DATA ERRORS IN URBAN TRAFFIC-CONTROL SYSTEMS

Philip J. Tarnoff, Federal Highway Administration

The presence of data errors in a traffic-control system is unavoidable. These errors result from the inadequacies of the surveillance system, inherent characteristics of the vehicle traffic, and inaccuracies in modeling the traffic system. If these errors are not controlled during system design and implementation, they can cause degradation of system operation to the point where it is less effective than that of a pretimed system. One of the measures that can be taken to prevent this is the design of a surveillance system that introduces errors that are no greater than the errors introduced by the other elements in the system. A second measure is the collection of data before system design that will permit identification of the parameters that must be varied on a time-of-day and link-specific basis in the prediction and optimization algorithms. This paper emphasizes the errors associated with the processing of vehicle volumes because the effectiveness of the control strategy depends most on the accuracy of this variable. Consideration is also given to the limitations inherent in the prediction process and the effect of system errors on vehicle delay at controlled intersections.

•THE URBAN Traffic Control System (UTCS) is a computer-controlled traffic signal system that has been installed by the Federal Highway Administration for developing advanced traffic-signal control strategies. The system development began in 1968 and continues at the present. A fully operational traffic-control system of 114 intersections has been installed in Washington, D.C. The system has been implemented to serve as a research facility to support the development of advanced control strategies that respond automatically to changes in traffic demand. To support these strategies, the design has included expanded detectorization, display, and data processing equipment beyond that that would be found in an operational system.

The surveillance system consists of approximately 500 loop detectors that have been installed to measure vehicle presence. From the detector outputs, the data processing system derives:

1. Volume—number of vehicles per lane per unit of time;
2. Occupancy—percentage of time of vehicle presence that is measured by the detectors;
3. Speed—average rate at which vehicles cross the detectors (this variable is proportional to occupancy divided by volume);
4. Queue length—number of vehicles waiting at the intersection approach at the end of the red phase;
5. Stops—number of vehicles on an approach that are required to wait for the red (this variable differs from queue length in that it represents the cumulative numbers of vehicles stopped over a 15-min period); and
6. Delay—estimated cumulative time that stopped vehicles are required to wait for the red.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Recent experiments related to the development of traffic-control strategies for UTCS have shown that these strategies are very sensitive to errors in input data and that large errors exist in these data. This has led to a comprehensive analysis that includes

1. Identification of the sources of data errors in the surveillance and prediction elements of the traffic-control system,
2. Quantification of the characteristics of individual errors,
3. Evaluation of the sensitivity of the control strategies to data errors,
4. Development of surveillance and prediction techniques for minimizing the effects of these errors, and
5. Evaluation of the magnitude of the fluctuations in traffic volumes to which the control strategies are to respond.

Hopefully this analysis will be successful. It is obvious that an error exceeding the variations in the quantity being controlled will reduce the control system to total ineffectiveness. In fact, based on observations to date, this might be the cause of the lack of success that past researchers have had in developing control strategies (1, 2). Many of these strategies have been developed without prediction techniques or effectiveness evaluation. This paper focuses on 1 aspect of this question—the errors in input data and their effect on control-system design. It also presents a brief summary of the control strategies being developed for the UTCS project. These control strategies are discussed in greater detail in other reports (3, 4, 5).

FIRST GENERATION CONTROL STRATEGY

The UTCS control-strategy development consists of the implementation of 3 generations of control (Table 1). The first generation of control is based on the use of signal timing patterns generated off-line and stored in a peripheral storage device. The system is capable of using 3 possible modes of pattern selection.

1. The operator select mode is one in which the system operator determines the operation pattern and makes a selection through the control panel. This selection can be made at any time during system operation.
2. The time-of-day mode is one in which the computer selects timing patterns every 15 min according to a predetermined schedule.
3. The traffic-responsive mode is one in which the computer attempts to select the pattern that is best suited for current traffic conditions every 15 min.

The first generation software also is capable of making adjustments in the timing patterns at selected intersections in response to fluctuations in traffic demand at each signal cycle [critical intersection control (CIC)]. The adjustment is accomplished by measuring vehicle volumes and modifying the signal split in such a way that the percentage of green time given to the competing demands is approximately proportional to the approach volumes.

SECOND GENERATION CONTROL STRATEGY

The principal difference between the first and second generation control strategies is that the second generation strategy computes the traffic signal timing on-line at a fixed rate of 4 to 7 min. (The exact rate has not yet been determined.) The optimization technique used for this computation is based on the SIGOP optimization, which computes and implements signal timing directly and does not require operator intervention. Obviously, under these circumstances, the traffic engineer loses the capability to make adjustments to the computed pattern that he or she would typically have when operating with the first generation system.

Table 1. Urban Traffic Control System strategies.

Strategy	Update Interval (min)	Prediction	Pattern Generation (selection)	Critical Intersection Control
First generation	15	None	Off-line pattern timing Time of day Traffic responsive Operator select	Comparison of A-phase and B-phase demand to determine A-phase yield point
Second generation	4 to 7 (precise value not yet determined)	Historically based	Employs modified version of SIGOP offset optimization	Both split and offset computation based on previous phase demand
Third generation	3 to 5 (variable)	Statistical predictor (form not completely determined)	Cycle-free optimizations for moderate flow and congested flow	Not applicable

Table 2. Control-strategy data requirements.

