

PETAR ANDRAŠIĆ¹

(Corresponding author)

E-mail: pandrasi@fpz.hr

TOMISLAV RADIŠIĆ, Ph.D.¹

E-mail: tradisic@fpz.hr

DORIS NOVAK, Ph.D.¹

E-mail: dnovak@fpz.hr

BILJANA JURJIČIĆ, Ph.D.¹

E-mail: bjuricic@fpz.hr

¹ University of Zagreb

Faculty of Transport and Traffic Sciences

Vukelićeva 4, 10000 Zagreb, Croatia

Science in Traffic and Transport

Original Scientific Paper

Submitted: 12 June 2018

Accepted: 4 Apr. 2019

SUBJECTIVE AIR TRAFFIC COMPLEXITY ESTIMATION USING ARTIFICIAL NEURAL NETWORKS

ABSTRACT

Air traffic complexity is usually defined as difficulty of monitoring and managing a specific air traffic situation. Since it is a psychological construct, best measure of complexity is that given by air traffic controllers. However, there is a need to make a method for complexity estimation which can be used without constant controller input. So far, mostly linear models were used. Here, the possibility of using artificial neural networks for complexity estimation is explored. Genetic algorithm has been used to search for the best artificial neural network configuration. The conclusion is that the artificial neural networks perform as well as linear models and that the remaining error in complexity estimation can only be explained as inter-rater or intra-rater unreliability. One advantage of artificial neural networks in comparison to linear models is that the data do not have to be filtered based on the concept of operations (conventional vs. trajectory-based).

KEY WORDS

air traffic complexity; complexity estimation; artificial neural networks; genetic algorithm; human-in-the-loop simulation; air traffic management;

1. INTRODUCTION

Air traffic complexity is usually defined as difficulty of monitoring and managing a specific air traffic situation [1]. In a given sector, air traffic controller (ATCO) has to

ensure that all aircraft reach their destinations or exit points in a safe and efficient manner. To achieve this, the ATCO has to monitor current traffic situation, predict aircraft trajectories, anticipate possible conflicts and solve them, all while taking care to accommodate the requests from pilots to the most possible extent. Complexity arises from interactions among the aircraft and between the aircraft and the airspace. Air traffic complexity is defined as that part of complexity which arises from interactions among the aircraft. However, since different types of interactions have different levels of difficulty in resolving them and interactions can be directly or indirectly related and connected, it is wrong to equate the number of interactions directly with complexity. It is also wrong to equate air traffic complexity and workload, although it has been proven that the increase in complexity results in the increase in workload which is the limiting factor of airspace sector capacity [2, 3]. Mogford et al. [4] reviewed many research papers searching for relationship between complexity and workload. Their conclusion was that the complexity is a crucial driver of workload but that the relationship between complexity and workload is not straightforward; it is mediated by other factors, such as equipment quality, individual differences, and controller cognitive strategies (Figure 1).

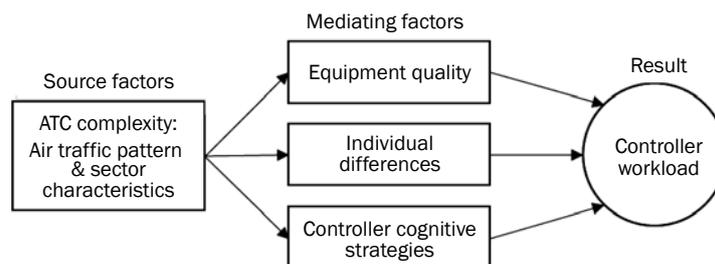


Figure 1 – Relationship between ATC complexity and workload [4]

Even though the source of complexity is known (interactions) and the consequences of complexity are also known (workload), the complexity itself remains firmly inside the mind of the ATCO. On one side, many methods of assessing workload were successfully used in the past [5–7]. However, since complexity is a psychological construct, it is impossible to assess it directly. It can only be inferred from interactions among aircraft or between aircraft and airspace. On the other side it is quite easy to detect interactions based on the traffic situation (defined as aircraft positions, velocity vectors, intent, route, airspace configuration, etc.). But, as mentioned previously, not all interactions are equal in terms of the complexity they generate and there are also possible multiplicative effects among interactions that make the complexity as a whole significantly greater than the sum of individual interactions. Therefore, to get from the traffic situation to complexity there is a leap that cannot be made without input from ATCOs. Multiple research teams were focused on using ATCO expert assessment of traffic situation complexity to determine which type of interactions influence the complexity most.

The main approach taken in such a research was to define metrics which can be used to determine the number of interactions (e.g. number of conflicts) and check them for correlation with ATCO subjective complexity assessment scores. One of the metrics or, as they will be called from here on, complexity indicators often used is traffic density which is a number calculated by dividing the number of aircraft with the volume of the airspace they occupy. In all research so far it was proven that it can by itself explain up to 50% of variance in subjective complexity scores [7]. Laedeman et al. [8] expanded on the notion of the traffic density by introducing Dynamic Density which they defined as a combination of 'both traffic density (a count of aircraft in a volume of airspace) and traffic complexity (a measure of the complexity of the air traffic in a volume of airspace)'. Authors used informal interviews with controllers to obtain a list of eight complexity factors to be used in dynamic density equation. Only criterion was that the factors could be calculated from the radar tracks or their extrapolations. The intention was to obtain an objective measure of controller workload based on the actual traffic. Their results showed that the dynamic density was able to account for 55% in controller activity variation.

