SCIENCE & MILITARY

VEDA A VOJENSTVO

No 2 | Volume 4 | 2009

Dear readers,

We are pleased to present newly updated and refurbished form of the Science and Military magazine issue. The body of editors comprises of the international top experts in both theoretical and practical fields of science and technology, who contribute to the magazine by their valuable advice and recommendations. Science and Military is published in printed as well as electronic version in form of abstracts, which can be found at http://www.aos.sk/casopisy/science/index.html; the official website of the Armed Forces Academy of General Milan Rastislav Štefánik. There you can also get in touch with the editor's office, get the information on possibilities to publish in the magazine. Broad awareness and accessibility positively influence scientific level and communicativeness at the workplaces and among employees in the field of science and technology. Science and Military is registered in ProQuest Central database and is published in full-text electronic version at the web site. Readers can thus see the latest issue or look for necessary information on older issues, where original scientific articles are published.

It is one of the basic measures for the level of the scientific work nowadays to quote the author of the work. The quality of scientific work outputs, interest in the work and use of information gained from the work may be inferred from the number of quotations. This information can be found for example in Web of Science database, which contains data records of publicly published periodicals reference. Our intention is to attain a permanent position within the database.

Another significant predicative marker in quality of particular work is impact factor, which is a certain measure of quality of scientific magazines.

Science and Military has ambitions to rank among the current magazines within the Current Contents system in the near future by featuring high-quality articles in English language and by maintaining high impact factor, which is the indicator of the scientific activities of the author of the article or even the whole work place. Our aim is to make sure the Science and Military serves the purpose of publishing original scientific articles on humanities, management, and science, technical, military and security-related topics.

We have decided to pay more attention to presenting practical experience, outputs and results of scientific research, its interconnection with theory and practice and exchange of experience among experts, educators, scientific workers, PhD. students and students.

Therefore we would like to hold a fruitful discussion between the editor's office and our readers. Let us hear from you (by mail or e-mail) – we appreciate knowing what our readers find worthwhile and interesting. Your prompts, comments, critique as well as articles will help us fulfil the purpose of our magazine.

Col. (GS) Assoc. Prof. Dipl. Eng. Pavel NEČAS, PhD.

Editor–in-Chief

Reviewers

Prof. Dipl. Eng. Vladimír ***BELLA****, CSc.*	*Armed Forces Academy Liptovský Mikuláš (SVK)*
Assoc. Prof. Dipl. Eng. ***JANOŠEC****, CSc.*	*Population Protection Institute Lázně Bohdaneč (CZE)*
Prof. Dipl. Eng. Tobiáš ***LAZAR****, DrSc.*	*Technical University of Košice (SVK)*
Col.(GS) Assoc. Prof. Dipl. Eng. Pavel ***NEČAS****, PhD.*	*Armed Forces Academy Liptovský Mikuláš (SVK)*
Prof. Dipl. Eng. Josef ***REITŠPÍS****, PhD.*	*University of Žilina (SVK)*
Prof. Dipl. Eng. Vladimír ***SEDLÁK****, PhD.*	*The University of Security Management in Košice (SVK)*
Assoc. Prof. Dipl. Eng. Milan ***SOPATA****, CSc.*	*Honeywell International s.r.o. Brno (CZE)*
Assoc. Prof. Dipl. Eng. Vladimír ***VRÁB,*** *CSc.*	*University of Defence Brno (CZE)*
Assoc. Prof. Dipl. Eng. Peter ***TEREM****, PhD.*	*The Mathias Bel University Banská Bystrica (SVK)*

THE DYNAMIC ERROR OF THERMOCOUPLES

František ADAMČÍK, Peter KRAJŇÁK

Abstract: The contribution is focused on solving the problem of the dynamic error at the thermocouples, used to measure temperature of gases aft of the turbine. The temperature of gases aft of the turbine is a parameter limited by the material used for manufacturing the turbine blades of aviation jet-engine. The time behaviour of the temperature the gases behind the turbine is also an important indicator in terms of the combustion chamber diagnostics. The dynamic error of thermocouples makes it impossible to use the temperature values of the gases aft of the turbine, recorded in flight onto an airborne engine diagnostic device, in various modes of engine operation, consequently, solving the problem of the dynamic errors of thermocouples is of primary importance.

Keywords: Jet-pipe temperature, parameter check, aircraft engine.

1 INTRODUCTION

In an aircraft engine, the temperature of the gases aft of the turbine t_4^* is measured by way of hermocouples equally distributed in a circle behind the turbine (Figure 1). All over the engine there are twelve bodies, each containing two independent thermocouples – for both the first and the second channel of measurement. For the electromotive force (EMF) of the thermocouple it holds that:

$$e = f\left(t_H\right) - f\left(t_{CH}\right) \tag{1}$$

there $f(t_H)$ and $f(t_{CH})$ – contact EMF in view of the hot and cold contacts of the thermocouples. To obtain an integral (medium) value of the temperatures for the gases aft of the turbine, all thermocouples are connected paralelly. As the „cold" contacts of the thermocouples are placed off the engine, their temperatures are substantially lower than those of the measured gases. The electro-motive force of the T-99 thermocouple is the linear function of the difference between the "hot" and "cold" contact and can be described in the form of:

$$e = k\left(t_H - t_{CH}\right) \tag{2}$$

where k = *constant*. The thermal electro-motive force of the thermocouples from the first channel of measurement is fed into the block of limiting regulators (BLR) and from the second channel is passing through two parallel circuits – on the indicator of jet-pipe temperature to the pilot cockpit and the other to the block of limiting commands.

2 DEFINING THE TASK

The most critical error of the thermocouple consists in the dynamic error in connection with the process of transferring the temperature of gas flow to the sensitive element of the thermocouples- to the hot contact. In general, it is characterized by a time constant τ, the magnitude of which is given by expression:

$$\tau = \frac{cm}{\alpha_Q S}, \tag{3}$$

where: c- specific thermal capacity of the material of the "hot" thermocouple contact, m – mass of the sensitive element of the "hot" thermocouple contact, S – area of the surface of the sensitive element subjected to the stream of hot gases, α_0 – coefficient of the temperature transfer of the "hot" thermocouple contact material.

The circuit of measuring the gas temperature aft of the turbine t_4^* represents, in view of the regulation, an aperiodical inertial element, which can be described in the operator's form by an equation:

$$t^*_{4\ mer} = \frac{K}{\tau P + 1} t^*_{4S}, \tag{4}$$

where $t^*_{4\,mer}$ – measured value of the temperature (on the thermocouples´ input) t_4^*gases aft of the turbine, t^*_{4S} – true value of the gases aft of the turbine t_4^*, coefficient of transfer, normal image K=1, p – Laplace operator.

From the expression (4) it follows that the process of measuring the temperature of gases aft of the turbine t_4^* can be described by a differential equation taking the form of:

$$\frac{d\, t^*_{4\ mer}}{d\, t} = K\ \omega\ t^*_{4S} - \omega\ t^*_{4\ mer}, \tag{5}$$

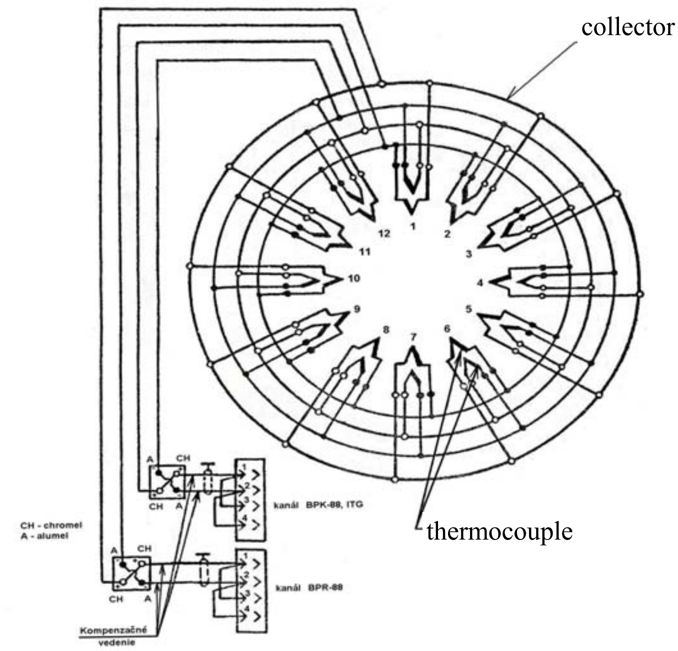


Figure 1 Connection plan of thermocouples

where $\omega = \frac{1}{\tau}$ the frequency proper to the process of delay. The magnitude of the time constant of the thermocouples used falls within the interval from 6 to 10 seconds in scope. Disregarding the dynamic error of measurement at the thermocouples when using them at the limiting channel for the temperature t_4^* of the gases aft of the turbine performed in the block of BLR could lead to overheating the engine turbine and compressor blades. In order to compensate for the effect of the dynamic delay in the process of measuring the temperature t_4^* of the gases af of the turbine, in the block of BMR there is a correction circuit – derivation element (Figure 2).

The value of the correctional signal Δt_4*_{mer} from the correctional element output can be expressed in the form of:

$$\Delta t^*_{4\,mer} = t^*_{4\,S} \frac{K_{kor} P}{\tau_{kor} P + 1} \qquad (6)$$

where K_{kor} - correctional circuit amplification coefficient, τ_{kor} – correctional circuit time constant.

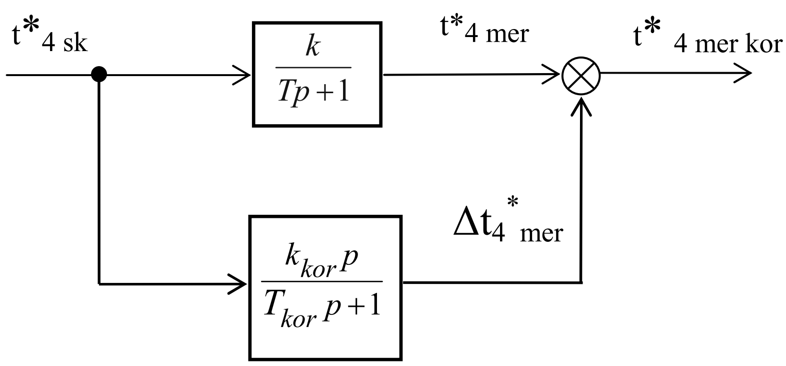


Figure 2 Correction circuit – derivation element

The equation (6) can be written in Cauchi´s form:

$$\frac{d\,\Delta t^*_{4\ mer}}{d\,t} = K_{kor}\ \omega_{kor}\ \frac{d\,t^*_{4\ S}}{d\,t} - \omega_{kor}\ \Delta t^*_{4\ mer}\ , \tag{7}$$

where $\omega_{kor} = \frac{1}{\tau_{kor}}$ the frequency proper to the correction process.

In practice (within the BLR block) the coefficient K_{kor} and the time constant τ_{kor} of the correction circuit is determined so that the resulting signal $t_4{}^*{}_{kor\ mer} = t_4{}^*{}_{mer} + \Delta t_4{}^*{}_{mer}$ should be closest to the true value of $t_4{}^*{}_S$ of the temperature of gases aft of the turbine:

$$t_4{}^*{}_S \approx t_4{}^*{}_{kor\ mer} = t_4{}^*{}_{mer} + \Delta t_4{}^*{}_{mer}. \tag{8}$$

The solution of the approximate equality is an engineering task, solved in the 50s of the last century in the time of mass introduction of jet engines. At the present time, when using systems of objective control, the topical tasts is the one of evaluating the true temperatures of gases $t_4{}^*{}_S$ aft of the turbine based on the information measured and recorded by a Tester system. The problem consists in the fact that, inside ethte BMR block, into which the thermo-electro-motoric voltage is fed from the thermocouples and are converted into DC voltage, and also in the electronic block there are no correctional circuits for measuring the temperature $t_4{}^*{}_{mer}$. Temperature $t_4{}^*{}_{mer}$ recorded by the Tester system therefore shows a dynamic error. The true value of the temperature $t_4{}^*{}_S$ is very important when inspecting dynamic processes:

- starting up the engine, when the blades of both turbines are cool and are subjected to the greatest thermal loading,
- intensive accelerations idle-run – maximal, idle-run – afterburner, at which a substantial thermal loading is also present.

3 POSSIBLE METHODS OF SOLVING THE TASK

Proceeding from the differential equation (5) one could obtain the true value of the temperature of gases aft of the turbine by way of solving the equation:

$$t^*_{4\,S} = \frac{1}{K}\left(t^*_{4\ mer} - \tau\,\frac{d\,t^*_{4\ mer}}{d\,t}\right). \tag{9}$$

The complexity of solving the given equation consists in the fact that the system lacks registration of derivation $\frac{d\,t^*_{4\ mer}}{d\,t}$ by the time of the measured temperature of gases $t_4{}^*{}_{mer}$ aft of the turbine. Therefore, first of all, it is inevitable to determine the value of this derivation in the moment of registrating the temperature of gases $t_4{}^*{}_{mer}$ aft of the turbine. In its general for, the equation of monitoring (recorded measurements) can be written in the vector form

$$\vec{z}_i = \vec{h}\left(\vec{x}_i, t_i\right) + \vec{\eta}_i, \tag{10}$$

where $\vec{z}_i$ – represents the vector with dimension of $\left(m \times 1\right)$ monitored in time t_i, $\vec{h}\left(\vec{x}_i, t_i\right)$ – in a general case a nonlinear smooth (can be derivated by vector $\vec{x}_i$ and time t_i) vector function, $\vec{x}_i$ – vector with dimension $\left(n \times 1\right)$ state of the object in time t_i, $\vec{\eta}_i$ – vector of noises of measurements, generally representing the stationary Gauss process of the white noise.

In practical exercises, more exactly when determining the values of partial derivation $\frac{d\,t^*_{4\ mer}}{d\,t}$, one can always determine the limited area of Γ state area, into which, with high probability (closing to 1), belongs the vector $\vec{x}_i \in \Gamma$. The area Γ lies in the sub-region $\Gamma_v \subset \Gamma$, in which, it is possible to approximate each of the l-th component of the vector function $\vec{h}$ using the polynom of second level:

$$\vec{h}_l(\vec{x}_i,t_i) \approx a_l^{\langle\nu\rangle} + \sum_{j=1}^{n} b_{lj}^{\langle\nu\rangle} x_{ij} + \sum_{j,k=1}^{n} b_{ljk}^{\langle\nu\rangle} x_{ij} x_{ik} + c_l^{\langle\nu\rangle} t_i + d_l^{\langle\nu\rangle} t_i^2, \quad \vec{x}_i \in \Gamma_\nu, \tag{11}$$

Expression (11) in the form of a matrix:

$$\vec{h}(\vec{x}_i,t_i) \approx \vec{A}^{\langle\nu\rangle} + B^{\langle\nu\rangle}\vec{x}_i + \vec{x}_i^T B^{\langle\nu\rangle}\vec{x}_i + \vec{C}^{\langle\nu\rangle} t_i + \vec{D}^{\langle\nu\rangle} t_i^2 \tag{12}$$

where $\vec{A}^{\langle\nu\rangle}$, $\vec{C}^{\langle\nu\rangle}$, $\vec{D}^{\langle\nu\rangle}$ - vectors (column matrix with dimension $(m\times 1)$), $B^{\langle\nu\rangle}$ - square matrix $(m\times n)$:

$$\vec{A}^{\langle\nu\rangle} = \begin{bmatrix} a_1^{\langle\nu\rangle} \\ \bullet \\ \bullet \\ \bullet \\ a_m^{\langle\nu\rangle} \end{bmatrix}, \quad \vec{C}^{\langle\nu\rangle} = \begin{bmatrix} c_1^{\langle\nu\rangle} \\ \bullet \\ \bullet \\ \bullet \\ c_m^{\langle\nu\rangle} \end{bmatrix}, \quad \vec{D}^{\langle\nu\rangle} = \begin{bmatrix} d_1^{\langle\nu\rangle} \\ \bullet \\ \bullet \\ \bullet \\ d_m^{\langle\nu\rangle} \end{bmatrix}, \quad B^{\langle\nu\rangle} = \begin{bmatrix} b_{11}^{\langle\nu\rangle} & \bullet & \bullet & \bullet & b_{1n}^{\langle\nu\rangle} \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ b_{m1}^{\langle\nu\rangle} & \bullet & \bullet & \bullet & b_{mn}^{\langle\nu\rangle} \end{bmatrix}. \tag{13}$$

Coefficients of the matrix (13) can be determined on the basis of experimental measurements, for example by way of smallest squares. Solution of the equation (12) in scalar form:

$$t_{4mer}^{*}(t_i) = h(t_i) \approx a^{\langle\nu\rangle} + c^{\langle\nu\rangle} t_i + d^{\langle\nu\rangle} t_i^2. \tag{14}$$

Then, the searched value of derivation to the first estimate of $\frac{dt_{4\,mer}^{*}}{dt}$ in the time of measuring the temperature of gases aft of the turbine t_4^* in the moment of t_i, can be expressed in the form of:

$$\frac{dt_{4\,mer}^{*}}{dt_i} \approx c^{\langle\nu\rangle} + 2d^{\langle\nu\rangle} t_i. \tag{15}$$

The Algorithm of calculating the ture temperature of the gases aft of the turbine t_4^* must be cyclical throughout the time interval of the controlled process. Then the area of Γ will be of rectangular from with dimensions on one axis from $t_{zač\_zap}$ to t_{kon_zap}, and on the other one from 0ºC to 1100ºC. As a sub-area of Γ_ν, we will look for rectangular forms falling within the area of Γ with the length of sides of 2 seconds in one axis and length from 0ºC to 1100ºC along the second axis. Thus, the condition of $\Gamma_\nu \subset \Gamma$ will be fulfilled. Proceeding form the cyclo-gram of the airborne recording system, the Tester, each of such sub-.areas will contain 9 measurements of the temperature of gases aft of the turbine t_4^*. In the given area, it is inevitable to determine the coefficients of $a^{\langle\nu\rangle}$, $c^{\langle\nu\rangle}$ an $d^{\langle\nu\rangle}$ for the expression of (14). To solve this task, considering the possible anomaly measurements (errors) in temperatures, a robust method of non-linear regessional analysis can be used.

The next step to perform in each cycle is the calculation of $\frac{dt_{4\,mer}^{*}}{dt}$ in the time of measuring the temperature of gases aft of the turbine t_4^* in time t_i according to the expression (15). Based on the elementary laws of regressional analysis, in order to obtain the most accurate value, it is inevitable to select for a point t_i on the time axis, corresponding to the centre of the sub-area Γ_ν.

The last step to perform is the calculation of the true temperature of the gases aft of the turbine t_{4S}^{*} using expression (9). In the further process, the next sub-area Γ_ν is selected by shifting along the time axis onto a value corresponding to the time interval until the next measurement of the temperature of the gases aft of the turbine t_4^* by the cyclo-gram of the airborne recording system.

4 METHOD OF MODELING THE INSPECTED MODE OF OPERATION OF AN TURBOJET ENGINE

The parametric monitoring of the operation of a turbo-jet aviation engine in dynamic modes of operation, such as: starting up the engine, acceleration, deceleration, makes it inevitable to employ mathematical models providing adequate description of the processis of changes in the parameters in the modes mentioned. It is about the gas-dynamic models developed by a succession of general differential equations (integro-diferencial ones) and those with partial derivations, the identification of characteristics by the values recorded by airborne, Tester-type devices, is impossible owing to the level of the monitorability of the processes. In practice, the monitored transitional processes at engines and their nodes and aggregates are described under the condition of the dynamic model coefficient remaining unchanged, in the extremal case, in the course of the monitored process. Then, the general model of the inspected mode of operation for the engine can be expressed in the form of:

$$\begin{cases} \dot{\vec{y}} = \vec{f}(\vec{y}, \vec{a}, \vec{u}, t) + \vec{w}(t) \\ \dot{\vec{a}} = 0 \end{cases} \tag{16}$$

where $\vec{y}$ - r-dimensional vector of output parameters of the engine; $\vec{a}$ - q-dimensional vector of parameters (coefficients); $\vec{u}$ - m-dimensional vector of input signals (equation); $\vec{w}(t)$ - normal random process of white- noise type with zero mean value and matrix ant the matrix of intensity $S_w(t)$; $\vec{f}(\vec{y}, \vec{a}, \vec{u}, t)$ - vector function that can be derivated continuously by arguments $\vec{y}, \vec{a}, \vec{u}$. Vectors $\vec{y}$ and $\vec{a}$ denote the n-dimensional vector of status $\vec{X}$. Coefficients of the model are considered constant in the course of the transitional process, which is expressed by the second vector differential equation $\dot{\vec{a}} = 0$. By analogy, by (2) measuring the parameters of the objects corresponds to the equation of monitoring:

$$\vec{z}_k = \vec{h}\left[\vec{y}(t_k), \vec{a}, \vec{u}(t_k), t_k\right] + \vec{v}_k, \tag{17}$$

where: $\vec{h}(\cdot)$ - r-dimensional by parameter $\vec{y}, \vec{a}, \vec{u}$, vector function to be derivated continuously; $\{\vec{v}_k\}$ - independent succession of r-dimensional normal vectors centralized with the correlation matrix R_k. Appart from this, monitoring is also extended to the vector of control $\vec{u}$. Its equation of monitoring can be written in the form of:

$$\vec{U}_k^* = \vec{U}(t_k) + \vec{\xi}_k, \tag{18}$$

where $\{\vec{\xi}_k\}$ - independent succession of normal errors of measurements with zero centre and correlation matrix of U_k. When describing the dynamic model of the process of starting up, acceleration and deceleration of the engine, the vector $\vec{y}$ of the engine parameter from the expression of (16) will contain: engine compressor rpm n_2 (high pressure compressor rotor), fan rotor n_1 (low pressure compressor of rotor), air pressure aft of the compressor P_K, fuel pressure delivered into the combustion chamber P_T, measured temperature of the gases aft of the turbine $t_{4\,mer}^*$. Then the equation (5) will comprise the vector of the parameters of the object $\vec{y}$ and, gradually, it will be possible to determine the value of $\frac{dt_{4\,mer}^*}{dt}$ in the moment of measuring the temperature t_4^* of gases aft of the turbine. Using the expressions (9) will then enable calculation of the true value of the temperatures of gases aft of the turbine t_{4S}^*.

5 CONCLUSION

The temperature of the gases aft of the turbine is a parameter the value of which is limited by the material used for manufacturing of the jet-engine rotor blades. The time behaviour of the temperature of gases aft of the turbine is also an important diagnostic indicator of the combustion chamber. The dynamic error of thermocouples makes it impossible to make use the values of gas temperatures aft of the turbines recorded in flight, with the recorder of engine diagnostics, in the dynamic modes and, therefore it is necessary to look for a solution of the problem of the dynamic error of thermocouples. The solution of the problem of processing the recording of the jet-pipe temperatures enables application of use of mathematical models of the engine operation in transitional modes. Extending the parametric control in view of the jet-pipe temperature is of great importance in assessing the status of the thermally loaded parts of the engine in order to identify its error or pre-error states. This extension should lead not only to more efficient diagnostics, but the output of the parametric control could become a tool assisting in determining the limits of the engine operation.

References

[1] ПОЛОЗОВ, А. А.: *Актуальные задачи параметричекого контроля сотояния сложных систем и методы их решения*, TOPAZ, Москва, 2006.

[2] ПОЛОЗОВ, А. А.: *Об одном подходе к решению задачи эффективного иагостического контроля сотояния жизненно-важных систем летательных аппаратов по информации бортовых устройств регистрации*, НИЭРАТ, Москва, 2005.

Assoc. Prof. Dipl. Eng. František ADAMČÍK, PhD.
Faculty of Aeronautics
Technical University in Košice
Rampová 7
041 21 Košice
Slovak Republic
E-mail: frantisek.adamcik@tuke.sk

Dipl. Eng. Peter KRAJŇÁK
PhD. student
Faculty of Aeronautics
Technical University in Košice
Rampová 7
041 21 Košice
Slovak Republic
E-mail: krajnak.peter.sl@gmail.com

DISCRETE MODELS OF OPTIMIZATION AND SAFENESS IN TRANSPORT AND TRANSFER

Štefan BEREŽNÝ

Abstract: Nowadays in most of the industries evaluation and ensuring various forms of security comes only out of statistical data processing and evaluating, eventually from specifications that are often designed on the statistical basis. In this paper I want to present view on the way, how to use possibilities of discrete optimization in this industry. With these information are used, that are collected from statistical data processing. I attempted to illustrate some findings from discrete optimization in various industries of transport and transfer.

Keywords: Mathematical model, optimization, statistical data processing, graph, objective function.

1 INTRODUCTION

This article gives some examples of the use of discrete optimization in traffic safety and transportation. I show except traditionally known models also the less standard models, which are based on previously mentioned known optimization problems. I show the discrete model, whose structure is based on statistical data processing. The information was processed for the Košice airport. The results are also in improving the safety of operations at the airport using discrete models. Inputs for discrete models can be a data, which are obtained by suitable statistics methods; interesting methods are described in [3]. In conclusion, mentioned some other directions in the use of discrete models of traffic.

2 DISCRETE MODELS

We know many discrete models. This article will focus on such models, which can be a way to represent the structure and using the solutions, which are subsets of sets, which form a graph. In the case of optimization problems I use the objective functions, which are defined over the values that are assigned to edges or vertices. At the outset, I describe the general model, which will be applied in various forms to specific examples of usage in the transport and transport safety.

Let the graph $G = (V, E)$ where V is the set of vertices of a graph G and E is a set of edges of graph G. Let $\boldsymbol{B}(G)$ is a system of feasible sets of graph G. Feasible set $D \in \boldsymbol{B}(G)$ ($D \subseteq E$) may be, for example, spanning tree of graph G, a perfect matching of graph G, a Hamiltonian circuit, etc . When this input is the most basic task to determine system sets $\boldsymbol{B}$ *(G)* is the empty set or not. A simple example of such tasks is whether there is between two cities the railway, bus, air or other link.

Modify the specified assumptions about the weighting function $w\colon E \to \mathbf{R}_0^+$. This function assigns each edge $e \in E$ nonnegative real number $w(e)$. Following the introduction of such functions makes sense considering the optimization problems that maximize respectively minimize the value of defined objective function. Objective function is a mapping $f\colon \boldsymbol{B}(G) \to \mathbf{R}_0^+$. The ordered pair ($\boldsymbol{B}$, f) consists of an optimization problem, which can be expressed in the form:

$$f(D) \xrightarrow[D \in \mathbf{D}(G)]{} \min \tag{1}$$

Among the best-known and most popular features includes the following objective functions:

$$SM(D) = \sum_{e \in D} w(e) \tag{2}$$

$$BT(D) = \max_{e \in D} w(e) \tag{3}$$

$$BL(D) = \max_{e \in D} w(e) - \min_{e \in D} w(e) \tag{4}$$

The problem with the objective function (2) we call the sum problem, with the objective function (3) as bottleneck problem and the objective function (4) as balanced problem. Many other problems have to maximize the form or the form of the recognition version and so on. This article deals only with optimization problems, whose objective function is minimized, i.e. we search the feasible set $D \in \boldsymbol{B}(G)$, for which the value of objective function $f(D)$ is minimal.

The algorithms that solve the problem with sum objective function often encountered in navigation systems, on web pages that look for variously types of connections (rail, bus, airplane, etc.).

Bottleneck problem must be resolved in cargo traffic, especially if the transports include heavy cargo or dangerous load. Vehicles with a high total weight cannot move in arbitrary ways and the planned route is limited by the maximum load capacity of bridges along the planned route. Similarly, the situation is with the oversized cargo. If the transports include excessive cargo, it cannot go under all the bridges, through tunnels or some areas and railway crossings with low-lying trolleys. When the routes are planning, we solve bottleneck problem. Balanced problem can be used in the case of planning the routes of security agencies, to that the transit was the safest. Formulation and solution of such problems is described in article [9].

How can we these or similar discrete models used directly or indirectly in providing security services, management and organization of transport (road, rail and air)?

The transport of dangerous load is the relatively large risk of road transport. In the transport of this kind of cargo, it is important to carefully plan the route and know the risks that on that route may occur.

Let graph G represents the current road network and let the weighting values assigned playing 0 and 1 as follows. If after a given road section, which is represented by edge e we can carry dangerous cargo, the value of w(e) = 0. If, on the edge it is not possible to carry dangerous cargo, the value of w(e) = 1. Thus we get valued graph G. Let a and b are vertices of a graph G representing the places between which it is necessary to carry dangerous cargo. For this problem we can use any objective function (2), (3) or (4). The problem is to find a path P in graph G between the vertices a and b, the value of objective function SM(P) = 0 (respectively BT(P) = 0 or BL(P) = 0). If we do not find such path P then it is not possible transported the dangerous cargo between cities, a - b. Moreover, if we find that such path P does not exist then the value of objective function SM(P) gives us the minimum number of road sections which do not meet the standards for the transport of such dangerous cargo. Similarly, we can simulate various restrictions and test whether there is a route that satisfies the restriction (e.g. only highway connection etc.).

We cannot forget to rescue and security forces. They must get to the destination in the short time. The important is the length of the route, but we cannot forget the risk of the route, which is provided car must pass, even though it has turned on the warning sound and light equipment for the vehicle.

The ideal is to avoid the sections with high accident rate or junctions and crossroads where traffic density is high and there is greater probability of being trapped in a jam. In such cases, it is most important which route is selected for the vehicle by dispatcher. It should have a device that would solve the problem. Let graph G represents a road network. Let the value at the edges represent the degree of congestion on the road or field, respectively the likelihood of an accident on that road section. Then using the objective functions (2) and (3) we can solve problem: we have found such *a* - *b* path that minimizes the objective function of *SM* or *BT*, so in the first case, we find a way where the sum of degrees of density of traffic will be as low as possible and thus assumes the fastest passing. In the latter case, we minimize the cross section with the largest probability of an accident to minimize risk that the car crashes in a traffic accident in carrying out the intervention.

All of the above problems and built models came from the most dedicated functions (2), (3) and (4). In solving many other problems we can also use dedicated functions, whose shape is:

$$CF_1(D) = \max_{1\le i\le p} \sum_{e\in S_i\cap D} w(e) \tag{5}$$

$$CF_2(D) = \sum_{1\le i\le p} \left[\max_{e\in S_i\cap D} \left(w(e) \right) \right] \tag{6}$$

$$CF_3(D) = \max_{1\le i\le p} \sum_{e\in S_i\cap D} w(e) - \min_{1\le i\le p} \sum_{e\in S_i\cap D} w(e) \tag{7}$$

$$CF_4(D) = \sum_{i=1}^{p} \left(\max_{e\in S_i\cap D} w(e) - \min_{e\in S_i\cap D} w(e) \right) \tag{8}$$

These functions are only one of many that have been defined and described for example in works [2, 3, 4, 5, 6, 7, 8, 10, 11]. In the above works are described the 8 dedicated functions whose value depends not only on the weights of edges, but also that to which category was added the edge. These articles describe these problems, their complexity and possibly algorithms that address some of the problems with the categorization defined by the edges of the graph *G*. Also, there are some examples of such models in practice. We describe above problems in general.

Suppose a graph *G* = (*V*, *E*) with edge weights *w*(*e*) for $e \in E$ is given. An optimisation problem ***P*** on a graph *G* is given by a family of feasible sets ***B***(*G*) and an objective function *f*: ***B***(*G*) → R_0^+. The ordered pair (***B***, *f*) determines a particular optimization problem:

$$f(D) \xrightarrow[D\in \mathbf{D}(G)]{} \min .$$

Let us now describe a special optimization problem with categorization. Suppose that edges of the graph are partitioned into disjoint categories $S_1, \dots, S_p$. We shall consider the following objective function (5), where $\max_{e\in S_i\cap\varnothing} = 0$.

I show only one simple example of the use of such objective functions. In this time, it is often tested the ability to perform the control of various buildings and institutions in the shortest possible time. If we split the various control points on individual departments, which we have control. Let the graph represents the town network to be controlled (or security) and the vertices are the control points. Let weights of edges represent the times that are necessary for the implementation of controls to move between two checkpoints. Let the edges, which are controlled by a department belonging to the same category. The problem is to find a scheme of transfer of control that the total time will be minimal. When you create a discrete model it will correspond to minimizing the objective function (5) in the appropriately defined graph *G*.

3 THE EXAMPLE OF DESCRETE MODEL WITH USING THE STATISTICS

In practice in creating models we often meet with the problem how to determine the number and value of constants, which are stated in the model. Many values are the technical specifications and workers with long-standing practice can easily estimate other values. However, many times we need the value that nobody knows or their estimates can cause the problems. Therefore often in construction, while accurate discrete models we need to use values that can only be obtained by statistical processing of existing data. Quite the current processing of data on flight delays in the Kosice airport can be found in the work [1]. These are the statistics, that airport has not had yet and had an interest in the processing of such data. I do not have the usage of these statistics by the airport, but in this chapter I describe, how the data can be used in the operation of the airport.

Currently, every international airport must ensure the safe operation of the airport, but also provide sufficient security of buildings and the airport area, which it is composed. Physical security is often linked with arrivals and departures of aircraft. Such trip can be planned on the basis of the flight plan. The problem arises in the way that this plan is not often respected and there is a relatively large amount of time deviations from the scheduled arrivals and departures. On this basis, it is not possible to make many checks under the plan and must be tailored to the real situation. Therefore it is possible to create a model that assigns persons to exercise control with respect to the probable deviation of the flight. Such revised plan will ensure more efficient use of security personnel at the airport and ensure the stitch level of its premises and grounds.

Let graph *G* represents a bipartite graph, where a set of vertices corresponding to workers of security service and the second set of vertices corresponding to the checks to be made. Edge worker states where the checks can take place while the entire edges dame in the same category, if the check is run in parallel. Let the value at the edges correspond to average deviations of estimated time for the flight. Then we are trying to find the assignment to take control, to the sum of the differences in the various categories to a minimum. This will ensure that the inspections are completed as possible with the smallest time deviations between parallel controls. To solve the problem we can use the objective function (8).

Based on this model we can also consider other alternatives that can be used not only temporal variations, but also other statistical values that can be obtained in the processing of data, which were be provided by airport. Scope and diversity of data is described in the work [1, 12].

4 WHAT IS NEXT?

The main question includes, what are other possibilities for using such models to ensure traffic safety and transportation. Very often we require optimization by more than one dedicated function. Such problems have not been very detailed; especially by using the functions (5) - (8), but you can find abstracts of the conference where these results were presented. Hint of such solution of one problem can be found in the work [9].

Very often the requirements appear to value in the optimal solution are not very far apart, but also requires that the maximum value contained in the agreed solutions was kept to a minimum. It is obvious that the optimization problem in which the search for such feasible solution *D* that minimizes simultaneously two different objective functions. In such cases, it may take optimum exists, but more frequently it may be that such an optimum exists. However, it may be admissible solution that would be optimum with respect to at least one objective function. Such a solution may not be the best for practical use and there is a problem, which of the possible solutions to declare admissible the optimum of such a problem with multi-objective functions.

For some problems it is possible to define so-called additional criterion, we dedicated two features create one and then looking optimum is only given to the newly created objective function. One of these criteria is called the distance from the optimum. Then the optimum solution may be regarded as acceptable solution, which the distance from optimums is minimal.

5 CONCLUSION

In this article I suggested the possibilities of using various discrete models to ensure the safety of traffic and transportation, whether directly or indirectly. I show a basic model without the dedicated and best–known function optimization problems, which are used in practice in many industries of transport and not only there. I suggested the use of objective functions, which are the basis for the so-called problems with the categorization of edges. These modified-objective functions increase the chances of resolving the many problems of real life.

I also pointed to the need to interlink the knowledge of individual disciplines and a simple example shows how to reconcile mathematical statistics, and discrete optimization. This example is only one of the few examples of how this can be. Usage is considerably wider. In conclusion, I have also pointed to the possibility of construction multi-criteria models whose solution may have been more difficult, not even an efficient algorithm, which we find appropriate solutions to such multi-criteria problems.

References

[1] BEREŽNÝ, Š., ANDREJKOVIČ, M.: *Statistical processing of flights on Košice Airport.* Acta Avionica, Košice : TU-LF (submited). ISSN 1335-9479.
[2] AVERBAKH, I., BERMAN, O.: *Categorized bottleneck-minisum path problems on networks*, Operation Research Letters 16 (1994) 291-297.
[3] SOTÁK, M.: *Determining stochastic parameters using an unified method.* In: Acta Electrotechnica et Informatica. Vol. 9, No. 2 (2009), p. 59-63. ISSN 1335-8243.
[4] BEREŽNÝ, Š., CECHLÁROVÁ K. and LACKO, V.: *Optimization problems on graphs with categorization of edges.* In: Proc. SOR 2001, eds. V. Rupnik, L. Zadnik-stirn, S. Drobne (Predvor, Slovenia, 2001) 171-176.
[5] BEREŽNÝ, Š., CECHLÁROVÁ, K. and LACKO, V.: *A polynomially solvable case of optimization problems on graphs with categorization of edges*, in: Proc. of MME′99 (Jinřichúv Hradec, 1999) 25-31.
[6] BEREŽNÝ, Š., LACKO, V.: *Balanced problems on graphs with categorization of edges*, Discussiones Mathematicae Graph Theory 23 (2003) 5-21.
[7] BEREŽNÝ, Š., LACKO, V.: *Special problems on graphs with categorization*, in: Proc. of MME′2000. Praha : 2000. 7-13.
[8] BEREŽNÝ, Š., LACKO, V.: *Easy (polynomial) problems on graphs with categorization*, in: Proc. of New trends of aviation development (Air Force Academy of General M. R. Štefánik, Košice, 2000) 36-46.
[9] BEREŽNÝ, Š., LACKO, V.: *The Color-balanced Spanning Tree Problems*, Kybernetika vol. 41 (2005) 539 –546.
[10] BEREŽNÝ, Š.: *Further options of transport*. In: MOSATT 2007: Modern Safety Technologies in Trasportation: Proceedings of the International Scientific Conference: 25th - 27th September 2007, Zlata Idka. Košice : Robert Breda, 2007. s. 19-21. ISBN 978-80-969760-2-7.
[11] BEREŽNÝ, Š.: *Overview of problems on graphs with categorization*. In: 22nd European Conference on Operational Research: Book of Abstracts: Prague, July 8-11, 2007. Praha: VŠE, 2007. p. 35.
[12] RICHEY, M. B., PUNNEN, A. P.: *Minimum weight perfect bipartite matchings and spanning trees under categorizations*, Discrete Appl. Math. 39 (1992) 147-153.

RNDr. Štefan BEREŽNÝ, PhD.
Technical University of Košice
Rampová 7
041 21 Košice
Slovak republic
E-mail: Stefan.Berezny@tuke.sk

SELECTED ASPECTS OF DRIVERS MICROCLIMATE EVALUATION

Štefan ČORŇÁK

Abstract: The traffic safety is a serious issue regarding both the society and general transport. This problem has increasing intensity in terms of increasing density of traffic and motorisation. The transport means number has been growing constantly which has negative impact onto traffic accidents numbers. Looking at the traffic accidents statistics which has been collated for long time period it is remarkable that the main contributors to the traffic accidents are drivers (more than 90%). An environment quality is one of the essential contributors to the driver's reactions apart from the good life style (feeding, corporal activities, sleeping, resting, etc.). The working place/an environment - microclimate in other words is a limited space of the environment. Such space is affected and its state is modulated by contributors represented by energy stream and mass stream between two environments. We can distinguish the internal microclimate according to the contributors - influencing factors such as: thermal, optical, acoustic, air quality of the interior, etc. No other situation in terms of the microclimate influence is at the vehicles´ area. The air quality of the driver's environment is the main contents of the contribution.

Keywords: Traffic safety, driver's environment, driver's microclimate, carbon dioxide.

1 INTRODUCTION

Traffic safety is a crucial problem of nowadays society. A number of means of transport have been increasing constantly which has a negative effect upon the number of road accidents. The development of basic figures regarding a number of vehicles, road accidents and their consequences in 1980 – 2008 is shown in Figure 1. [1], [7].

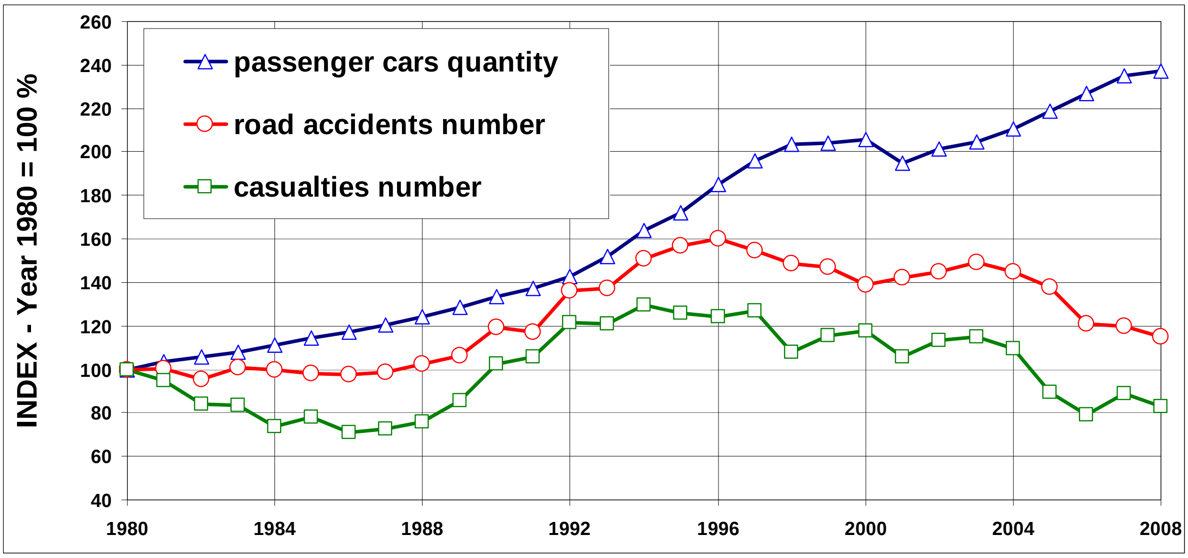


Figure 1 Development of basic figures regarding a number of vehicles and road accidents in the Czech Republic in 1980 - 2008

Main causes of accidents are well known and might be generally divided into the accidents caused by technical and human errors. It turns out that systematic reliability and safety growth of technical parts and vehicles systems as a whole can significantly reduce the purely technical errors of transport means. Up to 50% of all traffic accidents have been caused due to drivers´ attention decrease [2].

All drivers regardless of their age, sex and previous experience might face the attention decrease. It is generally known that a human being (a driver) can focus his attention on a certain number of items in his view range for a limited period of time only. After some time, usually after an hour, the level of attention of most human beings drops. Thus the capability of performing a required activity in a reliable and safe way is significantly limited. The attention decrease of drivers is a common phenomenon and of course each driver is subject to it in a different way. The research of driver attention is therefore observed systematically. Researchers all around the world focused their work on the development of the systems which help to control, support and limit driver actions. A concise overview of the methods and measured parameters is introduced in the Table 1 [3].

Table 1 Methods and selected measured parameters of driver's attention observation systems.

Methods of drivers attention observation	Measured parameter of driver's attention
Technical	-vehicle trajectory (LDW systems)
	- vehicle velocity
	- steering wheel turning
	- reaction time
	- quality of required tasks fulfilment
Psycho-physiological	- eye ball motion and blinking
	- EEG signals analysis
	- face observation
	- muscular activity
	- blood pressure, heart beating frequency
	- skin resistance
	- breathing frequency
	- blood alcohol measurement

It results from the information stated above that car manufacturers pay considerable attention to the driver's attention issue.

A driver's working environment, or its final condition in a vehicle's interior in terms of affecting the human being, has also significant effect onto the driver's attention level. The environment condition in the homosphere (i.e., the place where the human stays) is called a microclimate. It is the microclimate of the vehicle´s interior then. The vehicle environment modification enables the passengers to be transported in nice climate conditions and significantly decreases the driver's tiredness, thereby increasing the total vehicle safety.

The vehicle microclimate is determined by [4]:

- Air temperature;
- Air humidity;
- Air flow velocity;
- Air quality (air change, O_2 content, concentration of CO, CO_2, NO_X, dust, etc.).

The air temperature and air humidity are the main contributors to the microclimate quality. Systems like ventilation, heating and air conditioning help to provide thermal comfort in a vehicle.

Harmful substances elimination produced by passengers, i.e., CO_2 and the air humidity as a result of breathing, CO as a result of smoking, different smells etc., are meant by the air cleanness. Keeping the air cleanness in a vehicle interior is often neglected even in the most modern cars and is limited only to the monitoring of CO which gets into the vehicle from the outside.

It turns out that poor attention is paid to the fact that a lot of drivers spend quite a long time in the vehicle's closed space. Closed (airtight) spaces have many advantages, e.g. they are resistant to the noise incursion from the outside, etc. The main disadvantage is the automatic air change impossibility which causes the air quality worsening.

In the graph in Figure 2 there is an approximate composition of the "pure" outside atmospheric air and in the graph in Figure 3 there is an approximate composition after breathing out by a human [5].

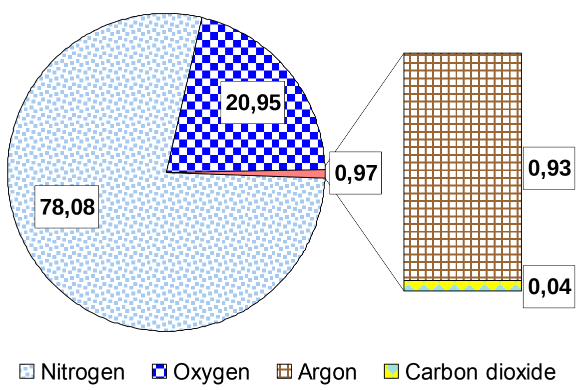


Figure 2 Approximate composition of the "pure" outside atmospheric air

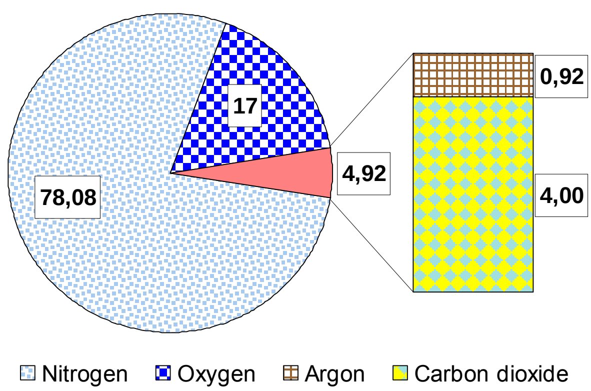


Figure 3 Approximate composition after breathing out by a human

It results from the graphs shown above that oxygen is changed to carbon dioxide due to respiration. The inhaledair contains approximately 0,04% of carbon dioxide (CO_2), which is approx 400 ppm of CO_2, however the exhaled air of an adult contains approx 4% of CO_2 on average, which is approx 40 000 ppm of CO_2 (which is approx 100 times higher concentration than in the surrounding air). When breathing in the open air the 4% concentration of CO_2 is likely to be insignificant. However, a different situation might occur in the closed (almost airtight) space. Driver environment is a typical example where a small space and insufficient ventilation might be the main cause of the increase in CO_2 concentration. A driver's sleepiness, lethargy, tiredness might increase when a CO_2 level is higher. Driver attention decline might be finally the main cause of a traffic accident. The microclimate in a vehicle and measuring of CO_2 in vehicles are supposed to be the main subject of the article.

2 CARBON DIOXIDE CONCENTRATION MEASUREMENT

2.1 Theoretical assumptions of the calculation

A number of approaches are used for measurement of CO_2 concentration in the air. The most frequently used method is using sensors working with infra red emission absorption (so called NDIR – Non Dispersive InfraRed Method), then sensors working on the electro-acoustic principles and sensors working on the electro-chemical principle. Each of the principles mentioned above has its own advantages and disadvantages [6].

The NDIR sensors are the most frequently used and their principle of work is based on the Lambers Beer law [4]:

$$I = I_0 . e^{-\varepsilon . c . l} \qquad (1)$$

where: I_0 - incident emission rate,
I - elapsed emission rate,
ε_0 - molar absorption coefficient,
l - absorption environment thickness,
c - observed substance concentration.

While using the absorbance formulae:

$$A = log\left(\frac{I_0}{I}\right) \qquad (2)$$

the Lambert-Beer law might be simplified this way:

$$A = \varepsilon_0 . c . l \qquad (3)$$

It results from the formulas stated above that the absorbance is directly proportional to the concentration of the observed substance *c* when the environment thickness *l* is known.

2.2 Device used and measurement principles

The device Testo 435 with a registration part for measured values was used to determine CO_2 concentration in the air. This device works with the non-dispersive infrared sensor (NDIR). Such sensor has the working range from 0 up to 50% volume units of CO_2. This sensor works with so called one channel double ray method.

In a miniature box (see Figure 4) there are placed two infrared receivers and two different optical filters (a measuring and a reference filter). The measuring filter (see Figure 4, position 2) transmits only the emission with the wave length of approx 4,27 μm. The reference filter (see Figure 4, position 4) transmits the emission with the wave length of approx 4,0 μm. The emission with such wave length is not absorbed by gas. By comparison of both signals the electronic device evaluates the concentration of CO_2 in gas.

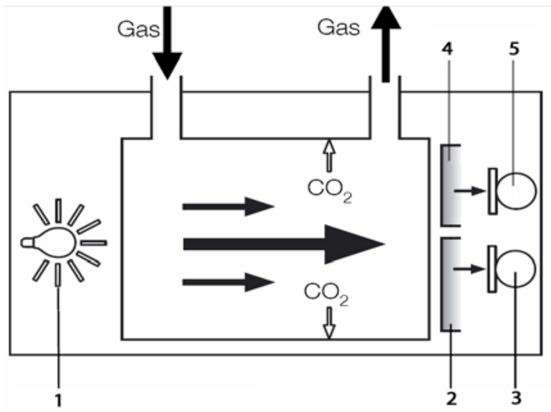


Figure 4 NDIR CO_2 infrared sensor where:
1-Infrared projector, 2- Interference filter I = 4,27 μm Absorption CO_2, 3 – Infrared detector filter I = 4,27μm Absorption CO_2, 4- Interference filter I = 4,0 μm,
5 – Infrared detector filter I = 4,0 μm

The measuring sampler IAQ of the Testo 435 device has besides a CO_2 sensor a humidity sensor, a temperature sensor, and a pressure sensor [7] too. It results from the information stated above that the Testo 435 device is capable to evaluate objectively the total microclimate quality in a vehicle.

2.3 Experiment conditions

The sampling spot was placed at the breathing zone level, which means the place where a driver and a passenger inhale its content.

The measuring itself was taken on the way from Jeseník to Javorník on 3rd August 2008 from 14:03

to 14:57. It was a low middle class car with internal volume of 2 550 dm^3 with two passengers on board. The vehicle was equipped with manually controlled air conditioning. The total cumulative mileage was 40 508 km and the vehicle did not crash before the measuring. Outside climate conditions at the beginning of the measuring were as follows: the CO_2 concentration was 370 ppm, the outside temperature was 28 °C, and the outside air humidity was approximately 51%. The concentration values were recorded every second during the measuring. The record evaluation was carried out withthe application of the Komfort-Software Testo V 3.4.

3 RESULTS AND DISCUSSION

In Figure 5 there is a record of climate conditions inside the vehicle using the Testo 435 device. The whole record might be divided into two main phases: the phase with no people on board (t_1), and the phase with the passengers on board (t_2).

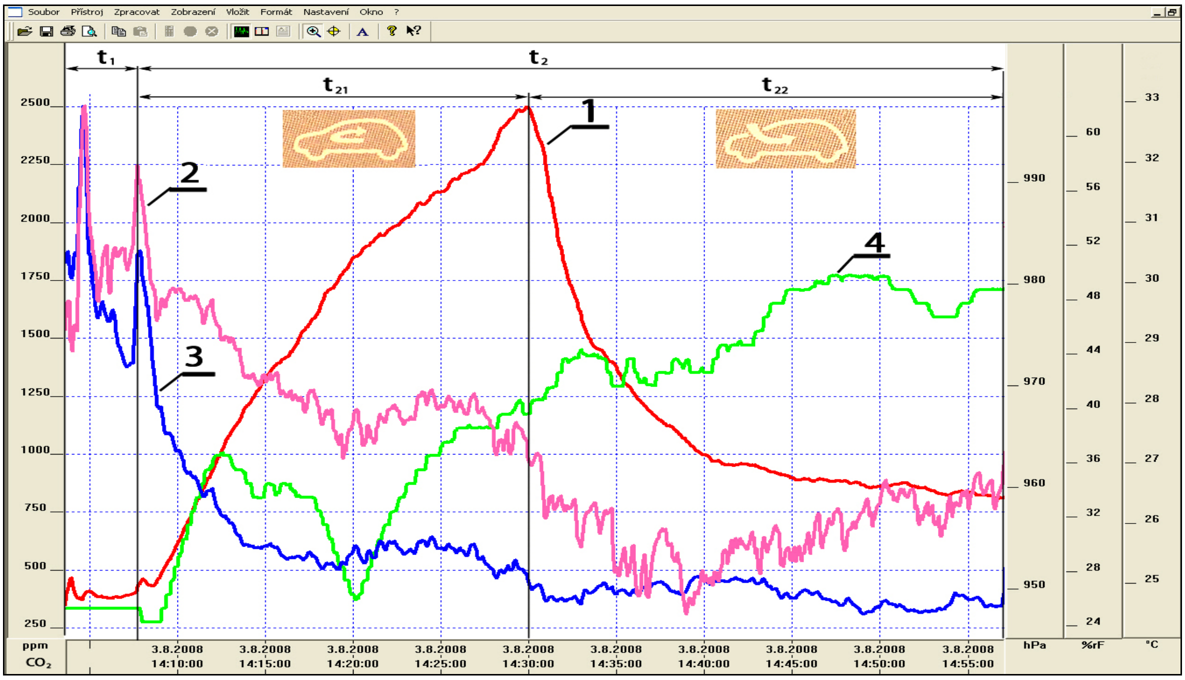


Figure 5 Record of climate conditions measurement inside the vehicle using the Testo 435 device where: 1 – time course of carbon dioxide, 2 – time course of the vehicle internal temperature, 3 – time course of the humidity, 4 – time course of the air pressure, t_1 – the period with no people on board, t_2 – the period with passengers on board, t_{21} – the period with passengers on board and with internal air circulation, t_{21} – the period with passengers on board with external air circulation

The first phase, in the time period of t_1, started by switching on the Testo 435 device, then the opening of the vehicle and installation of the device inside the vehicle came after.

