On-Line Diversion Prediction for Dynamic Control and Vehicle Guidance in Freeway Corridors

Yorgos J. Stephaneides, Eil Kwon, and Panos Michalopoulos

The effectiveness of route guidance and advanced control strategies in a corridor is a function of the time-dependent interaction between freeway and arterial flows in response to traffic conditions and information available to drivers. Although demand diversion is an essential element of optimal corridor management, no on-line demand predictor has considered explicitly the effect of control on the demand to be predicted. As a result, existing control strategies are mostly empirical and can address only known or anticipated peak-period loads. A new approach to modeling on-line demand diversion in freeway corridors is directly applicable to dynamic control and vehicle guidance. This new method treats freeway demand as a utility-maximizing, decision-making unit with ramp diversion behavior dependent primarily on the travel delay caused by congestion, as modified by on-line information and traffic control strategies. The method combines behavioral modeling with an extended Kalman filter to identify the diversion model parameters recursively using the most recent prediction error in real time. Because diversion prediction does not require upstream flow information, system instrumentation requirements are substantially avoided. The method was used for on-line prediction of demand and diversion at freeway entrance ramps of the I-35W corridor in the Minneapolis–St. Paul metropolitan area. Test results from 5-min prediction on several weekdays indicate the entrance ramp diversion proportion is predicted with 90 to 95 percent accuracy.

The ever-increasing demand for individual mobility, coupled with economic and environmental restrictions on increases in physical road capacity, are making freeway congestion a top-priority issue in most urban areas. In the United States, freeway congestion already causes 1.2 billion vehicle-hr of delay, wastes 1.3 billion gal of fuel, and increases user cost $9 billion per year (1). With no drastic solution in sight, the most cost-effective alternative is to maximize the efficiency with which the existing capacity of the entire freeway corridor (freeway and arterials) is used. This result can be accomplished by optimizing management and control of traffic flow in the corridor.

Optimal corridor control is based on the diversion of excessive demand from a congested link of the corridor to a less-congested alternative in space or time. For instance, peak freeway demand could be diverted to corridor arterials or drivers could be induced to depart earlier or later than usual. Such diversion is possible through the use of advanced real-time control strategies, including on-line coordination of ramp meters with arterial signals and, more recently, advanced driver information systems.

Although demand diversion is an essential element of optimal corridor management, no dependable diversion models exist that can describe the time-dependent behavior of flow responding to control strategies and congestion patterns (2–4). Existing corridor control methods adopt oversimplified assumptions such as driver omniscience, user-optimized equilibrium flow, and uniform reaction of drivers to the control (5,6). Further, no existing on-line demand predictor has explicitly considered the effect of control on the demand to be predicted. As a result, existing control strategies are mostly empirical, designed to address known or anticipated peak-period loads.

For effective control, simulation, and traffic management in freeway corridors, prediction of the time-dependent interaction between freeway and arterial flows in response to control and management actions is of critical importance. A new approach to modeling the on-line demand diversion in freeway corridors, which is based on results from recent research and experimentation (7), treats freeway demand as a utility-maximizing, decision-making unit. The ramp diversion behavior of this unit depends primarily on the travel delay caused by congestion as modified by on-line information and traffic control strategies. This method combines behavioral modeling with an extended Kalman filter (8) to identify recursively the diversion model parameters. By updating the prediction error at every time slice, the filter makes this method appropriate for application in real time. Unlike some earlier approaches (9), diversion prediction does not require upstream flow information and thus substantially limits instrumentation requirements.

This method was used with encouraging results to predict on-line demand diversion at freeway entrance ramps (equivalent to Location A of the typical ramp shown in Figure 1) of the I-35W corridor in the Minneapolis–St. Paul metropolitan area. This approach is currently being extended to predict diversion at corridor intersections.

BACKGROUND

Most freeway corridor control strategies aim to improve freeway flow by diverting drivers away from congested freeways to other routes within the corridor. The critical question is whether the benefit from reducing the mainline congestion offsets the problem created in the rest of the network. For example, entrance ramp metering, the most widely used type of control, seeks to reduce and divert freeway-entering demand.
by transferring the delay from the freeway to the entrance ramp. However, inappropriately low metering rates may cause severe congestion at the adjacent intersection by creating extensive spillback, thereby increasing the total delay in the corridor. Further, overestimation of freeway ramp demand could cause unnecessarily high metering rates, actually encouraging freeway use rather than inducing diversion.

