ANALYSIS OF SUGARCANE YIELD PRODUCTIVITY TRENDS IN THE WET TROPICS AT A DISTRICT LEVEL

By

R.N. ELLIS, K.E. BASFORD, M. COOPER, J.K. LESLIE and D.E. BYTH
The University of Queensland, School of Land and Food, Brisbane, Qld

Abstract

Data on tonnes of sugarcane per hectare (TCH) are now available to assess varietal improvement over time on a district basis. This study was conducted to develop a methodology for regular analysis of productivity data for mill districts to assess productivity trends over time, adjusted for variety and variety × environment interaction effects. In a 1996 Sugar Research and Development Corporation Report, Leslie and Wilson investigated district level data using a single varietal indexing method. This methodology was found to be subjective in the selection and calibration of the reference varieties. Restricted maximum likelihood (REML) methodology, a more rigorous analysis for district level data, was used to overcome this subjectivity. Best linear unbiased predictors (BLUPs) for random effects and best linear unbiased estimates (BLUEs) for fixed effects were computed for the mixed model analysis. Two analyses were conducted using combined district data and separate district data. Varieties were ranked according to their BLUP values, and district trends in TCH were plotted for the period 1958–1995. In the combined analysis, the BLUPs for TCH for each variety have shown little improvement since the introduction of Q99 in 1976. Separate district analyses indicated more recent advances in Babinda and Tully, but not Mulgrave. In general, the district production estimates from the separate district analyses were similar to those from the combined district analysis. Overall production of TCH in each district has increased over time, in part because a greater proportion of the area in the regions has been allocated better yielding varieties.

Introduction

Productivity trends for sugarcane yield (TCH) in the wet tropics were investigated at a district level. In a previous study, Leslie and Wilson (1996) had analysed district data from the Tully to the Mossman Mill districts of the far north Queensland Coast. This study provided an outline for future research and development of sugarcane in the wet tropics. Leslie and Wilson (1996) looked at six mill districts and the variables associated with sugarcane. Specifically, they found TCH had an upward trend for Mossman, Mulgrave, Mourilyan and Tully mill districts, sugar content (CCS) had a downward trend for each district, and sugar yield (TSH) had a significant upward trend only in the Mulgrave district. Among some of the explanations given for these trends was the impact that varieties have on TCH, CCS, and TSH. The purpose of this paper is to concentrate on the varieties and how they specifically affect district production of TCH.

Leslie and Wilson (1996) investigated the effect varieties and varietal improvement had on TCH at a district level using a single variety indexing method. They found Mulgrave to have little improvement in TCH from 1960 until 1983–84. After 1984, varieties such as Q99, Q117, and Q138 contributed to a relative gain in TCH. Excluding the early 1970s, production of TCH in Tully showed a substantial increase over the 38-year period. Q99, Q117 and later varieties were top yielding in Tully. In Babinda, there was an increase in TCH until 1976 followed by a decrease in 1977, which bottomed out in 1980 and recovered in the later years. Diseases such as Pachymetra and rust were cited as the causes of the reduced production levels from 1977–1984 in Mulgrave and Babinda (Leslie and Wilson, 1996). The introduction of Q77, Q113, Q107, and Q117 in 1980 helped in Babinda's production recovery, and later varieties such as Q120, Q122, Q138, and Q152 have kept Babinda's production increasing.

KEYWORDS: REML, Mixed Model, BLUP, Variety, Mill, Combined Analysis.

An investigation of Leslie and Wilson's (1996) methodology used to calculate the varietal estimates for TCH, CCS, and TSH found that approach to be somewhat subjective. This is due primarily to the unbalanced nature of the data set as the varieties grown changed over time across districts, and to the assumption that the performance of single reference varieties relative to other varieties is stable across years. This paper proposes a more rigorous statistical methodology for regular analysis of productivity data. Restricted maximum likelihood (REML) methodology was used to calculate best linear unbiased predictors (BLUPs) and estimates (BLUEs) for the varieties, districts, years, and their interaction terms. REML was used to overcome the subjectivity created by calculating single varietal indices, and to help accommodate the missing values in the data set.

Methods

Data for the mill districts of Mulgrave, Babinda, and Tully were investigated. The data consisted of tonnes of cane per hectare (TCH) measurements for each variety grown during the years 1958–1995 in Mulgrave, 1961–1995 in Babinda, and 1960–1995 excluding 1988 in Tully. A separate estimate of TCH for each year in each district was also given.