Algorithm	Measurement Interval	Critical Lane Variable	Range of Error[a]
First Generation			
Traffic-responsive pattern selection[b]	15 min (total)	Volume	Must be consistent indicator of traffic conditions
		Occupancy	Must be consistent indicator of traffic conditions
Critical intersection control[c]	Each phase	Volume	1 to 3 vehicles per cycle
		Queue	1 to 2 vehicles
		Speed	5 to 10 percent
Second Generation			
Network optimization	4 to 7 min	Primary volume	1 to 3 vehicles per cycle
		Queue	1 to 3 vehicles
		Speed	5 to 10 percent
Critical intersection control[d]	Start of each signal phase	Primary volume	1 to 3 vehicles per cycle
		Queue	1 to 3 vehicles
		Speed	5 to 10 percent
Third Generation			
Undersaturated control	3 to 5 min	Primary volume	1 to 3 vehicles per cycle
		Secondary volume	1 to 3 vehicles per cycle
		Speed	5 to 10 percent
Saturated intersection control[e]	Continuously monitored	Link content	1 to 3 vehicles

[a]Computed on the basis of keeping timing errors below 2 to 5 sec.
[b]Measured at locations that provide data representative of need for timing pattern selection.
[c]Volume and queue updated continuously on minor phase.
[d]For saturated intersections, only total approach volume is required.
[e]Data needed at all saturated intersection control intersections.

The second generation strategy also has CIC. In this case, the CIC adjusts both split and offset for every signal cycle. As with first generation, split is adjusted in response to relative approach volumes. Offset is adjusted to accommodate queues that have built up because of secondary flow and variations in vehicle speeds.

THIRD GENERATION CONTROL STRATEGY

The most complex of the control strategies being developed for UTCS is the third generation of control. This strategy consists of 2 levels of control selected on the basis of traffic demand.

1. Medium flow control computes timing patterns at intervals of approximately 5 min. This control mode permits cycle length to vary at adjacent intersections. Cycle length also can vary at a given intersection from one cycle to the next. Under these conditions, split and offset also will vary constantly in both time and space. Thus, it is no longer convenient to treat signal timing in terms of the variables cycle, offset, and split. In both this mode and the congested mode, optimization computes signal timing as green-on and green-off times for each approach.
2. Congested flow control operates when traffic at either a single intersection or a group of intersections builds up to the point at which the intersection can no longer accommodate all of the vehicles arriving during a signal cycle. In this mode of operation, congested intersections are identified and cycle lengths are increased to maximize their throughput. In addition, traffic from upstream intersections is gated into the congested intersection in a manner that will prevent spillback across the upstream intersection. The gating is designed to prevent buildup of congestion around a closed loop of streets; in effect, traffic backs up around the block. Signal timing is computed continuously in this mode of operation to determine green switching times. Because of its cycle-free characteristics, third generation strategy does not require a critical intersection control capability.

CONTROL-STRATEGY DATA REQUIREMENTS

It is evident that each of these control strategies will have differing data requirements (Table 2). In the table, queue length can be replaced by secondary flow because either of these variables can be used to determine the number of vehicles that must be discharged before the main group for the upcoming cycle.

Both first and second generation control strategies have areawide control as well as single-intersection control, which is intended to fine-tune the areawide control settings. This implies 2 distinctly different levels of data requirements existing within the same control strategy. As a result, the costly deployment of large numbers of detectors can be limited to those intersections requiring critical intersection control. Error range is provided in the table as an indication of the level of accuracy that can be anticipated from the surveillance system rather than from a reflection of the actual requirements of the control strategies. Other data requirements not included in this table fall into the following 2 categories:

1. Threshold values used to determine the mode of operation of the control strategies and
2. Parameters used to model the traffic system in the optimization process of the control strategies.

In the first case, threshold values are most often used to identify the existence of saturation. For example, the first generation of control defines the existence of saturation as the buildup of the standing queue past the furthest upstream detector at any time during the red signal state for that link. This threshold is necessary because the CIC algorithm requires the measurement of B-phase demand during A-phase green.

Figure 1. Start-up delay frequency distribution.

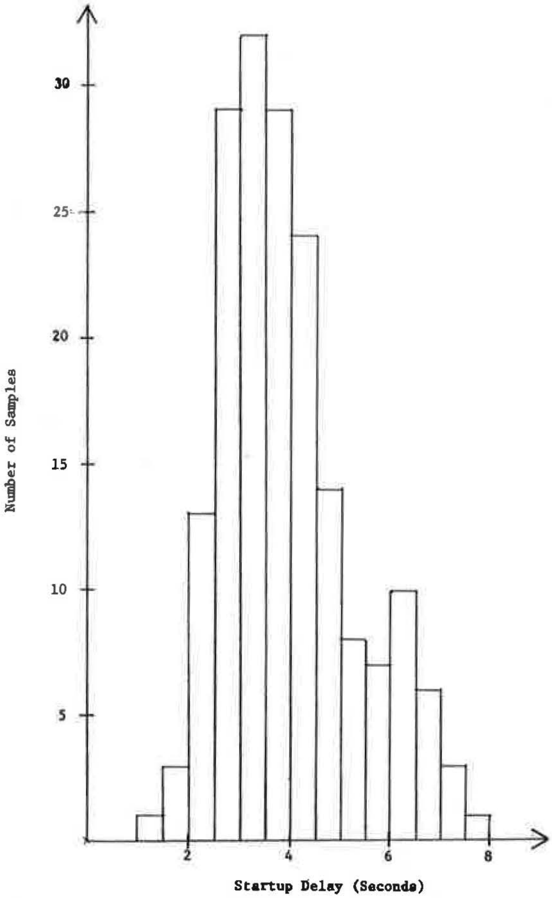
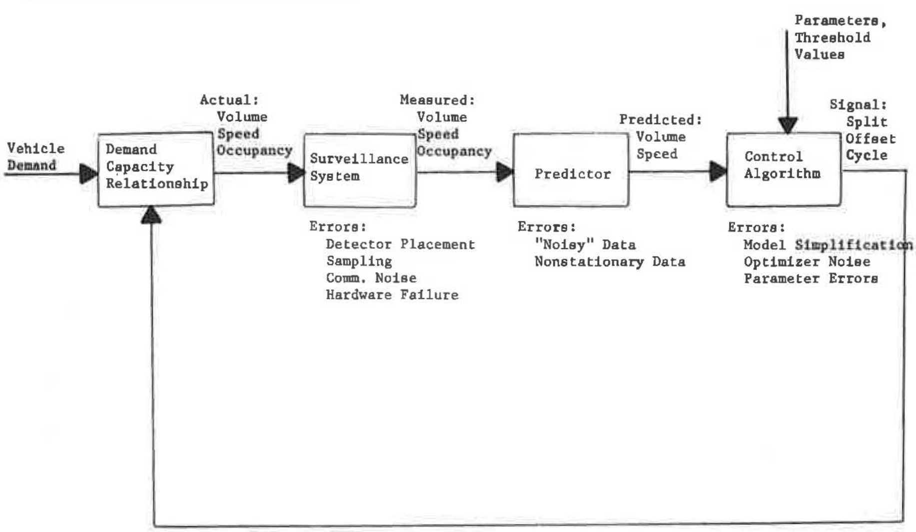


