Investigation of Shift in Decay Hazard (Scheffer) Index Values over the Period 1969-2008 in the Conterminous United States

Patricia K. Lebow
Charles G. Carll
USDA Forest Products Laboratory
Madison, Wisconsin

ABSTRACT

A statistical analysis was performed that identified time trends in the Scheffer Index value for 167 locations in the conterminous United States over the period 1969-2008. Year-to-year variation in Index values was found to be larger than year-to-year variation in most other weather parameters. Despite the substantial yearly variation, regression equations, with time (year) as the independent variable could be fit to the Scheffer Index data. A linear regression was found to provide appropriate fit when square-root transformations of Scheffer Index values were used as the dependent variable, and error terms were fitted as being spatially dependent. The 167 locations were segregated into ten climate groupings (termed "regions") based on annual precipitation and number of Fahrenheit Heating Degree Days. In most "regions" the calculated time coefficient value for the regression equation was positive, indicating a 40-year trend toward increase in Scheffer Index value. In three of the ten "regions", the time trend was moreover found to be statistically significant. All three of these "regions" were moist, with annual precipitation exceeding 20 inches (≈ 510 mm). The region with the most compelling increase in Index value was the moist region with in excess of 7000 Heating Degree Days. This region encompassed northern New England, the northern portion of the Great Lakes region, the Upper Mississippi valley, northern Minnesota, and the eastern part of North Dakota.

INTRODUCTION

A parameter, termed the Climate Index value, was devised by Scheffer (1971) to estimate the decay hazard, by geographic location within the conterminous United States (all states except Alaska and Hawaii), for wood exposed above ground to exterior conditions. The parameter was devised to be easily calculated from climatic data available from the U.S. Weather Bureau. The formula for the Climate Index value is:

$$Index = \sum_{Jan}^{Dec}[(T-35)(D-3)]/30 \quad [1]$$

where T is mean monthly average temperature (expressed in degrees F)
and D is the mean number of days in the month with 0.01 inches or more of precipitation
and where the term (T-35) is set equal to 0 if $T<35°F$
and where the term (D-3) is set equal to 0 if $D<3$

The Index value may alternatively be expressed as:

$$Index = \sum_{Jan}^{Dec}[(T-2)(D-3)]/16.7 \quad [2]$$

where T is mean monthly average temperature (expressed in degrees C)
and D is the mean number of days in the month with 0.25 mm or more of precipitation
and where the term (T-2) is set equal to 0 if $T<2°C$
and where the term (D-3) is set equal to 0 if $D<3$

The offset value for D (-3) was chosen to yield an Index value of zero in the driest regions of the southwestern United States. The offset value for T was chosen based on a temperature close to freezing (below which fungal activity was recognized as being inhibited). The divisor in the index formula was chosen to yield index values that would not exceed 100 for essentially any location in the conterminous United States (with the exception of locations on the Florida peninsula).

The Climate Index value has become widely-recognized, and is commonly termed “the Scheffer Index”. For the conterminous United States, no alternative metric has been shown to be as reliable as the Scheffer Index (SI) for characterizing climatic decay hazard (Carll, 2009). It is worth noting that Scheffer Index values, being based on readings taken at Government weather stations, do not address micro-climate effects. Microclimate effects can apparently be significant in mountainous regions and in the West Coast fog belt.

