearest per cent and the H-B scale, and the results suggested that use of the H-B scale resulted (in some situations) in a failure to reject the null hypothesis, (H_0), when H_0 was false (a type II error). A type II error leads to incorrect conclusions and/or actions as the statistical test fails to detect a significant difference, when in fact there is a difference.

Berger (1980) discusses various ways to assess disease intensity, and the importance of good quality estimates to ensure projections of yield loss and epidemic development, and addresses particular problems regarding estimates of a mid-range severity and very low disease severity using the H-B scale. Horsfall and Barratt (1945; Horsfall and Cowling 1978) stipulated multiple readings were required per plot (or plant) which were to be averaged, and never advocated a single reading. Thus, a concern Berger (1980) raised regarding ascribing a single category to a sample that has, for example, severity between 37.5 and 50% (6 on the H-B scale) is somewhat moot, as sequential samples taken from multiple plants and averaged will almost invariably result in a composite estimate of mean disease, and could rest anywhere on the percent scale as deemed by the range of category mid-point values averaged from the sample—it is unlikely all leaves in the sample will have the same disease severity

category (Redman *et al.*, 1968; Bock *et al.*, 2009c). Secondly, Berger (1980) commented that plots with actual disease of 0.0001% or 0.01% disease would receive a rating of "2." Although a rating of "2" might be given to an individual reading, again, Horsfall and Barratt state the need for large sample size. Thus if severity is very low, the majority of the estimates in the sample (leaves or plants) will likely be healthy (0%), and thus on average the estimate of disease would be low as well. Furthermore, although limits have not been tested, most raters will probably not be able to resolve disease severity much less than 0.1% on average-size leaves, and the larger the sample the more realistic the mean estimate of disease is likely to be (Horsfall and Cowling, 1978), assuming there is not a further source of error causing over or underestimation. However, it is important to note that any error in estimates of disease can have dramatic effects on estimates of future disease progress (Berger, 1980).

There is the issue that Horsfall and Barratt set out to address a perceived generic source of rater error in disease assessment by applying a logarithmic scale. The problem with this "one size fits all" approach is that not all raters demonstrate the same characteristics of error in assessing disease severity. Some over estimate, others underestimate, and yet others might be remarkably accurate and precise across the range of severity; furthermore, ability changes with each assessment, and probably with the symptom characteristics of the pathosystem (Nutter *et al.*, 1993; Nutter and Schulz, 1995; Nita *et al.*, 2003; Bock *et al.*, 2008a). With so much inter- and intrarater variation it seems unlikely a scale adapted to a single type of error is going to be useful to all raters, particularly not to those who do not demonstrate that error pattern. Nonetheless, based on data where severity estimates to the nearest percent were converted to H-B scale mid-points, the effect on the estimate of the mean values appeared to be negligible, although the H-B scale might result in greater variability of the mean estimate (Nita *et al.*, 2003; Bock *et al.*, 2009c and d).

It is of interest to note that the logarithmic nature of the H-B scale echoed the logarithmic nature of several other scales and field keys being developed earlier or around that time (Rusakov, 1927; Anon., 1947; Chester, 1950; Horsfall and Cowling, 1978). The Weber-Fechner law and logarithmic type intervals remain the stated basis for developing many of the SADs in use today (Hock *et al.*, 1992; Godoy *et al.*, 1997; Leite and Amorin, 2002; Gomes *et al.*, 2004; Andrade *et al.*, 2005; Belasque *et al.*, 2005; Godoy *et al.*, 2006).Whether there is an underlying psychophysical cause for the early similarity in scale structure, or just coincidence is not known (Campbell and Madden, 1990), but the H-B category scale, and its derivatives, are still in use. Indeed, disciplines other than plant pathology have adopted and use the H-B scale for estimating area (Bergh, 2001; Bussotti *et al.*, 2003; Copes *et al.*, 2003; Pernezney *et al.*, 2003).

Interval scales have some advantages. They are purportedly quick. Applied at a field or plot level they are allegedly more rapid than a percent estimate. They continue to have a role in plant disease assessment and will doubtless continue to be widely used for various applications, and as we understand how imprecision, inaccuracy and rater error relate to the characteristics of the mean estimate, perhaps scales can be constructed that minimize inaccuracy and imprecision, and yet are sufficiently sensitive to the range of disease severity likely to be encountered in the field, and rapid enough to apply for the task in hand. Nita *et al.*, (2003) showed that a category scale based on 5% increments was more precise and accurate compared to the H-B scale for assessment of *Phomopsis* on strawberry. Furthermore, equal-sized category scales can be treated as continuous data. The number of categories that a severity scale should have is discussed by Kranz (1970; 1988) and Hau *et al.*, (1989) in relation to scales with unequal-sized categories. Kranz (1970) believed that about seven classes were optimal, and observed that the ability of inexperienced raters to categorize disease correctly was poor based on results by Amanat (1977). Inexperienced assessors were asked to rate disease and only 15.8% of the estimates were placed in the correct category. Furthermore, the underlying distribution of severity on infected leaves has ramifications for the number of categories needed in a scale to accurately estimate the mean disease severity (Hau *et al.*, 1989).

Regarding the accuracy and precision of the estimate, the empirical data suggest that category-based scales are no better than direct estimation for assessing severity (Forbes and Korva, 1994; Nita *et al.*, 2003, Nutter and Esker, 2006; Bock *et al.*, 2009c). In situations where time is critical and it is appropriate to use a category scale, they should be designed with particular sensitivity to disease at the low end of the severity spectrum to prevent potential overestimation (Sherwood *et al.*, 1983). Categories of 0.1, 0.5, 1.0 and 5.0% might fulfill this requirement. The range of categories above this level should probably be based on a linear scale (5% or 10%) with raters offering their best estimate, preferably aided by SADs (Nita *et al.*, 2003; Nutter and Esker, 2006), taking a suitable sample size, so as to generate an accurate mean (other sources of bias notwithstanding). The maximum likely severity to be encountered should be considered (Kranz, 1977), and the scale set accordingly. Too many divisions and the assumed advantages of speed and simplicity offered by the scale will be lost. There remains an argument for using scales for extensive, rapid surveys of disease severity at plot, field or regional scales, and for recording disease in large field experiments or where large numbers of breeding lines are being assessed for disease resistance. However, a comparison of visual rating methods for time saving has yet to be demonstrated. It is the decision of the scientist to make an informed choice when adopting a scale to measure disease severity, depending on the aims of the study and the desired level of accuracy and precision.

A special type of interval scale is the field key, an example being one used to assess late blight of potato in Great Britain (Anon, 1947; Table 6). Field keys are a scale based on percent severity (for example, 0.1, 1.0 2.0, 5.0%. . . 100%) and used in conjunction with a descriptive or diagrammatic portion of the scale that offers explanation as to the likely

TABLE 6
A qualitative disease key used to estimate late blight of potato in the field (after Anon, 1947).

Rating	Severity characteristics
0.0	Not seen in field
0.1	Only few plants affected here and there; up to 1-2 spots in 12 yd. radius.
1.0	Up to 10 spots per plant or general light spotting.
5.0	About 50 spots per plant or up to 1 leaflet in 10.
25.0	Nearly every leaflet with lesions; plants still of normal form; field may smell of blight but looks green though every plant affected.
50.0	Every plant affected and about one-half of leaf area destroyed; field looks green, flecked with brown.
75.0	About three-fourths of leaf area destroyed; field looks neither green nor brown. In some varieties the youngest leaves escape infection, so the green color is more conspicuous than in varieties like King Edward, which commonly show severe shoot infection.
95.0	Only few green leaves remaining, but stems green.
100.0	All leaves dead; stems dead or dying.

distribution/frequency of the symptom in the field or on plants, which contains quantitative information (Anon, 1947). The late-blight type key was generally used on a whole field basis and apparently received wide usage (Chester, 1950). A similar key for apple scab based on the percent area infected (0.01–75%) was developed by Croxhall *et al.* (1952a and b), who felt that a SAD-assisted assessment of apple scab developed by Tehon and Stout (1930) was too slow for survey work.

4. Ratio Scales

The percent scale is a ratio scale and has been widely used to estimate disease severity in numerous pathosystems in plant disease assessment. James (1971, 1974) outlined some of the advantages of the continuous percent scale in assessing disease severity to the best of the rater's ability. He states i) that the upper and lower limits of a percent scale are consistently defined (0 and 100%), ii) the scale is universally familiar, iii) it is easily divided and subdivided, and iv) it is widely accepted as a way to measure area coverage. However, its effective usage does require that raters should arrive at similar estimates of any given actual disease, and that this estimate should be achieved simply and quickly. James (1974) discussed some of the merits of the percentage scale in relation to a logarithmic basis like the H-B scale, and agrees with Chester (1950) that additional divisions were highly desirable in the middle severity ranges as experienced raters tended to be able to discern severity at this level. This has subsequently been shown to be the case (Nutter and Esker, 2006). Indeed, James (1971) goes on to comment (although not backed by data) that with the aid of SADs raters can do better than attempting to assess based on a log scale, the exact ability being dependent on the individual. In addition, he maintained that it should be used as it allows observed differences to be recorded and used that might otherwise be missed by other rating systems. He concludes that it is always advisable to use a scale with equal divisions like the percent scale because even in the event a logarithmic relationship is determined, a transformation can be applied and will still be more accurate and reliable than applying *a priori* equal divisions on a log scale, a view borne out in other studies (Forbes and Korva, 1994; Nita *et al.*, 2003; Nutter and Esker, 2006; Bock *et al.*, 2009b).

Various authors have investigated the accuracy and precision of raters using the percent ratio scale and demonstrated ways in which raters can improve the quality of their estimates, and these are considered in the following sections. Raters vary substantially in the accuracy and reliability of assessment (Nutter et al, 1993; Nita *et al.*, 2003; Madden *et al.*, 2007; Bock *et al.*, 2008a, 2009b). As already noted, apart from variability in inter and intrarater reliability, other aspects influence the ability of the rater including lesion number and size (the tendency to overestimate at low disease severity, particularly acute with numerous, small lesions), use of SADs, training and concentration (Sherwood *et al.*, 1983; Nutter and Schulz, 1995; Godoy *et al.*, 2006; Bock *et al.*, 2008b). For example, when measuring the precision of four raters assessing dollar spot (*Sclerotinia homeocarpa*) on bentgrass, Nutter *et al.* (1993) showed a range in quality of assessment, with no one rater being in perfect agreement with the actual severity ($r^2 = 0.79$-0.97). The inter- and intrarater reliability was estimated and interrater reliability ($r^2 = 0.70$-0.89) and intrarater reliability ($r^2 = 0.83$-0.93) were also found to be variable (see Table 2), and as with other systems has provided base-line information on the unaided ability of assessors to measure disease severity (Nita *et al.*, 2003; Bock *et al.*, 2008a, 2009a). Based on the slope and intercept, accuracy of raters estimates compared to the actual value was similarly variable (intercept = 1.3-19.6, slope = 0.9-1.4), with all raters showing a bias or tendency to over or under estimate disease severity. Similarly in other large scale studies with multiple raters the reliability and accuracy of estimates has been variable as illustrated by raters estimating severity of *Septoria* on wheat and *Melampsora* on willow (Parker *et al.*, 1995a), *Septoria* and *Erysiphe* on wheat (Parker *et al.*, 1995b), various diseases on alfalfa (Guan and Nutter, 2003), *Phomopsis* on strawberry (Nita *et al.*, 2003), and citrus canker on grapefruit (Bock *et al.*, 2008a, 2009a).

E. Ways to Improve Visual Estimates of Disease Severity

To identify and quantify characteristics of error, estimated values must be compared to actual values. Once demonstrated, ways can be developed to address the error and then subsequent improvement can again be measured against the original data. Thus, establishing a baseline of understanding error is important in this whole process, not only to identify the sources, but

also for quantifying improvement. Several studies over the last thirty years have demonstrated the range of ability among both experienced and inexperienced visual assessors in many different pathosystems that exhibit a wide range of disease symptoms (Amanat, 1976; Sherwood *et al.*, 1983; Kranz, 1988; Weber and Jorg, 1991; O'Brien and van Bruggen, 1992; Nutter *et al.*, 1993; Nita *et al.*, 2003; Bock *et al.*, 2008a, 2009b).

1. General Field/Lab Training

Training raters helps improve both reliability and agreement, and traditionally this might be done with field material (Nutter and Schultz, 1995). Disadvantages are that it is often a busy time of the year and requires field visits to perform training. Samples kept for long lose typical traits of disease as they senesce. Doubtless training raters in the field is of value but this has not really been explored, probably because of the logistical issues and the ease of training offered by SADs and computer assessment programs. Although not in the field, Amanat (1976) used images of leaves with different sized lesions and areas infected and showed that repeated training with feedback was beneficial to improving precision of disease estimation by multiple raters.

2. Using Standard Area Diagrams (SADs)

The original Cobb scale was the first SAD used in plant pathology (Cobb, 1892). The rater compared samples to the five standard disease areas illustrating rust pustules on the leaf and placed it in the category to which it was most similar in severity. Many researchers have since developed and applied SADs to improve ability to assess disease on various plant organs, either in interval scales (Tehon and Stout, 1930; Ullstrup *et al.*, 1945; Large and Honey, 1955; Large, 1966) or for aiding in assessment on the percent ratio scale (James, 1971, 1974; Godoy *et al.*, 1997, 2006; Batzer *et al.*, 2002; Nutter *et al.*, 2006; Pethybridge *et al.*, 2007).

As Large (1966) notes, SADs should be distinguishable by the eye, and thus categories should not be too numerous. With specified use for application with a percent ratio scale, James *et al.* (1968) developed a SAD to aid assessment of *Rhynchosporium* leaf blotch assessment in barley showing 1, 2 and 5% infection and although he did not provide empirical data showing the value of SADs, went on to produce several other SAD sets for various disease (Figure 7) (James, 1971), with the recommendation that the rater use them as a guide and endeavor to provide a best estimate of percent area covered. At about the same time, Dixon and Dodson (1971) also developed a set of SADs for various diseases, although depending on the key they suggested using an interval scale, or interpolating percent area infected.

James (1971) believed that SADs "calibrated" a rater. Since that time various studies have demonstrated the value of SADs. Parker *et al.* (1995b) found that SADs tended to result in raters categorizing disease around that represented by the SADs, although this has not been reported in other studies (Hock *et al.*, 1992; Gomes *et al.*, 1996; Godoy *et al.*, 2004). The results from an experiment with multiple raters assessing images of diseased alfalfa leaves with and without the aid of SADs showed they improved the reliability and accuracy of the raters (Nutter *et al.*, 1998). Belasque *et al.* (2005) developed a series of citrus canker SADs representing small, medium and large lesion sizes, and lesions with and without leaf miner damage present, and used these scales to improve reliability and accuracy of rater estimates of disease severity. Many other reports show the benefits of SADs. Andrade *et al.*, (2005) developed a set of standard area diagrams to assess leaf spot (*Quambalaria eucalypti*) of *Eucalyptus* and found improvements in the precision, accuracy and reproducibility of disease assessments against actual values measured by image analysis. Gomes *et al.*, (2004) developed standard area diagrams of *Cercospora* leaf spot of lettuce caused by *Cercospora longissima* using Autocad and gave these to assessors and found that the SADs provided improved levels of accuracy and reliability as measured by the coefficient of determination, slope and intercept. Hock *et al.* (1992) used two different standard area diagrams for assessing small vs. large lesions of the tarspot disease complex on maize and Godoy *et al.* (2006) developed a diagrammatic scale (Figure 8) to assess soybean rust (*Phakospora pachyrhizi*) and both found overestimation was common among raters, but using the scale reduced error, and reduced error even more in those inexperienced raters.

Leite and Amorin (2002) used a diagrammatic scales set designed on a logarithmic basis (0.03, 0.2, 0.6, 3, 7, 12, 25, 40, and 66%) to assess *Alternatria* leafspot (*Alternaria helianthii*) of sunflower and they showed that use of the scale improved both accuracy and precision. The H-B scale casts a long shadow. SADs are often prepared based on the nonexistent Weber-Fechner law and the logarithmic relationships implied by Horsfall and Barratt (1945) (Godoy *et al.*, 1997; Andrade *et al.*, 2005; Belasque *et al.*, 2005). Although suspected for a long time (Chester, 1950; James, 1971, 1974; Herbert, 1982), only recently has data emerged that suggests the human eye can differentiate disease in the mid range more accurately and precisely than previously assumed (Nita *et al.*, 2003; Nutter and Esker, 2006; Bock *et al.*, 2009c and d). There is now at least one example of a SAD with a stated basis of a linear scale (Pethybridge *et al.*, 2004). However, it might make sense to concentrate SADs at the lower end of disease severity (<40%) as many infections are found in this range (Kranz, 1977), but using the rationale of the so-called Weber-Fechner law to support this no longer makes sense (see section IV.D.3). Early SADs developed by James (1971) did not incorporate a logarithmic component and recent suggestions are that a linear scale, and therefore a more linear range of SADs might be more apt for estimating disease (Nita *et al.*, 2003; Nutter *et al.*, 2006), even though some raters doubtless have a tendency to be less precise or accurate over certain disease severity ranges (Forbes and Jeger, 1987; Hau *et al.*, 1989; Bock *et al.*, 2008b). Logarithmic and linear SADs have not been compared, nor has optimum step size (perhaps every 5%), which is likely to be dependent on a number of factors including rater ability, disease severity range and pathosystem

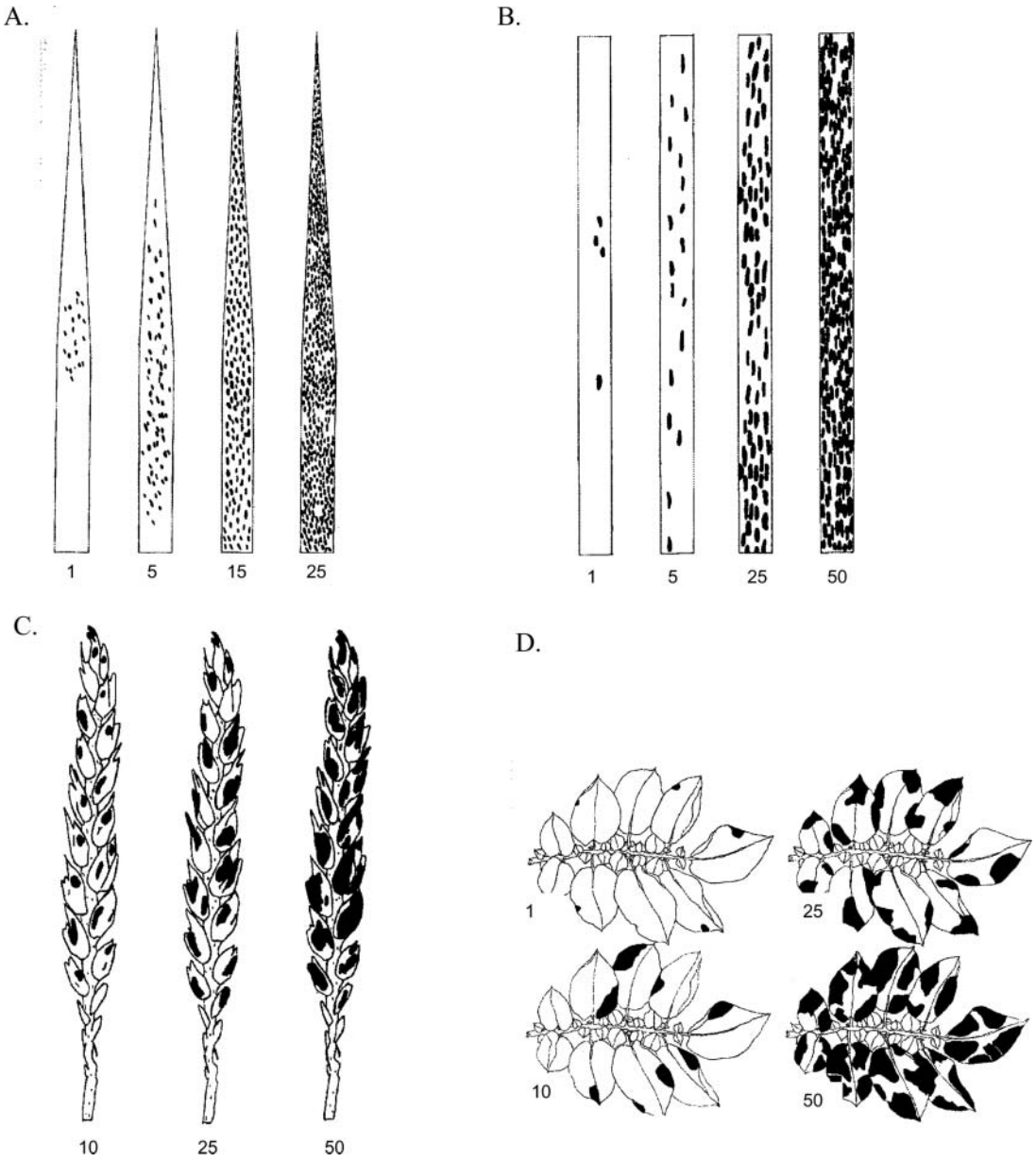

FIG. 7. Standard area diagrams used to estimate A. leaf rust of cereals, B. stem rust of cereals, C. *Septoria* glume blotch of wheat, D. late blight of potatoes (after James, 1971).

