Integrated Real-Time Ramp Metering Model for Nonrecurrent Congestion: Framework and Preliminary Results

GANG-LEN CHANG, JIFENG WU, AND STEPHEN L. COHEN

An integrated real-time ramp metering model for nonrecurrent freeway congestion among link flows has been developed and tested in this study. The core concept of the proposed algorithm is to capture the dynamic traffic state evolution with a two-segment linear flow-density model. To be implemented in real time, an effective solution algorithm has been proposed for determining the time-varying metering rates. The entire algorithm has also been integrated with INTRAS, the most well-known freeway simulation model, for conducting simulation experiments. Preliminary research results indicate that the proposed integrated control model is promising because its effectiveness increases with the severity of accidents and the level of congestion. The model execution time is also sufficiently short for potential real-time operations.

Ramp metering is a widely recognized potential control strategy for alleviating freeway congestion. Over the past several decades, traffic engineers have proposed and designed various ramp metering algorithms. One of the pioneering studies on this subject is the so-called time-of-day control proposed by Wattleworth and Berry (*1*) and further developed by Wattleworth (*2*) and Papageorgiou (*3*). The time-of-day control uses a linear programming model to generate the pretimed metering rates on the basis of the freeway capacity and regular daily traffic demands. Most other existing ramp metering studies are based on the concept of local traffic-responsive control, such as the percent occupancy strategy (*4*). The two well-known local traffic-responsive strategies are the demand-capacity method that is similar to the traditional occupancy-based strategy and the linear feedback strategy in ALINEA (*5*). Other conventional local traffic-responsive strategies include speed control metering and gap acceptance merge control. These two methods, along with the pretimed metering and the demand-capacity strategy, have been implemented in the microscopic freeway simulation model INTRAS (*6*). Although they all can be used to improve freeway congestion to some extent, all of these strategies have some limitations. Because the time-of-day strategy is based on past traffic patterns without consideration of actual current traffic condition, it obviously cannot be expected to be effective if nonrecurrent congestion occurs because of incidents. Although local traffic-responsive strategies do respond to actual traffic conditions, they do not impose metering rates on those ramps far upstream of the incident location because they cannot respond until the congestion reaches the detectors that control them.

In view of the deficiencies of time-of-day as well as local control, several studies on integrated traffic-responsive strategies have been proposed in recent years. Most of them have been grounded on optimal control theory, which usually leads to a large-scale nonlinear optimization problem. The most common approach is to employ the linear-quadratic optimization technique based on the minimization of a quadratic performance functional that penalizes deviations from nominal values of traffic status (*7–9*). Another way to approximately solve such a large-scale nonlinear optimization problem is the hierarchical decomposition algorithm presented by Papageorgiou (*10,11*). This method consists of three functional layers: an optimization layer based on steady-state traffic distribution patterns, an adaptation layer, and a direct control layer that implements local feedback controls. Because both link density and speed have been used as status variables and the dynamic model describing mean link speed evolution is rather complex, these nonlinear optimal control-based methods, although accurate in addressing the problem, generally require considerable computation effort for solution. Through analytical linearization of the nonlinear models, a linear regulator formula has been proposed by Payne et al. (*7*) for interconnected traffic-responsive ramp metering control. However, this linear regulator model (and linearization schemes for the nonlinear models described) is not applicable to the incident control case because the deviations from nominal conditions are large, hence the controls required to return to the nominal condition are also large and not describable with linear approximations.

Because the existing optimization-based models are generally too complex for on-line application, some heuristic areawide ramp metering algorithms also have been proposed. For instance, Koble et al. (*4*) developed an incident-specific ramp metering strategy by explicitly predicting the shock wave frontage induced by the incident. More recently, Nihan et al. (*12*) reported a predictive ramp metering algorithm that has been tested on line with very good accuracy and is especially effective for lightly congested flow conditions. Other heuristic strategies include the extended local traffic-responsive control in the FRECON2 model developed by University of California at Berkeley and the areawide demand-responsive ramp metering system of the I-5 corridor in Seattle, Washington. Although so many approaches have been developed, so far not one has been proved adequate for real-time freeway control and operations. Hence, it is still a challenging task to develop more effective real-time control strategies.

This paper reports the development of a new integrated real-time ramp metering algorithm. The first section describes the problem in integrated ramp metering control; the next section presents dynamic traffic models and an optimal control process; and the following section addresses some critical issues, such as the optimization of ramp metering rates, the estimation of the

G-L. Chang and J. Wu, Department of Civil Engineering, University of Maryland, College Park, MD 20742. S. L. Cohen, IVHS Division, HSR-10, FHWA, 6300 Georgetown Pike, McLean, Va. 22101.

traffic status, and the prediction of system model parameters. The next section presents an efficient algorithm for real-time applications. For preliminary evaluation and implementation, the proposed algorithm is integrated with INTRAS. The following section reports some numerical results with respect to its performance evaluation against several other ramp metering strategies. The final section summarizes conclusions and recommendations.

PROBLEM DEFINITION

An effective real-time ramp metering system should be (*a*) responsive and capable of determining the metering rates in real time according to the current traffic status and (*b*) effective in coordinating all interacting ramps so as to achieve a global system optimum. Conceivably, to be responsive, the on-line traffic status information must be obtainable; to be effective in coordination, the employed optimization model should realistically capture the dynamic interactions among ramp and freeway flows.

Consider a general freeway section, including a number of on ramps and off ramps as shown in Figure 1. Suppose that it can be conceptually divided into N small segments (links), and that each small segment contains at most one on ramp and one off ramp. The control time period is divided into a series of equal intervals. Before the presentation of the modeling structure, the definitions of all variables involved are summarized in Table 1.

Generally, previous and current traffic volume data $q_i(k)$, $r_i(k)$, $s_i(k)$ can be obtained from any surveillance systems. However, the traffic status variables and model parameters need to be estimated indirectly from the surveillance data, including volume and occupancy. Hence, given the dynamic travel demand pattern $\{q_0(k), D_i(k), \theta_i(k)\}$ and incident factors $\{\sigma_i(k)\}$ in advance, the ultimate challenge is to determine the real-time metering rates $R_i(k)$ so as to achieve a global system-optimum status.

MODEL FORMULATION

Freeway Traffic Status Dynamics

Suppose that an equilibrium flow-density relation $Q_i(k) = Q_i[\rho_i(k)]$ exists for each freeway link i; then the traffic status on a segment can be described simply by the mean link density. A dynamic equation for density evolution according to the flow conservation law can be formulated as follows:

$$\rho_i(k) = \rho_i(k-1) + [q_{i-1}(k) + \delta_i^{on}R_i(k) - \delta_i^{off}\theta_i(k)Q_i(k) - q_i(k)]T/L_i l_i \qquad i = 1, 2, \ldots, N \tag{1}$$

where $\delta_i^{on}R_i(k)$ and $\delta_i^{off}\theta_i(k)Q_i(k)$ are, respectively, the expected flow rates entering and exiting link i through on ramps or off ramps.

