Intersection Capacity Analysis: Are You Doing It Wrong?

By Tom Creasey, P.E., Ph.D. (M) and Bill Sampson, P.E. (F)

The *Highway Capacity Manual (HCM)* Signalized Intersection Analysis method has been one of the most often used analysis tools since the "modern" method was introduced in the 1985 edition of the *HCM*.[1] The method is applied to evaluate the sufficiency of intersection operations for varying levels of traffic demand, both undersaturated (i.e. demand less than physical capacity) and oversaturated. The sad truth is that the method is often wrongly applied when conditions are oversaturated.

Shutterstock/TierneyMJ

The method is wrongly applied in a few ways:

1. A single period analysis is applied when a multiple period analysis should be used; or
2. Traffic demand is underestimated.

Single Period vs. Multiple Period Analyses

Most of the procedures in the *HCM* are based on the peak 15-minute rate of flow, which is then converted to an equivalent hourly rate using the peak hour factor (PHF), where:

$$PHF \frac{Hourly\ Volume}{Peak\ rate\ of\ flow\ within\ the\ hour}$$

For a 15-minute peak flow rate, this equation becomes:

$$PHF \frac{V}{4xV_{15}}$$

where V is the hourly volume (in vehicles per hour) and V_{15} is the volume during the peak 15 minutes of flow. As part of the analysis, the demand volume is divided by the PHF to represent an equivalent hourly volume for the peak 15-minute period of the hour. In other words, *the* HCM *method is an analysis of the heaviest or worst 15 minutes during the peak hour.*

What happens if there are multiple periods when traffic conditions are bad? Furthermore, how might congested traffic conditions in one 15-minute period affect traffic conditions in an adjacent, subsequent period? The *HCM* Signalized Intersection delay equation (delay is the performance measure upon which intersection level of service is based) was modified in the 1997 update to the *HCM* to account for this. In the 1985 *HCM*, the delay for each lane group was expressed:

$$d = d_1 + d_2$$

Where d is the average stopped delay (in seconds) per vehicle for the lane group, d_1 is the uniform delay that occurs if arrival demand in the subject lane group is uniformly distributed over time, and d_2 is the incremental delay of random arrivals over uniform arrivals and for the additional delay due to cycle failures. Both terms where multiplied by a progression adjustment factor (PF) in computing average stopped delay and corresponding level of service.

The method was modified with the 1997 update to the *HCM* (published in 1998).[2] The progression factor, PF, was applied to the uniform delay (d_1) component only and a third term was introduced:

$$d = d_1\,PF + d_2 + d_3$$

The delay definition was changed from stopped delay (i.e. delay from being stopped at an intersection) to control delay (total delay including stopped delay plus delays incurred during deceleration and acceleration). Most importantly, delay was defined to include delays incurred beyond the analysis period when the lane group is oversaturated. The d_3 term was defined as, "residual demand delay to account for oversaturation queues that may have existed before the analysis period." A method for estimation of the d_3 initial queue delay term was included in an appendix to the method.

The method was clearly improved in its ability to address oversaturated conditions. This improvement, however, raised questions in its application:

- If multiple 15-minute periods within a peak hour are oversaturated, which one should be the focus of the analysis?
- What if the entire peak hour is oversaturated?

Beginning with the *HCM* in 2000,[3] specific guidance was given directing the analyst to study the entire period during which volumes approach and exceed capacity, even if the duration of the period was greater than one hour. Furthermore, lane group volumes should reflect the actual demand and not a measured or counted volume, as the demand is not entirely served during periods of oversaturation. A greater emphasis was placed on computing the initial queue delay (d_3) as the procedure was extended to analyze delay over multiple time periods. As stated in the current *HCM*, 6th Edition:[4]

> "If the analysis period's demand volume exceeds capacity, then a multiple time-period analysis should be undertaken when the study period includes an initial analysis period with no initial queue and a final analysis period with no residual queue. … This approach provides a more accurate estimate of the delay associated with the congestion."

What are the results of performing a single-period analysis when conditions are oversaturated?