Although optimal corridor control is intended to solve these problems, no control strategy yet developed is comprehensive enough to reflect drivers' responses to the control. The most advanced freeway corridor control methods are based on determining the control that results in the most desirable traffic assignment in a network. These methods assign traffic by determining the equilibrium flow pattern that will satisfy Wardrop's principle within a given time slice (10), or they apply a self-assignment procedure that assumes omniscient drivers can find the quickest route at each decision point of their trips (11,12). However, the case has been made that Wardrop's principle is not applicable to the dynamically changing traffic environment, mainly because drivers are not sufficiently well informed or skilled to choose the best route (5).

Most on-line demand predictors adopt a statistical trend-tracking approach. Prediction is based on the past trend of volume measurements and does not consider the effect of traffic control on driver decisions. For example, the original demand predictors in the second and third urban traffic control system generations (13) predict the turning volumes at intersections without considering the effects of signal timing on the turning movement. The prediction is based only on historical information and the recent prediction error. Although time series models have been developed for certain applications, such as short-term traffic demand forecasting (14,15) and freeway occupancy estimation (16,17), these models do not explain the interrelationship between the control and the resulting flow patterns. More important, the vast majority of models use constant parameters computed off-line, which significantly restricts the models' adaptability to the ever-changing traffic environment. Although Okutani and Stephanedes (9) tried to update model parameters on the basis of the most recent traffic demand prediction error by applying Kalman filtering, this approach, like most others, does not explicitly treat the effects of control policies on the predicted volume.

MODEL FORMULATION

Two major hypotheses were made to model the diversion of freeway demand at entrance ramps. The first hypothesis draws on the conclusions of recent experimental work that was based on disaggregate analysis of the diversion behavior of 1,100 drivers (7). These conclusions indicate that perceived trip time, including waiting time on the ramp, is the dominant factor determining diversion at entrance ramps. Socioeconomic factors, such as income, do not appear to be relevant. The findings suggest a utility-based model structure treating traffic demand as a homogeneous flow with diversion behavior depending primarily on the perceived delay caused by congestion. The first hypothesis underlying the new model, then, is that freeway demand can be treated as a decision-making entity that responds to ramp control and the resulting on-ramp traffic condition according to the principle of utility maximization. For an individual commuter, the utility of diversion is a function of the delay caused by the traffic condition at the ramp; drivers trade off the ramp traffic condition for the expected delay.

The second hypothesis reflects findings from extensive surveys in the Minneapolis–St. Paul metropolitan area indicating that, for every 10- to 15-min peak interval, an overwhelming portion of traffic consists of the same commuters following the same route every weekday. On the basis of this observation, the second hypothesis is that for a short (5-min) time interval, the ramp condition does not affect the total freeway demand approaching the ramp.

These hypotheses suggested two models. The survey findings indicate that drivers perceive three variables as the major causes influencing waiting time on the ramp: the freeway-entering rate, the number of cars on the ramp, and the ramp-entering rate. The first model predicts the ramp-entering proportion \( P_k \); that is, the ratio of the cars entering the ramp to the cars arriving at the ramp entrance at every time interval \( k \). This model states that the ramp-entering proportion is a function of the current freeway-entering rate \( C \) and the number of cars \( X + R \) on the ramp in the previous time interval. This formulation is as follows:

\[
P_k = \frac{1}{1 + \exp(\theta_{1k} + \theta_{2k}C_k + \theta_{3k}(X_{k-1} + R_{k-1}))}
\]

where \( X_k \) is the number of cars on the ramp in the beginning of the \( k \)th time interval, and \( R_k \) is the ramp-entering rate. On
Stephanedes et al.

where the \( \theta_{ae} \) are parameters to be updated in real time and \( k \) represents a 5-min interval.

The second model uses historical information to predict the number of cars arriving at the ramp entrance \( (Q_k) \) at every time interval \( k \):

\[
Q_k = \theta_{1k} E(Q_k) + \theta_{2k} E(Q_{k-1}) + \theta_{3k} Q_{k-1}
\]

where \( E(Q_k) \) is the past average rate during the \( k \)th time interval.

