57829

ON THE RELATIONSHIP BETWEEN SUNSPOT NUMBER AND THE FLARE INDEX

ROBERT M. WILSON

Space Sciences Laboratory, NASA Marshall Space Flight Center, AL 35812, U.S.A.

Abstract. During the years 1976-1991, sunspot number and the Kleczek flare index have displayed a strong linear correlation (r = 0.94), one that can be described by the equation $\hat{y} = -0.15 + 0.10\,x$, where x denotes annual sunspot number. While true, the temporal behaviors of the two parameters have differed, with sunspot number peaking first in 1979 and the flare index peaking much later in 1982 during cycle 21 and with more contemporaneous behavior in cycle 22 (both peaking in 1989, with a secondary peak in 1991). The difference appears to be directly attributable to the way in which the Kleczek flare index has been defined; namely, the annual flare index is the sum of the product of each flare's intensity (importance) times its duration (in minutes) divided by the total number of flares during the year. Because the number of 'major' flares (those of importance ≥ 2) and flares of very long duration (duration ≥ 100 min) both peaked after sunspot maximum (1982/81, respectively) in cycle 21, one should have expected the flare index to also peak (which it did). Likewise, because the number of major flares and flares of very long duration peaked simultaneously with sunspot number (1989) in cycle 22, one should have expected the flare index to also peak (which it did).

(NASA-TM-110643) ON THE RELATIONSHIP BETWEEN SUNSPOTS NUMBER AND THE FLARE INDEX (NASA. Marshall Space Flight Center) 12 p

N95-30467

Unclas

G3/92 0057829

1. Introduction

A simplified flare index, based strictly on the intensity (importance) and the duration (in min) of the Hα flare, was first introduced by Kleczek (1952). As such, it is believed that the flare index gives roughly the total energy emitted by the flare. Ataç (1987) extended the flare index through 1986, noting that chromospheric activity was significantly higher in cycle 21 as compared to cycle 20 and that activity began and reached peak earlier in the northern hemisphere as compared to the southern hemisphere. In passing, Ataç further noted that Kleczek's flare index was the only one that was being actively maintained (in contrast to the daily flare index that had been routinely reported in SOLAR GEOPHYSICAL DATA - Part II, Comprehensive Reports, through 1981). More recently, Özgüç and Ataç (1994) have again extended the flare index, now through 1991. They also compared the flare index with sunspot number and found the two parameters to be well correlated.

The purpose of this paper is to reinvestigate the relationship between the flare index and sunspot number for the years 1976–1991, in particular, the temporal behavior of the two data sets. Additionally, other solar-cycle related parameters (e.g., F10.7, corrected sunspot area, number of groups, number of flares, number of 'major' flares, and number of very long duration flares) are compared with the flare index in order to ascertain which if any might more appropriately describe the solar cycle-flare index relationship than the one using sunspot number.

2. Results and Discussion

Figure 1 displays annual averages of sunspot number (SSN) and the flare index (FI) for 1976–1991, similar in structure and appearance to the one shown in Özgüç and Ataç (1994), except that other pertinent statistical information has also been included. As in Özgüç and Ataç, the flare index values have been multiplied by 10 to make the comparison easier. Near the top of the chart and running horizontally is a comparison of the two parameters in terms of their 'runs' data (i.e., the current value being above or below its respective median value), their local 'trends' (i.e., the current value up or down as compared to the previous value), and their respective 'ranks' (from 1—lowest value to 16—highest value). Based on their respective ranks, one easily computes the Spearman rank correlation coefficient rs (Lapin, 1978, p. 633) to be 0.90, indicating that the annual averages of sunspot number and flare index appear to be well correlated. Analysis of the runs and trends data indicate that both data sets display statistically significant nonrandom variations (i.e., the solar cycle) and that the chance of obtaining the observed above/below or up/down combinations or those more suggestive of a departure from independence (on the basis of Fisher's exact test for 2 x 2 tables; Everitt, 1977; p. 15) is $P < 0.5$ percent. The circled terms identify those specific years when the behavior of the two parameters differed (runs and trends) or, in the case of ranks, when the two parameters differed the most in terms of ranking (1982).

