

The Pennsylvania State University

The Graduate School

College of Agricultural Sciences

**STUDIES ON THE EPIDEMIOLOGY AND MANAGEMENT OF BACTERIAL  
SPOT OF PEACH AND NECTARINE IN PENNSYLVANIA**

A Thesis in

Plant Pathology

by

Sarah Bardsley

© 2010 Sarah Bardsley

Submitted in Partial Fulfillment  
of the Requirements  
for the Degree of

Master of Science

December 2010

The thesis of Sarah Bardsley was reviewed and approved\* by the following:

Henry K. Ngugi  
Assistant Professor of Plant Pathology  
Thesis Advisor

Timothy W. McNellis  
Associate Professor of Plant Pathology

James W. Travis  
Professor Emeritus of Plant Pathology

James Schupp  
Associate Professor of Pomology

Frederick E. Gildow  
Professor of Plant Pathology  
Head of the Department of Plant Pathology

\*Signatures are on file in the Graduate School

## ABSTRACT

The purpose of this research was to develop methods to quantify bacterial spot severity in order to improve the quality of data collected as well as to gain a better understanding of the epidemiology of bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) in Pennsylvania peach and nectarine orchards. Based on Lin's concordance analysis as well as the categorical data analysis, the direct estimation method was a superior method of estimating bacterial spot severity compared to a 1 to 7 ordinal rating scale (1 = 0% lesion area, and 7 = >45%). Moreover, the reliability and accuracy of estimates of bacterial spot severity made by inexperienced raters using the direct estimation method were equal to those made by experienced raters. In both 2008 and 2009, cultivar and bactericide treatment were the significant factors affecting the rate of bacterial spot progress based on analysis of 84 temporal disease progress curves from each year. Survival analysis of data collected in 2009 showed that the mean time to leave abscission ( $T$ ) ranged from 41.8 to 56.3 days, and was significantly ( $P < 0.0001$ ) affected by cultivar, initial disease severity at the onset of the epidemic, and leaf age, but not by bactericide treatment. Results indicate that bacterial spot epidemics do not follow standard disease progress curves, and that strategies for bacterial spot management should focus on reducing initial disease. In addition, estimates of bacterial spot severity can be made by inexperienced raters without loss of accuracy and reliability with the use of the direct estimation method. The results are discussed in relation to their implication to bacterial spot management in the eastern US.

## TABLE OF CONTENTS

|   |      |
|---|------|
| LIST OF TABLES .....  | vi   |
| LIST OF FIGURES .....   | viii |
| ACKNOWLEDGEMENTS .....  | x    |
| Chapter 1 Introduction .....  | 1    |
| Background .....  | 1    |
| Temporal Progression of Bacterial Spot Epidemics on Stone Fruits .....  | 4    |
| Reliability and Accuracy of Methods Used to Evaluate Bacterial Spot<br>Severity .....   | 5    |
| Comparison of Assessments of Foliar Bacterial Spot Symptoms made by<br>Experienced and Inexperienced Raters .....   | 6    |
| Goals and Objectives .....  | 7    |
| Literature Cited .....  | 9    |
| Chapter 2 Evaluation of Direct Estimation and an Ordinal Scale Used to<br>Quantify Severity of Bacterial Spot of Peach and Nectarine .....  | 11   |
| Abstract .....  | 11   |
| Introduction .....  | 12   |
| Materials and Methods .....   | 14   |
| Results .....   | 18   |
| Discussion .....  | 22   |
| Literature Cited .....  | 26   |
| Chapter 3 Accuracy and Inter-rater Reliability of Assessments of Bacterial Spot<br>Severity Made by Experienced and Inexperienced Raters .....  | 41   |
| Abstract .....  | 41   |
| Introduction .....  | 42   |
| Materials and Methods .....   | 44   |
| Results .....   | 47   |
| Discussion .....  | 49   |
| Literature Cited .....  | 57   |
| Chapter 4 Temporal Bacterial Spot Progress and Factors Affecting the Timing of<br>Abscission of Peach and Nectarine Leaves with <i>Xanthomonas arboricola</i> pv.<br><i>pruni</i> ..... | 68   |
| Abstract .....  | 68   |
| Introduction .....  | 69   |
| Materials and Methods .....   | 72   |
| Results .....   | 76   |

|                                  |     |
|----------------------------------|-----|
| Discussion.....                  | 83  |
| Literature Cited.....            | 90  |
| Chapter 5 Discussion .....       | 106 |
| Literature Cited.....            | 112 |
| Appendix A: Weather – 2008 ..... | 114 |
| Appendix B: Weather – 2009 ..... | 115 |

## LIST OF TABLES

|  |    |
|--|----|
| Table <b>2-1</b> : Summary statistics for bacterial spot severity assessed by one experienced rater on the three sets of peach and nectarine leaves with direct estimation (DE) method or a disease severity rating scale (Scale).....   | 30 |
| Table <b>2-2</b> : Intra-rater reliability, determined by the correlation coefficient ( $r$ ) and Lin's concordance correlation coefficient ( $\rho_c$ ), for the direct estimation (DE) method and a disease severity rating scale (Scale) for the three sets of peach and nectarine leaves with symptoms of bacterial spot ..... | 31 |
| Table <b>2-3</b> : Intra-rater reliability, determined with concordance analysis, for the untransformed rating scale for 3 sets of peach and nectarine leaves with symptoms of bacterial spot .....  | 32 |
| Table <b>2-4</b> : Lin's concordance analysis statistics for comparisons between actual bacterial spot severity and estimated bacterial spot severity for three sets of peach and nectarine leaves.....  | 33 |
| Table <b>2-5</b> : Bootstrap analysis of accuracy statistics comparing data obtained with the 1-7 ordinal rating scale and the direct estimation method.....   | 34 |
| Table <b>2-6</b> : Concordance statistics for comparisons between actual bacterial spot severity and estimated bacterial spot severity for 3 sets of peach and nectarine leaves based on categorical scale data without transformation.....  | 35 |
| Table <b>3-1</b> : Summary statistics for bacterial spot severity assessed by 10 raters on the three sets of peach and nectarine leaves with direct visual estimation.....   | 61 |
| Table <b>3-2</b> : Intra-rater reliability, determined using the correlation coefficient ( $r$ ) for each of the 10 raters, for the direct visual estimation of bacterial spot severity of peach and nectarine leaves, for each of the 3 leaf sets .....   | 62 |
| Table <b>3-3</b> : Inter-rater reliability was determined for the direct visual estimation of bacterial spot severity of 3 sets of leaves rated twice by 10 raters based on mean pairwise correlation coefficient ( $r$ ), the intra-class correlation coefficient ( $\rho$ ), and the repeatability coefficient (RC) .....        | 63 |
| Table <b>3-4</b> : Rater accuracy was determined by comparing the actual bacterial spot severity and estimated bacterial spot severity for 3 sets of peach and nectarine leaves.....   | 64 |
| Table <b>3-5</b> : Comparison of accuracy and intra-rater reliability between experienced and inexperienced raters.....  | 65 |
| Table <b>4-1</b> : Comparison of five probability distributions tested for goodness-of-fit in describing time-to-defoliation of leaves displaying symptoms of bacterial spot of peach and nectarine .....  | 93 |

|  |    |
|--|----|
| Table 4-2: Mean estimates of initial disease onset for bacterial spot disease progress curves from the 2008 growing season .....   | 94 |
| Table 4-3: Mean estimates of initial disease onset for bacterial spot disease progress curves from the 2009 growing season. ....   | 95 |
| Table 4-4: Mean estimates of rate of disease progression for bacterial spot disease progress curves from the 2008 growing season. ....   | 96 |
| Table 4-5: Mean estimates of rate of disease progression for bacterial spot disease progress curves from the 2009 growing season. ....   | 97 |
| Table 4-6: Parameter estimates and associated test statistics for the accelerated failure time (AFT) models relating cultivar, leaf age, and initial disease onset with time to defoliation ( $T$ ) in peach and nectarine affected by bacterial spot in Pennsylvania in the 2009 season. .... | 98 |

## LIST OF FIGURES

|   |     |
|---|-----|
| Figure 2-1: Leaves typical of the 1-7 rating scale used to assess foliar bacterial spot severity.....   | 36  |
| Figure 2-2: Estimated and actual disease severity values for bacterial spot of peach and nectarine for three sets of leaves.....  | 37  |
| Figure 2-3: A comparison of the first (X) and second (Y) assessments using the direct estimation method and the rating scale for 3 sets of leaves .....   | 38  |
| Figure 2-4: Lin's concordance analysis for the three sets of leaves assessed using the rating scale.....  | 39  |
| Figure 2-5: Lin's concordance analysis for the three sets of leaves assessed using the direct estimation method.....  | 40  |
| Figure 3-1: Lin's concordance analysis for the first assessment of the first set of leaves assessed using direct visual estimation by experienced raters A through D... ..  | 66  |
| Figure 3-2: Lin's concordance analysis for the first assessment of the first set of leaves assessed using direct visual estimation by inexperienced raters E through J... ..  | 67  |
| Figure 4-1: Actual bacterial spot disease progress and cumulative disease progress on the peach cultivar Beekman in response to the treatments Mycoshield, Kocide 3000, and untreated check in the 2008 growing season.....                   | 99  |
| Figure 4-2: Actual bacterial spot disease progress and cumulative disease progress on the peach cultivar Beekman in response to the treatments Mycoshield, Kocide 3000, and untreated check in the 2009 growing season.....                   | 100 |
| Figure 4-3: Fitted cumulative bacterial spot disease progress and associated standardized residuals on the peach cultivar Beekman in response to the treatments Mycoshield, Kocide 3000, and untreated check in the 2008 growing season.....  | 101 |
| Figure 4-4: Fitted cumulative bacterial spot disease progress and associated standardized residuals on the peach cultivar Beekman in response to the treatments Mycoshield, Kocide 3000, and untreated check in the 2009 growing season. .... | 102 |
| Figure 4-5: Kaplan-Meier estimates of survival functions indicating time to abscission of peach and nectarine leaves of the cultivars Easternglo, Beekman, Snow King, and Sweet Dream with foliar symptoms of bacterial spot.....             | 103 |
| Figure 4-6: Kaplan-Meier estimates of survival functions indicating time to abscission of peach and nectarine leaves of the cultivars Easternglo, Beekman, Snow King, and Sweet Dream with foliar symptoms of bacterial spot.....             | 104 |

Figure 4-7: Kaplan-Meier estimates of survival functions indicating time to abscission of peach and nectarine leaves of the cultivars Easternglo, Beekman, Snow King, and Sweet Dream with foliar symptoms of bacterial spot. Leaves were classified into seven categories based on treatment received in the product evaluation trial ..... 105

## ACKNOWLEDGEMENTS

This thesis research was supported in part with funds from the United States Department of Agriculture Southeast Regional Integrated Pest Management, the Pennsylvania Peach and Nectarine Marketing Board, and the State Horticultural Association of Pennsylvania (SHAP).

I thank my major advisor, Dr. Henry K. Ngugi and the members of my advisory committee, Drs. Timothy W. McNellis, James W. Travis, and James Schupp for their guidance, time, and helpful suggestions. I appreciate all of the support and help I have received from the members of the Penn State Plant Pathology Department including Brian Lehman, Katie McAnlis, Mattie Kuntz, Bashar Jarjour, Emily Pfuefer, Terry Salada, and Noemi Halbrendt. Finally, I am grateful for the love and encouragement I have received from my family and friends.

## Chapter 1

### Introduction

#### BACKGROUND

Peach (*Prunus persica* (L.) Batsch) is the second most important fruit crop in the eastern United States after apple. In 2008, the eastern US peach crop was valued at over \$159 million with significant production from the states of Georgia, South Carolina, Pennsylvania, Michigan, New Jersey, North Carolina, Alabama, Virginia, West Virginia, New York, Ohio, and Maryland (Anonymous, 2009). In the same year, over 30,000 hectares of farm land were dedicated to the production of peaches in the eastern US.

Bacterial spot of stone fruits is one of the most important diseases of peach and nectarine in the eastern United States, especially in the southeastern US and the mid-Atlantic regions where the weather is warm and wet. Because of the high market demand for California-bred varieties, producers are forced to grow these highly susceptible, but highly marketable, cultivars. In years where weather conditions favor bacterial spot development, 100% fruit loss has been observed. For example, in 2005 bacterial spot reduced Georgia's peach crop value by 15% – a loss of almost \$5 million (Buttner et al., 2002). Such economic losses are not uncommon in other eastern states such as New Jersey and Pennsylvania.

*Xanthomonas arboricola* pv. *pruni* (*Xap*) is the causal agent of bacterial spot of cultivated plants belonging to the genus *Prunus*. Susceptible hosts include cultivars of peach, nectarine, Japanese plum, apricot, and almond (Ritchie, 1995). *Xap* is a Proteobacterium belonging to the genus *Xanthomonas*. It is a Gram-negative, rod shaped, aerobic, and flagellated bacterium with a single polar flagellum (Buttner et al., 2002). The species *X. arboricola* consists of the pathovars *corylina* (bacterial hazelnut blight), *fragariae* (bacterial leaf blight of strawberry), *juglandis* (walnut blight), *poinsetticola* type C strains, *populi* (bacterial canker of poplar), and *pruni* (bacterial spot of stone fruits) (Vauterin et al., 1995). Originally known as *Xanthomonas pruni*, the species has undergone numerous reclassifications. In the mid – 1990's, it was reclassified from

*Xanthomonas campestris* pv. *pruni* to *Xanthomonas arboricola* pv. *pruni* (Vauterin et al., 1995).

The geographical distribution of bacterial spot is limited to where stone fruit are grown and in places where the environment favors growth of the bacteria. The disease occurs in countries of Europe, North America, Africa, Asia, South America, and Oceania (Australia and New Zealand) including but not limited to the United States, Canada, Mexico, Brazil, China, Japan, France, Italy, Romania, and parts of Russia (Data Sheets on Quarantined Pests).

Bacterial spot was first described by F.E. Smith in 1903 on Japanese plum in a Michigan fruit orchard (Smith, 1903). Italian scientists were the first Europeans to describe the disease in 1920 (Battilani et al., 1999). Overall, there is very little genetic diversity among strains of *Xap* from different geographic locations and different hosts, suggesting that *Xap* is a relatively new plant pathogen (Zaccardelli et al., 1999). Bacterial populations have the highest diversity in North America compared to the low diversity of European populations. Bacterial populations from Italy and France are nearly genetically identical but are different from populations in North America. Because of the low diversity of the European population, it is likely that the population had undergone a recent bottleneck – the introduction of the pathogen aided by humans (Brannen, 2006).

Symptoms of bacterial spot occur on leaves, twigs, and fruit. On leaves, angular lesions at the leaf tip, mid-rib, and/or along the leaf margin occur from early spring to fall (Zehr et al., 1996). Initially, foliar lesions appear water-soaked and angular in shape as bacteria are delimited by the veins. As foliar infection progresses, the centers of lesions become dark in color and abscise from the leaf, developing a “shot hole” appearance. Young leaves, not fully expanded, are most susceptible as bacteria-infested water may become trapped in the leaf, prolonging the leaf wetness period. Premature leaf drop is common, effectively defoliating the tree when disease severity is high (Ritchie, 1995). Severe premature defoliation reduces fruit quality and size, but crop loss also results from severely damaged fruit which are unmarketable. On fruit, the earliest lesions occur about three weeks after petal fall (Ritchie, 1995). Lesions initially appear water-soaked and eventually become brown to dark-brown as they enlarge. Early season fruit infections

usually become large, open lesions that extend deep into the fruit. Lesions that develop after pit-hardening are usually superficial and may cause the skin to crack if the lesions coalesce (Ritchie, 1995). Cracks often favor secondary infection by the brown rot fungus, *Monilinia fructicola*. Twig symptoms consist of three types of cankers: spring and summer cankers as well as black tip. Summer cankers form during the season and are usually visible in early to mid summer. Black tip refers to a canker at the end of a twig or branch (Battilani et al., 1999). Initially, cankers have a black, water soaked appearance when environmental conditions are wet. As the growing season progresses and conditions become warmer and dryer, cankers enlarge and the bark surface may crack.

Bacterial spot of stone fruit is a poly-cyclic disease with many infection cycles throughout the growing season that are highly dependent upon temperature and leaf wetness (Zehr et al., 1996). Bacteria overwinter in spring cankers that are first visible during bloom (Ritchie, 1995). Although bacteria primarily overwinter in cankers, infected buds and leaf scars, and infected leaves on the soil surface below the tree can also harbor viable bacteria (Ritchie, 1995; Zaccardelli et al., 1998). *Xap* can survive in infected leaves on the ground for at least 7 months, and such leaves are most likely the primary source of inoculum in Italian and French peach orchards where overwintering cankers are not usually found (Zaccardelli et al., 1998). In Pennsylvania and other eastern US states, however, cankers are thought to play a major role as a source of primary inoculum. Under certain conditions, *Xap* can survive as an epiphyte on hosts and non-hosts for several weeks without causing disease (Battilani et al., 1999; Zehr et al., 1996).

As temperatures increase to between 14°C and 30°C, bacterial populations that have overwintered in the intercellular spaces of the cortex, phloem, and xylem, begin to grow within cankers and leaf scars (Battilani et al., 1999; Zehr et al., 1996). Bacteria from these overwintering sites are spread from cankers by wind driven rain to leaves and fruit (Battilani et al., 1999; Ritchie, 1995). Disease incidence and severity is higher for leaves and fruit located around cankers. Foliar infections provide additional inoculum for secondary infections of leaves and fruit. New infections become less frequent as temperatures rise above 30°C and as rain become less frequent (Ritchie, 1995). Late

season shoot infections generally become spring cankers, providing the primary inoculum source for the next growing season.

There are few effective controls for bacterial spot. Generally, the antibiotic oxytetracycline, marketed as Mycoshield® or FlameOut®, has been used with significant disease suppression. Copper compounds have been used as dormant sprays providing a prophylactic protection of trees against infection by *Xap* (Ritchie, 1995; Ngugi et al., 2009). They are also used as cover sprays during the season and are marketed as inorganic copper formulations such as Kocide 3000 or the organometallic compound Tenn-Cop 5E. At high enough rates, these copper compounds are phytotoxic and often results in leaf discoloration, curling, and premature defoliation (Lalancette and McFarland, 2007). The proper timing of repeated sprays is also essential as these chemical bactericides are not systemic or have limited residual activity in the field. Resistant peach and nectarine cultivars are available but are often less marketable than the highly susceptible varieties developed by breeding programs in the western US where bacterial spot is not a problem (Ritchie, 1995). Because stressed trees are more susceptible to bacterial spot, removing sources of tree stress such as ring nematode infestations, weeds, poorly drained soils, and nutrient deficiencies, will ultimately reduce disease severity (Shepard et al., 1999; Zehr et al., 1996). Cultural practices such as proper pruning, orchard sanitation, and removal of cankers may also reduce severity.

#### TEMPORAL PROGRESSION OF BACTERIAL SPOT EPIDEMICS OF STONE FRUITS

Knowledge of disease progression aids in making critical disease management decisions such as timing of chemical applications. Without such knowledge successful disease management may be hindered. Although the general epidemiology of bacterial spot is well understood, specific details of temporal disease progress in the eastern US are lacking. Data on bacterial spot severity are generally limited to observations made on a single point in host phenology (e.g. Zehr et al., 1996; Shepard et al., 1999). Where multiple bacterial spot evaluations have been made over time, data have usually been analyzed by means of area under disease progress curves (AUDPC) (e.g. Lalancette and

McFarland, 2007). Although useful for comparing treatments (e.g., bactericides evaluated for disease control), analyses based on AUDPCs fail to account for other epidemiologically important factors such as the timing of defoliation as well as the role of cultivar effects in disease progress.

For example, it is not known how disease progression differs with respect to different cultivars. In particular, it is not known if disease progress ceases early in the summer given that some cultivars mature as early as July while others remain in vegetative growth. Documenting such a result would support recommendations for early termination of bactericide applications for certain cultivars. It could also be that some cultivars delay the onset of bacterial spot, resulting in reduced disease severity at the end of the season. Such a result would support delayed bactericide applications. Defoliation is a common symptom of bacterial spot but the factors that influence it, such as cultivar, treatment, disease severity, and leaf age, are not well understood. There is an urgent need to understand the temporal progression of bacterial spot in Pennsylvania stone fruit orchards.

#### RELIABILITY AND ACCURACY OF METHODS USED TO EVALUATE BACTERIAL SPOT SEVERITY

Studies on the epidemiology of bacterial spot of stone fruits rely on precise and accurate estimates of disease intensity (incidence and severity) within the observational units (e.g., treatments plots such as cultivars or plots subjected to different bactericide treatments). However, evaluating bacterial spot intensity (incidence and severity) on peach and nectarine is a challenging task. Estimates of foliar disease incidence may not be very informative because nearly all plants within a plot will have at least one infected leaf, thereby necessitating the determination of “the degree of plant infection” (i.e. disease severity; Madden et al., 2007), which is difficult to quantify for bacterial spot. For most leaf spot diseases, disease severity at the individual leaf scale can be quantified by counting the number of lesions or spots per leaf (Madden et al., 2007). However, bacterial spot lesions often coalesce making disease counts difficult and prone to errors during visual estimation. Rating scales have often been used to quantify foliar disease

severity in epidemiological studies including those on bacterial spot (Battilani et al. 1999; Zehr et al., 1996). With the exception of the study by Citadin et al. (2008), bacterial spot rating scales have not been assessed for reliability (i.e., reproducibility of an estimate) and accuracy (i.e., the closeness of an estimate to the actual value), despite the importance of these scales in making disease management decisions, determining differential cultivar susceptibility, and comparing the effectiveness of chemical treatments. Poorly designed rating scales can bias data that can lead to false conclusions (Bock et al., 2009a).

The Horsfall-Barratt (H-B) disease scale is commonly used to assess foliar plant diseases and was designed to compensate for the error associated with direct visual estimation of disease severity (Horsfall and Barratt, 1945). The H-B scale is a 12 point disease severity scale and assesses the percentage of leaf area affected. Studies on Phomopsis leaf blight of strawberry (caused by *Phomopsis obscurans*) and Citrus Canker (caused by *Xanthomonas axonopodis* pv. *citri*) have shown that the H-B scale was not more reliable or accurate than direct estimation as Horsfall and Barratt originally suggested (Bock et al., 2009a; Nita et al., 2003). These studies also found that the H-B scale was less precise than direct estimation. Moreover, in the case of bacterial spot, diseased leaves often abscise early, usually before they attain the high scores on the H-B scale. Not surprisingly, most studies have utilized alternative disease rating scales for assessing bacterial spot severity. Because of the apparent inadequacies revealed with use of the H-B scale in assessing other bacterial diseases, it is critical to reassess the way foliar symptoms of bacterial spot are currently quantified in the eastern US.

#### COMPARISON OF ASSESSMENTS OF FOLIAR BACTERIAL SPOT SYMPTOMS MADE BY EXPERIENCED AND INEXPERIENCED RATERS

Often, individuals vary in their ability to estimate disease severity (Newton and Hackett, 1994). When multiple raters with varying levels of experience are needed to assess disease severity, the variability in their ability to assess disease may hinder and even bias important epidemiological data (Parker et al., 1995; Bock et al., 2010). However, multiple raters are often required to complete assessment of experimental

treatments. For example, bactericide trials are used to determine the most efficacious chemical or chemical combinations to manage bacterial spot. A rater with poor ability or multiple raters with inconsistent assessments may fail to detect differences in chemical treatments. Failure to statistically recognize a difference between treatments when indeed there is a difference would result in a Type II error (Everitt, 1999). This critical error would also result in incorrect recommendations to growers (Bock et al., 2009b). There is a need, therefore, to assess the reliability and accuracy of bacterial spot assessments made by raters with different levels of experience in order to standardize the collection of data in epidemiological studies for this disease.

## GOALS AND OBJECTIVES

The overall goal of this study is to develop methods to quantify bacterial spot severity in order to improve the quality of data collected as well as to gain a better understanding of the epidemiology of bacterial spot in Pennsylvania peach and nectarine orchards. The specific objectives are to:

- 1. Assess the reliability and accuracy of visual methods (direct estimation and use of an ordinal rating scale) used to quantify bacterial spot severity by comparing results to those of visual estimation and computer image analysis.**

The aim of this project is to compare the accuracy (i.e., the closeness of an estimate to the actual value) and reliability (i.e., reproducibility) of bacterial spot severity estimates obtained using the direct estimation method with estimates obtained with the use of a 1 to 7 ordinal rating scale. Data obtained from three sets of leaves rated by the same experienced rater were used for this analysis and details are described in Chapter 2.

- 2. Compare the reliability and accuracy of assessments of foliar bacterial spot severity made by experienced and inexperienced raters using the direct estimation method.** The aim of this project is to compare the accuracy and reliability of bacterial spot severity estimates made by raters with varying levels of experience using the direct estimation method. Data obtained from three sets of leaves (310 leaves total) rated by 4 experienced raters and 6 inexperienced raters were used for this analysis and details are described in Chapter 3.

- 3. Determine the temporal progression of bacterial spot epidemics of peach and nectarines in Pennsylvania orchards.** The aim of this project is to model the temporal progression of bacterial spot epidemics in Pennsylvania peach and nectarine orchards in order to determine the rate at which epidemics progress with respect to different cultivars and chemical treatments. It is also the aim of this study to determine what factors affect the timing of severe defoliation associated with this disease and details are described in Chapter 4.

## LITERATURE CITED

- Anonymous. 2007. Specified Fruits and Nuts by Acres: 2007 and 2002. 2007 Census of Agriculture – State Data – Pennsylvania. USDA-NASS, Washington, DC.
- Anonymous. 2009. Noncitrus Fruits and Nuts, 2008 Summary. Fr Nt 1-3(09)a, USDA NASS, Washington, DC.
- Battilani, P., Rossi, V., and Saccardi, A. 1999. Development of *Xanthomonas arboricola* pv. *pruni* epidemics on peaches. J. Plant Pathol. 81: 161-171.
- Bock, C.H., Gottwald, T.R., Parker, P.E., Cook, A.Z., Ferrandino, F., Parnell, S., and van den Bosch, F. 2009a. The Horsfall-Barratt scale and severity estimates of citrus canker. Eur. J. Plant Pathol. 125:23-38.
- Bock, C.H., Poole, G.H., Parker, P.E. and Gottwald, T.R. 2010b. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. Critical Reviews in Plant Sciences, 29:59-107.
- Bock, C.H., 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.
- Brannen, P. 2006. Peach. Page 10 in: 2005 Georgia Plant Disease Loss Estimates. A. Martinez, ed. CAES Special Bullentin 41-08, Univ. of Georgia, Athens.
- Buttner, D., Nennstiel, D. Klusener, B., and Bonas, U. 2002. Institut fur Genetik. Functional Analysis of HrpF, a Putuative Type III Translocon Protein for *Xanthomonas campestris* pv. *vesicastoria*.
- Citadin, I., Assmann, A.P., Mazaro, S.M., de Gouvêa, A., Danner, M.A., Malagi, G. 2008. Diagramatic scale to evaluate the severity of bacteriose in peach trees. Revista Brasileira de Fruticultura 30:327-330.
- Data Sheets on Quarantined Pests: *Xanthomonas arboricola* pv. *pruni*. Prepared by CABI and EPPO for the EU under Contract 90/399003.
- Everitt, B.S. 1999 Dictionary of Statistics, Cambridge University Press, U.K.Flaherty.
- Horsfall, J.G. and Barratt, R.W. 1945. An improved grading system for measuring plant disease. Phytopathology 35:655.

- Lalancette, N., and McFarland, K.A. 2007. Phytotoxicity of copper-based bactericides to peach and nectarine. *Plant Dis.* 91:1122-1130.
- Madden, L.V., Hughes, G., van den Bosch, F. 2007. *The Study of Plant Disease Epidemics.* The American Phytopathological Society, St. Paul.
- Newton, A.C. and Hackett, C.A. 1994. Subjective components of mildew assessment on spring barely. *Eur. J. Plant Pathol.* 100:395-412.
- Ngugi, H.K., Bardsley, S.J., and Halbrendt, N.O. 2009. Refining management strategies for bacterial spot of peach. *Pennsylvania Fruit News* Vol. 89 (2):18-24.
- 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.
- Parker, S.R., Shaw, M.W., and Royle, D.J. 1995. Reliable measurement of disease severity. *Aspects Appl. Biol.* 43:205-214.
- Ritchie, D.F. 1995. Bacterial spot. Pages 50-52 in: *Compendium of Stone Fruit Diseases.* J.M. Ogawa, E.I. Zehr, G.W. Bird, D.F. Ritchie, K. Uriu, and J.K. Uyemoto, eds. The American Phytopathological Society, St. Paul, MN.
- Shepard, D.P., Zehr, E.I., and Bridges, W.C. 1999. Increased susceptibility to bacterial spot of peach trees growing in soil infested with *Croconemella xenoplax*. *Plant Dis.* 83:961-963.
- Smith, E.F. 1903. Observations on a hitherto unreported bacterial disease, the cause of which enters the plant through ordinary stomata. *Science* 17:456-457.
- Vauterin, L., Hoste, B., Kersters, K., and Swings, J. 1995. Reclassification of *Xanthomonas*. *Int. J. Syst. Bacteriol.* 45:472-489.
- Zaccardelli, M., Malaguti, S., and Bazzi, C. 1998. Biological and epidemiological aspects of *Xanthomonas arboricola* pv. *pruni* on peach in Italy. *J. Plant Pathol.* 80:125-132.
- Zaccardelli, M., Ceroni, P., and Mazzucchi, U. 1999. Amplified fragment length polymorphism fingerprinting of *Xanthomonas arboricola* pv. *pruni*. *J. Plant Pathol.* 81:173-179.
- Zehr, E.I., Shepard, D.P., and Bridges, W.C., Jr. 1996. Bacterial spot of peach as influenced by water congestion, leaf wetness duration, and temperature. *Plant Dis.* 80:339-341.

