Effect of Weather on the Relationship Between Flow and Occupancy on Freeways

Fred L. Hall and Deanna Barrow

The relationship between flow rates and roadway occupancies on freeways has been investigated in several recent papers. However, all the data in those investigations have come from ideal conditions. The purpose of this paper is to investigate the same relationship under adverse weather conditions. An earlier study found that rain reduced freeway capacity, but it did not address the effects of rain on the nature of the function relating the two variables. Three possible effects on the function are investigated, and it is found that, in essence, the slope of the line relating flow to occupancy (in the uncongested regime) decreases as weather conditions deteriorate. In other words, different parameters are needed for the function to describe the relationship under different weather conditions. This finding has some important implications for efforts to develop a new incident detection algorithm for freeways, based on the nature of the flow-occupancy relationship. This paper addresses the effect of bad weather on freeway traffic operations. In particular, does rainy weather change the nature of the relationships among speed, flow, and roadway occupancy? Although it might be objected that the existence of such an effect is obvious and hardly merits detailed attention, it turns out to be important to develop a quantitative description of the effect for use with an automatic incident detection logic. This paper begins by briefly describing the background for the incident detection logic, in order to explain the rationale for developing a solid quantitative treatment of the effect of bad weather. The second section describes the data that were used for the analysis, with particular reference to weather conditions. The third section, dealing with the analysis, includes both a discussion of the methods used and a presentation of the results. The fourth section presents the conclusions that have been drawn from the investigation.

Recently, Navin (1) and Hall (2) proposed models of freeway operations based on catastrophe theory. One practical consequence of this proposed model is that it provides the possibility for a new logical basis for incident detection, described by Persaud and Hall (3), and suggested earlier by Athol (4). However, for that logic to function effectively, it is first necessary to ensure that the theoretical picture encompasses the full range of operating conditions that are likely to be encountered. Jones et al. (5) suggested that capacity is reduced during rain, although their functions appeared to fit rather poorly at capacity. If their functions are correct, however, this finding could arise in several ways, each of which implies a different consequence for the theoretical picture of traffic operations as represented by catastrophe theory. Athol (4) sketched one of these ways, but it appears to be only an impression, not fully analyzed.

The catastrophe theory model portrays freeway operations data as falling on a partially folded surface. Figure 1 portrays the general shape of this surface, the location of freeway data on it, and the two-dimensional projections from that surface, in the form of speed-flow, speed-occupancy, and flow-occupancy graphs. This model offers a number of improvements over the more conventional model, including the ability to account for the sudden jump in speeds that takes place during transitions to and from congested operations. In addition, the model makes clear that operations do not have to pass through capacity flows in moving to or from congested conditions.

One important empirical finding in the context of the catastrophe theory model is that uncongested operations occur fairly close to the "edge" on the upper fold; that is, small increases in occupancy for constant volume cause operations to "drop over the edge" into congestion. Although this observation needs additional confirmation, it appears to hold true for two different sets of Ontario data and holds the key to the automatic incident detection logic proposed by Persaud and Hall (3). That logic depends on being able to specify the functional form for uncongested operation and then identifying departures from the general pattern that would indicate movement to congestion. The advantage of this logic is that comparison with other locations would not be needed and detection of incidents could thereby be speeded up.

However, for such a logic to be practical, it has to cover more than just good weather days. Before this paper, all the analyses of traffic data in the context of catastrophe theory were intentionally limited to data collected under ideal conditions. From the results of Jones et al. (5), it is clear that capacity is affected. It is not clear whether the underlying relationships among the three variables are also affected, or whether data simply arise over only part of the normal range. In terms of the flow-occupancy curve (which earlier studies using the catastrophe theory model found to have an inverted V shape), three plausible situations could give rise to the results of Jones et al. First, the slope of the lines making up the inverted V may decrease, so that flow is lower at any occupancy. This situation would be a natural result of lower speeds and is the picture sketched by Athol (4). Second, it may be that the location of the function is unchanged at low flows but that for higher flows the slope is decreased. And third, it may be that the location of the line is unchanged but that operations simply do not achieve the higher flow rates during rainy weather. In this case, the data might tend to show an inverted U shape rather than the sharper point of the inverted V.