Recently, Carll (2009) and Morris and Wang (2008) have calculated SI values for various locations in North America from relatively recent climate normal data, and compared the calculated values to those calculated in the 1970’s or 1980’s from older weather data. Each of these investigators observed an increase in Index values. In the conterminous United States, the values generally increased, with some exceptions. When averaged over 228 locations, the overall increase was small (1.5 points), but nonetheless appeared to be statistically significant. Carll (2009) noted that there was apparent geographical variation in how index values changed relative to the values presented in 1971 by Scheffer. On a redrawn Climate Index map (Carll 2009), the isopleths for index values of 40 and 50 on the Climate Index map were noticeably shifted northward in the Upper Midwest and Great Lakes regions, whereas the isopleths for index values of 70 and 80 in the southeastern U.S. were barely shifted. Morris and Wang (2008) reported more consistent and larger increases in index values at various locations in Canada than in the United States; of 58 Canadian locations at which comparisons were made, 55 showed an increase in Index value (an average value increase of 4.7 points) while 3 showed a decrease in Index value (with an average value decrease of 1.8 points). In some Canadian locations, Scheffer Index values increased by as much as 10 points. Morris and Wang attributed the change in Index values to climate change (which can be argued as patently obvious, inasmuch as Scheffer Index values are based on climate data). They suggested that the underlying climate parameters underwent either a directional or a cyclical change, or both, and made reference to a particular cyclic temperature fluctuation known as the Pacific Decadal Oscillation (PDO). Morris and Wang thus suggested that periodic re-calculation of Scheffer Index values would be prudent.

OBJECTIVE

The purpose of this work was to better understand the characteristics of the Scheffer Index by investigating temporal trends and spatially-related temporal variation in SI values over a recent 40-year period for a comprehensive range of locations in the conterminous United States. The period was the most recent period for which weather data existed at the start of the investigation (during calendar year 2009).

WEATHER DATA

Data for each of 167 locations in the conterminous United States were obtained from the National Climatic Data Center (www.ncdc.noaa.gov). The data were downloaded as .pdf files. Files for each of the years 1969 through 2008 were downloaded for each of the locations. The locations are shown in Figure 1. Selected data from each year’s record were entered into Excel spreadsheets. The most important of the selected data were mean monthly temperature for each month, and the number of days in each month with measurable precipitation (precipitation of $\geq$ 0.01 inches water equivalent), as these data are needed to compute yearly Scheffer Index values. Data for total annual precipitation (expressed as water equivalent), and for annual heating degree days (HDD), and annual cooling degree days (CDD), each in Fahrenheit Degree Days with base temperature of 65°F (18.3°C), were also entered into the spreadsheets. The yearly precipitation values and HDD values were used for segregating individual locations into different “climate groupings”, as defined in Table 1. The groupings, being based on average yearly precipitation and average

HDD are arguably simplified; they do not distinguish locations with a “marine climate”, from locations with similar values for heating degree days and annual precipitation, but not having a “marine climate”[1].

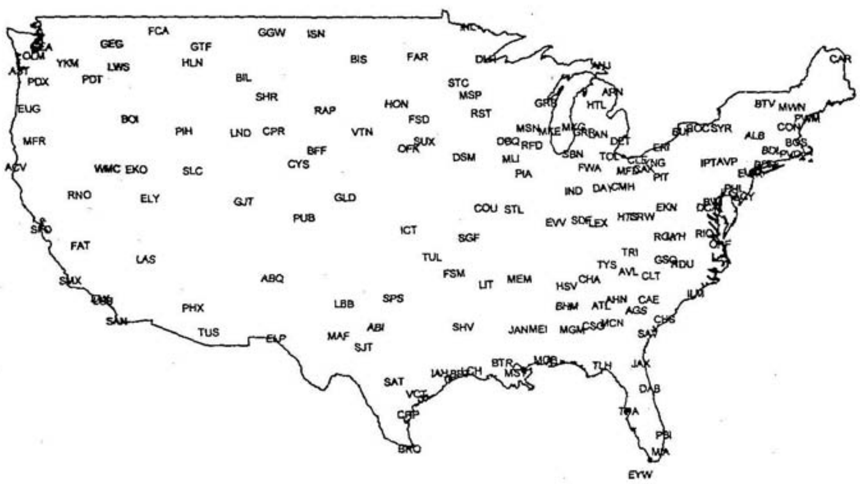

Figure 1. Airport codes for locations in which weather records were obtained for the years 1969 through 2008

In total, Scheffer Index values for 6653 location/year combinations were calculated. The arithmetic product of 167 (locations) and 40 (years) is 6680. For 27 location/year combinations, Scheffer Index values could not be calculated. One missing monthly value for either average temperature or number of days of measurable precipitation within a year prevented calculation of an SI value for that year.

