Description of Macroscopic Relationships Among Traffic Flow Variables Using Neural Network Models

TAKASHI NAKATSUJI, MITSURU TANAKA, POURMOALLEM NASSER, AND TORU HAGIWARA

The relationships between traffic flow variables play important roles in traffic engineering. They are used not only in basic traffic flow analyses but also in some macroscopic traffic flow simulation models. For many decades, various mathematical formulations that describe the relationships among density, flow, and speed have been proposed, including multiregime models. Previously, the best mathematical curve was determined by trying several different formulas and applying regression analysis. In these processes, one must specify in advance which mathematical formula should be adopted and where it should be shifted to another in a multiregime model. Neural network models have some promising abilities to represent nonlinear behaviors accurately and to self-organize automatically. A procedure for describing the macroscopic relationships among traffic flow variables using some neural network models is presented. First, a Kohonen feature map model was introduced to convert original observed data points into fewer, more uniformly distributed ones. This conversion improved regression precision and computational efficiency. Next, a multilayer neural network model was introduced to describe the two-dimensional relationships. The model was effective in describing the nonlinear and discontinuous characteristics among traffic flow variables. It was unnecessary to specify the regression curves and the transition points in advance. The multiple correlation coefficients resulting from the model were better than those resulting from a conventional nonlinear equation.

The relationships among traffic flow variables play important roles in traffic engineering. They are used not only in analyses of traffic flow behavior but also in some macroscopic traffic flow simulation models. For many decades, traffic-flow analysts have studied various mathematical formulations that describe the relationships among density, flow, and speed of uninterrupted traffic flows (1–3). The best mathematical formerly was determined by trying several formulas and applying regression analysis techniques. In some cases, one equation may be most appropriate; another may be better in the others. Moreover, multiregime models that use a few functions have been proposed, too (1–3). Normally they include discontinuity points not only in original functions but also in their derivatives—that is, in applying such models, one must specify in advance which mathematical formula should be adopted in each region and where it should be shifted to another.

As a matter of fact, the authors are now engaged in the development of a traffic flow simulation model (4) in which a characteristic curve prescribes the average traffic states. The curve is updated using traffic detector data at each observation point. The authors used to face the aforementioned difficulties in establishing a macroscopic relationship in a computer.

Some neural network models, such as a multilayer model (5), have the promising ability to describe nonlinear behaviors very well. So, it is expected that when they are applied to the regression problem, they can self-adjust the curvature of characteristic curves automatically while responding to the distribution of observed data. Above all, they require no preliminary knowledge of the mathematical formulas and the transition points. Another difficulty in regression analysis lies in the trimming of excessive observed data. When traffic flows are observed, often one comes across unequally distributed traffic data—distributed densely in a few restricted regions and sparsely in the others. This unequal distribution of observed data would affect the regression results badly. Excessive observed data in a region decrease the computational efficiencies, too. One must determine in advance which data should be retained and which should be trimmed. Some statistical criteria, such as AIC (Akaike information criteria) and FPE (final prediction error) (6), may provide useful knowledge about how *much* data should be retained, but they provide no information about *which* should be retained.

Some neural network models, such as a Kohonen feature map (KFM) model (7), have the ability to convert original observed data into fewer, more representative data automatically. The KFM model does not require any preliminary knowledge about the data structure. All one must do is specify the number of data points to which the original data set should be reduced.

BACKGROUND

Characteristic Curves

There are many characteristic curves proposed so far for describing the relationship between density and speed. In this study the authors used the formula derived from the car-following theory (3):

$$v = v_f \left[1 - \left(\frac{k}{k_j} \right)^{l-1} \right]^{\frac{1}{1-m}} \tag{1}$$

Civil Engineering Department, Hokkaido University, Kita 13, Nishi 8, Kitaku, Sapporo, 060, Japan.

where

k = density (veh/km),
v = speed (km/hr),
k_j = jam density,
v_f = free speed, and
l, m = sensitivity factors from car-following theory.

Substituting Equation 1 into the relationships $q = kv$, the other relationships among density, flow, and speed can be obtained as follows:

$$q = v_f k \left[1 - \left(\frac{k}{k_j} \right)^{l-1} \right]^{\frac{1}{1-m}} \tag{2}$$

$$q = k_j v \left[1 - \left(\frac{v}{v_f} \right)^{1-m} \right]^{\frac{1}{l-1}} \tag{3}$$

where q denotes the traffic flow rate in vehicles per hour. The unknown parameters in those equations are subject to some constraints (3):

$l > 0$
$m > 1$
$v_f^{min} \leq v_f \leq v_f^{max}$
$k_j^{min} \leq k_j \leq k_j^{max}$ (4)

Regression Analysis

Equations 1, 2, and 3 are expressed in a general form

$$y = f(x, a_1, a_2, a_3, a_4) \tag{5}$$

where

x = control variable,
y = state variable, and
a_j (j = 1,2,3,4) = unknown parameters of $l, m, v_f,$ and k_j in Equations 1–3, respectively.

