Cost-Effective Level of Service and Design Criteria

Al Werner and Terry Willis, Strategic Planning and Research Branch, Alberta Transportation, Edmonton

Research was initiated to determine the appropriate levels of service that could be applied to the design of rural highways in Alberta under a variety of conditions. By using Highway Capacity Manual speed-volume curves for two-lane rural roads and passenger-vehicle road user costs and average construction and maintenance costs for Alberta, cost-volume relations (unit average cost supply functions) are derived and applied in the analysis of annual hourly traffic data obtained at permanent counter locations in Alberta. To identify the demand function, determination of design hourly volumes from curves for the highest hour of the year is discussed (all 8760 hours are ranked). The supply and demand functions are then interrelated. Preliminary findings indicate that (a) a minimum unit cost per passenger vehicle per kilometer can be correlated with volume-capacity ratios depending on capital, maintenance, and user costs and (b) a design hourly volume K-factor based on the knee of the curve for the highest hour of the year is more consistent than the traditional 30th highest hourly volume K-factor. Although a methodology has been developed to determine a cost-effective volume-to-capacity ratio and the findings indicate that it may be appropriate to set a range of levels of service for different road characteristics, further work is required to refine cost relations so as to reflect terrain, traffic composition, and current road user costs.

In considering the planning and design of a new highway or improvements to an existing one, both supply and demand must be considered. On the supply side, three major factors are normally considered in designing a new facility or in evaluating an existing one: (a) safety, (b) structural adequacy, and (c) level of service (traffic congestion). On the demand side, the major factor considered is existing and future demand.

In practice, agencies try to design a facility to provide an acceptable level of service to the user. In the past, design criteria were adopted that provided a very high level of service and thus low levels of congestion. However, in recent years, because of the increasing costs associated with continuing to provide a relatively high level of service, agencies have had to reconsider their policies and associated design criteria.

Although supply and demand are often evaluated separately, they are interrelated. This paper offers a methodology for evaluating these relations based on concepts previously developed by Haritos (1), Cameron (2), and Winfrey and Zellner (3). We hope it makes a modest contribution toward clarifying and understanding some of the unresolved issues in transportation economics.

This paper first presents a methodology for deriving the most economical level of service for the supply side. Then, on the demand side, a design hourly volume and its selection are discussed. The paper concludes with the presentation of suggested design criteria.

MEASUREMENT OF ECONOMIC EFFICIENCY

Generally, highways are provided at public expense and little or no direct income is derived. Therefore, the agencies involved should design roadways not to maximize proft but to minimize costs to the public while providing good standards of safety and mobility.

In any attempt to define the costs attributable to providing a highway link, one might include costs for right-of-way, construction, maintenance, environmental disruption, motor-vehicle running costs, accidents, and travel time. In the analysis presented here for the Canadian province of Alberta, the following cost factors were used: (a) construction, (b) maintenance, (c) motor-vehicle running cost, and (d) travel time.

Although right-of-way costs can be a major factor, it has been assumed that right-of-way has full terminal value and therefore it is not considered here. Quantifiable costs related to environmental disruption—e.g., costs of erosion control, noise attenuation, and other measures to protect the environment—can be included in construction costs. However, unquantifiable costs, such as those for wildlife disruption, are not included. Accident costs have not been included here because no Alberta data were readily available and because accident costs can be considered part of the safety analysis that some agencies prefer to handle separately.

Fixed capital costs for roads are high, and annual maintenance costs are often also significant. If the road carries little traffic, the unit cost of providing the roadway is very high; as volume increases, however, unit cost decreases.

For road user costs (time plus running costs), lower traffic volumes usually provide the least unit cost and, as volumes increase, the cost to the user increases because of congestion. These relations are shown in Figure 1. Merging these two curves should result in a relationship in which, at some volume of traffic, a minimum cost of travel will occur.

To compute this relationship, it is necessary to relate capital and maintenance costs and road user costs to a common base. Since capital and maintenance costs are a function of volume and road user costs are a function of speed, the speed-volume relations presented in the 1965 Highway Capacity Manual (HCM) (4) were used to determine costs in dollars per vehicle kilometer. The following values were used:

1. Capital cost: \$181 250/km,
2. Maintenance cost: \$1000/km,
3. Discount rate: 8 percent over 15 years, and
4. Road user costs: 1976 Alberta running costs (5) for 100 percent passenger vehicles on level tangent sections and a value of time of \$3.70/h/passenger vehicle.