Three other teams [9–11] working under the Dynamic Density Programme developed additional 35 complexity indicators, which were later successfully validated as a group by Kopardekar et al. [12]. Unfortunately, it was later shown that the complexity indicator weights were not universal to all airspace sectors, i.e. they had to be adjusted on a sector by the sector basis [13]. Furthermore, the same authors [13] suggested that, due to possibly non-linear interactions between

complexity factors, the Dynamic Density performance could be improved by using non-linear techniques such as non-linear regression, genetic algorithms, and neural networks, which is exactly what was tested in this paper.

In our previous work, similar approach was used to test the effect of the trajectory-based operations (TBO) on air traffic complexity [14]. ATCOs were recruited to perform human-in-the-loop (HITL) simulations during which they were asked to provide real-time assessment of air traffic complexity. Linear regression model was used to select, among 20 most used complexity indicators, those indicators which correlated best with subjective complexity scores. Six indicators were used to generate a predictive linear model that performed well in conventional operations but less so under TBO. Therefore, two of seven novel TBO-specific complexity indicators were defined and experimentally validated. A second correlation model combining these two novel indicators with four already in use generated much better predictions of complexity than the first model. Nonetheless, the best correlation that was achieved was $R=0.83$ (R^2 -adjusted = 0.691). This paper will show how artificial neural networks and genetic algorithms can be used to estimate the complexity in a novel manner.

2. METHODOLOGY

This Section presents a brief overview of the experiment which produced data for training artificial neural networks (ANNs). The details of the experiment and analysis can be found in [14] and [15].

2.1 Apparatus

Human-in-the-loop simulator trials were deemed the best method of collecting subjective complexity score because such trials are more representative of the real operations than the fast-time simulations or observational studies. Additionally, review of the literature showed that many organizations and researchers had successfully used HITL simulations for complexity assessment (e.g. [10, 16]). Since the simulator that was available did not support TBO (i.e. did not have the options of generating, de-conflicting, and executing the 4D trajectories) and did not have the options to record all of the necessary data (i.e. all aircraft states, complexity indicator values, human-machine interface interactions), it has been decided that the best course of action was to develop a custom HITL ATC simulator with minimum required features (*Figure 2*). More details on the simulator development can be found in [15].

The following aims for the development of the ATC simulator have been set:

- Accurate and versatile aircraft model;

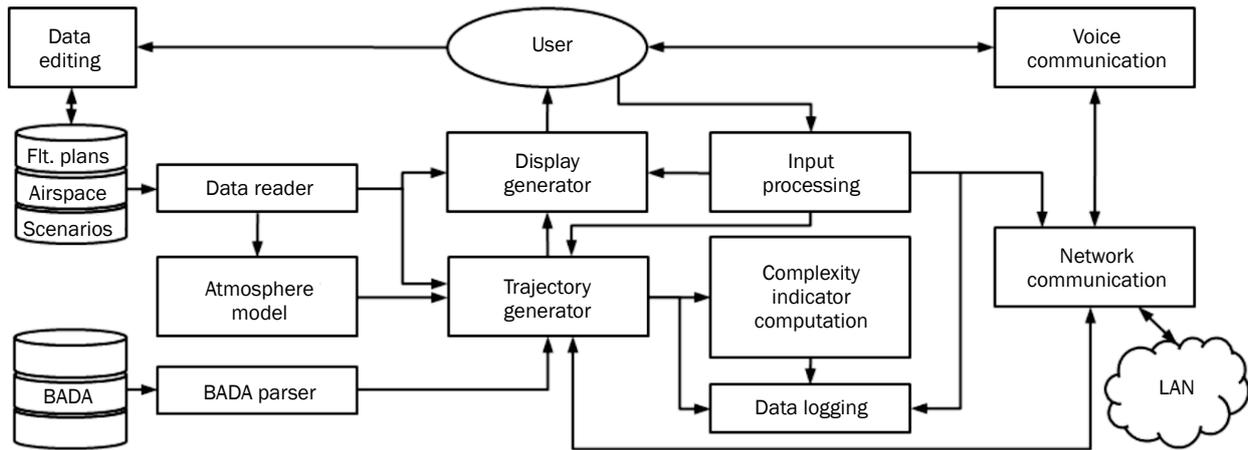


Figure 2 – Simulator software outline [15]

- Realistic working environment;
- Ability to record all necessary data;
- Support for TBO;
- Simple meteorological model;
- Simple surveillance model;
- Easy data editing;
- Voice and datalink communication.

Before the simulator was used in the research, it was validated using a series of tests. The user interface and tool operation were validated by comparison with operational ATC systems and by expert assessment. Aircraft models were validated by comparison with real-life flight data collected from quick access recorders, and working environment was validated by the licensed ATCOs [15].