The precipitation, HDD, and CDD data were also used to illustrate the relative degree to which different climate parameters varied from year to year (Table 2). The coefficient of variation (CV)[2] values presented in Table 2 can be used to compare relative year-to-year variability of different weather parameters, even though the parameters are measured on different scales with different measurement units. The values in Table 2 indicate large relative variability in the Scheffer Index value compared with most other weather parameters. Judged by the CV values presented in Table 2, Scheffer Index values show greater year-to-year variation than the number of HDD in heating climates, than the number of CDD in

[1] Locations with marine climates typically show limited seasonal temperature variation, but substantial seasonal variation in precipitation (with precipitation being most prevalent during winter). Locations with marine climates are all within a fairly narrow band along the Pacific Coast (meaning that a grouping for “marine climate” would include a fairly small number of locations). There also is some debate regarding where along the California coast a marine climate ends and a Mediterranean climate begins. The groupings outlined in Table 1, while arguably ‘insensitive”, are straightforward. Using the grouping criteria of Table 1, in no case is the classification of a location subject to dispute.

[2] Coefficient of variation (CV) is the standard deviation divided by the mean, and thus is proportional (dimensionless). CV values are often expressed as percentages.

cooling climates, or than the number of days per year with measurable precipitation, but slightly less year-to-year variation than annual precipitation (total water equivalent).

Table 1: Climate groupings

Climate grouping designation	Annual precipitation (as water equivalent, inches)	Heating Degree Days (65° F basis)
M500	> 20 (moist)	< 1000
M2000	> 20	1000 - 3000
M4000	> 20	3000 - 5000
M6000	> 20	5000 - 7000
M8000	> 20	> 7000
D500	< 20 (dry)	< 1000
D2000	< 20	1000 - 3000
D4000	< 20	3000 - 5000
D6000	< 20	5000 - 7000
D8000	< 20	> 7000

Table 2: Coefficient of variation values for Scheffer Index and selected weather parameters

Coefficient of variation (median CV values over indicated # of locations)				
Scheffer Index	Heat Deg. Days (where HDD>CDD)	Cool Deg. Days (where CDD>HDD)	Annual days with precipitation	Annual precipitation
18.5% (167 locations)	7.6% (138 locations)	8.3% (29 locations)	10.7% (167 locations)	19.0% (167 locations)

STATISTICAL ANALYSIS

Models were developed in an attempt to determine if there were shifts in the Scheffer Index over time, and if so, how the time trends were spatially related. Development of the models involved numerous steps. The first step in the process involved fitting a simple linear regression with the single independent variable of time (year) for each location, *i*. The regression had the form:

$$SI_{ij} = \beta_{0i} + (\beta_{1i} * time_j) + \varepsilon_{ij} \qquad [3]$$

where SI_{ij} is the Scheffer Index observed at site *i* at time *j*
and the ε_{ij} terms are the individual stochastic differences between yearly values and the values given by the deterministic portion of the regression equation. The ε_{ij} terms are commonly called "residuals".

In fitting data to the model, the time value was centered (a time value of zero was selected for mid-point of the 40-year interval). Thus the year 1988 was assigned a time value of -0.5, while the year 1989 was assigned a time value of +0.5, and the range of yearly values was from -19.5 to +19.5.

As is common in statistical analyses, the analysis included investigation of transformations of the SI values. One of the transformations involved calculating the square roots of yearly SI values for each location, *i*, and again fitting the transformed values to a simple linear regression of the form:

$$\sqrt{SI_{ij}} = \beta_{0i} + (\beta_{1i} * time_j) + \varepsilon_{ij} \qquad [4]$$

with the time value also centered over the 1969-2008 period

Plots of observed residual values indicated that when the SI values for each location were fit to [3] and the locations ranked in order of increasing mean value, the distribution of ε_{ij} terms was related to mean 40-year SI value (left plot of Figure 2). In contrast, when the SI values were transformed and fit to [4], the distribution of ε_{ij} terms showed no indication of being related to the mean 40-year value (right plot of Figure 2). The stabilizing influence of [4] on the distribution of residuals led us to choose this model for further analysis of the data.