Figure 2. Traffic-control system and sources of error.



This approach is used to determine A-phase duration in approximately the same manner as the computation of a yield point in a semiactuated controller. Street surveys have determined that queue buildup past the upstream detector on a 2-detector link [with a detector at the stop line and a second detector 210 ft (64 m) upstream in the same lane] can be identified reliably by a value of occupancy of 35 percent. In this case, occupancy is computed as the percentage of time that vehicle presence is measured by the upstream detector. These data are smoothed by using first-order smoothing as follows:

$$\overline{O}_i = \overline{O}_{i-1} + k\,(O_i - \overline{O}_{i-1})$$

where

k = smoothing constant (a value of 0.5 currently is being used),
O_i = value of occupancy measured during signal cycle i, and
$\overline{O}_i$ = smoothed value of occupancy at signal cycle i.

Second generation software uses a similar technique to control the operation of its critical intersection control algorithm; third generation requires this type of threshold to change from moderate-flow to congested-flow modes of operation.

In the second case, data requirements for control strategies are often overlooked in the design of the strategies. These requirements are the parameters used in the optimization process. The following are examples of these parameters:

1. Start-up delay,
2. Discharge headways,
3. Number of lanes,
4. Link lengths (intersection spacing), and
5. Group dispersion.

Because a surveillance system rarely is designed to measure these parameters in real time, the developer of the control strategy must treat these input parameters as systemwide, link-specific, or time-of-day constants. A link-specific constant is rarely selected because the cost of detailed link-by-link data collection for all times of day is extremely high. Yet to treat these parameters as systemwide constants can result in serious errors in the optimization process. For example, Figure 1 is a histogram of the start-up delay measured at 14 locations in Washington, D.C. This figure indicates that start-up delays of between 2 and 7 sec are common. The variance in this parameter can result in offset errors that will cause increases in stops within the network because inadequate queue discharge times will be used to account for the larger start-up delays. It will not have as great an effect on delay unless use of incorrect start-up delay causes inadequate green time to be assigned to a phase resulting in saturation. This study was undertaken as an attempt to determine the cause of numerous incorrect values of offset arising from the TRANSYT optimization of the first generation signal timing. Obviously, this type of problem, which arises in an off-line optimization, is equally likely to occur in an on-line control strategy.

SOURCES OF ERROR

Some of the potential sources of error in an on-line traffic control system have been discussed. They occur throughout the control process and can be controlled only by more complex surveillance, off-line data gathering, and sophisticated processing techniques. All of these measures will result in increased system cost, which must be balanced against the potential benefits of a responsive control system. None of these measures will completely eliminate error in the control process.

Figure 2 shows a summary of the various sources of error in a traffic-control system. Each of these errors results in a control computation that is suboptimal for the "actual" conditions on the street and will result in a degradation of system effectiveness. Such a degradation can easily result in a responsive system whose operation is less effective than that of a first generation system that could be implemented at a much lower cost.

It is convenient at this point to select vehicle volume as the variable that will be emphasized in the remaining discussion because in most cases it will have the most significant effect on degrading the quality of the control. Furthermore, it is clearly beyond the scope of this paper to analyze the effects of each of the many other variables in the traffic-control-system operation.

It can be seen from Figure 2 that the sources of error related to vehicle volume are detector placement in the surveillance system and the characteristics of the data processed by the surveillance system (noisy data and nonstationary data). It is difficult to present the generalized statistics of errors that can be expected from the limitations of detector placement because these errors are closely related to the characteristics of the street on which the detectors are installed. Assuming that the correct critical lane (flow lane with maximum volume) has been instrumented, volume errors will result from lane changing, channelization, midblock sources and sinks, and queue buildup. Many of these factors will cause errors that are not zero mean errors (because these errors are correlated serially with the volume). This is significant because a zero mean process is often assumed when the effect of volume errors on operation is analyzed.