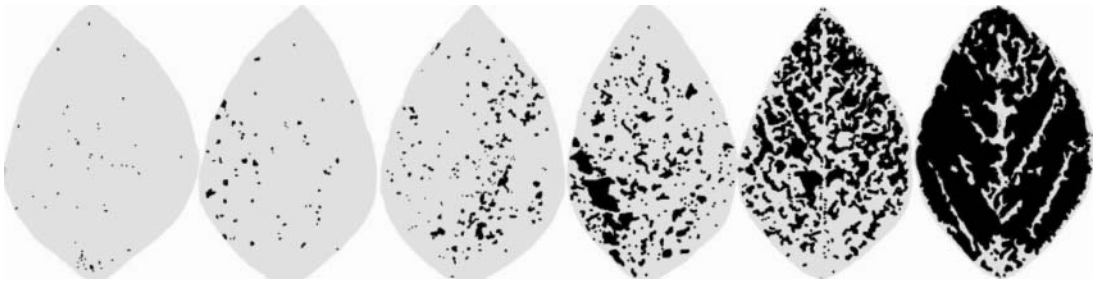

FIG. 8. A SAD developed to aid assessment of soybean rust severity (percent of diseased leaf area indicated) using a non-linear scale having unequal intervals (after Godoy *et al.*, 2006).

(pathogens with lots of small lesions being more difficult to estimate precisely and accurately, Sherwood et al, 1983; Godoy *et al.*, 1997).

Standard area diagrams can be generated in various ways. Computer programs can generate standard area diagrams for disease assessments, such as SEVERITY.PRO (Nutter and Litwiller, 1998). This allows choice of leaf and disease symptom (type and lesion size). Computer generated SADs can be used to improve estimates of disease severity (Nutter and Esker, 2001; Nutter *et al.*, 2006). Leaf area diagrams can also be constructed from scanned images of diseased leaves (Pethybridge *et al.*, 2007).

The use of SADs does not eliminate error, and relative error still appears greatest at low levels of disease (Beresford and Royle, 1991). In studying brown rust (*Puccinia hordei*) on barley they found that despite using a set of SADs (Anon., 1976) there was still greatest loss in relative accuracy at low severity. There are also some disadvantages to SADs – the observer has to fit the pattern on the actual leaf to a severity in the diagram set where the distribution of disease is different, and there is no feedback on the closeness of estimated to actual which is important where learning or training to assess disease.

3. Computer-based Training

Various studies have demonstrated the value of disease assessment training using a computer based system, significantly improving the accuracy and precision and intrarater reliability of raters, but training does require significant effort – particularly time to practice – although an advantage is it can be done in a class room rather than in the field. The earliest assessment training program, AREAGRAM (Shane *et al.*, 1985) was developed to train assessors by generating SADs rather than a full range of possible disease severities. Tomerlin and Howell, (1988) were also pioneers with this technology for disease assessment and developed an assessment program called DISTRAIN and Weber and Jorg (1991) used a program called ESTIMATE to train raters. Newton and Hackett (1994) used DISTRAIN to train a group of raters, and found that most raters showed a significant improvement in accuracy and precision when rating powdery mildew on barley. Parker *et al.* (1995b) also used DISTRAIN and showed similar improvement in ability to assess powdery mildew on barley, although they suggested that the benefits of training were short lived and regular training was required. The most sophisticated and widely used disease assessment training program is DISEASE.PRO (Nutter and Worawitlikit, 1989; Nutter and Schultz, 1995; Nutter and Gaunt, 1996). The DISEASE.PRO training program has been the basis for other programs, including host specific ones (ALFALFA.PRO; Nutter and Litwiller, 1993). Results from DISEASE.PRO demonstrate the advantage of computer-based training (Nutter and Schultz, 1995). In one study 80 raters assessed disease prior to training, and were retested after training. Subsequent to training the coefficient of determination (r^2) values improved for most raters by more than 10%, accompanied by lower coefficients of variation and improved slope and intercept fits. DISEASE.PRO is available as a training program from APS press (APS, St. Paul, MN), and as part of an epidemiology training package (Nutter, 1997a; APS, St Paul, MN). Nutter and Litwiller (1998) subsequently developed SEVERITY.PRO which allows for choice of a wider range of leaf shapes and lesion types that broadened its applicability to a greater range of host-pathogen systems. The system allows raters to i) set whether they want an immediate response for actual severity, ii) choose number of leaves and lesion characteristics (small, medium, or large, random or not), iii) view various charts that show estimation error, regression of estimated disease against actual disease. This allows pre and post-test results of individual raters to be compared in various ways to monitor ability and improvement as well as understand generic sources of error and those unique to individual raters (Nutter *et al.*, 2006). In tests performed on grapevine downy mildew with six raters, measures of accuracy (bias of the estimate, described as slope and intercept) were improved in half the test individuals, and measures of reliability were improved in five of the six individuals (described as the coefficient of determination) by up to 10%. Using a computer-based system, digital photographs of stem sections of rapeseed infected with phoma stem canker (*Leptosphaeria maculans*) have been used to train raters to rate diseased stems in the correct severity scale based on percent area infected (Aubertot *et al.*, 2006). Training improved subsequent to rating. A further advantage of computer training programs is that they can provide assured "actual" values against which to estimate disease (Nutter *et al.*, 2006). The drawback of computer training is that the advantages might be short-lived (Parker *et al.*, 1995b), assessors requiring regular re-training.

4. Using Leaf Grids

Not widely used or reported on, these have been used in the past. Parker *et al.* (1995b) reported using a leaf grid to assess disease. The generic leaf grid was made up of 1% divisions. They found that it was less good than computer training as there was a tendency for some raters to misuse this assessment aid. When used correctly it gave similar improvements to computer training using DISTRAIN. Disease grids have been used to aid assessment in the field (Lovell *et al.*, 1997).

F. A Comparison of Visual Rating Methods for Disease Severity Assessment

Considering how fundamental it is, there are few published studies objectively comparing visual rating methods of disease severity on the same plant units, and only recently have different methods been compared. There are more studies that have compared assessments of different parts or characteristics of disease on the plant (Townsend and Heuberger, 1943; Shokes *et al.*, 1987; Lipps and Madden, 1989; Christ, 1991; Vereijssen *et al.*, 2003; Danielsen and Munk, 2004), which although important, is more to do with sampling alternatives than to the aim of this review that is concerned with comparing quality

of severity estimation on the same plant part by individuals or different methods.

Early on different reports commented on the value of particular scales for rating disease but did not provide data that actually demonstrated advantage in reliability or accuracy, or in the amount of time taken (with quantitative evidence of concomitant loss – or gain - in reliability or accuracy). Chester (1950) and James (1971) make general comments on the value of visual assessment methods, and Couture (1980) discussed the relative merits of the percent scale and category scales and described their advantages, although at the time no data as to their relative reliability or agreement was available. Herbert (1982) also raised the issue of visual assessment in relation to rationale for the H-B scale.

By 1989 plant pathologists were actively starting to compare methods quantitatively (Slopek, 1989). By comparing various category scales to direct percent leaf area infected estimates, a 5-category scale was found to provide comparable accuracy to both the direct estimates and the H-B scale mid-points. Using three different interval scales to estimate severity of corky root of lettuce, O'Brein and van Bruggen (1992) related estimated severity to yield loss. The scales compared were a 7-category scale, a 10-category scale and a 12-category H-B type scale. Experienced and inexperienced raters used each scale and accuracy was determined. The 7-level scale was found to be the most accurate and precise, but the H-B scale was best in the 20–80% severity range, and experienced raters tended to be less biased compared to inexperienced raters, although in these studies no actual values were used (the actual value was based on scale originator estimates).

A particularly important study compared direct percent estimation of disease severity, and use of the H-B scale, and showed that the H-B scale did not improve the quality of severity estimation of late blight of potato and estimation by percent area was superior (Forbes and Korva, 1994). Similarly, Nita *et al.* (2003) compared direct estimation of severity of *Phomopsis* of strawberry with estimates based on the H-B scale, and although interrater reliability was generally slightly better for direct estimation, accuracy was not improved by applying the H-B scale to the severity estimates. A scale comprised of 5% categories improved reliability and accuracy compared to the H-B scale. In comparisons of direct estimation and the H-B scale, the H-B scale was less reliable and precise for estimates of citrus canker, but there was no consistent effect on accuracy (Bock *et al.*, 2009c). Increased replication and averaging improved the overall agreement with the actual data and stabilized the variance, but the H-B scale was not superior to direct estimation. Furthermore, using simulation modeling of the disease assessment process and hypothesis testing, Bock *et al.* (2009d) found that there were situations when using the H-B scale resulted in higher probability of a type II error, compared to direct estimation to the nearest percent.

Apart from these studies very little has been done to explore the characteristics of different assessment methods on the same plant units – or to quantify them relative to each other in terms of reliability, agreement with actual values and rapidity of assessment. Only when this information is available will plant pathologists be able to base their decision of scale choice on an objective footing.

G. A Note on the Analysis of Data From Ratio, Interval (Category) and Ordinal Scales

Data which forms a continuous pattern and is evenly spaced between categories might be analyzed using normal parametric methods (e.g., ANOVA), which includes ratio scales and some interval scales (Snedecor and Cochran, 1989; Madden *et al.*, 2007). The data should also be normally distributed.

To satisfy the needs of parametric statistics, many interval and ordinal scales are first converted to a *disease index*, which should provide a measure of the actual percent disease by taking into account the number of observations in each disease category in that sample. With the logarithmic-based H-B scale, categories are first converted back to a percentage midpoint before analysis is performed (Redman *et al.*, 1968; Madden *et al.*, 2007). With other category scales there are various ways of calculating weighted and unweighted disease indices (Chaube and Singh, 1991), but the basic calculation prior to analysis is performed by multiplying the number of sampling units in each category by the category number (or mean %) for each category, and summing the products and dividing by the total number of samples, although the precise way in which the index is calculated has varied and can depend on whether the scale is arithmetic or geometric, and whether it is expressed as a percent or in arbitrary numbers (Walker *et al.*, 1938; Chester, 1950; Croxhall *et al.*, 1952a and b; Large, 1966; Tinline and Ledingham, 1979; Chaube and Singh, 1991; Tinline *et al.*, 1994; Bruton *et al.*, 2000; Harveson and Rush, 2002). For example, Tinline *et al.* (1994) developed a scale to ascribe disease severity on wheat and barley. Plants were assigned a severity rating from a 4-category scale where 0 = healthy, 1-25% = slight, 26-50% moderate and >50% severe, with values of 0, 1, 2 and 5 being applied to wheat and 0, 2, 5 and 10 being assigned barley for the four categories prior to the percent disease ratings being calculated through a disease severity index, and an ANOVA performed on the resulting data. Non parametric methods can always be used for both ordinal and interval-based data, if necessary (Shah and Madden, 2004; Madden *et al.*, 2007).

H. The Future of Visual Rating Methods

Visual rating methods are, and will continue to be the single most important way of assessing plant disease for the foreseeable future. Other technologies being developed will no doubt continue to make a greater contribution. Thus, the need to elucidate the best ways to assess disease severity, identify and gauge the magnitude of error in estimates, and reduce that error is highly desirable. The reliability and agreement that

is required for the disease assessment task at hand should be considered when choosing an assessment system (surveys, rating germplasm, predicting disease progress, etc.). Wherever possible studies should adopt methods that maximize both rater reliability and agreement with the actual value.

V. DIGITAL IMAGERY AND IMAGE ANALYSIS IN THE VISIBLE SPECTRUM

Digital cameras are an inexpensive and widely used resource, and used for various applications in plant pathology. Photography (digital, and previously film) has been used to detect, quantify and study diseases and pathogens for many decades. Aerial film-photography of diseases fields started in the 1920s (Neblette, 1927) and has continued to be used and developed in plant disease detection and quantification (Colwell, 1956; Brenchley, 1964; Manzer and Cooper, 1967; Bauer *et al.*, 1971; Wallen and Jackson, 1971; Schneider and Safir, 1975; Toler *et al.*, 1981; Johnson *et al.*, 2003; Jones *et al.*, 2006). At the microscopic scale digital image analysis has also been used in plant pathology to measure and observe pathogen and host physiology and development (Smith and Dickson, 1991; Hilber and Schuepp, 1992; Dieguez-Uribeondo *et al.*, 2003; Seiffert and Schweizer, 2005) and pathogen dispersal (Fitt *et al.*, 1982).

It is in the mid range of plant disease severity estimation that that pertains specifically to measurement of symptoms on individual plant organs, plants and quadrats which is in concert with the resolution of most digital cameras at the sub-plot level. Early work was performed with video cameras sensitive to the visual spectrum or using digitized images of film (Nilsson, 1995), but as the technology progressed, and digital cameras became available and less expensive, from the late 90s onwards, most studies have used these.

A. Digital Cameras and Other Image-Acquiring Devices

Imaging devices include traditional film cameras, slide scanners, flat-bed scanners, video cameras and digital cameras. Digital cameras have fast become the primary device for imaging samples. There is a tremendous range of capability, and it would be unrealistic to attempt a breakdown of their pros and cons. This section will provide only a brief overview of digital imagery in relation to image acquisition. The popular photographic press runs numerous reviews of currently available digital cameras cataloguing their capabilities and a fairly recent review on digital imagery and image acquisition based on digital cameras was published in *Plant Disease* (Ricker, 2004), and describes some aspects of this technology relevant to plant pathologists. Digital images can also be obtained from a flat bed scanner used to digitize old photographic prints and negatives, or even plant leaves directly, and slides can be scanned and digitized by film scanners. Once obtained, there is a flow of information from the choice of the sample unit to record to the measurement of diseased area in the imaging process (Price and Osborne, 1990; Nilsson, 1995).

B. Image Acquisition

Digital cameras consists of a lens, a viewfinder (and/or an LCD display), and a light sensitive screen on which the light from the image falls. First considering monochromatic digital cameras (sensing only one color), the screen, comprised of an array of photosensors, measures the intensity of the incoming light. There are two common screen types. The CCD (charge-coupled device) screen which is an array of semi-conductor photosensors and the CMOS (complementary metal oxide semiconductor) screens which are more energy efficient and cheaper to manufacture, as well as acquiring the image data in parallel, which is faster. The photosensors are memory cells and convert the incoming light into electrons – an electric charge - the accumulated charge is released and is proportional to the intensity of the light. The contents of the photosensors on the CCD screen are then converted from an analogue to a binary digital signal (1, 0) by a frame grabber (an analogue-to-digital converter), and transmitted to the computer where they are drawn on the screen based on the 1 and 0 readings (Ricker, 2004). CMOS screens are designed so they produce a digital output and do not need to be converted. Thus, each photosensor is represented on the computer monitor by a pixel, with brightness and location identified through image processing. Color digital camera images are acquired by incorporating sensors for each of the primary colors red, green and blue in the screen. The photosensors can be co-arranged in different ways, from a regular mosaic to a layered arrangement. Each photosensors records only a single color and the actual color for the image for an individual pixel is obtained by an interpolation algorithm that compares it with the surrounding pixel color measurements thereby generating an estimate of color. The algorithm uses the RGB color model to accurately portray the original color (Cope, 2002).

C. Image Resolution and Subject Orientation

It is useful to understand the factors that influence digital image quality, and how these influence disease severity measurement during image analysis. These include focus, reflection (glare) of light on the object, uniformity of lighting prior to image acquisition (Blasquez and Edwards, 1985; Price and Osborne, 1990; Price *et al.*, 1993; Tucker and Chakraborty, 1997; Steddom *et al.*, 2005a), and resolution and compression of the image file subsequent to its acquisition and storage (Steddom *et al.*, 2005b). Various studies have stressed possible problems caused by reflection (Price *et al.*, 1993; Steddom *et al.*, 2005a) and the need to keep the image in uniform focus and sharp, avoiding variation in lighting or coloring that might be picked up by the imaging software and could result in error differentiating healthy and diseased areas (which can be particularly problematic where automation of the process is desired). Kampmann and Hansen (1994) found leaf veins and reflections on cucumber leaves degraded the accuracy of color image analysis. Martin and Rybicki (1998) also found this with leaves of maize. Reflection leads to low saturation, and

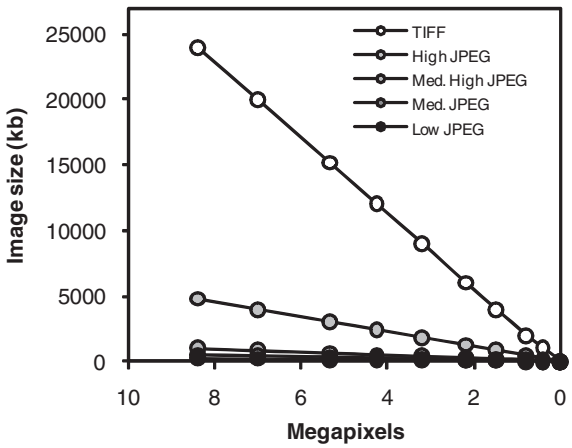

FIG. 9. Image size was dramatically reduced by using JPEG compression or reducing image size. The greatest difference was seen between TIFF images and JPEG images at the highest resolutions (after Steddom *et al.*, 2005b).

shadowing leads to low intensity, which might interfere with the process of segmentation (separation of areas of interest based on chosen criteria, for example, color or intensity). However, in a study of wheat plots infected with rust and imaged overhead, to minimize glare and shadow, and at an oblique angle, to maximize glare and shadow, removal of pixels representing glare resulted in a reduced correlation (8%) compared to estimates of severity by raters (Steddom *et al.*, 2005a). So although reflection and shadow can influence the result, it may cause greater error when these pixels are removed from the image.

The number of bits also influences the image quality. The bit depth is the grayscale or number of values from white to black. Thus each bit can be 1 or 0, white or black respectively, so a two-bit system has four values, black, white and two shades of gray: 00 01 10 11, and so on. The number of possible categories in a gray scale is 2^a where a is the number of bits (Ricker, 2004). The human eye is thought to be able to differentiate 10 million colors (Judd and Wyszecki, 1975), and most digital cameras are 8-bit per color (2^{8+8+8}) or 16.8 million colors.

The resolution and size of the image affects storage space for the data. Images are inherently large files (TIFF, BMP, JPG etc.), and there is little point in having an image that is larger than required to provide a good quality measurement (Steddom *et al.*, 2005b). Thus image resolution can be reduced, or files compressed, which saves space (Figure 9), but can influence the image quality. The resolution of a digital camera (or other digital imaging devices or printers) is defined by the numbers of photosensor (equals pixel) rows and columns on the CCD screen, summarized as megapixels for digital cameras, thus a camera with a 1600 x 1200 photosensors screen is described as a 2-megapixel camera (Ricker, 2004), and a camera with 4064 x 2704 photosensors is an 11.1 megapixel camera, which is considered a pretty good resolution at the time of writing. Cameras with over 16 million pixels are available, and for comparison, there are estimates that a good 35mm film image has a resolution of about 20 million pixels (Anon., 2004).

There are ways of reducing file size without impairing the ability to perform digital analysis, and this can be achieved by file compression and reducing resolution (Ricker, 2004). Thus size of an image file is related to resolution and bit depth and doubling the resolution quadruples the file size (for an 8-bit digital camera):

$$\text{Number of bytes} = (\text{total pixels} \times \text{bits per pixel})/8 \text{ bits/byte}$$

Compression of digital files can be in a "lossey" format or a "lossless" format. The "lossless" (for example, a TIFF image – tagged image file format) preserves all information in the original image and compression is completely reversible, while the "lossey" discards information, for example a JPEG (Joint Photographers Expert Group). JPEGs can be saved at various compression levels, the greater the compression, the poorer the image quality, and the process is irreversible (see Figure 9). Basically, compression involves blocking 8x8 groups of pixels which at high compression leads to a blocky image. Various artifacts can develop in the image analysis procedure as a result (Steddom *et al.*, 2005b). For plant disease severity measurement using image analysis, various image formats, resolutions and compressions including TIFF, and JPEG (high, medium-high, medium, and low) have been compared (Steddom *et al.*, 2005b). Measuring the severity of rust and tan-spot of wheat TIFF images were saved at a resolution of 8.4 million pixels and reduced progressively to 858 pixels per image. The TIFF images were converted to JPEGs with compression of 100, 75, 50, and 25 percent. The disease severity (percent necrotic leaf area) was measured for each image, and correlated with the percentage of necrotic leaf area of the original 8.4 million pixel TIFF images. Image format had little effect, and image resolution had to drop below the point it was difficult to discern lesions by eye (21,000 pixels per image) for resolution to become an issue (Steddom *et al.*, 2005b, Figure 10), although with resetting,

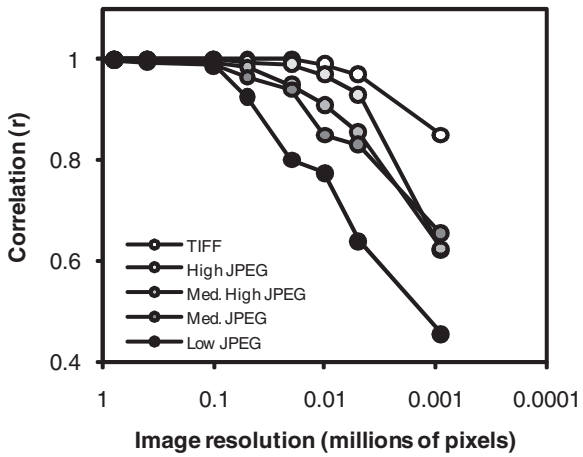

FIG. 10. Correlations between images with reduced resolution and reduced image quality against the original 8.4 million pixel images (after Steddom *et al.*, 2005b).

the image analysis could still perform well. Thus fairly low-resolution digital cameras and high JPEG compression levels are acceptable for plant disease measurement using image analysis. Once acquired, compressed and stored, further processing might be required.

