AEROSPACE RESEARCH IN BULGARIA

Volume 21, 2007, Sofia
Space Research Institute
Bulgarian Academy of Sciences

Address
AEROSPACE RESEARCH IN BULGARIA
Space Research Institute
6, Moskovska St., Sofia 1000, Bulgaria

e-mail: office@space.bas.bg

Language Editor
Lubomira Kraleva

Technical Editor
Milen Zamphirov

Pre-Publication Processing
Tsveta Srebrova

ISSN 1313 - 0927

Bulgarian Academy of Sciences **ISSN 1313 – 0927**

Aerospace Research in Bulgaria

21
Sofia, 2007

C o n t e n t s

ÑÚÄÚÐÆÀÍÈÅ

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

THIN VISCOUS ELLIPTICAL ACCRETION DISCS WITH ORBITS SHARING A COMMON LONGITUDE OF PERIASTRON II. POLYNOMIAL SOLUTIONS TO THE DYNAMICAL EQUATION FOR INTEGER VALUES OF THE POWERS IN THE VISCOSITY LAW

Dimitar Dimitrov

Space Research Institute - Bulgarian Academy of Sciences

Abstract

For integer exponents n (n = -1, 0, 1, 2 and 3) in the viscosity law $\eta = \beta \Sigma^n$ with Σ – surface density of the accretion disc, we have investigated the polynomial approach to the solutions of the dynamical equation in the accretion disc model of Lyubarskij et al. [12]. Power series expansions of the eccentricity e(u), its derivatives $\dot{e}(u)$ and $\ddot{e}(u)$, their powers $e^2(u)$, $e^3(u)$,...; $\dot{e}^2(u)$, $\dot{e}^3$, etc. are truncated at appropriate values of the exponents and then substituted in the dynamical equation. Making additional truncations in the intermediate products and results, we have nevertheless achieved accuracy of the solution better than ~ 10% for large enough domains in the plane (e, $\dot{e}$). These results are established graphically by comparing the polynomial approximation of the eccentricity $e_{polynomial}$ (where $u \equiv \ln p$; p is the focal parameter of the ellipse) and the exact values $e_{exact}(u)$ derived by means of numerical solving of the equation. The solutions of the second order ordinary differential equation of motion are parameterized by means of the boundary conditions $e_0 \equiv e(u = 0)$ and $\dot{e}_0 \equiv \dot{e}(u = 0)$. The coefficients in the power series expansions are evaluated in explicit form, but because of their complexity we give only their "sizes", considered as lengths of the files which represent them. Problems referred to the possible singularities of the results are also discussed.

Introduction

Both analytical and numerical studies during the last two decades reveal many of the observational characteristics of the accretion discs around compact objects with high masses (black holes in the active galactic nuclei) and stellar masses (white dwarfs, neutron stars and black holes in binary stellar systems). The comparison between the theoretical models and the observations provides to extract information not only about the physical properties of these objects but also about the structure and physical conditions in the accretion discs themselves. In particular, it is evident that approximately half of the young stellar objects, recognized as binaries, are associated with geometrically thin and optically thick circular accretion discs. For such cases the geometry and physics of the accretion flows are considerably complicated in comparison with the identical phenomena around single compact bodies. Similar situations may arise when a planet system forms inside the disc and the protoplanets locally deplete the material along their orbits. The recent observational evidence about the existing of extrasolar planet systems have given rise to another aspect concerning the radial structure of the accretion discs. As a rule, the suspected extrasolar planets have orbits with high eccentricities. Although the eccentricity values and also the radii of the protoplanet orbits may change for different stages of the planet-forming evolution process [1], it is suggested that the elongated eccentric orbits arise because of the eccentric shapes of the progenitor accretion discs. For example, Marcy et al. [2] have investigated two G-type main-sequence stars HD 210 277 and HD 168 443, and have concluded the existence of companions orbiting around these stars and having masses comparable with the mass of Jupiter. Their orbits are with large eccentricities: $e = 0.45$ and $e = 0.54$, respectively. Such eccentric orbits may result from gravitational perturbations imposed by other orbiting planets or stars, by passing stars in a dense star cluster, or by the eccentric protoplanetary disc. While HD 168 443 exhibits a long-term velocity trend, consistent with an undetected yet (directly) close stellar companion, HD 210 277 appears to be a single star. Consequently, the later possibility is most likely the explanation for its large eccentricity. This picture also takes a support from the widespread explanation of the superhump events in the light-curves of SU UMa type stars, that explores the eccentric structure of the discs. It is natural that under such circumstances the interest to the theory of accretion discs with non-circular orbits of their