By obtaining sets of observed data (x_i, y_i) (i = 1,2,...,N), one can identify the parameters by a regression technique. Since Equation 5 is in nonlinear form and is subject to some constraints given by Equation 4, the problem here reduces to a nonlinear constrained least mean square problem. That is, the unknown parameters are estimated so as to minimize the objective function J as follows:

$$J = \sum_i [y_i - f(x_i, a_1, a_2, a_3, a_4)]^2$$

Subject to

$$G_j \leq a_j \leq H_j \qquad j = 1, 2, 3, 4 \tag{6}$$

The authors used Box's complex algorithm to solve this problem. A detailed discussion of this algorithm can be found elsewhere (11)..

Multilayer Neural Network Model

Figure 1 shows a multilayer neural network model for describing the macroscopic relationships between traffic variables. It consists of three layers: an input layer, an intermediate layer, and an output layer. The strength of the connections is called synaptic weight. The normalized control variable x_i^B was entered into the input layer, such as k/k_j in Equation 1. The input signals were transmitted in sequence from the input layer to the output layer while the neural operations were repeated. The output layer produces the normalized objective variable y_k^D, such as v/v_f in Equation 1. This is the forward signal process in Figure 1. Next, the synaptic weights were adjusted so that the error between the output signals and the target signals is minimized. The backpropagation method (5) produces the adjustments of synaptic weights in each layer. In actual computation the synaptic weights are adjusted by the momentum method to smooth the adjustment and urge the convergence.

Kohonen Feature Map

The KFM model is a two-layered neural network that can organize a topological map from a random starting point. It has the ability to classify input patterns into several output patterns. Figure 2 depicts the basic network structure of a KFM model. the authors used a one-dimensional structure for this analysis. It consists of two layers: an input layer and a competitive layer. The interconnections (synaptic weights) are adjusted in a self-organizing manner without any target signals. the authors briefly explain how this can be done. An input pattern to the KFM is denoted here as

$$E = [e_1, e_2, \ldots, e_n] \tag{7}$$

Since the observed traffic variables are adopted as the input signals, the input layer has three neurons in it ($n = 3$). The weights from the input neurons to a single neuron in the competitive layer are denoted as

$$W_i = [w_{1i}, w_{2i}, \ldots, w_{ni}] \tag{8}$$

where i identifies the ith neuron in the competitive layer. The number of neurons there can be specified arbitrarily.

The first step in the adjustment of synaptic weights is to find a winning neuron c in the competitive layer whose weight vector
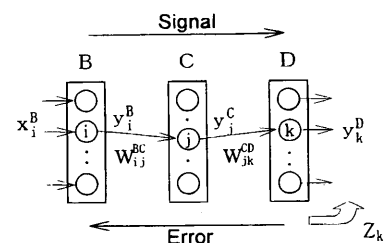

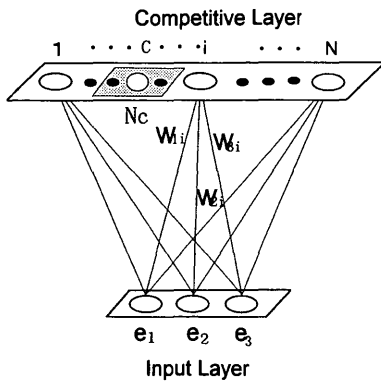**FIGURE 1 Basic structure of multilayer neural network.**

FIGURE 2   Basic structure of KFM model.

FIGURE 3   Overview of Yokohane Line and locations of traffic detectors.

matches most to each input vector E. The matching value is defined by the distance between vectors E and W_i:

$$\sqrt{\sum_j (e_j - w_{ij})^2} \qquad (9)$$

The neuron with the lowest matching values wins the competition. After the winning neuron c is identified, weights are updated for all neurons that are in the neighborhood N_c of the winning neuron. The adjustment is

$$\Delta W_{ij} = \begin{cases} \beta(e_j - w_{ij}) & \text{if } i \in N_c \\ 0 & \text{otherwise} \end{cases} \qquad (10)$$

where β is the learning rate, which is decreased over a span of many iterations. This adjustment results in the winning neuron becoming more likely to win the competition when the same or similar input pattern is presented subsequently. In other words, the synaptic weight vector W_i consequently represents those input patterns that resemble each other. This is what is called the integration of observed data. See work by Dayhoff (7) for more details.

