Modeling Pedestrian Volumes on College Campuses

LAURA L. COVE AND J. EDWIN CLARK

A study was undertaken to develop a reliable method for obtaining reasonable estimates of pedestrian volumes on college campuses using short-term volume counts. Pedestrian volume data were collected on the campuses of five colleges and universities in the southeastern United States. The short-term pedestrian volume data collected on the five college campuses were expanded using existing models and analyzed to test the validity of these models. The average sums of the errors between the observed and predicted pedestrian volumes calculated from the existing models were compared with the percentage errors calculated for the additional data collected from the college campuses. The percentage errors were not found to be statistically the same, thus the existing models are not valid for use in expanding short-term pedestrian volume counts on college campuses. Using a random sample of the pedestrian volume count observations from the college campuses, additional expansion models were developed. These models were chosen on the basis of the evaluation of the parameters of the coefficient of determination, R^2, and the standard error about the mean, SE_y. These new models were validated using the remaining data that had not been incorporated into the development of the new models. On the basis of the analysis of the data from the college campuses, the following conclusions were made: (a) existing expansion models were not valid for predicting pedestrian volumes on college campuses, (b) the count interval should be started 10 min before the beginning of class periods, (c) accuracy increased as the volume count period increased, and (d) accuracy increased as the prediction volume range increased.

The increased national awareness on behalf of the pedestrian in recent years has prompted an increase in the amount of research dealing with pedestrian safety issues. However, only a few such research projects were conducted in order to develop a reliable method for measuring pedestrian volumes. Estimates of pedestrian volume counts are needed to evaluate pedestrian safety. To obtain an effective analysis and to determine the relative hazard of various pedestrian behaviors, comparisons must be made between pedestrian behavior during accidents and normal, non-accident pedestrian behavior. The normal, non-accident behavior is designated as pedestrian exposure information. Usually to obtain this exposure information, manual pedestrian volume counts must be taken. Since these counts are typically undertaken for an entire day, they are very labor intensive and thus very costly. Other techniques for obtaining pedestrian volume counts include sampling over shorter time periods and using automated counting devices and analytical methods. However, with the exception of manual pedestrian counting, none of these techniques has been universally accepted by the research or user community.

L. Cove, Federal Highway Administration, 310 New Bern Avenue, Suite 410, Raleigh, N.C. 27611. J. E. Clark, Department of Civil Engineering, Clemson University, 110 Lowry Hall, Clemson, S.C. 29634.

PURPOSE AND OBJECTIVES

In an attempt to reduce the costs of collecting data on pedestrian volume, a method developed by Mingo et al. uses the practice of making short sample counts of pedestrian volumes that can be expanded to represent daily volumes through the use of appropriate expansion factors (*1*). The purpose of this study was to examine actual daily pedestrian volume counts of pedestrians on selected college campuses and test the validity and reliability of pedestrian volumes obtained through the use of the existing expansion models.

To accomplish the purpose of this study, the following objectives were established:

1. Identify specific sites and collection of field data consisting of pedestrian volumes for 5-min increments at five college campuses,
2. Expand sample field data through the use of existing expansion models based on the count interval and level of accuracy desired,
3. Perform statistical analysis of the actual data and the expanded data to check the validity of the expansion model, and
4. Develop additional expansion models that accurately predict pedestrian volumes on college campuses from short-term volume counts.

METHOD OF INVESTIGATION

The study was conducted by first identifying a particular segment of the population on the basis of one characteristic of the population. This chosen characteristic is age, and the ages being studied range from 17 through 24. The individuals within this age range make up 11 percent of the total population of the nation.

Normally, it is extremely difficult to conduct a study that requires the isolation of a particular segment of the population, including the aforementioned age range. However, most pedestrians 17 through 24 are concentrated on the campuses of colleges and universities throughout the country. According to the *Almanac of Higher Education*, of the 247,732,000 persons in the United States, 12,768,307, or 5.2 percent, attended a 2- or 4-year college or university in 1989 (*2*). The primary mode of travel on the campuses of colleges and universities is that of walking, so there is a high density of pedestrian movement at these locations. Pedestrians on campuses exhibit fairly uniform pedestrian characteristics. In addition to similar ages, these pedestrians have common trip purposes and predictable flow variations by time of day on the basis of class schedules. Because of these similar characteristics and the fact that these pedestrians are isolated within a relatively small geographic area of a region, city, or town on a campus, similar traffic patterns exist.

