The Influence of Soil Environment on the Incidence of Sorghum Downy Mildew: A Principal Component Analysis

W. Schuh, M. J. Jeger, and R. A. Frederiksen

Former graduate student, associate professor, and professor, Department of Plant Pathology and Microbiology, Texas Agricultural Experiment Station, College Station 77843.
Present address of first author: Department of Plant Pathology, Oklahoma State University, Stillwater 74078-0285.
Accepted for publication 25 July 1986 (submitted for electronic processing).

ABSTRACT

Schuh, W., Jeger, M. J., and Frederiksen, R. A. 1987. The influence of soil environment on the incidence of sorghum downy mildew: A principal component analysis. Phytopathology 77:128-131.

The influence of weather variables on disease incidence of sorghum downy mildew at four locations and over 4 yr was investigated by using principal component analysis. Two principal components, derived from the weather data only and representing wet and dry soil conditions, were used as independent variates in regression analysis and explained 41% of the variation in disease incidence. Regression of the disease incidence on the original soil environment variables did not provide significant equations. Results of this analysis could be used to classify sorghum-growing areas, based solely on weather, for their propensity to downy mildew.

Sorghum downy mildew, caused by *Peronosclerospora sorghi* (Weston & Uppal) C. G. Shaw (12,13), is endemic in the Coastal Bend area of south Texas (3). Several major epidemics have been reported (4). Seasonal fluctuations in disease incidence are likely to be caused by weather variations, resulting in soil conditions favorable or unfavorable for infection and disease expression. Soil temperature and soil moisture have a pronounced effect on the incidence of sorghum downy mildew (1,6), but quantitative relationships between soil environment conditions and disease are not well understood.

The purpose of this study was to determine the influence of soil environment variables on the incidence of sorghum downy mildew in the field and compare these results with those obtained under controlled conditions (11). The intention is to use this information to classify sorghum-growing areas in terms of disease propensity and to recommend planting periods based solely on records of soil environment that are unfavorable for subsequent disease development.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. § 1734 solely to indicate this fact.

MATERIALS AND METHODS

A CR21 Micrologger (Campbell Scientific, Inc., Logan, UT) was used to record soil moisture (bars) and soil temperature (C). These variables were selected for further investigation based on results of controlled condition experiments (11). Soil moisture was determined using a Model CEL-WFD soil moisture block (Campbell Scientific, Inc.). Soil temperature was recorded using a 101 Thermistor (Campbell Scientific, Inc.). The soil environment variables were recorded for the years 1982–1985 at four locations where sorghum downy mildew is endemic (2): Orange Grove (O), Beeville (B), Laward (L), and Robstown (R) in the Texas Coastal Bend. Test plots at each location had a documented history of sorghum downy mildew. The daily minimum and maximum soil temperatures (24 readings per day) and mean daily soil moisture (average of 24 readings per day) were computed from the logged data. Variables were recorded and computed for 4 wk after each planting date. From these recorded variables, seven environmental variables were derived, which subsequently formed the basis for a principal component analysis: 1) number of days with wet soil in week 1 (above 1/3 bar) (S1); 2) number of days with wet soil in week 1–2 (S2); 3) number of days with wet soil in week 1–3 (S3); 4) number of days with dry soil (below −1 bar) in week 1–3 (D3); 5) number of days with a minimum soil temperature below 20 C in week 1–3 (B3); 6) number of days with minimum or maximum soil temperature between 20–30 C in week 1–3 (B23); and 7) number of days with maximum soil temperature above 30 C in week 1–3 (A3).