It results from the record that the CO_2 concentration did not change that much (the period t_1) with no people on board – see the curve 1. However, the time course of the temperatures (see the curve 2) and the humidity (see the curve 3) changed significantly (they increased spontaneously). This is perhaps because of the temperature and humidity differences outside and inside the vehicle (the outside temperature was 28 °C and the inside temperature was approx 33 °C, the humidity outside was approx 51% and the inside humidity was approx 61%). It results from the record that a vehicle is supposed to be properly ventilated in the summer time before the journey starts.

The second phase, in the period of t_2, started with two passengers getting into the vehicle, then the starting of the engine, the closing of the doors and the switching on air conditioning came after. In the time period t_{21} the air throttle was adjusted to the position of internal air circulation in the vehicle, and in the time period t_{22} the air throttle was adjusted to the position of outside air income. During the whole entire period of t_2 the ventilator velocity switch was adjusted to the middle level (position 2).

It results from Figure 5 that after switching on air conditioning a nice temperature and humidity climate was created (the indoor temperature dropped from 32 °C to approx 28 °C in six minutes and the humidity settled on the approximate level of 31%). The sharp progressive rise of CO_2 which was caused by breathing of the passengers inside the car contrasts with the settled state. The edge limit of 1 000 ppm of CO_2 was reached approximately 4 minutes after getting into the car. CO_2 was rising continuously. The air throttle was adjusted to the position of external air income when reaching the CO_2 concentration of 2 495 ppm. The ventilator velocity switch was left in the middle level (position 2).

It results from Figure 5 that the CO_2 concentration inside the vehicle dropped down rapidly when opening the air throttle to the position of external air income and letting the outside air come in. However, the air temperature and humidity conditions almost did not change. With the view of this information it is obvious that for keeping good air quality in a vehicle (without CO_2) the air ventilation throttle should be closed for a limited period of time only (e.g., when entering tunnels, driving in heavy traffic in cities, etc.).

4 CONCLUSION

The results of the research showed that small space in a vehicle and insufficient ventilation are the main cause of CO_2 rise. High CO_2 concentration can lead to a driver's drowsiness, tiredness, lethargy, etc. Decrease of driver's attention might be the main cause of traffic accidents. Therefore we suggest developing this matter deeper as well as cooperating with vehicles manufacturers, research institutes, and medical institutions. Implementing an on-board diagnostics for CO_2 concentration inside a vehicle could be a good solution too.

This work was completed as a part of the research project of the Faculty of Military Technologies Nr. MO0FVT0000401 and project ME 949 in frame of the programme KONTAKT MŠMT "Analysis of the negative impact onto the drivers´ attention".

References

[1] TESAŘÍK, J.: *Driving licence points conviction in other words summary of the year 2006 on Czech roads.* Proceedings of the conference: Safety in road transport. Stará Lesná, Slovak republic : 2007. ISBN 978-80-85418-64-4. In Czech language.

[2] NOVÁK, M., VOTRUBA Z.: *Human Factor Influencing the Reliability of Car Driving*. Proceeding of the conference: Wearing, dependability and diagnostics 2007, Brno, Czech republic : 2007. Nr. 1, pp. 197 -204. ISBN 978-80-7231-294-8.

[3] TICHÝ T.: *Methods for exact detection and level of driver attention*. Second book thesis, Praha : ČVUT, 2008. In Czech language.

[4] ČORŇÁK, Š., BRAUN, P.: *On selected aspects of driver's. Proceeding of the conference*: Wearing, dependability and diagnostics 2008. Brno, Czech republic : 2008. Nr. 1, pp. 21-24. ISBN: 978-80-7231-558-1. In Czech language.

[5] ŠŤÁVOVÁ, P. at all.: *Study on Ventilation in a Single-Family House* - Measurement by CO2 Method and Simulation in CONTAM 2.4. Proceedings of conference on Buildings simulation and environment techniques. Praha : 2006. ISBN 80-01-03577-8. In Czech language.

[6] ROSICKÁ, Z.: *Reliable and Safe Management of Hazardous Materials – Asbestos Focused*. Proceeding of the conference: Wearing, dependability and diagnostics 2005. Brno, Czech republic : University of Defence, 2005. pp. 155-162. ISBN 80-7231-026-7.

[7] Sites www: http://www.besip.cz, http://www.testo.cz

Assoc. Prof. Dipl. Eng. Dr. Štefan ČORŇÁK
University of Defence
Kounicova 65
612 00 Brno
Czech Republic
E-mail: stefan.cornak@unob.cz

AIR TRAFFIC COMPLEXITY AS A SAFETY PERFORMANCE INDICATOR

Jelena DJOKIC, Hartmut FRICKE, Michael SCHULTZ, Christoph THIEL

Abstract: Numerous studies emphasize the significant influence of air traffic complexity on controller's workload and consequently on the overall safety level. This paper describes a detailed investigation into these air traffic complexity metrics. The evaluated data originate from traffic samples of a recently conducted real-time simulation in air traffic control at EUROCONTROL CRDS Budapest. In these traffic samples both flight characteristics of individual aircraft and the interactions between pairs of aircraft were taken into consideration. With an attempt to quantify potential changes of the safety level, relevant indicators as the position, collision- respectively conflict probabilities were used. A traffic simulation environment was developed in order to determine these quantitative criteria more accurately. Additional emphasis was put onto the consideration of actual navigation performance of each aircraft to allow correlating air safety with performance based navigation (PBN) concepts.

Keywords: Air Traffic complexity, safety, collision probability, Actual Navigation Performance (ANP).

1 INTRODUCTION

According to the air traffic analyses in the last decade, continuous growth was anticipated in the long term, at least until 2030. Nevertheless, since august 2008 situation surprisingly evolved in the opposite direction for the first time since 2002. While in 2007 a growth of 5% was recorded, in 2008 it reached only 0.4% as a consequence of the global financial crisis and economic downturn. However, on the daily basis, traffic went three times beyond the numbers recorded in 2007 [1].

Taking all this into consideration, greater ATC flexibility is needed for handling all possible future scenarios efficiently, while at least maintaining, rather improving, safety level according to SES key performance indicators [2]. Therefore, numerous studies investigate into enhanced automation, tools and procedures to achieve this goal. Even so, it is very difficult to measure the safety "performance" of ATC, since safety is the absence of accidents, and it is hard to measure absence [3]. Therefore, it is sought for an indicator of safety that would reflect quantified characteristic able to capture variations in safety assessment through different traffic scenarios in order to provide enough data for anticipation and mitigation of safety critical events.

This paper deals exclusively with the endangering of the involved air traffic expressed through the factors defined in order to capture complexity of the traffic and as a specific level of risk or safety for an aircraft conflict or accident per period of time [4].

Numerous studies investigate into safety through controller's workload, as prime factor driving safety, determined by both directly measurable air traffic factors (number of aircraft in the sector, speed, distance between aircraft, etc.) and controller's activity mediated by the controller's abilities, age, fatigue, level of experience, etc. [5]. As the investigation into the correlation between subjective controller workload and safety level was previously conducted and strong connection was found [6] this paper focuses only on air traffic complexity as the objectively measurable prime driver of subjective controller workload and collision probabilities as an attempt to quantify the changes in air traffic safety level.

Therefore, a detailed investigation into these air traffic complexity metrics and collision probability is described, followed by an analysis of their relationship. Moreover, the objective of the current study is to investigate into their interdependencies and examine whether they contribute to the overall safety as two separate independent entities, or they could be considered as unity in which they are partially congruent in depiction of the air traffic complexity.

2 COMPLEXITY

As the number of aircraft within a sector here called 'sector load' is increasing, the routine tasks associated with handling each aircraft safely (monitoring, communication and coordination) are increasing accordingly. Still, the level of difficulty experienced by the controller depends on other traffic factors beyond the sector load. These traffic factors refer both to the flight characteristics of each aircraft in the sector individually and the interactions between different aircraft.

Flight characteristics of the aircraft are referred to as the transition factors which relate to instantaneous changes of the state and position of the aircraft, e.g. changes in altitude, heading and speed. In this manner, interactions between aircraft can be captured by the degree of disorder among aircraft, i.e. the variability in headings and speeds.

The straightforward factor derived from the sector load taking into consideration changes in horizontal and vertical distances between aircraft within time, is called 'density'. Nevertheless, density itself is not affecting controllers feeling of difficulty of the situation as much as if aircraft are converging or diverging. In case of convergence, potential conflicts can emerge driven by sensitivity of both the relative distance and transition factors of the aircraft in the sector.

Though significant number of studies (see [7] for a recent review) considers ATC complexity matter, with many complexity factors proposed, there is still no set of comprehensive and generally accepted measures recognized.

This study presents a list of complexity factors selected from the literature on the basis of continuously recognized relevance (importance) and detailed calculation formula reported.

There are 14 selected complexity factors listed in Table 1, for which more thorough review readers are referred to the indicated source literature, since their detailed description and explanation is out of the scope of this paper.

Table 1 List of the complexity factors

	Complexity Factors	**Used in**
1	number of aircraft	[9], [11], [13]
2	number of climbing aircraft	[9], [11], [13]
3	number of descending aircraft	[9], [11], [13]
4	number of aircraft with lateral distance between 0-25nm and vertical separation less than 2000ft above 29000ft	[11], [10]
5	variance of ground speed	[9], [11], [13]
6	density indicator (mean)	[8], [12], [13]
7	variability in headings (track disorder) (mean)	[8], [12], [13]
8	variability in speed (speed disorder) (mean)	[8], [12], [13]
9	divergence between pairs of aircraft (mean)	[8], [12], [13]
10	convergence between pairs of aircraft (mean)	[8], [12], [13]
11-12	sensitivity indicator (a/c converging-mean; a/c diverging-mean)	[8], [12], [13]
13-14	insensitivity indicator (a/c converging-mean; a/c diverging-mean)	[8], [12], [13]

3 SAFETY CORRELATOR MODEL

The Safety Correlator (SC) is a fast- and real-time simulation tool, which calculates a Risk index – the 'Level of Safety' (LOS) – on the basis of the quantitative criteria collision and conflict probability, where the collision probability is defined as the probability of a collision of two aircraft within the terminal area (TMA). The second metric, the conflict probability, is expressed as the probability of a contact or penetration of a predefined safety zone around two aircraft. As the basic principles of the underlying model are comprehensively explained in [14] and [15], the following section only gives a short introduction to the theoretical background.

The RTCA CD&R Working Group has defined the term „Protected Airspace Zone" (PAZ) in its work about collision and conflict probability [16]. The PAZ represents airspace around aircraft in which no other aircraft shall penetrate. The dimension and shape of this safety zone has obviously a significant impact onto the resulting conflict probability. For this reason, the PAZ chosen from the RTCA CD&R Working was modified as follows to achieve an appropriate and still conservative metric for safety measurement: The original PAZ consists of a cylinder with a radius equal to the separation minima. So, each separation infringement leads mandatorily to a conflict. This modus operandi would lead to an excessive large conflict probability in the here observed airspace with its characteristic high traffic density. The consequently extended safety zone used here is illustrated in Figure 1.

The dimensions of this safety zone are deduced from the dimensions of the aircraft, from the values of the actual navigation performance of the reference aircraft, and from the velocity difference to the potential conflict partner, the intruder aircraft. The ANP values (differentiated in longitudinal, lateral, and vertical direction) are used analogue to the RNP values [17] as the standard deviation σ of a normal probability density function (PDF) to calculate the position probability by integration. The calculation refers to a space grid of pre-defined, three-dimensional raster elements in which the observed airspace is subdivided. A raster element represents the two dimensional integration boundaries when calculating the position probability of each aircraft every time interval. This is a valid simplification as a raster element corresponds to the dimension of the largest aircraft within the observed airspace. During each simulation cycle (typically this equals one second in real time) the assumed position for each reference aircraft in the observed airspace is calculated for all raster elements.

Because of the chosen probability approach, it is virtually possible, that an aircraft is expected in multiple raster elements. At this, the probability

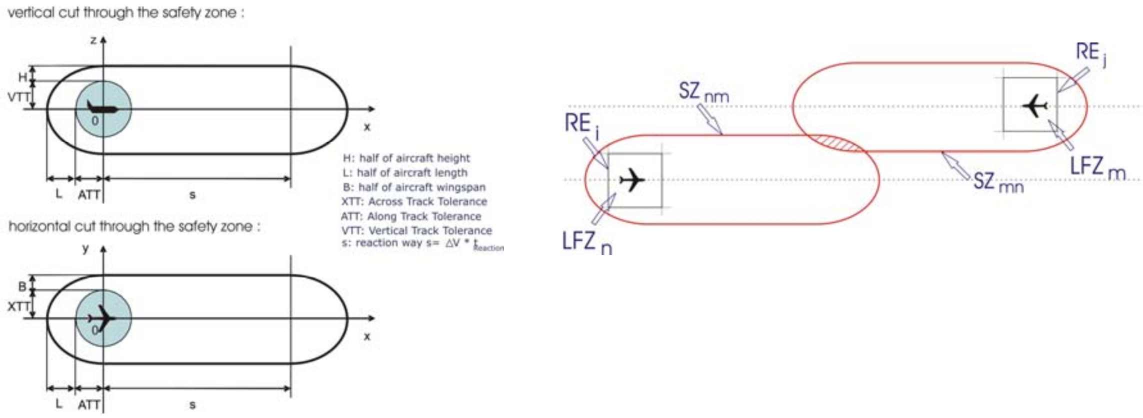


Figure 1 a) Adapted Safety Zone Model; b) Conflict Probability Scheme

decreases heavily with increasing distance of the raster element to the reference position of this aircraft with respect to its actual navigation performance, modeled through the PDF. So, it is necessary to calculate the conflict probability for every possible spatial configuration of any pair of aircraft. The sum of these conflict probabilities is the probability of a conflict between that pair of aircraft.

The described model is implemented as a prototype called *Safety Correlator* (SC) using the JAVA programming language. Beside modules for modeling the ANP of each individual aircraft and the modules for calculating the conflict probabilities, the SC also features a graphical "radar screen like" user interface. With this interface, the user is able to set the operational and technical constraints and to review the results of the calculations as a histogram "conflict probability over time". Furthermore, the SC has standard interfaces to connect to external sources of traffic data (e. g. radar data using the ASTERIX protocol).

4 THE REAL-TIME SIMULATION EXPERIMENT

4.1 Simulation

The data used for the analysis were collected during a two week IAA RTS1 (Irish Aviation Authority Real Time Simulation 1) real-time simulation experiment that aimed at investigating the operational feasibility, efficiency and benefits of solutions based on a package of modifications in the provision of the IAA Air Traffic Service (ATS) in the Shannon Control Area (CTA) below FL245; so in the low airspace [18].

For the present investigation, IAA RTS1 data was reduced to the busiest sector of the considered airspace (SHLOW). As a baseline, data was used from a scenario in which no operational changes were introduced, thus reflecting the current situation but with varying traffic load over time. Data was so used from 13 exercises, each lasting 1hour and 20 minutes in total, out of which approximately one hour was recorded.

4.1 Complexity and Safety measures

The flight plans and flown trajectories were used as input data for both the complexity factors calculations and the conflict and collision probability over time as a technical safety indicator. In particular the following data was used: Position of the aircraft (latitude, longitude and altitude), speed, heading and the next waypoint within a sector. Synchronously to the SC collision probability calculations, the complexity factors were updated every 5 seconds. This was achieved with additional software that was developed for this experiment. So a completely synchronised data set was obtained for the experiment declaring evolving complexity and safety metrics over time.

5 STATISTICAL EVALUATION AND RESULTS

The data used for the statistical analysis were derived from the recordings of the exercise covering a 1 hour timeframe divided in 5 second time steps. This lead typically to 9,360 measurements for each indicator (12measurements/minute → 12x60=720 measurements/hour → total: 720meas./h x 13 exercises= 9,360). However, due to recordings that had duration longer than 1hour in some exercises, the overall set comprised 9,573 measurements, forming the basis for the statistical analysis presented next.

5.1 Application of the Principal Component Analysis

As the initial set of the chosen complexity factors aimed at capturing to a large extent all aspects of complexity, a high dimensionality was reached, being difficult to analyse. Further, interdependencies were expected to exist between the variables.

Therefore, a Principal Component Analysis (PCA) was performed on this set of complexity factors trying to keep most of the (expected) information of the original set of variables while reducing its absolute number. The PCA is a method to reduce data dimensionality by performing a covariance analysis between factors.

Principal components with an Eigenvalue higher than 1 were extracted and subsequently rotated using the VARIMAX method. The rationale for using the criterion of an Eigenvalue>1 lies in fact that such a component is accounting for a significant variance in the overall data set. Consequently, components with Eigenvalues less than 1 were considered to be neglected and were not retained. Beside that "importance judgement", the VARIMAX rotation invokes an orthogonal rotation resulting in independent components.

There were 4 principal components extracted as the result of this statistical method. They accounted for 77.6% of the total variance in the overall data set. Table 1 displays these components sorted by descending Eigenvalues. It further depicts the percentage of variance they are accounting for. Taking into consideration loadings of a given factor on a given component in the 4 component x 14 factors matrix, the component meanings are derived. As the loading is equivalent to the correlation between that factor and its component, the factor with the highest loading generally leads the interpretation of the component.

Table 2 Results of the PCA performed on 14 complexity factors

Components	Eigenvalue	% of Variance	Cum. % of Variance
Comp.1	5.543	35.591	35.591
Comp.2	1.950	17.930	53.521
Comp.3	1.740	12.429	65.950
Comp.4	1.630	11.645	77.595

Comp.1 – aircraft distribution – highly correlated to the density factor and the number of aircraft in the sector (0.953 and 0. 942 respectively).
Comp.2 – aircraft transitioning – component that is strongly related to the variance of the ground speed on one side (0.830) and the number of climbing aircraft (0.777).
Comp.3 – convergence – this component describes convergence between aircraft in the sector as it shows relation to both the convergence factor and the insensitivity factor associated to the convergence between aircraft (0.963 and 0.966 respectively).
Comp.4 – divergence –analogously to the previous component, but in the opposite direction this component relates to divergence factor and associated insensitivity (0.957 and 0.954).

In a second step of the analysis the correlation between complexity factors and the collision probability taken from the SC as safety indicator was examined. In order to reveal this relationship, a second PCA application was performed. The hypothesis for this tep was: If those two entities to a certain extent describe the same aspect of safety- i.e. the complexity of the instantaneous geometry of the air traffic situation, the consideration of the collision probability as additional factor would reach a higher percentage of the total variance contained in the factor set (at 77.6% with the 1. PCA process). Additionally, the collision probability would contribute with a given loading on one of the components and by that reveal its interdependence with specific complexity factors with loadings within the same component. On the other hand, if those two entities refer to different aspects of safety, the total variance would be less.

Therefore, the same procedure to extract the relevant components was performed (Eigenvalues>1, VARIMAX rotation). Interestingly, with the same 4 principal components a lower value of the total variance in the metrics, now found with 72.95% did result, each of it contributing as follows:
Comp.1 (aircraft distribution) – density factor (0.938) and number of the aircraft in the sector (0.922),
Comp.2 (aircraft transitioning) – variance of the ground speed (0.815), number of the climbing aircraft (0.799), *Comp.3 (convergence*)– konvergence (0.963), and associated insensitivity (0.966) and
Comp.4 (divergence) – divergence (0.954) and associated insensitivity (0.956).

However, the collision probability factor did not provide loading on any of the extracted components.

Table 3 Results of the PCA performed on 14 complexity factors and collision probability

Components	Eigenvalue	% of Variance	Cum. % of Variance
Comp.1	5.598	37.318	37.318
Comp.2	1.953	13.021	50.339
Comp.3	1.741	11.603	61.942
Comp.4	1.651	11.004	72.946

6 CONCLUSIONS

With an attempt to quantify characteristics able to capture variations in safety of air traffic operations through different traffic scenarios, two aspects were investigated: Complexity of the instantaneous geometry of the traffic situation on one side and so-called collision/conflict probability on the other side. Further, in order to investigate into their interdependencies and examine whether they contribute to the overall safety as two separate independent entities, or they can be treated as unity as being partially congruent in depiction of the air traffic complexity, a two-step Principal Component Analysis (PCA) analysis was performed. In the first step, the PCA was performed using the initial set of complexity factors without collision probability as a baseline scenario, while in the second-step the conflict/collision probability factor was considered additionally. The intention of the first PCA was to reveal interdependencies within the set of complexity factors and consequently, reduce its dimensionality. The second PCA intended to investigate the level of correlation between the baseline complexity factors and the conflict/collision probability.

The analysis resulted in the extraction of the same components in both steps of the analysis, but with a lower total variance for the second one. Additionally, the conflict/collision factor did not add any loading on any of the derived components. As the loading is equivalent to the correlation between that factor and the component, it can be concluded that there is no interdependency between components identified based on the complexity factors and the conflict/collision probability factor. It also explains that conflict/collision probability does not give added information to the complexity of the air traffic situation, but targeting another scope of the safety assessment.

Namely, based on the analysis performed on the available dataset, the conclusion drawn out is that each of them captures different aspects of safety and provides separately added value in quantification of the changes in the safety level. Further research into both complexity and conflict/collision probability is recommended, as well as a deeper investigation into their correlation with the changes in safety level.

Acknowledgements – The authors would like to thank to EUROCONTROL CRDS, Budapest, and Irish Aviation Authority for providing us with data used within this study.

References

[1] EUROCONTROL, Performance Review Commission: *An Assessment of Air Traffic Management in Europe during the Calendar Year 2008*, Performance Review Report, 2009.

[2] SESAR Consortium: Air Transport Framework: *The Performance Target*, D2, 2006.

[3] PILON, N.: *ATM and Society – Demands and Expectations*, European air traffic management: principles, practice, and research, Edited by Andrew Cook, University of Westminster, UK, pp. 199-220, Ashgate 2007.

[4] DJOKIC, J., KIETZMANN, M., FRICKE, H.: *Assessing the Collision Risk in High Density Airspace - A Quantitative Model*, Proceeding of The International Scientific Conference MOSATT 2007, Zlata Idka, Slovak Republic, 2007, pp. 83-88.

[5] LOFT, S., SANDERSON, P., NEAL, A. and MOOIJ, M.: *Modeling and predicting mental workload in en route air traffic control: critical review and broader implications*, In: Human Factors: The Journal of the Human Factors and Ergonomics Society, 2007, vol. 49, pp. 376-399.

[6] DJOKIC, J., LORENZ, B., FRICKE, H.: *ATC Complexity as workload and safety driver*, Proceeding of The International Conference on Research in Air Transportation – ICRAT 2008, Fairfax, Virginia, U.S.A, 2008.

[7] HILBURN, B.: *Cognitive complexity in air traffic control - a literature review*, EEC Note No. 04/04)", EUROCONTROL, 2004.

[8] DELAHAYE, D., PUECHMOREL, S.: *Air traffic complexity: towards intrinsic metrics*, In: Proceedings of the 3rd USA/Europe Air Traffic Management 2000 R&D Seminar, Napoli, Italy, 2000.

[9] CHATTERJI, G. B., SRIDHAR, B.: *Measures for air traffic controller workload prediction*, In: Proceedings of the First AIAA Aircraft Technology, Integration, and Operations Forum, Los Angeles, CA, 2001.

[10] LAUDEMAN, I.V., SHELDEN, S. G., BRANSTROM, R., BRASIL, C. L.: *Dynamic density: an air traffic management metric*, NASA-TM-1998-112226, San Jose State University Foundation, San Jose, CA, 1998.

[11] KOPARDEKAR, P., MAGYARITS, S.: *Measurement and prediction of dynamic density*, In: Proceedings of the 5th USA/Europe Air Traffic Management 2003 R&D Seminar, Budapest, Hungary, 2003.

[12] CHATTON, F.: *Etudes de nouvelles métriques de complexité de la circulation aérienne,* Master's thesis, Ecole Nationale de l'Aviation Civile (ENAC), 2001.

[13] GIANAZZA, D., GUITTET, K.: *Evaluation of air traffic complexity metrics using neural networks and sector status*, In: Proceedings of the 2nd International Conference on Research in Air Transportation (ICRAT), Belgrade, Serbia, 2006.

[14] FRICKE, H.: *Sicherheit im Luftverkehr: Eine Methodik zur Quantifizierung luftverkehrlich induzierter Risiken an Flughäfen*; DGLR Jahrestagung 2001, Hamburg.

[15] KIETZMANN, M., FRICKE, H.: *Entwicklung eines quantitativen Bewertungsmodells zur objektiven Beurteilung des Einflusses neuer Betriebsverfahren auf die Sicherheit im Luftverkehr*; DGLR Jahrestagung 2004, Dresden.

[16] RTCA, RTCA Task Force 3, *Free Flight Implementation*, Final Report of RTCA Task Force 3, RTCA, Washington, DC, 1996.

[17] ICAO: ICAO-Doc. 9613- AN/937: *Manual on Required Navigation Performance (RNP),* Second Edition, Montreal, 1999.

[18] EUROCONTROL, *IAA Real Time Simulation*, Final report (CRDS Note No. 32), EUROCONTROL, 2008.

Dipl. Eng. Jelena DJOKIC, PhD.
Technical University Dresden
Institute of Aviation and Logistics
010 62 Dresden
Germany
E-mail: djokic@ifl.tu-dresden.de

Prof. Dipl. Eng. Hartmut FRICKE, DrSc.
Technical University, Dresden
Institute of Aviation and Logistics
010 62 Dresden
Germany
E-mail: fricke@ifl.tu-dresen.de

Dipl. Eng. Michael SCHULTZ, PhD.
Technical University, Dresden
Institute of Aviation and Logistics
010 62 Dresden
Germany
E-mail: schultz@ifl.tu-dresden.de

Dipl. Eng. Christoph THIEL
Technical University Dresden
Institute of Aviation and Logistics
010 62 Dresden
Germany
E-mail: thiel@ifl.tu-dresden.de

NOTES ON VEHICLE-BASED VISUAL TRAFFIC SURVEILLENCE FOR CRASH-PREDICTION

Péter GÁSPÁR, Zoltán FAZEKAS, Alexandros SOUMELIDIS

Abstract: Though some sections of motorways and multi-lane roads have been built in the last few years in Hungary, a good portion of the domestic and transit road traffic is still served by two-lane roads. Some of these roads are burdened with intense and relatively slow truck and lorry traffic. The traffic safety situation is aggravated by the impatience and carelessness of a sadly high number of drivers. These factors render the mentioned roads extremely unsafe. It is questionable whether the conventional measurement approaches – including also the fixed-camera measurements – are adequate for traffic safety monitoring purposes and whether they can provide sufficient insight into the traffic safety situation. Building upon the experience with camera-based traffic lane tracking systems, relying on reports on the applicability of surrogate safety measures for crash-prediction, and on the application of omni-directional vision systems in road vehicles, a mobile omni-directional camera-based approach of assessing the safety of the mentioned roads is outlined and motivated herein.

Keywords: Traffic safety, traffic surveillance, dynamic surround map, surrogate safety measures, catadioptric cameras, omni-directional vision systems.

1 INTRODUCTION

A large body of statistical data on traffic accidents is available from the EU countries and from OECD member states [15]. Comprehensive research studies analyze the causes, the fatalities, and the locations of the road accidents, as well as their perceptible trends [12], and estimate their economic consequences [17]. In Hungary, fatal traffic accidents number approximately 1000 per annum with about 7,000 traffic accidents per year resulting in serious injuries. With only a relatively small portion of the budget being spent on building new motorways, and in general on the maintenance and reconstruction of the existing road infrastructure which is in a fairly poor shape already, it is of utmost importance to use the available moneys in a sensible way.

Building upon the authors' experience gained during the development of a wide-baseline stereo vision-based traffic lane tracking system [2], [3], [4] and upon their experience in modelling free-form specular surfaces [6], furthermore, relying on reports on the surrogate safety measures [10] and particularly on the Surrogate Safety Assessment Model (SSAM) [11], and finally relying on the domestic and international experience with omni-directional vision systems [7], [8], [9] a viable approach of assessing the traffic safety of these roads is outlined herein. In this approach, the imaging device is an omni-directional camera mounted on a probe-vehicle that moves with the road traffic. In the present paper, the viability of this approach is considered for the particular road type and the main processing steps of this approach are outlined. It is probably worthwhile to mention why the aforementioned experiences are relevant to the present topic. First of all, the application field (i.e., road traffic related measurement) and the sensing modality (i.e., video-stream from camera) are common with those of the traffic lane tracking system mentioned above that was implemented in MATLAB Although, the real-time implementation of traffic lane tracking system is still under development, the real-time FPGA-based early-vision processing carried out in smart cameras has already been reported [4]. In the present application, both a real-time implementation, and an offline (i.e., slower) implementation would be meaningfully, though obviously the real-time implementation would facilitate a wider range of safety-related intervention into the traffic. In Figure 1, the cameras of the mentioned wide-baseline stereo lane tracking system is shown with its cameras mounted on the side-mirrors of the host vehicle.

Another experience that comes handy in developing the mentioned vehicle-based traffic safety measurement system is to do with the mathematical modelling light-reflection at free-form surfaces. This is because non-spherical – in cases custom-designed – mirrors are often used in catoptric and catadioptic omni-directional systems.

Figure 1 The cameras of the stereo lane tracking system – mentioned in the text – mounted on the host vehicle

The structure of the paper is as follows. In Section 2, the most important traffic surveillance approaches are summarised, then the use of the surrogate safety measures for assessing dynamic traffic scenes is motivated and the traffic situation on the roads to be assessed are described. Still in this section, the relevant features of the omni-directional cameras are summarised. In Section 3, the main processing steps involved in the vehicle-based traffic safety measurement are outlined. Finally, Section 4 concludes the paper and further work leading to realization of this traffic safety measurement is summarised.

2 BACKGROUND

2.1 Traffic Surveillance Approaches

Though, herein the safety related aspects of the traffic surveillance systems are emphasised, traffic surveillance systems are also used for other traffic related purposes, as well. These include traffic management, and travel time collection. The collected data could then be forwarded either in an offline, or an online manner to data collecting centres. The mode of data management obviously affects the way the collected data can be utilised. There are two main types of traffic surveillance systems, namely road-based and vehicle-based traffic surveillance systems.

The road-based traffic surveillance systems, such as inductive loops, have been a principal means of road surveillance and incident detection for many years mainly because of their reliability and inexpensive operation. However, the camera-based and other recent roadside vehicle detection technologies are now used extensively to measure high-volume road traffic [8], [9].

The technological advances seen in vehicle sensors [18] and the increase in speed and reliability of detection algorithms [1] make it possible to implement vehicle-based surveillance.

Vehicle-based traffic surveillance systems involve probe-vehicles – equipped with some tracking devices – that allow the vehicles to be tracked by a central computer facility. Although, vehicle-based traffic surveillance systems are not yet in wide use, these systems show promise of providing rich data on travel times and as a means to detect incidents. They can also be used to estimate flows and origin-destination patterns. Recently, a visual vehicle-based traffic surveillance system, interestingly, but not surprisingly using an omni-directional vision approach, was proposed for the above outlined applications [7]. The paper's most relevant aspect to the particular safety assessment approach suggested herein is the generation and the use of dynamic panoramic surround maps by the described system. These surround maps can be used to calculate the surrogate safety measures [10] covered briefly in Subsection 2.3. In the present paper, the visual vehicle-based traffic surveillance approach – such as the one described in [7] – is considered for certain types of roads.

As an illustration, a bird's eye view of the traffic scene is shown in Figure 2. It was derived using an inverse perspective mapping of the traffic scene shown on the left. This bird's eye view can be taken as an instantaneous surround map with a car – shown as a fairly large green region because of the inverse perspective mapping applied – in front of the host vehicle. The dynamic surround map is used to describe the movement and the trajectory of the host car and other vehicles with respect to the road surface.

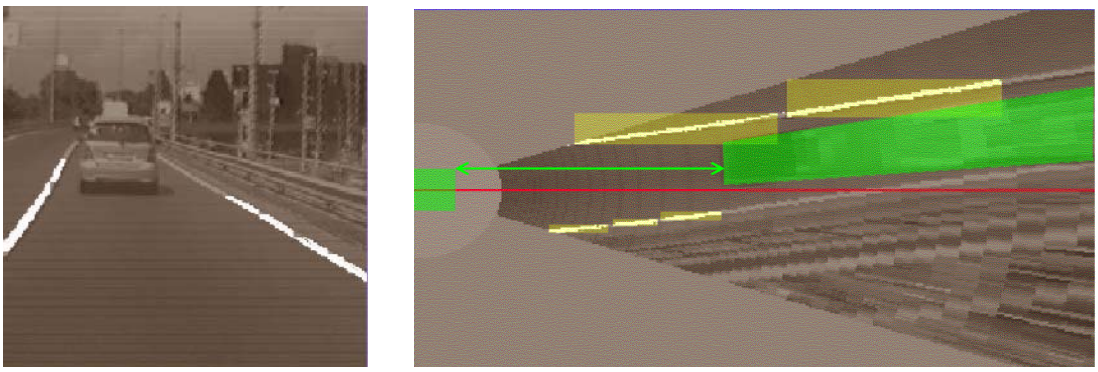


Figure 2 A traffic scene with lane marking detected (a) and its bird's eye view (b)
The distance between the host car and the car followed is marked.

2.2 Target Roads for Visual Vehicle-based Traffic Surveillance

Though some sections of motorways and multi-lane roads have been built in the last few years in Hungary, a good portion of the road traffic is still served by two lane roads. Lengthy sections of the national main road network belong to this road category. Some of these sections are burdened with intense and relatively slow truck and lorry traffic; these sections are frequently oversaturated. The situation is aggravated by the impatience and carelessness of some drivers. All these factors contribute to make these sections extremely unsafe. In the following, such roads will be referred as target roads.

There is a growing governmental intention to monitor, regulate and control the traffic on these roads, e.g. by deploying speed cameras and other traffic surveillance equipment, and by using traffic-dependent traffic light control regimes that optimise the throughput of junctions. However, as frequent users of these roads, the authors' opinion is that there is much to be done in respect of accident prevention on these roads.

Indeed, the drivers' irresponsibility has reached an astonishing and frightening level, manoeuvres such as risky and life-threatening overtaking has become the 'norm' on these roads. In many cases, it is not the speed-limit though that is violated, but the prohibition, or the safety rules of overtaking. In view of this rankling traffic safety situation, it is questionable that viable measurements and preventive measures – with any hope of success – can be solely stationary in respect of the target roads. In our view, under the mentioned circumstances, some measurement devices should move together with the traffic. Therefore, we propose the use of some mobile visual vehicle-based traffic surveillance system – such as described in [10] – for the target roads. The data collected in this manner could be processed, analyzed and data-mined with the aim of identifying non-trivial unsafe geographic locations (i.e., road sections other than junctions) and other conditions, where and under which unusual vehicular and driver behaviour frequently occur. This can be done in a similar manner to the validation of the crash-prediction potential of micro-simulations of intersections reported in [11]. This way, potential non-trivial accident hotspots could be identified and appropriate preventive actions could be taken.

2.3 Applying the SSAM Approach for Dynamic Traffic Scene Evaluation

The Surrogate Safety Assessment Model (SSAM) was developed and validated in the United States [11]. This assessment method augments the analysis of the expensive and unreliable field measurements and actual crash data and uses runs of appropriately tuned micro-simulations of the road traffic at junctions or road segments to assess traffic safety in these locations. Though, some traffic volume-based accident-prediction models – e.g., [20] – show better correlation with factual crash data than the SSAM-based predictions [11], the micro-simulation-based models provide a good insight into the underlying traffic patterns and accident types. The importance of micro-simulation-based models is evident in case of uncommon road geometries and arrangements, and in case designs, i.e., junctions not yet built. Furthermore, there is a wide range of differences between countries in the usual/local behaviours on roads. E.g., in the SSAM projects the American researchers were hesitant to use accident prediction models from other countries. They only did so in the case of some recently introduced traffic layout. Micro-simulation-based modelling seems to be more tuneable toward these local behaviour patterns. As declared in [11], the SSAM is expected to be used on real data acquired from visual traffic surveillance systems. So, as the camera coverage of the busiest intersections in Hungary increase, it is foreseen that surrogate safety measures can be collected via automated video processing methods at least for research and experimental purposes. Note that in case of vehicle-based traffic safety surveillance opted for in this paper, the surrogate safety measures need to be adapted to non-stationary measurements, i.e., when also the ego-motion of the vehicle hosting the surveillance camera must be considered.

2.4 Image Acquisition with Omni-directional camera

The advantage of using an omni-directional camera in the given context can be verified by considering e.g., the monocular lane detection approach depicted in Figures 2. In the figure, a traffic scene with detected lane markings is shown; the distance between the host car – equipped with the camera – and the leading car is also marked. The original unprocessed image was taken with a normal, relatively narrow-angle monocular camera, as it can be verified from the bird's eye view of the scene shown in Figure 2b. With similar objectives and cameras eight or more synchronized cameras would be necessary to cover the 360 degree field-of-view – required in this dynamic traffic safety assessment task – in a monocular fashion. To cover this field-of-view in a stereoscopic fashion, (i.e., doubly to enable point correspondences), further cameras would be necessary. Using a fisheye stereo vision systém as suggested in [9], still at least four synchronised cameras would be necessary.

Note that the cameras would need separate mountings, connections, and possibly separate calibration procedure.

Various omni-directional vision systems are known from the literature. The dioptric systems and catoptric systems are based on special lenses and on non-planar mirrors, respectively; while the cata-dioptric systems use both of these optical elements. Some of these systems are reviewed in [7], [18].

Using such optical systems, non-customary, strongly distorted images – similar to the one shown in Figure 3a – are taken and recorded. In Figure 3b, a possible arrangement of a catadioptric vision system is sketched for the car appearing in Figure 1. The upward looking conventional camera and the non-planar mirror of the vision system are rigidly mounted on a rod. Figure 3c illustrates how a mapping is generated via reflection of light at a free-form surface. By choosing an appropriate non-spherical specular surface for the catadioptric vision system, a better use of the image sensor area could be achieved. Such a catadioptric arrangement is shown in Figure 4.

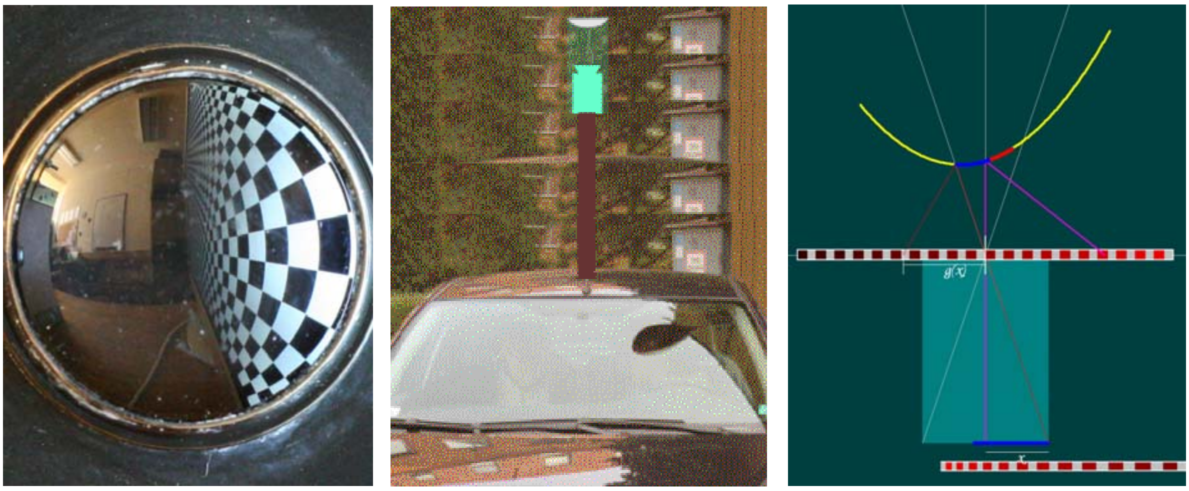


Figure 3 A spherical mirror reflecting a lab environment (a). The sketch of a catadioptric vision system mounted on the host vehicle (b). The mapping created by a non-planar specular surface reflecting a marked plane into a camera (c)

Using two or more appropriately designed and placed specular surfaces – facing a single camera, or a number of cameras – stereo imaging can be achieved, that is, a stereoscopic reconstruction of the scene is possible.

3 VEHICLE-BASED TRAFFIC SURVEILLANCE FOR CRASH-PREDICTION

Though, the main processing steps of the vehicle-based traffic safety assessment task are not markedly different from the general scheme of the lane detection algorithms – based on cameras and on additional sensors – given in [3], some additional processing steps must be included in the mentioned scheme. From the image data – in this case taken by the omni-directional imaging set up – the road and vehicle features are extracted.

Based on the road features and considering the road surface model (e.g., planar road) and the sensor model (in this case the model describing the rigid arrangement of a known specular surface and a calibrated camera), the outliers are removed from the road features. Then a lane model is fitted on the blobs representing the inlier road features.

Then taking ego-vehicle dynamics into account, tracking is carried out with respect to the road and – after proper filtering – to other vehicles. The surrogate safety measures are calculated based on these tracking results.

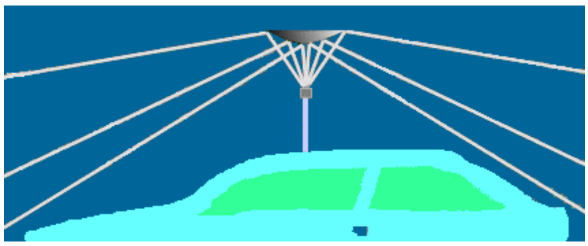


Figure 4 The sketch of a catadioptric vision system – mounted on the host vehicle – comprising a non-spherical mirror

4 CONCLUSION

In this paper, a particular visual vehicle-based traffic surveillance approach was proposed for certain types of roads (referred herein as target roads). Two modes of use were considered: with online and offline data collection.

For evaluating the safety risk of the traffic events extracted from the image sequences and to use these categorised incidents for crash prediction, the surrogate safety assessment measures and approach was proposed. The use of a single omni-directional camera with an on-board computer equipped with appropriate video, navigational and vehicle dynamics data recording capabilities is suggested.

The financial support provided by the Hungarian National Office for Research and Technology in the frame of grant TECH_08_2/2-2008-0088 is gratefully acknowledged by the authors.

References

[1] BÉCSI, T., PÉTER, T.: *A Mixture of Distributions Background Model for Traffic Video Surveillance,* Periodica Polytechnica. Transportation Engineering, Vol. 34, No. 1-2, 2006, pp. 109-117. ISSN 0303-7800.

[2] BÓDIS-SZOMORÚ, A., DABÓCZI, T., FAZEKAS, Z.: *A Far-Range Off-line Camera Calibration Method for Stereo Lane Detection Systems*, Proceedings of the IEEE Instru-mentation and Measurement Technology Conference. Warsaw, Poland : 2007, pp. 1-6. ISBN 1-4244-0588-2.

[3] BÓDIS-SZOMORÚ, A., DABÓCZI, T., FAZEKAS, Z.: *A Lane Detection Algorithm Based on Wide-Baseline Stereo Vision for Advanced Driver Assistance*, Proceedings of the 7th Conference of the Hungarian Association for Image Processing and Pattern Recognition, Budapest, Hungary, Paper-id. 142-2, 2009, pp. 1-10.

[4] BÓDIS-SZOMORÚ, A., VIRÁG, B., WAHL, I., FAZEKAS, Z., DABÓCZI, T.: *Stereo vision system for automatic detection and tracking of traffic lanes*, (in Hungarian), A Jövő Járműve, ISSN 1788-2699, 2008/3-4, 2008, pp. 48-55.

[5] COIFMAN, B. BEYMER, D., McLAUCHLAN, P. and MALIK, J.: *A Real-time Computer Vision System for Vehicle Tracking and Traffic Surveillance*, Transportation Research Part C: Emerging Technologies, ISSN 0968-090X. Vol. 6, No. 4, 1998, pp, 271-288.

[6] FAZEKAS, Z., SOUMELIDIS, A., BÓDIS-SZOMORÚ, A., SCHIPP, F.*: Specular surface reconstruction for multi-camera cornea topographer arrangements*, Proceedings of the 30th Annual International IEEE EMBS Conference, Vancouver, British Columbia, Canada, ISBN 978-1-4244-1815-2, 2008, pp. 2254-2257.

[7] GANDHI Tarak, TRIVEDI Mohan Manubhai: *Vehicle Surround Capture: Survey of Techniques and a Novel Omni Video Based Approach for Dynamic Panoramic Surround Maps*, IEEE Transactions on Intelligent Transportation Systems, Vol. 7, No. 3, 2006, pp. 293-308.

[8] GEHRIG, S. K.*: Large-Field-of-View Stereo for Automotive Applications*, The Proceedings of the Sixth Workshop on Omni-directional Vision, Camera Networks and Non-classical Cameras, Beijing, People's Republic of China, 2005.

[9] GEHRIG, S. K., RABE, C., KRUEGER, L.: *6D Vision Goes Fisheye for Intersection Assistance*, Canadian Conference on Computer and Robot Vision, Windsor, Ontario, Canada, ISBN 978-0-769-53153-3 2008, pp. 34-41.

[10] GETTMAN, D., HEAD, L.: *Surrogate Safety Measures from Traffic Simulation Models,* Transportation Research Record: Journal of the Transportation Research Board, ISSN 0361-1981, Vol. 1840, 2003, pp. 104-115.

[11] GETTMAN, D., PU, L., SAYED, T., SHELBY, S.: *Surrogate Safety Assessment Model and Validation*: Final Report, FHWA-HRt-08-051, Federal Highway Administration, Washington, D.C., USA, 2008.

[12] HOLLÓ, P: *Is the EU Road Safety Target Realistic for Hungary?* 4th Conference on Safe Roads in the 21st Centrury, Budapest, Hungary, 2006.

[13] HU, W., XIAO, X., XIE, D., TAN, T., MAYBANK, S.: *Traffic Accident Prediction Using 3D Model-based Vehicle Tracking*, IEEE Transactions on Vehicular Tech-nology, ISSN 0018-9545, Vol. 53, No. 3, 2004, pp. 677-694.

[14] LAYERLE, J. F., SAVATIER, X., MOUADDIB, M., ERTAUD, J. Y.*: Catadioptric Vision System for an Optimal Observation of the Driver Face and the Road Scene,* IEEE Intelligent Vehicles Symposium, Eindhoven, The Netherlands, 2008, pp. 410-415.

[15] OECD: FACTBOOK 2009, ISBN 978-92-64-05604-6, OECD, 2009.

[16] SAUNIER, N., SAYED, T., LIM, C.: *Probabilistic Collision Prediction for Vision-Based Automated Road Safety Analysis*, IEEE Intelligent Transportation Systems Conference, Seattle, Washington, USA, 2007, pp. 872-878.

[17] SESZTAKOV, V., TÖRÖK, A.: *Monetary value of road accidents in Hungary*, Proceedings of the International Scientific Conference on Modern Safety Technologies in Transportation, Kosice, Slovakia, ISBN 978-80-969760-2-7, 2007, pp. 250-253.

[18] TRIVEDI, M. M., GANDHI, T., McCALL, J.: *Looking-In and Looking-Out of a Vehicle: Computer-Vision-Based Enhanced Vehicle Safety,* IEEE Transactions on Intelligent Transportation Systems, Vol. 8, No. 1, 2007, pp. 108-120.

[19] WUERZ, A., GEHRIG, S. K., STEIN, F. J.: *Enhanced Stereo Vision Using Free-Form Surface Mirrors*, In: Robot Vision, R. Klette, S. Peleg, G. Sommer (Eds.), Lecture Notes in Computer Science, Vol. 1998, ISBN 978-3-540-41694-4, Springer Verlag, Heidelberg, 2001, pp. 91-98.

[20] ZEGEER, C. V., HUMMER, J., HERF, L., REINFURT, D., HUNTER, W.: *Safety Effects of Cross- Section Design for Two-Lane Roads*, Report No. FHWA-RD-87-008, Federal Highway Administration, Washington, D. C., USA, 1986.

Prof. Péter GÁSPÁR, MMEng., PhD., DSc.
Computer and Automation Research Institute
Hungarian Academy of Science
Kende u. 13-17
H-1111 Budapest
Hungary
E-mail: gaspar@sztaki.hu

Dr. Zoltán FAZEKAS, MEEng. PhD.
Computer and Automation Research Institute
Hungarian Academy of Science
Kende u. 13-17
H-1111 Budapest
E-mail: zoltan.fazekas@sztaki.hu

Dr.Alexandros SOUMELIDIS, MEEng. PhD.
Computer and Automation Research Institute
Hungarian Academy of Science
Kende u. 13-17
H-1111 Budapest
E-mail: soumelidis@sztaki.hu

HISTORICAL DATA BASED WIND SEGMENTATION

Eva GELNAROVÁ, Petr KRUPANSKÝ, Jiří SVOBODA, Petr MEJZLÍK, Tomáš NEUŽIL

Abstract: The 4D trajectory concept requires the precise guidance of aircraft along a predefined flight path. In particular, for the concerned longitudinal trajectory predictions are necessary, among others, reliable weather data, even in case the weather forecast is not available. In some applications, it might be useful to know the typical weather trends along the trajectory. This contribution describes the methodology how historical upper wind trends along a trajectory may be constructed. The analysis of effects influencing the segmentation of winds along the selected trajectory has been done using cluster analysis. The probability of transitions between wind magnitude states along the selected trajectory has been investigated using transiogram. As a data source, the stored predictions of U.S wind magnitudes and bearings from Rapid Update Cycle (RUC-2) have been used.

Keywords: Wind magnitude and bearing, transiogram, data clustering, the Rapid Update Cycle (RUC).

1 INTRODUCTION

The 4D trajectory concept requires the precise guidance of aircraft along a predefined flight path. In particular, for the concerned longitudinal trajectory predictions are necessary, among others, reliable weather data, even in case the weather forecast is not available. The recently proposed enhanced weather algorithm, for which Honeywell is seeking patent protection, combines the on-board sensor wind measurements with wind trends discovered by analysis of historical upper wind data. This paper presents the possible methodology for derivation of such trends and indicates which factors influence the segmentation (daytime or effect of the season are investigated). The particularity of this weather analysis is that the processes in the atmosphere are not analysed, instead the weather patterns are described from the aircraft perspective.

2 METHODOLOGY

2.1 Weather data source

The RUC-2 database [1] is a reliable source of forecasts of atmospheric parameters such as temperature, humidity, wind magnitude from the north and east. These are available on the 40km*40km lattice (Lambert conformal projection), at 37 pressure layers and at one-hour intervals. The database of RUC data available at Honeywell was created by the periodic downloading of the up-to-date, publically available weather forecasts. The data used for the analysis cover 406 days in the period from 21st August 2007 to 29th September 2008 and the percentage of missing records in the sample trajectory is approximately 10-20%, the records are missing in blocks due to download interruptions.

2.2 Selected altitude and trajectory for the study

The aircraft performance in cruise phase of the flight is affected mostly by the wind magnitude and bearing [5]. Thus this analysis is focused on these two parameters at pressure altitude 200HPa, which corresponds to height of 38 662 ft in the ICAO international standard atmosphere. The analysis was conducted on the sample of RUC lattice intersecting the cruise phase of the flight plan (FP) presented at Figure 1. The selected trajectory contains the original waypoints of the FP and additional points of RUC lattice.

2.3 Statistical methods

The effect of the hour and day on magnitude for all waypoints (WPTs) was tested by the nonparametric Kruskal-Wallis test and Friedman test on days with complete records only.

The wind bearing was partly analyzed by means of circular statistics [2], partly the circular data were "unrolled" to the interval (-179°, +180°] and treated with the classical statistical methods. The circular one-way analysis of variance (ANOVA) was performed to investigate the effect of the day and daytime.

The seasonal patterns have been explored with the agglomerative hierarchical clustering (Ward's method, Euclidean distance) of monthly median values (records only from complete days) in all analyzed WPTs [3]. The number of clusters and clustering tendency has been explored with average Silhouette coefficient.

The wind magnitude was transformed into categorical variable (bin width 10 kts). The need for high order-Markov chains have been overcame with the usage of combination of magnitude states and artificial states indicating: increase, decrease and stability of wind magnitude between first two waypoints. Based on these states the transiogram [4] for the WPT1 was constructed. Transiogram is the graphical display of the transition probabilities between the states as a function of distance/lag.