On the basis of similar characteristics, five universities were selected for field data collection: the University of Miami (Coral Gables, Florida), the University of Georgia (Athens), the College of Charleston (Charleston, South Carolina), Clemson University (Clemson, South Carolina), and the University of South Carolina (Columbia).

The data on pedestrian volumes were collected on Fridays during March and April 1991. Pedestrian count volumes were totaled and recorded in 5-min intervals for 12 hr, beginning at 7:00 a.m. and continuing through 7:00 p.m., on each campus. The data were then input into the models developed by Mingo et al., and the percentage errors were calculated by comparing the actual data with the expanded model data (*1*). These errors were then compared with the errors computed by Mingo et al. (*1*). Significance tests were used to determine whether the errors were statistically the same, therefore testing the validity of the models. If the models were not found to be valid, additional expansion models were to be developed using random samples of the observations from the college campuses. The validity of these additional models would be tested using data from the college campuses that had not been incorporated into the development of the models.

METHODOLOGY

Today three methods are generally recognized for measuring pedestrian volumes: mechanical counts, mathematical models, and manual counts. Several mechanical pedestrian counters have been developed and tested, but unfortunately they have not been widely accepted because of their excessive cost and various installation problems. Mathematical models are usually site specific, and the accuracy of the model depends on the type and amount of data collected. As the complexity and accuracy of the model increases, the data collection costs rise rapidly. Thus, the most commonly used method for obtaining pedestrian volume counts is manual counting, a procedure that is labor-intensive and therefore expensive.

College campuses throughout this country and others exhibit a unique problem. Large volumes of pedestrian traffic on these campuses must compete with automobile traffic. Most college campuses are designed with the intention of creating a park-like area within the boundaries of the university community. The presence of automobile traffic presents potential hazards to pedestrians and threatens the relaxed atmosphere.

Pedestrians on college campuses have different characteristics from pedestrians in a central business district or shopping district. The pedestrians' ages generally range from 17 to 24. Their flow variations and the purpose of their trips are predictable depending on the time of day.

Because of the large volume of pedestrian traffic that occurs on most university campuses, special techniques must be applied to provide for safe and efficient movement of persons traveling through the campus by this mode.

COLLECTION OF DATA

Introduction

The data used by Mingo et al. in developing the pedestrian volume expansion models were collected entirely in Washington, D.C., which creates the possibility of a limitation in the models by using data from only one city. There has been no determination that the expansion models are valid when used in another city that exhibits different characteristics than Washington does. To test the validity of these expansion models, data were collected at several sites and then input into the models, thus testing the validity by comparing the percentage errors calculated in the study by Mingo et al. with the percentage errors calculated for the additional data. If these percentage errors are found to be statistically similar, then the models developed previously would be validated. It was decided then to collect pedestrian volume data on the five college campuses in an attempt to verify the validity of the expansion models.

Data Collection Method

Pedestrian volume data were collected on Fridays during March and April 1991 at the five universities. Fruin suggested counting pedestrian volumes on "typical" days, free from the distortions of weather and other seasonal effects (*3*,p.122). Unusually hot, cold, or inclement weather keeps people off the streets and away from the counting area.

The locations of the collection sites at the universities were selected to ensure that pedestrian volumes were large enough to allow enough data to be collected within the limited resources of the study. A 100 percent sample of pedestrians crossing the location was taken at each site during each 12-hr data collection period. These 12-hr samples consisted of continuous counts that were made at each site by one data collector from 7:00 a.m. until 7:00 p.m. Pedestrian volumes were counted for 12-hr periods to ensure that all volumes during the peak travel times would be encompassed within the data. Figure 1 shows the variation in pedestrian volumes by hour of day for each of the campuses. Pedestrian volumes exhibited maximum peaking from 10:00 a.m. to 1:00 p.m.

Pedestrian volumes were recorded every 5 min of the 12-hr period. The crossing volumes were recorded either by crosswalk or by each leg of the intersection, and a total volume for the intersection was then calculated.