The incidence of sorghum downy mildew at each location in each year was assessed by using the disease reaction of the susceptible cultivar ATX399 × RTX2536. Plants were rated as either healthy or systemically infected. By using this classification system, the percent diseased plants was computed for each 6-m row. At each location, a minimum of 10 replication (6-m rows), which were uniformly distributed through the field, were used to compute the average disease incidence. At the site/year combination R831 and R832, five replications were evaluated. The average number of plants per 6-m row was 60. Principal component analysis was performed with the SAS Princomp procedure (9). Principal component analysis was used because the original variables were highly correlated (Table 1). Principal component analysis produces a set of new variables (linear combinations of the original variables) that are independent of each other and ranked according to the amount of variation accounted for. For each principal component, a score can be computed for each site × year combination included in the analysis. Full details of the technique are given by Kendall (5), Madden and Pennypacker (7), and Stynes and Veitch (14–17). The principal component scores for each site × year combination were then used as independent variates in a regression analysis, using the SAS GLM procedure (9), with disease incidence the dependent variable. Additionally, disease incidence was regressed on S1, S2, S3, D3, B3, B23, and A3, the originally derived variables. A total of 14 data sets, each consisting of the disease incidence and values for the seven variables, formed the basis for the statistical analysis.

TABLE 1. Correlations between the seven weather variables used in the principal component analysis

Variable[a]	S1	S2	S3	D3	B2	W23	A3
S1	1.00	0.92	0.83	−0.44	0.18	0.18	0.38
S2	0.92	1.00	0.95	−0.50	0.06	0.01	0.18
S3	0.83	0.95	1.00	−0.49	0.08	−0.01	0.14
D3	−0.44	−0.50	−0.49	1.00	0.00	0.03	−0.11
B2	0.01	0.06	0.08	0.00	1.00	−0.79	−0.25
W23	0.18	0.01	−0.01	0.03	−0.79	1.00	0.45
A3	0.38	0.19	0.14	−0.11	−0.25	0.45	1.00

[a] S1 = Number of days with saturated soil in week 1. S2 = Number of days with saturated soil in week 1–2. S3 = Number of days with saturated soil in week 1–3. B2 = Number of days with soil temperature below 20 C in week 1–3. B23 = Number of days with soil temperature between 20–30 C in week 1–3. A3 = Number of days with soil temperature above 30C in week 1–3. D3 = Number of days with soil moisture below −1 bar in week 1–3.

TABLE 2. Values for the seven weather parameters at four locations and 4 yr used in the principal components analysis

Locations and year[a]	S1[b]	S2	S3	D3	B2	W23	A3
B85	4	8	15	0	10	16	2
O85	1	6	13	2	11	12	0
B84	0	0	0	9	9	19	1
O84	0	0	0	21	11	15	0
L84	0	3	3	10	14	14	0
B83.1[c]	5	12	19	0	14	6	0
B83.2[c]	7	14	18	0	12	14	0
B83.3[c]	7	10	10	0	9	17	2
R83.1[c]	6	12	19	0	14	14	0
R83.2[c]	0	4	5	0	14	8	0
O83	0	0	1	2	21	4	0
L83	1	3	3	0	15	11	0
B82	1	1	1	1	8	15	0
O82	0	2	3	0	1	19	0

[a] Locations and year. B = Beeville, O = Orange Grove, L = Laward, and R = Robstown.

[b] S1 = Number of days with saturated soil in week 1. S2 = Number of days with saturated soil in week 1–2. S3 = Number of days with saturated soil in week 1–3. B2 = Number of days with soil temperature below 20 C in week 1–3. B23 = Number of days with soil temperature between 20–30 C in week 1–3. A3 = Number of days with soil temperature above 30 C in week 1–3. D3 = Number of days with soil moisture below −1 bar in week 1–3.

[c] Staggered planting dates.

RESULTS

The untransformed values for the seven weather variables (Table 2) were used to compute seven principal components (Table 3). Principal component 1 (PC1) had an eigenvalue of 3.24 and explained 46% of the variation generated by the derived variables. The eigenvalues and the percentage of the total variation accounted for were 2.04 and 29%, 0.76 and 11%, 0.66 and 9%, 0.18 and 3%, 0.11 and 1%, and 0.01 and 1% for principal components 2–7 successively.

Principal components 1–4 explained about 95% of the total variation and were selected as independent variables for the regressions analysis with disease incidence as the dependent variable.