An indication of the magnitude of surveillance system errors can be seen from a limited study of 4 locations in the UTCS network; the results showed hourly volume errors with mean error values of 38.5 vehicles per hour and standard deviations of 65 vehicles per hour (9). This type of error should not necessarily be considered typical because it was measured at some of the worst locations in the UTCS network and instrumentation changes are currently under way to reduce their effects. They are presented as an indication of the potential magnitude of the problem and to point out that surveillance errors can have large mean values. Surveillance system errors can often be controlled by increasing the number of detectors in a network. Prediction errors are a result of the characteristics of the data being processed. These data can be described in the following terms:

1. Volume data contain both a time-varying mean and variance (nonstationary); and
2. Spatial and temporal correlations of volume data are low and might also be time varying.

Typical spatial and temporal correlations are shown in Figure 3 (5, 10). All data in this figure refer to the L Street approach to the intersection of L and 15th Streets. This is the reason for the correlation of 1.00 in this link. The correlations of 4 cycles indicate the value of upstream data for predictor operation. Obviously, the poor correlation shown in Figure 3 implies that there are inadequate data on which to base the prediction. For this reason, the most successful predictors developed to date have relied heavily on historic data, that is, data derived from previous days with similar characteristics. Complete reliance on historical data would eliminate the need for a traffic-responsive system because the use of the same data from 1 day to the next would result in the same signal-timing patterns each day. This would be, in effect, a fixed-time operation.

There have been approximately 9 different predictors developed for the UTCS project (4, 5). Each of these predictors has been developed on a different basis, yet most have resulted in error distributions of the type given as follows for 100 links, 46 predictions per link (4):

Figure 3. Spatial and temporal correlations of link volumes.

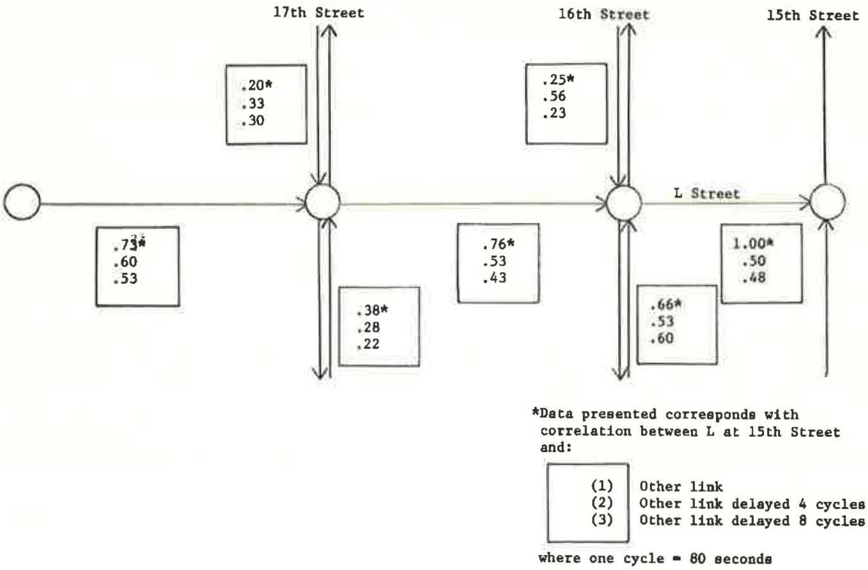
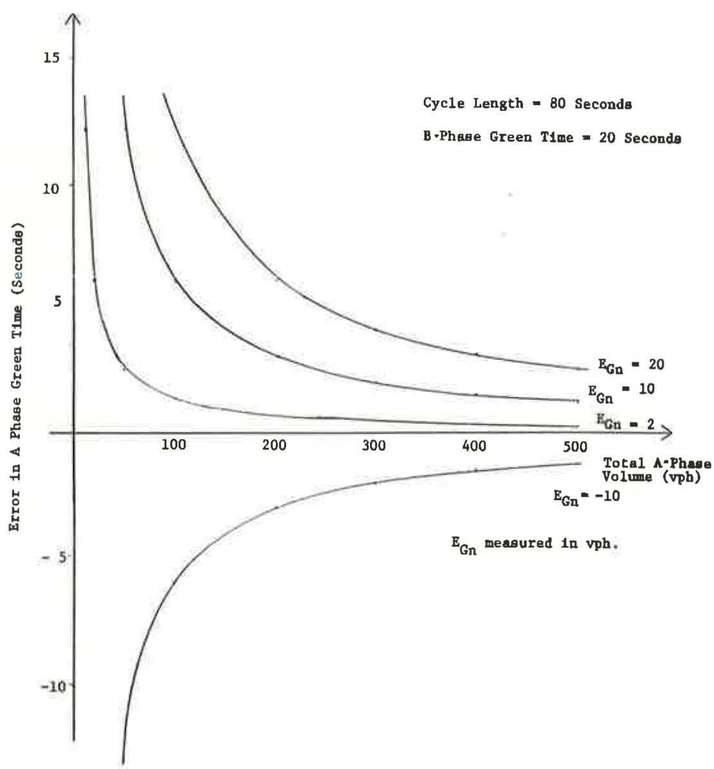


Figure 4. Errors in green time resulting from volume errors.



Volume Predictor	Error
$\overline{AM1}$	0.123
P05	0.719
P10	0.466
P20	0.175

where

$\overline{AM1}$ = mean AM1, which is $\left|\frac{f(t) - p(t)}{f(t)}\right|$,
P05 = Pr(AM1 ≥ 0.05),
P10 = Pr(AM1 ≥ 0.10),
P20 = Pr(AM1 ≥ 0.20),
f(t) = actual volume, and
p(t) = predicted volume.