ANALYSIS OF DATA

Introduction

The data collected on the five college campuses were expanded using the models developed by Mingo et al. and an-

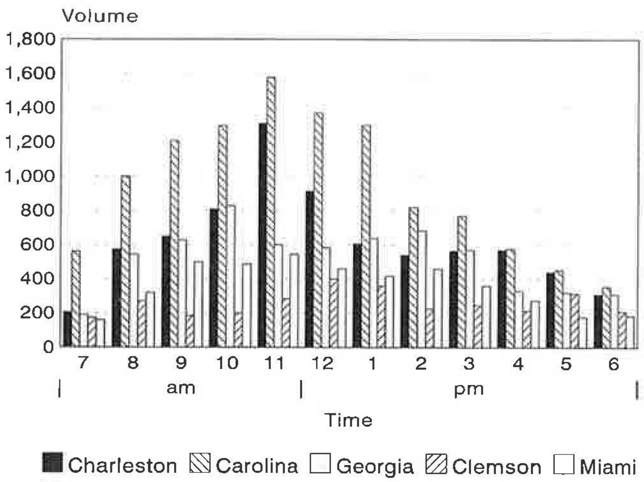

FIGURE 1 Pedestrian activity on college campuses.

alyzed to test the models' validity. The percentage errors calculated in the study by Mingo et al. were compared with the percentage errors calculated for the additional data. If these percentage errors were found to be statistically the same, then the models developed previously would be valid. Because these models did not produce suitable results, additional models were developed using a random sample of the pedestrian volume count observations from the college campuses. The new models were validated using the remaining data from the college campuses that had not been incorporated into the development of the models. Dependent on the results obtained through the use of either the existing models or the newly developed models, a reliable method for obtaining reasonable estimates of pedestrian volumes on college campuses was endorsed.

Pedestrian Volume Count Expansion Procedure

A seven-step procedure was used in the study of the five college campuses to develop the estimating procedure:

1. Select time period to estimate pedestrian volumes.
2. Select the count interval.
3. Develop data collection plan.
4. Collect data.
5. Select expansion model.
6. Compute estimated pedestrian volumes.
7. Determine estimated pedestrian volume ranges.

Verification of Validity of Existing Models

The collection of pedestrian volume data on the five campuses resulted in 60 hr of pedestrian volume data. Data were collected continuously every 5 min for 12 hr at each site, yielding a data base of 720 observations. In testing the validity of the 1-, 2-, 3-, and 4-hr expansion models developed by Mingo et al. (*1*), random intervals of 5, 10, 15, and 30 min were input into the appropriate models and 1–, 2-, 3-, and 4-hr volume predictions were calculated. For various predicted volume ranges, the average percentage differences were calculated for each count interval expansion model.

The research by Mingo et al. evaluated several count intervals and the position of the events within the interval. For all count intervals, the middle event produced a slightly better model since it exhibited the highest coefficient of determination (R^2) and the lowest standard error about the mean (SE_y). Since these models were developed in a large metropolitan area, the pedestrian volumes were fairly uniform throughout the hour and the entire day. On college campuses, the pedestrian traffic is dependent on class schedules, thus it is conceivable that the placement of the interval within the hour could have a significant effect on the accuracy of the expanded count. Haines et al. analyzed pedestrian volumes on the Boulder campus of the University of Colorado and determined that in all the volume counts, the peak periods were from 10 min before the beginning of the class periods until 5 min after classes began (*4*). Because fewer classes are scheduled in the late afternoon than in the morning, the afternoon volumes were less variable and thus a better predictor of normal flow.

On the basis of the recommendation by Haines et al., the placement of the intervals of the 5-min volume counts to be input into the models was selected to be 10 min before the beginning of classes. The predicted volumes were calculated using the expansion models and procedures of Mingo et al. described earlier and then compared with the observed pedestrian volumes. Percentage errors were calculated by subtracting the predicted volumes from the observed volumes and dividing by the observed volumes. The absolute value of the average sums of the errors between the observed and predicted pedestrian volumes for each of the count intervals that corresponded to the given volume ranges was computed. Basically, these factors are percentages that indicate the relative levels of accuracy of an expanded sample crosswalk count.