Because principal components are a linear combination of all variables entered into the analysis, the corresponding values in the eigenvectors for each variable were used to interpret the principal components. Interpretation was based on the absolute value of each variate in the eigenvector. Principal component 1 was dominated by high positive weights on S1, S2, S3, and to a smaller extent negative weights for D3 (Table 4). Principal component 1 can be interpreted as representing the effects of wet soil conditions. A site × year combination that has wet soils will result in a high positive score on this principal component; conversely, a site × year combination that is dry will result in a low or negative score.

Principal component 2 was dominated by high positive weights for B23 and to a lesser extent A3, and high negative weight for B2. This principal component represents the effect of both favorable and unfavorable conditions for disease incidence. Therefore, scores on this principal component will influence disease incidences less because of counterbalancing effects, and made interpretation difficult.

Principal component 3 was dominated by high positive weights

TABLE 3. The principal components, as computed through the seven weather variables

Principal component (PC)[a]	Eigenvalue	Proportion[b]	Cumulative[c]
PC1[c]	3.24	0.46	0.46
PC2	2.04	0.29	0.75
PC3	0.76	0.11	0.86
PC4	0.66	0.09	0.95
PC5	0.18	0.03	0.98
PC6	0.11	0.01	0.99
PC7	0.01	0.01	1.00

[a] Principal components 1–7.

[b] Proportion of variation explained in model.

[c] Cumulative proportion of variation explained in model.

TABLE 4. The eigenvector variation as related to the seven weather variables

	Eigenvectors						
Variable[a]	PC1[b]	PC2	PC3	PC4	PC5	PC6	PC7
S1	0.53	0.02	0.17	0.17	0.36	−0.57	−0.46
S2	0.54	−0.11	−0.02	0.24	−0.11	−0.14	0.78
S3	0.52	−0.13	−0.06	0.24	−0.31	0.64	−0.38
D3	−0.34	0.08	0.52	0.77	−0.12	−0.02	0.00
B2	−0.01	−0.61	−0.44	−0.13	0.54	0.31	0.11
B23	0.09	0.65	−0.10	0.14	0.61	0.38	0.13
A3	0.21	0.40	0.71	−0.46	−0.28	0.05	0.04

[a] S1 = Number of days with saturated soil in week 1. S2 = Number of days with saturated soil in week 1–2. S3 = Number of days with saturated soil in week 1–3. B2 = Number of days with soil temperature below 20 C in week 1–3. B23 = Number of days with soil temperature between 20–30 C in week 1–3. A3 = Number of days with soil temperature above 30 C in week 1–3. D3 = Number of days with soil moisture below −1 bar in week 1–3.

[b] Principal components 1–7.

for D3, B2, and A3. This principal component represents the effect of soil conditions that are disease suppressive, either too cool (< 20 C) and dry (< −1.0 bar) or too warm (> 30 C) and dry. Either or both effects will lead to high positive scores on PC 3.

Finally, principal component 4 was dominated by high positive weights on D3, and to a lesser degree, a negative weight on A3. This principal component represents the variation caused by warm or dry soils. A site × year combination that has dry soils will result in a high positive score of this principal component; alternatively, warm soils will result in a negative score of this principal component. Variation accounted for by the remaining principal components was marginal (about 4%). Thus the dimensionality of the weather variables was effectively reduced from seven to four. Only the scores of PC1-PC4 were entered into the regression of the principal components on the disease incidence (Table 5).

When a forward selection procedure (9) for the dependent variable infection was used, the model containing the independent variables PC1 and PC4 gave the best prediction of disease incidence ($P = 0.05$); this model explained about 41% of the variation in disease incidence. When all variables (PC1-PC4) were entered into the regression model, the R^2 value rose to 0.50, as could be expected, but the validity of the model decreased ($P =$ 0.07).

When single PC scores were regressed on the disease incidence, PC1 had the highest R^2 value of 0.29, followed by PC4 and with R^2 value of 0.11 and PC3 with an R^2 value of 0.09.

Disease incidence was regressed on the original derived variables using stepwise regression. No statistically significant model, either with multiple or single variables, was obtained. The most significant model included D3 as a single variable ($P = 0.29$).