particles has increased during the recent years, because the explanation of the properties of such objects is a necessary condition for understanding the superhump events as a whole. The complexity of this problem may be illustrated by the comparison between the cataclysmic variables and low-mass X-ray binaries. As pointed out by Haswell et al. [3], in cataclysmic variables superhumps are believed to result from the presence of 3:1 orbital resonance in the accretion disc. Then the accretion disc becomes non-axisymmetric and precesses. The variations of luminosity in cataclysmic variables are caused by a tidally-driven modulation of the viscous dissipation into the disc, depending on the beat between the orbital and disc precession period. By contrast, in low-mass X-ray binaries the tidal dissipation in the outer parts of the accretion disc is unimportant because the optical emission is dominated by reprocessing of the X-rays emitted from the compact object. Consequently, in these two cases the superhump modulation is caused by two distinct mechanisms. Similarly, detailed hydrodynamic simulations of the superoutburst events in dwarf novae, including the full tidal potential of the binary system, are performed in the work of Truss et al. [4]. Their theoretical (using numerical methods) investigations of the mass flux through the disc, the growth rate of the superhumps and the disc eccentricity show that the superoutburst-superhump phenomenon is a direct result of tidal instability. Other studies of such events demonstrate that the stabilization of the superhump period at low values favours model, in which period changes arise strictly from eccentricity changes rather than mean radius changes in the disc [5]. This explains why decreasing period and decreasing amplitude are strongly linked in the superhumps of dwarf-nova.

Other observational studies of the dwarf nova WZ Sagittae [6] whose eclipses permit measuring the location and brightness of the mass-transfer hot spot imply that the disc must be very eccentric and nearly as large as the white dwarf's Roche lobe. Because the hot-spot luminosity exceeds its quiescent value by a factor of up to 60, this indicates that the enhanced mass transfer from the secondary plays a major role in the eruption, determining the geometrical shape of the accretion disc as well.

Accretion discs are expected to occur in a large number of various selestial objects, for example around protostars, accreting compact objects on stellar binary systems and also around supermassive black holes at the cores of galaxies. The importance of these phenomena for astrophysics

lies in the circumstance that they are related with local hydrodynamical or magnetohydrodynamical processes which can allow outward transport of the angular momentum of the infalling matter. So, a considerable part of the surrounding matter is able to reach the surface of the compact object, increasing its mass and changing its spin velocity. The significance of the above properties naturally explains why a large number of theoretical (both analytical and numerical) models of accreting discs have been developed during the last two decades and have been compared with the available observational astronomical data. A lot of theoretical works have revealed many of the subtle properties of the accretion discs in order to accommodate the models to these data. In particular, Ogilvie [7] stresses that an important and widely neglected aspect of the interaction between an accretion disc and a massive companion with a coplanar orbit is the vertical component of the tidal force. The response of the disc to the vertical forcing is resonant at certain radii, at which a localized torque is exerted and at these radii a compressive wave may be emitted. The $m = 2$ inner vertical resonance in a binary star is typically located within the tidal truncation radius of a circumstellar disc. This resonance contributes to angular momentum transport and produces a potentially observable non-axisymmetric structure. Larwood and Kalas [8] numerically investigate a close stellar fly-by encounter of a symmetrical circumstellar planetesimal disc and derive that this mechanism can give rise to the many kinds of asymmetries and substructures attributed to the edge-on dusty discs of β Pictoris. Their conclusions are supported by the optical coronographic observations of the outer perts of the disc of β Pictoris whose asymmetry was found to be approximately 25 %.

Eclipses in the binary stellar systems are often used techniques for obtaining numerical estimates of the perameters of these objects, including the characteristics of accretion discs when they are present in such binaries. For example, the use of the hot-spot eclipse times of the deeply eclipsing dwarf nova IY UMa enables to trace out the shape of its disc during the late superhump era. The result is an eccentric disc [9]. The analysis of the high-speed photometry of the dwarf nova EX Draconis through its outburst cycle reveals that the disc expands during the rise phase until it fills the most of the primary Roche lobe and one-armed spiral structure present in the disc at the stages of the outburst [10].

The above mentioned papers are only a small part of the numerous theoretical and observational evidence illustrating that the accretion disc may