The orientation of a subject can influence the measurement, so it is important that the image provides the cleanest view possible of the area of interest. Kokko *et al.* (1993) studied disease severity on subcrown internodes of wheat infected with *Cochliobolous sativus*. A 4-category visual scale showed that digital gray scale image analysis of wheat was relatively reliable ($r = 0.99$), but they found a significant effect of orientation on the severity measurement. Mean intensity was variable due to uneven distribution of disease around the stem section. They stressed the need for standardization to achieve repeatable results; including lighting, orientation, and three-dimensional shape and using as broad a range of gray scale as possible to help with differentiation of symptoms.

D. Image Processing

Once the image has been obtained, it can be edited in various ways in many different image analysis and image processing software packages. Color and contrast can be corrected, images rotated sharpened inverted or further manipulated. Programs such as Adobe Photoshop (Adobe Systems Inc., San Jose, CA) are powerful software package that offer many options for enhancing images. Most image analysis programs also offer image editing and modification including enhancing edges and geometric corrections.

E. Image Analysis Software and Image Measurement

Many different proprietary and custom image analysis software programs have been used in plant disease severity assessment. Some examples are shown in Table 7. Before exploring the application of these it is worth going through the image analysis protocol, which has many common processes regardless of the software program being used. Firstly, with color images the image is composed of three colors and each pixel in the image has a particular value for each of the primary colors red, green and blue based on the RGB color model, which is a three-dimensional color space used to generate the correct color (Figure 11A) in the color of the perceived image (Russ, 2002). The pixel color is described by hue, saturation and intensity (HSI, Figure 11B). The hue is the pure color of the pixel, while the saturation of a pixel is the amount of color (pure hue, to white, which contains a wider range), and the intensity of the pixel which relates to its brightness (from the pure hue to completely black). An image of a citrus leaf infected with citrus canker is shown to demonstrate the RGB components of the original image (Figure 12). However, the HSI characteristics are

TABLE 7
Various image analysis software used in various studies to measure disease severity on leaves, plants and small plots.

Software source	Name	References
Commercially available	ASSESS, APS Press, St Paul, MN	Tjosvold and Chambers, 2006; Donzelli and Churchill, 2007; Jackson *et al.*, 2007; Vicent *et al.*, 2007; Bock *et al.*, 2008a and b; Grunwald *et al.*, 2008; Bock *et al.*, 2009a and b
	Image Pro Software (Media Cybernetics, Silver Springs, MD)	Mian *et al.*, 1998; Diaz-Lago *et al.*, 2003
	JLGenias (JL Automation, Sunderland, UK)	Martin and Rybicki (1998)
	SigmaScan (Jandel Scientific Software, San Rafael, CA)	Niemera *et al.*, 1999; Olmstead *et al.*, 2001
	Sigma Scan Pro software (SPSS Inc, Chicago)	Biernacki and Bruton, 2001
	Skye-Probetech,	Venette and Venette, 1991
	Soft Imaging Systems GmBH,	Boso *et al.*, 2004
Custom software	Microsoft C compiler (Microsoft Corporation, Redmond, WA) and Matrox color board and color (Matrox Electronic systems, Ltd, Dorval, Quebec)	Ahmad *et al.*, 1999
	Image09 (on line, University of Cape Town, South Africa)	Martin and Rybicki (1998)
	Visual C++.	Tucker and Chakraborty (1997)
	Using BASIC with a DS-65 digitiser (Micro Works, Del Mar, CA).	Lindow and Webb (1983)
	QUANT, Vale *et al.*, (2003).	Andrade *et al.* (2005).

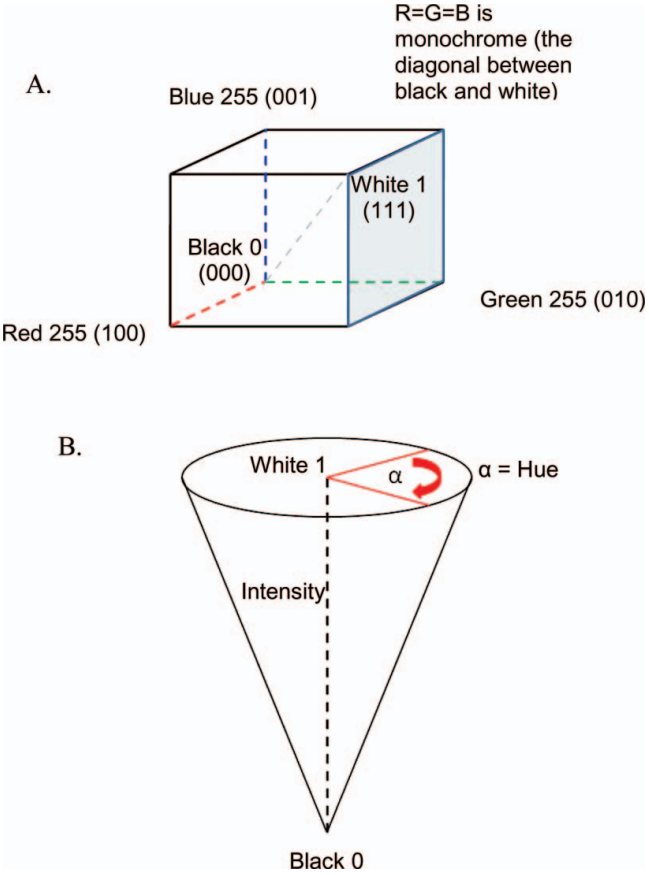

FIG. 11. The basis of the image models and the characteristics of the images that are used in image analysis procedures. A. The RGB model can be considered a cube, with colors being defined by the coordinates of the RGB triplet in 3D space. B. The HSI color model is a mathematical transformation of the RGB cube, with intensity on the vertical axis of the cone (equivalent to the black to white diagonal in the RGB cube), saturation is the radius of the cone, and hue is the angle around the intensity axis. The hue is the pure color; increasing saturation makes color more vivid and increasing intensity makes the color brighter without changing the color or saturation. Hue and saturation are the two most commonly used criteria for separating leaves and diseased areas in analysis of color images (after Lamari, 2002).

most often used to separate the objects of interest in the image (Figure 13). The action of separating areas of interest based on these criteria is called segmentation (in color images pixels are selected that match specific criteria based on hue or saturation, rarely intensity). Once the background has been removed from a leaf, the hue has been found to be the most effective characteristic for delineating healthy from diseased areas with color images, while the intensity plane must be used for black and white images (Lamari, 2002; Steddom *et al.*, 2005b).

A popular and readily available image analysis software for plant pathologists is ASSESS© (ASSESS©: Image Analysis Software for Plant Disease Quantification, APS Press, St. Paul, MN, Lamari, 2002). For thresholding, a histogram 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 lesion areas of the image. The "thumbnails" (the program controls which determine the color range to be selected) are adjusted until the best fit of the chlorosis/necrosis is selected. Once the thresholds are set, the program calculates the area in pixels (Figure 14). In ASSESS©, a percent area function can be applied and the color images measured for area of leaf and area showing symptoms. The thumbnail settings for the leaf and lesion are retained by ASSESS©, providing a guideline for the subsequent image measurements. The process of pixel thresholding is applied in most image analysis systems that are used to measure plant disease severity (Blanchette, 1982; Lindow and Webb, 1983; Newton, 1989; Price *et al.*, 1993; Kampmann and Hansen, 1994). Inevitably setting the thresholds requires some subjectivity and will vary among operators, although this source of error has not been explored in image analysis. When measuring "actual" values, some image enhancement might be necessary to optimize measurements because of inconsistency in lighting amongst images, slight differences in the color of healthy leaf tissue, and variation in necrosis color (Bock *et al.*, 2008a, 2009b). As with most image analysis software programs, ASSESS© has various filters, contrast and color saturation functions, as well as color balancers which can be applied to consistently enhance the area of interest to maintain measurement accuracy for recording actual values. Automation can be achieved by choosing specific threshold settings for images and then running using a macro, but automation can result in error when based on a single image's specific threshold values (although image preprocessing can improve the analysis, Bock *et al.*, 2009a). A recently developed version of ASSESS© (ASSESS© V2.0) that incorporates an automatic threshold feature (each image is treated separately, no common threshold) by applying an automatic algorithm to recalibrate thresholds for each image has been released (2008). There are no published data on automated disease measurement using this system yet available.

An algorithm to automatically set the threshold on successive images has previously been incorporated into other software systems. Lindow (1983) and Lindow and Webb (1983) quantified severity of *Alternaria solani* on tomato and *Aschochyta pteridium* on bracken using black and white digitized images using a 64 gray scale. Algorithms corrected for variation in the background from image to image and subsequent automated measurements of 10 to 20 individual leaves of each species with 0-100% necrosis showed intra-method reliability of +/- 0.8%, and it provided for rapid processing, requiring just 4.1 sec/image. By incorporating this type of self correcting system sensitive to the variation in the background, they were able to achieve a high level of reproducibility at relatively low cost. Other image analysis systems have used algorithms and a step-wedge or photographic gray scale standard for calibration (Martin and Rybicki, 1998). A gray scale image analysis system that automatically assessed leaf area infected gave reliable results, in agreement with actual values – and was more precise and less biased than estimates by raters (Table 8). The reference

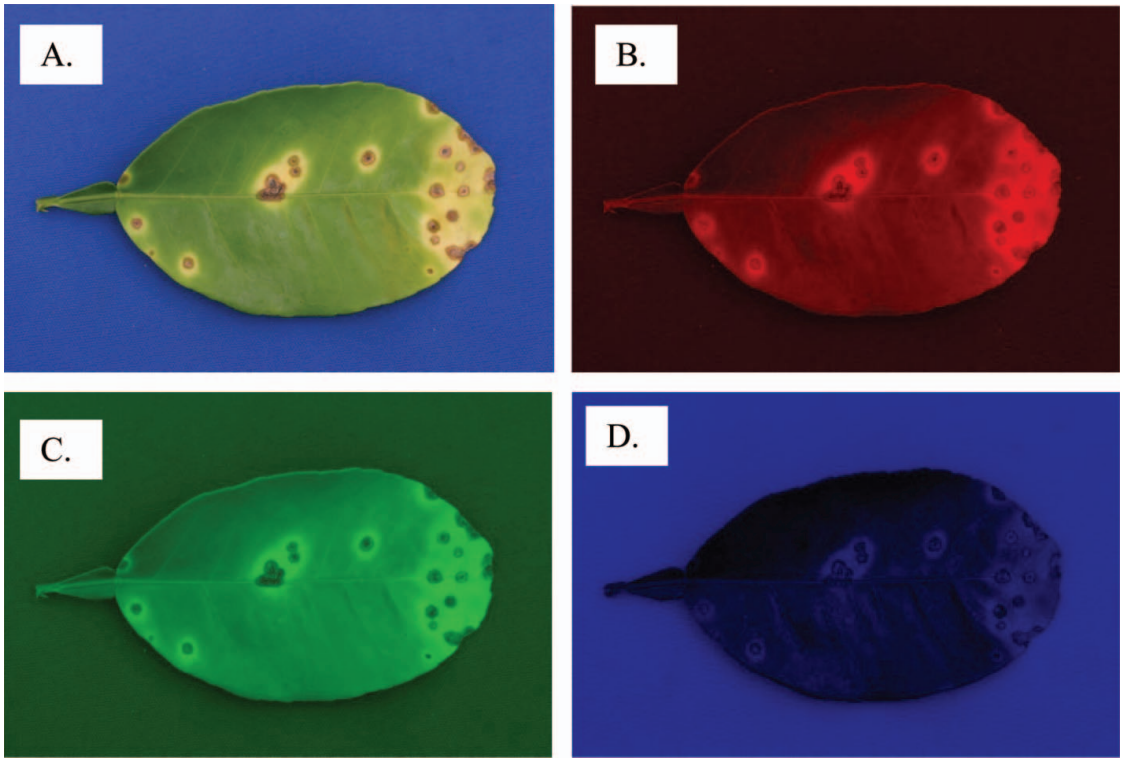

FIG. 12. Sensors in most digital cameras record images in three colors, red green and blue, and thus each image is a composite of the combinations of color in the 3D color space. Thus an original image (A) is comprised of a red component (B), a green component (C) and a blue component (D), the relative proportions of which result in the actual color of the image. The color, along with the hue, saturation and intensity are useful characteristics for delineating objects of interest in an image.

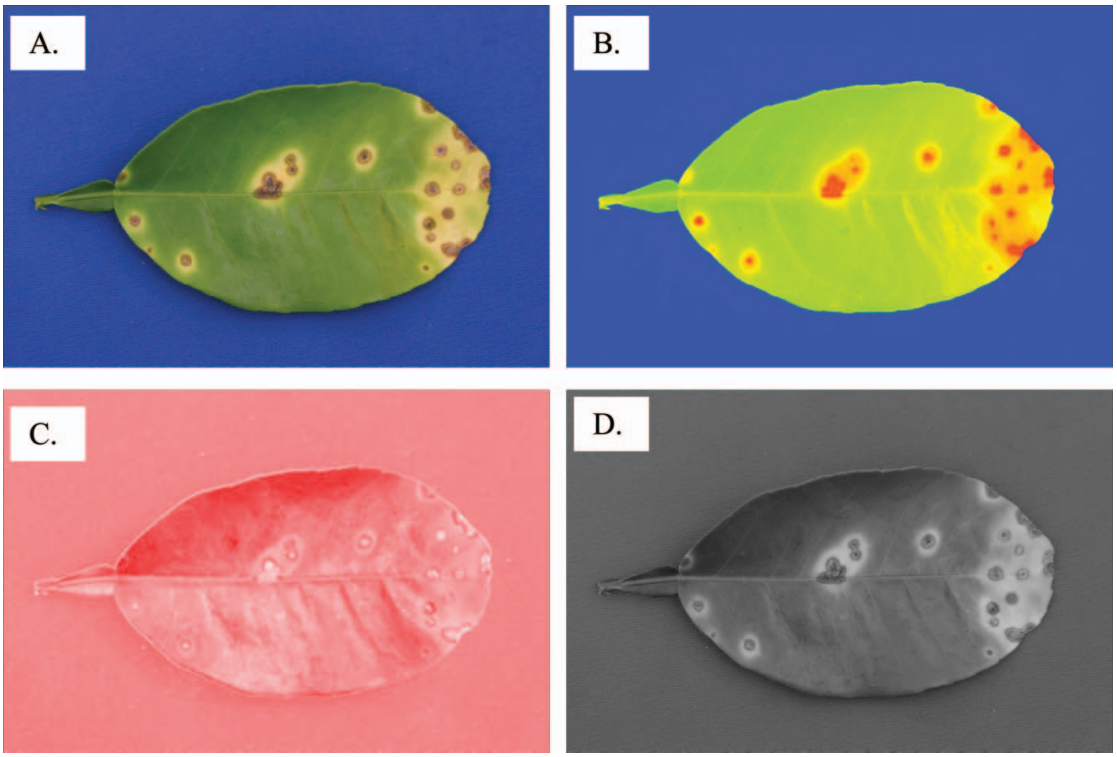

FIG. 13. The original image of a grapefruit leaf with symptoms of citrus canker (A), and an image showing each of the hue (B), saturation (C) and intensity planes (D) that are used to separate areas of interest in images of leaves or plants.

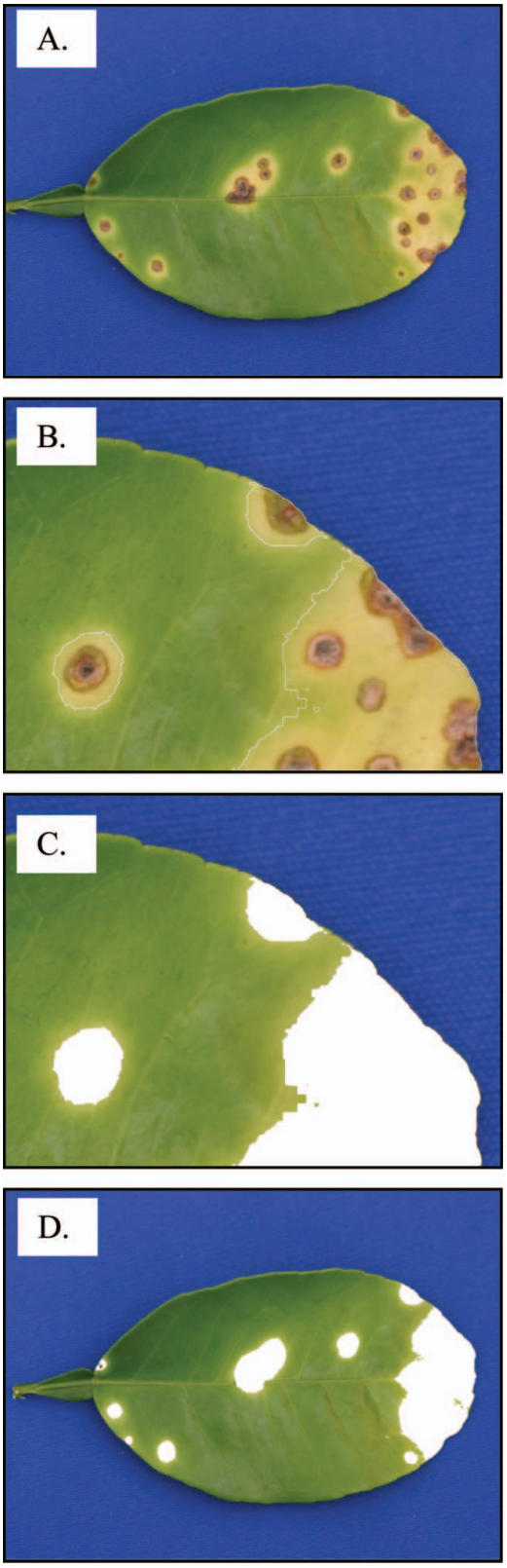

FIG. 14. The original image and detail of a canker infected grapefruit leaf (A) showing the process of image outlining in detail (B) and area demarcation of the symptoms (C, D) in ASSESS© (Lamari, 2002).

TABLE 8
Determination of the accuracy of disease assessment techniques (image analysis and visual assessment[1]) based on linear regression analysis of chlorotic area estimates of the area of maize leaves showing symptoms of maize streak virus. The estimates of disease were plotted against actual chlorotic areas (after Martin and Rybicki, 1998).

		Regression parameters[2]		
Analysis technique	Scorer	r^2	*Slope*	*y-intercept*
Commercial image analysis system	Manual	0.973	1.040	−0.877
	Automatic	0.976	1.045	0.346
Custom image analysis system	Scorer 1	0.985	1.013	0.275
	Scorer 2	0.971	1.032	−0.782
	Scorer 3	0.980	1.020	−0.190
	Scorer 4	0.979	0.999	0.610
Visual assessment	Scorer 1	0.913	1.066*[3]	12.958*[4]
	Scorer 2	0.960	1.245*	1.064
	Scorer 3	0.957	1.135*	5.201*
	Scorer 4	0.927	1.029	8.385*

[1]Two image analysis systems, a commercially available system operated both manually and automatically, and a custom made system operated by each of the raters were compared to visual assessment of the leaf images for accuracy.

[2]Linear regression analysis ($y = a + bx$) was used to compare estimates to the actual values. The coefficient of determination, (r^2) indicates the precision of the estimates, and the slope and y-intercept indicate the bias.

[3]Slope values that are significantly different from 1 ($P \leq 0.01$) are followed by*.

[4]y-Intercept values that are significantly different from 0 ($P \leq 0.01$) are followed by*.

step-wedge function allowed full automation and a custom system took 36.2 sec per image compared to 10.1 sec per leaf for a commercial one. Subsequent application of the custom system showed that it was better able to discern resistance to maize streak virus in maize compared to visual estimates (Martin *et al.*, 1999).

F. The History and Application of Image Analysis in Plant Disease Measurement

There have been several reviews that have included consideration of images from digital cameras in agriculture and other plant sciences (Nutter, 1990; Price and Osborne, 1990; Nilsson, 1995), but none that have considered it solely in the context of disease severity measurement. In the early days of photography and image analysis of plant disease, most of the images were acquired from a substantial distance, generally with the use of camera systems mounted in aircraft (Taubenhaus *et al.*,

1929; Brenchley, 1964; Jackson and Wallen, 1975; Toler *et al.*, 1981). The use of videography and image digitization as a tool to measure disease severity at the scale of a leaf did not really receive attention until the 1970s and 80s. The advent of desk top computers and relatively inexpensive video and digital cameras and the availability of image analysis software programs resulted in greater interest and application of image analysis for measuring disease severity on individual plants, roots, leaves or fruit, or small areas (quadrats). There are now many studies where image analysis of plant leaves, roots stems or fruit has been used, and/or compared with visual rating, although not all have described an "actual" value for agreement comparison which inevitably precludes comment on their accuracy. Thus some caution is needed when image analysis results are compared only to those of a visual rater, whose assessment might be biased to an unknown degree.

