



THE PROBLEM OF MODELLING A TRUNK AIR ROUTE NETWORK

Tatiana Tseytlina¹, Victor Balashov², Andrey Smirnov³

*N. E. Zhukovsky Central Aerohydrodynamic Institute (TsAGI), Zhukovsky, Russian Federation
E-mails: ¹ttseytlina@gmail.com; ²balashov@tsagi.ru; ³smirnov@tsagi.ru (corresponding author)*

Received 10 January 2013; accepted 22 January 2013



Tatiana TSEYTLINA, MSc.

Education: Moscow Institute of Physics and Technology (MIPT), 2008.

Research interests: air route networks, passenger air transport, neural network modelling.

Publications: author of one scientific article, eight conference papers, co-author of three scientific articles.

Present position: junior researcher, N.E. Zhukovsky Central Aerohydrodynamic Institute (TsAGI).



Victor BALASHOV, Assoc. Prof, PhD.

Education: Moscow Aviation Institute (MAI), 1963.

Research interests: research and forecasting of passenger air transport.

Publications: author and co-author of more than 200 scientific publications.

Present position: leading researcher, TsAGI.



Andrey SMIRNOV, Assoc. Prof, PhD.

Education: Moscow Institute of Physics and Technology (MIPT), 1984.

Research interests: strategy of aviation industry in Russia, research and forecasting the development of the air transport system.

Publications: author and co-author of more than 150 scientific publications.

Present position: head of department at TsAGI.

Abstract. In this work we developed a fuzzy neural network-based model of the conditions for the existence of air routes, i.e. the rules underlying the emergence, existence and elimination of air routes (direct links between cities). The model belongs to the class of information models: the existence or non-existence of an air route is considered dependent on a complex of parameters. These parameters characterise the transport link, as well as the generational and target capabilities of the connected cities. The model was constructed using genetic algorithm techniques and self-organising Kohonen maps (implemented by software features of the STATISTICA package), as well as software tools of the Fuzzy Logic Toolbox and the Neural Network Toolbox of the MatLab development environment. The model is used to forecast the development of the topology of the network. The forecast is a necessary component of long-term forecasts of demand in the aircraft market.

Keywords: air route network, conditions for the existence of air routes, direct link, fuzzy modelling, network development forecast.

1. Introduction

An adequate toolset is needed in order to assess the contribution of new technologies to achieving the target indicators of advanced airplanes within the Russian air transport system. The set of tools in question should have two levels. The first level incorporates basic models that will enable us to formulate scenarios for the development of passenger air transport. These scenarios should characterise the forecasted traffic intensity for various classes of passenger aircraft within the route network in order to meet distributed public demand for passenger air transport (Vasermanis *et al.* 2004). The second level of the toolset being developed must include models and programs for estimating the contribution of new technologies in the reduction of contaminants and greenhouse gas emissions, as well as the improvement of the fuel efficiency of new trunk and regional passenger aircraft fleets.

The object of the study is the Russian trunk air route network. The goal of this research is to create a software complex for modelling and predicting the development of the Russian trunk air route network (Blinova 2007). The significance of this research is in that it allows solving applied problems connected with long-term prediction of the Russian market of long-distance and regional aircraft based on changes in topology of the air route network.

In order to solve the problem of predicting the development of air routes, we developed a software complex (Fig. 1) performing the following functions: preparing source data, creating a model for the development of air route networks, and predicting the development of air route networks.

2. Model of conditions for existence of air routes

We developed a fuzzy neural network-based model (Borisov *et al.* 2007; Haykin 1999; Kosko 1994; Piegat 2001) of the conditions for the existence of air routes, i.e. direct links between Russian cities. This model is used to forecast the development of the topology of the network. The model belongs to the class of information models. The model was based on the following hypothesis: there are universal rules governing the appearance, existence and dissolution of an air route, and these rules can be determined using a limited number of measurable parameters. We developed this hypothesis based on the analysis of changes in the topology of Russia’s trunk air route network over the past 15 years.