To analyse the productivity trends of TCH over time, two approaches were taken. The first approach was to analyse the data as a complete data set by combining the data from each of the separate mills. The estimates generated by this analysis are based on information from three mills instead of one. For example, if a particular variety yielded low in one district, but high in the others an adjustment is made to obtain a general estimate over all districts for that variety. Another reason for combining the data is to obtain mill, variety × mill, and year × mill effects for TCH. This further information may enable a better interpretation of the production trend data. The second approach undertook separate analyses for each district. Varieties that grow better or worse than others in specific districts will be identified and comparisons can be made of varietal improvement within districts. However, the separate district model is only based on variety and variety × year effects. Any improvement in the district would appear as varietal improvement since mill estimates are not present in this model.

Restricted maximum likelihood (REML) methodology was used in the analysis of both combined and separate data sets. This procedure estimates the variance components associated with each effect, producing best linear unbiased predictors (BLUP) for random effects and best linear unbiased estimates (BLUE) for fixed effects. The BLUPs and BLUEs can then be ranked accordingly to investigate varietal improvement.

Using the combined data set, the REML model chosen was:

$$\text{TCH} = \text{year} + \text{mill} + \text{variety} + (\text{variety} \times \text{mill}) + (\text{variety} \times \text{year}) + \text{error}. \quad (1)$$

The varieties were considered random effects, while years and districts (mill) were treated as fixed effects in a mixed model analysis in order to make varietal comparisons for short periods. An error term was also added to account for the residuals in the model. Another model was run with variety, mill, and year effects treated as random. It showed no differences from the mixed model in the estimates produced. For the separate mill data, a similar approach was taken, treating varieties as random effects and years as fixed effects in the model:

$$\text{TCH} = \text{year} + \text{variety} + (\text{variety} \times \text{year}) + \text{error}. \quad (2)$$

The percentage of tonnes for each variety grown was used as a weight in determining the estimates for TCH. More weight was given to the varieties that had grown a greater percentage of the total tonnage for that year. These calculations were carried out using the ASREML software (Gilmour *et al.*, 1996).

Graphs for TCH estimates across years were constructed, using S-Plus software (Statistical Sciences, 1998), by taking the BLUP estimates for the appropriate effects, and weighting them by the tonnage of cane grown for a particular variety in each year. In the combined analysis, the overall mean for TCH was added to the effects, while in the separate analysis, the means for each district were added. TCH estimates incorporating the effects, variety + mill + (variety × mill) + (variety × year), are superimposed on TCH estimates incorporating variety + mill + (variety × mill). The latter give smoother trends for TCH changes over time. These graphs were chosen as the most informative ways of looking at the varietal response over years. In the separate analyses, the same procedure was followed to provide TCH estimates for the variety (smooth trend) and variety + (variety × year) (rough trend) effects.

Results and discussion

Variety evaluation

Table 1 gives the estimated effect of the varieties calculated by the BLUPs plus a mean of 79.21 with a standard error of 1.01 in the combined analysis. Estimated effects calculated with the BLUPs using the separate analyses plus means of 81.85, 72.09, and 83.80 with standard errors of 1.12, 2.21, and 1.59 in Mulgrave, Babinda, and Tully, respectively, are also given.

In the combined analysis, the TCH BLUPs for the varieties have improved since the varieties current in 1958 in Mulgrave, 1961 in Babinda, and 1960 in Tully, but little or no improvement is indicated since the introduction of Q99 in 1976 (Table 1). The separate analyses gave similar results in that three of the top five varieties in the combined analysis are also superior performers in each of the mill districts. Q99 was the top variety in Mulgrave. Q101 was the top in Babinda, and Q152 was the top in Tully. Q99 ranked seventh in both Babinda and Tully (Table 1). Differences between the combined and separate BLUPs for individual varieties can be seen. Separate BLUPs are generally based on fewer mill/years. For example, the BLUP for Q101 in Babinda is based on three mill/years while that for Q130 uses five mill/years. The same holds true for the BLUPs for varieties Q152 and Q149 in Tully which are based on two and three mill/years, respectively. Thus, these estimates have larger standard errors and are less reliable than those estimates with smaller standard errors. By combining the mill districts, a

Table 1—A ranking of the estimated effect calculated by the BLUPs + mean for the top five varieties in the combined and separate analysis for the districts of Mulgrave, Babinda, and Tully. When Q99 is not included in the top five varieties its BLUP + mean and rank are included.