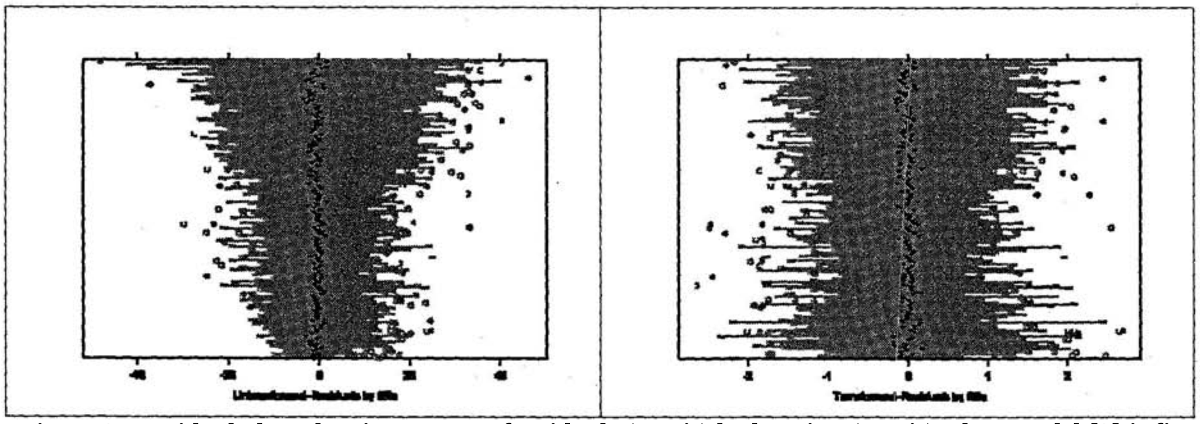

Figure 2. Residual plots showing ranges of residuals (x-axis) by location (y-axis) when model [3] is fit (left) and when model [4] is fit (right). Residuals are plotted with respect to location, with locations ranked in increasing order of Scheffer Index or transformed Scheffer Index,

The ε_{ij} terms in regression equations such as [3] or [4] are commonly assumed (at least initially) to be identically and independently distributed. In this study, an assumption of identically and independently distributed residuals would translate to the assumption that there was no systematic spatial or temporal influence on their distribution. This assumption is counter-intuitive. Most persons would expect that yearly fluctuation in a weather-related parameter (and by extension, that yearly deviation from an underlying time trend in that parameter) would be more similar at nearby locations than at more distant locations. For example, if year X was an abnormally hot (or dry, or wet) year in St. Louis, most people would expect the same year to also have been abnormally hot (or dry or wet) in Peoria, but not necessarily in Toledo, Atlanta, or Cheyenne. During each of the 40 years, the differences between residuals at nearby locations were (not unexpectedly) smaller than the differences between residuals at more distant locations . Within a distance of approximately 450 miles, the spatial influence on distribution of residuals was approximately exponential with respect to distance, with the influence tending to be especially strong at shorter distances. The transformed linear model outlined in [4], accompanied by a spatially-related exponential term for description of residual values, was therefore judged as appropriate for the dataset.