The existence or non-existence of an air route is considered dependent on a complex of parameters. These parameters characterise the transport link, as well as the generational and target capabilities of the connected cities (that is, how likely a linked city becomes a starting point and a destination point, respectively). In this work, we consider the backbone network of 144 trunk airports located in 139 Russian cities. Eight input parameters corresponding to the year 2006 are considered:

- Characteristics of air connections:
 1. Length of air route.
 2. Class of air route (according to the aircraft received).
 3. Characteristic of air route direction.
 4. Existence or non-existence of a rail link between the cities.

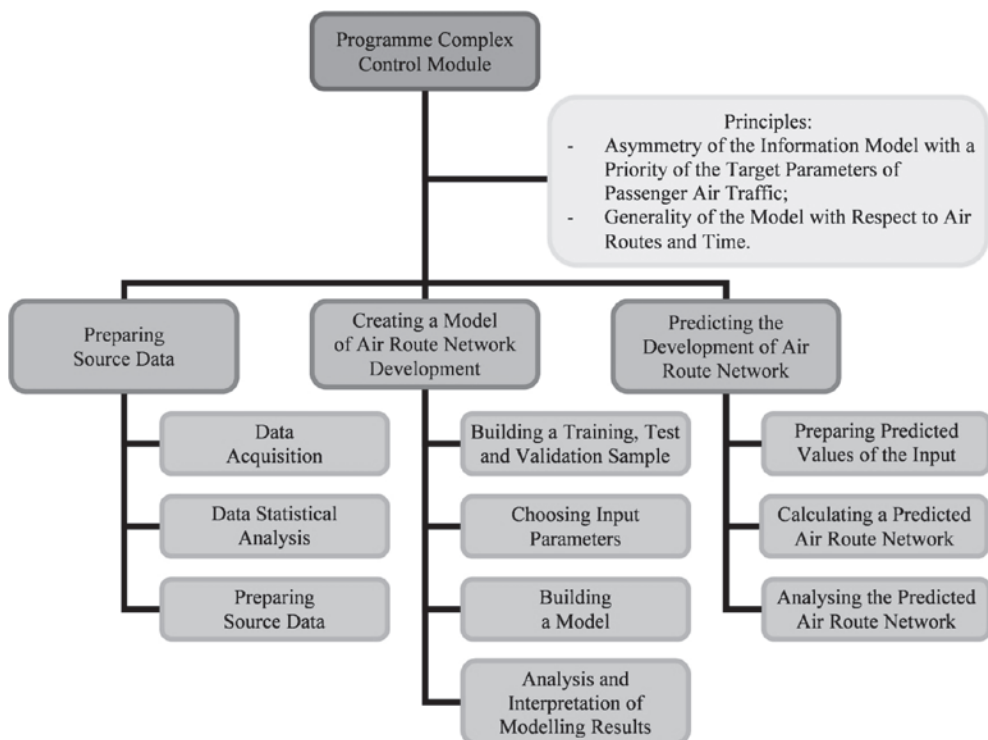


Fig. 1. Flowchart of the software package

- Characteristics of generational capabilities:
 1. Population size of the constituent entity.
 2. Gross regional product (GRP) of the constituent entity.
- Characteristics of target capabilities:
 1. Amount of available accommodation.
 2. City’s status.

They are used to build representative sets of *training*, *test* and *validation* data. For this purpose, we used the method of self-organising Kohonen maps (Haykin 1999; Kohonen 1990, 2001) implemented in the STATISTICA software package (Neural networks ... 2008). Use of this method enabled a significant reduction in the amount of training data while retaining the information value of the original data essentially unchanged.

