Analysis of Complex Networks: An Overview of Methodologies and a Neural Network Application to Intermodal Transport in Italy

Nijkamp, P.; Reggiani, A.; Tritapepe, T.

published in
Networks in Transport and Communication: A Policy Approach
1997

document version
Publisher's PDF, also known as Version of record

15 Analysis of complex networks: an overview of methodologies and a neural network application to intermodal transport in Italy

Peter Nijkamp, Aura Reggiani and Tommaso Tritapepe

15.1 Modelling complex networks: an introduction

Spatial systems are manmade multi-component systems with a high degree of variability and unpredictability. Seemingly simple issues like the assessment of infrastructure impacts on regional development or the estimation of information provision on the behaviour of road users are already difficult to handle. Lack of precision is an inherent feature of spatial research and a source of frustration for planners.

It should be noted however, that the analysis of complex spatial systems - with a variety of interactions among diverse components and layers - is in principle not different from the research task in other disciplines, such as physics or biology. Clearly, spatial systems have an additional degree of complexity compared to natural science systems, viz. human intervention at both an individual and a collective level. But this - basic and farreaching – characteristic should not prevent us from searching for universal analytical principles which may be helpful in better understanding the degree of variation and stochasticity in spatial systems.

An important methodological step forward in this context has been offered by synergetic theory (see amongst others Haken 1983a, 1983b). Synergetics aims to study the complex relationships between interactive components and their consequences for the evolving macro structures by looking for simple universal principles that govern the dynamics of flows and morphology in the organization of a system. Such dynamic forces may generate qualitatively new structures in space and time, while also new functions of the systems may be generated.

Clearly, this idea provokes many new research issues, e.g. regarding resilience, robustness, equilibrium or sustainability of such systems. Synergetics aims at mapping out such changes and their implications by focusing on the qualitative changes in a complex system. The main feature is that the complex nature of a macro system can be characterized by only a few parameters that govern the movement or behaviour of the micro-components that make up particular subsystems. This also explains the search for simple, dynamic and nonlinear equations of motion that can adequately describe (and forecast) the behaviour of a multi-component multiactor system (see also Domanski 1994 and Weidlich and Haag 1983).

Clearly, such research endeavours are not entirely new in regional science. Typical examples of earlier attempts can be found in gravity theory, entropy theory or spatial equilibrium theory. The main feature of modern synergetic methods is their emphasis on qualitative, nonlinear dynamic macro structures, described by a limited set of basic equations. As a result, there is much focus on dissipative structures, phase transitions, slow and fast dynamics, catastrophic behaviour, chaotic dynamics, self-organizing dynamics, and ecologically-oriented evolution (see also Allen and Sanglier 1981; Domanski 1992; Nicolis and Prigogine 1977; and Nijkamp and Reggiani 1992, 1993, 1995).

Therefore, the analysis of complex systems has in recent years become an important research issue in many disciplines. It has intensified the search for universal principles governing nonlinear dynamics systems with a particular interest in methodological approaches. This also explains the current popularity of niche theory, Volterra-Lotka dynamics and predator-prey models in biological and socioeconomic sciences. In a similar vein, we also observe much emphasis on network evolution, as networks are prominent examples of interacting dynamic multi-component systems, in which flow dynamics and morphological evolution play a crucial role. In spatial sciences these systems also have a strong human behavioural aspect (flow behaviour, regulatory regime).

Networks are not only specific, organized spatial structures (i.e. based on node edge interconnectivity), but offer user functions which aim at improving the efficiency of spatial interactions. Thus, morphological structure and spatial interaction are dual phenomena. This is also echoed in the recent conceptual interest in networks as complex sociotemporal systems (see e.g. Batten et al. 1994, and Reggiani and Nijkamp 1996). In this literature, networks are interpreted as spatial economic or socioeconomic systems characterized by evolutionary processes governed by a multilayer, multi-component organization of interdependent subsystems. Clearly, a hierarchical system is only a special case of a more general network ramification.

It should be noted that real world network systems – emerging from the complex interaction between slow and fast dynamics of the various system's components – do not only reflect the complex nature of dynamic systems, but may also mirror auto-organizing powers of the system at hand, which may be depicted by various typologies of evolutionary non-regular behaviour (Reggiani and Nijkamp 1995). A well-known popular concept in this context has become the term 'sustainability', which describes the continuity potential of a dynamic system under changing external conditions and countervailing powers and behaviour inside the system (see Nijkamp 1991). In a recent study, Beckenbach and Pasche (1994) have defined sustainability as a stable domain/corridor in a dynamic spatial economic system. Also related concepts like dynamic fragility and resilience may be mentioned in this context. Resilience may be conceived of as the potential of the system to resist internal/external perturbations. It is thus clear that – once the complex relationship of fragile or resilient system has been identified and formalized for some relevant domains or corridors – it is necessary to investigate for a relevant time horizon (including sometimes multiple generations) the degree of sustainability of such a system in the light of various internal and external shocks to the system at hand. Only in this way is it possible to get a better grasp of the structural tendencies of the system concerned and its dynamic functioning. An illustrative example in this context is offered by transport systems, whose nonlinear evolution may induce a (more or less stabilizing) perturbation within the entire spatial network of regions and cities (including interregional connections).

Clearly, there is a multitude of methodological approaches and analytical tools for the above issues which aim to measure or quantify – or at least formalize or typify – the above mentioned concepts in synergetics. Complex decision problems are increasingly analyzed from the viewpoint of a similarity with the functions and actions of our brains. This is also revealed in concepts like biocomputing and artificial intelligence. In the past decade, much scientific literature has been devoted to neural networks (NNs) as alternative models of information processing (see for a review, Reggiani et al. 1997). Therefore, the theoretical aspects of this new methodology will briefly be described in Section 15.2.

An application to interaction patterns of 67 Italian areas will then be analyzed by using the NN model and the logit model in Section 15.3. This last part refers to previous experiments (see Nijkamp et al. 1996) carried out for modelling the modal split problem between rail and road transport modes in Italy in relation to the introduction of a recent technological breakdown, the High Speed Train. In particular the performance of the models, by varying the attributes, will be compared by testing the models in terms of spatial forecasts. Finally, Section 15.4 contains some concluding remarks.

15.2 Neural networks for the analysis of complex systems

15.2.1 Introduction

Neural network approaches are able to *generalize* from experience, without fixing – *a priori* – any behavioural rule/model among variables. Applications of neural networks (NNs) are nowadays abundant in many disciplines. In regional science this approach has also gained popularity, e.g. in transport and spatial economic interactions (see, for a review, Reggiani et al. 1995). In this approach it is necessary – like in the case of complex networks – to adopt tools that are able to represent connectivity, communication, adaptivity, control and prediction patterns.

The principal inspiration of NNs is the human brain whose structure is made up of a large number of connected neurons. Every neuron is a cell body which receives electrochemical input signals through the dendrites[1] and transmits the resulting electrochemical output signal through the axon,[2] information transmission (electrochemical signals) is then spread by means of spatial connections, the 'synapses', between axons and dendrites of different neurons. In particular, when the combined signals are strong enough, the neuron is 'activated' and will produce an output signal.

The structure of NNs is generally represented by logical units ('neurons') connected by channels of communication ('synapses') which intercompute independently, since each unit cooperates in the transmission of information by means of a different 'weight'.[3] This differentiation in the weights thus corresponds to different values in the synapses. The above phenomenon takes place in particular during the phase of learning in order to allow adaptation to new conditions. In fact, just like cerebral behaviour, NNs are also able to recognize patterns which they have never observed before. This characteristic of 'generalizing'[4] identifies the behaviour of the system as 'intelligent'. In other words, since 'real' events never repeat themselves in the same manner, intelligent systems are able to observe, by means of past experience, continuity and similarity of such events, by offering the possibility of correctly predicting future events.

Within the context of Artificial Intelligence, NNs are distinguished by their ability to elaborate and create information by means of Parallel Distributed Processing systems. This is an important aspect of NNs, since massive parallelism provides, on the one hand, the possibility of significantly increasing computer speed (see Kosko 1992) and, on the other hand, provides a great 'fault tolerance' (since interconnection between units is essentially local).