Appropriate description of the residual terms for [4] was necessary for correct set-up of inferential tests concerning the deterministic (non-stochastic) parts of the model. The issue of greatest interest concerned the time coefficient (β_{1i}) values. The first hypothesis tested (by Analysis of variance) was whether all time

[3] This can be visualized with semivariogram plots. For reasons of manuscript brevity, semivariaogram plots are not included in this manuscript. Various types of semivariograms can be studied, but based on our model formulation, each of the 40 years would need to be presented to fully and adequately depict the relationships. Further details on this type of modeling can be found in Pinheiro and Bates (2000).

coefficient values were zero. This hypothesis was rejected ($F_{(167,6319)}$ = 1.8251, p-value <0.0001)[4]. This was followed with a test of the hypothesis of common $ß_{1i}$ values. This test (of the hypothesis that the time coefficient was similar for all locations) was also rejected ($F_{(166,6319)}$ = 1.8110, p-value < 0.0001). The time trend (in square root of SI value) was thus clearly dependent on location.

The location-specific time coefficient estimates from model [4] were segregated into each of the ten "climate groupings" outlined in Table 1. For brevity, the "climate groupings" are hereafter referred to as "regions". Figure 3 contains "box plots", which depict variability in the time coefficient values within each of the "regions". The white bar, in a more-or-less central location within each shaded "box" indicates the "regional" median time coefficient value. The boxed areas indicate interquartile ranges[5], the whiskers incorporate the entire range of coefficient values, provided that the highest or lowest value is within 1.5 times the interquartile range of the upper and lower quartile values, and the single horizontal lines indicate outlier coefficient values (Cleveland, 1994). The D500 "region" only contained one location, making the interquartile range meaningless for the region; thus no "box" is depicted in Figure 3 for the D500 region.

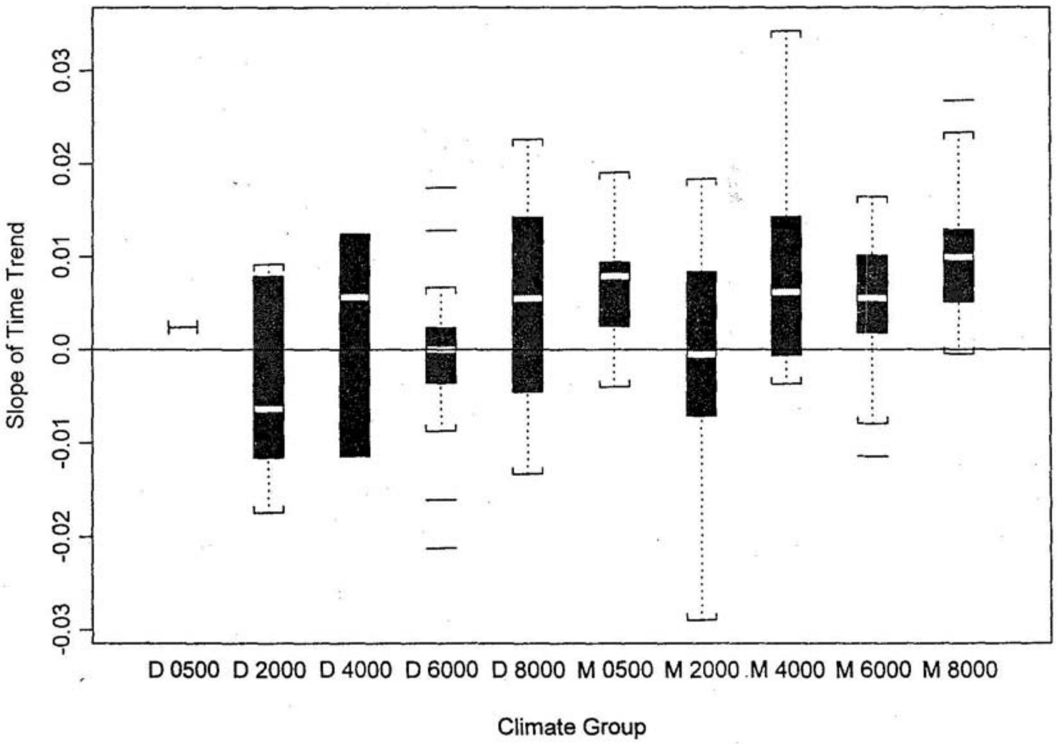

Figure 3. Box plots of time coefficient parameter estimates for individual locations, grouped by regions as given in Table 1. Coefficient values calculated for each location by model [4].