2.1. Constructing the training set

In this work, the traditional approach to constructing the training set, which takes into account both generational and destination capabilities of both cities of a given air route, is referred to as *symmetrical*. This approach implies that the parameters of both cities are equally important for the existence of the air route and that they must be considered when forming the elements of symmetrical training sets. At the same time, the existence of the air route can be assumed to be determined either by the generational capabilities or the destination capabilities of one of the cities. This approach forms the basis for the elements of the so-called *unsymmetrical* training sets.

In this work, eight versions of the training set were considered while building the model of the conditions of the existence of air routes (Tab. 1). These versions of training sets differ in two ways: in the principle underlying the formation of training sets and in the principle underlying the formation of training elements per se. Training sets have a two-letter designation. The first letter (A, B, C, D) corresponds to the type of training set. The second letter (S, P, C, D) corresponds to the type of training set element.

Version **A** implies developing a model based on the traditional approach to generating a training set. Generational and destination parameters of cities (airports) in training set elements are formed as sums (**S**) or products (**P**) of the corresponding parameters of the two cities of one air route. Thus, versions **AS** and **AP** of the training set are built. They fall in the category of symmetrical versions. Other versions of training sets considered in this work are unsymmetrical. The dissymmetry of these versions lies in the fact that a certain criterion is used to pick a primary city from the two cities of an air route.

In version **B**, the set of training elements is created based on pre-ranking the cities in terms of the air routes that pass through them. At each ranking step, a city with the maximum number of air routes is determined, whereupon these air routes are excluded from subsequent

consideration. Thus, each air route is thought of as belonging to one city only, the primary city in a given city pair. Training set versions **BS** and **BP** are formed similarly to versions **AS** and **AP**. The values of generational and destination parameters of the elements of these sets are calculated as sums (**S**) or products (**P**) of the corresponding parameters of the two cities. In versions **BS** and **BP**, the training set dissymmetry lies in the determination of the primary city. Version **BC** of the training set uses generational parameters of the primary city as generational parameters of the training set elements and uses destination parameters of the secondary city as the destination parameters of the training set elements. Version **BD** of the training set uses destination parameters of the primary city as destination parameters of the training set elements and uses generational parameters of the secondary city as the generational parameters of the training set elements.

Table 1. Eight versions of the training set

| Training set | Element parameters of the training set | |
|---|---|---|
| Symmetrical training set (A) | | |
| AS | Sum (S) of the values of the corresponding measurable parameters of both cities | |
| AP | Product (P) of the values of the corresponding measurable parameters of both cities | |
| Ranking the cities in terms of the number of airlines (B) | | |
| BS | Sum of the values of the corresponding measurable parameters of both cities | |
| BP | Product of the values of the corresponding measurable parameters of both cities | |
| Unsymmetrical training set | | |
| Training set | Generational (C) parameters of training set elements | Destination (D) parameters of training set elements |
| BC | Generational parameters of the primary city | Destination parameters of the secondary city |
| BD | Generational parameters of the secondary city | Destination parameters of the primary city |
| CC | Generational parameters of the primary generating city | Destination parameters of the secondary generating city |
| DD | Generational parameters of the secondary destination city | Destination parameters of the primary destination city |

In version **C**, the training set is formed after comparing the generational capabilities of the two cities of the air route. The one with greater generational parameter values is picked as the primary city. The procedure implies comparing the population of the two cities and subsequently (if these parameters are equal) the GRP values for these cities. Version **C** of the training set correlates best with the

similarly named version of the formation of training set elements, where generational parameters of the primary city are used as generational parameters of the training set elements, and destination parameters of the secondary city (the one with lower generational capabilities) are used as the destination parameters of the training set elements. Thus, the CC training set version is built.

In version **D**, the training set is constructed after comparing the destination capabilities of the two cities of the air route. The one with greater destination parameter values is picked as the primary city. The procedure implies comparing the number of accommodation units for both cities and (if these parameters are equal) the administrative status of these cities. Version **D** of the training set, like version **C**, agrees best with the similarly-named version of training set elements, where destination parameters of the primary city are chosen as destination parameters of the elements, and generational parameters of the secondary city (the one with less significant destination capabilities) are picked as the generational parameters of the elements. Thus, version **DD** of the training set is built.