It is difficult to define NNs in formal terms: in principle they can be considered as non-linear dynamic systems with many freedom degrees as well as 'free' models of estimation (see Kosko 1992). Consequently, the common element in the above definitions is the concept of 'freedom'; in other words, there is 'free biological behaviour' within NN which cannot be subjected to any mathematical model (usually creating logical bounds between output and input). Thus, in contrast to the necessity to program computers (which requires knowledge of the mathematical model which represents the reality concerned), NNs are trained; that is, they learn from a set of 'examples' with input and output. In other words, NNs can be defined as connected systems by means of which a wide range of complex problems, 'not formalizable' by logical mathematical models, can be solved.

Consequently it is interesting to note that NN are suitable in a *forecasting context*, given their ability to *generalize*, i.e., to elaborate new situations for different scenarios. In this framework, it should be noted that this peculiarity strictly depends on both the chosen training set and on the architectural configuration of the network (number of hidden levels, number of units on these levels, etc.) (see e.g. Fischer and Gopal 1994). The first point presupposes the choice of a representative sample in order to get an '*unbiased*' distribution of observations, while the second point implies a 'careful' mapping of NN architecture, since the optimal configuration is reached only by means of experimental methodologies (see also Malliaris and Salchenberger, 1992).

Starting from the above considerations we will briefly describe, in the next subsection, the typology of a 'Two-layer-Feedforward' network, which will be adopted for our NN experiments in transport (see Section 15.3).

15.2.2 The two-layer feedforward neural network

The above-mentioned NN methodology has been widely applied in different fields of regional economics and particularly in transportation science (see, for a review, Reggiani et al. 1997), with reference to the construction of origin-destination matrices, the forecasting of vehicle-flows, the simulation of users' behaviour, etc. The main elements of a feedforward NN can then be illustrated by analyzing the following characteristics:

- NN structure;
- activity of the hidden unit;
- NN learning.

Neural network structure The NN structure can be considered as a weighted and oriented graph, constituted by modes (units) and weighted links (see Figure 15.1).

It is subdivided into layers and each layer contains a different number of units. Essentially, units are linked in three ways: a) 'forward', when the direction of the unit's connection is from the lower layer to the upper layer; b) 'backward', when the direction is inverse with respect to the forward direction; and c) 'lateral', when the units of a layer are connected.

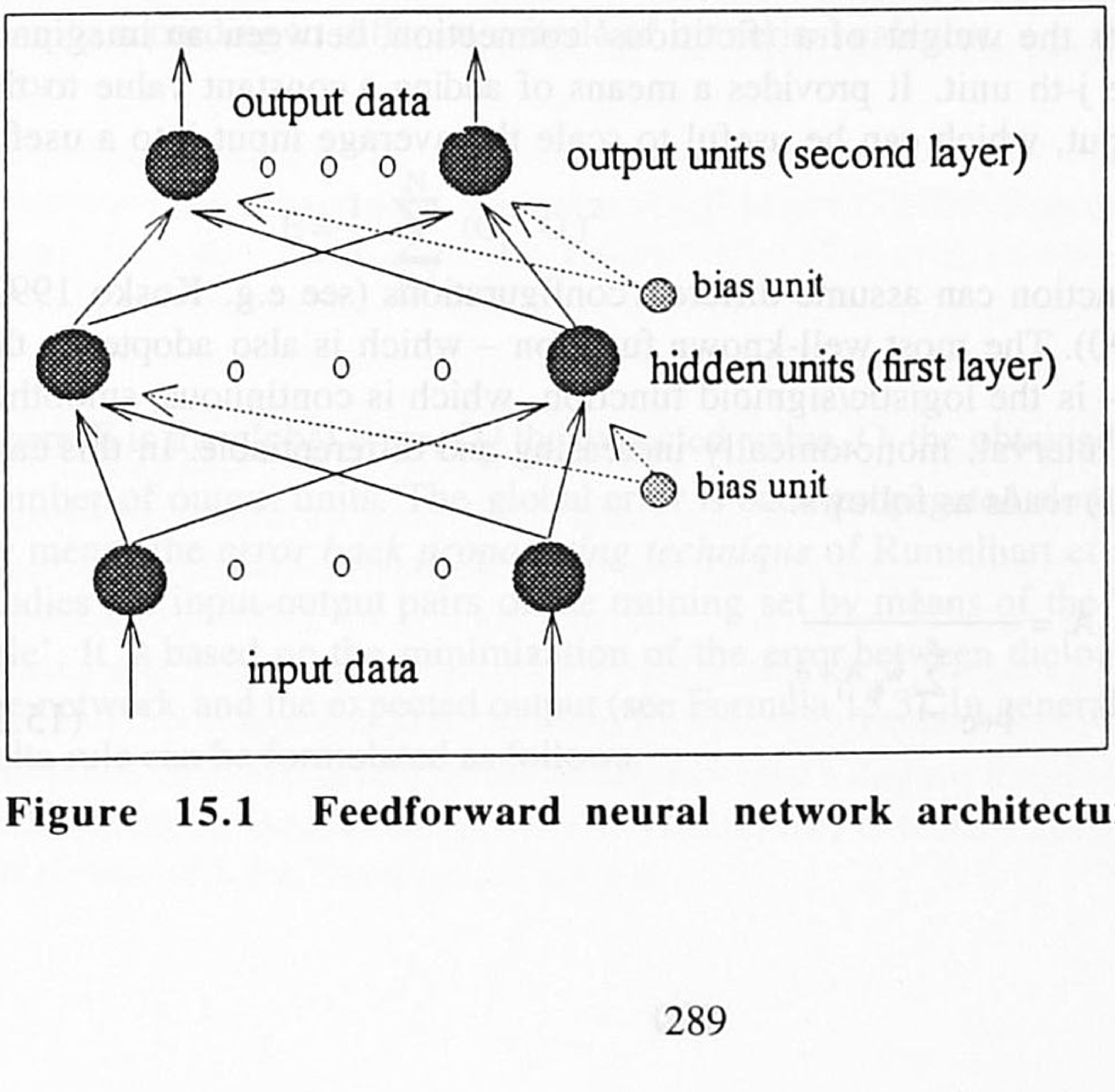


Figure 15.1 Feedforward neural network architecture

A two-layer feedforward network as subsequently adopted, is characterized by a layer of input units (1st layer), connected to a second layer of units (2nd (hidden) layer), which is then connected to the layer of output units (3rd layer). The properties of this type of network are the following: total connection among units; there are no self, lateral or back connections, only unidirectional connections from input to output layers; a certain convergence to a solution. For these reasons, this NN typology is regarded as being very simple from a computational viewpoint. Its ability depends on the learning algorithm (see Subsection 15.2.2.3).

Activity of the hidden/output unit Each hidden/output unit has its own activation state determined by the activity of the units linked to it (Figure 15.2 shows a generic hidden unit), based on the following relationship:

$$A_j = f[\sum_{i=1}^{n} (w_{ij} A_i) + \partial_j] \tag{15.1}$$

where:

- A_i, (i = 1,....,n) are the activations of the units linked to j-th unit (n is the number of units of the layer above the j-th unit);
- w_{ij} (i = 1,...,n; j = 1,...,m) are the weights of connections between the j-th unit and the units linked to it (m is the number of units of layer of the j-th unit);
- the expression 'f[]' denotes the 'activation' function (known as the transfer function) which allows the transmission of the signals only when the activation of a unit exceeds a particular threshold value;
- the symbol ∂ in f denote the 'bias', which determines the threshold value of the j-th unit activation. Such a value is different for each unit and can be considered as the weight of a 'fictitious' connection, between an imaginary unit and the j-th unit. It provides a means of adding a constant value to the summed input, which can be useful to scale the average input into a useful range.