Standard deviations of predictor error range between 7 and 18 vehicles per hour depending on the variability of the volume data on which the predictions are based. This value of error for the predictor operation, which is not controllable, provides the system designer of a traffic-responsive system with an indication of the acceptable level of surveillance errors. Thus it can be concluded that a good surveillance system design is one that results in volume measurement errors with a standard deviation that is less than 7 vehicles per hour when it is used in the modes given in Table 2. This information would be applied to detector location at an intersection approach in the following manner:

1. Identify through lane carrying the largest volume;
2. Select detector location as far back from intersection as possible but downstream from any major sources or sinks such as parking garages;
3. Measure lane volume at intersection and compare it with volume passing over selected detector location; and
4. Compute standard deviation of difference between measurement of volume entering intersection and volume at selected detector location at each signal cycle.

Perform test for 30 signal cycles during both peak periods and midday. Standard deviation should be less than 7 vehicles per hour (vph). Use of 30 samples is recommended based on past experience with similar measurements.

SENSITIVITY OF SPLIT COMPUTATION TO VOLUME ERRORS

An example of the effect of volume errors on control-strategy operation is the relationship between these errors and split computation. If the split for each intersection is computed by using green demand, which is assumed to be equivalent to total approach volume, the time for the nth phase (t_n) can be written

$$t_n = \frac{G_{Dn} \cdot C}{G_{Dt}}$$

where

C = cycle length,
G_{Dn} = green demand on phase n, and
G_{Dt} = total green demand on all phases.

For the purpose of this discussion, green demand is flow-lane volume in vehicles per cycle arriving at the approach to the intersection serviced during phase n. An error in volume measurement is equivalent to an error in green demand on phase n (E_{G_n}) that causes error in t_n (E_{t_n}); the following expression results:

$$t_n + E_{t_n} = \frac{(G_{D_n} + E_{G_n}) \cdot C}{G_{Dt} + E_{G_n}}$$

Assuming that E_{G_n} is much less than G_{Dt}, we can solve this expression for E_{t_n} by substituting the expression for t_n that yields the result

$$E_{t_n} = \frac{E_{G_n} \cdot (C - t_n)}{G_{Dt}}$$

If this equation is applied to the major phase of a 2-phase intersection, $C - t_n$ is equivalent to the minor-phase time at the intersection. The results are plotted in Figure 4.

One approach to relating the effect of split errors to network performance is to compute the increase in intersection delay resulting from the incorrect computation of green time. If the split error does not cause the intersection to become oversaturated, and the offset is not affected, only the random vehicle delay at the intersection will be changed. Random delay is defined as the correction added to computation of vehicle delay to allow for cycle-by-cycle variations from average behavior. The random delay correction is modeled in the TRANSYT signal timing program as follows (6):

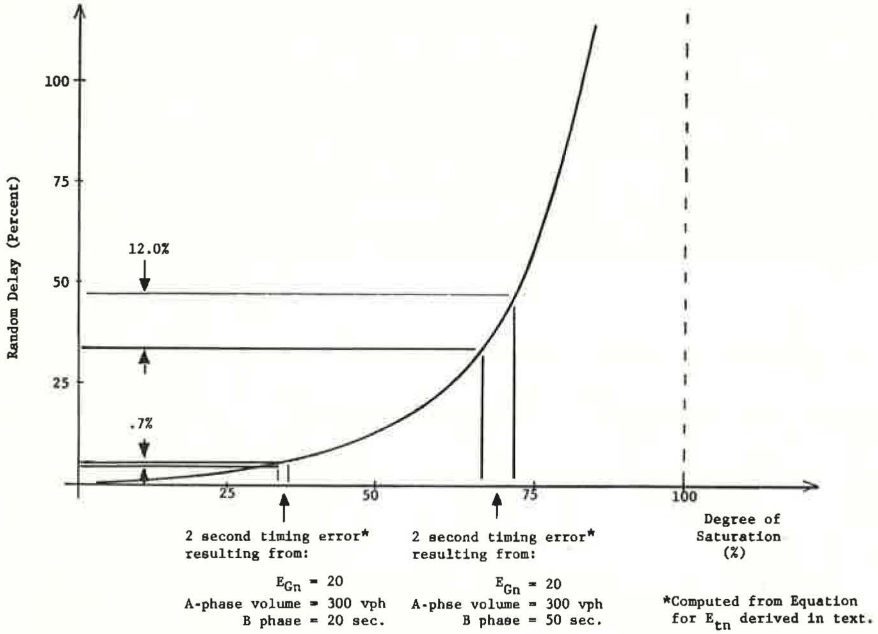
$$D_R = \frac{1}{4}\frac{x^2}{1 - x}$$

where x = the degree of saturation or, in other words, the fraction of green time during which vehicles are discharged through an intersection. From this relationship, which has been considered by other investigators (5, 7), it can be seen that a split computation error resulting in a reduction of available green time and an increase in degree of saturation will have a greatly increased effect on the delay experienced by motorists at that intersection. These effects are shown graphically in Figure 5, which indicates that the sensitivity of delay to errors in green time depends on the degree of saturation existing at the intersection. This is not a surprising result because it is equivalent to the statement that incorrect signal timing at an intersection will have a more serious effect on the intersection's operation under heavy traffic. What is surprising about this result is that a split error of only 2 sec for a degree of saturation of 75 percent can produce an increase in delay of 12 percent. This is equivalent to the level of improvement anticipated from a traffic-responsive system (Fig. 5). The 2-sec error was produced from an error in estimated green demand of 20 vehicles per hour, which is a value that is probably less than the standard deviation of that total error resulting from combined surveillance and prediction errors.