Analysing the modelling results, we identify four air routes classes. These classes are introduced through a two-digit index, where the first digit corresponds to the real air route, and the second one, to the modelled one:

- 11 – air route exists and is correctly modelled;
- 00 – air route does not exist and is correctly modelled;
- 10 – air route exists and is incorrectly modelled;
- 01 – air route does not exist and is incorrectly modelled.

The first two classes correspond to modelling results that match the actual network. The third and fourth classes manifest the discrepancies between the network model and the real network. Figure 2 shows the numerical comparison results of eight model versions of the conditions for the existence of air routes. It displays data on the number of city pairs with correctly modelled existing (class 11) and non-existing (class 00) air routes as of 2006, expressed in percentage points. The most efficient method of model evaluation is pairwise comparison based on the formation principle of training set elements.

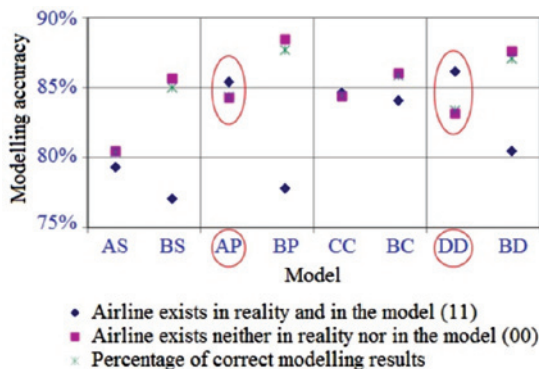


Fig. 2. Comparison results of eight model versions of the conditions for the existence of air routes

Models BS, BP, BC and BD simulate the existing air routes worse than other models. This is primarily because the principle underlying the training set construction does not correspond to the principle used to generate training set elements, which means that the sets in these models are less informative. At the same time, these models are better at modelling non-existent air routes. The latter is because the training sets did not include many city pairs that were not connected by air routes. The best modelling results were achieved by using models BC and CC. This is because despite the difference in training set construction principles, the sets of elements in them were largely similar. The comparison of model versions AS and AP shows that for a symmetrical model, the product of similar parameters is preferred to their sum. Even though the DD model achieves somewhat lower success rates at modelling the existing air routes than the AP model, the DD model is believed to be more promising. This model simulates 86% of existing air routes correctly, which is the maximal number for all models. Besides, the correctly modelled air routes total over 95% of annual passenger traffic.

The model of the conditions for the existence of air routes that treats the purpose of the journey as the key factor determining the existence of a direct air route between two Russian cities (airports) is believed to be the most promising. In this case, the existence of the air route is primarily determined by passenger traffic bound for the city with the greatest destination parameters.

2.2. Building output parameter values

The term *existing air route* means here a direct air link between a pair of cities. A *non-existent* air route is understood as a lack of a direct air link between a pair of cities. At the initial stages of model development, the notion that an *air route exists* was unambiguously interpreted based on the aircraft flight timetable for 2006. The output parameter could assume two values: 1 (the air route exists) or 0 (the air route does not exist). Value 1 was assigned if there was at least one aircraft flight along the route. Value 0 was assigned if there were no flights during the year in question. We found that this interpretation of the notion of *air route existence* prevents the model from demonstrating high value in terms of generalisation, so we had to expand the output parameter set. For this purpose, we used data of flight timetables for three years (2005 to 2007). For each air route, the total number of aircraft flights in each year was determined. These data were normalised to the maximum value for the air route over three years. Normalised data on the total number of flights for each air route in 2005, 2006 and 2007 can be represented on the sides of a three-dimensional cube (Fig. 3). This allows changes in aircraft flight intensity to be visualised for each air route over three years.