1. The estimate of delay associated with congestion will be less accurate, much more inaccurate as demands increase.
2. The estimate of delay will be less than the delay computed from a multiple-period analysis where the initial queue is computed for each individual analysis period.
3. Resulting selected mitigation measures may not be sufficient due to the underestimation of delay.

Why is this important? Underestimating delay can result in signal timing with shorter cycle lengths and phase times that do not process the actual demand. Another outcome of these incorrect analyses would include inadequate turn bay lengths due to the underestimation of queues. Where developer impact fees are charged, underestimating the delay can mean the traffic impacts are not fully mitigated and the collected fees insufficient to provide the proper improvements.

The following example illustrates the point. The source for this example is Example Problem 1 from Chapter 30 Urban Streets: Supplemental in the *Highway Capacity Manual,* 6th Edition. Hourly traffic volumes from Intersection 1 are used but have been inflated by 50 percent so that some operations for some of the lane groups are oversaturated. These are illustrated in Figure 1.

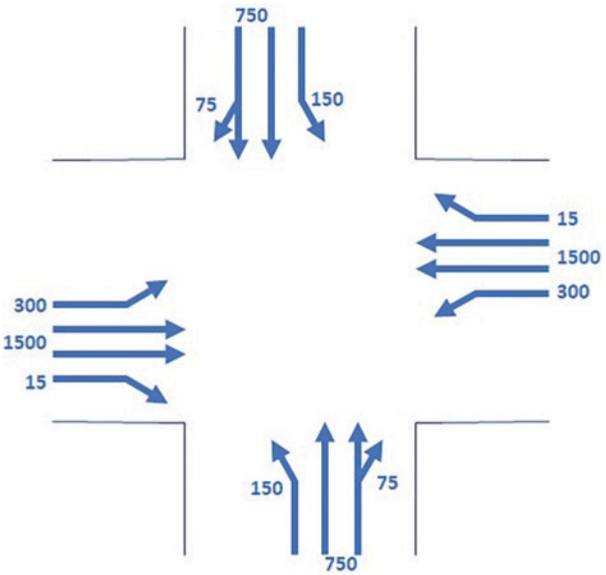

Figure 1. Example of Peak Hour Intersection Movement Demand Volumes.

The example focuses on the eastbound left turn and through movements. The left turn movement is slightly under capacity (with a volume-to-capacity ratio slightly less than 1.00), while the through movement is oversaturated. Two analyses were performed: 1) a single period analysis using PHF = 1.00 (i.e. an equal distribution of traffic demand across the peak hour); and 2) a multiple period analysis consisting of four 15-minute periods with traffic demand equally distributed.

For selected performance measures, results of the comparison between the two approaches is summarized in Table 1.

The eastbound left turn lane group is undersaturated. There is no Initial Queue Delay (d_3) component and the results are nearly identical. This is not the same for the eastbound through lanes, where the volume-to-capacity ratio is greater than 1.00. Compared with the single period analysis, the Incremental Delay (d_2) increases with successive time periods. The most remarkable difference is in the d_3 term, the unmet demand that carries over from the end of one 15-minute time period to the beginning of the next. This value is zero for the single period analysis, assuming no unmet demand exists when the analysis period begins. Even if an initial queue is recorded at the beginning of a single period analysis, the cumulative effects of cycle failures are not carried forward and the delay is underestimated. In many cases, it can be grossly underestimated, as this illustration shows. The table for this example shows a control delay of 900 seconds per vehicle (s/veh) in TP 4 of the multiple-period analysis, compared with 141 s/veh in the single-period analysis, for a difference of 574 percent. Similar differences occur for back of queue with 32 vehicles per lane (veh/ln) in the single-period analysis compared with 159 veh/ln in TP 4 of the multiple-period analysis for a 408 percent difference. These dramatic differences occur for this lane group with a volume-to-capacity ratio of only 1.3, which can be much higher at many signalized intersections during peak periods.

Underestimating the Demand

Counting vehicles as they cross the stop line is not adequate for collecting data to analyze congested conditions. When conditions become congested, stop line counts reflect capacity and not actual

Table 1. Summary Comparison of Single Period and Multiple Period Analysis Results.