Capital cost was brought back to an equivalent uniform annual cost, and annual maintenance cost was added. This cost was then brought down to an average hourly cost and divided by the volume of vehicles for given speeds obtained from the speed-volume curves.

Vehicle running costs for Alberta are empirically derived values presented in tabular form that give the cost to run a vehicle at various speeds. Travel time costs were simply divided by the desired speeds to obtain the cost to travel 1 km at that speed and were added to vehicle running costs. Combining these costs resulted in the curve shown in Figure 2, which indicates that total unit cost minimizes at a volume-to-capacity (V/C) ratio of approximately 0.28 (level of service B) for passenger vehicles on level tangent sections.

Although this lends some credibility to providing level of service B as a design criterion, there are several factors that will affect the analysis and cause a shift of

Figure 1. Road user and capital costs versus traffic volume.

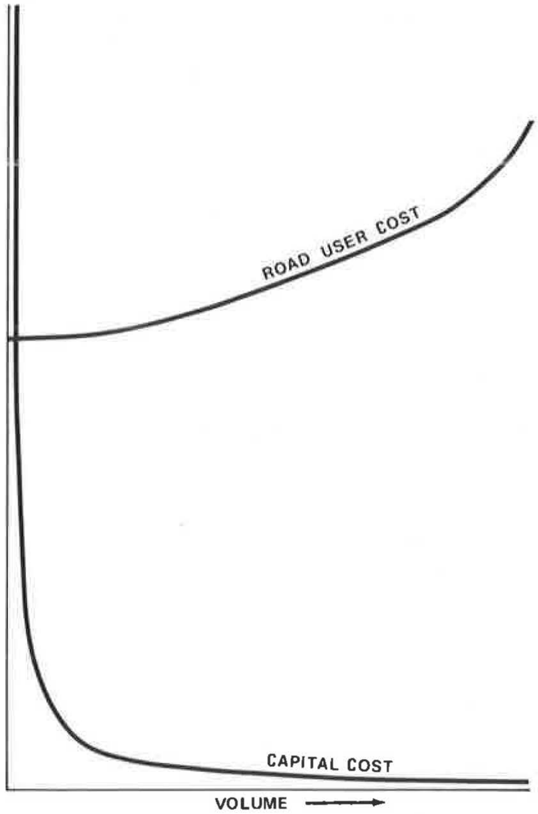
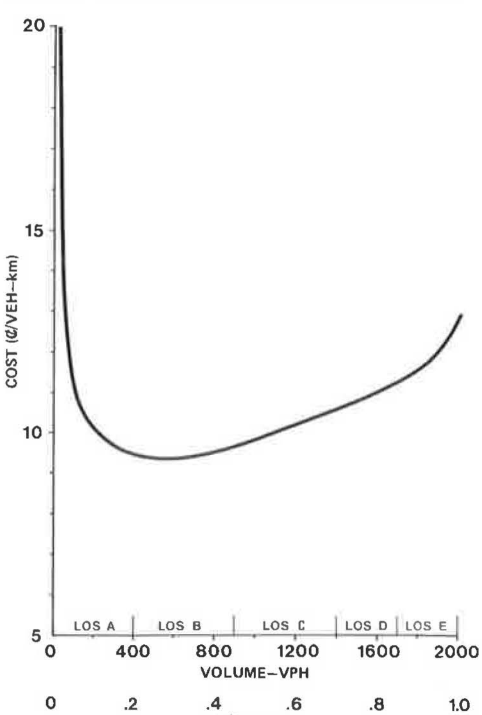


Figure 2. Combined cost versus levels of service.



the minimum cost point: much higher costs as a result of constructing a roadway in difficult terrain; increased running costs on grades and curvatures; the effect of trucks, buses, and recreational vehicles in the traffic stream; and the value of time.

SENSITIVITY TESTING

To reflect varying topography, traffic composition, and

Figure 3. V/C ratio cost minimas versus value of time.

