**MODEL-BASED TRAFFIC MONITORING BY MEANS OF NEURAL NETWORKS**

Dimitri Lefebvre[1,2], Philippe Thomas[1], Jean Marc Thiriet[3], Nadhir Messai[1] and Abdellah El Moudni[1]
[1]UTBM – SET - Site de Belfort - 90010 Belfort - France
[2]GEII - IUT Belfort Montbéliard - 90016 – Belfort – France
[3]CRAN - ESSTIN - 54500 – Vandoeuvre - lès - Nancy - France
E-mail:dimitri.lefebvre@utbm.fr

**Abstract:** This article is about traffic monitoring by the use of neural networks. Magnetic sensors are used in order to extract on-line the flow and density variables. Such variables are processed by the monitoring network in order to detect and isolate incidents that disturb the traffic. Our approach is based on a macroscopic model, and more precisely on the fundamental flow-density diagram that characterizes the traffic. An admissible region is defined in the flow-density space from this diagram. Incidents are detected when the measured data are out of the admissible region. The classification properties of neural networks are used to design the monitoring network. The proposed method is applied to the monitoring of a complex road junction in the city of Nancy in France.

**Key words:** traffic monitoring, incidents detection, neural networks, classification.

**Acknowledgements :** the authors thank the city of Nancy for its contribution.

## 1. Introduction

Increasing demands on safety for transportation systems has created an increasing interest in the traffic monitoring that concerns the on-line detection and isolation of incidents. Incidents are defined as non-permitted deviations of certain characteristic variables of the system. The principle of model-based monitoring methods is to compare the system behavior with that of a reference model. When their behaviors are different, an incident is detected and isolated (Isermann 1984). Models of the traffic were developed by many authors during the two last decades. They are divided into two classes. On the one hand, mesoscopic affectation models were proposed to describe the paths assignment. CONTRAM (Leonard *et al.* 1989) and DYNASMART (Jayakrishnan *et al.* 1994) are two famous examples for affectation models. Such models are not well adapted for traffic monitoring. On the other hand, macroscopic models were proposed to describe the flows evolution. INTEGRATION (Van Aerde 1995). STRADA (Buisson *et al.* 1996). SIMAUT (Morin 1984) and META (Papageorgiou *et al.* 1990) are among the most famous ones. Such models are useful for monitoring purposes.

The model-based traffic monitoring requires intermittent or continuous collection of measured data. Such data are numerous and must be processed on-line. Because of their massive parallelism, neural networks are well adapted for real time processing. Moreover, neural networks are good classifiers (Lippmann 1987). For these reasons neural networks are useful for monitoring purposes. The goal of our work is to propose a model-based method for traffic monitoring by means of neural networks. Our approach is based on a macroscopic model of the traffic flow in order to detect and isolate the defaults. Our contribution is to adapt the fundamental density – flow diagrams for monitoring purpose. The incident detection and isolation is described as a classification problem in the flow-density space. A neural implementation of the proposed monitoring function is detailed. The second section is a brief introduction to the traffic flows representations that are used as reference models. In the section 3, the monitoring neural network is described. The section 4 is a concrete application for the proposed method.

## 2. Traffic flow modelling

Models of the traffic are generally composed of three sub-models that are strongly connected (Bell *et al.* 1997, Buisson 1996). Directed graphs are required to represent the network structure. Nodes stand for the junctions and links stand for the section of roads. Such models are based on node-link incidence matrices. Model of the traffic flows are also required. Macroscopic models are based on hydrodynamics theory and mesoscopic ones are based on queuing theory. The first class treats the traffic as a compressible fluid, and the second class quantizes traffic and concentrate on the arrival and departure processes at particular points of the network. At last, models are required that describe the traffic affectation. Such models are useful to determine the paths that are chosen by the vehicles drivers. They are based on origin-destination-path incidence matrices, and on the evaluation of cost functions.