The transition flow rate, $q_i(k)$, between adjacent links i and $i + 1$ can be approximated with the weighted sum of two segment boundary flows: $[1 - \delta_i^{off}\theta_i(k)]Q_i(k)$ and $Q_{i+1}(k) - \delta_i^{on}R_{i+1}(k)$. The two weight factors are denoted by $\alpha_i(k)$ and $l - \alpha_i(k)$, respectively. Thus,

$$q_i(k) = \alpha_i(k)[1 - \delta_i^{off}\theta_i(k)]Q_i(k) + [1 - \alpha_i(k)][Q_{i+1}(k) - \delta_{i+1}^{on}R_{i+1}(k)] \qquad i = 1, 2, \ldots, N-1 \tag{2}$$

and $q_N(k)$ is just the downstream boundary flow rate of freeway link N, so it can be computed with

$$q_N(k) = Q_N(k) - \delta_N^{off}\theta_N(k)Q_N(k) \tag{3}$$

The parameters $\alpha_i(k)$ in Equation 2 play an important role in capturing the interrelations between adjacent segment flows.

Incorporating Equations 2 and 3 into Equation 1 leads to the following equation:

$$\begin{aligned}\rho_i(k) = \rho_i(k-1) &+ [T/(L_i l_i)]\alpha_{i-1}(k)[1 - \delta_{i-1}^{off}\theta_{i-1}(k)] \cdot Q_{i-1}(k)\\ &+ [T/(L_i l_i)\{1 - \alpha_{i-1}(k) - \alpha_i(k) - \delta_i^{off}\\ &\times [1 - \alpha_i(k)]\theta_i(k)\} \cdot Q_i(k)\\ &+ [T/(L_i l_i)][\alpha_i(k) - 1] \cdot Q_{i+1}(k)\\ &+ [T/(L_i l_i)]\alpha_{i-1}(k\delta_i^{on} \cdot R_i(k)\\ &+ [T/(L_i l_i)][1 - \alpha_i(k)]\delta_{i+1}^{on} \cdot R_{i+1}(k)\\ &i = 2, 3, \ldots, N-1\end{aligned} \tag{4a}$$

For $i = 1$ and N,

$$\begin{aligned}\rho_1(k) = \rho_1(k-1) &+ [T/(L_1 l_1)] \cdot q_0(k)\\ &+ [T/(L_1 l_1)][\alpha_1(k) - \delta_1^{off}\theta_1(k)\\ &+ \delta_1^{off}\alpha_1(k)\theta_1(k)] \cdot Q_1(k)\\ &+[T/(L_1 l_1)][\alpha_1(k) - 1] \cdot Q_2(k)\\ &+ [T/(L_1 l_1)]\delta_1^{on} \cdot R_1(k)\\ &+ [T/(L_1 l_1)]\delta_2^{on}[1 - \alpha_1(k)] \cdot R_2(k)\end{aligned} \tag{4b}$$

$$\begin{aligned}\rho_N(k) = \rho_N(k-1) &+ [T/(L_N l_N)]\alpha_{N-1}(k)[1 - \delta_{N-1}^{off}\theta_{N-1}(k)]\\ &\cdot Q_N(k) - [T/(L_N l_N)]\alpha_{N-1}(k) \cdot Q_N(k)\\ &+ [T/(L_N l_N)]\alpha_{N-1}(k)\delta_N^{on} \cdot R_N(k)\end{aligned} \tag{4c}$$

Hence, given the parameter values $\{\alpha_i(k)\}$, it can be seen that if $Q_i(k)$ is a linear function of $\rho_i(k)$, the density evolution equation (Equation 4) should also be a linear dynamic system. Now, assuming that a linear flow-density relation holds,

$$Q_i(k) = [w_i + u_i\rho_i(k)]\sigma_i(k) \tag{5}$$

in which the parameters w_i and u_i depend on the range of density $\rho_i(k)$, and $\sigma_i(k)$ is the incident factor representing capacity reduction as a result of incidents. If no incident occurs on link i, $\sigma_i(k)$ should equal 1.

To facilitate the presentation, the following two new vector variables are defined:

$$\rho(k) = [\rho_1(k), \ldots, \rho_N(k)]^T$$

$$\mathbf{R}(k) = [R_1(k), \ldots, R_N(k)]^T$$

Then, from Equations 4 and 5, the following matrix form for the dynamic linear density equation is obtained:

$$\rho(k) = \rho(k-1) + \mathbf{A}_1(k)\rho(k) + \mathbf{A}_2(k)\mathbf{R}(k) + \mathbf{a}(k) \tag{6}$$

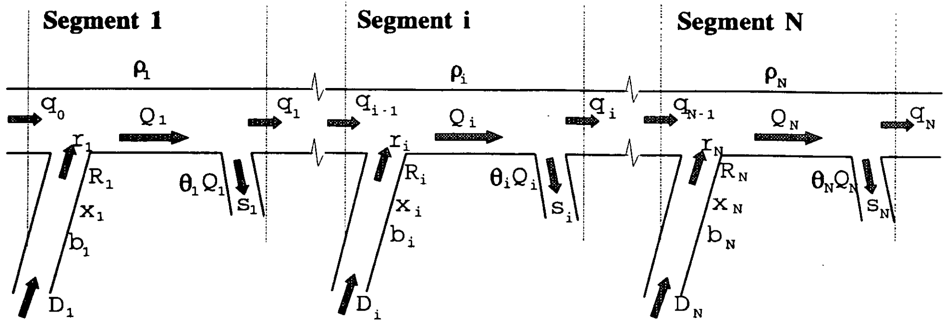

FIGURE 1 General freeway section and its major arguments.