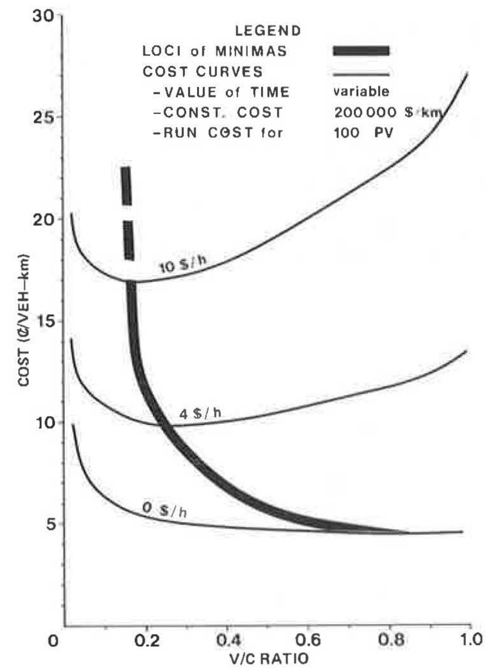
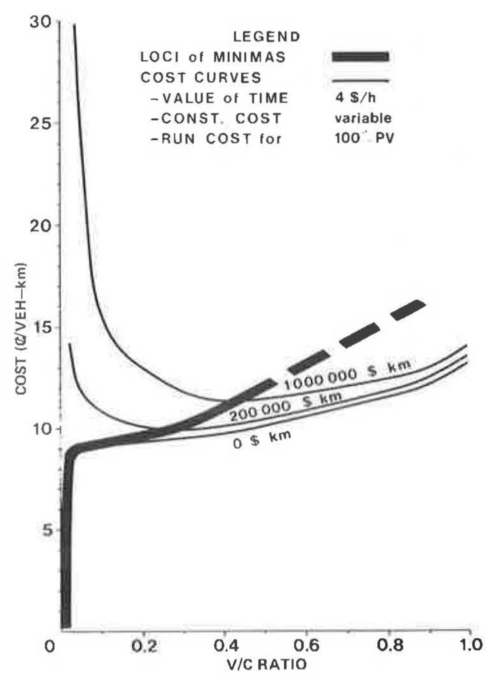


Figure 4. V/C ratio cost minimas versus construction cost.



time values, each cost parameter was varied while the others were held constant. The value of time was varied from $0 to $10/h, construction costs from $0 to $1 million/km, and running costs from $0.043 to $0.063/vehicle-km (as represented by a vehicle composition that varies from 100 to 60 percent passenger vehicles with 20 percent trucks and 20 percent recreational vehicles).

Figures 3-5 show the plots of the loci of the minimas for each varying parameter. Whereas both increased running costs and construction costs tend to shift the point of minimization toward the lower level of service, the value of time is the most sensitive variable—especially for the lower time values—and tends to shift the point of minimization toward the higher level of service with increasing values. This testing indicates that no single V/C ratio can be defined as the most economical and suggests instead a level-of-service range that depends on topography, traffic composition, and trip purpose.

Although the technique presented here provides an economical V/C ratio, it is inherent in the calculations that uniform hourly volumes occur for every hour of the year. Since this is not the case in reality, hourly, daily, weekly, and seasonal variations were investigated by considering (a) daily traffic volumes for two permanent counter locations in the Alberta Primary Highway System and (b) average annual daily traffic (AADT) for 29 permanent counter locations in the province.

The procedure followed was to calculate the total cost for each hour of the year based on the cost-volume relation shown in Figure 2, accumulate these costs for each day and for the year, and then divide by the daily or annual volumes. Each cost-volume data pair was then plotted, and a hand-fit curve was drawn.

Figures 6 and 7 show the daily cost-volume relations generated and show that there is a leveling off of unit costs when the daily volumes approach 7000-8000 vehicles/d. Although minimization is not clearly evident, it appears that minimization occurs at approximately 8000 vehicles/d in Figure 7.

Figure 8 shows the AADT cost-volume relationship. Note that minimization does not occur within the range of available two-lane data. To extend the curve, selected multilane counters were analyzed. Because it was assumed that the volumes carried on these multilane facilities could be accommodated on a two-lane roadway, the costs were calculated by using two-lane capacities. Although there are insufficient data to plot a curve so as to determine the point of minimization with confidence, the same leveling-off trend as that found in Figures 6 and 7 is observed. If minimization did occur, it would not be expected before 8000 vehicles/d (Figure 7). It appears that the traveling public in rural Alberta would not accept this level of service as satisfactory. In fact, Provincial

Figure 5. V/C ratio cost minimas versus vehicle running cost.