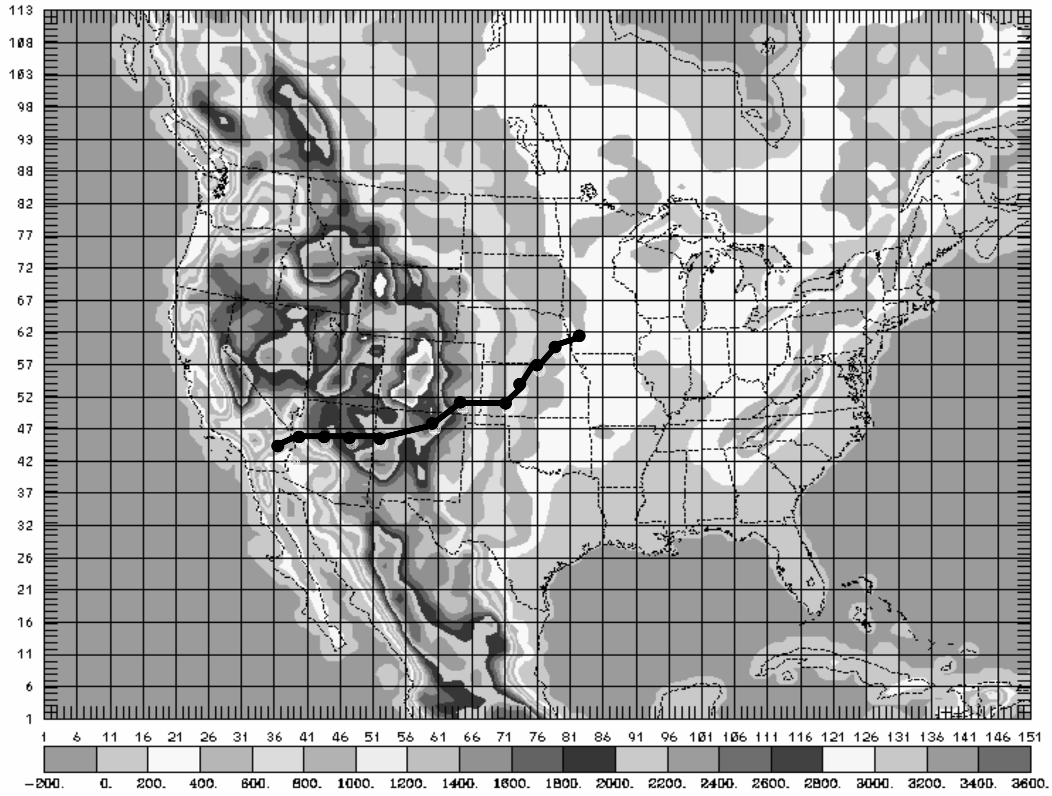


Figure 1 The cruise phase of selected flight plan. The additional RUC points are not plotted

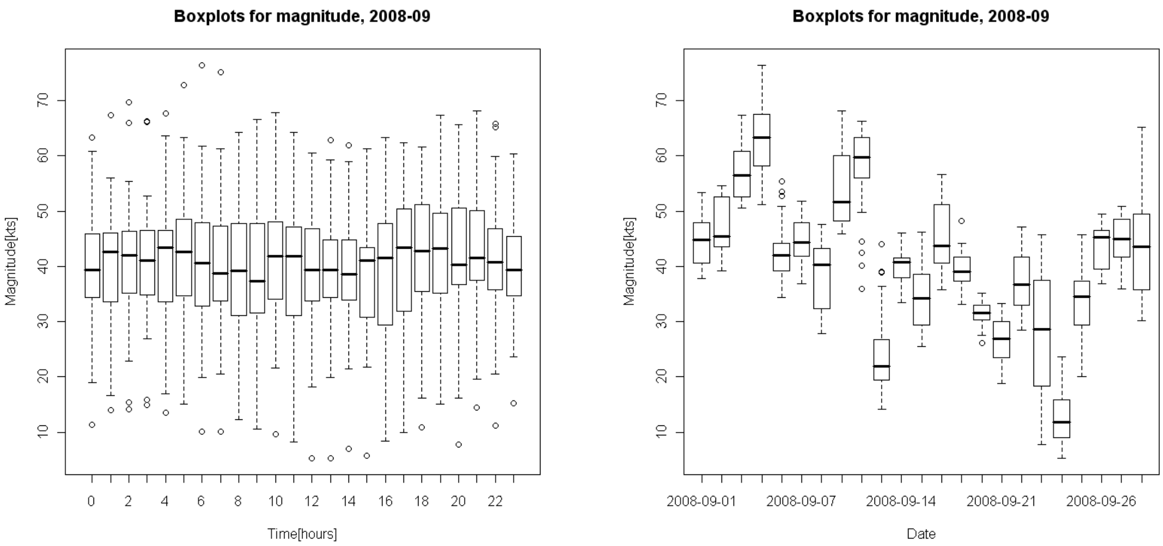


Figure 2 Magnitude at WPT1, September 2008

Left – boxplots for hours, right-boxplots for days (Only days with complete records.)

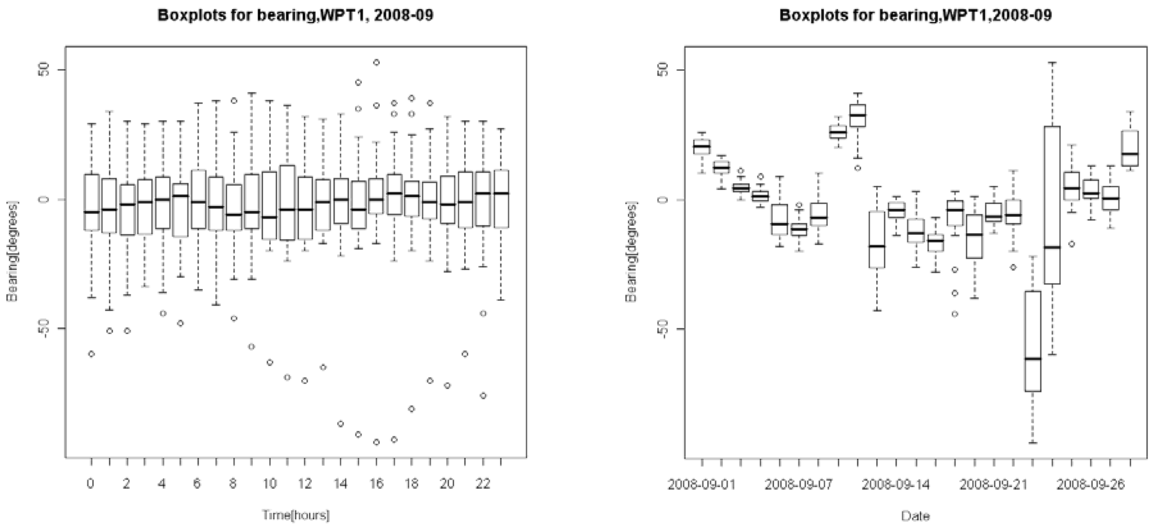


Figure 3 Bearing at WPT1, September 2008

Left – boxplots for hours, right-boxplots for days (Only days with complete records.)

3 RESULTS

3.1 Wind seasonal patterns – time series along the trajectory

3.1.1 Influence of the day-time

First, the influence of the day-time and the day on wind magnitude and bearing has been analyzed for all selected trajectory WPTs in separate months. There were significant differences among the days in all analyzed months and WPTs (all p-values of Kruskal-Wallis tests and circular ANOVA were smaller then 0.05) On the other hand, there were no significant differences in wind magnitude and bearing with respect to the time of day (all p-values larger then 0.05). The situation is apparent from boxplots for a selected WPT at Figure 2 and Figure 3. But the effect of the day time on wind magnitude may also be analyzed when the effect of the day is treated as the nuisance block effect. In this case, the effect of the day-time was for some months and WPTs significant (p-values of Friedman test are in some cases smaller then 0.05, results not shown). The results of Friedman test show that in some months, the hour wind magnitudes could follow a pattern within a day. But when hour-specific data are analysed together (without respect to the influence of day), there is no significant difference. As a result, the day-time effect has not been taken into account for wind magnitude and bearing.

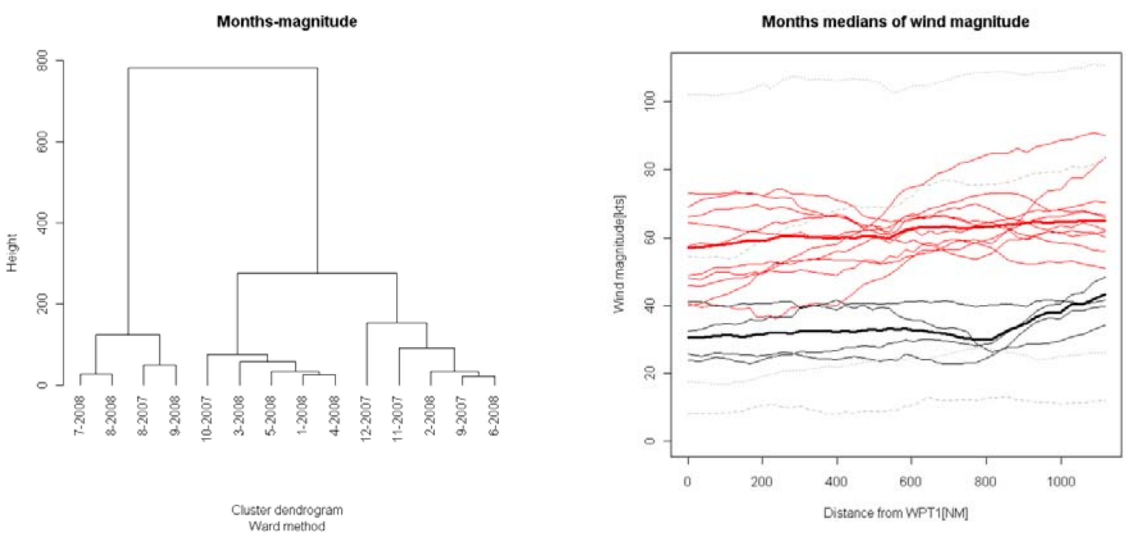


Figure 4 Finding patterns in month medians of wind magnitudes: left – dendrogram, right – The resulting clusters of months medians. The thin lines represent the months medians, the thick lines represent the medians of seasons. The grey lines are 5 resp. 95 quantiles of the data at selected season

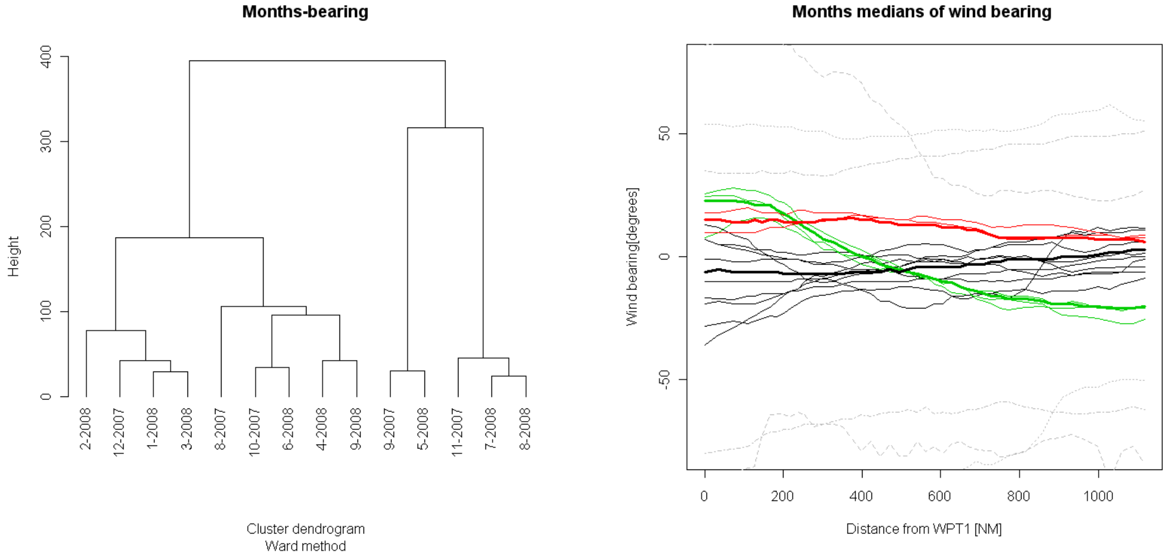


Figure 5 Finding patterns in month medians of wind bearing: left – dendrogram, right – The resulting clusters of months medians. The thin lines represent the months medians, the thick lines represent the medians of seasons. The grey lines are 5 resp. 95 quantiles of the data at selected season

Seasonal effect

To assess the seasonal effect, the medians for months and WPTs were computed. The clustering of the months into segments is based on the data from all 46 WPTs. The resulting dendrogram for wind magnitude (Figure 4 - left) shows the cluster of summer months. The resulting number of clusters was set to two (maximum of average Silhouette coefficient was 0.59 for 2 clusters). The average Silhouette coefficient is not high,[1] thus we can not count with well separate clusters.

Clustering median month values may be seen too rough, thus the clustering was also performed on median day values of wind magnitude for complete days. The maximal average Silhouette coefficient was only 0.35 for 2 clusters. Based on this clustering, the probability of being in a summer cluster may be computed as the proportion of days with “summer wind” in a 10 adjacent days.

The resulting dendrogram for wind bearing is shown at Figure 5 - left, the resulting number of clusters is set to 3. For 3 clusters the average Silhouette coefficient was the highest, but still only 0.49, thus with the clusters are not well separated, as may be also seen from Figure 5 - right.

3.2 Probabilities of transition between the states in dependency on a lag/distance

With increasing distance (lag), the conditional probability distribution of wind magnitude given the wind magnitude at initial WPT is independent of the wind magnitude at this point (graphs not shown).

To characterize spatial heterogeneity along the trajectory the transiogram was constructed. To show the characteristic behaviour, the transiograms for one particular WPT and selected wind magnitude 41-50 kts are presented at Figure 6. The transiogram for all data (Figure 6 top-left) shows that, when no additional information about wind dynamics is known, the most probable is the situation that wind magnitude stays at the same state. This is the situation of current wind prediction algorithm in FMS, when currently measured value is used for all WPTs of the trajectory. When the wind dynamics is known, this algorithm is no longer optimal. The probability of staying at the same state is not the highest one (Figure 6 top-right, bottom right).

4 DISCUSSION AND CONCLUSION

4.1 Comments on the methodology choice

In literature, the most suitable models regarded to the surface wind speed time series due to high autocorrelation are the Markov chain models [6,7]. However, the required outputs on our upper-wind analysis are specific, we are not interested in modelling of the most probable performance in time and thus the data are treated with other tools. The segmentation of time series is closer to the classical time-series decomposition and spectrum analysis. But because of missing values and nonstacionarity, the Fourier analysis was finally not used. Moreover, the extraction of the trend using moving-averages induced spurious frequencies (Yule-Slutsky effect [8]).

[1] The value of the *Silhouette coefficient* can vary between − 1 and 1.

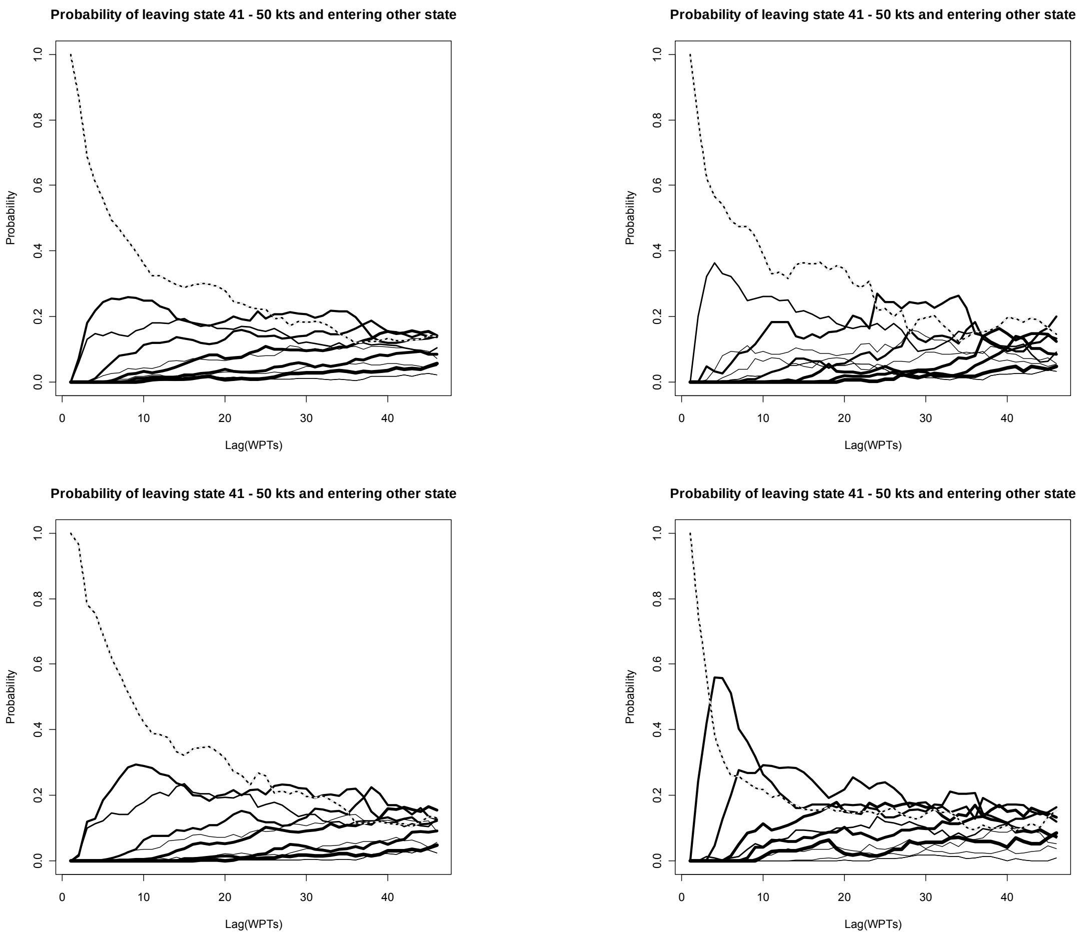


Figure 6 Transiograms - The probabilities of leaving the state 41-50 kts and entering other states (the thickness of the line correspond to wind magnitude) at WPT1are depicted for all data (top left), decreasing magnitude at WPT1(top right), stable (bottom left) and increasing (bottom right). Dotted lines represent the probability of staying at state 41-50kts

4.2 Results summary

For selected trajectory, the overall cluster structure in wind data is relatively weak. Important factors are the local spatial and temporal correlation structure. The wind magnitude ranges from 0 to 163kts, but in most cases the values are between 12 and 105 kts (detailed results not shown). The bearing value is mostly in a strip between -90 and +50 degrees (detailed results not shown).

The effect of the daytime has not been confirmed. The cluster structure in monthly medians has been detected for the magnitude and bearing, but especially in bearing case the structure is weak and must be treated with caution.

Because the available data covers only approximately one year, the resulting segments can not be generally understood as representatives of the typical seasonal behaviour. They reflect only the development in the selected year along the sample trajectory.

The presented results are the product of the study of selected trajectory and thus they are not exhaustive and valid generally for the whole airspace.

This work has been supported by Honeywell International, s.r.o.

References

[1] The Rapid Update Cycle (RUC): http://ruc.noaa.gov/.

[2] UPTON, Graham J.G, FINGLETON, Bernard: *Spatial data analysis by example.* Volume 2, Categorical and directional data, Wiley, 1989.

[3] TAN, P. N., STEINBACH, M., KUMAR, V.: *Introduction to data mining.* Addison-Wesley, 2006.

[4] LI, W.: *Transiograms for characterizing spatial variability of soil classes*, Soil Science Society of American Journal 71(3), 2007 pp. 881-893.

[5] ERASMUS - WP1 - D1.1 - V1.0- Air Trajectory Prediction.

[6] ETTOUMI F. Y., SAUVAGEOT H., ADANE A. E. H.: *Statistical bivariate modeling of wind using first-order Markov chain and Weibull distribution*, Renewable Energy, 2003, vol. 28, pp. 1787-1802.

[7] SAHIN, A. D., SEN, Z.: *First-order Markov chain approach to wind speed modelling*, Journal of Wind Engineering and Industrial Aerodynamics, 2001,Vol. 89, pp. 236-269.

[8] LEGENDRE, P., LEGENDRE, L.: *Numerical ecology*, Elsevier, 1989.

RNDr. Eva GELNAROVÁ
Honeywell International s. r. o.
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: eva.glenarova@heneywell.com

Dipl. Eng. Petr KRUPANSKÝ, Ph.D.
Honeywell International s. r. o.
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: petr.krupansky@honeywell.com

Dipl. Eng. Jiří SVOBODA, Ph.D.
Honeywell International s. r. o.
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: jiri.svoboda@honeywell.com

RNDr. Petr MEJZLÍK, Ph.D.
Honeywell International s. r. o.
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: petr.mejzlik@honeywell.com

RNDr. Tomáš NEUŽIL, Ph.D.
Honeywell International s. r. o.
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: tomas.neuzil@honeywell.com

ANALYSIS OF THE CHARACTERISTICS OF A PILOT IN THE AIRCRAFT FLIGHT

Rudolf JALOVECKÝ, Přemysl JANŮ

Abstract: The paper represents an attempt to analyze the properties of human flight in the aircraft. The approach is shown to create a replacement model of the behaviour of humans - pilot in the creation of a replacement model in terms of automatic control. There is not much information on how to access the analysis features of humans in terms of automatic control in history. Simulation of "response" on the human impulse input is an essential building block in the design of circuit auto flight aircraft control. Particular attention is paid to simulations of pilot - aircraft group in curbing the fast oscillations aircraft.

Keywords: Aircraft automatic control, Autopilot, Aircraft, Matlab, Simulink.

1 INTRODUCTION

To describe human behaviour by suitable mathematic functions, which are possible to further use to particular analysis, is very difficult procedure. During getting leastwise approximate replacement block scheme of regulation circuit of human behaviour is supposed his behaviour in occurrence case of exactly defined stimulations. Orientation reflections about the possible replacement structures of a human pilot from automatic regulation view point appear too little in the literature [2], [3]. Models which characterize replacement of pilot scheme are mentioned at [1]. With development of computer technology, when advanced simulation systems are available [8], is possible to assemble replacement models of human behaviour and then to prosecute their analysis more comfortable. Questions of human behaviour - the pilot and his definition from the perspective of the automatic regulation with usage of simulation tools dedicates article [4], which would be understood as an introduction to analysis of human - pilot behaviour from automatic regulation view point with usage of modern simulation systems.

For this paper authors chose other way of imaging waveforms of simulated quantities. Imaging of waveforms at 2D graphs presents conventional standard with sufficient survey of required waveforms. Interesting and more clearly is imaging of the waveforms at 3D graphs, when alternation of chosen simulation parameter is carried out on the third axis. In MATLAB – SIMULINK programme command MESHGRID (time, time constant) is used for creating grid of 3D graph. Command MESH(time, time constant, response) is used for depiction of 3D graph. Data before values depiction of response must be saved to two dimensional matrix (time, time constant), with matches to third axis of 3D graph (response).

All sequentially presented graphs use these possibilities of MATLAB environment. Specified third parameter is time constant alternation of one of inertial member describing variation of human characteristic. Certain disadvantage of black and white print at papers is lower clearness against depiction in electronic – colour form.

2 MODEL OF HUMAN CHARACTERISTICS SIMULATION

Authors of this paper have already published in [5] and [6] initial approaches to creation of models simulating human characteristics. Complete information about possible behaviour of human during different situations would present greatly extensive model, which would have been changing according to actual abilities of human. Human behaviour model is approached to a lot of simplifications during its creating.

Defined types of human-pilots are used from mentioned papers [5] and [6]. For other simulated cases human-pilots type „A“ and „C“ are taken into account. There are linear models with transport delay characterized by transfer functions.

Pilot type „A“

$$F_{(p)} = \frac{Y_{(p)}}{X_{(p)}} = K \frac{(T_3 p + 1)}{(T_1 p + 1)(T_2 p + 1)} e^{-\tau p} \qquad (1)$$

where:

- K - increasing of force on the steers in relation to their deflection (from 1 to 100)
- T_1 - time constant, i.e. reaction ability to rate of change of input signal (5 to 20s) (prediction time constant)
- T_2 - dynamics properties of the pilot power members components (0.1 to 0.2s) (neuromuscular time constant)
- T_3 - integrating time constant, i.e. pilot’s ability to realize varying activity (0.2 to 1s)
- τ - transfer delay (0.1 to 0.4) (time of pilot reaction).

Pilot type „C“

$$F_{(p)} = \frac{Y_{(p)}}{X_{(p)}} = K \frac{1}{(T_2 p + 1)} e^{-\tau p} \qquad (2)$$

With mentioned “types” models of human pilot’s dynamics, the simulation of responses to single input impulse according to Figure 1 has been provided.

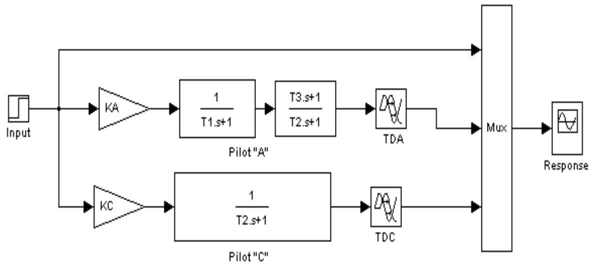


Figure 1 Block scheme of man characteristics simulation

Values of time constants for instance of simulations are mentioned at Table 1 [7].

Table 1 Ranges of time constants for definition of human pilot characteristic

T1 = 5 ÷ 20 s	T2 = 0.1 ÷ 0.2 s	KA = 2	TDA = 0.15 s
T3 = 0.2 ÷ 1 s		KC = 1	TDC = 0.40 s

There are even graphic outputs from simulations of human response on „unit“ input signal at references [5] let’s say [6].

3 HUMAN ABILITIES DURING HEADING CONTROL OF AN AIRCRAFT

Pilot performs various structures of tasks during aircraft control. Heading control of an aircraft is e. g. one of these tasks. Analysis of possible characteristics and consequently human - pilot skills during heading control of an aircraft is possible to implement with usage of simulation environment MATLAB – SIMULINK. Basic simulation scheme is depicted in the Figure 2, which contains submodul of aircraft movement in lateral plain with oscillation dumper at aileron channel (Figure 3) and both human models according to equation (1) and (2) at submoduls, which have nearly same shape as two parallel signal channels in the Figure 1.

Block types „Sink Block Parameters“ are stepwise assembled under each other at the right side of block scheme in the Figure 2, which make possible sequentially filling of result matrixes of simulation for selected third parameter. Simulation programme is written in MATLAB editor and simulink model is cyclically called (by function **sim()**) always for actual values of simulated plot.

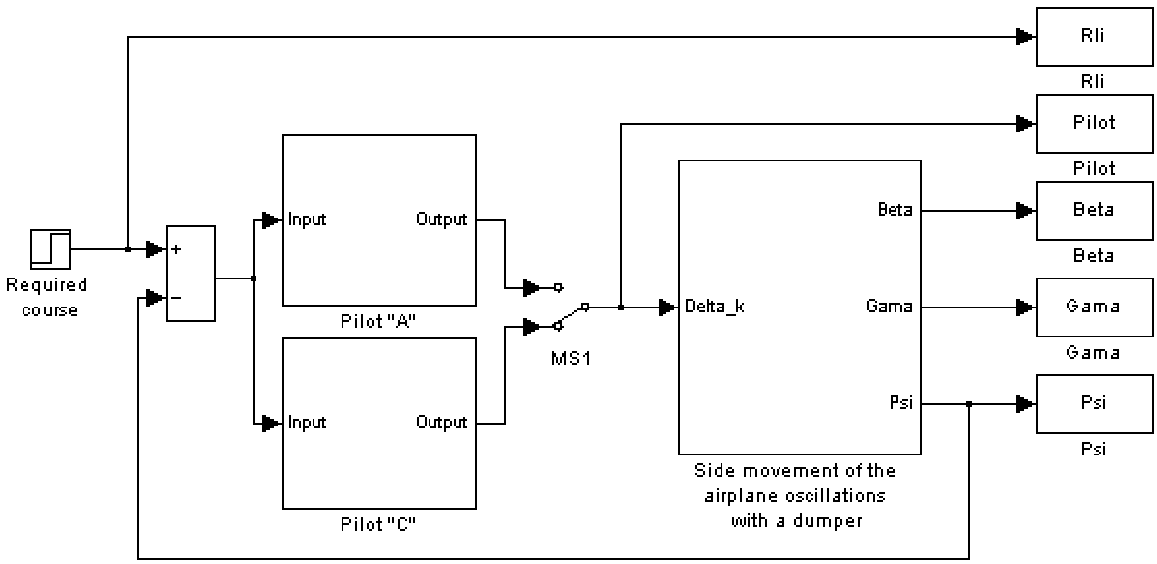


Figure 2 Simulation block scheme of human behaviour during heading control of and aircraft

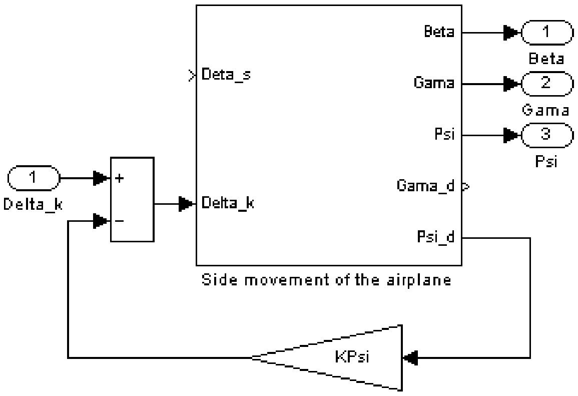


Figure 3 Block scheme of lateral movement model submodul with oscillation dumper at heading channel

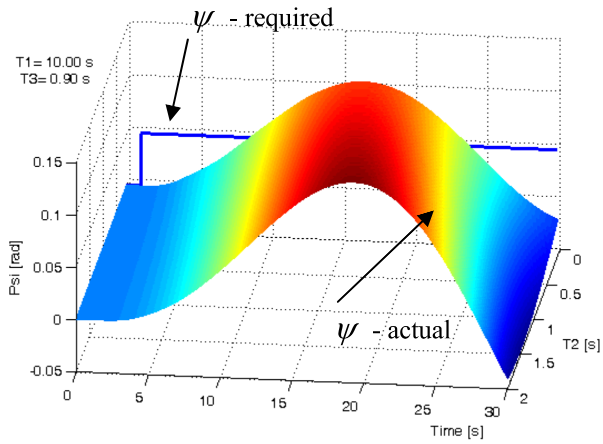


Figure 4 Character of heading control of aircraft flight by pilot type „A"

3.1 Analysis of selected simulation control results with pilot type "A"

During simulation of lateral movement control by pilot with transmission characteristic type "A" is gathered that this pilot is nearly not able to realize heading control of aircraft flight. In the Figure 4 is possible to see required heading angle of aircraft flight (full line from 0 value to 0.05 radian value at time 1s) and time variation of real heading angle with synchronous time constant T_2 variation at the replacement pilot model. This inability to control required flight heading inheres at existence of real value of transport delay – pilot's reaction time and probably in replacement model of human pilot, that doesn't fully describe his real abilities. This reality is furthermore seen in Figure 5, that represents action quantities from rudder deviation, thus pilot's reaction on stimulus of heading variation of aircraft flight. From waveform is possible to deduce, that pilot is getting late during variation of aircraft heading and with time increasing of regulation is making rudder deviation larger and larger. With simulation continuing emphatic and technical non-productive rudder deviation would come, because of linearized model of aircraft movement at space and linearized pilot's model is used.

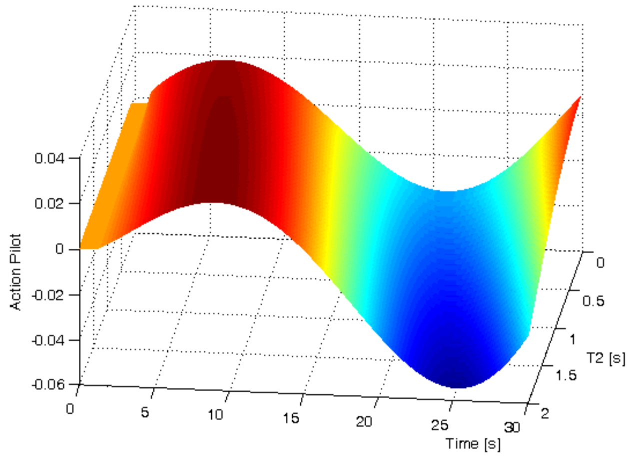


Figure 5 Waveform of pilot's type „A" intervention during heading control of aircraft flight

3.2 Analysis of selected simulation control results with pilot type "C"

During simulation of lateral movement control by pilot type C transmission characteristic is dramatically different situation. In the Figure 6 is again depicted required heading variation of aircraft flight and time variation of real heading of aircraft flight with synchronous time constant T_2 variation. From simulation is clear, that pilot in this case is able to realise heading control of an aircraft. Control quality is changing with human characteristics - neuromuscular time constant. Time constant variation determinates reaction ability on stimulus. The smaller is constant value, the faster is pilot in reaction on change of aircraft movement. Human physiology doesn't make this value to be zero or very small number. However this value negatively affects aircraft behaviour in form of "vibrating" heading value during conscious control its change.

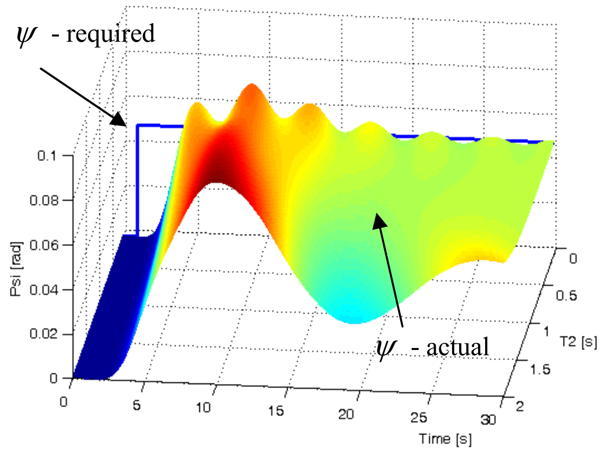


Figure 6 Character of heading control of aircraft flight by pilot type „C“

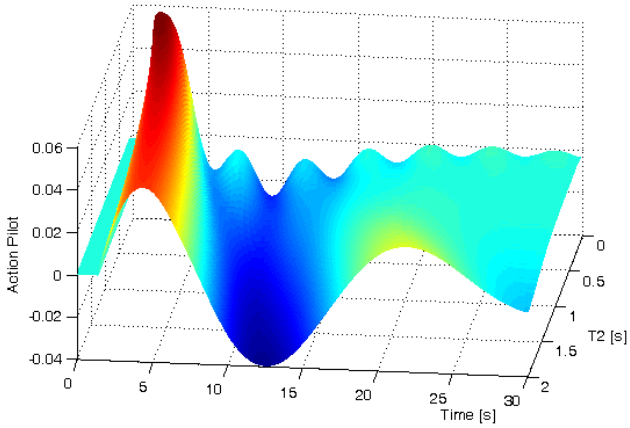


Figure 7 Waveform of pilot type „C“ intervention during heading control of aircraft flight

It is interesting to observe intervention of the pilot during this control. It is time variation signal of rudder deviation mentioned in the Figure 7. If time constant T_2 is small, time variation of rudder deviation is very large. Controlled parameter – flight heading is very often "overshot". But pilot is able to control aircraft and to lead it to required heading in relatively short time. If time constant T_2 is large, pilot doesn’t overshoot required heading very fast and required heading reach in the same time instant (slightly slower) as in the case of small time constant T_2. In addition in this case regulation is more fluently, thus mechanical parts loading of aircraft, attachment of wing to fuselage is extensively smaller.

4 CONCLUSION

In the paper is mentioned possible way of behaviour pilot’s simulation during his trial of lateral movement control of an aircraft, clearly heading change of aircraft flight with aileron deviation. It is necessary to give stress on all used linearized models and then in the case of aircraft on simplified model of side movement with oscillation dumper in aileron channel. Orientation simulation showed, that is possible to use computer technology to analyse behaviour and supposed human characteristic. Range of the paper can’t cover whole problem of human characteristic simulation from automatic regulation view point.

The work presented in this paper has been supported by the Ministry of Defence of the Czech Republic (Research Plan No. MO0FVT 0000403).

References

[1] ZIHLA, Z.: *Automatic control of aerospace,* Part 1. (in Czech). Provisional textbook, U-1220/I, Brno: Military academy, 1987. 395p.

[2] HAVLIKOVA, M.: *Models of driver’s behaviour* (in Czech). [online] Elektrorevue, Volume 22 2004, [cited 2009-1-20]. Available from: <http://www.elektrorevue.cz/clanky/ 04022 /index.html>

[3] HAVLIKOVA, M.: *Human reliability and hazard of his failure in systems MMS* (in Czech). Automatization, Volume 49, No. 7-8, p. 461-462.

[4] JALOVECKY, R.: *Man in the Aircraft Flight Control System* (in Czech). In: Measurement, Diagnostics and Dependability of Aircraft Systems, Brno : University of Defence, 2008, p.67-74.

[5] JALOVECKY, R.: *Man in the Aircraft Flight Control System*, Advances in Military Technology, Vol. 4, No. 1, Brno : University of Defence, June 2009.

[6] JALOVECKY, R., JANU, P.: *The features model of a pilot during of dumping the rapid oscillations in an aircraft.* In: JALOVECKY, R. and STEFEK, A. (ed.) Proceedings of the International Conference on Military Technologies 2009. Brno : University of Defence, 2009, p. 333-344. ISBN 978-80-7231-649-6.

[7] JALOVECKY, R.: *Onboard systems flight control I.* (in Czech). Brno : University of Defence, 2008. 82p.

[8] SOPATA M., SOTÁK M., BRÉDA R.: Modelling and simulation from verification and validation viewpoint (in Slovak). In: New trends in aviation development: Proceeding of the 6 th International conference, Kosice : Military Aeronautical Academy General´s M. R. Stefánik, 2004, ISBN 80-7166-050-7, p. 28-33.

Assoc. Prof. Dipl. Eng. Rudolf JALOVECKÝ, CSc.
Faculty of Military Technology
University of Defence
Kounicova 65
612 00 Brno
Czech Republic
E-mail: rudolf.jalovecky@unob.cz

Dipl. Eng. Přemysl JANŮ
Faculty of Military Technology
University of Defence
Kounicova 65
612 00 Brno
Czech Republic
E-mail: premysl.janu@unob.cz

OBSTACLE DETECTION SYSTEMS USING EO CAMERAS

Ondřej KOTABA

Abstract: This paper describes the possible approaches to the two distinct tasks of the sense&avoid capabilities using the visual sensors. For the obstacle mapping, we discuss the feasibility of stereo reprojection, optical flow, and structure from motion approaches. For the traffic collision detection and avoidance, the classification approaches, stereo reconstruction, structure from motion, and heterogeneous systems are being evaluated.

Keywords: Visual sense and avoid, vision-based navigation, sense and avoid, S&A, object avoidance, traffic collision avoidance, obstacle detection.

1 INTRODUCTION

One of the important requirements posed for the unmanned aerial vehicles (UAVs onwards) is their capability to autonomously sense threats and to avoid them, commonly referred to as Sense & Avoid (S&A onwards) capabilities.

The set of specific requirements does change depending on the expected operational scenarios of the UAV, its parameters and other aspects. For example, one can expect completely different situations to pose threats for the small, low altitude, low endurance UAVs, than for the high altitude flying, high endurance heavy fliers. Needless to say, the mentioned requirements are vital for the ability of the autonomous systems to coexist in the civil airspace together with the manned aircraft.

In general, it is possible to separate the S&A capabilities into two major tasks, each having a different objectives and means to achieve them. Not all the UAVs need to be able to perform both:

- Static obstacle mapping & avoidance - capability of the UAVs to see the surrounding, mostly but not exclusively static, objects and to plan the trajectory accordingly.
- Traffic collision avoidance – capability of the UAVs to sense the surrounding traffic that either is or might become flying on the collision course, further the capability to process that information and to perform the evasive manoeuvre, if needed, in accordance with the valid flight rules.

The following paragraphs of this paper detail the possible approaches of using the visual information, gathered primarily by conventional electro-optical sensors, to fulfil the tasks described above.

2 ELECTRO-OPTICAL SENSORS

The category of electro-optical (EO) sensors can be defined as any sensors that emulate the function of the mammal optical apparatus by transforming the ambient or artificially induced electromagnetic radiation reflected from the objects to electronic signals. Examples of such systems are regular visible light cameras, line scanning cameras, infrared and ultraviolet cameras, intensified imaging devices, bolometers and others. In the broader meaning of the term, LIDARS, especially the flash lidars then, might be understood to be the members of the EO family as well, but for the purpose of this paper, we leave them out as a separate category.

The main advantages of the electro-optical sensors, compared to others, are further detailed in [1] but a short list might include:

- Extremely mature technology;
- Broad usage, resulting in a relatively low cost;
- Very small spatial and weight load requirements;
- Naturally passive – low power consumption;
- Long time in use, ~50 years of associated algorithms' research.

The main disadvantages are then the inherent incapability to provide the range information and a relatively high sensitivity to weather conditions. One more disadvantage is a relatively low dynamic range but an extensive research is being conducted and first products with logarithmic dynamic scale are available on the market already [1].

3 OBSTACLE AVOIDANCE

3.1 Overview

The task which we call the obstacle mapping & avoidance refers to the capability of autonomous vehicle to observe its surroundings and to plan its own flight-path the way that is not in collision with any obstacles. The main problem of using the EO sensors for this task is their dependency on the amount of salient points in the observed scene. While there are many approaches how to detect the range to an object (described below), in any case this needs to be somehow distinct from it surroundings and/or background. There need to be some "clues" to lock to ... there is little to analyze in the scene of white wall. In other words, we say that

(1) A sensing system relying only on the 2D sensors provides only a sparse data-set, as opposed to the dense data obtained by other sensors.

One of the ways to overcome this problem is by taking an assumption that a local area of a scene without any salient features is part of a continuous surface. In such a case, using a supplementary sensor, for example single-point laser range finder, can be used to obtain the range information for several points of the continuous surface and to perform the interpolation for the rest of the surface. (See Figure 1 for an example.)

3.2 Stereo pair

Practically the most common approach to derive the range information from plane imaging sensors, like cameras, is a utilization of the stereo reconstruction. As described in the paragraph 3.1, this approach delivers the sparse depth information. Indeed, there are a number of algorithms performing a dense stereo reconstruction, but it is usually assumed that the behaviour of the surface in between the points is specific. For example, most common assumption is the planarity of the continuous surfaces. While this is completely acceptable for visualization purposes, the approach is not rigid enough for the flight trajectory planning purposes.

The example of such reconstructed scene information is depicted on Figure 1.

3.3 Optical flow

The optical flow algorithms [6], [7] estimate the per-pixel motion vectors in the image between consecutive frames. While not capable to derive the distance to the objects in the scene, the optical flow algorithms are able to return the rational distance of the objects, thus theoretically allowing constructing a map of the scene excluding its scale, given the vehicle is moving. The biggest disadvantage of the algorithms is their relatively high computational complexity requiring HW implementation for any real time applications. The example depicting results of the Lucas-Kanade optical flow estimation [6] is on Figure 2.

3.4 Structure from motion

Similar to the stereo reconstruction, as well to the synthetic aperture radar approach, is the method using only one camera and its successive frames. It is important to say, however, that the mentioned solution would again provide a sparse dataset. To obtain the dense map, we propose to complement such a system by additional sensor to fill the gaps between the observed points and then using the interpolation without having to assume the planarity of interpolated surfaces.

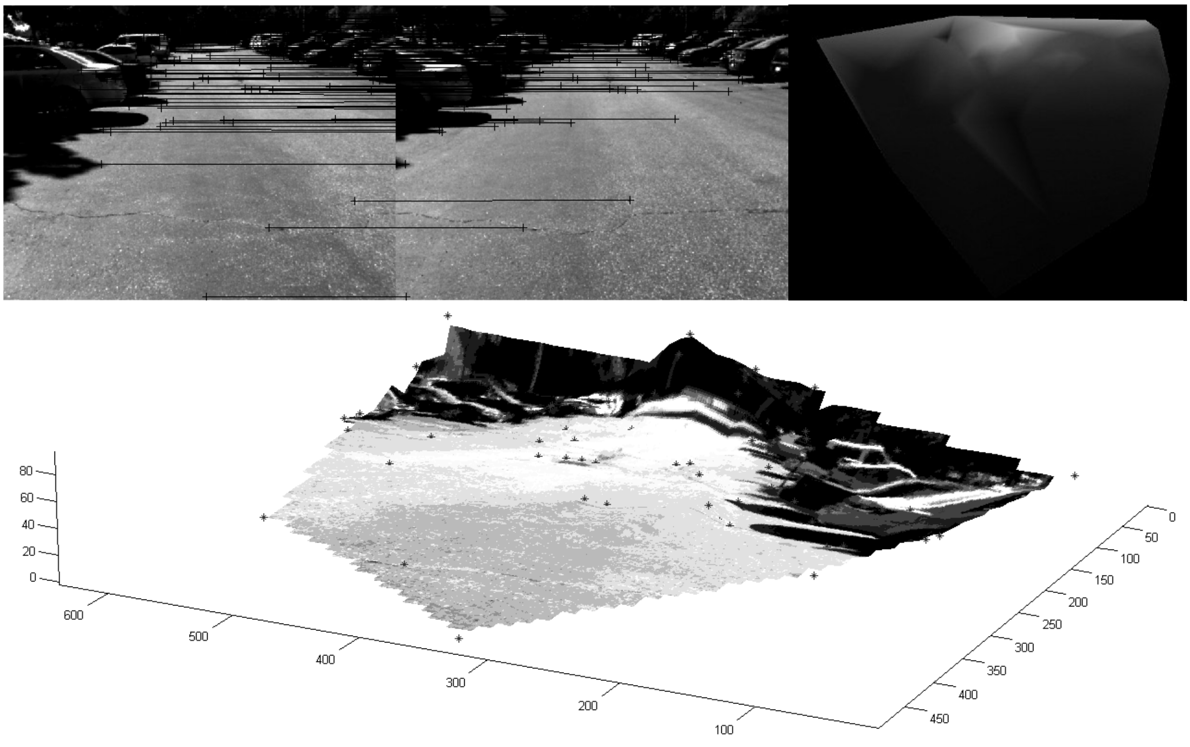


Figure 1 Figure of the stereo reconstruction example, along with the linear interpolation of surfaces between the feature points
Top left: Left and right image respectively, along with the point correspondences outlined. Top right: Resulting interpolated depth map. Bottom row: Original left picture mapped as a texture on the interpolated surface reprojected from feature points. Camera lies at (320, 480, 1), units on axes are in pixels. Note the distortion at the far end of the scene, across the camera, caused by decreased depth resolution of the stereo rig.

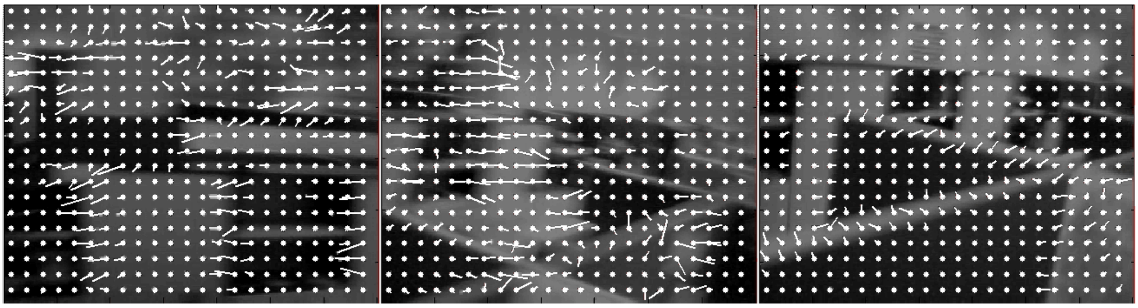


Figure 2 Example images of the optical flow calculated by the Lucas-Kanade method
The motions are following (L to R): translation leftwards, translation rightwards, clockwise rotation + translation rightwards

4 TRAFFIC COLLISION AVOIDANCE

4.1 Requirements

The main requirements for the system capable to detect any possible airborne objects are:

- Minimal range of the detection - depending on the maximum mutual speed of the objects and on the time needed to perform the evasive manoeuvre;
- Field of view – given by legislation [4];
- Reliability and false detection rate;
- Latency.

It is important to emphasize, that while many aim to detect the range to the possibly colliding objects, it is not the primary objective for the system.

(2) The primary objective is to detect the time to collision,

along with the mutual trajectory in some space, not necessarily Cartesian. For example, having the knowledge of an object on 30 heading and 5 degrees elevation with a specific estimated time to collision is completely enough to perform an evasive manoeuvre, even without knowing how fast it flies and how far it is.

4.2 Monocular camera

One of the first approaches that were used for this task in the prior art were the classification-based approaches [8], [9], [10]. While these approaches basically work, they are not descriptive enough, meaning that the classification benevolent enough to detect all planes with high enough accuracy is bound to trigger an extreme number of false alarms on the other hand [8]. One of the biggest challenges when detecting incoming airborne threat is, that

(3) Objects on a collision course appear stationary against the background in the infinity.

Given the above statement, while it is possible to use the motion detection algorithms directly to detect distant airborne objects,

(4) they tend to detect only the objects that do not pose a threat.

4.3 Stereo pair

While using the stereo set of cameras might appear to be an appealing option, there are problems with this. The main problem is the depth resolution of the stereo rig. If we assume that the sub-pixel processing is not used (which in reality should be used, indeed, for example the SURF [11] and SIFT [12] description and matching algorithms inherently use sub-pixel correction, as well as the sub-pixel cross-correlation), the depth resolution of the stereo set is a function of

- Distance;
- Observation angle;

and depends mainly on the following parameters:

- Stereo rig baseline – a distance between the two cameras;
- Focal length;
- Pixel size;

according to the following equations:

$$\bar{r} = v\tan(\zeta) - d_2$$

$$\zeta = \underline{\delta} + \arctan\left(\frac{d_2}{v}\right)$$

$$d_1 = \varepsilon\cos(\underline{\alpha})$$

$$v = \varepsilon\sin(\underline{\alpha})$$

$$\underline{b} = 2\varepsilon$$

$$\underline{d} = d_1 + d_2$$

further resulting in

$$\bar{r} = \frac{\underline{b}}{2}\sin(\underline{\alpha})\tan\left(\underline{\delta} + \arctan\left(\frac{\underline{d} - \frac{\underline{b}}{2}\cos(\underline{\alpha})}{\frac{\underline{b}}{2}\sin(\underline{\alpha})}\right)\right) - \underline{d} + \frac{\underline{b}}{2}\cos(\underline{\alpha})$$

Where $\bar{r}$ is a resolution (m) at a given point, $\underline{d}$ and $\underline{\alpha}$ are the distance (m) and observation angle (rad) of the point, $\underline{\delta}$ is the angular span (rad) of each pixel and $\underline{b}$ is the camera rig baseline (m). Note that the effect of the lens field of view is already included in $\underline{\delta}$. See Appendix 1 for a diagram.

For the exemplary values of 1 m baseline, 112 μrad pixel span (equivalent approximately to the 640x480 camera with 50mm lens), the achieved resolution can be seen at the Figure 4. As you can clearly see, the 1% resolution precision is obtained approximately 50 meters far. At 100 meters, the depth resolution is already over two meters and at 200 meters, it is already around 10 meters! If we want to achieve the resolution of 5% of a distance at a distance of 2000 meters (considered reasonable for the avoidance manoeuvre) the required baseline between cameras would have to be at least 5.2 meters with the given camera parameters, which is not applicable for small UAVs. Indeed, cameras with smaller pixel size should achieve better results but would be larger in size, on the other hand, not even mentioning the fact that the horizontal field of view, as required by several sources, needs to be at least 220 degrees.

For that reason, but not only, the system of horizontally scanning vertically mounted line cameras seems to be a valid approach, but the rotating rig of two linescan cameras would be again too bulky for many of today's low-cost UAVs.

Using the sub-pixel processing, it is possible to enhance the depth resolution by approximately one order but that still is not enough to estimate the distance to the airborne objects.

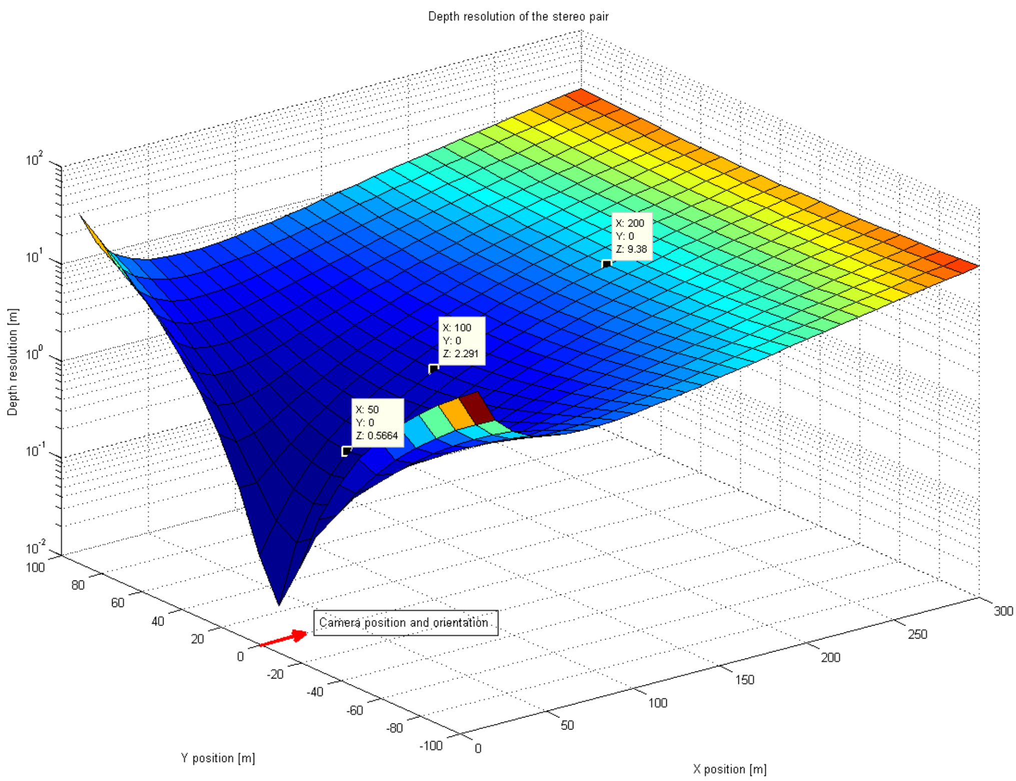


Figure 3 The depth resolution of the stereo camera pair with 1 meter baseline, 640x480 resolution of each camera and lens focal length equivalent to 50mm. Camera is positioned at the origin and pointing at the positive direction of the X axis

4.4 Structure from motion

Again, as well as for the obstacle mapping, a solution might be brought by the structure from motion approach. In this case, only one camera, or equivalently multiple cameras pointing in different directions to span over the required 240 degrees of horizontal field of view, would be used to capture multiple frames from different locations. As this approach is working on a similar principle as synthetic-aperture radar [13], we can thereby refer to it as a synthetic-baseline multi-frame 3D reconstruction [4], [5], and [3]. The largest disadvantage of this approach is the need for the UAV to modify its trajectory so that it captures the scene from a slightly different orientation, perpendicular to its flight trajectory. This brings a need for the UAV to perform a sinusoidal or elliptical motion changes and slightly increases the fuel consumption. However, along with the approach described in the paragraph 4.5, this method might yield the required performance while keeping the mass of the S&A system small.