Combined analysis TCH Variety	Estimated Effect	BLUP-se	Number of Mill/Years	Cumulative % of Tonnage
Q99	87.66	2.55	37	217
Q120	86.53	2.45	34	444
Q124	86.34	2.99	17	125
Q138	85.89	2.64	22	294
Q117	85.88	2.33	36	884
Separate analysis TCH: Mulgrave Variety	**Estimated Effect**	**BLUP-se**	**Number of Mill/Years**	**Cumulative % of Tonnage**
Q99	93.99	2.28	11	89
Q113	91.85	1.94	15	336
Q124	89.30	2.45	9	96
Q100	88.90	2.89	8	22
Q117	88.55	2.00	14	340
Separate analysis TCH: Babinda Variety	**Estimated Effect**	**BLUP-se**	**Number of Mill/Years**	**Cumulative % of Tonnage**
Q101	91.13	5.30	3	10
Q120	89.09	3.50	11	187
Q138	87.79	3.98	7	125
Q117	86.16	3.39	12	250
Q130	84.18	4.42	5	26
*Q99 (7th)	81.40	3.65	10	33
Separate analysis TCH: Tully Variety	**Estimated Effect**	**BLUP-se**	**Number of Mill/Years**	**Cumulative % of Tonnage**
Q152	99.32	4.09	2	8
Q149	96.87	3.64	3	7
Q124	95.73	2.58	7	27
Q117	94.48	2.27	10	294
Q138	92.92	2.55	7	65
*Q99 (7th)	92.13	2.14	16	96

larger number of mill/years are used and this decreases the standard errors of the estimates.

Table 2 represents the cumulative percent of tonnage grown in each year for the top five ranking varieties listed in Table 1. The combined tonnage was calculated by averaging the variety tonnage for top five varieties (listed under the combined analysis in Table 1) over the three mill districts. The cumulative percent of tonnage for the top five varieties in each year for each district are also listed. The increased percent tonnage over years indicates that greater proportions of the areas in each district are being allocated better yielding varieties.

Table 2—The cumulative percent of tonnage grown in each year for the top five ranking varieties listed in Table 1. The tonnage for the combined analysis varietal rankings are averaged over the three mill districts.

Year	Combined analysis % of tonnage	Mulgrave % of tonnage	Babinda % of tonnage	Tully % of tonnage
58–74	0	0	0	0
75	0	0	3.6	0
76	2.8	1.3	3.7	0
77	5.1	7.1	2.7	0
78	8.0	11.0	0	0
79	8.9	11.2	0	0
80	9.1	17.1	0	0
81	8.1	28.7	0	0
82	9.6	49.7	0	0
83	12.2	59.0	0	0
84	16.8	68.7	3.5	0
85	24.5	69.5	17.1	0
86	31.9	66.9	28.1	12.2
87	39.3	63.2	37.3	27.9
88	31.7	49.1	47.0	NA
89	58.9	62.9	56.2	52.9
90	63.9	62.1	61.6	55.8
91	65.0	58.5	63.6	54.4
92	67.2	54.4	65.7	51.0
93	67.9	49.7	72.1	49.1
94	62.1	42.2	60.3	46.9
95	62.4	38.3	65.7	48.0

Time trend analysis

Figure 1 shows the TCH trend over time for each of the three mills as explained in the methods section for the combined analysis. Mulgrave remained stable from 1958 to 1970 with a mean TCH of 79.2. The TCH then increased from 1971 to 1985 and then levelled off around 89.2. Tully, starting with a TCH value of 75.6 in 1960, showed a gradual increase from 1960 to 1987 levelling off around 87.8. In Babinda, TCH values started at 71.6 in 1961 and remained stable through 1971, decreased through 1972 to 1979 with a mean TCH of 70.3, and then increased from 1980 to 1993 peaking at 79.1 and seems to be levelling out.

Figure 2 represents the graphical interpretation of the separate analyses. The trend in Mulgrave is similar to the combined analysis. TCH remained stable from 1958 to 1970 with an average TCH of 79.4, then increased from 1971 to 1985 with an average TCH of 85.1, and levelled off in 1986 to 1995 with TCH averaging 87.7. Tully also had a similar trend from 1960 to 1980 when compared to the combined analysis.