1. General Studies on the Application of Image Analysis

Nilsson (1980, 1995) was an early pioneer in the application of remote sensing and image analysis in plant pathology, and considered the applicability of image processing for disease assessment. He described using a Leitz Texture Analyzing System (Ernst Leitz-Westar GmbH, Wetzlar, Germany) to measure disease severity, and concluded that image analysis had great potential in plant pathological research. Another early application was that by Blanchette (1982), who used a VP-8 image analyzer system to demonstrate the potential of the technology to measure the severity of wood decay in pine. Image analysis readily differentiated healthy from decayed wood using an eight gray-scale system, providing good correlation ($r > 0.91$) with actual areas measured from tracings. A pioneering study that demonstrated the capacity of image analysis to measure disease severity on complex leaves with good intra-method reliability and agreement was that of Lindow and Webb (1983) and Lindow (1983). Other early investigative studies of image analysis included that of Bronson and Klittich (1984) who developed PHYTOSCAN 83. Blasquez and Edwards (1985) used transparencies digitized by recording on a monochrome video camera and compared development of severity of late blight on tomato using image analysis and an arbitrary visual rating scale. Reliability compared to raters was found to be good. Several studies subsequently demonstrated either black and white or color image analysis to be superior to visual rating in various different pathosystems (Kokko *et al.*, 1993; Price *et al.*, 1993; Kampmann and Hansen, 1994). Price *et al.* (1993) compared actual values to color and monochrome digital image analysis and visual assessments of severity of coffee leaf rust (*Hamileia vastatrix*) on coffee and found color image analysis was more accurate than visual assessment or black and white image analysis. Garling et al (1999) compared image analysis, multispectral radiometry, grids and visual assessment for measuring dollar spot (*S. homeocarpa*) on turf grass. All four methods were significantly different from each other – with image analysis apparently providing the most accurate measurements. Bock *et al.*, (2008a and 2009b) used image analysis to measure actual disease severity on a leaf-by-leaf basis and found image analysis was more reliable when repeated compared to visual raters. These data were used to investigate various sources of error in visual assessment of citrus canker (Bock *et al.*, 2008b) and a system was developed to automate measurement of citrus canker severity (Bock *et al.*, 2009a) using ASSESS© (Lamari, 2002). Agreement of the automated system to actual values was similar to visual raters, but took longer per image.

However, not all studies on the application of image analysis have found it produces results in close agreement to actual values or rater estimates. Nutter *et al.* (1993) reported poor agreement between image analysis and actual values. Tucker and Chakraborty (1997) used image analysis to count lesion number and measure severity of leaf blight (*Alternaria helianthi*) of sunflower (*Helianthus annuus*) and oat leaf rust (*Puccinia coronata* f.sp. *avenae*) on oats (*Avena sativa*). Although there was good agreement between actual disease measured by a planimeter and lesion numbers from image analysis, raters consistently provided a better estimate of leaf area infected and lesion number for both diseases when compared to image analysis. Olmstead *et al.*, (2001) found that image analysis of powdery mildew (*Podosphaera clandestine)* on sweet cherry was inferior to rater estimates when compared to actual values. Despite these studies, the vast majority of recent work suggests that image analysis most often provides a more accurate and precise, but generally more time-consuming way of rating disease. Image analysis has now been widely tested and explored as a tool in plant pathology, and also for applicability in sectors within the discipline having various research goals.

2. Specific Applications of Image Analysis in the Visible Spectrum for Disease Severity Assessment.

a. Quantifying host resistance. There have been several studies testing the application of image analysis to measure disease severity in disease resistance studies. Among the first study to investigate its application at discerning germplasm resistance was that of Newton (1989), who used image analysis to measure the sporulating area of powdery mildew (*Erisyphe graminis*) on barley leaves of various cultivars and found good correlation with most components of partial resistance, but image analysis was unable to discern a reduced colony size component, which visual raters could do. In contrast, Todd and Kommendahl (1994) found image analysis more discerning than raters at differentiating germplasm reaction of corn to *Fusarium* spp. causing stalk rot, and for differentiating among *Fusarium* spp., although no actual values were used in the study. Image analysis was as good as, and complimentary to visual assessment comparing resistance of potatoes to late blight based on internal images of symptoms in the tuber (Niemira *et al.*, 1999). One study in particular that produced results in close agreement with actual values was used to measure host reaction to maize streak virus (MSV) in differentially resistant corn genotypes (Martin and Rybicki, 1998; Martin *et al.*, 1999). Using a grayscale image

analysis system they automatically assessed leaf area infected and were able to differentiate degrees of resistance to MSV that were not distinguishable using other methods of assessment (it also allowed identification of virulence types of MSV on the differential series). The custom system was developed to fully automate image analysis that could set threshold cut-off values and measure area infected without operator input (Martin and Rybicki, 1998), which is analogous to that developed by Lindow and Webb (1983). Kokko *et al.* (2000) used image analysis to measure total area and mean intensity of common root rot (*Cochliobolus sativus*) of barley and demonstrated the image analysis was sensitive to cultivar differences, the relationships between disease intensity, root area and root weight, and was more objective and reliable than visual assessment based on a scale. Using color images of oat leaves infected with oat crown rust (*Puccinia coronata*), Diaz-Lago *et al.* (2003) differentiated leaf areas using segmentation based on color (green = healthy, orange-red-brown = sporulating, dark-brown-black = telia, and white = reflection) and assessed uredinium density, uredinium size, relative infection frequency, latent period, days to first pustule appearance and disease progress rates based on images acquired every 1–2 days. Image analysis measurements were effective at differentiating the components of partial resistance. Resistance of clones of grape (*Vitis vinifera*) to downy mildew (*Plasmopara viticola*) was compared using image analysis of foliar symptoms (Boso *et al.*, 2004), with differences in the leaf area infected identified among cultivars. Evans and Pope (2006) used image analysis to compare disease severity on leaves of two wheat cultivars inoculated with *Fusarium graminearum* – analysis of the results showed differences in measured severity between the two varieties. Using a slightly different method (whole pot digital image analysis) several lines of tall fescue (*Festuca arundinacea)* germplasm were successfully screened for resistance to *Rhizoctonia* spp (Sykes *et al.*, 2008).

In contrast to some of the resistance screening studies where image analysis has proven discerning, Olmstead *et al.*, (2001) found that image analysis of powdery mildew (*P. clandestine)* infection on sweet cherry leaves was inferior to visual assessments when compared to the actual values. They concluded that visual assessments were adequate for assessing powdery mildew severity on cherry. Furthermore, digital image analysis was inferior to real-time PCR measurement of fDNA to measure oat genotype resistance to *P. coronata* (pathogen severity),and both image analysis and visual assessment failed to identify some measures of resistance identified by fDNA quantification (Jackson *et al.*, 2006) showing that although disease symptoms are informative, in some cases quantification of the pathogen is more powerful. Thus image analysis has proven to be a useful addition to the range of tools available for plant disease measurement in plant breeding, although it has yet to realize its full potential.

b. Pathogen population biology. Image analysis has been used to study pathogen population biology and virulence, which inevitably has some overlap with studies in plant breeding. Based on image analysis of disease severity on leaves of wheat infected with isolates of *Pyrenophora tritici-repentis* from different geographic locations, Sah and Fehrmann (1992) were able to discern differences in isolate and cultivar interactions among assessment data, and the lack of geographic association of isolates confirmed no geographic pools of virulence. Black sigatoka (*Mycosphaerella fijiensis*) is a very damaging disease on banana (*Musa* spp). Donzelli and Churchill (2007) measured severity of black sigatoka caused by different isolates and were able to demonstrate differences in virulence or "aggressiveness" based on image analysis results, and also discerned effects of different inoculation methods.

c. Pathogen effects on different host species. The effect of pathogens on various host species has been investigated with image analysis. The severity of infection can be used to compare the impact of different pathogens on particular host plants. Biernacki and Bruton (2001) used image analysis to compare the impact of three different root pathogens (*Monosporascus connonballus*, *Acremonium cucurbitacearum* and *Rhizopycnis vagum*) on various root dimensions of muskmelon (*Cucumis melo* var *cantalupensis*) and demonstrated image analysis was a useful way to provide quantitative assessments of plant injury of these root-rot pathogens. In an unrelated study, using image analysis host differences were explored by Grunwald et al (2008). Species of *Viburnum* differed in symptom severity and thus susceptibility to sudden oak death (*Phytopthora ramorum)*.

d. Relating disease severity to yield loss and fungicide efficacy. One of the major outcomes of any disease assessment activity is to relate the damage to yield loss – or disease control to improved yield. Few studies have actually used image analysis results in relation to yield loss. Shaw and Royal (1989) used image analysis to accurately measure the area (actual values) of wheat leaves infected by *S. tritici* to understand the sources of error in individual rater estimates (by regression of estimates of individuals against actual severity, individual error functions were developed). On this basis they were able to estimate and validate a function describing the rate at which *Mycosphaerella graminicola* caused yield loss in winter wheat, and demonstrate yield was best predicted by the integral of the square root of *M. graminicola* severity on the flag leaf alone. Although this demonstrated a way of possibly obtaining better estimates of disease, the intrarater reliability of individuals makes any function developed from one assessment questionable relative to further assessments – i.e., the function is really only good for the assessment on which it is based.

In measuring disease control treatments, image analysis has been used to gauge the efficacy of fungicides. For example, Vicent *et al.* (2007) used image analysis to assess the percent leaf area diseased on mandarins (*Citrus reticulata*) caused by brown spot of citrus (*Alternaria alternata*) and based on the image measurements they were able to show clear differences in rain fastness and persistence among various fungicides used in the study. In an earlier study, image analysis was a tool used

to help estimate sample size in a fungicide experiment (Steddom *et al.*, 2005b). The optimum numbers of leaves to sample was calculated based on a preesample of ten wheat leaves from each plot where strobilurin fungicide was being tested against rust. These data were used to demonstrate that optimal sample size was affected by fungicide treatment – some treatments had more variable ranges of disease severity, requiring a larger sample to ensure an equally good estimate of the mean.

e. As a tool in developing assessment aids. Image analysis has been used to develop assessment aids for plant disease assessment, particularly SADs, the advantages of which have already been discussed in a previous section. Once developed, they can be used to aid visual estimation of disease severity in the field. Bacchi *et al.* (1992) used color digitization of video images of infected bean leaves to prepare a photographic scale for the assessment of *Uromyces appendiculatus* on bean, and Pfender (2004) used image analysis to develop a SAD set to aid estimating leaf area infected with stem rust (*Puccinia graminis* subsp. *graminicola)* on perennial rye grass (*Lolium perenne*). Pethybridge *et al.* (2004) developed a series of SADs for ray blight disease (*Phoma liguilicola*) on pyrethrum (*Tanacetum cinerariaefolium*) in Australia, and based these on image analysis of scanned leaves which were then used to aid estimate of the disease in the field and compare various disease measurements for relationships to disease control and subsequent yield (Pethybridge *et al.*, 2007). Belasque *et al.* (2005) used image analysis to measure the actual range of severity of citrus canker on citrus leaves and used the results to provide the limits for a series of SADs representing small, medium and large lesion sizes, and lesions associated with leaf miner damage. These were used to demonstrate rater error before and after training which resulted in improved agreement by raters. Similarly, Andrade *et al.*, (2005) developed a set of standard area diagrams to assess leaf spot (*Q. eucalypti*) of *Eucalyptus* and found improvements in agreement of disease assessments against actual values measured by image analysis. Based on a 6-point severity rating scale, an image based system was also developed to train raters for assessment of phoma stem canker on oilseed rape (Aubertot *et al.*, 2006), with subsequent improvement in classification of diseased stem sections.

f. To compare components of disease. Various components and measures of plant disease severity have been investigated with the aid of image analysis. In 1975, Eyal and Brown described using image analysis to measure the density of *Septoria tritici* pycnidia on wheat leaves and were able to describe the relationship between pycnidial number, mean area and percent coverage – and demonstrated that as density of pycnidia increased, size decreased. More recently Tjosvold and Chambers (2006) were able to demonstrate using image analysis on leaves of camellia and rhododendron infected with *P. ramorum* that leaf area diseased was highly correlated with lesion number, thus demonstrating the association of the two measures of disease. On the other hand, Bock *et al.* (2008a and 2009b) found that with the citrus canker pathosystem there was a generally poor relationship between the percent area infected and lesion numbers. Different pathosystems are likely to vary, and image analysis can provide a useful way with which to measure various symptom components (pycnidia, lesion number, chlorosis, necrosis, etc.) objectively.

g. Measuring disease on fruit and seed. Image analysis has also been used to measure disease on fruit and seed. Corkidi *et al.* (2006) measured the severity of mango anthracnose (*Colletotrichum gloeosporoides*) on mango. A three-dimensional image analysis system was developed to assess the whole fruit surface and good reliability was found when compared to actual severity. Although generally seed infection is an incidence measure, the potential application to measure severity on seed also makes this worthy of mention. Image processing has been used to classify seed as healthy, diseased, or immature. This was achieved based on a color analysis of symptoms incorporating a model comprising six color features and provided 88% accuracy overall (Ahmad *et al.*, 1999).

An apparently unexplored aspect of image analysis is application to differentiate multiple disease symptoms, which is probably due to the fact that image analysis does not offer an easy way of discerning actual symptoms (necrosis or chlorosis caused by different pathogens). Although in the crop or field we might be interested in a single disease causing an epidemic, there are occasions when other symptoms will be evident (or physiological disorders or pest injury). Image analysis in the visible spectrum has not been tested with regard to its resolving power for multiple diseases, but it is likely not to be good. Trained visual raters probably provide the most discerning approach to assessing multiple diseases.

G. The Future of Image Analysis Methods for Disease Measurement

Image analysis has become increasingly widely used as a research tool for measuring severity on plants or plant organs over the last three decades. As the technology improves and the software becomes more flexible and better capable of differentiating disease and coping with variation among samples there is likely to be a continued increase in its application and use. The lure of speed and automation in assessment, non-destructive sampling and a permanent record are appealing. The reliability and accuracy potentially offered by image analysis will ensure its continued use and an expanding role in plant disease assessment.

VI. HYPERSPECTRAL IMAGING

A. Use of Hyperspectral Imaging

Of the three methods being considered in this review, hyperspectral imaging (HSI) is the most recent and least explored for application in measurement of plant disease. HSI, also known as imaging spectroscopy, is a technology that has received broad

interest in agricultural research, particularly because of its potential application, and the type of data it can generate (Campbell, 2007). Over the past two decades, HSI has been investigated at many different scales and for many applications in the agricultural community. Applications have included monitoring fruit quality, fecal detection on fruits and animal carcasses, creating land use/land cover maps, detecting insect and weed infestations in crops, and disease detection and quantification in crops (Hirano *et al.*, 2003; Uno *et al.*, 2005; Jensen, 2007; Okamoto *et al.*, 2007; Ye *et al.*, 2007). This is illustrated by a search of the literature that brings up over 200 publications in the past 10 years that deal with the use of hyperspectral imaging in agriculture, although only relatively few of these have focused on plant pathology (Bravo *et al.*, 2003; Apan *et al.*, 2005; Mishra *et al.*, 2007), and even fewer with disease severity assessment (Coops *et al.*, 2003; Huang *et al.*, 2007; Larsolle and Muhammad, 2007; Qin *et al.*, 2008). Nonetheless, characteristics of HSI give it a novel and useful application in plant pathology, as it generates a lot of information on the spectral characteristics of the leaf surface that can be used to detect and quantify symptoms.

B. History and Background to HSI

HSI has its roots in NASA's Jet Propulsion Laboratory with the development of the Airborne Imaging Spectrometer (AIS) in the early 1980s (Campbell, 2007). There were earlier systems that produced similar data, but these systems were "non-imaging," such as the Geophysical Environmental Research Spectroradiometer (GERS) (Kruse *et al.*, 1999). HSI technology combines the science of spectroscopy with imaging to acquire both spectral and spatial information of an object or scene simultaneously. Briefly, spectroscopy is the study of how matter interacts with electromagnetic radiation, and how the matter of interest absorbs and reflects this energy (Green *et al.*, 1998). A radiometer measures the intensity of radiated, or thermal energy from an object (Campbell, 2007), usually in just one "band" of wavelengths with frequencies longer than 3 microns. A photometer is similar in function, in that it measures energy from a single wavelength source (which might be fairly broad, such as 50 nm wide), but is typically measures in the shorter, higher frequency portions of the spectrum. A spectrometer uses a spectrograph and detector to measure a wide range of the electromagnetic spectrum in discrete, contiguous "bands." A spectroradiometer is, therefore, a combination of a radiometer and a spectrometer. It measures radiant energy over a large range of the spectrum and separates the information into discrete, contiguous bands, thus giving a spectral curve for the object in front of the optics. This term has been used synonymously for instruments that measure in the higher frequency wavelengths (such as the visible and near infra-red, NIR), especially those used for ground truthing airborne data (Campbell, 2007).

A hyperspectral image is created from many high spectral resolution, contiguous spectral wave bands or sections of the electromagnetic spectrum, but on an image scale. The hyper-

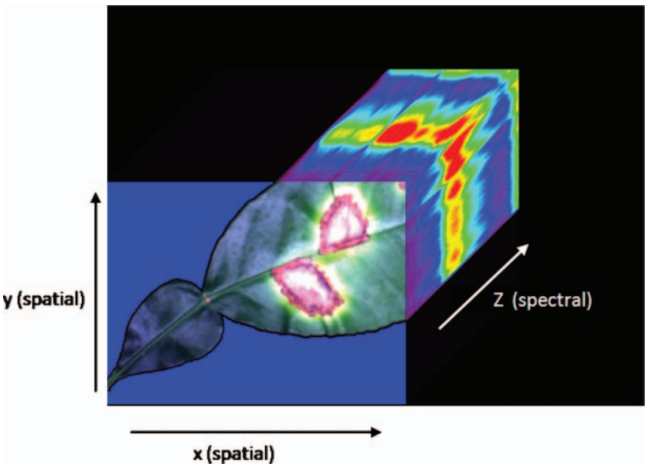

FIG. 15. Hyperspectral datacube (or "hypercube") of a grapefruit leaf with lesions of citrus canker. Notice the two spatial dimensions (X and Y) and the spectral dimension (Z).

spectral image is acquired using a spectrograph coupled with a digital sensor (CCD, CMOS, InGAs, etc.) effectively enabling each picture element (pixel) of the integrated circuit to become a separate spectroradiometer. This produces a three-dimensional image, sometimes called a data-cube or "hypercube," with spatial X and Y axes, and a spectral Z-axis (Figure 15). Typically, hyperspectral images consist of hundreds of registered, contiguous spectral bands such that for each pixel it is possible to derive a complete reflectance spectrum.

An object within the spatial image can often be delineated by its spectral "signature," so the end result is as if each object had been independently run through a standard spectrophotometer. Once identified, objects can be classified, and differentiated via their spectral signatures. The term "spectral signature," also called the "spectral response," "spectral curve," or even just "spectra," refers to the recorded spectral energy from the sample. The term "spectral signature" is actually a misnomer because the spectra aren't as consistent as the term suggests. Biological samples in particular tend to be heterogeneous, both within and between samples, resulting in spectral variation of signatures. In addition, there can be other constituents, sometimes called spectral endmembers, in a given pixel that will influence the spectral response. For example, if one was imaging a corn field from above, the "signature" of that field would differ among images acquired when the seeds were planted, versus early growth, versus harvest, due to the biological changes that are occurring as the plant grows, as well as the amount of ground that is visible between the plants (Campbell, 2007).

As with other remote sensing techniques, HSI has been applied to traditional applications, including geography, geology, biomass and chlorophyll concentrations, and land cover classification (Jensen, 1996). Only relatively recently has it been explored for use in plant disease detection and severity measurement (Bravo *et al.*, 2003, 2004; Apan *et al.*, 2005; Huang *et al.*, 2007; Larsolle and Muhammad, 2007; Qin *et al.*, 2008).

TABLE 9
Different methods of producing a hypercube (after Fisher and Welch, 2006, with additions)

	Snapshot	Staring 2-Dim	Pure Spectra	Wide FOV[1]	Field Stop Required	High Throughput
Grating	Yes	No	Yes	Yes	Yes	Moderate
Computed Tomographic Imaging Spectrometer (CTIS)	Yes	Yes	No	No	Yes	Moderate
Spinning Prism	No	Yes	No	Yes	No	Yes
Linear Variable Filter (LVF)	Yes	No	Yes	Yes	No	No
Liquid Crystal Tunable Filter (LCTF)	No	Yes	Yes	Moderate	No	Moderate
Acousto-Optic Tunable Filter (AOTF)	No	Yes	Yes	Moderate	No	Moderate
Digital Array Scanned Interferometer (DASI)	Yes	No	No	Yes	No	Yes
Scanning Michaelson	No	Yes	No	No	No	Yes
Hadamard Transform	No	Yes	Yes	Yes	Yes	Yes
Prism-Grating-Prism (PGP)	No	No	Yes	Moderate	No	Yes

[1] Field of View

Because hyperspectral imagery possesses high spectral resolution it can often detect spectral subtleties that allow a target to be delineated from its background. However, due to its high spectral resolution, the imagery that these sensors can generate is quite large when compared to standard RGB digital images. Therefore, more computing power is required to analyze the data. In addition, large amounts of disk space are also required to store the data. Obviously, file sizes have many variables (number of pixels on the CCD, spectral resolution of the system, spatial size of the image, etc.) but HSI image sizes in the hundreds of megabytes (Mb) are not unusual.

C. Acquisition of the "Hypercube" Image

There are several ways in which the three dimensional hypercube can be acquired (Table 9) (Fisher and Welch, 2006). Two of the more common techniques rely on taking two of the dimensions of the hypercube at a time and then layering the subsequent images together to form the third dimension (Kim, *et al.*, 2001). The first is to take multiple spatial images, each with a different spectral input, and "stack" them together. This technique has been referred to as "staring imaging" (Gowen, *et al.*, 2007). A staring image can be accomplished in several ways, including (but not limited to) filter wheels, tunable filters, and beam splitters (Figure 16). The second common technique for creating a hypercube is by taking one line of spatial data, with all its spectral information. This technique, called "pushbroom scanning" is done by taking a two-dimensional $X-Z$ image, and then changing the Y-axis (i.e., moving the camera or moving the object) and then taking the next $X-Z$ image. To do this, a spectral dispersion element with a very fine slit (typically 7–80μm) is placed behind the front lens, so that only this small section of the focal plane continues on through to the CCD screen. The single line of data is passed through a dispersion element, such as a prism or grating, which separates out the spectral data before it is projected onto a digital sensor. As the sensor and/or object moves, the next spatial line of data is captured, processed, appended to the previous lines, and the 3-dimensional hypercube is built. Using the "stack of cards" analogy from Figure 16, with a "pushbroom" HSI system, as the cards are placed on the stack, the spatial data (i.e., visible image) would be on the edge of a card, while the spectral data would be along the face of the card. Since there is relative motion between the sensor and the target, this technique is useful for spaceborne and airborne sensors. However, there have been other applications where pushbroom systems have been used in laboratory environments, where motion is added either to the target (Qin, *et al.*, 2008), or to the sensor (Moshou *et al.* (2006) (Figure 17).