4.5 Multi-UAV cooperation for S&A

One of the other ways how to achieve a larger baseline between the cameras is the possibility to operate a set of two and more UAVs simultaneously, each in proximity to the others, and to transmit the image information between them using a radio link. The GPS time signal can be used for the camera triggering and thus to achieve the time -synchronous image acquisition. Of course, the acquired images need to be transformed to match the camera projection planes and so precise navigation solution is again a critical aspect of the system. The rather nice advantage of the multi-UAV system containing three and more UAVs is, that if the distance between the vehicles is not minimal, only one of them at most can be considered to be on the collision course with the threat airborne object and thus the statement (3) only applies to this one, allowing the others to detect the moving airborne objects by performing the simple motion analysis of the scene from successive frames. See Figure 5 for an example of motion detection method usable for moving cameras.

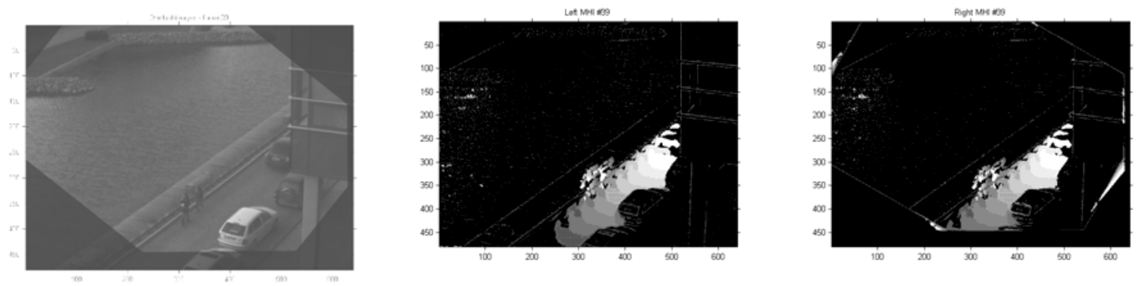


Figure 4 Example of the motion detection from two non-calibrated cameras

a) Automatic estimation of the fundamental matrix [14] is performed to "overlay" the images from two independent cameras. b), c) Pair of the MHIs (Motion History Images) [15] representing the areas in the scene with motion. Note the car passing, as well as the two walking persons. Residual edges are an effect of the transformation imprecision and could be avoided by simple Gaussian blurring.

Problem arises with the reliability of such multi-UAV system as the probability of a failure is much worse than the one of the stand-alone system.

5 CONCLUSIONS

In this paper, we have discussed several approaches to achieve the S&A capabilities. Based on our discussion we believe the viable way of performing both the obstacle mapping and traffic collision avoidance, can be achieved using the structure from motion approach; that is by 3D reprojection from arbitrary frames of the same camera using the precise trajectory information provided by the navigation system. Currently, we believe that other approaches, while appealing at the first sight, are less suitable for the requested goals with the performance and system parameters required.

6 APPENDIX 1 – DEPTH RESOLUTION OF THE STEREO CAMERA PAIR WITH COLLINEAR OPTICAL AXES

$$\bar{r} = v\tan(\zeta) - d_2$$

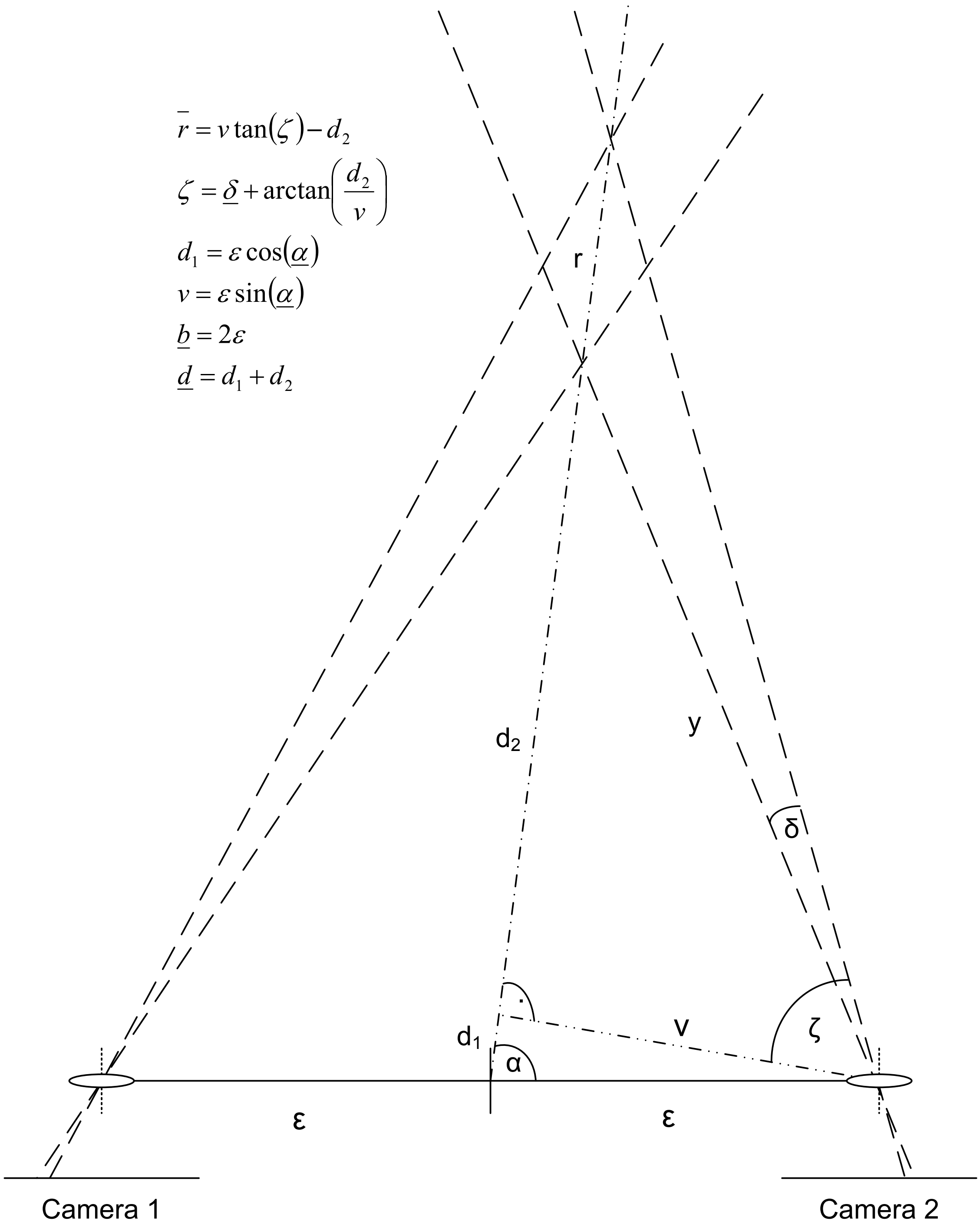
$$\zeta = \underline{\delta} + \arctan\left(\frac{d_2}{v}\right)$$

$$d_1 = \varepsilon\cos(\underline{\alpha})$$

$$v = \varepsilon\sin(\underline{\alpha})$$

$$\underline{b} = 2\varepsilon$$

$$\underline{d} = d_1 + d_2$$



$$\bar{r} = \frac{\underline{b}}{2}\sin(\underline{\alpha})\tan\left(\underline{\delta} + \arctan\left(\frac{\underline{d} - \frac{\underline{b}}{2}\cos(\underline{\alpha})}{\frac{\underline{b}}{2}\sin(\underline{\alpha})}\right)\right) - \underline{d} + \frac{\underline{b}}{2}\cos(\underline{\alpha})$$

References

[1] PhotonFocus, AG: http://www.photonfocus.com.

[2] KANOVSKY, P., KOTABA, O.: *Sense & Avoid Thoughts*, Proceedings of 1st National Conference Civil Unmanned Aerial Systems 2008, UAS 2008, Prague : Czech Republic, November 12-13, 2008.

[3] LUKAS, J., KOTABA, O.: *GPS-denied Navigation concept*, Proceedings of 1st National Conference Civil Unmanned Aerial Systems 2008, UAS 2008, Prague, Czech Republic : November 12-13, 2008.

[4] KANOVSKY, P., KOUKOL, O.: *European UAS legislation*, Proceedings of 1st National Conference Civil Unmanned Aerial Systems 2008, UAS 2008, Prague, Czech Republic, November 12-13, 2008.

[5] HARTLEY, R. I.: *In Defence of the 8-point Algorithm*, Proceedings of the Fifth International Conference on Computer Vision, 1995.

[6] LONGUET-HIGGINS, H. C.: *A computer algorithm for reconstructing a scene from two projections,* Nature, vol. 293, 1981.

[7] DELLAERT, F., SEITZ, S., THORPE, C., and THRUN, S.: *Structure from Motion without Correspondence*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2000.

[8] HARTLEY, R., ZISSERMAN, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, 2003.

[9] LUCAS, B. D., KANADE, T.: *An iterative image registration technique with an application to stereo vision*, 1981, Proceedings of DARPA Imaging and understanding workshop, pp. 121-130, Washington D.C., USA, April 1981.

[10] HORN, B. K. P., SCHUNCK, B. G., *Determining optical flow*. Artificial Intelligence, vol. 17, pp 185-203, 1981.

[11] UTT, J., McCALMONT, J. F., and DESCHENES, M.: *Development of a Sense and Avoid System*, AIAA Infotech Aerospace, Arlington, VA, USA, September 26-29 2005.

[12] MCGEE, T. G., SENGUPTA, R., and HEDRIC, J. K.: *Obstacle Detection for Small Autonomous Aircraft Using Sky Segmentation*, In: Proc. of the IEEE International Conference on Robotics & Automation, Barcelona, Spain, April 18-22 2005.

[13] McCALMONT, J. F., UTT, J., DESCHENES, M., and TAYLOR, M. J.: *Sense and avoid Technology for Global Hawk and Predator UAVs,* In: Andresen, B. F. and Fulop G. F. ed. Infrared Technology and Applications XXXI., Proc. of the SPIE, Vol. 5783, pp. 684-692, 2005.

[14] BAY, H., TUYTELAARS, T., VAN GOOL, L., SURF: *Speeded Up Robust Features*, in Proceedings of 9th European Conference on Computer Vision Graz, Austria May 7 – 13.

[15] LOWE, D. G.: *Distinctive image features from scale-invariant keypoints*. International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

[16] Wikipedia article on SAR at http://en.wikipedia.org/wiki/Synthetic_aperture_radar.

[17] LUONG, Q. T., FAUGERAS, O.: *The fundamental matrix: Theory, algorithms and stability analysis*, International Journal of Computer Vision, 17(1):43-76, 1996.

[18] DAVIS, J.: *Hierarchical Motion History Images for Recognizing Human Motion*, IEEE Workshop on Detection and Recognition of Events in Video, Vancouver, Canada, July 8, 2001.

Dipl. Eng. Ondřej KOTABA
Honeywell International, s.r.o.
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: ondrej.kotaba@honeywell.com

UTILIZATION OF SIMULATION AT THE EVALUATION OF FATIGUE PROPERTIES OF MATERIALS

Mariana KUFFOVÁ, Václav KRÁLÍK

Abstract: The fatigue process is very sensitive to a great amount of both external and internal factors that, each of them on its own, but especially when acting simultaneously, can affect the fatigue resistance of materials, parts and constructions. The fatigue tests can be performed in various experimental conditions at different loading frequencies. We have all come across examples of materials fatigue failure, whether it is the broken rail which caused trains to be delayed or the broken shaft which caused car crash, etc. The importance of fatigue is tied foremost to safety of persons, which life is dependent on the reliability of given device operation. Lowering costs and in substantial extent also increasing efficiency allows us progress in computer technology usability and in application of numerical techniques onto large amount of problems in mechanical engineering.

Keywords: Simulation, finite element software ADINA, safety, construction materials.

1 INTRODUCTION

Fatigue of metallic structural components is widespread failure in the different industrial branches. Unlike other degradation phenomena, which in most instances can be controlled with proper design and maintenance, fatigue has been unavoidable since cyclic loading occurs. A critical issue is how to monitor structural fatigue. Fatigue has two major classifications: low-cycle and high-cycle. Low-cycle fatigue and high-cycle fatigue are two different mechanism, each causing crack growth, und ultimately failure in mechanical components subjected to cyclic loading. The main difference between the two is the amount of loading applied to the structure, and consequently the number of loading cycles to failure. Low-cycle fatigue occurs in materials under cyclic load where each loading incrementinduces a small degree of plastic deformation, eventually forming a crack. Continued loading will cause the crack to propagate, accelerating until catastrophic failure occurs. Conversely, high-cycle fatigue occurs under less loading than low-cycle, where each loading cycle only deformsthe material elastically, thus requiring many more cycles to induce and propagate a crack.

The type of fatigue that can ultimately affect a structure is dependent upon the type of loading to which it is subjected. Cracks initiate and propagate from preexisting flaws, material defects, or design features (fater holes or sharp corners). In the fact, most fatigue is widespread, as hundreds, or even thousands of cracks are manifested in cyclic loading. The net effect of numerous fatigue cracks located in the same general area is that they synergistically interact reducing the structure's residual strength. However, the single-crack concept is still important, because ultimately, catastrophic failure can occur when a single crack goes critical and in the process envelops other adjacent cracks in zippering effect [1].

Two basic methods are employed to predict and determine potential fatigue locations. The first is full-scale fatigue testing, which is performed during the development process. These tests are employed to ascertain the expected durability. Decreasing of expenses as well as increasing of efficiency of whole process of mechanical design and providing of operability during the overall lifetime of parts and machineries allow us to make progress in the field of the utilization of the computational technologies and the application of numerical methods for the solution of huge amount of mechanical engineering praxes' problems. Nowadays we have several commercial programs at our disposal which allow us to solve the crack propagation. Many authors have dealt with influence of the crack growth on the functionality of the particular parts from global point of view [1, 2, 3, 4].

2 NUMERICAL METHODS

The analysis of cracks within structures is an important application if the damage tolerance and durability of structures and components are to be predicted. As part of the engineering design process engineers have to assess not only how well the design satisfies the performance requirements but also how durable the product will be over its life cycle. Often cracks cannot be avoided in structures; however the fatigue life of the structure depends on the localization and size of these cracks. In order to predict the fatigue life for any component, a fatigue life and crack growth study needs to be performed. Linear and nonlinear fracture mechanics can be performed with the ADINA (Automatic Dynamic Incremental Nonlinear Analysis) system. The ADINA fracture mechanics capabilities include the computation of conservation criteria (J – integral, energy release rate) in 2-D and 3-D finite element models, and the analysis of actual crack propagation in 2-D finite element models. Two different numerical methods are available for the computation of the conservation criteria. The line contour method and the virtual crack extension method. For the analysis of crack propagation, the "node shift/release" technique is used in ADINA. With this technique, which includes local remeshing of the crack front during crack propagation, a smooth and continuous advance of the crack can be modeled. The fracture mechanics allows performing an analysis with only one crack however. The crack line or surface can be located on the boundary or inside of the finite element model. The specimen and loading must be symmetric.

The crack propagation surface must be planar, which typically corresponds to the ligament in a fracture mechanics test specimen, is defined in ADINA as the set of nodes which may possibly be released when the crack opens. In 2-D models, the crack propagation surface reduces to a line (Figure. 1).

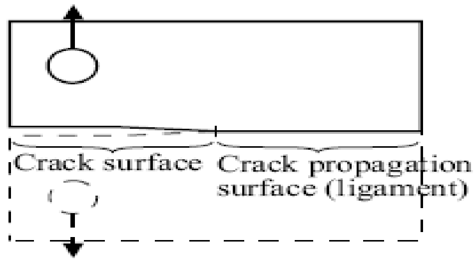


Figure 1 The crack propagation surface for 2-D analysis

In order to evaluate the amount of crack advance from the value of the energy release rate a resistance curve must be input. This curve (Figure 2) is a material property. When the crack growth control parameter is a displacement parameter, then this curve relates the value of the displacement paremeter to the crack advance. When the crack growth control parameter is an energy release rate, then this curve relates the value of the energy release rate to the crack advance.

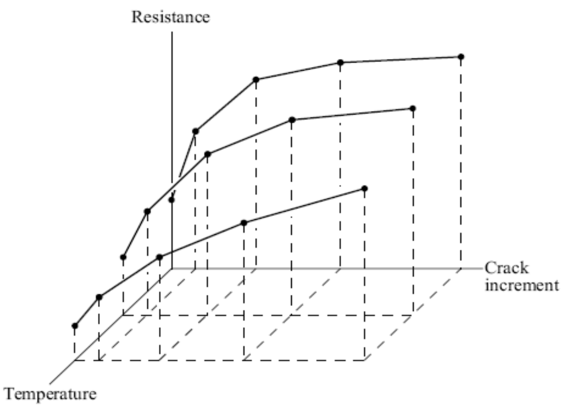


Figure 2 Resistance curves

A typical mesh for 2-D crack propagation analysis is shown in Figure 3. Quadrilateral elements with any number of nodes can be used. To ensure good results, the mesh in the area where the crack propagates should be made of a regular density of elements with regular shapes.

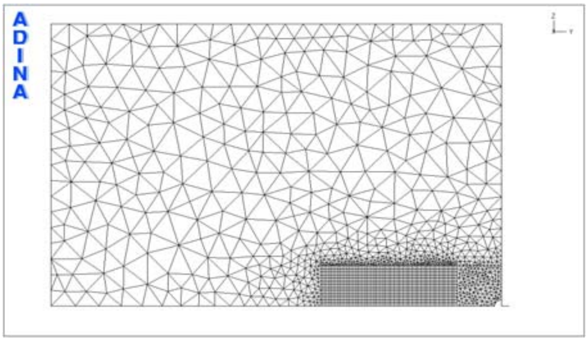


Figure 3 Mesh for crack propagation analysis

Transition meshes are used to connect the portion of refined mesh at the crack tip zone to the mesh used for the rest of the structure.

3 EXPERIMENTAL MATERIALS

The simulations were carried out on two different materials which are widespread in the construction in the sea and railway transport. Aluminium alloy AlMg3 EN AW-5754 (DIN 1745) is weldable and corrosionproof, it is used for construction of means of sea transport, shipbuilding, tanks, mechanical parts and containers. Construction steel S235JRG1 (EN 10025A1- STN 41 1373) is weldable, low-carbon, soft steel used for building transport machinery and devices, for chosen mechanical parts, parts of constructions, frames and suspension of rail vehicles and other transport vehicles. Chemical composition and chosen mechanical properties of aluminium alloy and low-carbon steel are shown in Table 1, 2 (EN AW-5754) and Table 3, 4 (S235JRG1).

Table 1 Chemical composition (mass. %) , EN AW-5754

Al	Cu	Mg	Si	Fe
95.52	0.1	2.6	0.4	0.4

Table 2 Mechanical properties, EN AW-5754

Re [MPa]	Rm [MPa]	A [%]	E [MPa]	HV
Min. 80	190-240	Min.18	$6.5.10^4$	78

Table 3 Chemical composition (mass. %) , S235JRG1

C	Mn	Si	P	S
0.15	0.93	0.44	0.01	0.009

Table 4 Mechanical properties, S235JRG1

Re [MPa]	Rm [MPa]	A [%]	Z [%]	E [MPa]	HV
235	372	24	57	2.06.105	132

4 NUMERICAL RESULTS

Using the finite-element program ADINA a fatigue test was simulated for given materials - steel S235JRG1 as well as aluminium alloy EN AW-5754. It was a low cycle fatigue and a model of corresponding test specimen was used on which the experimental tests were performed. The test specimen shows double symmetry and this knowledge was effectively used to reduce the model and modelled was only one quarter of the specimen together with boundary conditions for symmetry. The loading was given the same as in the case of experiment as a cyclic stress with maximum equivalent to force 100 000 N and with asymmetry of the loading cycle R=0.35; 0.45; 0.5; 0.6 for steel S 235JRG1 and R = 0.4; 0.6, 0.7; 0.8 for Al-alloy AW-5754. Quadratic elements were used to discretize the model. In the area of crack propagation and in the vicinity finer mesh was produced in order to achieve higher accuracy. It turned out that the quality of the mesh has significant influence on the accuracy of the results. The material is modelled as an ideal elastic-plastic material with no defects or imperfections, except for the defined initial crack.

For processing of the numerical results the starting point was the dependence of crack length 2a on number of cycles N. From the dependence 2a – N was numerically obtained the dependence of fatigue crack growth rate da/dN on the crack length 2a and also numerically obtained was the dependence of fatigue crack growth rate da/dN on the applied value of stress intensity factor amplitude ΔK_{apl} [3]. Obtained dependence of crack growth rate da/dN on ΔK_{apl} for steel S235JRG1is shown in figure 4 and for EN AW-5754 is shown in figure 5.

.

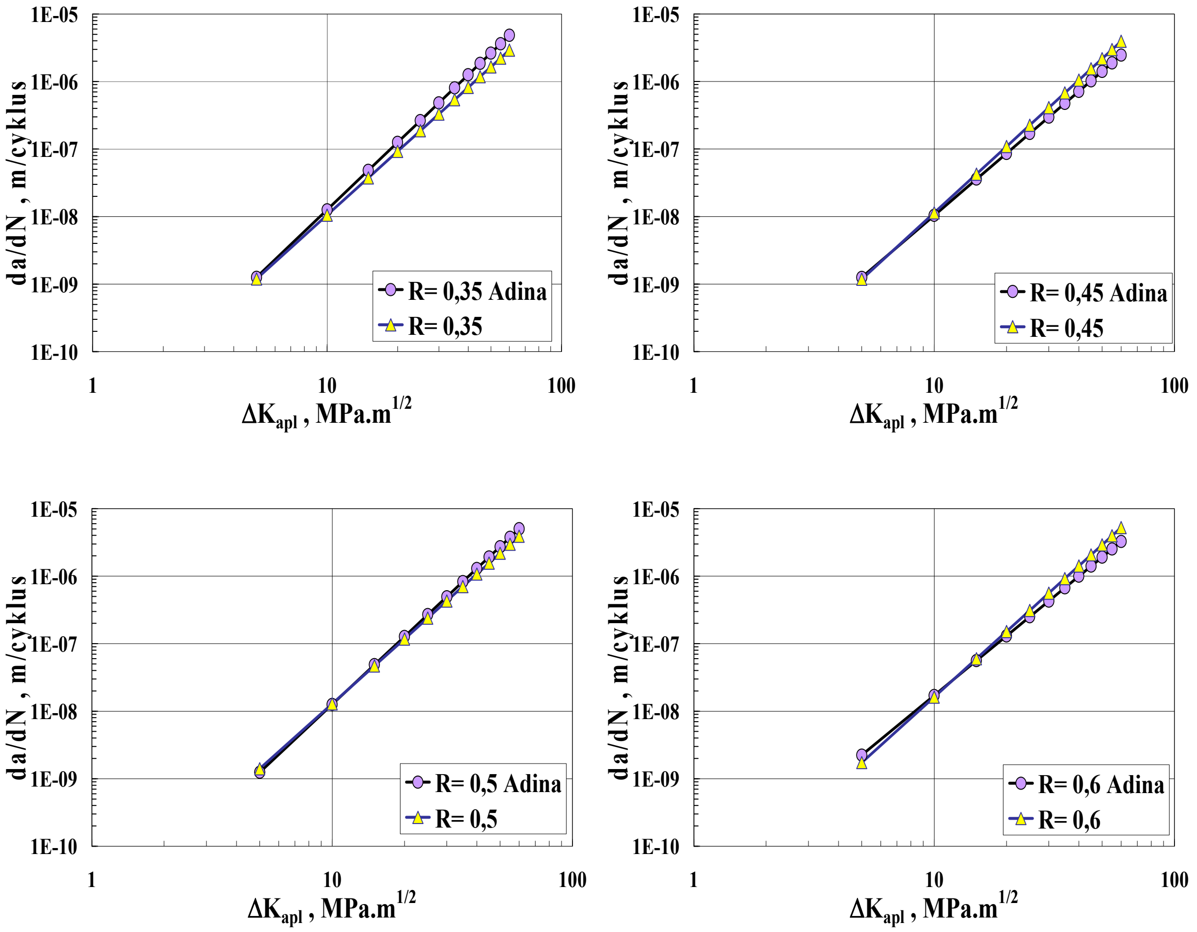


Figure 4 Dependence da/dN - ΔK_{apl} for steel S235JRG1

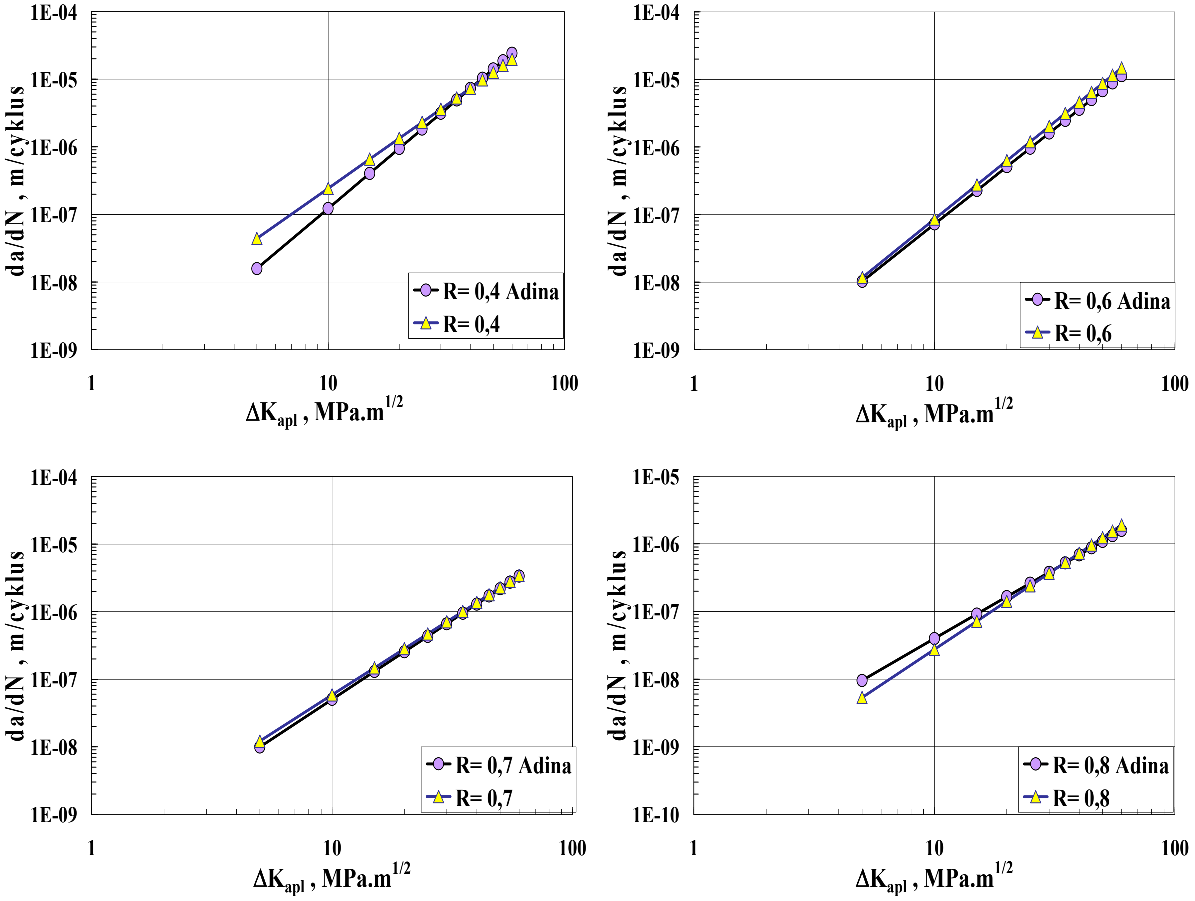


Figure 5 Dependence da/dN - ΔK_{apl} for Al-alloy EN AW-5754

There are listed acquired values of coefficients C, m Paris dependency equation da/dN - ΔK_{apl}, in table 5 for steel S235JRG1 valid for length period of fatigue cracks 2a = 20 till 35 mm as well as in table 6 for Al-alloy EN AW-5754 valid for length period of fatigue cracks 2a = <15, 35>. Values of coefficients C, m are for particular asymmetry coefficient R.

Table 5 Value of coefficients C, m Paris dependency equation, steel S235JRG1

R	**S235JRG1**	
	F_{max} = 100 kN, f = 10Hz	
	C	**m**
0,35	$7.59.10^{-12}$	3.141
0,45	$6.14.\ 10^{-12}$	3.266
0,5	$8.25.\ 10^{-12}$	3.192
0,6	$9.89.\ 10^{-12}$	3.236

Table 6 Value of coefficients C, m Paris dependency equation, Al-alloy EN AW-5754

R	**EN AW-5754**	
	F_{max} = 100 kN, f = 10Hz	
	C	**m**
0,4	$5.72.10^{-11}$	3.309
0,6	$8.95.\ 10^{-11}$	3,236
0,7	$9.49.\ 10^{-11}$	3,312
0,8	$1.05.\ 10^{-10}$	3,283

5 CONCLUSION

In the base of information analysis which consequent from theoretical analyze of solved problem and discussion of numerical results it is possible to say following conclusions:

- For used low -carbon weldable steel S235JRG1 and Al-alloy AW-5754 were obtained:

- the dependences da/dN - ΔK_{apl} which were compared to experimental dependences,
- values of constants C, m Paris dependence equation da/dN - ΔK_{apl} for interval of fatigue breaches 20 till 35 mm for S235JRG1 and for interval 15 till 35 mm for AW-5754 and individual asymmetries of loading R.

- With increasing of loading cycle asymmetry R, it is increased the fatigue crack growth rate.
- According to comparison of experimental and numerical results, it is possible to state a good correlation between the experimental measurement and numerical simulation.
- Variance in results is strongly dependent on fineness of the mesh in the crack propagation area and also on material model.
- Finite element software ADINA is able to solve the problem of the fatigue crack propagation and its results are fully comparable to the experimental results.

This work has been supported by the Grant Agency of Slovak Republic grant No. 852_08-RO02_RU21-240 „Indoor Navigation and Mapping Robot System“.

References

[1] CARPENTER, M.: *Managing the Fleet:Materials Degradation and its Effect on Aging Aircraft.* AMPTIAC, Vol. 5, No. 4, 2001, pp. 1-20.

[2] CHAROENPHAN, S., PLESHA, M. E., BANK, L. C.: *Simulation of crack growth in composite material shell structures*. Int. J. Numer. Meth. Engng 2004; 60, p.2399–2417.

[3] URAL, A., WAWRZYNEK, P. A., INGRAFFEA, A. R.: *Simulating Fatigue Crack Growth in Spiral Bevel Pinion NASA Glenn Research Center*: Cleveland OH, 2003.

[4] NGUYEN, O., REPETTO, E. A., ORTIZ, M., RADOVITZKY, R. A.: *A cohesive model of fatigue crack growth.* Int. Journal of Fracture, 2001, 110: 351–369.

[5] KUFFOVÁ, M. , ŠTIAVNICKÝ, M.: *Analysis of the fatigue crack growth using ADINA.* In: Materials Engineering, 15, 1, 2008.s. 32, ISSN 1335-0803.

[6] KUFFOVÁ, M. , ŠTIAVNICKÝ, M.: *Rýchlosť šírenia únavovej trhliny zvárateľnej ocele STN 41 1373.* In Zb. „Opotřebení, spolehlivost, diagnostika 2008, Brno, Česká republika, 2008, s. 111-116, ISBN 978-80-7231-558-1.

Dipl. Eng. Mariana KUFFOVÁ, PhD.
Department of Mechanical Engineering
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 01 Liptovský Mikuláš
Slovak Republic
E-mail: mariana.kuffova@aos.sk

Dipl. Eng. Václav KRÁLÍK, PhD.
Department of Mechanical Engineering
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 01 Liptovský Mikuláš
Slovak Republic
E-mail: vaclav.kralik@aos.sk

COMMUNICATION ANTENNA EMITTING IN TRANSVERSE VERTICAL PLANE OF AIRCRAFT

Ján LABUN, Pavol KURDEL, Pavol LIPOVSKÝ

Abstract: In professional and laymen circles there is a prevailing idea that in order to enlarge the boundary of the radio communication range, it is appropriate to bank an aircraft so that to improve the line of sight of the antenna with reference to the ground communication point. This theory is true on condition that the angle of banking does not exceed a certain stated angle, which is specific for each type of aircraft and is dependent on various factors. It is about the factors such as dimensions and the shape of the aircraft, location of the antenna on the aircraft surface, operating frequency etc. From this point of view, the article is a treatise on the causes and rules underlying of the frequency drops in the radiation pattern of the communication antenna located on the body of the aircraft from the point of the angle of bank with the minimum level of the signal received. As it is generally known, the change in the level of signal reception is of substantial effect on the range of communication link.

Keywords: Asymmetric dipole, aircraft antenna, range of radio communication.

1 INTRODUCTION

For air communication, the aircraft usually uses the antenna of non-symmetrical dipole type – unipole, which is commonly fitted to the upper axis part of aircraft fuselage. Given position of this dipole has the task to ensure circular direction diagram in horizontal plane and by this guarantee quality connection of the aircraft with control center in any direction.

From the point of view of securing certain – wider range of frequency band, the dipole has usually larger – wider dimensions. As the antenna has to have also good aerodynamical features, this antenna is then shaped (recently) into the shape of a fin. In spite of these conditions, the antenna has to have required electric parameters and radiation features available.

In professional as well as amateur air community there is a prevailing opinion that to increase the range of connection it is advantageous to bank the aircraft so that the antenna would be "better" uncovered towards the ground communication point. This theory is right, however, the bank angle should not extend certain given value. This given value of banking is specific for each type of aircraft and depends on numerous influences, such as:

- dimensions and shape of aircraft,
- position of antenna and its distance from certain important elements on aircraft fuselage,
- working frequency,
- antenna dimensions, shape and polarization,
- antenna radiation characteristics, etc.

As a result of these specific characteristics of antenna position on the metal aircraft fuselage, there occurs deformation of original, almost ideal shape of direction diagram of antenna. One of the negative impacts of these deformations is the occurrence of outlets (considerable minimums) of direction diagram, by which there can occur a loss ofconnection even before the border of radio range. From this viewpoint, the paper discusses the reasons and laws of the occurrence of outlet of direction diagram of communication antenna placed on the metal aircraft fuselage from the point of view of determining the bank angle with minimal level of received signal.

2 ANTENNA ON SPACIOUS CONDUCTIVE SURFACE

Necessary and objective requirement of placing antenna of radio communication means on the surface of conductive metal aircraft fuselage presents a situation of placing the antenna in the vicinity of spacious, well conductive surface. To calculate the radiation of antennas placed within spacious, well conductive surface, we use a well-known principle of reflection. The field of radiating antenna induces currents, which also contribute to the total radiation of the antenna, in the conductive – metal aircraft fuselage.

The share of induced currents is equivalent to the radiation of reflection image of antenna according to the interface plane. Induced currents in reflection image have the same value (amplitude) as the currents in the antenna itself. The phase of these induced currents depends on the antenna polarization. Induced currents of reflection image evoked by vertically polarized antenna have a phase equivalent to the current in the evoking antenna. (The current in the reflection image of the vertical antenna itself is evoked reversely, but externally it has the same phase, Fig. 1a). Horizontally polarized antenna, however, evokes induced currents with opposite phase. (The current in reflection image of horizontal antenna itself is evoked also reversely, but it has the opposite phase also externally, Fig. 1b). Frequent example also of aircraft antenna is presented by so-called refract L antennas, where horizontal part presents capacity load. The phase of currents in different parts of antenna and its reflection image is shown in Fig. 1c.

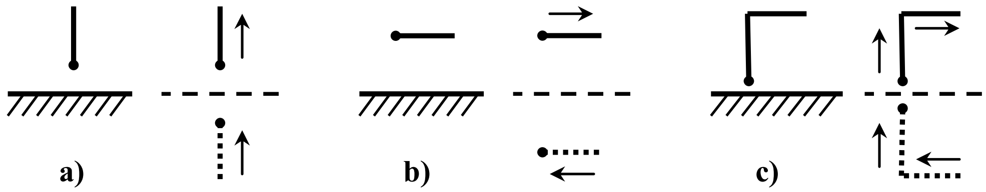


Figure 1 The example position of antenna above conductive aircraft fuselage with reflection image

To determine the radiation of antenna above well conductive surface means to add the radiation of real antenna to the radiation of its reflection image. The results obtained by this are, however, valid only in half-space above conductive plane surface. In Fig. 2 there is a vertical dipole in the height „h“ above conductive plane. Further, the paper deals only with vertical polarization, because in communication in air service, only vertical polarization has been used. The intensity of the field in the point of receiving „P“ is the sum of intensity E_1 evoked by the dipole and intensity E_2 evoked by reflection image. Bothe intensities, however, differ in phase, because the courses of waves r_1 and r_2 into the receiving point have different length.

$$r_1 = r - h \cdot \cos\vartheta, \qquad r_2 = r + h \cdot \cos\vartheta \qquad (1)$$

Resulting intensity of the field of vertically polarized antenna placed in the height *„h“* above well conductive flat surface is given by the following relation:

$$E_{\vartheta}^{(P)} = 60\, I_{\max} \frac{e^{-jkr}}{r} \cdot j \frac{\cos(kl\cos\vartheta) - \cos kl}{\sin\vartheta} \cdot 2\cos(kh\cos\vartheta) \qquad (2)$$

Contribution of reflection image of the antenna into receiving point shows in the characteristic function of radiation by the factor *2 cos(kh cos ϑ)*. At the same time, the resulting shape of direction diagram depends on the height of the placement of the antenna, more specifically, on the rate of its height h to the wavelength λ *“h/λ“*. Several examples of height dependence of direction diagrams of vertical dipole above plane conductive surface given by the function of radiation from the relation (2) is shown in Fig. 3.

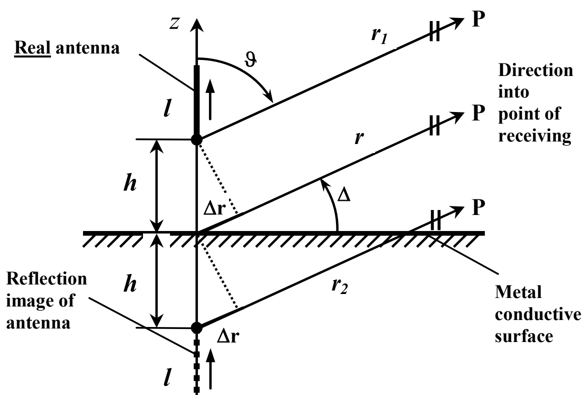


Figure 2 Radiation out of real antenna and its reflection image into point of receiving

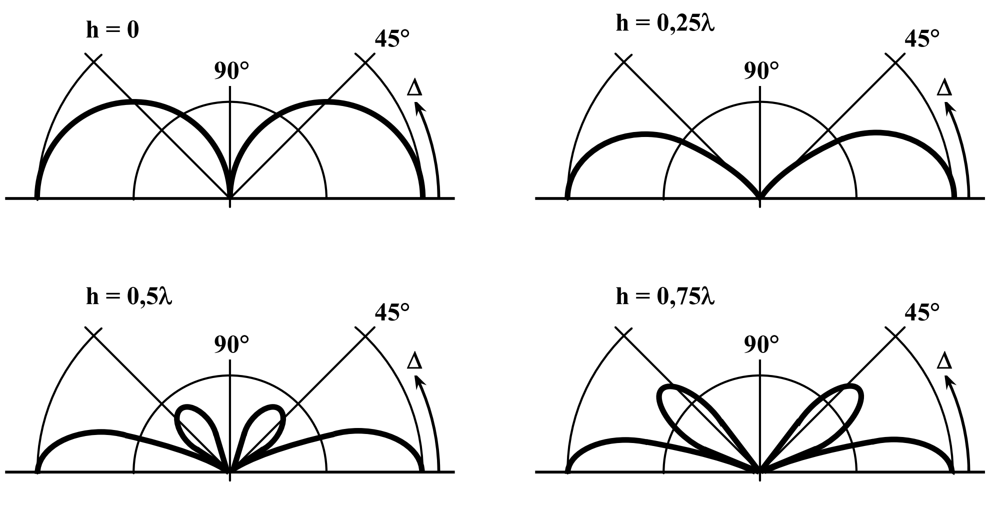


Figure 3 Directional diagram of vertical dipole antenna above conductive area for differently height “h”

3 ANTENNA PLACED ON AIRCRAFT

So far, the paper considered placing of the antenna on the upper part of aircraft fuselage an also its radiation above this surface – above the aircraft, has been evaluated. It seems that towards the direction below the aircraft, as a result of shading effect of metal fuselage, wings and other construction elements, the radiation will not occur. However, the opposite is truth, because as a result of diffraction, the radio wavelength will get also into the shaded space – under the aircraft.

Diffraction is the change of direction of spreading the radiation as a result of bending radio wavelengths on the metal parts (obstacles) of the aircraft. The phenomenon of diffraction occurs in all types of waving (sound, electromagnetic, light) as well as other kinds of obstacles.

Because the communication in air service operates in the field of radio wavelengths, we are mainly interested in the diffraction of radio waves in cylinder, because the shape of aircraft fuselage can be approximated to it as well as obstacles, near which the radio waves spread.

Instead of complicated mathematical expression of diffraction, in Fig. 4 there is a direction diagram of communication antenna above and under the fuselage of aircraft MiG-29. The following Fig. 5 shows the orientation of aircraft and plane of measuring of the direction diagram of the antenna on the aircraft MiG- 29. The plane of pictured direction characteristics „y-z“ is perpendicular to the flight direction „x“.

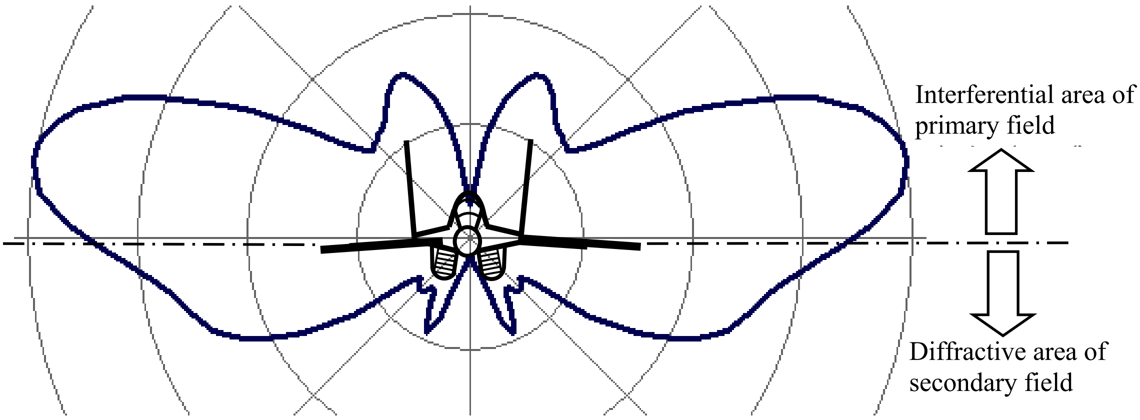


Figure 4 Direction diagram of vertical dipole antenna below conductive area of aircraft MiG - 29

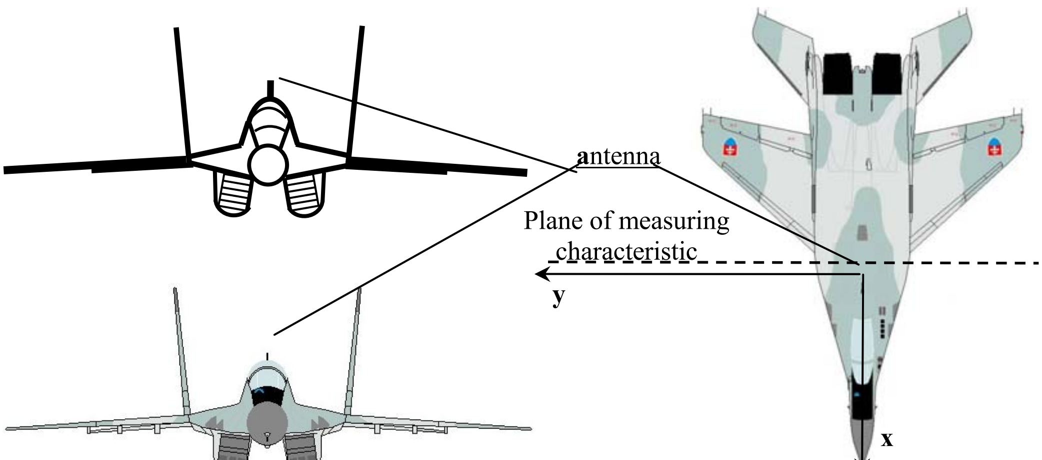


Figure 5 Orientation and plan of measuring directional diagram on aircraft MiG - 29

4 EFFECT OF AIRCRAFT BANK ON THE CONNECTION RANGE

To create mathematical model of direction diagram of vertically polarized antenna fixed on conductive surface of a real aircraft, an image according to Fig. 6 has been created. In this figure, there is a scheme of conductive surface of the aircraft MiG-29 with indicated real height of antenna 1,1m. Operating frequency has been determined to 120MHz, which is one of the frequencies of determined air service frequency ranges (118MHz ÷ 137MHz). Fig. 8 shows calculated mathematical model of direction diagram.

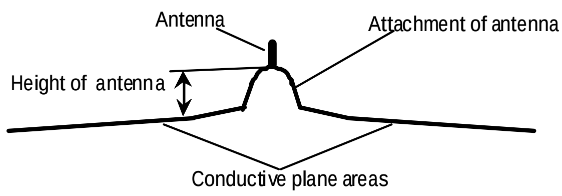


Figure 6 Simplified image of antenna and position its on conductive fuselage aircraft MiG - 29

For the purpose of measurement and following evaluation of real direction diagram, a model of an aircraft MiG-29, 1:10, has been made so that it would enable measurement in given range. The measurement has been made in the depression chamber, where the standard condition determined for such kind of measurement (see Fig. 7) has been kept. Also, Fig. 8 depicts measured direction diagram of realized model.

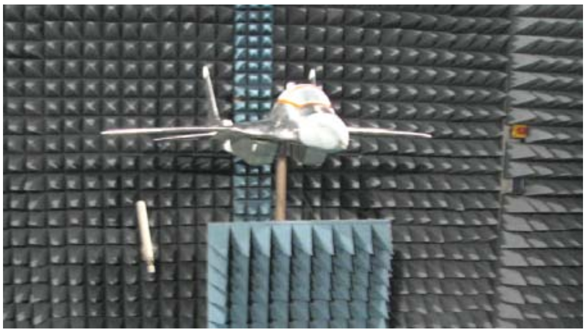


Figure 7 Measuring of direction diagram on the aircraft model - MiG 29

The comparison of measured direction diagram on the model of real aircraft with direction diagram created by above given mathematical model is shown in Fig. 8. For comparison, the following measured angles will serve:

a) maximum of the main lobe (max HL),
b) maximum of the side lobe (max BL),
c) minimum between the given lobes (min H-B).

Mathematical model:

max. HL ⇒ 0°, max BL ⇒ 56°, min H-B ⇒ 40°.

Measurement on the model:
max. HL ⇒ 10°, max BL ⇒ 66°, min H-B ⇒ 42°.

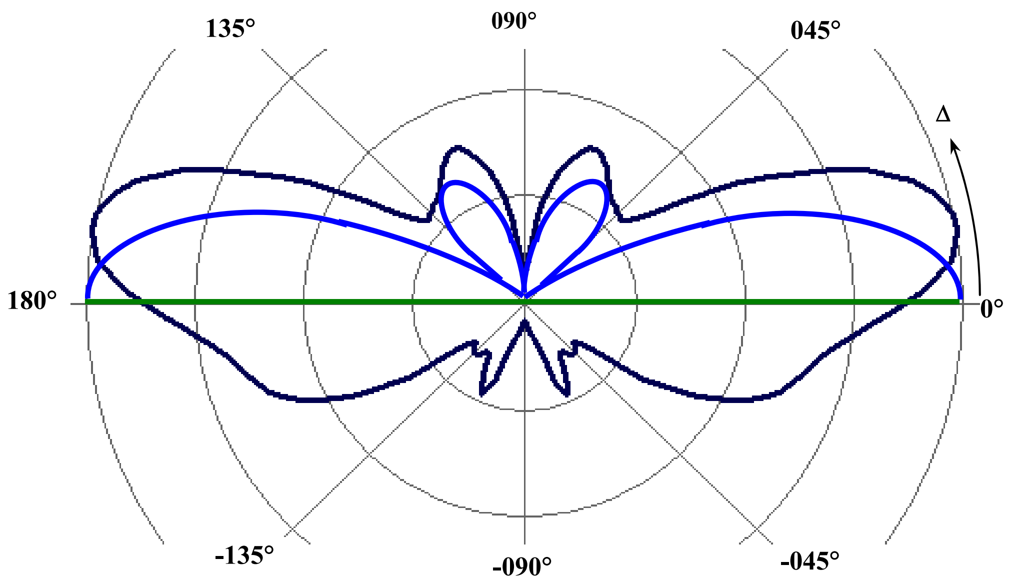


Figure 8 The comparison of measured direction diagram with mathematical model

By the comparison of both types of direction diagrams it is possible to conclude that the diagram measured on the model of a real aircraft can be compared with the mathematical model of a real aircraft. Certain ambiguity can be seen in the angle of the maximum of the main lobe, which is approximately 10°, in the comparison of the maximum of the main lobe of the mathematical model, where the value is 0°. The difference can be explained by the fact that the plane surface of metal wings of the real aircraft is not horizontal, but it has a negative angle - 5°. This negative lobe of wings will rise the main lobe of direction diagram in mentioned 10°. As a result, also measured maximum of side lobe is raised in 10°. The maximum of mathematical model of side lobe is 56°, and measured value of this lobe is 66°.

Comparing the angles of minimal value of direction diagram in the area between the main and side lobes is not so distinctive, it presents only 2°. Besides, the decrease of the minimum on the measured diagram is not so distinctive as on mathematical model. Given differences follow the fact that on the real aircraft there is no univocal transition between the vertical position of the antenna and horizontal surface of the wings. The shape of direction diagram is influenced by the aircraft fuselage, which is compared to "antenna attachment" and different bank of wings in the area of their fixing to the fuselage, see Fig. 5 and 6.

The fact is that during aircraft maneuvers, during its banking in about 40° there occurs significant decrease of signal in broadcasting or receiving, by which the range of connection is decreased as well. This phenomenon will be important, if the aircraft is located near the border of radio range. Then, when the connection is lost, the pilots can think that by banking the aircraft – by better „uncovering“ of the antenna, the communication can be renewed. It would be suitable if pilots knew this phenomenon and that by greater aircraft banking the connection will not be renewed. The model situation with the frequency 120MHz shows that this critical angle is 40°. Analogically, with the use of knowledge from Fig. 3 and mathematical modeling it is possible to conclude that with increased frequency this critical angle decreases. The change is that with operation frequency 240MHz the value of this critical angle is only 20°.

5 CONCLUSION

The paper deals with the problem of up to which rate it is suitable to bank the aircraft to increase the limits of the range of radio connection. The authors of the paper have faced an opinion in practice, that in order to increase the range, it is suitable to „better“ uncover the antenna towards the ground communication point by banking the aircraft. It seems that this theory is valid under the condition that the bank angle does not exceed certain determined value, in our model case it is the angle of 10°. This value of banking is specific for each aircraft and depends on several influences, as it is described in the introduction in more details. In our model case the critical angle of minimal reception is 40°. It is important to know that the change of decrease with increasing angle is quicker than it could seem at first sight. By this paper, the author would like to inform professional public about this phenomenon so that it could contribute to increased flight safety.

References

[1] ADAMČÍK, F.: *Artificial intelligence technology on board of aircraft*. In: Advances in military technology. vol. 1, no. 1 (2006), p. 137-142. ISSN 1802-2308.

[2] BÁLINT, J., BEREŽNÝ, Š.: Mathematic model of temporary segregated areas (TSA). In: Zeszyty Naukowe. no. 1 (5) (2004), p. 63-67. ISSN 1641-9723.

[3] JALOVECKÝ, R.: Palubní systémy řízení letu 1. Brno : Univerzita obrany, 2008. 82 p. ISBN 978-80-7231-500-0.

[4] KVASNICA, P., KVASNICA, I.: Simulation of flying parallel modelling in computer. In: International Review of Aerospace Engineering. Vol. 1, no. 5 (2008), p. 441-446. ISSN 1973-7459.

