March, 1979 LIDS-P-889

# DYNAMIC MODEL-BASED TECHNIQUES FOR THE DETECTION OF INCIDENTS ON FREEWAYS*

A.S. Willsky, E.Y. Chow, S.B. Gershwin, C.S. Greene, P.K. Houpt, and A.L. Kurkjian *

Laboratory for Information and Decision Systems
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

## Abstract

In this paper we discuss an approach to the detection of incidents on freeways. Our techniques are based on the use of a macroscopic dynamic model describing the evolution of spatial-average traffic variables (velocities, flows, and densities) over sections of the freeway. With such a model as a starting point we develop two incident detection algorithms based on the multiple model and generalized likelihood ratio techniques. We also describe a new and very simple system for processing raw data from presence-type vehicle detectors to produce estimates of the aggregate variables, which are then in turn used as the input variables to the incident detection algorithms. Simulation results using a microscopic simulation of a two-lane freeway indicate that: (1) our algorithms are robust to the differences between the dynamics of actual traffic and the aggregated dynamics used to design the detection systems; and (2) our methods appear to work as well as existing algorithms in heavy traffic conditions and work better in moderate to light traffic. Areas for future work are outlined at the end of the paper.

---

* This work was supported by the U.S. Department of Transportation under the OST Program of University Research, Contract No. DOT-OS-60137.

## I. Motivation and Overview

The problem addressed in this paper is the development of a systematic approach to the detection of freeway incidents (accidents, stalled cars, debris on the road, etc.). The goal of our work was the design of algorithms that: (1) directly use data from conventional presence detectors which provide binary information at each point in time, indicating the presence or absence of a vehicle directly over the detector; and (2) minimize human operator requirements in detection, classification, and isolation of incident events. The consideration of this problem is of obvious importance both for the efficient dispatching of emergency services and for the design of advanced technology traffic control systems, which require accurate knowledge of existing traffic conditions in order to provide effective on-line control decisions.

Our work represents a new approach to the traffic incidents detection problem in that we have based our analysis on a dynamic model that describes the temporal evolution of key traffic variables (flows, densities, velocities) representing aggregate traffic conditions over links of the freeway. Having such a model, we can then bring into play a variety of modern system-theoretic procedures for estimation, detection, and identification.

Prior to our investigations, several researchers [1-8] had studied the problem of reliable incident detection on freeways and had developed a number of automatic detection systems. All of these techniques directly utilize information available from presence detectors. While many of

these algorithms take into account the temporal evolution and temporal or spatial correlation of observables derived from detector data, none of these techniques involve the systematic utilization of nonlinear differential equations that relate key traffic variables. In the most comprehensive study of incident detection systems of this type [8], Payne, et.al., have indicated that these techniques have false alarm problems when traffic compression waves occur. Intuitively, the use of dynamic models that capture such phenomena should help to alleviate this problem. In addition, we have found that previously developed algorithms do not do well in detection capacity-reducing incidents in light or moderate traffic. Again the use of dynamics should be of use in extracting information concerning such incidents in which the direct effect on the observables may not be dramatic.

Motivated by the preceding observations and by the successful studies of dynamic models for traffic behavior [9,40] and of freeway traffic control based on such models [10-13], we have considered the problem of incident detection based on the model proposed by Payne [9,40], which is reviewed in Section II. Several key issues arise immediately when one considers such an approach:

(a) The incident detection techniques in [1-8] are, in general, extremely simple to implement and are directly amenable to decentralized implementation. Thus, any increase in complexity that may arise in implementing a model-based algorithm must be justified by an accompanying improvement in overall system performance.

(b) The incident detection algorithms in [1-8] operate directly on quantities available from the output of loop detectors. On the other hand, the Payne model [40] describes the evolution of aggregate variables. Hence, in order to implement the model-based incident detection algorithms, we must develop a method for estimating the aggregate variables when provided with loop detector data.

(c) The aggregate or macroscopic dynamic model in [9,40] clearly represents an approximation to real traffic behavior. Thus, any incident detection algorithm based on this model must be insensitive to the discrepancies between actual traffic behavior and that predicted by the dynamic model.

With these issues in mind, we have developed incident detection systems using two different dynamic model-based hypothesis testing techniques, the multiple model (MM) and generalized likelihood ratio (GLR) algorithms. The MM and GLR methods are briefly reviewed in Section III and IV, and the testing of these algorithms using direct measurements of the aggregate variables is discussed in Section V.

The next step in our study concerned the problem of using presence detector data to produce estimates of the aggregate variables needed as inputs for our detection algorithms. The nature of presence detector data is discussed in Section VI, and in Section VII we develop an extremely simple system for the estimation of traffic variables (specifically density and flow) from presence detector data. We feel that

this system is of interest in itself. In Section VIII we combine this system with our detection algorithms and describe the results of microscopic simulations in which presence detector data were generated and used.

The results described in this paper indicate that dynamic model-based detection algorithms do offer the potential for performance improvements over existing algorithms. A number of questions remain to be examined, perhaps the most critical of which is a precise assessment of how much improvement is possible at what cost in terms of increased detection system complexity. Issues such as this and others related to the implementation of MM- or GLR - based incident detection algorithms are presented in Section IX. Because of limitations on space, some of the details of our work have been omitted. The interested reader may find them in references [22-25].

## II. The Aggregate Traffic Model

The detection algorithms we have developed are based on the equations proposed by Payne [9,40] for the dynamics of freeway traffic flow. This model captures basic aspects of both the fluid flow and car-following models of traffic dynamics. The variables in the dynamic model are spatial mean velocities ($v$, in miles per hour), densities ($\rho$, in cars/mile/lane), and flows ($\phi$, in cars/hour/lane) over links of the freeway between presence detector locations. This yields a spatially discretized set of coupled equations.

$$\frac{d\rho_i}{dt} = \frac{\phi_{i-1}-\phi_i}{\delta x_i} \tag{2.1}$$

$$\frac{dv_i}{dt} = -\frac{v_i(v_i-v_{i-1})}{\frac{1}{2}(\delta x_i+\delta x_{i-1})} + \frac{v_i^e(\rho_i)-v_i+\omega_i}{T} - \frac{\nu}{T}\frac{1}{\rho_i}\left[\frac{\rho_{i+1}-\rho_i}{\frac{1}{2}(\delta x_i+\delta x_{i+1})}\right] \tag{2.2}$$

where $\phi_i = v_i\rho_i$. Here subscripts are used to denote the link number with which each variables is associated, $\delta x_i$ is the length of link i, $\omega_i$ represents acceleration noise (used to model normal variations in velocities due to the statistical behavior of individual drivers), and $\nu$ and T are parameters introduced by Isaksen and Payne to model driver response characteristics. The $v^e(\rho)$ term represents the driver's desired equilibrium speed as a function of the density of traffic. A number of shapes for this curve have been proposed, and we have used a form which yields the correct properties at high and low density and yields reasonable maximum capacities. Our techniques could be easily adapted to any other choice for the $v^e$-curve. The general form of our $v^e$-curve depicted in Figure 2.1, is determined by 3 free parameters; $v_{free}$, the equilibrium velocity under light traffic conditions; $\rho_{free}$, the density at which the equilibrium velocity begins to decrease; and $\rho_{jam}$, the maximum density of cars that the freeway can hold. The curve between $\rho_{Free}$ and $\rho_{jam}$ is logarithmic

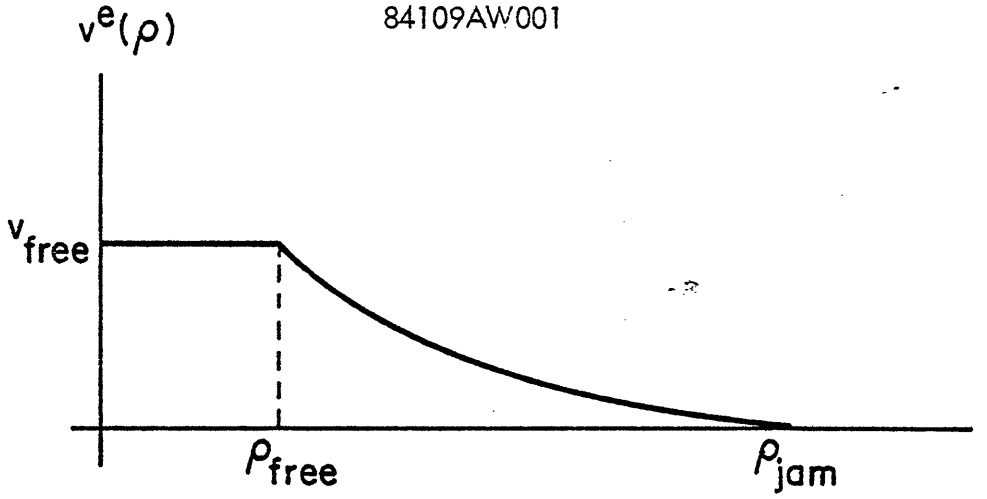


Figure 2.1: THE $v^e$-curve used.

$$v^e(\rho) = \frac{v_{free} \ell n[\rho_{jam}/\rho]}{\ell n[\rho_{jam}/\rho_{free}]} \tag{2.3}$$

$$\rho_{free} \leq \rho \leq \rho_{jam}$$

As we will discuss in a moment, the parameters of the $v^e$-curve can be adjusted on each link (thus the notation $v_i^e$ in (2.2)) to reflect freeway conditions on that link.

Assuming we are modelling M links of the freeway (i=1 is the upstream link, i=M is the downstream link), we must modify the equations (2.1), (2.2) on the end links to account for boundary conditions. Specifically, on link 1 we use the equations

$$\frac{d\rho_1}{dt} = \frac{\text{Flow} - v_1\rho_1}{\delta x_1} \tag{2.4}$$

$$\frac{dv_1}{dt} = \frac{v^e(\rho_1)-v_1}{T} - \frac{\nu}{T}\frac{1}{\rho_1}\left[\frac{\rho_2-\rho_1}{\frac{1}{2}(\delta x_1+\delta x_2)}\right] \tag{2.5}$$

For our simulations, "Flow" was assumed to be a Poisson arrival process with a specified mean value, which was used to control the overall level of traffic. For link M, we assume a zero density gradient across the last boundary, leading to the equation

$$\frac{dv_M}{dt} = \frac{v_M(v_M-v_{M-1})}{\frac{1}{2}(\delta x_{M-1}+\delta x_M)} + \frac{v^e(\rho_M)-v_M + \omega_M}{T} \tag{2.6}$$

Links 1 and M essentially establish the boundary conditions, and thus our primary concern is with results on the M-2 internal links.

Note that the specification of $v^e(\rho)$ implicitly defines the capacity as the maximum allowable steady-state flow on the freeway. Specifically, using the definition

$$\phi^e(\rho) = \rho v^e(\rho) \tag{2.7}$$

we obtain the "fundamental diagram of traffic" depicted in Figure 2.2. From this curve, we see that by adjusting $v_{free}$ or $\rho_{free}$ we can parametrize the capacity (as defined above) on each link of the freeway. Specifically, some algebra yields capacity as a function of the parameters:

$$\text{Capcity} = \frac{\rho_{jam} v_{free}}{e \, \ell n(\rho_{free}/\rho_{jam})}$$

Having this basic model, one can then consider the modelling of abrupt changes in the model that correspond to particular incidents or other inhomogeneities and problems that one wishes to distinguish from capacity-reducing events. Specifically, we have modeled three types of events for each link of the freeway:

(a) A capacity reducing incident on link i. This was modeled as a decrease in the size of $\rho_{free}$ on the ith link, leading to the appropriate decrease in capacity (e.g., a one-third reduction for loss of one lane on a three-lane freeway).

(b) A pulse of traffic, lasting for a specified duration, entering link i. This model was included for two reasons. First, one may want

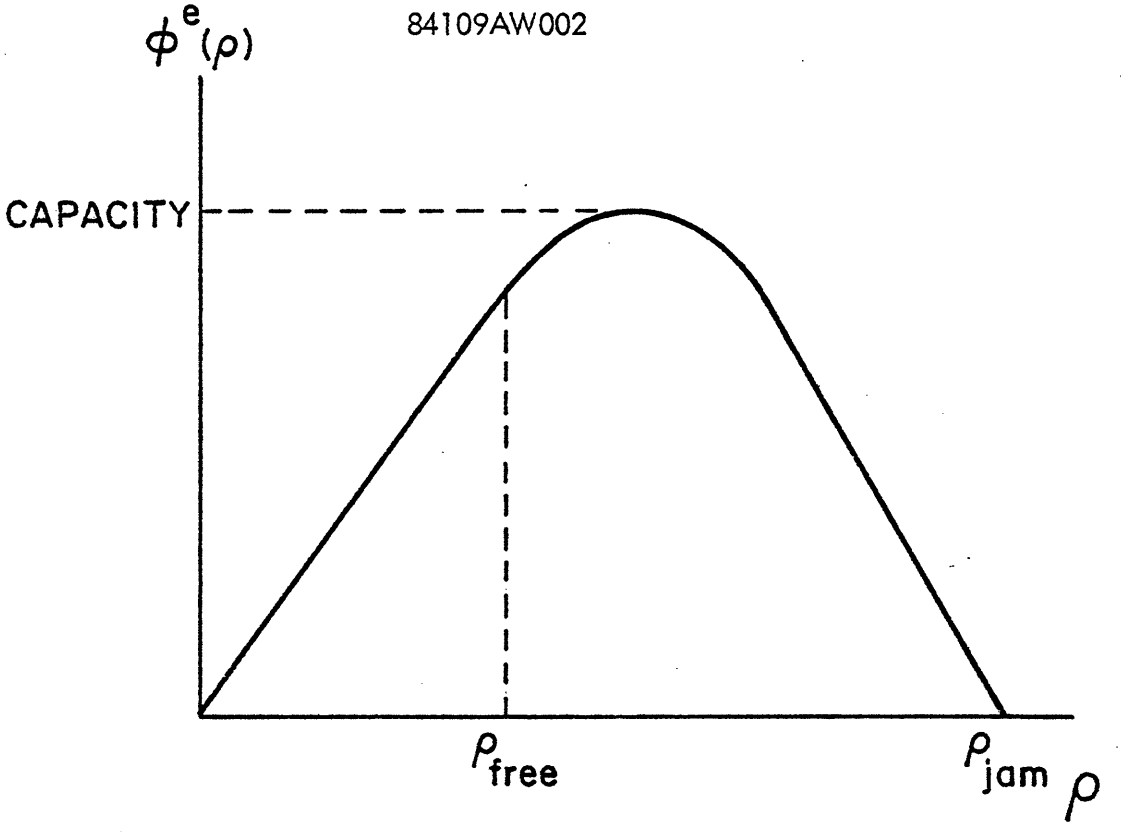


Figure 2.2 THE FUNDAMENTAL DIAGRAM OF TRAFFIC.

to detect large disturbances caused by, say, a sporting event letting out, in order to adjust a freeway control algorithm. Secondly, and most importantly, normal random fluctuations of traffic are a possible cause of false alarms for an incident detection system (e.g., the compression wave problems discussed in Section I), and our inclusion of this pulse model was based on a desire to determine if our detection methods were capable of distinguishing such events from real incidents.