The proposed incident detection logic would be simplest to implement if the possibility holds, but whether it holds can

Department of Geography, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada.
be ascertained only by rejecting the other possibilities. Regardless of which one is the case, it is obviously necessary to resolve this question before attempting to implement an automatic incident detection system based on this logic. Similar questions arise for geometric issues, one of which (the effect of grade) was examined by Persaud and Hall (6).

DATA

The traffic data were obtained from the Burlington Skyway Freeway Traffic Management System (FTMS), which encompasses approximately 10 km on the Queen Elizabeth Way (QEW) near Hamilton, Ontario. Three aspects of the data acquisition are described: first, the selection of appropriate locations in the FTMS; second, the weather data and selection of appropriate days from the record for analysis; and third, the choice of variables to be used.

Locations

The Burlington Skyway FTMS currently includes 12 data collection stations, 6 northbound and 6 southbound, in addition to closed circuit television and changeable message signs. Several of the 12 stations are located on significant grades, either on the structure over the ship canal connecting Hamilton Harbour to Lake Ontario or in the vicinity of overpasses for city streets. Because it seemed likely that grades would have their own effect on the underlying relationships at issue in this study (see Persaud and Hall (6)), it was necessary to select only locations on relatively level sections of the roadway. Two stations, northbound 7 and northbound 12, met these conditions. Station 7 was selected as the location for the extraction of data from the FTMS records.

Weather Data

The Monthly Meteorological Summary (7) for the Hamilton Airport weather station, produced by Environment Canada, provided a good basis for determining weather statistics. The monthly reports contained hourly data, including measures of temperature, wind direction, wind speed, and precipitation.

August 1986 and November 1986 were chosen for analysis. Two types of days were selected. Days having prolonged periods of rain, snow, or freezing rain were used to identify the effect
of adverse weather conditions on traffic. Days having no precipitation occurrences were used for comparison.

Because the FTMS site is located about 20 km (12 miles) from the Hamilton Airport weather station, there was some question whether the summary of weather conditions recorded at the airport accurately reflected the weather conditions at the Burlington Skyway. The Ontario Ministry of Transportation and Communications (MTC) kept detailed road condition reports for the winter months, mid-November to mid-April. However, for the remainder of the year, the only weather information available at FTMS was a handwritten log kept when the weather had a direct effect on traffic. We did not wish to rely on the FTMS log because there might well have been times when the weather had not affected the traffic, with the result that relying on the log would give a biased picture of the effect. Hence the Monthly Meteorological Summary (7) was the primary source for determining weather conditions at the Burlington Skyway.

Of the 31 days in August, 4 days had consistent reports of rain throughout the day. The 4 days (August 1, 6, 7, and 23) had 8 or more hours of rainfall each. Limiting the analysis to these days was intended to minimize the chance that the rainfall at Hamilton Airport was the result of a localized summer thunderstorm that may not have affected the skyway. (Extended periods of rainfall almost certainly indicate widespread rain.)

To assist in determining the road conditions for the month of November 1986, a comparison was made between MTC's independent data on road conditions and Environment Canada's monthly weather summary. In November 1986, 14 days were recorded by Environment Canada as having some form of precipitation at the airport. Comparison with MTC's records showed only 2 days with wet or slippery road conditions on the skyway: November 20 had continuous snowfall, poor visibility, and slippery road conditions; and November 26 had continuous rainfall and wet road conditions but above-freezing temperatures. Unfortunately, when the November 26 data were retrieved, no usable data were found, apparently because of detector malfunctions. Consequently, all the rainy weather data for the analysis were obtained from August.

It was essential to analyze several days when weather conditions were ideal. For consistency, the clear weather data were also extracted from August. Three days were chosen (August 4, 20, and 22), all of which had reports of no precipitation in Environment Canada's monthly weather summary.