[5] LABUN, J.: Širokopásmová komunikačná anténa na palube lietadla. In: Nové trendy rozvoja letectva : 7. medzinárodná vedecká konferencia : Košice September 6-8, 2006, Slovakia : Zborník príspevkov. Košice : TU, 2006. 6 s. ISBN 80-8073-520-4.

[6] MATOUŠEK, Z., OCHODNICKÝ, J., HYKEL, A.: Šírenie elektromagnetických vĺn a vodičové antény: Lipt. Mikuláš : Akadémia ozbrojených síl generála M. R. Štefánika, 2006. 140 s. ISBN 80-8040- 306-6.

[7] SZABO, S., OLEJNÍK, F., KANDRÁČ, P., ŠTEFIK, S.: Lietadlá dopravných spoločností. 1. vyd. Košice : LF TU, 2007. 161 s. ISBN 978-80-8073-740-5.

Dipl. Eng. Ján LABUN, PhD.
Dipl. Eng. Pavol KURDEL
Dipl. Eng. Pavol LIPOVSKÝ
Technical University of Košice
Rampová 7
041 21 Košice
Slovak Republic
E-mail: jan.labun@tuke.sk, pavol.kurdel@tuke.sk
pavol.lipovsky@tuke.sk

DETECTION AREA ANALYSIS IN THE ELINT SYSTEM

Zdeněk MATOUŠEK, Pavel BUČKA

Abstract: In this article we describe a possible analysis technique of detection area in the Electronic Intelligence - ELINT system depending upon the values of detection probability P_D, false alarm probability P_F and the change of output within ELINT source P_Z. By this analysis it is assumed that there is a monitoring of mobile ELINT sources, which are located on the margin of ELINT system radio horizon, whereby it is necessary to carry out the radio-locating of these sources in the main lappet of directivity pattern source axis and ELINT device.

Keywords: Electronic Intelligence, reconnaissance system, the Doppler effect, transmitter, power agility, the probability detection.

1 INTRODUCTION

Electronic Intelligence – ELINT is one of the basic types of signal intelligence – SIGINT, which uses the principles of electromagnetic energy in the atmosphere propagation when operating. The aim of ELINT devices is to detect, locate, analyze, identify and pursue the sources of electronic signals in the detection area. This area is defined by directivity pattern external border and its size depends on more parameters of ELINT reconnaissance.

The source of electronic signals and ELINT devices create so called "ELINT reconnaissance system". Its most important parameter is the shape and size of detection area, in which the ELINT device is able to detect the electronic signals sources with required index figures. Exactly these index figures are limited by parameters of the ELINT reconnaissance system, as for example:

- maximum distance detection R of ELINT source,
- detection probability P_D,
- false alarm probability P_F,
- the output of ELINT source P_Z.

The given parameters limit the practical use of ELINT devices. In real life there are cases when it is inevitable to review, whether the given type is appropriate for detecting certain types of sources or the given reconnaissance information (from the point of incidence and given probabilities) suits the user.

The correct analysis of these cases is not easy and requires complex access to facts which look in some cases incomparable. The base of solving given problem is to evaluate the change of range in ELINT device depending on change of detection P_D, false alarm probability P_F, and the output of ELINT source P_Z. When analyzing the problem given above it is assumed that there is detection of ELINT sources in the axis main beam of directivity pattern source and ELINT device. The configuration of ELINT system is shown on the Figure 1.

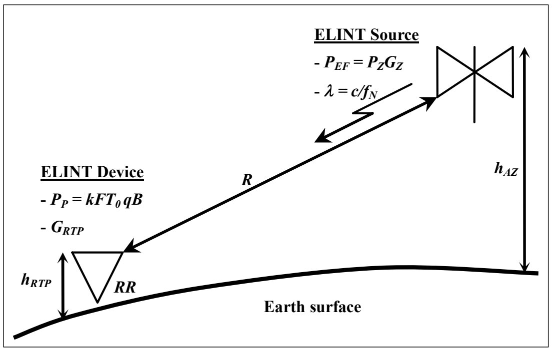


Figure 1 ELINT system configuration

2 DISTANCE OF DETECTING SOURCES IN ELINT SYSTEM

The maximum distance of detecting ELINT sources belongs to the most important parameters of ELINT device, whereby its value limits the maximum distance of ELINT device space detection. When counting the maximum range R_{max} [km] of ELINT device, we start from Radio horizon equation in the following form [1]:

$$R_{\max} = 4{,}123\left(\sqrt{h_{RTP}} + \sqrt{h_{AZ}}\right) \quad (1)$$

where:

h_{RTP} is the height of ELINT device antenna *[m]*,

h_{AZ} is the height of ELINT source antenna system *[m]*.

Theoretical distance of detection R ELINT source is given by Beacon Equation in the following form [2]:

$$R = \sqrt{\frac{P_Z G_Z G_{RTP} \lambda^2}{(4\pi)^2 kT_0 FqBL}}, \quad (2)$$

where:

P_Z is the output of ELINT source *[W],*
G_Z is the antenna gain of ELINT source,
G_{RTP} is the antenna gain of ELINT device,
λ is the wavelength *[m],*
q is the power signal to noise ratio on the output of ELINT device receiver,
F is the ELINT device receiver noise figure,
k is the Boltzmann constant $1{,}38.10^{-23}$ *[J/K],*
T_0 is the absolute temperature *[K],*
B is the receiver bandwidth of ELINT device *[Hz]*,
R is the distance of locating ELINT source *[m],*
L is the loss when processing and transmitting the pulse electronic signal.

From the Beacon Equation it is obvious that the distance of detecting ELINT source changes not only with the change of its output P_Z, but also with the change of power signal to noise ratio q on the output of ELINT device receiver [3].

The least favourable conditions for detecting signals in the ELINT system are for the distances of ELINT sources, which are moving on or in the close proximity to the margin of radio horizon.

3 THE INFLUENCE OF SOURCE PROBABILITY DETECTION ON RANGE IN ELINT SYSTEM

The level of power P_Z transmitted by the ELINT source in the direction of ELINT device or the power signal to noise ratio on the output of ELINT device receiver q define the probability detection of ELINT source P_D and the false alarm probability P_F [3].

The given detection probabilities of ELINT source depend on the distance of detection R, as well as on its power P_Z.

Even if the ELINT device will be stationary and the ELINT source will be moving in the close proximity to the margin of radio horizon, it is possible to characterise the ELINT system as non-stationary transmission system.

Due to the movement of ELINT source, the Doppler effect will occur in received signals, whereby their phase will change. The signal at the input of the ELINT receiver will be the combination of more elements, which spread due to the reflections from near surroundings onto the antenna system from various directions. This fact will become evident in the random changes of the received signals amplitude. From the above given reasons it is possible to consider such system to be a transmission channel with Raylegh distribution amplitude and balanced distribution of received signal phase [4, 5]. For such defined ELINT system it is possible to formulate the target detection probability P_d by the power signal-to-noise ratio at the receiver output q and false alarm probability of the target P_F in this formula [6]:

$$P_D = P_F^{\frac{1}{1+q}}. \tag{3}$$

a) Influence of source detection probability on the range in the ELINT system.

The detection probability of sources P_D expresses the percentage of periodically repeated signals transmitted by the ELINT source will be correctly detected by ELINT device.

Let's assume that ELINT device monitors two sources. For ELINT reconnaissance system when detecting the first source let the following parameters mean:

- detection probability $P_{D1,}$
- false alarm probability $P_{F1,}$
- ELINT source output $P_{Z1.}$

When detecting the second source in the same ELINT system then these parameters mean:

- detection probability P_{D2},
- false alarm probability P_{F2},
- ELINT source output P_{Z2}.

According to [6] it is possible to express the ratio of detecting distance R_2 / R_1 by the following formula:

$$\frac{R_2}{R_1} = \sqrt{\frac{P_{Z2}}{P_{Z1}} \frac{\frac{\ln P_{F1}}{\ln P_{D1}} - 1}{\frac{\ln P_{F2}}{\ln P_{D2}} - 1}}. \tag{4}$$

From the formula (4), provided that $P_{F1} = P_{F2} = P_F$ a $P_{Z1} = P_{Z2}$, it is possible to set the formula for detection distance R_2 of the second ELINT source this way:

$$R_2 = R_1 \sqrt{\frac{\frac{\ln P_F}{\ln P_{D1}} - 1}{\frac{\ln P_F}{\ln P_{D2}} - 1}}. \tag{5}$$

On the grounds of formula (4) it is possible to express the relative percentage difference ΔR_D of ranges R_2 and R_1 by adequate ratio P_{D2} / P_{D1} with this formula [7]:

$$\Delta R_D = \frac{R_2 - R_1}{R_1} 100. \tag{6}$$

Then it is possible to express the range of ELINT device to the second ELINT source by the following formula:

$$R_2 = \left(1 + \frac{\Delta R_D}{100}\right) R_1. \qquad (7)$$

According to the formulas (5) and (6) the calculations and simulations dependency $\Delta R_D = f(P_{D2})$ for the chosen values P_{D1} (discrete values ranging from 0,5 to 0,95) were made. The result of these calculations and simulations are the graphs on the figure 2. [7].

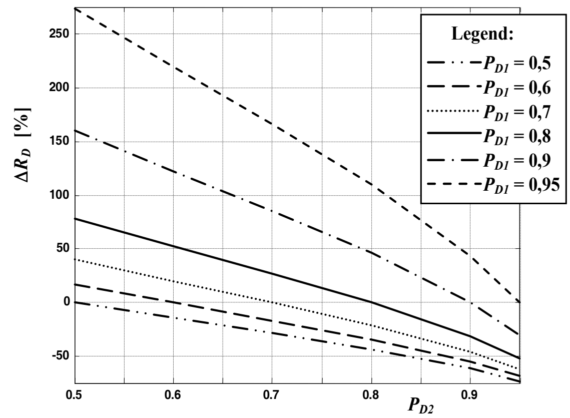


Figure 2 Graph of dependency ΔR_D from P_{D2} for the chosen values P_{D1}

From the graph on figure 2 it is obvious that if the condition $P_{D1} > P_{D2}$ applies, then $R_2 > R_1$. If $P_{D1} < P_{D2}$, then $R_2 < R_1$.

b) Influence of source false alarm probability on the range in the ELINT system.

False alarm probability P_F expresses how many per cent of randomly repeated noise peaks at the input of ELINT device receiver will be incorrectly detected by this device as a useful signal.

For the analysis of ELINT source false alarm probability influence on the change of ELINT device range according to formula (3) and the assumption that $P_{D1} = P_{D2} = P_D$ and $P_{Z1} = P_{Z2}$, it is possible to state the formula for range detection R_2 of the second ELINT source this way:

$$R_2 = R_1 \sqrt{\frac{\dfrac{\ln P_{F1}}{\ln P_D} - 1}{\dfrac{\ln P_{F2}}{\ln P_D} - 1}}. \qquad (8)$$

According to the formula (8) it is possible to state the relative percentage difference ΔR_F of the ranges R_2 and R_1 by the adequate P_{F2} / P_{F1} with this formula [7]:

$$\Delta R_F = \frac{R_2 - R_1}{R_1} 100. \qquad (9)$$

Then it is possible to state the ELINT device range to the second source by the following formula:

$$R_2 = \left(1 + \frac{\Delta R_F}{100}\right) R_1. \qquad (10)$$

According to the formulas (8) and (9) the calculations and simulations of dependency $\Delta R_F = f(P_{F2})$ were made for the chosen values P_{F1} (discrete values from 10^{-6} to 10^{-9}). The result of these calculations and simulations are the graphs on the Figure 3 [7].

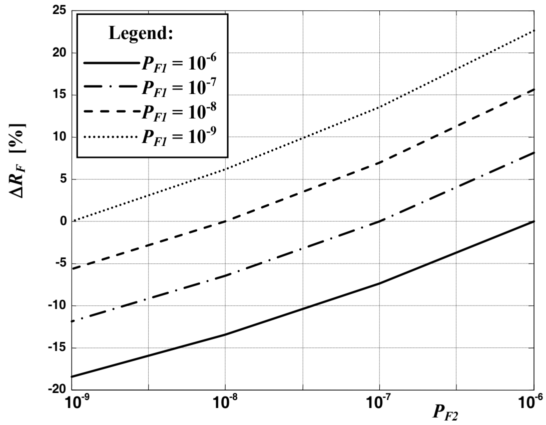


Figure 3 Graph of dependency ΔR_D from P_{F2} for the chosen values P_{F1}

From the graph on the Figure 3 it is obvious that if inequality $P_{F1} > P_{F2}$ is valid, then $R_2 < R_1$. If $P_{F1} < P_{F2}$ then $R_2 > R_1$.

4 THE INFLUENCE OF ELINT SOURCE POWER CHANGE ON RANGE IN ELINT SYSTEM

The ELINT source power P_Z is given by the used transmitter, by which every ELINT source is equipped with. Let's assume that ELINT device monitors two sources. The power of these sources is P_{Z1} and P_{Z2}. The other possibility is that it monitors the source with so called power agility (the change of power from P_{Z1} to P_{Z2} and vice-versa). Then according to the formula (4) possible to state the ratio of transmitting outputs P_{Z1} / P_{Z2} by the following formula:

$$\frac{P_{Z1}}{P_{Z2}} = \frac{R_1^{\,2}}{R_2^{\,2}} \frac{\dfrac{\ln P_{F1}}{\ln P_{D1}} - 1}{\dfrac{\ln P_{F2}}{\ln P_{D2}} - 1}. \qquad (11)$$

From the formula (11) provided that $P_{D1} = P_{D2}$ and $P_{F1} = P_{F2}$, it is possible to state the dependence of change in ELINT sources transmitting powers P_{Z1} or P_{Z2} on the ELINT device range R with this formula:

$$R_2 = R_1 \sqrt{\frac{P_{Z2}}{P_{Z1}} \frac{\frac{\ln P_F}{\ln P_D} - 1}{\frac{\ln P_F}{\ln P_D} - 1}}. \qquad (12)$$

According to (12) it is possible to state relative percentage difference ΔR_P of ranges R_2 and R_1 by the adequate ratio P_{Z2} / P_{Z1} with the following formula [7]:

$$\Delta R_P = \frac{R_2 - R_1}{R_1} 100. \qquad (13)$$

Then it is possible to state the ELINT device range to the second source by this formula:

$$R_2 = \left(1 + \frac{\Delta R_P}{100}\right) R_1. \qquad (14)$$

According to the formulas (12) and (13) the calculations and simulations dependency $\Delta R_P = f(P_{Z2})$ for chosen values $P_{Z1} = (0{,}5 \div 20)$ kW were made. The example of part of calculation results and simulations (discreet values P_{Z1} in range from 4 kW to 7 kW) is displayed by graph on Fig 4 [7].

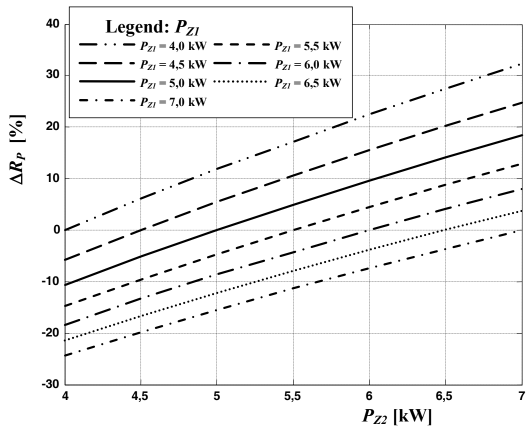


Fig. 4 Examples of part of results of dependency ΔR_P from P_{Z2} for the chosen values P_{Z1}

From Fig. 4 it is obvious that if the condition $P_{Z1} > P_{Z2}$ will be valid, then $R_2 < R_1$. Provided that $P_{Z1} < P_{Z2}$, then $R_2 > R_1$.

5 THE RANGE IN ELINT SYSTEM WHEN CHANGING SEVERAL PARAMETERS AT THE SAME TIME.

In the previous sections of this article there were several cases being solved, when congruent changes of range in the ELINT system by changing one chosen parameter of this system (P_D, P_F or P_Z) were analysed. However, in real life scenarios there are cases, when several parameters in ELINT system change at the same time.

It is possible to solve these cases on the basis of the results of the above mentioned simulations [8] and defined formulas (5, 8 and 12). Provided that we know all individual relative percentage differences (ΔR_D, ΔR_F and ΔR_P), the changes in ELINT system range, which change at the same time, it is possible to state the ELINT device range onto the second ELINT source by this formula [7]:

$$R_2 = \left(1 + \frac{\Delta R_D + \Delta R_F + \Delta R_P}{100}\right) R_1. \qquad (15)$$

6 CONCLUSION

On the basis of the above given results of calculations and simulations it is possible to judge with sufficient credibility the possibilities of detecting sources in ELINT system. When doing this it is necessary to consider distribution of amplitude and phase of received signals and by predicted agility of transmitted outputs by ELINT sources.

From the analysis and synthesis of the above given calculations and simulations results it is possible to state that by partial change of some of the chosen ELINT system parameters (P_D, P_F and P_Z) there is a change in the area dimensions of ELINT device detection. Nevertheless, this change is the most noticeable by the increase or decrease of detection probability P_D. Less noticeable differences, even if they are still very strong, are observable by the change of transmitted output P_Z. The false alarm probability P_F influences the area dimensions of ELINT device detection the least. The change in the area of detection is shown here only for bigger multiples of its basis value. It is inter alia possible to make use of the shown approach of ELINT system area detection analysis for indirect comparison of different ELINT devices parameters; that is for example when making operational calculations.

The authors regard as the original contribution of this article the method of analysis and synthesis of ELINT device range change by monitoring various sources in ELINT system, defining relative percentage differences or ranges ΔR_D, ΔR_F

and ΔR_P, as well as performed simulations [8] and attained results by modelling ELINT systems with changes of selected parameters in the software environment MATLAB.

References

[1] BARTON, D., LEONOV, S.: *Radar Technology Encyclopedia*. Artech House, Inc. Boston, 1997.

[2] PEYTON, Z., PEEBLES, Jr.: *Radar Principles*. John Wiley, New York,1998.

[3] TKÁČ, J. , OCHODNICKÝ, J., ŠPIRKO, Š.: *Vyhodnotenie dosahov radaru s využitím digitálnej mapy*. In: Zborník „Nové trendy v rozvoji letectva", Sekcia IV. - Letecká elektronika. Košice : 2002, str. 41-47. ISBN 80-7166-034-5.

[4] LEE, W. C. Y.: *Mobile Communications Design Fundamentals. Second Edition*, John Wiley & sons, New York, USA, 1993.

[5] STEELE, R.: *Mobile Radio Communication*. Pentech Press Publishers, London, Great Britain : 1994, ISBN: 0-7803-1102-7.

[6] ŠPIRKO, Š., TKÁČ, J.: *The comparison of propability of target detection*. In: Proceedings of 10th Internetional Symposium on Mechatronics. Trenčianske Teplice, June 6.-8. 2007, Trenčianska univerzita, Trenčín :. 141 – 144p. ISBN 978-80-8075-210-1.

[7] MATOUŠEK, Z.: *Analýza priestoru detekcie v rádiotechnickom prieskume*. [Dizertačná práca], Liptovský Mikuláš: AOS v Liptovskom Mikuláši, 2008, 125 s.

[8] PETZ, I., NEČAS, P.: *Simulation techniques and modelling in training*. In: International conference on Military Technologies 2009 "ICMT 09", [Elektronický zdroj], Brno, May 5.-8. 2009, University of Defence Brno, 482-485p, ISBN 978-80-7231-648-9.

Assoc. Prof. Dipl. Eng. Zdeněk MATOUŠEK, PhD.
Electronic Department
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 01 Liptovský Mikuláš
Slovak Republic
E-mail: zdenek.matousek@aos.sk

Assoc. Prof. Dipl. Eng. Pavel BUČKA, CSc.
Vice-rector for Education
Armed Forces Academy of General M. R. Štefánik
031 01 Liptovský Mikuláš
Slovak Republic
E-mail: pavel.bucka@aos.sk

MPM 20 MAGNETIC AURA RESEARCH AND ITS UTILIZATION FOR MPM 20 SITUATIONAL CONTROL

Jana MODROVIČOVÁ, Ladislav MADARÁSZ, Ján KABÁT

Abstract: This article is aimed at in-depth research and experimental confirmation of aircraft turbojet engines magnetic aura existence, more exactly magnetic aura of small turbojet engine MPM 20. MPM 20 magnetic aura is one of its important properties and its utilization possibilities for control purpose until now aren't closely explored. Worked article contains motives and references of its utilization for MPM 20 situational control methodical application.

Keywords: MPM 20, magnetic aura, measurement, cross-section.

1 INTRODUCTION

Magnetism is a physical property, but also some kind of physical performance of some particles as well as objects. Often magnetism serves for early mechanic defect detection of observed objects. Small turbojet engine MPM 20 is observed object and has its own magnetism which is called MPM 20 magnetic aura. In more details, magnetic aura is described in [7,8,9].

This article is concerned with small turbojets engine MPM 20 magnetic aura next research and possibilities in design of situational control based on its magnetic aura. Previous measurements as well as their results can be found in [7,8]. Measurements described in this article are focused on magnetic aura examination and its relation to engine revolutions. Measurements were executed on built measuring construction by 4-channel magnetometer, revolution counter and compressor, which was used for the crank up of the engine. Mentioned magnetometer is more detailed described in [2,3,5,6].

2 MEASURING CHAIN

As was already mentioned before in introduction, created measuring chain was composed of wooden construction on which, the motor was placed, and of wooden frame with 4 probes of magnetometer placed on it. Probes were rotated by 90° each. Sketch of probes placement is shown in Figure 1.

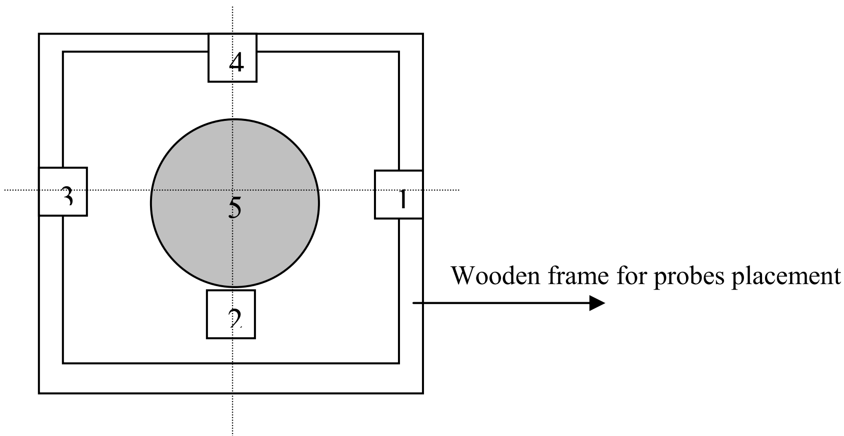


Figure 1 Schematics of wooden frame for probes fixing and placement (cross-section)

Legend: 1 – probe of magnetometer No.1 placed in horizontal plain,
2 – probe of magnetometer No.2 placed in vertical plain,
3 – probe of magnetometer No.3 placed in horizontal plain,
4 – probe of magnetometer No.4 placed in vertical plain,
5 – cross-section of MPM 20 engine.

In next figure, described construction can be seen

Figure 2 Final wooden construction for measurement (front view)

3 MEASUREMENTS

Measurements described in this article were done by already mentioned magnetometer. They were conducted in four cross-sections of MPM 20 engine. In first run, results were gained when the engine was standing still. Next results are while engine was running with speed 1800-2000 RPM and the last results were recorded with lower speed of engine 900-1050 RPM. Sampling frequency was 1 kHz.

Sections, where measurements were conducted are depicted in next figure.

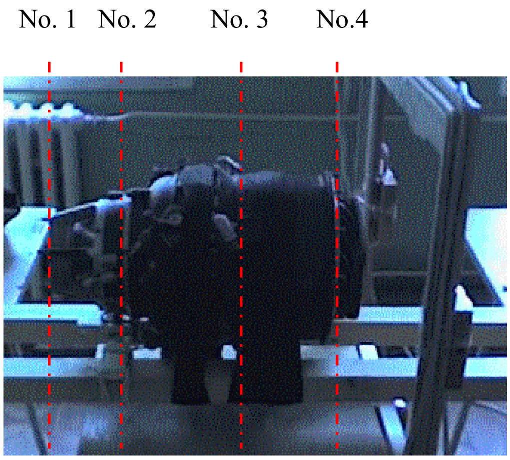


Figure 3 Cross-sections of engine in which the measurements were conducted

In next table, meteorological conditions during measurements are stated.

Tabel 1 Meteorological conditions for measurements done on 23.6.2009

Meteorological conditions					
Date of measurement	**Dew point [^{0}C]**	**Atmosph. pressure [hPa]**	**Avg. outdoor temperature [^{0}C]**	**Avg. indoor temperature [^{0}C]**	**Ozone [Dobson unit]**
23.6.	15	1010	26	24	279 (-12%)

In following graphs, behaviors for each section of engine are shown. In each section, 3 measurements were conducted. When motor was standing still and then when it was turning with lower and higher RPM. Each triplet of figures has corresponding table with boundary values of each run for better readability of measured range. Results of measurements in cross-section 1:

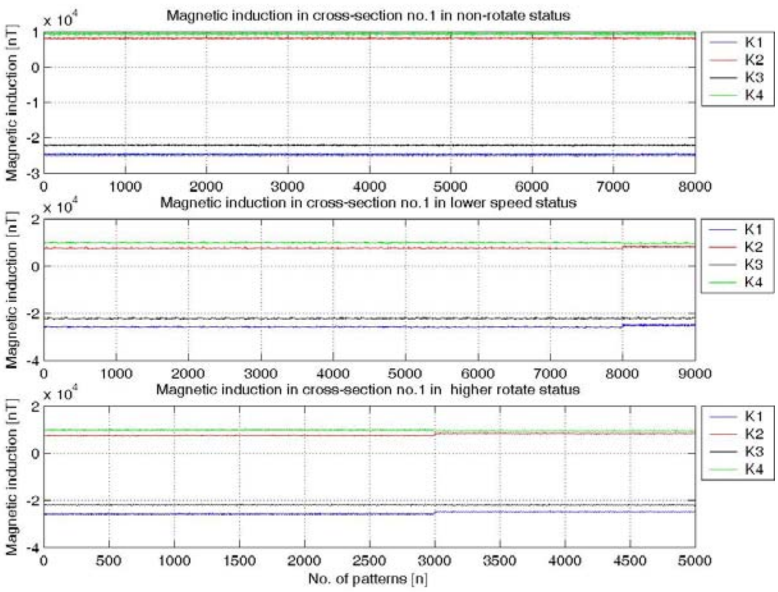


Figure 4 Results of measured magnetic induction in cross-section 1

Table 2 Boundary values for each run in section 1

	Stands till		**Lower RPM**		**Higher RPM**	
	min [nT]	**max [nT]**	**min [nT]**	**max [nT]**	**min [nT]**	**max [nT]**
Channel No.1	-25 254	-24 411	-26 326	-24 898	-26 260	-24 673
Channel No.2	7 864	8 466	7229	8483	6994	8584
Channel No.3	-22581	-21815	-22740	-21780	-22599	-21725
Channel No.4	8841	9765	9204	10388	8975	10201

Results of measurements in cross-section 2:

Table 3 Boundary values for each run in cross-section 2

	Stands till		**Lower RPM**		**Higher RPM**	
	min [nT]	**max [nT]**	**min [nT]**	**max [nT]**	**min [nT]**	**max [nT]**
Channel No.1	-25492	-24580	-26125	-24880	-26011	-24909
Channel No.2	8659	9402	7738	8927	8466	9426
Channel No.3	-22862	-22190	-22731	-21761	-22837	-21987
Channel No.4	8594	9501	9193	10095	9004	9953

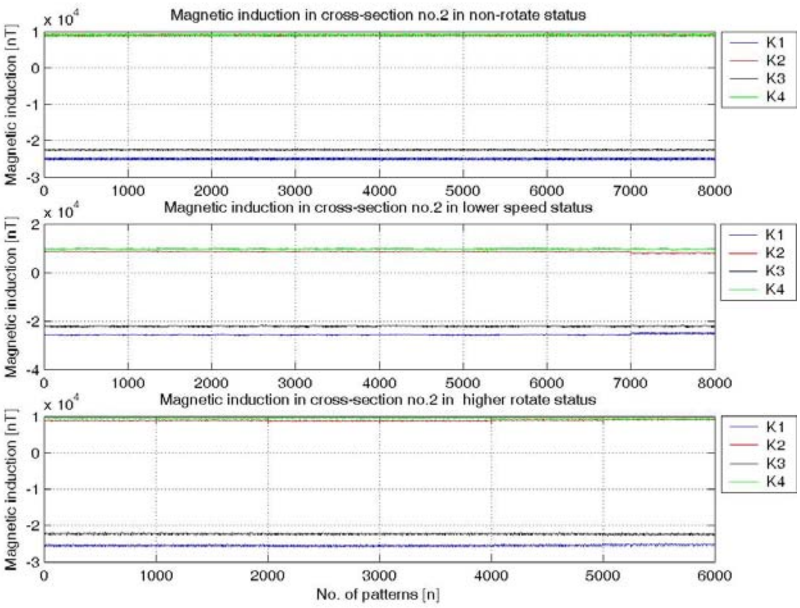


Figure 5 Results of measured magnetic induction in cross-section 2

Results of measurements in cross-section 3:

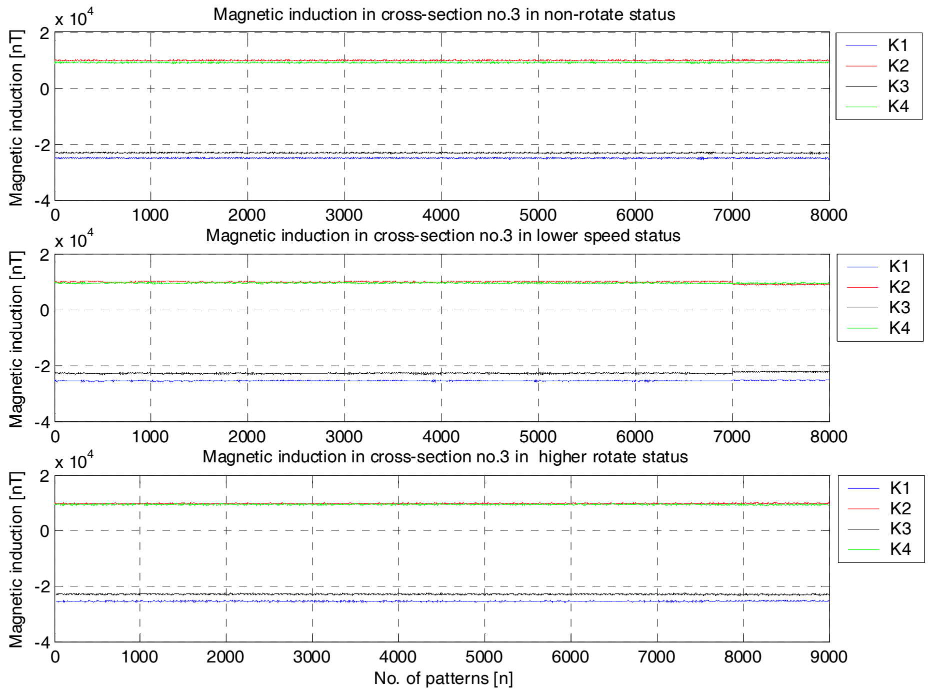


Figure 6 Results of measured magnetic induction in cross-section 3

Table 4 Boundary values for each run in cross-section 3

	Standstill		**Lower RPM**		**Higher RPM**	
	min [nT]	max [nT]	min [nT]	max [nT]	min [nT]	max [nT]
Channel No.1	-25425	-24624	-25914	-25067	-25826	-25030
Channel No.2	9607	10249	8755	10239	9365	10148
Channel No.3	-23567	-22876	-23231	-21983	-23450	-22524
Channel No.4	8764	9553	9117	9913	8937	9874

Results of measurements in cross-section 4:

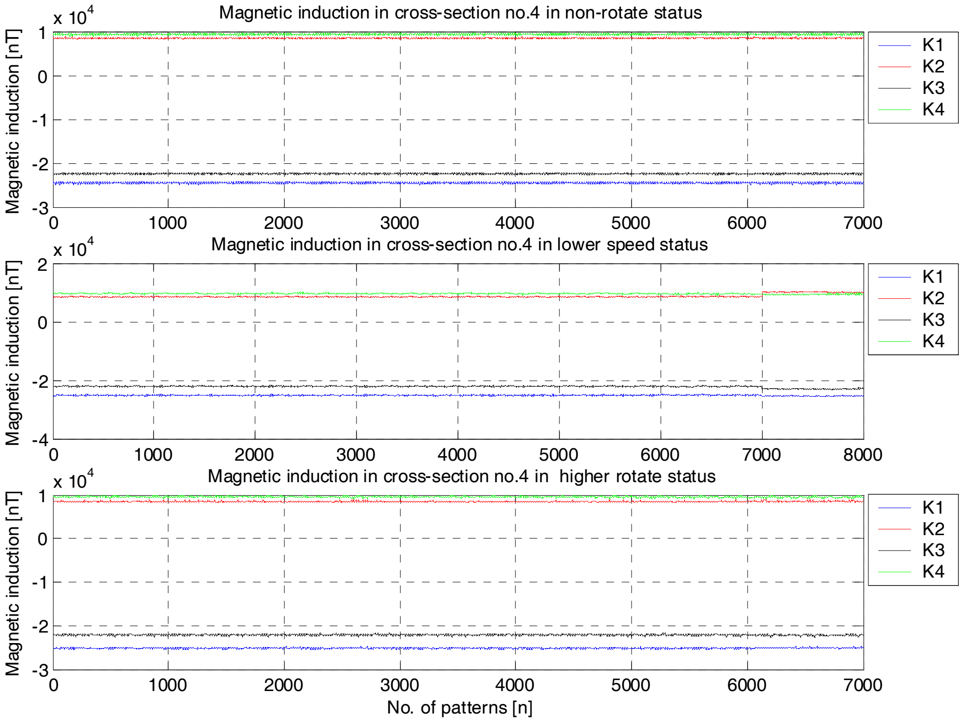


Figure 7 Results of measured magnetic induction in cross-section 4

Table 5 Boundary values for each run in cross-section 4

	Stand still		Lower RPM		Higher RPM	
	min [nT]	max [nT]	min [nT]	max [nT]	min [nT]	max [nT]
Channel No.1	-24933	-24149	-25749	-24693	-25484	-24782
Channel No.2	8216	8820	8190	10365	8195	8841
Channel No.3	-22716	-22109	-23373	-21688	-22633	-21703
Channel No.4	8902	9735	8982	10044	9098	9992

From measured results depicted in graphs as well as from the tables, where maximal and minimal values for each run are shown, it is clear, that reached revolutions didn't have the influence on MPM 20 magnetic aura.

Because the maximal reached RPM were 2000, it is not possible to consider, how the magnetic aura of MPM 20 will react in higher revolutions. Therefore it will be necessary to conduct similar measurements on MPM 20 placed directly on the stand during cold overspeed.

4 CONCLUSION

As was already stated before, after these measurements, it' s impossible to evaluate the influence of the revolutions of engine on its magnetic aura. From these measurements it's possible to come to the conclusion, that the revolutions don't have the influence on engine's magnetic aura. Therefore it's impossible to consider the possibilities of improvement for algorithms for situational control of MPM 20. This control was already designed and is described in [1,4].

For better understanding of relation between MPM 20 revolutions and its magnetic aura, it will be necessary to execute other measurements in given sections during cold overspeed when higher RPM could be attained. Based on them, the influence of magnetic aura can be more precisely determined, which will allow the modification of now available situational control of MPM 20.

The work was supported by projects: VEGA no. 1/0394/08 – Situational control algorithms and large scale systems modeling.

References

[1] ANDOGA, R.: *Hybridné metódy situačného riadenia zložitých systémov*. Doktorandská dizertačná práca, FEI TUKE, Košice : 2006, pp.120.

[2] ČOPÍK, J.: *Monitorovania magnetického poľa v záujmovom priestore*., Doktorandská dizertačná práca, LF TUKE, Košice : 2005, pp.135.

[3] ČOPÍK, J.: *Autonómne meranie vybraných fyzikálnych polí*., Písomná práca k dizertačnej skúške, Vojenská letecká akadémia gen. M. R. Štefánika, Košice : 2002, pp.49.

[4] FŐZŐ, L.: *Využitie matematického modelu rovnovážneho a nerovnovážneho chodu motora MPM20 pri návrhu algoritmu riadeniav každom čase,* Doktorandská dizertačná práca, FEI TUKE, Košice : 2008, pp.142.

[5] HUDÁK, J.: *Magnetometria, vybrané problémy vývoja a využitia*, Habilitačná práca, Vojenská letecká akadémia gen. M. R. Štefánika, Košice : 1999, pp.107.

[6] KABÁT, J., MADARÁSZ L., MODROVIČOVÁ J.: *Introduction to problem, basis and measurement of magnetic aura of turbojet aircraft engines,* In: Computional Intelligence and Informatics: Proceeding of the 9th International Symposium of Hungarian Researchers, November 6-8, Budapest : ISBN 978-963-7154-82-9, 2008, pp. 379-386.

[7] MODROVIČOVÁ J.: *Štúdia možnosti vplyvu magnetickej aury leteckého motora MPM 20 na jeho situačné riadenie.*, Písomná práca k dizertačnej skúške, FEI TUKE, Košice, 2009, pp.85.

[8] MODROVIČOVÁ, J., MADARÁSZ, L., KABÁT, J.: *Magnetic aura study of small turbojet engine MPM 20*. In: SAMI 2009 : 7th International Symposium on Applied Machine Intelligence and Informatics : January 30-31, 2009, Herľany, Slovakia. [S.l.] : IEEE, 2009. s. 25-28. ISBN 978-1-4244-3802-0.

[9] MODROVIČOVÁ, J.: *Magnetic aura of Small Turbojet Engine MPM 20*. In: SCYR 2009 : 9th Scientific Conference of Young Researchers : Proceedings from conference : May 13th, 2009 Košice, Slovakia. Košice : FEI TU, 2009. s. 213-216. ISBN 978-80-553-0178-5.

Dipl. Eng. Jana MODROVIČOVÁ
Technical University of Košice
Letná 9
041 21 Košice
Slovak Republic
E-mail: jana.modrovicova@tuke.sk

Prof. Dipl. Eng. Ladislav MADARÁSZ, PhD.
Technical University of Košice
Letná 9
041 21 Košice
Slovak Republic
E-mail: ladislav.madarasz@tuke.sk

Dipl. Eng. Ján KABÁT
Technical University of Košice
Rampová 7
041 21 Košice
Slovak Republic
E-mail: jan.kabat@tuke.sk

MULTIPLE SENSOR INTEGRATION FOR AUTONOMOUS VEHICLE NAVIGATION

Martin Ernesto OREJAS, Ľuboš VÁCI, Milan SOPATA

Abstract: This paper presents a software solution for multiple sensor integration for autonomous vehicle navigation. The system unique features and shortcomings are taken into account to implement the fusion of an inertial navigation system (INS) with measurements from global navigation satellite system (GNSS), altimeter, odometer and magnetometer employing a complementary extended Kalman filter to produce an accurate, reliable and robust navigation solution. The aim of this work is to present specific problems and solutions of sensor integration.

Keywords: INS, GNSS, Kalman filter, navigation system integration, sensor fusion.

1 INTRODUCTION

Nowadays there is an increasing demand for low-cost, low-weight precise navigation systems, a clear example being unmanned vehicles (UAVs), and the development of such systems has been boosted by the advent of low-cost MEMS inertial sensors. This has had a profound impact in the development of integrated GNSS/INS systems; while the equations to integrate INS with GNSS are well known, new challenges arise when using low-cost COTS sensors [1]. The inclusion of other sensors as barometer, odometer and magnetometer in the navigation solution becomes of major importance when trying to overcome some of the inherent problems that appear when working with MEMS sensors.

The integration of inertial navigation systems with global navigation satellite systems is a logical step due to the complementary characteristics of each system. An INS is precise for a short span of time, can be run at high frequency and doesn't depend on external signals however, will drift over time if it is not corrected due to bias accumulation in the process of integrating accelerations and angular rates. On the other hand, a GNSS has a bounded error but run at slower frequencies and is dependent on external signals, therefore can suffer outages. In these situations, a filtering technique called the complementary filter allows an optimal filter to be designed to minimize the effect of the errors on the signal estimate [2].

In the complementary-filter approach, the INS is the primary navigation system that calculates the navigation states at high rate and uses the measurements from aiding sensors coming at lower rates. In our system, a loosely coupled integration was used for the INS/GNSS integration, expanded with additional sensors like barometer, magnetometer and odometer. A complementary extended Kalman filter was implemented to estimate the error of the INS states.

The paper is organized as follows: In chapter 2 the equations necessary to implement INSs are described. In chapter 3 we present the error equations that will be necessary to perform the error state estimation. The models used for the inertial sensors are introduced in charter 4. Chapter 5 deals with the equations used to develop the complementary EKF and the next chapter give a short description of the fault detection and exclusion algorithm used evaluate the occurrence of a wrong measurement. Chapter 7 contains some brief comments about the integration of additional aiding sources to our system. The following chapter outlines the basic steps for the online initialization and alignment of the IMU platform. The results of our simulations are showed in charter 9. Finally some remarks about the results and the future plans are provided.

2 INS MECHANIZATION EQUATIONS

The basic concept [2], [3], [4], of an INS is to integrate accelerations to determinate velocity and position in a desired coordinate frame. Considering the possibility of relative angular motion between frames, gyroscopes are required to maintain the sensor-to-navigation frame transformation. Figure 1 illustrates the functional and operational concept of INS mechanization.

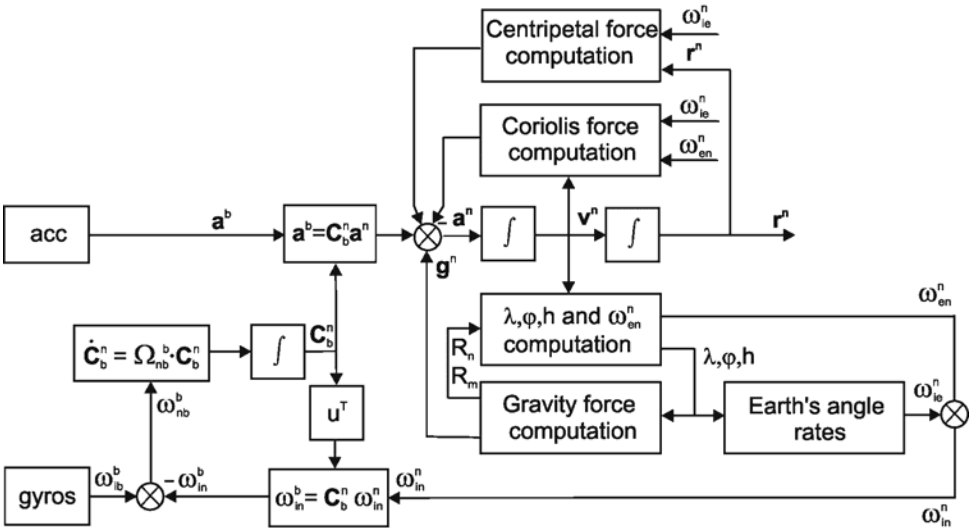


Figure 1 INS block diagram

The continuous time INS mechanization equations are expressed by following set of equations [2].

$$\dot{\boldsymbol{r}}^n = \boldsymbol{D}^{-1}\boldsymbol{v}^n$$
$$\dot{\boldsymbol{v}}^n = \boldsymbol{C}_b^n\boldsymbol{a}^b - \left(2\boldsymbol{\Omega}_{ie}^n + \boldsymbol{\Omega}_{en}^n\right)\boldsymbol{v}^n + \boldsymbol{g}^n - \boldsymbol{\Omega}_{ie}^n\boldsymbol{\Omega}_{ie}^n\boldsymbol{r}^n$$
$$\dot{\boldsymbol{C}}_b^n = \boldsymbol{C}_b^n\left[\boldsymbol{\Omega}_{ib}^b - \boldsymbol{C}_n^b\left(\boldsymbol{\Omega}_{en}^n + \boldsymbol{C}_e^n\boldsymbol{\Omega}_{ie}^e\right)\right] \tag{1}$$

Where

$$\boldsymbol{D}^{-1} = \begin{bmatrix} 0 & \dfrac{1}{(R_n + h)\cos\varphi} & 0 \\ \dfrac{1}{(R_m + h)} & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{2}$$

Where

$\boldsymbol{r}^n$ = position in navigation frame,

$\boldsymbol{v}^n$ = velocity in navigation frame,

$\boldsymbol{g}^n$ = gravity in navigation frame,

$\boldsymbol{a}^b$ = acceleration in body frame,

$\boldsymbol{\Omega}_{ie}^n$ = antisymmetric matrix of angular rate between inertial and Earth frame expressed in navigation frame (Earth rotation),

$\boldsymbol{\Omega}_{en}^n$ = antisymmetric matrix of angular rate between Earth and navigation frame expressed in navigation frame (transit rate),

$\boldsymbol{\Omega}_{ib}^b$ = antisymmetric matrix of angular rate between inertial and body frame expressed in body frame (sensors measurement),

$\boldsymbol{C}_b^n$ = Direction cosine matrix (DCM) from body to navigation frame,

$\boldsymbol{C}_e^n$ = Direction cosine matrix from Earth-Centred-Earth-Fixed to navigation frame.

The attitude equation was implemented using quaternion algebra to avoid singularities in attitude angles computation. The attitude dynamics expressed by quaternions are

$$\dot{\boldsymbol{q}} = -\frac{1}{2}\boldsymbol{\Omega}_{nb}^b \cdot \boldsymbol{q} \tag{3}$$

3 STRUCTURE

A classical approach for the development INS error equations is by perturbation analysis, where the navigation parameters are perturbed with respect to the true navigation values. This approach applies the Taylor series expansion on the continuous INS mechanization equations and retains only the constant and linear terms. The derivation of the perturbation model is described in many studies and literature [2], [3], [4], [5], [6], therefore only a brief derivation will be presented.

Velocity errors are simply the error in the velocities computed in the navigation frame

$$\tilde{\boldsymbol{v}}^n = \boldsymbol{v}^u + \delta\boldsymbol{v}^n \tag{4}$$

Position errors are obtained from the Earth-Centred-Earth-Fixed to navigation frame DCM, where the computed or perturbed DCM is represented as

$$\tilde{\boldsymbol{C}}_e^n = \left[\boldsymbol{I} - (\delta\boldsymbol{\theta}\times)\right]\boldsymbol{C}_e^n \tag{5}$$

In a similar way attitude errors are obtained from the navigation to body frame DCM.

$$\tilde{\boldsymbol{C}}_b^n = \left[\boldsymbol{I} - (\phi\times)\right]\boldsymbol{C}_b^n \tag{6}$$

If we assume the error equations for measured acceleration, computed gravity vector and computed and measured angular rates as

$$\begin{aligned}
\tilde{\boldsymbol{a}}^{u} &= \boldsymbol{a}^{u} + \delta\boldsymbol{a}^{u} \\
\tilde{\boldsymbol{g}}^{n} &= \boldsymbol{g}^{n} + \delta\boldsymbol{g}^{n} \\
\tilde{\boldsymbol{\omega}}_{iu}^{u} &= \boldsymbol{\omega}_{iu}^{u} + \delta\boldsymbol{\omega}_{iu}^{u} \\
\tilde{\boldsymbol{\omega}}_{ie}^{n} &= \boldsymbol{\omega}_{ie}^{n} + \delta\boldsymbol{\omega}_{ie}^{n} \\
\tilde{\boldsymbol{\omega}}_{in}^{n} &= \boldsymbol{\omega}_{in}^{n} + \delta\boldsymbol{\omega}_{in}^{n}
\end{aligned} \tag{7}$$

We get the following set of equations describing the INS perturbation model [3]

$$\begin{aligned}
&\delta\dot{\boldsymbol{\theta}} = \delta\boldsymbol{\omega}_{en}^{n} - \boldsymbol{\omega}_{en}^{n} \times \delta\boldsymbol{\theta} \\
&\delta\dot{\mathbf{v}}^{n} = \mathbf{v}^{n} \times (2\delta\boldsymbol{\omega}_{ie}^{n} + \delta\boldsymbol{\omega}_{in}^{n}) - (2\boldsymbol{\omega}_{ie}^{n} + \boldsymbol{\omega}_{in}^{n}) \times \delta\mathbf{v}^{n} + \mathbf{a}^{n} \times \phi + \mathbf{C}_{b}^{n}\delta\mathbf{a}^{b} \\
&\dot{\phi} = \delta\boldsymbol{\omega}_{en}^{\mathbf{n}} + \boldsymbol{\omega}_{\mathbf{ie}}^{\mathbf{n}} \times \delta\boldsymbol{\theta} - \boldsymbol{\omega}_{in}^{n} \times \phi - \mathbf{C}_{b}^{n}\delta\boldsymbol{\omega}_{ib}^{b}
\end{aligned} \tag{8}$$

4 SENSORS MODELING

Although the importance of having a good model for the sensors is often neglected, the way the sensors are modeled has a great influence in the overall system performance. Which errors will be taken into account will not only depend on the sensor type and quality but also the application and the environment where the sensors and system are to operate. For low-grade sensors, as the ones used in our system, usually three sources of errors are modelled: a constant bias, a varying bias and white noise. There are many possibilities when modelling a varying bias and a common method is to use a Gauss-Markov (GM) process, or as a particular case of this, a random walk. Thus, the sensor output can be expressed as

$$\begin{aligned}
&\tilde{\boldsymbol{y}} = \boldsymbol{y} + \delta\boldsymbol{y} \\
&\delta\boldsymbol{y} = \boldsymbol{b}_{\text{const}} + \boldsymbol{b}_{\text{GM}} + \boldsymbol{w} \\
&\dot{\boldsymbol{b}}_{\text{GM}}(t) = \frac{-1}{\tau}\boldsymbol{b}_{\text{GM}}(t) + \boldsymbol{w}_{\text{GM}}
\end{aligned} \tag{9}$$

where w is zero mean white noise and can be determined by method in [7], w_{GM} is white noises which drives GM process. Time constant of GM is defined by τ and can be determined by method in [8]. This model can be further simplified if an estimation of the constant bias is done in the initialization routine leaving only the varying bias and the white noise as the error sources of the sensor.

5 COMPLEMENTARY EXTENDED KALMAN FILTER

In an INS implemented in the complementary filter structure, the output of the INS provides the navigation solution and the EKF estimates the INS errors. The INS error vector is fed back to correct the INS internal states [2] and [3]. This is illustrated in Figure 2.

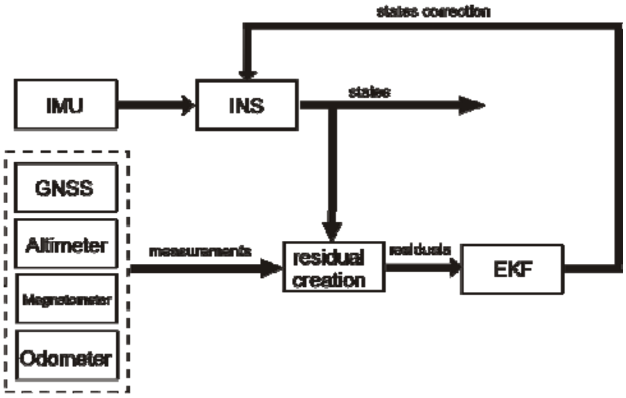


Figure 2 Feedback implementation of the complementary extended Kalman filter

The nonlinear extended Kalman filter is by far the most common method used for integrating INS with GPS. In fact, the error equations presented in chapter 2 are already a linear version of the actual INS error equations so it is not necessary to compute any Jacobean to obtain *H* and *F* matrices, required for the Kalman filter. Instead, *F* and *H* matrices can be obtained directly from the errors equations such that they satisfy

$$\begin{aligned}
\partial\dot{x}(t) &= F\big(x(t)\big)\cdot\partial x(t) + \Gamma\big(x(k)\big)w(t) \\
y(t) &= H(x(t),t)\cdot\partial x(t) + v(t)
\end{aligned} \tag{10}$$

where y represent the residuals created from the INS states and the measurements (coming from the external sensors) and *w* and *v* are white noises with respective covariances

$$\operatorname{cov}\big(w(t)\big) = Q \qquad \operatorname{cov}\big(v(t)\big) = R \tag{11}$$

Before introducing the Kalman filter equations we need to discretize the continuous system described above. It is assumed that the transition matrix *F* is constant over the sampling period *Ts*), therefore

$$\Phi\big(x(k)\big) = e^{F(x(k))\cdot Ts} \tag{12}$$

A truncated Taylor series can be used to compute the previous equation. To obtain the covariance $Q_d(k)$ of the equivalent process noise $w(k)$ many approximations are available. One of the most simple and common is

$$Q_d(k) \approx \Gamma(x(k)) \cdot Q \cdot \Gamma(x(k))^T \cdot T_s \tag{13}$$

A slightly more complicated formula is implemented in our system but its description is out of the scope of this work.

The Kalman filter has two main steps: the *time update* and the *measurement update*. The *time update* only depends on the system and computes how the estimated errors $\Delta\hat{x}^-$ and their covariances $P^-(k)$ propagate through one sampling period.

$$\Delta\hat{x}^-(k) = \Phi(k-1)\Delta\hat{x}(k-1) \tag{14}$$

$$P^-(k) = \Phi(k-1)P(k-1)\Phi(k-1)^T + Q_d(k-1) \tag{15}$$

The second step is where the measured data (the computed residuals in the complementary filter) is used to improve the estimation of the errors and to compute the new covariance *P(k)*. First the Kalman filter gain *K(k)* is computed as

$$K(k) = P^-(k)H(k)^T\left[R(k) + H(k)P^-(k)H(k)^T\right]^{-1} \tag{16}$$

K is then used to compute *P*. There are different techniques to compute the update of *P* and although, theoretically, they are all equivalent they have different numerical properties. The equation implemented in our system was chosen because of its numerical stability.

$$P(k) = [I - K(k)H(k)]P^-(k)[I - K(k)H(k)]^T + K(k)R(k)K(k)^T \tag{17}$$

Finally the error estimation is computed as

$$\Delta\hat{x}(k) = x^-(k) + K(k)\left[\tilde{y}(k) - \hat{y}(k)\right] \tag{18}$$

This estimated error will be used to correct the INS states.