(c) Sensor failures. As mentioned in Section I, the macroscopic-model-based detection systems we have developed assume that one has measurements of $\rho_i$ and $v_i$ on each link. At this level, a sensor failure is defined as any condition such that the sensed density or velocity differs systematically from the actual value of the variable being measured. We will say more about sensor failure models when we discuss the MM and GLR methods in the next two sections.

For our simulation studies, we have considered a 6-link freeway, where, under normal conditions, each link was assumed to have the same number of lanes. The parameters $\rho_{free}$ and $v_{free}$ were set at 23.1 cars/mile/lane and 55 miles per hour which corresponds to a capacity of 2000 cars/hour/lane. The variable FLOW was taken as a mean flow (which was varied in our study) plus a zero mean fluctuation of variance 50 $(\text{cars/hour/lane})^2$. In our initial macroscopic simulation of the MM and GLR detection algorithms we used a reduction of capacity by 1/3

(from 2000 to 1333 cars/hour/lane) to model an incident causing the loss of one lane an a three-lane highway. This corresponds to a reduction in $\rho_{Free}$ to 7.4. However, our later tests with microscopic simulations led us to modify this model, since an incident in one lane does effect traffic in other lanes through the lateral lane switching of vehicles away from the lane on which the incident occurred. The value of $\rho_{free}$ finally chosen was .03 cars/mile/lane. This value translates into a capacity of 500 cars/mile/lane -- a reduction of 75% of capacity.

For the pulse of traffic model on link i, a valued of 1600 cars/hour/lane was used as an input flow to the density equation on link i. Finally, we have assumed that our measurements of $\rho_i$ and $v_i$ are corrupted by white noise. Initially for the macroscopic studies, the standard deviations of these noises were taken to be 3.33 cars/mile/lane and 5 m.p.h., respectively. Later microsimulations, directly using presence detector data, led to larger values for these noises: 10 cars/mile/lane and 28 m.p.h. (see Section VII).

Finally, both the MM and GLR methods require the use of state estimation techniques. As the dynamics of traffic (2.1), (2.2) are highly nonlinear, an approximate filtering technique has been used. Let our state be denoted by x(t):

$$x'(t) = [\rho_1(t), v_1(t), \ldots, \rho_M(t), v_M(t)] \tag{2.8}$$

and our measurement vector by $z(t_k)$

$$z(t_k) = x(t_k) + N(t_k) \tag{2.9}$$

where the components of N are assumed to be independent white noise processes.

The estimation technique we have used is the nominal linearized Kalman filter. Let $\hat{x}(t|t_k)$ be our estimate of $x(t)$ given $Z_k \triangleq \{z(t_1),\ldots,z(t_k)\}$. Then, given $\hat{x}(t_k|t_k)$, we predict ahead to obtain $\hat{x}(t_{k+1}|t_k)$ by integrating (2.1), (2.2) assuming zero noise and an initial condition of $\hat{x}(t_k|t_k)$. The new measurement $z(t_{k+1})$ is then incorporated according to

$$\hat{x}(t_{k+1}|t_{k+1}) = \hat{x}(t_{k+1}|t_k) + H[z(t_{k+1}) - \hat{x}(t_{k+1}|t_k)] \tag{2.10}$$

where the gain H is determined off-line by solving for the Kalman filter gain for the traffic model linearized about some equilibrium mean flow-density-velocity point. This linearization has been done about many operating point. Comparing the Kalman gains from a wide variety of such points revealed very little variation with flow. Thus, it appears that one set of gains adequately handles most levels of flow. It should be mentioned that little work has been done under congested conditions and a second set of gains may be needed under this scenario. However, in spite of using only a single set of gains good performance has resulted under widely varying traffic flow conditions.

## III. The Multiple Model Method

The Multiple Model (MM) method for system identification has been considered by several researchers, and we refer the reader to [26-28]

and the references cited therein for a detailed development of the technique. The method addresses the problem of identifying a linear-Gaussian system

$$\dot{x}(t) = Ax(t) + w(t) \tag{3.1}$$

$$z(t_k) = Cx(t_k) + v(t_k) \tag{3.2}$$

given the measurements $Z_k$ and a set of hypothesized models (i=1,...,N)

$$\dot{x}_i(t) = A_i x_i(t) + w_i(t) \tag{3.3}$$

$$z(t_k) = C_i x_i(t_k) + v_i(t_k) \tag{3.4}$$

The output of the MM method is the set $p_i(t_k)$ of conditional probabilities for the validity of each of the models given $Z_k$. A Kalman filter is implemented for each of the N models and the measurement residuals

$$\gamma_i(t_{k+1}) = z(t_{k+1}) - C_i \hat{x}_i(t_{k+1}|t_k) \tag{3.5}$$

from each filter are used to update the $p_i(t_k)$ according to the equation

$$p_i(t_{k+1}) = \frac{F_i(\gamma_i(t_{k+1}))p_i(t_k)}{\sum_{j=1}^{N} F_j(\gamma_j(t_{k+1}))p_j(t_k)} \tag{3.6}$$

where $F_i$ is the probability density for $\gamma_i(t_{k+1})$ assuming the ith model is correct. If hypothesis i is true, then $\gamma_i$ is a white, zero mean, Gaussian sequence with covariance

$$E[\gamma_i(t_k)\gamma_i(t_k)'] = V_i(t_k) \tag{3.7}$$

which can be determined off-line as part of the Kalman filter calculations (see [16]). Thus

$$F_i(\gamma_i(t_{k+1})) = \frac{\exp\{-\frac{1}{2}\ \gamma_i(t_{k+1})'V_i^{-1}(t_{k+1})\gamma_i(t_{k+1})\}}{[(2\pi)^m \det(V_i(t_{k+1}))]^{1/2}} \tag{3.8}$$

where m=dim $\gamma_i$. Essentially, the MM method consists of a set of one-step predictors, each based on a different model for the observed data. The probabilities, which are functions of the prediction errors $\gamma_i$, i=1,...,N, reflect the relative performance of each of these predictors.

The MM method has been adapted for use with the Payne model as described in the preceding section. A number of comments need to be made about this design and about the MM method in general:

(1) We have implemented a nominal-linearized Kalman filter* for each of our hypotheses:

- For the normal model, the dynamics (2.1),(2.2) used the normal $v^e$-curve on each link.
- For the model representing an incident on link i, the dynamics (2.1),(2.2) are modified by replacing the normal $v^e$-curve on link i with the reduced capacity curve.

---

* From now on we will call all our filters "Kalman filters." It should be understood that they are all nominal linearized Kalman filters.

. For the model representing a pulse of traffic on link i, the dynamics (2.1),(2.2) are modified by including an input flow in the equation for $\dot{\rho}_i$.

Theoretically, one set of Kalman gains is needed for each hypothesis. However, although there are differences among the various models, the Kalman gains are very similar. Thus, the same Kalman gain has been used for the Kalman filters for each of these models.

(2) In addition to the above, there are also a set of models and associated filters representing sensor failures. We have modeled a failure in our ability to measure a particular state variable by modifying the measurement equation (2.9):

$$z(t_k) = Cx(t_k) + N(t_k) \tag{3.8}$$

where C is diagonal, with 1's along the diagonal except for a zero in the location corresponding to the particular state measurement which is hypothesized to be faulty. Note that (3.8) corresponds to modeling a failure as a measurement that contains only noise and is uncorrelated with the state. For each of these models, we have used the same filter gains as for the normal model except for zeroing the gains on the measurement hypothesized to be faulty.

(3) The residuals from the Kalman filters are used together with (3.6) to compute the probabilities for each hypothesis. Note that (3.6) was derived assuming that

(a) the actual system and all of the hypotheses are linear-Gaussian;

(b) one of the hypotheses matches the true system;

and

(c) the true system does not switch from one hypothesis to another (corresponding, for example, to the onset of an incident).

None of these assumptions is valid, and thus some comments are in order. Assumption (a) essentially addresses the problem of the utility of the nominal-linearized Kalman filter -- i.e. assuming the dynamic model is correct, is it valid to postulate that the filter residuals will be zero-mean, white, with precomputed covariance? The second assumption implies that (under assumption (a)) the residuals from one of the filters will be white and zero mean. In practice this is never precisely the case, but our experience has been that neither of these assumption has caused great problems. A number of explanations can be given to account for this, but there are no general results that predict when these filters will work well. Based on our experience it is our feeling, however, that, while the estimates from the filters may be sensitive to linearization and model uncertainties, the performance of a detector based on the residuals from these filters may be far more robust. Basically, a discrete decision process should work well, as long as the models for the several hypotheses are sufficiently different. When this is true, the effect on the observables of the differences in the models will be greater than the effect of the approximations introduced in applying the MM approach. Intuitively, this can be thought of as a signal-to-noise ratio problem, where the effects of the assumptions add uncertainty. In this

sense, assumptions (a) and (b) will limit the minimum size incident that can be detected, where size is to be interpreted as the magnitude of the effect of the incident on the dynamics. For example, we may be able to detect a stalled car, which causes severe and localized capacity reduction, but the smaller effect caused, say, by debris on the road may not be detectable. Also, as we will see, the effect of an incident increases in magnitude as the level of traffic increases. Thus, one might expect these to be a minimum flow level, such that it is impossible to detect incidents in traffic lighther than that level. We will see precisely this effect in our simulations.

It is worth noting that assumption (b) can lead to difficulties if none of the hypothesized models are near the true dynamics. We will point out several problems of this type which lead to particular issues that must be considered in modifying our MM system to obtain improved performance.

Assumption (c) can lead to difficulties in the ability of MM to detect incidents as they occur -- i.e. before the occurrence of an incident on link i, the probability for this hypothesis may become so small that the system will not be able to respond quickly after the incident has occurred. (this is the obliviousness problem discussed in [30]). In order to overcome this problem, one could implement a MM system that included models that switched from normal traffic dynamics to the various incident conditions, but the problem with this is that one would have to implement a growing bank of filters to account for

the different times at which the switch could occur. Therefore we have not adopted this approach. Rather, the remedy employed in our work is a relatively common one -- a lower bound is set on any probability (we have used .01). As we will see, this leads to good response characteristics. We note also that the Kalman filter based on a pulse of traffic on link i is unstable if no such pulse is there (the Kalman filter has a constant driving term in the $\dot{\rho}_i$ equation not present in the true system). Thus, if such a pulse were to develop at some point in time, the filter estimate for this hypothesis might already be so much in error that the MM system might not detect the pulse. To overcome this, whenever the probability of a pulse model falls below .05, the estimate produced by this filter is reset to the estimate for the most probable model.

## IV. The Generalized Likelihood Ratio Method

The Generalized Likelihood Ratio (GLR) method for detecting abrupt changes in dynamic systemsis described in [17] for a specific case and in [30] for a larger class of problems. The basic idea is the following. We assume that a dynamic system under normal conditions is described by the model

$$\dot{x}(t) = Ax(t) + w(t) \tag{4.1}$$

$$z(t_k) = Cx(t_k) + v(t_k) \tag{4.2}$$

A Kalman filter based on this model is implemented. We then hypothesize that an abrupt change (the $\ell$th of, say, N possible abrupt changes) in the system occurs at time $\theta$ and that this change can modeled by an additive term in (4.1) or (4.2). In this case linearity yields the following model for the residuals of the normal model filter

$$\gamma(t) = \alpha g_{\ell}(t,\theta) + \tilde{\gamma}(t) \tag{4.3}$$

where $\tilde{\gamma}(t)$ is the normal zero-mean, white residual and $g_{\ell}(t,\theta)$ is the precomputable deterministic signature describing the bias induced in $\gamma$ at time t by a type $\ell$ abrupt change occurring at time $\theta$. The parameter $\alpha$ is an unknown scalar magnitude for the abrupt change (e.g., the size of a bias in a sensor, the effect of a stalled car on the residuals, which, as we've stated, may depend on the flow level).

Given the model (4.3) for each of the N hypotheses, we compute a set of correlations of the actual filter residuals with the various signatures

$$d_i(t_k,\theta) = \sum_{m=1}^{k} g_i'(t_m,\theta) V^{-1}(t_m) \gamma(t_m) \tag{4.4}$$

The generalized log-likelihood ratio for a type i incident occurring at time $\theta$ then is

$$\ell_i(t_k,\theta) = \frac{d_i^2(t_k,\theta)}{s_i(t_k,\theta)} \tag{4.5}$$

where

$$S_i(t_k,\theta) = \sum_{m=1}^{k} g_i'(t_m,\theta)V^{-1}(t_m)g(t_m,\theta)$$

which can be precomputed. If we define

$$\ell_i(t_k) = \max_{\theta} \ell_i(t_k,\theta) \tag{4.6}$$

we then have a measure of the likelihood that a type i incident has occurred sometime in the past. A variety of decision rules can then be devised based on these quantities (see, for example, [25,31]).