The transfer function can assume different configurations (see e.g. Kosko 1992; Maren et al. 1990). The most well-known function – which is also adopted in the feedforward NN – is the logistic/sigmoid function, which is continuous, smoothly defined over the interval, monotonically increasing and differentiable. In this case formulation (15.1) reads as follows:

$$A_j = \frac{1}{1+e^{-\sum_{i=1}^{n} w_{ij} A_i + \partial_j}} \tag{15.2}$$

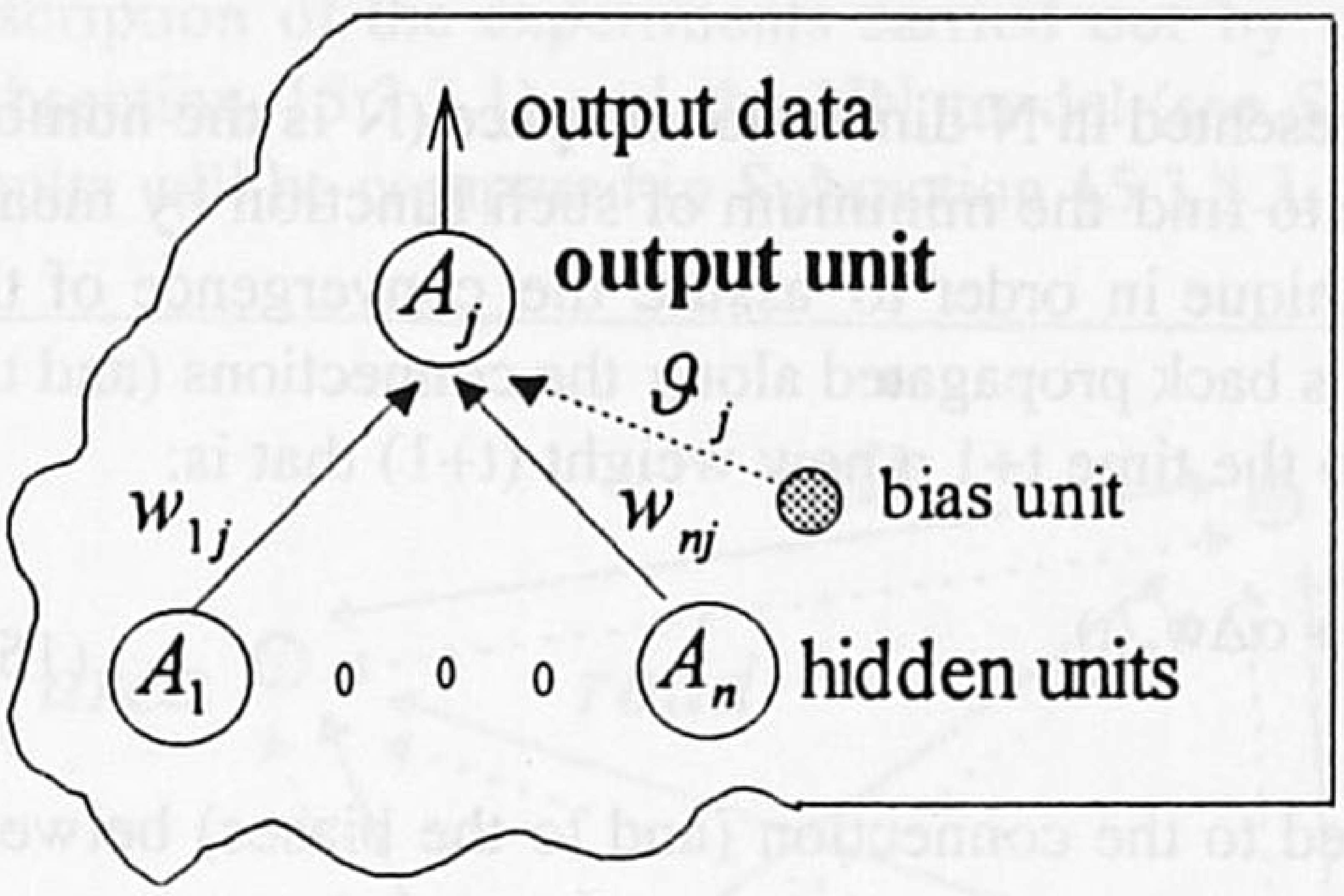


Figure 15.2 Generic hidden unit

Neural network learning The behaviour of an NN depends mainly on the learning algorithm, since this phase provides the basic information on the environment at hand, by preparing the neural network to recognize and classify the patterns.

In particular, during the learning phase, an NN system may adapt or organize itself each time it examines a new sample (or a set of samples), by changing the weights of each connection. Consequently, since NNs only learn a part of the possible samples (which can be infinite), the choice of the samples is very important (since the samples have to reflect the whole environmental pattern). Indeed, during the learning phase, internal 'representations'[5] of the samples take shape and define a 'quantization'[6] of the patterns by means of vectors of real numbers,[7] which can be modified in order to minimize the estimated error of the sample according to different criteria of numerical calculus, e.g. the mean squared error :

$$E = \frac{1}{2} \sum_{j=1}^{N} (O_j^{\,i} - O_j)^2 \tag{15.3}$$

where E is the global error, $O_j^{\,i}$ the expected value, O_j the obtained result and N the number of output units. The global error is back propagated along the connections by means the *error back propagating technique* of Rumelhart et al. (1986) which studies the input-output pairs of the training set by means of the 'generalized delta rule'. It is based on the minimization of the error between the output provided by the network and the expected output (see Formula 15.3). In general, the generalized delta rule can be formulated as follows:

$$\Delta w_{ij}(t) \propto -\frac{\delta E}{\delta w_{ij}(t)} \tag{15.4}$$

where the error function E is represented in N-dimensional space (N is the number of the connections). It is possible to find the minimum of such function by means of the 'descent of gradient' technique in order to assure the convergence of the algorithm. Then the global error is back propagated along the connections (and the biases) between units assigning to the time t+1 a new weight (t+1) that is:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \Delta w_{ij}(t) \tag{15.5}$$

where $w_{ij}(t)$ is the weight assigned to the connection (and to the biases) between the i-th and the j-th unit at the t time, while α is the learning coefficient fixed *a priori.*

Another formulation of the propagation of the error may be (see again Rumelhart et al. 1986):

$$w_{ij}(t+1) = w_{ij}(t) - \alpha \frac{\delta E}{\delta w_{ij}(t)} + \lambda \Delta w_{ij}(t) \tag{15.6}$$

where $\lambda (0 \leq \lambda < 1)$ is defined as the factor moment that can contribute to the variation of the weight with reference to the variation which had taken at the preceding time t (see further Fischer and Gopal 1994).

Finally, it is then clear that in this learning phase a great deal of attention should be paid to the choice of all the 'strategic' elements which can influence an 'optimal' network - for example, the training choice (random or sequential), the objective error choice, the initial conditions of the parameters, etc.

15.3 Experiments with neural networks in transport

15.3.1 Introduction

In order to explore the potential of the NN approach, some experiments will be carried out with reference to the modal split problem between two transport modes in 67 Italian areas. This study is a continuation of the analysis begun in the paper by Nijkamp et al. (1996) aimed at exploring the modal split between rail and road transport modes in Italy in relation to the introduction of a new technological innovation, the High-Speed Train (see Figure 15.3 for a scheme of the case study).

For our purpose, two methodologically different models, the logit model and the NN model, will be investigated and compared. Particularly, we will focus our attention on the performance of two models by varying the number of attributes.

In Subsection 15.3.2, the data-set and the statistical indicator used in order to compare the models will briefly be described. Subsection 15.3.3 contains a description of the experiments carried out by using both the logit model (see Subsection 15.3.3.1) and the NN model (see Subsection 15.3.3.2). Finally, the results will be compared in Subsection 15.3.3.3.