### 2.1 Flows models

Macroscopic models consider the traffic as a compressible fluid that circulates on the $p$ links $(L_j)_{j=1\ldots p}$ between the $q$ nodes $(N_i)_{i=1\ldots p}$ of the network (Leutzbach 1988, Lightill *et al.* 1955). For each link $(L_j)_{j=1\ldots p}$ the traffic is described in term of speed $v_j(x,t)$, flow $q_j(x,t)$ and density $d_j(x,t)$. The variable $x$ stands for the spatial measure on the considered link, and $t$ stands for the time. Such models are based on the 3 following equations.

$$\frac{\partial d_j(x,t)}{\partial x} + \frac{\partial q_j(x,t)}{\partial t} = 0 \text{ , j = 1…p.} \quad (1)$$

Equation (1) is a conservative relationship of first order. Let us notice that second order relationships are sometimes used instead equation (1) in order to take into account the drivers reactive delays (Papageorgiou *et al.* 1990, Payne 1971).

$$q_j(x,t) = d_j(x,t).v_j(x,t) \quad , j = 1\ldots p. \tag{2}$$

Equation (2) is a dimensional consistency relation: the flow (measured in vehicles per hour) equals to speed (measured in kilometers per hour) multiplied by density (measured in vehicles per kilometer).

$$q_j(x,t) = F(d_j(x,t)) \quad , j = 1\ldots p. \tag{3}$$

Equation (3) is known as the fundamental flow-density diagram (FDD), that is generally obtained by direct observation. Additional relations, related to the queuing theory describe the traffic flow around the nodes $(N_i)_{i=1\ldots p}$ (Bell *et al.* 1997). Such aspects are not considered in this work.

**2.2 Fundamental flow-density diagram**

The FDD given by equation (3) has the following properties. According to the hydrodynamic models, zero flow is encountered at zero density and again at maximum of queue density. Such models suggest two densities and also two speeds for any flow lesser than the maximal flow: a lower speed corresponding to the higher density and a higher speed corresponding to the lower density. The lower density (or higher speed) refers to non-congested flow, and the higher density (or lower speed) refers to congested flow. In non-congested conditions, flow is generally found to be stable, and in congested conditions instability prevails. Because of the relation (2), the slope of the line connecting the origin to any point on the FDD gives the speed corresponding to the flow and density by that point. The shape of the FDD depends on the flow model. INTEGRATION FDD is composed of two polynomial curves (Van Aerde 1995). STRADA FDD is composed of two parabolic curves (Buisson *et al.* 1996). SIMAUT FDD is composed of polynomial curve of degree 3 and a line (Morin 1984). At last, META FDD is composed of an exponential curves (Papageorgiou *et al.* 1990). Among the existing flow models, STRADA was proved to be suitable not only for urban highway, but also for complex junctions (figure 1) (Buisson *et al.* 1996). For this reason, we use STRADA FDD in the following, to work out the monitoring network.

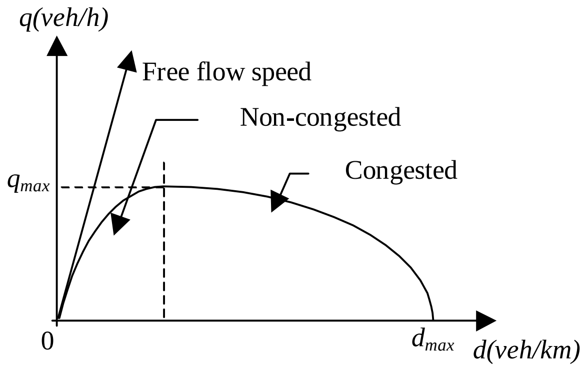



Figure 1: FDD for STRADA

**2.3 Traffic sensors**

In most urban and interurban networks, traffic sensors are magnetic loops implanted in the causeway. Video camera, pneumatic sensors, acoustic sensors and so on are also used, but are not considered in this work (Thiriet *et al.* 1995). A magnetic loop delivers an analog signal that is first transformed into a binary output extracted with a sampling period and sent to a central computer (figure 2).

