Abstract
The proliferation of fixed freeway sensor data has opened the door to more detailed analysis of freeway operations and capacity. The availability of more data (usually from fixed points on the network) also brings data management and processing challenges. Past research has often used arbitrary temporal aggregation as a means of smoothing data measured at single points. Comparisons are difficult to draw across segments and details of interest are often filtered out. This paper describes how the spatial and temporal evolution of traffic conditions on a freeway can be diagnosed using sensor data retained in their most raw form. The diagnostic tools used in this paper include curves of cumulative vehicle arrival number versus time and cumulative occupancy versus time constructed from data measured by neighboring freeway loop detectors. Once suitably transformed, these cumulative curves provide the measurement resolution necessary to observe the transitions from freely-flowing to queued conditions and to identify some notable, time-dependent traffic features surrounding freeway bottlenecks. Given that little is known about how traffic flows through bottlenecks, a greater understanding is required to formulate, to enhance or to verify mathematical models of vehicular traffic, so that they are consistent with the actual traffic features that are found to be reproducible. This understanding is also required before one can conclude whether or not bottleneck flows can be increased by eliminating or postponing freeway queues with control measures such as ramp metering.

Keywords bottleneck, freeway, queueing, traffic dynamics.

Introduction
Bottlenecks are critical components of freeway systems and understanding their behavior is important for improving system operations. Figure 1 shows that an “active” bottleneck arises when vehicles discharge from an upstream queue (guaranteeing that the bottleneck serves vehicles at a maximum rate) and vehicles are unimpeded by traffic conditions emanating from further downstream (Daganzo 1997). Bottlenecks can be difficult to detect since they are temporally and spatially variable. Detection can be further hampered by large spacing of freeway surveillance detectors (see Figure 1).

1 Associate Professor, Department of Civil and Environmental Engineering, Portland State University, P.O. Box 751, Portland, Oregon 97207-0751 USA
Inductive loop detectors are very common types of sensors that can automatically record traffic flow parameters. Single loop configurations can count vehicles, measure occupancy and with a simple assumption about average vehicle length, estimate the time mean speed of passing vehicles. The use of a pair of detectors spaced at a known distance can facilitate vehicle counting as well as direct measurement of occupancy, vehicle length, and time mean speed. This is true at the level of individual vehicles; however most inductive loop detectors are connected to traffic controllers that are programmed to aggregate traffic parameter measurements over longer periods such as 20 or 30 sec or one minute. These sensor systems rely on communications systems to transfer the traffic measurements to a central traffic management center. If tuned and maintained well, and with robust power and communications systems, inductive loop detectors can provide valuable information about the traffic state at a single point and over freeway segments. Detectors should be placed in individual freeway lanes at a reasonable longitudinal spacing on the order of 500 meters. They should also be installed on on-ramps and off-ramps, so that they can provide a reasonable picture of how freeway traffic is operating.

With the proliferation of fixed freeway sensor data on freeways around the world, more detailed analysis of freeway operations and capacity has become possible. Despite the availability of more data (usually from fixed points on the network), there are also significant challenges related to data management and processing. The objective of this paper is to describe how the spatial and temporal evolution of traffic conditions can be diagnosed by careful processing of inductive loop detector (or other sensor) data that has been retained in its most raw form. Using data from a freeway in Germany, several diagnostic tools will be explained, including the use of curves of cumulative vehicle arrival number versus time and cumulative occupancy versus time constructed from data measured by neighboring freeway loop detectors. Once suitably transformed, these cumulative curves provide the measurement resolution necessary to observe the transitions from freely-flowing to queued conditions and to identify some notable, time-dependent traffic features surrounding freeway bottlenecks. Given that little is known about how traffic flows through bottlenecks, a greater understanding is required to formulate, to enhance or to verify mathematical models of vehicular traffic, so that they are consistent with the actual traffic features that are found to be reproducible. This understanding is also required before one can conclude whether or not bottleneck flows can be increased by eliminating or postponing freeway queues with control measures such as ramp metering.
Data