	Eastbound Left Turn					***Eastbound Through***				
	Single	Multiple 15-Minute Periods				Single	Multiple 15-Minute Periods			
Performance Measure	Period	TP1	TP2	TP3	TP4	Period	TP1	TP2	TP3	TP4
Volume-to-Capacity Ratio (X)	**0.911**	**0.911**	**0.911**	**0.911**	**0.911**	**1.239**	**1.243**	**1.305**	**1.305**	**1.305**
Uniform Delay (d_1), s/veh	39.6	39.6	39.6	39.6	39.6	26.4	26.5	28.7	28.7	28.7
Incremental Delay (d_2), s/veh	25.4	25.8	25.8	25.8	25.8	114.6	116.6	143.6	143.6	143.6
Initial Queue Delay (d_3), s/veh	0.0	0.0	0.0	0.0	0.0	0.0	0.0	229.8	504.2	778.6
Control Delay (d), s/veh	65.0	65.4	65.4	65.4	65.4	141.0	143	402.1	676.5	950.9
*Control Delay Difference, %**							*1%*	*185%*	*380%*	*574%*
Back of Queue (veh/ln)	9.4	9.4	9.4	9.4	9.4	31.3	31.6	71.4	115.2	159.0
*Back of Queue Difference, %**							*1%*	*128%*	*268%*	*408%*
Initial Queue (veh)**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	73.7	161.0	248.6

* Compared to Single Period Anaysis

** For the entire EB Through lane group

No significant difference for undersaturated

Significant differences in successive time periods for oversaturated demand, as witnessed by queues that form and grow during the congested period. If actual demand data are not obtained when conditions are congested, the flow rate cannot exceed capacity (by definition) and the analysis can significantly underestimate delay and queues. Once demand exceeds capacity, the arrival rate upstream of the stop line must be known in order to accurately estimate the demand.

An illustrative example is provided in Table 2, where hypothetical volumes for an intersection lane group are shown. The lane group capacity is 500 vehicles per 15-minute period. Stop line departures are counted and totaled every 15 minutes. Counts for time periods (TP) 1 and 2 are 300 and 400 vehicles, respectively, which is less than capacity. Accordingly, there is no residual queue at the end of these time periods.

At the end of Time Period 3, 500 vehicles have been observed to cross the stop line and there is a queue of 25 vehicles. The queue continues to grow over successive time periods until it reaches a maximum (325 vehicles) at the end of TP 7. It remains for three more time periods before finally disappearing at the end of TP 11. To accurately compute the demand, the queued vehicles for the period in question (*t*) must be added to the stop line count for the same period, but the queued vehicles at the end of the previous period (*t-1*) are subtracted, as they were the first vehicles to be served in period *t* and are included in the stop line count.

The graph in Figure 2 shows the disparity between stop line count and demand for this illustrative example. Beginning with TP 3 and extending through TP 7, the actual demand exceeds the stop line count, which results in underestimating the d_3 term in the *HCM* delay equation. Because delay increases exponentially when conditions are oversaturated, the delay can be grossly underestimated.

When oversaturation is reached, stop line counts will be the same (or very similar) for each time period and there will be

Table 2. Tabulation of Demand Over Multiple Time Periods.

Time Interval	Time Period No.	Stop Line Count (veh)	Period End Queue Length (veh)	Demand Arrival Volume* (veh)
4:00 - 4:15	1	300	0	300
4:15 - 4:30	2	400	0	400
4:30 - 4:45	3	500	25	525
4:45 - 5:00	4	500	75	550
5:00 - 5:15	5	500	150	575
5:15 - 5:30	6	500	250	600
5:30 - 5:45	7	500	325	575
5:45 - 6:00	8	500	300	475
6:00 - 6:15	9	500	200	400
6:15 - 6:30	10	500	100	400
6:30 - 6:45	11	400	0	300
6:45 - 7:00	12	300	0	300