where $\mathbf{A}_1(k) = [a^1_{ij}(k)]$ is an $N * N$ matrix with its elements given by

$$a^1_{11}(k) = [T/(L_1 l_1)][-\alpha_1(k) - \delta_1^{\text{off}}\theta_1(k) + \delta_1^{\text{off}}\alpha_1(k)\theta_1(k)]u_1\sigma_1(k)$$

$$a^1_{12}(k) = [T/(L_1 l_1)][\alpha_1(k) - 1]u_2\sigma_2(k)$$

$$a^1_{1j}(k) = 0 \qquad \text{for } j = 3, 4, \ldots, N$$

$$a^1_{i,i-1}(k) = [T/(L_i l_i)]\alpha_{i-1}(k)[1 - \delta^{\text{off}}_{i-1}(k)]u^{i-1}\sigma_{i-1}(k)$$

$$a^1_{i,i}(k) = [T/(L_i l_i)][(1 - \alpha_{i-1}(k) - \alpha_i(k) - \delta_i^{\text{off}}\theta_i(k) + \alpha_i(k)\delta_i^{\text{off}}\theta_i(k)]u_i\sigma_i(k)$$

$$a^1_{i,i+1}(k) = [T/(L_i l_i)][\alpha_i(k) - 1]u_{i+1}\sigma_{i+1}(k)$$

$$a^1_{i,j}(k) = 0 \qquad \text{for } j = 1, \ldots, i-2, i+2, \ldots, N; \quad i = 2, \ldots, N-1$$

$$a^1_{N,N-1}(k) = [T/(L_N l_N)]\alpha_{N-1}(k)[1 - \delta^{\text{off}}_{N-1}\theta_{N-1}(k)] \times u_{N-1}\sigma_{N-1}(k)$$

$$a^1_{N,N}(k) = -[T/(L_N l_N)]\alpha_{N-1}(k)u_N\sigma_N(k)$$

$$a_{N,j}{}^1(k) = 0 \qquad \text{for } j = 1, 2, \ldots, N-2$$

$\mathbf{A}_2(k) = [a^2_{ij}(k)]$ is an $\mathbf{N} * \mathbf{N}$ matrix with its elements given by

$$a^2_{1,1}(k) = [T/(L_1 l_1)]\delta_1^{\text{on}}$$

$$a^2_{1,2}(k) = [T/(L_1 l_1)][1 - \alpha_1(k)]\delta_2^{\text{on}}$$

$$a^2_{1,j}(k) = 0 \qquad \text{for } j = 3, 4, \ldots, N$$

$$a^2_{i,i}(k) = [T/(L_i l_i)]\alpha_{i-1}(k)\delta_i^{\text{on}}$$

$$a^2_{i,i+1}(k) = [T/(L_i l_i)][1 - \alpha_i(k)]\delta^{\text{on}}_{i+1}$$

$$a^2_{i,j}(k) = 0 \qquad \text{for } j = 1, \ldots, i-1, i+2, \ldots, N; \quad i = 2, \ldots, N-1$$

$$a^2_{N,N}(k) = [T/(L_N l_N)]\alpha_{N-1}(k)\delta_N^{\text{on}}$$

$$a^2_{N,j}(k) = 0 \qquad \text{for } j = 1, 2, \ldots, N-1$$

$a(k) = [a_1(k), \ldots, a_N(k)]^T$ is an $\mathbf{N} * 1$ vector with its elements given by

$$a_1(k) = [T/(L_1 l_1)] \cdot q_0(k) + [T/(L_1 l_1)] \times [-\alpha_1(k) - \delta_1^{\text{off}}\theta_1(k) + \alpha_1(k)\delta_1^{\text{off}}\theta_1(k)]w_1\sigma_1(k) + [T/(L_1 l_1)][\alpha_1(k) - 1]w_2\sigma_2(k)$$

$$a_i(k) = [T/(L_i l_i)]\alpha_{i-1}(k)[1 - \delta^{\text{off}}_{i-1}\theta_{i-1}(k)]w_{i-1}\sigma_{i-1}(k) + [T/(L_i l_i)][1 - \alpha_{i-1}(k) - \alpha_i(k) - \delta_i^{\text{off}}\theta_i(k) + \alpha_i(k)\delta_i^{\text{off}}\theta_i(k)]w_{i\delta i}(k) + [T/(L_i l_i)][(\alpha_i(k) - 1]w_{i+1}\sigma_{i+1}(k)$$

$$\text{for } i = 2, 3, \ldots, N-1$$

$$a_N(k) = [T/(L_N l_N)]\alpha_{N-1}(k)[1 - \delta^{\text{off}}_{N-1}\theta_{N-1}(k)] \times w_{N-1}(k)\sigma_{N-1}(k) - [T/(L_N l_N)]\alpha_{N-1}(k)w_N(k)\sigma_N(k)$$

Suppose that matrix $[\mathbf{I} - \mathbf{A}_1(k)]$ is invertible; then the dynamic density relation (Equation 6) can be reformulated into a canonical form

$$\rho(k) = \mathbf{A}(k)\rho(k-1) + \mathbf{B}(k)\mathbf{R}(k) + \mathbf{d}(k) \tag{7}$$

where

$\mathbf{A}(k) = [\mathbf{I} - \mathbf{A}_1(k)]^{-1}$;

$\mathbf{B}(k) = [\mathbf{I} - \mathbf{A}_1(k)]^{-1}\mathbf{A}_2(k)$; and

$\mathbf{d}(k) = [\mathbf{I} - \mathbf{A}_1(k)]^{-1}\mathbf{a}(k)$.

Ramp Traffic Status Dynamics

Another major traffic status variable is the mean number of vehicles (content), $X_i(k)$, occupying each on ramp i. According to flow conservation law, the content change from Interval $k - 1$ to k is given as

$$x_i(k) = x_i(k-1) + [d_i(k) - R_i(k)]T$$

In a matrix form, it becomes

$$\mathbf{X}(k) = \mathbf{X}(k-1) + \mathbf{E}(k)\mathbf{R}(k) + \mathbf{D}(k) \tag{8}$$

TABLE 1 Definitions of Relevant System Variables

Network geometric and physical data	
L_i:	physical length of segment i
l_i:	number of lanes contained in segment i
ρ^{cr}:	critical mean density value at which link flow rate reaches its maximum
ρ^{max}:	jam density value for freeway links
w_i, u_i:	parameters of equilibrium freeway flow-density model for link i
R_i^{max}:	maximum metering rate for on-ramp i
R_i^{min}:	minimum metering rate for on-ramp i
b_i:	vehicle storage capacity of on-ramp i
Designed parameters for modeling analysis	
T:	duration of one time interval
M:	number of time intervals involved in a time horizon for optimization
N:	number of subsegments divided for the entire freeway section
Dynamic traffic demands	
$q_0(k)$:	flow rate entering the upstream boundary of the freeway section during interval k
$D_i(k)$:	flow rate entering the upstream on-ramp i during interval k
$\theta_i(k)$:	proportion of turning traffic at off-ramp i during interval k
Traffic volumes	
$q_i(k)$:	flow rate entering freeway link i+1 from link i during interval k
$Q_i(k)$:	mean flow rate of freeway link i during interval k
$r_i(k)$:	actual flow rate entering the freeway from on-ramp i during interval k
$s_i(k)$:	actual flow rate exiting the freeway at off-ramp i during interval k
Incident information	
$\sigma_i(k)$:	capacity reduction parameter, with $[1-\sigma_i(k)]100\%$ representing the reduced percentage of capacity for link i due to incidents
Dynamic model parameters	
$\alpha_i(k)$:	parameter to represent the interaction between flows of link i and i+1 during interval k
Traffic status variables	
$\rho_i(k)$:	mean density of link i during interval k
$x_i(k)$:	mean number of vehicles (content) at on-ramp i during interval k
$Q_i(k)$:	mean flow rate on freeway link i during interval k
Control variables to be solved	
$R_i(k)$:	metering flow rate for on-ramp i during interval k