The observations that follow were taken from the 14 km segment of northbound Autobahn 9 (A9) near Munich Germany, illustrated in Figure 2 (the traffic flow direction is toward the left). Data were used from June 27, 2002. Dual inductive loop detectors recorded vehicle count, occupancy (the percent time a detector was covered by a vehicle) and time mean speed in each lane over 1-minute intervals. The data were available separately for trucks for autos. This freeway has no ramp metering, clear weather prevailed and a unique variable speed limit system was in operation on the facility. The next section describes several basic ways that the detector data can be analyzed, along with some possible pitfalls. The next section contains the bottleneck’s definitive diagnosis, including its location, time of activation and deactivation and some discharge features that were uncovered (for a single day).

Detector Data Analysis Tools

Detector data are often available from fixed points on freeways, locations that were not chosen in relation to actual traffic conditions or for research purposes. Early attempts at processing loop detector data used arbitrary temporal aggregation as a means of smoothing data measured at single points. It is well understood that traffic data exhibit both statistical fluctuations and time dependent trends. The challenge is how to differentiate between the two. Figures 3 and 4 show two examples of how data can be plotted from single points along a freeway. Figure 3 shows a bivariate plot of one-minute data from the A9, with flow (in vehicles per hour) on the x-axis and velocity (km/hr) on the y-axis. As shown there is a wide scatter of data along the uncongested (upper) branch, with only a few points on the congested (lower) branch. The important point here is that each of these one-minute points does not represent stationary traffic conditions, and in fact many of these points may include vehicles traveling under different prevailing conditions (e.g., stop and go). Since the one-minute time intervals are arbitrary, it is also possible that points shown here were measured during a transition from uncongested to congested conditions (or vice versa). For this reason, such a plot can be misleading. In addition, the plot reveals nothing about what is occurring over a relevant freeway segment, nor does it indicate the temporal relationships between traffic states.
Figure 3: Bivariate Plot

Figure 4 shows a sample of time series count data from the A9—including the one-minute data (large fluctuations visible), 5-minute and 15-minute aggregations. Similar to Figure 3, the spatial aspects of traffic conditions are not visible and it is difficult to discern time dependent flow changes from the statistical fluctuations.

Figure 4: Time Series Plot

As an illustration of how both temporal and spatial dimensions can be revealed, Figure 5 shows a sample of speed data plotted on a time-space plane for one afternoon. As shown, some features are visible, and the peak period can be identified between
14:00 and about 19:00. However, comparisons are difficult to draw across segments and details of interest are often filtered out using this technique. What is needed is a robust mechanism for taking advantage of nearly ubiquitous sensor data to inform a theoretical underpinning describing traffic dynamics. In addition, any improved method should process the sensor data without losing resolution and reveal parametric changes over time. An improved method should also be used for count (flow), speed and other parameters.

**High Resolution Traffic Analysis Method**

Previous research (Cassidy and Windover 1995, Cassidy and Bertini 1999, Munoz and Daganzo 2002) have described a method for visually comparing transformed curves of cumulative vehicle arrival number versus time and cumulative occupancy versus time measured at neighboring (loop detector) locations. This treatment of the data can reveal how some time-dependent traffic features propagated over time and space. With these graphical presentations, it is possible to identify the detectors that were located downstream of active bottlenecks and to study in detail the flows measured by these detectors.

![Figure 6: Cumulative Count Curve](image)

![Figure 7: Oblique Cumulative Count Curve](image)

Figure 6 shows a cumulative count curve, $N(x,t)$, for two hours’ data (1 minute level of granularity) across all lanes from station 340 on the A9 on July 4, 2002. Using dimensional analysis, one can see that the slope of the $N(x,t)$ at any time is the flow passing station 340. The raw $N(x,t)$ appears to be essentially a straight line, revealing little about changing traffic conditions. This curve can be transformed without losing information by constructing an oblique cumulative count curve. One way to visualize this is to construct a straight line (Figure 6) of slope $q_0$ (in this example, $q_0$ is 5180 vehicles/hour). The value of $q_0$ was obtained by trial and error to enhance visual inspection. The next step is to plot just the difference $N(x,t)-q_0t'$, were $t'$ is the elapsed time from the beginning of the curve. Figure 7 shows the result, with a re-scaled y-axis on the right side. Now it is clear that changes in prevailing flow did occur during this period. By comparing slopes, it is clear that a sizable flow increase at about 14:35 is visible followed by a flow decrease at about 15:00. For reference the raw 1 minute flow (left hand y-axis) is also constructed as $Q(x,t)$, clearly indicating the improvement that the oblique plotting method provides. The oblique plotting technique can also be used to plot occupancy and speed data, which reveals times at which notable changes in the...
parameters occur.