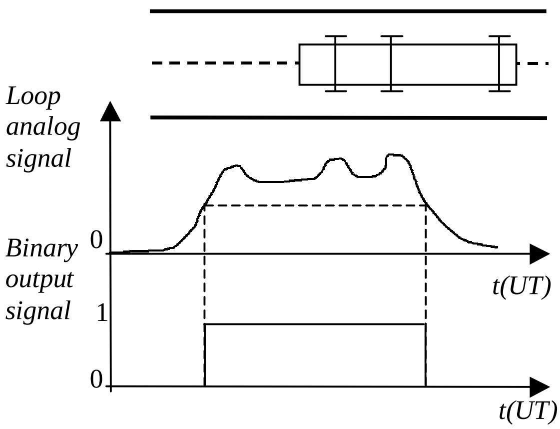



Figure 2: Magnetic loop output signal

This computer receives information emitted by each link $(L_j)_{j=1\ldots p}$ of the network and elaborates the macroscopic parameters that represent the traffic evolution. The shape of the magnetic loop is mostly a rectangular one. The size and implantation depend on the parameters that must be measured. Long loops are useful to measure the occupation rate of vehicles and short loops are preferable to measure the flow. The flow is evaluated as the numbers of impulses in the binary output signal during a time unit (TU). The occupation rate ($R_O$) is evaluated as the ratio of time during which the binary output signal equals 1. According to some assumptions, the occupation rate gives an estimation of the density in the particular point where the loop is placed (Buisson 1996):

$$d(t) = \frac{R_O(t)}{(L_V + L_l)}. \tag{4}$$

where $L_V$ stands for the average length of the vehicles and $L_l$ stands for the length of the loop.

The speeds estimation can be obtained with long loops as short loops. The speed can be obtained directly from the slope of the analog signal when this signal is available. When not, it could be obtained from the duration of each impulse. But in this case, the average length of the vehicles must also be fixed.

## 3. Monitoring network

### 3.1 Artificial non-linear static neural networks

An artificial non-linear static neuron (Lipmann 1987, Rosenblatt 1962, Widrow *et al.* 1990) is composed by an adaptive linear combiner $\Sigma$ and an activation function *f(.)* (figure 1). The node receives an input signal vector ($x_1$, $x_2$,…, $x_n$), and produces an output signal $y$.

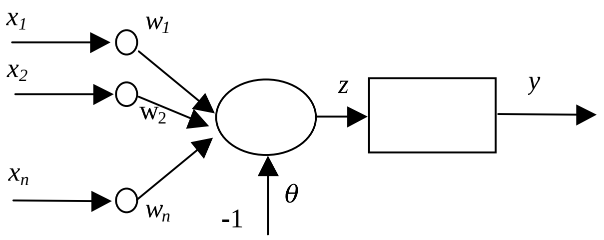


Figure 3: Structural scheme of an artificial static neuron

The output of the adaptive linear combiner $\Sigma$ is given by equation (5):

$$z = \sum_{i=1}^{n} w_i . x_i - \theta \ , \quad (5)$$

where $(w_i)_{i=1,..,n}$ are the weights of the node. The bias weight $\theta$ that is connected to the constant input -1. The most useful activation functions *f(.)* are linear or threshold linear ones for approximation problems and hard limiting quantizer or sigmoid ones for classification problems (Lippmann 1987, Hornik *et al.* 1989). The use of hard limiting quantizers results in adaptive threshold logic elements that produce a binary output. $\theta$ effectively controls the threshold level of the quantizer. An adaptive threshold logic element is able to classify an input pattern vector $(x_1, x_2, \ldots, x_n)$ into two categories A ($y$ = 0) and B ($y$ = 1). For example, a node with 2 inputs classifies the vectors $(x_1, x_2)$ into categories A and B that are separated by the line *D* (figure 4). The classification depends on the value of the weights according to (6). The output verifies $y = 1$, if $(x_1, x_2)$ is in A, or $y = 0$, if $(x_1, x_2)$ is in B:

$$z = w_1 . x_1 + w_2 . x_2 - \theta ,$$
$$y = f(z). \quad (6)$$

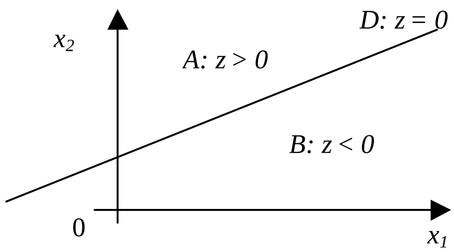


Figure 4: Separating line in space $(x_1, x_2)$

A critical thresholding condition occurs when $z$ = 0. Thus, the equation of the line *D* is given by (7):

$$x_2 = -\frac{w_1}{w_2} x_1 + \frac{1}{w_2} \theta \ . \quad (7)$$

Hard limiting quantizers are usually replaced by sigmoid or semi-sigmoid functions as the one given in equation (8):