6 RESIDUAL TEST

The inclusion of a wrong measurement in the EKF can cause a severe degradation of the navigation solution, therefore, when developing a system to be implemented in a real application it is essential to have some kind of Fault Detection and Exclusion (FDE) algorithm to avoid incorporating erroneous measurements to the filter. This was done using the residuals between the aiding sensors measurements and their predicted values obtained by the INS. From the residuals and their statistical properties a scalar test statistic chi-square distributed with n degrees of freedom is created, where n is the number of measurements used for creating the test statistic. This statistic is later compared with a predefined threshold to evaluate if a failure has occurred. Using the chi-square distribution allows us to test together a group of measurements that are correlated to each other improving the chances to successfully detect a failure.

7 ADDITIONAL SENSORS

As mentioned earlier, when working with low-grade inertial sensors the addition of redundant sensors not only has the potential to greatly reduce the characteristic drift of the unaided INS but also to improve the navigation solution even when GPS is available. For this project three additional sensors were added to our integrated system: a barometer, an odometer and a magnetometer. The barometer is getting a standard addition in this kind of systems and helps to stabilize the vertical channel. The magnetometer is used to assist with the attitude estimation, especially in the absence of GPS signal (nevertheless tests showed that even if GPS is available, the improvement in heading estimation due the addition of the magnetometer is substantial). The whole 3D magnetic vector is utilized for the integration and extra states were added to the Kalman filter to estimate and compensate the iron distortions. The significant benefit of using the magnetometer is apparent when testing the system under extreme conditions as it is shown in charter 9. For the odometer a one wheel model was used for the integration and currently the two wheels model is being developed. Another extra state was included in the Kalman filter to estimate the scaling.

8 INITIALIZATION – ALIGNMENT

A large contribution to INS performance degradation is caused by wrong initialization. Moreover, large initial system uncertainties could lead to Kalman filter divergence and destabilize the closed loop system. The alignment often consists of two modes: *coarse levelling mode*, used to estimate tilt (roll and pitch), and a *heading alignment mode*, used to estimate heading. The coarse levelling mode is enabled and runs when the vehicle is stationary and uses information from the accelerometers to estimate tilt. MEMS precision makes gyro compassing unfeasible, therefore heading can't be estimated from the IMU alone while stationary, additional sensor aids have to be incorporated in order to obtain heading. Heading alignment mode is dividend in two: *in-motion alignment mode* and *extended coarse alignment mode*. In-motion alignment is enabled when the platform starts moving and the GPS velocity is used to estimate the heading. Finally, in the extended coarse alignment mode a complementary filter is used to refine the initial attitude estimation with information from additional sensors (gyroscopes and magnetometer).

9 SIMULATIONS –TESTS

Several simulations were done in order to analyze the effect of each component of the system on the overall performance. A trajectory generator was used to create the true trajectory used for comparison and to simulate the data from the different sensors. The errors for position, velocity and attitude under normal conditions are shown in Figure 3.

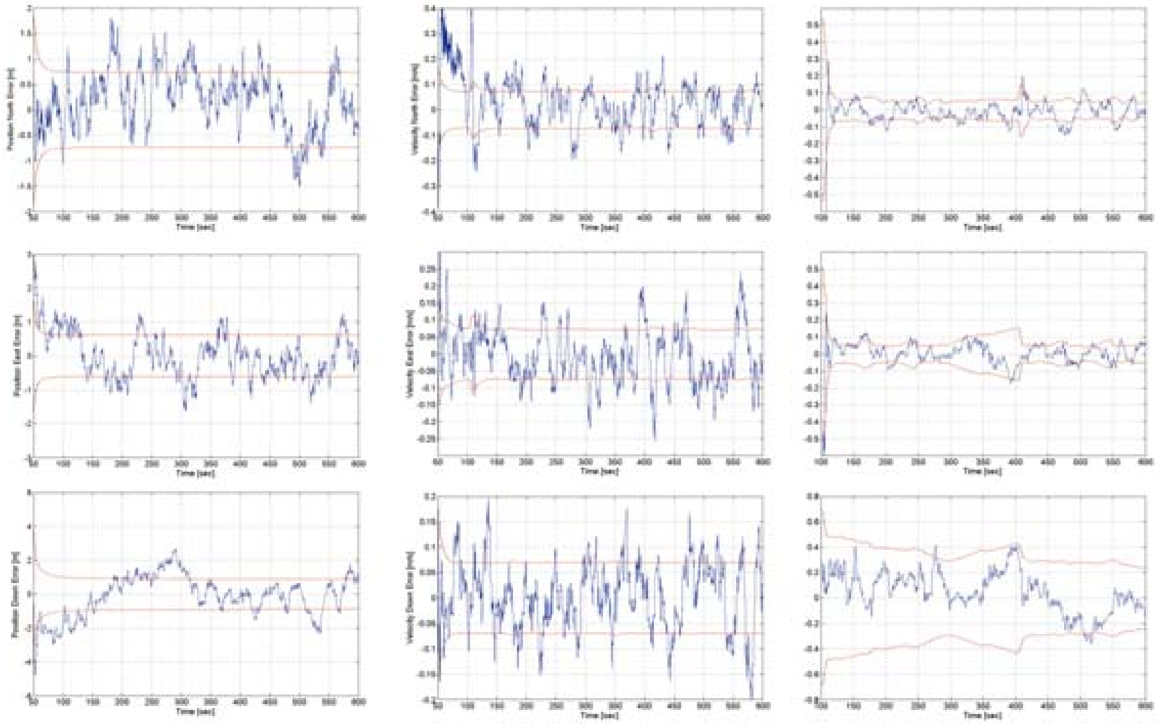


Figure 3 Error of Navigation Solution for position, velocity and attitude. The red lines represent the standard deviation of the error computed by the Kalman filter.

Sensors characteristics are listed in table below.

Table 1 Sensor characteristics

Sensors	Noise characteristics		
	Bias	White noise standard deviances	GM process time constant 1/999,95 [s]
Accelerometers	7.10^{-2} [ms^{-2}]	1.10^{-2} [ms^{-2}]	$9,00018.10^{-6}$ [ms^{-2}]
Gyroscopes	2.10^{-4} [$rads^{-1}$]	5.10^{-5} [$rads^{-1}$]	$2,25.10^{-10}$ [$rads^{-1}$]
GNSS position	none	5 [m]	none
GNSS velocity	none	0,17 [ms^{-1}]	none
Altimeter	none	8 [m]	none
Magnetometer	3.10^{-2} [Gauss]	$6,5.10^{-3}$ [Gauss]	none

Simulations showed that the steady-state bounds for position and velocity errors were primarily defined by the GPS errors, i.e. if GPS signal was available the addition of low-cost inertial sensors did not have a significant influence in the position and velocity errors. On the other hand, attitude errors arehighly dependent on the inertial sensor characteristics. Also, in case of GPS outage, the inertial sensors quality will define the rate of drift of the navigation solution. This is illustrated in Figure 4 and Figure 5, where a GPS outage if 70 seconds was simulated. The fact that our residual test is able to detect this outage/failure explains the increase in the computed standard deviation and also point out the importance of having such a test, to avoid the inclusion of wrong measurements that could potentially destabilize the system.

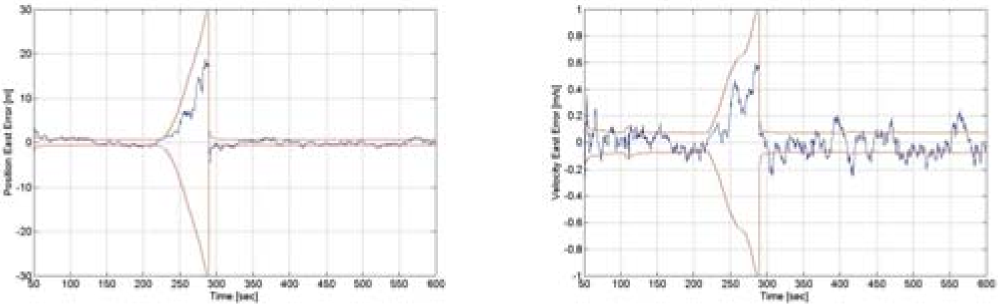


Figure 4 Position and velocity errors (east) when GPS is unavailable

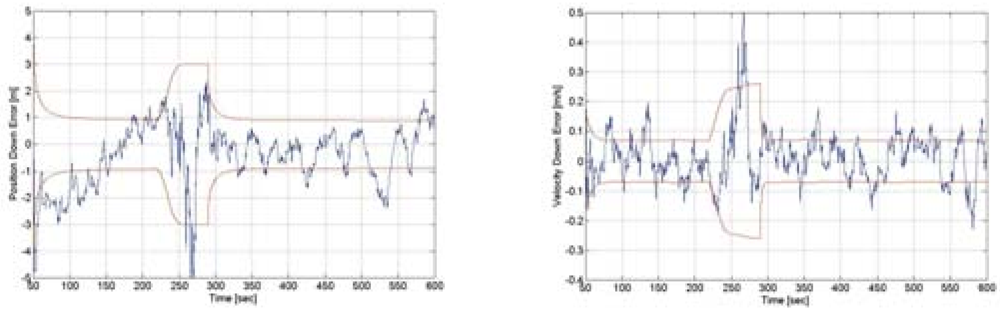


Figure 5 Position and velocity errors (down) when GPS is unavailable

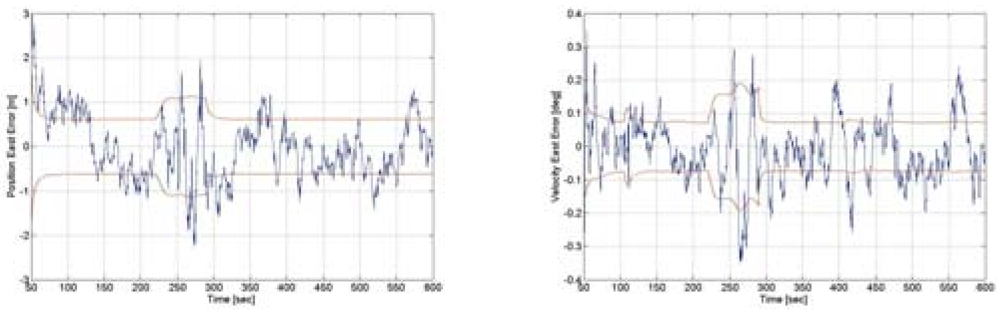


Figure 6 Position and velocity errors (east) when only GPS velocity is unavailable

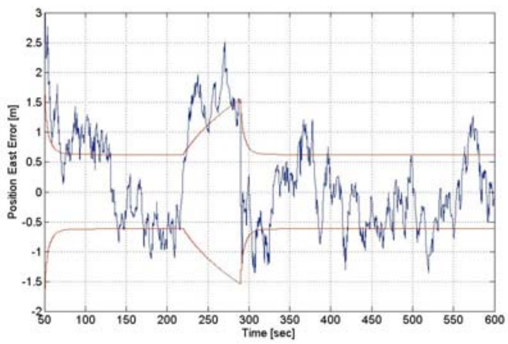


Figure 7 Position errors (east) when GPS is unavailable and odometer is enabled

It's interesting to notice the effect of the barometer in the vertical position and velocity estimation and how, even in the absence of GPS, position and velocity errors are bounded for this channel due to the redundancy provided by this sensor.

The advantage of having a FDE algorithm capable of testing different groups of measurements together becomes apparent in Figure 6, where a failure in GPS velocity was simulated (odometer was switched off) while GPS position was still available. Clearly the Kalman filter still uses the GPS position data to compute the corrections and avoid a drift in the navigation solution and therefore only a slight increase in position and velocity errors is observed.

Another clear example of the advantage of having redundant measurements and being able to test them separately is shown in Figure 7 where a failure was simulated for the GPS signal while the odometer was enabled. As expected, the position error starts growing but the availability of the velocity measurement coming from the odometer greatly reduce the rate of this growth.

Finally, one of the experiments that were performed under extreme conditions is shown to depict the influence of the magnetometer in the attitude estimation, particularly the heading estimation. The experiment was carried out over an extremely irregular terrain with a heavily oscillating lever arm. Figure 8 describes the heading error (the error was obtained comparing with the heading computed from GPS velocity data) of the integrated system when the magnetometer is enabled and disabled. The three small plots on the right show the extremely noisy measurements collected with the gyroscopes, accelerometers and magnetometer. It is clear that under these conditions the system is unable to track the heading unless the magnetometer is enabled.

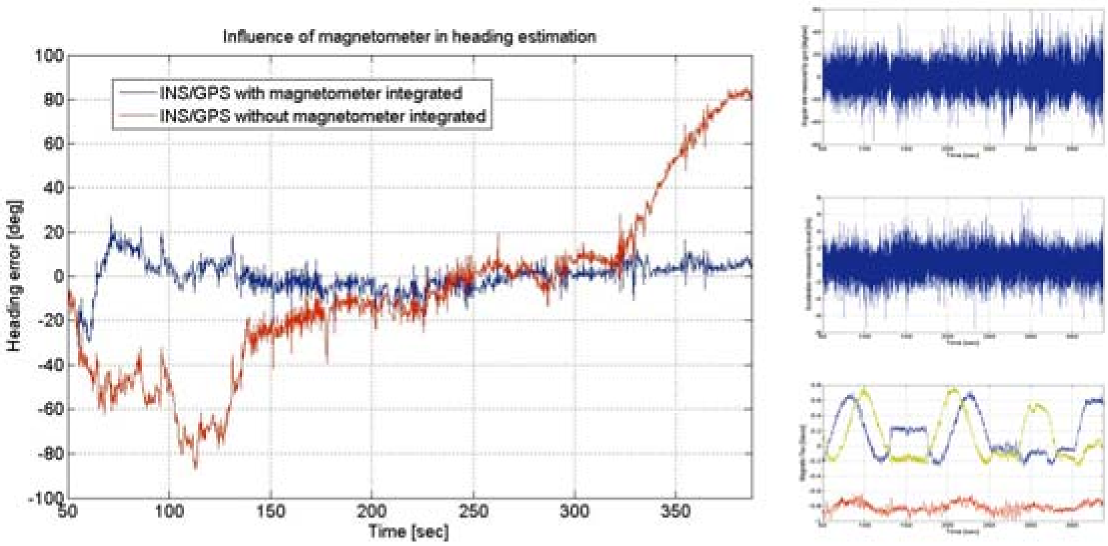


Figure 8   Heading error when magnetometer is switched on/off. On the right: Gyroscope, accelerometer and magnetometer data

10 CONCLUSION

In this paper we discussed the basic principles of integrating inertial navigation systems with global navigation satellite systems and the issues that arise when implementing this in a real system with low-cost COTS sensors. Many design and implementation issues, not covered in the literature, emerge when applying the somehow standard equations use for INS/GNSS integration. Both, simulations and tests, confirmed the importance of adding additional sensors and a FDE algorithm to deal with real application issues like outage of GPS signal or failure in the GPS receiver and also noisy and poor environmental and dynamic conditions. In the near term future, a tightly coupled integration itis planned to be developed, this will allow not only a more accurate solution, due to the increase in the amount of observables and the improvement in the tracking loops of the GNSS receiver, but also will permit the GNSS keep aiding the INS even if less than the minimum four satellites, usually needed to obtain a PVT solution, are available.

References

[1] BROWN, K. A.: *GPS/INS uses Low-Cost MEMS IMU*, Aerospace and Electronic System Magazine, IEEE, Volume 20, Issue 9, pp 3-10.

[2] FARRELL, A. J., and BARTH, M.: *The Global Positioning system and Inertial Navigation System*, McGraw Hill, 1998, ISBN: 0-07-022045-X.

[3] ROGERS, M. R.: *Applied Mathematics in Integrated Navigation Systems*, 2d ed., AIAA Education series, 2003, ISBN: 1563476568.

[4] SHIN H. E.: *Estimation technique for Low Cost Inertial Navigation*, Ph.D. thesis, Department of Geometrics Engineering, University of Calgary, May 2005.

[5] TITTERTON, H. D. and WESTON, J.: *Strapdown Inertial Navigation Technology*, The Institution of Electrical Engineers, 2d ed., 2004.

[6] ZARCHAN, P. and MUSOFF, H.: *Fundamentals of Kalman filtering: A Practical Approach,* AIAA, 2005, ISBN: 1-56347-694-0

[7] SOTAK, M.: *Determining stochastic parameters using an unified method*, Acta Electrotechnica et Informatica, Volume 9, No. 2 (2009), pp 59-63, ISSN: 1335-8243.

[8] SOTAK, M.: *The parameters evaluation for inertial sensors models*, Acta avionica, Volume 11, No. 17 (2009), pp 46-49, ISSN: 1335-9479.

M.Sc. Martin Ernesto OREJAS
Honeywell International s.r.o.
Spielberk Office Centere
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: martin.orejas@honeywell.com

Dipl. Eng. Ľuboš VÁCI
Honeywell International s.r.o.
Spielberk Office Centere
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: lubos.vaci@honeywell.com

Assoc. Prof. Dipl. Eng. Milan SOPATA, Ph.D.
Honeywell International s.r.o.
Spielberk Office Centere
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: milan.sopata@honeywell.com

PASSIVE SAFETY OF CAR FLEET IN THE CZECH REPUBLIC

Jan POKORNÝ, Michal ŠELMÁT

Abstract: Safety of car fleet is usually estimated according to its age. But this estimation is quite inaccurate and gives only information about the average age of common cars in a territory. It does not consider passive and active safety structure of that car fleet. This article shows possibility involving some of these parameters to this estimation. A method introduced in this paper is based on fuzzy sets application to this problem and appears from data obtained from Central Register of Vehicles and EuroNCAP.

Keywords: Passive safety, car fleet, fuzzy set.

1 INTRODUCTION

In the Czech Republic there are annually published articles about average age of car fleet based on Central register of vehicles. This average age is usually associated with general level of safety and ecological aspects of cars included in this car fleet. Such estimation does not work deeply with the structure of this car fleet. It only says how old on average the common car is in a territory. From the view of safety, it does not take such important aspects as passive and active safety and weight of cars into consideration.

For better safety estimation of the car fleet it would be useful to append to data from the Central register of vehicles at least some information about weight and safety aspects of each car model involved in the car fleet. So each car model may be characterized by four basic parameters:

- number of registered cars,
- age,
- level of safety,
- weight.

These parameters, especially age, level of safety and weight, are close to description of total safety of each car model.

2 DATA SOURCES

2.1 Central register of vehicles

The best information source about structure of car fleet in the Czech Republic should be Central register of vehicles (CRV). This is directly connected to all vehicle registration offices in the country and keeps statistics of all registered vehicles.

Data in CRV are organized in a number of files and divided according to various criteria. One of these files is useful for next work. Vehicles in this file are organized according:

- Kind (OA - passenger vehicle, NA – goods vehicle).
- Category (for OA next dividing on M1, M2, M3).
- Car make (name of the vehicle producer – e.g. Škoda).
- Type (name of the vehicle model – e.g. Octavia).
- Model year (number of vehicles produced in the given year, divided by single years from 1945 to 2007; earlier made vehicles are put in column “older”).
- Sum (sum of all vehicles of given kind, category, car make and type for all years).

All of these data are entered to the register by employees of vehicle registration offices and nobody else has the capacity to modify it. On this account there are lots of mistakes in the register and these are good to repair before further work with this data. There are usually key mistakes in the category or the car make, which create a new category or car make that doesn’t really exist. Sometimes there are two or more rows for one car make, but they are named by another way (for example “Volkswagen” and “VW”), which are important to be link together. It is also important to class vehicles which have no mark on position of kind or category to the right place. Some records are not possible to be repair because they are unidentifiable. These are not able to be used for further work. All of these and more corrections must be made before filtering of the data to receive only a file of passenger vehicles.

In our situation we processed the data of CRV from the year 2007. We set a condition to obtain 95% car fleet in the beginning. Originally we had 4 253 153 vehicles in the category M1. After all corrections we obtain 4 018 207 cars, which is 94.5%. We decided to exclude cars which are extremely old (made before 1945 – veteran cars) or cars, which are registered only in a few pieces (under 500 pieces). These cars made up 3% of all vehicles in the category M1. The final file contains 281 various models of cars. [5]

2.2 Level of safety

Passive or active safety is difficult to measure and thanks to this there is an opportunity for polemics about safety of concrete vehicles.

Generally it is only possible to assume the quality in this way according to the car outfit that should protect passengers.

We decided to use results of crash tests done by EuroNCAP organization as a basic method for evaluation of passive safety. Results of these tests are regularly published and in one place it is possible to find all tests that have ever been done. The Test procedure and results are clear and single models of cars are possible to compare.

Of course, we know about some limits and problems of these tests in context of car fleet safety. One of them is a problem concerning the difference between the tested cars and selling cars. For this test the car chosen is the one in the best selling version. But later on the road there are lots of versions more or less different from the tested model. Other problems make changes in the test procedures. They are not so often, there have been only two changes since the 1996, but the results before and after the change are not fully comparable. However, we assume that they still well reflect the level of passive safety of the cars.

As known, the results of these tests are published with a number of obtained stars (from 1 to 5, where 5 stars is the best classification) which corresponds to the number of points obtained during particular test procedures. These results were associated to concrete model years of cars in final file from CRV.

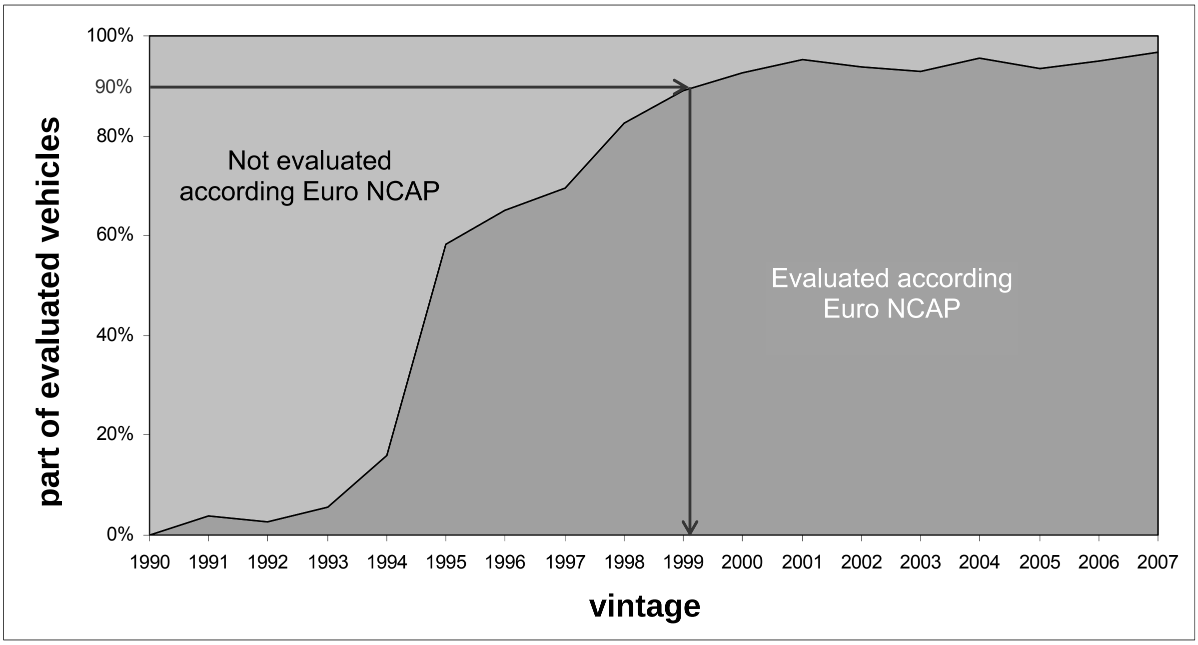


Figure 1 Part of evaluated vehicles according Euro NCAP [5]

But there is another problem which rises from the history of these tests. The first tests were made in 1996. But the final file contains lots of vehicles made before 1996. The Number of vehicles that could be associated with the final file was in our case 1 514 267. This number was increased to 2 035 730 after an estimation of the passive safety of some of older car models according to crash tests done by other companies. Although it was 90% of 1999´s cars which were evaluated (Figure 1), this was still only half of all vehicles. Here is no other way to solve this problem than to make a rough estimate of the rest of the car fleet. Most of these cars were made before the first tests of Euro NCAP were published. So it is possible to suppose that these cars are on the same or worse level of passive safety than the first tested cars. In some cases there are known some older two vehicles crash tests or there are also other possibilities how to classify older cars - for example classification according to Volksam [7]. It has lots of data from two real car accidents which date many years back. So it is not so difficult to find some differences in the safety among the older cars. It is important to keep in mind that the number of vehicles that are not able to be evaluated according to the Euro NCAP, will be decreasing. We worked with data from 2007, but the situation today is better from this view.

2.3 Influence of weight and age on safety

Euro NCAP test procedures are based on single car crash to the barrier or the pole or on the collision with movable barrier, which are the same for all categories of cars. But in real accidents there are more influences to be taken into account:

- technical state and age of vehicles,

- weight of involved vehicles,
- size compatibility of involved vehicles.

Technical state of cars should be all right thanks to regular technical inspections. But it is difficult to imagine that a very old car will have the same safety potential as if it was new. Also, lifetime and reliability of some safety parts is declared only to some period. Such influence like rust also has a negative effect on the car body. All these things can be associated with the age of the car. Of course it depends on the owner, driver, environment and so on, but generally each subject is degraded during the time. Not only physical obsolescence, but also moral or better said technical and technological obsolescence is very important here. There are big technological differences between new and ten years old cars. Also this influence of incompatibility is possibly connected to the age.

Also, the weight of involved cars has physically clear influence on consequences of accidents, especially in two car crashes. It is known that in such crashes a bigger car has an advantage. On the other side in two cars collisions heavier cars are more aggressive than the smaller ones. [6] For this reason each car model from the final file of CRV was associated with its weight.

3 ESTIMATION OF THE SAFETY OF CAR FLEET

Because some of the mentioned parameters are not exact or accurate numbers and parameters and relations between these parameters could be described by verbal expressions, man can use fuzzy system for solving this problem.

Our fuzzy inference system was built in Matlab Fuzzy Logic Toolbox. It contains three incoming variables – EuroNCAP, weight and age (Figure 2).

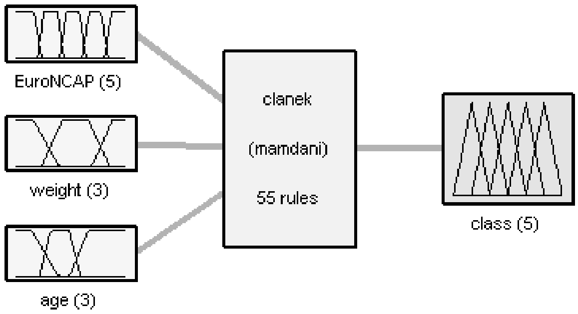


Figure 2 Fuzzy inference system

Fuzzy set EuroNCAP has five trapeze membership functions describing the passive safety of a car – no, little, middle, good and well. Univerzum of this set is a number of points that a car can obtain during the test. Other two fuzzy sets (weight and age) both have three membership functions. Input variable weight is described by these membership functions: light, middle and heavy. Univerzum of this set is made by curb weights of common cars. Input variable age contains membership functions: new, middle-age and old. Univerzum is a set of years of car life from zero to thirty. Output variable was named “class” and it is made by five triangular membership functions, so the results are expected as a real number between one and five. Number one means that the car is not safe, number five is a really safe car – it is like number of stars in Euro NCAP. All fuzzy sets and its membership functions are shown in Figure 3.

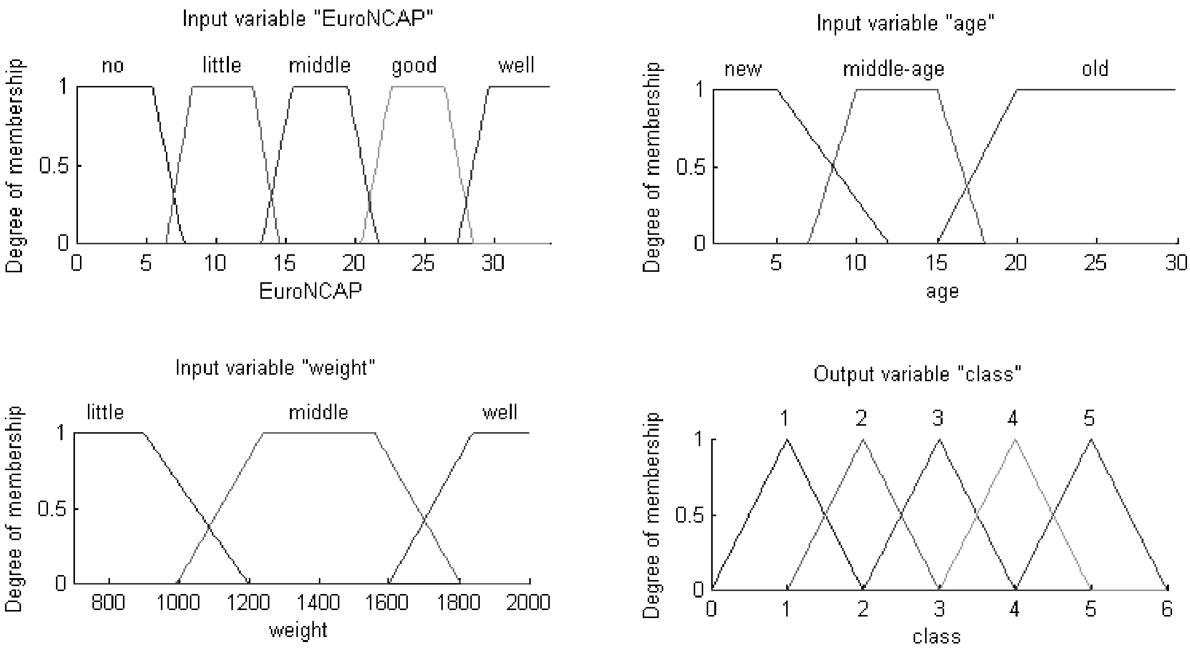


Figure 3 Fuzzy sets – input and output variables and their membership functions

This fuzzy system was built on Mamdani's fuzzy inference system. The base of conditional rules contains 55 rules and for defuzzyfication a centroid method was used. The rules were built from the view of a passenger of a car – it means a heavier car is safer than a light one. From the global view on safety of car fleet rules should be built little bit differently, because heavy cars are usually more aggressive toward smaller and lighter cars and in two vehicle crashes they can cause worse consequences. Relationships between input variables and output variable are shown in F igure 4.

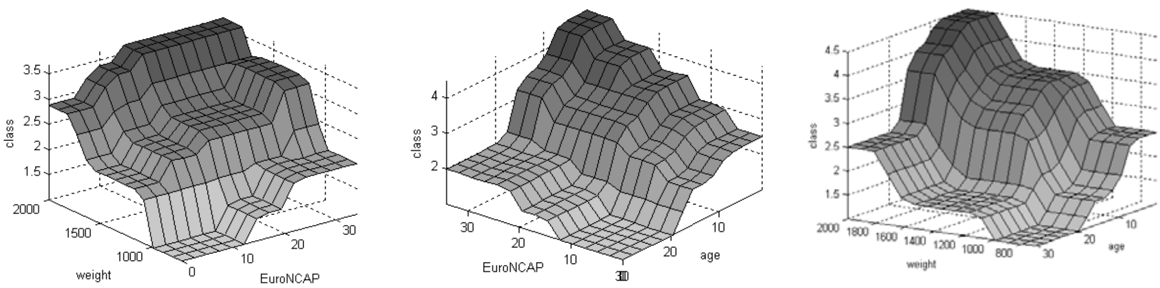


Figure 4 Relationships between input variables and output variable

4 CONCLUSION

Estimation of car fleet safety based only on the average age is not accurate. For better estimation of this safety it is important to think over more parameters like the level of passive safety, weight and size compatibility. Because it is not usually possible to find accurate and perfect data of these parameters for all vehicles from the car fleet, it seems to be good to use a method which is able to work with inaccurate or estimated data. These conditions are fulfilled in the method based on fuzzy sets. We have made such basic fuzzy system which is able to evaluate safety of car fleet in the Czech Republic. Although here is still a large area for improvement of this model, first results seems to be satisfactory. The biggest disadvantage of this method could be seen in lots of mistakes in Central register of vehicles. It takes too much time and energy to repair these mistakes and make other corrections.

References

[1] COX, E.: *The Fuzzy Systems Handbook*, Academic Press, San Diego, USA, 1998, 716 p., ISBN 0-12-194456-5.

[2] European New Car assessment Programme: Test Procedures [online]. [cit. 2008-01-15]. Dostupné z www: <http://www.euroncap.com/testprocedures.aspx>.

[3] OLEJ, V.: Modelovanie ekonomických procesov na báze výpočtovej inteligencie, Miloš Vognar M&V, Hradec Králové 2003, 160 stran, ISBN 80-90324-9-1.

[4] Statistiky: Centrální registr vozidel [online]. [cit. 2007-12-10]. Dostupné z www: <http://www.mvcr.cz/statistiky/crv.html>.

[5] ŠELMÁT, M.: *Hodnocení bezpečnosti vozového parku České republiky*, diplomová práce, Univerzita Pardubice, Pardubice 2008.

[6] TAPIO, J., PIRTALA, P., ERNVALL, T.: *The accident involvement and injury risk rates of car models*, University of Oulu, Oulu, Finland 1995, 95 p., ISBN 951-42-4144-4.

[7] Folksam: *How safe is your car?* 2007 [online]. [cit. 2008-03-19]. Dostupné z www: <http://www.folksam.se/english/reports>.

Dipl. Eng. Jan POKORNÝ
University of Pardubice
Jan Perner Transport Faculty
Studentská 95
532 10 Pardubice
Czech Republic
E-mail: jan.pokorny@upce.cz

Dipl. Eng. Michal ŠELMÁT
University of Pardubice
Jan Perner Transport Faculty
Studentská 95
532 10 Pardubice
Czech Republic
E-mail: m.selmat@seznam.cz

EVALUATION OF A NEW TIMETABLE CONCEPT ON THE TRACK PRAHA-KLADNO USING A SIMULATION TOOL

Ondřej POLÁK, Vít JANOŠ, Karel BAUDYŠ

Abstract: On behalf of České dráhy (the Czech Railways) a railway operation model for the railway track Praha – Kladno was compiled using the Swiss simulation tool „Opentrack“. The aim of the study was to prove the feasibility and robustness of a new timetable concept, which slightly raised the traffic intensity. The study proved the feasibility and pointed out the critical moments of the planned concept. It especially simulated the possible effects of operation irregularities at the critical points. To prove the robustness of the timetable a 5-15 minutes entry delay was simulated for every train in the planned work-day timetable. The conclusions of the simulation were later projected into specific dispatching issues for every station dispatcher as well as the overall operations control. Apparently thanks to the results of the study the influence of delays on this track was minimized. This was an important contribution to improve the marketing image of the new timetable.

Keywords: Operations simulation, Opentrack, robustness of a timetable, railway operations control.

1 INTRODUCTION

The aim of our presentation is at first to introduce the main aspects and possibilities of using the railway operations simulation generally and to describe the most common methods and tools used for railway operations simulation.

The second part of the presentation should mention the results and difficulties of practical application of railway operations simulation using the Swiss simulation tool “OPENTRACK” (ETH Zurich) in Czech Republic. The described experience and conclusions were obtained mainly during the development of the Praha-Kladno railway operations simulation model. This model was developed in co-operation with the Czech Railways (České dráhy) – the operator of this track.

2 THE EVOLUTION OF RAILWAY OPERATIONS SIMULATION

The first one to experimentally use a software tool to plan the railway operations was the University of Hannover (Germany) in late 70’s. But these early tools were able just to support single steps of the planning process (e.g. timetable construction). In the 80’s were introduced first more complex simulation tools which were integrating more planning steps efficiently and the 90’s the performance of hardware and the development of complex simulation software tools enabled the functional usage of railway operations simulation to construct and prove the timetable as well as to specify and demonstrate the contribution of infrastructural or technological improvements especially in Germany and in Switzerland.

In Czech Republic the main railway operator – České dráhy – and later also the infrastructure manager – SŽDC – has been using a timetable planning system SENA JR VT (developed by České dráhy and the Universities of Pardubice and Žilina) since 1997. But this tool is not a simulation tool, it’s used just for quite simple timetable construction.

In a similar way our Faculty of Transportation Sciences (FD ČVUT, Prague) has used the software FBS (iRFP Dresden – Germany) since 1999. And, since 2003, FD ČVUT has been finally using also a really complex simulation tool – the OPENTRACK (ETH Zurich – Switzerland).

3 THE BASIC PRINCIPLE OF RAILWAY OPERATIONS SIMULATION AND THE MOST IMPORTANT METHODS

The basic principle of railway operations simulation can be described in following way:

The real railway operations - as a co-operation of man, infrastructure and vehicles - is presented by a model that enables to experimentally imitate those parts of the processes in the transportation system which are important for the performed research. The statistical processing of its outputs (the range and accuracy may slightly differ according to needs and the used tool) brings us respectable results [1].

Every simulation tool consists of following components:

- Entry data input:
    - Infrastructure data,
    - Vehicle data,
    - Transport demand data (proposed timetable, needed interval, needed interchange possibilities, needed departure or arrival settings).
- Specific simulation algorithm.
- Results output enabling efficient interpretation of the achieved data.

The key component is obviously the simulation algorithm. There are two basic algorithm arts – the synchronous and the asynchronous one.

3.1 The asynchronous simulation algorithms

The asynchronous simulation tools take the operating trains one by one as they appear in the proposed timetable and calculate the whole ride of the train in dependence on the infrastructure and vehicle's traction parameters. The calculated routes are compared afterwards and detected conflicts are being solved.

3.2 The synchronous simulation algorithms

The synchronous simulation tools calculate the movement of all trains "in the real time" and the appearing conflicts are being solved "ad hoc". It is usually possible to set a specific period in which the conflicts are being forecasted. The OPENTRACK is one of the synchronous systems.

3.3 Multiple simulation

The multiple simulation is a specific feature developed in the last years which is very efficient especially to prove the robustness of a timetable.

This component enables to enter some time elements (entering delay, stop duration, interchange duration) or incidents (traction performance, appearing of infrastructure obstructions) in form of a mean value with a probability distribution.

Later the simulation can be set to be performed many times when the mentioned time elements are sampled according to the probability distribution. When the number of simulation runs is high enough (tens till hundreds of runs according to the complexity of the model), the results can prove the robustness of a timetable very accurately and also determine the critical system points.

4 THE OPENTRACK SIMULATION TOOL

The OPENTRACK system is a synchronous simulation tool that has been developed since the end of the 90's by the Institute of transport planning ETH Zurich. Today it is used by many railway operators as well as railway infrastructure managers in Europe as well as overseas (e.g. SBB, DB, ÖBB).

It is also used by many consultants and research institutes concerning in railway operations and planning (ETH Zurich, TU Wien,…). In Czech Republic it is currently used for research purposes by the Czech Technical University (ČVUT v Praze) and the University of Pardubice.

OPENTRACK *More detailed features descriptions and references are available at www.opentrack.ch. [2, 3].*

5 THE SIMULATION MODEL OF THE RAILWAY OPERATION – TRACK PRAHA – KLADNO - CIRCUMSTANCES OF THE PROJECT

The timetable for the period 2008/9 should introduce a new operation concept on the track Praha-Kladno. As a part of the Praha suburban transport system, the track should offer a much more intensive traffic with morning and afternoon interval of 15-30 minutes.

This was quite a challenging task, as the track is almost in full length single tracked and the traffic is mostly operated locally by an elderly electromechanical signal box.

That is why the operator as well as the infrastructure manager proposed the timetable to be quite unstable and set up this project that should meet three main requirements:

- if the timetable is feasible,
- which way the timetable will react on irregularities,
- what measures can be taken to minimize the influence of irregularities.

The project was planned in four steps:

- to create the infrastructure model, enter the vehicle's traction data and the proposed timetable,
- to simulate the operation according to the proposed timetable at standard conditions to prove the feasibility,
- to simulate the influence of an entry delay up to 15 minutes for all trains during a work day,
- to propose some basic rules to minimize the influence of the entry delays.

6 SOME BASIC FACTS ABOUT THE RESULTING MODEL

The entire project was solved in about five months of work. The first two months were needed to create the infrastructural model and enter the vehicle's traction and timetable data. The most important work - the simulation - was performed in the third month of work. The results were interpreted in the last two months and, according to the results, the basic rules were formed and verified. As soon as the results were presented to České dráhy, the simulation and the resulting rules were being revised for the successive month.

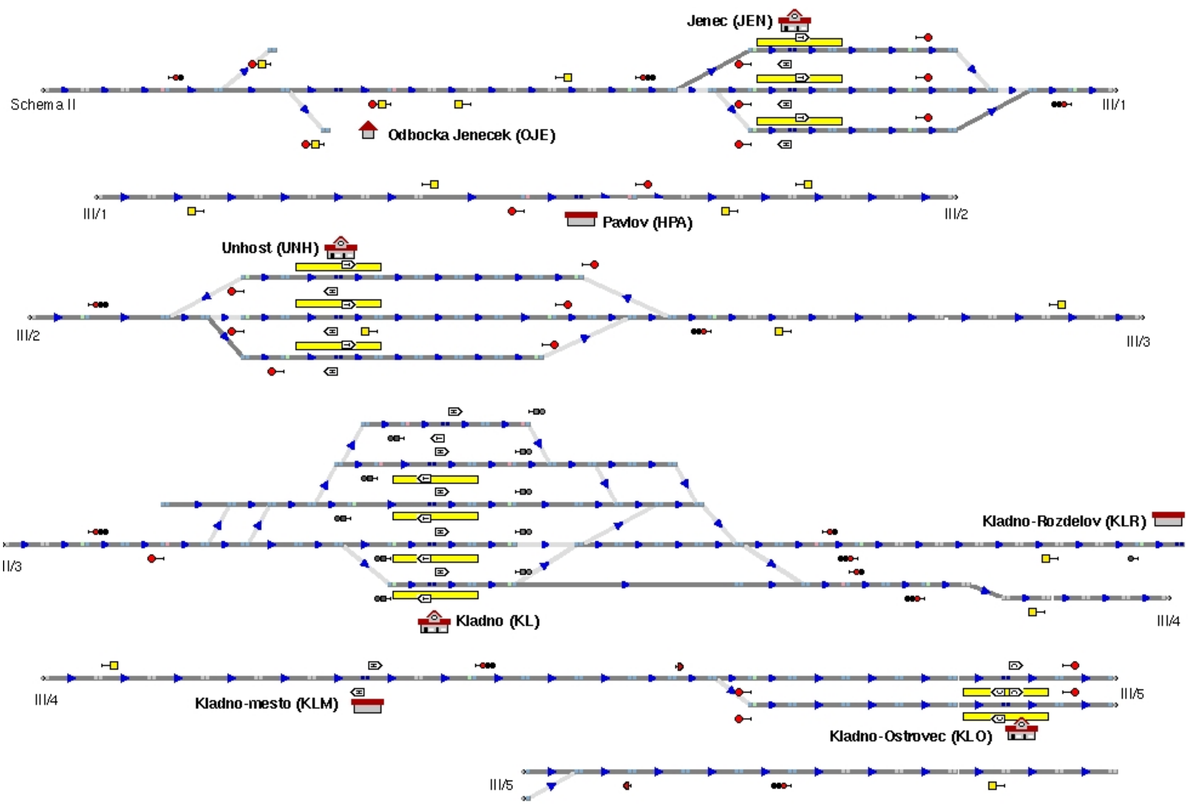


Figure 1 One third of the infrastructural model of the track Praha-Kladno

Table 1 Basic facts about the created simulation model

The timetable length of the simulated track	35, 65 km
The accurate length of the tracks in the model	111, 34 km
Number of signals used in the model	about 220
Number of trains simulated during a workday	80 trains (40 pairs)
The number of timepoints entered in the timetable	about 1100
Used vehicles	Type 810, 814.2, 814.3 and 854 ČD
Approximate length of the one entire simulation run (0:00-2:00 following day)	about 12 minutes when using the simulation accuracy step of 2 sec.

7 RESULTS OF THE PROJECT

There have been three basic results of the project:

- a prove of the timetable feasibility,
- the proposal of three basic rules that should support the robustness of the timetable,
- an individual report of the influence of the entry delay for every train (or periodical group of trains) during a workday with some proposed measures for specific trains.

7.1 A prove of the timetable feasibility

The first important result of the operations simulation was the conclusion, that the proposed timetable is feasible. Under standard conditions (calculated performance reserve of 4%) there were no delays longer than 20 seconds (at the 30 sec. accuracy of the timetable). Under bad conditions (calculated performance reserve of 10%) some trains reached a delay of up to 130 seconds but the timetable still remained robust in a way, that no trains have to be cancelled and the order of trains passing the stations did not have to change even slightly.

7.2 Proposed basic operation rules

Based on the simulation of bad conditions as well as the simulation of train entry delays three basic rules for the case of traffic irregularities were proposed:

- to minimize waiting for delayed trains (for the purpose of interchange) – on the simulated track absolutely and on the following part of track (Kladno – Rakovnik) to a maximum of 7 minutes,
- to become very flexible in changing the stations were the trains cross or overtake,
- to prefer trains having a delay up to 7 minutes not respecting their specific preference class instead of the typical preference of fast and other specific class trains.

7.3 Individual reports about the influence of entry delay for every train

For every train (or tact group of trains) the entry delays of 5, 10 and 15 minutes were simulated. According to the results of these simulations the critical delays were determined for every train. The critical values were the periods over which the delay came over to other trains, the train crossing points had to be changed or the traffic caused the delay to rise further on in the simulated part of track. Because the track does not have a central dispatcher, it was later needed to rewrite the train reports into sets of rules for the operator of every station. This was done by the experts of České dráhy.

8 CRITICAL POINTS OF THE PROJECT

During the project work and the interpretation of the results, two main groups of critical issues have been discovered. We maintain that there is a possibility of future improvements at these points, which could bring a slightly better accuracy of the results and minimize the needed effort.

The first group involves the differences in the law, regulations and habits between the Czech and Swiss railway operations. The second one involves the missing possibility to use the multiple simulation within the conditions of Czech Republic.

8.1 The differences between Swiss and Czech railway operations

The SW Opentrack is based on the Swiss railway operation law and rules (actually, the system offers also settings for German, Austrian and partially Dutch conditions. So as to keep some specific Czech rules, we had to find some improvised solutions (the most crucial one was to keep the appropriate period between two trains entering one station almost at the same time, which was finally solved by using the overlaps/skid distances, which are practically used very rarely on Czech railways).

These measures took quite much time to set up and especially to calibrate to reach the needed period and - what is even worse – they brought a slight risk, that in critical situations (e.g. when the train tries to shorten the existing delay) the reaction of the model will not be absolutely adequate.

Since we can't suppose the OPENTRACK or any other simulation tool to be soon adapted to the Czech legal and technological conditions, we will probably have to optimize our alternative solutions further on, and be aware of the possible inaccuracy caused by these processes.

8.2 The missing possibility of using the multiple simulation

In this case the trouble is not caused by the software. The OPENTRACK system includes all needed features for the multiple simulation and they are ready to use. The reasons why we are not able to use this feature consist in the lack of experience in the Czech Republic as well as abroad and the missing efficient tool for the statistical analysis of the results.

Probably the main obstacle is the lack of any system studies concerning a statistical analysis of irregularities in the Czech Railway operating system (which would depend on the sort of track and used signal box, track length, daytime, weekday…). Without this knowledge we can't find any adequate probability distribution of the time elements, which is one of the most crucial points to reach the aim of multiple simulation.

The usage of the multiple simulation would enable a very efficient analysis not only with regard to the influence of single train delays, but also the influence of sampled combinations of more irregularities at once, occurring with a realistic probability. This was impossible in the way we performed the simulation today, because of the enormous number of possible combinations and quite complicated preparation of every simulation run.

9 CONCLUSION

The project of THE SIMULATION MODEL OF THE RAILWAY OPERATION – TRACK PRAHA – KLADNO became after five years of theoretical research the first practical application of the simulation modelling method of railway operations in the Czech Republic basing on the operator's demand. This experience revealed the above mentioned obstacles, which complicate the use of this method in the Czech conditions. These obstacles needed an extra effort from the solving team as well as understanding and co-operation of the operator.

In spite of these obstacles we think the model gave us, and especially the operator, important conclusions, which would otherwise have to be determined in practical operations with a slight risk. This enabled the operator to take measures to minimize the influence of delays in advance. These measures supported the marketing image of the new timetable concept, especially in the first days and weeks after its long before promoted introduction, when the passengers are abnormally sensitive to delays and other irregularities.

10 ABBREVIATIONS

ETH Zürich – Eidgenössische Technische Hochschule Zürich.

SENA JŘ VT – Software used by the Czech Railways and later SŽDC for railway timetable construction (orig, Sestava Nákresného Jízdního Řádu Výpočetní Technikou).

ČVUT v Praze, FD – Czech Technical University, Faculty of Transportation Sciences.

iRFP Dresden – Institut für Regional- und Fernverkehrsplanung Dresden.

SŽDC – Czech Railway Infrastructure Manager (orig. Správa železniční a dopravní cesty s.o.).

The project of THE SIMULATION MODEL OF THE RAILWAY OPERATION – TRACK PRAHA – KLADNO was founded by České dráhy – the operator of the track Praha-Kladno, as well as the most tracks in the Czech Republic. The operator also supported the project with the needed entry data. We would like to thank especially to Ing. Jan Hrabáček, PhD. from České dráhy, who has a lot of experience with the OPENTRACK SW and provided us with many important suggestions and feedbacks.

References

[1] BÄR, M.: *Betriebsführung des Bahn- und ÖPN-Verkehrs*, Kapitola V. Leistungsermittlung undBemessung von Betriebsanlagen, interní studijní materiály TU Dresden. Dresden (SRN): Professur für Bahnverkehr, öffentlicher Stadt- und Regionalverkehr, Institut für Verkehrssystemtechnik, Fakultät Verkehr-swissenschaften "Friedrich List", Technische Universität Dresden, Germany : 2004, p. V-2, pp. 14-18.

[2] POLÁK, O., JANOŠ, V., BAUDYŠ, K.: *Simulační model železničního provozu trati Praha Masarykovo nádraží – Kladno Ostrovec*, project based on demand of České dráhy a.s., Fakulta dopravní ČVUT, Praha, Czech Republic, 2008.

[3] HÜRLIMANN, D.: *Opentrack – Simulation of Railway Networks [on-line]*. [ETH Zürich],[2009],[2009-06-10].

[4] URL: < http://www.opentrack.ch/opentrack /opentrack_e/opentrack_e.html >

[5] PASTOR, O., TUZAR, A.: *Teorie dopravních systémů*. ASPI, Praha, Czech Republic 2007.

Dipl. Eng. Ondřej POLÁK
Czech Technical University in Prague
Zíkova 4
166 36 Prague
Czech Republic
E-mail: xpolako@fd.cvut.cz

Dip. Eng. Vít JANOŠ, Ph.D.
Czech Technical University in Prague
Zíkova 4
166 36 Prague
Czech Republic
E-mail: janos@fd.cvut.cz

Dipl. Eng. Karel BAUDYŠ, Ph.D.
Czech Technical University in Prague
Zíkova 4
166 36 Prague
Czech Republic
E-mail: baudys@fd.cvut.cz

TEST ANALYSES OF DAM DEFORMATIONS FOR SECURITY OF PEOPLE AND ENVIRONMENTAL PROTECTION

Vladimír SEDLÁK, Miloš JEČNÝ, Marián MESÁROŠ, Miroslav KELEMEN, Pavel NEČAS

Abstract: Deformations on buildings and structures due to own weight, water pressure, inner temperature, contraction, atmospheric temperature and earth consolidation occur. Especially, it is necessary to embark on monitoring and analysing of deformation effects and movements of any sizeable dams and water basins and so to prevent of their prospective catastrophic effects into the environment and also human lives. The paper is centred on stability of the bulk (roc-fill) dam of the water basin Pod Bukovcom near Košice in the East Slovak Region. Results and analyses of the geodetic terrestrial and GPS measurements on the rock-fill dam are undergone by to test-statistics, the model of stability or prospective movement of the rock-fill dam with time prediction. The paper outputs are incorporated into GIS and information system of U.S. Steel Košice.

Key words: Deformation, dam, security, GPS.

1 INTRODUCTION

Deformations and movements of buildings and construction by effect of own weight, water pressure, inside temperature, retraction, atmospheric temperature and earth consolidation, are occurred. These deformations and movements are necessary to investigate according to the philosophy that "all is in the continual movements". Especially, it is necessary to go into monitoring and analysing deformations and movements of some sizeable building works of the human. The dams belong to the major building works, where the monitoring of these deformations and movements must be done. Some dams have also strategic signification and many of them must be protected by reason of military strategy (Kelemen et al. 2009, Nečas 2005, 2007).

The bulk dam Pod Bukovcom is built on the river Ida between the villages Bukovec and Malá Ida in the East Slovakia (Figure 1). The bulk fagot dam is situated in the morphologically most advantageous profile, in the place of the old approximately 7 m high and approximately 220 m length fagot dam, which was liquidated following the building-up of the up-to-date bulk fagot dam. The industrial water supply for cooling the metallurgical furnace equipments in the company US Steel Košice in a case of damages is the purpose of the dam. The water basin is also for flattening the flow waters and for recreational purposes during the summer time.