where

$$\mathbf{X}(k) = [x_1(k), \ldots, x_N(k)]^T$$

$$\mathbf{E}(k) = \text{diag}[-T\delta_1(k), \ldots, -T\delta_N(k)]$$

$$\mathbf{D}(k) = [T \cdot D_1(k)\delta_1(k), \ldots, T \cdot D_N(k)\delta_N(k)]^T$$

So far, Equations 7 and 8 have represented the interrelations between freeway evolution dynamics and the ramp control variable $\{R_i(k)\}$. The appropriate objective function will now be defined.

Objective for Ramp Metering Control

As in most traffic control strategies, the primary purpose of ramp metering is to alleviate freeway congestion, both recurrent and nonrecurrent, and thus improve its performance. Several measures of effectiveness (MOEs) have been proposed in the literature to evaluate the operational performance, such as total throughput, total vehicle-miles, average speed, and total delay. Theoretically, the total traffic throughput (TTT) is relatively more appealing than others and thus is chosen as the control objective of this study. TTT is defined as the total number of vehicles discharging from the freeway section over the control period of interest. For the example freeway section shown in Figure 1, the total throughput is given by

$$TTT = \sum_k \left[\sum_i^{N-1} \delta_i^{\text{off}} \theta_i(k) Q_i(k) + Q_N(k) \right] T$$

$$= T \sum_k \left[\sum_{i=1}^{N-1} \delta_i^{\text{off}} \theta_i(k) u_i \sigma_i(k) \rho_i(k) + u_N \sigma_N(k) \rho_N(k) \right]$$

$$+ T \sum_k \left[\sum_{i=1}^{N-1} \delta_i^{\text{off}} \theta_i(k) w_i \sigma_i(k) + w_N \sigma_N(k) \right] \quad (9)$$

Model Constraints

The essential constraints for optimizing ramp metering rates are the dynamic traffic evolution equations (Equations 7 and 8). The other constraints are the physical lower and upper bounds for the mean link density values, the metering rates, and the ramp contents:

$$0 \le \rho_i(k) \le \rho^{\max} \qquad i = 1,2, \ldots, N \quad (10)$$

$$R_i^{\min} \le R_i(k) \le R_i^{\max} \qquad i = 1,2, \ldots, N \quad (11)$$

$$0 \le x_i(k) \le b_i \qquad i = 1,2, \ldots, N \quad (12)$$

Note that an additional operational constraint is required that pursues the implementation of sufficiently large, if necessary, metering rates so as to prevent ramp queues from spilling back to surface streets. This objective may conflict with the total freeway throughput when traffic demand is high. It has to be temporarily ignored if the mainline freeway operational MOE is the primary consideration. When improvement on the entire network is the ultimate goal, ramp metering control only is not enough, and additional control measures, including real-time diversion control and signal timing coordination at surface street intersections, must also be considered for achieving optimal control. This type of more complex control issue at the corridor network level has been approached by Chang et al. (*13*).

Optimal Control Model

Theoretically, the optimal time-varying ramp metering rate for the entire control period can be solved in one step. In practice, however, considering both the required computational effect for real-time operations and the dynamic nature of all key time-varying parameters, it is recommended that each optimal control mode be executed over a relatively shorter period and updated with the feedback information from surveillance systems. Supposing that a time horizon comprising M consecutive intervals is chosen as one control period, an optimal control model can be formulated as follows:

Maximize objective function Equation 9

subject to Equations 7, 8, 10, 11, and 12 (13)

However, before applying this optimization model, all unknown model coefficients and parameters involved in Model 13 must be identified. In particular, current link densities must be estimated so that a proper linear-form flow density model (Equation 5) can be calibrated.

STATUS ESTIMATION AND DYNAMIC PARAMETER PREDICTION

Traffic Status Estimation

It is notable that traffic status information [$\rho(t-1)$ and $\mathbf{X}(t-1)$] at the beginning of interval t must be estimated for the optimization model Equation 13. Several methods are available in the literature for dealing with such issues. For instance, Kalman filtering is one of the most efficient approaches for system status identification and has been extensively applied in traffic control. The single segment estimation (SSE) approach developed by Payne et al. (*14*) is an example application of the Kalman filtering technique for estimating link densities and on-ramp queuing lengths with point volume and occupancy data from surveillance detectors. The SSE approach can be directly utilized in this study.

Parameters Updating and Prediction

In addition to the traffic status information, the time-varying parameters $\{\alpha_i(k)\}$ in Model 13 must be identified before the execution of the models. At each instant, since the parameters' current and previous values can be obtained from the traffic surveillance data, some type of time-series model such as the autoregressive moving average (ARMA) model (*15*) can be calibrated and applied to predict the future parameter values.

More specifically, at interval t, given the $r_i(k)$, $s_i(k)$, $q_i(k)$, and $Q_i(k)$ for all i from the on-line traffic surveillance system, the parameters α_i can be updated according to Equation 2:

$$\alpha_i(t) = \frac{q_i(t) - [Q_{i+1}(t) - \delta_{i+1}^{\text{on}} r_{i+1}(t)]}{[Q_i(t) - \delta_i^{\text{off}} s_i(t)] - [Q_{i+1}(t) - \delta_{i+1}^{\text{on}} r_{i+1}(t)]}$$

With the following simple autoregressive model AR(m) of m lags, the future parameter values $\{\alpha_i(k)\}$ can be obtained through a time-series recursive prediction:

$$\alpha_i(k) = \sum_{j=1}^{m} \delta_j(k) \alpha_i(k-j) \quad (14)$$

However, because of the dynamic nature of the model coefficients $\delta_j(k)$, before performing prediction, these $\delta_j(k)$, $j = 1, \ldots, m$, should be updated with the current $\alpha_i(k)$. Such an updating process can be executed with the application of a linear least-squares algorithm or the Kalman filtering technique.