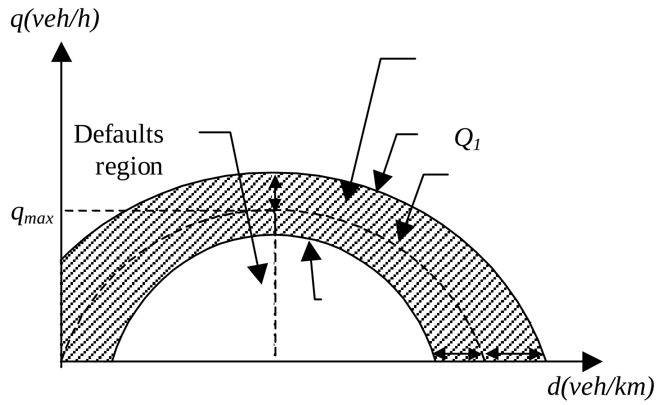
$$y = f(z) = \frac{1}{1 + e^{-\omega . z}} \in \ ]0, 1[, \quad (8)$$

where $\omega$ characterizes the slope of the semi-sigmoid. The semi-sigmoid function (8) is continuously differentiable, and bijective, in comparison with the hard limiting quantizer that is not continuous and not bijective. For this reason, semi-sigmoid functions are used in the sequel instead of hard limiting quantizers.

Adaptive threshold logic elements are useful to solve classification problems. The propagation of information is very fast, because of the massive parallelism of the net architecture. Thus, they are suitable for real time applications as monitoring. Moreover, they are able to tackle unknown situations. This generalization property is interesting for classification problems like diagnosis ones. For all these reasons, neural networks are well adapted for monitoring and diagnosis purposes.

### 3.2 Admissible region

Let us consider the link $L_j$ where the flow and the density are measured with a sampling period $\Delta t$ from the magnetic loops output signals. Let us define $\hat{d}_j^k = \hat{d}_j(k.\Delta t)$ as the measured value of the density in this section at time $t = k.\Delta t$. Let us also define $\hat{q}_j^k = \hat{q}_j(k.\Delta t)$ as the traffic flow in the section at the same instant. The complete traffic model is not required for monitoring and diagnosis applications. The fundamental diagram provides all information that is required for this purpose. A default is detected when the measured data are too far from the FDD. The defaults isolation results from the separated modeling of each link.




The fundamental diagram is used to separate the space *(d, q)* into 2 regions: a default region and an admissible region (figure 5, index *j* is omitted). According to the FDD of STRADA (figure 1), the admissible region can be bounded by two elliptic curves $Q_0$, and $Q_1$. The default region corresponds to the rest of the space *(d, q)*. $Q_0$, and $Q_1$ are given by equation (9) where the parameters $d_{max}$ and $q_{max}$ are related to the FDD. The admissible region depends on 4 complementary parameters $(e_i^+)_{i=0,1}$, and $(e_i^-)_{i=0,1}$, that correspond to the maximal permitted variations of the measured flow $\hat{q}_j^k$ and measured density $\hat{d}_j^k$ ( figure 5).

$$Q_1 \ : \ \frac{(d^k)^2}{(d_{max} + 2.e_0^+)^2} + \frac{(q^k)^2}{4.(q_{max} + e_1^+)^2} - \frac{d_{max}.d^k}{(d_{max} + 2.e_0^+)^2} - \frac{e_0^+.(e_0^+ + d_{max})}{(d_{max} + 2.e_0^+)^2} = 0, \quad (9a)$$

$$Q_0 : \quad \frac{(d^k)^2}{(d_{max}-2.e_0^-)^2} + \frac{(q^k)^2}{4.(q_{max}-e_1^-)^2} - \frac{d_{max}.d^k}{(d_{max}-2.e_0^-)^2} - \frac{e_0^-.(e_0^- - d_{max})}{(d_{max}-2.e_0^-)^2} = 0. \tag{9b}$$

### 3.3 Monitoring network

The monitoring network is composed of 3 parts: an arithmetic pre-processor, a layer of two adaptive threshold logic elements and a logical post-processor (Lefebvre 1999).

The arithmetic pre-processing is required to process multiplicative operations of the input signals. Because the admissible region is bounded by elliptic curves, terms $(\hat{d}^k)^2$, $(\hat{q}^k)^2$, and $(\hat{q}^k).(\hat{d}^k)$.must be worked out from the measured variables $(\hat{d}^k)$ and $(\hat{q}^k)$.