Data on all model variables, such as the number of cars approaching, entering, and leaving the ramp, can be measured by a traffic engineer using modern equipment. Thus the proposed models could become operational. For instance, one video-based detector, appropriately located over a ramp, could measure \( P, C, X, R, \) and \( Q \) in real time even when spillback on the entrance ramp causes the traffic queue to extend beyond the ramp storage area. This measurability is an advantage over recent specifications (7) based only on variables reported by drivers, such as perceived trip time.

However, the two new models cannot be applied in real-time control in this form, because they contain an error. This error is not bounded and, if left unchecked, could make the diversion prediction meaningless. To address this issue, an on-line predictor is developed by incorporating the extended Kalman algorithm into the models, thus allowing the model parameters to be adjusted dynamically. As part of the prediction formulation, the values of model parameters \( \theta_{ae} \) are assumed to represent the behavioral state of the demand at the \( k \)th time interval following the nonstationary random walk process. The random walk process provides more flexibility than ARIMA-type models and has been successfully applied to model physical systems that are subject to rapid variations (18). The model parameters are continuously updated by the extended Kalman filter using the prediction models as the observation equations. The resulting filter formulation and the detailed procedure for the real-time prediction of the demand diversion at entrance ramps are summarized as follows (see also Figure 2).

- **Filter Formulation**

  \[
  F-I: \quad \theta_{1k+1} = \theta_{1k} + w_{1k}
  \]

  \[
  F-II: \quad \{ P_k = \frac{1}{\omega + 1} E(Q_k) + \theta_{3k} E(Q_{k-1}) + \theta_{3k} Q_{k-1} + v_{3k} \}
  \]

where \( w_{1k}, v_{3k} \) are state and observation noise vectors, respectively, assumed to be white noise.

- **On-Line Prediction Using the Discrete-Time Extended Kalman Filter**

  1. Initialize \( (k = 0) \): \( \theta_k = \theta_0 \), where any prior knowledge can be used to initialize \( \theta \).
  2. Let \( k = k + 1 \).
  3. Predict the ramp-entering proportion \( (P_k) \) and the number of cars \( (Q_k) \) arriving at the ramp entrance, using the filter formulation and the model parameters \( \theta_k \).
  4. Measure the actual \( P_k \) and \( Q_k \), and obtain the prediction error \( (e_k) \), the difference between actual and predicted values.
  5. Update \( \theta \) on the basis of the most recent prediction error \( e_k \): \( \theta_{k+1} = \theta_k + K_k e_k \), where \( K_k \) the Kalman gain, is

![FIGURE 2 Framework of proposed on-line prediction.](image-url)
computed on-line from the covariance of the observation noise (v), the covariance of the model parameters (θ), and the sensitivity of the prediction to the model parameters (f).

6. Go to Step 2.

In particular, these steps are implemented in the computer using the following index incrementing notation:

1. Initialize (k = 0): θ_{k+1} = θ_0 and G_{k+1} = G_0, where G is the covariance matrix of θ.
2. Predict P_{k+1} and Q_{k+1} using updated θ_{k+1/k}, where
θ_{k+1/k} = θ_{k+1/k}.
3. Measure actual P_{k+1}, Q_{k+1} and obtain error η_{k+1}.
4. Update θ_{k+1/k} on the basis of the η_{k+1}:
θ_{k+1/k+1} = θ_{k+1/k} + K_{k+1}η_{k+1},
G_{k+1/k+1} = (I - K_{k+1}G_{k+1})G_{k+1/k},
where
K_{k+1} = G_{k+1/k}(S_{k+1} + G_{k+1/k}(S_{k+1} + R_{k+1})^{-1},
prime denoting transposed matrix
G_{k+1/k} = G_{k+1} + q_k
S_{k+1} = \frac{∂θ_{k+1}/∂θ_{k+1}}{∂θ_{k+1}/∂θ_{k+1}} θ_{k+1} = θ_{k+1/k}
G_k = E[(θ_k - θ_{k/k})(θ_k - θ_{k/k})']
q_k = E[(w_k)(w_k')'] = q_kδ_{ij}, covariance of state noise vector
R_k = E[(η_k)(η_k')'] = R_kδ_{ij}, covariance of observation noise vector
δ_{ij} = 1 if k = j, 0 otherwise, the delta function
h(θ) = \{P_k, Q_k\}, the filter equations
I = identity matrix
S = sensitivity matrix.