D. Wavelength Ranges

The wavelength range of an individual HSI system is dependent upon several factors. Firstly, each digital sensor has a range

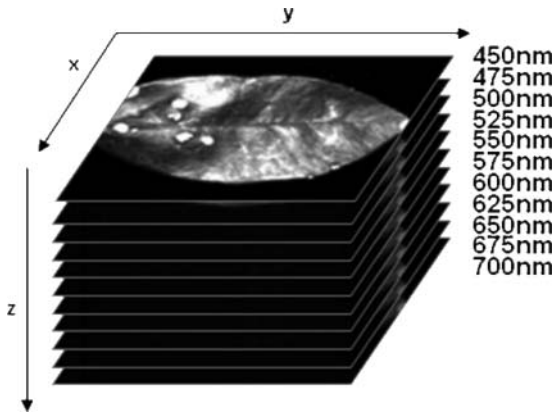

FIG. 16. Representation of how a Staring Image hypercube can be made. Each "layer" is a picture of the same spatial image (in this case, a leaf with lesions) taken with the filter shown to the right. The images are stacked together like a deck of playing cards to form the hypercube. The X and Y axes are spatial, while the Z axis is spectral.

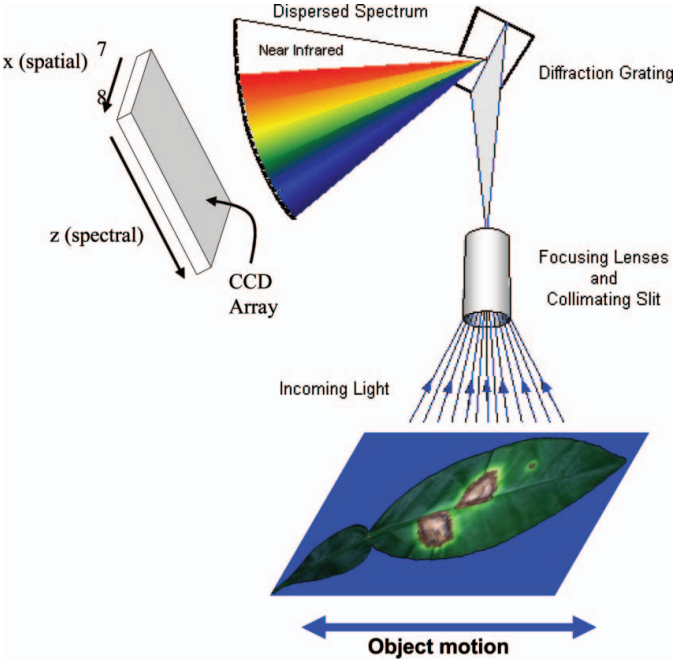

FIG. 17. Graphical representation of how a "pushbroom" hyperspectral scanner operates. As the object is in front of the lens, a small spatial line of data is taken from the center of the image via the collimating slit. This data is diffracted, and then projected on the CCD array. The object moves, and the next spatial line is processed.

to which it is sensitive. For example, a common CCD is sensitive in the visible and near-infrared portions of the spectrum (specifically, 400–1000 nm), while InGaAs (Indium Gallium Arsenide) and MCT (Mercury Cadmium Telluride) arrays are better suited for the shortwave IR (or NIR) region (1000–2500 nm). Other sensors can operate at shorter or longer wavelengths in the electromagnetic spectrum, ranging from the UV (200–400 nm) to thermal IR (up to 30,000 nm/12 μm), but these sensors are usually airborne, or spaceborne systems and have not been used to detect or measure plant disease. Within plant disease detection and severity measurement a wide range of wavelengths have been explored to study various pathosystems including 350–2500 nm (Apan *et al.*, 2005; Delalieux *et al.*, 2007; Liu *et al.*, 2007; Mishra *et al.*, 2007; Yang *et al.*, 2007), 400–850 nm (Huang *et al.*, 2007), and 400–900 nm (Qin *et al.*, 2008).

Another factor that will limit the spectral range of a system is the spectral dispersion optics. Liquid Crystal Tunable Filters (LCTFs) are usually limited in their spectral range (for example, visible only (420–720 nm), or NIR (650–1050 nm). Another popular tunable filter, the Acousto-Optic Tunable Filter (AOTF), also has limited spectral ranges based on the crystal used, but the AOTF has a much faster tuning speed than a LCTF. Prism-Grating-Prisms (PGPs) and holographic grating systems can record a larger wavelength range (400–1000 nm), again, based on the prisms and grating elements used, but the time to take the image is now dependant on how fast the sensor can scan the image area.

E. Spatial Resolution

The spectral and spatial resolution can vary between different systems, based on several factors. For spatial resolution, as with any imaging system, three of these factors are the distance from the object, the number of picture elements (pixels) on the sensor, and the focal length. With commercial quality CCD sensors now over 14-megapixels, and higher quantum efficiency CCD cameras getting up to 4-megapixels, the spatial resolution is getting much higher. For example, the images of the leaf used in Figures 16 and 17 were taken with a laboratory based hyperspectral camera system consisting of a 1.3-megapixel camera, at a height of 45 cm over the leaf, with a 17 mm lens. This produced a spatial size of about 89.1 x 89.8 μm/pixel in the X and Y dimensions, or about 111 x 112 pixels/cm^2. It should be pointed out that this is for one camera set-up in a laboratory, and an airborne system typically would not record such high resolution. However, an airborne system can image a much larger area per image. For example, Huang *et al.* (2007) looked at yellow rust infestation of wheat using an airborne sensor flown at 1000 m and achieved a spatial resolution of 1 x 1 m. Lawrence and Labus (2003) used a HSI system that was flown in a helicopter at 500 m above a Douglas fir stand, and also achieved a 1 x 1 m spatial resolution, with a swath of 500 m to cover an area of 650 x 418 m. Lass and Prather (2004) used a HSI system to detect Brazilian pepper trees in the Florida everglades. By flying at an altitude of 2500 m, they obtained a swath of 2.5 km, and a pixel resolution of 5 x 5 m. On a smaller scale, Qin *et al.* (2008) used a pushbroom type scanner to inspect individual grapefruits for citrus canker. Although the authors did not publish their spatial resolution, based on the reported spatial dimensions (280 x 658 pixels) and given an approximate size of a grapefruit of 10-15 cm, it can be estimated that their spatial resolution was high. Moshou *et al.* (2006) used an HSI system to inspect wheat in the field for yellow rust. Their system was suspended on a cart approximately 1 m above the plants, and had a spatial resolution of 0.07 mm. There are advantages and disadvantages for all systems and system setups – and as with visual assessment and digital photography, care must be taken to get the best system for the application or research to be done.

F. Image Processing Software and Algorithms for Analyzing Hyperspectral Data

As HSI has become more and more popular in different research environments, there are several vendors that have developed imaging processing software for hyperspectral data (Table 10). As mentioned earlier, HSI creates very large amounts of data, of the order of hundreds of Mb per image, so analysis can be difficult. Common image processing algorithms are often not suited for analysis of the highly correlated data contained in a hypercube. As such, several new routines for pre-processing and analyzing these data have become available. Pre-processing typically involves such steps as atmospheric correction, calibration, noise removal, and replacing any known bad data points. It may

TABLE 10
Some commercially available hyperspectral image processing programs.

Hyperspectral imaging processing software	Company	Web Site
ER Mapper	Earth Resources Mapping	www.ermapper.com
EASI/PACE	PCI Geomatics	www.pci.on.ca
ENVI	ITT-Visual Information Systems	www.ittvis.com/envi
ERDAS Imagine	Leica Geosystems Geospatial Imaging, LLC	www.erdas.com
GRASS GIS	Center for Applied Geographic and Spatial Research	http://grass.osgeo.org
IDRISI	Clarke Labs	www.clarklabs.org
PG Steamer	Pixoneer Geomatics	www.pixoneer.com
TNT Mips	MicroImages	www.microimages.com
Image Intelligence™ Suite	Definiens	www.definiens.com
RemoteView	Overwatch Geospatial	www.geospatial.overwatch.com/#

also involve spatial and/or spectral subsetting. Once these steps are complete, analysis of the data can proceed. Although analysis of the data may take many forms, one of the more common techniques is to classify the image pixels based on their statistical similarity to a known spectral "endmember" of interest to the analyst. This is referred to as supervised classification.

1. File Reduction and Subsetting

Due to the large file sizes and highly correlated data contained in hyperspectral imagery, often one of the first steps in analyzing the data is to reduce the data dimensionality. Noise removal and image size reduction can be accomplished in many ways, but two popular methods are principle components analysis (PCA) and the minimum noise fraction (MNF, or Rotation) transformation routines. PCA is a common statistical multivariate analysis that compresses the data by drawing out maximum covariance and removes correlated elements, leaving only uncorrelated output bands. It enables the identification of combinations of the original data wavelength bands that have the most impact on the variation in pixel values within the image (Campbell, 2007). When a PCA is run on an image, the first PC accounts for the most variation, and each subsequent PC accounts for less variation. The MNF is a linear transformation consisting of two sequential PCA rotations. The first decorrelates and rescales the noise in the data (also called "noise whitening"), and the second PCA rotation is applied to the noise-whitened data (Anon., 2006). This process maximizes the signal-to-noise ratio of each PC, so that the noise can be effectively removed from the dataset. Following the MNF transformation, a reverse-MNF returns the data to its original data space. The reverse MNF incorporates only those bands containing coherent images, determined by examining the images and their associated eigenvalues.

2. Spectral Library Definition

After the data reduction, the final step before supervised classification is the definition of the spectral library. The spectral library is a database composed of the "known" spectra of interest that need to be identified in the HSI process. This can be accomplished in a variety of ways. The simplest approach is to select a portion of the image from "known" test data, and average the pixels together. This works well if you have a fairly homogenous sample, but for more heterogeneous samples, or samples with mixed pixels, it is not as effective. One method of finding the most spectrally pure pixels within an image is called the pixel purity index (PPI). The PPI is a technique where random vectors are drawn through the data cloud, and the extreme pixels are identified (Boardman *et al.*, 1995). This process is repeated through thousands of iterations, and the purity of the pixel is expressed as how many times that pixel was classified as "extreme" from the vectors. These extreme pixels are called "endmembers," and are considered to be the most spectrally "pure" pixels in the image. If one of these endmembers in the training data is part of the object that is of interest (such as the disease symptom on the plant), this endmember spectra could be used in the spectral library. The library can include as few as one spectra, or an unlimited number from which an analyst can chose to apply to the test data. For example, if looking to differentiate between two diseases and two other biostressors, as well as the ground, there could be five spectra in the library. Another alternative is to create a system that only looks for the objects that are made up of that one spectrum of interest, i.e., a particular disease symptom, while ignoring everything else in the image.

3. Classifiers

Image classifiers are algorithms or techniques that separate the pixels of a digital image into different groups, or "classes" (Campbell, 2007). There are many different classification routines available in the software listed in Table 10, and there is also the option of creating user-defined algorithms that may incorporate various classifiers, statistical routines, etc. Some algorithms are very simple "band math" type of functions, such as the Normalized Difference Vegetation Index (NDVI), which

is used to discriminate healthy biomass from soil. More detailed information on many of these vegetative indices (VI), and comparisons under different circumstances can be found elsewhere (Elvidge and Chen 1995; Haboudane *et al.*, 2004; Payero *et al.*, 2004; Tilling *et al.*, 2006). Although these VIs are important to the analysis of biomass and leaf area, they are not hyperspectral, but multispectral applications, and therefore will not be addressed in this review. They do represent a whole class of "simple" algorithms that can be done with small sub-sets of HS data. Since there can be hundreds of bands within a hyperspectral image, there are many "band math" style algorithms that can be applied to parts of the data set, using simple arithmetic functions.

Classifiers can be loosely grouped into two types: "supervised" and "unsupervised." Unsupervised classifiers separate an image based on the statistical similarity of the pixels within an image. These classifiers assume no *a priori* knowledge of the scene's endmembers. Thus, the analyst need only enter a few general parameters, such as the number of spectral classes desired. Supervised classifiers require more input from the analyst, as the person analyzing the image must inform the software what to look for. Supervised classification uses either pre-defined spectral libraries (already discussed above) or areas within the image that are "known" to the user to classify the "unknown" pixels within an image (Campbell, 2007). Another way of thinking about a supervised classifier is as an analysis of the image using a training set, which in this case could be a known area of the image, or from another image, in the case of a spectral library. To differentiate a "supervised" from an "unsupervised" classifier, imagine an overhead hyperspectral image of a forest. With an unsupervised classifier, the analyst might tell the computer to break the image into four categories, thinking that the image would have trees, bare ground (earth), water, and rock. Each pixel is then assigned to one of the four classes based on the spectra of that pixel. The grouping of the pixels does not necessarily need to relate to any information that the analyst has; for example, the image might not have any water in it, but the pixels would still be grouped into 4 classes. So, the classification may not give the analyst the information they are looking for.

With a supervised classifier the analyst is able to tell the program to look for a specific spectrum within an image (in the above example, look for water), given that the spectra is already defined, either through a spectral library, or by choosing several pixels within the image to define the spectra to look for. This can be done for as many spectra as desired, so you can search an image for trees, ground, water and rock, assuming the spectra have been defined for those four components. Once the supervised classifier is launched, it will segment the image into either "n" or "n+1" classes, where "n" is the number of spectra defined by the analyst. Some classifiers will also have an unclassified class, for those pixels that don't fit into any of the defined classes.

Some common supervised classifiers that have been used in the processing of hyperspectral data are described in the following subsection. However this list is not meant to be exhaustive. An excellent source of more information and a detailed description of many of these, as well as the methods of calculation is Richards and Jia (2006). Once identified, test images can be classified using these different algorithms, and/or combinations of algorithms.

a. Parallelepiped classifier. One of the simplest classifiers is the parallelepiped classifier. Using the means from each class in the training data as well as their upper and lower limits, boundaries are drawn in the data set that segments the test data into the different classes (Richards and Jia, 2006). For a pixel to be classified into any specific class, it must fit within the boundaries in all the bands within the image. This method does leave open areas where data points (pixels) cannot be classified into any of the defined classes and therefore assigned to an "unclassified" class. Although this method can be used on hyperspectral data, it is better suited for multispectral data.

b. Maximum likelihood classifier. Maximum likelihood (ML) is a classifier which uses the means and variances estimated from the training data to determine the probability that any particular pixel within an image belongs to a particular class (Campbell, 2007). It is considered one of the most commonly used classifiers in image analysis (Richards and Jia, 2006). In this classifier, all pixels are assigned to a class, with no "unclassified" data. One assumption made by this classifier is that the vector pixel is normally distributed with both the mean and variance, where both of these are unknown (Lugo-Beauchamp, *et al.*, 2004).

c. Mahalanobis distance classifier. A derivative of the ML classifier is the Mahalanobis distance classifier. This algorithm works in a similar fashion to the ML, but its main assumption is that the covariance matrix of the classes is equivalent. Because of this assumption, the Mahalanobis distance classifier is much faster than the ML (Anon., 2006).

d. Linear spectral unmixing classifier. Linear spectral unmixing (LSU) is an analysis based on the linear spectral mixing model, which holds that any individual pixel is made up of composite spectra, not pure spectra. This is due to the fact that sensor resolution is not fine enough to allow for only one object in a pixel (Campbell, 2007). If the dispersal of objects within the pixels field of view is high, the composite is said to be nonlinear. LSU assumes that it is possible to identify the components and their quantities of any mixed pixel, if you know the pure spectra of the components, such as from laboratory work.

e. Minimum distance classifier. The minimum distance classifier, also referred to as the Euclidean distance classifier, calculate the distance of each vector pixel to the mean vector of each class, using Euclidean metrics. Each vector pixel is then assigned to the closest class, and as with the ML classifier, there is no unclassified data (Lugo-Beauchamp, *et al.*, 2004).

f. Matched filter classifier. The matched filter classifier is a statistical algorithm that allows for the ability to identify a single, known class without the influence, or even knowing, any other endmember signatures (Boardman, 1998). Boardman

(1998) combined this classifier with the linear spectral mixing model, creating the mixture tuned matched filter classifier (MTMF), which "combines the best parts of the LSM model and the statistical MF model while avoiding the drawbacks of each parent method." The MTMF has the ability to classify a single known target vector pixel without knowing the background endmember signatures, without being constrained to using just pure pixels.

g. Spectral angle mapper classifier. The spectral angle mapper classifier maps the pixel vectors to a known vector from a spectral library, by converting the image to an n-dimensional data space, where n is the number of bands in the image (Kruse *et al.*, 1993). Once the data is in this n-D space, the pixel vectors are compared to the known vectors, and if the pixel vector is within a certain angle of the known, it is classified as that substance. If the pixel vector is not near enough to any of the known vectors, it is unclassified. This method is not confused by illumination differences, as the brightness of the pixel will only affect the length of the pixel vector when that data is converted in to n-D space, and the angle between the vectors will remain constant.

h. Neural net classifier. The neural net (NN) classifier is a layered feed-forward network classification that incorporates supervised learning (Richards and Jia, 2006). NN's can be seen as a more attractive system for classifying images based on their spectra when compared to the previously mentioned supervised classifiers, based on the learning ability of the system (Egmont-Petersen *et al*., 2002). One major difference between an image processing NN and a standard NN is that the threshold logic units (TLU) customarily seen are actually processing elements, which are similar but do not apply a thresholding operation. Rather, they use a mathematically differentiable operation instead (Richards and Jia, 2006). The network learns by minimizing the difference between the output node activation and the output (Anon., 2006).

i. Binary encoding classification. Binary encoding is an algorithm that converts the spectra in the library into 0's and 1's by comparing each point along the spectral curve to the spectral mean. If the point falls below the mean, it is assigned a value of "0", and if above the mean, a "1". Then, after all the data is encoded in this method, test spectra are compared to the library spectra based on how many bands (points) match the library spectra (Anon., 2006).

G. Application in Plant Pathology

HSI is a relatively young science, and its application in plant pathology is particularly recent, even more so in plant disease severity assessment. Several of the studies that have been done have been concerned more with disease detection, rather than quantification and these are considered in this review as they demonstrate the current state of this science as it is applied and is developing. The use of HSI for quantifying disease has yet to be fully explored, or reach its full potential, but the relatively few occasions where it has been used suggest that it can be a powerful disease assessment tool.

1. Detection of Disease

Using HSI to detect diseases in plants has now been tested on several plant species and HSI has shown a convincing ability to detect disease in infected crops. This work has been done mostly over the past ten years. In one of the earliest studies, HSI was used to detect yellow rust in wheat fields (Bravo, *et al.*, 2003), using a hand pushed rolling cart that held the spectrometer (a PGP system) approximately 1 m above the ground, with a detection success rate of 96%. Bravo *et al.*, (2004) developed their detection system further, by adding multispectral fluorescence imaging (MSFI) capability to the HSI system previously used. This was accomplished by having a 4-band multispectral camera, a Xenon arc lamp with a low pass filter, and a shade, to minimize the effects of ambient light. The ability of each system to detect disease within a plot was tested independently, using quadratic discriminant analysis, which is based on the Mahalanobis distance between a single observation and the class (healthy or diseased) mean. Finally, they combined the best wavelengths from each system, analyzed that data with the quadratic discriminant analysis, and differentiated healthy plants 98% of the time and diseased plants 91% of the time, demonstrating a promising ability to detect disease. In another study, Apan *et al.* (2005) used a spectrophotometer (350-2500 nm) to detect early blight in tomatoes. After removing the noise in the spectra, they analyzed the data using a Partial Least Squares regression and demonstrated good spectral separability between infected and healthy tissue, based on the "red-edge" and green regions of the visible spectrum. The red edge is a spectral phenomenon for plants that produce chlorophyll, and is defined by a massive increase in reflectance at approximately 700 nm, or just on the border of red and infrared in the electromagnetic spectrum (Seager *et al.*, 2005). As a leaf is stressed, this rapid uprise in the reflectance shifts downward, towards the red region of the spectrum. This is called the "blue shift" of the red edge (Rock, *et al.*, 1988). Apan *et al*. (2005) also suggested that parts of the NIR region would also be useful in differentiating diseased plants from healthy plants. Further work, using a first derivative showed reasonable prediction ability ($\sim$80%). Similar results were found for ladybird infestation in eggplant, in the same experiment. In other studies, HSI has been applied to detect and quantify systemic diseases in the field, which is an incidence measure rather than a severity measure but still quantifies disease in the case of systemic pathogens. For example, Mishra *et al.* (2007) used a spectroradiometer to look at the spectral characteristics of a Huanglongbing (HLB) infection in a citrus field. Using information from the 350–2500 nm range of the electromagnetic spectrum, they identified the most important wavelengths and used two different analyses (a spectral derivative analysis, and a spectral ratio analysis) to find the wavelengths that provided the most discriminability between healthy citrus and HLB infected plants. Two ratios were sensitive to differences between

HLB positive and negative trees, with an ability to discriminate HLB positive from negative trees using wavelengths between 585–705 nm. Qin *et al.* (2008) developed a custom HSI system based on PGP technology, which gave them a usable spectral range of 400–900 nm. They used it to find and differentiate citrus canker on grapefruit, emulating a processing line set-up, using a simple PCA and threshold for the classification. Cankered fruit were differentiated from healthy fruit and other diseases 93% of the time, even after treatments of chlorine, soap washes, and a wax coating. For a new technique, HSI has thus shown an encouraging capability to detect diseases in the field, and on individual plants. Only recently has this discerning ability to detect disease been harnessed for disease quantification as well.