The basic idea behind the GLR method is that distinct incident modes lead to distinct, systematic trends in the prediction errors produced by the Kalman filter. The GLR algorithm compares the observed residuals to the precomputed signatures for these systematic trends and from this comparison produces measures of the likelihood that each hypothesized incident type has occurred.

The GLR algorithm as described above has been adapted for use with the Payne model. Again a number of comments are in order:

(1) In this case we have implemented a single nominal-linearized Kalman filter based on the normal dynamics (2.1), (2.2).

(2) Since the true system and the Kalman filter are nonlinear, in principle the decomposition of the filter residuals as in (4.3) is not valid. However, we assume a decomposition of this form and thus must calculate the incident signatures, which correspond to

to the deterministic response to an incident of the true system-normal mode filter combination. The nonlinearity of the system and filter necessitated the computation of these signatures via simulations: the macroscopic model and normal model Kalman filter were simulated without any stochastic effects in the dynamics or measurements, and, for each incident type, the model (2.1),(2.2) was chosen to correspond to the particular incident. (e.g., a reduced capacity $v^e$-curve on link i for the link i incident hypothesis). The resulting filter residuals constituted the signature for that incident type. Note that this approach would yield the correct signatures in the linear case. The models used for capacity reducing incident and traffic pulses were the same as those used in the MM method. Sensor failures, however, were modeled by the development of a bias in one component of z

$$z(t_k) = x(t_k) + N(t_k) + \alpha e_i \sigma(t_k,\theta) \tag{4.7}$$

where $\alpha$ is the unknown bias size, $\theta$ is the time at which the failure occurs, $e_i$ is the ith standard basis vector, and $\sigma(t,\theta)$ is the unit step (=0 for $t<\theta$, =1 for $t\geq\theta$).

(3) As with the MM algorithm, the GLR system is based on several assumptions that do not hold in the application to incident detection. We have already commented on the effect of nonlinearities on the decomposition of the residuals into a signature component and a white noise part. Similarly, modelling errors imply that none

of the hypotheses are precisely correct. As we discussed in the preceding section, the issue then becomes one of signal-to-noise: is the effect of the incident on the observables significantly larger than the effect of the approximations and uncertainties.

(4) Note that, unlike the MM algorithm, the GLR method explicitly considers the shifting of the system from normal to incident conditions at an unknown time. Thus, the obliviousness problem of the MM method is not encountered here. The price one pays for this, however, is the calculation of (4.4) for a number of hypothesized times $\theta$. In principle we should calculate $\ell_i(t_k,\theta)$ for $\theta=t_1,\ldots,t_k$ -- a growing computational load. We have employed a standard method for overcoming this -- we compute $\ell_i$ only for a "sliding window" of the most recent past $t_{k-M} \leq \theta \leq t_k$. With a sampling time $t_k - t_{k-1} = 5$ seconds, we have kept a 250 second window, (51 points).

## V. Simulations of the Macroscopic Detection Systems

The first two sets of tests of the GLR and MM systems were designed to determine the performance characteristics of these methods and their robustness in the presence of both modeling errors and the effects of the linearizations involved in their design. For each of these sets of tests, the GLR and MM systems were designed about a single, fixed operating

point with fixed values for all parameters, such as the assumed measurement noise covariances, the postulated effects of different incidents on traffic dynamics, etc. The parameter values used were those given in Section II. In addition, the linearized Kalman filters were designed about a high mean flow operating point of 1667 cars/hour/lane.*

The first set of tests involved the use of data obtained from simulating the model (2.1), (2.2) using parameter values that differed from those assumed in the MM and GLR designs. The parameters that were varied were

- The actual mean flow onto link 1. This was varied from a low flow of 900 cars/hour/lane up to capacity (2000 cars/hour/lane), a very large range.
- The sensor noise variances. Significantly larger values for these variances were used in some of the experiments.
- The initial estimation error. Large values were used for this in order to observe the transient behavior of the algorithms.

The performance of both systems was encouraging:

- Detection performance was uniformly good over the entire range of actual mean flows used (900 to 2000 cars/hour/lane). No false alarms were observed, no incorrect detections (e.g. declaring a pulse on link 3 when the true event was an incident on link 4) occurred, and the response time of the systems was small. Figures 5.1, 5.2 illustrate typical

---

* Note that this mean flow not only affects the gain, but it also is a driving term in the prediction step of the filters, as it enters in as a driving term in the equation for $\dot{\rho}_1$.

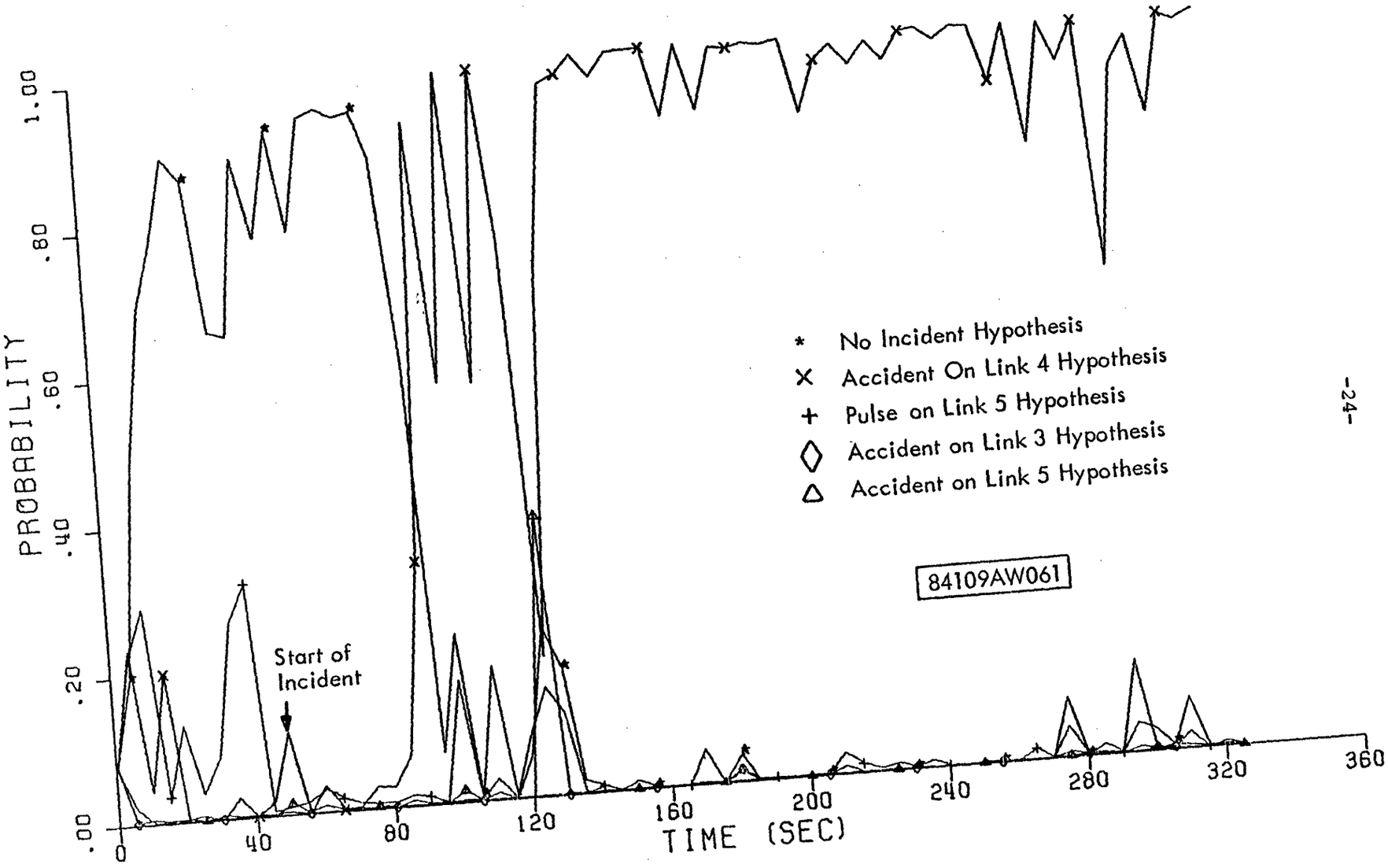


Figure 5.1: MM Probability Plot - Accident on Link 4.

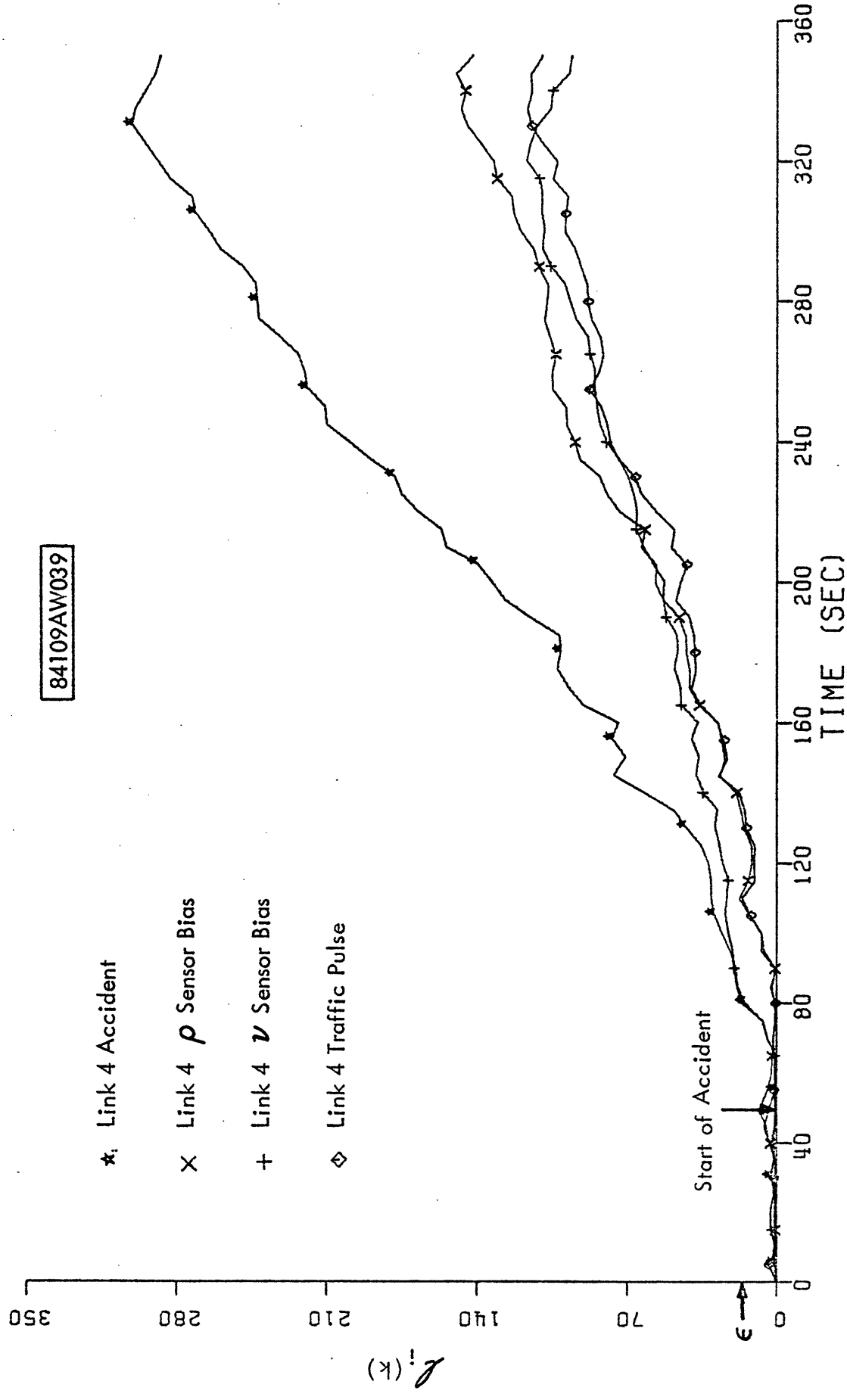


Figure 5.2: GLR, $\ell_i(k)$: Link 4 Accident.

performance of the GLR and MM systems. The value of $\varepsilon$ indicated in Figure 5.2 and in later figures is 16. Under all of the assumptions used in the derivation of the GLR method (such as the whiteness of the residuals), this threshold implies a false alarm probability at any instant of time of less than .0002.

- Performance is somewhat degraded when the actual measurement variances are a factor of 16 larger than nominal. No false alarms were observed, and all incidents were correctly identified with, however, an increased detection delay (compare Figures 5.3 and 5.4 with 5.1 and 5.2).

- Large initial estimation errors cause only transient effects on GLR and MM. Performance is excellent after the initial start-up.

These tests, while indicating a certain level of robustness of the GLR and MM systems, do not provide information about system robustness to the details of the Payne model. To provide this type of information, we have used a microscopic traffic simulation [15]. This program is based on the St. John car-following equations [14] and can be used to simulate traffic under almost any conditions. The program simulates two lanes of flow, operates in discrete time and offers the following features:

(a) a variety of vehicle and driver types can be modelled

(b) presence detectors can be placed as desired

(c) on-ramp and input flow rates can be specified

(d) accidents can be simulated by stopping a vehicle at any desired time and location.