5. Let k = k + 1 and go to Step 2.

Implicit in specification F–I is the assumption that the model parameters follow a normal distribution. A more general specification would assume θ_{k+1} = θ_0 + w_{k+1}, where G represents the fixed characteristics of the process; ongoing research seeks to identify G. For the data in this analysis, the three parameters θ_{a0}, θ_{a}, and θ_{a} in F–II are almost identical. In this survey, this finding was not surprising, for two reasons. First, the variation of drivers’ departure times was less than 10 min. In addition, for every 10- to 15-min interval during the rush hour, the overwhelming portion of traffic consisted of the same commuters. Therefore, Q_k is the weighted average of the past and current volume measurements, and F–II is simplified by assuming the three parameters are identical, representing the aggregated effect of departure time variations.

The major feature of the proposed framework is that it can be used for real-time traffic management and control. It combines the behavioral modeling approach with a filtering technique so that the causal relationship between the control and diversion flow can be captured and updated through time. Use of the most recent prediction error improves the prediction at every time interval.

DATA

The data for this study were collected in the I-35W freeway corridor in Minneapolis, Minnesota, a heavily traveled and often congested corridor. (Figure 3 shows the corridor crossing the Twin Cities in a north-south direction.) First, an extensive questionnaire survey was conducted of 1,100 drivers approaching the entrance ramps in the sample corridor during the morning rush hour for 1 month. The information collected from each driver includes trip origin and destination, detailed description and evaluation of alternate routes, maximum ramp waiting time the driver could tolerate before diverting to an arterial, and socioeconomic and other data related to the commuting trip.

Traffic movements at the 35th Street ramp, a typical entrance ramp in the corridor (see Figure 1 for ramp configuration), were measured during the morning rush hour for a 1-month period in October 1988. Five-minute data from these measurements are shown in Figure 4 for a typical weekday in October 1988. They include

• Ramp-approaching rate (Q_a), that is, the number of cars arriving at the entrance of the freeway ramp during a 5-min interval.
• Ramp-entering rate (R_a), that is, the number of cars entering the ramp in 5 min.
• Diverting rate (D_a), that is, the number of cars diverting from the ramp entrance to the adjacent arterial in 5 min.
• Freeway-entering rate (C_a), that is, the number of cars entering the freeway from the ramp in 5 min. This rate is assumed not to exceed the ramp-metering rate.
• The number of cars on the ramp at the beginning of the kth time interval (X_k).

![Figure 3 I-35W Freeway Corridor, Minneapolis](image-url)
The survey data were used to analyze the behavioral characteristics of drivers diverting at entrance ramps. Survey details and resulting static disaggregate models of the ramp diversion have been reported by Stephanedes et al. (7). Cross-correlation analysis of the measured ramp flow data indicates that the current freeway-entering rate ($C_k$) and the total cars on the ramp during the previous time interval ($X_{k-1} + R_{k-1}$) have highly significant nonzero correlations with the current ramp-entering proportion ($P_k$). Although cross-correlation analysis alone cannot provide a functional relationship between the relevant variables, these correlations suggest that freeway demand flow can be treated as an entity sensitive to the control and the on-ramp traffic condition.

**MODEL TEST AND VALIDATION**

The proposed model and on-line prediction procedure were first tested with real data collected from the 35th Street ramp in the sample corridor. To avoid the large initialization error, the initial values of the model parameters ($O$s) were determined by fitting the proposed model with the past observations. This procedure produced initial parameter values of $b_1 = -5.0$, $b_2 = -0.05$, $b_3 = 0.05$, and $b_4 = 0.33$.

These values were used to test the proposed models and prediction method with the real data collected in October 1988. The resulting predictions for a typical weekday, along with the observed values, are shown in Figure 5. The average error rate for the diversion prediction (the ramp-entering proportion) ranges from 5.4 to 8.8 percent. The average prediction error of demand approaching the ramp has higher values, ranging from 6.1 to 13.4 percent. The high end of the latter error may reflect the fact that no upstream information is considered to predict the approaching demand. This error could be reduced by considering the volumes in the upstream links related with the ramp under analysis, along the lines developed by Okutani and Stephanedes (9).