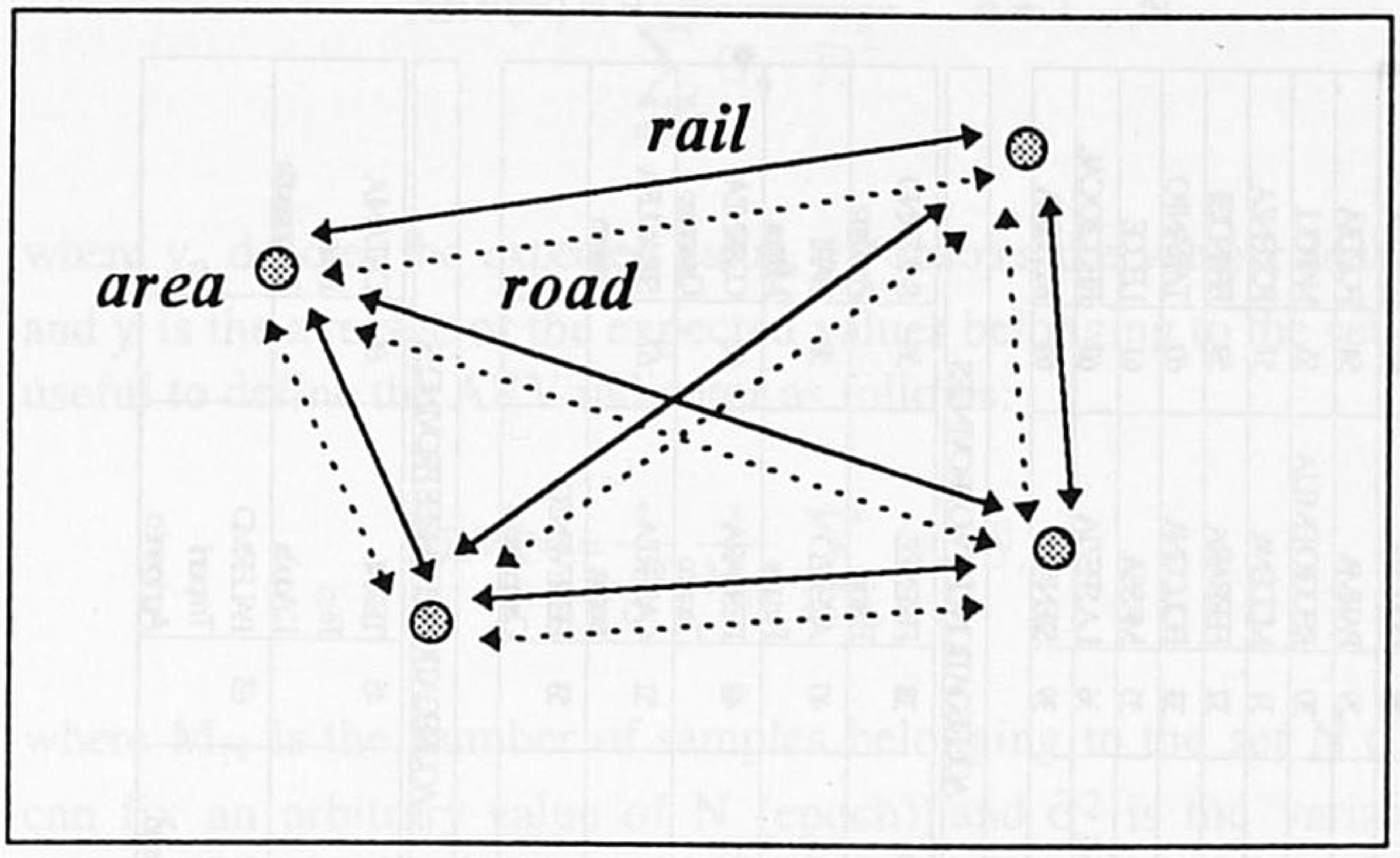


Figure 15.3 Scheme of the case study

15.3.2 The basic element for the experiments

The data set The database[8] refers to transport flows for the two different modes (rail and road), as well as the related attributes 'distance', 'time' and 'cost'. The Italian territory is divided into 67 areas corresponding to both single provinces and an aggregation of two or three provinces (the island of Sardinia is not considered); Figure 15.4 shows a map of such a subdivision.

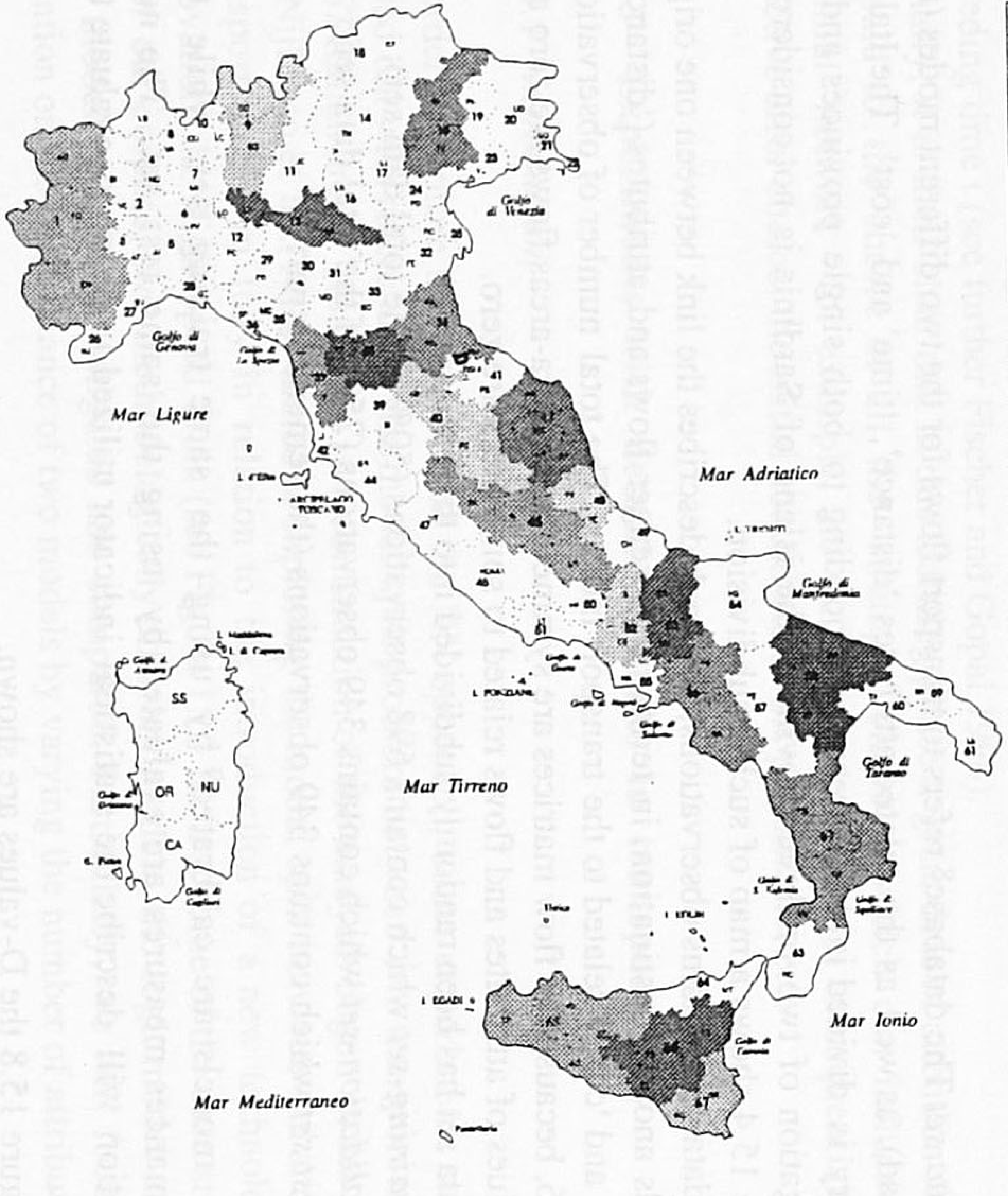
The data set contains observations, and each describes the link between one origin towards another destination, in terms of passenger flows and attributes ('distance', 'time' and 'cost') related to the transport modes. The total number of observations is 1396, because the flow matrices are symmetric, the intra-areas flows are zero and the values of attributes and flows related to some links are zero.

The data set has been randomly subdivided into three sub-sets:

- *training-set* which contains 698 observations (50% of the total data set);
- *validation-set* which contains 349 observations (25% of the total data set);
- *test set* which contains 349 observations (the remaining part).

Both models are calibrated[9] by using the same training set, while the performances measures are evaluated by using the same test set. The next subsection will describe the statistical indicator utilized in order to evaluate the results.

In Figure 15.8 the D-values are shown.