## Chapter 2

### **Evaluation of Direct Estimation and an Ordinal Scale Used to Quantify Severity of Bacterial Spot of Peach and Nectarine**

#### ABSTRACT

Bacterial spot caused by *Xanthomonas arboricola* pv. *pruni* is the most important bacterial disease of peach and nectarine in the eastern US where severe epidemics can result in 100% yield loss on susceptible cultivars. Epidemiological studies on bacterial spot depend on the quality of visual estimates of disease severity. The objective of this study was to assess the reliability and accuracy of visual methods used to quantify bacterial spot severity by comparing results of visual estimation with those of computer image analysis. Three sets of peach or nectarine leaves ( $n = 103, 103,$  and  $104$  leaves; disease severity levels from 0 to 100%) were assessed twice by one experienced rater using direct visual estimation (percent leaf area covered by necrotic and chlorotic lesions) and a 1 to 7 rating scale (1 = 0% lesion area, and 7 = >45%). The same leaves were also assessed with the APS Assess image analysis software. Based on Lin's concordance analysis of continuous data, the direct estimation method was superior to the rating scale due to systematic bias introduced when converting scale data into percentage. The Lin's concordance coefficient ( $\rho_c$ ) values for data obtained with the rating scale were 0.41 to 0.11 lower than data obtained with the direct estimation method when assessing accuracy and were 0.026 to 0.033 lower when assessing reliability. Although with untransformed data, accuracy and reliability values associated with the rating scale were improved, comparison of  $\rho_c$  and  $r$  values for the two methods based on bootstrap analysis indicated that direct estimation resulted in significantly ( $P < 0.05$ ) more accurate estimates than the ordinal rating scale. Because comparable levels of reliability and accuracy cannot be retained in data obtained with the scale after mid-point transformation, estimates based on direct estimation methods are ultimately superior to those obtained with the scale. These results together with the statistical properties of the resulting data indicate direct estimation is a more desirable method for assessing bacterial spot severity on stone fruits.

## INTRODUCTION

Bacterial spot of stone fruits, particularly peach and nectarine, is a yield limiting disease caused by the gram-negative bacterium *Xanthomonas arboricola* pv. *pruni*. Considered the most important bacterial disease of peach and nectarine, symptoms occur on leaves, twigs, and fruit. Foliar lesions are angular in shape, delimited by the veins, and often surrounded by a chlorotic halo (Zehr et al. 1996). As lesions age, the centers abscise from the leaf, developing a “shot-hole” appearance (Ritchie, 1995). Yellowing of all or part of the leaf is also common. Severe foliar infection ultimately reduces yield due to reduced photosynthetic competence and carbohydrate uptake (Crisosto et al., 1995).

Evaluating bacterial spot intensity (incidence and severity) is a challenging task. Estimates of foliar disease incidence may not be very informative because nearly all plants within a plot will have at least one infected leaf, thereby necessitating the determination of “the degree of plant infection” (i.e. disease severity; Madden et al., 2007), which is difficult to quantify for bacterial spot. As with most leaf spot diseases, disease severity at the individual leaf scale can be quantified by counting the number of lesions or spots per leaf (Madden et al., 2007). However, although symptomatic area remains distinguishable from asymptomatic tissue, bacterial spot lesions often coalesce making disease counts difficult and prone to errors during visual estimation. Moreover, in severe epidemics, bacterial spot symptoms are often characterized by chlorosis and leaf yellowing, which is more difficult to quantify. While not a true lesion, leaf yellowing is a symptom of bacterial spot (Ritchie, 1995) and reduces the photosynthetic competence of the leaf area affected (*unpublished data*). The gradations of yellow and light green do not easily contrast with visually healthy leaf tissue or lack clearly marked boundaries, a situation that greatly complicates the estimation of disease severity.

Common methods of visual estimation of disease severity include direct estimation with or without the aid of disease diagrams, and the use of disease rating scales (Nita et al., 2003; Bock et al., 2008; Madden et al., 2007). In direct visual estimation, a person rating disease severity (hereafter referred to as the rater) examines a sample (e.g., a leaf with disease symptoms) and assigns it a severity value from 0 to

100% based on the perceived area of diseased tissue relative to the total area of the sample. With the aid of disease diagrams, the rater assigns the severity value (0-100%) based on the closeness of the perceived severity of the sample to one of the values in the diagram, a graphic depiction of selected levels of disease severity. A disease rating scale typically involves the division of the continuous severity values (0-100%) into a fixed number of classes so that a rater assigns the sample a value indicating the perceived disease class. An example of a disease rating scale is the Horsfall-Barratt (H-B) scale (Horsfall and Barratt, 1945) which is widely used in phytopathological studies (Bock et al., 2008; Nita et al., 2003). Disease severity classes in most rating scales tend to be inherently ordered and therefore data obtained with a rating scale comprise ordinal categorical variables (Stokes et al., 2000).

No matter what method of visual estimation is used, disease severity assessments should be reliable and accurate (Campbell and Madden, 1990). Reliability is the degree to which the measurements of the same diseased individuals obtained under different conditions produce similar results (Everitt, 1999). Intra-rater reliability is the agreement between measurements (e.g., disease severity) taken repeatedly by the same raters (Madden et al. 2007), while inter-rater reliability is the agreement between measurements of the same diseased specimen as rated by multiple raters. Accuracy is the closeness of an estimate to the actual value (Everitt, 1999). An accurate estimate of disease severity is close to if not the same as the actual disease severity value. Actual disease severity is difficult to measure and often a “gold standard”, such as computer image analysis, is used as an accepted measurement of actual disease severity (Lin, 1989; Bland and Altman, 1999; Shoukri, 2004).

Rating scales have typically been used to quantify foliar disease severity in epidemiological studies including those on bacterial spot (Battilani et al. 1999; Zehr et al., 1996). Preliminary studies in our lab also suggest that using a rating scale was a faster method of obtaining estimates of disease severity. However, recent studies with other bacterial diseases have suggested that direct estimation often results in more reliable and accurate estimates of disease severity (Bock et al., 2009a; Bock et al., 2009b). With the exception of the study by Citadin et al. (2008), bacterial spot rating scales have not been assessed for reliability (i.e., reproducibility of an estimate) and

accuracy (i.e., the closeness of an estimate to the actual value), despite the importance of these scales in making disease management decisions, determining differential cultivar susceptibility, and comparing the effectiveness of chemical treatments. Poorly designed rating scales can bias data that can lead to false conclusions (Bock et al., 2009). The objectives of this study are to assess the intra-rater reliability and accuracy of visual methods used to quantify bacterial spot severity by comparing results to those of visual estimation and computer image analysis.

## MATERIALS AND METHODS

*Plant Pathology Peach and Nectarine Block at the Penn State Fruit Research and Extension Center, Biglerville, PA:* This is a 4-cultivar (Easterglo, Beekman, Snow King, and Sweet Dream) peach and nectarine block that was planted in 2006 and will hereafter be referred to as 'FREC block'. This experimental block received standard fungicide and insecticide applications according to the commercial practices in the northeastern United States. However, no dormant copper sprays were applied in order to permit the buildup of *X. arboricola* pv. *pruni* inoculum. In 2008, this block was inoculated on 8 May with a suspension of  $1 \times 10^7$  CFUs of *X. arboricola* pv. *pruni*. The suspension consisted of a mixture of strains Xap 42 and Xcp 88301 and was applied to leaves until dripping using a spray bottle. In 2009, this block was inoculated on 1 May with a suspension of Xap that consisted of strains Xap 42, Xcp 88-301, and DSLB Xap2. The suspension was applied to leaves until dripping using spray bottles. On May 14, 2009, the antibiotic oxytetracycline (Mycoshield) which is the standard treatment for bacterial spot, was accidentally sprayed but infection still occurred.

*Leaf Collection:* Leaves used for this experiment were obtained from three cultivars in non-treated control plots of a product evaluation experiment testing the efficacy of bactericides for the management of bacterial spot in the FREC block. Before collection, leaves were rated based on a 1 to 7 disease severity scale (1 = 0% lesion area, 2 = 1-3%, 3 = 4-8%, 4 = 9-15%, 5 = 15-25%, 6 = 25-45%, 7 = >45% (Fig. 2-1). This scale was designed to replace the H-B scale that has commonly been used to evaluate foliar disease symptoms (Flaherty et al., 2000; Al-Dahmani et al., 2003). Experience

with bacterial spot development in the eastern US indicated that on most cultivars, leaves with about 50% diseased area become chlorotic and abscise. Bacterial spot ratings based on the H-B scale would therefore not accurately represent leaves with high disease severity. A total of six sets of 105 leaves were collected from the cultivars Beekman, Sweet Dream, and Easternglo, with two sets taken from each cultivar. The sets were comprised of 15 leaves of each scale rating (i.e., 15 leaves assigned a rating of 1, 15 leaves that received a rating of 2, etc.). All of the leaves were collected and rated by one individual familiar with bacterial spot symptoms. Leaves were pressed using large books in order to ensure a completely flat surface to facilitate digital scanning. Flattened leaves were scanned at 300 dpi and saved as Tagged Image File Format (tiff) files.

*Image Analysis:* The APS Assess 2.0 image analysis software for plant disease quantification (Lakhdar Lamari. American Phytopathological Society) was used to determine the actual percent of diseased area of each leaf. This program uses image analysis algorithms including the Hue-Saturation-Intensity (HSI) color model with the mathematical transformation of the RGB model where,  $y = \text{blue}$ ,  $x = \text{green}$ , and  $z = \text{red}$ . The HSI model first separates the image into the leaf and the background and then separates the lesions from the leaf. While assessing the leaves using Assess 2.0, the color space was set to 'HSI' and the color plane was set to 'high'. Threshold levels for leaf and lesion were set accordingly for each individual scanned leaf and were recorded for purposes of reproducibility. To ensure the most accurate disease severity assessment, lesions were first colored black using the Microsoft Paint program.

*Visual Assessment:* The scanned images of three sets of leaves (a total of 315 leaves), one set from each cultivar, were uploaded to a Microsoft Power Point presentation. On each slide, there was only one leaf with a random number indicating the position of the leaf in the list (i.e., the order in which they were displayed to an assessor). Displayed on a computer, the same experienced rater visually rated each leaf for disease severity using the rating scale twice with a 10 to 15 minute break between ratings of each set. The rater then twice evaluated the same leaves using direct visual estimation of the percent leaf area covered by lesions with a similar break between the two repeats.

*Data Analysis:* For both reliability and accuracy, data obtained with direct estimation method was analyzed without further transformation using methods for

continuous variables as described below (Rousson et al., 2002). Two approaches were used for the analyses of data obtained with the disease severity rating scale. First, the data were converted from the 1 to 7 ordinal scale to percent disease severity and subjected to the analyses for continuous measurements used for the direct estimation data. However, several studies have shown that transforming data from ordinal scales to percentage introduces an error as mid-points of each percentage severity must be used (Bock et al. 2009). Non-transformed data obtained with the rating scale were therefore also assessed for accuracy and reliability using concordance statistics for categorical data (Svensson, 2000). To facilitate the assessment of accuracy in data obtained with the rating scale, data on actual disease severity obtained with the APS Assess image analysis software were converted to the 1 to 7 ordinal scale used for actual ratings with a minor modification. In this transformation, values from the APS Assess  $\geq 0.25\%$  were considered to represent leaves with no disease and assigned a score of 1 on the scale.

For continuous variables, intra-rater reliability was assessed based on Pearson's moments correlation coefficient ( $r$ ) which quantified the variability around the straight line relationship between severity estimations from the first and second assessments (Nita et al., 2003; Rousson et al., 2002). The correlation coefficient was a measurement of precision. Intra-rater reliability and accuracy were also determined by Lin's concordance correlation coefficient ( $\rho_c$ ), a type of intra-class correlation (Lin, 1989) that measures the extent to which two sets of observations align on the line of unity (i.e.,  $y = x$ ) also known as the line of concordance. Lin's concordance correlation coefficient is a product of  $r$  and  $C_b$ , the bias coefficient which measures how much the best-fit line deviates from the concordance line and where a  $C_b$  of 1 represents a relationship with no bias (Bock et al., 2008). The coefficient  $v$  measures the difference between the slopes of the concordance line and the best fit line where a  $v$  of 1 represents a perfect relationship between the two slopes. The coefficient  $u$  is the location of the best-fit line relative to the concordance line (Nita et al., 2003). The coefficients  $v$  and  $u$  are both used to calculate  $C_b$ :

$$C_b = \frac{2}{(v + 1/v + u^2)}$$

and

$$v = \sigma_1/\sigma_2 \text{ and } u = (\mu_1 - \mu_2)/\sqrt{\sigma_1 \times \sigma_2}$$

The terms  $\mu_1$  and  $\mu_2$  are the means for the estimated values and the actual values, respectively. The terms  $\sigma_1$  and  $\sigma_2$  are the standard deviations for the estimated values and the actual values, respectively. A perfect agreement between the actual and estimated severity is represented by  $\rho_c = 1$ . The actual relationship, or the best fitting line, was determined with linear regression analysis where the estimated severity values ( $X$ ) were predictors of the actual values ( $Y$ ). For analysis of the disease rating scale, the midpoints of each class values were used. All concordance analysis were computed using GenStat 12 (VSN International, UK) while graphs were plotted using SigmaPlot 10 software (Systat Software, DE).

Reliability and accuracy of the categorical data obtained with the rating scale were assessed with several concordance statistics commonly used for categorical data (Svensson, 2000; Fleiss, 1981), including the Cohen's weighted kappa coefficient ( $\kappa_w$ ) and associated 95% confidence interval, the Goodman and Kruskal's gamma, Kendall's tau-b, Stuart's tau-c, and Somer's D. Lin's concordance correlation coefficient ( $\rho_c$ ), the bias coefficient ( $C_b$ ), Pearson correlation ( $r$ ), and the Spearman rank correlation coefficient were also determined. The Cohen's weighted kappa coefficient, determines the degree of agreement between two ratings and is used for categorical variables with more than two classes (Cohen, 1968) such as the 1 to 7 scale in the present study. The weighed statistic has values of 0 to 1 where a value of 0 reflects no agreement and a value of 1 indicates perfect agreement (Cohen, 1968). The Cohen's weighted kappa coefficient  $\kappa_w$  was computed as:  $\kappa_w = \frac{P_{a(w)} - P_{e(w)}}{1 - P_{e(w)}}$ , where  $P_{a(w)}$  is the weighted probability of agreement (i.e., concordance) among observations and  $P_{e(w)}$  is the hypothetical agreement due to chance. The weighted Cohen's kappa was computed with the Fleiss-Cohen weights (Fleiss and Cohen, 1973).

The gamma statistic, which may be less than or equal to 1, is the percentage of difference between the two estimated assessments or the actual and estimated severity values and is expressed as  $G = \frac{S_C - S_D}{S_C + S_D}$ , where  $S_C$  and  $S_D$  refer to the number of concordant and discordant pairs, respectively (Goodman and Kruskal, 1954). Kendall's tau-b is a correlation coefficient for rank-ordered categorical data, adjusted for tied observations, may be equal to or less than 1, and is expressed as  $T_b = \frac{2(S_C - S_D)}{\sqrt{\{n(n-1) - t^{(x)}\}(n(n-1) - t^{(y)})}}$ , where

$S_C$  and  $S_D$  refer to the number of concordant and discordant pairs, respectively,  $n$  is the number of paired observations, and  $t^{(x)}$  and  $t^{(y)}$  refer to the number of tied observations on  $X$  and  $Y$ , respectively (Kendall, 1945). Stuart's tau-c, which may be greater than or equal to -1 and less than or equal to 1, is used to compare assessments made by scales with dissimilar numbers of categories and can be expressed as  $T_C = \frac{2(S_C - S_D)}{n^2 \frac{(m-1)}{m}}$ , where  $S_C$  and  $S_D$ , again, refer to the number of concordant and discordant pairs, respectively,  $n$  is the number of paired observations, and  $m$  is the number of categories in the scale (Stuart, 1953). Somers' D is an asymmetric measure of association where the distinction is made between the independent and dependent variables and is expressed as  $d_{YX} = \frac{2(S_C - S_D)}{n(n-1) - t^{(x)}}$ , where  $S_C$  and  $S_D$  refer to the number of concordant and discordant pairs, respectively,  $n$  is the number of paired observations, and  $t^{(x)}$  refers to the observations tied on  $X$  when  $X$  is regarded as the independent variable (Somers, 1962). When  $Y$  is regarded as the independent variable,  $d_{XY}$  is calculated accordingly (Svensson, 2000). The Spearman rank correlation coefficient is a non-parametric measure of correlation between two variables and was computed with adjustment for ties. The Lin's concordance correlation coefficient ( $\rho_c$ ), the bias coefficient ( $C_b$ ), and the Pearson correlation ( $r$ ) are as previously described for continuous variables (Nita et al., 2003). A bootstrap analysis of accuracy statistics ( $\rho_c$  and  $r$ ) comparing data obtained with the rating scale and the direct estimation method was performed in order to compute 95% confidence intervals necessary to determine if values based on direct estimation were significantly higher than values obtained with the rating scale. A total of 2000 balanced bootstrap samples were created for the analysis in GenStat 12.

## RESULTS

*General:* Each of the three sets of leaves was assessed twice using the direct estimation method and twice using the rating scale. Average estimated severity values did not differ much within each set (Table 2-1). However, average estimated severity tended to differ between the sets of leaves. The median was usually higher for severity estimated using the rating scale than those estimated directly. Disease severity estimates ranged from 0.0% to 100.0% and the minimum estimated value was the same across all

leaf sets. The maximum, however, was always higher for assessments estimated directly due to the limited nature of calculating statistics using the midpoint for each level of the rating scale. For all three sets, and for both direct estimation and the disease severity rating scale, the estimated severity was closely correlated with the actual severity values (Fig. 2-2). There was a slight trend of overestimation in the second set while; visual estimates for set 3 displayed a slight trend of underestimation. These effects were most pronounced for data obtained with the direct estimation method. Individual severity estimates based on the rating scale resulted in a series of horizontal lines due to the limited number of possible values (the midpoint of only 7 class intervals) (Fig. 2-2).

*Reliability of visual estimation:* In general, intra-rater reliability determined by the correlation coefficient,  $r$ , was high for all 3 sets of leaves and for both the direct estimation and the rating scale (Table 2-2). The  $r$  values were greater than 0.90 for all assessments using direct estimation and the rating scale. For the direct estimation,  $r$  values ranged from 0.948 to 0.990 while for the rating scale,  $r$  values ranged from 0.919 to 0.965. Across all three sets of leaves, the  $r$  value was 0.024 to 0.029 greater for the direct estimation than it was for the rating scale. The Lin's concordance correlation coefficient ( $\rho_c$ ), was also high for all 3 leaf sets and for both the direct estimation and the rating scale (Table 2-2; Fig. 2-3). The  $\rho_c$  values for all assessments using the direct estimation and the rating scale were also greater than 0.90. For the direct estimation,  $\rho_c$  values ranged from 0.947 to 0.990 and for the rating scale,  $\rho_c$  values ranged from 0.914 to 0.962. For all 3 leaf sets, the  $\rho_c$  values were between 0.026 and 0.033 greater for the direct estimation than it was for the rating scale.

Analysis of concordance for categorical variables, determined by the concordance statistics for categorical data and Lin's concordance analysis for the untransformed data obtained with the rating scale, was high for all three sets of leaves (Table 2-3). Values of Cohen's weighted kappa ranged from 0.954 to 0.975. The Z-tests for all 3 leaf sets were significant (two-tailed  $P < 0.0001$ ), indicating a high level of agreement between the two assessments for each leaf set. Pearson correlation coefficient ( $r$ ) ranged from 0.960 to 0.975, which were from 0.01 to 0.041 higher than  $r$  values for the transformed rating scale. The Spearman rank correlation coefficient ranged from 0.964 to 0.974. The gamma statistic ranged from 0.990 to 0.993. This indicated that the number of

concordant pairs of disease severity scores as a percentage of all pairs of scores, when ignoring ties, was greater than 99%. Kendall's tau-b, Stuart's tau-c, Somers' D C|R, and Somers' D R|C ranged from 0.918 to 0.937, from 0.894 to 0.937, from 0.918 to 0.935, and from 0.917 to 0.939, respectively, indicating that the probability that any two observations were concordant (i.e. the same leaf was assigned the same score at the two separate ratings) exceeded 90% for all three sets after adjusting for ties (Table 2-3). Lin's concordance correlation coefficient ( $\rho_c$ ) ranged from 0.955 to 0.975, which was from 0.009 to 0.041 higher than  $\rho_c$  values for the same data transformed to percentage before concordance analysis. The bias coefficient ( $C_b$ ) ranged from 0.994 to 1.000, which is also slightly higher than the  $C_b$  values obtained for the transformed data (Table 2-2).

*Accuracy of Visual Estimation Methods:* Accuracy was determined for the 1<sup>st</sup> and 2<sup>nd</sup> assessments of the 3 sets of leaves with direct estimation and the rating scale using Lin's concordance analysis. Based on Lin's concordance analysis, the  $\rho_c$  values were generally high for all 6 assessments (2 assessments for each set of leaves) and for both methods of assessment; i.e., disease severity estimates using the direct estimation and for data obtained with the rating scale transformed to a percentage before the analysis (Table 2-4). However, accuracy was, in general, less than reliability. For all 3 leaf sets and both assessments, the  $\rho_c$  values for data obtained with direct estimation, were all greater than 0.90. Only 2 out of the 6 assessments using the rating scale followed by transformation, however, had a  $\rho_c$  value greater than 0.90, although the remaining values were greater than 0.80. For the direct estimation,  $\rho_c$  values ranged from 0.908 to 0.982 while  $\rho_c$  values ranged from 0.816 to 0.921 for the rating scale (Table 2-4). For all 6 assessments, the  $\rho_c$  values were greater for the direct estimation than they were for the rating scale with  $\rho_c$  values for direct estimation being 0.41 to 0.11 higher than  $\rho_c$  values for the rating scale. Results of the bootstrap analysis indicate that  $\rho_c$  values assessing the accuracy of the rating scale were significantly ( $P < 0.05$ ) lower than those obtained for the direct estimation method. This was regardless of whether data obtained from the scale was first transformed before analysis or analyzed as a categorical variable without transformation (Table 2-5).

The results for the accuracy of the three leaf sets for both the direct estimation and the rating scale are shown in Figures 2-4 and 2-5 and indicate that there was a linear

relationship between actual and estimated disease severity for all assessments. As the accuracy is the product of the bias ( $C_b$ ) and precision ( $r$ ), the  $C_b$  was greater for the direct estimation than it was for the rating scale. The  $C_b$  values ranged from 0.977 to 0.995 for the direct estimation with 0.885 to 0.963 for the rating scale. The estimates of  $r$  were also generally higher for the direct estimation than it was for rating scale. The  $r$  values ranged from 0.929 to 0.988 for the direct estimation and from 0.922 to 0.974 for the rating scale.

Accuracy, determined by the categorical concordance statistics and Lin's concordance analysis for the 1<sup>st</sup> and 2<sup>nd</sup> assessments of the untransformed rating scale, was high for all three sets of leaves (Table 2-6). Values of Cohen's weighted kappa ranged from 0.916 to 0.943 and the Z-test indicated that these statistics were significantly greater than zero (two-tailed  $P < 0.0001$ ) for all 6 assessments. The Pearson correlation coefficient ( $r$ ) ranged from 0.936 to 0.954, which was slightly higher than the  $r$  values calculated for the same data following transformation to a percentage scale (Table 2-4). Values of  $r$  for the direct estimation method, however, were still significantly ( $P < 0.05$ ) greater than  $r$  values calculated for the non-transformed rating scale, and results of the bootstrap analysis indicated that the difference between these values was significant (Table 2-5). Likewise, the values of Spearman rank correlation coefficient ranged from 0.921 to 0.956. The gamma statistic ranged from 0.966 to 0.992, indicating that the number of concordant pairs of disease severity scores as a percentage of all pairs of scores, when ignoring ties, was greater than 96%. Kendall's tau-b, Stuart's tau-c, Somers' D C|R, and Somers' D R|C ranged from 0.854 to 0.902, from 0.815 to 0.871, from 0.865 to 0.9142, and from 0.844 to 0.892, respectively, indicating that the probability that any two observations were concordant exceeded 81% for all three sets after adjusting for ties (Table 2-6). Lin's concordance correlation coefficient ( $\rho_c$ ) ranged from 0.916 to 0.943, which was from 0.0149 to 0.0997 higher than  $\rho_c$  values for the transformed scale. Values of  $\rho_c$  for the direct estimation method, however, were greater than those for the non-transformed rating scale and results of the bootstrap analysis indicated that these values were significantly different from each other (Table 2-5). The bias coefficient ( $C_b$ ) ranged from 0.969 to 0.989, which was from 0.0159 to 0.0928 higher than the  $C_b$  values for the same data when subjected to mid-point transformation to a percentage (Table 2-4).

## DISCUSSION

Successful disease management relies heavily on high quality epidemiological studies where accurate and reliable data are essential. Estimates of severity for bacterial leaf spot diseases are commonly obtained with the use of disease rating scales (Battilani et al., 1999; Flaherty et al., 2000; Al-Dahmani et al., 2003; Pataky et al., 1997), or by direct estimation methods presumably because lesions coalesce making their counts impractical. The Horsfall-Barratt (H-B) disease scale, commonly used to assess foliar plant disease symptoms, is a 12 point disease severity scale designed to compensate for the error associated with direct visual estimation of disease severity (Horsfall and Barratt, 1945). Due to the nature of foliar bacterial spot symptoms, however, the H-B scale was impractical as leaves with more than 50% area covered with bacterial spot are generally uncommon (Battilani et al. 1999), or often abscised after reaching more than 50% disease severity. Indeed, previous studies have generally not used the H-B scale for bacterial spot evaluations, instead preferring to use scales with fewer categories and lower maximum disease levels about 55% (Battilani et al., 1999; Zehr et al., 1996; Citadin et al., 2008). Because of these considerations, the 1-7 rating scale was created to allow for small intervals for severity levels less than 45% and to terminate at the point where leaves would abscise. The resulting severity scale was similar to those of Battilani et al. (1999), or Zehr et al. (1996) but while non-linear, could be linearized with a natural logarithm transformation.

Regardless of the method of assessment, an important issue relating to disease measurements is that of validity. That is, how *accurate* and *reliable* are the measurements in depicting the true degree of the extent to which the observed sampling units are diseased. With a new method, such as the scale developed in the present study, it is imperative that the method is evaluated for compliance with the 'standard' or other available methods (Bland and Altman, 1999). In this study, the accuracy and reliability of estimates of bacterial spot severity on peach and nectarine leaves obtained with direct estimation or the use of an ordinal 1-7 disease rating scale were assessed based on statistical procedures suited for assessing agreement in continuous and ordinal variables.

Lin's concordance analysis (Lin 1989) and linear regression were used to determine the accuracy and reliability of the bacterial spot severity assessment obtained with the direct estimation method as well as the ordinal 1-7 rating scale after converting the data to a continuous percentage scale. Direct estimation was more reliable and more accurate than the rating scale as a means of measuring bacterial spot disease severity of peach and nectarine based on data analyzed as continuous variables. The correlation coefficient ( $r$ ) and Lin's concordance correlation coefficient ( $\rho_c$ ), intra-rater reliability, were higher for the data obtained with the direct estimation method than for the rating scale. Furthermore, bootstrap analysis of  $\rho_c$  and  $r$  values for accuracy confirmed that direct estimation was significantly more accurate than the rating scale (Table 2-5).

To facilitate direct comparison of visual and direct estimation methods, it was necessary to use the mid-points of the rating scale to convert data from the ordinal scale to a continuous variable. This is a popular and widely used approach to concordance analysis in phytopathological literature (e.g., Nita et al., 2003; Bock et al., 2008; Madden et al. 2007). However, this practice ultimately adds bias to the analysis (Bock et al., 2009). When actual disease severity is similar to or the same as the scale midpoint, the scale provides high levels of accuracy, but when the actual disease is closer to the scale interval boundaries, accuracy is substantially reduced (Bock et al., 2009; Bock et al., 2010). This bias is greater for rating scales with fewer intervals, such as the scales developed for bacterial spot estimation including the one presented here, due to the limited number of possible severity values that could be assigned to each leaf (Battilani et al., 1999; Zehr et al., 1996; Citadin et al., 2008). Therefore, it was no surprise that in analyses based on continuous data, the direct estimation method was consistently more accurate and reliable than the rating scale.

These results of the assessment of reliability and accuracy based on continuous variables were also consistent with those of other studies where the H-B scale was found to be less reliable and accurate than the direct estimation method (e.g.: Nita et al., 2003; Bock et al., 2009b; Bock et al., 2010). In this study,  $r$  values, measuring reliability, ranged from 0.965 to 0.919 for data obtained by the rating scale and from 0.948 to 0.990 for data obtained by direct estimation and were comparable to the values obtained by Nita et al. (2003) (0.61 to 1.0 for the H-B scale and from 0.71 to 1.0 for direct estimation).

Similarly, the values of  $\rho_c$  obtained in assessment of accuracy (0.816 to 0.921 for data obtained by the rating scale and from 0.908 to 0.982 for data obtained by direct estimation) were within the range of those obtained previously by Nita et al. (2003) (0.43 to 0.97 for the H-B scale and from 0.51 to 0.99 for direct estimation) and Bock et al. (2009b) (0.23 to 0.88 for the H-B scale and from 0.54 to 0.94 for direct estimation). The higher variation in the  $\rho_c$  and  $r$  values in these earlier studies relative to those found in this study is likely due to the use of multiple raters, with and without experience. Regardless, our results concur with the key findings of these studies in that the values are higher for data obtained by direct estimation than those obtained by a rating scale that had been transformed into continuous data due to introduction of bias during the transformation.