Tully yields gradually increased from 74.7 TCH in 1960 to 81.6 TCH in 1980. TCH increased greatly from 82.3 in 1981 to 89.3 in 1987, and then levelled off with a mean of 91.4 in the last seven years. The separate analysis trend for Babinda showed the greatest difference from the combined trend. Babinda showed a gradual increase in TCH from 63.6 in 1961 to 74.8 in 1984, more rapid increase from 1985 to 1990 with TCH going from 77.3 to 84.0, and then levelled off with a mean of 84.3 in the last five years. Both the combined and separate analyses show evidence of TCH depression in 1977–1984 period for all of the districts. Leslie and Wilson (1996) identified this depression for Mulgrave and Babinda, and attributed the depression to the effects of Pachymetra and rust on Q90.

There were apparent differences between the combined and separate analyses of productivity for the Babinda district. Although the TCH trends between the two analyses are not statistically different in any given year, there appears to be an upward trend in the separate analysis in Babinda that is not apparent in the combined analysis. This difference may be real but it could be due to the large standard error of the estimates based on the separate mill analyses. The BLUPs estimates for Babinda had inherently higher standard errors, for reasons unknown, when compared to Mulgrave and Tully. Another reason for this disparity could be in the estimation of the mill and mill interaction effects. In the separate analyses, if the BLUPs for the varieties remain constant, i.e. no improvement, improvement of the mill implies improvement of the varieties. In the combined analysis, differences of mill improvement would be accounted for in the variety, mill, and (variety × mill) effects.

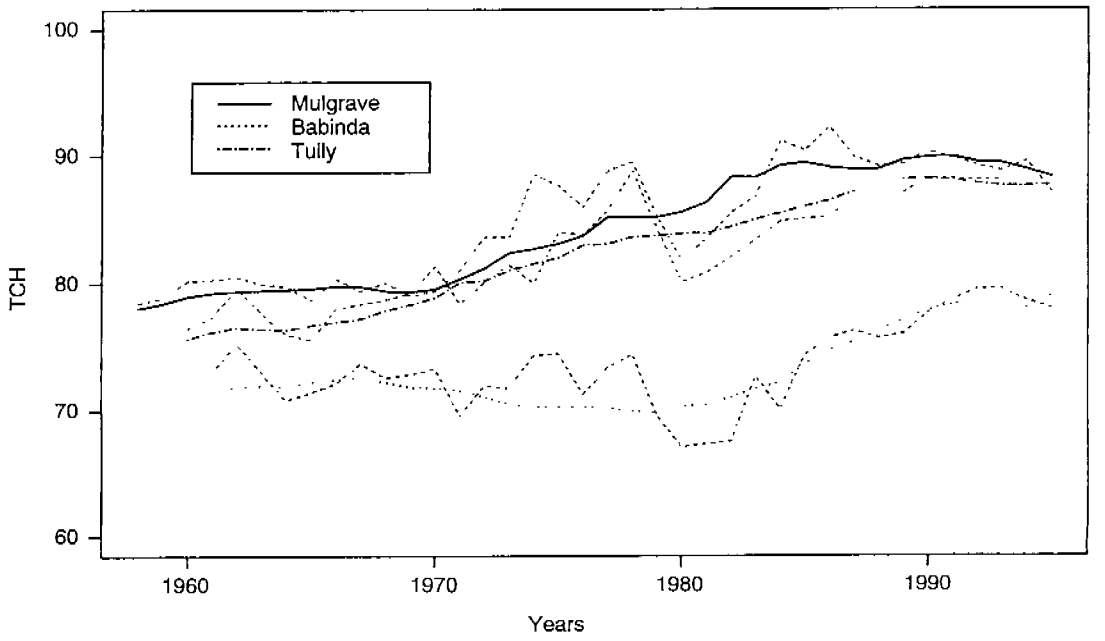

Fig. 1—TCH productivity over a 38-year period for the districts of Mulgrave, Babinda, and Tully using a combined analysis. The smoothed trends are denoted by the legend, while the rough trends are plotted around the smoothed trends in each district.