Figure 2 depicts the scatter plot of flare index versus sunspot number. Based on standard linear correlation analysis (e.g., Lapin, 1978, p. 317), one easily computes the regression equation to be $\hat{y} = -0.15 + 0.10\ x$ (where x is SSN), the Pearson correlation coefficient r to be 0.94, the coefficient of determination (r squared) to be 0.89, and the standard error of estimate se to be 2.0. Thus, one infers that FI and SSN appear to be well correlated, with about 89 percent of the variance in FI being explained by the regression against SSN. The box in the lower part of Figure 2 shows the results of Fisher's exact test for the two parameters, indicating that the inferred relationship should be considered strong. The circled data point identifies the values of FI and SSN encountered in 1982 and reveals that it lies about three standard errors above the regression line. Therefore, while a strong linear correlation appears to exist between FI and SSN, on occasion, the inferred (predicted) flare index differs greatly from its observed value. The question now is why is this so.

Figure 3 compares FI and SSN against other important solar-cycle related parameters (cf. Wilson, 1993), including F10.7 (the 10.7-cm solar radio flux), CSSA (the corrected total sunspot area), N(G) (the number of 'groups' or regions on the Sun), N(f) (the number of grouped optical flares, as reported in SOLAR GEOPHYSICAL DATA - Part II, Comprehensive Reports), N(Mf) (the number of 'major' flares, where a major flare is one of Hα importance ≥ 2; Dodson-Prince and Bruzek, 1977), and N(LDf) (the number of flares of very long duration, duration ≥ 100 min). Identified with each curve is the Spearman rank correlation coefficient rs (comparing each parameter against FI). Thus, one infers that FI correlates well with all the various solar-cycle related parameters, especially with N(f) and N(LDf), both having rs = 0.96.

Figure 4 displays the scatter diagrams of FI versus N(f) (left panel) and FI versus N(LDf) (right panel). As in Figure 2, each panel gives the regression equation, the correlation coefficient, the coefficient of determination, and the standard error of estimate, as well as a measure of the strength of the inferred correlation (P). Again, the values associated with the year 1982 are circled. It is apparent that either regression is found to be superior to that based on SSN (Figure 2) and that the spread about the regression line is tighter (smaller se). From this, one infers that the rather large discrepancy found for 1982 between observed and predicted FI (based on SSN) essentially disappears when one uses a different proxy. Thus, while sunspot number peaked early in cycle 21 as compared to FI, the lack of simultaneity is more to be expected, since FI appears to better correlate against other parameters, in particular N(f) and N(LDf), than against SSN.

Returning briefly to Figure 3, one notes that both N(Mf) and N(LDf) have later occurring peaks during cycle 21, as compared to either SSN or N(f), more in line with the observed temporal behavior of FI. Perhaps a bivariate regression (Lapin, 1978, p. 369; Ehrenberg, 1982, p. 200) may provide an even better fit. Figure 5 shows the results of a bivariate analysis, comparing FI against both N(Mf) and N(LDf). Such an analysis makes sense because by definition the FI of an individual flare is the product of its intensity (importance) and duration. If major flares (flares of large importance) and flares of very long duration more strongly contribute to the overall FI average (although the number of such flares does vary over the solar cycle, the proportion of N(Mf) to N(f) and N(LDf) to N(f) varies little, averaging 1.03 percent and 1.65 percent, respectively, during 1976–1991), then the later occurring maxima of N(Mf) and N(LDf) might explain why FI peaks 3 years after sunspot maximum for cycle 21. Shown in Figure 5 is the bivariate regression equation, the coefficient of multiple correlation and its square, the Spearman rank correlation coefficient, and the standard error of estimate. While the bivariate fit may slightly improve the overall performance of predicting FI as compared to using SSN, it still does not do as well as using either N(f) or N(LDf) alone. (The rather strong correlation of FI and N(LDf) may be indicative of a suspected solar-cycle variation of the average duration of a solar flare; cf. Wilson, 1987.)

While securing values for N(f), N(Mf), and N(LDf), it became apparent that the way flares are grouped has changed. Prior to 1982, flares were grouped according to their start-max-end times by Hale plage region. Flares occurring within the overall duration from the same Hale plage region were grouped together as one event, even though the location within the region and the time of maximum may have differed. Thus, one event (flare) often had multiple peaks for time of maximum occurrence. Beginning in 1982, a new format was introduced, one which grouped flares on the basis of start-max-end times by NOAA group number and location. Flares occurring closely in time and location, occasionally from the same NOAA region, were generally treated as separate events. So, the concept of what constitutes a flare count has subtly changed. An effort was made to make the N(Mf) and N(LDf) counts after 1981 more like those occurring earlier. The greatest number of changes (several) found was for 1982. So, it may be that the reported FI may be slightly overvalued, especially for 1982. It is beyond the scope of this paper to address this issue.