have not only eccentric shape, but also a complicated internal structure like gaps and spiral density waves. This situation makes reasonable the studying of discs composed by particles moving on *elliptical* orbits around a compact gravity center. In the present work we continue an earlier investigation [11] of an accretion disc model developed by Lyubarskij et al. [12]. Their analysis of the accretion flow appears to some extent as a generalization of the standard α-disc accretion [13] to the case of non-circular (i.e., elliptical) orbits. Here, our goal is to obtain analytical solutions to the dynamical equation describing the accretion flow according to the model of Lyubarskij et al. [12] and to derive the domain where our results are valid. For the latter reason it is worthy to note also some of the limitations of the α-disc model of Shakura and Sunyaev [13]. The standard model of disc accretion assumes that the gravitational energy is locally efficiently radiated from both sides of the outer disc surface and the gas keeps its (nearly) *Keplerian* rotation because the interactions between the neighbouring radial annuli are neglected. However, there may exist an important process which leads to a structure different from that picture – namely, the advection. Physically, the advection process means that the generated energy via viscous dissipation is restored as entropy of the accreting gas rather than being radiated. As stressed in [14], the advection effect may be very important both for the cases of low and high rates, since radiation decreases efficiently under these circumstances. The angular velocity of the gas is much lower than the Keplerian, i.e. the sub-Keplerian velocity is one of the general properties of the advection-dominated flows. In the model of an optically-thick disc considered in [14] , the emission of blackbody radiation from the disc surface is so inefficient that the advection cooling dominates over surface cooling because of the high accretion rate that leads to photon trapping in the disc. While in the standard model of Shakura and Sunyaev [13] in the radiation-pressure dominated region thermal instability exists, in the advection-dominated case the accretion flow is thermally stable in the same range. The reason why the advection cooling stabilizes the radiation-pressure dominated region of the disc is that it plays two important roles: balancing and lowering the generated energy.

Another important property of the standard α-disc model is that turbulent stresses leading to outward angular momentum transport in accretion discs are treated as resulting from isotropic effective viscosity, related to the pressure through the α-parametrization of Shakura and Sunyaev [13]. This

simple approach may be adequate for the simplest aspects of accretion disc theory and was historically necessitated by an incomplete understanding of the origin of turbulence [15]. Recently Balbus and Hawley [16-19] have shown that the magnetorotational instability provides a mechanism of generating turbulent Reynolds and Maxwell stresses in sufficiently ionazed discs for which the α-viscosity model is not able to provide satisfactory description of many aspects of this process. The new generation of models taking into account these properties of the accretion flows should be particularly useful in understanding the dynamics of warped, eccentric and tidally distorted discs and also non-Keplerian flows (which are expected, for example, close to black holes).

In the above mentioned notes we have touched some of the unresolved problems of the standard α-disc accretion model. These deficiencies must be kept in mind when the results obtained in the next sections of the present paper are considered. Our solutions to the particular cases of the dynamical equation, governing the eccentric accretion flow, are in fact, solutions of a problem treating the accretion picture on the base of the Shakura-Sunyaev model for circular discs [13], extended to the case of eccentric orbits by Lyubarskij et al. [12]. All the restrictions concerning the applicability of such theories in reality (tested by means of observations) will be valid for our solutions even when they are mathematically exact in the considered domain of parameter space).

Numerical Solutions to the Dynamical Equation of the Accretion Flow

In a previous paper [11] we have obtained the explicit form of the dynamical equation valid for a *stationary* accretion, as considered by Lyubarskij et al. [12]. These results are obtained for particular values of the exponent n (namely, n = -1, 0, 1, 2, 3) in the accepted power-law relation $\eta = \beta \Sigma^n$ between the viscosity η and the surface density Σ of the eccentric accretion disc. Following the notations in [11], this equation can be written as a homogeneous second order ordinary differential equation:

$$(1) \qquad A(e, \dot{e}, n)\, \ddot{e} + B(e, \dot{e}, n)\, \dot{e} = 0 ,$$

where A(e, $\dot{e}$, n) and B(e, $\dot{e}$, n) are already known functions of e, $\dot{e}$ and n (their derivation is the main outcome of [11]) and the dot (.) denotes differentiation with respect to $u \equiv ln\, p$; p is the focal parameter of elliptical trajectories of the gas particles. In their paper Lyubarskij et al. [12] have numerically solved equation (1) for some particular values of n (see Figs. 2-5 from [12]) and have obtained three classes of solutions. In order to verify our analytical derivations we have also repeated the solution of differential equation (1) using a numerical method and have found agreement between our graphics and the graphics in [12]. We shall further use the exact results of the numerical integration of dynamical equation (1) as standards with respect to which we shall compare the validity of our analytical approximations to the solutions of (1). This will give us the opportunity to establish the domain where our analytical approach is successful and also to estimate the precision of the approximations. We shall test the most simple and the most suitable for the analytical applications fitting of the eccentricity dependence $e = e(u \equiv ln\, p,\ n)$ – the polynomial approximation.

Polynomial Approximation to the Solutions of the Dynamical Equation

We shall try to find a solution to equation (1) using the following power-law expansion for the unknown eccentricity function $e = e(u, n)$

$$(2) \qquad e(u, n) = \sum_{i=0}^{M} a_i(n)\, u^i ,$$

where the coefficients $a_i(n)$ (i = 0, ..., M) are unknown functions on the parameter n and are subject to further determination. Because our investigation of the problem of finding solutions to equation (10) is restricted only to five fixed integer values of n (n = -1, 0, 1, 2, 3), in what follows we shall omit the explicit notation of the dependence on n. The meaning of n will be clear from the heading of the considered case. Generally speaking, the power series (2) may contain an infinite number of terms (i. e., M = ∞). But in order to obtain practically effective and usable computational procedure, we must truncate series (2), assuming some finite value M (i. e., M < ∞) . In the present work we have confined ourselves to the value M = 5. This was done for computational reasons. When we attempted to determine higher order coefficients a_6, a_7, etc., the analytical expressions for these functions became so long and complicated that the available memory of the computer was not