THE PROVINCES					
2	VERCELLI	21	GORIZIA	41	PESARO
3	ASTI	22	TRIESTE	42	LIVORNO
4	NOVARA	23	VENEZIA	44	GROSSETO
5	ALESSANDRIA	24	PADOVA	46	ROMA
6	PAVIA	25	ROVIGO	47	VITERBO
7	MILANO	26	IMPERIA	49	CHIETI
8	VARESE	27	SAVONA	50	FROSINONE
10	COMO	28	GENOVA	51	LATINA
11	BRESCIA	29	PARMA	54	FOGGIA
12	PIACENZA	30	REGGIO EMILIA	55	NAPOLI
14	TRENTO	31	MODENA	57	POTENZA
15	BOLZANO	32	FERRARA	59	BRINDISI
16	VERONA	33	BOLOGNA	60	TARANTO
17	VICENZA	35	MASSA	61	LECCE
19	PORDENONE	36	LA SPEZIA	63	REGGIO CAL.
20	UDINE	39	SIENA	64	MESSINA

AGGREGATION OF TWO PROVINCES					
9	BERGAMO Sondrio	38	FIRENZE Pistoia	56	SALERNO Avellino
13	CREMONA Mantova	40	AREZZO Perugia	58	BARI Matera
18	TREVISO Belluno	48	PESCARA Teramo	62	COSENZA Catanzaro
34	FORLI' Ravenna	52	CASERTA Isernia	67	SIRACUSA Ragusa
37	PISA Lucca	53	BENEVENTO Campobasso		

AGGREGATION OF THREE PROVINCES					
1	TORINO Aosta Cuneo	45	TERNI Rieti L'Aquila	66	CATANIA Enna Caltanissetta
43	ANCONA Macerata Ascoli Piceno	65	PALERMO Trapani Agrigento		

Figure 15.4 Italian land subdivision

The statistical indicator In Subsection 15.4.5, the results of two models, which are methodologically different, will be compared by using the Average Relative Variance (ARV (N)) indicator, whose mathematical expression is:

$$ARV(N) = \frac{\sum_{n \in N} (y_n - \hat{y}_n)^2}{\sum_{n \in N} (y_n - \bar{y})^2} \qquad n = 1,...,N \tag{15.7}$$

where y_n denotes the expected value, y_n denotes the value calculated from the model and y is the average of the expected values belonging to the set of data N. It is also useful to define the ARV indicator as follows:

$$ARV(N) = \frac{1}{\hat{\sigma}^2} \frac{1}{M_N} \sum_{n \in N} (y_n - \hat{y}_n)^2 \tag{15.8}$$

where M_N is the number of samples belonging to the set N (i.e., using NN, one can fix an arbitrary value of N (epoch)) and $\hat{\sigma}^2$ is the variance of the expected values of the entire data set (this is a proxy, since the variance of each epoch would have been considered).

This statistical indicator is widely used in the NNs literature (see e.g. Fischer and Gopal 1994; Nijkamp et al. 1996) primarily for two reasons:

- it ensures an independence - both quantitatively and qualitatively - of the error estimated from the selected samples (see expression (15.3.2));
- it is rather easy to understand the goodness of the indicator values, since it will tend to zero, the more exact the estimate is.[10]

Finally, it is important to underline that all future experiments will be evaluated by using the absolute values of the predicted flows according to the following expression:

$$T_{ij}^{train} = p_{ij}^{train} * T_{ij} \tag{15.9}$$

where:

T_{ij} :	Total flow related to link (ij);
T_{ij}^{train} :	Total rail mode flow related to link (ij);
p_{ij}^{train} :	Probability of choosing the rail mode with reference to link (ij).

15.3.3 Observations on the experiments

As mentioned in Subsection 15.3.1, this work continues the analysis started in a previous study by Nijkamp et al. (1996); we will particularly analyze the modal split problem between rail and road transport modes in Italy by varying the attributes. The initial case study (developed in the previous paper) concerned a comparative analysis between the NN model and the logit model and used three attributes, 'distance', 'time' and 'cost' related to each transport mode (and referring to each link (ij)). Different possible NN architectures were investigated and finally, the configuration shown in Figure 15.5 resulted as the best one (by comparing the values of ARV indicator):

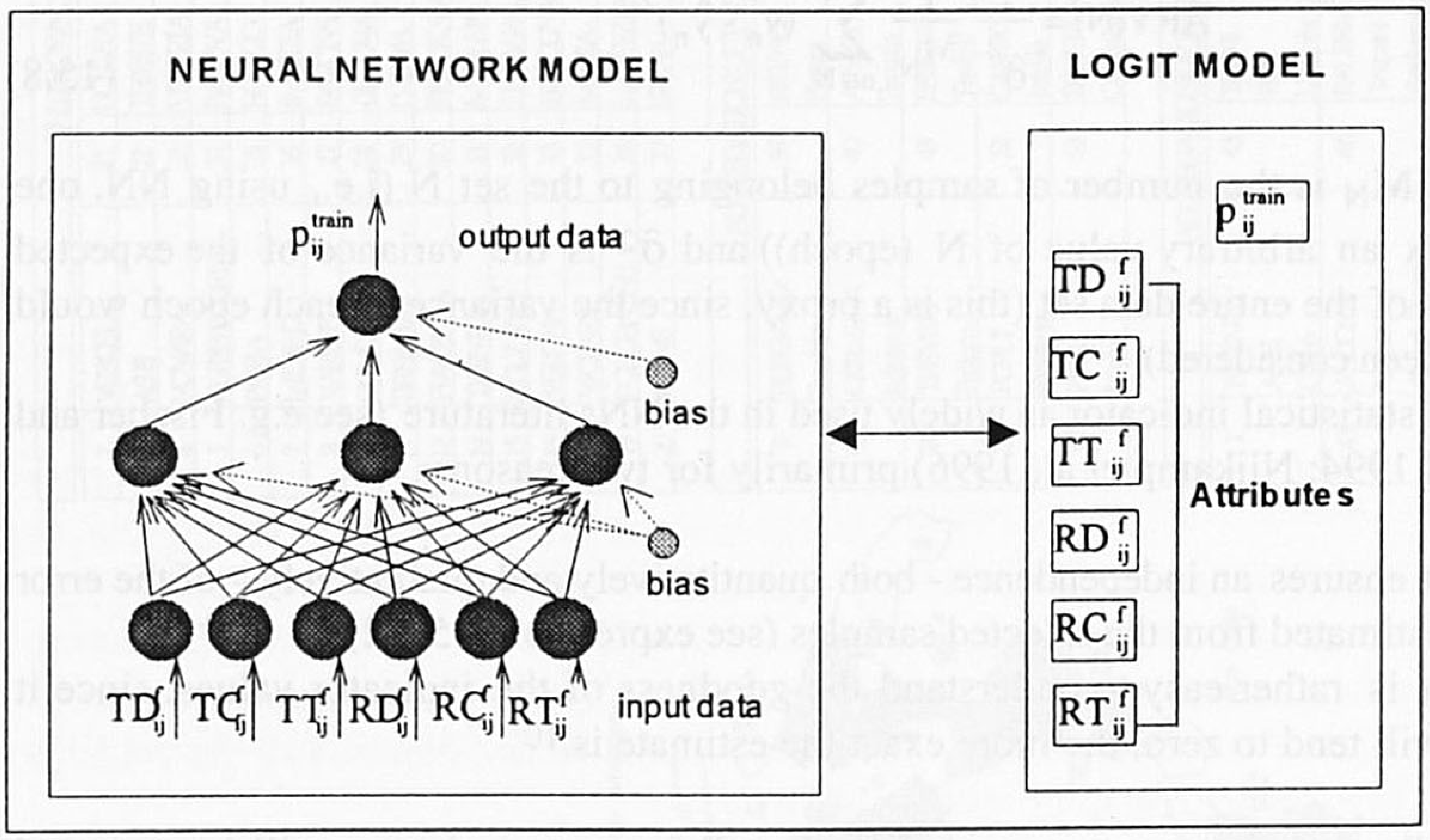


Figure 15.5 Configuration of the models adopted for the problem under investigation

The variables in Figure 15.5 have been transformed (in values between [0-1]) by means of the following functions:

$$V_j^f = \exp(-0.005 * V_j) \qquad (15.10)$$

$$V_j^f = \exp(-0.0002 * V_j) \qquad (15.11)$$

where in the previous formulas V_j (j = 1,..,9) denote the general variables utilized in the experiments and where (i) formula (15.10) (identified by the index $_f$) has been adopted for transforming the values of the attributes and (ii) formula (15.11) (identified by the index $_t$) is used for transforming the flow values. Furthermore, the variables, related to the link (ij), are defined as follows:

TD_{ij}^f : transformed train distance (according to 15.10));
TC_{ij}^f : transformed train cost (according to (15.10));
TT_{ij}^f : transformed train time (according to (15.10));
RD_{ij}^f : transformed road distance (according to (15.10));
RC_{ij}^f : transformed road cost (according to (15.10));
RT_{ij}^f : transformed road time (according to (15.10));
T_{ij}^t : transformed total passenger flow (according to (15.11));
$T_{ij}^{train(t)}$: transformed rail mode flow (according to (15.11));
p_{ij}^{train} : rail mode probability (see (15.9)).

Since we want to investigate the performance of the models by varying the attributes, in the first experiment (Case A1), the 'distance' attributes for both transport modes have been eliminated (i.e. 4 attributes in total have been used); in the second and third experiments (Case A2 and A3), the 'cost' and the 'time' attributes (again for both transport modes), have been eliminated (i.e. 2 attributes in total have been used), respectively. Figure 15.6 illustrates a scheme of the three experiments which have been conducted.