Because SSN (N(f)) is already known for 1992 and 1993, being 94.3 (3952) and 54.6 (2541), respectively, one can easily estimate FI for these years on the basis of the regression equations shown in Figures 2 and 4. For 1992, one computes FI = 9.3±4.0, based on SSN, and FI = 6.9±2.8, based on N(f), both estimates being the ±2 standard error prediction interval. For 1993, one computes FI = 5.3±4.0 (SSN) and FI = 4.4±2.8 (N(f)). Thus, the peak FI for cycle 22 appears to have occurred in 1989.

In conclusion, while FI and SSN do appear to be highly correlated, stronger correlations are found to exist between FI and N(f) and FI and N(LDf). The latter occurring maximum in FI found during cycle 21 (1982) as compared to sunspot maximum (1979) seems to naturally follow from the fact that more flares, especially those of very long duration and those of greater importance, peaked later in the cycle after sunspot maximum. During cycle 22, a more contemporaneous behavior occurred with FI, SSN, N(f), N(Mf), and N(LDf) all peaking in 1989.

References

Ataç, T.: 1987, *Astrophys. Space Sci.* **135**, 201.

Dodson-Prince, H.W. and Bruzek, A.: 1977, in *Illustrated Glossary for Solar and Solar-Terrestrial Physics*, A. Bruzek and C. J. Durrant (eds.), D. Reidel Publ. Co., Dordrecht, Holland, p. 81.

Ehrenberg, A.S.C.: 1982, *A Primer in Data Reduction*, 305 pp., John Wiley & Sons, Inc., New York.

Everitt, B.S.: 1977, *The Analysis of Contingency Tables*, 128 pp., John Wiley & Sons, Inc., New York.

Kleczek, J.: 1952, *Publ. Centr. Astron. Inst. Czech.*, No. 22.

Lapin, L.L.: 1978, *Statistics for Modern Business Decisions*, 2nd ed., 778 pp., Harcourt Brace Jovanovich, Inc., New York.

Özgüç, A. and Ataç, T.: 1994, *Solar Phys.* **150**, 39.

Wilson, R.M.: 1987, NASA TP 2714, 38 pp., Marshall Space Flight Center, Huntsville, Alabama.

Wilson, R.M.: 1993, *J. Geophys. Res.* **98**(A7), 11,477.