TRAFFIC DATA

Observed Data

The observed data used here come from the Metropolitan Expressway in Tokyo. The data were collected on the Yokohane Line between Tokyo Haneda Airport and Yokohama in October 1993. Supersonic traffic detectors are installed in each of the two directional lanes every 300 m, and traffic data on flow, occupancy, and average speed are compiled every 1 min. Figure 3 depicts the schematic drawing of the freeway section and the location of the traffic detectors. Traffic data on both lanes in the eastbound direction from Yokohama to Tokyo Airport were used. This road section experiences incessant congestion in the daytime on weekdays. The authors chose such time periods that include extensive traffic situations, ranging from free-flow to congested states. In this analysis, assuming that density is proportional to time occupancy, the authors used time occupancy directly rather than converting it to density (8). This requires a minor change in the nonlinear equations from Equa-
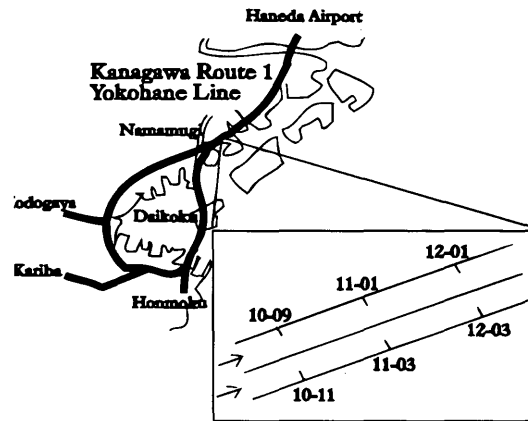
tions 1 through 3. Assuming homogeneity around observation points, the authors treated the time-mean speed as identical with the space-mean speed. However, it should be noted that this assumption is not always valid. One must examine carefully what has been analyzed, in particular when traffic is congested.

Training

By using the KFM model in three-dimensional space, the original observed data were converted to fewer points of more integrated data. Figure 4 depicts the schematic drawing of the conversion. Iterative trainings by the model produce the neurons whose weights correspond to integrated data. They were projected on each two-dimensional plane for two-dimensional analysis. Next, by using a multilayer neural network model, the input-output relationships between the control and the state variables were built up. The completion of training by the backpropagation method brings a stable regression between them.

Kohonen Feature Map

To convert observed data to sets of integrated data, the authors prepared a KFM model consisting of an input layer with three neurons and a competitive layer with neurons that correspond to the number of integrated data points. Before the training, all observed data are normalized. After having given a set of observed data to the input layer in Figure 2, the authors selected a winning neuron in the competitive layer and adjusted the weights of neurons in the neighborhood of the winning neuron. This process is iterated for all input patterns consecutively. Training iteration continues until the change of synaptic weights becomes sufficiently small. Finally, a stable formation of integrated data can be obtained.

The most important problem in this process is how to determine the number of integrated data points. Generally, the appropriate number of data points depends on the use of a characteristic curve; for interpreting traffic flow behaviors, the number must be determined carefully so as to not lose the original data properties. One must determine it while checking the information statistics based on a criterion, such as AIC or FPE. On the other hand, for using a char-
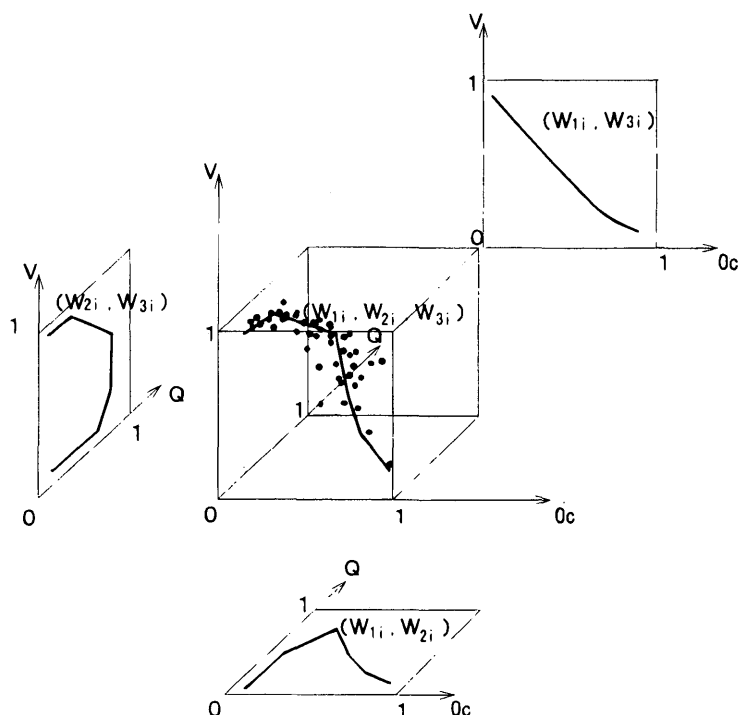
FIGURE 4   Schematic drawing of integration of observed data by Kohonen feature mapping.