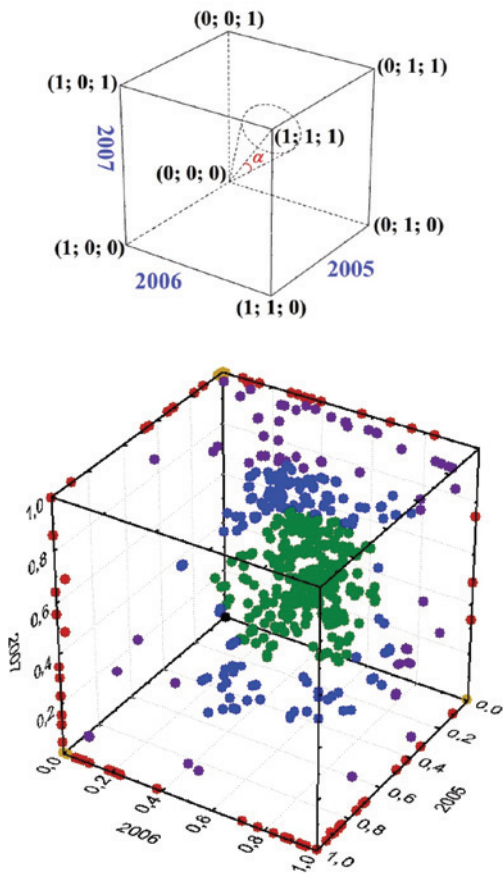


Fig. 3. Air routes on sides of three-dimensional cube

Each air route (depending on the stability of its state and the rate of change in aircraft traffic) is assigned a specific output parameter value: 0.0, 0.4, 0.6, 0.8, 0.9 and 1.0 (Tab. 2). This is determined based on the stability of this air route’s status over three years. To do this, we estimate the proximity of this air route’s point on the sides of the 3D cube to the diagonal of this cube, which is calculated using equation (1):

$$\cos \alpha = \cos(\vec{a} \wedge \vec{b}) = \frac{\vec{a} \times \vec{b}}{|\vec{a}| \times |\vec{b}|} = \frac{1 \cdot x_{2005} + 1 \cdot x_{2006} + 1 \cdot x_{2007}}{\sqrt{1^2 + 1^2 + 1^2} \times \sqrt{x_{2005}^2 + x_{2006}^2 + x_{2007}^2}}, \tag{1}$$

where $\vec{a} = \vec{a}(1;1;1)$ — cube diagonal, $\vec{b} = \vec{b}(x_{2005}; x_{2006}; x_{2007})$, x_i – total (normalised) number of flights in the i -th year.

The more stable the air route over the three years under investigation, the closer its point lies to the diagonal of the 3D cube. Thus, the analysis of air route stability over three years allows coming up with a more suitable interpretation of the notion *air route exists* and expands the output parameter value set that is used in developing the model.

Table 2. Output parameter values

| Existence of an air link in 2005–2007 | Angular range for α | Value of the output parameter | Interpretation of values |
|---------------------------------------|----------------------------|-------------------------------|--------------------------|
| 3 years | $[0 \div 10^\circ]$ | 1 | Existing air route |
| | $(10 \div 20^\circ]$ | 0.9 | |
| | $(20 \div 35^\circ]$ | 0.8 | |
| 2 years | | 0.6 | Non-existent air route |
| 1 year | | 0.4 | |
| 0 years | | 0 | |

2.3. Developing the model

The main factor defining the existence of an air route is *the purpose of the trip*: destination parameters of the primary city are chosen as destination parameters of the elements, and generational parameters of the secondary city (the one with less significant destination capabilities) are picked as the generational parameters of the elements.

We made a quantitative assessment of the efficiency of various combinations of parameters used to develop the model. To do this, we use a genetic algorithm method that is implemented in the STATISTICA software package (Neural Networks ... 2008; Rutkovskaya *et al.* 2006). As a result, it was decided to use six out of eight parameters as significant input parameters. The least significant parameters turned out to be the ones characterising the transport link, i.e. air route length and characteristic of air route direction.