Figure Captions

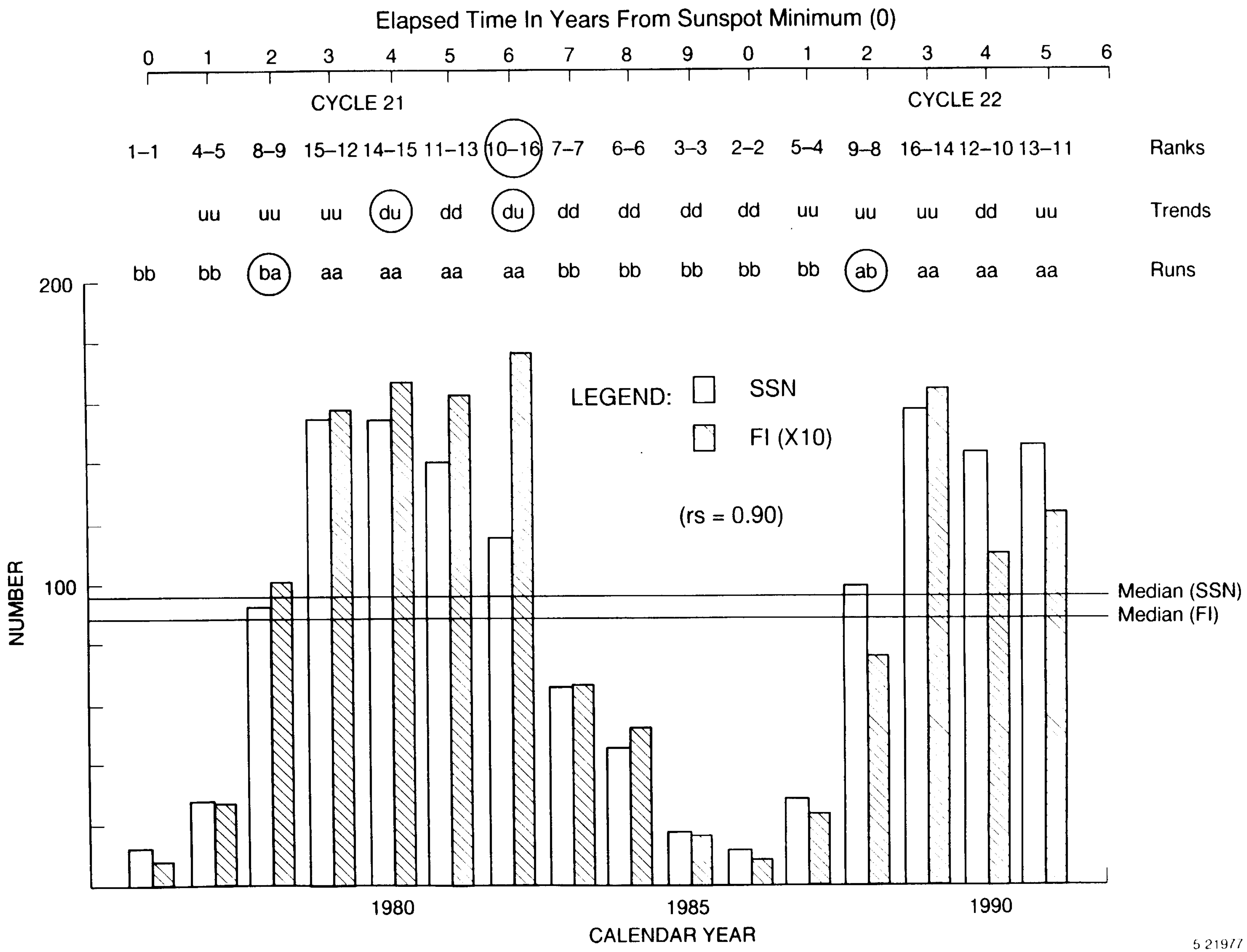
Figure 1 Annual averages of sunspot number (SSN) and the flare index (FI) for the interval 1976-1991. Shown are the median values for both parameters, the 'runs' combinations (where 'a' means above the median and 'b' means below the median), the 'trends' combinations (where 'u' means up and 'd' means down), the 'ranks' (where each parameter's value is ranked from 1, lowest, to 16, highest), and the Spearman rank correlation coefficient rs (see text for details). Elapsed time in years from sunspot minimum year is identified across the top, as well as the particular sunspot cycle number.

Figure 2 Scatter plot of FI versus SSN. The median values are identified as the vertical and horizontal lines. The regression is the heavy diagonal line, indicating a positive or direct correlation between the two parameters. Identified are the regression equation $\hat{y}$, the Pearson correlation coefficient r, the coefficient of determination (r squared), and the standard error of estimate se. Also shown is the probability (P) of obtaining the observed distribution or one more suggestive of a departure from independence.