acteristic curve in a simulation model, excessive data should be trimmed because such data affect the regression badly. In this case, one may be able to determine the number experimentally because only the average characteristics of traffic flow states are needed. In this paper, assuming the usage in a simulation model, the authors determined it experimentally: 20 points for each data set containing 120 points of observed data.

Figure 5 shows how original observed data are integrated as the training proceeds. For simplicity, the evolution process is projected on the occupancy-speed plane, and, for convenience, it is enlarged to the real scale. White circles in Figure 5a are original observed data, and black ones in the center of the graph are the initial weights that are set to the value 0.5 plus a small, within 10 percent, randomized value. Figures 5b–d show the distribution of the neuron weights after 50, 130, and 200 training iterations, respectively. It can be seen that the weights spread out gradually over the original space as the training proceeds. As shown in Figure 5, the KFM model requires nearly 100 to 300 iterations to complete the training.

Multilayer Neural Network

As mentioned before, the authors prepared a multilayer network with a neuron in the input layer and a neuron in the output layer for two-dimensional analysis. The synaptic weights were adjusted by the back-propagation method.

In this paper the authors adopted a training procedure (9) that is somewhat different from the usual one. Here, the authors adjust the weights thoroughly for an input pattern until the error between the output signal and the target signal becomes sufficiently small. The adjustment is repeated for all input patterns. The completion of adjustment for an input pattern deteriorates the synaptic weights for the other patterns, so that those training processes are iterated hundreds or thousands of times, normally 10,000 to 30,000 times. The training method adopted here was effective in avoiding entrapment into a local minimum and converged steadily to a global minimum.

RESULTS

In presenting how well the neural network models describe the nonlinear phenomena without any specific functions, the authors compare two methods: an analytical one by nonlinear equations, and one using artificial intelligence through neural network models. However, the authors refrain from interpreting the curves from the traffic flow viewpoints because there is much to do before doing so, including determining the appropriate number of integrated data points.

Occupancy-Speed Curve

First, the methods are compared using the traffic data observed at Detector Station 1201. The period is 2 hr. Figure 6 shows three regression curves: (a) a curve by a nonlinear equation for original observed data, (b) one by a neural network model without the KFM model, and (c) one by a neural network model with the KFM model.
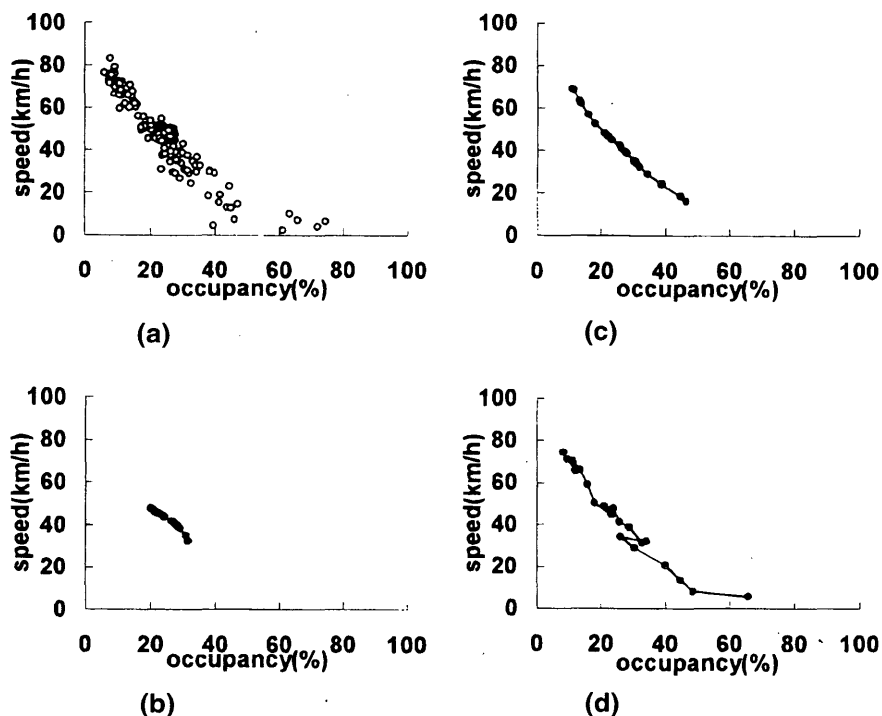
FIGURE 5 Evolution into integrated data by Kohonen feature mapping: *a,* 0 iterations; *b,* 50 iterations; *c,* 130 iterations; *d,* 200 iterations.