One method used to test the ability of the expansion models developed by Mingo et al. to predict accurate pedestrian volumes on college campuses is to compare these range factors for both sets of data. If the percentage errors, represented by the range factors, are found to be statistically the same, then the existing models are valid.

Because the sample sizes of the data are different, a statistical test that incorporates sample size had to be used. For the 1-, 2-, 3-, and 4-hr expansion models, hypothesis testing was performed. For the 1- and 2-hr models, z-tests were performed because the sample size was greater than 30, and t-tests were performed for the 3- and 4-hr models because the sample size was less than 30. To determine which of the expansion models were valid, these tests were done for each range factor corresponding to the count interval and the predicted pedestrian volume. Therefore, 27 individual hypothesis tests were performed.

Development of Additional Models

Even though the existing models appeared to produce fairly good results as far as the range factors were concerned, they were not proven to be valid for predicting pedestrian volumes on college campuses. The flow of pedestrians on college campuses is predictable on the basis of variations of time of day, but it is not as uniform as pedestrian flow in a large metropolitan area, such as Washington, D.C., where the existing models were developed. Additional models should be developed with a primary focus on these models' ability to accurately predict pedestrian volumes on college campuses.

In the initial stages of the model development, the data distributions were reviewed and it was determined that all the variables showed positive skewness. In order to perform a regression analysis on the data, the data must be normally distributed. The data could be analyzed using a distribution free nonparametric test, or the data could be transformed so that parametric tests could be performed. The use of parametric tests is more desirable since such tests are more powerful than nonparametric tests. Some statisticians have used transformation processes to normalize their data when normality of data is required, even though other statisticians argue that this transformation process is not completely understood and is nothing more than a manipulation of the data to fit the model. For this study, the logarithms were calculated for all observations for all of the variables in order to transform the data to produce a normal distribution.

Regression analysis was performed on all four count intervals. Two-thirds of the observations of the pedestrian volume data from the college campuses were used to develop these models, and the remaining third were used to verify the models and to develop the range factors associated with the expansion models. The count interval beginning 10 min before classes started was used in the development of all the expansion models. Using the ability of the Quattro Pro software package (*5*) to perform regression analysis, 16 models were developed using the pedestrian volume data from the five college campuses. Table 1 was constructed to examine the R^2 and SE_y for each count interval for all of the models.

By examining the values represented in Table 1, an estimate of how well the data fit the models can be determined. SE_y is the estimated standard error of the y-values and represents the deviation of the observed y-values from the values of the linear combinations represented in the models. R^2 is a statistic that measures the validity of the model. It ranges from 0 to 1, 1 being perfect correlation. It is apparent that as the count interval increased from 5 to 10 to 15 to 30 min, the prediction models became more reliable, as expected since the variation between the count intervals decreased as the count interval increased.

TABLE 1 Regression Output of College Campus Prediction Models

Prediction Model	Count Interval (min)			
	5	10	15	30
1 hr				
R^2	.80	.88	.92	.91
SE_y	.11	.08	.07	.07
2 hr				
R^2	.86	.91	.93	.93
SE_y	.09	.07	.06	.06
3 hr				
R^2	.83	.91	.96	.96
SE_y	.09	.06	.04	.04
4 hr				
R_2	.86	.93	.95	.96
SE_y	.09	.06	.05	.05

RESULTS

Models Developed for Use on College Campuses

Models were developed for the purpose of predicting pedestrian volumes on college campuses by expanding short-term counts. These models were chosen on the basis of the evaluation of the parameters of R^2 and SE_y. The expansion models developed for the count interval beginning 10 min before the start of classes are presented in Figure 2.

Validation of College Campus Expansion Models

To have data available to validate these new models, several observations were excluded from the modeling effort. Twenty observations were used in the validation of the 1-hr models, 10 observations in the 2-hr models, 10 observations in the 3-hr models, and 5 observations in the 4-hr models. All four counting intervals—5, 10, 15, and 30 min—were studied for each model.