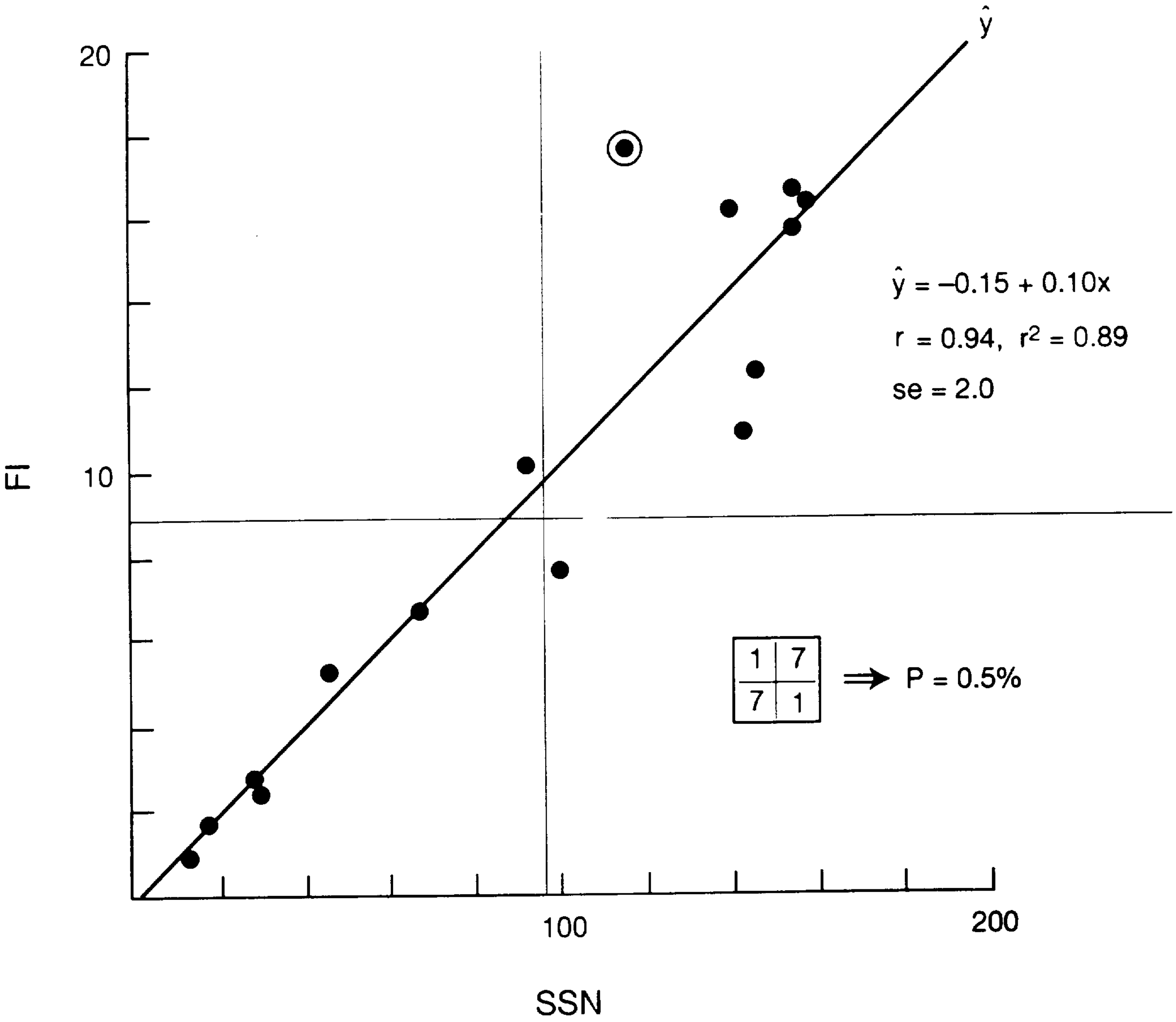
Figure 3 Comparison of FI with SSN and other solar-cycle related parameters. The Spearman rank correlation coefficient is identified for each parameter (see text for details).

Figure 4 Scatter plot of FI versus N(f) (left panel) and