Beyond the oblique plotting technique, it is also possible to take advantage of data measured at neighboring loop detector stations to gain an understanding of how traffic features propagate over space (Cassidy and Windover 1995, Newell 1982, Newell 1993). This can provide a vast improvement over plots of data measured at single points.

An unaltered cumulative plot of 1-minute count data from one point would actually appear like a stair-step function with equal step widths (note that a raw cumulative plot of individual vehicle actuations at the detector would be a stair-step function with equal step heights). Such a plot could be smoothed into an approximate, continuous, increasing function by connecting the tip of each stair step, and referred to as $N(x,t)$. The slope of $N(x,t)$ at any time is the prevailing flow at that time. Referring to Figure 8, it is clear by examining $N(x_1,t)$ that flow changes, increases and decreases, are visible as slope increases and decreases respectively. As shown in Figure 8, it is possible to plot two cumulative curves in series. Assuming that there is vehicle conservation between $x_1$ and $x_2$, a queueing diagram can be constructed by plotting vehicle counts from two successive measurement locations, with the curves separated by the free flow trip time between the two locations. In Figure 8 the $N(x_1,t)$ and $N(x_2,t)$ curves have been constructed. By definition the horizontal distance between the two curves is the vehicular trip time between $x_1$ and $x_2$ for any vehicle $j$, and the vertical distance is the number of vehicles present on the segment between the two measurement locations at some time $t_1$. Figure 9 shows that it is possible to magnify the presence of changing conditions on the segment between $x_1$ and $x_2$ by shifting the upstream curve horizontally to the right by the free flow trip time. Now Figure 9 reveals the excess travel time (delay) and the excess vehicle accumulation between $x_1$ and $x_2$ due to the presence of a queue.

**Diagnosing the Bottleneck in Detail**

To illustrate the use of the method described above and in order to pinpoint bottleneck activation and deactivation, A9 traffic features were analyzed using data from June 27, 2002. Figure 10 presents oblique curves of $N(x,t)$, cumulative vehicle arrival number, constructed from counts measured across all lanes at detectors 380, 390 and 420, over a 55-minute period during the afternoon. The $N$ were constructed by taking linear interpolations through the 1-minute counts so that a curve’s slope at time $t$...
would be the flow past location $x$ at that time. The counts for each curve in Figure 10 began ($N=0$) relative to the passage of a hypothetical reference vehicle (curve 420 includes counts from the München-N on-ramp) so all curves describe the same collection of vehicles. Any horizontal or vertical separations between curves would have been the trip times and vehicle accumulations between detectors, respectively (Newell 1982, Newell 1993).

Each curve in Figure 10 was shifted horizontally to the right by the average free-flow trip time between the respective detector and downstream detector 380. Resulting vertical displacements between curves are the excess vehicle accumulation between detectors due to vehicular delays. In order to magnify the curves’ features, an oblique coordinate system was used where $N(x,t) - q_0 t$ where $q_0$ was an oblique scaling rate and $t'$ was the elapsed time from the beginning of each curve. The same value of $q_0$ was used for all curves and therefore did not affect the vertical separations (Cassidy and Windover 1995).