enough in order to perform the analytical evaluation of these quantities. Such a circumstance must not be considered only as a technical problem, but also as an evidence that the analytical approach we have selected is not productive in view of high accuracy evaluation of the eccentricity $e(u)$ along the radius of the accretion disc. Consequently, the applied method for approximation of $e(u)$ by means of a polynomial is working effectively only when the truncation M = 5 is appropriate to give the desired precision. We stress that the main aim of our investigation is to find analytical expressions for $e(u)$ which are suitable for further analytical manipulations and which should not produce too complicated mathematical formulae as final results. Unfortunately, if the lower order coefficients a_2, a_3, a_4 and a_5 are tedious expressions, there is not any reasonable hope to expect that the application of the approximation (2) will ensure this optimistic outcome. So, we have to confine ourselves to the more particular case of simplifying of the task, namely, to use in the intermediate analytical calculations the suitable for analytical work polynomial representation (2), without introducing in it the explicit form of the coefficients a_i (i = 0, 1,..., M) and only arriving at the final results to do these replacements.

In the case n = 1, when dynamical equation (1) has relatively simple form (see eq. (14) from [11]) we have made comparison for two different cut-offs of series (2): for M= 5 and M = 8 (a better degree of approximation). The results show that we cannot establish an evident increase of accuracy when M = 8 is used instead of M = 5. Transferring this conclusion to the cases n = 0, 1, 2 and 3 (without an *explicit* proof !), we are challenged to believe that assuming the cut-off M = 5 is a reasonable compromise between the complexity of the coefficients a_i (i = 0, 1,..., M) and the accessibility of a higher accuracy of the approximation (2) by means of greater values of the cut-off M. Taking in advance the results in the next section of this chapter, we give the "size" of the coefficients a_i (i = 2,..., M_5(or M_8)) as evaluated by the occupied computer memory. Of course, this is a very rough measure of the complexity of these quantities, but their explicit formulae are too long and complicated to be given in this paper. For this reason, we prefer to prepare only a brief sketch of their length, which of course does not reveal their internal structure. Such a description of the coefficients a_i (i = 2,..., M_5 (or M_8).) may seem as a very fictive picture of their real mathematical properties. Nevertheless, the data clearly demonstrate the increasing complication of the computational procedure when higher order terms are taken into account into power series (2).

It also supports the unavoidable need to cut off this series at M = 5. Consequently, the derived domain of validity of approximation (2) can be hardly extended (for a value of the accuracy fixed in advance) of the solution $e = e(u, n)$) to a wider region in the parameter space of the initial conditions ($e_0 = e(u_0 \equiv ln\, p_0, n)$, $\dot{e}_0 = \dot{e}_0(u_0 \equiv ln\, p_0, n)$). We stress that we have chosen to investigate/derive the solutions to dynamical equation (1) by using polynomial approximation (2) for its simplicity in view of its application to analytical evaluations (easy differentiation and integration, simple procedures for finding roots and singularities, etc.). However, we *do not state* that there are not more complicated *exact* analytical solutions to the homogeneous second order ordinary differential equation (1). Our inability to find such solutions $e = e(u \equiv ln\, p, n)$, though they may be very awkward, does not mean that they cannot be found at all ! Of course, we shall confine ourselves not to the maximal purpose to obtain these solutions, but to put into use more complicated approximate expressions than (2), in order to expand the domain of validity of the approach. However, this eventually may not be the successful position. Out of the complicated analytical applications, the more exact solutions (in the sense of an extended domain of validity) do not ensure in advance a better coincidence with the observations. This is because of the inaccuracy of the physical description of the reality inherent to the model of Lyubarskij et al. [12], as discussed earlier. Therefore, the use of more precise approximations than (2) does not guarantee that the higher price to be paid for that is an acceptable decision. We only mention in that sense, that we have tried to transform the most simple of the considered dynamical equations (namely, eq. (14) from [11]; Case $n = -1$) substituting $e(u, n = -1) = cos[\psi(u, n = -1)]$, but the introduction of the angular coordinate $\psi(u, n)$ did not produce the expected result – the equation was not reduced to a form supposing an easier to find exact solution.