2 THE NETWORK OF THE BULK DAM POD BUKOVCOM

Six reference points stabilized outside of the dam bulk fagot dam. The reference points are situated about 50-100 m from the dam (Technické podklady…, 1965-98). The reference points have the labelling from A1 up to F1. These points supplied the old reference points from who's the measurement are performed since 1985. The stabilization of these reference points is realised by the breasting pillars with a thread for the exact forced centring of the surveying equipment (total stations and GPS).

The object points on the bulk fagot dam are set so as they represented the fagot dam geometry and the assumed pressures of the water level on the fagot dam at the best. The points are set in six profiles on the fagot dam. So as the object points transmit of the fagot dam deformations, they had to be approximately stabilized deep 1.8 m. Generally 26 object points are set on the fagot dam (Figure 2). Two of them are destroyed.

Figure 1 The bulk dam Pod Bukovcom

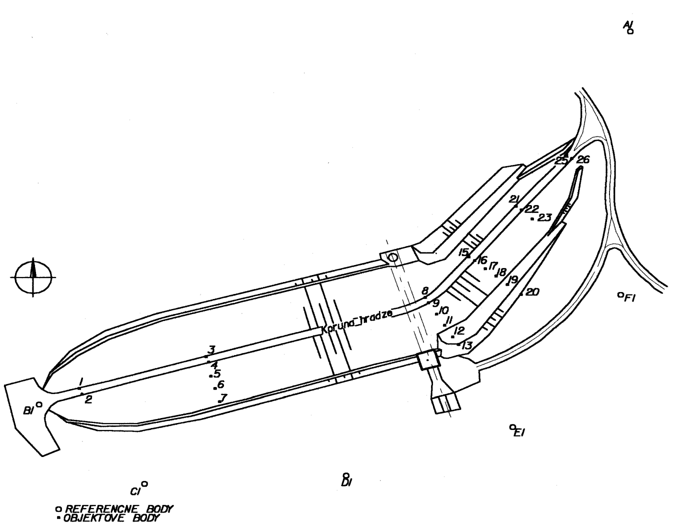


Figure 2 The network point field of the bulk dam
○ reference points • object points

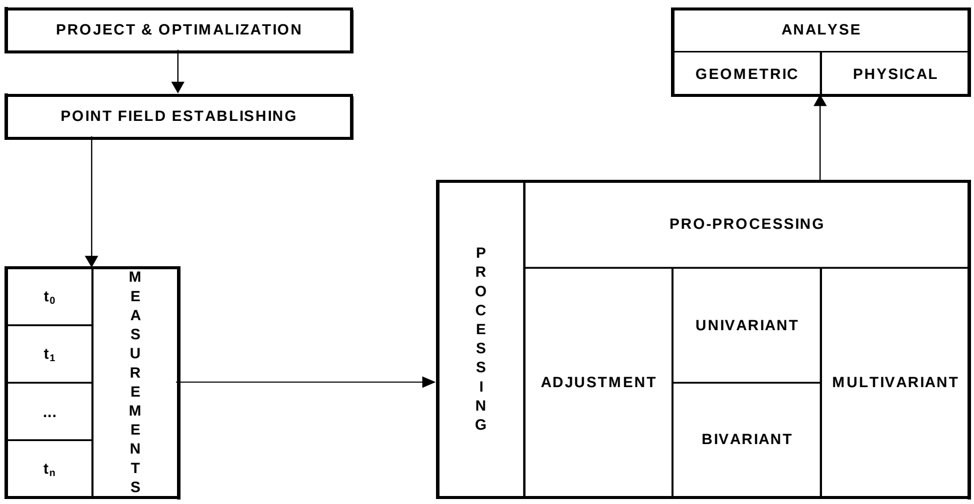


Figure 3 Scheme of deformiy detection algorithm

3 THE DEFORMITY DETECTION ALGORITHM

Deformity detections are performed according to the concrete procedure technique. This procedure is called the algorithm (Figure 3). From the scheme in Figure1 results, that full procedure since the project trough the measurement ends by the obtained adjustment results analyse. The processed results are analysed from the aspect of geometrical or physical properties of the examined object.

3.1 The deformity detection alalyse

Analyse of the deformation network processed data can be done by the analytic or the analytic-graphic ways. It depends on the used middles for the network congruence. The used methods are varied asunder by the result shape of the results presentation. However, from the point view of the deduction analyse the results presentation are equivalent. From the point of view of the congruence testing analyse is divided into the statistical and deterministic analyses.

The congruence method of the geodetic networks follows out from the base of examination and analyse of the positional co-ordinates from the individual epochs. From the point of view of the tested values the deformity detection analyse methods are divided into the parametric and nonparametric methods.

The parametric testing methods make use of the co-ordinate differences of the tested points, while the nonparametric methods test the invariant differences of the network elements. Values for the network structures testing are obtained by means of the estimative model *LSM* (the last square method) or by means of the robust statistic models.

The statistic testing practices are the most frequently used for a purpose of the deformation networks congruence testing? Arbitration whether the network co-ordinate or invariance differences are statistically meaningful or not meaningful is the task of the testing. For this purpose it is necessary to form the null-hypothesis, which has the shape (Ječný 2000, Sedlák 1996, Sedlák and Ječný 2004).

$$H_0 : E(\hat{\boldsymbol{C}}^1) = E(\hat{\boldsymbol{C}}^2) \tag{1}$$

or in the shape respectively

$$H_0 : E(\boldsymbol{L}^1) = E(\boldsymbol{L}^2) \tag{2}$$

where $\hat{\boldsymbol{C}}^i$ is the vector of the adjusted co-ordinates of the object points in the epoch i, $\boldsymbol{L}^i$ is the vector of the measured values in the epoch i.

It means that the middle values of the vector of the adjusted co-ordinates or measurements from the first epoch are equalled to the middle value of the vector of the adjusted co-ordinates or measurements from the second epoch.

For the co-ordinate differences $\delta\hat{\boldsymbol{C}}^i$ is valid the equation

$$H_0 : E(\delta\hat{\boldsymbol{C}}^1) = E(\delta\hat{\boldsymbol{C}}^2)\,. \tag{3}$$

The often register for the adjusted co-ordinates of the object points is in the adnichiled form

$$\hat{\boldsymbol{C}}^1 - \hat{\boldsymbol{C}}^2 = 0\,. \tag{4}$$

For the null-hypothesis H_0 the equation is also used in the shape

$$H_0 : H.\boldsymbol{\Theta} = \boldsymbol{h} \tag{5}$$

where $\boldsymbol{h}$ is the null-vector, $\boldsymbol{\Theta}$ is the matrix of the estimate parameters.

The test statistics T is compared with the null-hypothesis. The universal test statistics is the most frequently composed on the tested value and middle error s ratio.

$$T = \frac{\left|\delta\hat{\boldsymbol{C}}\right|}{s.\delta\hat{\boldsymbol{C}}}\,. \tag{6}$$

The null-hypothesis $H_0{:}H.\boldsymbol{\Theta} = 0$ is composed for the co-ordinate differences vector. According to it the test statistics T will be in the shape

$$T = \frac{\dfrac{\delta\boldsymbol{C}^T.\boldsymbol{Q}_{\delta\hat{C}}^{\,1}.\delta\boldsymbol{C}}{k}}{\dfrac{\boldsymbol{v}^T.\boldsymbol{Q}_L^{\,1}.\boldsymbol{v}}{f}} \tag{7}$$

where $\boldsymbol{Q}$ is the deformation vector matrix, $\boldsymbol{v}$ is the vector of the corrections.

The quadratic form of the co-ordinate divergences is in the numerator and the empirical variation factor s_0 is in the denominator. The test statistics shape after arrangement is

$$T = \frac{\delta\boldsymbol{C}^T.\boldsymbol{Q}_{\delta\hat{C}}^{\,1}.\delta\boldsymbol{C}}{k.s_0^2} \approx F(1-\alpha, f_1, f_2) \tag{8}$$

where $1-\alpha$ is the reliability coefficient, α is the confidence level (95% or 99%), f_1, f_2 are the stages of freedom of F distribution (Fischer's distribution) of the accidental variable T, k is the co-ordinates number accessioning into the network adjustment.

The stages of freedom are selected according to the adjustment type. For the free adjustment, they are the equations are valid

$$f_1 = n - k + d\,, \qquad f_2 = k - d \tag{9}$$

and for the bonding adjustment

$$f_1 = n - k\,,\ \ f_2 = k \tag{10}$$

where n is number of the measured values entering into the network adjustment, d is the network defect at the network free adjustment.

The test statistics T should be subjugated to a comparison with the critical test statistics T_{CRIT}. T_{CRIT} is found in the tables of F distribution according the network stages of freedom.

Two occurrences can be appeared:

- $T \le T_{CRIT}$: The null-hypothesis H_0 is accepted. It means that the differences vector co-ordinate values are not significant.
- $T \ge T_{CRIT}$: The null-hypothesis H_0 is refused. It means that the differences vector co-ordinate values are statistically significant. In this case we can say that the deformation with the confidence level α is occurred.

3.2 Analytic process of testing

Definition of the null-hypothesis H_0 is the first step according to the equation

$$H_0 = E(s_0^{2\,I}) = E(s_0^{2\,2}) = \sigma_0^2\,, \tag{11}$$

where σ_0 is the selected variation.

F distribution is used at the testing. *F* distribution has the stages of freedom f_1 and f_2. Full testing is in progress in three phases. The first phase, it is the comparison testing, which tests whether the measurements in the epochs were equivalent. The second phase, it is the realisation of the global test, which will show whether the statistically meaningful data are occurred in the processed vector. The third phase, it is the identification test. This test is realised only in a case when the null-hypothesis is not confirmed at the global test. The identification test will check the statistic significance of each point individually.

To check the reference points at first is suitable at the testing. If some of the reference points do not pass over the test, it will mean that the point is moved with the certainty α. Such point will be changed up among the object points or it will be eliminated from the next processing.

If we have a safety that the reference points are fixed then the object points are only submitted to the testing. The comparison test operates with the test statistics *T* according to the equation

$$T = \frac{s_0^{2\,I}}{s_0^{2\,II}} \approx F(f_1, f_2) \tag{12}$$

where *I,II* are the measurement epochs

The critical value T_{KRIT} is searched in the *F* distribution tables according to the degrees of freedom $f_1=f_2=n-k$ or $f_1=f_2=n-k+d$.

The test statistics *T* is compared with the critic value T_{CRIT} and the null-hypothesis H_0 is considered:

- $T \leq TCRIT$: the null-hypothesis H_0 is accepted and it means that measurements in the epochs are equivalent themselves.
- $T \geq TCRIT$: the null-hypothesis H_0 is refused and it means that measurements in the epochs are not equivalent themselves.

The global test operates with the test statistics T_G according to the equation

$$T_G = \frac{\delta\hat{\boldsymbol{C}}^T . \boldsymbol{Q}_{\delta\hat{C}}^{\,1} . \delta\hat{\boldsymbol{C}}^T}{k . s_0^2} \approx F(f_1, f_2) \tag{13}$$

where

$$s_0^2 = \frac{(v^T . \boldsymbol{Q}_L^{\,1} v)^I + (v^T . \boldsymbol{Q}_L^{\,1} . v)^2}{f_1 + f_2}\,. \tag{14}$$

The critic value T_{KRIT} is found in *F* distribution tables according to the degrees of freedom $f_1=k$, $f_2=n-k$ or $f_1= k+d, f_2=n-k+d$.

The test statistics *T* is compared with the critic values T_{CRIT} and the null-hypothesis is considered:

- $T \leq T_{CRIT}$: The null-hypothesis H_0 is accepted and it means that the co-ordinate differences vector values are petit.
- $T \geq T_{CRIT}$: The null-hypothesis H_0 is refused and it means that the co-ordinate differences vector values are meaningful. In this case the third phase must be operated at which to be found which points allocate any displacement.

The identity test operates with the test statistics T_i according to the following equation

$$T_i = \frac{\delta\hat{\boldsymbol{C}}_i^T . \boldsymbol{Q}_{\delta\hat{C}}^{\,1} . \delta\hat{\boldsymbol{C}}_i}{s_0^2} \approx F(f_1, f_2)\,. \tag{15}$$

The critic value T_{CRIT} is chosen in the *F* distribution tables according to the degrees of freedom $f_1=n$ a $f_2=n-k$ or $f_1=1$ a $f_2=n-k+d$.

The test statistics *T* is compared with the critic value T_{CRIT} and the null-hypothesis H_0 is taken into consideration:

- $T \leq T_{CRIT}$: The null-hypothesis H_0 is accepted and it means that the adjusted co-ordinate difference values of the tested point is statistical petit.
- $T \geq T_{CRIT}$: The null-hypothesis H_0 is refused and it means that the adjusted co-ordinate difference values of the tested point is statistical meaningful. This point is moved with an expectation α.

After detection of the point displacement this point is excluded from the following testing and whole file is submitted to testing once more.

3.3 Determining the co-factor matrix of the deformation vector

So as the testing the co-ordinate differences could be operated, it is needed to determine the co-factor matrix of the co-ordinate differences $\boldsymbol{Q}_{\delta\hat{C}}$. Its scale will determine by the following equation

$$\boldsymbol{Q}_{\delta\hat{C}} = \boldsymbol{Q}_{\hat{C}}^{I} + \boldsymbol{Q}_{\hat{C}}^{II} - (\boldsymbol{Q}_{\hat{C}}^{I,II} + \boldsymbol{Q}_{\hat{C}}^{II,I})\,. \tag{16}$$

This equation is valid at the network simultaneous adjustment. At the deformation network separate adjustment the following equation is valid

$$\boldsymbol{Q}_{\delta\hat{C}} = \boldsymbol{Q}^{I}_{\hat{C}} + \boldsymbol{Q}^{II}_{\hat{C}} \,. \tag{17}$$

From this follows that it is necessary to choose a respectable structure and a follow-up procedures in the deformation network processing.

3.4 Analytic and graphic way of testing

The graphic shape of point displacement is a result and we can used the following equation

$$\delta\hat{\boldsymbol{C}}^{T} .\boldsymbol{Q}^{\,1}_{\delta\hat{C}} .\delta\hat{\boldsymbol{C}} = T.k.s_0^2 \,. \tag{18}$$

This equation presents the ellipse equation. The ellipse half-axle values and the ellipse swing out angle values round a co-ordinate system are necessary to know for a purpose of the ellipse depict. The following equation can be used for the ellipse half-axle values $a_{i\alpha}, b_{i\alpha}$

$$a^2_{i\alpha} = ((\boldsymbol{Q}_{\delta\hat{x}i} + \boldsymbol{Q}_{\delta\hat{y}i}) + \sqrt{(2\boldsymbol{Q}_{\delta\hat{x}i} - \boldsymbol{Q}_{\delta\hat{y}i})^2 + 4.(\boldsymbol{Q}_{\delta\hat{x}i\delta\hat{y}i}{}^2)}).F(1-\alpha,2,n-k).s_0^2 \,, \tag{19}$$

$$b^2_{i\alpha} = ((\boldsymbol{Q}_{\delta\hat{x}i} + \boldsymbol{Q}_{\delta\hat{y}i}) - \sqrt{(\boldsymbol{Q}_{\delta\hat{x}i} - \boldsymbol{Q}_{\delta\hat{y}i})^2 + 4.(\boldsymbol{Q}_{\delta\hat{x}i\delta\hat{y}i}{}^2)}).F(1-\alpha,2,n-k).s_0^2 \,, \tag{20}$$

where $a_{i\alpha}$ is the ellipse main half-axle in mm,
$b_{i\alpha}$ is the ellipse adjacent half-axle in mm.

The swing out angle of φ is determined according to the equation

$$tg2\varphi_a = \frac{2.\boldsymbol{Q}_{\delta\hat{x}i\delta\hat{y}i}}{\boldsymbol{Q}_{\delta\hat{x}i} - \boldsymbol{Q}_{\delta\hat{y}i}} \,. \tag{21}$$

These ellipses are named the confidence (relative) ellipses. It is possible to form them only in a case if the deformation network simultaneous processing procedure is appointed. The confidence ellipse is depicted according to the design elements with a centre in the point from the second epoch. The positional vector between the point position from the second and the first epoch is also depicted. The null-hypothesis is definable by the confidence ellipse, which covers whole positional vector in a full scale. The ellipse does not characterise a displacement of the considered point if it covers the positional vector in a full scale. The null-hypothesis is accepted. The ellipse characterises the displacement of the considered point if it does not cover the positional vector in a full scale. The null-hypothesis is refused.

3.5 Results of the analytic and graphic analyse

GPS measurements and data processing were realized in the epochs: spring 1999, 2000, 2001, 2002 and 2003 (Sedlák et al. 2008, 2009). Twelve months were the time period between the epochs. The positional survey of deformation of the dam Pod Bukovcom was carried out. A free unit adjustment of the deformation network of the object points was realized. The network was processed by means of using LSM. Gauss-Markov mathematic model was applied into the processing procedure. In respect thereof the significance levels and the degrees of freedom were determined. The selected network was an adequate redundancy (measurements redundancy).

The position (2D) accuracy of the points of the network Pod Bukovcom was appreciated by the global and the local indices.

Global indices were used for an accuracy consideration of whole network, and they are numerically expressed. The network, which indicates have the last number, means that its observed elements were the most exactly observed, and the equal adjustment has also a high accuracy degree.

The following global indices were considered:

- *the variance global indices*: $tr\,(\boldsymbol{\Sigma}_{\hat{C}})$, i.e. a track of the covariance matrix $\boldsymbol{\Sigma}_{\hat{C}}$,
- *the volume global indices*: $det\,(\boldsymbol{\Sigma}_{\hat{C}})$, i.e. a determinant.

Local indices were as the matter of fact the point indices, which characterize the reliability of the network points.

The local indices were in the following expressions:

- *the middle 2D error:* $\sigma_p = \sqrt{\sigma^2_{\hat{X}_i} + \sigma^2_{\hat{Y}_i}}$,
- *the middle co-ordinate error:* $\sigma_{XY} = \sqrt{\dfrac{\sigma^2_{\hat{X}_i} + \sigma^2_{\hat{Y}_i}}{2}}$

- *the confidence absolute ellipses* which were served for a consideration of the real position in the point accuracy. We need know the ellipsis constructional elements, i.e. the semi-major axis a, the semi-minor axis b and the bearing φ_a of the semi-major axis. We had to also determine the signification α (*Table 1, Figure 4*).

Table 1 The analytic-graphic testing results –the confidence ellipses elements (2003)

Point	***a*** [mm]	***b*** [mm]	φ_a [^{g}]
1	10.7	4.6	289.1788
2	11.9	4.5	271.1377
3	7.2	4.3	316.2047
4	7.7	4.4	304.6711
5	18.4	5.1	194.1328
6	11.8	4.5	238.1088
7	12.1	4.5	238.6154
8	6.1	5.4	246.6188
9	5.9	5.5	249.0293
10	5.8	5.5	253.5228
11	5.9	5.3	257.2947
12	6.1	5.2	260.2185
13	6.2	5.1	261.8504
15	6.3	5.4	239.4482
16	6.3	5.3	239.4007
17	6.3	5.3	239.5853
18	6.3	5.3	240.3191
19	6.4	5.2	241.8465
20	6.7	5.1	244.1852
21	8.0	4.7	215.0106
22	8.2	4.7	213.8748
23	8.4	4.6	212.3949
25	13.1	4.3	199.6478
26	19.2	4.2	197.5029

The analytic analyse was implemented for a comparison after the results processing. According to this analyse the global test value T_G responded to 1.5498 and the value T_{CRIT} responded to 1.8284. From this follows that neither objects point did not note down statistically meaningful displacement during a period between the measurement epochs.

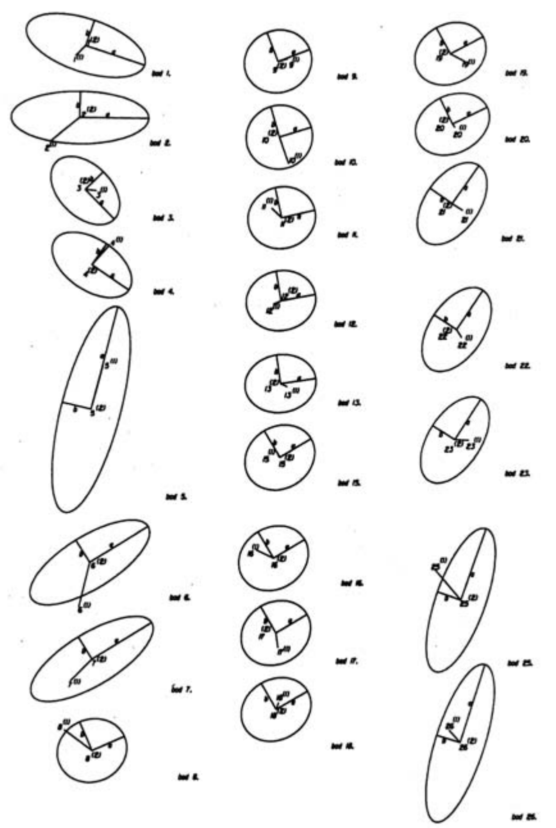


Figure 4 The confidence ellipses; the deformation vectors: 1999-2003

The analytic analyse was implemented for a comparison after the results processing. According to this analyse the global test value T_G responded to 1.5498 and the value T_{CRIT} responded to 1.8284. From this follows that neither objects point did not note down statistically meaningful displacement during a period between the measurement epochs.

4 CONCLUSIONS

The independent results from the analytic and analytic-graphic analyses confirmed an assumption that the object points and thereby also the dam object did not note down any statistically meaningful displacement with the definiteness on 95 %. The confidence ellipses of the points No: 6, 8 and 25 do not verify the null-hypothesis because the deformation vector does not exceed of an ellipse. Shrillness of the positional vector is indeed insignificant from which a conclusion was deducted that the displacement at these points was not occurred.

The observation of the bulk dam of the water work Pod Bukovcom is performed since its construction finishing as yet. The observations are periodical. A time period between epochs is gradually elongated since a half of year till two years time after a fixed course of the dam object movements. The results just confirmed this fixed trend. From geodetic analyses processed after each observation the obtained knowledge are applied at a designing and observation of similar water works deformations. Thereby an assurance is increased for population living nearby of the dam and also thereby economic and ecological damages caused by any emergency on the water work can be forestalled.

The paper followed out from the research project KEGA No. 3/6203/08 researched at the University of Security Management in Košice, Slovakia.

References

[1] JEČNÝ, M., 2001: Analyticko-grafické testovanie deformácií na vodnom diele Pod Bukovcom. (In Slovak). (Analytic and graphic testing of deformations on the bulk dam Pod Bukovcom). *Acta Montanistica Slovaca*, Vol.6, No.3/2001, Košice : 2001, 182-189.

[2] JEČNÝ, M., SEDLÁK,V., 2004: Deformation Measurements on the Bulk Dam in East Slovakia. *Transections of the VŠB – Technical University of Ostrava*, Mining and Geological Series, Vol.L, No.2/2004, 1-10.

[3] KELEMEN, M, NEČAS, P., HALÁDIK, J., SEDLÁK,V., 2009: Projekt edukácie odborníkov V-4 v oblasti bezpečnosti, obrany a bezpečnostních technológií. (In Slovak). (Project of education of Experts in countries V-4 in a field of security, defence and security technologies). In: *Zborník z Mezinárodní bezpečnostní konference: Bezpečnostní technologie, Systémy a Masnagement 2009*. Zlín, 9.-11.9.2009, 16/44 (CD).

[4] NEČAS, P., 2005: Global Security Environment: Future Challenges and Risks. *Institut bezpečnostných a obranných štúdií*, Bratislava : 2005.

[5] NEČAS, P., 2007: Wanted More Security: Towards a Better World. *Technical University of Košice,* Košice : 2007.

[6] SEDLÁK,V., 1997: Matematické modelovanie lomových bodov v poklesových kotlinách. (In Slovak). (Mathematical modelling breakpoints in subsidence). *Acta Montanistica Slovaca*, Vol. 1, No. 4/1996, 317-328.

[7] SEDLÁK,V., LOŠONCZI, P., KISS, I., 2008: Bezpečnostné informačné technológie. (In Slovak). (Security information Technologies). *University of Security Management in Košice* (ed.), Košice, 2008.

[8] SEDLÁK,V., LOŠONCZI, P., PODLESNÁ, I., 2009: Družicové navigačné systémy. *University of Security Management in Košice* (ed.), Košice : 2009.

[9] STN 73 0405: Meranie posunov stavebných objektov. (In Slovak). (Measurement of the building objects displacements). Slovak Technical Norms.

[10] Technické podklady a výsledky pozorovania deformácií vodného diela pod Bukovcom. (In Slovak). (Technical data and results of the deformations survey on the bulk dam Pod Bukovcom). Tech. reports, VSŽ, a.s. Košice (U.S. Steel Košice), 1965-98.

Prof. Dipl. Eng. Vladimír SEDLÁK, PhD.
University of Security Management in Košice
Kukučínova 17
040 01 Košice
Slovak Republic
E-mail: vladimir.sedlak@vsbm.sk

Dr.h.c. Prof. Dipl. Eng. Marián MESÁROŠ, PhD.
Univesity of Security Management in Košice
Kukučínova 17
040 01 Košice
Slovak Republic
E-mail: marian.mesaros@vsbm.sk

Dipl. Eng. Miloš JEČNÝ, PhD.
GET s.r.o, Geology, Ecology and Mining Service
Korunovační 630/29
170 00 Prague
Czech Republic
E-mail: jecny@tiscali.cz, get@get.cz

Brig. Gen. Assoc. Prof. Dipl. Eng. Miroslav KELEMEN, PhD.
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 01 Liptovský Mikuláš
Slovak Republic
E-mail: miroslav.kelemen@aos.sk

Col. (GS), Assoc.Prof. Dipl. Eng. Pavel NEČAS, PhD.
Armed Forces Academy of General M. R. Štefánik
Liptovský Mikuláš
Slovak Republic
E-mail: pavel.necas@aos.sk

ALLIANCE'S STRATEGIC CONCEPT

Peter SPILÝ, Pavel NEČAS

Abstract: Alliance's Strategic Concept (hereinafter „SC") represents the most important strategic frame, which determines the main direction and areas of its activity. SC has always responded to an actual global security environment situation and to changes, which have had direct influence over Alliance's security. Due to this document importance, validity of each SC was longer, thus only six SC were adopted throughout Alliance history. Nowadays we are witnesses to intensive discussions of the new SC creation necessity as reflected in the conclusions from the recent NATO Summit in Strasbourg/Kehl when the heads of the member states tasked the Secretary General to establish and lead a group of qualified experts, whose task is to develop a new SC by the next NATO summit.

Keywords: Alliance's strategic concept, global security environment, challenges and risks, massive retaliation, flexible response, deterrence, partnership.

1 INTRODUCTION

The North Atlantic Treaty (hereinafter "the Treaty") signatories came to an agreement on safeguarding the freedom, common heritage and civilisation of their people, founded on the principles of democracy, individual liberty and the rule of law. They seek to promote stability and well-being in the North Atlantic area [1]. Further, the North Atlantic Treaty Organization (hereinafter „NATO") member states resolved to unite their efforts for collective defence and for the preservation of peace and security.

In the context of the Treaty, SC could be characterized as its real application in the particular historic situation or as "the Alliance's operational and dynamic view of its founding treaty" "[2]. SC regulates Alliance activity in the political and military sphere, suggests a way of its reaction to threats and challenges which influence Alliance's security. "In a simplified way SC is some equivalent to what in individual states represents security and defence strategy" [3].

The evidence of SCs strategy significance is the fact that only six SCs were adopted throughout Alliance sixty-year history. A new SC arrangement was determined by necessity to react to changes. SCs were created irregularly and their time validities were different.

2 SC FROM THE BEGINNIG OF THE ALLIANCE TO THE END OF THE COLD WAR

The Treaty laid the fundamentals of the first SC creation. In Article 3 the Parties bound to maintain and develop their individual and collective capacity to resist armed attack. In Article 5 they agreed that armed attack against one or more of them will be considered an attack against them all and each contractual country offers adequate help to attacked party including the use of armed forces. The key element which enforces the Treaty clauses is the North Atlantic Council (hereinafter "NAC"). The establishment of NAC is anchored in Article 9 and in this article is declared that "the Council shall set up such subsidiary bodies as may be necessary; in particular it shall establish immediately a defence committee that shall recommend measures for the implementation of Articles 3 and 5" [1]. The Treaty put in force after ratification process completion in August 1949. NAC acted without delay and besides the Defence Committee set up the Military Committee and the Standing Group. By that time the Alliance did not have any integrated military command structure. Instead of it, planning was done by five Regional Planning Groups. The main objective of all executive bodies was a collective defence planning and SC represents a core of this job. The first SC was issued in December 1949.

2.1 DC 6/1 the Strategic Concept for the Defence of the North Atlantic Area December 1, 1949

It is interesting that the first SC, coded DC 6/1, was issued only three months after the Treaty ratification. Its root was a document SG 1 drafted by the Standing Group. SG 1 was submitted to the Military Committee and subsequently to the Defence Committee for approval (documents are coded as per execution body that issued or approved them) [4].

DC 6/1 includes a chain of implemented military measures. Two of them are notable:

a) Insure the ability to carry out strategic bombing promptly by all means possible with all types of weapons, without exception. This is primarily the USA responsibility assisted as practicable by other nations.
b) Arrest and counter as soon as practicable the enemy offensives against the Treaty powers by all means available, including air, naval, land and psychological operations. Initially, the hard core of ground forces will come from the European nations. Other nations will give aid with the least possible delay and in accordance with over-all plans [5].

In presented clauses, decision to use nuclear weapons ("all types of weapons, without exception") is expressed for the first time in history, although only in a hidden form. While strategic bombing is primary directed by the USA, a ground forces core creation is ceded to Alliance's European members.

2.2 MC 3/5 (Final) the Strategic Concept for the Defence of the North Atlantic Area December 3, 1952

One of the most significant events since the first SC adoption was the Korean War (from 1950 to 1953). The War imposed fundamental structural changes to the Alliance. An integrated military command structure with two strategic commands was constituted. The Allied Command Europe (ACE) with its headquarters at SHAPE in Paris was created. General Dwight D. Eisenhower was appointed as the first Supreme Allied Commander Europe (SACEUR) [6]. The second strategic command in Norfolk in the USA named Allied Command Atlantic was led by Supreme Allied Commander Atlantic (SACLANT).

The first NATO's enlargement was implemented in 1952 and Greece and Turkey became the Alliance's new members.

Structural changes and new members were the main causes for previous SC reassessment. In essence the MC 3/5 military measures copied the first SC. Articles of strategic bombing capability with using any weapons was a basis of this defence concept [7].

2.3 MC 14/2 (Rev) (Final Decision) Overall Strategic Concept for the Defence of the North Atlantic Area May 23, 1957

In 1953 SACEUR established the "New Approach Group" tasked with new way of nuclear weapons integration into NATO strategy. The Group's job resulted to a document „The Most Effective Pattern on NATO Military Strength for the Next Few Years". It was the first Alliance's document explicitly discussing the use of nuclear weapons [4].

In mid-fifties, talks concerning defence expenses saving graduated. Arguments were ranging from almost total nuclear weapons reliance to more flexible using of all military forces. NAC finally made a decision in a document "Directive to the NATO Military Authorities from the North Atlantic Council". In the Directive there is a requirement "in accordance with the concept of forward strategy, counting on the use of nuclear weapons at the outset, and sustaining operations, without any intention to make a major withdrawal, until the strategic counter-offensive has achieved its objective" [8]. Thus NAC inclined to more adaptive employment of deployable forces "Shield" and strategic bombing forces "Sword".

Under authority of the Directive, the third SC named "Massive Retaliation" was developed. The most distinctive feature of it was massive use of nuclear weapons for the North Atlantic area defence. Its main objective puts a stress on a deterrent to aggression. "The principal elements of the deterrent are adequate nuclear and other ready forces and they manifest determination to retaliate against any aggressor with all the forces at our disposal, including nuclear weapons, which the defence of NATO would require" [9].

Military measures were issued in a separate document „Measures to Implement the Strategic Concept". They confirmed principles set by the Directive.

2.4 MC 14/3 Overall Strategic Concept for the Defence of the North Atlantic Area January 16, 1968

After adoption of "Massive Retaliation" SC, NATO's European members started to indicate their worry about the USA "willingness" to risk a nuclear war on its own territory. These fears sprang from the USA nuclear hegemony deprivation. The Soviet Union developed its own nuclear weapons and particularly the means of delivery of these weapons. Since a launch of the first artificial earth satellite there was no doubt about the Soviet Union capability to hit any point on the world by nuclear warheads.

Another essential problem was posed by Alliance's vagueness about reaction to adversary act of hostility below the level of a major aggression. This problem broke out openly during the Berlin Crisis and later the Cuban Missile Crisis, when NATO's nuclear reaction could lead to the fatal consequences.

As a response to this situation Alliance launched preparation of a new SC. The USA played a substantial task and "started advocating a stronger non-nuclear posture for NATO and the need for a strategy of "flexible response" [10]. This drift caused resistance particularly from France, which was afraid of its sovereignty loss. Finally arguments over nuclear weapons importance caused the withdrawal of France from NATO's integrated military command structure in 1966.

The fourth NATO's SC was issued on 16th January 1968. It defined three types of military response to the aggression [11]:

- direct defence,
- deliberate escalation,
- general nuclear response.

Flexibility and escalation were two key features of this SC.

Considering the MC 14/3 strategy flexibility its validity was preserved till the Cold War termination.

3 SC AFTER THE END OF THE COLD WAR

From 1989 radical changes occurred in the Euro-Atlantic area which required the Alliance reaction. The fall of the Berlin Wall in November 1989 and the end of the Cold War (officially in November 1990) were the most striking changes. In the second half of 1989 communist regimes in the Eastern Europe collapsed. Consequently the Warsaw Pact was dismantled in June 1991. Germany was united in October 1990 and in December 1991 the Soviet Union was dissolved.

These realities were reflected in new SC creation. Because of putting stress on adopted arrangements for defence transparency, SC of 1991 was for the first time issued as an unclassified document released to the public.

3.1 The Alliance's New Strategic Concept November 8, 1991

The SC of 1991 was based on an assumption that full-scale attack on all of NATO's European fronts has effectively been removed and thus no longer provides the focus for Allied strategy. On the other hand new security threats appeared, prediction of which was complicated. Instabilities may arise from the serious economic, social and political difficulties, including ethnic rivalries and territorial disputes, which are faced by many countries in central and Eastern Europe [12].

The SC of 1991 fixed four basic security tasks for the Alliance:

- to provide one of the indispensable foundations for a stable security environment in Europe,
- to serve as a transatlantic forum for Allied consultation on any issues that affect their vital interests,
- to deter and defend against any threat of aggression,
- to preserve the strategic balance within Europe.

Military repercussions for the SC of 1991 were issued in the document “MC Directive for Military Implementation of the Alliance's Strategic Concept (MC 400)”. Unlike the SC of 1991, this document was classified.

3.2 The Alliance's Strategic Concept April 24, 1999

At the Madrid Summit in 1997 the heads of NATO states raised a demand for SC revision. It was substantiated by the political and security changes which had become even more topical since the SC of 1991 adoption.

The SC of 1999 specified security risks as „multi-directional and often difficult to predict“[13]. These risks have occurred in and around the Euro-Atlantic area and include, inter alia, economic difficulties, ethnic conflicts, religious rivalries, dissolution of states, WMD proliferation, global spread of technology which can be of use in the production of weapons.

NATO transformation, initiated after the Cold War ending, was influenced by two main changes. The first was NATO enlargement when the first three countries from the former Soviet bloc obtained an invitation and the second was a settlement of situation in Bosnia and Herzegovina. Just the second change evoked the Treaty Article 5 reinterpretation and was an impulse for mission outside the Alliance’s area.

The principal security tasks were modified as follows:

- security,
- consultation,
- deterrence and defence,
- in order to enhance security and stability of the Euro-Atlantic area:
  - crisis-management on case-by-case basis and by consensus,
  - partnership, cooperation, and dialogue in the Euro-Atlantic area.

The SC of 1999 is the latest one up to now. Like previous SC-s, it has been complemented by a document concerning military matters. In this case it was “MC Guidance for the Military Implementation of the Alliance Strategy” (MC 400/2)”. Guidance is classified.

4 THE NEW SC

Each time new challenges have arisen; NATO nations have sought a new consensus on the changing strategic environment and how to address it together by crafting a guidance document for the Alliance. The current SC was adopted before the September 11 assaults and anthrax attacks in the United States and major terrorist attacks in Europe, before transatlantic dissonance over the invasion of Iraq, before Alliance engagement in Afghanistan, before additional waves of NATO and EU enlargement, before cyber attacks on Estonia, the reappearance of an assertive Russia and many other global trends [14]. These events are only some examples which have occurred since April 1999 and evoked necessity to work on the “new” SC.

Even though the Alliance tackled a new SC after the Strasbourg/Kehl Summit in April 2009 (instructions for the NATO Secretary General were given in „Declaration on Alliance security“[15]), the high level documents have been approved during the past ten years. They have guided the current SC. Moreover some documents and projects have been worked up. Their purpose was to provoke wide discussions concerning future global security environment.

In this sense, the measures regarding new capabilities adapting Alliance to contemporary challenges as follows:„The Prague Capabilities Commitment“, NATO Response Force creation and NATO military command structure streamlining were affirmed at the Prague Summit 2002.

In August 2004 NATO's two strategic staffs developed document „Strategic Vision: the Military Challenge“in order to provide a vision of Alliance operations for the next 15 years.

At the Riga summit 2006, another distinctive document "Comprehensive Political Guidance (CPG)" was adopted. „CPG provides a framework and political direction for NATOs continuing transformation, setting out, for the next 10 to 15 years, the priorities for all Alliance capability issues, planning disciplines and intelligence" [16]. CPG set, among other things, the following capability requirements:

- the ability to conduct and support multinational joint expeditionary operations far from home territory with little or no host nation support and to sustain them for extended periods,
- the ability to deter, disrupt, defend and protect against terrorism,
- the ability and flexibility to conduct operations in circumstances where the various efforts of several authorities, institutions and nations need to be coordinated in a comprehensive manner.

In April 2009 a report of the Danish institute for international studies "Come home, NATO? The Atlantic Alliance's New Strategic Concept" [2] was issued. From NATO's political point of view the Report states threats and fundamental security tasks and from military aspect the necessary measures for response to future challenges and threats. In replay of a question „Global or regional NATO?" the authors expect NATO to keep trajectory of globalization.

In May 2009 the Allied Command Transformation (ACT) completed the „Multiple Futures Project (MFP)" [17]. A wide range of politicians, academicians and experts from all transatlantic community participated in the Project. It provides an identification of the Alliance's future threats and challenges. The Project assessed nine drivers of change (friction, economic integration, asymmetry, changing state capacity, resource allocation, competing ideologies and world views, climate change, use of technology and demographics) which have an impact on the future security environment. The core of the next step was to combine these drivers and create four plausible futures (dark side of exclusivity, deceptive stability, clash modernity and new power politics). These futures represented a basis for describing 40 risk conditions. Finally 33 security implications and 26 military implications were derived.

Under the authority of the NATO Secretary General, a major security conference in Brussels in July 2009 [18] was organized. This event has formally launched the process leading to the new SC and brought together a broad range of representatives from Allied and Partner governments, NATO structures, international organizations, civil society, including parliaments, the corporate sector, NGOs, think tanks, academia and the media. In his speech, a former director of the German Institute for International and Security Affairs and the International Institute for Strategic Studies in London, Christoph Bertram described NATO as one existing multilateral transatlantic forum. He highlighted NATO's engagement in Afghanistan as a major direct concrete challenge and also advised to perceive NATO's enlargement not as the highest priority but more peaceful and less dangerous, to reduce Alliance's reliance on nuclear weapons and to put stress on deterrence as a NATO's traditional not only nuclear concept.

One of the viewpoints how to meet 21st century challenges and risks was published in the NATO project "Alliance Reborn: An Atlantic Compact for the 21st Century". Among others, the project resonates with idea of building transatlantic resilience by the protection of the Alliance's connectedness, not just its territory. In addition, the Alliance is considered indispensable but not insufficient for dealing with all global threats what calls for the comprehensive approach. "NATO's effectiveness depends on solid partnerships. NATO should establish a truly strategic partnership with the EU and meaningful partnerships with the UN, the OSCE and the African Union" [15].

In his speech delivered at Chatham House in London in July 2009, the NATO Secretary General Jaap de Hoop Scheffer stressed the fact that we try to develop a 21st century Alliance, but we do so with a 20th century mindset. "Another truth of the new strategic environment is that NATO can no longer be a solo-player. Reality calls for a new understanding of solidarity among the Allies. And they also call for new forms of cooperation between the Alliance and the wider world" [19].

5 CONCLUSION

SC has various functions: codification of past decisions and practices, public diplomacy during communication the NATO's strategy effort to wider national audiences and to provide strategy direction [3]. The last attribution has to reflect widely varied strategy environment with new challenges and risks.

The future is not foreseeable. Nevertheless some development trends are obvious even now. If we want to specify the most important elements of Alliance's reform they will be represented by considerable strengthening of Alliance's global face and building of strategic partnership between Alliance and EU and other international organizations. This process has to reach a new higher qualitative level.

References

[1] *The North Atlantic Treaty* [online]. NATO, Apr. 1949 [quoted 2009-07-13]. Accessible on the Internet: < http://www.nato.int/docu/basictxt/ treaty.htm >.

[2] RINGSMOSE, J, RYNNING, S.: *Come home, NATO? The Atlantic Alliance's New Strategic Concept* [online]. [Copenhagen (Denmark)]: DIIS, Apr. 2009 [quoted 2009-07-13]. Accessible on the Internet: <http://www.diis.dk/graphics/Publications/Reports2009/DIIS_report_200904_RingmoseRynning_NATO_web.pdf>.

[3] MAJER, M.: Potrebuje NATO novú strategickú koncepciu? In: *Euro – Atlantic Quarterly*, mar. 2006, issue EAQ 1/2006.

[4] PEDLOW, W. G (ed.): The evolution of NATO strategy 1949 – 1969 [online]. In: *NATO Strategy documents 1949 – 1969.* [Brussels (Belgium)]: SHAPE, Oct. 1997 [quoted 2009-07-14]. Accessible on the Internet: <http://www.nato.int/docu/stratdoc/eng/intro.pdf>.

[5] DC 6/1 The Strategic Concept for the Defence of the North Atlantic Area. In: *NATO Strategy documents 1949 – 1969.* [Brussels (Belgium)]: SHAPE, Oct. 1997 [quoted 2009-07-14]. Accessible on the Internet: <http://www.nato.int/docu/stratdoc/eng/ a491201a.pdf>.

[6] *History of SHAPE and Allied Command Operations* [online]. [Brussels (Belgium)]: SHAPE, Jul. 2009 [quoted 2009-07-14]. Accessible on the Internet: < http://www.nato.int/shape/about/background2.htm#1>.

[7] MC 3/5 (Final) The Strategic Concept for the Defence of the North Atlantic Area. In: *NATO Strategy documents 1949 – 1969.* [Brussels (Belgium)]: SHAPE, Oct. 1997 [quoted 2009-07-14]. Accessible on the Internet: <http://www.nato.int/docu/stratdoc/eng/ a521203a.pdf>.

[8] CM (56) 138 (Final) Directive to the NATO Military Authorities from the North Atlantic Council. In: *NATO Strategy documents 1949 – 1969.* [Brussels (Belgium)]: SHAPE, Oct. 1997 [quoted 2009-07-14]. Accessible on the Internet: <http://www.nato.int/docu/stratdoc/eng/a561213a.pdf>.

[9] MC 14/2 (Rev) (Final Decision) Overall Strategic Concept for the Defence of the North Atlantic Area**.** In: *NATO Strategy documents 1949 – 1969.* [Brussels (Belgium)]: SHAPE, Oct. 1997 [quoted 2009-07-14]. Accessible on the Internet: <http://www.nato.int/docu/stratdoc/eng/ a570523a.pdf>.

[10]*Strategic Concept* [online]. NATO, Aug. 2009 [quoted 2009-07-14]. Accessible on the Internet: <http://www.nato-otan.org/cps/en/SID-7415F406-737E1397/natolive/topics_56626. htm?selectedLocale=en>.

[11] MC 14/3 (Final) Overall Strategic Concept for the Defence of the North Atlantic Treaty Organization Area**.** In: *NATO Strategy documents 1949 – 1969.* [Brussels (Belgium)]: SHAPE, Oct. 1997 [quoted 2009-07-14]. Accessible on the Internet: <http://www.nato.int/docu/stratdoc/eng/a680116a.pdf>.

[12] *The Alliance's New Strategic Concept* [online]. NATO, Nov. 1991 [quoted 2009-07-14]. Accessible on the Internet: <http://www.nato.int/cps/en/natolive/official_ texts_23847.htm>.

[13] *The Alliance's Strategic Concept* [online]. NATO, Apr. 1999 [quoted 2009-07-14]. Accessible on the Internet: <http://www.nato.int/cps/en/ natolive/official_texts_27433.htm>.

[14] HAMILTON, D., et. al.: *Alliance Reborn: An Atlantic Compact for the 21st Century* [online]. The Washington NATO Project, Feb. 2009 [quoted 2009-07-22]. Accessible on the Internet: <http://transatlantic.sais-jhu.edu/ Publications/nato_report_final.pdf>.

[15] *Declaration on Alliance Security* [online]. NATO, Apr. 2009 [quoted 2009-07-13]. Accessible on the Internet: <http://www.nato.int/cps/en/natolive/news_ 52838.htm?mode=pressrelease>.

[16] *Comprehensive Political Guidance* [online]. NATO, Nov. 2006 [quoted 2009-07-23]. Accessible on the Internet: <http://www.nato.int/docu/basictxt/b061129e. htm>.

[17] *The Multiple Futures Project* [online]. NATO ACT, 2009 [quoted 2009-07-23]. Accessible on the Internet: < http://www.act.nato.int/multiple futures/20090503_MFP_finalrep.pdf>.

[18] *Launching NATO's New Strategic Concept* [online]. NATO, Jul. 2009 [quoted 2009-07-23]. Accessible on the Internet: <http://www.nato.int/cps/en/natolive/events_55992.htm>.
[19] *The Future of the Atlantic Alliance* [online]. NATO, July 2009 [quoted 2009-07-23]. Accessible on the Internet: http://www.nato.int/cps/en/natolive/ opinions_56498.htm>.

Assoc. Prof. Dipl. Eng. Peter SPILÝ, PhD.
Armed Forces Academy of General M. R. Štefánik
Security and Defence Department
Demänová 393
031 01 Liptovský Mikuláš
Slovak Republic
E-mail: peter.spily@aos.sk

Col. (GS) Assoc. Prof. Dipl. Eng. Pavel NEČAS, PhD.
Vice-rector for Science
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 01 Liptovský Mikuláš
Slovak Republic
E-mail: pavel.necas@aos.sk

TRAFFIC LIGHTS CONTROL USING RECURRENT NEURAL NETWORKS

Michal TURČANÍK

Abstract: The paper presents theoretical and experimental investigation of traffic lights control using recurrent neural networks. The structure of the recurrent neural network for traffic lights control and a learning algorithm for recurrent neural networks is presented. Verification of the functionality of the designed recurrent neural network was done through dynamic simulation model of selected traffic lights.

Keywords: Feed forward and recurrent neural networks, control system.

1 INTRODUCTION

In the past, when there were few vehicles on the roads, the time-of-day (TOD) traffic signal worked very well. The TOD signal operates on a preset signal-cycling scheme independent of traffic conditions. It cycles on the basis of the average number of cars passed the traffic lights that is stored in the memory device of an electric signal unit. Today, with the increasing traffic and congested roads, the conventional traffic light creates start-up delay time and end-lag time. It is estimated that this way 30 to 45% efficiency in traffic handling is lost, as well as increased fuel costs, since it is not optimized for today's traffic conditions.

2 ARTIFICIAL NEURAL NETWORKS

To solve this problem, traffic lights using neural networks are investigated. This scheme uses a traffic lights control, which changes signal based on the passing vehicle through passing area. Through computer simulation, this method has been proven to be much more efficient than fixed time interval signal since the average waiting time, average vehicle speed, and fuel consumption will be improved (1, 2, 3, 4).

Neural networks have been trained to perform complex functions in various fields, including pattern recognition, identification, classification, speech, vision, and control systems. Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in the nature, the connections between elements largely determine the network function. A neural network can be trained to perform a particular function by adjusting the values of the connections (weights) between elements. Typically, neural networks are adjusted, or trained, so that a particular input leads to a specific target output. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically, many such input/target pairs are needed to train a network.

Many variations of the perceptron were created by Rosenblatt (5). One of the simplest was a single-layer network whose weights and biases could be trained to produce a correct target vector when presented with the corresponding input vector. The training technique used is called the perceptron learning rule. The perceptron generated great interest due to its ability to generalize from its training vectors and learn from initially randomly distributed connections. Perceptrons are especially suited for simple problems in pattern classification. They are fast and reliable networks for the problems they can solve.

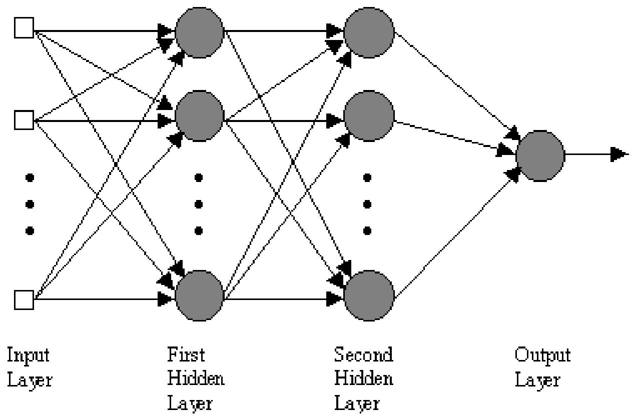


Figure 1 Multi- layer perceptron

Another variation on the multilayer perceptron was invented by Jeff Elman. A three-layer network is used, with the addition of a set of "context units" in the input layer (Figure 2). There are connections from the middle (hidden) layer to these context units fixed with a weight of one. At each time step, the input is propagated in a standard feed-forward fashion, and then a learning rule is applied. The fixed back connections result in the context units always maintaining a copy of the previous values of the hidden units (since they propagate over the connections before the learning rule is applied). Thus the network can maintain a sort of state, allowing it to perform such tasks (e.g. sequence-prediction) that are beyond the power of a standard multilayer perceptron (6).

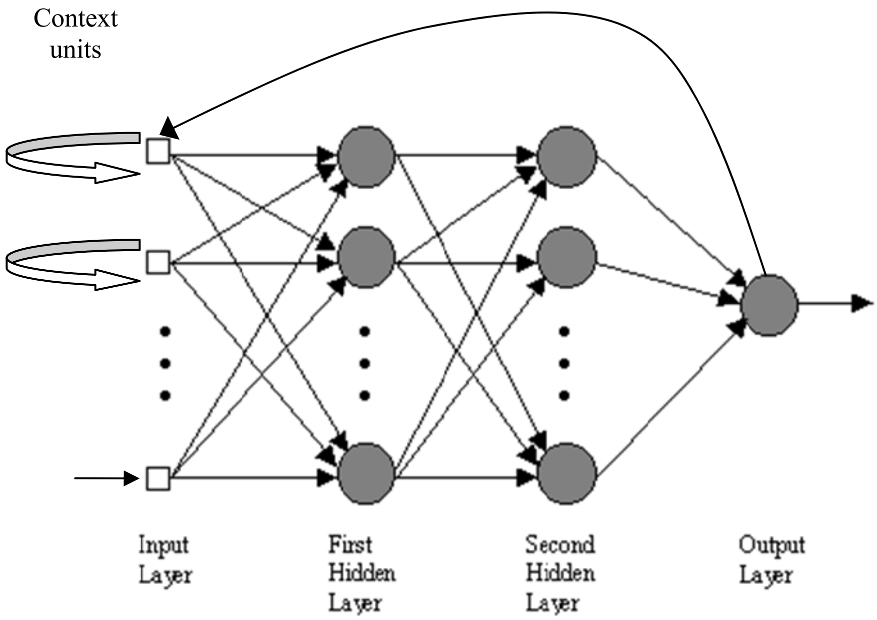


Figure 2 Recurrent multi- layer perceptron

3 NEURAL NETWOKS DESIGN FOR TRAFFIC LIGHTS CONTROL

3.1 Model of the traffic lights controlled by neural network

The central part of the new controller is a neural network. It has to be designed to be able to deal with many different (traffic) situations quickly and autonomously (without requiring human interference), interacting with its environment through sensors and effectors. Figure 3 depicts the model of the traffic lights controlled by a neural network. The first step in the adaptive traffic lights control is collecting of information about traffic intensity. It means that the number of passing vehicles through all passing areas that are controlled by the traffic lights is measured by sensors. The second step is the preprocessing of the traffic information. The neural network computes timing information for the control unit of the traffic lights controller on the basis of the information coming from the sensors. The neural network does not control the traffic lights directly. The outputs of the neural network are time intervals (timing) to control opening and closing specific directions of the traffic lights. Direct control of the traffic lights is done by control unit which works in an infinity loop until new setting is applied.