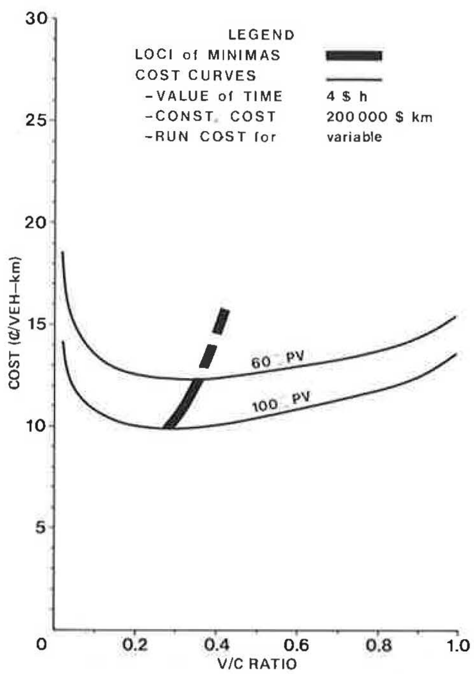
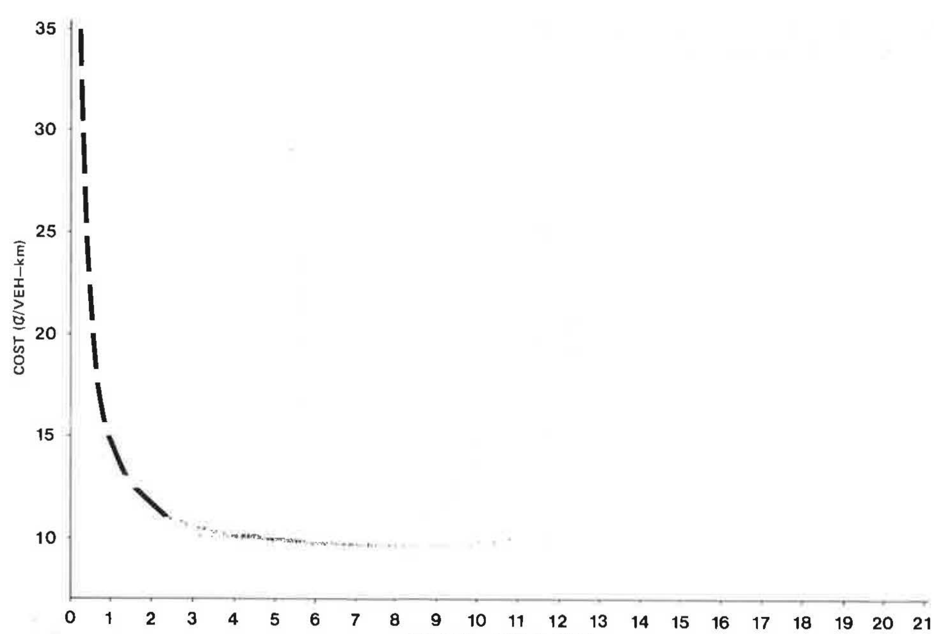


Figure 6. Daily cost-volume relations at counter 63.



Highway 11 from Red Deer to Sylvan Lake (counter 63 shown in Figure 6), which has an AADT of 5600 vehicles/d, is being proposed for upgrading (possibly four lanes or an alternate two-lane route) within the next three years.

The preceding technique can be used to determine the most cost-effective V/C ratio for any given set of circumstances in analyzing the need to upgrade a highway facility or provide an entirely new route. There remains, however, the unresolved issue of relating the cost-effective V/C ratio, determined on an hourly basis, to the usual practice of using a design hourly volume, which is often expressed as a percentage of AADT. This paper attempts to define a rational approach to determining and correlating the relations.

DESIGN HOURLY VOLUME

The design hourly volume (DHV) is the volume of traffic during 1 h that is used as an acceptable operating condition for design purposes. The American Association of State Highway Officials (AASHO) has stated that the DHV represents the load that the highway must accommodate and largely determines the type of facility required and its characteristics (6). The DHV is selected in such a way that the highway under design should not experience extreme congestion at any time or unacceptable congestion for extended periods. However, the DHV must not be such that traffic would rarely be great enough to cause even minimal congestion because the facility would then be overdesigned and uneconomical.

Figure 7. Daily cost-volume relations at counter 36.