**Variables**

After the locations and the days for analysis were selected, it was necessary next to select the variables. At each station, there are pairs of speed measuring detectors in each lane. The FTMS records upstream and downstream flow rates, upstream and downstream occupancies, mean speed, and mean vehicle length for 30-second intervals, 24 hours a day, 7 days a week for each detector pair. Flow rates and occupancies were selected as the variables to work with. The decision to use occupancies rather than to calculate densities was made in earlier work and was discussed in Hall et al. (8). The two main advantages for the present work were that occupancies is the variable provided by most freeway management systems (and therefore most familiar to those managing such systems), and that it provides a tighter fit against flow than does density. The flow-occupancy relationship was selected for analysis over the other two relationships because the sharply peaked nature of the flow-occupancy relationship makes it easier to separate congested and uncongested operations.

**ANALYSIS**

Several steps were followed in the analysis. The first was a simple visual inspection of scatterplots of the data, to see whether they conformed generally to the picture developed earlier using 5-minute data (8, 9). Because the data did conform, the second step was to remove the congested data from the file, in order to concentrate on the relationship for uncongested operations. The third step was to fit functions to those data, for each day separately and then for combinations of days. Several procedures were used in this step. A fourth step was to look for differences only in the high flow ranges, by fitting functions to a restricted range of occupancy values. In these ways, all three possibilities raised earlier were investigated.

Inspection of data plots for flow versus occupancy indicated general similarities between the rainy weather (fig. 2) and clear weather data (fig. 3), as well as with plots based on 5-minute averages (fig. 4). The left-hand, or uncongested, side of the curve is represented by a tight cluster of points, whereas the right-hand side is represented by a large scatter of points. This correspondence in the general pattern allowed for a more detailed, statistical comparison. Had the overall patterns been dissimilar, there would have been no need to proceed.

Maximum flow rates are higher in figures 2 and 3 than in figure 4 because in all cases the flow rates are calculations of hourly values, based on short interval observations. Figures 2 and 3 represent 30-second observations, whereas figure 4 is based on 5-minute counts. In fact, the maximum flows observed in figures 2 and 3 are not sustainable for more than 30 seconds. Detailed inspection of the data shows that no two consecutive observations have such high flow rates. Despite the volatility of the 30-second data, we have chosen to analyze the data at that level of detail, in order for this analysis to contribute to an efficient incident detection logic.

For two reasons it was decided to limit the analysis to the uncongested portion of the data. First, the rationale for this investigation arises from the need to identify the limits on uncongested operation, to detect incidents that represent departures. Hence the uncongested data are of primary interest. Second, the scatterplots suggest that this portion of the data can be depicted sensibly by a line. The congested data, on the other hand, are so widely scattered as to suggest that it would be misleading to represent them by a single line. Further, if the underlying relationship between these two variables is an inverted V, as has been suggested (8, 9), then the two arms can be fitted separately. Indeed, to include the congested data may well distort the relationship identified for the uncongested portion of the curve.

To ensure that only uncongested data were included in the analysis, the following procedures were used. First, periods of congestion were identified on the basis of speeds. The definition of congestion was based on a speed change of more than 18 km/h over a 30-second interval. The period of congestion begins with a sudden drop in speeds and ends with a
jump of this magnitude back from low to high speeds. At the stations of interest on the skyway, there are no physical bottlenecks that would cause recurrent congestion. Instead, congestion occurs as a result of an incident. In order to be sure to remove all data that might represent transitions between congested and uncongested operations, we also deleted from consideration 5 minutes of data before the drop in speed and 5 minutes of data after the speed jump. Any interval containing missing data for one or more variables was also deleted on the assumption that the variables that were reported for the interval also may have been affected by the temporary malfunction in the detector. This screening of the data began at 6 A.M. on each day and continued until roughly 500 data points were obtained for each day, or until 6 P.M.