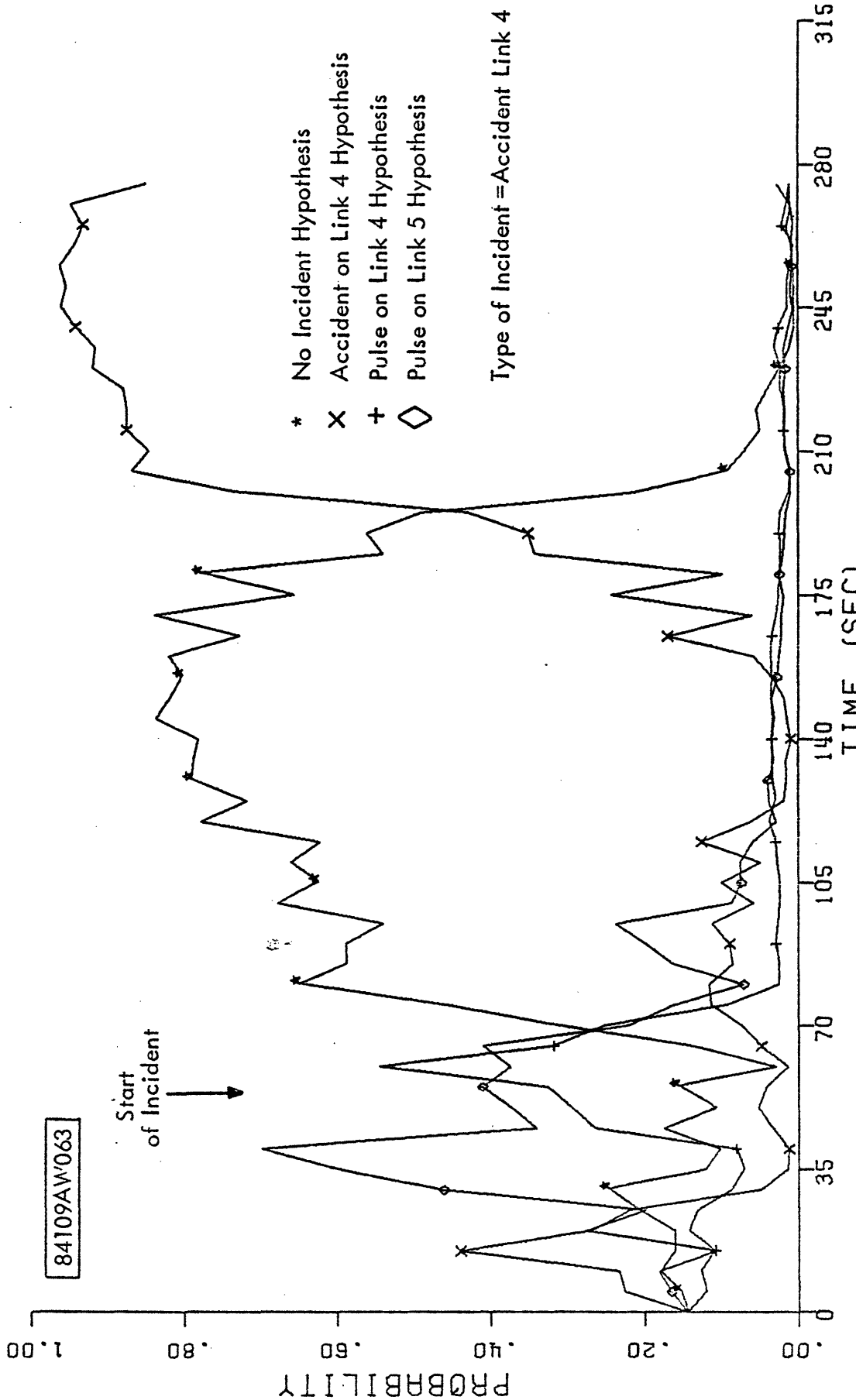


Figure 5.3: MM, Probability Plot - Increased Observation Noise.

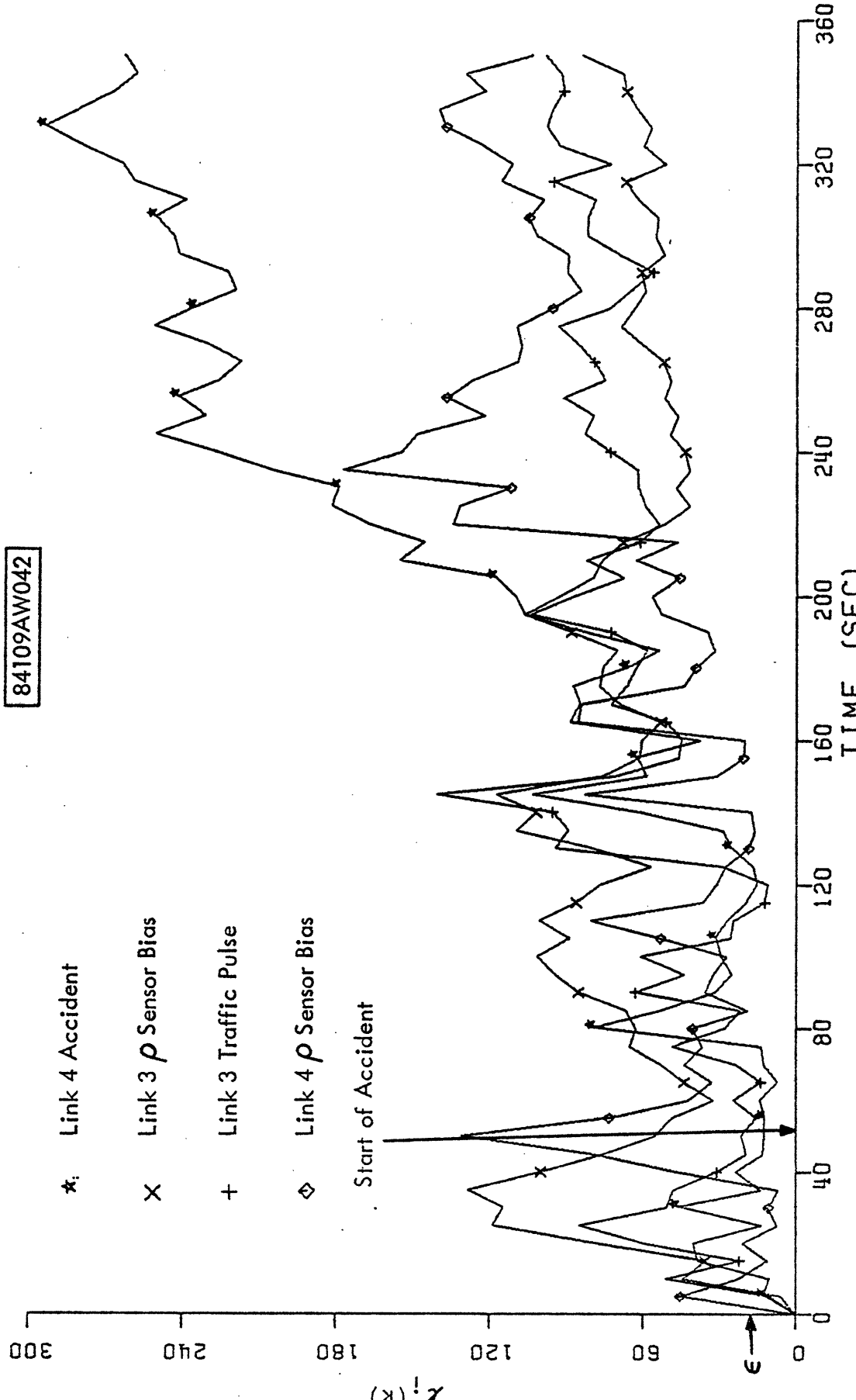


Figure 5.4: GLR, $\ell_i(k)$: Link 4 Accident with Increased Sensor Noise.

Because the simulation is based on a microscopic model, the position, speed, acceleration, driver type and vehicle type for each vehicle on the road are known and available in the program. Thus, it is possible to compute the density and average velocity of traffic over links of the simulated freeway. We have done this in order to generate measurements (which we have then corrupted with noise) of the aggregate variables used in the GLR and MM algorithms. These data were then fed into the detection systems. The following are the conclusions that can be drawn from this study:

- Both the GLR and MM systems performed well in detecting incidents down to flow levels of 900 cars/hour/lane. This gives us an idea of the fundamental limitations of our algorithms -- at flows less than this the various noise sources and approximations are stronger than the "signal" due to the incident.

- The GLR approach has some difficulties in distinguishing incidents from sensor biases. We will discuss this problem in Section IX. See Figures 5.5, 5.6 for typical MM and GLR responses.

- Short-term spatial inhomogeneities in traffic cause transient responses in the GLR and MM systems. Figures 5.7, 5.8 are for a simulation in which two slowly moving vehicles disrupt traffic flow. As expected, this looks like a "traveling incident". This could be alleviated by a persistence requirement on the probabilities or log-likelihood or by incorporating a traveling incident model for MM and signature for GLR.

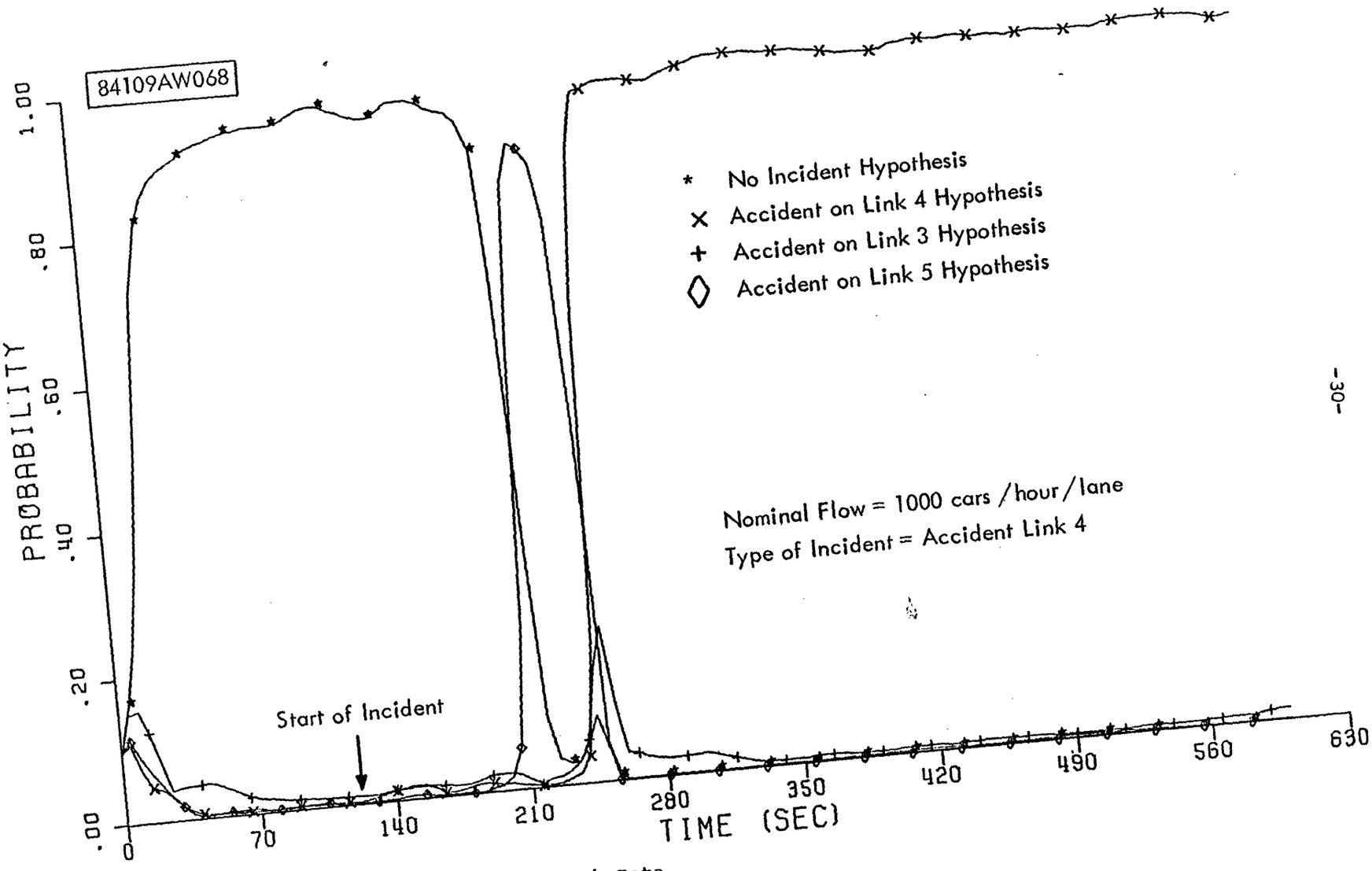


Figure 5.5: MM, Probability Plot - Link Data.