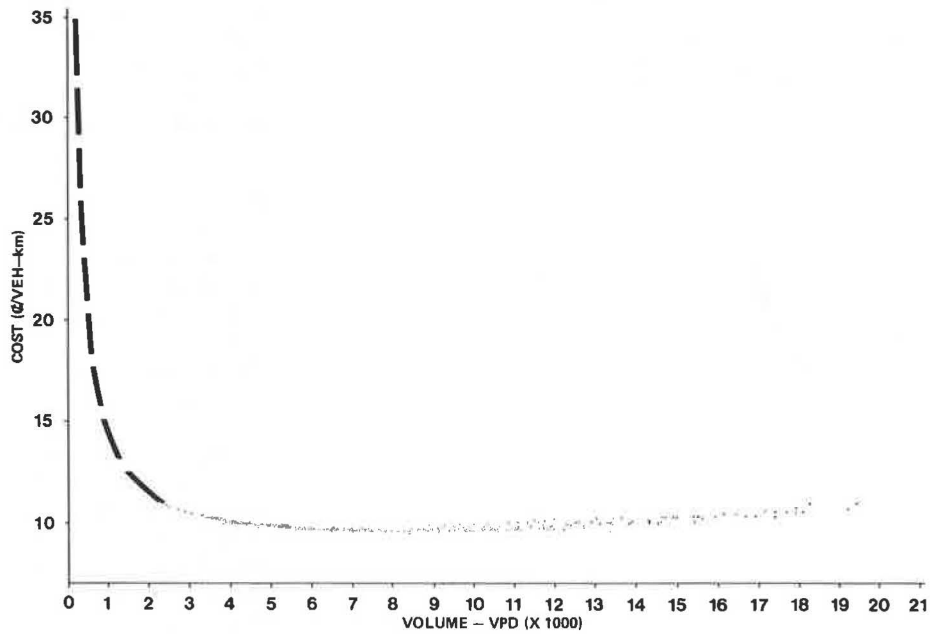
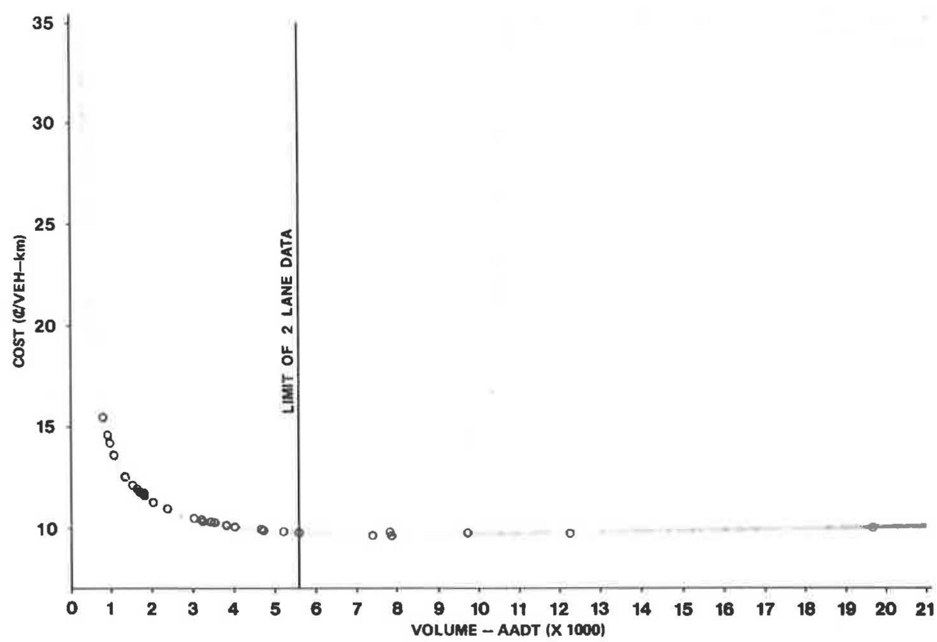


Figure 8. Annual cost-volume relations at Alberta permanent counters.



A common method of determining the DHV, proposed in the 1950 Highway Capacity Manual (7), involves the use of a graph that shows the highest hourly traffic volumes of the year according to rank. The 30th highest hourly volume is used by a number of agencies as the DHV for rural highways on the basis that the slope of the curve changes rapidly at that point and it is there that the ratio of benefits to expenditures is near the maximum. In a

Figure 9. Curves for 5500 highest hours of the year at four Alberta counter locations.