A fuzzy neural network model of the conditions for the existence of an air route was developed using the software tools of the MatLab (Fuzzy Logic Toolbox and Neural Network Toolbox) (Leonenkov 2005; Shtovba 2007). The neuro-fuzzy network of the model is synthesised and refined using an adaptive neuro-fuzzy inference system (ANFIS) (Jang 1993) editor based on an input-output pair sample (Fig. 4). The training of this neuro-fuzzy network results in building a set of fuzzy rules that determine whether or not an air route must exist between a pair of cities under the given values of measurable parameters.

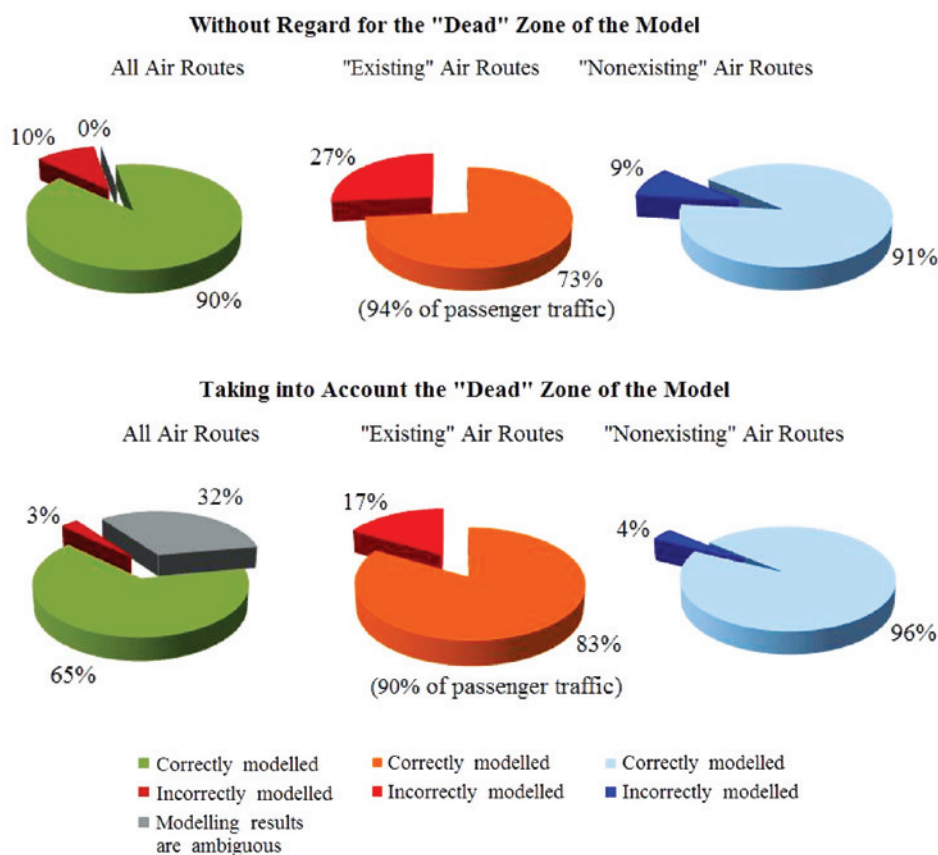


Fig. 4. Modelling results

3. Analysis of modelling results

Analysis of modelling results shows that air routes for which the modelled output parameter values are close to 0.5 are modelled unambiguously. We decided to establish a 'dead' zone corresponding to modelled output parameter values in the range of (0.3–0.7).

The results of modelling with and without taking into account the 'dead' zone are shown in figure 4. The model yields ambiguous results for 32% of air routes relative to the total number. Introduction of the 'dead' zone substantially decreased the number of incorrectly modelled air routes. Among the correctly modelled air routes, correct results are produced for 83% of air routes existing in 2006 and for 96% of non-existent air routes. Thus, introducing a 'dead' zone for the output variable allows us to enhance modelling quality.