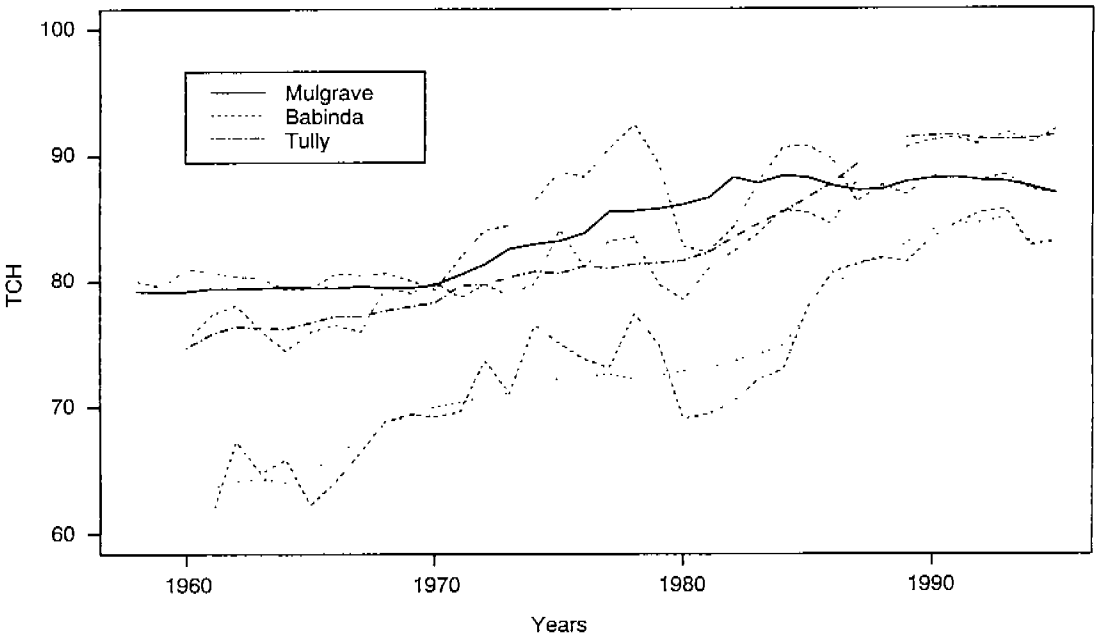

Fig. 2—TCH productivity over a 38-year period for Mulgrave, Babinda, and Tully using separate mill analyses. The smoothed trends are denoted by the legend, while the rough trends are plotted around the smoothed trends in each district.

Raw data analysis

A separate average of TCH over varieties for each year in each district was plotted in Figure 3. This estimate is based on the raw data giving a single TCH value for each year in the each district. As expected, the variation among years is high. Fitting non-linear trends associated with these raw data proved to be non-significant. Leslie and Wilson (1996) did find linear trends to be significant, but only explained 5–37% of the total variation. In these circumstances, the estimates obtained by Leslie and Wilson were reported as percent increases in TCH from 1960 to 1995 of 22.9% in Mulgrave, 20.8% in Tully, and 10.8% in Babinda. They also derived estimates of the gains due to varietal improvement in TCH of 12.1%, 30.1%, and 26.1% for each district, respectively. Although it is not statistically rigorous to make direct comparisons using the REML analyses, the results do not dispute the findings of Leslie and Wilson (1996). The REML analysis found Mulgrave generally had higher TCH values followed by Tully and Babinda. Mulgrave and Tully had a slight increase over the 38 years, while Babinda showed less evidence of an increase. Year estimates contributed a greater percent of the total variation in the model followed by variety × year, mill, variety × mill, and variety estimates.

Data limitations

Varietal productivity data aggregated at district levels confound true effects of variety on TCH with unknown effects due to other factors and to varietal interactions with those factors, where the weighted distributions of varieties in respect of those factors are not fully and equally random. An example is where one variety is grown predominantly on well drained, fertile soils of a district and another predominantly on poorly drained infertile soils of the same district. In this example, some of the apparent TCH superiority of the former variety will be due solely to soil fertility differences, some to variety × soil interactions and some to real differences between the yielding abilities of the two varieties.

Historical productivity data do not allow removal of the confounding effects of factors such as soil fertility, drainage, microclimate, stand density, time of harvest and plant and ratoon classes from varietal comparisons. Nor can it be determined whether varieties have been differentially distributed in their exposures to any of these factors.