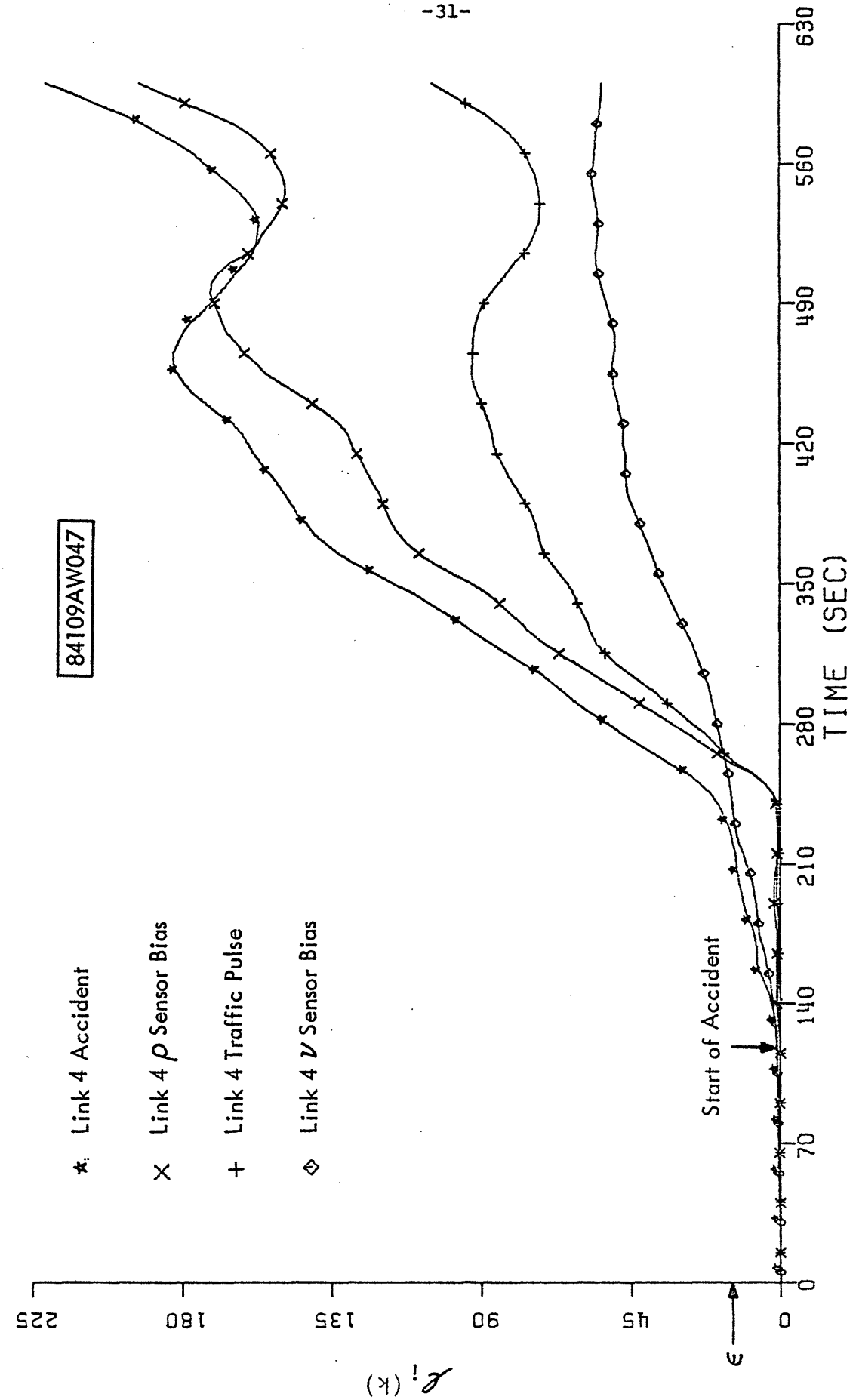


Figure 5.6: GLR, $\ell_i(k)$: Accident on Link 4.

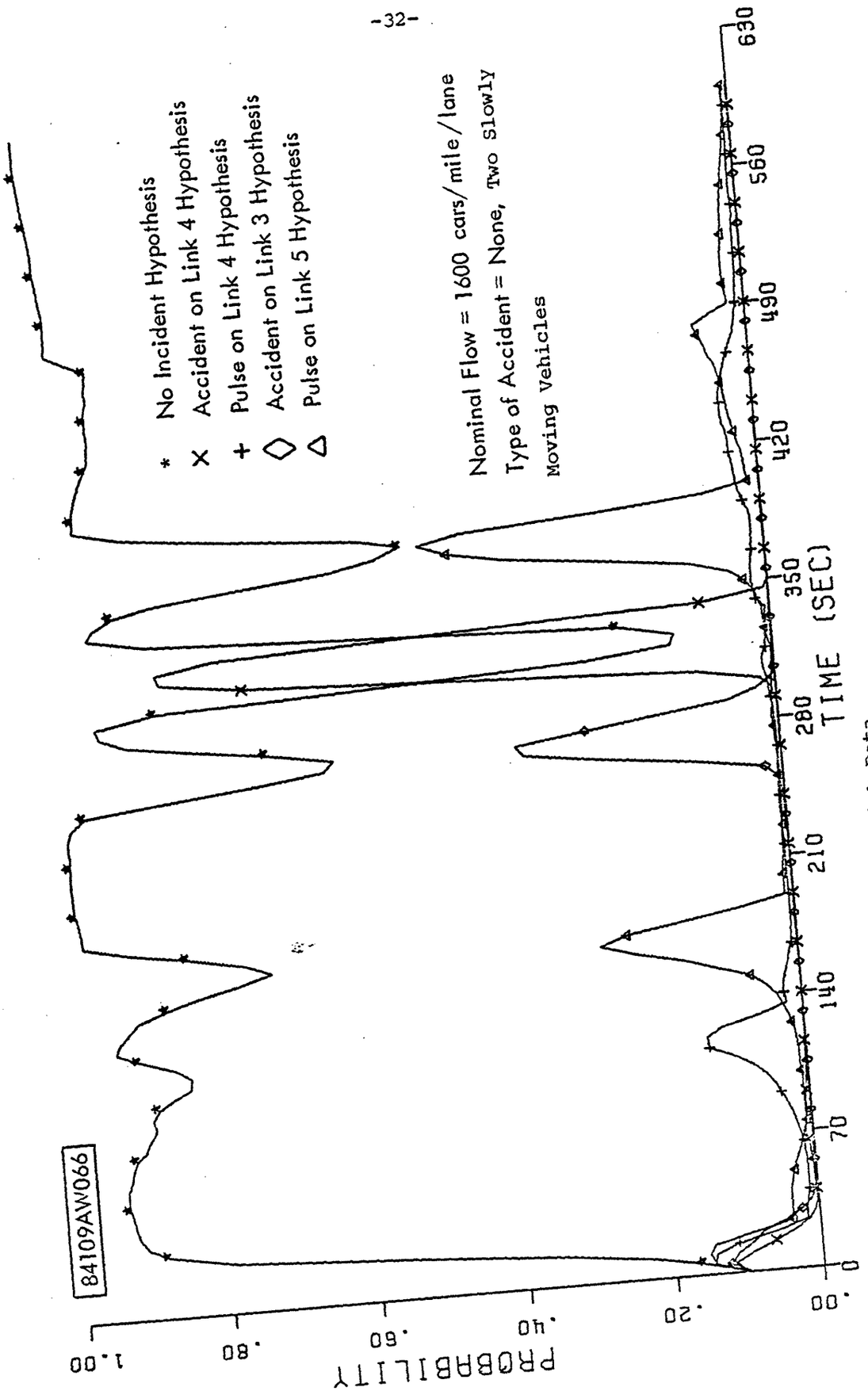


Figure 5.7: MM, Probability Plot - Link Data.

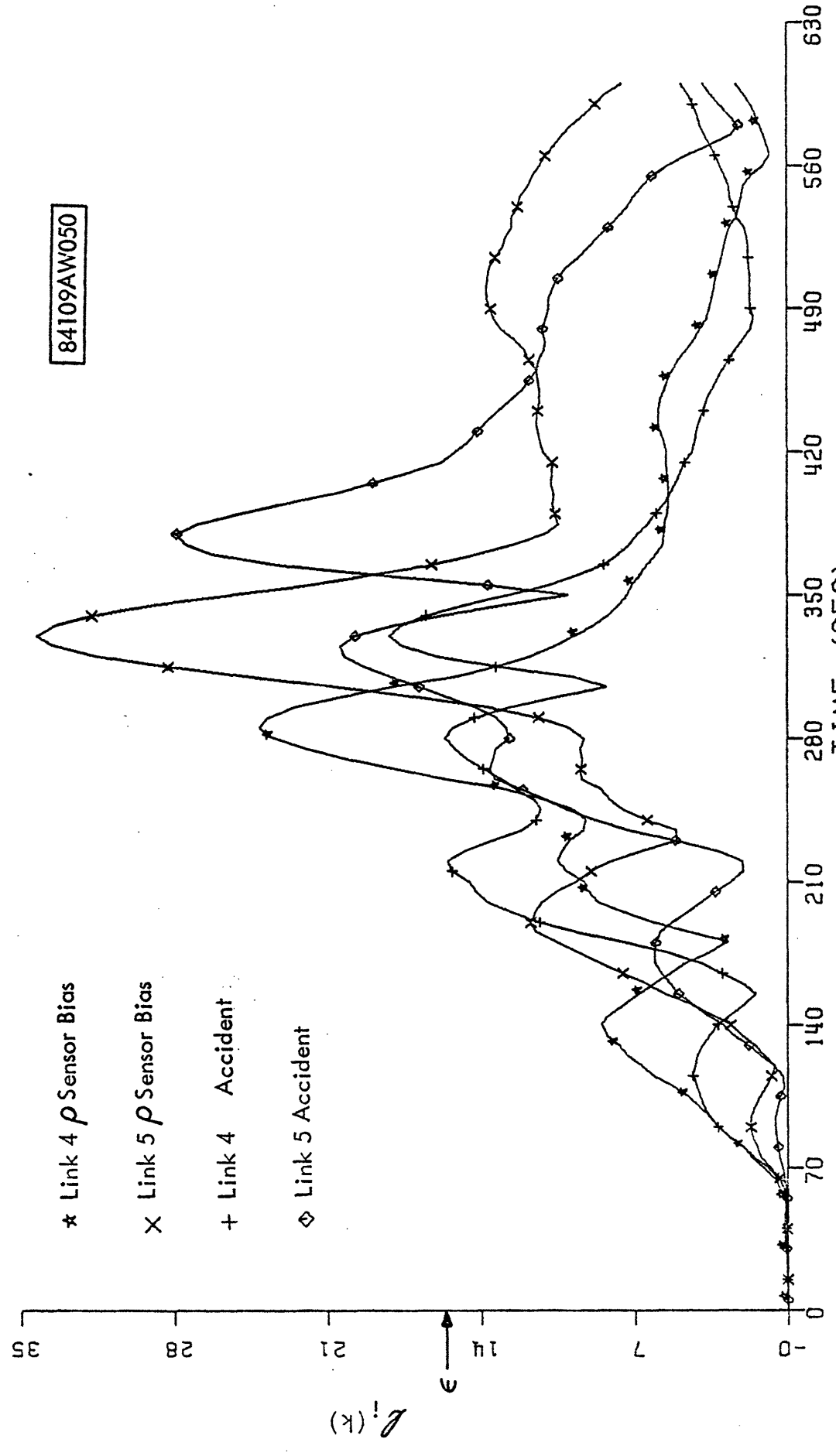


Figure 5.8: GLR, $\ell_i(k)$: Link Data, No Accident, Two Slowly Moving Vehicles.

Thus, while some questions are raised by these simulations, the basic conclusion of this study is that the MM and GLR algorithms appear to be insensitive to the details of the dynamic models used in their design and to the precise parameter values used. On the other hand, the data used in these simulations is not realistic: in the first set we used a macroscopic simulation, while in the second we used a microscopic simulation and actually computed the aggregate variable measurements needed by counting cars on each link and averaging their velocities. What is still missing is a system for taking presence detector data and producing estimates of $\rho_i$ and $v_i$ that can then be used as inputs to the GLR and MM systems. This is the topic of the next few sections.

## VI. The Nature of Presence Detector Data

In this section we briefly discuss the type of information contained in the outputs of presence detectors. A detailed discussion is contained in [23]. The ideal output of a detector at any time is 0 if no car is above the detector and 1 if a car is present. Hence the time history of a detector output is a sequence of unit pulses, whose pulse width is the time that an individual car is over the detector. From such data, one can directly compute two quantities of importance in estimating aggregate traffic variables and in detecting incidents:

(1) Car count data -- the number of cars passing a given detector station in a prescribed interval of time. This directly yields flow

information -- i.e. cars/hour/lane.

(2) Occupancy data -- the percentage of time that cars are over the given detector during the prescriped time interval. This is related to traffic density.

These quantities may contain errors due either to imperfections in the detectors or to a car being counted twice, by detectors in adjacent lanes, as the car switches lanes (see [23]).

Intuitively, when a capacity reducing blockage occurs somewhere on a freeway, the density of traffic upstream of the incident location increases, while it decreases downstream. Hence, one can consider designing incident detection algorithms that look for low occupancy at the downstream detector and high occupancy at the upstream detector. In fact, this is the basis of many of the algorithms discussed [1-8]. Some 24 algorithms, all based on simple functionsof flow and occupancy crossing calibrated thresholds, were studied by Payne, et.al., [8]. One particular algorithm of this type, the so-called Algorithm #7, has emerged as the most widely accepted incident detection system. This algorithm employs an occupancy difference test along the lines of the one mentioned above, together with a persistence requirement (to reduce false alarms) and a test at adjacent detectors to reduce the number of alarms caused by traveling compression waves. The details of this algorithm are described in [8] and in [23].

We have used our microscopic simulation, equipped with presence detectors at half-mile intervals to generate a number of scenarios

for the testing of Algorithm #7 and of the systems described in the next two sections. These results indicate that Algorithm #7 detects incidents well in heavy traffic but has difficulty in detecting incidents in low or moderate traffic. In our simulations this algorithm was unable to detect incidents in flows below 1400 cars/hour/lane. In fact, in [23] it is argued that the probability is .5 that Algorithm #7 will be able to detect (with arbitrarily large detection delay) an incident at a flow of 1300 cars/hour/lane. This is not a light traffic condition. In fact, for the $v^e$-curve described in Section II, this flow corresponds to a high enough density ($> \rho_{free}$) so that the equilibrium velocity is 48 miles/hr.

Although the above comments indicate a limitation to the performance of Algorithm #7, it should be remembered that: (1) this algorithm directly uses presence detector outputs (averaged, typically, over 20-60 second time intervals); and (2) the algorithms is extremely simple to implement. In the next two sections we develop a method for using presence detector data with the MM and GLR systems. In Section IX we discuss the complexity issue.

## VII. The Estimation of Aggregate Variables from Presence Detector Data

As described in the preceding section, a presence detector provides time-averaged information about traffic conditions at a fixed spatial location. On the other hand, the variables in the models on which the GLR and MM systems are based are spatial-averaged quantities at fixed times. In this section we discuss the problem of processing data of

the first type to produce estimates of variables of the second kind.

A number of authors [32-35] have considered density estimation systems. The simpliest of these is discussed by Nahi and Trivedi [32,33]. Their recursive estimation system is based upon counting vehicles as they enter and exit the link. Given a good initial density estimate, Nahi's method showed the ability to track the density very closely in spatially homogeneous conditions. No results were presented for inhomogeneous conditions. The issue of imperfect vehicle count information was not considered. The performance with poor initial estimates was also not discussed. Furthermore, an explicit homogeneity assumption was made in the development of the system. This type of assumption is clearly not valid for incident conditions and can be expected to lead to large estimation errors.