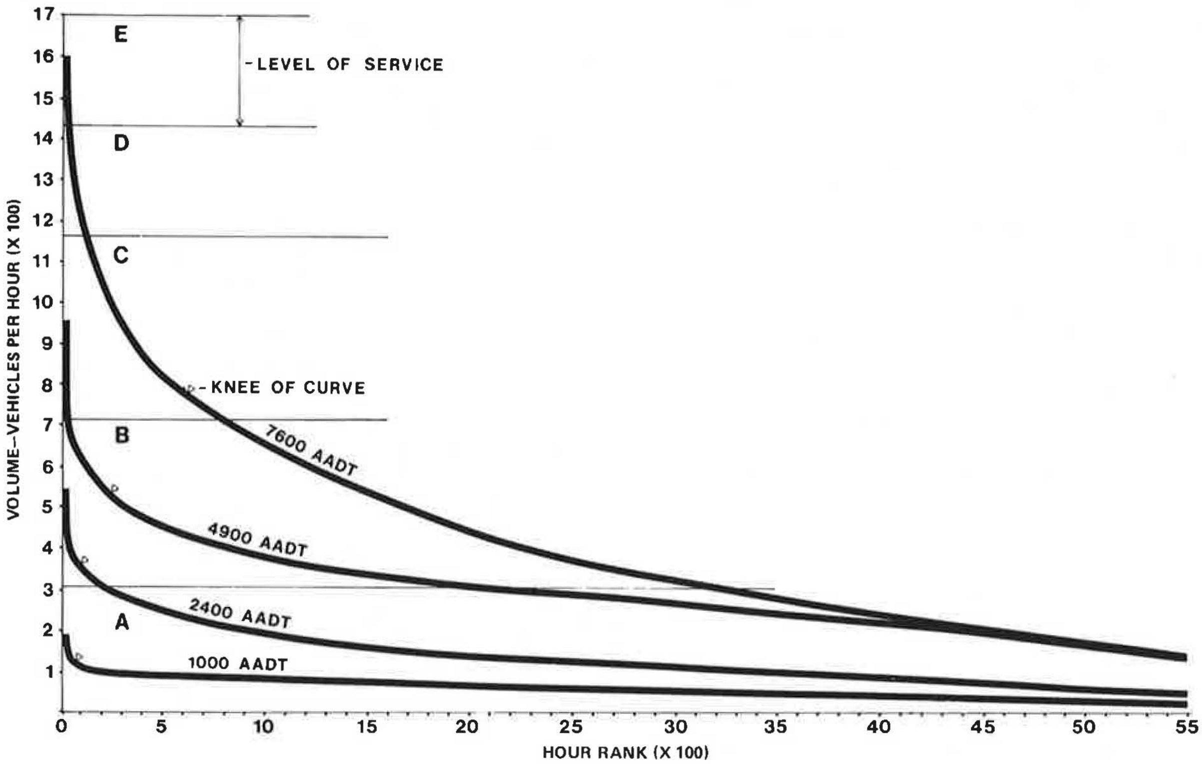


Table 1. K-factors for 30th highest hour of the year and knee of the curve.

Counter	Location	Type of Facility	AADT	Year	Hour at Knee	K-Factor Knee	K-Factor 30th Highest Hour
78	Highway 14	Two-lane undivided	1 000	1976	50	13.5	14.5
111	Highway 16	Two-lane undivided	2 400	1976	100	14.2	16.2
63	Highway 11	Two-lane undivided	4 900	1976	260	10.8	14.8
63	Highway 11	Two-lane undivided	5 600	1977	290	10.0	14.2
36	Trans-Canada Highway	Four-lane divided	7 600	1976	600	10.2	18.1
36	Trans-Canada Highway	Four-lane divided	7 800	1977	620	10.5	17.6
60	Highway 2	Four-lane divided	9 750	1977	570	9.7	17.0
57	Highway 2	Four-lane divided	12 250	1977	440	9.6	14.0
102	Highway 16	Four-lane divided	19 700	1977	480	10.1	12.7

Table 2. Tentative design standards for rural highways in Alberta.

Design Classification[a]	Type of Highway	No. of Lanes	Surface Width[b] (m)	Maximum Posted Speed (km/h)	20-Year Design AADT	Level of Service	DHV K-Factor
RFD-4[c]	Rural freeway divided	4		110	30 000-60 000	B-C	0.10
RAD-4[c]	Rural arterial divided						
RAU-213[d]	Rural arterial undivided	2	13	100	6000-10 000	B-C	0.10
RAU-211[d]	Rural arterial undivided	2	11	100	5000-9000	B-C	0.11
RAU-209[d]	Rural arterial undivided	2	9	90	4000-8000	B-C	0.12
RCU-209[d]	Rural collector undivided						
RCU-208[d]	Rural collector undivided	2	8	90	3000-7000	B-C	0.13
RLU-208[d]	Rural local undivided						

[a]Alberta Transportation classifications.
[b]Outside shoulder to outside shoulder.
[c]Design volume varies because of traffic composition and directional split.
[d]Design volume varies because of traffic composition and passing sight distance.

case in which the slope changes rapidly at some point other than the 30th highest hour, the DHV is chosen at the "knee of the curve".