An important note should be made. The general solution of differential equation (1) depends on two integration constants which are subject to determination from the initial and boundary conditions. The considered problem deals with a *stationary* accretion disc, so it is sufficient to give the values of two physical characteristics of the disc at a given fixed value of the focal parameter p_0 (respectively, $u_0 \equiv ln\, p_0$). The most natural choice seems to select the eccentricity e and its derivative $\dot{e} \equiv \partial e / \partial u \equiv \partial (ln\, e) / \partial p$. In terms of the polynomial approximation used here, the later variable may be written as

(3) $$\dot{e}(u, n) = \sum_{i=0}^{M-1} (i + 1)\, a_{i+1}\, u^i \;;\quad (n = -1, 0, 1, 2, 3).$$

In what follows in this paper, we choose the value of the focal parameter at which we give the boundary conditions, to be equal to 1. Respectively, $u_0 \equiv ln\, p_0 = 0$. Because $p = b^2/a$, this means that we set the boundary conditions on an ellipse with major and minor semiaxes a and b, respectively, satisfying the relation $b = a^{1/2}$. Then, omitting the notation of n, from (2) and (3) we obtain the simple relations $e_0 \equiv e(u_0) = a_0$ and $\dot{e}_0 \equiv \dot{e}(u_0) = a_1$. In other words, using the above gauge, we in fact assign values to the coefficients a_0 and a_1. Consequently, these quantities must be considered as independent *input* parameters of the problem and all other variables are functions of them. In particular, the higher order coefficients a_i (i = 2,..., M_5 (or M_8)) also depend on $a_0 = e_0$ and $a_1 = \dot{e}_0$ (and on the exponent n in the viscosity law $\eta = \beta \Sigma^n$, of course). For each fixed value n = -1, 0, ..., 3, we have derived a_i (i = 2,..., M_5 (or M_8)) in an explicit form as functions on a_0 and a_1. And then the lengths of these expressions were approximately evaluated as shown by the data on Table 1.

Table1. Approximate evaluations of the "size" of the coefficients a_i (e_0, $\dot{e}_0$, n).

Powers / Coefficients	n = - 1	n = 0	n = + 1	n = + 2	n = + 3
a_2	3.22 kb	8.68 kb	14.3 kb	6.13 kb	8.16 kb
a_3	3.83 kb	46.4 kb	111 kb	28.7 kb	40.2 kb
a_4	6.38 kb	143 kb	369 kb	76.7 kb	127 kb
a_5	9.04 kb	2.31 Mb	856 kb	168 kb	281 kb
a_6	12.07 kb				
a_7	19.0 kb				
a_8	28.7 kb				

Selecting the polynomial approximation (2) and the boundary conditions $e_0 \equiv e(u_0 = 0)$ and $\dot{e}_0 \equiv \dot{e}(u_0 = 0)$, we can also compute easily the second derivative of the eccentricity:

$$(4) \qquad \dot{e}(u) = \sum_{i=0}^{M-2} (i+2)(i+1)\, a_{i+2}\, u^i \ ,$$

the difference $e(u) - \dot{e}(u)$ between the eccentricity $e(u)$ and its derivative $\dot{e}(u)$:

$$(5) \qquad e(u) - \dot{e}(u) = \sum_{i=0}^{M-1} [a_i - (i+1)\, a_{i+1}]\, u^i \ ,$$

the powers of the eccentricity $e(u)$ and $\dot{e}(u)$:

$$(6) \qquad e^2(u) = \sum_{i=0}^{M} \Big(\sum_{k=0}^{i} a_k\, a_{i-k} \Big)\, u^i \ ,$$

$$(7) \qquad e^3(u) = \sum_{i=0}^{M} \Big(\sum_{k=0}^{i} a_k \sum_{p=0}^{i-k} a_p\, a_{i-k-p} \Big)\, u^i \ ,$$

$$(8) \qquad e^4(u) = \sum_{i=0}^{M} \Big[\sum_{l=0}^{i} \Big(\sum_{k=0}^{l} a_k\, a_{l-k} \Big) \Big(\sum_{k=0}^{i-l} a_k\, a_{i-l-k} \Big) \Big]\, u^i \ , \text{ etc.} \ ,$$

$$(9) \qquad \dot{e}^2(u) = \sum_{i=0}^{M} \Big[\sum (k+1)(i-k+1)\, a_{k+1}\, a_{i-k+1} \Big]\, u^i \ ,$$

$$(10) \quad \dot{e}^3(u) = \sum_{i=0}^{M} \Big\{ \sum_{k=0}^{i} \Big[\sum_{l=0}^{k} (l+1)(k-l+1)\, a_{l+1}\, a_{k-l+1} \Big]\, (i-k+1)\, a_{i-k+1} \Big\}\, u^i, \text{ etc.}$$

Obviously, the higher powers of the eccentricity $e(u)$ and its derivative $\dot{e}(u)$ become more and more bulky, but nevertheless they may be computed in an explicit manner. We shall not, of course, give them here. But we shall write a formula, also used in our computations. According to [20] (see expression 5.3):

(11) $(1 - e^2)^{1/2} = \sum_{i=0} d_i e^{2i}$; $d_0 = 1$, $d_1 = - ½$;

$d_i = - [(2i - 3)!!] / (2i)!!$ for $i = 2, 3, \ldots$;

(this means that $d_2 = - 1/8$, $d_3 = - 1/16$, $d_4 = - 5/128$, etc.).