The logit model The application of logit model in modal split problems is rather common in the transport economic literature (see e.g. Ben-Akiva and Lerman, 1985). In general, if we denote the number of attributes by χ_k and the random component of the j-th possible choice by δ_j, then it is possible to define a utility function, related to the j–th transport mode, as follows:

$$u_j = \sum_k^K \beta_k \chi_k + \delta_j \qquad k = 1,...,K \tag{15.12}$$

where β_k are parameters to determine.

The logit model assesses the probability p_j of choosing the j-th transport mode by maximizing (under certain assumptions) the above utility u_j; its general form is:

$$p_j = \frac{\exp^{u_j}}{\sum_{i=1}^{n} \exp^{u_i}} \qquad i,j = 1,...,n \tag{15.13}$$

In the present experiments, j = 1, 2 (n = 2), i.e. only two choices are possible: road mode or rail mode. There are six related attributes χ_k (k = 1,...,6); they are the following:

$$TD^f_{ij} ; TC^f_{ij} ; TT^f_{ij} ; RC^f_{ij} ; RD^f_{ij} ; RT^f_{ij}$$

which have already been defined above.

In order to calibrate the model, the procedure Blogit in the STATA software has been used. Such a procedure allows the calculation of logit estimates respecting aggregate data, adopting a maximum likelihood technique and applying the Newton–Raphson algorithm as the convergence procedure.

Learning phase of the feedforward neural network model The calibration phase of the logit model corresponds to the learning phase of the feedforward NN model. This phase allows NNs to 'learn' an 'inner' rule of the variables at hand, without fixing *a priori* relationships/models between the variables. The only base is constituted by representative 'examples' of the problem at hand.

In this application, the training set (defined in Subsection 15.3.2.1) contains 698 observations whose variables are shown in Figure 15.6. In this figure it is also easy to observe that the signals of the input-layer are propagated forward according to formulations (15.1) and (15.2).

The NN architecture which has been used is a two-layer feedforward, totally connected NN with the back-propagation algorithm as learning procedure (see Subsection 15.2.1.3). According to this procedure, the values of the output unit are compared with the real values in order to adjust the weights of the network by thinking of equations (15.3), (15.4) and (15.5).

The parameters defining the neural architectures were set out (after some experiments) in the previous work by Nijkamp et al. (1996). The results are as follows:

- learning rate $\alpha = 0.5$
- momentum factor $\lambda = 0.2$
- epoch size: 1
- initial weight values: randomly between [-0.1-0.1]

No useful mathematical rule exists to decide on the number of hidden units. It is usually determined empirically. In the current experiments, it has been defined without changing the degrees of freedom[11] of the NN model (according to the reasoning on the sample size in the paper by Nijkamp et al., 1996). Therefore, in the Case A1, the hidden units are 4 and in the Case A2 and A3 are 6.

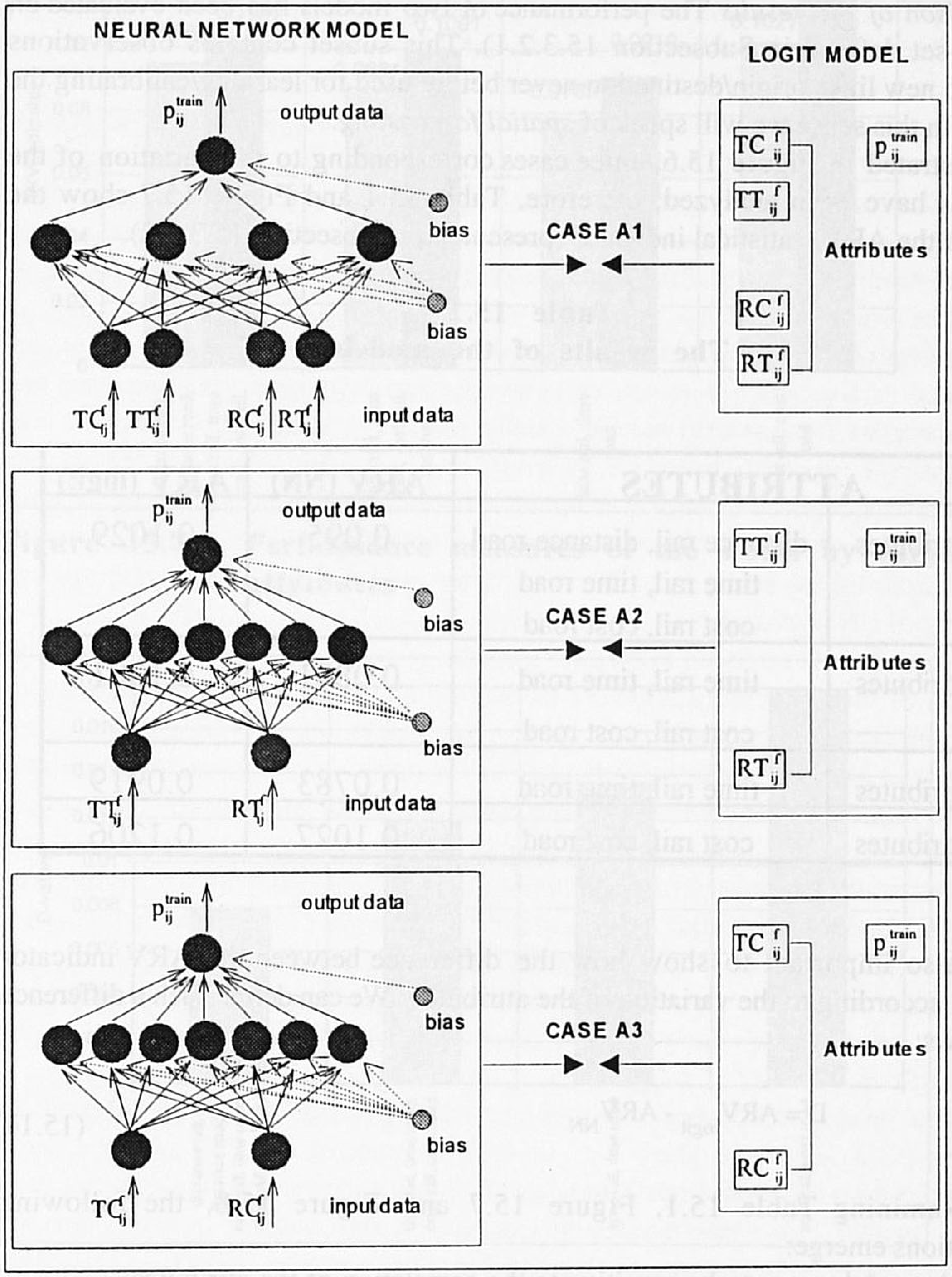


Figure 15.6 Scheme of the experiments

Finally, it should be noted that by using a feedforward NN it is necessary to cope with the overfitting problem. Consequently, in all three cases A1, A2 and A3, the cross-validating technique (by using the cross-validation subset defined in Subsection 15.3.2.1) has been used in order to avoid such a problem (for details on the overfitting problem and the cross-validating technique, see e.g. Fischer and Gopal 1994; Reggiani and Tritapepe 1997).

Comparison of the results The performance of two models has been evaluated on the test-set defined in Subsection 15.3.2.1). This subset contains observations related to new links origin/destination never before used for learning/calibrating the models. In this sense we will speak of *spatial forecasting*.

As illustrated in Figure 15.6, three cases corresponding to the variation of the attributes have been analyzed; therefore, Table 15.1 and Figure 15.7 show the values of the ARV statistical indicator (presented in Subsection 15.3.2.2).

Table 15.1
The results of the models

ATTRIBUTES		ARV (NN)	ARV (logit)
6 Attributes	distance rail, distance road time rail, time road cost rail, cost road	0.095	0.1029
4 Attributes	time rail, time road cost rail, cost road	0.0881	0.1001
2 Attributes	time rail, time road	0.0783	0.0919
2 Attributes	cost rail, cost road	0.1027	0.1206

It is also important to show how the difference between the ARV indicator changes according to the variation of the attributes. We can define such a difference as follows:

$$D = ARV_{logit} - ARV_{NN} \tag{15.14}$$

By examining Table 15.1, Figure 15.7 and Figure 15.8, the following observations emerge:

- both models seem to be sensitive to the correlation of the attributes;
- by reducing the number of attributes, the NN seems to perform better than the logit model.