The adaptive threshold logic elements have semi-sigmoid activation functions *f(.)* defined by equation (8). Each node has 5 analog inputs $(\hat{d}^k)$, $(\hat{d}^k)^2$, $(\hat{q}^k)$, $(\hat{q}^k)^2$, and $(\hat{q}^k).(\hat{d}^k)$.and one analog output $(y_i)_{i=0,1} \in$ ]0, 1[. Each curve $(Q_i)_{i=0,1}$, separates the pattern space $(d, q)$ into 2 regions, and each node is capable to classify the pattern vector $(\hat{d}^k, \hat{q}^k)$ in these regions according to equation (10).

$$z_i^k = w_{0i}.\hat{d}^k + w_{0i}^2.(\hat{d}^k)^2 + w_{1i}.\hat{q}^k + w_{1i}^2.(\hat{q}^k)^2 + w_{01i}.\hat{d}^k.\hat{q}^k - \theta_i, \quad i=0,1. \tag{10}$$

$$y_i^k = f(z_i^k),$$

Weights of the adaptive threshold logic element are given in the table 1.

| | $Q_1$ | $Q_0$ |
|---|---|---|
| $w_{0i}$ | $\frac{-d_{max}}{(d_{max}+2.e_0^+)^2}$ | $\frac{-d_{max}}{(d_{max}-2.e_0^-)^2}$ |
| $w^2_{0i}$ | $\frac{1}{(d_{max}+2.e_0^+)^2}$ | $\frac{1}{(d_{max}-2.e_0^-)^2}$ |
| $w_{1i}$ | 0 | 0 |
| $w^2_{1i}$ | $\frac{1}{4.(q_{max}+e_1^+)^2}$ | $\frac{1}{4.(q_{max}-e_1^-)^2}$ |
| $w_{01i}$ | 0 | 0 |
| $\theta_i$ | $\frac{e_0^+.(e_0^+ + d_{max})}{(d_{max}+2.e_0^+)^2}$ | $\frac{e_0^-.(e_0^- - d_{max})}{(d_{max}-2.e_0^-)^2}$ |

Table 1: Weights of the monitoring network

The analog outputs $(y_i^k)_{i=0,1}$ must first be rounded off to the nearest integer value and then combined with the logical function (11) in order to obtain the monitoring function $y^k$. A default is detected at time $t = k.\Delta t$, $k>0$, if and only if the monitoring function verifies $y^k = 1$.

$$y^k = y_1^k + (y_0^k)' . \tag{11}$$

where "+" and " ' " stand for the logic operators OR and NOT. The monitoring function can be rewritten as:

$$y^k = max(y_1^k, 1 - y_0^k). \tag{12}$$

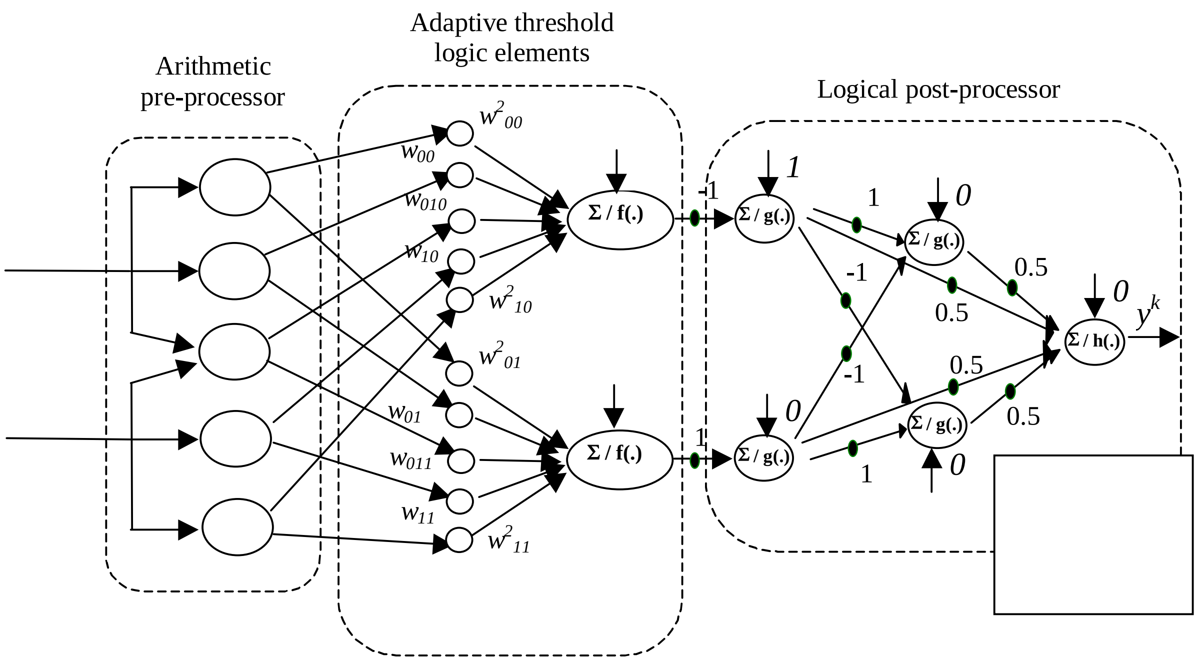



Figure 6: Monitoring network of a link

Such a function can be obtained with neural networks (Lippmann 1987). A single-input single-output node with a linear activation function *h(.)* is able to build the logical NOT operator. A comparator subnet, called MAXNET selects the maximal value of two inputs *a* and *b*, to build the logical OR operator. Such a net requires two threshold linear activation functions *g(.)* and one linear one *h(.)*. It is efficient even if the inputs of the nets are not exact binary values as the ones obtained with semi-sigmoid activation functions *f(.)*. Thus rounding functions are not necessary when using the MAXNET as OR operator. A default is detected at time $t = k.\Delta t$, $k>0$, if and only if the monitoring network verifies $y^k > 0.5$. The architecture of this net consists in an arithmetic pre-processor cascaded with 2 linear classifiers and with 1 MAXNET to build the logical post-processor (figure 6).

## 4. Application

Let us consider the road junction C38 in the city of Nancy in France (figure 7) as an example.

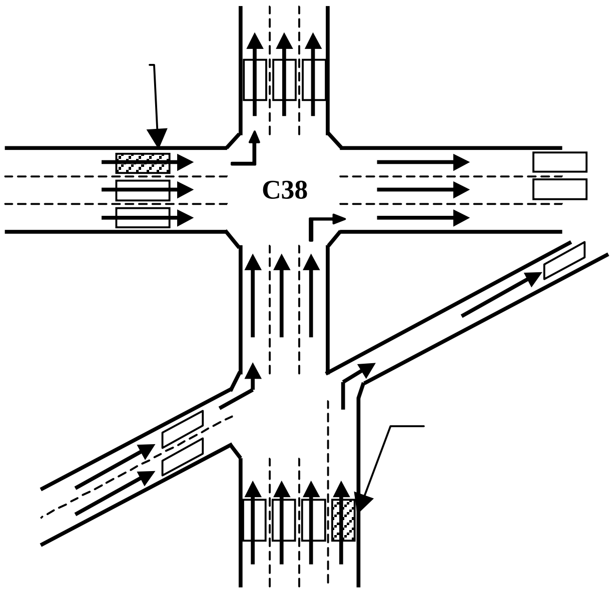



Figure 7: Road junction C38

This junction is monitored with 14 magnetic loops that are also used for the regulation of the traffic signals in junction C38. Among these sensors, let us focus on the loops B51 and B45. The loop B51 is placed on a turn to left line that avoids the center of the city. This line is generally non-congested. On the contrary, the loop B45 stands on a turn to right line that drives to the city center. This line is often congested.