Functions were fitted to the uncongested data in order to test the first possibility raised earlier, namely, that the slope of the relationship is reduced by bad weather. The approach taken for the curve fitting was as follows. First, a functional form had to be selected. Then equations were fitted to each day’s data separately, using all of the available uncongested data. Tests were run with dummy variable regressions to see if the observable differences in coefficients were significant. On the basis of these results, it was recognized that it was necessary to draw a sample from each day’s data, in order to meet one of the assumptions of regression analysis more closely, namely that of a uniform distribution of observations across the independent variable. On the basis of this sample, all of the equations were reestimated.
In the process of extracting the uncongested data from the larger set, it became obvious that the critical occupancy (i.e., that at which maximum flows are achieved) is lower for the rainy weather (about 28%) than for the clear weather (about 30%). This was not tested mathematically, but it is worth noting. For the regression analyses, both sets are restricted to values less than or equal to 28% for comparability. Several tests were run to ensure that this limitation did not change the nature of the function for the clear weather days. (Although critical occupancy is an important parameter in some incident detection approaches, it appears to be a more conservative indicator than is necessary and is therefore not pursued in this analysis.)

From the visual inspection, two functional forms seemed plausible: a power function and a linear one. The linear model was rejected because it would not necessarily go through the origin, which is mandatory for an equation representing flow and occupancy. Consequently, the function used was

\[ \text{flow} = b_0 \times \text{occupancy} b_1 \]

and the model actually estimated was

\[ \ln(\text{flow}) = a + b_1 \times \ln(\text{occupancy}) \]

where \( a = \ln(b_0) \).

The results in table 1 based on the full set of data show that all of the equations are quite similar, and there is indeed a fair amount of overlap among the \( b_i \) coefficients for the two weather categories. These results also show that in each weather category there is one outlier from the general cluster of \( b_i \) values.

The obvious next step is to test whether the difference in equations within one weather condition is significant. If it is, then there is little point in looking for significant differences between weather conditions. Initial investigation of this issue showed that before it could be properly resolved, the data needed some further treatment. The observations are clearly not uniformly distributed across the range of densities (see figs. 5 and 6). This distribution is contrary to one of the assumptions of regression analysis and might well cause misleading results when statistical inference is important. To overcome this possible problem, it was necessary to sample from within the available data.

A random sampling procedure was carried out on the full data set for each day. From the full set, 10 flows were randomly selected, without replacement, at each occupancy for each day. When there were fewer than 10 flows at any occupancy, as is the case with the extreme occupancies, all were retained to ensure as close to a uniform distribution as possible without drawing too small a sample.

The equations for each day were then reestimated on the basis of the sampled data (table 2). The most noticeable point is that the equation for what had been the outlying day has been changed and is no longer the farthest from the group average.

To test whether there are significant differences among the days in one condition, a dummy (i.e., 0, 1) variable was introduced in table 3, and an expanded equation analyzed:

\[ \ln(\text{flow}) = a + b_1 \times \ln(\text{occ}) + b_2 \times D_i + b_3 \times D_i \times \ln(\text{occ}) \]

where \( D_i \) is 1 for the extreme day (August 23 for rainy weather;
TABLE 1  REGRESSION ANALYSIS ON UNCONGESTED DATA SETS

<table>
<thead>
<tr>
<th>DAY</th>
<th>COEFFICIENTS (T-RATIO)</th>
<th>R^2</th>
<th># OF DATA POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAINY WEATHER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 AUG. 1986</td>
<td>5.411 (204.6)</td>
<td>0.7768</td>
<td>0.942</td>
</tr>
<tr>
<td>6 AUG. 1986</td>
<td>5.342 (217.5)</td>
<td>0.8064</td>
<td>0.928</td>
</tr>
<tr>
<td>7 AUG. 1986</td>
<td>5.325 (317.3)</td>
<td>0.8124</td>
<td>0.964</td>
</tr>
<tr>
<td>23 AUG. 1986</td>
<td>5.341 (445.6)</td>
<td>0.7969</td>
<td>0.972</td>
</tr>
<tr>
<td>CLEAR WEATHER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 AUG. 1986</td>
<td>5.386 (464.1)</td>
<td>0.8000</td>
<td>0.978</td>
</tr>
<tr>
<td>20 AUG. 1986</td>
<td>5.349 (276.5)</td>
<td>0.8308</td>
<td>0.954</td>
</tr>
<tr>
<td>22 AUG. 1986</td>
<td>5.364 (346.5)</td>
<td>0.8271</td>
<td>0.971</td>
</tr>
</tbody>
</table>

FIGURE 5  Uncongested flow-occupancy data, for all four rainy days, showing areas of heavier concentration of data.