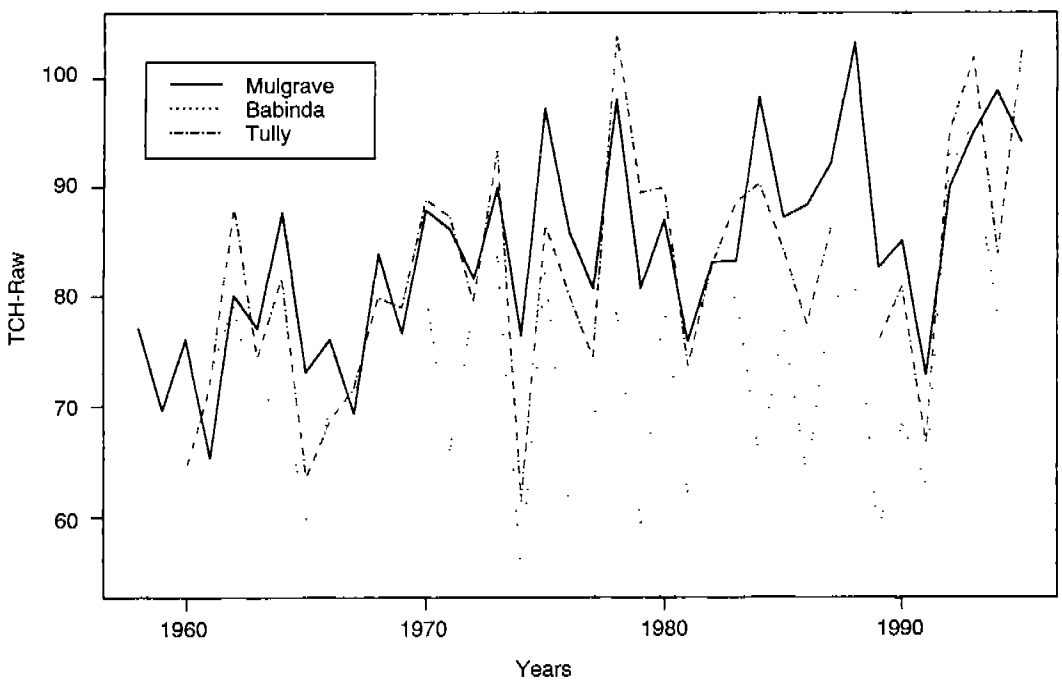

Fig. 3—TCH raw data over a 38-year period for Mulgrave, Babinda, and Tully.

The analyses reported in this paper, and most other analyses or uses of productivity data, make an assumption that all varieties have been similarly distributed. One must therefore caution interpretations of productivity data which attribute differences or changes solely to variety. Conversely, there have been no other analyses to quantify the responses or adaptations of all varieties to locational or husbandry variations within mill districts. Caution should be applied, therefore, to claims that specific adaptations have caused differences in varietal performance as measured from productivity data.

The problem is one that may be resolved by further research and if future annual block data includes and retains soil and other information on relevant factors. There will, however, be a limit imposed by the increase in standard errors of estimates as the number of factors being differentiated increases. That limit is not yet known.

Conclusions

Restricted maximum likelihood methodology (REML) has been shown to be applicable in the analysis of productivity data at the district level. It enables a more rigorous estimation of productivity trends over time when adjusting for variety and variety × environment interactions than the previous method used by Leslie and Wilson (1996). TCH was found to have increased in each district since 1960, but seems to be levelling off in recent years. It appears that the increased TCH production is mainly due to a greater proportion of the area in the regions being allocated better yielding varieties. The utility of these findings should be interpreted with caution due to unknown confounding of varietal effects of other factors.

Acknowledgements

A special thanks to the Sugar Cane Research and Development Corporation (SRDC) for funding this research.

REFERENCES

Gilmour, A., Thompson, R., Cullis, R.B. and Welham, S. (1996). ASREML. Technical Report, NSW Agriculture, Wagga Wagga, NSW, Australia.

Leslie, J.K. and Wilson, G.L. (1996). Productivity Trends in Sugarcane in the Wet Tropics, Technical Report, Sugar Research and Development Corporation, Queensland, Australia.

Statistical Sciences (1998). S-PLUS Guide to Statistical and Mathematical Analysis, Version 4.5. StatSci, a division of MathSoft, Inc., Seattle.