The above formula is also applied for expansions in powers of the expressions $(1 - \dot{e}^2)^{1/2}$ and $[1 - (e - \dot{e})^2]^{1/2}$, replacing $e(u)$ by $\dot{e}(u)$ or by $[e(u) - \dot{e}(u)]$, respectively. Combining the expansions like written above, we have also computed the power series expansions of $(1 - e^2)^{3/2}$, $(1 - e^2)^{5/2}$, ...; $(1 - \dot{e}^2)^{3/2}$, $(1 - \dot{e}^2)^{5/2}$, ... ; $[1 - (e - \dot{e})^2]^{3/2}$, $[1 - (e - \dot{e})^2]^{5/2}$, ... As a final result, we have obtained the necessary power series expansions of the coefficients $A(e, \dot{e}, n)$ and $B(e, \dot{e}, n)$ for every fixed value of $n = -1, 0, 1, 2$ and 3. Substituting these into dynamical equation (1) and reducing after some algebraic manipulations its left-hand side to a series in powers of u, we transform (1) to the following form:

(12) $\sum_{i=0}^{M} c_i(e_0, \dot{e}_0, a_2, a_3, \ldots, a_M, n)\, u^i = 0$, $(n = - 1, 0, 1, 2, 3 ; M = 5 \text{ or } 8)$.

This nullification must be fulfilled for arbitrary values of u and, consequently, all the coefficients $c_i(e_0, \dot{e}_0, a_2, a_3, \ldots, a_M, n)$ also must be equal to zero:

(13) $c_i(e_0, \dot{e}_0, a_2, a_3, \ldots, a_M, n) = 0$, $(n = - 1, 0, 1, 2, 3 ; M = 5 \text{ or } 8)$.

It should be mentioned that multiplying different kinds of expressions like these given by formulae (2) – (11), we shall obtain terms including powers of u greater than M. Such circumstance leads to undesirable complications of the intermediate calculations of the left-hand side of (12). For example, we may get terms proportional to u^{15}, u^{16}, u^{17}, etc. To avoid this objectionable situation, we cut off the powers of u which are much greater than M (M = 5 or 8). Of course, these manipulations of the intermediate analytical expressions must be done carefully in order to avoid in the final sum erroneous truncations of terms proportional to u^i ($i \leq$ M). In that sense, we insured ourselves by formally preserving the terms in (12) up to order u^{M+3}, which do not create too much problems with respect to the complexity of the formulae.

Deriving explicitly the coefficients $c_i(e_0, \dot{e}_0, a_2, a_3, \ldots, a_M, n)$, it turns out that for each fixed value of n = -1, 0, 1, 2 and 3, relations (13) give linear equations for a_i (i = 2, 3,..., M = 5 or 8) provided that the low order coefficients a_i are already known. That is to say, equalities (13) are in fact recurrence formulae for a_i (i = 2, 3,..., M = 5 or 8). In more details, we have that the free term of (11) depends only on e_0, $\dot{e}_0$, a_2 and n, but not on a_3, a_4, etc. Equating it to zero

(14) $\quad c_0(e_0, \dot{e}_0, a_2, n) = 0$,

we obtain *a linear* equation for the unknown quantity a_2 , which is easily solved. Further we see that the coefficient c_1 depends only on e_0, $\dot{e}_0$, a_2, a_3 and n. Similarly, it turns out that its nullification

(15) $\quad c_1(e_0, \dot{e}_0, a_2, a_3, n) = 0$

appears as *a linear* equation for the unknown quantity a_3 under the condition that the value of a_2 is already computed from (14). This sequence of operations can be extended, because

$c_2 \equiv c_2(e_0, \dot{e}_0, a_2, a_3, a_4, n) = 0$, $c_3 \equiv c_3(e_0, \dot{e}_0, a_2, a_3, a_4, a_5, n) = 0$, etc., and these relations are *linear* equations for a_4, a_5, etc. realizing all these steps, we obtain finally the polynomial approximation (2) for the eccentricity $e(u)$ of the orbits of the accretion disc particles. We again stress that (2) is only *a fitting* to the exact solution of the dynamical equation (1), regardless of whether the latter is obtained by numerical or (may be ?) analytical methods.