Note that to update the above model parameters with Kalman filtering, it is convenient to assume that all $\delta_j(k)$ follow a random walk model. In this way, a canonical status space dynamic model can be set up as

$$\delta_j(k) = \delta_j(k-1) + w_j(k) \qquad j = 1,2, \ldots, m$$

$$e_i(k) = \sum_{j=1}^{m} \delta_j(k) e_i(k-j) + v_i(k)$$

where $w_j(k)$ and $v_i(k)$ are Gaussian white noise with known variance-covariance matrices. Then the parameter updating is given by

$$\delta_j(k) = \delta_j(k-1) + K_j\left[e_i(k) - \sum_{n=1}^{m} \delta_n(k-1)e_i(k-n)\right]$$

where K_j is the associated Kalman gain.

With the new coefficients $\{\delta_j(k)\}$ the future parameter values $\{\alpha_i(t)\}$ can thus be predicted with Equation 14.

SUCCESSIVE LINEAR PROGRAMMING ALGORITHM

Having established the modeling concepts, an integrated algorithm procedure for computing real-time ramp metering rates is shown in Figure 2. The basic logic governing this algorithm is the rolling horizon concept that was first introduced to traffic control by Gartner (*16*) and further utilized by Chang et al. (*17*).

One of the most important aspects of this real-time algorithm framework is the feedback control. At each step, the model computes the optimal ramp metering rates for the next M intervals, that is, a set of $R(t)$, $R(t+1)$, . . . , $R(t+M-1)$ is computed. The control system will then process the comparison between the predicted and detected traffic conditions so as to determine the

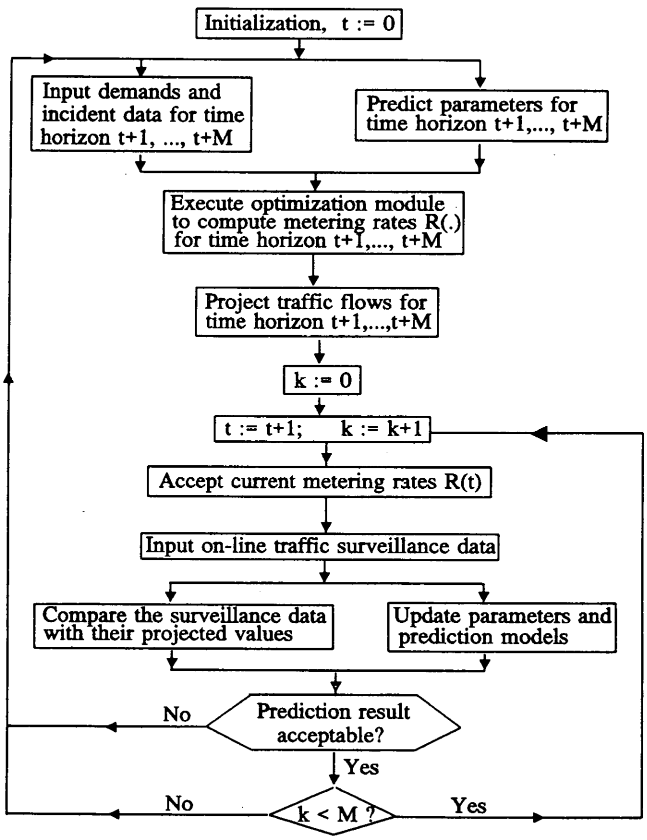

FIGURE 2 Flow chart of the real-time ramp metering control logic.

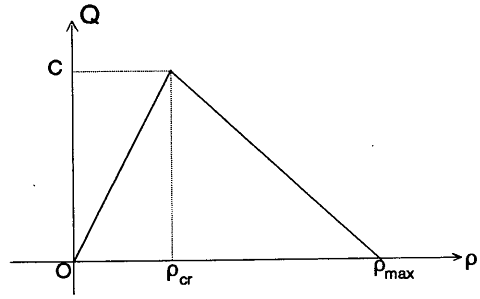

FIGURE 3 Two-segment linear flow-density model.

acceptability of those future metering rates. As shown in Figure 2, if the projected traffic flows at interval t are consistent with the on-line surveillance data, one may accept the computed metering rates for subsequent intervals; otherwise, the optimization procedure for the next time horizon is repeated after the parameters are updated and prediction with the new surveillance data is executed.

Now turning to the central part of the ramp metering algorithm, consider the optimal control model Equation 13 and the dynamic constraints Equations 7 and 8. It seems to be a linear programming problem with respect to variables $\rho(k)$, $\mathbf{X}(k)$, and $\mathbf{R}(k)$, $k = t+1$, . . . , $t+M$. However, Equation 13 is not pure linear programming because the coefficient matrices of the density dynamic constraint Equation 7 depend on the range to which the density $\rho(k)$ belongs. A linear-form dynamic equation is derived under the condition that a linear flow-density relation (Equation 5) holds. Therefore, before the linear programming techniques are applied, a piecewise linear flow-density model must be calibrated.

According to recent research studies reported in the literature (*18,19*) a two-segment linear flow-density function, as shown in Figure 3, is reasonable for representing freeway traffic flows. Thus for any freeway link i, Equation 5 has two sets of parameters $w_i(k)$ and $u_i(k)$ corresponding to the two density ranges $[0,\rho^{cr}]$ and $[\rho^{cr},\rho^{max}]$ for calibrating the linear function.

Now it is clear that the corresponding boundary constraints for $\rho_i(k)$ should be added when a linear dynamic Equation, such as Equation 7, is used. To apply a linear programming model for such a unique optimal metering model and dynamic constraints, a special technique is proposed, successive linear programming (SLP) algorithm, which enables the model to be executed sufficiently fast for real-time applications. All principal steps of the proposed SLP algorithm are summarized as follows:

Step 1.1 According to the current traffic status and metering flows, add the corresponding lower- and upper-bound constraints for $\rho_i(k)$, $i = 1,2, \ldots, N$; $k = t+1, \ldots, t+M$, to the LP Model 13.

Step 1.2 Compute the coefficient matrices of the linear density dynamic equation according to the current linear flow-density models.

Step 1.3 Solve the resulting linear programming model to obtain a set of solutions $\rho(k)$, $X(k)$, and $R(k)$ and the corresponding objective function value.

Step 2 Check whether some $\rho_i(k)$ equals ρ^{cr}. If not, stop with the current LP solution. Otherwise, go to Step 3.

Step 3 Change the lower- and upper-bound constraints for those $\rho_i(k) \approx \rho^{cr}$ into the other range and modify the corresponding parameters in the linear flow-density models and the coefficient matrices of the linear density dynamic equation constraints.

Step 4 Solve the updated LP model to obtain a new set of solutions and its corresponding objective function value.

Step 5 Check whether the objective function value has been improved. If not, stop with the current LP solution; otherwise, return to Step 2.