2. Using HSI to Assess Disease Severity

The earliest attempts to use HSI for measuring disease severity were performed on a plantation scale. Coops *et al.* (2003) used the CASI-2, airborne based HSI system to assess severity of needle blight in Australian Pine, and applied a six-point category scale (1 = no visible signs of infection, to 6 = severe defoliation with 80-100% of the whorls showing evidence of blight) for severity based on visual assessment from the ground. Comparing the visual data to data obtained from the CASI-2, they tracked the blight score versus three indices, upper and lower slope red edge indices, and a red edge vegetation stress index (RVSI). Since the spatial resolution of the CASI-2 is not at a leaf level (they reported it to be 0.8 m), they looked at 2 sections of the canopy – a halo which was the pixels around the center of the tree (essentially eliminating the trunk region), and the mean for pixels that covered the canopy. The halo measurement seemed to have a higher correlation with the higher blight scores (higher infection), whereas the mean measurement were better correlated to the lower scores, although not as significant. Of the three indices, the RVSI had the most significant correlation. They operationally defined accuracy as the proportion of omission and commission errors in the different classes, and found that accuracy of the HSI system was dependent on the disease class, with the greatest accuracy (40%) in class 4. When they combined the data into three categories (low, medium and high infection) instead of six, the accuracy of classification increased to $> 70\%$ in all classes. At a smaller scale, Muhammed and Larsolle (2003) did a detailed examination at the effect of fungal growth on the spectral signature of wheat plants, using four different severity levels on a visual scale. Firstly, they reduced the dimensionality of the data by running either a PCA or an independent component analysis (ICA), a different multivariate linear transformation on the training datasets. They then ran these transformations through a multivariate analysis method - feature vector based analysis (FVBA) - which produced one-dimensional vectors, which could then be compared against unknown datasets (images). They found that the coefficient of determination for the relative disease intensity was 95.7% for a second order polynomial, on either a PCA or the ICA, followed by the FVBA when comparing the spectral classification to visual assessment. As the severity of disease increased, there was a flattening of the spectra at the green peak, and a decrease in the shoulder of the plateau of the NIR region. In a later article, Larsolle and Muhammed (2007) did a similar study on wheat and barley, but simplified their analyses by removing the PCA and FVBA, and classifying the images using a nearest neighbor classifier. This classifier compared the known data vectors to the unknowns, and looked for the highest correlation coefficient with the lowest sum of squares between the two vectors. Allowing a 10% error in the results, the ability to correctly classify the measured disease severity compared to the visual assessments was 86.5%. Investigating yellow rust disease in wheat, Huang, *et al.* (2007) used an airborne system that acquired VIS-NIR data (400–850 nm) and compared the resulting photochemical reflectance index (PRI) (a type of VI) against a visual nine-point classification system, based on percentage of the plant covered in rust. With regression analysis they were able to show that PRI was fairly reliable compared to visual assessments for determining severity (coefficient of determination, $r^2 = 0.91$). Using the HSI system at the scale of individual leaves, Delalieux *et al.* (2007) measured foliar symptoms of apple scab using a spectroradiometer (350–2500 nm) and a 100 W halogen lamp on individual apple leaves. The results were compared against a previously developed 6-level visible scale (Chevalier *et al.*, 1991). Using various statistical analyses, and multiple wavelength ranges, they were able to show that "early detection of biotic plant stress using hyperspectral remote sensing has potential," with ability to correctly classify ranging from 80 to 90%. Liu *et al.* (2007) used a spectroradiometer to determine brown spot disease severity in rice at 350–2500 nm. After reducing the data to just 3 wavelengths using multiple stepwise linear regressions, they were able to get coefficients of determination between observed and predicted disease severity of 0.94, with a root mean square error (RMSE) of 5.8%. Although not directly comparable, as these data were not compared to actual values, this regression suggests similarity to reports by Nutter *et al.* (1993) using a radiometric method to measure disease severity in plots of bent grass infected with dollar spot. Liu *et al.* (2007) also examined the data using PCA, and found that when using just the first two PCs, the coefficient of determination was 0.63 between observed and predicted disease severity. However, a PLS analysis of the data gave the best results ($r^2 = 0.97$, RMSE of 2%). It is worth mentioning that spectroradiometers have also been used to measure the intensity of infestation with insect pests, which requires the similar needs of relative quantification of particular spectra to estimate disease severity (Yang *et al.*, 2007). Although not a phytopathological example, a spectroradiometer was used to examine leaf-folder infestations in rice plants, and compared the spectral responses against a 9 point percent scale. After recording the data between 350–2500 nm, the information was analyzed using a multiple linear regression, and various VI's. Their results indicated that the best linear regression for plants in the active tillering stage included six different wavelengths between 757 and

1963 nm, and produced a coefficient of determination of 0.86. When examining the plants during the heading stage of development, the linear regression to correctly score the infestation intensity used only four wavelengths between 517–974 nm, and showed good precision in relation to the 9 point percent scale ($r^2 = 0.96$).

3. Ability to Discern Multiple Disease Symptoms

Most of the research on the ability of HSI to detect biotic stress in plants only dealt with a single, specific biotic stressor, either a single disease or a single nutrient deficiency, when compared to healthy plant tissue. One exception is the study by Moshou *et al.*, (2006). They were able to distinguish between a yellow rust infection and a nitrogen deficiency in wheat using up to five different wavelengths in the visible and NIR sections of the electromagnetic spectrum, using a neural net classifier. Disease was measured approximately 1 m above the plants while on a rolling cart, and the data analyzed with a standard multispectral algorithm (NDVI) to separate leaf material from the rest of the canopy, followed by an image normalization calculation to normalize the light intensity at each leaf, and the data compressed by determining the important bands via ANOVA. From these wavelengths, the three different criteria (healthy, diseased, and nitrogen deficient) were separated with two different classification techniques. First, they applied a quadratic discriminant analysis (QDA), and then used a second classification technique, a self-organizing map neural net. The best results were obtained using five wavelengths, which led to agreement of 95–100%, although they mention that "diseased canopies show high spectral variations, causing the necessity of spatial averaging" (Moshou *et al.,* 2003). Multiple diseases have also been detected on citrus with HSI. Qin *et al.* (2008) were not only able to differentiate citrus canker on grapefruit from healthy fruit 93% of the time, they also were able to differentiate healthy from unmarketable conditions such as copper burn, and various diseases including greasy spot and melanose 92% of the time. This demonstrates the potentially very useful ability of HSI to differentiate at least some disease stresses based on spectral reflectance. Qin *et al.* (2009) improved their ability to differentiate canker infected grapefruit from several other types of disease symptoms by using an image classification technique called spectral information divergence (SID). SID uses the spectra vector like the SAM, but in the SID, the vector is modeled as a probability distribution. With this technique, Qin *et al.* (2009) were able to correctly separate the canker infected fruit 96% of the time. This research shows that there is potential in certain areas of plant disease detection to use HSI to identify one disease from another.

Thus although HSI has been investigated as a research tool in the discrimination and detection of plant diseases, and has the potential to be an important tool for the assessment of disease severity, it is yet to be applied as a research tool or used in a commercial application for plant disease detection or quantification. This could be due in part to the fact that when a plant is stressed, either by an infection or by other biostressors such as a nutrient deficiency, insect invasion, and physical damage, the spectral response can be similar. Therefore, it remains challenging to differentiate the cause of the stress to the plant, based solely on the spectral response.

H. The Future of HSI for Disease Severity Measurement

The case for hyperspectral imaging for measuring disease severity is not yet proven, although this technology has potential that has yet to be fully explored. For example, the neural net algorithms have shown some promise in discrimination. Other algorithms can also be developed that are more specialized for the detection of plant diseases. It has been shown to have the capacity to differentiate some diseases. Furthermore, there are also many other areas of the electromagnetic spectrum, including UV, mid- and far-IR, and even up in the thermal bands, that have not been tested, and which could turn out to be more discriminating in the differentiation and assessment of plant diseases using HSI.

VII. ADVANTAGES AND DISADVANTAGES OF VISUAL RATING, IMAGE ANALYSIS, AND HYPERSPECTRAL IMAGING

Advantages of visually assessed disease

a) The process can be quick.
b) With some training it is relatively easy to recognize and differentiate multiple diseases.
c) The use of assessment aids and training markedly improves results.
d) There are several techniques that can be used to suit a particular need (ordinal scales, interval scales, category scales and ratio scales).
e) No equipment required.

Disadvantages of visually assessed disease

a) Raters may tire and lose concentration, thus decreasing their accuracy.
b) There can be substantial inter- and intrarater variability (subjectivity).
c) There is a need to develop standard area diagrams to aide assessment.
d) Training may need to be repeated to maintain quality. Raters are expensive.
e) Visual rating can be destructive if samples are collected in the field for assessment later in the laboratory.
f) Raters are prone to various illusions (for example, lesion number/size and area infected).

Advantages of digital photography and image analysis in the visible spectrum

a) Image analysis can be quick, accurate and reliable when automated.

b) If a good automated system can be developed (there are a few reported instances of agreement verification) then it can be extremely powerful.
c) Technology exists to make the assessment both reliable and accurate.
d) Image analysis equipment is relatively inexpensive.
e) There is specific software adapted for the applications and specific needs and issues in plant disease measurement.

Disadvantages of digital photography and image analysis in the visible spectrum

a) Coping with plant-to-plant variation in color and various image artifacts or flaws is not straightforward.
b) Not established how to deal with multiple diseases, damage or physiologic conditions on sample leaves.
c) It requires some training in the program to become proficient.
d) Truthing is often required to ensure the quality of the measurement.

Advantages of hyperspectral imagery

a) Massive amounts of data and information about the target is acquired at one time.
b) If a good automated system can be developed (very few reported) then it can be particularly discerning.

Disadvantages of hyperspectral imagery

a) Enormous file size, which can be slow to capture and process,
b) Still a new technology, not fully tested or adapted to the needs of plant disease severity assessment.
c) Not established how to deal with multiple diseases but offer more possibilities.
d) It is expensive.
e) Substantial training and expertise is required to use it to its full potential.

VIII. SOME FUTURE RESEARCH PRIORITIES IN VISUAL ASSESSMENT, THE APPLICATION OF HYPERSPECTRAL IMAGING AND IMAGE ANALYSIS FOR MEASURING DISEASE SEVERITY

1. A comparison or different interval and ratio scales, to compare reliability, agreement and the time taken to assess disease so as to provide a more objective framework for choosing scales for specific purposes, and to confirm advantages and disadvantages of these systems.
2. What is an optimum scale? Can simulation modeling approaches be used to help investigate the effects of scale and rater ability on the mean and standard deviation of the estimates to investigate and develop ideal scales for different purposes?
3. How long do the benefits of a computer training session last? Is this individual dependent? Are there ways to make this more effective and permanent?
4. SADs are often logarithmically based. The ability to compare different approaches to choosing SAD severities exists. Is there any difference in the quality of the estimate if they are linear compared to logarithmic, and is there any merit to having a lower disease severity more highly represented?
5. Do visual raters show a constant estimation error with actual disease? Raters clearly vary tremendously in ability for several reasons, but to understand what underlying propensities exist will require a comprehensive approach involving multiple raters and using standard psychophysical methods across the full spectrum of disease severity. It remains unconfirmed which disease severities are less well estimated.
6. Is there any difference in error when estimating disease severity based on taking digital images or visual assessment at different scales? Are leaves, plants or quadrats the best units? How does the mean severity, and error in the severity estimate compare?
7. Compare different methods of automating disease assessment (direct thresholding/using algorithms to recalibrate thresholds for each image) to gain an insight into the best ways to measure disease, and the benefits accrued by incorporating any new techniques.
8. Can ways be developed that will allow image analysis to differentiate diseases on the same leaf or within the same canopy using the RGB model?
9. How many samples are required to gain a realistic estimate of the mean using different methods? Assuming an adequate sampling strategy, the more accurate and precise a technique, presumably the smaller the samples needed to provide an accurate estimate of the mean disease, while reflecting the actual variability of the population.
10. Can hyperspectral imaging, either in its current form factor or in other systems, provide a platform by which several diseases can be differentiated simultaneously, with no interference from other biostressors?
11. Can multispectral systems (3–20 bands) that can differentiate disease be developed? These systems might be significantly cheaper than current hyperspectral systems.

IX. CONCLUSIONS

Over the last thirty years there have been huge advances in understanding plant disease assessment and applying new technologies, particularly image analysis, and more recently hyperspectral imagery. Both of these are likely to play an increasing role in the assessment of disease severity. The science is becoming defined and the concepts of disease measurement are being explored, and many of the common sources of error in visual rating have been identified. Inter- and intrarater variability exists, and the advantages and disadvantages of visual disease assessment methods have begun to make themselves understood on an empirical basis. The value of training and SADs has been demonstrated to improve the quality and consistency of visual rating. Objective, newer methods of

measuring severity of disease are now more commonplace and being explored and include image analysis and hyperspectral imagery. Image analysis can deliver results with high agreement to actual values, even when automated. Its objectivity and ability to deliver high throughput is advantageous. Both are useful additions and have applications for the measurement of disease severity at several different scales from individual leaves and other organs to remote aerial and satellite images.

ACKNOWLEDGMENTS

We appreciate the time, useful comments and helpful suggestions of several reviewers in reading draft versions of this review. Professor Forrest W. Nutter was also extremely generous with his time reviewing the manuscript and making numerous suggestions that have improved the content and readability – we are particularly grateful for his input. Ute Albrecht is thanked for help with the German translation of various documents. We also appreciate permission from different sources to reproduce figures that were used to illustrate specific points in the article.

REFERENCES

Ahmad, I. S., Reid, J. F., Paulsen,M. R., and Sinclair, J. B. 1999. Color classifier for symptomatic soybean seeds using image processing. *Plant Dis.* **83:** 320–327.

Amanat, P. 1976. Stimuli effecting disease assessment. *Agric. Conspectus Scientificus* **39:** 27–31.

Amanat, P. 1977. Modellversuche zur Ermittlung individueller und objektabhgiger schätzfehler bei pflanzenkrankenheiten. Diss. Universität Gießen. Cited in: Hau, B., Kranz, J., and Konig, R. 1989. Fehler beim Schätzen von Befallsstärken bei Pflanzenkrankheiten. *Z. Pflkrankh. Pflschutz.*: **96:** 649–674.

Andrade, G. C. G., Alfenas, A.C., Mafia, R.G., Maffia, L.A., and Gonçalves, R.C. 2005. Diagrammatic scale for assessment of eucalyptus leaf spot severity caused by *Quambalaria eucalypti*. *Fitopatologia Bras*. **30:** 504–509.

Anon. 1917. Miscellaneous business. *Phytopathology* 7: 149.

Anon. 1947. The measurement of potato blight. *Trans. Brit. Mycol. Soc.* **31:** 140–141.

Anon. 1976. *Manual of Plant Growth Stages and Disease Assessment Keys*. Ministry of Agriculture, Fisheries and food (Publications). Middlesex. UK.

Anon. 2004. *Your ACD guide to digital photography*. 142 p. http://files.acdsystems.com/english/registration/going-digital.pdf

Anon. 2006. *ENVI User's Guide*. Ver. 4.3., ITT Visual Information Solutions, Boulder, CO.

Apan, A., Datt, B., and Kelly, R. 2005. Detection of pests and diseases in vegetable crops using hyperspectral sensing: a comparison of reflectance data for different sets of symptoms. **In**: *Proceedings of SSC 2005 Spatial Intelligence, Innovation and Praxis: The National Biennial Conference of the Spatial Sciences Institute*. pp. 10–18. Spatial Sciences Institute, Melbourne, Australia.

Aubertot, J. N., Sohbi, Y., Brun, H., Penaud, A., and Nutter, F. W. 2006. Phomadidacte: a computer-aided training program for the severity assessment of phoma stem canker of oilseed rape. "Integrated Control in Oilseed Crops" *IOBC Bulletin* **29:** 247.

Bacchi, L. M., Berger, R. D., and Davoli, T. A. 1992. Color digitization of video images of bean leaves to determine the intensity of rust caused by *Uromyces appendiculatus*. *Phytopathology* **82:** S1162.

Baird, J. C. and Norma, E. 1978. *Fundamentals of Scaling andPpsychophysics.* Wiley, New York, NY.

Barnhart, H. X., Haber, M. J., and Lin, L. I. 2007. An overview on assessing agreement with continuous measurements. *J. Biopharm. Stat.* **17:** 529–569.

Batzer, J. C., Gleason, M. L., Weldon, B., Dixon, P. M., and Nutter, F. W. 2002. Evaluation of postharvest removal of sooty blotch and flyspeck on apples using sodium hypochlorite, hydrogen peroxide with peroxyacetic acid, and soap. *Plant Dis.***86:** 1325–1332.

Bauer, M. E., Mroczynski, R. P., MacDonald, R. B., and Hoffer, R. M. 1971. Detection of southern corn leaf blight using color infrared aerial photography. **In:** *Proc. Third Biennial Workshop on Color Aerial Photography in the Plant Sciences*, Gainsville, FL, pp. 114–126.

Beaumont, A, Marsh, R. W., Bescoby, H. B., and Brierley, W. B. 1933. Symposium and discussion on the measurement of disease intensity. *Trans. Brit. Mycol. Soc.* **18:** 174–186.

Belasque Junior, J., Bassanezi, R. B., Spósito, M. B., Ribeiro, L. M., Jesus Júnior, W.C. de, and Amorim, L. 2005. Escalas diagramáticas para avaliação da severidade do cancro cítrico. *Fitopatologia Bras*. **30:** 387–393.

Beresford, R. M. and Royle, D. J. 1991. The assessment of infectious disease for brown rust (*Puccinia hordei*) of barley. *Plant Path.* **40:** 374–381.

Berger, R. D. 1980. Measuring disease intensity. **In:** *Proc. E.C. Stakman Commemorative Symposium on Crop Loss Assessment*. University of Minnesota Misc. Publ. 7, St Paul. pp. 28–31.

Bergh, J. C. 2001. Ecology and aerobiology of dispersing citrus rust mites (Acari:*Eryophyidae*) in central Florida. *Environ. Entomol*. **30:** 318–326.

Biernacki, M. and Bruton, B., 2001. Quantitative response of *Cucumis melo* L. inoculated with root rot pathogens. *Plant Dis.* **85:** 65–70.

Birnbaum, M. H. 1994. Psychophysics. **In:** *Encyclopedia of Human Behavior*, San Diego: Academic Press.

Blanchette, R. A. 1982. New technique to measure tree defect using an image analyzer. *Plant Dis.* **66:** 394–397.

Bland, J. M. and Altman, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **68:** 307–310.

Bland J. M. and Altman, D. G. 1999. Measuring agreement in method comparison studies. *Stat. Meth. Med. Res.* **8:** 135–160.

Blasquez, C. H. and Edwards, G. J. 1985. Image analysis of tomato leaves with late blight. *J. Imag. Tech.* **3:** 109–112.

Boardman, J. W. 1998. Leveraging the high dimensionality of AVIRIS data for improved sub-pixel target unmixing and rejection of false positives: mixture tuned matched filtering. **In**: *Summaries of the 7th Annual JPL Airborne Geoscience Workshop*. JPL Publication 9721, p. 55.

Boardman, J. W., Kruse, F. A., and Green, R. O. 1995. Mapping target signatures via partial unmixing of AVIRIS data: **In**: *Summaries, Fifth JPL Airborne Earth Science Workshop*. JPL Publication 95-1, v. 1, pp. 23–26.

Bock, C. H. and Jeger, M. J. 1996. Downy mildew of sorghum. *Int. Sorgh. and Mill. News.* **37:** 33–51.

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

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

Bock, C. H., Parker, P. E., Cook, A. Z., and Gottwald, T. R. 2009a. Automated image analysis of the severity of foliar citrus canker symptoms. *Plant Dis.* **93:** 660–665.

Bock, C. H., Parker, P. E., Cook, A. Z., Riley, T., and Gottwald, T. R. 2009b. Comparison of assessment of citrus canker foliar symptoms by experienced and inexperienced raters. *Plant Dis.* **93:** 412–424.

Bock, C. H., Gottwald, T. R., Parker, P. E. and Cook, A. Z., Ferrandino, F., Parnell, S., and van den Bosch, F. 2009c. The Horsfall-Barratt scale and severity estimates of citrus canker. *Euro. J. Plant Path.* **125:** 23–38.

Bock, C. H., Gottwald, T. R., Parker, P. E., Ferrandino, F., Welham, S., van den Bosch, F. and Parnell, S. 2009d. Some consequences of using the Horsfall-Barratt scale for estimating disease severity compared to nearest percent estimation. **In:** *Proc. of the 10th Int. Workshop Plant Dis. Epid.*, University of Cornell, Geneva, NY, Pp 20–22.