2.2 Airspace and traffic

Airspace in which the simulations were performed has been chosen according to the following requirements [14]:

- Participants (ATCOs) had to be familiar with the airspace. This ensures that they can accurately assess the air traffic complexity. It also saves time on pre-simulation training and removes the possibility of different learning rates affecting the results.
- Airspace data must be available. Most European air navigation service providers (ANSPs) nowadays share their airspace data on-line for free.
- Available flight data upon which the flights for the simulation scenarios were created must include flights passing through the desired airspace. In this way realistic flight data can be used.
- Since this research considers only en-route operations, a sector of en-route (upper) airspace will be used.

One of the sectors which met all these criteria and the one chosen for the experiment is Zagreb Upper North sector. Geographically, the sector consists of airspace over northern Croatia. Vertically, the sector, as used in this research, starts at FL 285 and ends at

FL 660. In reality, due to traffic demand, the sector is often vertically divided into several elemental sectors depending on the traffic loads and in that case 'Upper' is used to describe the sector from FL 325 - FL 355. For this research the complete vertical expanse was used. Most flights flying through this sector are over-flights while minority of flights depart or arrive in Croatia [14].

The transfer of traffic between neighbouring Area Control Centres (ACC) and Zagreb ACC is regulated by Letters of Agreement (LoA). For this research the relevant parts of LoAs were Flight Level Allocation and Special Procedures sections which state the conditions that have to be met for all flights crossing the boundary of the controlled area (called Flight Level Allocation Scheme - FLAS). The purpose of FLASes is to ensure that flights will cross the controlled area (CTA) boundary at required flight levels that enable them to land at the desired airport or to be seamlessly joined with the existing traffic. It also states what the coordination points (COP) or transfer of control (TOC) points are. The participants were required to adhere to these procedures during the simulation runs [14].

To ensure representativeness of the traffic flows, actual historic traffic data were used to create simulation scenarios. Off-peak traffic was used to generate traffic for low-traffic scenarios; peak traffic was used to generate traffic for high-traffic scenarios, and additional flights were generated in addition to peak traffic to create traffic for future-traffic levels scenarios. An analysis was made of historic traffic data to determine the aircraft type distribution and most commonly used routes. The results of this analysis were used to make the traffic which was generated for future-traffic scenarios more representative than the actual traffic [14].

2.3 Participants

Participants in the experiment were recruited among the active licenced air traffic controllers working with the Croatia Control Ltd., Croatian air navigation

service provider (ANSP). All controllers had at least 4 years of experience working at ANSP and a minimum of 2 years had passed since they acquired the air traffic control license. All 10 of them had prior experience of controlling the traffic in the airspace sector which was used for the experiment [14].

All the participants were briefed before the commencement of the experiment. The briefing included the following topics: air traffic complexity, subjective complexity rating scale, trajectory-based operations, tools and functions of the simulator, airspace, simulator scenarios, and operational procedures. Also, participants were given a short simulator training of a minimum of 90 minutes (two scenarios, one with conventional and one with trajectory-based operations).

Besides ATCOs, one of the authors participated as the pseudo-pilot in all experiments. The controller could communicate with the pseudo-pilot only via a headset. Since the pseudo-pilot had to take on the role of air traffic controllers in other air traffic control units, to facilitate coordination, an assistant to the pseudo-pilot participated in the scenarios involving high and future traffic levels. The experiments did not involve planner controllers, only the executive ones [14].

3. DATA ANALYSIS AND RESULTS

This Section will present data gathered during the initial experiment, how they were processed and analysed. More details on the initial analysis, which was done using multiple step-wise linear regression, can be found in [14].

3.1 Data recording and processing

Throughout the complexity measurement experiment, three categories of data were recorded [14]:

- Raw aircraft state data. It records the complete aircraft state of each aircraft in a scenario at a 1-second interval. These data can be used to replay the simulation runs and to calculate values of complexity indicators in post-processing. This was useful for calculating the values of the new complexity indicators that were developed after the initial analysis of the current complexity indicators.
- Values of complexity indicators. A set of 20 complexity indicators were coded in the simulator itself. This code produced values for each of the indicators at a 1-second interval. Their values were recorded for post-experiment analysis.
- Subjective complexity scores. During the simulator runs, the research participants (air traffic controllers) assessed the air traffic complexity using a scale from 1 to 7.

The following is a complete list of complexity indicators which were used in the analysis [14]:

- Aircraft count;

- Aircraft density I (based on sector volume);
- Aircraft density II (based on convex hull volume);
- Aircraft density II squared (based on convex hull volume and the squared number of aircraft);
- Separation criticality index;
- Number of aircraft with horizontal separation less than 8 NM;
- Inverse of minimum horizontal separation in the same vertical neighbourhood;
- Inverse weighted mean of aircraft vertical separation;
- Ratio of standard deviation to mean value for ground speed;
- Ratio of mean aircraft distance to number of aircraft;
- Altitude variance;
- Fraction of aircraft with fewer than 600 seconds to conflict;
- Fraction of aircraft climbing;
- Fraction of aircraft descending;
- Fraction of aircraft either climbing or descending;
- Number of aircraft pairs at a 3D Euclidean distance less than 5 NM;
- Number of aircraft pairs at a 3D Euclidean distance of 10-15 NM;
- Ground speed variance;
- Fraction of TBO aircraft; and
- Standard deviation of aircraft headings.