No apparent attempt was made by the proposers of the method discussed here to justify or prove that these points in the curve do in fact provide the greatest economic benefit. Further, there is no clear indication as to what level of service should be chosen for the DHV. This has been left to the agency to determine and usually has been a policy decision of one form or another that is often made on a very obscure basis.

It is standard practice to determine a future DHV by multiplying the estimated or forecast AADT by a value K (i.e., the ratio of DHV to AADT), the hour used often being the 30th highest hour that is expected to occur in some future design year.

The 1965 HCM (4) recognizes the problem of selecting a measured or predicted traffic volume to be used for design purposes and, although the 30th highest hour is discussed, the HCM states the following: "This frequent reference to the 30th highest hour should not be misconstrued as a recommendation for rigid adoption, but rather as an example of typical highest hour relationship and trends." The following discussion is intended to provide some further insight into the process of selecting the DHV K-factor.

Highest-Hour-of-the-Year Signatures and K-Factors

It appears that highway agencies have traditionally ranked only the first 100 to 250 highest hours of the year. The remainder have been considered of little importance because the knee of the curve was usually evident within the first 100 hours. Based on a limited sample of Alberta counter locations, where all 8760 h were ranked, it appears that this is not the case. The highest 5500 h of four of these locations are shown in the graph in Figure 9. The knee of the curve is very evident for the lower-volume road; however, as the AADT increases, the knee disappears from within the first 100 h and shifts to somewhere in the 200- to 600-h range. This shift, of course, results in different K-factor values for the knee of the curve than for the 30th highest hour. Table 1 compares K-factors based on the hour at the knee of the curve with those based on the 30th highest hour. The knee-of-the-curve values are lower, within a narrower range, and tend to decrease in value as AADT increases. There is also a tendency for the knee of the curve to occur at a higher-ranked hour as AADT increases.

Formulation of Design Criteria

In arriving at a basis for choosing a DHV and a level of service for the DHV, the following guidelines are proposed:

1. The DHV chosen for the highway under design should be such that traffic demand for other higher hours of the year will not exceed the capacity of the facility for even short intervals of time except under rare or very exceptional circumstances.
2. The level of service chosen should provide the driver with various degrees of choice of speed and freedom from tension consistent with the length, duration, and purpose of the trip.
3. The attitude of motorists toward adverse operating conditions is influenced by their awareness of the environment in which they are traveling (e.g., difficult topography and built-up areas) and their recognition of associated practical cost limitations that preclude the design of the ideal facility.

The computations presented in this paper are based on a limited analysis that has given some further insight into formulating design criteria. Based on the rationale that the most economical DHV occurs where the slope of the curve for the highest hours of the year changes most rapidly, the knee of the curve and associated K-values appear to be most appropriate even though there is no known quantitative basis for their use. It follows that the level of service for the DHV should be equivalent to the V/C ratio where total unit costs are minimized.

Although no clear mandate has been presented, the approach suggested here will permit planners and designers to develop and select criteria on a more sound economic basis. This, of course, results in a wide range of DHVs and levels of service. These are given in Table 2, which has been formulated based on the work presented here. Since the table represents a very limited number of site-specific cases, it is by no means final.

CONCLUSIONS AND RECOMMENDATIONS

Although this analysis has been rather limited in scope because of the lack of data and the use of manual methods, we feel that the work is sufficiently valid to make some preliminary conclusions and recommendations for further research. The following conclusions can be made:

1. A cost-effective V/C ratio can be computed for various supply conditions by using the technique described.
2. Evidence presented on the demand side further supports the use of K-values for the knee of the curve rather than use of the 30th highest hour for the DHV because the values are more consistent.
3. Daily and yearly cost-volume relations do not indicate a cost-effective volume as clearly as does the hourly measure.
4. Economic justification for converting existing two-lane facilities into multilane facilities (as the public now demands) does not appear evident based on the measures of operating efficiency presented. However, the technique is felt to hold some merit as one of the parameters for priority rating.