[4] The subscripted values in parentheses indicate the numerator and denominator degrees of freedom (in that order). The p-value indicates the probability of observing as high a value of the test statistic (in this case as high an F value) given that the null hypothesis is true. The p-value is often used to compare to a pre-determined probability of "type I error" (risk of incorrectly rejecting the null hypothesis). A p-value of <0.05 (1 in 20 or lower chance) is generally (conventionally) considered an appropriate level for type I error.

[5] The observed interquartile range is the difference (range) between the estimated population value at the 25th percentile and the estimated population value at 75th percentile (with values in ranked order).

Finally, a model was fit to the data within each "region" as:

$$\sqrt{SI_{ij}} = \dot{\beta_{0i}} + (\beta_{1k_i} * time_j) + \varepsilon_{ij} \qquad [5]$$

where the time coefficients β_{1k_i} are specific to the region k of a location i as opposed to the location itself and with the ε_{ij} terms spatially dependent.

For each of the ten "regions", a null hypothesis test was conducted to test whether the regional time coefficient value was significantly different than zero. For this, conditional two-tailed t-tests were performed; see Pinheiro and Bates (2000) for details. Calculated time Coefficient values, and their corresponding standard error and hypothesis-test values are shown in Table 3[6,7].

Table 3: Fitted time coefficient values for [5] by the climate groupings outlined in Table 1

"Region"	Time coefficient	Std. Error	t-value	p-value=P(t>$\|t_{6476}\|$)
D 0500	0.01432	0.0084757	1.689	0.0912
D 2000	-0.00049	0.0053067	-0.093	0.9259
D 4000	0.00647	0.0056485	1.145	0.2522
D 6000	-0.00067	0.0045919	-0.146	0.8842
D 8000	0.00560	0.0047214	1.186	0.2357
M 0500	0.00836	0.0055039	1.519	0.1287
M 2000	0.00831	0.0042343	1.962	0.0498
M 4000	0.00902	0.0040067	2.252	0.0244
M 6000	0.00526	0.0042479	1.238	0.2159
M 8000	0.01004	0.0045268	2.218	0.0266

t_{6476} indicates a t-statistic value with 6476 degrees of freedom. The degrees of freedom were determined by the total number of observations (6653) minus the degees of freedom for the model (167 for intercept values plus 10 for slope values).

RESULTS AND DISCUSSION

For only two of the ten "regions" in Table 3 was the time coefficient negative. These two "regions" were each classified as dry (having limited precipitation), and in both the p-value (for rejection of the hypothesis that the slope value did not differ from zero) exceeded 0.85. In these two regions, the null hypothesis would thus be accepted. In contrast, for eight of the ten "regions" the time coefficient was positive. Using a customary type I error level of 0.05, the null hypothesis (of time coefficient not differing from zero) would be rejected for three of the eight "regions" with positive time coefficient values. In each of these three "regions" annual precipitation exceeded 20 inches.

---

[6] The distribution of residuals for the data in each of these "regions" appeared satisfactory, although it generally was not as obviously normal as the distribution of residuals of the whole (undivided) data set.

[7] t-values in the table are numerically equal to the time coefficient values divided by their respective standard errors. P-values indicate the probability of observing as high a value of the test statistic (in this case as high a t-value) given that the null hypothesis is true. P-values are dependent on the degrees of freedom of the t-distribution.

Of the five dry "regions", three showed a positive regional time coefficient whereas two showed a negative regional time coefficient. The p-value (for rejection of the null hypothesis) was consistently higher for the two dry "region" with a negative coefficient value than for the three dry "regions" with a positive coefficient value. The overall trend in SI value in the United States' dry regions thus appears to have been more towards increasing Index values than towards decreasing Index values. Using a customary level for type I error however, the analysis does not support the contention that there was a systematic increase in decay hazard index over the period of 1969-2008 in the dry regions of the conterminous United States.