Once the introduction of bias through the use of mid-points is acknowledged, it is questionable whether comparisons between the direct estimation and the rating scale methods based on continuous data are valid. Because our observations of the same sets of leaves were repeated, it was possible to assess, at least the intra-rater reliability of the scale using methods suitable for categorical concordance analysis (Svensson, 2000). The principle behind statistics describing concordance for categorical data (Gamma, Kendall's  $\tau$ , Sommer's, Stuart's and Cohen's kappa coefficients) is that they assess the probability that a single observational unit (in this case a diseased leaf) is assigned the same score in two or more assessments. The analysis of agreement for categorical variables resulted in improved levels of reliability and accuracy of the non-transformed data obtained with the rating scale. The probability that any two pairs of assessments were concordant was greater than 99% for reliability and greater than 96% for accuracy (when ignoring ties) and Spearman's rank correlation coefficients exceeded 0.92 in all cases (Tables 2-3 and 2-6). Likewise, the Lin's concordance coefficients ( $\rho_c$ ), Pearson's correlation coefficients ( $r$ ), and the bias coefficients ( $C_b$ ) increased in this analysis for both reliability (based on repeated estimates of disease severity using the scale) and accuracy (based on converting data obtained with image analysis software to the scale format). Indeed, for one assessment of the 3<sup>rd</sup> leaf set, the rating scale was more reliable and accurate than the direct estimation method (Tables 2-3 and 2-6). The remaining assessments made by the rating scale have levels of accuracy and reliability comparable to those obtained with the direct estimation method. However, the results of the

bootstrap analysis of accuracy statistics ( $\rho_c$  and  $r$ ) for categorical and continuous data obtained with the rating scale and the direct estimation method indicate that, overall, data obtained with direct estimation had significantly greater levels of accuracy than did the rating scale (Table 2-5). Nevertheless, where the objective for disease assessments precludes analyses of variance techniques that necessitate conversion of data from an ordinal to continuous scale, estimates based on the severity scale may be a sufficient and faster method for obtaining data. An example of this would be data intended to rank treatments such as germplasm being evaluated for susceptibility to bacterial spot in breeding programs. In the present study and as noted by Madden et al. (2007), the use of the scale was by far a much faster method of assessing disease severity. The time taken to assess approximately 100 leaves with the rating scale ranged from 7 to 12 minutes compared to the 11 to 15 minutes needed to assess the same leaves using direct estimation. A disease diagram supporting the rating scale may also improve the reliability and accuracy of the rating scale (Nutter et al., 1998; Dos Santos et al., 2009).

The direct estimation method, however, is a more desirable method of disease assessment due to the statistical properties of the resulting data. In botanical epidemiology, disease severity data are almost always intended for use in various types of models (Campbell and Madden, 1990; Madden et al., 2007). Whereas, statistical methods for categorical data analysis are available (e.g., Agresti, 2002; Stokes et al., 2000), many plant disease models assume data are from continuous variables. Data obtained with the use of ratings scales therefore must first be converted into percentage for use in the models, thereby introducing systematic error (Bock et al., 2009b). Thus, because comparable levels of reliability and accuracy cannot be retained in data obtained with scale after mid-point transformation, estimates of bacterial spot severity on peach and nectarine leaves based on direct estimation methods are ultimately superior to those obtained with the scale.

## LITERATURE CITED

- Agresti, A. 2002. Categorical data analysis, 2<sup>nd</sup> ed. New York, NY: Wiley-Interscience.
- Al-Dahmani, J.H., Abbasi, P.A. Miller, S.A. and Hoitink, H.A.J. 2003. Suppression of bacterial spot of tomato with foliar sprays of compost extracts under greenhouse and field conditions. *Plant Dis.* 87:913-919.
- Bastiaans, L. 1991. Ratio between virtual and visual lesion size as a measure to describe reduction in leaf photosynthesis of rice due to leaf blast. *Phytopathology* 81:611-615.
- Battilani, P., Rossi, V., and Saccardi, A. 1999. Development of *Xanthomonas arboricola* pv. *pruni* epidemics on peaches. *J. Plant Pathol.* 81: 161-171.
- Bland, J.M. and Altman, D.G. 1999. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8:135-160.
- Bock, C.H., Parker, P.E., Cook, A.Z., and Gottwald, T.R. 2008. 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. 2009a. Automated image analysis of severity of foliar citrus canker. *Plant Dis.* 93:660-665.
- Bock, C.H., Gottwald, T.R., Parker, P.E., Cook, A.Z., Ferrandino, F., Parnell, S., and van den Bosch, F. 2009b. The Horsfall-Barratt scale and severity estimates of citrus canker. *Eur. J. Plant Pathology* 125:23-38.
- Bock, C.H., Gottwald, T.R., Parker, P.E., Ferrandino, F., Welham, S., van den Bosch, F., and Parnell, S. 2010. Some consequences of using the Horsfall-Barratt scale for hypothesis testing. *Phytopathology* 100: in press.
- Campbell, C.L. and Madden, L.V. 1990. *Introduction to Plant Disease Epidemiology*, John Wiley and Sons, New York.
- Cohen, J. 1968. Weighted kappa: Nominal scale agreement with provision for scaled disagreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* 70:213-220.

- Citadin, I., Assmann, A.P., Mazaro, S.M., de Gouvêa, A., Danner, M.A., Malagi, G. 2008. Diagrammatic scale to evaluate the severity of bacteriose in peach trees. *Revista Brasileira de Fruticultura* 30:327-330.
- Crisosto, C.H., Mitchell, F.G., and Johnson, S. 1995. Factors in fresh market stone fruit quality. *Postharvest News and Information* 6: 17N – 21N.
- Dos Santos, W.R., Alves, M.C., de Carvalho, L.G., Rios, F.A. and Fraga, A.C.J. 2009. Elaboration and validation of a diagrammatic scale for epidemiological studies of *Cercospora* leaf spot of castor bean. *J. Phytopathology* 10: 1-6.
- Everitt, B.S. 1999 *Dictionary of Statistics*, Cambridge University Press, U.K. Flaherty, J.E., Jones, J.B., Harbough, B.K., Somodil, G.C., and Jackson, C.E. 2000. Control of bacterial spot on tomato in the greenhouse and field with H-mutant bacteriophages. *HortScience* 35:882-884.
- Fleiss, J.C. The measurement of interrater agreement. In: *Statistical Methods for Rates and Proportions*. New York, NY: John Wiley & Sons Inc.; 1981:212-236.
- Fleiss, J.C. and Cohen, J. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement* 33:613-619.
- Goodman, L.A. and Kruskal, W.H. 1954. Measures of association for cross classifications. *Journal of the American Statistical Association* 49:732-764.
- Horsfall, J.G. and Barratt, R.W. 1945. An improved grading system for measuring plant disease. *Phytopathology* 35:655.
- Kendall, M.G. 1945. The treatment of ties in ranking problems. *Biometrika* 33:239-251.
- Lin, L.I.-K. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 45:255-268.
- Lin, L., Hedayat, A.S., Sinha, B., and Yang, M. 2002. Statistical methods in assessing agreement: Models, issues, and tools. *J. Am. Stat. Assoc.* 97:257-270.
- Lindow, S.E. and Webb, R.R. 1983. Quantification of foliar plant disease symptoms by microcomputer-digitized video image analysis. *Phytopathology* 73:520-524.
- Madden, L.V., Hughes, G., van den Bosch, F. 2007. *The Study of Plant Disease Epidemics*. The American Phytopathological Society, St. Paul.

- Moya, E.A., Barrales, L.R., Apablaza, G.E. 2005. Assessment of the disease severity of squash powdery mildew through visual analysis, digital image analysis and validation of these methodologies. *Crop Protection* 24: 785-789.
- 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.
- Nutter, F.W. and Schultz, P.M. 1995. Improving the accuracy and perception of disease assessments: Selection of methods and use of computer-aided training programs. *Can. J. Plant Pathol.* 17: 174-184.
- Nutter, F.W. Jr. and Esker, P.D. 2006. The role of psychophysics in phytopathology: the Weber–Fechner Law revisited. *European Journal of Plant Pathology* 114:199-213.
- Nutter, F.W., Litwiller, D. and Wegulo, S.N. 1998. Do standard area diagrams improve the accuracy and precision of disease assessments? *Proceedings of the 7<sup>th</sup> International Congress of Plant Pathology*. p2.1.16.
- Parker, S.R., Shaw, M.W., and Royle, D.J. 1995. The reliability of visual estimations of disease severity on cereal leaves. *Plant Pathology* 44:856-864.
- Pataky, J.K., du Toit, L.J., and Kerns, M.R. 1997. Bacterial leaf blight on *shrunk-2* sweet corn. *Plant Disease* 81:1293-1298.
- Ritchie, D.F. 1995. Bacterial spot. Pages 50-52 in: *Compendium of Stone Fruit Diseases*. J.M. Ogawa, E.I. Zehr, G.W. Bird, D.F. Ritchie, K. Uriu, and J.K. Uyemoto, eds. The American Phytopathological Society, St. Paul, MN.
- Rousson, V., Gasser, T., and Seifert, B. 2002. Assessing the intrarater, interrater and test retest reliability of continuous measurements.
- Sherwood, R.T., Berg, C.C., Hoover, M.R., and Zeiders, K.E. 1983. Illusions in visual assessment of *Stagonospora* leaf spot on orchardgrass. *Phytopathology* 73: 173-177.
- Shoukri, M.M. 2004. *Measure of Interobserver Agreement*, Chapman and Hall/CRC, Boca Raton, FL.
- Shoukri, M.M and Pause, C.A. 1999. *Statistical Methods for Health Science*. 2<sup>nd</sup> ed. CRC Press, Boca Raton, FL.
- Somers, R. 1962. A new asymmetric measure of association for ordinal variables. *American Sociological Review* 27:799-811.

- Stokes, M.E., Davis, C.S., Koch, G.G. 2000. Categorical data analysis using the SAS system, 2<sup>nd</sup> ed. Cary, NC: SAS Institute Inc.
- Stuart, A. 1953. The estimation and comparison of strengths of association in contingency tables. *Biometrika* 40:105-110.
- Svensson, E. 2000. Concordance between ratings using different scales for the same variable. *Statistics in Medicine* 19:3483-3496.
- Zehr, E.I., Shepard, D.P., and Bridges, W.C., Jr. 1996. Bacterial spot of peach as influenced by water congestion, leaf wetness duration, and temperature. *Plant Dis.* 80:339-341.

**Table 2-1.** Summary statistics for bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) severity assessed by one experienced rater on the three sets of peach and nectarine leaves with direct estimation (DE) method or a disease severity rating scale (Scale)

| Statistic | Set 1 (n=103)     |      |         |         | Set 2 (n=103) |      |         |         | Set 3 (n=104) |      |         |         |
|-----------|-------------------|------|---------|---------|---------------|------|---------|---------|---------------|------|---------|---------|
|           | DE-1 <sup>a</sup> | DE-2 | Scale-1 | Scale-2 | DE-1          | DE-2 | Scale-1 | Scale-2 | DE-1          | DE-2 | Scale-1 | Scale-2 |
| Mean      | 16.1              | 16.1 | 20.5    | 19.5    | 26.3          | 26.8 | 24.4    | 27.5    | 19.5          | 19.3 | 22.3    | 24      |
| Median    | 3                 | 5    | 12      | 12      | 7             | 11   | 12      | 20      | 10            | 7    | 20      | 20      |
| Minimum   | 0                 | 0    | 0       | 0       | 0             | 0    | 0       | 0       | 0             | 0    | 0       | 0       |
| Maximum   | 97                | 99   | 72.5    | 72.5    | 100           | 100  | 72.5    | 72.5    | 100           | 100  | 72.5    | 72.5    |

<sup>a</sup>Each set was assessed twice by direct estimation (DE) and the rating scale (Scale) by an experienced rater. The first assessment is noted by 1 and the second assessment is noted by 2.

<sup>b</sup>For the scale, the midpoint of each class was used to calculate the statistics.

**Table 2-2.** Intra-rater reliability, determined by the correlation coefficient ( $r$ ) and Lin's concordance correlation coefficient ( $\rho_c$ ), for the direct estimation (DE) method and a disease severity rating scale (Scale) for the three sets of peach and nectarine leaves with symptoms of bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*). Also included are related statistics including the bias coefficient ( $C_b$ ), the coefficient  $v$ , and the coefficient  $u$ <sup>a</sup>

| Statistics | Set 1 ( $n = 103$ )    |                    | Set 2 ( $n = 103$ ) |         | Set 3 ( $n = 104$ ) |         |
|------------|------------------------|--------------------|---------------------|---------|---------------------|---------|
|            | DE                     | Scale <sup>b</sup> | DE                  | Scale   | DE                  | Scale   |
| $r$        | 0.986                  | 0.962              | 0.990               | 0.965   | 0.948               | 0.919   |
| $\rho_c$   | 0.983                  | 0.957              | 0.990               | 0.962   | 0.947               | 0.914   |
| $C_b$      | 0.997                  | 0.995              | 1.000               | 0.997   | 1.000               | 0.995   |
| $v$        | 1.076                  | 1.0864             | 1.0201              | 0.944   | 0.971               | 0.946   |
| $u$        | $-7.80 \times 10^{-4}$ | 0.0524             | -0.0142             | -0.0538 | 0.00722             | -0.0849 |

<sup>a</sup> One experience rater assessed each set of peach and nectarine leaves twice using the direct estimation (DE) method and the rating scale (Scale). Each assessment was done without notes on the previous assessments.

<sup>b</sup> For the rating scale, mid-points were used to calculate the statistics.

**Table 2-3.** Intra-rater reliability, determined by kappa statistics and Lin's concordance analysis, for the untransformed<sup>a</sup> rating scale for 3 sets of peach and nectarine leaves with symptoms of bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*). Statistics include Cohen's Kappa (with associated 95% confidence interval), Z-statistic, Pearson Correlation, Spearman Correlation, Gamma, Kendall's tau-b, Stuart's tau-c, Somers' D C|R, Somers' D R|C, Lin's concordance coefficient ( $\rho_c$ ), and the bias coefficient ( $C_b$ )<sup>b</sup>

| Measures of concordance         | Leaf Set 1      | Leaf Set 2      | Leaf Set 3      |
|---------------------------------|-----------------|-----------------|-----------------|
| Cohen's Kappa (SE) <sup>c</sup> | 0.975 (0.0064)  | 0.974 (0.0057)  | 0.954 (0.0099)  |
| 95% Conf. Intervals             | 0.962-0.987     | 0.963-0.985     | 0.935-0.974     |
| Z statistic (P-value)           | 9.899 (<0.0001) | 9.886 (<0.0001) | 9.837 (<0.0001) |
| Pearson Correlation (SE)        | 0.975 (0.0061)  | 0.974 (0.0056)  | 0.960 (0.0082)  |
| Spearman Correlation (SE)       | 0.971 (0.0082)  | 0.974 (0.0060)  | 0.964 (0.0068)  |
| Gamma (SE)                      | 0.991 (0.0048)  | 0.993 (0.0042)  | 0.990 (0.0060)  |
| Kendall's tau-b (SE)            | 0.937 (0.0153)  | 0.935 (0.0122)  | 0.918 (0.0126)  |
| Stuart's tau-c (SE)             | 0.907 (0.0209)  | 0.913 (0.0155)  | 0.894 (0.0181)  |
| Somers' D C R (SE)              | 0.935 (0.0158)  | 0.931 (0.0127)  | 0.918 (0.0129)  |
| Somers' D R C (SE)              | 0.939 (0.0152)  | 0.939 (0.0139)  | 0.917 (0.0131)  |
| $\rho_c$                        | 0.975           | 0.974           | 0.955           |
| $C_b$                           | 0.999           | 1.000           | 0.994           |

<sup>a</sup> For these statistics, the mid-points were not used to calculate the statistics.

<sup>b</sup> One experienced rater assessed each of the 3 sets of peach and nectarine leaves twice using the rating scale. Each assessment was done without notes on the previous assessments.

<sup>c</sup> Standard error for associated statistic.

**Table 2-4.** Lin's concordance analysis statistics for comparisons between actual<sup>a</sup> bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) severity and estimated bacterial spot severity for three sets of peach and nectarine leaves. The concordance coefficient ( $\rho_c$ ) measures the accuracy of the estimated disease severity. The related statistics include the correlation coefficient ( $r$ ), the bias coefficient ( $C_b$ ), the coefficient  $v$ , and the coefficient  $u$ <sup>b</sup>

| Statistics | Set 1 ( $n = 103$ ) |        |                      |         | Set 2 ( $n = 103$ ) |       |         |         | Set 3 ( $n = 104$ ) |        |         |         |
|------------|---------------------|--------|----------------------|---------|---------------------|-------|---------|---------|---------------------|--------|---------|---------|
|            | DE-1                | DE-2   | Scale-1 <sup>c</sup> | Scale-2 | DE-1                | DE-2  | Scale-1 | Scale-2 | DE-1                | DE-2   | Scale-1 | Scale-2 |
| $r$        | 0.988               | 0.978  | 0.971                | 0.952   | 0.967               | 0.964 | 0.974   | 0.934   | 0.929               | 0.962  | 0.922   | 0.963   |
| $\rho_c$   | 0.982               | 0.962  | 0.902                | 0.852   | 0.962               | 0.957 | 0.921   | 0.899   | 0.908               | 0.945  | 0.816   | 0.883   |
| $C_b$      | 0.994               | 0.983  | 0.930                | 0.895   | 0.995               | 0.993 | 0.946   | 0.963   | 0.977               | 0.982  | 0.885   | 0.916   |
| $v$        | 0.899               | 0.835  | 0.691                | 0.636   | 1.0143              | 0.994 | 0.721   | 0.764   | 0.813               | 0.837  | 0.636   | 0.672   |
| $u$        | -0.285              | -0.029 | -0.118               | -0.166  | 0.102               | 0.117 | 0.0738  | -0.026  | -0.066              | -0.072 | -0.227  | -0.153  |

<sup>a</sup>Actual severity was obtained using APS Assess software.

<sup>b</sup>Each set was assessed twice by the direct estimation (DE) method and the rating scale (Scale) by one experienced rater. The first assessment is noted by 1 and the second assessment is noted by 2.

<sup>c</sup>For the scale, the midpoint of each class was used to calculate the statistics.

**Table 2-5.** Bootstrap<sup>a</sup> analysis of accuracy statistics comparing data obtained with the 1-7 ordinal rating scale and the direct estimation method

| Statistic <sup>c</sup>       | Continuous Data <sup>a</sup> |                  | Categorical Data <sup>b</sup> |                   |
|------------------------------|------------------------------|------------------|-------------------------------|-------------------|
|                              | $\rho_c$                     | $r$              | $\rho_c$                      | $r$               |
| Mean difference <sup>d</sup> | 0.0738                       | 0.0120           | 0.0257                        | 0.0192            |
| Standard Error               | 0.0172                       | 0.0106           | 0.00973                       | 0.00782           |
| 95% Conf. Int. <sup>e</sup>  | 0.0418 to 0.109              | -0.009 to 0.0323 | 0.00525 to 0.0428             | 0.00267 to 0.0330 |

<sup>a</sup> Data obtained from the scale that had been converted to percentage by the midpoint method.

<sup>b</sup> Data obtained from the non-transformed scale.

<sup>c</sup> Based on 2000 balanced bootstrap samples.

<sup>d</sup> Mean difference between statistics associated with the rating scale and the direct estimation method.

<sup>e</sup> 95% confidence interval. If the mean falls between the upper and lower limits, the difference is statistically significant at  $P = 0.05$  level. However, if the interval includes 0.0, or if the mean falls outside of the interval, the difference is not statistically significant.

**Table 2-6.** Concordance statistics for comparisons between actual<sup>a</sup> bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) severity and estimated bacterial spot severity for 3 sets of peach and nectarine leaves based on categorical scale data without transformation

| Measures of concordance         | Set 1                 |                 | Set 2           |                 | Set 3           |                 |
|---------------------------------|-----------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                 | Scale-1 <sup>bc</sup> | Scale-2         | Scale-1         | Scale-2         | Scale-1         | Scale-2         |
| Cohen's Kappa (SE) <sup>d</sup> | 0.943 (0.0094)        | 0.926 (0.0127)  | 0.936 (0.0142)  | 0.922 (0.0208)  | 0.916 (0.0160)  | 0.918 (0.0119)  |
| 95% Conf. Intervals             | 0.924-0.961           | 0.902-0.951     | 0.908-0.964     | 0.881-0.9624    | 0.884-0.9471    | 0.895-0.942     |
| Z statistic (P-value)           | 9.669 (<0.0001)       | 9.528 (<0.0001) | 9.728 (<0.0001) | 9.557 (<0.0001) | 9.596 (<0.0001) | 9.716 (<0.0001) |
| Pearson Correlation (SE)        | 0.953 (0.0075)        | 0.939 (0.0109)  | 0.954 (0.0105)  | 0.942 (0.0168)  | 0.937 (0.0147)  | 0.948 (0.0078)  |
| Spearman Correlation (SE)       | 0.937 (0.015)         | 0.921 (0.0176)  | 0.957 (0.007)   | 0.943 (0.0114)  | 0.9472 (0.0092) | 0.956 (0.0066)  |
| Gamma (SE)                      | 0.982 (0.0099)        | 0.966 (0.0139)  | 0.989 (0.0075)  | 0.971 (0.0142)  | 0.983 (0.0116)  | 0.992 (0.0050)  |
| Kendall's tau-b (SE)            | 0.879 (0.0177)        | 0.854 (0.0211)  | 0.902 (0.0117)  | 0.879 (0.0172)  | 0.887 (0.0133)  | 0.899 (0.0094)  |
| Stuart's tau-c (SE)             | 0.841 (0.0267)        | 0.815 (0.0288)  | 0.871 (0.0154)  | 0.846 (0.0184)  | 0.857 (0.0186)  | 0.869 (0.0151)  |
| Somers' D C R (SE)              | 0.892 (0.0170)        | 0.865 (0.0209)  | 0.914 (0.0137)  | 0.888 (0.0183)  | 0.892 (0.0164)  | 0.905 (0.0127)  |
| Somers' D R C (SE)              | 0.867 (0.0202)        | 0.844 (0.0233)  | 0.890 (0.0123)  | 0.870 (0.0178)  | 0.881 (0.0129)  | 0.892 (0.0098)  |
| $\rho_c$                        | 0.943                 | 0.926           | 0.936           | 0.922           | 0.916           | 0.919           |
| $C_b$                           | 0.989                 | 0.987           | 0.981           | 0.979           | 0.978           | 0.969           |

<sup>a</sup> Actual severity was obtained using APS Assess software and was then converted to the 1 to 7 rating scale with the exception that the values of the APS Assess less than or equal to 0.25% were considered to represent leaves with no disease and assigned a score of 1 on the scale.

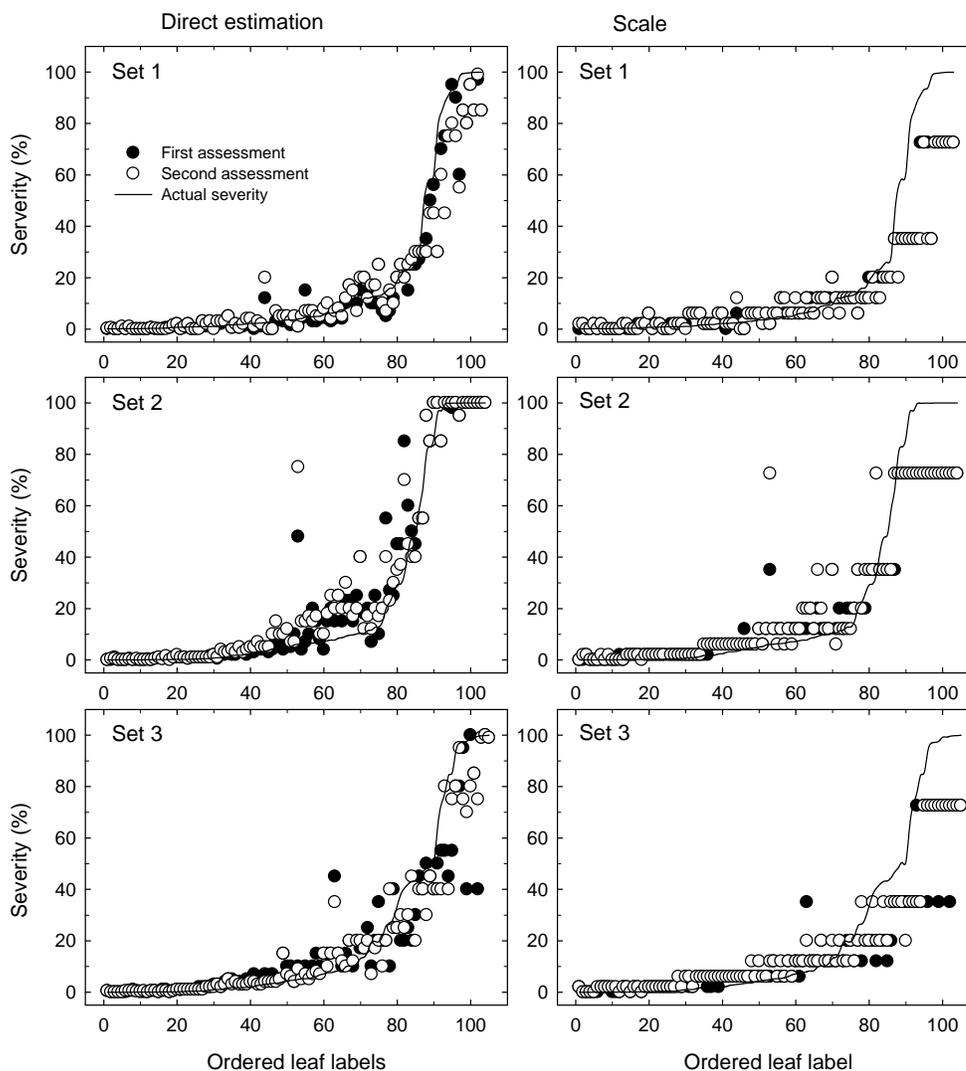
<sup>b</sup> Each set was assessed twice by the rating scale (Scale) by one experienced rater. The first assessment is noted by 1 and the second assessment is noted by 2.

<sup>c</sup> For these statistics, the scale mid-points were not used to calculate the statistics.

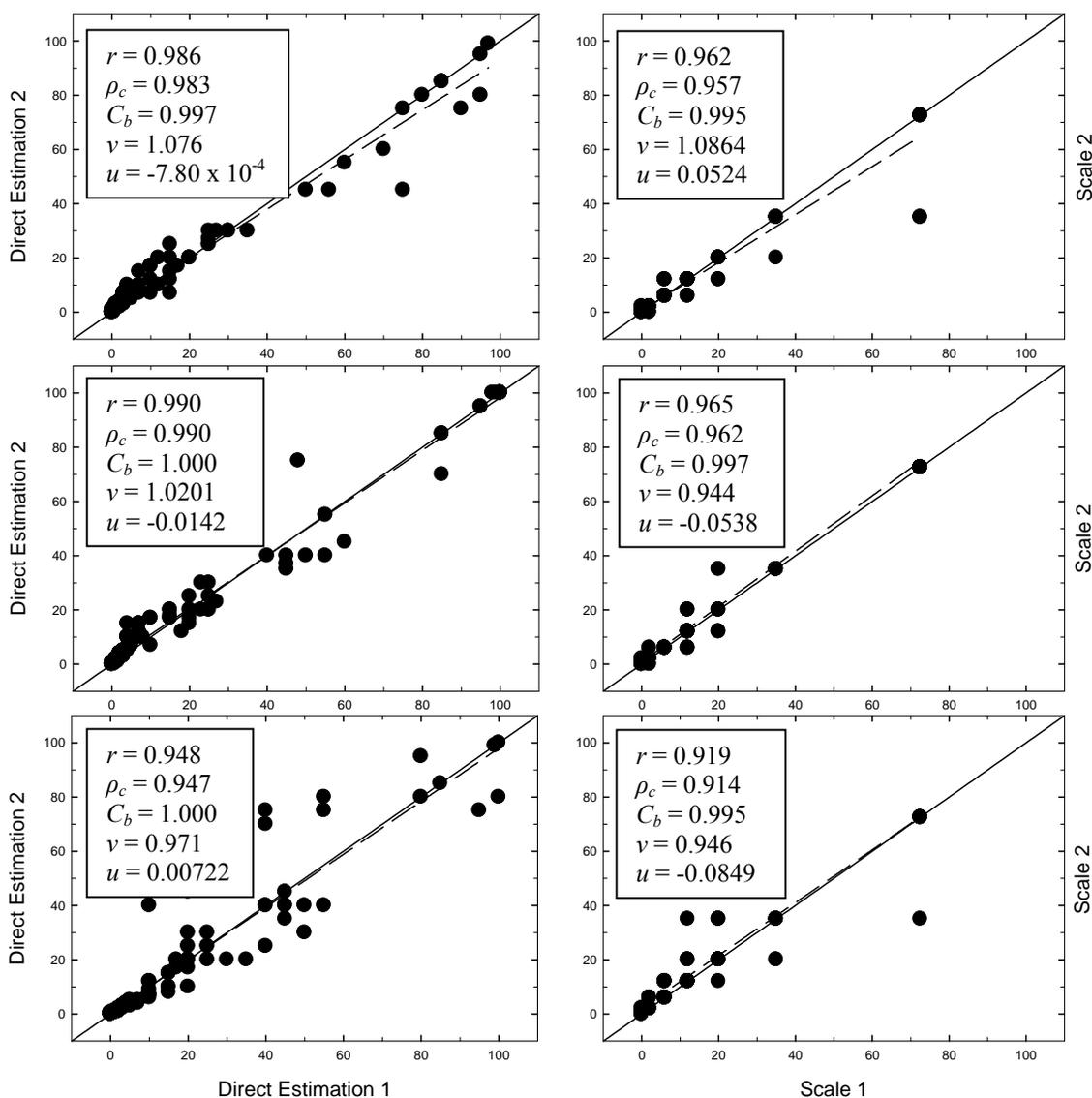
<sup>d</sup> Standard error for associated statistic.