August 22 for good weather), and 0 otherwise. In such an equation, the coefficients for the condition represented by the dummy variable can be found by summing the relevant coefficients. For example, in the equation for the rainy days, the pooled estimate for the three days is

\[ \ln(\text{flow}) = 5.37 + 0.789 \times \ln(\text{occ}) \]

and the estimate for the extreme day, August 23 can be found, setting the dummy variable equal to 1, as

\[ \ln(\text{flow}) = (5.372 - 0.049) + (0.7891 + 0.0198) \times \ln(\text{occ}) \]

which is exactly the equation for August 23 given in table 2.

If the extreme day requires a separate equation, either \( b_2 \) or \( b_3 \) or both will be significant. Because these coefficients have been estimated on the basis of the full sample size for that day (207 points in the case of August 23), the sample is large enough that the Student's \( t \) distribution can be approximated by the normal distribution. Hence the critical value for a 5% confidence level is 1.96. The results (table 3) show that neither of the extreme days requires an equation that is significantly different from the mean of the others in the group.

The final step was to run a regression using three dummy variables, one for each of the extreme days as just identified and one for the weather condition itself. This approach would
show clearly whether the difference between weather conditions is more important than the difference within one condition. The \( t \) statistics for the resulting equation show that only one of the coefficients involving a dummy variable is significant—the coefficient on the logarithm of occupancy for rainy days. Thus this segment of the analysis supports the hypothesis that different functions are needed for the two weather conditions and that there is no significant variation within one weather condition. Because separate equations are needed, they were estimated individually, without any dummy variables for separate days. The resulting functions and data are shown in figure 7. The equations found were, for clear weather days,

\[
\ln(\text{flow}) = 5.385 + 0.81371 \times \ln(\text{occ})
\]

and for rainy weather days,

\[
\ln(\text{flow}) = 5.352 + 0.7969 \times \ln(\text{occ})
\]

The possibility remains that the functions are different not because of an overall shift in the data but because of a shift in only one part of the range. There appears to be some visual support for this for the rainy-days data in figure 7, in that for occupancies above 20% the data may not be evenly distributed about the line but instead lie predominantly below it. Data for the clear days, however, seem uniformly distributed about the line.
TABLE 3 REGRESSION ANALYSIS ON SAMPLED DATA SET USING DUMMY VARIABLES

<table>
<thead>
<tr>
<th>RAINY WEATHEREXTREME DAY 23 AUG. 1986</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1 = 1 extreme day</td>
</tr>
<tr>
<td>= 0 remaining rainy weather days</td>
</tr>
<tr>
<td>ln(flow) = 5.372 + 0.789<em>ln(occ) - 0.049</em>D1 + 0.0198<em>D1</em>ln(occ) (t-statistic) (279.17) (102.36) (-1.63) (1.56)</td>
</tr>
<tr>
<td>CLEAR WEATHER EXTREME DAY 22 AUG. 1986</td>
</tr>
<tr>
<td>D2 = 1 extreme day</td>
</tr>
<tr>
<td>= 0 remaining clear weather days</td>
</tr>
<tr>
<td>ln(flow) = 5.37 + 0.816<em>ln(occ) + 0.0395</em>D2 - 0.0076<em>D2</em>ln(occ) (t-statistic) (148.98) (1.74) (-0.80)</td>
</tr>
<tr>
<td>COMBINED RAINY AND CLEAR WEATHER DATA</td>
</tr>
<tr>
<td>D1 = 1 extreme rainy weather day</td>
</tr>
<tr>
<td>D2 = 1 extreme clear weather day</td>
</tr>
<tr>
<td>D3 = 1 all rainy weather days</td>
</tr>
<tr>
<td>= 0 all clear weather days</td>
</tr>
<tr>
<td>ln(flow) = 5.37 + 0.789<em>ln(occ) - 0.049</em>D1 - 0.0001<em>D3 + 0.0395</em>D2 + 0.0198<em>D1</em>ln(occ) (t-statistic) (315.13) (115.54) (-1.84) (-0.00) (1.40)</td>
</tr>
<tr>
<td>+ 0.0212<em>D3</em>ln(occ) + 0.0212<em>D3</em>ln(occ) - 0.0076<em>D2</em>ln(occ) (1.76) (2.82) (-0.64)</td>
</tr>
</tbody>
</table>