The subjective complexity score was entered by clicking on an on-screen panel which opened once every 120 seconds. Each participant did nine simulation scenarios, each lasting approximately 50 minutes, which means that in theory there should be 25 subjective complexity scores per participant per scenario. This totals to 225 complexity scores per participant, or 2,250 complexity scores overall. In reality, only 1,997 complexity scores were entered. This is due to three reasons. First, one participant only did 7 scenarios (this accounts for 50 missing scores) before leaving the research for personal reasons. Second, some participants could not finish all of the future-traffic level scenarios because they were designed to be very difficult (this accounts for 57 scores). Third, some participants did not enter the scores as soon as they were prompted. In some cases this was due to intense focus on controlling the traffic. This accounts for the remaining 146 missing scores [14].

3.2 Genetic algorithm search

The purpose of the analysis was to determine whether artificial neural networks (ANN) could be used to estimate air traffic complexity based on the values of the complexity indicators, and if they could, how good was the performance of ANN compared to linear regression methods. As mentioned previously, in our earlier experiment the best correlation between

the values produced by linear prediction model and the actual subjective complexity scores was $R=0.83$ (R^2 -adjusted = 0.691). Previously, Gianazza and Guitet had also used ANNs to evaluate air traffic complexity metrics by comparing their values with the sector status [17]. Their assumption was that the sector was split at the time when the workload was too high and, on the other hand, two sectors were merged when the workload was low. They also assumed that the sole reason for the increased workload was an increase in complexity. Compared to their method, we have used actual complexity scores provided by the licensed air traffic controllers during HITL simulations as a measure of air traffic complexity.

Due to relatively small number of input parameters (20 complexity indicators), multi-layer perceptron (MLP) network architecture was selected for the ANN which will be trained to estimate the complexity (Figure 3). MLPs are proven to be universal function approximators which means that, among other applications, they can be used to approximate non-linear relationships between data [18]. Here, MLP was used to estimate air traffic complexity based on the values of air traffic complexity indicators.

In this case, three layers were used: input, hidden and output layer. Input layer initially had 20 neurons to match the number of input parameters. There was only one output parameter, the subjective complexity score estimation; therefore, there was only one output neuron. To allow for possible performance improvements which could be achieved by changing the number of neurons in the hidden layer, a genetic algorithm was used to search for the best-performing hidden layer configuration. Also, the genetic algorithm was used to try out different combinations of complexity indicators

General genetic algorithm which was used to optimize the ANN consisted of the following steps:

- *Generate a population of neural networks* with uniformly distributed number of neurons in the hidden layer and randomly selected number and combination of complexity indicators. In this case, 8,000 randomly generated ANNs were used for initial population. Since the number of possible

combinations of 20 complexity indicators is $2^{20} - 1 \approx 1$ million, and the number of neurons in the hidden layer was limited to 40, the complete search space included approximately 4×10^7 combinations. This means that the initial population included less than 0.02% of all possible combinations.

- *Train the ANNs on the 80% of the data set.* ANNs were trained using supervised learning where backpropagation was used to adjust the weights of neurons by calculating the gradient of the loss function. This was performed using Levenberg-Marquardt algorithm [19].
- *Evaluate the ANNs on the remaining 20% of the data set.* Using different set of data to evaluate the ANNs than the one used to train them makes it possible to test for generalization of the learning. In this way, those ANNs which produced overfitted model with poor general performance will not be favoured over the rest of ANNs.
- *Sort the ANNs according to performance.* As a measure of performance mean squared error (MSE) was used. ANNs with the lowest error were put on the top of the list.
- *Multiply ANNs according to their rank.* Bottom half of the list (worst performing networks) was discarded and the rest were duplicated according to their rank: ANNs from 2,000th-4,000th position were copied once, those from 1,000th-2,000th position were copied twice, those from 500th-1,000th position were copied three times, etc. This enabled the most successful ANNs to increase their numbers so that those parts of the search space were more thoroughly explored in search of the best solution. Due to rounding down each time the remaining list of the ANNs was halved, this method of ANN duplication did not generate 8,000 networks as in the beginning of the algorithm. Therefore, the rest of the spots (12) were filled with randomly selected ANNs from the bottom half of the list.
- *Mutate the ANNs.* In this part of the algorithm the two variable parameters, the number of hidden neurons and the number of complexity indicators, were

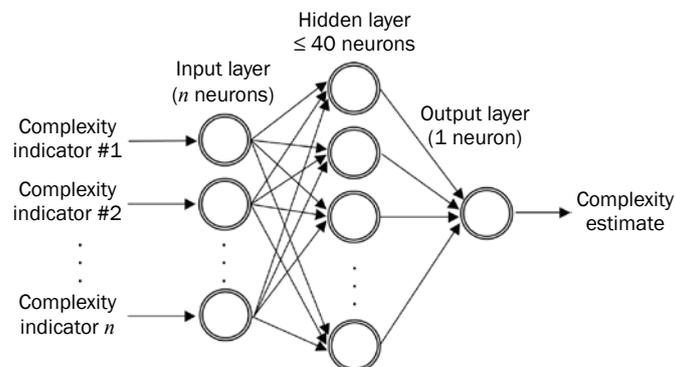


Figure 3 – Multilayer perceptron as used for complexity estimation

semi-stochastically modified. With equal probability, the number of neurons in the hidden layer was increased by one, decreased by one, or not modified at all. Same with the number of complexity indicators. In 1/10th of the ANNs crossbreeding occurred, where the number of neurons in the hidden layer in network A was copied to network B, and vice versa.