CONCLUSIONS

This paper has attempted to present some recent results of research resulting from the UTCS project. The research has demonstrated the existence of rather large data errors within a traffic-control system that have the potential for significantly degrading the operation of that system. These errors can be minimized only through careful surveillance of system design and creation of a large and detailed data base to serve the control-strategy operation.

Figure 5. Random delay at undersaturated intersections.



If the system designer is not willing to undertake these measures in the implementation of a real-time-responsive control system, the resulting system operation could be less effective than that of a pretimed system.

REFERENCES

1. J. Holroyd and J. A. Hillier. The Glasgow Experiment: Plident and After. Transport and Road Research Laboratory, England, RRL Rept. LR384, 1971.
2. Improved Control Logic for Use With Computer-Controlled Traffic. Final Rept., NCHRP Project 3-18(1), Nov. 1973.
3. Urban Traffic Control and Bus Priority System—Operator's Manual, Volume 2. FHWA, NTIS No. PB 214 641, Sept. 1972.
4. Urban Traffic Control System and Bus Priority System Traffic Adaptive Network Signal Timing Program—Software Description. FHWA, Aug. 1973.
5. Variable Cycle Signal Timing Program—Volume 3, CYRANO: Cycle-Free Responsive Algorithms for Network Optimization: Moderate, Congested, and Light Flow Regimes. FHWA.
6. D. I. Robertson. TRANSYT Method for Area Traffic Control. Traffic Engineering and Control, Oct. 1969, pp. 275-280.
7. R. Wormleighton. Queues at a Fixed Time Traffic Signal With Periodic Random Input. Canadian Operations Research Society Journal, Vol. 3, No. 3, Nov. 1965, pp. 129-141.
8. K. Keitt and H. Lum. Study of the Characteristics of Startup Delay at Urban Intersections. FHWA, Internal Rept., Sept. 1974.
9. G. Smith, H. Lum, and J. MacGowan. Traffic Volume on UTCS Links. FHWA, Internal Memorandum, Feb. 1974.
10. N. R. Draper and H. Smith. Applied Regression Analysis. John Wiley and Sons, Inc., New York, 1966, p. 33.

DISCUSSION

Dale W. Ross, DARO Associates, Inc.

Tarnoff has made a number of important points on how various errors can degrade the performance of a computerized traffic surveillance and control system. It is indeed time that some serious research was devoted to this subject. The paper does not, however, substantiate the implied hypothesis that data errors can cause a real-time, traffic-responsive control system to be less effective than a pretimed system. The paper tends to exaggerate these effects beyond what may be the actual situation. Some other points should be considered to see that the effects of data errors may have been exaggerated.

RANDOM DELAY FORMULA IS UPPER BOUNDING ONLY

The analysis and example in the paper assume that the formula

$$D_R = \frac{1}{4}\frac{x^2}{1 - x}$$

is an accurate or representative model of random delay at a signalized intersection. Careful reading of the field experimental results from which this formula was developed shows that this relation is not actually a model for random delay itself (11), but rather that the formula represents an upper bound or envelope for the field data taken on random delay. Robertson points out that there is considerable scatter in field observations of random delay (11). This is particularly the case at higher degrees of saturation. The effect on the paper is that the use of this formula must be considered a worst-case analysis, and that the actual sensitivity of random delay to data errors can be an order of magnitude less than that derived by using the above formula. The use of this formula in TRANSYT was based not so much on its being an accurate model of random delay as it was on its being a means of forcing the TRANSYT model to select phase durations and cycle lengths that led to saturation levels of less than 90 percent (6, 10). The deliberate exaggeration of the random delay by this formula at large degrees of saturation thus served as a built-in means of ensuring that the TRANSYT model would not select unreasonable phase durations. The paper misconstrues the use and meaning of the formula.

RANDOM DELAY IS LESS THAN TOTAL DELAY

The paper considers only the random delay component of total intersection approach delay. The other primary component is the deterministic delay due to offset and phase durations. Robertson shows that even when degree of saturation is as large as 90 percent and the offset is the best possible, it is typical to find that the random delay is no more than half of the total delay (11, Fig. 9). Consequently, the sensitivity of total delay due to timing errors is less than the paper indicates.

ASSUMED VOLUME ERRORS ARE LARGE

The text table on distribution of prediction errors indicates that the UTCS has had a volume prediction mean error of about 12 percent. This seems to be about twice as large as results that are being obtained in other current experimental work (12). In fact, according to J. Lam and D. Kaufman of the Corporation of Metropolitan Toronto, experimental results with a predictor similar to that used in the ASCOT system, which has been described elsewhere (2), have shown prediction to be on the order of 4 to 5 percent. Thus the inferences made in the paper may be overstated because of the assumption of fairly large prediction errors.

EXAMPLE USES CYCLE LENGTH THAT EXAGGERATES ERRORS

If one carefully examines the example given in the paper pertaining to Figure 5, one finds that the example assumptions (an A-phase degree of saturation of 0.75, an A-phase volume measurement error of 20 vehicles per hour, a B-phase duration of 50 sec, an A-phase timing error of 2 sec, and an A-phase volume of 300 vehicles per hour) are consistent among themselves only if the sum of the A-phase and B-phase volumes is 500 vehicles per hour, and if the cycle length is 125 sec. The assumption of a 125-sec cycle length for the 2-phase signal is somewhat unrealistic. If one had used Webster's method (13) to select a near-optimum cycle length for the intersection, the cycle length would have been chosen to satisfy the relation

$$C_o = \frac{1.5\ L + 5}{1 - Y}$$

where

C_o = optimum cycle length,
L = total lost time for the intersection, and
Y = sum of the volume-to-saturation flow ratios for the phases.