With regard to the first consideration - although expected for the logit model – one might expect different results for the NN model; this is due mainly to the feature of NNs, which should be able to learn and then forecast – even on the basis of incomplete, noisy and fuzzy information.

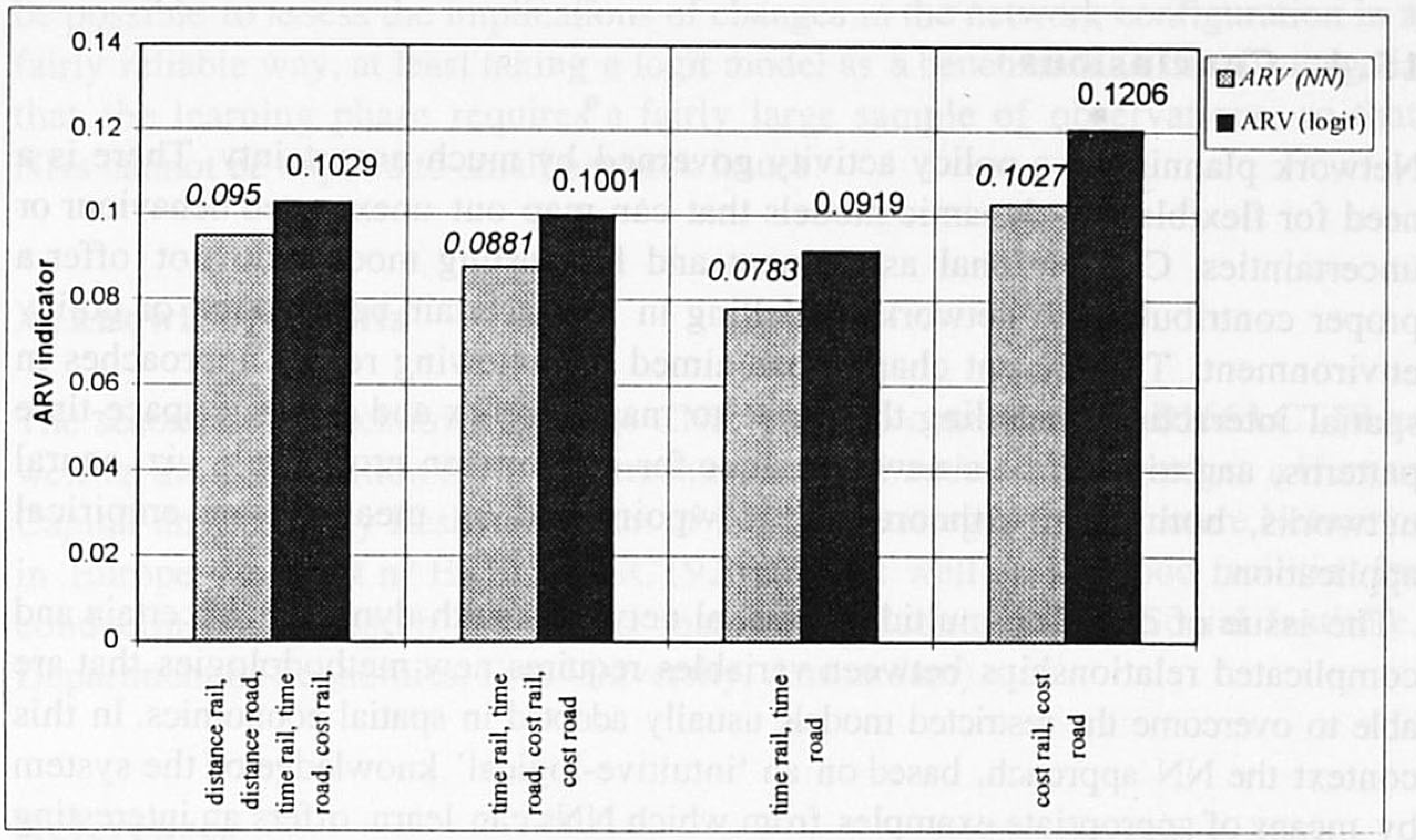


Figure 15.7 Performance measures of the model by varying the attributes

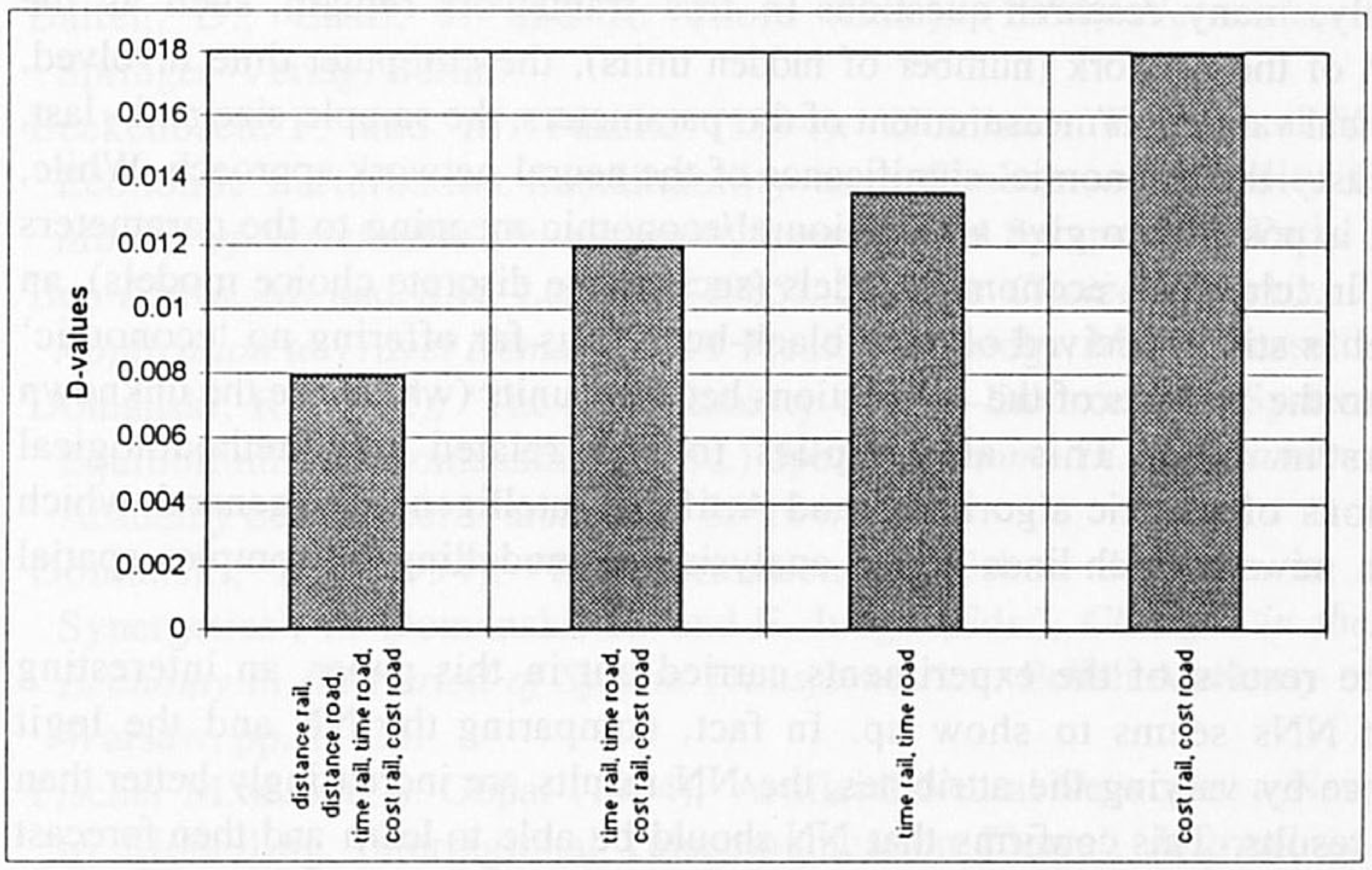


Figure 15.8 Differences of the performance measures by varying the attributes

Since this characteristic seems to be proven from our second consideration, we may conclude that the correlation of the attributes does not have similar effects – other than noisy and fuzzy information – on the NN performance.

15.4 Conclusions

Network planning is a policy activity governed by much uncertainty. There is a need for flexible and dynamic models that can map out unexpected behaviour or uncertainties. Conventional assessment and forecasting models do not offer a proper contribution to network modelling in an uncertain behavioural or policy environment. The present chapter has aimed at reviewing recent approaches in spatial interaction modelling that serve to map complex and dynamic space-time patterns, and to describe a new technique for information processing, viz. neural networks, both from a theoretical viewpoint and by means of an empirical application.