Site × year scores on principal component 1 were plotted against the corresponding scores on principal component 4; no distinct trend was found, although variation was mostly higher within years than with location.

DISCUSSION

The influence of soil moisture and soil temperature on the incidence of sorghum downy mildew has been investigated by several researchers. Balasubramanian (1) described an optimum temperature of 26.3 C with a considerable reduction in disease incidence when the soil temperature was below 21.3 C. He found soil moisture in the range of 44–47% (available soil moisture) to be optimal for disease incidence; higher soil moistures (77–79% available soil moisture) were detrimental. Pratt (8) reported a failure of oospores to germinate when the soil was watered daily for 21 days after planting. Kenneth (6) determined in Wisconsin-style temperature tanks that the optimum temperature was between 24–29 C for infection. He observed no infection below 20 C. Schuh et al (11) determined the optimum soil temperature for infection to be 27 C, with no downy mildew observed below 20 C and above 30 C. Saturated soils and dry soils were both suppressive for disease.

The principal component analysis was executed using the seven derived variables. The best regression model contained the principal components 1 and 4 as independent variates and explained about 41% in the variation of disease incidence. This degree of predictive power is satisfactory when taking into consideration the fact that the model does not contain variables such as sand content and inoculum density, which have a documented influence on the disease incidence levels (11). Tedious isolation procedures exclude the assessment of inoculum density in a commercial production situation (10). These findings demonstrate the strong influence of soil environment on disease incidence and explain fluctuations between years. The principal components used in the regression model represent variation in soil conditions unfavorable for sorghum downy mildew. PC1 represents the variation caused by wet soils, and to a lesser extent, by cold soils. PC4 depicts variation caused by dry soils. Both conditions, wet and dry, have a disease suppressive effect, as has been shown in field and greenhouse trials (2,6,11). Even though PC4 only had a *P*-value of 0.17, it was included in the model for two reasons. First, the model *P*-value was 0.05 and secondly, the suppressive effect of dry soils was proven in controlled condition experiments. The high *P*-value could have been caused by the fact that dry soil conditions were only found during the 1984 growing season.

PC3 represents the influence of soils that are cold (< 20 C) or warm (> 30 C). Even though both conditions had a disease suppressive effect in experiments under controlled conditions (11), the principal component was not significant ($P = 0.22$) in the regression analysis. This could be explained by the way temperature data were obtained, i.e., the use of minimum and maximum temperatures for each day. These temperatures were present in the soil for only a small number of hours per day. Thus, these conditions did not persist long enough in the soil to have a significant influence as compared with the controlled condition experiments, where these temperatures were maintained for 24 hr each day.

The failure to obtain statistically significant regression models with the original variables demonstrates the usefulness of the principal component analysis when dealing with sets of variables that are highly intercorrelated. By reducing the dimensionality in a set of variables, principal component analysis can also be used to reduce the number of variables entered into regression models, which is especially important when weather variables are only a subset of all the variables used (14,16,17). Moreover, the preliminary screening for independent variables based on previously obtained information is a much better procedure than simply using all measured variables in the hope that a model of significance will emerge. Often the latter procedure produces correlations that, although high, are illogical and spurious.

The variation of the PC scores is greater between years than in between locations. This is because the weather stations were relatively close together (about 160 km) and that, in general, the area is subjected to the same average weather conditions.

The results obtained in this study could be used to classify areas according to their disease potential, when average weather conditions around planting time are known. They help explain why sorghum downy mildew is endemic in the Coastal Bend area, where favorable weather conditions for disease expression are fairly common during planting time.