Figure 5 demonstrates an OD matrix for the model produced by taking into account the 'dead' zone when

interpreting modelling results. The OD matrix is used to visualise the air route network. Each matrix cell corresponds to a probable air route. Different colours denote correctly and incorrectly modelled air routes that existed or did not exist in 2006. Air routes that were not considered or modelling results that are ambiguous are shown in white.

We analysed the incorrectly modelled air routes. Air routes in the class (10) 'air route exists in reality but not according to the model' (red marker) are mostly confined to inner regional routes, i.e. they are essentially local air routes. Air routes in the class 'air route exists according to the model but not in reality' (01) (blue marker) are mostly located in columns corresponding to cities with substantial target capabilities and can be interpreted as promising air routes. Otherwise, the model correctly represents the structure of Russia's trunk air routes.

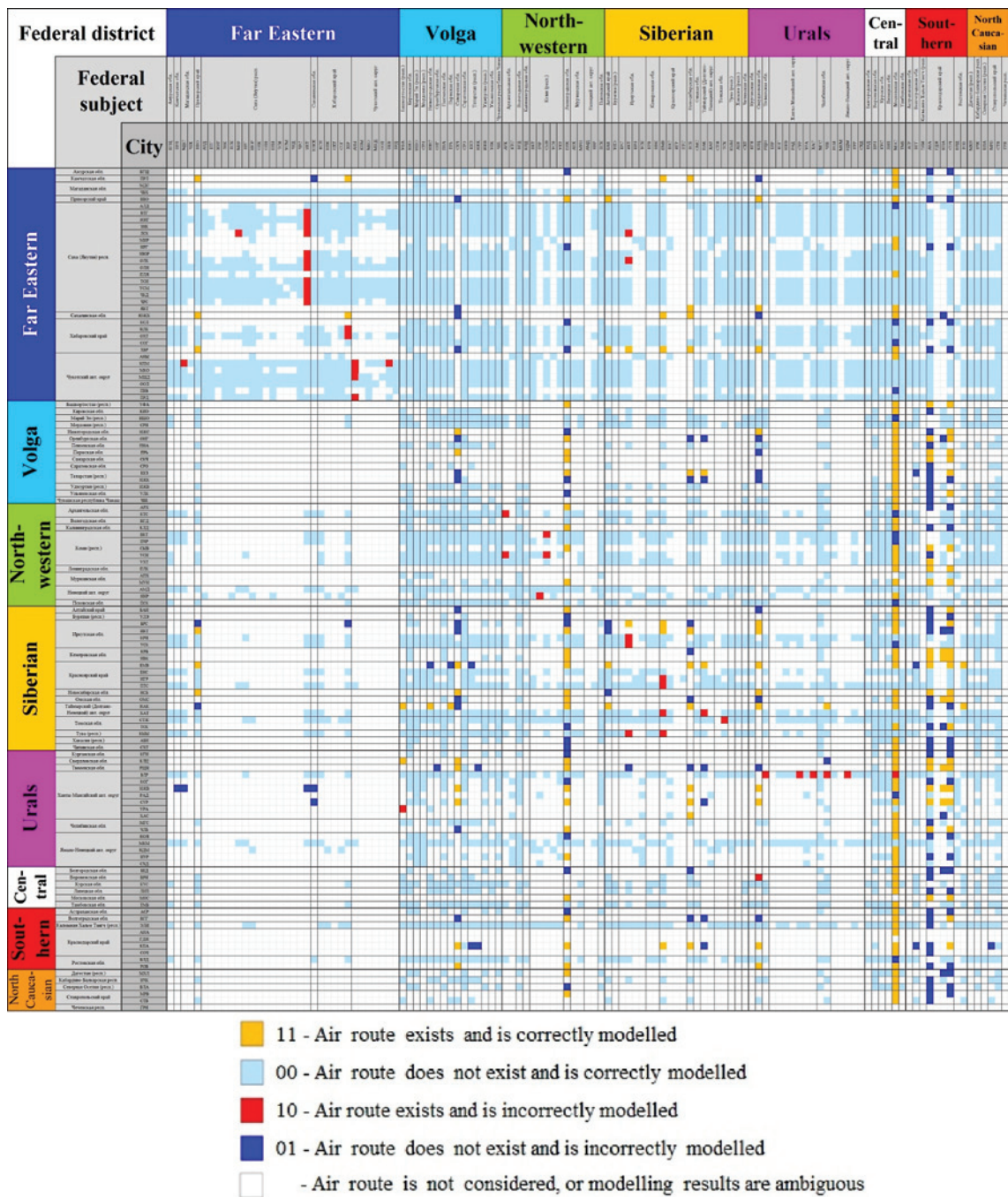


Fig. 5. Model OD matrix (taking into account the 'dead' zone of the model)

4. Conclusions

The technology of adaptive neuro-fuzzy inference systems was used to develop an information model for conditions for the existence of air routes. The model unambiguously models the existence or non-existence of direct air links for 70% of city pairs. Among air routes that are modelled unambiguously, correct results are obtained for 83% of existing (as of 2006) and 96% of non-existing

trunk air routes between Russian cities. These air routes account for 90% of annual passenger traffic. Thus, the model that is developed reflects the structure of the core of Russian trunk air routes correctly. The model of the conditions for the existence of air routes can be used to predict the topology of the development of the Russian trunk air route network through to 2020.