The binary output signals of B51 and B45 are extracted with a sampling period $\delta t = 0.1$ s during a time interval of 3 hours. To avoid the queuing behaviors due to the traffic signals, densities and flows are worked out for variable time periods $\Delta t \in [55, 95]$ seconds that are synchronized with the cycles of the traffic signals. The flows $\hat{q}_j^k = \hat{q}_j(k.\Delta t)$ result directly from the number of impulses in the loops binary output signals during $\Delta t$. The densities $\hat{d}_j^k = \hat{d}_j(k.\Delta t)$ are obtained with equation (4) from the evaluation of the occupation rates during $\Delta t$. Both loops have a length $L_l = 1$ m. The average length of vehicles is fixed such that $L_V = 3.5$ m. The proposed monitoring concerns only the neighborhood of the loops. Admissible regions for both loops are similarly defined with the following parameters: $d_{max} = 220$ veh/km , $q_{max} = 12$ veh/min, $e_0^+ = 15$, $e_0^- = 5$, $e_1^+ = 3$ and $e_1^- = 2.5$ (figures 8 and 9). The couples of measured variables $(\hat{d}_j^k, \hat{q}_j^k)$ are reported in the admissible and defaults regions. Let us notice that traffic conditions are different in the neighborhood of B45 and B51. Data that are reported for B51 correspond to non-congested traffic conditions, while data that are reported for B45 correspond to congested traffic conditions.

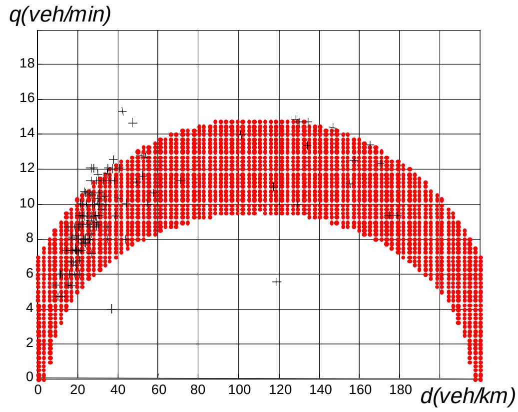



Figure 8: Measured data for B51

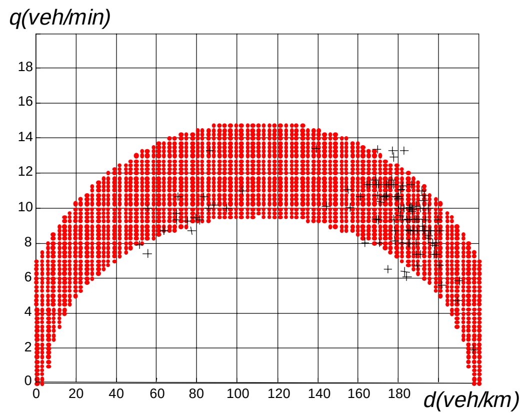



Figure 9: Measured data for B45

Defaults are detected when couples of data $(\hat{d}_j^k, \hat{q}_j^k)$ are reported outside the admissible regions. A monitoring network is designed for each line that is monitored. Due to the semi-sigmoid activation *f(.)* function, the outputs $y^k$ of the networks are analog ones (figures 10a and 11a). Analog signals are rounded off to detect defaults (figures 10b and 11b) Let us distinguish intermittent defaults that are no significant and persistent ones that correspond to traffic incidents. During the considered time interval, no incident was detected for the junction C38, but several intermittent defaults occur.

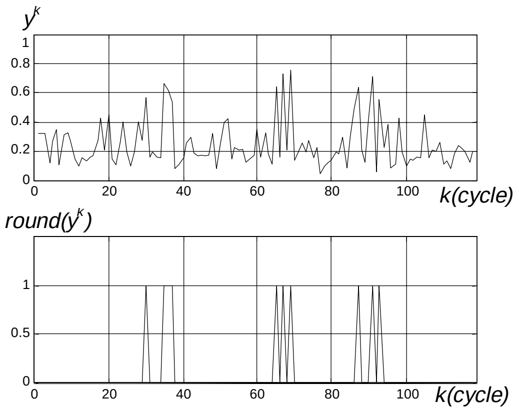



Figure 10: Monitoring of B51

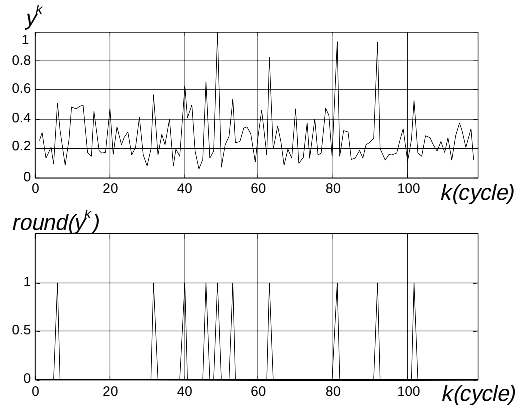



Figure 11: Monitoring of B45

As a conclusion let us notice that the performances of the monitoring network are strongly related to the model of the traffic flow. Such performances can be improved with a more accurate identification of the FDD. Another important remark concerns the sensitivity of the monitoring function. The slope parameter $\omega$ ($\omega$ = 30 for our monitoring network) of the activation function in equation (8) can be adapted to modify the analog output.