- Return to the training step.

This procedure was repeated 45 times or, as it is often called in reference to genetic algorithms, generations. This number of generations was chosen to enable the best performing ANNs to change their initial configuration (e.g. number of neurons in the hidden layer) by the amount equal to the maximum possible change from any starting point (e.g. maximum change in number of neurons in the hidden layer is 39, from 1 to 40). Later analysis showed that 45 generations was more than enough because the algorithm reached showed no sign of further improvement after only 15 generations.

3.3 Results of the genetic algorithm search

For each generation of the genetic algorithm, a population average was calculated (Figure 4). It can be seen that the population average MSE has stopped decreasing in the 15th generation. Since there is only one output (complexity score in range from 1 to 7), the MSE was intentionally non-normalized, so it can be used to directly assess the error in terms of complexity scores.

Initial distribution of the number of neurons was uniform in the range from 1 to 40. It can be seen in Figure 5 that the population-wide average number of neurons in the hidden layer decreased from initially around 20 to only 4 by generation 25. Better performing networks were those with smaller number of neurons in the hidden layer even though no cost was associated with having more neurons. This is probably because ANNs with more neurons overfitted the training data and thus performed poorly on the validation data set.

The number of complexity indicators used as inputs to the ANN remained high (average of 17.55 indicators out of possible 20) throughout the search indicating that the best performing networks did find some utility in using more indicators. However, not all indicators were used with equal frequency as can be seen in Table 1 which shows how often each indicator was used in best performing ANNs for each of the 45 generations. Though frequency of use is not the best measure of utility and it is not directly comparable to multiple step-wise linear regression results from the previous work [14], an overlap between those indicators selected by linear regression and those most frequently used by ANNs can be seen (indicators used in linear regression model are marked with an asterisk in Table 1).

Lastly, the ANN with best complexity estimation performance will be presented. It was first identified in generation 30; it had 4 neurons in the hidden layer, used 19 complexity indicators and performed with MSE of 0.5249 on the validation data set. The training and validation performance can be seen in Figure 6.

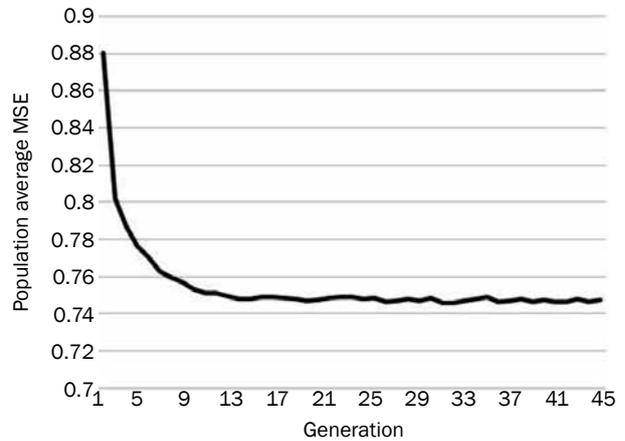


Figure 4 - Population average MSE (non-normalized)