The white circles in Figures 6*a* and *b* present 120 points of original observed data, and the black ones in Figure 6*c* present 20 points of data integrated by the KFM model.

It is seen in Figures 6*a* and *b* that the observed data are excessively distributed in both regions where time occupancy is from 5 to 15 percent and from 20 to 40 percent. Those excessive data points affect the regression curve very badly. It should be noted that the shape of the curves is quite different in the high-occupancy region (occupancy is more than 40 percent), although there is little difference in the correlation coefficients, as presented in Table 1. This means that the densely distributed data in the low- and middle-occupancy regions almost govern the curve, and to the contrary, the data in the high-occupancy region have little effect on it.

On the other hand, Figure 6*c* shows 20 points of integrated data and the regression curve by the neural method with the KFM model. It is seen that by introducing the KFM model, the authors were able to make the original data more uniformly distributed. In particular, the five original data in the high-occupancy region in Figure 6*a* are reduced to two sets of data in Figure 6*c*. This favorably improved the regression in the region. One can see that the regression curve with the KFM is located in the middle of the original observed data in the high-occupancy region. This appears to be desirable for applying the curve in a traffic simulation model. However, for interpreting traffic flow phenomena in the region, the overtrimmed curve is not adequate. In such cases, one should increase the number of integrated data or use original raw data.

Figure 7 shows the regression curves for the other detector stations. As in Figure 6, the white circles are the original data, the black are the integrated data, and the thick line is the regression curve pro-

duced by the multilayer neural method. One realizes at a glance that the regression curves are more complicated than those of the nonlinear equation in Figure 6*a*. It is seen in the low-occupancy region that the curves have a "snake head": they are nearly flat where time occupancy is less than 15 percent. In addition, the regression curves consist of a few convex parts. In other words, they are discontinuous in their derivatives. Likewise, a small gap can be seen around the time occupancy of 20 percent in Figure 7*b*.

In this way, the neural network method has the promising ability to describe a discontinuous relationship more precisely. It needs neither to divide the whole region into several nor to introduce an individual function for each region. Unfortunately, those features of the neural network models are not easy to evaluate quantitatively. However, the correlation coefficients reflect those features indirectly. Table 1 presents the coefficients produced by both of the neural methods along with those produced by the nonlinear equation for all cases. It is seen that the neural methods are better than the nonlinear equation. Also, there is little difference between both of the neural methods. This means that the neural network models can flexibly self-adjust the curvature of regression curves according to the number of data points. Needless to say, the neural method with the KFM model is more efficient in the computation than that without the KFM model.
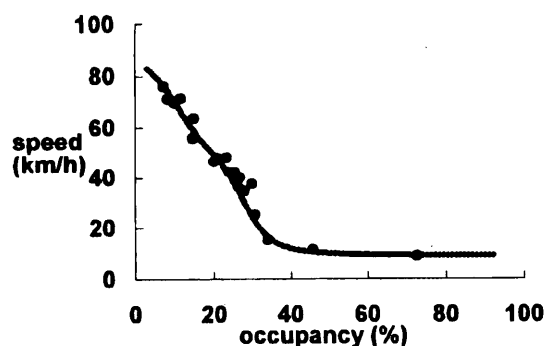
Occupancy-Flow Curve