FIGURE 7 Overlaid sampled flow-occupancy data for both clear and rainy conditions, with regression lines.

To test whether the differences occur only at high flows, additional dummy variable regressions were run (using only the one dummy variable related to weather). These were done sequentially, eliminating all data for occupancies above a given value, starting with 27% and decreasing the threshold 1% each time, until there remained no significant difference due to weather. Only for occupancies below 8% did the difference between the two equations disappear entirely. Because that result meant that the difference in function is found over two-thirds of the data range, we deemed it more appropriate to use two separate equations for the full range rather than attempt to differentiate them only above 8%.

It is clear from the comparison of figure 2 with figures 3 and 4 that the third possibility, that of a U-shaped function for rainy days, is not supported. Consequently, it appears from the analysis of the rainy-days data that there is a clear shift downward during adverse weather conditions, in the function representing uncongested flow-occupancy data.

To test this result when carried to an extreme, we also investigated the one November day (the 20th) with severe weather conditions (snow, poor visibility, and slippery roads), attempting to analyze it the same way. The most difficult task proved to be separating the congested and uncongested data. Speed drops of the magnitude used for the earlier data were not found here, undoubtedly because uncongested speeds were so much lower. The difficulties can be seen in the flow-occupancy plot of these data (figure 8). Consider the left-hand side data—i.e., the cluster around the diagonal from (0, 0) to (20, 2000). It seems likely that some of the data on the lower part of this cluster represent operations in the early stages of congestion. Including speed as a decision variable did not make the distinction any more obvious. After trying a number of decision criteria, we decided to use the simple one that speeds greater than or equal to 70 km/h represented uncongested flow, simply for comparison purposes. The best fit line to those data is

\[ \ln(\text{flow}) = 4.9868 + 0.8883 \ln(\text{occ}) \]

with an \( R^2 \) of 0.946, based on 367 observations. Because of the arbitrariness of this decision criterion, the difference between this equation and those for clear or rainy days was not tested statistically. However, a visual inspection (fig. 9) shows that the estimated flow-occupancy line for the snowy conditions falls below the other two, as might be expected. Note also that to the extent that the 70 km/h criterion has excluded points which in fact represent uncongested operations, the line for snowy conditions is closer to the other two than it really should be.
The immediate conclusions from this analysis seem clear: adverse weather affects the flow-occupancy function by reducing the slope of the curve that describes uncongested operation; as a result, maximum flows are also reduced during such conditions. Neither of these results comes as any great surprise. Indeed, Athol (4) suggested this kind of picture more than 20 years ago. Nevertheless, these results have helped to clarify the nature of the changes.

The implications of these results for the incident detection logic described by Persaud and Hall (3) appear to require a more complicated logic than might have been needed otherwise. Because bad weather will cause a downward shift in the flow-occupancy line, the incident detection/trigger cannot rely on these two variables alone but must simultaneously include speed. Yet as conditions worsen (e.g., the November 20 snowstorm), speeds naturally decline until it is difficult to identify speed drops.

One possibility is that the proposed incident detection logic will be of most benefit for climates that are more temperate than Ontario’s, or for fair weather only. An alternate possibility is that an adaptive logic could be developed that builds the “expected” picture of operations anew every few hours or on demand. These, however, are the consequences of the results. The conclusion itself seems clear enough: bad weather affects the flow-occupancy relationship by pushing down the tip of the inverted V and making it flatter.

REFERENCES