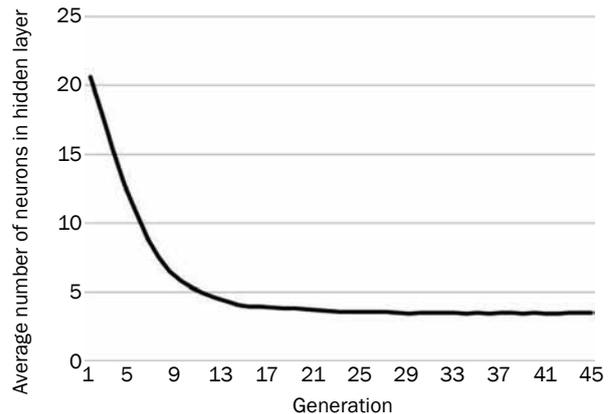


Figure 5 - Population-wide average number of neurons in the hidden layer

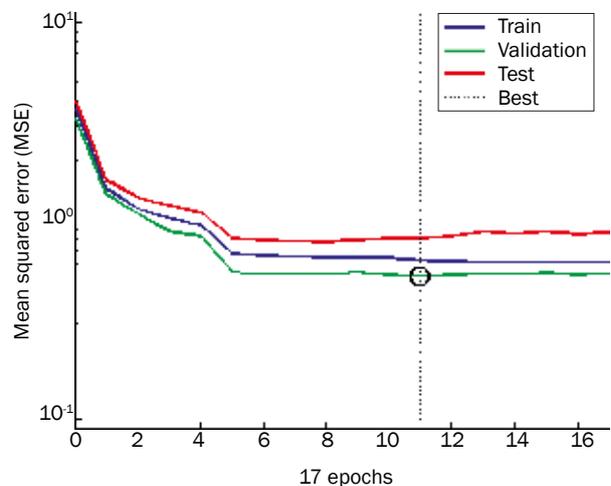


Figure 6 - Training and validation performance of best ANN

Table 1 – Complexity indicators frequency of use

Complexity indicator	Frequency of use	Complexity indicator	Frequency of use
Fraction of TBO aircraft *	100%	Fraction of aircraft climbing	84%
Standard deviation of aircraft headings *	91%	Aircraft density II (based on convex hull volume)	84%
Aircraft count *	91%	Separation criticality index	82%
Fraction of aircraft with fewer than 600 seconds to conflict	89%	Aircraft density II squared (based on convex hull volume and the squared number of aircraft)	82%
Number of aircraft pairs at a 3D Euclidean distance less than 5 NM *	87%	Ground speed variance	80%
Fraction of aircraft either climbing or descending *	87%	Ratio of standard deviation to mean value for ground speed	80%
Inverse weighted mean of aircraft vertical separation	85%	Number of aircraft pairs at a 3D Euclidean distance of 10-15 NM	78%
Ratio of mean aircraft distance to number of aircraft *	85%	Aircraft density I (based on sector volume)	78%
Inverse of minimum horizontal separation in the same vertical neighbourhood	84%	Altitude variance	78%
Fraction of aircraft descending	84%	Number of aircraft with horizontal separation less than 8 NM	73%

*used in linear regression model

This ANN was also retested with the complete data set and MSE was equal to 0.6415 which is still below the population mean for all ANNs. To make the comparison with the linear regression model possible, the correlation coefficient was calculated and it is $R=0.783$ which is somewhat lower than the correlation achieved by linear regression model ($R=0.83$). Likewise, R^2 is 0.6087 which is lower than the one achieved by linear regression model ($R^2=0.693$). However, these results have been achieved on the complete data set as opposed to the TBO-only data set used for linear regression. The data were not filtered in this case because ANNs were expected to adjust automatically to the difference in concept of operations (conventional vs. TBO) which was only partially the case. Linear regression model, when used on the complete set of data, achieved performance of $R=0.746$ and R^2 -adjusted=0.554 [14]. The comparison of performance of all models can be seen in Table 2.

Standard error of the estimate of the subjective air traffic complexity is the smallest in the ANN model but not by a large margin. It can be seen in Figure 7 that in 49% of cases the ANN estimates the complexity score to be equal to the one proposed by air traffic

controllers. In 44% of cases the ANN estimates differ from the controller's by one. There are still approximately 7% of cases in which the ANN estimates differ from the controller's by 2 or more levels. Some ideas on the cause of this can be found in the Discussion.

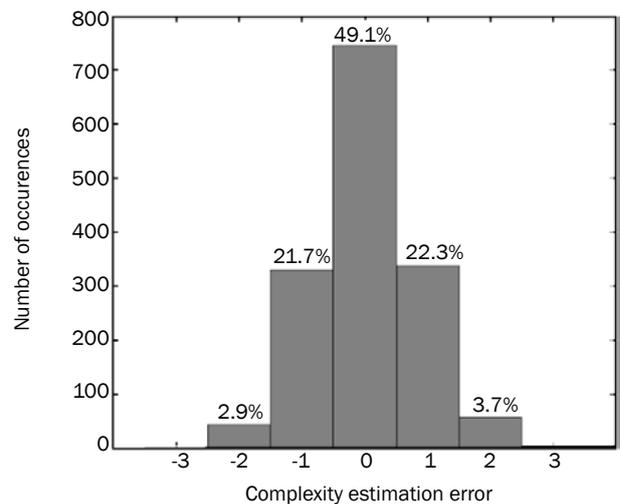


Figure 7 – ANN complexity estimation error distribution

Table 2 – Comparison of performance of different complexity estimation models

	Kopardekar et al. [16]	Linear model [14]	TBO-specific linear model [14]	ANN model
R	0.83	0.746	0.833	0.783
R^2	0.69	0.556	0.693	0.609
R^2 -adjusted	N/A	0.554	0.691	N/A
Std. error of the estimate	N/A	0.851	0.889	0.801