Motivated by the results in [32,33] we have developed a new link density estimation system that is very simple and also overcomes the limitations of Nahi's system. Consider a single link of a freeway with detectors in each lane at both ends of the link. Let $\Delta$ be the time interval over which the temporal averaging of loop detector data is performed, and let k denote the discrete-time index --e.g. $\rho(k)$ denotes the spatial average density at time $k\Delta$. Let

$C_u(k)$ = the number of cars counted in the time interval $(k-1)\Delta \leq t \leq k\Delta$ at the upstream detector station

$C_d(k)$ = the number of cars counted in the time interval $(k-1)\Delta \leq t \leq k\Delta$ at the downstream detector station

$OCC_u(k)$ = the occupancy measured at the upstream detector station over the time interval $(k-1)\Delta \leq t \leq k\Delta$

$OCC_u(k)$ = the occupancy measured at the downstream detector station over the time interval $(k-1)\Delta \leq t \leq k\Delta$.

First note that $C_u(k)/\Delta$ is a direct measure of the flow (cars per unit time) at the upstream detector, and $C_d(k)/\Delta$ is an analogous measure at the downstream detector. Thus, a reasonable estimate of flow on the link is the average of these quantities:

$$\hat{\phi}(k) = \frac{C_u(k) + C_d(k)}{2\Delta} \tag{7.1}$$

Also note that the difference $C_u(k) - C_d(k)$ measures the change in the number of cars on the link during the time interval $(k-1)\Delta \leq t \leq k\Delta$. Hence we directly obtain the following equation for the evolution of link density

$$\rho(k) - \rho(k-1) = \frac{C_u(k) - C_d(k)}{h} + w(k-1) \triangleq u(k-1) + w(k-1) \tag{7.2}$$

where h is the length of the link and $w(k-1)$ is a noise process used to model the possible discrepancies between the actual change in the number

of cars on the link and the number obtained from car count information. This discrepancy is caused by a detector missing a car or a car being counted by detectors in two different lanes at the same station (see [23] for details). The calculation of the variance Q of w(k-1) is discussed in [23], and a value of Q on the order of .1 cars/mile/lane was found to be valid over a wide range of traffic conditions.

Equation (7.2) implies that we can use car count data to keep track of changes in density, but by itself such data cannot reduce any initial uncertainty in $\rho$, and the accumulation over time of the noise process w(k) will lead to further deterioration in our estimate of density. Thus we would like to use occupancy data to provide a direct measurement of density.

Intuitively, if traffic is spatially and temporally homogeneous, one should be able to relate temporal averages to spatial averages. Using results of this type [36-38], it is shown in [23] that under spatially homogeneous conditions density is proportional to occupancy. The constant of proportionality is

$$\alpha = 528\ E\left(\frac{1}{\ell+d}\right) \tag{7.3}$$

where $\ell$ is the length (in feet) of a vehicle picked at random, d is the effective length of the loop detector, which is assumed known*, and the

* Detectors have nonzero length. In addition, as discussed in [23], detectors respond to vehicles in a somewhat larger region. In general, d is a function of car length, $\ell$ (see [23]).

expectation in (7.3) is with respect to a given distribution of vehicle lengths. See [23] for a discussion of the calculation of $\alpha$. Based on a specific distribution for $\ell$ and a value of d=8 feet, we obtain the value used in our studies:

$$\alpha = 17.95 \tag{7.4}$$

This analysis leads to the following processing algorithm. Let

$$z(k) = \frac{OCC_u(k) + OCC_d(k)}{2\alpha} \tag{7.5}$$

Then, under spatially homogeneous conditions, we have

$$z(k) = \rho(k) + \eta(k) \tag{7.6}$$

where $\eta(k)$ is an unbiased sequence of errors, which are assumed to be white with known variance R. Results in [23] indicate that the variance of $\eta(k)$ should be taken in the range from 50 to 100 cars/mile/lane.

Given the model (7.2), (7.6) we can design a one-dimensional Kalman filter as a density estimation system that works well under spatially homogenous conditions

$$\hat{\rho}(k+1|k) = [1-H]\hat{\rho}(k|k-1) + Hz(k) + u(k) \tag{7.7}$$

where the filter gain is given by

$$H = \frac{Q + \sqrt{Q^2+4QR}}{Q + 2R + \sqrt{Q^2+4QR}} \tag{7.8}$$

and the steady-state variance of the residual

$$r(k) = z(k) - \hat{\rho}(k|k-1) \tag{7.9}$$

is given by

$$V = \frac{Q + 2R + \sqrt{Q^2+4QR}}{2} \tag{7.10}$$

This system has been studied using our microscopic simulation, and, under normal conditions, large initial errors in the estimate of $\rho$ can be reduced significantly within one minute with measurements taken every 5 seconds. This is a major improvement over previously developed techniques. Figure 7.1 depicts a typical simulation result.

Recall that the underlying assumption behind (7.6) is the homogeneity of traffic. Clearly any inhomogeneity, such as an incident, may cause the relationship (7.6) to fail. This leads directly to the following idea: can we monitor the residuals (7.9) in order to <u>detect</u> such inhomogeneities and to compensate the estimate (7.7) to correct for errors in (7.6)? To this end, we consider a simple model for the breakdown of (7.6) due to the development of spatial inhomogeneities -- the onset of a bias of unknown size $\nu$:

$$z(k) = \rho(k) + \eta(k) + \nu\sigma(k-\theta) \tag{7.11}$$

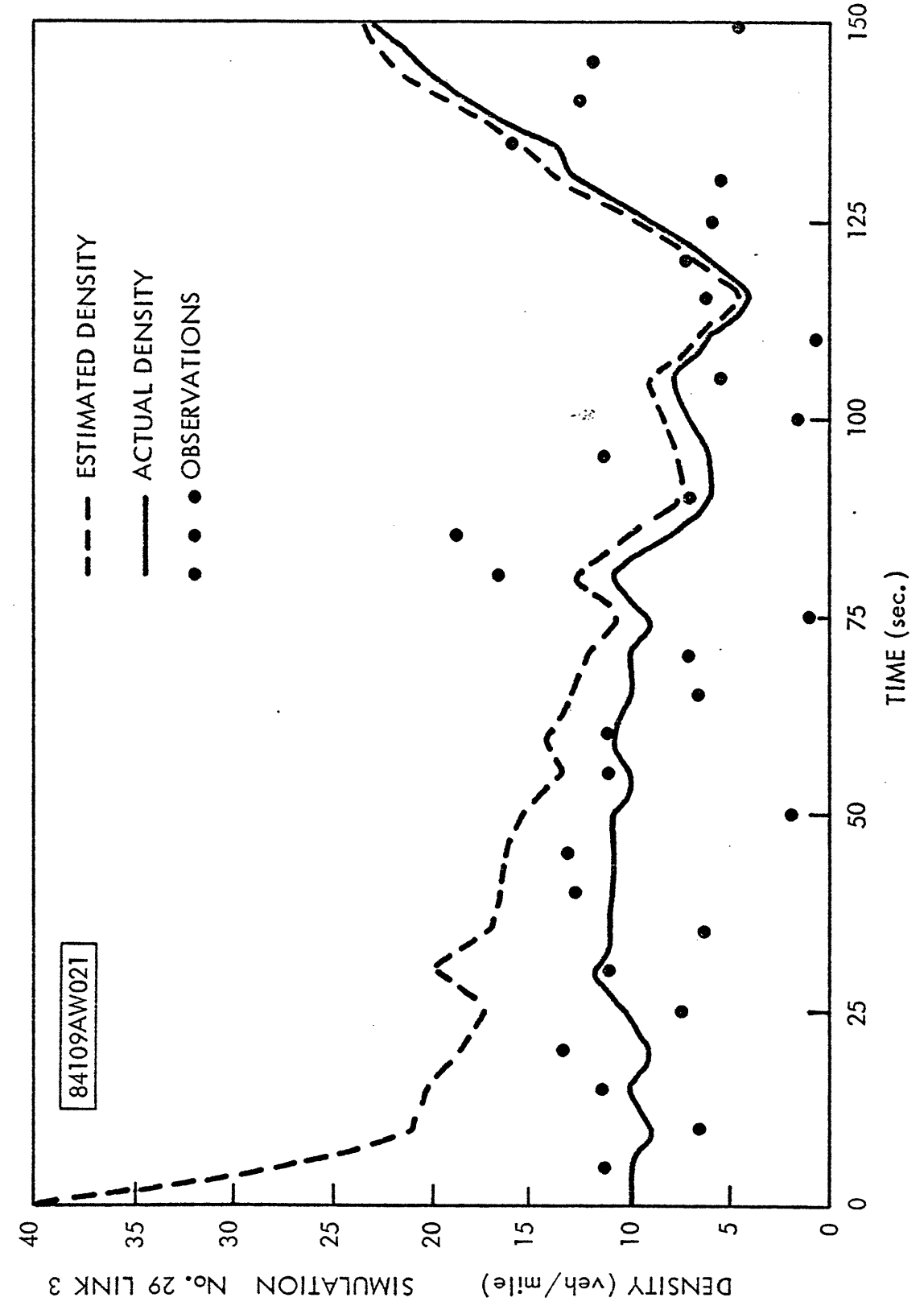


Figure 7.1: Density Estimation Performance.

(where $\sigma(m)=1$, $m\geq 0$, $=0$, $m<0$).

Using (7.2), (7.11) we can devise a simple density estimation system for use under <u>any</u> traffic conditions. Design a Kalman filter as in (7.7). Then, following the discussion of the GLR algorithm in Section IV, the residuals (7.9) can be written as

$$r(k) = \nu g(k-\theta) + \tilde{r}(k) \tag{7.12}$$

where $\tilde{r}(k)$ is a zero mean, white process with variance given by (7.10) The signature $g(k-\theta)$ can be easily calculated in this case as

$$g(j) = (1-H)^j \tag{7.13}$$

We now can implement a GLR-based algorithm (not to be confused with the incident detection algorithm of Section IV). We compute

$$d(k,\theta) = \frac{1}{V} \sum_{j=\theta}^{k} g(j-\theta) r(j) \tag{7.14}$$

$$\ell(k,\theta) = \frac{d^2(k,\theta)}{s(k-\theta)} \tag{7.15}$$

where

$$s(j) = \frac{1}{V} \sum_{k=0}^{j} g^2(k) = \frac{1-(1-H)^{2(j+1)}}{[1-(1-H)^2]V} \tag{7.16}$$

Here s(j) is a measure of the amount of information available in $r(\theta)$. $r(\theta+1),\ldots,r(\theta+j)$ about a bias initiated at time $\theta$. Note from (7.13) that the signature g decreases geometrically, and it is easy to see that (7.16) reaches a limiting value

$$s(\infty) = \frac{1}{V[1-(1-H)^2]} \tag{7.17}$$

The implications of these observations is that there is no reason to calculate $d(k,\theta)$ for $k-\theta$ too large, since $r(k)$ contains essentially no information about possible biases occurring at time $\theta$. Also, for $k-\theta$ too small, $s(k-\theta)$ will be small, and we will not have accumulated enough data to make a useful decision. In [23] these calculations are carried out for the numerical values cited earlier in this section. Based on these, we chose the window

$$k-13 \leq \theta \leq k-9 \tag{7.18}$$

for the calculations (7.14), (7.15). In this range $s(j) \simeq .08$, and we used this constant value in our implementation of (7.15).

Let

$$\hat{\theta}(k) = \underset{k-13\leq\theta\leq k-9}{\arg\max}\ \ell(k,\theta) \tag{7.19}$$

The decision rule used was

$$\ell(k,\hat{\theta}(k)) \underset{\text{Normal conditions}}{\overset{\text{Bias Detected}}{\gtrless}} \varepsilon \tag{7.20}$$

where the threshold $\varepsilon$ is chosen to provide a reasonable tradeoff between false alarms and correct detections (see [17,23,25,31] for treatments of the evaluation of GLR performance). Note that the probability of detection depends upon the size $\nu$ of the bias. In [23] a tradeoff analysis is carried out to determine detection performance as a function of postulated size for $\nu$ and values of $\varepsilon$. Also, we refer the reader to [23] for a discussion of expected sizes for $\nu$ under different traffic conditions. The results of this analysis are that in low or medium flows, values of $\nu$ on the order of 5 to 20 vehicles/mile/lane) can be expected to occur following an incident, while values as large as 80 can occur in heavy traffic. Thus, one can set $\varepsilon$ at a higher value in heavy traffic, reducing false alarms while maintaining a high correct detection probability. Note that flow-scheduled thresholds are easily implemented, as the estimate of $\phi$ given by (7.1) is good. In our simulations, we found that thresholds ranging from 2.5 at low flow levels to 3.6 in heavy flow produced good detection performance.

Following the detection of a bias, we want to compensate the filter estimate $\hat{\rho}$ to correct for the effect of the bias. In a manner similar to the calculation of $g(k-\theta)$, we can calculate the bias $\rho_b$ in $\hat{\rho}(k|k)$ caused

by a measurement bias $\nu$ occurring at time $\theta$.

$$\rho_b(k) = [1-(1-H)^{k-\theta+1}]\nu \triangleq F(k-\theta)\nu \tag{7.21}$$

Then, given that a detection is made at time k and given the most likely time $\hat{\theta}$ from (7.19) and the most likely magnitude of the measurement bias*

$$\hat{\nu} = \frac{d(k,\hat{\theta})}{s(k-\hat{\theta})} \tag{7.22}$$

we obtain an estimate of $\rho_b$:

$$\hat{\rho}_b = F(k-\hat{\theta})\hat{\nu} \tag{7.23}$$

which we use to correct our estimate

$$\hat{\rho}(k|k)_{new} = \hat{\rho}(k|k)_{old} - \hat{\rho}_b \tag{7.24}$$

Having done this and having compensated the measurements by subtracting $\hat{\nu}$ from all incoming measurements, we are in a position to detect further changes, such as the return to homogeneous conditions.