With the parameter values used in the paper, one can verify that this formula would have yielded an optimum cycle length of 125 sec only if L had been approximately 17.5 sec. This is an inordinate amount of lost time for a 2-phase signal. If a more reasonable lost time of 6 to 8 sec per cycle were used, one would find that approximately 60 sec would be the optimum cycle length. Thus a more reasonable cycle length for the parameter values given in the paper would have been 60 sec instead of 125 sec. If the 60-sec cycle had then been used in the analysis of the paper, it would have been seen that the effect of the 20-vph volume measurement would then have been only a 0.96-sec timing error instead of the 2.0-sec error in the paper. Thus the cycle length assumed in the paper perhaps overstates the magnitude of the error by a factor of about 2.

These 4 points indicate that the paper probably exaggerates the effects of data errors. The paper makes a good point that these errors need further study, but one should not make hasty conclusions regarding the effect of such errors. In particular, the analysis in the paper should not be misconstrued as meaning that real-time, traffic-responsive control systems are likely to be less effective than pretimed systems. One needs to be careful to draw such conclusions only from well-founded research results.

RECOMMENDATIONS

The introduction of the paper identifies 5 factors that UTCS researchers are investigating regarding the sensitivity of real-time control to data errors. It is suggested that the list of factors be expanded to 7; the 2 additional factors in the analysis would be

1. Investigation of the surveillance and control algorithms that provide the best compromise between good signal timings and insensitivity to data and parameter errors, which would require cross-testing of the UTCS work, British work (1), Canadian work (12), and ASCOT work (2), and
2. Investigation of programming and programming-induced errors such as

round-off and truncation errors in computations.

The last factor is one that should not be overlooked. Real-time software systems for traffic control are fairly intricate, and even the most brilliant programmers and engineers can make several subtle errors in the programming that do not evidence themselves in an obvious, consistent manner.

The issue of software or programming errors was, in fact, a significant factor in the ASCOT field results (2), and because I was principal investigator for ASCOT development and have continued to apply ASCOT techniques, I rebut Tarnoff's comment that the ASCOT development and tests met with a "lack of success." First, "success" can be measured in different ways. In some respects the ASCOT development was quite successful. It demonstrated that it is possible to achieve highly flexible methods of traffic control by using limited computational resources and that such control was well within the capabilities of most minicomputers. Also, the city of San Jose, California, continues to use the ASCOT system on a day-to-day basis, and has even expanded its use of ASCOT from a 12-hour control day to an 18-hour control day. Furthermore, techniques and methods used in ASCOT are finding application in other cities, including Chicago (14) and Toronto (12).

Tarnoff is correct in stating that in the San Jose tests of ASCOT, results were inconclusive regarding the effectiveness of ASCOT versus the effectiveness of pretimed operation. Some of the lack of improvement has been attributed to deficiencies in the offset optimization logic of ASCOT, and these are reported elsewhere (2). It is now known that several major software errors have been discovered in ASCOT, and these are major reasons for the lack of improvement.

The field tests of ASCOT were conducted in the summer of 1973, and at that time every possible effort was made to ensure that the software had been carefully screened and tested for programming errors. In the spring of 1974, a study was begun to develop documentation of ASCOT for San Jose's operating and engineering personnel, to develop additional programs for the evaluation of the system by using surveillance data, and to conduct a review of the software system to identify possible improvements and errors. The work revealed several software errors that had not been known at the time of the field tests and later (2). Here are some of the major errors that were found.

1. The ASCOT logic for computing offsets depends on the TRANSYT traffic-flow model for modeling the platooning of traffic and choosing offsets tailored to the platooning. It was found that in programming this model, link IN-patterns were incorrectly computed from the sum of upstream OUT-patterns. If one refers to the equations given by Robertson (11, p. 18), the correct equation is

$$q'(i + t) = F \cdot q(i) + (1 - F) \cdot q'(i + t - 1)$$

Instead of that equation, ASCOT had been programmed with the equation

$$q'(i + t) = F \cdot q(i) + (1 - F) \cdot q(i + t - 1)$$

(Primed variables are IN-pattern variables; unprimed variables are upstream OUT-pattern values.) The consequence of this error was that platooning was not correctly represented, and offsets could not be selected properly.

2. TRANSYT GO-patterns were organized in disk memory in groups of 10 links. One indexing error prevented any GO-patterns that had been computed for the last group of 10 links in any intersection group (subset) from ever being written to disk. Consequently, GO-pattern data for such links were missing, which led to erroneous TRANSYT platoon modeling.

3. The ASCOT CIC method depended on a subroutine to add the variable controller intervals on the current phase of each CIC-controlled intersection to determine the optimum time to switch from 1 phase to the next. This summing subroutine was programmed incorrectly, which led to incorrect estimates on the best time to switch phases.

Further investigation of ASCOT beyond the initial field tests has revealed that programming errors existed that had major consequence. The programming errors discovered were subtle errors and others may still exist. The point of all this is that data errors are only a part of the picture and that software errors also should be recognized as important. It often takes years to completely "iron out" a new software system, and further research should be devoted to improving means of reducing such errors.

REFERENCES

11. D. I. Robertson. TRANSYT: A Traffic Network Study Tool. Transport and Road Research Laboratory, Harmondsworth, England, RRL Rept. LR253, 1969.
12. Improved Operation of Urban Transportation Systems. Ministry of Transport, Canada, Ministry of Transportation and Communications, Ontario, and Corporation of Metropolitan Toronto, study in progress.
13. F. V. Webster. Traffic Signal Settings. Transport and Road Research Laboratory, Harmondsworth, England, Technical Paper 39, 1958.
14. Specifications and Contract Document 80.65-73-1. Dept. of Purchases, Contracts, and Supplies, city of Chicago, Sept. 1973.