Boso, S., Santiago, J. L. and Martínez, M. C. 2004. Resistance of eight different clones of the grape cultivar Albariño to *Plasmopara viticola*. *Plant Dis.***88:** 741–744

Bravo, C., Moshou, D., West, J., McCartney, A., and Ramon, H. 2003. Early disease detection in wheat fields using spectral reflectance. *Biosystems Engineering*. **84:** 137–145.

Bravo, C., Moshou, D., Oberti, R., West, J. McCartney, A., and Ramon, H. 2004. Foliar disease detection in the field using optical sensor fusion. *Agricultural Engineering International: the CIGR Journal of Scientific Research and Development*. Manuscript FP 04 008. **Vol. VI**.

Brenchley, G. H. 1964. Aerial photography for the study of potato late blight epidemics. *World Rev. Pest Cont*. **3:** 68–84.

Breusch, T. S. and Pagan, A. R. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **47:** 1287—1294.

Bronson, C. R. and Klittich, W. M. 1984. Phytoscan 83: a computer program for quantitative disease assessment. *Phytopathology* **74:** 871.

Bruton, B. D., Garcia-Jimenez, J., Armengol, J., and Popham, T. W. 2000. Assessment of virulence of *Acremonium cucurbitacearum* and *Monosporascus cannonballus* on Cucumis melo. *Plant Dis.* **84:** 907–913.

Burdon, J. J., Thrall, P. H., and Ericson, L. 2006. The current and future dynamics of disease in plant communities. *Ann. Rev. Phytopathol.* **44:** 19–39

Bussotti, F., Schaub, M., Cozzi, A., Krauchi, N, Ferretti, M, Novak., K., and Skelly, J. M. 2003. Assessment of ozone visible symptoms in the field: perspectives of quality control. *Env. Pollut.* **125:** 81–89.

Campbell, C.L. and L.V. Madden. 1990. *Introduction to Plant Disease Epidemiology*. John Wiley and Sons, New York, NY..

Campbell, J. B. 2007. *Introduction to Remote Sensing*. 4^{th} ed., Guilford Press, New York, NY.

Chaube, H. S. and Singh, U. S. 1991. *Plant Disease Management: Principles and Practices*. CRC Press, Boca Raton, Florida.

Chester, K. S. 1950. Plant disease losses: their appraisal and interpretation. *Plant Dis. Rep.* Supplement **190–198** (S193): 190–362.

Chevalier, M., Lespinasse, Y., and Renaudin, S. 1991. A microscopic study of the different classes of symptoms coded by the *Vf* gene in apple for resistance to scab (*Venturia inaequalis*). *Plant Pathol*. **40:** 249–256.

Christ, B. J. 1991. Effect of disease assessment method on ranking potato cultivars for resistance to early blight. *Plant Dis.* **75:** 353–356.

Cobb, N. A. 1892. Contribution to an economic knowledge of the Australian rusts (*Uredinae*). *Agric. Gazt*. (NSW) **3:** 60.

Colwell, R. N. 1956. Determining the prevalence of certain cereal crop diseases by means of aerial photography. *Hilgardia* **26:** 223–286.

Cooke, B. M. 2006. Disease assessment and yield loss. **In:** *The Epidemiology of Plant Diseases*. B. M. Cooke, D. Gareth Jones and B. Kaye (Eds.) Second edition. The Netherlands: Springer.

Coops, N., Stanford, M., Old, K., Dudzinski, M., Culvenor, D., and Stone, C. 2003. Assessment of Dothistroma needle blight of Pinus radiate using airborne hyperspectral imagery. *Phytopathology*. **93:** 1524–1532.

Cope, P. 2002. *The Digital Photographer's Pocket Encyclopedia*. Silver Pixel Press, Rochester, NY.

Copes, W. E., Chastagner, G. A., and Hummel, R. L. 2003. Toxicity responses of herbaceous and woody ornamental plants to chlorine and hydrogen dioxides. Online. *Plant Health Progress* doi:10.1094/PHP-2003-0311-01-RS.

Corkidi, G. Balderas-Ruiz, K.A. Taboada, B. Serrano-Carreon, L. and Galindo, E. 2006. Assessing mango anthracnose using a new three-dimensional image-analysis technique to quantify lesions on fruit. *Plant Path.* **55:** 250–257.

Couture, L. 1980. Assessment of severity of foliage diseases in cooperative evaluation tests. *Can. Plant Dis. Surv* **6:** 8–10.

Critchley, L.A., Peng, Z. Y., Fok, B. S., Lee, A., and Phillips, R. A. 2005. Testing the reliability of a new ultrasonic cardiac output monitor, the USCOM, by using aortic flowprobes in anesthetized dogs. *Anesth Analg*. **100:** 748–753.

Croxhall, H. E., Gwynne, D. C. and Jenkins, J.E.E. 1952a. The rapid assessment of apple scab on leaves. *Plant Path.* **1:** 39–41.

Croxhall, H. E., Gwynne, D. C. and Jenkins, J.E.E. 1952b. The rapid assessment of apple scab on fruit. *Plant Path.* **1:** 89–92.

Danielsen, S. and Munk, L. 2004. Evaluation of disease assessment methods in quinoa for their ability to predict yield loss caused by downy mildew. *Crop Prot.* **23:** 219–228.

de Jong, S. M. and Van de Meer, F.D. (Eds.). 2006. *Remote Sensing Image Analysis: Including the Spatial Domain. Bookseries on Remote Sensing Digital Image Processing* Vol.5. Kluwer Academic Publishers, Dordrecht. ISBN: 1-4020-2559-9, 359 pp.

Deen, H. G., Zimmermann, R. S., Lyons, M. K., Mcphee, M. C., Verheijde, J. L., and Lemens, S. M. 2000. Test-retest reproducibility of the exercise treadmill examination in lumbar spinal stenosis. *Mayo Clin. Proc.* **75:** 1002–1007.

Delalieux, S, van Aardt, J., Keulemans, W., Schrevens, E., and Coppin, P. 2007. Detection of biotic stress (*Venturia inaequalis*) in apple trees using hyperspectral data: non-parametric statistical approaches and physiological implications. *Eur. J. Agronomy*. **27:** 130–143.

Diaz-Lago, J. E., Stuthman, D. D., and Leonard, K. J. 2003. Evaluation of components of partial resistance to oat crown rust using digital image analysis. *Plant Dis.* **87:** 667–674.

Diéguez-Uribeondo, J., Förster, H. and Adaskaveg J. E. 2003. Digital image analysis of internal light spots of appressoria of *Colletotrichum acutatum*. *Phytopathology* **93:** 923–930.

Dixon, G. R. and Dodson, J. K. 1971. Assessment keys for some diseases of vegetable, fodder and forage crops. *J. Nat. Inst. Agr. Bot.* (GB) **23:** 299–307.

Donzelli, B.G.G. and Churchill., A.C.L. 2007. A quantitative assay using mycelial fragments to assess virulence of *Mycosphaella fijiensis*. *Phytopathology* **97:** 916–929.

Edwards, G. J., Balzquez, C. H. and Miller, J. 1985. Preliminary experiments with remote sensing to detect citrus canker. *Proc. Fla. State. Hort. Soc.* 98: 16–18.

Egmont-Petersen, M., deRidder, D., and Handels, H. 2002. Image processing with neural networks. *Pattern Recognition*. **35:** 2279–2301.

Ehrenstein, W. H. and Ehrenstein, A. 1999. Psychophysical methods. **In:** *Modern Techniques in Neuroscience Research*. Berlin: Springer. Pp 1211–1241.

Elvidge, C. D. and Chen, Z. 1995. Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sens. Env.* **54:** 38–48.

Evans, C. K. and Pope, J. 2006. Digital image analyses of primary leaf lesions on wheat seedlings of Frontana and Alsen inoculated with *Fusarium graminearum*. *Phytopathology* **96:** S33.

Everitt, B. S. 1998. *The Cambridge Dictionary of Statistics*. pp 360. Cambridge University Press. Cambridge, UK.

Eyal, Z., and Brown, M. B. 1975. A quantitative method for estimating density of *Seprtoria tritici* pycnidia on wheat leaves. *Phytopathology* **66:** 11–14.

FAO. 1971. *Crop loss assessment methods. FAO manual on the evaluation and prevention of losses by pests, diseases and weeds.* Ed. L. Chiarappa. UNFAO, Rome Italy/ Commonwealth Agricultural Breaux, Slough, England.

Fisher, J., and Welch, W. C., 2006. Survey and Analysis of Fore-Optics for Hyperspectral Imaging Systems. **In**: *Proceedings of SPIE Defense and Security Symposium Conference, Infrared Technology and Applications XXXII*, Orlando, Florida, Vol. 6206(1):6206R.

Fitt, B.D.L., Lysandrou, M., and Turner, R. H. 1982. Measurement of spore-carrying splash droplets using photographuic film and an image analyzing computer. *Plant Path.* **31:** 19.

Forbes, G. A. and Jeger, M. J. 1987. Factors affecting the estimation of disease intensity in simulated plant structures. *Z. Pflkrankh. Pflschutz.* **94:** 113–120.

Forbes, G. A. and Korva, J. T. 1994. The effect of using a Horsfall-Barratt scale on precision and accuracy of visual estimation of potato late blight severity in the field. *Plant Path.* **43:** 675–682.

Garling, J., Rimelspach, R., and Boehm, M. J. 1999. A comparisoin of four methods to accurately measure dollar spot severity of turfgrass. *Phytopathology* **89:** S26.

Gaunt, R. E. 1995. The relationship between plant disease and severity and yield. *Ann. Rev Phytopath.* **33:** 119–144.

Gerten, D. M. and Weise, M. V. 1984. Video image analysis of lodging and yield loss in winter wheat relative to foot rot. *Phytopathology* **74:** 872.

Godoy, C. V., Koga, L. J., and Canteri, M. G. 2006. Diagrammatic scale for assessment of soybean rust severity. *Fitopat. Brasileira* **31:** 63–68.

Godoy, C. V., Carneiro, S.M.T.P.G., Iamauti, M. T., Dalla Pria, M., Amorim, L., Berger, R. D., and Bergamin Filho, A. 1997. Diagrammatic scales for bean disease: development and validation. *Z. Pflkrankh. Pflschutz. Pfanzen*. **104:** 336–345.

Gomes, A.M.A., Michereff, S. J., and Mariano, R.L.R. 2004. Elaboração e validação de escala diagramática para cercosporiose da alface. *Summa Phytopath.* **30:** 39–43.

Gottwald, T. R., da Graça, J. V., and Bassanezi, R. B. 2007. Citrus huanglongbing: the pathogen and its impact (online). *Plant Health Prog.* doi:10.1094/PHP-2007-0906-01-RV.

Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G., and Frias, J. M. 2007. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. *Trends in Food Sci. and Tech*. **18:** 590–598.

Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B.J., Faust, J. A., Pavri, B. E., Chovit, C. J., Solis, M., Olah, M. R., and Williams, O. 1998. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sens. Environ.* **65:** 227–248.

Grunwald, N., Kitner, M., McDonald, V., and Goss, E. M. 2008. Susceptibility in *Viburnum* to *Phytopththora ramorum*. *Plant Dis.* **92:** 210–214.

Guan, J. and Nutter, F. W., Jr. 2003. Quantifying the intrarater repeatability and interrater reliability of visual disease and remote sensing assessment methods in the alfalfa foliar disease pathosystem. *Can. J. Plant Path.* **25:** 143–149.

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B. 2004. Hyperspectral vegetation indices and novel algorithms for predicting green lai of crop canopies: modeling and validation in the contect of precision agriculture. *Remote Sens. Env.* **90:** 337–352.

Harveson, R. M. and Rush, C.M. 2002. The influence of irrigation and cultivar blends on the severity of multiple root diseases of sugar beets. *Plant Dis.* **86:** 901–908.

Hau, B., Kranz, J., and Konig, R. 1989. Fehler beim Schätzen von Befallsstärken bei Pflanzenkrankheiten. *Z. Pflkrankh. Pflschutz.*: **96:** 649–674.

Hebert, T. T. 1982. The rational for the Horsfall-Barratt plant disease assessment scale. *Phytopathology* **72:** 1269

Hilber, U. W. and Scheupp, H. 1992. Accurate and rapid measurement of lengths of fungal germ tubes by image analysis. *Can. J. Plant Path.* **14:** 185–186.

Hirano, A., Madden, M., and Welch, R. 2003. Hyperspectral image data for mapping wetland vegetation. *Wetlands* **23**(2):436–448.

Hock, J., Kranz, J., and Renfro, B. L. 1992. Tests of standard diagrams for field use in assessing the tarspot disease complex of maize. *Trop. Pest Manag.* **38:** 314–318.

Holmes, G. J., Brown, E. A., and Ruhl, G. 2000. What's a picture worth? The use of modern communications in diagnosing plant diseases. *Plant Dis.* **84:** 1256–1265.

Horsfall, J. G. 1945. Fungicides and their action. *Ann. Crypto. Phytopath.* Vol II. *Chron. Bot.*, 239 pp. Waltham, MA.

Horsfall J. G. and Barrat, R. W. 1945. An improved grading system for measuring plant disease. *Phytopathology* **35:** 655 (Abstract)

Horsfall, J. G. and Cowling, E. B. 1978. Pathometry: the measurement of plant disease (pp 120–136). **In:** *Plant Disease: An Advanced Treatise. Vol II.* J. G. Horsfall and E. B. Cowling, (eds.). Academic Press, New York.

Horsfall, J. G. and Heuberger, J. W. 1942. Measuring magnitude of a defoliation disease of tomatoes. *Phytopathology* **32:** 226–232.

Huang, W., Lamb, D. W., Niu, Z., Zhang, Y., Liu, L., and Wang, J. 2007. Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. *Precision Agric.* **8:** 187–197.

Hubel, D. H. 1995. *Eye, Brain and Vision.* Scientific American Library, No. 22., W.H. Freeman, New York, NY. 242 pp.

Jackson, E. W., Avant, J. B., Overturf, K. E., and Bonman, J. M. 2006. A quantitative assay of *Puccinia coronata* f.sp. *avenae* DNA in *Avena sativa*. *Plant Dis.* **90:** 629–636.

Jackson, E. W., Obert, D. E., Menz, M., Hu, G., Avant, J. B., Chong, J., and Bonman, J. M. 2007. Characterization and mapping oat crown rust resistance using three assessment methods. *Phytopathology* **97:** 1063–1070.

Jackson, H. R. and Wallen, V. R. 1975. Microdensitometer measurements of sequential aerial photographs of field beans infected with bacterial blight. *Phytopathology* 65: 961–968.

Jackson, H. R., Wallen, V. R., and Downer, J. F. 1978. Analysis and electronic area measurement of complex aerial photographic images. *J. App. Photo. Eng.* **4:** 101–106.

James, W. C. 1971. An illustrated series of assessment keys for plant diseases, their preparation and usage. *Can. Plant Dis. Surv*. **51:** 39–65.

James, W. C. 1974. Assessment of plant disease losses. *Ann. Rev. Phytopath.* **12:** 27–48.

James, W.C., Jenkins, J.E.E., and Jemmett, J. L. 1968. The relationship between leaf blotch caused by *Rhynchosporium secalis* and losses in grain yield of spring barley. *Ann. Appl. Biol.* **62:** 273–288.

Jensen, J. R. 1996. *Introductory Digital Image Processing – A Remote Sensing Perspective*. 2nd Edition, Prentice Hall, Upper Saddle River, NJ. ISBN 0-13-205840-5.

Jensen, J. R. 2007. *Remote Sensing of the Environment – An Earth Resource Perspective*. 2nd Edition, Pearson Prentice Hall, Upper Saddle River, NJ. ISBN 0-13-188950-8.

Johnson, D. A., Alldredge, J. R., Hamm, P. B., and Frazier, B. E. 2003. Aerial photography used for spatial pattern analysis of late blight infection in irrigated potato circles. *Phytopathology* **93:** 805–812.

Jones, D. C., Rush, C. M., Bredehoeft, M., Cattanach, A. 2006. Use of digital imagery to estimate crop loss due to *Rhizomania* in fields planted to *Rhizomania* resistant cultivars. *Phytopathology* **96:** S56.

Judd, D. B. and Wyszecki, G. 1975. *Color in Business, Science and Industry*. 3rd edition, Wiley Series in Pure and Applied Optics, New York: Wiley-Interscience, 388. ISBN 0471452122.

Kapmann, H. H. and Hansen, O.B. 1994. Using color for quantitative assessment of powdery mildew on cucumber. *Euphytica* **79:** 19–27.

Kim, M. S., Chen, Y. R., and Mehl, P. M. 2001. Hyperspectral reflectance and fluorescence imaging system for food quality and safety. *Trans. ASAE*. **44:** 721–729.

Klassen, P., Fürst, P., Schulz, C., Mazariegos, M., and Solomons, N. W. 2001. Plasma free amino acid concentrations in healthy Guatemalan adults and in patients with classic dengue. *Amer. J. Clin. Nutr.* **73:** 647–652

Koch, H. and Hau, B. 1980. Ein psychologischer aspect beim schatzen von pflanzenkrankheiten. *Z. Pflkrankh. Pflschutz* **87:** 587–593.

Kokko, E. G., Conner, R. L., Kozub, G. C., and Lee, B. 1993. Quantification by image analysis of subcrown internode discoloration in wheat caused by common root rot. *Phytopathology* **83:** 976–981.

Kokko, E. G., Conner, R. L., Lee, B., Kuzyk, A. D., and Kozub, G. C. 2000. Quantification of common root rot symptoms in resistant and susceptible barley by image analysis. *Can. J. Plant Prot.* **22:** 38–43.

Kranz, J. 1970. Schätzklassen für Krankheitsbefall. *Phytopath. Z.* **69:** 131–139;

Kranz, J. 1977. A study on maximum severity in plant disease. *Travaux dédiés à G. Viennot-Bourgin*, 169–173.

Kranz, J. 1988. Measuring plant disease. Pages 35–50 **In**: *Experimental Techniques in Plant Disease Epidemiology* (J. Kranz and J. Rotem, eds.), Springer-Verlag, New York.

Krueger, L. E. 1989. Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behav. Brain Sci*. **12:** 251–320.

Kruse, F. A., Boardman, J. W., and Huntington, J. F. 1999. Fifteen Years of Hyperspectral Data: northern Grapevine Mountains, Nevada. **In**: *Proceedings of the 8th JPL Airborne Earth Science Workshop: Jet Propulsion Laboratory*. JPL Publication 99-17, p. 247–258.

Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., and Goetz, A. F. H. 1993. The spectral image processing system (SIPS) – interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Env.***44:** 145–163.

Lamari, L. 2002. *ASSESS: Image Analysis Software for Plant Disease Quantification*. APS Press, St. Paul, MN.

Large, E. C. 1953. Some recent developments in fungus disease survey work in England and Wales. *Ann. App. Biol.* **40:** 594–599.

Large, E. C. 1955. Methods of plant disease measurement and forecasting in Great Britain. *Ann. App. Biol.* **42:** 344–354.

Large, E. C. 1966. Measuring plant disease. *Ann. Rev. Phytopath.* **4:** 9–26.

Large, E. C. and Honey, J. K. 1955. Survey of common scab of potatoes in Great Britain, 1952 and 1953. *Plant Path.* **4:** 1–8.

Larrsolle, A. and Muhammed, H. H. 2007. Measuring crop status using multivariate analysis of hyperspectral field reflectance with application to disease severity and plant density. *Precision Agric*. **8:** 37–47.

Lass, L. W. and Prather, T. S. 2004. Detecting the locations of Brazilian pepper trees in the everglades with a hyperspectral sensor. *Weed Tech.***18:** 437–442.

Lawrence, R. and Labus, M. 2003. Early detection of douglas-fir beetle infestation with subcanopy resolution hyperspectral imagery. *West. J. App. For.* **18:** 202–206.

Lee, Y. J. 1989. Aerial photography for the detection of soil-borne disease. *Can. J. Plant Path.*, **11:** 173–176.

Leite, R.M.V.B.C. and Amorin, L. 2002. Development and validation of a diagrammatic scale for *Alternaria* leaf spot of sunflower. *Summa Phytopath.* **28:** 14–19.

Lillesand, T. M., Meisner, D. M., French, D. W., and Johnsson, J. L. 1981. Evaluation of digital photographic enhancement for dutch elm disease detection. *Photogramm. Eng. Remote Sens.* **48:**1581–1592.

Lin, L. I. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45:** 255–268.

Lin, S. J., Brown, P. A., Watkins, M. P., Williams, T. A., Lehr, K. A., Liu, W., Lanza, G. M., Wickline, S. A., and Caruthers, S. D. 2004. Quantification of stenotic mitral valve area with magnetic resonance imaging and comparison with Doppler ultrasound. *J. Am. Coll. Cardiol*. **44:** 133–137.

Lindow, S. E. 1983. Estimating disease severity of single plants. *Phytopathology* **73:** 1576–1581.

Lindow, S. E. and Webb, R. R. 1983. Quantification of foliar plant disease symptoms by microcomputer-digitized video image analysis. *Phytopath.* **73:** 520–524.

Lipps, P. E. and Madden, L. V. 1989. Assessment of methods of determining powdery mildew severity in relation to grain yield on winter wheat cultivars in Ohio. *Phytopathology* **79:** 462–470.

Liu, Z., Huang, J., Shi, J., Tao, R. Zhao, W., and Zhang, L. 2007. Characterizing and estimating rice brown spot disease severity using stepwise regression, principle component regression, and partial least-squares regression. *J. Zhejiang Univ. Sci. B*. **8:** 738–744.

Lovell, D. J., Parker, S. R., Hunter, T., Royle, D. J., and Coker, R. R. 1997. Influence of crop growth and structure on the risk of epidemics by *Mycosphaerella graminicola* (*Septoria tritici*) in winter wheat. *Plant Path.* **46:** 126–138.