A more detailed discussion of the properties and performance of the proposed SLP algorithm can be found elsewhere (*13*). Not all LP problems generated by the algorithm can guarantee a feasible solution. If there is no feasible solution, any local traffic-responsive strategy can be applied instead at this iteration step, so as to continue the algorithm procedure. In the simulation tests performed here, it was found that such infeasible LPs occur only rarely; hence it has only a slight impact on the performance of the algorithm. To show the potential effectiveness of the models and strategy developed in this study, a simulation experiment is presented in the next section.

NUMERICAL TESTS

Field Network and Surveillance Detectors Assignment

To evaluate the proposed model and algorithm, a section of the I-5 corridor in Seattle, Washington, was selected as the field network for simulation tests. As shown in Figure 4, this corridor network contains 9 on ramps, 6 off ramps, and one parallel arterial (SR-99) as well as 14 crossing surface streets. In this simulation test, exactly one full set of loop detectors was placed close to each node for freeway links. To minimize the use of detectors, each detector station was located at the on-ramp upstream merging point or the off-ramp downstream diverting point, depending on the node configuration. In addition, two detectors were placed near the upstream and downstream boundaries of each on ramp, and one detector was placed near the upstream boundary of each off ramp. Because all ramps in this network have one lane, one detector is sufficient for each ramp station.

Simulation Design

In this simulation test, an incident was assumed to occur on freeway Link 7 → 8. Simulation tests were then performed for four different traffic conditions with different demand levels, incident severity, and duration. This simulation plan was based on a research report of a simulation study of coordinated signal control strategies by Farradyne Systems Inc. (*20*).

The following six on-ramp control strategies were specified to investigate their MOEs regarding freeway performance, given an identical control operation on the surface streets:

1. Strategy O: baseline case operation, no ramp metering;
2. Strategy A: close one on ramp immediately upstream of the incident site;

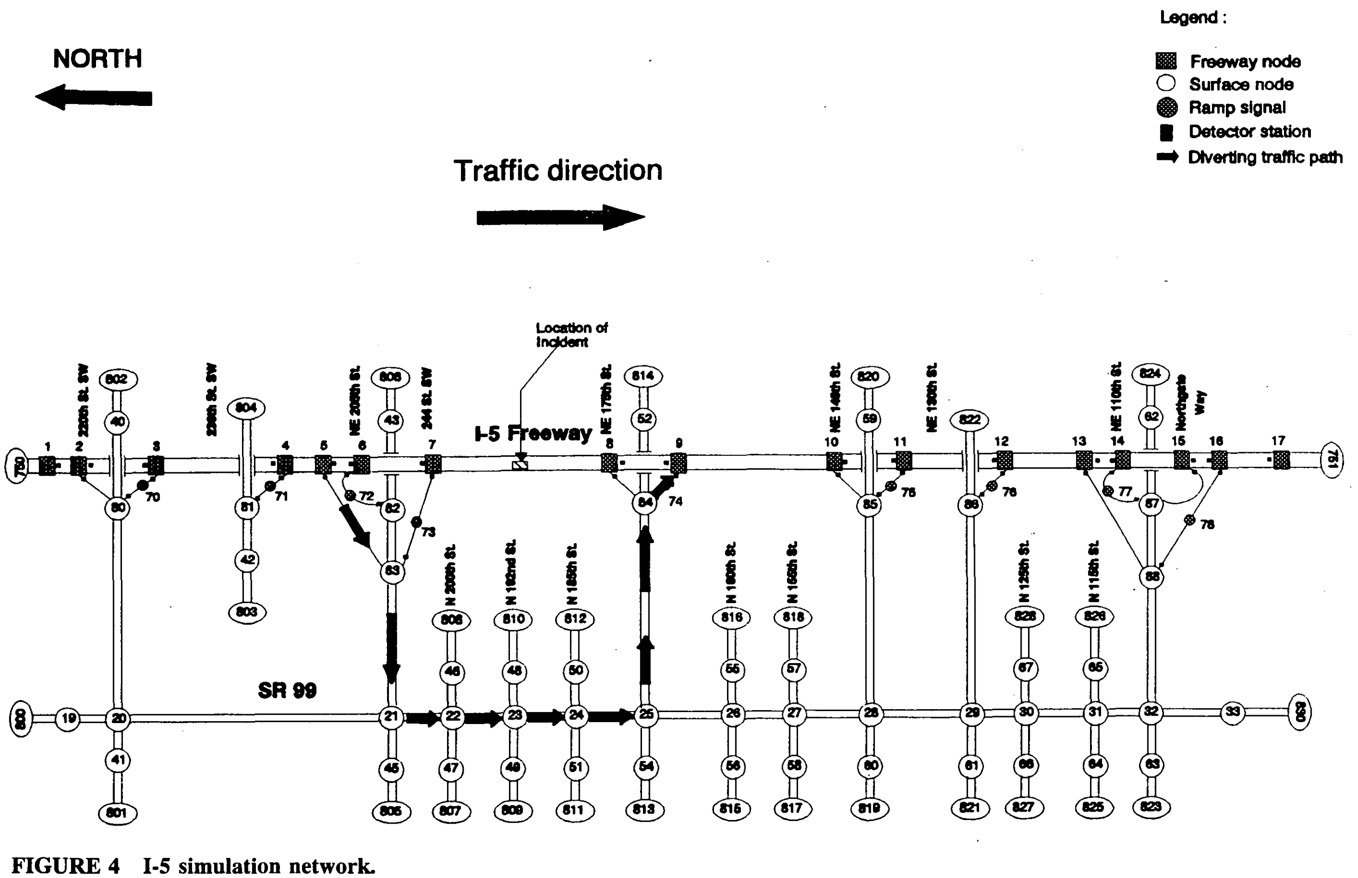

FIGURE 4 I-5 simulation network.

3. Strategy B: close two on ramps immediately upstream of the incident site;
4. Strategy C: local demand-capacity ramp metering;
5. Strategy D: Areawide ramp metering algorithm integrated with INTRAS and tested by Farradyne Systems, Inc. (*20*); and
6. Strategy S: Real-time metering with the SLP algorithm developed by the authors.

Each ramp metering strategy was implemented over four kinds of traffic conditions that are of different entry volume levels or different incident and traffic diversion patterns, or both. The following four test cases of traffic conditions are given:

Case 1. A total of 60 percent of the peak-hour volume [more concrete data and test results can be found elsewhere (*21*)], low incident level (i.e., one lane blockage, 10 percent rubberneck factor for other lanes, 10-min duration), 10 percent diversion from mainline to the diversion route as highlighted in Figure 5.