It remains to check the precision of our approach, comparing the graphics of the solutions of type (2) with the results of the numerical integration of (2). It is clearly seen that in the parameter space $(e, \dot{e})$ (for every fixed value of n) there are regions where the polynomial approximation (2) gives an excellent agreement with the exact solutions of equation (1). There the difference $e(u)_{\text{exact}} - e(u)_{\text{polynomial}}$ may be less than 10^{-6} or even better!). But there are also domains where it is fully unacceptable. The transition between these two regions is, however, too steep as a rule. Consequently, it is very important to determine precisely the boundaries of the validity domain of the tested (in this paper) truncated power-low series approximation (2). We remind that the eccentricity $e(u)$ and its derivative $\dot{e}(u)$ must satisfy the following restrictions for all u: $|e(u)| < 1$, $|\dot{e}(u)| < 1$ and $|e(u) - \dot{e}(u)| < 1$ [11], which in turn determine the shape of the overall domain, where we are seeking solutions of dynamical equation (1). The above equalities will be discussed in more details in a forthcoming paper. We also note that (for our illustrative purposes) we have chosen to reject all the polynomial solutions at the level where

$|e(u)_{exact} - e(u)_{polynomial}| \leq 1$, i.e., we accept that satisfactory accuracy is achieved when the deviation of the tested analytical approximation (2) from the exact (numerically obtained) solution is better than 0.1. It is evident that the domain where polynomial approximation (2) turns out well has a complex shape, which we shall not try to evaluate analytically. But nevertheless, this domain is large enough to say that approach (2) makes sense and the considered problem is solved at least particularly.

Singularities of the Power - Law Expansion Coefficients a_i

Evaluating the applicability of truncated (M = 5 or M = 8) power series (2) relative to the possibility to solve the equation governing the stationary accretion flow in the model of Lyubarskij et al.[12], we have to study in more details the behaviour of coefficients a_i (i = 2, 3,..., M) in the parameter space $(e, \dot{e})$ for each n = -1, 0, 1, 2 and 3 in separate. Except for the case n = -1, the explicit expressions for these quantities (functions on e and $\dot{e}$) are so long and complicated, as discussed earlier, that we are forced to apply graphical argumentation rather than purely analytical one. Our aim is not to investigate these coefficients as a whole, but only their denominators. The nullifications of the latters dangerous sources, generating divergences of series (2). Fortunately, this problem is essentially simplified by the property that the denominators of a_3, a_4, a_5,..., a_8 are integer powers of the denominator of a_2 within to a non-zero factor. This will be explicitly shown in the following example, illustrating the case n = -1. Therefore, it is enough to concentrate our efforts on finding the roots of the denominator of a_2.

The case n = -1 is the most simple situation. For n = -1 the dynamical equation is relatively simple in comparison with the other cases ([11], see eq. (14)]:

(16) denominator $a_2 = A(n = -1)$,

where $A(n = -1) = (1 - e^2)(144 - 80e^2 - 16e^4 - 8e^6 - 5e^8)$. This expression does not depend on $\dot{e}(u)$. Provided that $|e(u)| < 1$ for all u, the roots of the denominator a_2 (including their multiplicity) are:

$$(17)\quad \begin{cases} e_{1,2} = -e_{3,4} = -1.209\,912 \pm 1.138\,902\, i\ ; \\ e_{5,6} = \pm 1.097\,309\ ;\quad e_{7,8} = \pm 1\ ; \\ e_{9,10} = -4.422 \times 10^{-17} \pm 1.771\,346\, i\ , \end{cases}$$

where i is the imaginary unit. Obviously, there are not troubles generated

by the nullification of the denominator of a_2, because all the roots are complex or greater than (or equal to) 1 by absolute value. The higher order coefficients have denominators as follows:

(18) denominator $a_3 = 96 \times [A(n = -1)]^2$,

(19) denominator $a_4 = 1536 \times [A(n = -1)]^3$,

(20) denominator $a_5 = 30720 \times [A(n = -1]^4$, etc.

In *all* the coefficients a_i (i = 2, 3,..., 8), the denominators are multiple to the powers of $A(n = -1)$ and the other multipliers in each factorization are not equal to zero. In the other cases $n = 0$, 1, 2 and 3, the situation is not so *analytically* clear because of the complexity of the coefficients a_i (i = 2, 3,..., 8), but the *graphical* representations of their denominators indicate that there is no evidence of nullification of the latters (*at least*) for extended domains in the plane $(e, \dot{e})$. Of course, the domains of validity of polynomial approximation (2) must agree with the restrictions $|e(u)| < 1$, $|\dot{e}(u)| < 1$ and $|e(u) - \dot{e}(u)| < 1$ [11], as mentioned earlier.