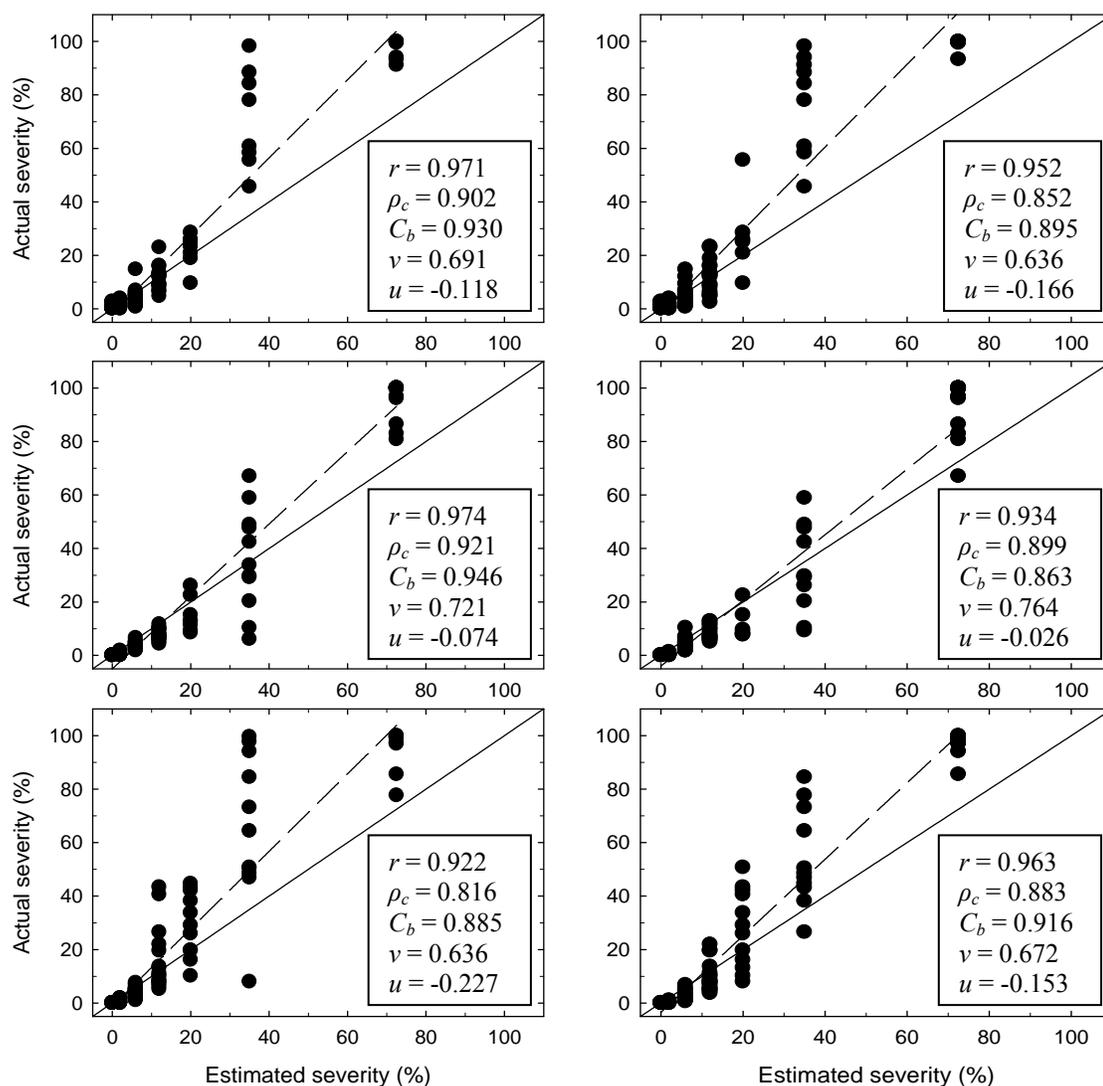
**Fig 2-1.** Leaves typical of the 1-7 rating scale used to assess foliar bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) severity. The 7 leaves represent the rating scale and symptomatic area of 1 = 0%, 2 = 1-3%, 3 = 4-8%, 4 = 9-15%, 5 = 15-25%, 6 = 25-45%, 7 = >45%, respectively, as estimated by the experienced rater.



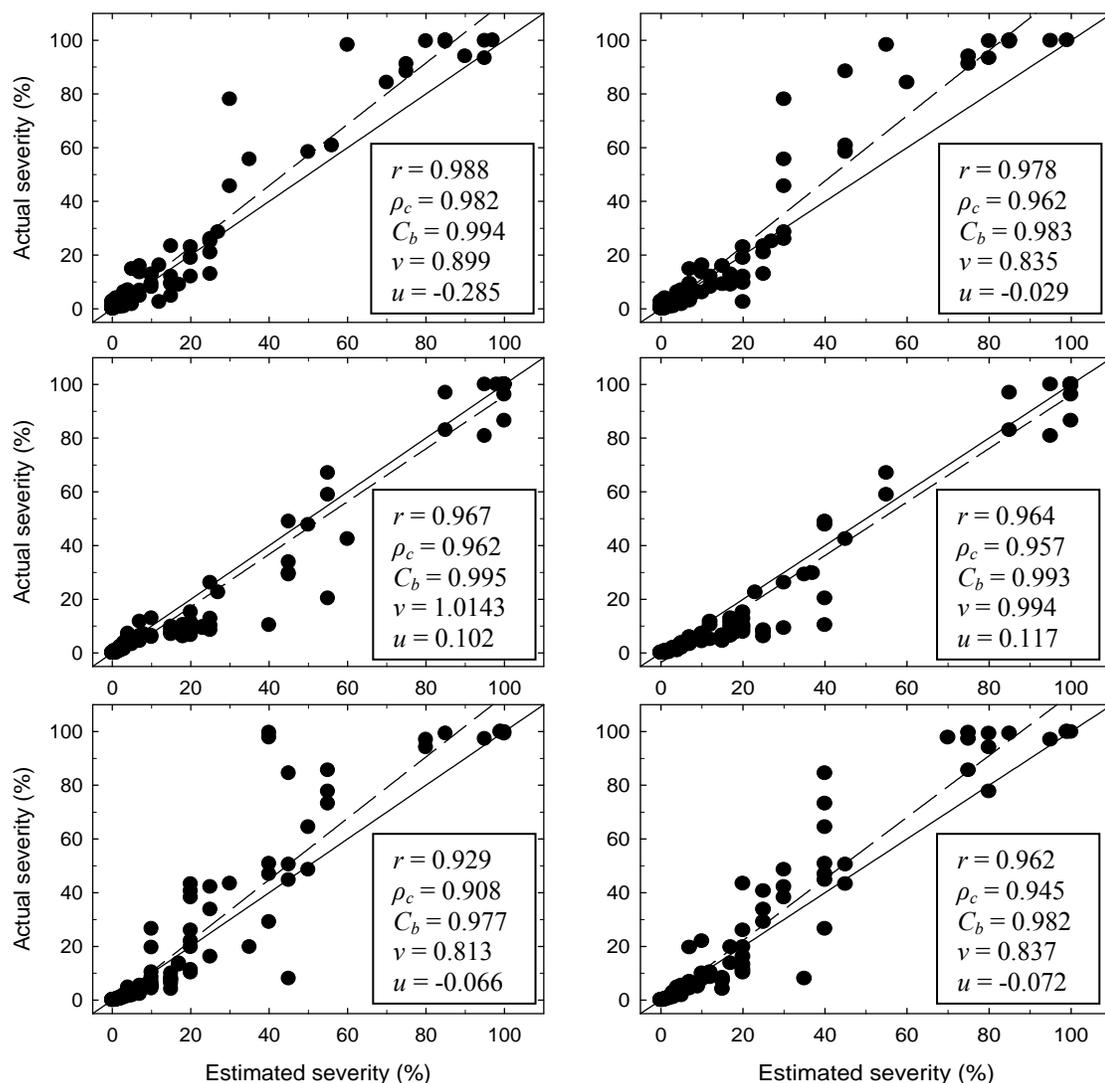
**Fig 2-2.** Estimated and actual disease severity values for bacterial spot of peach and nectarine for three sets of leaves. Each point represents the estimated severity of leaves rated by the same rater at two different times while the line represents actual disease severity for the same leaf based on computer image analysis. The midpoints were used for severity estimations using the rating scale. This limited the possible severity degrees so the points appear to form horizontal lines. In general, there was good agreement between visual direct estimates and actual disease severity.



**Fig 2-3.** A comparison of the first (X) and second (Y) assessments using the direct estimation method (left) and the rating scale (right) for 3 sets of leaves. The line of concordance, representing a theoretical perfect agreement between actual and estimated severity, is indicated by the solid line and the best-fit line, the actual relationship between the estimated severity and actual severity, is represented by the broken line. The correlation coefficient,  $r$ , is a measure of precision and Lin's concordance coefficient,  $\rho_c$ , measures accuracy by combining the effects of precision and bias,  $C_b$ , the difference between the best-fit line and the concordance line. Variance,  $v$ , measures the relationship between the actual and estimated severity where the variance of 1 represents a perfect relationship. The coefficient  $u$  defines the location of the best-fit line relative to the concordance line.



**Fig. 2-4.** Lin's concordance analysis for the three sets of leaves assessed using the rating scale. Each set of leaves was rated twice by an experienced rater. The first assessment is on the left and the second assessment is on the right. The concordance line, representing a theoretical perfect agreement between actual and estimated severity, is indicated by the solid line and the best-fit line, the actual relationship between the estimated severity and actual severity, is represented by the broken line. The correlation coefficient,  $r$ , is a measure of precision and Lin's concordance coefficient,  $\rho_c$ , measures accuracy by combining the effects of precision and bias,  $C_b$ , the difference between the best-fit line and the concordance line. Variance,  $v$ , measures the relationship between the actual and estimated severity where the variance of 1 represents a perfect relationship. The coefficient  $u$  defines the location of the best-fit line relative to the concordance line.



**Fig. 2-5.** Lin's concordance analysis for the three sets of leaves assessed using the direct estimation method. Each set of leaves was rated twice by an experienced rater. The first assessment is on the left and the second assessment is on the right. The concordance line, representing a theoretical perfect agreement between actual and estimated severity, is indicated by the solid line and the best-fit line, the actual relationship between the estimated severity and actual severity, is represented by the broken line. The correlation coefficient,  $r$ , is a measure of precision and Lin's concordance coefficient,  $\rho_c$ , measures accuracy by combining the effects of precision and bias,  $C_b$ , the difference between the best-fit line and the concordance line. Variance,  $v$ , measures the relationship between the actual and estimated severity where the variance of 1 represents a perfect relationship. The coefficient  $u$  defines the location of the best-fit line relative to the concordance line.

### Chapter 3

## Accuracy and Inter-rater Reliability of Assessments of Bacterial Spot Severity Made by Experienced and Inexperienced Raters

### ABSTRACT

The objective this study was to compare the reliability and accuracy of visual assessments of foliar bacterial spot severity made by experienced and inexperienced raters. Because in chapter 2 it was determined that the 1-7 ordinal rating scale (1 = 0% lesion area and 7 = >45%) did not provide estimates of bacterial spot severity as accurate or reliable as those obtained with the direct estimation method, only the direct estimation method of disease assessment was investigated in this chapter. Three sets of peach or nectarine leaves ( $n = 103, 103, \text{ and } 104$  leaves; disease severity levels from 0 to 100%) were assessed twice by 10 raters (4 considered experienced and 6 considered inexperienced) using direct visual estimation (percent leaf area covered by necrotic and chlorotic lesions). The same leaves were also assessed with the APS Assess image analysis software. Based on Lin's concordance analysis and the correlation coefficient ( $r$ ) as well as the results of a  $t$ -test and a Kolmogorov-Smirnov test, there was no significant difference between the accuracy and reliability of assessments made by experienced and inexperienced raters. The mean Lin's concordance coefficient ( $\rho_c$ ) value for assessments made by experienced raters was 0.941 compared with the  $\rho_c$  value of 0.925 obtained by the inexperienced raters. The mean correlation coefficient ( $r$ ), assessing reliability, obtained by the experienced raters was 0.953 compared with the  $r$

value of 0.942 obtained by inexperienced raters. Levels of inter-rater reliability were also high. The mean pairwise correlation coefficient ( $r$ ) ranged from 0.929 for the 1<sup>st</sup> repetition of the 3<sup>rd</sup> set of leaves to 0.970 for the 2<sup>nd</sup> repetition of the 1<sup>st</sup> set of leaves. The repeatability coefficient (RC) ranged from 14.521 for the 2<sup>nd</sup> repetition of leaf set 1 to 23.641 for the 1<sup>st</sup> repetition of leaf set 3 with corresponding  $\chi^2$  values of 7742.148 ( $P = 0.000$ ) and 20924.394 ( $P = 0.000$ ), respectively. Therefore, disease assessment of bacterial spot of peach and nectarine may be made by multiple raters with varying levels of experience without loss of accuracy and reliability.

## INTRODUCTION

In the previous chapter, it was determined that estimates of foliar bacterial spot severity based on the direct estimation method were ultimately superior to those obtained with the 1-7 ordinal rating scale (1 = 0% lesion area, and 7 = >45%). Based on Lin's concordance analysis of continuous data,  $\rho_c$  values for data obtained with the rating scale were 0.41 to 0.11 lower than data obtained with the direct estimation method when assessing accuracy and were 0.026 to 0.033 lower when assessing reliability. Although levels of accuracy and reliability were improved when determined by the concordance analysis for categorical data where rating scale estimates were not transformed to midpoint percentages, results of the bootstrap analysis comparing  $\rho_c$  and  $r$  values for the two methods indicated that direct estimation resulted in significantly ( $P < 0.05$ ) more accurate estimates than the ordinal rating scale. These results collaborated those of several other studies working with different diseases (Nita et al., 2003; Bock et al.,

2009b; Bock et al., 2010a; Nutter et al., 2003). Based on this outcome, no further evaluations of the estimation methods based on the scale were attempted. Instead, only the direct estimation method was further investigated.

As stated in the previous chapter, epidemiological studies on bacterial spot depend on the quality of disease assessment. In many research settings, the task of collecting field data on bacterial spot severity may be split among multiple raters with different ability and experience. Examples of such personnel may include regular experienced personnel such as the research scientists, or persons with limited experience such as temporary student workers on summer break. Numerous phytopathological studies have documented large differences among raters evaluating the same set of diseased samples (Bock et al., 2010a; Bock et al., 2010b). In many pathosystems, reduced levels of accuracy and reliability have been associated with many factors including (i) rater intrinsic ability (Nita et al., 2003; Bock et al., 2009b; Nutter et al., 1993), (ii) time interval between disease assessments (Parker et al., 1995b), (iii) complexity of symptoms being assessed (Bock et al., 2008; Bock et al., 2009b), and (iv) interaction among multiple factors (Bock et al., 2010b). For instance, foliar bacterial spot symptoms are difficult to identify and may be confused with foliar injuries caused by copper and Captan spray applications (Agnello et al., 2006; Lalancette et al., 2007).

A common limitation in previous phytopathological studies that have evaluated the effect of rater ability or experience on disease assessment is that most have used descriptive narratives about rater ability without providing specific tests. Comparisons of quality of disease assessment deal with the concepts of *consistency* and *conformity*, which in turn require equivalence hypothesis tests, as opposed to subjective judgment (Yi

et al., 2008). With varying levels of experience and training, the students' ability to provide accurate and reliable estimates of disease using the direct estimation method is critical (Bock et al., 2009b). Knowledge of inter-rater reliability and accuracy when assessing bacterial spot symptoms is required to assess the quality of the data obtained from multiple raters. The objective of this study therefore was to assess the intra-rater and inter-rater reliability as well as the accuracy of estimates of bacterial spot severity obtained from multiple raters and to compare the quality of assessments made by experienced and inexperienced raters.

## MATERIALS AND METHODS

*Visual Assessment:* The details of leaf collection and image analysis used for this assessment are outlined in Chapter 2. Before the visual assessment took place, inexperienced raters were given the opportunity to practice assessing disease severity with the computer-based program DISTRAIN (Tomerlin and Howell, 1988). Each rater received 15 minutes to practice disease severity assessment with the computer program. They also received a detailed explanation of foliar bacterial spot symptoms from an experienced rater. Displayed on a computer, 10 raters, 4 considered experienced and 6 considered inexperienced, assessed each set of leaves twice using direct visual estimation without the aid of another rater.

*Data Analysis:* Intra-rater reliability was assessed based on Pearson's moments correlation coefficient ( $r$ ) which quantified the variability around the straight line relationship between severity estimations from the first and second assessments of all 10

raters (Nita et al., 2003). The correlation coefficient was a measurement of precision. In addition, 95% confidence intervals were calculated. Inter-rater reliability was determined in 3 ways. First, it was assessed based on the mean pairwise correlation coefficient, ( $r$ ). Using the proc CORR procedure from SAS 9.1, rater assessments of each leaf set were compared to each other and the resulting pairwise correlation coefficients were averaged. Second, the intra-class correlation coefficient ( $\rho$ ) was calculated for all 10 raters combined based on the variance components formulated from a two way random effects ANOVA model (Nita et al., 2003). Variance components were calculated using the proc GLM produce in SAS 9.1 and values of  $\rho$  were determined as follows:

$$\rho = \sigma_{\text{leaf}}^2 / (\sigma_{\text{leaf}}^2 + \sigma_{\text{rater}}^2 + \sigma_{\text{error}}^2)$$

where  $\sigma_{\text{leaf}}^2$ ,  $\sigma_{\text{rater}}^2$ , and  $\sigma_{\text{error}}^2$  are the variances for leaf, rater, and error, respectively. These values were calculated for each repetition of each leaf set for the combined 10 raters. Third, a repeatability coefficient was calculated based on the within subject variance (calculated from a two way random effects ANOVA model), and was estimated as  $1.96\sqrt{2\hat{\sigma}_W}$  (Bland and Altman, 1986 & 1999). The allowed difference limit was 5 so the corresponding within-subject variance ( $\sigma_U^2$ ) was equal to  $5^2 / (1.96 \times 1.96 \times 2) = 3.25$ . The chi-square value was found by dividing the sum of the squared deviation (SSW) by the within-subject variance ( $\sigma_U^2 = 3.25$ ). The  $p$ -value was then calculated from the chi-square value and degrees of freedom.

Accuracy was determined by Lin's concordance correlation coefficient ( $\rho_c$ ), a type of intra-class correlation (Lin, 1989) that measures the extent to which two sets of observations, the 1<sup>st</sup> and 2<sup>nd</sup> repetitions, align on the line of unity (i.e.,  $y = x$ ), also known

as the line of concordance. Details of the calculations of Lin's concordance coefficients are as indicated in Chapter 1.

In order to determine if experienced raters were more accurate and reliable than inexperienced raters, a bootstrap analysis of accuracy ( $\rho_c$ ,  $C_b$ , and  $r$ ) and reliability ( $r$ ) statistics comparing data obtained by experienced and inexperienced raters was performed in order to create a large number of "re-samples" from the observed data to permit the calculation of the  $t$ -test statistic (assuming parametric data) and the Kolmogorov-Smirnov (K-S) statistic (assuming nonparametric data) which otherwise would not have been possible with so few values. A total of 2000 random bootstrap samples were created for the analysis in GenStat 12.

## RESULTS

*General:* Each of the three sets of leaves was rated twice by the ten raters using direct visual estimation. Raters A through D were considered experienced while raters E through J were considered inexperienced. The average estimated bacterial spot severity varied from 15.045 to 24.398 for leaf set 1, from 25.769 to 36.625 for leaf set 2, and from 19.300 to 31.981 for leaf set 3 across both repetitions (Table 3-1). The median disease severity ranged from 3.0 to 15.0 for leaf set 1, from 5.0 to 25.0 for leaf set 2, and from 3.0 to 20.0 for leaf set 3.

*Intra-rater reliability:* Intra-rater reliability was determined using the correlation coefficient ( $r$ ). Values of  $r$  were greater than 0.90 for all 3 leaf sets and all 10 raters (Table 3-2). For experienced raters,  $r$  values ranged from 0.973 to 0.998 for leaf set 1,

from 0.941 to 0.998 for leaf set 2, and from 0.948 to 0.999 for leaf set 3. For inexperienced raters,  $r$  values ranged from 0.958 to 0.991 for leaf set 1, from 0.965 to 0.994 for leaf set 2, and from 0.935 to 0.982 for leaf set 3. Although the highest  $r$  value from each leaf set was produced by an experienced rater,  $r$  values from inexperienced raters were not consistently lower than those of experienced raters.

*Inter-rater reliability:* Inter-rater reliability was determined based on the mean pairwise correlation coefficient ( $r$ ), the intra-class correlation coefficient ( $\rho$ ), and the repeatability coefficient. Mean values of  $r$  were generally high and ranged from 0.929 for the 1<sup>st</sup> repetition of leaf set 3 to 0.970 for the 2<sup>nd</sup> repetition of leaf set 1 (Table 3-3). Values of  $\rho$ , also high, ranged from 0.900 for the 1<sup>st</sup> repetition of leaf set 3 to 0.961 for the 2<sup>nd</sup> repetition of leaf set 1. Repeatability coefficient ranged from 14.521 for the 2<sup>nd</sup> repetition of leaf set 1 to 23.641 for the 1<sup>st</sup> repetition of leaf set 3 with corresponding  $\chi^2$  values of 7742.148 ( $P = 0.000$ ) and 20924.394 ( $P = 0.000$ ), respectively.

*Rater accuracy:* Accuracy was determined for the 1<sup>st</sup> and 2<sup>nd</sup> assessments of all 3 leaf sets made by direct visual estimation by all 10 raters using Lin's concordance analysis. The mean values of the two assessments were generally high with 83% and 72% of  $\rho_c$  values greater than 0.90 for experienced and inexperienced raters, respectively (Table 3-4). For experienced raters, values of  $\rho_c$  ranged from 0.964 to 0.980 for leaf set 1, from 0.858 to 0.960 for leaf set 2, and from 0.922 to 0.955 for leaf set 3. For inexperienced raters,  $\rho_c$  values ranged from 0.940 to 0.987 for leaf set 1, from 0.802 to 0.924 for leaf set 2, and from 0.880 to 0.944 for leaf set 3. Values of  $\rho_c$  produced by inexperienced raters were not consistently lower than those of experienced raters.

The results of accuracy for the 1<sup>st</sup> assessment of leaf set 1 for both experienced and inexperienced raters are shown in Figures 1 and 2, respectively. For all assessments, there was a linear relationship between actual and estimated severity. The bias coefficient ( $C_b$ ) ranged from 0.989 to 0.998 for leaf set 1, from 0.968 to 0.994 for leaf set 2, and from 0.974 to 0.996 for leaf set 3 for experienced raters (Table 3-4). Values of  $C_b$  for the inexperienced rater ranged from 0.971 to 0.998 for leaf set 1, from 0.925 to 0.986 for leaf set 2, and from 0.951 to 1.000 for leaf set 3. The correlation coefficient ( $r$ ) ranged from 0.966 to 0.984 for leaf set 1, from 0.881 to 0.966 for leaf set 2, and from 0.946 to 0.960 for leaf set 3 for experienced raters. For inexperienced raters, values of  $r$  ranged from 0.967 to 0.988 for leaf set 1, from 0.866 to 0.939 for leaf set 2, and from 0.924 to 0.947 for leaf set 3.

*Effects of rater experience:* Mean values of accuracy ( $\rho_c$ ,  $C_b$ , and  $r$ ) and intra-rater reliability ( $r$ ) were compared between experienced and inexperienced raters using the  $t$ -test and the Kolmogorov-Smirnov (K-S) test. The mean values of  $\rho_c$ ,  $C_b$ , and  $r$  were 0.941, 0.953, and 0.987, respectively, for experienced raters and were 0.925, 0.942, and 0.981, respectively, for inexperienced raters (Table 3-5). Similarly, the mean value of  $r$ , measuring intra-rater reliability, was 0.977 and 0.975 for experienced and inexperienced raters, respectively. Despite higher average accuracy and reliability values produced by experienced raters, there was no significant difference between rater experience for any of the mean accuracy and reliability values (Table 5). The  $t$  statistics for  $\rho_c$ ,  $C_b$ , and  $r$  were 1.480 ( $P = 0.143$ ), 1.350 ( $P = 0.183$ ), 1.670 ( $P = 0.101$ ), respectively. The K-S statistics for  $\rho_c$ ,  $C_b$ , and  $r$  were 5.880 ( $P = 0.053$ ), 5.880 ( $P = 0.053$ ), and 2.180 ( $P =$

0.337), respectively. The  $t$  and K-S statistics for  $r$ , measuring intra-rater reliability, were 0.280 ( $P = 0.782$ ) and 1.800 ( $P = 0.407$ ), respectively.

## DISCUSSION

In numerous studies, rater experience, or familiarity and prior knowledge of disease assessment, often affected the reliability and accuracy of disease assessments (Nita et al., 2003; Bock et al., 2009b; Nutter et al., 1993). In general, estimates made by experienced raters were more accurate and reliable than those made by inexperienced raters (Newton and Hackett, 1994; Bock et al., 2009b; Nita et al., 2003). For example, Bock et al. (2009b) found that experienced plant pathologists were more precise than people with little or no prior experience in disease severity assessment when assessing percent necrotic area (mean  $r^2 = 0.61$  for experienced raters and mean  $r^2 = 0.45$  for inexperienced raters), and percent chlorotic and necrotic area (mean  $r^2 = 0.73$  for experienced raters and mean  $r^2 = 0.58$  for inexperienced raters). Newton and Hackett (1994) also found that experienced raters were better at assessing symptoms of powdery mildew on barley than were inexperienced raters. In this study, however, there was no difference in the accuracy and reliability of disease severity estimates made by experienced and inexperienced raters (Table 3-5). Although numerically higher values of coefficients measuring accuracy and reliability more often were obtained from assessments made by experienced raters (Tables 3-2 and 3-4), parametric and non-parametric statistical tests did not support a hypothesis for differences among the two groups. In the previous studies comparing the reliability and accuracy of disease

assessments from multiple raters, various statistical tests have been used. Analysis of variance (ANOVA), assuming normally distributed data, has been used to determine sources of variation in disease estimates. Used with an *F*-test to test for statistical significance, several studies have identified factors that contribute to error and variation or otherwise influence disease assessment. For example, Sherwood et al. (1983) used ANOVA to determine that, among other things, individual rater and rater experienced significantly influenced assessment of *Stagonospora* leaf spot on orchardgrass. ANOVA has also been used to determine inter-rater reliability by calculating the intra-class correlation coefficient ( $\rho$ ) (Nita et al., 2003; Madden et al., 2007). Used in this study, values of  $\rho$  were generally high and ranged from 0.900 to 0.961 (Table 3-3) compared to similar values found by Nita et al. (2003) which ranged from 0.800 to 0.960. The correlation coefficient (*r*) was used to measure intra and inter-rater reliability in this study as in other studies (Nita et al., 2003; Bock et al., 2008; Bock et al., 2009b). Measuring only the association between two estimates, it is considered a less powerful test of reliability than the intra-class correlation coefficient ( $\rho$ ) (Lin, 1989; Madden et al., 2007, Bock et al., 2010b). Nevertheless, values of *r* measuring inter-rater reliability were generally high and ranged from 0.929 to 0.970 (Table 3-3) compared to similar values found by Nita et al. (2003) which ranged from 0.800 to 0.970. Values of *r* measuring intra-rater reliability were also high and included over-lapping 95% confidence intervals among experienced and inexperienced raters, indicating values were not significantly different from each other.

A common feature in previous phytopathological studies comparing reliability and accuracy of disease assessments from multiple raters is that most did not provide

equivalence tests, a key requirement for agreement studies (Yi. et al., 2008). Instead, comparisons were based on the magnitude of correlation coefficients ( $r$  and  $\rho$ ) as well as  $r^2$  and GLM analysis (Nita et al., 2003; Bock et al., 2008; Bock et al., 2009b). Without explicit equivalence tests, such descriptive comparisons were subjective; that is, there were generally no statistical tests to evaluate the hypothesis of agreement in the estimate coefficients obtained from different groups. In this study, two approaches were developed to allow computation of statistical equivalency tests, the repeatability coefficient and bootstrap confidence intervals. Most often used in clinical studies, an equivalence test based on the repeatability coefficient (RC) (based on the within subject variance) was calculated as a measurement of inter-rater reliability (Bland and Altman, 1986 & 1999; Yi et al., 2008). A  $\chi^2$  test was used to test for statistical significance. Significant  $P$ -values ( $P < 0.0001$ ; Table 3-5) indicated that the repeated estimates had a small amount of variation (no larger than the variation limit that was set to 5% in this study), implying good agreement and in this case, significantly high levels of inter-rater reliability (Yi et al., 2008). The value of 5% used here while arbitrary is practical because most hypothesis tests are based on a 5% significance level. Results indicated that on average, the reliability of disease severity estimates obtained by experienced and inexperienced raters did not differ by more than 5% indicating that hypothesis tests based on data from the two groups would yield the same outcome.

The second equivalence test was based on a bootstrap analysis of the means of the various agreement statistics. In equivalence testing for agreement studies, the goal is to provide evidence that assessments obtained from the two types of raters are statistically equivalent. The null hypothesis, therefore, was that there was a difference in agreement

statistics ( $r$ ,  $\rho$ ,  $C_b$ , etc.) obtained with data from the two groups of raters, that was outside the zone of scientific indifference (i.e. statistically significant). This zone is best demonstrated with the use of confidence intervals. The use of bootstrap analysis allowed us to compute 95% confidence intervals around the mean differences of the agreement statistics for the experienced and inexperienced raters (Table 3-5). Based on 2000 random bootstrap samples, in this study, results indicated that for all the agreement statistics evaluated, the 95% confidence intervals for the differences in values obtained from the two groups of raters included zero, indicating that there was sufficient evidence to reject the null hypothesis (Table 3-5).

The results of the bootstrap analysis were also in agreement with those of a  $t$ -test and the non-parametric K-S test which also indicated non-significant differences. Non-significant  $P$ -values inferred that there was no difference between values of accuracy ( $\rho_c$ ,  $C_b$ , and  $r$ ) and reliability ( $r$ ) obtained by experienced and inexperienced raters (Table 3-5). In this case, the bootstrap analysis was more conclusive because the  $P$ -values obtained from the parametric  $t$ -test especially for accuracy comparisons based on  $r$  and  $\rho$  were borderline ( $P = 0.053$ ; Table 3-5). In the absence of the confidence intervals (only possible through the bootstrap analysis), comparisons based on the  $t$ -test in this case would have concluded the evidence for the difference in statistics obtained with the two rater groups was weak. Altogether, the results of the equivalence test as well as the  $t$ -test and K-S test indicated that not only were the estimates made by experienced and inexperienced raters not significantly different from each other, they were also significantly similar to each other.

The lack of differences between experienced and inexperienced raters in accuracy and reliability of disease severity estimates observed in this study was intriguing given the vast evidence from previous studies that rater experience is an important factor in quality and consistency of disease assessments (Sherwood et al., 1983; Newton and Hackett et al., 1994; Parker et al., 1995b Nita et al., 2003; Bock et al., 2009b). A potential explanation for this observation is that the inexperienced raters in this study were given training prior to commencing disease assessments. The inexperienced raters received brief training on visual disease assessment using the computer program DISTRAIN (Tomerlin and Howell, 1988), as well as a detailed explanation of bacterial spot symptoms. This training, quickly followed by the assessment, may explain why values of accuracy and reliability did not significantly differ between experienced and inexperienced raters (Table 3-5). Various training methods have been created and successfully used to improve the accuracy and reliability of assessments made by inexperienced raters (Nutter and Schultz, 1995; Amanat, 1976; Tomerlin and Howell, 1988; Newton and Hackett, 1994; Parker et al., 1995b; Nutter and Worawitlikit, 1989; Nutter and Schultz, 1995; Nutter and Gaunt, 1996; Nutter and Litwiller, 1993). Field training, perhaps the most obvious way to improve disease assessment is also the most time consuming (Nutter and Schultz, 1995). Trainees may gain experience with the actual disease and its symptoms throughout the season. Indeed, Amanat (1976) found that precision of several raters was improved by repeated training sessions. Improved inter-rater reliability also resulted from training in assessment of late leaf spot of peanut (Shokes et al., 1987). More recently, computer based training has been used to improve disease assessment and include but are not limited to such programs as DISTRAIN

(Tomerlin and Howell, 1988), DISEASE.PRO (Nutter and Worawitlikit, 1989; Nutter and Schultz, 1995; Nutter and Gaunt, 1996), ALFALFA.PRO (Nutter and Litwiller, 1993), and SEVERITY.PRO (Nutter and Litwiller, 1998). After using DISTRAIN, Newtown and Hackett (1994) found that the accuracy and reliability of assessments of powdery mildew on barley made by inexperienced raters improved but warned it was not a replacement for actual field experience. Parker et al. (1995b) also found the effects of DISTRAIN to be beneficial but short-lived. Our data therefore suggest that when inexperienced raters must be used to evaluate bacterial spot, training and instruction prior to commencing disease assessments will improve the quality of the data obtained.