Figure 8 presents the regression results by both methods for the occupancy-flow curve at Detector Station 1201. Compared with the occupancy-speed curve in Figure 6, the behaviors are a bit more
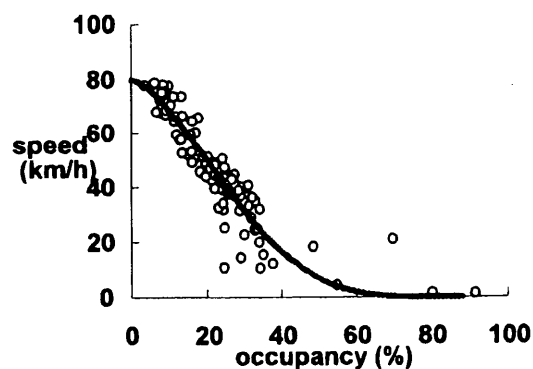
(a)
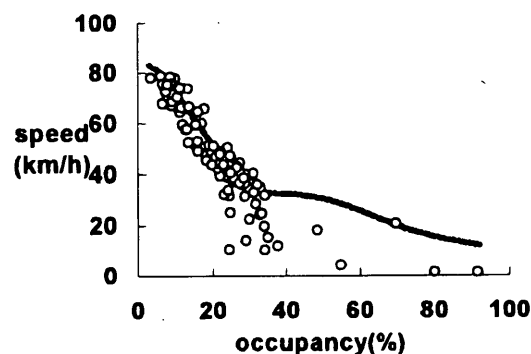
(b)

(c)

FIGURE 6 Comparison of neural network models with nonlinear equation on occupancy-speed curve: *a,* nonlinear equation; *b,* neural network without KFM; *c,* neural network with KFM.

TABLE 1 Comparison of Multiple Correlation Coefficients on Occupancy-Speed Curve

Detector Point	Non-linear Equation	Neural Network	
		without KFM	with KFM
1009	0.94	0.97	0.97
1011	0.87	0.92	0.94
1103	0.91	0.94	0.95
1201	0.88	0.91	0.92
1203	0.92	0.96	0.97

not describe such data. The description for such data is the most difficult subject in the mathematical formulations.

Figure 9, similar to Figure 7, shows the regression curves by the neural method for the other cases. One can see that the distribution of integrated data is more complicated than that of those in the occupancy-speed curves in Figure 7: the thin curve that connects the integrated data in sequence has two peaks. It should be noted that the regression curve (thick line) in Figure 9*a* corresponds well to the movement of the data. In this way, the neural method is able to describe such a complex relationship, too. Here also, one must care-
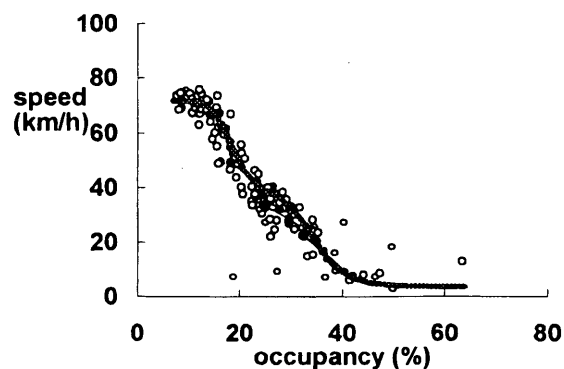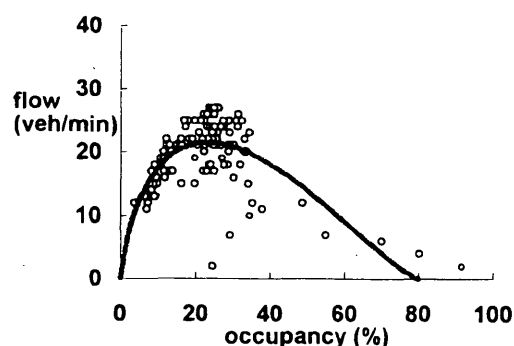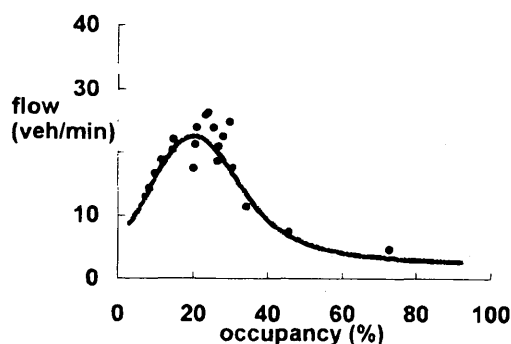

(a)
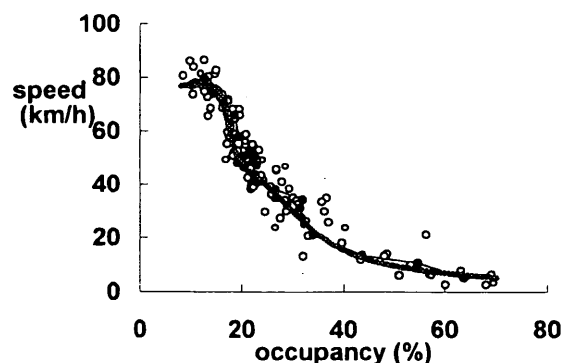
(b)