FI versus N(LDf) (right panel). Statistical parameters shown follow format given in Figure 2 (see text for details).

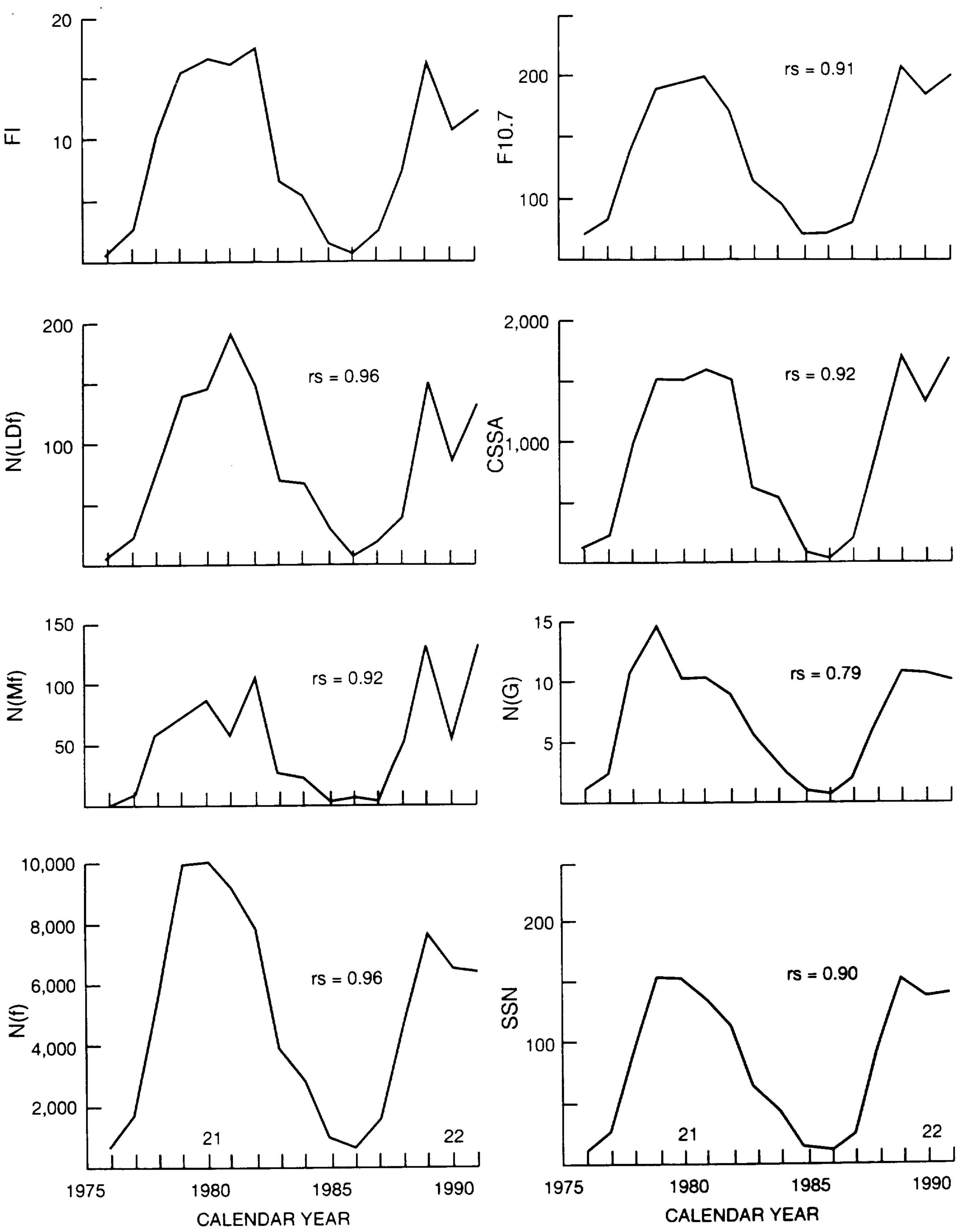
Figure 5 Scatter plot of FI (observed) versus FI (predicted), based on a bivariate fit of FI, N(Mf), and N(LDf).