With or without experience, raters commonly vary in their ability to assess disease severity (Sherwood et al., 1983; Newton and Hackett, 1994; Nita et al., 2003; Bock et al., 2008). Nutter et al. (1993) found that raters varied between and among each other in respect to inter- and intra-rater reliability of assessments of dollar spot severity on spent grass. Bock et al. (2008) also found that the reliability of estimates of citrus canker symptoms on grapefruit leaves varied among raters. Besides rater ability and experience, variability may be due to several factors including but not limited to value preference and symptom type. For instance, Bock et al. (2008) found that some raters chose certain values more often than would be expected, creating “knots”. These “knots”, are ultimately a source of error and variability as different numbers may be preferred by certain raters. Moreover, the multiple symptoms, including necrotic lesions, shot holes, and chlorosis, associated with bacterial spot may be additional sources of variability. Chlorosis and gradations of light green and dark yellow are often difficult to quantify. Bock et al. (2008) found that raters were more accurately able to assess

necrotic lesions and chlorotic area associated with foliar citrus canker symptoms than necrotic lesions alone. Parker et al. (1995a) found that raters often confused symptoms of *Septoria tritici* on wheat with natural senescence of the plant. Together, factors like these may have contributed to the variability of assessments found in this study (Table 3-3).

Foliar symptoms of bacterial spot are difficult to identify (Travis, J.W. *personal communication*). Often, such symptoms are confused with foliar injury caused by Captan and copper spray applications (Agnello et al., 2006; Lalancette et al., 2007). Both commonly used chemicals may cause shot-hole injury as well as yellowing. The leaves used in this study were only affected by bacterial spot and were obtained from non-treated plots that were not treated with Captan since it was one of the experimental treatments. Moreover, because all leaves were obtained and prepared for assessment by one person, the comparisons of accuracy and reliability made in this study assess only the rater's ability to estimate bacterial spot severity. Effects of symptom expression and confounding effects of treatments on disease expression which can affect rater performance in field evaluations of bacterial spot severity were not examined. In the field where most assessments are made, it is not uncommon for leaves to have symptoms of bacterial spot as well as chemical or mechanical injuries. It is not known whether or not raters, experienced or not, would be able to discern symptoms of bacterial spot from these common injuries. If, in fact, such raters cannot, accuracy of bacterial spot assessments would be greatly compromised. Based off of previous studies (Parker et al., 1995a), however, it may be assumed that experienced raters can better distinguish bacterial spot symptoms from other injuries and that field experience may be the best method of training inexperienced raters to discriminate disease symptoms from these

common injuries (Nutter and Schultz, 1995). Future studies are needed to verify these assumptions.

Nevertheless, the results of this study indicate that estimates of bacterial spot severity on peach and nectarine leaves may be made by multiple raters with varying levels of experience without loss of accuracy and reliability. The results also imply that little training is needed before disease assessment is made by inexperienced raters. Nevertheless, it is recommended that increased periods of computer-based training as well as field training should be implemented before inexperienced raters are tasked with assessing disease symptoms for the purpose of epidemiological studies (Nutter and Schultz, 1995; Newtown and Hackett, 1994). Ways to further improve assessment of bacterial spot should also be investigated.

## LITERATURE CITED

- Agnello, A., Chouinard, G., Firlej, A., Turechek, W., Vanoosthuyse, F., Vincent, C. 2006. Bacterial Blossom Blast of Pear. Page 161 in: Tree Fruit Field Guide to Insect, Mite, and Disease Pests and Natural Enemies of Eastern North America. Natural Resource, Agriculture, and Engineering Service (NRAES), Ithaca, NY.
- Amanat, P. 1976. Stimuli effecting disease assessment. *Agric. Conspectus Scientificus* 39: 27-31.
- 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.
- Bock, C.H., Parker, P.E., Cook, A.Z., and Gottwald, T.R. 2008. Characteristics of the perception of different severity measures of citrus canker and the relationships between the various symptom types. *Plant Dis.* 92:927-939.
- Bock, C.H., Gottwald, T.R., Parker, P.E., Cook, A.Z., Ferrandino, F., Parnell, S., and van den Bosch, F. 2009a. The Horsfall-Barratt scale and severity estimates of citrus canker. *Eur. J. Plant Pathol.* 125:23-38.
- 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., Ferrandino, F., Welham, S., van den Bosch, F., and Parnell, S. 2010a. Some consequences of using the Horsfall-Barratt scale for hypothesis testing. *Phytopathology* 100: in press.
- Bock, C.H., Poole, G.H., Parker, P.E. and Gottwald, T.R. 2010b. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical Reviews in Plant Sciences*, 29:59-107.
- Dos Santos, W.R., Alves, M.C., de Carvalho, L.G., Rios, F.A. and Fraga, A.C.J. 2009. Elaboration and validation of a diagrammatic scale for epidemiological studies of *Cercospora* leaf spot of castor bean. *J. Phytopathology* 10: 1-6.
- Lalancette, N. and McFarland, K.A. 2007. Phytotoxicity of copper-based bactericides to peach and nectarine. *Plant Disease* 91:1122-1130.
- Lin, L.I.-K. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 45:255-268.
- Madden, L.V., Hughes, G., van den Bosch, F. 2007. *The Study of Plant Disease Epidemics*. The American Phytopathological Society, St. Paul.
- Newton, A.C. and Hackett, C.A. 1994. Subjective components of mildew assessment on spring barley. *Eur. J. Plant Pathol.* 100:395-412.
- 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.
- Nutter, F.W., Jr. and Esker, P.D. 2006. The role of psychophysics in phytopathology: The Weber-Fechner law revisited. *Eur. J. Plant Pathol.* 114:199-213.

- Nutter, F.W., Jr., Gleason, M.L., Jenco, J.H., and Christians, N.C. 1993. Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. *Phytopathology* 83:806-812.
- Nutter, F.W., Jr. and Guant, 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. Chakaborty et al. (ed.). ASA, CSSA, and SSSA, Madison, WI.
- Nutter, F.W., Jr. and Littwiller, D. 1993. Alfalfa.Pro – a computerized disease assessment training program for foliar diseases of alfalfa. p. 15 In: *Proceedings of the 23<sup>rd</sup> Central Alfalfa Improvement Conference, 20-22 June 1993*. Univ. of Nebraska, Lincoln, NE.
- 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., Litwiller, D. and Wegulo, S.N. 1998. Do standard area diagrams improve the accuracy and precision of disease assessments? *Proceedings of the 7<sup>th</sup> International Congress of Plant Pathology*. p2.1.16.
- Nutter, F.W., Jr. and Schultz, P.M. 1995. Improving the accuracy and precision of disease assessment: Selection of methods and use of computer-aided training programs. *Can. J. Plant Pathol.* 17:174-178.
- Parker S.R., Shaw, M.W., and Royle, D.J. 1995a. The reliability of visual estimates of disease severity on cereal leaves. *Plant Pathology* 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.

- 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.
- 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.
- Tomerlin, J.R. and Howell, T.A. 1988. DISTRAIN: A computer program for training people to estimate disease severity on cereal leaves. *Plant Disease* 72:455-459.
- Yi, Q., Wang, P.P. and He, Y. 2008. Reliability analysis for continuous measurements: Equivalence test for agreement. *Statist. Med.* 27: 2816-2815.

**Table 3-1.** Summary statistics for bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) severity assessed by 10 raters on the three sets of peach and nectarine leaves with direct visual estimation<sup>a</sup>

| Rater <sup>b</sup> | Set 1  |        |         | Set 2  |        |         | Set 3  |        |         |
|--------------------|--------|--------|---------|--------|--------|---------|--------|--------|---------|
|                    | Mean   | Median | Std Err | Mean   | Median | Std Err | Mean   | Median | Std Err |
| A                  | 18.117 | 8.0    | 2.708   | 28.990 | 12.0   | 3.643   | 20.114 | 7.0    | 2.754   |
| B                  | 19.853 | 8.0    | 2.857   | 31.865 | 17.5   | 3.625   | 29.162 | 13.0   | 3.180   |
| C                  | 15.045 | 3.0    | 2.894   | 30.750 | 6.5    | 3.800   | 23.331 | 3.0    | 3.293   |
| D                  | 16.129 | 3.0    | 2.623   | 26.343 | 7.0    | 3.288   | 19.493 | 10.0   | 2.529   |
| E                  | 20.786 | 10.0   | 2.743   | 30.962 | 13.5   | 3.536   | 31.400 | 15.0   | 3.149   |
| F                  | 18.398 | 8.0    | 2.703   | 28.250 | 11.0   | 3.558   | 20.000 | 7.0    | 2.858   |
| G                  | 22.699 | 12.0   | 2.880   | 33.164 | 17.0   | 3.629   | 23.962 | 10.0   | 2.959   |
| H                  | 19.806 | 8.0    | 2.743   | 30.856 | 16.0   | 3.569   | 31.514 | 13.0   | 3.275   |
| I                  | 19.330 | 6.0    | 2.939   | 28.702 | 7.5    | 3.756   | 21.886 | 6.0    | 3.102   |
| J                  | 24.398 | 15.0   | 2.953   | 37.481 | 25.0   | 3.648   | 31.981 | 20.0   | 3.071   |
| A                  | 18.447 | 6.0    | 2.872   | 25.769 | 10.0   | 3.463   | 21.114 | 6.0    | 2.850   |
| B                  | 19.553 | 6.0    | 2.873   | 31.798 | 16.5   | 3.630   | 28.410 | 11.0   | 3.179   |
| C                  | 16.837 | 3.0    | 3.084   | 28.679 | 5.0    | 3.892   | 23.209 | 3.5    | 3.332   |
| D                  | 16.121 | 5.0    | 2.436   | 26.813 | 11.0   | 3.420   | 19.300 | 7.0    | 2.610   |
| E                  | 18.359 | 7.0    | 2.697   | 28.019 | 11.5   | 3.438   | 24.076 | 9.0    | 2.938   |
| F                  | 18.097 | 6.0    | 2.744   | 29.394 | 13.0   | 3.531   | 23.143 | 10.0   | 2.882   |
| G                  | 19.330 | 7.0    | 2.841   | 29.394 | 10.5   | 3.549   | 23.314 | 9.0    | 3.016   |
| H                  | 18.534 | 6.0    | 2.667   | 30.385 | 18.0   | 3.457   | 26.171 | 12.0   | 3.019   |
| I                  | 17.272 | 5.0    | 2.930   | 27.721 | 8.0    | 3.697   | 21.152 | 6.0    | 2.989   |
| J                  | 23.903 | 15.0   | 2.889   | 36.625 | 21.0   | 3.708   | 31.381 | 15.0   | 3.272   |

<sup>a</sup> Each of the 3 sets of leaves was assessed twice by direct visual estimation (DE) by 10 raters. The first assessment is above the second assessment.

<sup>b</sup> Raters A to D were experienced in assessing foliar disease severity and raters E to J had less experience rating disease severity.

**Table 3-2.** Intra-rater reliability, determined using the correlation coefficient ( $r$ ) for each of the 10 raters, for the direct visual estimation of bacterial spot (caused by *Xanthamonas arboricola* pv. *pruni*) severity of peach and nectarine leaves, for each of the 3 leaf sets<sup>a</sup>

| Rater <sub>b</sub> | Set 1 ( $n = 103$ ) |                     | Set 2 ( $n = 103$ ) |             | Set 3 ( $n = 104$ ) |             |
|--------------------|---------------------|---------------------|---------------------|-------------|---------------------|-------------|
|                    | $r$                 | 95% CI <sub>c</sub> | $r$                 | 95% CI      | $r$                 | 95% CI      |
| A                  | 0.983               | 0.973-0.987         | 0.972               | 0.953-0.977 | 0.967               | 0.951-0.977 |
| B                  | 0.998               | 0.996-0.998         | 0.998               | 0.997-0.999 | 0.999               | 0.998-0.999 |
| C                  | 0.973               | 0.952-0.977         | 0.941               | 0.911-0.958 | 0.966               | 0.950-0.977 |
| D                  | 0.986               | 0.976-0.988         | 0.990               | 0.985-0.993 | 0.948               | 0.923-0.964 |
| E                  | 0.982               | 0.967-0.985         | 0.970               | 0.950-0.977 | 0.935               | 0.870-0.935 |
| F                  | 0.977               | 0.967-0.985         | 0.991               | 0.986-0.993 | 0.966               | 0.943-0.973 |
| G                  | 0.980               | 0.961-0.982         | 0.978               | 0.960-0.981 | 0.976               | 0.964-0.983 |
| H                  | 0.958               | 0.935-0.970         | 0.965               | 0.948-0.976 | 0.943               | 0.896-0.949 |
| I                  | 0.990               | 0.982-0.992         | 0.994               | 0.990-0.996 | 0.981               | 0.971-0.986 |
| J                  | 0.991               | 0.986-0.993         | 0.992               | 0.987-0.994 | 0.982               | 0.971-0.986 |

<sup>a</sup> Raters assessed the same 3 sets of leaves twice using direct visual estimation and assigned each leaf a severity between 0 and 100% based on the percentage of diseased leaf area out of the total leaf area.

<sup>b</sup> Raters A to D were experienced in assessing foliar disease severity and raters E to J had less experience rating disease severity.

<sup>c</sup> 95% confidence intervals were determined for the correlation coefficients.

**Table 3-3.** Inter-rater reliability was determined for the direct visual estimation of bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) severity of 3 sets of leaves rated twice (Rep I and II) by 10 raters based on mean pairwise correlation coefficient ( $r$ ), the intra-class correlation coefficient ( $\rho$ ), and the repeatability coefficient (RC)

| Leaf set | Rep I |          |                 |           |             | Rep II |        |        |           |             |
|----------|-------|----------|-----------------|-----------|-------------|--------|--------|--------|-----------|-------------|
|          | $r$   | $\rho^a$ | RC <sup>b</sup> | $X^2^c$   | $P$ - value | $r$    | $\rho$ | RC     | $X^2$     | $P$ - value |
| I        | 0.967 | 0.956    | 14.829          | 7623.830  | 0.000       | 0.970  | 0.961  | 14.521 | 7742.148  | 0.000       |
| II       | 0.956 | 0.950    | 21.606          | 17527.162 | 0.000       | 0.966  | 0.957  | 19.260 | 13753.686 | 0.000       |
| III      | 0.929 | 0.900    | 23.641          | 20924.394 | 0.000       | 0.954  | 0.936  | 19.365 | 14038.892 | 0.000       |

<sup>a</sup> Calculated based on variance components formulated from ANOVA (Nita et al. 2003).

<sup>b</sup> Calculated based on the within subject variance (Bland and Altman, 1986 & 1999).

<sup>c</sup> Calculated based off of an allowable difference limit of 5 and the sum of squared deviation (Bland and Altman, 1986 & 1990).

**Table 3-4.** Rater accuracy was determined by comparing the actual<sup>a</sup> bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) severity and estimated bacterial spot severity for 3 sets of peach and nectarine leaves. The concordance coefficient ( $\rho_c$ ) measures the accuracy of the estimated disease severity. The related statistics include the correlation coefficient ( $r$ ) and the bias coefficient ( $C_b$ )<sup>b</sup>

| Rater <sup>c</sup> | Set 1 ( $n = 103$ ) |       |       | Set 2 ( $n = 103$ ) |       |       | Set 3 ( $n = 104$ ) |       |       |
|--------------------|---------------------|-------|-------|---------------------|-------|-------|---------------------|-------|-------|
|                    | $\rho_c$            | $r$   | $C_b$ | $\rho_c$            | $r$   | $C_b$ | $\rho_c$            | $r$   | $C_b$ |
| A                  | 0.980               | 0.983 | 0.997 | 0.939               | 0.949 | 0.990 | 0.955               | 0.960 | 0.994 |
| B                  | 0.980               | 0.984 | 0.995 | 0.888               | 0.918 | 0.968 | 0.922               | 0.946 | 0.974 |
| C                  | 0.964               | 0.966 | 0.998 | 0.858               | 0.881 | 0.974 | 0.952               | 0.956 | 0.996 |
| D                  | 0.972               | 0.983 | 0.989 | 0.960               | 0.966 | 0.994 | 0.926               | 0.946 | 0.980 |
| E                  | 0.978               | 0.986 | 0.992 | 0.917               | 0.935 | 0.981 | 0.901               | 0.924 | 0.974 |
| F                  | 0.971               | 0.975 | 0.996 | 0.924               | 0.939 | 0.985 | 0.935               | 0.940 | 0.995 |
| G                  | 0.962               | 0.973 | 0.989 | 0.894               | 0.921 | 0.970 | 0.944               | 0.947 | 0.997 |
| H                  | 0.982               | 0.967 | 0.994 | 0.894               | 0.916 | 0.975 | 0.898               | 0.925 | 0.971 |
| I                  | 0.987               | 0.988 | 0.998 | 0.910               | 0.923 | 0.986 | 0.935               | 0.935 | 1.000 |
| J                  | 0.940               | 0.968 | 0.971 | 0.802               | 0.866 | 0.925 | 0.880               | 0.925 | 0.951 |

<sup>a</sup>Actual severity was obtained using APS Assess software.

<sup>b</sup> Each set was assessed twice by direct visual estimation by 10 raters. These statistics are the mean of the first and second assessment.

<sup>c</sup> Raters A to D were experienced in assessing foliar disease severity and raters E to J had less experience rating disease severity.

**Table 3-5.** Comparison of accuracy and intra-rater reliability between experienced and inexperienced raters. Rater accuracy was determined by comparing the actual<sup>a</sup> bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) severity and estimated bacterial spot severity for 3 sets of peach and nectarine leaves. The concordance coefficient ( $\rho$ ) measures the accuracy of the estimated disease severity. The related statistics include the correlation coefficient ( $r$ ) and the bias coefficient ( $C_b$ )<sup>b</sup>. Intra-rater reliability was determined using the correlation coefficient ( $r$ )

| Rater Experience                                   | Accuracy          |                   |                   | Reliability       |
|--|-------------------|-------------------|-------------------|-------------------|
|  | $\rho$            | $r$               | $C_b$             | $r$               |
| Experienced  | 0.941             | 0.953             | 0.987             | 0.977             |
| Not Experienced                                    | 0.925             | 0.942             | 0.981             | 0.975             |
| Mean difference <sup>c</sup>                       | 0.0172            | 0.0112            | 0.0067            | 0.0018            |
| 95% Confidence interval of difference <sup>c</sup> | -0.0384 to 0.0049 | -0.0263 to 0.0058 | -0.0149 to 0.0005 | -0.0146 to 0.0104 |
| $t$ -statistic <sup>d</sup>                        | 1.480             | 1.350             | 1.670             | 0.280             |
| $P$ -value   | 0.143             | 0.183             | 0.101             | 0.782             |
| K-S statistic <sup>e</sup>                         | 5.880             | 5.880             | 2.180             | 1.800             |
| $P$ -value   | 0.053             | 0.053             | 0.337             | 0.407             |

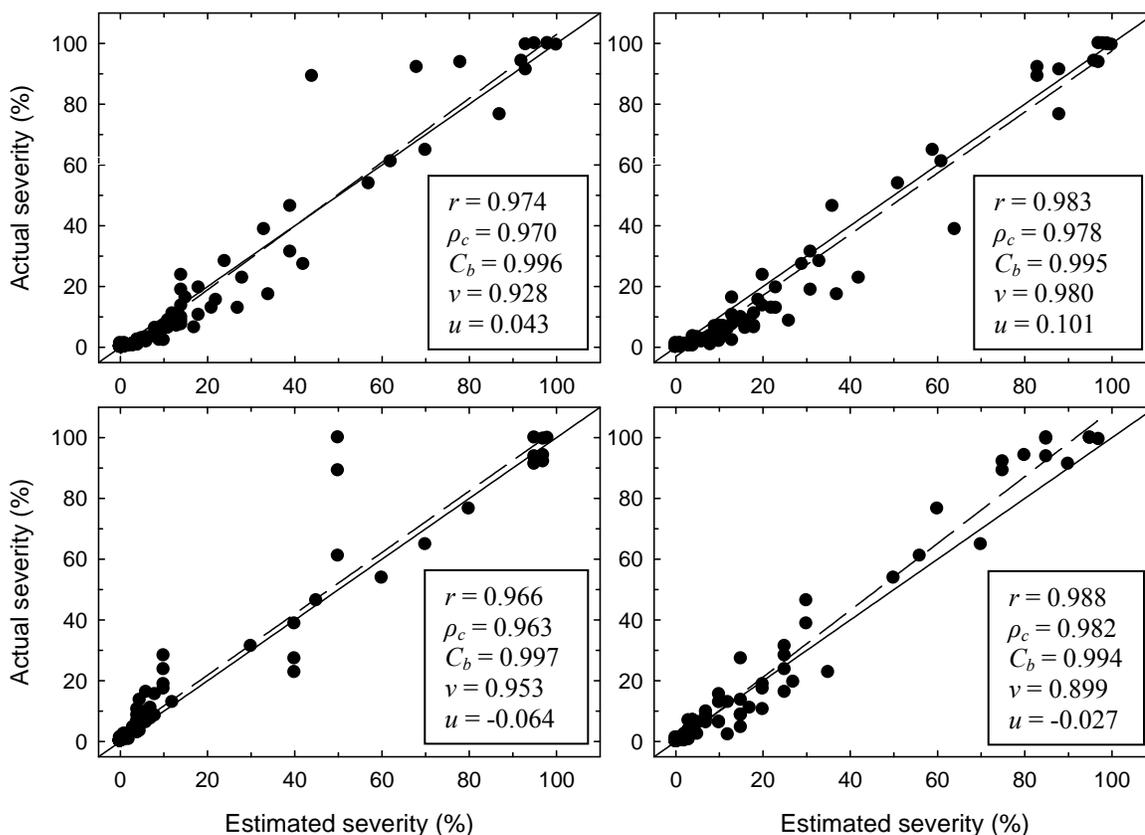
<sup>a</sup> Actual severity was obtained using APS Assess software.

<sup>b</sup> Each set was assessed twice by direct visual estimation by 4 experienced raters and 6 inexperienced raters. Accuracy and reliability statistics are the mean of each group.

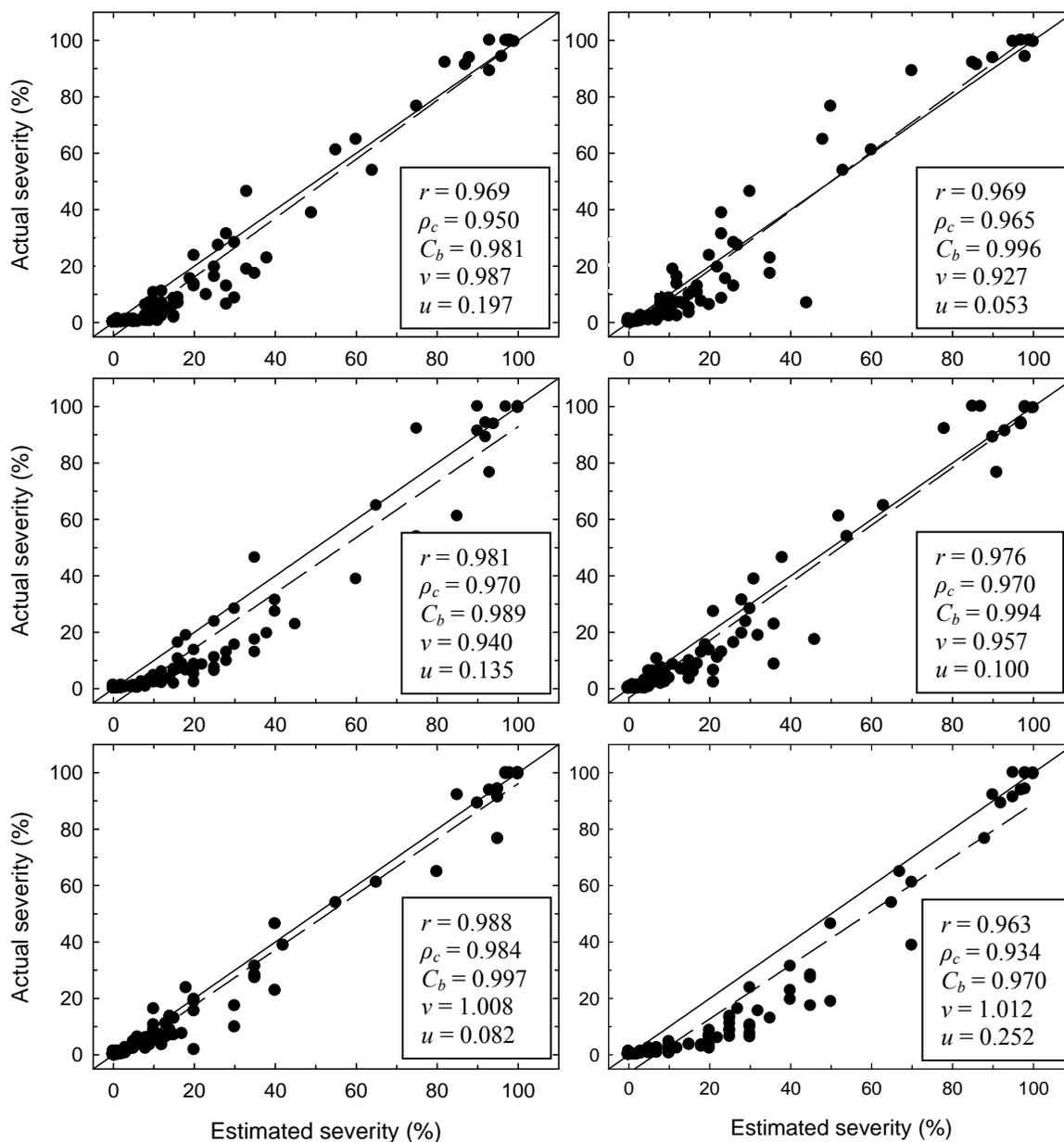
<sup>c</sup> Based on 2000 random bootstrap samples.

<sup>d</sup>  $t$ -test statistic and associated  $P$  value.

<sup>e</sup> Kolmogorov-Smirnov statistic and associated  $P$  value.



**Fig 3-1.** Lin's concordance analysis for the first assessment of the first set of leaves assessed using direct visual estimation by experienced raters A through D. The concordance line, representing a theoretical perfect agreement between actual and estimated severity, is indicated by the solid line and the best-fit line, the actual relationship between the estimated severity and actual severity, is represented by the broken line. The correlation coefficient,  $r$ , is a measure of precision and Lin's concordance coefficient,  $\rho_c$ , measures accuracy by combining the effects of precision and bias,  $C_b$ , the difference between the best-fit line and the concordance line. Variance,  $v$ , measures the relationship between the actual and estimated severity where the variance of 1 represents a perfect relationship. The coefficient  $u$  defines the location of the best-fit line relative to the concordance line.



**Fig 3-2.** Lin's concordance analysis for the first assessment of the first set of leaves assessed using direct visual estimation by inexperienced raters E through J. The concordance line, representing a theoretical perfect agreement between actual and estimated severity, is indicated by the solid line and the best-fit line, the actual relationship between the estimated severity and actual severity, is represented by the broken line. The correlation coefficient,  $r$ , is a measure of precision and Lin's concordance coefficient,  $\rho_c$ , measures accuracy by combining the effects of precision and bias,  $C_b$ , the difference between the best-fit line and the concordance line. Variance,  $v$ , measures the relationship between the actual and estimated severity where the variance of 1 represents a perfect relationship. The coefficient  $u$  defines the location of the best-fit line relative to the concordance line.

## Chapter 4

### **Temporal Bacterial Spot Progress and Factors Affecting the Timing of Abscission of Peach and Nectarine Leaves Infected with *Xanthomonas arboricola* pv. *pruni***

#### ABSTRACT

The objective of this study was to obtain basic knowledge on the disease severity at initial disease onset, rate of bacterial spot development in a growing season, and the progression of tree defoliation as they relate to the current management strategies in Pennsylvania. Data on bacterial spot severity were obtained on a 5 to 7 day interval in replicated field plots of four cultivars subjected to different bactericide programs during the 2008 and 2009 seasons and used to construct temporal disease progress curves. Additionally in the 2009 season, the time of leaf abscission was monitored for up to 20 leaves on each of three tagged shoots per tree in each of the experimental plots. In each year, a total of 84 disease progress curves were constructed by plotting the cumulated disease severity values against time of observation (days) and analyzed with linear regression methods. The effects of cultivar, treatment, and the interaction between the two factors were assessed based on parameters estimating initial disease severity at the onset of epidemics, and the rate of disease progress were compared. Survival analysis using the Kaplan – Meier and the accelerated failure time (AFT) models of time to leaf abscission was conducted for the 2009 data in order to assess the effects of bactericide treatment, cultivar, leaf age, and initial disease onset on the survival of leaves infected by *Xanthomonas arboricola* pv. *pruni*. In 2008 and 2009, cultivar and treatment were

significant factors affecting the rate of disease progress. The cultivar Sweet Dream and treatments containing Kocide 3000 (a copper containing compound) were consistently associated with high rates of disease progress while the standard antibiotic oxtetracycline (Mycoshield) treatment and the untreated check had the lowest rates of disease progress. In 2009, 48% of leaves assessed for disease severity abscised while the remaining leaves were censored. The mean time to leaf abscission ranged ( $T$ ) from  $41.8 \pm 0.58$  to  $56.3 \pm 0.30$  days and was significantly ( $P < 0.0001$ ) affected by cultivar, initial disease onset, and leaf age but not by the bactericide treatment. The shortest time to leaf abscission was noted for cultivar Beekman (47.7 days) while diseased leaves survived longest on the nectarine Easternglo (54.9 days). The AFT model indicated that every additional 1% initial disease severity decreased the time to leaf abscission by 10.7%. Results indicate that bacterial spot symptoms do not follow standard disease progress curves and that strategies for bacterial spot management should focus on reducing initial disease.