$* \, Demand_t = Departures_t + Queue\ Length_t - Queue\ Length_{t-1}$

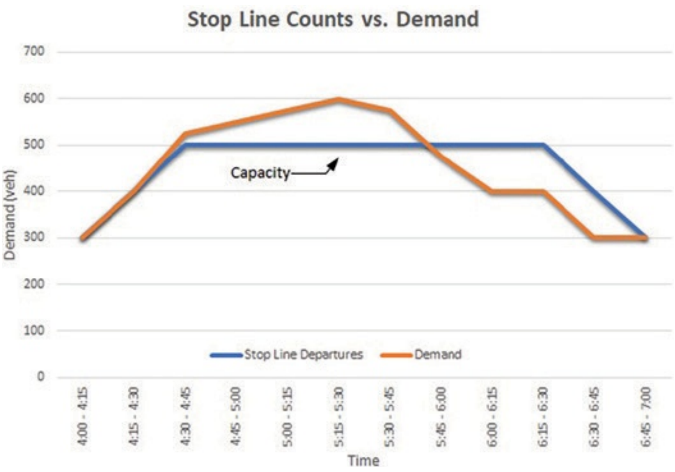

Figure 2. Comparison of Demand versus Capacity (Stop Line Departures).

residual queues. By counting the number of queued vehicles at the end of each time period, the actual demand can be estimated using the method provided in the example.

Application

Several commercial software tools implement the deterministic *HCM* methods in their analysis of signalized intersections. To accurately estimate delay when conditions are oversaturated, tools should be capable of:

- Performing multiple-period analyses when conditions are oversaturated; and
- Estimating unmet demand at the beginning of each analysis period.

Before using any particular tool, it is the obligation of the analyst to ensure the correct application of the tool when conditions are oversaturated. Failing to do so undermines the credibility of the tool, the analyst and the conclusions drawn from the evaluation. In reporting the results, performance measures (including delay) should be tabulated for each 15-minute time interval within the study period. While this may expand the reporting of results, it provides a more accurate picture of intersection operations within the entire peak period, regardless of its length, and more importantly leads to better decision making in mitigation efforts.

Some may choose to use microscopic simulation as an alternative tool. Simulation tools compute delay differently than the deterministic method documented in the *HCM*. However, a properly calibrated simulation model should provide similar results to the *HCM* method. Regardless of which approach is taken, the temporal variation in demand over the analysis period should be adequately reflected, whether in a multiple period deterministic analysis or a microsimulation analysis.

Conclusion

Either of the mistakes discussed can cause the analyst to underestimate delay when conducting an intersection analysis. When both mistakes are made (also a common occurrence), the errors are exacerbated.

How does one know if a multiple period analysis is needed and if actual demand is greater than stop line counts? A good rule-of-thumb is that a multiple-period analysis should be performed if any one lane group approaches a volume-to-capacity ratio of 1.00. This also would imply that stop line counts should be adjusted to reflect demand, either by using the method illustrated in this article or by collecting concurrent mid-block counts upstream of the stop line (beyond the back of the queue) and adjusting the stop line counts proportionally. **itej**

References

1. Transportation Research Board, *Highway Capacity Manual*, Special Report 209, Washington, DC, USA, 1985.
2. Transportation Research Board, *Highway Capacity Manual*, Special Report 209, Third Edition, Washington, DC, USA, 1998.
3. Transportation Research Board, *Highway Capacity Manual*, Washington, DC, USA, 2000.
4. Transportation Research Board, *Highway Capacity Manual*, 6th Edition: *A Guide for Multimodal Mobility Analysis*, Washington, DC, USA, 2016.

Tom Creasey, P.E., Ph.D. (M) *is vice president of Transportation Engineering for Caliper Corporation. He is a longtime member of the TRB Highway Capacity and Quality of Service Committee, which oversees the* Highway Capacity Manual, *and currently serves as chair of the committee. Tom received his bachelor's, master's, and Ph.D. degrees in civil engineering from the University of Kentucky.*

Bill Sampson, P.E. (F) *is the director of the McTrans Center at the University of Florida and is responsible for technical assistance and ongoing development of the Highway Capacity Software. He is a former member of the TRB Highway Capacity and Quality of Service Committee and has been continually active for more than 30 years. Bill received his bachelor's and master's degrees in civil engineering from the University of Kentucky.*