## 5. Conclusions

Our contribution was to propose a static monitoring network based on the fundamental flow-density diagram of the STRADA traffic model. An admissible region that is centered around this diagram was defined. The measured flow and density variables were reported in the flow-density space in order to detect and isolate the incidents that disturb the traffic. The main interest of the proposed method is to provide a local monitoring for each magnetic sensor. Several sensors are not required. Monitoring was first stated as a logical function, and then turned into a neural network. The advantages of neural networks for this purpose were also pointed out. The massive parallelism and the classification properties were used in order to satisfy the real time constraints of monitoring applications.

Our further works are to investigate the learning properties of neural networks in order to design data-based monitoring network that are not related to any traffic model.

## 6. Reference


Bell M.G., Iida Y., 1997, *Transportation network analysis*, Wiley.

Buisson C., 1996, Comparaison sur un réseau simple, de modèles macroscopiques d'écoulement du trafic routier, INRETS, no. 52 pp.15-35.

Buisson C. Lebacque J.P. Lesort J.B. Mongeot H., 1996, The STRADA dynamic assignment model. 3rd World Congress on Intelligent Transport Systems, Orlando, USA

Cremer M., Papageorgiou M., 1981, Parameter identification for a traffic flow model, *Automatica*, no. 17, pp. 837-844.

Hornik K., Stinchcombre M., White H., Multilayer feedforward networks are universial approximaters, Neural networks, vol. 2, no. 5, pp. 359-366, 1989.

Isermann R., 1984, Process fault detection based on modeling and estimation methods, *Automatica*, vol. 20, pp. 387 - 404.

Jayakrishnan R., Mahmassani H.S., Hu T.Y., An evaluation tool for advanced traffic information and management systems in urban networks, Transportation Res., no. 2C, pp. 129-147 1994

Lefebvre D. ,1999, Failure detection and isolation for manufacturing systems, *Int. J. of Mech. Prod. Syst. Eng.*, no. 2, pp.33 – 44.

Leonard D.R., Gower P. Taylor N.B., CONTRAM: structure of the model, Rapport TRRL, 1989

Leutzbach W., 1988, *Introduction to the theory of traffic flow*, Springer-Verlag.

Lightill M.J., Whitham G.B., 1955, On kinematic waves, II. A theory of traffic flow on roads. *Proc. of the Royal Society*, no. 229A, pp. 317-345.

Lippmann R., 1987, An introduction to computing with neural networks, *IEEE ASSP Mag.*, vol. 4, no. 2, pp.4-22.

Morin J.M., 1984, SIMAUT un programme de simulation du trafic autoroutier, *Rech. Trans. Sécu.*, no. 4, pp. 15 - 20.

Papageorgiou M., Blosseville J.M., Hadj-Salem H., 1989, Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris, *Transportation Res.* no. 23B, pp. 29-47.

Papageorgiou M., Blosseville J.M., Hadj-Salem H., 1990, Modelling and real time control of traffic flow on the southern part of Boulevard Périphérique in Paris, Part I Modelling, *Transportation Res. A*, no. 24, pp. 345-359.

Payne H.J., 1971, Models of freeeway traffic and control, in *Simulation Council.*

Rosenblatt F., 1962, *Principles of neurodynamics : Perceptrons and the theory of brain mechanisms*, Spartan Books.

Thiriet J.M., Noizette J.L., Robert M., Riviere J.M., 1995,A smart or intelligent sensor for road and urban traffic, 28th ISATA, Stuttgart, Germany.

Van Aerde M., 1995, INTEGRATION release 2: User's guide, Transportation Syst. Res. Group, Ontario, Canada

Widrow B., Lehr M.A., 1990, 30 years of adaptative neural networks: Perceptron, Madaline, and backpropagation, *IEEE Proceedings*, vol. 78, no. 9, pp. 1415-1442.

Zwingelstein G., 1995, Diagnostic des défaillances : théorie et pratique pour les systèmes industriels, Hermes