Elapsed Time In Years From Sunspot Minimum (0)
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6
CYCLE 21
CYCLE 22
1–1 4–5 8–9 15–12 14–15 11–13 10–16 7–7 6–6 3–3 2–2 5–4 9–8 16–14 12–10 13–11 Ranks
uu uu uu du dd du dd dd dd dd uu uu uu dd uu Trends
bb bb ba aa aa aa aa bb bb bb bb bb ab aa aa aa Runs
200
100
NUMBER
LEGEND:
SSN
FI (X10)
(rs = 0.90)
Median (SSN)
Median (FI)
1980
1985
1990
CALENDAR YEAR

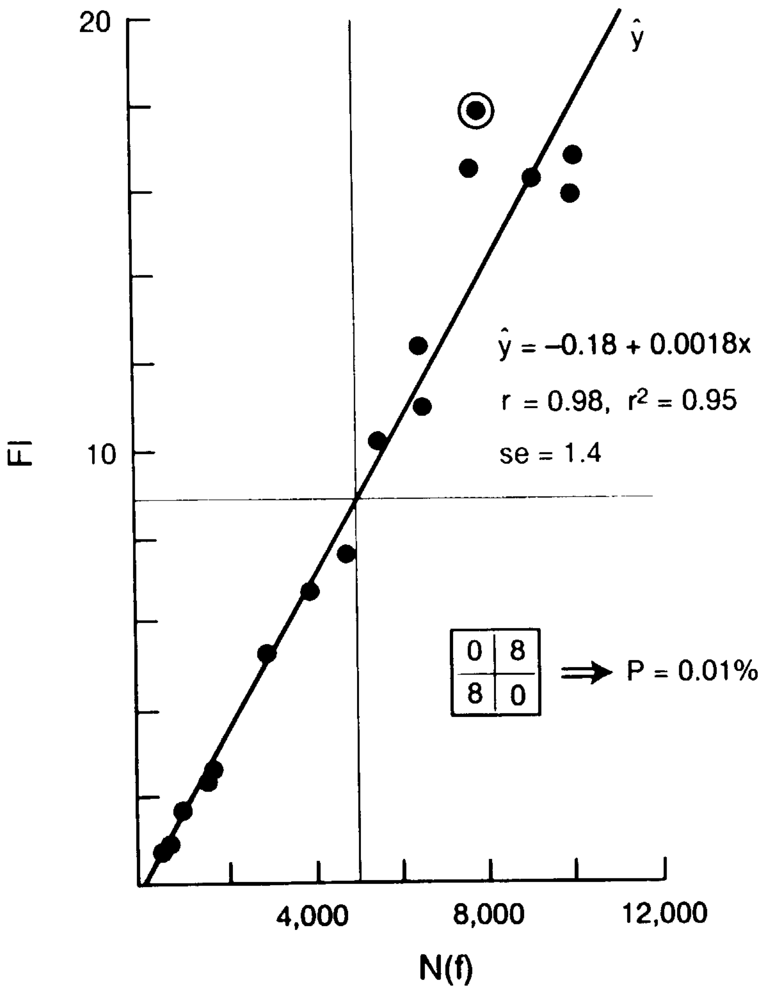
5 21977

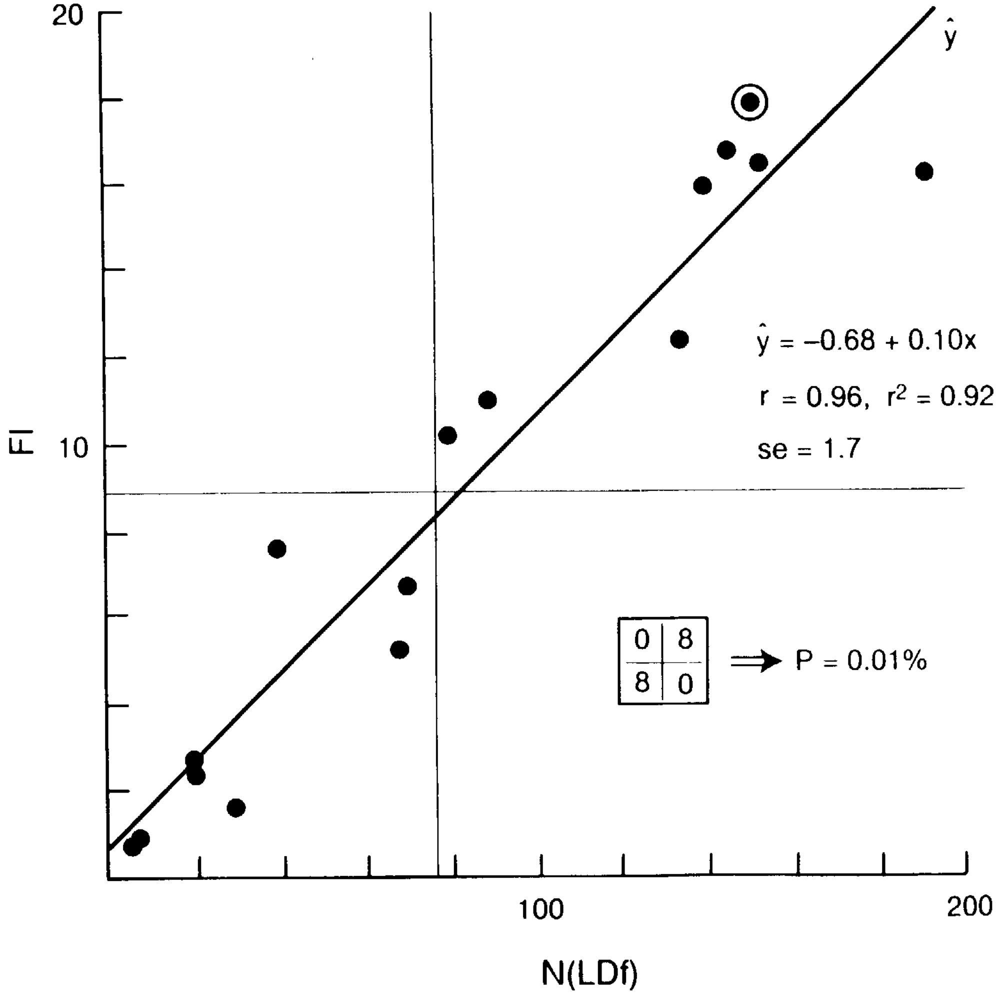


5-21980

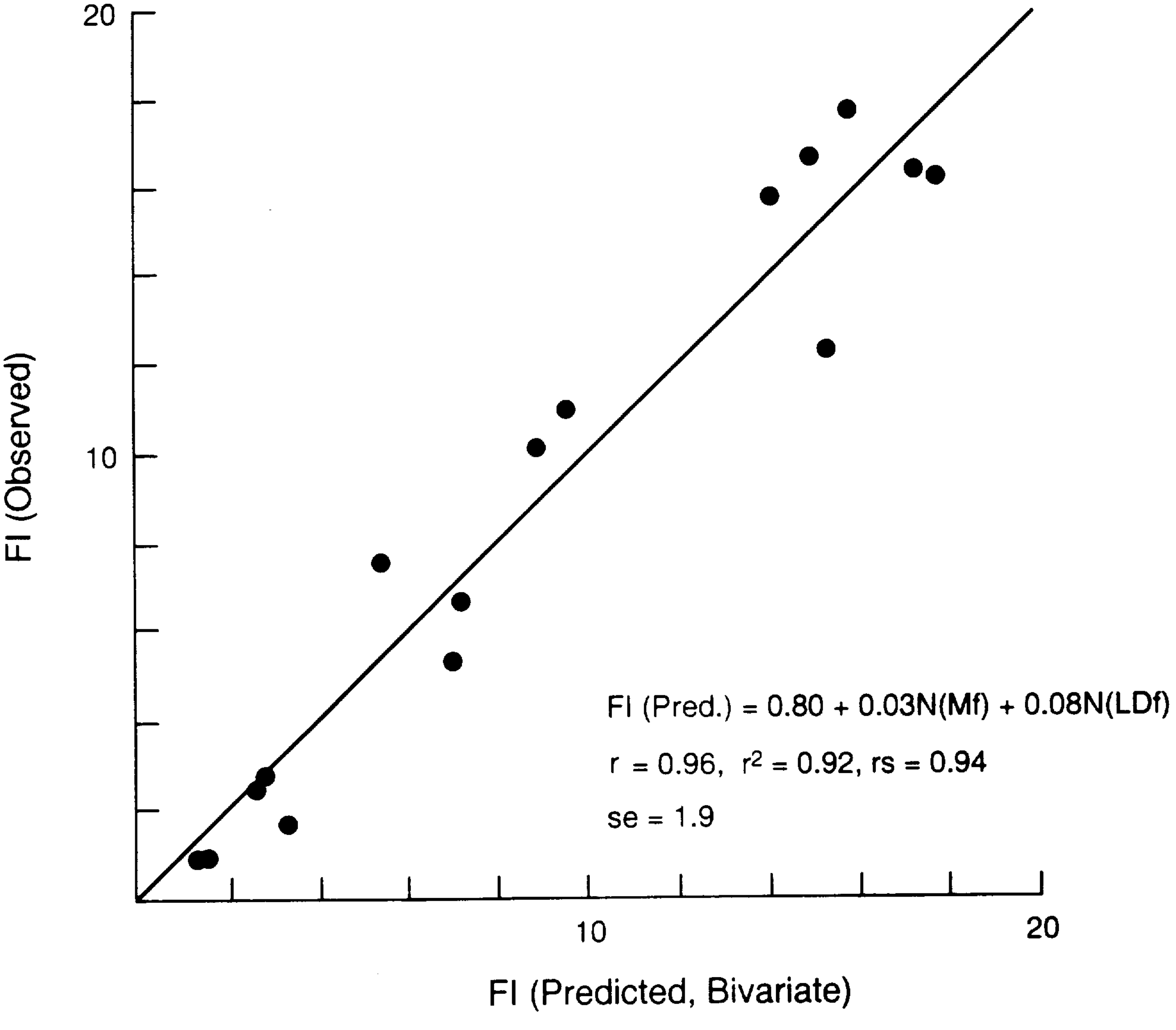

FI
20
10
F10.7
200
100
rs = 0.91
N(LDf)
200
100
rs = 0.96
CSSA
2,000
1,000
rs = 0.92
N(Mf)
150
100
50
rs = 0.92
N(G)
15
10
5
rs = 0.79
N(f)
10,000
8,000
6,000
4,000
2,000
rs = 0.96
21
22
1975
1980
1985
1990
CALENDAR YEAR
SSN
200
100
rs = 0.90
21
22
1975
1980
1985
1990
CALENDAR YEAR

5-21976


20
10
FI
4,000
8,000
12,000
N(f)
ŷ
ŷ = −0.18 + 0.0018x
r = 0.98, r² = 0.95
se = 1.4
0 | 8
8 | 0
⟹ P = 0.01%


20
10
FI
100
200
N(LDf)
ŷ
ŷ = −0.68 + 0.10x
r = 0.96, r² = 0.92
se = 1.7
0 | 8
8 | 0
⟹ P = 0.01%

5-21978



FI (Pred.) = 0.80 + 0.03N(Mf) + 0.08N(LDf)

$r = 0.96$, $r^2 = 0.92$, $rs = 0.94$

se = 1.9

5-21979