Case 2. A total of 60 percent of the peak-hour volume, high incident level (i.e., two lanes blockage, 20 percent rubberneck factor for other lanes, 20-min duration), 30 percent diversion from mainline to the diversion route.

Case 3. A total of 100 percent of the peak-hour volume, low incident level (i.e., one lane blockage, 10 percent rubberneck factor for other lanes, 10-min diversion from mainline to the diversion route).

Case 4. A total of 100 percent of the peak-hour volume, high incident level (i.e., two lanes blockage, 20 percent rubberneck factor for other lanes, 20 min duration), 30 percent diversion from mainline to the diversion route.

The simulation runs of all the six strategies over these four traffic conditions amount to 24 cases, which were named sequentially as follows:

O1, O2, O3, O4	A1, A2, A3, A4	B1, B2, B3, B4
C1, C2, C3, C4	D1, D2, D3, D4	S1, S2, S3, S4

Simulation Procedures

The simulation time for each of the 24 INTRAS runs was 35 intervals (35 min) over three periods as follows:

Period 1. A 5-min duration under normal traffic conditions without incident,

Period 2. A 20-min duration with an incident occurring at the beginning and lasting for 10 or 20 min. At the beginning of this period, diversion was performed at the off ramp immediately upstream of the incident location (freeway Node 5) by manual adjustment of turning percentages at intersections along the diversion route.

Period 3. A 10-min duration representing the recovery period after the incident has been removed from the freeway. After the removal of the incident, the turning percentages were reverted to those values before the incident, as specified in the first subinterval.

Simulation Results

As shown in Table 2, the TTT produced under Strategy 6 (with the SLP algorithm) is notably superior to those with other control strategies in all four cases of traffic conditions. The performance of the proposed SLP strategies can be examined further through the results shown in Figure 5, where the improvement in TTT increases with the level of congestion and the severity of the incident.

The simulation results with all other MOEs from the INTRAS output are also examined. For the mainline freeway operations, among the six strategies except baseline Strategy O, Strategy S has produced the following:

- The highest vehicle-miles and speed as well as the lowest vehicle-minutes and delay when both volume and incident levels are high (Case 4);
- The second-highest vehicle-miles but the lowest speed, highest vehicle-minutes and delay when volume level is high but incident level is low (Case 3);
- The second-highest vehicle-miles, medium speed and delay, but the second-highest vehicle-minutes when volume level is low but incident level is high (Case 2); and
- The highest vehicle-miles, second-highest speed and medium delay, but the second-highest vehicle-minutes when both volume and incident level are low (Case 1).

As a byproduct of the INTRAS output, these MOEs for the entire corridor network also have been obtained. Compared with the six strategies except baseline Strategy O, Strategy S has produced the following:

- The highest vehicle-miles, medium speed and delay, but the highest vehicle-minutes when both volume and incident levels are high (Case 4);
- The second-lowest vehicle-miles and speed as well as the highest vehicle-minutes and second-highest delay when volume level is high but incident level is low (Case 3);
- The second-highest vehicle-miles and medium speed but the second-highest vehicle-minutes and delay when volume level is low but incident level is high (Case 2); and
- The highest vehicle-miles, medium speed and delay, but the second-highest vehicle-minutes when both volume and incident level are low (Case 1).

In summary, the proposed SLP approach has shown convincing improvement over all other strategies for freeway operation in the case of heavy traffic and high incident level. However, its improvement under low congestion is not so significant as to justify the use of such a sophisticated method. In addition, with the entire corridor as the objective, no substantial improvement can be achieved with any algorithm under any of the cases. This actually implies that for contending with nonrecurrent congestion, one should view the entire corridor as a control system and perform both ramp metering and diversion control as studied by Chang et al. (*13*).

CONCLUSIONS

This study has developed an integrated ramp metering model with a piecewise linear dynamic optimal control function and an effi-