The issue of complex multidimensional networks with dynamic, uncertain and complicated relationships between variables requires new methodologies that are able to overcome the restricted models usually adopted in spatial economics. In this context the NN approach, based on an 'intuitive-logical' knowledge of the system by means of appropriate examples from which NNs can learn, offers an interesting potential, especially for prediction purposes. An empirical test of the NN approach may be undertaken using a proper indicator which may verify the spatial forecasting of NN approaches.

Obviously, many research questions in this framework remain, such as the dimension of the network (number of hidden units), the computer time involved, the empirical validation/measurement of the parameters, the sample size and - last, but not least - the 'economic' significance of the neural network approach. While, in fact, it is possible to give a behavioural/economic meaning to the parameters estimated in 'classical' economic models (such as the discrete choice models), an NN model is still conceived of as a 'black-box', thus far offering no 'economic' meaning to the weights of the connections between units (which are the unknown parameters in NN). This also applies to the related new methodological contributions of genetic algorithms and Artificial Intelligence in general, which may open new research lines in the analysis and modelling of complex spatial networks.

From the results of the experiments carried out in this paper, an interesting feature of NNs seems to show up. In fact, comparing the NN and the logit performance by varying the attributes, the NN results are increasingly better than the logit results. This confirms that NN should be able to learn and then forecast even on the basis of incomplete, noisy and fuzzy information (which, in this case, corresponds to a minor number of attributes).

Finally, the lack of intrinsic behavioural contents may be a major advantage of NN applications, as this eliminates a discussion of subjective modelling choices.

In any case, there seems to be a great potential for the application of NN methods in complex network analysis.

Transportation planning may be greatly supplied by the use of NN techniques. With a minimum of theoretical assumptions on network behaviour, it appears to be possible to assess the implications of changes in the network configuration in a fairly reliable way, at least taking a logit model as a benchmark. A disadvantage is that the learning phase requires a fairly large sample of observations, so that NNs cannot be applied to small network issues.

Acknowledgements

The second author acknowledges the CNR bilateral contract n° 96.01663.CT10 as well as the contribution MPI 40 percent. The third author acknowledges a Human Capital and Mobility Research Grant of the EC (Strategic Infrastructure Networks in Europe, Contract n° ERBCHBGCT920162), as well as the good facilities for conducting his research at the host institution (Economic and Social Institute, Department of Economics, Free University, Amsterdam).

References

Allen, P.M. and M. Sanglier (1981), 'A Dynamic Model of Central Place Systems', *Geographical Analysis*, Vol. 13, pp. 149-164.

Batten, D., Casti, J. and R. Thord (Eds.), (1994) *Networks in Action*, Springer–Verlag, Berlin.

Beckenbach, F. and M. Pasche (1994), 'Nonlinear Ecological Models and Economic Perturbation Sustainability as a Concept of Stability Corridors', *Models of Sustainable Development*, International Symposium: Paris.

Ben-Akiva, M. and S.R. Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press: Cambridge, Massachussetts.

Domanski, R. (1992), 'The Changeability of Man-Environment Systems far from Equilibrium', in Domanski, R. (Ed.) *Space Economy in the Transition*,, Polish Academy of Science: Warsaw, pp. 85-104.

Domanski, R. (1994), 'Transformation of Urban Systems in Terms of Synergetics', in Domanski, R. and E. Judge (Eds.), *Changes in the Regional Economy in the Period of System Transformation*, Polish Academy of Sciences: Warsaw, pp. 11-23.

Fischer M.M. and S. Gopal (1994), 'Artificial Neural Networks: a New Approach to Modelling Interregional Telecomunication Flows', *Journal of Regional Science*, Vol. 34, N. 4, pp. 503-527.

Haken, H. (1983a), *Synergetics*, Springer-Verlag: Berlin.

Haken, H. (1983b), *Advanced Synergetics*, Springer-Verlag: Berlin.

Hebb, D.O. (1949), *The Organization of Behaviour*, Wiley: New York.

Hertz, J., Krogh, A. and R.G. Palmer (1991), *Introduction to the Theory of Neural Computation*, Addison Wesley: Redwood City, Cal.

Kosko, B. (1992), *Neural Networks and Fuzzy Systems*, Prentice Hall Inc.: Englewood Cliffs, N.J.

Lambooy, J.G. (1993), 'The European City: From Carrefour to Organizational Nexus', *Tijdschrift voor Economische en Sociale Geografie*, Vol. 84, No. 4, pp. 258-268.

Malliaris, M. and L. Salchenberger (1992), 'A Neural Network Model for Estimating Option Prices', (mimeo).

Maren, A., Harston. C. and R. Pap (1990), *Handbook of Neural Computing Application*, Academic Press: San Diego.

Nicolis, G. and I. Prigogine (1977), *Self-Organisation in Non-Equilibrium Systems*, Wiley: New York.

Nijkamp, P. (Ed.) (1991), *Sustainability of Urban Systems*, Avebury: Aldershot, UK.

Nijkamp, P. and A. Reggiani (1992), *Interaction, Chaos and Evolution in Space*, Springer-Verlag: Berlin.

Nijkamp, P. and A. Reggiani (Eds.) (1993), *Non-Linear Evolution of Spatial Economic Systems*, Springer-Verlag: Berlin.

Nijkamp, P. and A. Reggiani (1995), 'Non-Linear Evolution of Dynamic Spatial Systems: The Relevance of Chaos and Ecologically-based Models', *Regional Science and Urban Economics*, Vol. 25, pp. 183-210.

Nijkamp P., Reggiani, A. and T. Tritapepe (1996), 'Modelling Inter-Urban Flows in Italy: a Comparison between Neural Network Approach and Logit Analysis', *Transportation Research C*, Vol. 4, No. 6, pp. 323–38.

Reggiani, A. and P. Nijkamp (1995), 'Competition and Complexity in Spatially Connected Networks', *System Dynamics Review*, Vol. 11, No. 1, pp. 51-66

Reggiani, A. and P. Nijkamp (1996), 'Towards a Science of Complexity in Spatial Economic Systems', in van den Bergh, J.C.J.M., Nijkamp, P. and P. Rietveld (Eds.), *Recent Advances in Spatial Equilibrium Modelling,* Springer Verlag: Berlin, pp. 359-378.

Reggiani A. and T. Tritapepe (1997), 'Neural Network and Logit Models to Commuters' Mobility in the Metropolitan Area of Milan', in V. Himanen, Nijkamp, P. and A. Reggiani (Eds.), *Neural Networks in Transport Systems*, Avebury: Aldershot (forthcoming).

Reggiani A., Romanelli, R, Tritapepe,T. and P. Nijkamp (1997), 'Neural Networks: an Overview and Applications in the Space-Economy', in Himanen, V., Nijkamp, P. and A. Reggiani (Eds.), *Neural Networks in Transport Systems,* Avebury: Aldershot, 1997 (forthcoming).

Weidlich, W. and G. Haag (1983), *Concepts and Methods of a Quantitative Sociology*, Springer-Verlag: Berlin.

Notes

1. Dendrites are usually defined as 'tree-like networks of nerve fibre' (see Hertz et al. 1991).

2. An axon is a 'single long fibre' (see Hertz et al. 1991).

3. Here weight is a real number assigned to a connection between two units.

4. 'Generalizing' is the capacity of a system to create new patterns in accordance with previously studied examples.

5. Prototype of sampled patterns.

6. 'Quantization' is the set of weights (real numbers) created by the prototypes.

7. The 'Adaptive Vector Quantization' (AVQ); see Kosko 1992.

8. The data set has been kindly provided by the Italian State Railways ('Ferrovie dello Stato') and it refers to Census data (1987).

9. Usually the verb 'calibrate' refers to logit model and the verb 'learn' refers to the NN model. We will use the two verbs without distinction referring to both models.

10. Note that in this case, because of the non-linear regression nature of the NN experiments, the well-known R^2 statistical indicator no longer has the value 1 as maximum value (the more exact the estimate is), but it may assume values greater than 1.

11. If we, in an NN model, denote the number of input units by i, the number of hidden units by h and the number of output units by o, then the degree of freedom of the model is: $(i*h*o + h + o)$, where the number of connections between the hidden and output units with the bias unit (see Figure 15.1) is taken into account by adding h and o.