TABLE 5. Disease incidence and principal component scores (PC1-PC4) for each location and year combination

Location and year[a]	DI[b] (%)	Principal component scores[c] PC1	PC2	PC3	PC4
B85	8	1.00	0.93	1.24	−1.00
O85	2	0.08	−0.20	−0.71	0.25
B84	9	−0.87	1.30	0.75	−0.10
O84	6	−1.45	0.50	1.14	2.32
L84	18	−0.91	−0.95	0.62	0.43
B83.1	12	1.08	−1.19	−0.24	0.53
B83.2	1	1.45	−0.30	−0.50	1.12
B83.3	0	1.25	1.16	1.36	−0.79
R83.1	5	1.26	−0.47	−0.35	0.89
R83.2	8	−0.43	−0.74	−0.46	−0.83
O83	18	−0.93	−1.63	0.66	−1.33
L83	15	−0.43	−0.52	−0.32	−0.83
B82	28	−0.61	0.57	−1.07	−0.48
O82	19	−0.50	1.54	−2.13	−0.17

[a] Locations and year. B = Beeville, O = Orange Grove, L = Laward, and R = Robstown.

[b] Disease incidence (average of 10 replications).

[c] Scores for principal components 1–4. Scores were rounded off to two digits.

LITERATURE CITED

1. Balasubramanian, K. A. 1974. Role of date of seeding, soil moisture, temperature, and pH in the incidence of downy mildew of sorghum. Plant Soil 41:233-241.
2. Frederiksen, R. A. 1980. Sorghum downy mildew in the United States: Overview and outlook. Plant Dis. 64:903-908.
3. Frederiksen, R. A., Amador, J., Jones, B. L., and Reyes, L. 1969.

Distribution, symptoms, and economic loss of downy mildew caused by *Sclerospora sorghi* (Kulk.) Weston & Uppal in grain sorghum in Texas. Plant Dis. Rep. 53:995-998.
4. Futrell, M. C., and Frederiksen, R. A. 1970. Distribution of sorghum downy mildew in the USA. Plant Dis. Rep. 54:311-314.
5. Kendall, M. 1975. Multivariate Analysis. Lifetime Learning Publications, Belmont, CA 458 pp.
6. Kenneth, R. 1970. Downy mildews of gramineae in Israel. Indian Phytopathol. 23:371-377.
7. Madden, L., and Pennypacker, S. P. 1979. Principal component analysis of tomato early blight epidemics. Phytopath. Z. 95:364-369.
8. Pratt, R. G. 1978. Germination of oospores of *Sclerospora sorghi* in the presence of growing roots of host and nonhost plants. Phytopathology 68:1606-1613.
9. SAS Institute, Inc. 1982. SAS User's Guide: Statistics. 1982 edition. Cary, N.C. 584 pp.
10. Schuh, W., Frederiksen, R. A., and Jeger, M. J. 1986. Analysis of spatial patterns in sorghum downy mildew with Morisita's index of dispersion. Phytopathology 76:446-450.
11. Schuh, W., Jeger, M. J., and Frederiksen, R. A. 1987. The influence of soil temperature, soil moisture, soil texture, and inoculum density on the incidence of sorghum downy mildew. Phytopathology 77:125-128.
12. Shaw, C. G. 1975. The taxonomy of graminaceous downy mildews, with emphasis on those attacking maize. Trop. Agric. Res. 8:47-55.
13. Shaw, C. G. 1978. *Peronosclerospora* species and other downy mildews of Gramineae. Mycologia 70:594-604.
14. Stynes, B. A. 1980. Synoptic methodologies for crop loss assessment. Pages 166-175 in: Crop Loss Assessment. Proc. of E. C. Stakman Commemorative Symp. P. S. Teng and R. E. Gaunt, eds. Misc. Publ. 7. Minn. Agric. Exp. Stn., St. Paul, MN. 327 pp.
15. Stynes, B. A., and Veitch, L. G. 1981. A synoptic approach for crop loss assessment used to study wheat. IV. The description of cultural practices and the combination of cultural data with soil data. Aust. J. Agric. Res. 32:1-8.
16. Veitch, L. G., and Stynes, B. A. 1979. A synoptic approach for crop loss assessment used to study wheat. I. Relationships between soil properties and traditional soil classifications. Aust. J. Soil Res. 17:227-236.
17. Veitch, L. G., and Stynes, B. A. 1981. A synoptic approach for crop loss assessment used to study wheat. II. Variables derived from soil properties and traditional soil classifications suitable for regression studies. Aust. J. Soil Res. 19:13-21.