The following recommendations are made for further research:

1. Procedures for measuring economic operating efficiency should be refined. Several areas require attention, namely (a) capacity and level-of-service volumes and corresponding speeds for two-lane roads (speed-volume curves) require validation (this is currently one of the greatest gaps in two-lane highway capacity theory), (b) vehicle operating costs for Alberta should be updated, and (c) value of time requires considerably more analysis and understanding and the derivation of values for different trip purposes and trip lengths.
2. Data for two-lane roads with higher AADTs should be analyzed to validate further the concept of the knee of the curve for DHV, the K-values derived so far, and the

unit average cost supply functions (Ontario may be one of the few Canadian sources for this information).

3. Although direct relations between the knee of the curve and the K-factor and cost-effective V/C ratios can be shown, the relation between economic level of service (supply) and DHV (demand) is still obscure and requires further research.

ACKNOWLEDGMENT

The observations and views presented in this paper are strictly our own. The design criteria presented are not formal Alberta Transportation policy.

REFERENCES

1. Z. Haritos. Rational Road Pricing Policies in Canada. Canadian Transport Commission, Ottawa, 1973.
2. N. Cameron. Determination of Design Hourly Volume. Univ. of Calgary, Alberta, thesis, May 1975.
3. R. Winfrey and C. Zellner. Summary and Evaluation of Economic Consequences of Highway Improvements. NCHRP, Rept. 122, 1971.
4. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
5. B. Ashtakala. Alberta Road User Costs. Alberta Transportation, 1976.
6. A Policy on Geometric Design of Rural Highways. AASHO, Washington, DC, 1965.
7. Highway Capacity Manual. U.S. Government Printing Office, 1950.

Freeway Level of Service: A Revised Approach

Roger P. Roess, William R. McShane, and Louis J. Pignataro, Polytechnic Institute of New York, Brooklyn

Concepts, philosophies, and standards for freeway level of service presented in the 1965 Highway Capacity Manual are reviewed. A revised approach is developed that incorporates density in the definition of standards. Speed-flow relations under ideal conditions are approximated based on secondary source data and a limited number of pilot field surveys associated with current work. The recommendations made for new level-of-service standards for freeways are based on recalibrated speed-flow relations and incorporate density as a parameter.

The basis for any technique of capacity analysis is the definition of quality-of-service criteria and the correlation of these criteria with operational and design parameters. The 1950 Highway Capacity Manual (1) defined service in terms of "possible" and "practical" capacity. Practical capacity represented the maximum traffic volume that could be accommodated (under prevailing roadway and traffic conditions) while an acceptable quality of service was provided.

The 1965 Highway Capacity Manual (HCM) (2) introduced the concept of level of service, which allows for a more detailed treatment of service quality. The 1965 HCM defines level of service as "a qualitative measure of the effect of a number of factors, which include speed and travel time, traffic interruptions, freedom to maneuver, safety, driving comfort and operating cost" on operations. It also defines six levels of service—A through F—which describe a wide range of conditions, from totally free at level A to forced flow at level F.

CURRENT STANDARDS FOR LEVEL OF SERVICE

Current standards for freeway level of service are given in Table 9.1 of the 1965 HCM (2, pp. 252-253). Each level is a range of operating conditions for which the table defines boundary conditions in terms of two parameters: (a) volume-to-capacity (V/C) ratio, which may be stated as a volume, and (b) operating speed. Table 9.1 gives minimum values of operating speed and maximum V/C values for each level of service. The standards in the table apply under "ideal" conditions, which include (a) no trucks or buses in the traffic stream, (b) 3.6-m (12-ft) minimum lane widths, and (c) no obstructions in the median or roadside area closer than 1.8 m (6 ft) to the pavement edge. The standards for the V/C ratio depend on average highway speed, which is a weighted average design speed for the highway segment under study.

For a highway segment to be said to operate under a particular level of service, the criteria for both V/C ratio and operating speed must be met. This is an important point. The standards in Table 9.1 of the 1965 HCM do not, nor were they intended to, represent a correlation between speed and V/C ratio. The existence of a V/C ratio appropriate for level of service C does not guarantee that the operating speed for that level will also be met. This characteristic of the standards leads to a number of problems in their use.

QUESTIONS, ISSUES, AND ALTERNATIVES

In formulating recommendations for level-of-service standards, a number of critical philosophic and practical issues must be raised. The resulting recommendations should meet two primary objectives:

1. Levels of service must be defined in terms that are meaningful for the driver who experiences them and meaningful to the planners, analysts, and designers who will use the standard.
2. Definitions of level of service must be consistent with each other and consistent in application to the various types of subsections that occur on a freeway (i.e., open sections, weaving areas, and ramp terminals).

A number of key issues concerning the concept of level