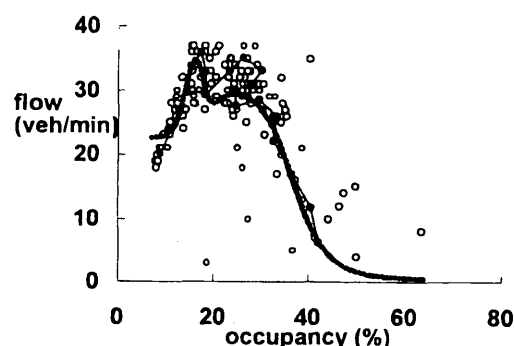**FIGURE 7 Occupancy-speed curves by neural network models: *a,* Station 1011; *b,* Station 1103.**

complex. Clearly, the nonlinear equation in Figure 8*a* fails to describe the relationship in the congested region. On the contrary, as shown in Figure 8*b,* one can recognize the good regression in the region. Integration of original data in the high-occupancy region into a few data points contributed to this improvement. Of course, it must be examined carefully if the number of data points in the region is sufficient or not, according to the purpose for which the curve is used. In addition, one can see that the curve is not so well regressed in the vicinity of capacity, apparently because of the data being scattered in the region. That is, even the neural method can-
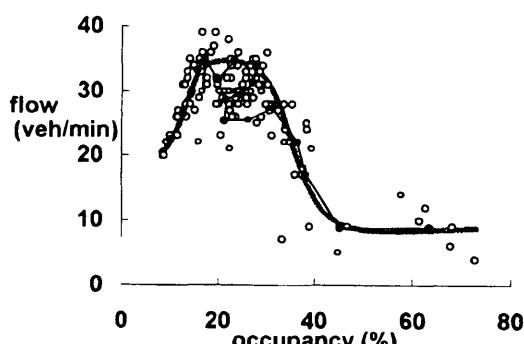
FIGURE 8 Comparison of (*a*) neural network
models with (*b*) nonlinear equation on occupancy-
flow curve, Station 1201.

FIGURE 9 Occupancy-flow curves by neural
network models: *a*, Station 1011; *b*, Station 1203.

fully examine the validity of the curve from traffic engineering
viewpoints.

On the other hand, the regression curve in Figure 9*b* is relatively
smooth, although the integrated data are distributed zigzag as in
Figure 9*a*. This is because even the neural network model is not able
to describe such a function that has two or more state values for a
control value. In this case, there are two or three flow values for an
occupancy value near capacity. Anyway, it should be noted that the
curves are not so well regressed yet in the vicinity of capacity in
both of the figures. For reference, the correlation coefficients for all
cases are given in Table 2. One can see that the neural method is
much better than the nonlinear equation.

Flow-Speed Curve

In general, flow-speed curves become more complicated because of
the transition of traffic states (*10*). They would take a different path
according to whether the traffic goes into congestion or recovers to
free-flow state. However, in this paper, neglecting those dynamic
behaviors, the authors treated traffic states as static ones. Figure 10,
similar to Figures 7 and 9, shows regression curves for two cases,
in which the authors treated speed as the control variable and flow
as the state variable. Because of the lack of observed data in the
free-flow state, the regression curve cannot be seen in the high-
speed region. The curve in Figure 10*a* presents a somewhat poor
regression with the integrated data points around capacity whereas

the one in Figure 10*a* follows them somewhat better. To trace the
data in Figure 10*a* more precisely, it may be necessary to change the
number of data points. But this should be done only if it is mean-
ingful from the viewpoint of traffic engineering. Here also, as indi-
cated in Table 3, the neural method gives better correlation coeffi-
cients than the nonlinear equation.

CONCLUDING REMARKS

The relationships among traffic flow variables play important roles
in traffic engineering. They are used not only in analyses of traffic
flow behaviors but also in some macroscopic traffic flow simulation
models. Noting that some neural network models have promising
abilities to represent nonlinear behaviors and to self-organize auto-
matically, the authors applied them to the description of the rela-
tionships. First, the authors introduced a KFM model to integrate
the original observed data points into fewer, more uniformly dis-
tributed ones. Next, a multilayer neural network model was used to
describe the relationships between traffic flow variables. the authors
investigated the applicability of the neural network models to the
regression problem and compared the results with those produced
by a conventional nonlinear equation. The major findings are as fol-
lows:

1. A KFM method served to integrate original observed data
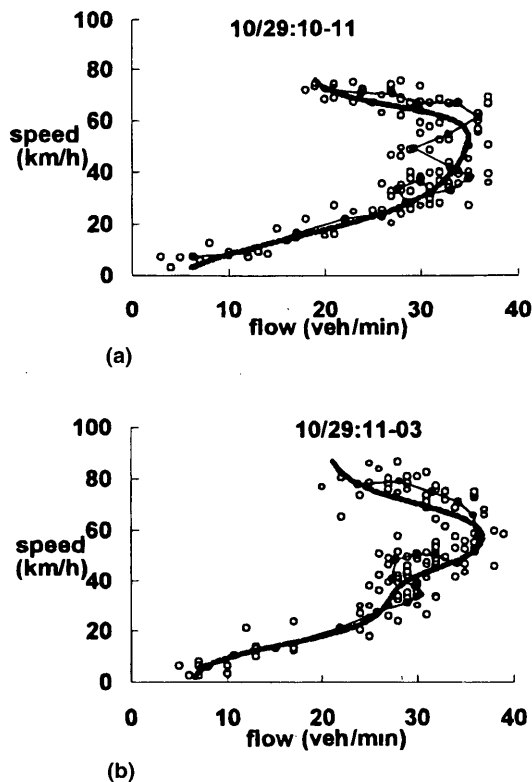points into fewer, more uniformly distributed data points. All that

(a)

(b)

FIGURE 10 Flow-speed curves by neural network models: *a,* Station 1011; *b,* Station 1103.

TABLE 2 Comparison of Multiple Correlation Coefficients on Occupancy-Flow Curve

Detector Point	Non-linear Equation	Neural Network with KFM
1009	0.60	0.78
1011	0.47	0.58
1103	0.74	0.79
1201	0.52	0.66
1203	0.61	0.80

must be done to specify the desired number of integrated data points. This integration contributes to the improvement of regression precision and computational efficiency.

2. A multilayer neural network model was effective in describing the nonlinear and discontinuous relationships between traffic flow variables. The model made it unnecessary to specify the regression curves and the transition points in advance. In addition, the multiple correlation coefficients produced by the model were better than those produced by a nonlinear equation.

TABLE 3 Comparison of Multiple Correlation Coefficients on Flow-Speed Curve

Detector Point	Non-linear Equation	Neural Network with KFM
1009	0.70	0.83
1011	0.74	0.85
1103	0.79	0.82
1201	0.63	0.81
1203	0.75	0.81

The method proposed here still has some disadvantages: it requires a bit of burdensome work to estimate some fundamental traffic parameters, such as maximum volume, which are significant for analyzing traffic flow behavior.

In this paper, the discussion was limited to the availability of neural network models. The interpretation of traffic phenomena using them is left to future work. Moreover, the availability of other neural models that might be more effective than those used here must be examined.

ACKNOWLEDGMENTS

REFERENCES

1. Gerlough, D., and M. Huber. *Special Report 165: Traffic Flow Theory.* TRB, National Research Council, Washington, DC., 1976.
2. McShane, W. R., and R. P. Roess. *Traffic Engineering.* Prentice-Hall, Englewood Cliffs, N.J., 1990.
3. May, A. D. *Traffic Flow Fundamentals.* Prentice-Hall, Englewood Cliffs, N.J., 1990.
4. Nakatsuji, T., and T. Kaku. Improvement of Traffic Flow Simulation Precision by Direct Usage of Traffic Detector Data. *Proc., Infrastructure Planning,* Vol. 16, No 1, 1993, pp. 115–120.
5. Wasserman, P. D. *Neural Computing.* Van Nostrand Reinhold, 1989.
6. Sakamoto, Y., M. Ishiguro, and G. Kitagawa. *Information Statistics.* Kyoritu, Japan, 1983.
7. Dayhoff, J. *Neural Network Architecture.* Van Nostrand Reinhold, 1990.
8. Acha-Daza, J. A., and F. L. Hall. Application of Catastrophe Theory to Traffic Flow Variables. *Transportation Research,* Vol. 28B, No. 3, 1994, pp. 235–250.
9. Nakatsuji, T., and T. Kaku. Development of a Self-Organizing Traffic Control System Using Neural Network Models. In *Transportation Research Record 1324,* TRB, National Research Council, Washington, D.C., 1991, pp. 137–145.
10. Gunter, M. A., and F. L. Hall. Transition in the Speed-Flow Relationship. In *Transportation Research Record 1091,* TRB, National Research Council, Washington, D.C., 1986, pp. 1–9.