## INTRODUCTION

Bacterial spot of peach and nectarine is a threat to eastern US stone fruit industry where environmental conditions are conducive to infection by *Xanthomonas arboricola* pv. *pruni* (*Xap*). Although the general epidemiology of bacterial spot of peach and nectarine is well understood (Ritchie, 1995), the specific details of temporal disease progression in the eastern US are lacking. For example, it is well documented that bacterial spot is a polycyclic disease with multiple infection cycles throughout the growing season that are favored by warm wet weather (Ritchie, 1995; Zehr et al., 1996).

However, specific details about the effects of host cultivar on the timing of leaf and fruit infection, the rate of temporal disease progress, and the timing of tree defoliation in the eastern US are lacking. Previous studies in the eastern US have often analyzed bacterial spot epidemics by means of area under disease progress curves (AUDPC) (e.g. Lalancette and McFarland, 2007). Although useful when summarizing and comparing treatments (e.g., bactericides evaluated for disease control), analyses based on AUDPCs fail to account for other epidemiologically important factors such as the timing of infection and tree defoliation, rates of disease progression, and the role of cultivar effects in disease progress. Knowledge of these factors is essential for developing effective and sustainable bacterial spot management practices. For example, knowledge about the effects of experimental bactericides on the timing of leaf abscission can help identify the efficacious products by focusing on products that delay tree defoliation.

Disease progress curves and models may be used to quantify and compare epidemics by summarizing and simplifying the average change in disease over time (Madden et al., 2007). Through the mathematical comparison of disease progress curves, it is possible to differentiate bactericide treatments and even cultivars in their effects on rate of disease progress, levels of initial disease at onset of epidemics, and maximum levels of disease in a statistically meaningful manner (Gilligan, 1990). While disease progress curves have been used to characterize bacterial spot epidemics in Italy (Battilani et al., 1999), strong evidence suggests there are major differences between epidemics occurring in Europe and the US, including differential formation of overwintering cankers. Nevertheless, no study to date has mathematically analyzed and compared

bacterial spot disease progress curves despite the impact on disease management such a study might have.

Premature tree defoliation is a common symptom of bacterial spot that is often overlooked in epidemiological studies (Battilani et al., 1999; Lalancette and McFarland, 2007; Zaccardelli et al., 1998). When disease severity is high, entire branches may defoliate which ultimately reduces fruit quality (Ritchie, 1995). The effects of bacterial spot severity on tree defoliation are likely to be more pronounced on the more susceptible cultivars which suffer more severe disease. Despite the widespread evidence for prevalence and severity of premature defoliation associated with severe bacterial spot epidemics in commercial peach and nectarine orchards in the eastern US (Lalancette and McFarland, 2007; Ritchie, 1995; Zaccardelli et al., 1998; Battilani et al., 1999), factors affecting this symptom have not been determined. For example, it is not well understood how cultivar, treatment, leaf age, and disease severity contribute to leaf abscission. Premature defoliation may be quantified and factors associated with this symptom may be compared using survival analysis technique (Madden and Nault, 1983; Westra et al., 1994; Scherm and Ojiambo, 2004; Ojiambo and Scherm, 2005; Esker et al., 2006; Pethybridge et al., 2010). This technique, often used in clinical studies to model data on time-to-death, may be effectively employed in plant epidemiology (Muenchow, 1986) and has recently been used to assess the effectiveness of management options for ray blight of pyrethrum caused by *Phoma ligulicola* (Pethybridge et al., 2010).

Effective integrated management of bacterial spot of peach and nectarine in the eastern US orchards requires a thorough understanding of factors moderating temporal disease progression as well as the effects of disease on the timing of tree defoliation.

There is a need, therefore, to develop models to quantify these epidemiological processes in order to identify the factors that can be responsive to disease management practices. The overall goal of this study is to obtain basic knowledge on the initial disease onset, rate of bacterial spot progression in season, and progression of tree defoliation as they relate to current management strategies in Pennsylvania. The specific objectives of the study was two-fold: (i) to model and compare the rate of progression and disease onset of bacterial spot epidemics based on disease progress curves with respect to cultivar and treatment, and (ii) to model the effects of bactericide treatment, cultivar, leaf age, and timing of initial disease onset on the time-to-abscission of peach and nectarine leaves caused by bacterial spot, and therefore determine the factors that affect premature tree defoliation.

## MATERIAL AND METHODS

*Experimental Chemicals:* Data on time to defoliation and the epidemiology of bacterial spot was obtained from plots that were used as part of a product efficacy test carried out in the FREC block (previously described in Chapter 2). In 2008, experimental treatments evaluated included: (i) the standard antibiotic oxytetracycline treatment applied as Mycoshield (354.88 ml); (ii) the copper containing compound Kocide 3000 (DF 49.30ml); (iii) Kocide 3000 (DF 49.30ml) + 0.1% solution of Garlic product; (iv) 0.1% solution of Garlic product; (v) the experimental antibiotic kasugamycin applied as Kasumin (1.89L/ 378.54L) + the fungicide Captan (1360.78g); (vi) Kasumin (1.89L/378.54L) + Kocide 3000 (49.30ml); and (vii) non-treated check. Treatments were

arranged in a split-plot design with chemical products assigned to main-plots and single-trees of each cultivar assigned to sub-plots and replicated four times. In 2009, the same treatments tested in 2008 were assessed except that the Garlic product was evaluated at a rate of 1% solution.

*Foliar Disease Severity Assessment:* In 2008, 15 random leaves per tree were assessed every 5 to 7 days for bacterial spot severity using the 1-7 ordinal rating scale (1 = 0% lesion area, 2 = 1-3%, 3 = 4-8%, 4 = 9-15%, 5 = 15-25%, 6 = 25-45%, 7 = >45%) (as previously described in Chapter 2). In 2008, defoliation was not evaluated. In 2009, three shoots per tree within each sub-plot were tagged using plastic flagging tape. Starting at the leaf closest to the main scaffold, up to 20 leaves per shoot were evaluated every 5 to 7 days for disease severity using the 1 to 7 rating scale. Defoliation of leaves on the tagged shoots was also evaluated, in 2009, by recording the date leaves were first observed to be missing.

*Data Analysis for Disease Progress:* For data analysis, the midpoint of each scale interval of the 1-7 rating scale was used (1 = 0% midpoint symptomatic leaf area, 2 = 2%, 3 = 6%, 4 = 12%, 5 = 20%, 6 = 35%, 7 = 72.5%). For each assessment date, data from each tree were averaged in order to obtain a single severity value, based on 15 leaves per tree in 2008 and up to 60 leaves per tree (3 shoots of 20 leaves) in 2009. The four replicates were treated separately. Disease progress curves were created by plotting the percent disease severity (y-axis) against time in days starting after the first disease assessment (x-axis). For each year (2008 and 2009) separate curves were created for each replicate of each treatment on each cultivar (i.e., sub-plot).

Because disease severity did not steadily increase or decrease over time (Fig. 4-1 and 4-2), several nonlinear disease progress models were tested for goodness of fit including the logistic and Gompertz models as well as the cumulative response model (Madden et al., 2007). Models were evaluated visually for goodness of fit by comparing the observed values with the fitted values from the model, and plots of standardized residuals. Models were also evaluated based on the coefficient of determination ( $R^2$ ) which is expressed as 1 minus the ratio of the sum of squares for error (SSE) to the total sum of squares (SST) and is considered the proportion of variability that is explained by the model (Madden et al., 2007). Preliminary analysis revealed that disease severity data from both the 2008 and 2009 experiments could not be adequately fitted to any of the standard growth curves commonly used in plant disease epidemiology (e.g. Gilligan, 1990; Ngugi et al., 2000; Madden et al., 2007). Based on this observation, models based on cumulative values of the disease severity were investigated. For each sub-plot data, cumulative values were computed by adding the value of the 1<sup>st</sup> assessment with the 2<sup>nd</sup> assessment, then by adding the value of the 2<sup>nd</sup> assessment with the 3<sup>rd</sup> assessment, and so on. Cumulative values were used as the response variable with time of assessment as the explanatory variable in a simple linear regression analysis. Treatment and cultivar effects on disease progression were compared by ANOVA based on estimated parameters as in Ngugi et al. (2000). All statistical analyses were completed using GENSTAT 12 (VSN International, Hemel Hempstead, UK).

*Data Analysis for Survival Analysis:* Preliminary analysis of factors affecting the defoliation of peach and nectarine leaves was carried out by calculating Kaplan – Meier (KM) estimates [ $\hat{S}(t)$ ] of survival (Pethybridge et al., 2010; Kaplan and Meier, 1958).

KM estimates are the nonparametric maximum likelihood estimates of the probability that a leaf will survive to time  $t_j$  and can be expressed as:

$$\hat{S}(t) = \prod_{j=1}^k \left( \frac{n_j - d_j}{n_j} \right)$$

where  $n_j$  is the number of leaves not yet defoliated right before time  $t_j$ , and  $d_j$  is the number of defoliated leaves at time  $t_j$ , where  $t_k \leq t \leq t_{(k+1)}$  for  $k = 1, 2, \dots, r$  (Kaplan and Meier, 1958). Difference in time to defoliation was assessed among chemical treatments, leaf age, and cultivar using the non-parametric log-rank, Wilcoxon, and -2Log(LR) tests using the LIFETEST procedure in SAS (SAS Institute).

Only factors with significant KM statistics were further analyzed using the accelerated failure time (AFT) model. The AFT model assumes a linear relationship between the logarithm of time to an event (i.e., defoliation) and the underlying covariates (i.e., cultivar, leaf age, and initial disease onset) and which can be expressed in its general form as:

$$\log Te_i = f(x_i) + \varepsilon_i (i = 1, \dots, n)$$

where  $Te_i$  is the time to event (i.e., abscission) of the  $i$ th individual,  $f(x_i)$  describes the relationship of covariates, and  $\varepsilon_i$  are independent error terms (Pethybridge et al., 2010). Unlike ordinary regression, AFT models do not assume a normal distribution of error terms. Therefore, it was necessary to identify the best distribution to fit the terms in the model because different distributions result in different accounts of the model (Allison, 1995; Collett, 2003). In order to identify the best distribution, five models were specified and included treatment, cultivar, age, initial disease onset, and cultivar + age + initial disease onset. The exponential, logistic, log-logistic, log-normal, and Weibull

distributions were tested. The distributions were assessed for goodness of fit using the LIFEREG procedure in SAS in order to find the log-likelihood. The log-likelihood was then used to calculate Akaike Information Criterion (AIC) values (Akaike, 1974; Afifi et al., 2004). The log-normal distribution most often had the smallest AIC values and therefore provided the best fit compared to the other distributions (Table 4-1). Therefore, the log-normal distribution was used to evaluate the effects of cultivar, age, and disease severity at initial disease onset on defoliation using the LIFEREG procedure in SAS.

## RESULTS

*Weather:* In both 2008 and 2009, weather was conducive to bacterial spot development. The 2008 growing season was particularly wet; a total of 501.90 millimeters of rain and 881 leaf wetness hours were recorded at the PSU FREC between April 1<sup>st</sup> and July 31<sup>st</sup> (Appendix A). The mean temperatures were 12.2, 14.7, 22.6, and 23.7 degrees Celsius for the months of April, May, June, and July, respectively. In 2009, the growing season was dryer; 270.73 mm of rain and 1,051 leaf wetness hours, in total, were recorded at the PSU FREC between April 1<sup>st</sup> and July 31<sup>st</sup> (Appendix B). Average temperatures were 12.1, 16.6, 20.8, and 21.8 degrees Celsius for the months of April, May, June, and July, 2009, respectively.

*Disease Progress:* A total of 84 bacterial spot disease progress curves from 4 replicates of 7 treatments, tested three cultivars, were analyzed in 2008 and again in 2009. Each year, bacterial spot developed on peach foliage and fruit. The disease progress curves depicted a sinusoidal epidemic with periods of rate increase followed by

rate decrease, and again by rate increase, for both the 2008 and 2009 epidemics (Figs. 4-1 and 4-2). There was also significant variability between repetitions of the same cultivar and treatment. This variability resulted either in poor fits for models based on standard growth curves and in many cases the regression algorithms failed to converge thereby necessitating the use of alternative models. The model based on cumulative disease severity values as the response variable provided steady rates of disease increase which could be readily analyzed and compared as well as provide a good fit for the disease progress data (Figs. 4-1 and 4-2). Fitted curves (Figs. 4-3 and 4-4) differed little from observed data that had been fitted with the model. The  $R^2$  values for all fitted curves were generally high and ranged from 0.920 to 0.992 in 2008 and from 0.975 to 0.995 in 2009 for the presented treatments of the cultivar Beekman (Figs. 4-3 and 4-4). In general, standardized residual plots showed randomized scatter although in some cases there was evidence for departure from randomness or independence (Figs. 4-3 and 4-4). Statistical tests revealed that for all data sets, the residuals passed the normality test (*data not shown*). Based on these results, treatment factors (bactericide treatments and cultivar response) were analyzed based on their effects on the regression intercept parameter (representing an estimate of the mean level of bacterial spot severity at the time of first assessment hereafter referred to as disease onset), and the rate parameter (which represents the instantaneous change in disease severity over time) (Madden et al., 2007).

Treatment effects on initial disease onset for bacterial spot were inconsistent between 2008 and 2009 (Tables 4-2 and 4-3). In 2008, there was no significant effect of treatment ( $F = 0.90$ ;  $P = 0.527$ ), cultivar ( $F = 1.39$ ;  $P = 0.251$ ), or the interaction of treatment and cultivar ( $F = 0.78$ ;  $P = 0.734$ ) on disease onset (Table 4-2). That is, levels

of disease severity at onset for all cultivars and treatments, including the untreated check, were not significantly different from each other. By contrast, in 2009 there was a significant effect of treatment ( $F = 5.47$ ;  $P = 0.002$ ) and the interaction between treatment and cultivar ( $F = 2.73$ ;  $P = 0.002$ ); however, there was no significant effect of cultivar ( $F = 2.22$ ;  $P = 0.094$ ) on initial disease onset. Interestingly, the untreated check did not produce the highest levels of disease severity at disease onset nor did the standard antibiotic oxytetracycline (Mycoshield) treatment produce the lowest levels of disease severity or rates of disease progress as might be expected.

In 2009, disease severity at onset for each cultivar depended on the treatment received and severity differed among treatments depending on the cultivar (Table 4-3). Overall, the highest level of disease severity at onset (9.97%) was noted on the highly susceptible cultivar Snow King for which there was no significant differences due to treatment. The mean disease severity at disease onset for Snow King ranged from 7.91% for the untreated check to 13.8% for the standard Mycoshield treatment. Among the treatments, the lowest mean disease severity at onset was noted for the treatments combining the antibiotic kasugamycin (Kasumin) with the bactericide copper or the fungicide Captan, although the effects on these treatments depended on the cultivar. Overall, the mean disease severity at disease onset for the mean treatment ranged from 5.17% for the Kasumin + Captan treatment to 12.99% for the Mycoshield treatment. Based on the least significant differences (LSD) test where the LSD was 3.744 for treatments and 6.270 for the interaction between treatments and cultivars, treatments consisting of Mycoshield, Kocide 3000, Kocide 3000 + Garlic, Garlic, and the untreated check were not significantly different from each other with respect to disease severity at

disease onset. The treatments consisting of Kocide 3000, Kocide 3000 + Garlic, Kasumin + Captan, Kasumin + Kocide 3000, and the untreated check were also not significantly different from each other (Table 4-3).

In 2008, there was a significant effect of treatment ( $F = 9.21$ ;  $P < 0.001$ ) and cultivar ( $F = 30.47$ ;  $P < 0.001$ ), but not the interaction between these factors ( $F = 1.66$ ;  $P = 0.060$ ), on the rate of disease progress. The mean rate of disease progress for cultivar mean ranged from 2.015 for the cultivar Beekman to 2.557 for the cultivar Sweet Dream and was significantly higher for the more susceptible cultivars (Snow King and Sweet Dream) than for the less susceptible cultivars Easternglo and Beekman (Table 4-4). Among the bactericide treatments, the mean rate of disease progress for the treatment mean ranged from 1.927 for the Mycoshield treatment to 2.716 for the Kocide 3000 treatment. Based on the least significant differences (LSD) test, treatments consisting of Kocide 3000, Kocide 3000 + Garlic, and Kasumin + Kocide 3000 had the highest rates of disease progress and were not significantly different from each other. Treatments consisting of Mycoshield, Garlic, Kasumin + Captan, and the untreated check had the lowest rates of disease progress and were not significantly different from each other (Table 4-4). Among the cultivars, Easternglo and Beekman had the lowest rates of disease progress and did not significantly differ from each other (Table 4-4). Cultivars Snow King and Sweet Dream recorded the highest rates of disease progress.

In 2009, there was a significant effect of treatment ( $F = 11.50$ ;  $P < 0.001$ ) and cultivar ( $F = 7.61$ ;  $P < 0.001$ ) but as was the case in 2008, there was no significant effect of the interaction between these factors on the rate of disease progress ( $F = 1.14$ ;  $P = 0.337$ ; Table 4-5). The mean rate of disease progress for cultivar mean ranged from 2.07

for the cultivar Beekman to 2.423 for the cultivar Snow King and was significantly higher for the most susceptible cultivar Snow King (Table 4-5). Among the bactericide treatments, the mean rate of disease progress for the treatment mean ranged from 1.929 for the untreated check to 2.716 for the Kasumin + Kocide 3000 treatment. Based on the least significant differences (LSD) test where the LSD was 0.2255 for the treatments and 0.1632 for the cultivar, treatments Kocide 3000, Kocide 3000 + Garlic, Garlic, and Kasumin + Captan were not significantly different from each other while treatments Mycoshield and Garlic were not significantly different from each other. Interestingly, there was no difference between the effects on the rate of disease progress between the standard bacterial spot treatment Mycoshield and the untreated check. The treatment Kasumin + Kocide 3000 had the highest rate of disease progress. For the effects of cultivar on rate of disease progress, cultivars Snow King and Sweet Dream were not significantly different, while cultivars Easternglo and Sweet Dream were also not significantly different, and lastly, the cultivars Easternglo and Beekman were not significantly different from each other (Table 4-5).

*Defoliation:* In 2009, a total of 3,052 leaves were evaluated for disease severity at time of abscission. By the end of the study, 1,460 leaves had abscised while 52.16% of the leaves were censored, i.e., had not abscised. For each sub-plot, leaves were divided into three age categories (young, middle-aged, and old) in order to quantify the effects of leaf age on abscission. During the study, 490 leaves were considered young, 715 leaves were considered middle-aged, and 1,847 leaves were considered old. Of the young leaves, 37 abscised by the end of the study compared to the 112 middle-aged and 1,311 old leaves that had abscised. The remaining 92.45% of the young, 84.34% of the middle-

aged, and 29.02% of the old leaves were censored. KM plots indicate that old leaves abscised faster than young and middle-aged leaves (Fig. 4-5).

The effects of cultivar and bactericide treatments on the time to leaf abscission were also evaluated. During the study, 722, 757, 801, and 772 leaves were monitored from cultivars Easternglo, Beekman, Snow King, and Sweet Dream, respectively. By the end of the study, 247, 395, 481, and 337 leaves had abscised, respectively. The remaining 65.79%, 47.82%, 39.95%, and 56.35% of leaves from each cultivar, respectively, were censored. Both Wilcoxon and the log-rank tests indicated that survival curves significantly differed ( $P < 0.0001$ ) among leaf age and cultivars, but not among treatments ( $P = 0.0904$  for the Wilcoxon test, and 0.1048 for the log-rank test). This was supported by plots of the KM survivor curves depicting the effects of age, cultivar, and treatment (Figs. 4-5, 4-6, and 4-7) which show significant separations for the different factors. Leaves from the cultivars Beekman and Snow King abscised faster than those of the cultivars Easternglo and Sweet Dream, as indicated by the KM plots (Fig. 4-6). The effects of bactericide treatment were not significant based on the Wilcoxon and log-rank tests ( $\chi^2 = 10.936$ ,  $P = 0.0904$  for the Wilcoxon test;  $\chi^2 = 10.508$ ,  $P = 0.1048$  for the log-rank test). That is, leaves from one treatment did not abscise faster than leaves from any other treatment, as depicted in the KM plot (Fig. 4-7).

The AFT model was a good fit for the time-to-abscission for peach and nectarine leaves displaying symptoms of bacterial spot (Table 4-1). Of the five distributions examined, the log-normal distribution was determined to be the best fit with AIC values between 4,818.2 and 6,328.3. On the contrary, the logistic model was a poor fit with AIC values between 14,632.4 and 17,267.7 (Table 4-1). Therefore, the AFT model with a log-

normal distribution was chosen to model the effects of cultivar, leaf age, and initial disease onset on the survival of peach and nectarine leaves affected by bacterial spot. The AFT model facilitated the evaluation of the effect of initial disease onset because it was measured on a continuous scale – a factor that could not be accounted for by the KM model. The effects of cultivar, leaf age, and initial disease onset were significant ( $P < 0.0001$ ) factors affecting leaf abscission.

The estimates of mean survival time from the AFT model indicated the young leaves had the longest survival time of 45.1 days ( $\pm 0.19$ ). Compared to young leaves, old leaves abscised significantly faster ( $\chi^2 = 265.80$ ;  $P < 0.0001$ ) with a mean survival time of 41.8 days ( $\pm 0.58$ ). Middle-aged leaves also abscised faster than young leaves but not significantly so ( $\chi^2 = 4.28$ ;  $P = 0.0385$ ) with a mean survival time of 56.3 days ( $\pm 0.30$ ) (Table 4-6).

The effects of cultivar on leaf abscission were also evaluated. The AFT estimates of mean survival time indicated that leaves of the cultivar Sweet Dream had the longest survival time of 54.8 days ( $\pm 0.83$ ). Leaves of the cultivar Easternglo abscised only slightly faster than those of Sweet Dream ( $\chi^2 = 3.90$ ;  $P = 0.0483$ ) with a mean survival time of 54.9 days ( $\pm 0.88$ ). Leaves of the cultivars Snow King ( $\chi^2 = 81.83$ ;  $P < 0.0001$ ) and Beekman ( $\chi^2 = 95.70$ ;  $P < 0.0001$ ) abscised significantly faster than those of Sweet Dream with mean survival times of 48.01 days ( $\pm 0.81$ ) and 47.7 days ( $\pm 0.92$ ), respectively (Table 4-6). Disease severity at initial disease onset was also a significant factor affecting leaf abscission ( $\chi^2 = 606.88$ ;  $P < 0.0001$ ; Table 4-6). Based on the AFT model, every additional 1% of severity of initial disease reduced the time to leaf

abscission by 10.78% (calculated as  $\{e^{\beta_1} - 1\} \times 100$  and  $\beta_1 = -0.1141$ , the parameter for the initial disease onset in table 4-6) (Pethybridge et al., 2010).

## DISCUSSION

In this study, bacterial spot epidemics were analyzed using statistical methods in order to establish the parameters that best illustrate the progression of bacterial spot in peach and nectarine, in response to different chemical treatments, cultivars, and the interaction between the two factors. A linear regression model based on cumulative disease severity values was used to allow the direct comparison among epidemics under these different conditions. The model provided a good fit to the data for both years as the fitted curves were close to the observed values resulting in high  $R^2$  values that exceeded 0.90 for all separate regressions (Figs. 4-3 and 4-4). Plots of standardized residuals generally indicated good fit with a few exceptions in which a systematic pattern was discernable. In linear regression analysis, two key assumptions are made about the variance of the data (i.e., the residuals): normal distribution and independence (i.e., no correlation) (Montgomery and Peck, 1992). Because all residual plot values passed a normality test, plots with obvious pattern suggested the likelihood of autocorrelation, although this was not investigated further. This is not at all surprising given the variability within the data and the wave-like nature of the temporal disease progress curves. The important implication of these results is that: (i) bacterial spot disease epidemics do not follow typical temporal disease progress curves that have been successfully used in many studies for comparative epidemiology; and (ii) the proposed

method of regression analysis based on cumulative disease severity values provides a good alternative to the use of standard disease progress curves for this disease. However, models based on cumulative disease severity values should be evaluated further to ensure assumptions of linear regression methods are not violated.

The effects of various factors were evaluated based on parameter estimates obtained from the model. As expected, the parameters depicting disease onset and rate of disease progression differed with respect to cultivar and chemical treatment. Parameters also varied between 2008 and 2009, which is likely due to varied weather and environmental conditions (Zehr et al., 1996) although the effects of weather patterns were not investigated. Also in this study, premature defoliation of peach and nectarine, associated with bacterial spot, was quantified using survival analysis. The time to leaf abscission was related to factors including cultivar, leaf age, and disease severity of bacterial spot at the initial onset of epidemics.

The factors affecting initial disease onset varied between years. In 2008, none of the treatments or cultivars was significantly different from each other (Table 4-2). By contrast, the effects of treatment and the interaction between cultivar and treatment were significant factors affecting the onset of disease in 2009. In botanical epidemiology, the intercept parameter, which measures the initial disease, biologically represents the initial amount of inoculum potentially available to initiate infection (Madden et al., 2007; Ngugi et al., 2000; Gilligan, 1990). The results of the 2008 season were therefore not surprising because data collection began soon after fields had been inoculated. It is possible that infections resulting from the inoculation masked whatever variation may have existed in disease severity due to initial overwintering inoculum. In 2009, the overwintering

inoculum likely played a more important role. Because 2008 was the first year the trees had been inoculated, the plots ended in different levels of infection and bacterial spot severity, due to the bactericide treatment they had received during the season. However, the purpose of examining initial disease onset was to determine whether or not the different cultivars grown in the FREC block could delay the onset of the epidemic and therefore reduce levels of disease severity when disease assessment began. Treatments used during the season are not likely to affect disease onset so the significant effects found in 2009 are likely just an artifact because the first disease assessment was made on 22 May while the first bactericide application was not made until 28 May (Madden et al., 2007). Differential susceptibility of peach leaves belonging to different cultivars has previously been demonstrated (du Plessis, 1988; Civerolo, 1975) and confirmed by our own observations. However, these results demonstrated that differences in susceptibility do not occur at initial disease onset for the cultivars tested in this study. Instead, cultivar affected the rate of disease progress in both 2008 and 2009 epidemics. Of the cultivars examined, Snow King and Sweet Dream had the highest average rates of disease progress for both years (Tables 4-4 and 4-5). These two cultivars were bred in California where the environment is not suitable for bacterial spot development (Schupp, J.R. *personal communication*).

As expected, bactericide treatments significantly affected the rate of disease progression for both years. The highest rates of disease progress for both years were associated with treatments containing Kocide 3000, a copper containing compound (Tables 4-4 and 4-5). The application of these treatments resulted in significantly greater rates of disease progress than the untreated check. These results imply that treatments of

Kocide 3000 aid the progression of bacterial spot. However, copper containing compounds are commonly used as a part of successful bacterial spot management (Ritchie, 1995). Nevertheless, high rates of copper are phytotoxic to peach and nectarine leaves and have been shown to produce symptoms of irregularly-shaped shot-holes (Lalancette and McFarland, 2007). Because treatment was not a significant factor affecting defoliation, it is not likely that the lack of defoliation (producing high levels of disease severity) was related to the high rates of disease progression. Instead, it may be that the shot-hole symptoms related to high rates of copper were difficult to distinguish during disease assessment, resulting in artificially high rates of disease progress for these treatments.

In 2008, the standard antibiotic oxytetracycline (Mycoshield) treatment had the lowest rate of disease progress; however, it was not significantly different from that of the untreated check (Table 4-4). In 2009, the lowest rate of disease progress was associated with the untreated check, and again, it was not significantly different from the standard antibiotic treatment (Table 4-5). On the face of it, these results would imply that no chemical treatment was useful for management of bacterial spot epidemics. However, that is incorrect. The method used here to assess disease severity and compare the effects of chemical treatments on the rate parameter cannot take into account the effects of defoliation which can be severe when bacterial spot pressure is high (Ritchie, 1995). This is because disease severity can only be assessed for leaves remaining on the tree. Heavily diseased leaves, which would receive the highest severity scores, may abscise between assessment dates. In 2008, severe leaf abscission was observed but not recorded; however, in 2009, abscission was assessed every 5 to 7 days. This enabled us

to conduct a survival analysis of time to abscission in order to determine the factors that affect abscission as well as obtain a more realistic depiction of disease progress.

Leaf abscission likely influenced the parameters of the disease progress of bacterial spot. Indeed, nearly half (48%) of all leaves assessed in 2009 prematurely abscised. As a result of the survival analysis, we know that severely diseased leaves abscise faster than leaves with low levels of severity that persist longer on the trees. Because of that, the remaining leaves that are assessed for the study of disease progression have lower levels of disease severity. Without the abscised leaves with high levels of disease severity to take into account, assessments of disease are artificially lowered. This may explain why the disease progress rates associated with the standard antibiotic oxytetracycline (Mycoshield) treatment and the untreated check were not significantly different from each other in both years. It is likely that leaves of the untreated check abscised quickly as high levels of disease developed, artificially lowering the rate of disease progress for the untreated check to levels similar to those of the Mycoshield treatment. A similar explanation can be made for the cultivar Sweet Dream. Not only was it associated with the highest rates of disease progress in 2008 and 2009, leaves of that cultivar abscised slower than those of the other cultivars. Because the diseased leaves were maintained on the trees longer than any of the other cultivars, higher levels of disease severity could be assessed. This example also indicates that rates of disease progress of bacterial spot found in this study, and likely in other epidemiological studies where leaf abscission occurred, underestimate the actual rate of disease increase in the field.

The effect of bactericide treatment, cultivar, leaf age, and initial disease onset were evaluated using survival analysis in order to determine if these factors affected abscission of peach and nectarine leaves. All but bactericide treatment significantly affected the time of leaf abscission (Table 4-6 and Figs. 4-5, 4-6, and 4-7). Leaves with high levels of disease severity when disease assessment began, abscised much earlier than those with low levels of severity. The AFT model revealed that small differences (1%) in initial disease severity resulted in a greater than 10% difference in the duration of leaf survival. These results suggest that levels of initial disease are the most important factor determining the likelihood of a leaf abscising. Bacterial spot management strategies that aim to reduce tree defoliation should therefore focus on reducing the level of initial disease, for example, through minimizing the amount of overwintering inoculum.