References

- Blinova, T. O. 2007. To forecast passenger traffic flows in Russia, *Aviation* 11(1): 28–34.
- Borisov, V. V.; Kruglov, V. V.; Fedulov, A. S. 2007. *Fuzzy Models and Networks*. Moscow: Goryachaya liniya – Telekom. 284 p.
- Haykin, S. 1999. *Neural Networks – a Comprehensive Foundation*. 2nd ed. New Jersey: Prentice Hall. 842 p.
- Jang, J.-S. R. 1993. ANFIS: adaptive-network-based fuzzy inference system, *IEEE Transactions on Systems, Man, and Cybernetics* 23(3): 665–685.
<http://dx.doi.org/10.1109/21.256541>
- Kohonen, T. 1990. The self-organizing map, in *Proceedings of the IEEE* 78(9): 1464–1480.
<http://dx.doi.org/10.1109/5.58325>
- Kohonen, T. 2001. *Self-Organizing Maps*. 3rd ed. Berlin, Heidelberg, New York: Springer-Verlag. 501 p.
<http://dx.doi.org/10.1007/978-3-642-56927-2>
- Kosko, B. 1994. Fuzzy systems as universal approximators, *IEEE Transactions on Computers* 43(11): 1329–1333.
<http://dx.doi.org/10.1109/12.324566>
- Leonenkov, A. V. 2005. *Fuzzy Modeling in the MATLAB and the FuzzyTECH*. St. Petersburg: BHV – Petersburg. 736 p.
- Neural Networks. *STATISTICA Neural Networks: Methodology and Technology of Modern Data Analysis*. 2008. 2nd ed. Moscow: Goryachaya liniya – Telekom. 392 p.
- Piegat, A. 2001. *Fuzzy Modeling and Control*. Heidelberg: Physica Verlag. 728 p.
- Rutkovskaya, D.; Pilinskiy, M.; Rutkovskiy, L. 2006. *Neural Networks, Genetic Algorithms and Fuzzy Systems*. Moscow: Goryachaya liniya - Telekom. 452 p.
- Shtovba, S. D. 2007. *Design of Fuzzy Systems by Means MATLAB*. Moscow: Goryachaya liniya - Telekom. 288 p.
- Vasermanis, E. K.; Nechval, N. A.; Nechval, K. N. 2004. Optimal allocation of airplanes to routes, *Aviation* 8(1): 3–7.