Of the five "regions" classified as moist (with more than 20 inches of annual precipitation), all five showed a positive regional time coefficient, and (as indicated previously) in three of these five "regions", the coefficient value was found to differ significantly from zero. The "M8000 region", which was the coldest of the moist "regions", had a regional Coefficient value of greater magnitude than any of the other moist "regions"; the accompanying p-value also clearly indicated statistical significance of the coefficient value (value appreciably less than 0.05). In this "region", the evidence thus compellingly points toward an upward shift of decay hazard Index values between 1969 and 2008. This "region" includes the most northerly locations in roughly the eastern half of the country (as far west as the eastern portion of North Dakota). The finding that decay hazard Index evidently changed most in the coldest moist region of the conterminous United States concurs with the observations of Moms and Wang (2008) that Scheffer Index values have evidently increased in Canada over the last few decades to a greater overall extent than in the conterminous United States. In the "region" designated "M4000" (with between 3000 and 5000 HDD) the magnitude of the time coefficient value was also greater than in most other "regions", and the p-value also clearly indicated statistical significance of the coefficient value (value appreciably less than 0.05). Upward shift in the Index value was thus compelling in this "region", although not quite as compelling as in the "M8000 region".

CONCLUSIONS

Statistical evaluation of Scheffer Index values over the period 1969-2008 provided no compelling evidence that there was a shift in values in the portions of the conterminous United States classified as dry (where average annual precipitation was less 20 inches). In contrast, the evaluation compellingly indicated an overall upward shift in values in three of five "regions" of the conterminous United States classified as moist (where average annual precipitation exceeded 20 inches). The most compelling upward shifts in Index values were observed in the moist "region" with between 3000 and 5000 Heating Degree Days and in the moist "region" with in excess of 7000 Heating Degree Days.

ACKNOWLEDGEMENTS

The authors acknowledge the assistance of Apolonia Bocanegra-Severson (now retired from FPL) and Torrance Caldwell (former summer FPL student-employee, enrolled at Clark-Atlanta University) in downloading data files and in spreadsheet data entry.

REFERENCES

1. Carll, C. G. 2009. Decay hazard (Scheffer) index values calculated from 1971-2000 climate normal data. General Technical Report FPL-GTR-179. Madison, WI: U. S. Department of Agriculture, Forest Service, Forest Products Laboratory. 11 pp.
2. Cleveland, W.S. 1994. The Elements of Graphing Data. Hobart Press. Summit, NJ. 297 pp.
3. Morris, P. I., and J. Wang. 2008. A new decay hazard map for North America using the Scheffer Index. IRG/WP 08-10672. (prepared for IRG Americas Regional Meeting, Nov 30 - Dec 2, 2008 in Guanacaste, Costa Rica). 13 pp. The International Research Group on Wood Protection. www.irg-wp.com
4. Pinheiro, J. C. and Bates, D. M. 2000. Mixed-Effects Models in S and S-PLUS. Springer-Verlag New York, Inc. New York, New York. 528 pp.
5. Scheffer, T. C. 1971. A climate index for estimating potential for decay in wood structures above ground. Forest Products Jour. 21(10):25-31.

PROCEEDINGS

One Hundred Sixth Annual Meeting

of the

AMERICAN WOOD PROTECTION ASSOCIATION

Hyatt Regency Riverfront
Savannah, Georgia
May 23-25,2010

VOLUME 106

AMERICAN WOOD PROTECTION ASSOCIATION
P O BOX 361784 • BIRMINGHAM, ALABAMA 35236-1784 • USA

ISSN 0066-1198

AMERICAN WOOD PROTECTION ASSOCIATION
P.O. BOX 361784
BIRMINGHAM, ALABAMA 35236-1784
USA

Colin McCown, *Editor*
Beth Williams, *Assistant Editor*