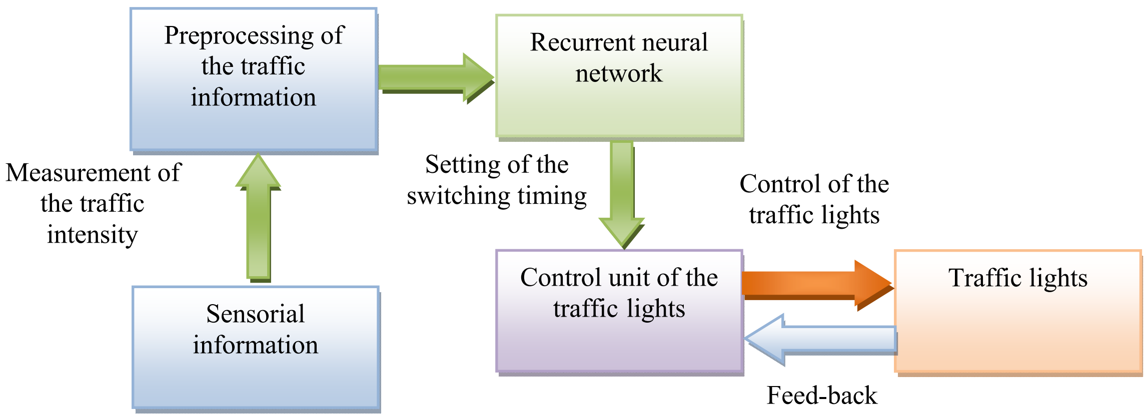


Figure 3 The model of the traffic lights controlled by recurrent neural network

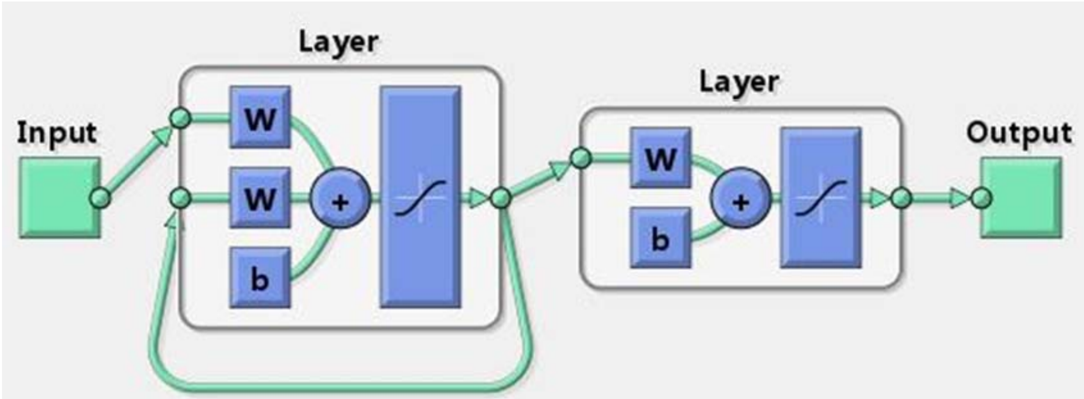


Figure 4 The structure of layer recurrent neural network

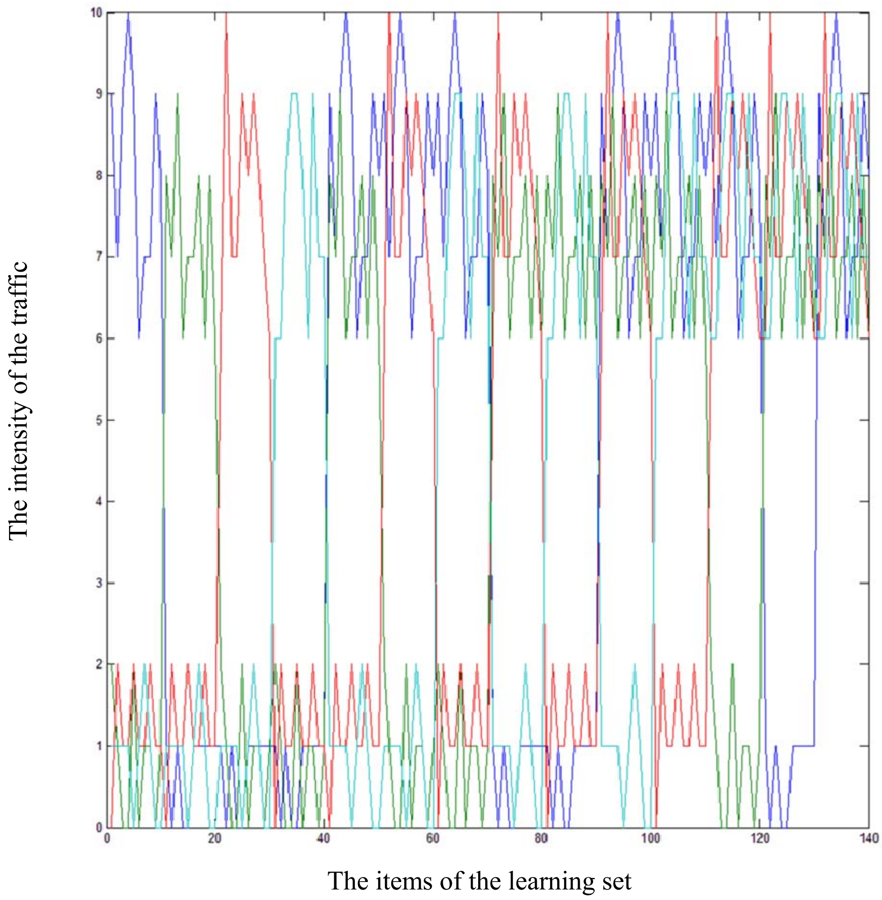


Figure 5 The learning set for four lines or directions

3.2 Structure of the neural network

The structure of the layer recurrent neural network is shown in Figure 4. The input layer of the neural network receives pre-processed traffic intensity information. The number of the input layer neurons corresponds to the number of the lines or directions that are controlled by traffic lights. The number of the neurons of the hidden layer is set to 20. This number could be a target of the next optimization. The number of the neurons of the output layer corresponds to the number of the neurons of the input layer. The value of the output layer´s neurons represents timing information for the traffic lights control. All layers are fully interconnected and there are feedback connections between the hidden and input layers.

3.3 Learning, validation and testing sets

Training set can be made easily directly from the time series. A certain number of measured values is used as inputs and the value to be predicted (i.e., the value in the future, in some chosen distance after these input measured values) is used as required output. The input part of the time series is called window, the output part is the predicted value. By shifting the window over time series the items of the training set are made. It is advised to left part of time series for testing, i.e., not to use this part during learning, but to use it to test how successfully the network learned to predict our data. The training set obtained in this way can be then adjusted for the needs of a particular neural network (7, 8, 9).

Available data are often divided into three set: learning set, validating set and testing set. These sets can overlap and do not have to be continuous. The learning set (see Figure 5) is a sequence that is shown to the neural network during the learning phase. The network is adapted to it to achieve required outputs (weights and bias in the network are changed according to this set). The difference from the required output is measured using the validating set and this difference is used to validate whether the learning of the network can be finished. The last set, testing set, is then used to test whether the network is able to work also on the data that were not used in the previous process.

3.4 Experiment

For this experiment a simple crossroad was designed. The infrastructure is shown in Figure 6. In this network, we created four streams of traffic, the first one moving from A1 to A2 and B2, the second one moving from A2 to A1 and B1, the third one moving from B1 to B2 and A1 and the last one moving from B2 to B1 and A2. For this experiment all routes had the same priority and there was no route with higher waiting time than the others. All vehicles that traveled through the crossroad were generated by a random generator. It was compared time needed for processing defined number of cars through the crossroad by two methods. In the first one fixed time division for controlling the traffic lights was used and the second one was based on neural networks.

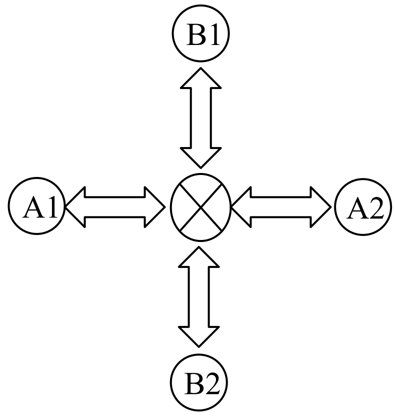


Figure 6 Model of the crossroad

3.5 Results

The results show that the layer recurrent neural network is better than fixed time division to control the traffic lights. The average number of cars in the waiting queue for the defined model of the crossroad for control by means of the neural network is 0.44 and by fixed time division is 2.1. Controlling traffic lights by neural network results in more than 70% lower waiting times when compared with a fixed time division for controlling the traffic lights.

4 CONCLUSION

In this article it was shown that traffic control is an important research area and its benefits make investments worthwhile. It was described how traffic lights can be controlled by neural networks and practical use of layer recurrent neural network models was shown. In section 2 artificial neural networks were introduced and their utilization as an optimization tool for various control problems was demonstrated. Then an experiment regarding traffic lights control comparison was described. One series of experiments was performed using the designed model of the traffic lights. The experiments were performed on the defined infrastructure. The results of the experiments show that layer recurrent neural networks on simple infrastructure outperform the fixed controllers by reducing waiting time by more than 70%. This number is very high and we can propose that with growing complexity of the infrastructure this number would be decreasing. The work in this area could follow on more complex and city-like infrastructure for the future experiments. Also the structure of the layer recurrent neural network could be a target of furtheroptimization. The number of the neurons of the hidden layer and the number of the time steps that are used for feedback are suitable for optimization.

References

[1] WIERING, M., KOOPMAN, A.: *Intelligent Traffic Light Control*, Technical Report UU-CS-2004-029, July 9, 2004, 31 pages.

[2] WIERING, M.: *Multi-agent reinforcement learning for traffic light control*. In Langley, P., editor, Proceedings of the Seventeenth International Conference on Machine Learning (ICML'2000), pp. 1151–1158.

[3] TUBAISHAT, M., SHANG, Y., HONGCHI, S.: *Adaptive Traffic Light Control with Wireless Sensor Networks*, Proc. of the Consumer Communications and Networking Conference, 2007. CCNC 2007. 4th IEEE, USA, Jan. 2007, 187-191.

[4] ROCHNER, F., MULLER, C.: *Adaptive Decentralized and Collaborative Control of Traffic Lights*. Informatik 2004, Ulm, 20. - 24. September 2004.

[5] ROSENBLATT, F.: *Principles of Neurodynamics*, Washington D.C., Spartan Press, 1961.

[6] ELMAN, J. L.: *Finding structure in time,* Cognitive Science, Vol. 14, 1990, pp. 179-211.

[7] TURČANÍK, M., MOKRIŠ, I.: *Possible Approach to Optimization of Neural Network Structure*. Proc. of Conf. SECON 97, Warsaw, pp. 326 - 335.

[8] PATTRESON, D.: *Artificial Neural networks - Theory and Applications*, Prentice Hall, 1996.

[9] MELACCI, S., SARTI, L., MAGGINI, M., BIANCHINI, M.: *A Neural Network Approach to Similarity Learning. Lecture Notes in Computer Science*, Artificial Neural Networks in Pattern Recognition, June 30, 2008, pp. 133-136.Airborne.

Dip. Eng. Michal TURČANÍK, PhD.
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 01 Liptovský Mikuláš
Slovak Republic
E-mail: michal.turcanik@aos.sk

AIRBORNE SIDE AND AUTOMATED EN-ROUTE CONFLICT DETECTION AND RESOLUTION

Jiří VAŠEK, Pavel KOLČÁREK, Petr KRUPANSKÝ, Christopher MISIAK

Abstract: This paper presents a summary of results gathered during the simulator based airborne impact assessment of the ERASMUS research project. The ERASMUS project is a European Commission FP6 project, which tries to significantly improve efficiency, effectiveness and safety of the European Air Traffic management system. Its goal is to provide automatic conflict detection and resolution up to 20 minutes ahead of the conflict during the en-route phase of the flight, while maintaining the advantage of actively involved Air Traffic Controllers. In spite of the major focus on the ground side of the Air traffic management, the implementation of ERASMUS in the future European airspace heavily depends on a number of airborne related issues such as acceptability, timely response or situation awareness. Results related to these areas are presented in the paper and based on them, modifications of the current onboard equipment related to the processing of the conflict resolution clearance onboard are proposed.

Keywords: Automated aircraft separation management, transaction time, acceptability, data-link communication.

1 INTRODUCTION

The forecasts of the International Civil Aviation Organization (ICAO) envisaged that the air traffic in Europe will double within the 2002/2020 timeframe [7]. The goal of the ERASMUS (En-Route Air traffic Soft Management Ultimate System) project is to help the European ATM to prepare for such growth in air traffic; i.e., to investigate the possibility of accommodating increased levels of aircraft operations without negative impacts on Air Traffic Controller (ATCO) workload or safety [2].

1.1 ERASMUS koncept

ERASMUS is considered to be implemented in the future airspace by year 2020 as an automated en-route medium term conflict detection and resolution (CD&R) tool. When a conflict is anticipated, a solution based on a speed adjustment of a selected aircraft will be generated [8]. This proposed solution will be delivered to the selected aircraft via the Controller Pilot Data Link Communication (CPDLC) channel in a form of clearance containing a Required Time of Arrival (RTA) command. Emerging from previous text, ERASMUS requires two key technologies for its airborne operation: (1) CPDLC and (2) the RTA functionality provided by some flight management systems (FMS).

In its current configuration, ERASMUS uses a combined clearance. First part of the clearance creates a new waypoint in the flightplan, and second part places a RTA constraint upon this new waypoint. These ERASMUS' generated solutions correspond to speed changes of -6% or +3% of the current aircraft's speed. These speed changes were proved not drawing the attention of controllers away from their normal tasks and make the effects of the system's operations undetectable for them [1]. ERASMUS will not inform the ATCO about sending a clearance to an aircraft, but the ATCO will be aware of ERASMUS operating within his/her sector.

2 AIRBORNE SIDE AND ERASMUS

As mentioned earlier, ERASMUS' main goal is to decrease the workload of ATCO. The ERASMUS' en-route separation will be managed without direct involvement of the ATCO, who will benefit from the ERASMUS actions only indirectly – by having less tactical interventions for conflict resolution.

In contrast, the pilot will be directly involved in the ERASMUS activities, because he/she is expected to process the ERASMUS clearance and, ideally, to accept it and follow its instructions. Various steps need to be performed by the pilot to process the ERASMUS clearance onboard.

Three main areas related to the airborne side of ERASMUS were identified: acceptability, time, and situational awareness.

2.1 Acceptability

The ERASMUS program assumes that the concept will be readily accepted by the aircrews, because its clearances are expected to minimally affect airborne en-route operations [2]. This is because ERASMUS clearances are assumed to (1) request minor speed changes, (2) not violate the aircraft's speed envelope, (3) have minimal impact on arrival times, (4) have minimal impact on fuel consumption, and (5) have minimal impact on pilots' workload. These expectations may be correct, but there are still several factors with a potential to significantly reduce the acceptability of ERASMUS concept for pilots (e.g. change in ATCO-pilot interaction, uncertainty that the detected conflict would really occur, uncertain costs for airlines, the magnitude ofspeed change may not be perceived minor by pilots) [3].

2.2 Time

The time plays very important role on the airborne side of ERASMUS, because the pilot should process the ERASMUS clearance onboard within a time interval of 3 minutes which is based on the ERASMUS' iteration cycle. If exceeding the time interval, ERASMUS will assume the proposed clearance was rejected and the conflict resolution was unsuccessful. A significant number of rejected/expired clearances would have a detrimental impact on ERASMUS' operational efficiency. Time related to the airborne side comprises from system-paced and human-paced parts forming together a total transaction time described on Figure 1.

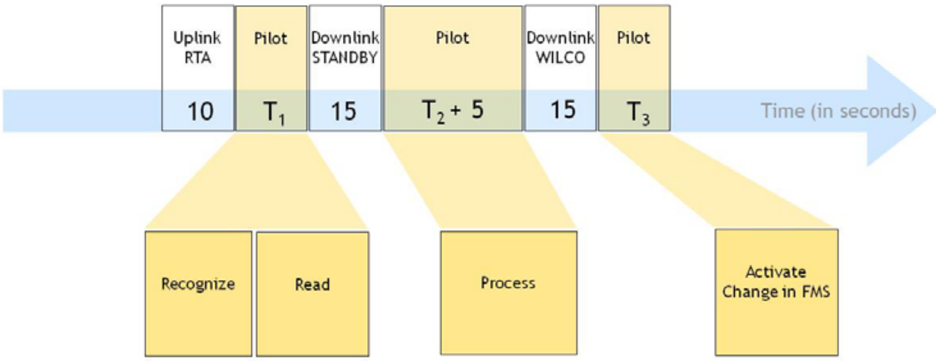


Figure 1 Transaction time components

1 st Block: Time required to transmit an ERASMUS generated RTA clearance to the aircraft (10 s).

2nd Block: T1 - Time required to notice and read the newly up-linked clearance.

3rd Block: Built-in delay of transmitting stand by (STDBY) response (15 s).

4th Block: T2 – Time required to process the clearance and make a decision plus 5 s for built-in delay of automatic loading into the secondary flight plan (F-PLN).

5th Block: Built-in delay of sending response to the ground (15 s).

6th Block: T3 – Time required to activate the secondary F-PLN.

The given system-paced delays are rather optimistic, because the time required for the ground-air communication may be longer [5]; this configuration was used in the assessment.

The areas subjected to assessment were the human-paced tasks labelled T_1, T_2, and T_3. Most of these tasks are relatively easy, but especially T_2 task requires complex decision-making and consideration of various factors.

2.3 Situation awareness

The ERASMUS' operations are not expected to change the components of situation awareness significantly. A potentially problematic component of situation awareness can be seen in the inability of pilots to ask the ATCO about any aspect of the ERASMUS clearance. More information about the conflict may be beneficial for pilots especially if a costly manoeuvre is required. ERASMUS is an automatic system which does not negotiate with the pilot nor informs pilot about the conflicting traffic. This may impair the pilots' ability to build a picture about the conflicting traffic. While situation awareness may support effective responding, it does not itself incorporate response selection and execution [9]. Hence, the decision-making process could be affected by the impaired situation awareness in the ERASMUS managed airspace, but such impact would probably not directly affect the decision-making process about the ERASMUS clearance.

3 METHOD

3.1 Participants

Eleven participants (all male) took part in the assessment. The age of participants ranged from 25 years to 50 years with an average age of 38.5 years, all were certified commercial transport aircraft pilots (A320, A330, A340, and B777).

A variety was found in the flying experience (hours spent flying the aircraft or certified simulator) between the participants. The distribution of their experience expressed as the number of total flying hours and a percent of that total flying hours spent on a FMS equipped aircraft is provided in Table 1. This table also describes the distribution of participants from the proficiency of a CPDLC and TA usage point of view; 1=always use CPDLC/RTA, 5=never use CPDLC/RTA.

Table 1 The experience of the assessment's sample

	Total flying hours	Percent of spent on FMS equipped aircraft	CPDLC use	RTA use
Mean	5173	50.5%	4,1	4,5
Minimum	300	10%	2	3
Maximum	10500	95%	5	5

Two subgroups of pilots were identified within the assessment's sample based upon total flying hours. For the purposes of discussion, five participants with the highest total hours (8000–10500 hours) were defined as "more experienced", while six participants with fewer total hours (300–3600) were defined as "less experienced"; the "less/more experienced" label is statistical and does not reflect the real experience of the pilot.

3.2 Equipment

A highly adjustable cockpit simulator connected to the traffic generator using real traffic samples was used in the assessment. Table 2 describes four assessment scenarios reflecting various conditions of ERASMUS operation. Each participant went through all exercises.

Table 2 The exercises description

Exercise	Speed change	Type of conflict	Number of RTAs
1	+3%	Crossing	2
2	–6%	Crossing	1
3	–3%	Catch-up	2
4	–6%	Catch-up	1

Objective and subjective metrics were used in the assessment. The objective measures were represented by logging various events (i.e., button presses) and by recording the display screens and the behavior of the participants and radio traffic communication.

The subjective data consisted of three types of questionnaires: (1) presented at the beginning of the assessment's day, (2) after each exercise, and (3) after all exercises. Mainly five-point scaled questions (including a "neutral" point) were used. The participants were also offered the opportunity to add any comments where they considered it to be valuable [3].

4 RESULTS

The assessment was conducted at DSNA (Direction de la Technique et de l'Innovation, DTI) in Toulouse (France).

4.1 Acceptability

A very important indicator of the acceptability of the ERASMUS clearances is the ratio of accepted (WILCO) versus rejected (UNABLE) ERASMUS clearances. Considering the overall results, 14 clearances were rejected (21%) and 52 were accepted (79%). Individual exercises' results showed the highest proportion of rejected ERASMUS clearances in Exercise 2 (45%) and in Exercise 4 (36%).

The most important factors participants considered before deciding to reject/accept the clearance were the following (given in descending order of importance): (1) speed impact, (2) fuel impact, and (3) impact on time/ETA (Estimated Time of Arrival).

Considering the acceptability components, it was found that the body of the clearance itself, the time of its arrival onboard, and the number of data-link clearances were not perceived as a problem. However, six pilots (55%) perceive the presence of the geographic coordinates in the body of the clearance as unnecessary and making the clearance difficult to read. The remaining two areas – not being able to negotiate an aspect of the clearance with ATCO and the magnitude of the speed change requested by the ERASMUS clearance were more acceptable in the Exercise 1 and 3, but it was rather the opposite in the Exercise 2 and 4. The reason lies in the –6% speed change requested. A two sample t-test showed a very significant difference in the acceptability of the magnitude speed change between Exercises 1—3 and 2—4; t(stat)=3.64, p=0.0004. The decreased acceptability of the magnitude of speed change also corresponds with the decreased acceptability of not being able to negotiate an aspect of the ERASMUS clearance with ATCO (r=0.45; p=0.01). Seven pilots (64%) also reported that they would rather change the heading or the flight level than accept the –6% speed change request. The pilots perceived the –3/+3% speed change exercises as the most acceptable (Exercises 1 and 3). Two sample t-test revealed a significant difference between the subgroups of pilots in the general acceptability of different areas, with more experienced pilots expressing significantly less general acceptability than less experienced pilots; t(stat)=3.79, p=0.0002.

4.2 Time

For the purposes of the research, the following events were considered as time stamps for the components of the transaction time:

- Notice time – clearance displayed on DCDU until the ATC message button was pressed.
- Reading time – ATC message button pressed until pressing SEND button of the STDBY response.
- Processing time – STDBY button pressed until WILCO/UNABLE button pressed (this time includes simulated delay of loading the clearance into secondary flightplan (5 s) and time required for delivering the clearance to the ground (15 s)).
- Secondary flightplan activation time – WILCO button pressed until Activate secondary flightplan button pressed.
- Total transaction time – clearance displayed on DCDU until WILCO/UNABLE button pressed (time required to for secondary flightplan activation was excluded as it is not important for the ERASMUS iteration cycle).

The percentile distribution of the total transaction time identified that 95th percentile was 158 s.

There were no significant differences between exercises neither in the total transaction time (or its components) nor between the type of the response (WILCO/UNABLE). This indicates that the transaction time is not affected by the ERASMUS' RTA characteristics (i.e., magnitude of speed change requested) nor by the number of clearances received.

Table 3 Transaction time results

Time (seconds)	Mean	Median	Min. value	Max. value
Notice time	8	6	3	29
Reading time	15	13	2	60
Processing time	68	59	26	148
Secondary flightplan activation time	10	6	2	58
Total transaction time	104	95	61	200

4.3 Situation awareness

A significant difference was identified in the ability to deduce the location of conflicting traffic between exercises in relation to the type of conflict involved. While the participants were not able to deduce the location when a crossing conflict was involved (no correct answers), it was the opposite when catch-up conflict was involved (73 % correct answers); including the certain, but also the uncertain answers.

Most of the pilots (77 %) would like to be able to see the location of the conflicting traffic on a display, in spite of the fact that importance this feature is perceived as rather neutral (~3.3).

5 CONCLUSIONS AND DISCUSSION

The ERASMUS concept was proved to be viable, feasible and generally acceptable. Nevertheless, the airborne impact assessment revealed number of issues which need to be addressed in the future. Some issues are related to the configuration of the ground side of the automated CD&R, some of them may require modification of the onboard equipment.

The automatic ground CD&R by minor speed changes is perceived as generally acceptable for pilots, but with an exception – magnitude of the speed change requested. This issue needs to be treated with caution, because only speed changes not exceeding 3 % of the current aircraft speed are readily accepted by pilots. The aspects associated with the decision-making process of the speed change request reflect the pilots' business concerns; these concerns are besides the impact on speed also ETA and fuel consumption.

Interesting finding was that more experienced pilots expressed significantly less acceptability than less experienced pilots. This is probably because the more experienced pilots are also more concerned about the airline's business issues (i.e., cost index).

Preceding its implementation, ERASMUS concept would probably have to prove that the speed changes requested (1) have globally a commercial benefit for an airline, and (2) stay in the aircraft's performance limits; especially due to airlines' business objectives.

The time plays a very important role in ERASMUS, because its efficiency highly depends on the ability of the aircrew to respond to the clearance within a 3 minutes expiration period. Besides three system-paced components (e.g., data-link communication delays), three human-paced components of the transaction time were identified; (1) „notice time", (2) „reading time", and (3) „processing time". Neither the total transaction time nor its components depend on the clearance characteristics (i.e., magnitude of speed change requested) or the number of clearances received.

While the mean value of the total transaction time (104 s) sounds promising, the 95th percentile (158 s) sounds different. The 22 s for the system-paced tasks seems not to be enough, because the time required for up-link and down-link is expected to take longer. One may object that the easiest way would be to increase the length of the ERASMUS' iteration cycle, but such increase would lead to a significant deterioration of the ERASMUS efficiency. If it is not possible to increase the length of the ERASMUS' iteration cycle, then improvement of the cockpit interface facilitating the decision-making process will be required.

The situation awareness was investigated through the pilots' awareness of the location of conflicting traffic. The ability of deducing the location of conflicting traffic depends on the type of the conflict. A very significant difference between the crossing conflict and the catch-up conflict was identified. No pilot was able to deduce the conflicting traffic location in the crossing conflict exercises contrary to 73% of correct responses in the catch-up conflict exercises.

Pilots would like to be able to have more information about the conflicting traffic location from various reasons e.g. safety, information about the conflict itself, certainty about the aircrafts in vicinity, etc. However, the participants did not perceive the ability of being able to see the conflicting traffic location as important, such information would be simply in a "nice to have" category.

5.1 Suggestions for airborne interface improvement

The suggestions for the improvement of the interface are based mainly on the acceptability and transaction time results. Most of the participants perceived the current human machine interface (HMI) as problematic, unnecessarily complicated, and therefore recommended for improvement. Several ways on how to facilitate the decision-making process, make HMI more comprehensible and user-friendly were identified.

First, it is possible to merge the loading of the clearance into the FMS with sending the STDBY response to the ground. Currently, both operations need to be done separately. It may even be possible to make sending the STDBY response to the ground unnecessary when dealing with clearance generated by ground automation. However, sending the STDBY response to the ground informs about the clearance being processed.

Second, the legibility may be improved. The presence of geographic coordinates (latitude, longitude) in the body of the clearance was perceived by majority of participants as unnecessary. Some pilots also made objections towards the font used, which was probably caused by a display used in the assessment (not a real aircraft display). The recommendation to improve legibility may be to remove the geographic coordinates from the clearance, but preserve them in a discreet mode (i.e., the coordinates would by displayed only on a pilot's request).

Third, the use of the secondary F-PLN was perceived as inappropriate, because the pilots usually use it in a different way. This temporary solution may be replaced by use of a special flightplan dedicated to the processing clearances generated by automated ground CD&R.

This work has been supported by the 6th Frame Programme of the European Commission "ERASMUS" - Contract No. TREN/06/ FP6AE/S07.58518/518276.

References

[1] AVERTY, P., JOHANSSON, B., WISE, J., Capsie, C.: *Could ERASMUS speed adjustments be identifiable by air traffic controllers?*, 7th ATM R&D Seminar. Barcelona, Spain, July 2-7, 2007.

[2] ERASMUS: D2.2.1 *Concept of operations*, Contract TREN/06/FP6AE/S07.58518/518276, 2007.

[3] ERASMUS: D4.2.1 Definition of the Experimental Plan.

[4] ERASMUS: Human Factor case: Airborne System Impact Assessment, internal Honeywell document, 2008.

[5] Flight Crew Data Link Guidance for Link 2000+ Services v.3.0, Eurocontrol, January 2007.

[6] SESAR Consortium. (30. November 2006). Air Transport Framework: The performance target. Available from: http://www.sesarju.eu/gallery/content/public/DLM-0607-001-02-00a.pdf.

[7] VILLIERS, J.: *ERASMUS - A friendly way for breaking the capacity barrier*, ITA, volume 58, June 2004.

[8] WICKENS, C. D., HOLLANDS, J. G.: *Engineering Psychology and Human Performance*, Prentice Hall, Upper Saddle River, New Jersey, 2000, ISBN 0-321-04711-7.

Jiří VAŠEK
Honeywell, s.r.o.
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: jiri.vasek@honeywell.com

Pavel KOLČÁREK
Honeywell, s.r.o.
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: pavel.kolcarek@honeywell.com

Petr KRUPANSKÝ
Honeywell, s.r.o.
Tuřanka 100
627 00 Brno
Czech Republic
E-mail: petr.krupansky@honeywell.com

Christopher MISIAK
Golden Valley
Minneaopolis
USA
E-mail: christopher.misiak@honeywell.com

APPRAISAL OF CRITICAL TRANSPORTATION INFRASTRUCTURE – RISKS AND SAFETY

Jaroslav ZAJÍČEK, Jiří HAVLÍČEK, Pavel FUCHS
Radovan SOUŠEK, Miroslav KELEMEN, Lucia KOVÁČOVÁ

Abstract: Risk management is a proper platform for solving of critical infratructure protection tasks. Risk Assessment as a substantial part of risk management contains processes and tools for risk evaluation. Transportation infrastructure disturbances are evaluated by risk measures and calculation of risk value is presented.

Keywords: Critical infrastructure, risk assessment, risk evaluation, probability, consequence, FMECA.

1 INTRODUCTION

Critical transportation infrastructure is a part of national infrastructure, which comprises, besides the transportation, the energy, telecommunication, banking and finance, public health and healthcare, public administration, etc. At the present time we miss a methodology we could apply to this issue directly without any adjustments.

2 ASSESSED FACTORS

Risk assessment has to be based on the principles of combination of probability and consequences. Further categorization of the probability in assessment process is not advisable, the attention should be focused on the consequences resulting from the assessed undesired occurrences.

It is advisable to substitute the non-dimensional probability by the frequency of occurrence, which, assuming the real values, is expressed by the time units [h^{-1}]. By doing this it won't be necessary to relate the assessed factor to the a particular time period, e. g. 1 year.

Evaluation of consequences could be separated to the infrastructure operator's costs related to an object reconstruction and the total expenses of road traffic participants varying in dependence on the length of a potential diversion route, traffic density and the time required for the particular object reconstruction.

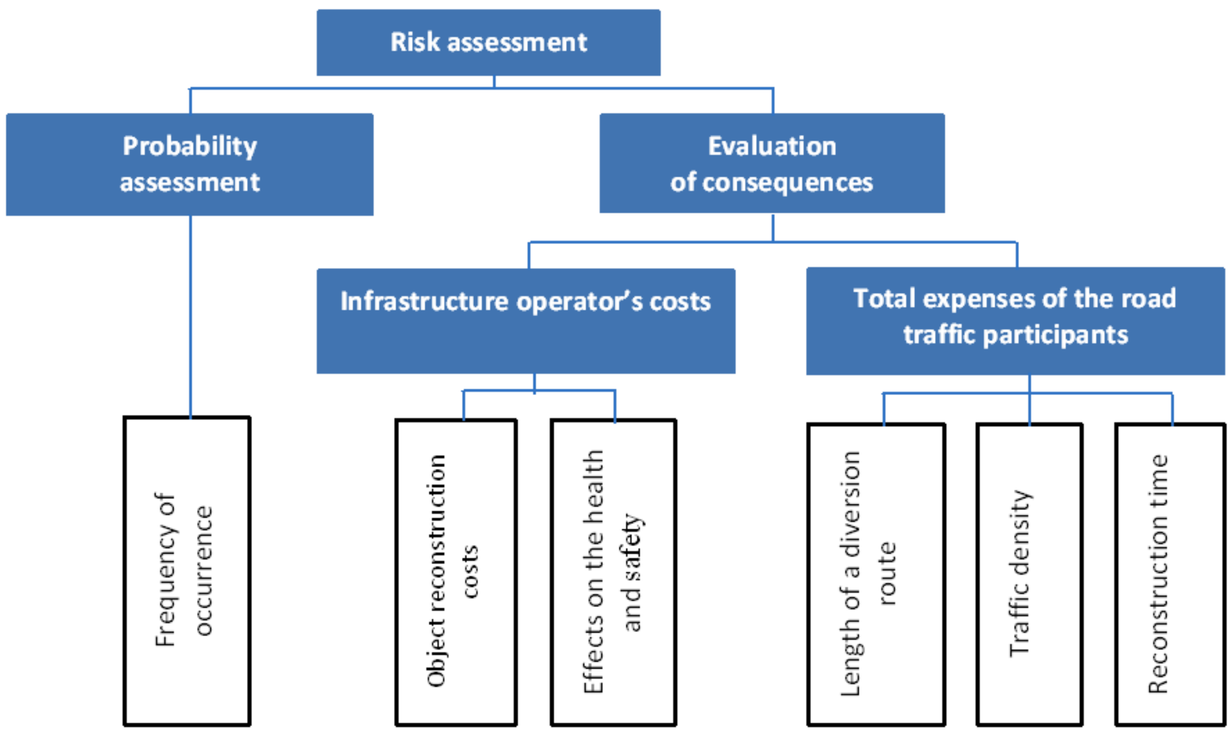


Figure 1 Risk assessment processes

3 ASSESSED OBJECT

Crisis management of the Czech Republic considers the road and railway infrastructure to be critical. From the national point of view the railway infrastructure is more critical than the road one, but regarding to the territorial division of a country, the road infrastructure appears more critical and will be further presented. Selected types of road infrastructure objects to be assessed are as follows: bridge, tunnel, junction, stretch of road, railway crossing.

Points and lines as the map symbols represent the particular types of objects. Line objects assessment has to respect the length of assessed object, but it is not necessary to separate the results of line objects analysis from those obtained by performing the point object analysis.

4 ASSESSED OCCURRENCES

Any undesired occurrence pertaining to the traffic infrastructure objects, such as ordinary wear and tear, wrong or defective construction, natural phenomena, traffic accidents, etc. can be assessed. Wear, accident, ... causes a partial or total closure. Considering the risk it is possible to compare undesired occurrences of examined objects globally as well as separately in defined categories in regard to a kind of undesired occasion.

5 VALUE SCALES

All of the valuation scales are of five points. Evaluation of the factors does not primarily take into account the complicated risk/risk value assessment. If there are more factors to be assessed than the standard calculation methods usually use, the calculation would not be a simple sum or product of values although a value scale would be adjusted. Primarily it is important to adapt the scales to the real values as the output data and also to the commonly used terms (timescale constructed of the time periods such as day, week, month, etc., instead of 1 hour, 10 hours, 100 hours, etc.), and also to the anticipated probability distribution identifying probability of the value falling into a defined interval.

All of the designed scales are five-point. Analysis and evaluation do not require the scales to be of the same number of levels, but it seems useful for those performing the evaluation. It is recommended to define a designation for each scale value.

Occurrence rate

- Designation of the factor (indicator): F_1
- Number of scale levels: 5
- Designation of real substitution value: X_1
- Real parameter measurement unit: h^{-1}
- Approximating real function: f_1

Designated time data have been chosen in consideration of commonly used ones, range and form of arithmetic progression is then identified in regard to the anticipated rate of evaluated occurrences.

Table 1 Approximating functions

F_1	Occurrence rate	X_1 [h^{-1}]	f_1
1	negligible (cca once per 100 years or less)	1,1E-06	$f_1 = 10^{F_1 - 7}$
2	low (cca once per 10 year)	1,1E-05	
3	medium (cca once per year)	1,1E-04	
4	high (cca once per month)	1,4E-03	
5	abundant (cca once per week or more)	6,0E-03	

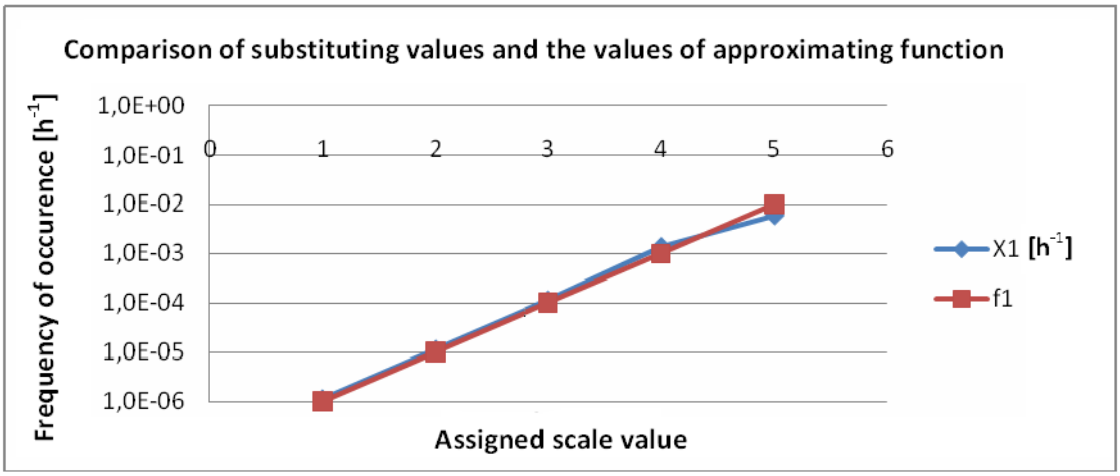


Figure 2 Frequency values comparison

Object reconstruction costs

- Designation of the factor (indicator): F2
- Number of scale levels: 5
- Designation of real substitution value: X2
- Real parameter measurement unit: CZK
- Approximating real function: . f_2

Values fall into the range of scale with the ratio equal to 10. This high-value coefficient allows the diametrically different occurrences in terms of costs to be evaluated, e. g. ice layer vs. reconstruction of a section.

Table 2 Approximating functions

F_2	Object reconstruction costs	X_2 [CZK]	Approximating functions
1	negligible (cca 1 000 CZK or less)	1,00E+03	$f_2 = 10^{F_2+2}$
2	low (cca 10 000 CZK)	1,00E+04	
3	medium (cca 100 000 CZK)	1,00E+05	
4	high (cca 1 000 000 CZK)	1,00E+06	
5	extremely high (cca 10 000 000 CZK or more)	1,00E+07	

Comparison of substitution values and the ones of approximating functions is not necessary, because these are equal for each point of evaluation.

Effects on the health and safety

- Designation of the factor (indicator): F3
- Number of scale levels: 5
- Designation of real substitution value: X3
- Real parameter measurement unit: CZK
- Approximating real function: f3

If we express the impact on healt in financial terms it can be included into the risk associated with undesired occassion as well. The price of health or life is disputable and markedly varies in different studies. There are many studies on the price of human life regarding the age of people or a country where they live, etc. Estimation of values for each level shown in the table bellow is not an issue to be worked out in this text so the data valuated by RRM software (Risk and Reliability Management) have been used without the detail analysis.

Table 3 Approximating functions

F_3	Effects on the health and safety	X_3 [CZK]	Approximating function
1	negligible (health threats without persistent effects)	1,0E+04	$f_3 = 8.10^{F_3+2}$
2	small (persistent injury)	8,0E+05	
3	medium (persistent injury of more people)	8,0E+06	
4	large (death)	8,0E+07	
5	death (death of more people)	2,5E+08	

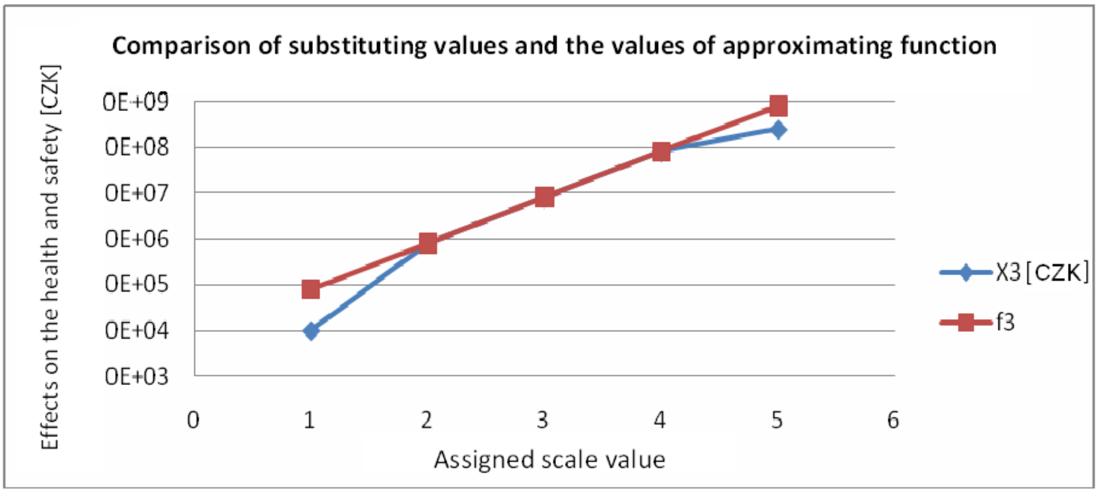


Figure 3 Effects values comparison

Length of a diversion route

- Designation of the factor (indicator): F4
- Number of scale levels: 5
- Designation of real substitution value: X4
- Real parameter measurement unit: km
- Approximating real function: f4

Table 3 Approximating functions

F_4	Length of a diversion route	X_4 [km]	Approximating function
1	negligible (cca 5 kms or less)	5	$f_4 = 3.\,2^{F_4}$
2	small (cca 10 kms)	10	
3	medium (cca 25 kms)	25	
4	large (cca 50 kms)	50	
5	extraordinary (cca 100 kms or more)	100	

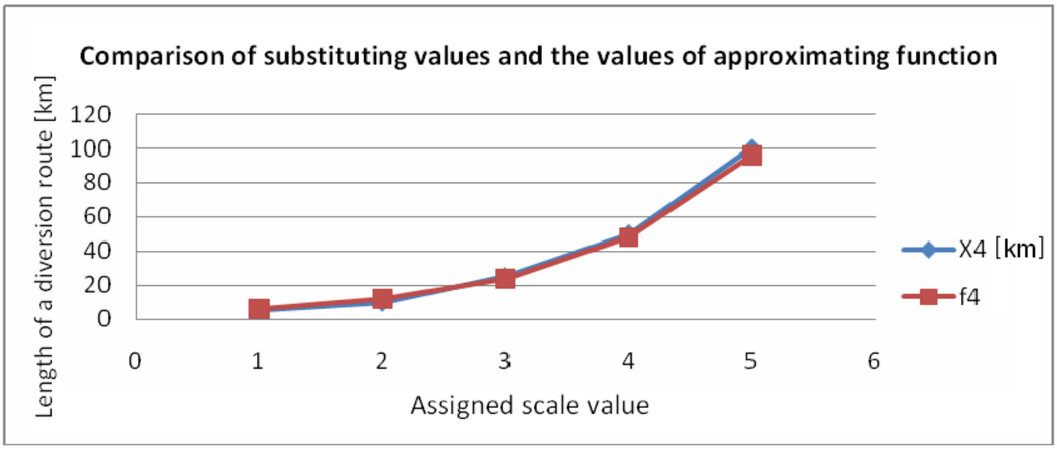


Figure 4 Length of a diversion route values comparison

Traffic density

- Designation of the factor (indicator): F5
- Number of scale levels: 5
- Designation of real substitution value: X5
- Real parameter measurement unit: ks.(24h)-1
- Approximating real function: f5

Value scale presented bellow has been designed in accordance with results of the periodical traffic.density statistics made by the Roads and Motorways Agency of the Czech Republic. Annual average of the vehicles on the motorways and main roads during the 24 hours period is being considered. If the exact values of traffic density on particular road stretches are known, it is better to abandon a semiquantification of this factor and use the statistical density values.

Table 4 Approximating functions

F_5	Traffic density	X_5 [vehicles/ 24 hrs]	Approximating function
1	negligible (cca 500 vehicles/24hrs or less)	500	$f_5 = 500.\,2{,}5^{F_5-1}$
2	low (cca 2 000 vehicles/24hrs)	2000	
3	medium (cca 6 000 vehicles/24hrs)	6000	
4	High (cca 12 000 vehicles/24hrs)	12000	
5	extremely high (cca 30 000 vehicles/24hrs or more)	30000	

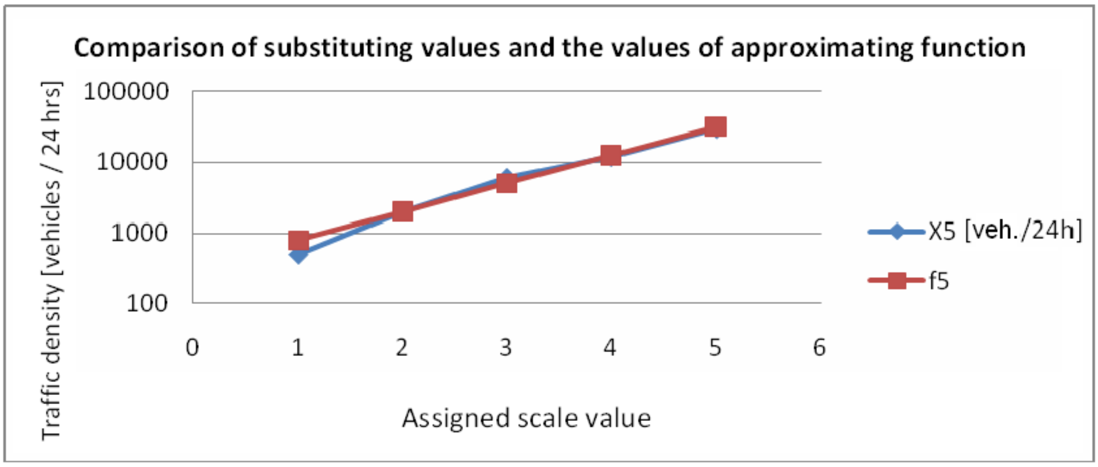


Figure 5 Traffic density values comparison

Reconstruction time

- Designation of the factor (indicator): F_6
- Number of scale levels: 5
- Designation of real substitution value: X6
- Real parameter measurement unit: h
- Approximating real function: f_6

Table 4 Approximating functions

F_6	Reconstruction time	X_6 [hrs]	Approximating function
1	negligible (cca 8 hours or less)	8	$f_6 = 6^{F_6}$
2	short (cca 1 day)	24	
3	medium (cca 1 week)	168	
4	Long (cca 1 month)	730	
5	extremely long (cca 1 year or more)	8760	

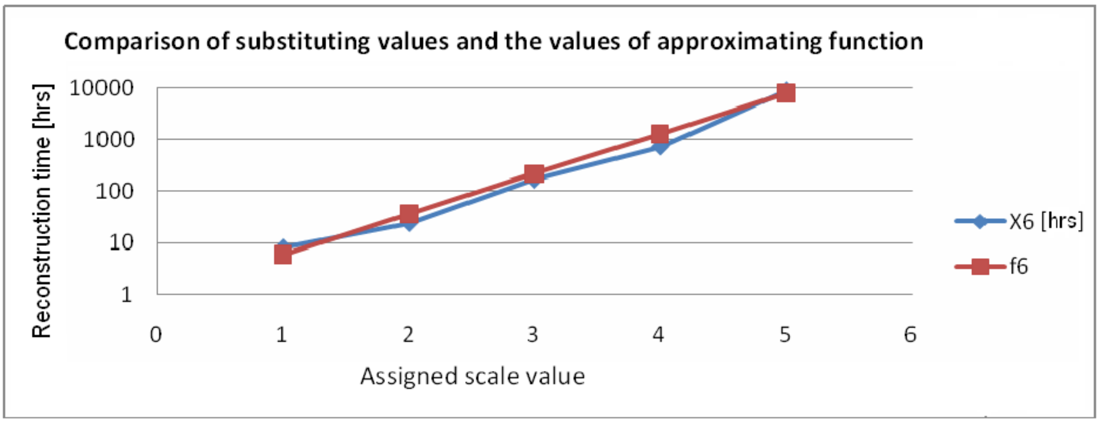


Figure 6 Reconstruction time values comparison

6 CALCULATION OF AN EVENTUAL RISK/RISK VALUE

Value scales including the substitution values and approximating functions enable us to define a function for estimating an eventual risk of undesired occurence. The risk is calculated by multiplying the occurence rate and financially estimated total consequence. Cumulative distance of the diversion of all vehicles during the reconstruction time is calculated by multiplying the X_4, X_5 and X_6 factors. For the calculation purposes it is necessary to determine the costs per 1 kilometer, symbolized by k_4. Example: following inputs can be used for k_4 calculation:

- Proportion of a freight traffic 0,1
- 1 km long freight traffic rate 35 CZK
- 1 km long passenger traffic rate 6 CZK

Average costs per one kilometer (k_4) are approx. 9 CZK (= 0,1.35+0,9.6).

Eventual risk is expressed by following term:

$$R = X_1.(X_2 + X_3 + k_4.X_4.X_5.X_6)$$

X_1 to X_6 values are the substituting real values. In case that requires the use of function, the eventual risk can be estimated by using the approximating functions f_1 to f_6, thus:

$$R \cong f_1.(f_2 + f_3 + k_4.f_4.f_5.f_6)$$

After the substitution of particular functions we get:

$$R \cong 10^{F_1-7}.(10^{F_2+2} + 8.10^{F_3+2} + 9.3.2^{F_4}.500.2{,}5^{F_5-1}.6^{F_6})$$

One option is to end the estimation after assigning the risk R to the undesired occurence. If necessary it is possible to transform the eventual risk back to the value scale, which is useful especially for the colour plotting on a map or easy comparison of assessed occurences by comparing two integral numbers falling into a defined range.

Regarding to a geometric character of the scale designed for the factor evaluation and probable distribution of a risk associated with assessed occurences, the value scales, after the transformation of a risk to a risk priority number (RPN), should be geometric as well.

Common ratio of a geometric scale q with a desired number of levels n is being derived from the following term, assuming that substituting value of a first level is *min(R)*, i. e. representing the minimum risk in the term

$$q = \sqrt[n]{\frac{\max(R)}{\min(R)}} = \sqrt[n]{\frac{X_1(10).[X_2(10)+X_3(10)+k_4.X_4(10).X_5(10).X_6(10)]}{X_1(1).[X_2(1)+X_3(1)+k_4.X_4(1).X_5(1).X_6(1)]}}$$

For example: for *n=100* the common ratio is approximately equal to 1,25 (*q=1,25)*.

7 COMBINATION OF SEMI-QUANTITATIVE AND QUANTITATIVE INPUTS

Results of the non-failure and risk analysis achieved from the statistic input data are always more exact and credible than the ones achieved by analyzing the estimated input data and therefore it would be counterproductive to use the scale evaluation instead of correct input data and consequently decrease the accuracy of the analysis. Suggested valuation and risk calculation methods afford the possibility to substitute any factor, for which the statistical data are available, by the exact real value.

8 CONCLUSION

Input data can be evaluated by using the method described before, but the evaluation regarding the certain output requirement can be performed as well. Deterministic approach could be used to evaluate and compare the consequences of undesired occurences disregarding the due probability or it is possible to separately evaluate the risk associated with the object reconstruction, the effects on the health and safety and also the risk of the third party associated with the other occurences, e. g. traffic diversion.

References

[1] BUBNÍK, J., KEDER, J., MACOUN, J., MAŇÁK. J.: SYMOS'97 *Systém modelování stacionárních zdrojů*. Metodická příručka pro výpočet znečištění ovzduší. Český hydrometeorologický ústav Praha, Praha : 1998.

[2] CROWL, D. A., LOUVAR, J. F. *Chemical process Safety: Fundamentals with Application, PTR Prentice* – Hall, Inc. A. Simon & Schuster Company, Englewood Cliffts, New Jersey 1990.

[3] DVOŘÁK, Z., 2006. *New aspects of transporting dangerous material and oversized shipments,* In: Transport 2006, XVIth scientific comference, Visše transportno učilišče "Todor Kableškov", 2006, p. I-39-I-42, ISBN 954-12-0130-X.

[4] KAMENICKÝ J., ZAJÍČEK J.: *Effectiveness optimization of RCM process in Risk*, Reliability

and Social Safety, Stavanger, 25.-27.6.2007, ISBN 978-0-415-44786-7.

[5] MDČR 1F44E/015/030 - *Dopravní infrastruktura jako kritický prvek národní infrastruktury z hlediska zabezpečení základních funkcí státu.*

[6] *Methods for the calculation of physical effects resulting from releases of hazardous materials* (liquids and gases), (Yellow Book). Third Edition, second print. Committee for the Prevention of Disasters (CPR), Directorate - General of Labour of the Ministry of Social Affairs. The Hague, 2005. CPR 14E.

[7] SOUŠEK, R. et.al.: *Doprava v krizových situacích.* Pardubice. Institut Jana Pernera, 2008. 248s. ISBN 80-86530-46-9.

MSc. Jaroslav ZAJÍČEK
Technical University of Liberec
Studentská 2
461 17 Liberec
Czech Republic
E-mail: jaroslav.zajicek@tul.cz

MSc. Jiří HAVLÍČEK
Technical University of Liberec
Liberec
Studentská 2
461 17 Liberec
Czech Republic
E-mail: jiri.havlicek@tul.cz

MSc. Pavel FUCHS, PhD.
Technical University of Liberec
Studentská 2
461 17 Liberec
Czech Republic
E-mail: pavel.fuchs@tul.cz

Assoc. Prof. Dipl. Eng. Radovan SOUŠEK, PhD.
University of Pardubice
Studentská 95
532 10 Pardubice
Czech Republic
E-mail: Radovan.Sousek@upce.cz

Brig. Gen. Assoc. Prof. Dipl. Eng. Miroslav KELEMEN, PhD.
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 01 Liptovský Mikuláš
Slovak Republic
E-mail: miroslav.kelemen@aos.sk

MSc. Lucia KOVÁČOVÁ, PhD.
University of Security Management in Košice
Kukučínova 17
040 01 Košice
Slovak Republic
E-mail: lucia.kovacova@vsbm.sk