Figure 7.2 indicates the typical performance of this system. In this simulation an incident introduces spatial inhomogeneities. Note that the GLR system does not remove the entire bias in one step. This

---

* See [17] for the derivation of (7.22).

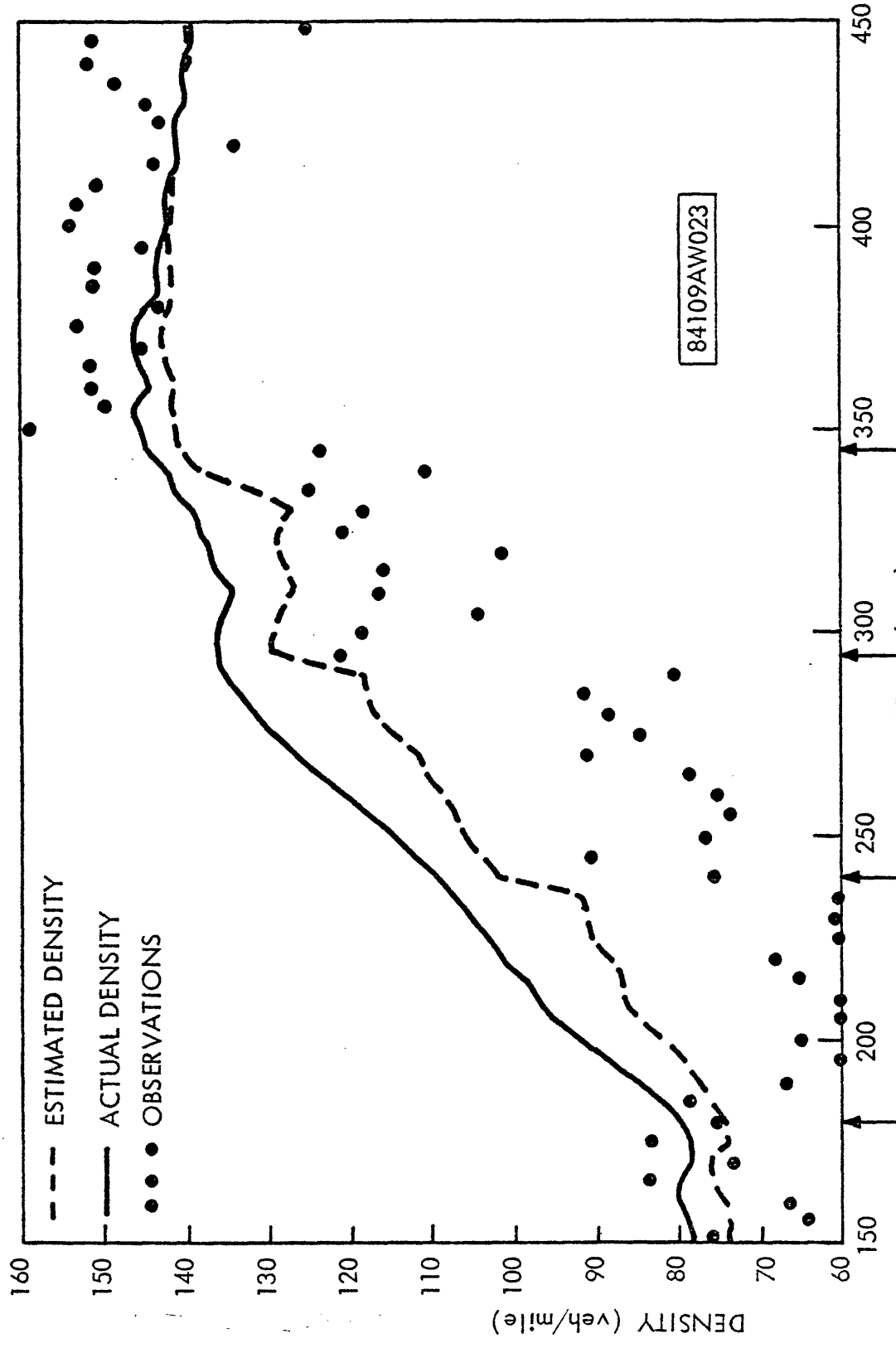


Figure 7.3: Density Estimation Performance Under Incident Conditions.

is not only due to errors in the estimate $\hat{\nu}$ but also can be traced to the fact that the model (7.11) is a very simple idealization of the relationship between z and $\rho$ under inhomogeneous conditions. Yet even with this simple model and the resulting very simple estimation - GLR system (it is one-dimensional), the performance shown in Figure 7.2 is extremely encouraging. The system has self-correcting characteristic that allows it to track density under all conditions that we have simulated.

Finally, recall that the GLR and MM systems require measurements of both $\rho$ and v. Using the approximation*

$$v = \frac{\phi}{\rho} \tag{7.25}$$

we can obtain an estimate of velocity

$$\hat{v}(k) = \frac{\hat{\phi}(k)}{\hat{\rho}(k|k)} \tag{7.26}$$

This estimate is not nearly as good as the estimates of $\phi$ and $\rho$, due to the errors in (7.25). The problem of estimating average velocity is a difficult one. Some issues involved in this problem are discussed in [23], and in Section IX of this paper.

---

* Note that for a compressible fluid (7.25) is an exact relationship if we use point values for v, $\phi$, and $\rho$. It is only approximate if we use spatial averages.

## VIII. The GLR and MM Algorithm Using Presence Detector Data

The system described in the preceding section for computing estimates of spatial mean densities and velocities was combined with the GLR and MM algorithms, with the estimates produced by the former being used as the measurements for the latter. This combined system was then tested using the same microscopic simulations as described in Section V, although in this case detector data were used directly. As discussed in Section II, the GLR and MM systems were modified slightly by using larger values for measurement noise variances to account for the errors in the estimates of $\rho$ and $v$ provided by our detector data pre-processing algorithm.

As one might expect the additional errors introduced by the increased uncertainties in our derived measurements of $\rho$ and $v$ lead to a slight increase in the minimum flow at which incidents can be detected. However, these algorithms still detected incidents in flows down to 1000 cars/hour/lane. Recall the Algorithm #7 required flows of at least 1400 cars/hour/lane in our simulations.

Aside from this increase in minimum flow required for detection, the simulation results using presence detector data are extremely similar to the results obtained using aggregate measurements directly computed from the microsimulation. Compare Figures 8.1, 8.2 with Figures 5.5, 5.6.

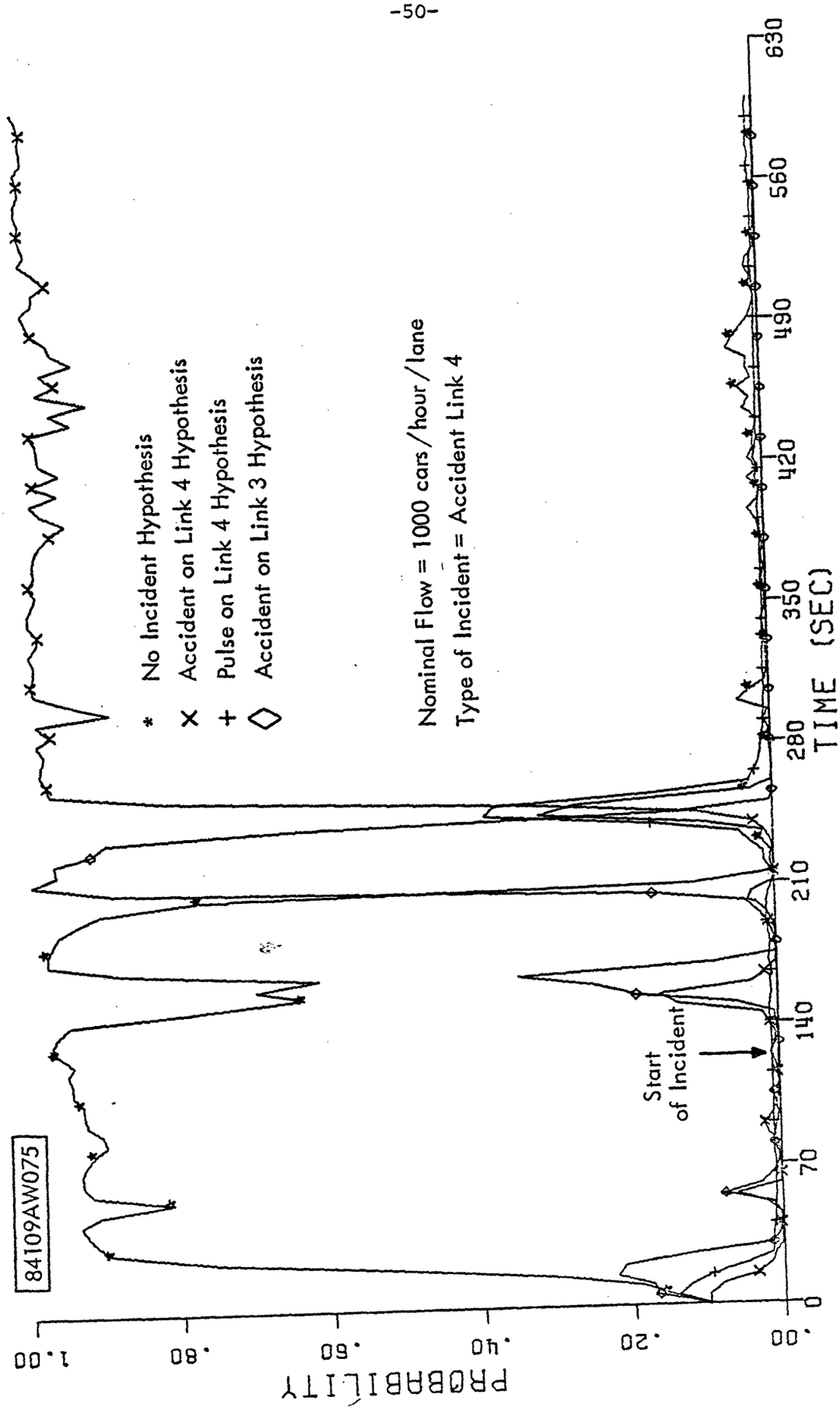


Figure 8.1: MM, Probability Plot - Presence Detector Data.

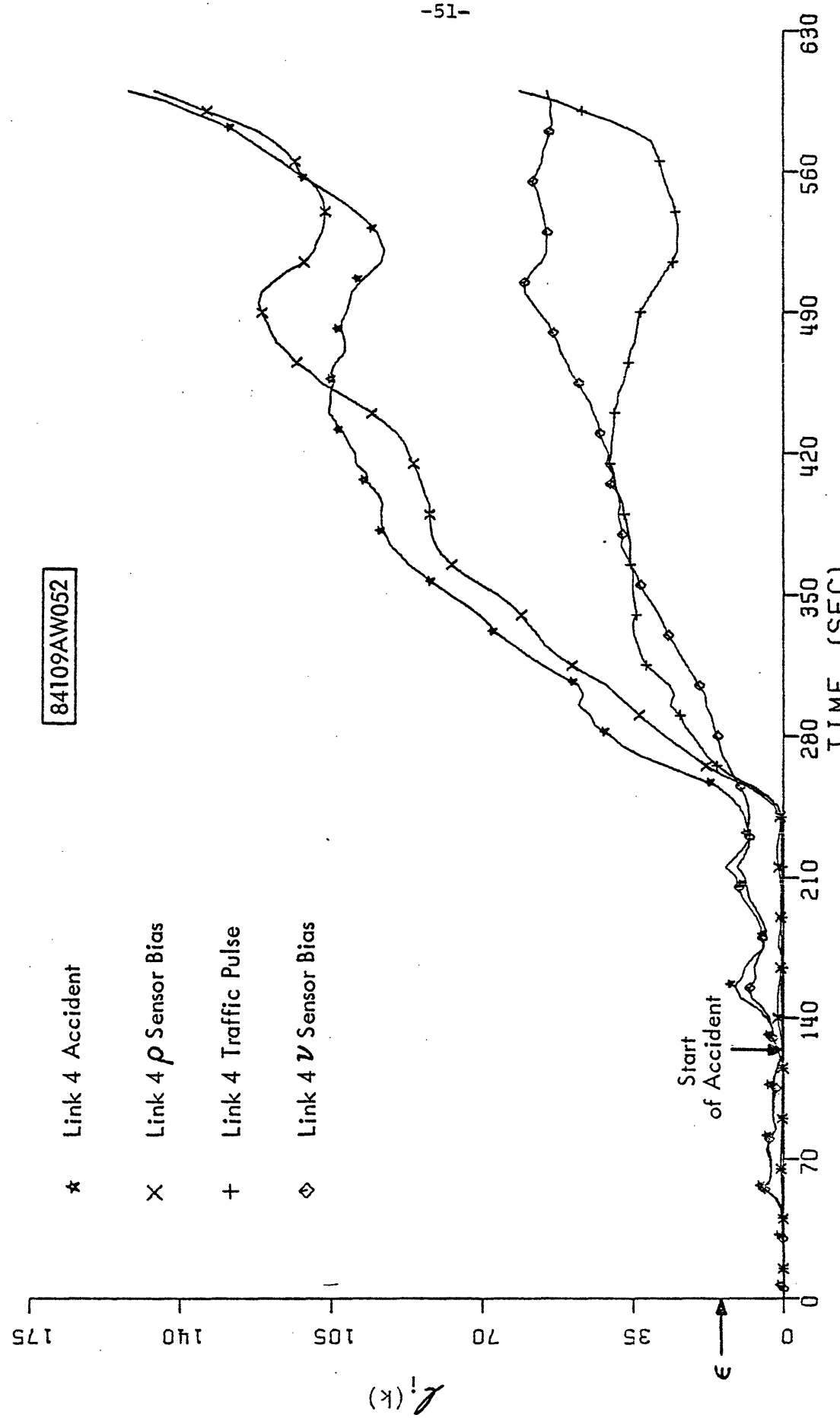


Figure 8.2: GLR, $\ell_i(k)$: Same Simulation as in Figure 8.1:

## IX. Discussion

In this paper we have described the application of modern estimation and detection techniques to the problem of detecting incidents on freeways. The algorithms we have developed have shown promise of providing performance improvements over existing algorithms. Moreover, the extremely simple system we have developed for estimating aggregate traffic variables from presence detector data should be of interest in traffic surveillance applications other than incident detection. We feel that the work described herein represents a good example of the successful interfacing of advanced systems techniques with the constraints and non-ideal nature of a real world problem.