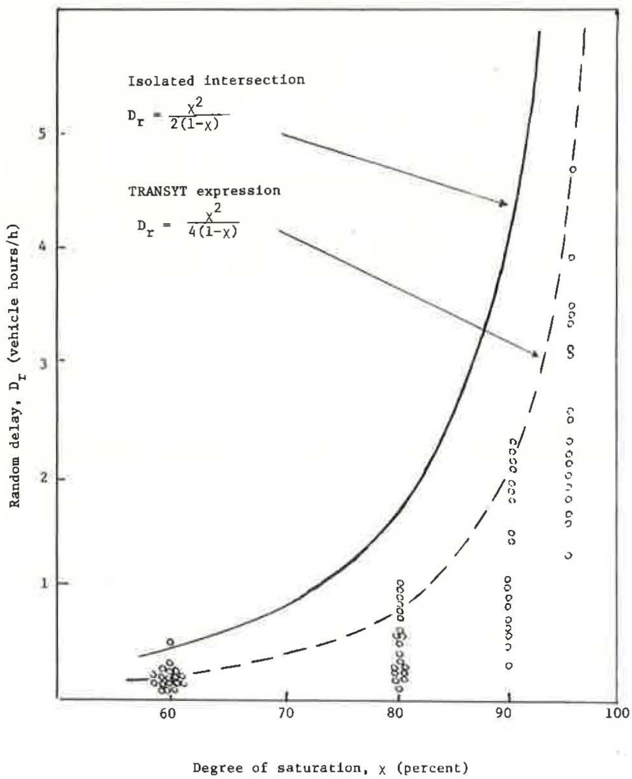
AUTHOR'S CLOSURE

Ross interpreted this paper as being a negative viewpoint on traffic-responsive strategies in general and the ASCOT program in particular. The conservative outlook expressed in this paper relative to potential improvements in traffic flow that might be possible with traffic-responsive strategies was not intended to reflect adversely on his work, which is recognized to have been performed under budget and time constraints. Nevertheless, the fact remains that neither ASCOT nor the British traffic-responsive control strategies have materially improved traffic flow in the cities where they were tested. Preliminary experience with the UTCS second generation control strategy, which has undergone both simulated and real-life testing, has produced similar results. Thus, this paper was written as an attempt to present an objective explanation for these results.

Ross stated that the random delay formula used in the paper exaggerates the effects of random delay. The basis on which this statement is made is a sentence by Robertson (11) taken out of context: "This curve... is seen to exaggerate the mean random delay at the higher saturation levels." As can be seen from the curve presented in this reference and reproduced here (Fig. 6), the higher saturation levels referred to are above 80 percent. Yet the example that Ross claims exaggerates the result uses a saturation level of 75 percent, a value that was selected specifically to avoid the possibility of exaggerated results.

Ross indicated that other delays are more significant than random delay and as evidence again referenced Robertson (11, Fig. 9). The accuracy of this depends on signal offset, degree of intersection saturation, and the ratio of primary to secondary vehicle flows. The use of an arbitrary example to support such a statement is hardly conclusive. Furthermore, for the Ross statement to be correct, the effects of the split errors discussed in the paper would have to be more pronounced than those that

Figure 6. Variation of random delay with saturation.



were presented in the example because the effects of nonrandom delay are additive with the random delay. Thus, the results presented here can be considered understated rather than exaggerated results as Ross implies.

Ross referred to research performed in Toronto as evidence that the UTCS prediction volume errors are quite large. The UTCS predictor was tested by using data from Toronto. The results of these tests were errors in the same 4 to 5 percent range experienced by Lam of Corporation of Metropolitan Toronto. However, volume errors with a historically based predictor depend on the daily variation in vehicle volume. The Toronto system has the relatively smooth repeatable traffic-flow characteristic of suburban arterials. In the UTCS network the opposite is true, and a degradation in predictor performance results. Therefore, the errors presented in the paper must be considered typical of those that would be experienced in the central business district of a major U.S. city.

Although Ross is correct in stating that the cycle length chosen for the 75 percent-saturated case is long compared to cycle lengths generally used in coordinated signal systems, his assumption that the cycle length at every intersection must satisfy Webster's equation for "optimum" cycle length is not correct in all cases. In a coordinated signal system, cycle length is selected to satisfy the intersection requirements of longest cycle length and minimum green times as dictated by pedestrian crossing times. Furthermore, Webster's equation produces optimum results only at an isolated intersection with random arrivals and is not applicable for a network with platooned arrivals.

Perhaps the most important point is the fact that the effect of the 2-sec timing error was the purpose of the example and was not the particular set of circumstances that produced it. For example, a similar 2-sec error could have been produced by a cycle length of 80 sec with A-phase time of 50 sec, A- and B-phase volumes of 500 and 300 vehicles per hour respectively, and an A-phase volume error of 53 vehicles per hour (less than 1.2 vehicles per cycle).

In conclusion, it must be stated that none of the specific points raised by Ross in any way detracts from the content or conclusions of the paper. Although there is still a place for traffic-responsive strategies in cities that plan to install additional hardware as a substitute for manual updating of traffic signal timing, the potential of these strategies for improvements of traffic flow is far from assured. On the basis of available information, it is the responsibility of every research organization to avoid raising the false hopes that traffic-responsive strategies in their current form can provide a major improvement in urban traffic flow.