Lugo-Beauchamp, W., Cruz, K., Carvajal-Jiménez, C. L., and Rivera, W. 2004. Performance of hyperspectral imaging algorithms using itanium architecture. **In:** *Proceedings of the Second IASTED International Conference on Circuits, Signals, and Systems* Paper 449-199. M. H. Rashid Ed., Clearwater Beach, FL, USA, November 28, 2004 - December 1, 2004. Acta Press, Calgary, Canada.

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

Manzer, F. E. and Cooper, G. R. 1967. Aerial photographic methods of potato disease detection. *Maine Agr. Exp. Stn. Bull. 646*. 1–14.

Marsh, R. W., Martin, H. and Munson, R. G. 1937. Studies upon the copper fungicides. III. The distribution of fungicidal properties among certain copper compounds. *Ann. App. Biol.* **24:** 853–866.

Martin, D. P. and Rybicki, E. P. 1998. Microcomputer-based quantification maize streak symptoms in *Zea mays*. *Phytopathology* **88:** 422–427.

Martin, D. P., Willment, J. A., and Rybicki, E. P. 1999. Evaluation of maize streak virus pathogenicity in differentially resistant *Zea mays* genotypes. *Phytopathology* **89:** 695–700.

McKinney, H. H. 1923. Influence of soil temperature and moisture on infection of wheat seedlings by *Helminthosporium sativum*. *J. Agr. Res.***26:** 195–218.

McRoberts, N., Hughes, G. and Madden, L. V. 2003. The theoretical basis and practical application of relationships between different disease intensity measurements in plants. *Ann. App. Biol.***142:** 191–211.

Melchers, L. E. and Parker J. H. 1922. Rust resistance in winter wheat varieties, USDA Bulletin 1046. 32 pp.

Mian, M. A. R., Boerma, H. R., Phillips, D. V., Kenty, M. M., Shannon, G., Shipe, E. R., Soffes Blount, A. R., and Weaver, D. B. 1998. Performance of frogeye leaf spot–resistant and –susceptible near-isolines of soybean. *Plant Dis.* **82:** 1017–1021.

Mishra, A., Ehsani, R., Albrigo, G., and Lee, W.S. 2007. Spectral characteristics of citrus greening (huanglongbing). In: *Proceedings ASABE Annual Meeting*, Paper number 073056. ASABE, St. Joseph, MI.

Moore, W. C. 1943.The measurement of plant disease in the field. *Trans. Brit. Mycol. Soc.* **26:** 28–35.

Moore, W. C. 1949.The significance of plant disease in Great Britain. *Trans. Brit. Mycol. Soc.* **36:** 295–306.

Moshou, D., Bravo, C., Wahlen, S., West, J., McCartney, A., De Baerdermacker, J., and Ramon, H. 2003. Simultaneous identification of plant stresses and diseases in arable crops based on a proximal sensing system and self-organising neural networks. **In:** *Precision Agriculture*, pp. 425–432. Stafford, J. and Werner, A., Eds., Wageningen Academic Publishers, Wageningen, The Netherlands.

Moshou, D., Bravo, C., Wahlen, S., West, J., McCartney, A., De Baerdermacker, J., and Ramon, H. 2006. Simultaneous identification of plant stresses and diseases in arable crops using proximal optical sensing. *Precision Agric*. **7:** 149–164.

Muhammed, H. H. and Larsolle, A. 2003. Feature vector based analysis of hyperspectral crop reflectance data for discrimination and quantification of fungal severity in wheat. *Biosystems Eng.* **86:** 125–134.

Neblette, C. B. 1927. Aerial photography for the study of plant disease. *Photo-Era Mag.* **58:** 346.

Newell, L. C. and Tysdal, H. M. 1945. Numbering and note taking systems for use in improvement of forage crops. *J. Amer. Soc. Agron.* **37:** 736–749.

Newton, A. C. 1989. Measuring the sterol content of barley leaves infected with powdery mildew as a means of assessing partial resistance to *Erysiphe graminis* f.sp. *hordei*. *Plant Path.* **38:** 534–540.

Newton, A. C. and Hackett, C. A. 1994. Subjective components of mildew assessment on spring barley. *Euro. J. Plant Path.* **100:** 395–412.

Niemira, B. A., Kirk, W. W., and Stein, J. M. 1999. Screening for late blight susceptibility in potato tubers by digital analysis of cut tuber surfaces. *Plant Dis.* **83:** 469–473.

Nilsson, H. E. 1980. Remote sensing and image processing for disease. *Protection. Ecol.* **2:** 271–274.

Nilsson, H.-E. 1995. Remote sensing and image analysis in plant pathology. *Annu. Rev. Phytopathol.* **15:**489–527.

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.

Norwich, K. H. and Wong, W. 1997. Unification of psychophysical phenomena: The complete form of Fechners law. *Percept. Psychophys.* **59:** 929–940.

Nutter, F. W., Jr. 1990. Remore sensing and image analysis for crop loss assessment. **In:** *Crop Loss Assessment in Rice* (pp 93–105). International Rice Research Institute, Manila, The Philippines.

Nutter, F. W., Jr. 1997a. Disease severity assessment training. **In:** *Exercises in Plant Disease Epidemiology*. Francl, L.F., and Neher, D.A. (eds.). American Phytopathological Society, APS Press.

Nutter, F. W., Jr. 1997b. Quantifying the temporal dynamics of plant virus epidemics: A review. *Crop Prot.* **16:** 603–618.

Nutter, F. W., Jr. 2001. Disease assessment terms and concepts. Pages 312–323 **In:** *Encyclopedia of Plant Pathology*, O. C. Maloy and T. D. Murray, eds. John Wiley and Sons, Inc., New York, NY.

Nutter, F. W., Jr. and Esker, P. D. 2001. Disease assessment keys. Pages 323–326 **In:** *Encyclopedia of Plant Pathology*, O. C. Maloy and T. D. Murray, eds. John Wiley and Sons, Inc., New York, NY.

Nutter, F. W., Jr., and Esker, P. D. 2006. The role of psychophysics in phytopathology. *Euro. J. Plant Path.* **114:** 199–213.

Nutter, F. W., Jr. and Gaunt, R. E. 1996. Recent developments in methods for assessing disease losses in forage/pasture crops. p. 93–118. **In:** *Pasture and Forage Crop Pathology*. S. Chakraborty *et al.* (ed.). ASA, CSSA, and SSSA, Madison, WI.

Nutter, F. W., Jr. and Litwiller, D. 1993. Alfalfa. Pro—a computerized disease assessment training program for foliar diseases of alfalfa. p. 15. **In:** *Proceedings of the 23rd Central Alfalfa Improvement Conference*, 20–22 June 1993. Univ. of Nebraska, Lincoln, NE.

Nutter, F. W., Jr. and Litwiller, D. 1998. A computer program to generate standard aerea diagrams to aid raters in assessing disease severity. *Phytopathology* **88:** S117.

Nutter, F. W., Jr. and Schultz, P. M. 1995. Improving the accuracy and precision of disease assessments: selection of methods and use of computer-aided training programs. *Can. J. Plant Path.* **17:** 174–185.

Nutter, F. W., Jr. and Worawitlikit, O. 1989. Disease.Pro: A computer program for evaluating and improving a person ability to assess disease proportion. *Phytopathology* **79:** 1135 (Abstract).

Nutter, F. W., Jr., Esker, P. D., and Coelho Netto, R. A. 2006. Disease assessment concepts in plant pathology. *Euro. J. Plant Path.* 115:95–103.

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

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

Nutter, F. W., Jr., Gleason, M. L., Jenco, J. H., and Christians, N. L. 1993. Accuracy, intrarater repeatability, and interrater reliability of disease assessment systems. *Phytopathology* **83:** 806–812.

O'Brein, R. D. and van Bruggen, A.H.C. 1992. Accuracy, precision, and correlation to yield loss of disease severity scales for corky root of lettuce. *Phytopathology* **82:** 91–96.

Okamoto, H., Murata, T., Kataoka, T., and Hata, S-I. 2007. Plant classification for weed detection using hyperspectral imaging and wavelet analysis. *Weed Biol. Manag.* **7:** 31–37.

Olmstead, J. W., Lang, G. A., and Grove, G. G. 2001. Assessment of severity of powdery mildew infection of sweet cherry leaves by digital image analysis. *HortScience* **36:** 107–111.

Parker S. R., Shaw, M. W., and Royle, D. J. 1995a. The reliability of visual estimates of disease severity on cereal leaves. *Plant Pathol.* **44:** 856–864.

Parker S. R., Shaw, M. W., and Royle, D.J. 1995b. Reliable measurement of disease severity. *Aspects Appl. Biol.* **43:** 205–214.

Payero, J. O., Neale, C. M. U., and Wright, J. L. 2004. Comparison of eleven vegetation indices for estimating plant height of alfalfa and grass. *Applied Engineering in Agriculture*. **20:** 385–393.

Pederson, V. D., and Nutter, F. W., Jr. 1982. A low-cost portable multispectral radiometer for assessment of onset and severity of foliar disease of barley. *Proc. Int. Soc. Optical Eng.* **356:** 126–130.

Pernezney, K., Nuessley, G., and Stall, W. 2003. *Integrated pest management for Florida snap beans*. Pamphlet PPP37, IFAS, University of Florida, Gainesville, FL.

Peterson, R. F., Campbell, A.B., and Hannah, A. E. 1948. A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. *Can. J. Res.* (C.) **26:** 496–500.

Pethybridge, S. J., Hay, F. S., and Wilson, C. R., 2004. Pathogenicity of fungi commonly isolated from foliar disease in Tasmanian pyrethrum crops. *Austral. Plant Path.* **33:** 441–444.

Pethybridge S. J., Esker, P., Dixon, P., Hay, F., Groom, T., Wilson, C., and Nutter, F. W. 2007. Quantifying loss caused by ray blight disease in Tasmanian pyrethrum fields. *Plant Dis*. **91:** 1116–1121.

Pfender, W. F. 2004. Effect of autumn planting date and stand age on severity of stem rust in seed crops of perennial ryegrass. *Plant Dis*. **88:** 1017–1020.

Price, T. V. and Osborne, C.F. 1990. Computer imaging and its application to some problems in agriculture and plant science. *Crit. Rev. Plant Sci.* **9:** 235–266.

Price, T. V., Gross, R., Ho Wey, J., and Osborne, C. F. 1993. A comparison of visual and digital image-processing methods in quantifying the severity of coffee leaf rust (*Hemileia vastatrix*). *Aust. J. Exp. Agric.* **33:** 97–101.

Qin, J., Burks, T. F., Kim, M. S., Chao, K., and Ritenour, M. A. 2008. Citrus canker detection using hyperspectral reflectance imaging and PCA-based image classification method. *Sens. Instrum. Food Qual. Safety* **2:** 168–177.

Qin, J., Burks, T. F., Ritenour, M. A., and Bonn, W. G. 2009. Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence. *J. Food Engineering*. **93:** 183–191.

Redman C. E. and Brown, I. F. 1964. A statistical evaluation of the Barratt and Horsfall rating system. *Phytopathology* 54: 904.

Redman C. E., King, E. P., and Brown I. F., Jr. 1968. *Tables for Converting Barratt and Horsfall Rating Scores to Estimated Mean Percentages.* Indianapolis, IN: Elanco Products. 8 p.

Richard, J., Daures, J., Parer-Richard, C., Vannereau, D., and Boulot, I. 2000. Of mice and wounds: reproducibility and accuracy of a novel planimetry program for measuring wound area. *Wounds* **12:** 148–154.

Richards, J. A. and Jia, X. 2006. *Remote Sensing Digital Image Analysis*. Springer-Verlag, Berlin, Germany.

Ricker, M. D. 2004. Pixels, bits, and GUIs: The fundamentals of digital imagery and their application by plant pathologists. *Plant Dis.* **88:** 228–241.

Rock, B. N., Hoshizaki, T., and Miller, J.R. 1988. Comparison of in situ and airborne spectral measurements of the blue shift associated with forest decline. *Remote Sens. Environ.* **24:** 109–127.

Rusakov, L. F. 1927. A combination scale for estimating the development of rusts. *Bolezni Rast.* **16:** 179–185.

Russ, J. C. 2002. *The Image Processing Handbook*. 4th Edition. CRC Press, NY.

Sah, D. N. and Fehrmann, H. 1992. Virulence patterns of geographically differing isolates of *Pyrenophora tritici-repentis* and sources of resistance in wheat. *Plant Dis.* **76:** 712–716.

SAS. 2004. SAS/STAT 9.1 *User's Guide.* SAS Institute. Cary, NC. 5136 pp.

Schneider, C. L. and Safir, G. R. 1975. Infrared aerial photography estimation of yield potential in sugarbeets exposed to blackroot disease. *Plant Dis. Rep*. **59:** 627–631.

Seager, S., Turner, E. L., Schafer, J., and Ford, E. B. 2005. Vegetation's red edge: a possible biomarker of extraterrestrial plants. *Astrobiology* **5:** 372–390.

Seem, R. C. 1984. Disease incidence and severity relationships. *Annu. Rev. Phytopathol.* **22:** 133–150.

Seiffert, U. and Schweizer, P. 2005. A pattern recognition tool for quantitative analysis of in planta hyphal growth of powdery mildew fungi. *J. Molec. Plant-Microbe Inter.* **18:** 906–912.

Shah, D. A., and Madden, L. V. 2004. Nonparametric analysis of ordinal data in designed factorial experiments. *Phytopathology* **94:** 33–43.

Shane, W. W., Thomson, C., and Teng, P. S. 1985. *AREAGRAM – a statistical area diagram computer program*. Epidemiology Report No. 3, Dept. of Plant Pathology, University of Minnesota, St. Paul, 34 pp.

Shaw, M. W. and Royle, D. J. 1989. Estaimation and validation of a function describing the rate at which *Mycosphaerella graminicola* causes yield loss in winter wheat. *Ann. Appl. Biol.* **115:**425–442.

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.

Sheskin, D. J. 1997. *Handbook of Parametric and Non-Parametric Procedures.* CRC Press, Boca Raton, FL.

Shokes, F. M., Berger, R. D., Smith, D. H., and Rasp, J. M. 1987. Reliability of disease assessment procedures: A case study with late spot of peanut. *Oleagineux* **42:** 245–251.

Slopek, W. S. 1989. An improved method of estimating percent area diseased using a 1 to 5 disease assessment scale. *Can. J. Plant Path.* **11:** 381–387.

Smith, P. M., Last, F. T., Kempton, R. J., and Gisborne, J. H. 1969. Tomato leaf mould: its assessment and effects on yield. *Ann. Appl. Biol.* **63:** 19–26.

Smith, S. E. and Dickson, S. 1991. Quantification of active vesicular-arbuscular mycorrhizal infection using image analysis and other techniques. *Austral. J. Plant Phys.* **18:** 637–648.

Snedecor, G. W. and Cochran, W. G. 1989. Statistical Methods. 8^{th} edition. Iowa State University Press, Iowa, USA. 503 pp.

Steddom, K., Jones, D., Rudd, J. and Rush, C. 2005a. Analysis of field plot images with segmentation analysis; effect of glare and shadows. *Phytopathology* **95:** S99.

Steddom, K., McMullen, M., Schatz, B., and Rush, C. M. 2005b. Comparing image format and resolution for assessment of foliar diseases of wheat. Online. *Plant Health Prog.* doi:10.1094/PHP-2005-0516-01-RS.

Stevens, S. S. and Galanter, E. H. 1957. Ratio scales and category scales for a dozen perceptual continua. *J. Exp. Psychol.* **54:** 377–411.

Stevens, S. S. 1961. To honor Fechner and repeal his law. *Science* **133:** 80–86.

Stonehouse, J. 1994. Assessment of Andean bean diseases using visual keys. *Plant Path.* **43:** 519–527.

Subrahmanyam, P., McDonald, D., Gibbons, R. W., Nigam, S. N., and Neville, D. J.1982. Resistance to rust and late leaf spot diseases in some genotypes of Arachis hypogaea. *Peanut Science* **9:** 6–10.

Sykes, V. R., Horvath, B. J., and Warnke, S. E. 2008. Resistance screening of *Festuca arundinacea* to both *Rhizoctonia solani* and *Rhizoctonia zeae* using digital image analysis. *Phytopathology* **98:** S154.

Taubenhaus, J. J., Ezekiel, W. N., and Neblette, C. B. 1929. Airplane photography in the study of cotton root rot. *Phytopathology* **19:** 1025–1029.

Tehon, L. R. and Stout, G. L. 1930. Epidemic diseases of fruit trees in Illinois., 1922–1928. Ill. *Nat. Hist. Surv., Bull.* **18:** 415–502.

Tilling, A. K., O'Leary, G., Ferwerda, J. G., Jones, S. D., Fitzgerald, G., and Belford, R. 2006. Remote sensing to detect nitrogen and water stress in wheat. **In:** *Proceedings of the 13th Australian Agronomy Conference,* Turner N.C., Acuna T. and Johnson, R.C., Eds., 10–14 September 2006, Perth, Western Australia. Australian Society of Agronomy.

Tinline, R. D. and Ledingham, R. J. 1979. Yield losses in wheat and barley cultivars from
common root rot in field tests. *Can. J. Plant Sci*. **59:** 313–320.

Tinline, R. D., Diehl, J. A., and Spurr, D. T. 1994. Assessment of methods for evaluating common root rot in spring wheat and infection of subterranean plant parts by the causal fungus *Cochliobolus sativus*. *Can. J. Plant Pathol.* **16:** 207 214.

Tjosvold, S. A. and Chambers, D. L. 2006. Correlation of lesion size with percent lesion coverage on camellia and rhododendron leaves inoculated with *Phytophthora ramorum*. *Phytopathology* **96:** S114.

Todd L. A. and Kommedahl, T. 1994. Image analysis and visual estimates for evaluating disease reactions of corn to *Fusarium* stalk rot. *Plant Dis.* **78:** 876–878.

Toler, R. W., Smith, B. D., and Harlan, J. C. 1981. Use of aerial colorinfrared photography to evaluate crop disease. *Plant Dis.* **65:** 24–31.

Tomerlin, J. R. and Howell, T. A. 1988. DISTRAIN: A computer program for training people to estimate disease severity on cereal leaves. *Plant Dis.* **72:** 455–459.

Townsend, G. R. and Heuberger, J. W. 1943. Methods for estimating losses caused by diseases in fungicide experiments. *Plant Dis. Rep.* **27:** 340–343.

Trager, G., Michaud, G., Deschamps, S., and Hemmerling, T. M. 2006. Comparison of phonomyography, kinemyography and mechanomyography for neuromuscular monitoring. *Can. J. Anesth.* **53:** 130–135.

Tucker, C.C. and Chakraborty, S. 1997. Quantitative assessment of lesion characteristics and disease severity using digital image processing. *J.Phytopath.* **145:** 273–278.

Ullstrup, A. J., Elliott, C., and Hopppe, P. E. 1945. Report of the committee on methods for reporting corn disease ratings. Mimeographed. Unnumbered Publication of the USDA, Division of Cereal Crops and Diseases. 5 pp.

Uno, Y., Prasher, S. O., Lacroix, R., Goel, P. K., Karimi, Y., Viau, A., and Patel, R. M. 2005. Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. *Comp. Elect. Agri.***47:** 149–161.

Vale, F.X.R., Fernandes-Filho, E.I., and Liberato, J. R. 2003. QUANT: A software for plant disease severity assessment. **In:** *Proceedings 8th International Congress of Plant Pathology.* Christchurch, New Zealand, p 105.

Venette, J. R. and Venette, R. C. 1991. Image analysis for evaluation of bean rust severity. *Phytopathology* **81:** S1213.

Vereijssen, J. Schneider, J.H.M., Termorshuizen, A. J., and Jeger, M. J. 2003. Comparison of two disease assessment methods for assessing *Cercospora* leaf spot in sugar beet. *Crop Prot.* **22:** 201–209.

Vicent, A., Armengol, J., and García-Jiménez, J. 2007. Rain fastness and persistence of fungicides for control of *Alternaria* brown spot of citrus. *Plant Dis.* **91:** 393–399.

Walker, J. C., Larson, R. H., and Albert, A. R. 1938. Studies of resistance to potato scab in Wisconsin. *Amer. Potato J.* **15:** 246–252.

Wallen, V. R. and Jackson, H. R. 1971. Aerial photography as a survey technique for the assessment of bacterial blight of field beans. *Can. Plant Dis. Surv.* **51:** 163–169.

Warke, T. J., Kamath, S., Fitch, P. S., Brown, V., Shields, M. D., and Ennis, M. 2001. The repeatability of nonbronchoscopic bronchoalveolar lavage differential cell counts. *Eur. Respir. J.* **18:**1009–1012.

Web of Science. 2009. http://isiwebofknowledge.com/products_tools/multidisciplinary/webofscience/

Weber, G. E. and Jorg, E. 1991. Errors in disease assessment – a survey. *Phytopathology* **81:** S1238.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica* **48:** 817–838.

Xu, Y., Kang, D., Shi, Z., Shen, H., and Wehner, T. 2004. Inheritance of resistance to zucchini yellow mosaic virus and watermelon mosaic virus in watermelon. *J. Hered.* **95:** 498–502.

Ye, X., Sakai, K., Manago, M., Asada, S-I., and Sasao, A. 2007. Prediction of citrus yield from airborne hyperspectral imagery. *Precision Agric.* **8:** 111–125.

Yang, C-M., Cheng, C-H. and Chen, R-K. 2007. Changes in spectral characteristics of rice canopy infested with brown planthopper and leaffolder. *Crop Sci.* **47:** 329–335.