Older leaves abscise significantly faster than middle-aged and young leaves presumably because older leaves have been exposed to multiple infection periods and have greater amounts of overall disease severity (Ojiambo and Scherm, 2005). However, preliminary studies on photosynthetic competence of peach and nectarine leaves infected with *Xap*, have shown that the net assimilation rate declines exponentially with respect to increasing disease severity (*unpublished*). Because of that, the effects of disease severity and leaf age are likely confounded and would require further investigation to separate. Moreover, cultivar was related to the progression of leaf abscission. Leaves of the cultivars Sweet Dream and Easternglo abscised slower than those of Beekman and Snow King (Table 4-6). The observation that leaves from Beekman abscised earlier than those of Sweet Dream and Easternglo was surprising given that Beekman is generally

considered less susceptible than the other cultivars. Although a second year of data on leaf abscission is required to confirm this observation, the present data suggest that cultivar susceptibility to infection is not a good indicator of the survival of leaves infected with *Xap*.

Survival analysis of time to defoliation was used as result of the observation of severe defoliation occurring in the FREC block in the 2008 season. This analysis quantifies previous qualitative observations of bacterial spot epidemics related to severe defoliation (Battilani et al., 1999; Ritchie, 1995; du Plessis, 1988). It proved to be a useful tool for assessing the effect of leaf age, disease severity at disease onset, treatment, and cultivar. The KM estimates provided preliminary results of factors associated with defoliation in which were again tested with the AFT model. The AFT model was able to account for continuous variables, such as initial disease onset, while the KM estimates could only model the effects of categorical variables (Allison, 1995; Collet, 2003; Scherm and Ojiambo, 2004). Survival analysis also helped explain some of the discrepancies found in the disease progress data, implying rates of disease progress obtained from temporal diseases analysis were probably low due to loss of heavily diseased leaves. Because of the profound effects of defoliation on bacterial spot epidemics, it is suggested that bacterial spot should not be assessed without considering this important symptom. In addition, future quantitative epidemiological studies on temporal progression of bacterial spot epidemics should develop or evaluate other statistical models that account for sinusoidal curves such as those observed in this study. Examples of such models might include polynomial regression models.

## LITERATURE CITED

- Afifi, A.A., Clark, V., and May, S. 2004. Computer-Aided Multivariate Analysis, 4<sup>th</sup> ed. Boca Raton, FL, USA: Chapman & Hall/CRC.
- Akaike, H. 1974. A new look at the statistical model identification. Institute of Electrical and Electronics Engineers Transaction and Automatic Control AC-19:716-23.
- Allison, P.D. 1995. Survival Analysis Using the SAS System. A Practical Guide. Cary, NC, USA: SAS Institute.
- Battilani, P., Rossi, V., and Saccardi, A. 1999. Development of *Xanthomonas arboricola* pv. *pruni* epidemics on peaches. J. Plant Pathol. 81: 161-171.
- Civerolo, E.L. 1975. Quantitative aspects of pathogenesis of *Xanthomonas pruni* in peach leaves. Phytopathology 65:258-264.
- Collett, D. 2003. Modelling Survival Data in Medical Research, 2<sup>nd</sup> ed. Boca Raton, FL, USA: CRC Press.
- du Plessis, H.J. 1988. Differential virulence of *Xanthomonas campestris* pv. *pruni* to peach, plum, and apricot cultivars. Phytopathology 78:1312-1315.
- Esker, P.D., Gibb, K.S., Padovan, A., Dixon, P.M., Nutter Jr, F.W. 2006. Use of survival analysis to determine the postincubation time-to-death of papaya due to yellow crinkle disease in Australia. Plant Disease 90:102-107.
- Gilligan, C.A. 1990. Comparison of disease progress curves. New Phytol. 115:223-242.
- Kaplan, E.L. and Meier, P. 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53:457-481.
- Kolnaar, R.W. and van den Bosch, F. 2001. Effect of temperature of epidemiological parameters of *Puccinia lagenophorae*. Plant Pathology 50:363-370.

- Lalancette, N. and McFarland, K.A. 2007. Phytotoxicity of copper-based bactericides to peach and nectarine. *Plant Disease* 91:1122-1130.
- Madden, L.V. and Nault, L.R. 1983. Differential pathogenicity of corn stunting mollicutes to leafhopper vectors in *Dalbulus* and *Balbulus* species. *Phytopathology* 73:1608-14.
- Madden, L.V. 1986. Statistical analysis and comparison of disease progress curves. In: Leonard, K.J., Fry, W.E., eds. *Plant Disease Epidemiology, Vol 1: Population Dynamics and Management*. New York, USA: Macmillan, 55-84.
- Madden, L.V., Hughes, G., van den Bosch, F. 2007. *The Study of Plant Disease Epidemics*. The American Phytopathological Society, St. Paul.
- Montgomery, D.C. and Peck, E.A. 1992. *Introduction to Linear Regression Analysis*. 2 ed, New York, USA: John Wiley & Sons, Inc.
- Muenchow, G. 1986. Ecological use of failure time analysis. *Ecology* 67:246-50.
- Ngugi, H.K., Julian, A.M., King, S.B., and Peacocke, B.J. 2000. Epidemiology of sorghum anthracnose (*Colletotrichum sublineolum*) and leaf blight (*Exserohilum turcicum*) in Kenya. *Plant Pathology* 49:129-140.
- Ojiambo, P.S. and Scherm, H. 2005. Survival analysis of time to abscission of blueberry leaves affected by *Septoria* leaf spot. *Phytopathology* 95:108-113.
- Pethybridge, S.J., Ngugi, H.K., and Hay, F.S. 2010. Use of survival analysis to assess management options for ray blight in Australian pyrethrum fields. *Plant Pathology* 59:480-491.
- Ritchie, D.F. 1995. Bacterial spot. Pages 50-52 in: *Compendium of Stone Fruit Diseases*. J. M. Ogawa, E.I. Zehr, G.W. Bird, D.F. Ritchie, K. Uriu, and J.K. Uyemoto, eds. American Phytopathological Society, St. Paul, MN.

- Scherin, H. and Ojiambo, P.S. 2004. Applications of survival analysis in botanical epidemiology. *Phytopathology* 94:1022-1026.
- Westra, A.A.G., Arneson, C.P., and Slack, S.A. 1994. Effect of interaction of inoculum dose, cultivar, and geographic location on the development of foliar symptoms of bacterial ring rot of potato. *Phytopathology* 84:410-415.
- Zaccardelli, M., Malaguti, S., and Bazzi, C. 1998. Biological and epidemiological aspects of *Xanthomonas arboricola* pv. *pruni* on peach in Italy. *Journal of Plant Pathology* 80:125-132.
- Zehr, E.I., Shepard, D.P., and Bridges, W.C., Jr. 1996. Bacterial spot of peach as influenced by water congestion, leaf wetness duration, and temperature. *Plant Dis.* 80:339-341.

**Table 4-1.** Comparison of five probability distributions tested for goodness-of-fit in describing time-to-defoliation of leaves displaying symptoms of bacterial spot of peach and nectarine (caused by *Xanthomonas arboricola* pv. *pruni*)

| Model                        | Akaike information criterion (AIC) <sup>a</sup> |          |              |            |         |
|------------------------------|---|----------|--------------|------------|---------|
|                              | Exponential                                     | Logistic | Log-logistic | Log-normal | Weibull |
| Treatment                    | 6,490.3   | 17,267.7 | 6,349.9      | 6,328.3    | 6,382.9 |
| Cultivar                     | 6,395.2   | 17,164.0 | 6,266.4      | 6,268.2    | 6,274.0 |
| Age                          | 5,707.8   | 16,368.4 | 5,498.9      | 5,463.8    | 5,596.2 |
| Initial disease              | 6,130.9   | 15,415.5 | 5,584.4      | 5,628.5    | 5,947.9 |
| Cultivar+Age+Initial disease | 5,306.4   | 14,632.4 | 4,771.7      | 4,818.2    | 5,077.3 |

<sup>a</sup> AIC =  $-2(\text{Log-likelihood}) + 2k$ , where  $2k$  accounts for the number of parameters. The smallest values of AIC indicate the simplest model with the best fit.

**Table 4-2.** Mean estimates of initial disease onset for bacterial spot (caused by *Xanthomonas aboricola* pv. *pruni*) disease progress curves from the 2008 growing season

| Treatment and rate   | Easternglo          | Beekma<br>n | Snow<br>King | Sweet<br>Dream | Treatment<br>Mean |
|--|---------------------|-------------|--------------|----------------|-------------------|
| 1. Mycoshield 17 WP 12 oz                                      | 20.29 <sup>ab</sup> | 14.25       | 15.41        | 13.4           | 15.84             |
| 2. Kocide 3000 30 DF 3.334 oz                                  | 12.73               | 17.43       | 14.00        | 10.46          | 13.66             |
| 3. Kocide 3000 30DF 3.334 oz + 0.1% solution of Garlic Product | 11.07               | 15.02       | 10.87        | 10.38          | 11.84             |
| 4. Garlic Product 0.1% solution                                | 15.2                | 15.25       | 14.84        | 11.37          | 14.17             |
| 5. Kasumin 2L 2 qts + Captan 80 WDG 3lb                        | 16.04               | 13.38       | 9.01         | 13.9           | 13.08             |
| 6. Kasumin 2L 2 qts + Kocide 3000 30 DF 3.334 oz               | 23.04               | 19.68       | 10.98        | 18.53          | 18.06             |
| 7. Untreated Check   | 10.73               | 16.47       | 9.22         | 15.8           | 13.05             |
| Cultivar mean  | 14.93               | 15.19       | 12.38        | 12.84          |                   |

<sup>a</sup> Values are means based on four replicate single tree plots. Disease severity was based on visual assessment using the 1-7 ordinal rating scale (1 = 0% lesion area, and 7 = >45%) (Chapter 2) every 5 to 7 days throughout the growing season.

<sup>b</sup> Means in the same column are not significantly different as indicated by the results of the analysis of variance,  $P < 0.05$  (calculated in GenStat 12, VSN International, Hemel Hempstead, UK).

**Table 4-3.** Mean estimates of initial disease onset for bacterial spot (caused by *Xanthomonas aboricola* pv. *pruni*) disease progress curves from the 2009 growing season

| Treatment and rate  | Easternglo           | Beekman | Snow King | Sweet Dream | Treatment Mean |
|---|----------------------|---------|-----------|-------------|----------------|
| 1. Mycoshield 17 WP 12 oz                                       | 7.73bc <sup>ab</sup> | 18.36a  | 13.8a     | 12.05a      | 12.99a         |
| 2. Kocide 3000 DF 1.667 oz                                      | 10.75ab              | 2.74c   | 9.13a     | 7.54abc     | 7.54ab         |
| 3. Kocide 3000 30 DF 1.667 oz + 0.1% solution of Garlic Product | 3.55c                | 9.44b   | 7.97a     | 12.06a      | 8.25ab         |
| 4. Garlic Product 0.1% solution                                 | 15.19a               | 10.2b   | 10.62a    | 11.18ab     | 11.8a          |
| 5. Kasumin 2 qts/100 gal + Captan 80 WDG 31lb                   | 2.01c                | 5.17bc  | 10.11a    | 3.41c       | 5.17b          |
| 6. Kasumin 2 qts/100 gal + Kocide 3000 DF 1.667 oz              | 1.49c                | 4.92bc  | 10.22a    | 5.46bc      | 5.52b          |
| 7. Untreated Check  | 9.57abc              | 10.81b  | 7.91a     | 3.92c       | 8.05ab         |
| Cultivar mean   | 7.18                 | 8.8     | 9.97      | 7.95        |                |

<sup>a</sup> Values are means based on four replicate single tree plots. Disease severity was based on visual assessment using the 1-7 ordinal rating scale (1 = 0% lesion area, and 7 = >45%) (Chapter 2) every 5 to 7 days throughout the growing season.

<sup>b</sup> Means in the same column followed by the same letter(s) are not significantly different according to least significant differences test ( $\alpha = 0.05$ ) calculated in GenStat 12 (VSN International, Hemel Hempstead, UK).

**Table 4-4.** Mean estimates of rate of disease progression for bacterial spot (caused by *Xanthomonas aboricola* pv. *pruni*) disease progress curves from the 2008 growing season

| Treatment and rate   | Easternglo         | Beekman | Snow King | Sweet Dream | Treatment Mean |
|--|--------------------|---------|-----------|-------------|----------------|
| 1. Mycoshield 17 WP 12 oz                                      | 1.82 <sup>ab</sup> | 1.468   | 2.246     | 2.176       | 1.927b         |
| 2. Kocide 3000 30 DF 3.334 oz                                  | 2.652              | 2.367   | 2.835     | 3.01        | 2.716a         |
| 3. Kocide 3000 30DF 3.334 oz + 0.1% solution of Garlic Product | 2.635              | 2.258   | 2.599     | 2.767       | 2.565a         |
| 4. Garlic Product 0.1% solution                                | 1.836              | 1.508   | 2.312     | 2.434       | 2.023b         |
| 5. Kasumin 2L 2 qts + Captan 80 WDG 3lb                        | 1.96               | 1.859   | 2.536     | 2.394       | 2.187b         |
| 6. Kasumin 2L 2 qts + Kocide 3000 30 DF 3.334 oz               | 2.443              | 2.346   | 2.444     | 2.651       | 2.471a         |
| 7. Untreated Check   | 1.8                | 1.822   | 2.338     | 2.37        | 2.083b         |
| Cultivar mean  | 2.134b             | 2.015b  | 2.48a     | 2.557a      |                |

<sup>a</sup> Values are means based on four replicate single tree plots. Disease severity was based on visual assessment using the 1-7 ordinal rating scale (1 = 0% lesion area, and 7 = >45%) (Chapter 2) every 5 to 7 days throughout the growing season.

<sup>b</sup> Means in the same column and the 'cultivar mean' row followed by the same letter(s) are not significantly different according to least significant differences test ( $\alpha = 0.05$ ) calculated in GenStat 12 (VSN International, Hemel Hempstead, UK).

**Table 4-5.** Mean estimates of rate of disease progression for bacterial spot (caused by *Xanthomonas aboricola* pv. *pruni*) disease progress curves from the 2009 growing season

| Treatment and rate  | Easternglo          | Beekman | Snow King | Sweet Dream | Treatment Mean |
|---|---------------------|---------|-----------|-------------|----------------|
| 1. Mycoshield 17 WP 12 oz                                       | 2.037 <sup>ab</sup> | 1.776   | 2.158     | 2.192       | 2.041cd        |
| 2. Kocide 3000 DF 1.667 oz                                      | 2.284               | 2.354   | 2.455     | 2.406       | 2.375b         |
| 3. Kocide 3000 30 DF 1.667 oz + 0.1% solution of Garlic Product | 2.258               | 2.045   | 2.445     | 2.327       | 2.269b         |
| 4. Garlic Product 0.1% solution                                 | 1.993               | 1.81    | 2.508     | 2.336       | 2.162bc        |
| 5. Kasumin 2 qts/100 gal + Captan 80 WDG 31lb                   | 2.242               | 2.221   | 2.516     | 2.241       | 2.305b         |
| 6. Kasumin 2 qts/100 gal + Kocide 3000 DF 1.667 oz              | 2.865               | 2.652   | 2.548     | 2.829       | 2.723a         |
| 7. Untreated Check  | 1.631               | 1.632   | 2.334     | 2.118       | 1.929d         |
| Cultivar mean   | 2.187bc             | 2.07c   | 2.423a    | 2.35ab      |                |

<sup>a</sup> Values are means based on four replicate single tree plots. Disease severity was based on visual assessment using the 1-7 ordinal rating scale (1 = 0% lesion area, and 7 = >45%) (Chapter 2) every 5 to 7 days throughout the growing season.

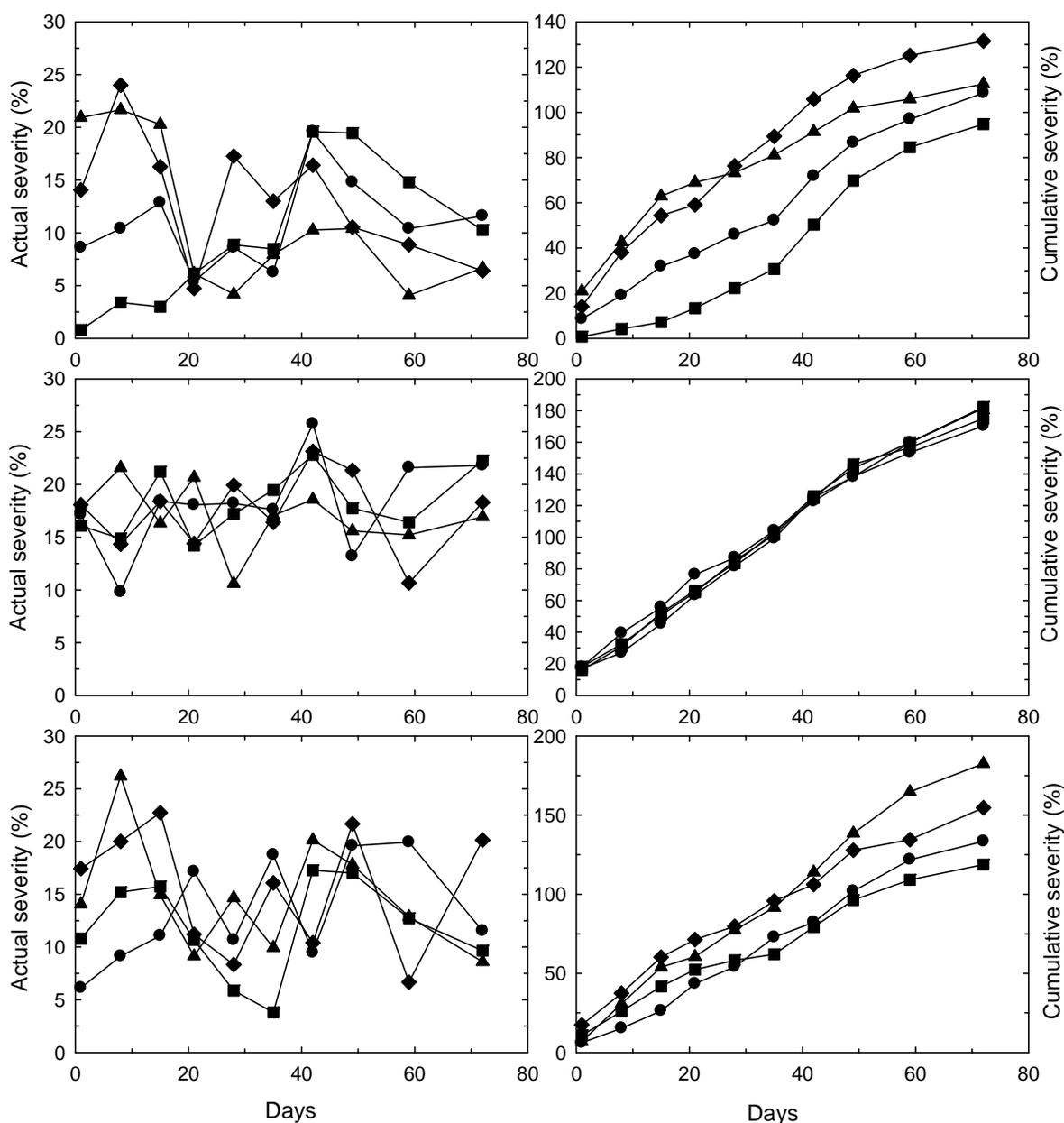
<sup>b</sup> Means in the same column and the 'cultivar mean' row followed by the same letter(s) are not significantly different according to least significant differences test ( $\alpha = 0.05$ ) calculated in GenStat 12 (VSN International, Hemel Hempstead, UK).

**Table 4-6.** Parameter estimates and associated test statistics for the accelerated failure time (AFT) models relating cultivar, leaf age, and initial disease onset with time to defoliation ( $T$ ) in peach and nectarine affected by bacterial spot (*Xanthomonas arboricola* pv. *pruni*) in Pennsylvania in the 2009 season

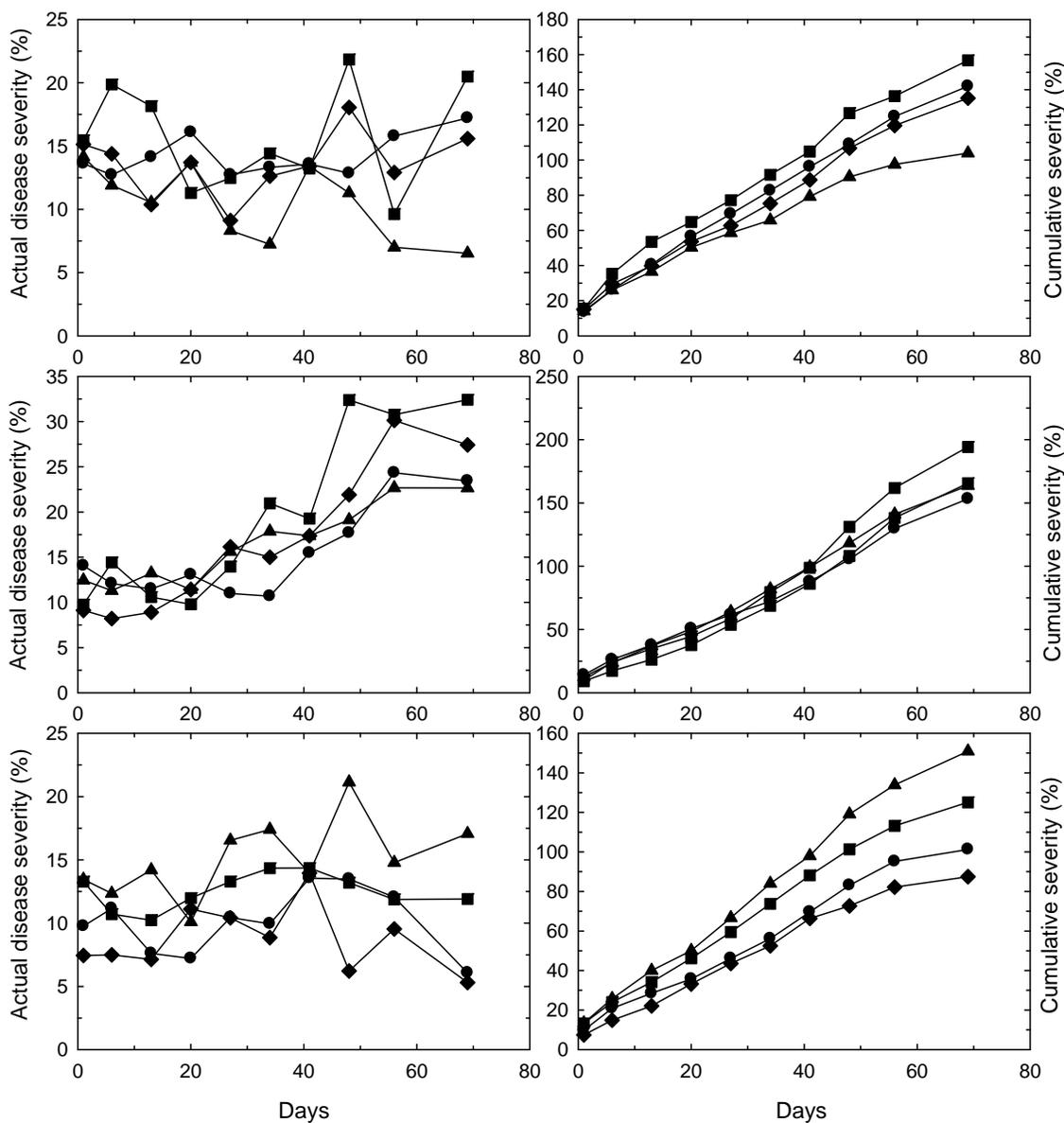
| Parameter             | d.f. | Estimate <sup>a</sup> | Standard Error | 95% Confidence limits |        | $\chi^2$ | P > $\chi^2$ | T (days), mean $\pm$ standard error <sup>b</sup> |
|-----------------------|------|-----------------------|----------------|-----------------------|--------|----------|--------------|--|
| Intercept             | 1    | 5.768                 | 0.0901         | 5.591                 | 5.945  | 4093.82  | <0.0001      | .....  |
| Cultivar Easternglo   | 1    | -0.114                | 0.0576         | -0.227                | -0.001 | 3.90     | 0.0483       | 54.9 $\pm$ 0.88                                  |
| Cultivar Beekman      | 1    | -0.538                | 0.055          | -0.645                | -0.430 | 95.70    | <0.0001      | 47.7 $\pm$ 0.92                                  |
| Cultivar Snow King    | 1    | -0.480                | 0.0531         | -0.584                | -0.376 | 81.83    | <0.0001      | 48.01 $\pm$ 0.81                                 |
| Cultivar Sweet Dream  | 0    | 0.000                 | .....          | .....                 | .....  | .....    | .....        | 54.8 $\pm$ 0.83                                  |
| Leaf Age: Middle      | 1    | -0.182                | 0.0878         | -0.354                | -0.010 | 4.28     | 0.0385       | 56.3 $\pm$ 0.30                                  |
| Leaf Age: Old         | 1    | -1.300                | 0.0797         | -1.456                | -1.144 | 265.80   | <0.0001      | 41.8 $\pm$ 0.58                                  |
| Leaf Age: Young       | 0    | 0.000                 | .....          | .....                 | .....  | .....    | .....        | 45.1 $\pm$ 0.19                                  |
| Initial Disease Onset | 1    | -0.114                | 0.0046         | -0.123                | -0.105 | 606.88   | <0.0001      | .....  |
| Scale                 | 1    | 0.861                 | 0.0177         | 0.827                 | 0.896  | .....    | .....        | .....  |

<sup>a</sup> Negative values indicate the cultivar and leaf age defoliated faster than the highest level in each variable (e.g. cultivar Sweet Dream among cultivar comparisons).

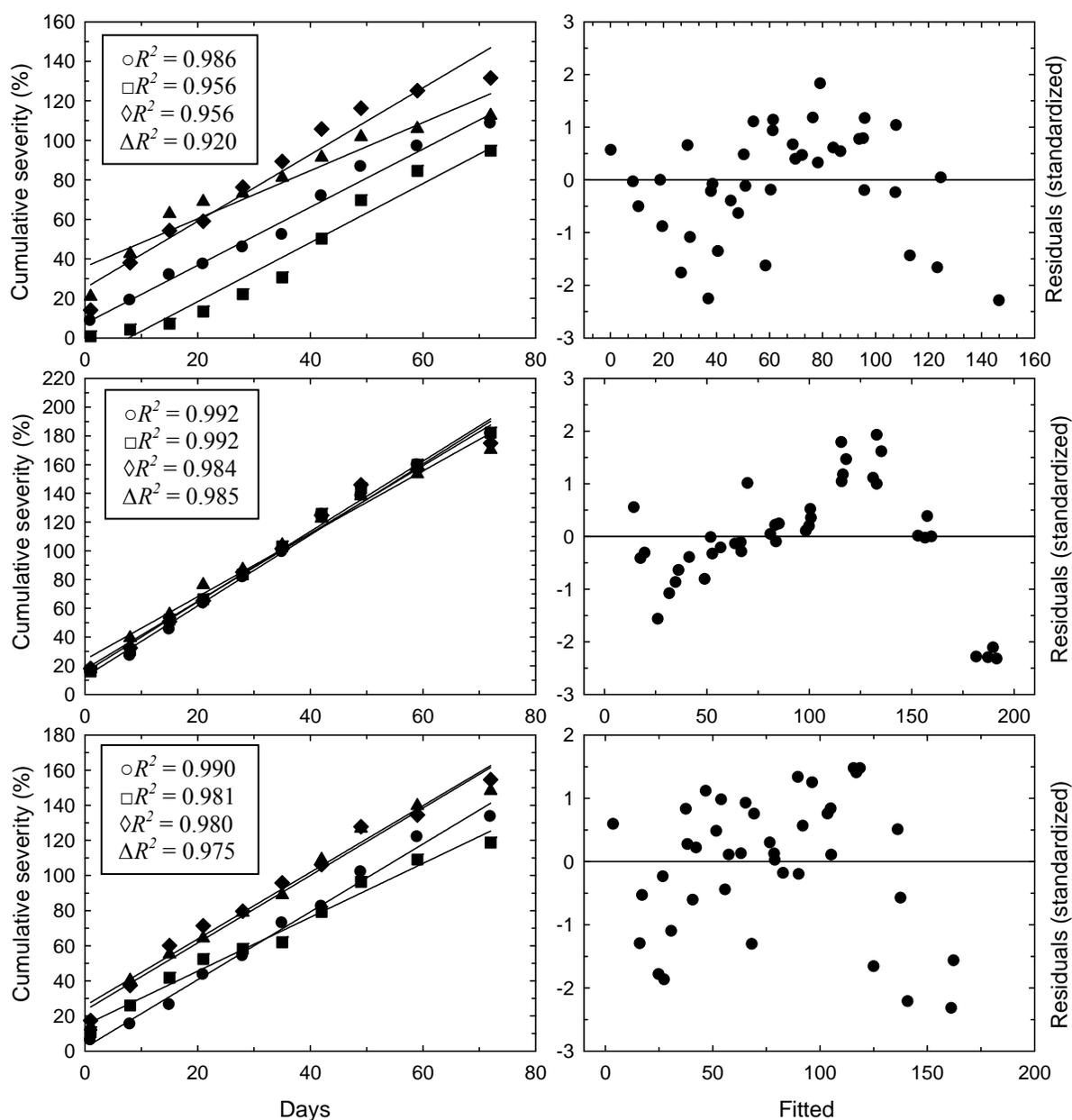
<sup>b</sup> Values are the means and standard errors of the median survival times obtained from the FREC block.