The primary use of the validation of the models was to determine the percentage error in the predictions of the pedestrian volume counts. The value of SE_y is used for this purpose, since SE_y bands diverge at the ends of the regression line as the values of X move away from the mean of X ($\bar{X}$). The SE_y bands may become extremely separated when X moves far away from $\bar{X}$, and thus renders the SE_y meaningless. Therefore, the use of percentage change between actual and predicted volume counts was used to determine empirically the error or prediction ranges associated with the expansion models.

For various predicted volume ranges, the average percentage differences were calculated for each count interval in each expansion model. These range factors are presented in Table 2. The volume ranges increased in size as the expansion model increased from 1 to 2 to 3 to 4 hr because of the increase of the volume sizes being predicted and the number of observations per range.

To use these prediction range factors, first select the volume level (row) that corresponds with the count period and the estimated volume from Step 6. Select the sample count interval (column) that was used. Read the prediction range factor, and the estimated volume range will be the estimated volume (Step 6) plus or minus the prediction range factor multiplied by the estimated volume. An example to illustrate this process is shown in the following.

Using a 5-min count, predict the 4-hr pedestrian volume prediction. A 5-min count might be equal to 50 pedestrians, therefore $I5 = 50$. The appropriate expansion model to use for a 4-hr pedestrian volume prediction on the basis of a 5-min volume count from Figure 2 is $V4$ = INVLOG [.74408 log ($I5$) + 2.047303]. The estimated volume will be equal to 2,049 persons per 4 hr. The range of values that one may assume the value will actually fall is equal to 2,049 plus or minus 23 percent of 2,049, where 23 percent is obtained from

```
One-Hour Prediction

        PED5M:    V1  = INVLOG [0.709564 log (I5) + 1.5108]
          where   V1  = one-hour prediction in persons per hour
                  I5  = the specified five-minute count

        PED10M:   V1  = INVLOG [0.749178 log (I10) + 1.241982]
          where   I10 = the specified ten-minute count

        PED15M:   V1  = INVLOG [0.808811 log (I15) + 0.996939]
          where   I15 = the specified 15-minute count

        PED30M:   V1  = INVLOG [0.902426 log (I30) + 0.55304]
          where   I30 = the specified 30-minute count

Two-Hour Prediction

        PED5M:    V2  = INVLOG [0.743682 log (I5) + 1.749562]

        PED10M:   V2  = INVLOG [0.76066 log (I10) + 1.514637]

        PED15M:   V2  = INVLOG [0.896754 log (I15) + 1.296608]

        PED30M:   V2  = INVLOG [0.897296 log (I30) + 0.864096]
          where   V2  = two-hour prediction in persons per two
                        hours

Three-Hour Prediction

        PED5M:    V3  = INVLOG [0.79884 log (I5) + 1.835829]

        PED10M:   V3  = INVLOG [0.840315 log (I10) + 1.541358]

        PED15M:   V3  = INVLOG [0.879492 log (I15) + 1.325787]

        PED30M:   V3  = INVLOG [0.992658 log (I30) + 0.807528]
          where   V3  = three-hour prediction in persons per
                        three hours

Four-Hour Prediction

        PED5M:    V4  = INVLOG [0.74408 log (I5) + 2.047302]

        PED10M:   V4  = INVLOG [0.762558 log (I10) + 1.811265]

        PED15M:   V4  = INVLOG [0.797503 log (I15) + 1.618377]

        PED30M:   V4  = INVLOG [0.908667 log (I30) + 1.138706]
          where   V4  = four-hour prediction in persons per
                        four hours
```

FIGURE 2 Prediction models used for expansion of short-term volume counts on college campuses.

Table 2. The 4-hr pedestrian volume prediction will therefore be between 1,577 and 2,520.

In reviewing the table, it is seen that the percentage error decreased as the count interval increased in all cases but two. As was determined earlier, during the modeling effort, the longer count intervals had higher values of R^2 and appeared to be better predictors of accurate pedestrian volumes. The previous finding was supported here, since the average percentage differences decreased as the count interval increased.