Although improvement is still needed, the test results indicate the effectiveness of the proposed prediction method. As Figure 5 indicates, the ramp-entering proportion prediction has acceptable accuracy, and more important, the initialization error does not propagate in time, demonstrating the adaptive nature of the proposed method.

In order to validate the proposed models, they were applied to the Diamond Lake Street ramp in the sample corridor (see Figure 1 showing the ramp configuration). The measured traffic movements for a typical weekday are shown in Figure 6. Using the same initial values for the model parameters as those for the 35th Street ramp, the ramp-entering proportion and the total approaching demand were predicted following the same process. The resulting prediction for a typical weekday in March 1989—5 months after the first test—is shown in Figure 7. The average error rate of the ramp-entering proportion prediction ranges from 5.0 to 5.6 percent, whereas the ramp-
RAMP ENTERING PROPORTION (P_x) PREDICTION

RAMP APPROACHING RATE (Q_x) PREDICTION

CONCLUSION

The effectiveness of route guidance and advanced control strategies in a corridor is a function of the time-dependent interaction (diversion) between the freeway and arterial flows in response to the information available to the drivers and actual traffic conditions such as incidents and control. Although demand diversion is an essential element of optimal corridor management, no on-line demand predictor now in use explicitly considers the effect of control on the demand to be predicted. As a result, existing control strategies are mostly empirical and can address only known or anticipated peak-period loads. A new approach to modeling on-line demand diversion in freeway corridors is directly applicable to dynamic management and control.

The new method treats freeway demand as a utility-maximizing, decision-making unit. The ramp diversion behavior of this unit depends primarily on the travel delay caused by congestion and modified by on-line traffic control strategies. Two models are formulated. The first model predicts the ramp-entering proportion, and the second, the number of cars arriving at the ramp entrance. In particular, the models state that the ramp-entering proportion is a function of the current freeway-entering rate and the number of cars on the ramp in the previous time interval. Prediction of the number of cars arriving at the ramp is done using historical information. All data can be collected from on-line measurements, although the use of modern video-based detecting equipment would make data collection easier.

The method combines the behavioral modeling approach with an extended Kalman filter. By relinearizing the system dynamics about each new state estimate, the Kalman filter does not allow large initial estimation errors to propagate through time. Further, the Kalman filter recursively identifies the model parameters using the most recent prediction error in real time. Because the prediction improves at every time interval, these models can be used for prediction in conjunction with on-line control and vehicle guidance.

Diversion prediction does not require upstream flow information, thus substantially limiting system instrumentation requirements. Although the accuracy of the prediction of the ramp-approaching rate could be improved if upstream flow information from intersections adjacent to the freeway ramp were available, any benefit from such an improvement should be weighed against the additional instrumentation cost required. Similarly, the ramp-entering proportion model could be improved by explicitly incorporating the characteristics of the alternate arterials in the specification. This information is included in the parameters estimated, but explicit modeling would ease the task of the on-line predictor and would decrease the prediction error.

The model formulation is based on the findings from a survey of 1,100 commuters in the I-35W corridor in the
FIGURE 7  Demand-diversion prediction at Diamond Lake Street ramp, March 22, 1989.
Minneapolis–St. Paul Metropolitan Area. The method was used to predict the on-line demand diversion at freeway entrance ramps of the test corridor. Test results from 5-min prediction on several weekdays indicate the entrance ramp diversion proportion on ramps is reasonable.

Although commuting trips in a corridor of a specific metropolitan area were measured, the behavioral principles underlying the models also should be applicable to similar trips in other urban areas. Further, judging from the test results, the filtering algorithm should be effective in reducing real-time prediction error. The prediction methodology was applied at freeway entrance ramps, but it could be extended to prediction at freeway exit ramps. Work on extending the prediction framework to intersection turning movements, while relaxing the assumptions made in the filter formation, is in progress.

ACKNOWLEDGMENT

The helpful comments made by two anonymous reviewers are particularly acknowledged. This work was partly supported by the National Science Foundation and the Center for Transportation Studies of the University of Minnesota.

REFERENCES


Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.