4. DISCUSSION

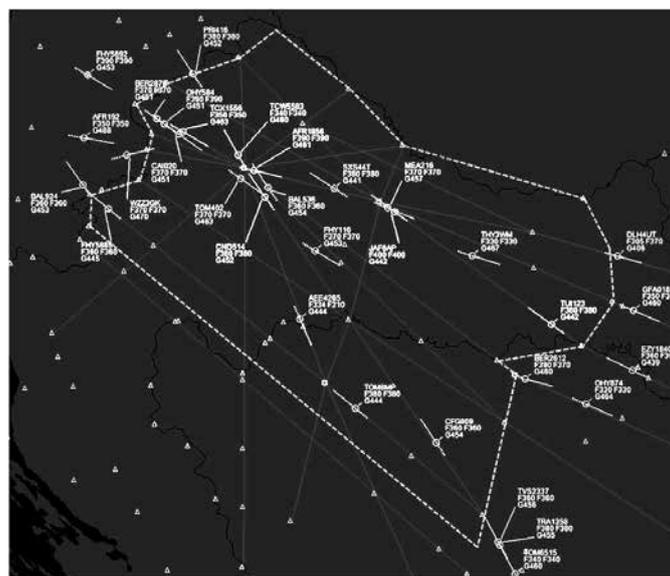
Results have shown that ANNs can be used to determine air traffic complexity with accuracy similar to linear models. The genetic algorithm search for the best ANN configuration has shown that smaller networks, with less neurons in the hidden layer, outperform the larger ones. Also, the genetic algorithm search has shown that the best performing ANNs used almost all complexity indicators available. This does not mean that all of them are equally useful (or useful at all), because weights between neurons could have been

adjusted to discard the data which do not improve accuracy of complexity estimation. However, the analysis of frequency of the complexity indicator usage in the population of the best performing ANNs shows that some of the indicators are very useful for complexity estimation. These are mostly the same indicators which were identified as useful by linear regression.

Overall, the performance of ANNs is mostly equal to linear models. This means that the issues with linear model accuracy did not stem from non-linear relations between indicators and subjective complexity scores. Instead, there might be some other source of



Figure 8 – Example of intra-rater unreliability; both situations were given the same complexity score (=1)



Complexity scores: ATCO 1: 4, ATCO 2: 3, ATCO 4: 2, ATCO 5: 5

Figure 9 – Example of inter-rater unreliability; complexity scores vary from 2 to 5 for the traffic situation presented

error and that source is not directly related to the complexity estimation model. Further investigation was done into other possible sources of error and one was identified – human expert inconsistency. The main issue that was encountered with the method used in this research was low intra-rater and inter-rater reliability. In this context, a rater is someone who rates or gives scores of something. Here, air traffic controllers were raters since they were the ones assessing the complexity in the given traffic situations. Intra-rater reliability (or consistency) is the degree of agreement among multiple ratings by a single rater, while inter-rater reliability is the degree of agreement among multiple raters. In the most extreme example of intra-rater unreliability, there were a few controllers who rated the whole scenario with the same complexity score. This led to the result of having the same scores for the situation with almost no aircraft (*Figure 8a*) compared to the situation with 12 aircraft simultaneously present in the airspace (*Figure 8b*).

Inter-rater reliability, or rather, unreliability, was also prominent in many cases. Same traffic situations were rated with very different complexity scores by different controllers (*Figure 9*). Neither linear model nor ANN can give different outputs on the basis of the same input.

5. CONCLUSION

In this paper a novel approach to air traffic complexity estimation has been presented. A genetic algorithm search was used to find the ANN configuration with best performance in terms of complexity estimation accuracy. It was found that the multi-layer perceptron needed only a few neurons in the hidden layer (4 in the best ANN) and that the ANNs routinely used almost all complexity indicators available. Overall, the performance of ANN is similar to the linear estimation model which proves that the issues with model accuracy could not be improved with non-linear methods as previously proposed. It was concluded that the remaining variance in subjective complexity scores cannot be explained by traffic characteristics. Another source of error was proposed, which is rater unreliability. ATCOs could not score (rate) traffic situations with perfect consistency among themselves or between different traffic situations. This makes any other attempt at modelling the complexity on the basis of complexity scores provided by experts bound to end with similar results.

Finally, in our future work we will try to find a method which could determine complexity in a way which requires a little less input from experts and relies more on the traffic situation data, though it is impossible to explore complexity without involving humans. One such method could be made by asking ATCOs to rate traffic situations relative to each other. We believe that this method could achieve higher inter-rater and intra-rater reliability than giving scores on a fixed scale. The problem with that method is that it cannot be used with real-time simulations. Another method could be to give the ATCOs a traffic situation and asking them to

state all interactions they see and which steps need to be taken to resolve all conflicts. The number of interactions and the number of steps to resolve the conflicts could be used as a measure of complexity. Overall, we find that the problem of estimating complexity, even after almost two decades of research on the global scale, has still remained an unsolved challenge.