As the volume ranges increased, the percentage error was reduced except for the 3-hr prediction model. At low-volume sites, the flow of pedestrians is often erratic, thus causing large peaks and valleys over short time intervals. Cameron estimated that the probability of sampling at a volume peak or valley was approximately 50 percent, thus decreasing the potential of obtaining a true representative sample of the overall volume (*6*). At a site with high pedestrian volumes, the flow is more uniform from one time interval to the next, therefore a sample taken from a high-volume site is often more representative of the accurate volume than a sample taken from a low-volume site.

It would also appear that as the prediction period increased from 1 to 2 to 3 to 4 hr, the prediction would become less accurate on the basis of the variation that exists with small sample intervals. However, in the college campus prediction models, the opposite was true, and as the prediction period

TABLE 2 Accuracy of Predicted Pedestrian Volumes Using College Campus Prediction Models for 1-, 2-, 3-, and 4-hr Predictions

Pedestrian Volume Level	Range Factor (%) Count Interval (min) 5	10	15	30
1-hr prediction				
0–500	33	32	20	16
> 500	22	18	16	8
2-hr prediction				
0–500	27	31	20	10
> 500	25	20	12	10
3-hr prediction				
0–1,500	24	14	6	18
> 1,500	28	27	26	14
4-hr prediction				
0–1,500	23	20	14	11
> 1,500	23	19	11	8

increased, the predictions became more accurate. The reasons for this occurrence are that for a 1-hr prediction, an individual 5-, 10-, 15-, or 30-min interval was input into the expansion model, whereas with a 2-hr prediction, two intervals were averaged and input into the expansion model. For the 3-hr prediction, three intervals were averaged and input into the model, and for the 4-hr predictions, four intervals were averaged and input. This averaging of the intervals practically eliminated the possibility of sampling during a peak or valley, thus reducing the variation of the pedestrian volumes and producing a better estimate of the pedestrian volumes than did the actual volumes.

CONCLUSIONS AND RECOMMENDATIONS

In conclusion, the use of expansion models in predicting pedestrian volumes on college campuses has shown promise. With respect to practicality, the pedestrian volume sampling method offers a technique that results in a significant savings of time and effort.

On the basis of the results obtained in this study, the following conclusions relative to the use of expansion modeling for accurately predicting pedestrian volumes on college campuses were developed:

1. Expansion modeling from short-term pedestrian volume counts is a promising alternative to direct observation manual counting on college campuses.
2. The counting interval position for use in expansion modeling should be 10 min before the beginning of class periods.
3. The volume predictions became more accurate as the counting period increased from 1 to 2 to 3 to 4 hr.
4. The accuracy of the prediction models also increased as the prediction volume range increased, because at low-volume sites, an erratic occurrence of volume peaks and valleys was apparent.
5. The 1-hr error estimates are fairly high, and thus since 1-hr counts are fairly economical to obtain, direct manual counting for this period is suggested.

On the basis of the study described herein, the following recommendations were developed:

1. Additional research should be undertaken by collecting data at several universities for further testing of the validity of the models developed in this study.
2. Data may also be collected at the five universities selected in the development of this model and used to develop additional models that could be compared to test the reliability of the models developed in this study.
3. The models may also be tested by selecting positions of the counting intervals that have positions different from the recommended period 10 min before the beginning of classes.

REFERENCES

1. R. Mingo, H. D. Robertson, and S. E. Davis. *Measuring Pedestrian Volumes and Conflicts.* Report FHWA-RD-88-036. FHWA, U.S. Department of Transportation, Dec. 1988.
2. *The Almanac of Higher Education, 1989–1990.* University of Chicago Press, Ill., 1989, pp. 1–40.
3. J. J. Fruin. *Pedestrian Planning and Design.* Metropolitan Association of Urban Designers and Environmental Planners, Inc., New York, 1971.
4. G. Haines, R. Kochevar, and V. H. Surti. Analysis of Campus Traffic Problems. In *Transportation Research Record 498*, TRB, National Research Council, Washington, D.C., 1974, pp. 1–12.
5. *Quattro Pro Software.* Borland International, Inc., 1989.
6. R. M. Cameron. Mechanical Measurement of Pedestrian Volumes. In *Transportation Research Record 498*, TRB, National Research Council, Washington, D.C., 1974, pp. 13–19.

Publication of this paper sponsored by Committee on Pedestrians.