Figure 10: Bottleneck Diagnosis

All three curves in Figure 10 were superimposed between 14:45 and 15:21, indicating the presence of freely flowing traffic throughout this freeway section. Beginning about 15:21, excess accumulation was visible between detectors 380 and 390 followed by flow reductions at detectors 380 and 390 at 15:21. The divergence of the detector 390 curve from the one at detector 380 (at 15:21) marked the arrival of a backward-moving queue at detector 390. A pronounced flow reduction at detector 390 accompanied this divergence. The insets in Figure 10 are oblique plots of speed measured at detectors 380, 390 and 420. The times at which marked speed reductions occurred (shown by the vertical arrows) verify that the queue passed station 390 at 15:212 and reached station 420 by 15:24. Further analysis of oblique speed curves revealed that the queue reached station 630 by 15:58 (this can be seen in Figure 5 also).

Next, Figure 11 shows oblique $N(x,t)$ for detectors 340-380. Constructed like the curves in Figure 10, along with insets showing the minimal speed changes, this figure reveals the presence of freely flowing traffic between all station pairs. Recalling the
bottleneck definition shown in Figure 1, it is clear that a bottleneck was activated between stations 380 and 390. This analysis would not have been possible with simply single detector plots using unaltered y-axes.

Figure 11: Bottleneck Diagnosis

This analysis has definitively illustrated the activation of a bottleneck between detectors 380 and 390 at 15:21 and that a queue propagated as far upstream as detector 630. Consistent with the bottleneck’s definition, freely flowing traffic remained downstream. Figure 12 shows the entire picture, using similar analysis, indicating the presence of four active bottlenecks on this day.

Figure 12: June 27, 2002 Bottlenecks

Now it is possible to focus on the period just before and just after 15:21,
surrounding the activation of the first bottleneck. Any predetermined data aggregation (e.g. 5- or 15-minute) would have masked the time at which the bottleneck was activated, leading to erroneous estimates of prevailing flows. Figure 13 shows the oblique $N(x,t)$ for detectors 380, 390 and 420 for a longer period. As shown, the presence of an upstream queue is visible due to the vertical displacement of the curves. Also shown on Figure 13 are the prevailing flows measured at station 380 just before bottleneck activation (5510 vph) and the bottleneck outflow measured while the bottleneck was active (5370 vph). Over 11 activations, the bottleneck at this location exhibited an average outflow of 5370 vph, which is viewed by some as the capacity. This bottleneck exhibited a mean difference between the pre-queue flow and the outflow of 5%. The phenomenon known as the two capacity theory, has been verified at several sites around the world (e.g., Cassidy and Bertini 1999, Bertini et al. 2005).

**Conclusions**

It is shown that it is possible to definitively diagnose a freeway bottleneck’s location, activation and deactivation times and other features using loop detector data. It is hoped that this paper will encourage further analysis of traffic phenomena around the world and that researchers will consider sharing their raw data in order to develop consensus (or at least open debate) about traffic dynamics. Careful and systematic analysis of detector data from many countries, across facilities with a wide variety of geometries can only heighten our understanding of traffic flow phenomena. In addition, it would be desirable to develop a clearinghouse of well-documented data sets from freeways, including such geometric features as merges, diverges, sags, crests, tunnel entrances, lane drops, and others. Some have suggested that a primary focus of traffic research should be to examine time-dependent flows, to determine velocities of propagations of disturbances, and to determine how traffic reacts to time- or space-dependent influences such as bottlenecks. It has been expected that modern data collection and analyses techniques would refine the traffic theories of the early 1960’s, yet it has been reported that little has been added to the understanding of dense traffic
flow (Newell 1995). It should be noted that any system for pinpointing recurrent freeway bottlenecks must be capable of distinguishing them from queues caused by incidents (Lin and Daganzo 1997). For example, an incident resulting in a blocked lane would reveal a much larger flow reduction (on the order of 25% for a three-lane section) upon queue formation than would a recurrent bottleneck (on the order of 10%). This is the subject of ongoing research.

Acknowledgment

Much of this research was funded by the BMW Group. The author would like to thank Dr. Klaus Bogenberger, BMW Group for his support. Steven Hansen and Steven Boice were graduate research assistants who were invaluable to the success of this project. Finally, Christian Mayr and Frank Frischeisen of Autobahndirektion Südbayern generously provided the data used herein.

References