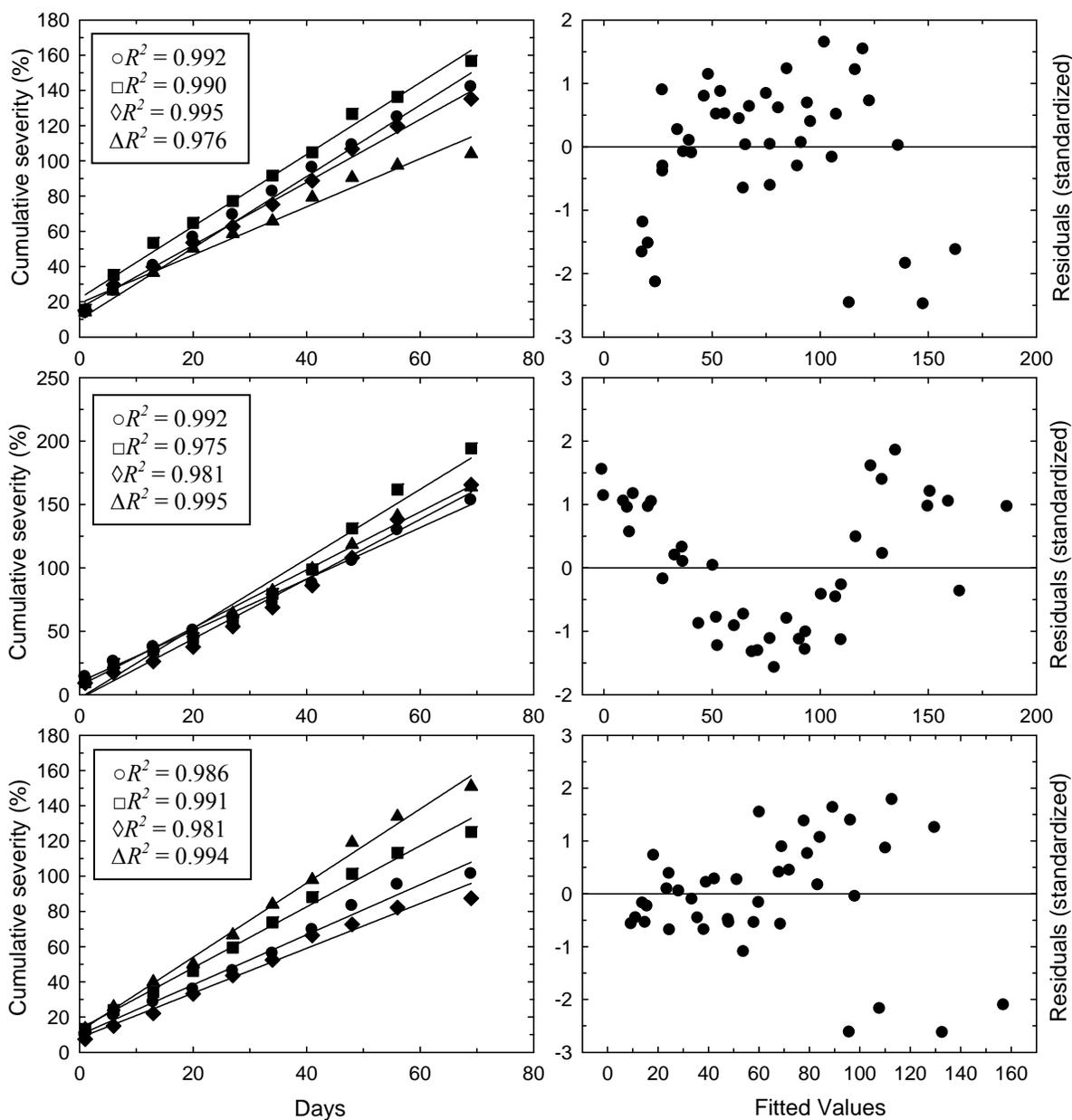
**Fig 4-1.** Actual bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) disease progress (left) and cumulative disease progress (right) on the peach cultivar Beekman in response to the treatments Mycoshield (top), Kocide 3000 (middle), and untreated check (bottom) in the 2008 growing season. Disease severity was assessed with the 1-7 ordinal rating scale (1 = 0% lesion area, and 7 = >45%) (Chapter 2) every 5 to 7 days throughout the 2008 growing season starting on day 0 (3 June 2008). Different lines represent the 4 replicate trees in each treatment (Replicate 1: circle; Replicate 2: square; Replicate 3: diamond; Replicate 4: triangle).



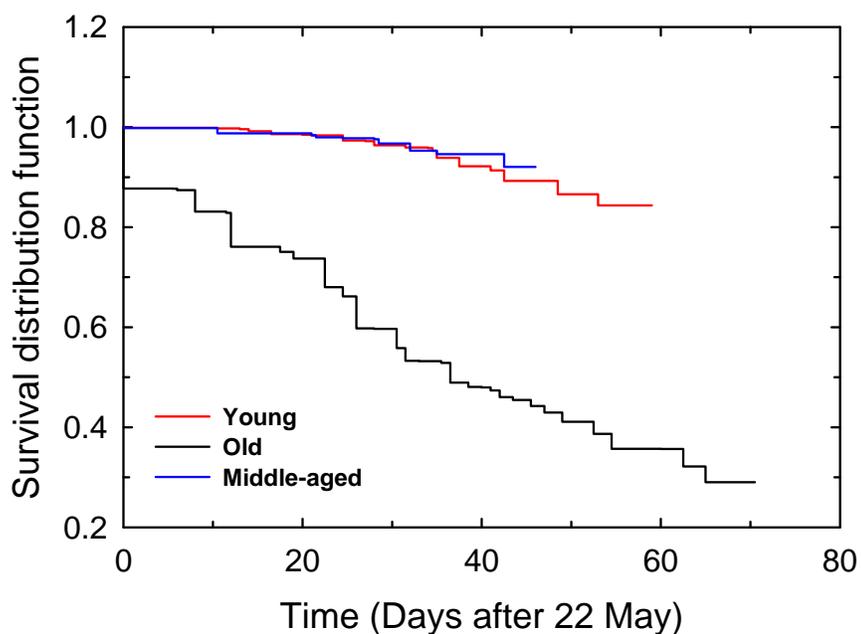
**Fig 4-2.** Actual bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) disease progress (left) and cumulative disease progress (right) on the peach cultivar Beekman in response to the treatments Mycoshield (top), Kocide 3000 (middle), and untreated check (bottom) in the 2009 growing season. Disease severity was assessed with the 1-7 ordinal rating scale (1 = 0% lesion area, and 7 = >45%) (Chapter 2) every 5 to 7 days throughout the 2009 growing season starting on day 0 (22 May 2009). Different lines represent the 4 replicate trees in each treatment (Replicate 1: circle; Replicate 2: square; Replicate 3: diamond; Replicate 4: triangle).



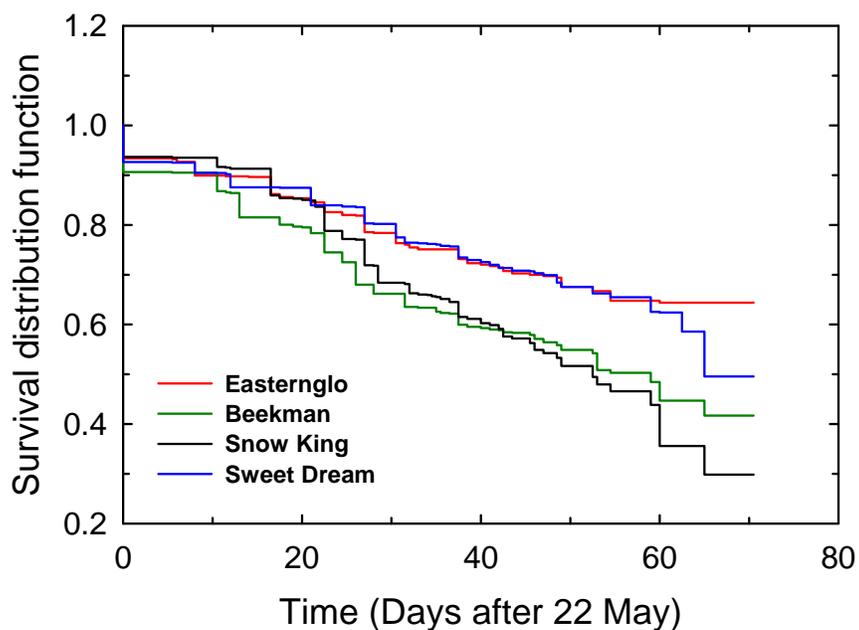
**Fig 4-3.** Fitted cumulative bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) disease progress (left) and associated standardized residuals (right) on the peach cultivar Beekman in response to the treatments Mycoshield (top), Kocide 3000 (middle), and untreated check (bottom) in the 2008 growing season. Straight lines represent fitted curves,  $R^2$  values indicate the proportion of variability of Y that is explained by the cumulative response model. Disease severity was assessed with the 1-7 ordinal rating scale (1 = 0% lesion area, and 7 = >45%) (Chapter 2) every 5 to 7 days throughout the 2008 growing season starting on day 0 (3 June 2008). Different lines represent the 4 replicate trees in each treatment (Replicate 1: circle; Replicate 2: square; Replicate 3: diamond; Replicate 4: triangle).



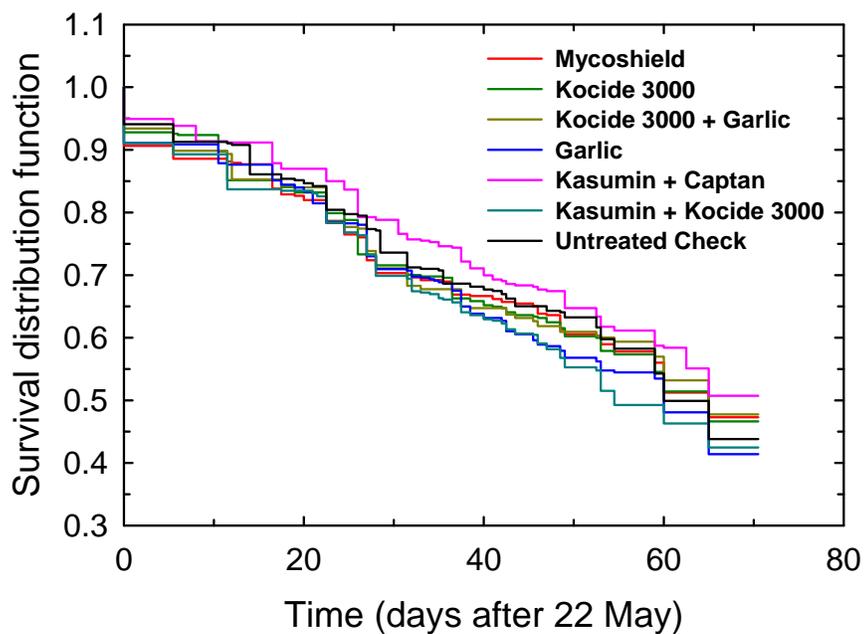
**Fig 4-4.** Fitted cumulative bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) disease progress (left) and associated standardized residuals (right) on the peach cultivar Beekman in response to the treatments Mycoshield (top), Kocide 3000 (middle), and untreated check (bottom) in the 2009 growing season. Straight lines represent fitted curves,  $R^2$  values indicate the proportion of variability of Y that is explained by the cumulative response model. Disease severity was assessed with the 1-7 ordinal rating scale (1 = 0% lesion area, and 7 = >45%) (Chapter 2) every 5 to 7 days throughout the 2009 growing season starting on day 0 (22 May 2009). Different lines represent the 4 replicate trees in each treatment (Replicate 1: circle; Replicate 2: square; Replicate 3: diamond; Replicate 4: triangle).



**Fig 4-5.** Kaplan-Meier estimates of survival functions indicating time to abscission of peach and nectarine leaves of the cultivars Easternglo, Beekman, Snow King, and Sweet Dream with foliar symptoms of bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) ( $n = 3,052$ ). Leaves were classified according to leaf age into three categories: young, middle-aged, and old based on timing of leaf emergence. Old leaves abscised quicker than did young and middle aged leaves.



**Fig 4-6.** Kaplan-Meier estimates of survival functions indicating time to abscission of peach and nectarine leaves of the cultivars Easternnglo, Beekman, Snow King, and Sweet Dream with foliar symptoms of bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) ( $n = 3,052$ ). Leaves were classified according to cultivar into four categories: Easternnglo, Beekman, Snow King, and Sweet Dream. Leaves of the cultivars Beekman and Snow King abscised faster than those of the cultivars Easternnglo and Sweet Dream.



**Fig 4-7.** Kaplan-Meier estimates of survival functions indicating time to abscission of peach and nectarine leaves of the cultivars Easternnglo, Beekman, Snow King, and Sweet Dream with foliar symptoms of bacterial spot (caused by *Xanthomonas arboricola* pv. *pruni*) ( $n = 3,052$ ). Leaves were classified according to treatment received in the product evaluation trial into seven categories: Mycoshield, Kocide 3000, Kocide 3000 + Garlic, Kasumin + Captan, Kasumin + Kocide 3000, untreated check, based on treatment received in the field. Treatment did not significantly affect timing of leaf abscission.

## **Chapter 5**

### **Discussion**

This research focuses on the epidemiology and management of bacterial spot of peach and nectarine, a particularly devastating disease in the eastern US. First, the reliability and accuracy of the 1-7 rating scale (1 = 0% lesion area, and 7 = >45%) used to assess bacterial spot severity was evaluated and compared to the direct estimation method. Second, the accuracy and reliability of estimates of bacterial spot severity made by experienced and inexperienced raters were compared. Lastly, the rate of disease progress and initial disease onset for bacterial spot epidemics on cultivars subjected to different bactericide treatments were modeled and compared. In addition, a survival analysis of time to abscission was conducted on peach and nectarine leaves affected by bacterial spot to determine factors that influence the survival of diseased leaves.

Studies on the epidemiology of any disease depend on the quality of data collected for statistical assessment. The 1 to 7 rating scale developed and evaluated in Chapter 2 had not previously been used or assessed in any other study on bacterial spot. Indeed, similar rating scales had been used in studies on bacterial spot but only one other rating scale had been assessed for accuracy and reliability (Citadin et al., 2008). The results of Chapter 2 indicated that the direct estimation method was superior to the rating scale in assessing bacterial spot severity. However, because the studies were conducted concordantly, the 1-7 rating scale was used to assess foliar disease severity in Chapter 4. Although data based on direct estimation would have been more accurate and reliable, the comparisons made among disease progress curves in this study remain valid. Indeed, numerous epidemiological studies have used rating scales, including the much disputed

Horsfall-Barratt (H-B) scale that has been cited well over 500 times in the past 70 years (Bock, et al., 2010). Nevertheless, future studies on the epidemiology of bacterial spot should utilize the direct estimation method to obtain estimates of foliar disease severity.

Most of the literature comparing disease estimates made by experienced and inexperienced raters indicates that estimates of disease severity made by inexperienced raters were less reliable and accurate than those made by experienced raters. Literature also suggests that extended periods of disease assessment training were needed to improve levels of accuracy and reliability of estimates made by inexperienced raters. Therefore, we were hesitant to enlist any inexperienced employees to assess bacterial spot severity for fear the quality of the data would greatly suffer. However, the results of Chapter 3 reveal that, with one short period of computer-based training, disease assessment of bacterial spot could be made, without the loss of accuracy and reliability, by an inexperienced rater. Results of this study differ from other studies perhaps due to the training given directly prior to disease assessment where as in other studies; training was given only after the initial assessments of disease were made. Still, there are factors in the field that may hinder the accuracy of bacterial spot assessment. Foliar injuries caused by Captan and copper spray applications may be confused for bacterial spot symptoms by experienced and inexperienced raters, alike. Only continued field training may help reduce this assessment error.

Throughout Chapters 2, 3, and 4, some of the statistical methods utilized in the studies are not common in plant pathology literature. For example, in Chapter 2, a categorical data analysis approach that included such statistics as the Gamma, Kendall's  $\tau$ , Somer's, Stuart's and Cohen's kappa coefficients assessed the likelihood that a

single observational unit (i.e., a diseased leaf) is assigned the same score in two or more assessments (Svensson, 2000) has not been used in studies on the accuracy and reliability of disease assessments. The use of categorical analysis methods indicated improved levels of accuracy and reliability for estimates made with the rating scale although this was not sufficient to overcome the limitations associated with the scale. In addition to the categorical analyses methods, the bootstrap analysis used in Chapter 2 statistically confirmed that the direct estimation method was significantly more accurate than the rating scale where other studies had merely qualitatively compared the two. Moreover, the bootstrap analysis performed in Chapter 3 permitted the computation of 95% confidence intervals around the mean differences of the agreement statistics for the experienced and inexperienced raters. Because the 95% confidence intervals for the differences in values obtained from the two groups included zero, there was no sufficient evidence to reject the null hypothesis, implying the two groups were not significantly different from each other with respect to the accuracy and reliability of disease severity estimates. As yet, no other phytopathological study on disease assessment has based their conclusions on such statistics. The survival analysis methods used in Chapter 4 are also generally novel. Only a few botanical studies have ever used the analysis of time to leaf abscission (Madden and Nault, 1983; Westra et al., 1994; Scherm and Ojiambo, 2004; Ojiambo and Scherm, 2005; Esker et al., 2006; Pethybridge et al., 2010) where such an analysis is primarily used in clinical studies of time to death of human patients. This approach enabled us to identify and quantify the factors associated with tree defoliation, an important symptom of bacterial spot of stone fruits. To the best of our knowledge, this is the first study to model and identify such factors for the peach-X.

*arboricola* pv. *pruni* pathosystem. Knowledge of the factors that are associated with premature defoliation such as leaf age, cultivar, and initial disease onset, will provide a base in which management of bacterial spot may be improved. Furthermore, such statistics mentioned above set these studies apart and provide stronger evidence on which to support such conclusions.

There was much difficulty developing a mathematical model to describe the disease progress of bacterial spot epidemics. While  $R^2$  values were high for models based on cumulative disease severity values indicating a good fit to the data and residual plots passed the normality test, discernable systematic patterns of some of the residual plots could not be ignored. These patterns suggested the presence of autocorrelation, a serious violation of the basic assumptions of linear regression analysis (Montgomery and Peck, 1992). Such difficulty fitting a model to disease progress curves may explain why such a study has not previously been completed for bacterial spot of stone fruits. Although this issue was not further investigated, it is clear that a further examination of this topic is required to fully resolve modeling of temporal bacterial spot progress. For now, however, the model based on cumulative disease severity will satisfy the needs of the current study which is to compare cultivar and bactericide effects.

The data from field trials provided a good opportunity to model the effects of bactericide treatment, cultivar, and the interaction between the two factors on the disease severity at initial disease onset and the rate of disease progress of bacterial spot. Taken together with the results of survival analysis, the data presented in Chapter 4 indicates that bactericide treatments and cultivars affect the rate of bacterial spot progress. That is, the rate parameter from the disease progress curves is the parameter consistently

modified by bactericides and host response to infection. The results also suggested that cultivars may affect the amount of overwintering inoculum but this will require further testing because an artifact in the data in the form of a significant effect due to bactericide treatment was also noted in 2009. The results also highlighted a difficulty associated with evaluating bactericides that may cause phytotoxicity such as copper treatments. Efforts should be made to better distinguish copper injury from bacterial spot symptoms in the future in order to obtain more accurate estimates of disease severity. In 2008 and 2009, the standard treatment Mycoshield and the untreated check were associated with the lowest rates of the disease progress and were not significantly different from each other. This was likely due to the premature abscission of heavily diseased leaves. Leaf abscission, therefore, likely influenced the results of disease progress.

The observation of severe defoliation therefore underscored the need to assess and analyze the progression of leaf abscission when studying bacterial spot epidemics. In this regard, the survival analysis techniques such as those used in Chapter 4 should prove useful. The survival analysis conducted through Kaplan – Meier (KM) estimates and the accelerated failure time (AFT) model, showed that all of the factors tested, except for treatment, significantly affected time to abscission. These results could heavily influence management strategies of bacterial spot in the future. For example, because higher levels of disease severity at initial disease onset are related to increased rates of abscission, control measures should be focused on reducing overwintering inoculum and therefore lowering initial levels of disease and increasing the lifespan of leaves. This would support the use of dormant copper applications, a bactericide treatment that has shown to reduce initial levels of inoculum in preliminary studies (Ngugi et al., 2009). The effects

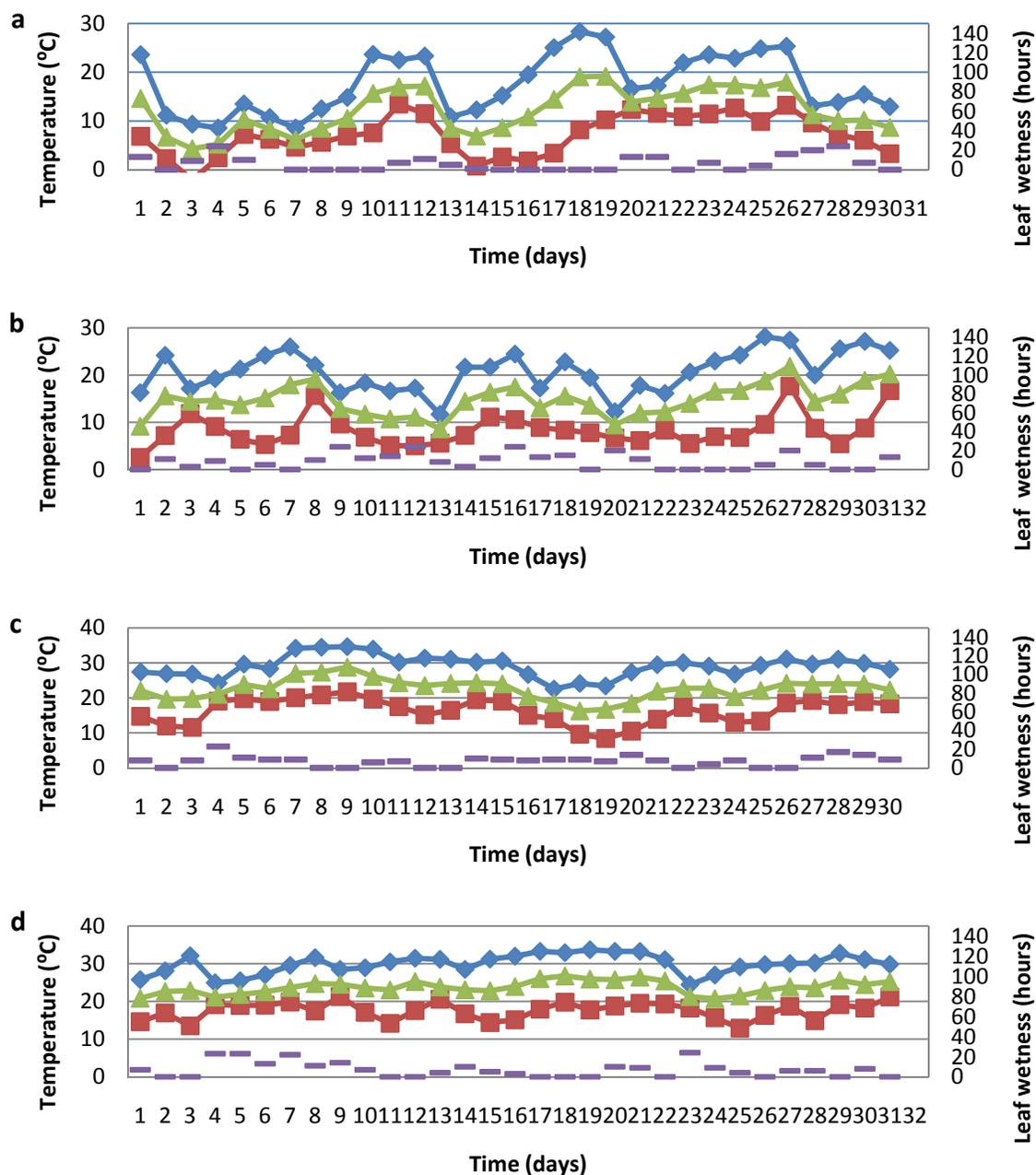
of leaf age and disease severity are difficult to separate and therefore complicate recommendations for disease management. Further evaluation is needed to determine if these factors are different from each other or are influenced by each other, and to investigate the physiological and yield implications of bacterial spot severity on peach and nectarine. The statistical methods developed and evaluated in this study should provide the tools for such a study and for other studies that require estimation of disease severity.

## LITERATURE CITED

- Bock, C.H., Poole, G.H., Parker, P.E. and Gottwald, T.R. 2010. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical Reviews in Plant Sciences*, 29:59-107.
- Citadin, I., Assmann, A.P., Mazaro, S.M., de Gouvêa, A., Danner, M.A., Malagi, G. 2008. Diagrammatic scale to evaluate the severity of bacteriose in peach trees. *Revista Brasileira de Fruticultura* 30:327-330.
- Esker, P.D., Gibb, K.S., Padovan, A., Dixon, P.M., Nutter Jr, F.W. 2006. Use of survival analysis to determine the postincubation time-to-death of papaya due to yellow crinkle disease in Australia. *Plant Disease* 90:102-107.
- Madden, L.V. and Nault, L.R. 1983. Differential pathogenicity of corn stunting mollicutes to leafhopper vectors in *Dalbulus* and *Balbulus* species. *Phytopathology* 73:1608-14.
- Montgomery, D.C. and Peck, E.A. 1992. *Introduction to Linear Regression Analysis*. 2 ed, New York, USA: John Wiley & Sons, Inc.
- Ngugi, H.K., Bardsley, S.J., and Halbrendt, N.O. 2009. Refining management strategies for bacterial spot of peach. *Pennsylvania Fruit News* Vol. 89 (2):18-24.
- Ojiambo, P.S. and Scherm, H. 2005. Survival analysis of time to abscission of blueberry leaves affected by *Septoria* leaf spot. *Phytopathology* 95:108-113.
- Pethybridge, S.J., Ngugi, H.K., and Hay, F.S. 2010. Use of survival analysis to assess management options for ray blight in Australian pyrethrum fields. *Plant Pathology* 59:480-491.

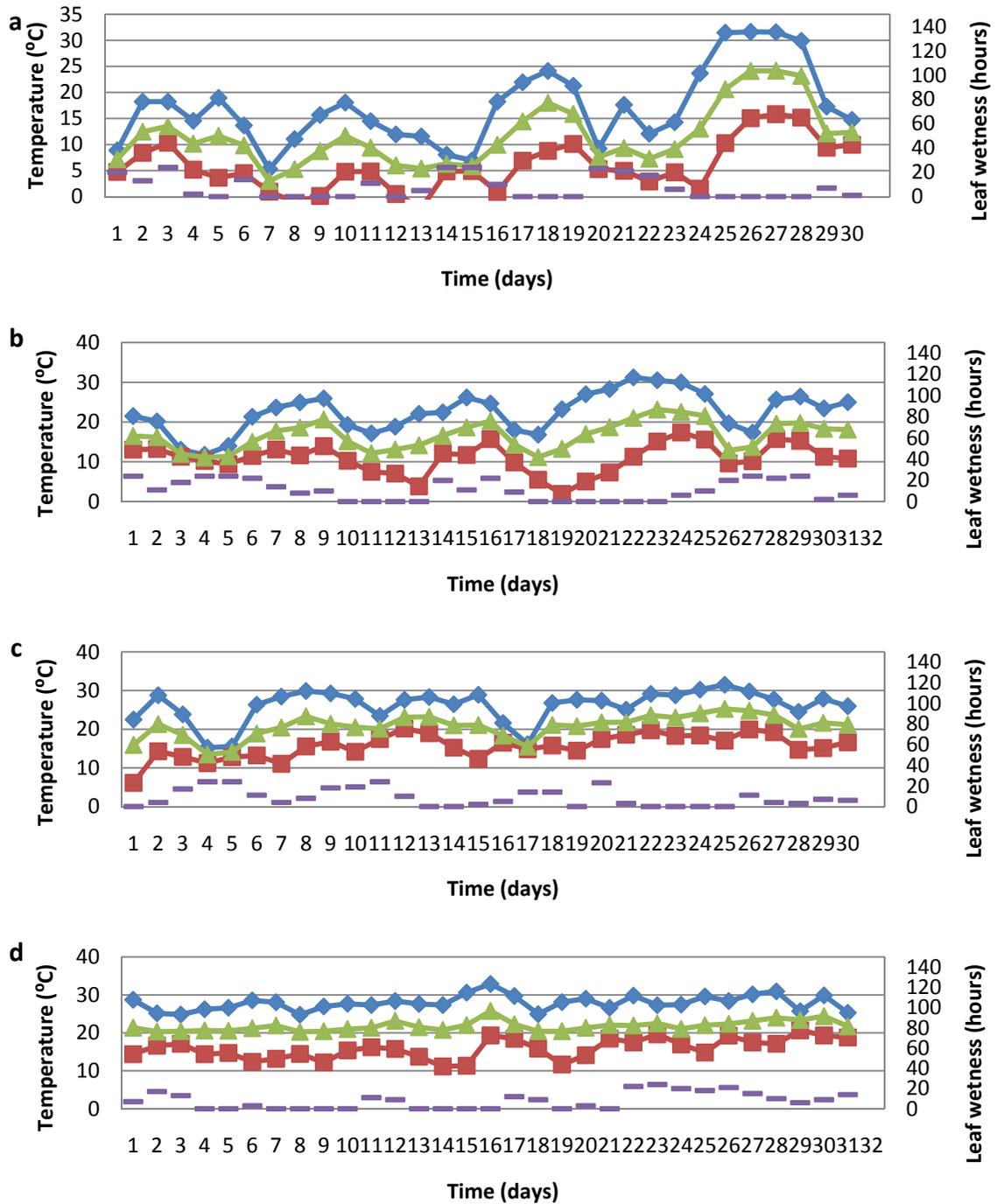
- Scherm, H. and Ojiambo, P.S. 2004. Applications of survival analysis in botanical epidemiology. *Phytopathology* 94:1022-1026.
- Svensson, E. 2000. Concordance between ratings using different scales for the same variable. *Statistics in Medicine* 19:3483-3496.
- Westra, A.A.G., Arneson, C.P., and Slack, S.A. 1994. Effect of interaction of inoculum dose, cultivar, and geographic location on the development of foliar symptoms of bacterial ring rot of potato. *Phytopathology* 84:410-415.

## APPENDIX A: Weather – 2008



**Fig A-1.** High (blue diamond), low (red square), and mean (green triangle) daily temperatures (°C) recorded at the PSU FREC in the 2008 months of April (a), May (b), June (c), and July (d). Hours of leaf wetness (purple bar) were also logged.

APPENDIX B: Weather – 2009



**Fig B-1.** High (blue diamond), low (red square), and mean (green triangle) daily temperatures (°C) recorded at the PSU FREC in the 2009 months of April (a), May (b), June (c), and July (d). Hours of leaf wetness (purple bar) were also logged.