While the techniques that we have developed have yielded encouraging results, a number of further questions must be addressed:

(1) The system described in Section VII provides good estimates of $\rho$ but rather poor ones of $v$. Alternate methods for estimating velocity should be investigated. For example, in [20] it is shown that the actual presence detector signal (which is not simply 0's and 1's) can be used to estimate vehicle length. Having vehicle length, one can estimate vehicle velocity by measuring the length of time that the vehicle is over the detector. These individual velocities can be averaged to form a time-averaged estimate of velocity at a fixed spatial location. Then, much as discussed in Section VII, under homogeneous conditions we can use the results in [23,36] to transform this quantity to an estimate of spatial-average velocity at a fixed time. The development of a system as in Section VII that will also work under inhomogeneous conditions remains for the future.

(2) Alternatively, to overcome the problem of the poor estimate of $v$, we could consider using $\rho$ and $\phi$ as the basic variables in the design of the GLR and MM systems.

This would require some design modifications but might lead to a more effective detection algorithm.

(3) One problem that must be attacked is that of sensor failures. As discussed in Sections V and VIII, the GLR techniques has some difficulties in distinguishing incident hypotheses and some of the sensor bias models. The main point involved in alleviating this problem is the observation that macroscopic sensor failure models have very little to do with the actual failure of a presence detector. Thus, one might expect better performance from a system that includes more realistic models. For example, if a presence detector on link i fails, it is reasonable to expect the "measurements" of both $\rho_i$ and $v_i$ to be faulty. Thus, instead of considering these separately, we might want to consider an hypothesis representing such a simulataneous failure. Examining Figures 5.6 and 8.2, we see that it is only the log-likelihood for a link 4 density bias that is comparable to that for the actual incident hypothesis. Since a link 4 velocity bias is less likely, it seems reasonable to expect that an algorithm based on a simultaneous failure model would have an easier time distinguishing this from an incident.

(4) The idea expressed in (3) above may very well work, but whatever is done along those lines, one always will have the problem that macroscopic sensor failure models do not accurately represent actual physical phenomena. Thus it might be preferable to perform sensor failure detection directly on the presence detector data. Fail to zero or full-on failures are not difficult to detect using simple logic directly on detector outputs. Also, it may be possible to design a GLR system, much as that discussed in Section VII, for the detection of sensor failures.

(5) Work is needed in the development of useful detection rules, based on the MM probabilities or GLR-log-likelihood ratios. Persistence requirements (i.e. $p_i$ or $\ell_i$ remaining above a threshold for a period of time) are clearly needed, and such a feature will eliminate many problems such as the transient response of GLR and MM in the simulation containing two slowly moving vehicles (see Figures 5.7,5.8). In addition, we may wish to consider flow-scheduled thresholds,

as were used in our aggregate variable estimation system. Here, one could use higher thresholds in high flows to avoid false alarms caused, for example, by compression waves. One would not be sacrificing very much in terms of detection performance, as the effect of an incident in heavy traffic is much larger than in light traffic. Conversely, a lower threshold in light traffic would improve detection performance without any drastic effect on the false alarm rate, since large amplitude spatial inhomogeneities do not occur under normal conditions when traffic is light. The tradeoff in terms of false alarm probability vs. detection delay must be considered for any such decision rule.

(6) The inclusion of additional hypotheses (and the deletion of some of the present) ones should be investigated for both the GLR and MM systems. For example, we may wish to have a "traveling incident" model. Basically, we should aim to include hypotheses for events that we want to detect plus hypotheses for events, such as slowly moving vehicles, which may confuse the detection algorithms unless accounted for. Clearly further thought and experimentation is needed in order to determine a preferred set of hypotheses.

(7) Computational issues in the implementation of the MM and GLR systems must be considered. For example, we may wish to consider a dual-mode system, in which the presence detector data processing system is used to signal an initial alarm (i.e. a GLR detection), and the GLR and MM systems are only engaged subsequent to such a detection. Also, the decentralization of our systems must be considered. The technique for converting presence detector data to measurements of aggregate variables is already decentralized (only adjacent detector stations need communicate with each other) and computationally very simple. The decentralization of the Kalman filters, using adjacent data can also be accomplished (see [13]), and the GLR algorithm we have developed is decentralized in that an incident on link i leads to non-zero signatures only on links i-1,i, and i+1, (see [25]). A similar decentralization should be possible for the MM algorithm. At this juncture, it appears that the GLR system may offer substantial computational advantages over the MM algorithm in that only a single Kalman tracking filter is required, and the basic GLR calculations, consisting of correlating observed tracking errors with several

innovations signatures, may be performed efficiently using fast correlation-convolution method [39].

(8) Finally, after the issues mentioned above have been addressed, the MM and GLR systems should be tested on real traffic data. It is clear that these algorithms are somewhat more complex than existing systems, and such tests will provide the basis for deciding if this increase in complexity is justified by a commensurate improvement in detection performance.

## REFERENCES


1. H.J. Payne, "Freeway Incident Detection Based Upon Pattern Classification," Proc. 1975 IEEE Conf. on Decision and Control, Houston, Texas, Dec. 10-12, 1975.

2. A.R. Cook and D.E. Cleveland, "Detection of Freeway Capacity-Reducing Incidents by Traffic Stream Measurements," in Transportation Research Record No. 495, Incidents and Freeway Operations, 1974.

3. C.L. Dudek, C.J. Messer, and N.B. Nuckles, "Incident Detection on Urban Freeways," in Transportation Research Record, No. 495, Incidents and Freeway Operations, 1974.

4. C.L. Dudek, "Better Management of Traffic Incidents -- Scope of the Problem," presented at the 1974 Summer Meeting of the Transportation Research Board, Jacksonville, Florida, August 1974.

5. C.L. Dudek, C.J. Messer, and A.K. Dutt, "Study of Detector Reliability for a Motorist Information System on the Gulf Freeway," in Transportation Research Record, No. 495, Incidents and Freeway Operations, 1974.

6. A.D. May, W.M. Chow, M. Eldor, C.K. Lu, and M. Sakasita, "Optimal Design and Operation of Freeway Incident Detection-Service Systems," M.S. Department of Transportation Report No. DOT-TST-75-92, February 1975.

7. J.M. McDermott, "Surveillance and Control for Incidents on Chicago Area Freeways," presented at the 1974 Summer Meeting of the Transportation Research Board, Jacksonville, Florida, August 1974.

8. H.J. Payne, E.D. Helfenbein, H.C. Knobel, "Development and Testing of Incident Detection Algorithms," Volume 2, Final Report, FHWA Contract FH-11-8278, February 1976.

9. L. Isaksen and H.J. Payne, "Freeway Traffic Surveillance and Control," Proc. IEEE, Vol. 61, No. 5, May 1973, pp 526-536.

10. P.K. Houpt and M. Athans, "Dynamic Stochastic Control of Freeway Corridor Systems, Vol. I: Summary," M.I.T. Electronic Systems Laboratory Report ESL-R-608, Cambridge, Massachusetts, August 1975.

11. S.B. Gershwin, "Dynamic Stochastic Control of Freeway Corridor Systems, Vol. II, Steady State Optimal Traffic Assignments Using the Accelerated Gradient Projection Method," M.I.T. Electronic Systems Laboratory, Report ESL-R-609, Cambridge, Massachusetts, August 1975.

12. D. Looze, P.K. Houpt, and M. Athans, "Dynamic Stochastic Control of Freeway Corridor Systems, Vol. III: Dynamic Centralized and Decentralized Control Strategies," M.I.T. Electronic Systems Laboratory, Report ESL-R-610, Cambridge, Massachusetts, August 1975.

13. D. Orlhac, M. Athans, J. Speyer, and P.K. Houpt, "Dynamic Stochastic Control of Freeway Corridor Systems, Vol. IV: Estimation of Traffic Variables Via Extended Kalman Filter Methods," M.I.T. Electronic Systems Laboratory Report ESL-R-611, Cambridge, Massachusetts, August 1975.

14. A.D. St. John, "Study of Traffic Phenomena Through Digital Simulation," Final Report, Research Grant AC-000106, Department of Health, Education, and Welfare, 1966.

15. W. Mitchell, "The Estimation and Simulation of Freeway Traffic Flow Using Car-Following and Fluid Analog Models," S.M. thesis, Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, Massachusetts, June 1977.

16. A.H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

17. A.S. Willsky and H.L. Jones, "A Generalized Likelihood Ratio Approach to the Detection and Estimation of Abrupt Changes in Linear Systems," IEEE Trans. Aut. Control, Vol. AC-21, February 1976, pp. 108-112.

18. M.-F. Chang and D.C. Gazis, "Traffic Estimation with Consideration of Lane Changing," IBM Research Report RC5250, Yorktown Heights, New York, February 1975.

19. W.F. Phillips, "Kinetic Model for Traffic Flow," Dept. of Mech. Eng., Utah State Univ., Logan, Utah, 1977, Final Report for U.S. Dept. of Transportation Research Contract DOT-OS-40097.

20. J.T. Olesik, "Use of Loop Detector Phase Signatures For Vehicle Speed and Type Classification," S.M. and E.E. thesis, Dept. of Electrical Engineering and Computer Science, M.I.T., 1976.

21. D. O'Mathuna, "Highway Traffic Kinematics and the Characteristic Relation," Dept. of Transportation, Rept. No. DOT-TSC-OST-76-48, March, 1977.

22. E.Y. Chow, S.B. Gershwin, C.S. Greene, P.K. Houpt, A.L. Kurkjian, and A.S. Willsky, "Dynamic Detection and Identification of Incidents on Freeways, Volume I: Summary," M.I.T. Electronic Systems Laboratory Report ESL-R-764, Cambridge, Massachusetts, Sept. 1977.

23. A. Kurkjian, S.B. Gershwin, P.K. Houpt, and A.S. Willsky, "Dynamic Detection and Identification of Incidents on Freeways. Volume II: Approaches to Incident Detection Using Presence Detectors," M.I.T. Electronic Systems Laboratory Report ESL-R-765, Cambridge, Massachusetts, September 1977.

24. C.S. Greene, P.K. Houpt, A.S. Willsky, and S.B. Gershwin, "Dynamic Detection and Identification of Incidents on Freeways. Volume III: The Multiple Model Method," M.I.T. Electronic Systems Laboratory Report ESL-R-766, Cambridge, Massachusetts, September 1977.

25. E.Y. Chow, A.S. Willsky, S.B. Gershwin, and P.K. Houpt, "Dynamic Detection and Identification of Incidents on Freeways. Volume IV: The Generalized Likelihood Ratio Method," M.I.T. Electronic Systems Laboratory Report ESL-R-767, Cambridge, Massachusetts, September 1977.

26. D.T. Magill, "Optimal Adaptive Estimation of Sampled Processes," IEEE Trans. Aut. Cont., Vol. AC-10, Oct. 1965, pp. 434-439.

27. D.G. Lainiotis, "Partitioning: A Unifying Framework for Adaptive Systems, I: Estimation," Proc. IEEE, Vol. 64, No. 8, 1976, pp. 1126-1142.

28. C.S. Greene, "An Analysis of the Multiple Model Adaptive Control Algorithm," Ph.D. Dissertation, M.I.T., Aug. 1978.

29. M.G. Safanov and M. Athans, "Gain and Phase Margins for Multiloop LQG Regulators," IEEE Trans. Aut. Cont., Vol. AC-22, No. 2, April 1977, pp. 173-179.

30. A.S. Willsky, "A Survey of Design Methods for Failure Detection in Dynamic Systems," Automatica, Vol. 12, 1976, pp. 601-611.

31. E.Y. Chow, "Analytical Studies of the Generalized Likelihood Ratio Technique for Failure Detection," S.M. thesis, M.I.T., Feb. 1976.

32. N.E. Nahi, "Freeway Data Processing," Proc. IEEE, Vol. 61, May 1973, pp. 537-541.

33. N.E. Nahi and A.N. Trivedi, "Recursive Estimation of Traffic Variables: Section Density and Average Speed," Univ. of Southern California, Los Angeles, California, pp. 269-286.

34. D.C. Gazis, and C.K. Knapp, "On-Line Estimation of Traffic Densities from Time-Series of Flow and Speed Data," Trans. Sci., Vol. 5, No. 3, pp. 283-301, August 1971.

35. D.C. Gazis and M.W. Szeto, "Design of Density Measuring Systems for Roadways," p. 44-52, Highway Research Board Record, No. 388, 1972.

36. J.G. Wardrop, "Some Theoretical Aspects of Road Traffic Research," Proc. Inst. Civil Engrs., Part II, 1, No. 2, pp. 325-362, 1952.

37. S.B. Gershwin, "On the Relation Between Vehicle Flow, Vehicle Density and Velocity Distribution," MIT Electronic Systems Laboratory, Cambridge, Technical Memorandum, ESL-TM-570, September 24, 1974.

38. L.Breiman, "Time Scales, Fluctuations and Constant Flow Periods in Uni-Directional Traffic," Transportation Research, Vol. 7, pp. 77-105, Pergamon Press, 1973.

39. T.G. Stockham, High Speed Convolution and Correlation," 1966 Spring Joint Computer Conference, AFIPS Proc., Vol. 28, 1966, pp. 229-233.

40. H.J. Payne, "Models of Freeway Traffic for Control," Math. Models of Public Systems, Simulation Council Proceedings, Vol. 1, No. 1, Jan. 1971.