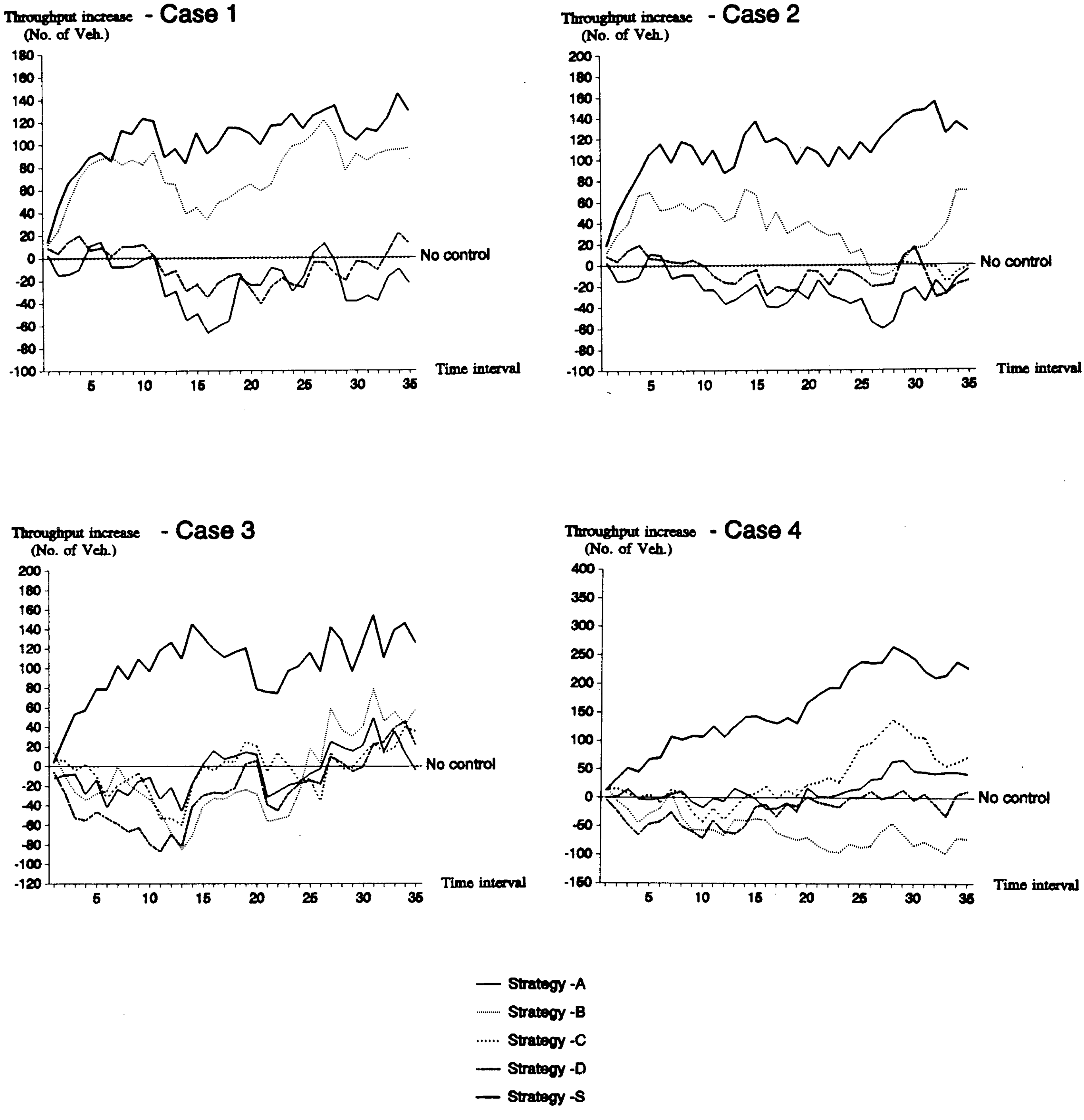

FIGURE 5 Cumulative throughput increases versus no control.

cient algorithm for real-time applications. Numerical tests of the proposed model and the SLP algorithm on a typical freeway corridor with INTRAS simulation demonstrated that the integrated control strategy with the proposed SLP algorithm outperforms all other strategies in total freeway throughput and in other MOEs, including total vehicle-miles, total vehicle-minutes, and average speed and delay under both high-volume and high-incident conditions. However, no significant performance can be achieved with any ramp metering strategies if the entire corridor network is concerned. Hence, although it is reasonable to conclude that the SLP algorithm is a promising strategy for real-time freeway nonrecurrent congestion control, integration with proper diversion control will be necessary to lead the entire corridor to optimum status.

TABLE 2 Simulation Results of Total Freeway Throughput

Control Strategy	Case 1	Case 2	Case 3	Case 4
Strategy -O	2927	2991	4535	4220
Strategy -A	2904	2987	4531	4263
Strategy -B	3024	3062	4593	4149
Strategy -C	2927	2990	4571	4293
Strategy -D	2939	2976	4557	4232
Strategy -S	3057	3120	4662	4449

NOTE: Values given are numbers of vehicles.

ACKNOWLEDGMENT

This research was supported by the Grants for Research Fellowships from FHWA, U.S. Department of Transportation.

REFERENCES

1. Wattleworth, J. A., and D. S. Berry. Peak Period Control of a Freeway System—Some Theoretical Investigations. In *Highway Research Record 89*, HRB, National Research Council, Washington, D.C., 1965, pp. 1–25.
2. Wattleworth, J. A. Peak Period Analysis and Control of a Freeway System. In *Highway Research Record 157*, HRB, National Research Council, Washington, D.C., 1967, pp. 10–21.
3. Papageorgiou, M. A New Approach to Time-of-Day Control Based on a Dynamic Freeway Traffic Model. *Transportation Research B*, Vol. 14B, 1980, pp. 349–360.
4. Koble, H. M., T. A. Adams, and V. S. Samant. *Control Strategies in Response to Freeway Incidents*, Vol. 2. Report FHWA/RD-80/005. ORINCON Inc., 1980.
5. Papageorgiou, M., H. Hadj-Salem, and J.-M. Blosseville. ALINEA: A Local Feedback Control Law for On-Ramp Metering. In *Transportation Research Record 1320*, TRB, National Research Council, Washington, D.C., 1991, pp. 58–64.
6. Wicks, D. A., and E. B. Lieberman. *Development and Testing of INTRAS, a Microscopic Freeway Simulation Model*, Vol. 1. Report FHWA/RD-80/106. KLD Associates, Inc. 1980.
7. Payne, H. J., D. Brown, and J. Todd. *Demand Responsive Strategies for Interconnected Freeway Ramp Control Systems. Vol. 1: Metering Strategies.* Verac Inc., 1985.
8. Papageorgiou, M., J.-M. Blosseville, and H. Hadj-Salem. Modeling and Real Time Control of Traffic Flow on the Southern Part of Boulevard Peripherique in Paris: Part II: Coordinated On-Ramp Metering. *Transportation Research A*, Vol. 24A, 1990, pp. 361–370.
9. Goldstain, N. B., and K. S. P. Kumar. A Decentralized Control Strategy for Freeway Regulation. *Transportation Research B*, Vol. 16B, 1982, pp. 279–290.
10. Papageorgiou, M. A Hierarchical Control System for Freeway Traffic. *IEEE Transactions on Automatic Control*, Vol. AC-29, 1983, pp. 482–490.
11. Papageorgiou, M. A New Approach to Time-of-Day Control Based on Dynamic Freeway Traffic Model. *Transportation Research B*, Vol. 14B, 1980, pp. 349–360.
12. Nihan, N. L., D. B. Berg, and L. N. Jacobson. *Predictive Algorithm Improvements for a Real Time Ramp Control System.* Final Report. Research Project GC8286, Washington State Department of Transportation, 1991.
13. Chang, G.-L., J. Wu, and H. Lieu. A Real-Time Incident-Responsive System for Corridor Control: A Modelling Framework and Preliminary Results. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1994.
14. Payne, H. J., D. Brown, and S. L. Cohen. Improved Techniques for Freeway Surveillance. In *Transportation Research Record 1112*, TRB, National Research Council, Washington, D.C., 1987, pp. 52–60.
15. Box, G. F. P., and G. M. Jenkins. *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco, 1976.
16. Gartner, N. H. OPAC: A Demand-Responsive Strategy for Traffic Signal Control. In *Transportation Research Record 906*, TRB, National Research Council, Washington, D.C., 1983.
17. Chang, G.-L., P. K. Ho, and C. H. Wei. A Dynamic System-Optimal Control Model for Commuting Traffic Corridors. *Transportation Research C*, Vol. 1, No. 1, 1993.
18. Hall, F. L., B. L. Allen, and M. A. Gunter. Empirical Analysis of Freeway Flow-Density Relationships. *Transportation Research A*, Vol. 20A, 1986, pp. 197–210.
19. Banks, J. H. Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations. In *Transportation Research Record 1225*, TRB, National Research Council, Washington, D.C., 1989, pp. 53–60.
20. Farradyne Systems, Inc. *Coordinated Operation of Ramp Metering and Adjustment Traffic System Control Systems.* Draft Final Report. FHWA, DTFH61-89-C-00006, 1992.
21. Wu, J. Development and Evaluation of Real-time Ramp Metering Algorithms for Alleviating Non-Recurrent Congestion. Research Report of GRF Project. IVHS Division, FHWA, U.S. Department of Transportation, Aug. 1993.

Publication of this paper sponsored by Committee on Freeway Operations.