PETAR ANDRAŠI¹

E-mail: pandrasi@fpz.hr

Dr. sc. **TOMISLAV RADIŠIĆ¹**

E-mail: tradisic@fpz.hr

Dr. sc. **DORIS NOVAK¹**

E-mail: dnovak@fpz.hr

Dr. sc. **BILJANA JURJIČIĆ¹**

E-mail: bjuricic@fpz.hr

¹ Sveučilište u Zagrebu, Fakultet prometnih znanosti
Vukelićeva ulica 4, 10000 Zagreb, Hrvatska

SAŽETAK

Kompleksnost zračnog prometa uobičajeno se definira kao težina praćenja i upravljanja određenom situacijom u zračnom prometu. Budući da je to psihološki konstrukt, najbolja mjera kompleksnosti je ona koju daju kontrolori zračnog prometa. Međutim, postoji potreba da se napravi metoda za procjenu kompleksnosti koja se može koristiti bez anketiranja kontrolora. Za tu su potrebu do sada korišteni uglavnom linearni modeli. Ovdje istražujemo mogućnost korištenja umjetnih neuronskih mreža za procjenu kompleksnosti. Korišten je genetski algoritam u potrazi za najboljom konfiguracijom umjetne neuronske mreže. Naš zaključak je da umjetne neuronske mreže djeluju jednako dobro kao i linearni modeli i da se preostala pogreška u procjeni kompleksnosti može objasniti samo kao nepouzdanost u procjenama kontrolora. Jedna od prednosti umjetnih neuronskih mreža u usporedbi s linearnim modelima je u tome što se podaci ne moraju filtrirati na temelju koncepta operacija (konvencionalnog naspram temeljenog na putanjama).

KLJUČNE RIJEČI

kompleksnost zračnog prometa; procjena kompleksnosti; umjetne neuralne mreže; genetski algoritmi; simulacije s čovjekom u petlji; upravljanje zračnim prometom;

REFERENCES

- [1] Meckiff C, Chone R, Nicolaon J-P. The tactical load smoother for multi-sector planning. *Proceedings of the 2nd usa/europe air traffic management research and development seminar*; 1998.
- [2] Christien R, Benkouar A, Chaboud T, Loubieres P. Air traffic complexity indicators & ATC sectors classification. *Proceedings of the 21st Digital Avionics Systems Conference, 27-31 Oct. 2002, Irvine, CA, USA*, vol. 1. IEEE; 2002. p. 2D3-2D3.
- [3] Majumdar A, Ochieng W. Factors affecting air traffic controller workload: Multivariate analysis based on simulation modeling of controller workload. *Transportation Research Record: Journal of the Transportation Research Board*. 2002;1788(1): 58-69.
- [4] Mogford RH, Guttman J, Morrow S, Kopardekar P. *The Complexity Construct in Air Traffic Control: A Review*

- and Synthesis of the Literature. CTA INC MCKEE CITY NJ; 1995.
- [5] Schmidt DK. On modeling ATC work load and sector capacity. *Journal of Aircraft*. 1976;13: 531-537.
- [6] Hurst MW, Rose RM. Objective Job Difficulty, Behavioural Response, and Sector Characteristics in Air Route Traffic Control Centres. *Ergonomics*. 1978;21(9): 697-708.
- [7] Stein E. *Air traffic controller workload: An examination of workload probe*. Atlantic City, New Jersey: FAA; 1985.
- [8] Laudeman IV, Shelden S, Branstrom R, Brasil C. *Dynamic density: An air traffic management metric*. Ames Research Center. Report No. A-99-10366, 1998.
- [9] Chatterji G, Sridhar B. Measures for air traffic controller workload prediction. *1st AIAA, Aircraft, Technology Integration, and Operations Forum*; 2001, p. 5242.
- [10] Pawlak WS, Brinton CR, Crouch K, Lancaster KM. A framework for the evaluation of air traffic control complexity. *AAIA Guidance, Navigation, and Control Conference, 29-31 July 1996, San Diego, CA, USA*; 1996. p. 3856.
- [11] Kopardekar P. Dynamic density: A review of proposed variables. FAA WJHTC Internal Document Overall Conclusions and Recommendations. Federal Aviation Administration, 2000.
- [12] Kopardekar P, Magyarits S. Dynamic density: measuring and predicting sector complexity. *Proceedings of the 21st Digital Avionics Systems Conference, 27-31 Oct. 2002, Irvine, CA, USA*, vol. 1. IEEE; 2002. p. 2C4-2C4.
- [13] Kopardekar P, Magyarits S. Measurement and prediction of dynamic density. *Proceedings of the 5th USA/Europe Air Traffic Management R & D Seminar*, vol. 139; 2003.
- [14] Radišić T, Novak D, Juričić B. Reduction of Air Traffic Complexity Using Trajectory-Based Operations and Validation of Novel Complexity Indicators. *IEEE Transactions on Intelligent Transportation Systems*. 2017;18: 3038-3048.
- [15] Radišić T, Novak D, Juričić B. Development and validation of an ATC research simulator. *Proceedings of the INAIR 2015 International Conference on Air Transport, 12-13 Nov 2015, Amsterdam, The Netherlands*; 2016.
- [16] Kopardekar PH, Schwartz A, Magyarits S, Rhodes J. Airspace complexity measurement: An air traffic control simulation analysis. *International Journal of Industrial Engineering: Theory, Applications and Practice*. 2009;16: 61-70.
- [17] Gianazza D, Guittet K. Evaluation of air traffic complexity metrics using neural networks and sector status. *Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAT 2006, 24-28 June 2006, Belgrade, Serbia and Montenegro*; 2006. p. 113-122.
- [18] Cybenko G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*. 1989;2: 303-314.
- [19] Levenberg K. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*. 1944;2: 164-168.