References

1. Del Popolo, A., K. Y. Ekşi. Migration of giant planets in a time-dependent planetesimal accretion disc., Monthly Not. Royal Astron. Soc., 332, 2002, № 2, p. 485.
2. Marcy, G. W., R. P. Butler, S. S. Vogt, D. Fischer, M. C. Liu. Two New Candidate Planets in Eccentric Orbits. Astrophys. J., 520,1999, № 1, pt.1, p. 239.
3. Haswell, C. A., A. R. King, J. R. Murray, P. A. Charles. Superhumps in low-mass X-ray binaries., Monthly Not. Royal Astron. Soc., 321, 2001, № 3, p. 475.
4. Truss, M. R., J. R. Murray, G. A. Wynn. On the nature of superoutbursts in dwarf novae., Monthly Not. Royal Astron. Soc., 324, 2001, № 1, p. L1.
5. Patterson, J., J. kemp, A. Shambrook et al. Superhumps in cataclysmic binaries. XII. CR Bootis, a helium dwarf nova., Publ. Astron. Soc. Pacific, 109, 1997, № 10, p. 1100.
6. Patterson, J., G. Masi, M. Richmond et al. The 2001 superoutburst of WZ Sagittae., Publ. Astron. Soc. Pacific, 114, 2002, № 6, p. 721.
7. Ogilvie, G. I. A non-linear theory of vertical resonances in accretion discs., Monthly Not. Royal Astron. Soc., 331, 2002, № 4, p. 1053.
8. Larwood, J. D., P.G. Kalas. Close stellar encounters with planetesimal discs: the dynamics of asymmetry in the β Pictoris system., Monthly Not. Royal Astron. Soc., 323, 2001, № 2, p. 402.
9. Rolfe, D. J., C. A. Haswell, J. Patterson. Late superhumps and the stream-disc impact in IY UMa., Monthly Not. Royal Astron. Soc., 324, 2001, № 3, p. 529.
10. Baptista, R., M. S. Catalán. Changes in the structure of the accretion disc of EX Draconis through the outburst cycle., Monthly Not. Royal Astron. Soc., 324, 2001, № 3, p. 599.

11. D i m i t r o v D. Thin viscous elliptical accretion discs with orbits sharing a common longitude of periastron. I. Dynamical equation for integer values of the powers in the viscosity law., Aerospace Research in Bulgaria, 19, 2006.
12. L y u b a r s k i j, Yu. E., K. A. P o s t n o v, M. E. P r o k h o r o v. Eccentric accretion discs., Monthly Not. Royal Astron. Soc., 266, 1994, № 2 , p. 583.
13. S u n y a e v, R. A., N. I. S h a k u r a. Black holes in binary systems. Observational appearance., Astron. & Astrophys., 24, 1973, p. 337.
14. W a n g, J.- M., J.- J. Z h o u. Self-similar solution of optically thick advection-dominated Flows., Astrophys. J., 516, 1999, № 1, pt.1, p. 420.
15. O g i l v i e, G. I. On the dynamics of magnetorotational turbulent stresses., Monthly Not. Royal Astron. Soc., 340, 2003, № 3, p. 969.
16. H a w l e y, J. F., S. B a l b u s, W. W I n t e r s. Local hydrodynamic stability of accretion Disks., Astrophys. J., 518, 1999, № 1, pt.1, p. 394.
17. B a l b u s, S. A., M. R i c o t t i. On nonshearing magnetic configurations in differentially rotating disks., Astrophys. J., 518, 1999, № 2, pt.1, p. 784.
18. B a l b u s, S. A., J. C. B. P a p a l o i z o u. On the dynamical foundations of α-disks., Astrophys. J., 521, 1999, № 2, pt.1, p. 650.
19. H a w l e y, J. F., J. M. S t o n e. Nonlinear evolution of the magnetorotational instability in ion-neutral disks., Astrophys. J., 501, 1998, № 2, pt.1, p. 758.
20. D w i g h t, G. B. Tables of Integrals and Other Mathematical Data. New York, Mc Millan Company, 1961.

ТЪНКИ ВИСКОЗНИ ЕЛИПТИЧНИ АКРЕЦИОННИ ДИСКОВЕ С II. ПОЛИНОМНИ РЕШЕНИЯ НА ДИНАМИЧНОТО УРАВНЕНИЕ ЗА ЦЕЛОЧИСЛЕНИ СТОЙНОСТИ НА СТЕПЕНИТЕ В ЗАКОНА ЗА ВИСКОЗИТЕТА

Д. Димитров

Резюме

За целочислени степенни показатели n (n = -1, 0, 1, 2 и 3) в закона за вискозитета $\eta = \beta \Sigma^n$, където Σ е повърхностната плътност на акреционния диск, ние сме изследвали полиномния подход за намиране

решенията на динамичното уравнение в модела на акреционен диск на Любарски и др.[12]. Разложенията в степенни редове на ексцентрицитета $e(u)$ и неговите производни $\dot{e}(u)$ и $\ddot{e}(u)$, техните степени $e^2(u)$, $e^3(u)$,...; $\dot{e}^2(u)$, $\dot{e}^3(u)$ и т.н. са орязани при подходящи стойности на степенните показатели и след това те са поставени в динамичното уравнение. Извършвайки допълнителни орязвания в промеждутъчните произведения и резултати, въпреки това ние сме достигнали точност на решението по-добра от ~ 10% за една достатъчно голяма област в равнината $(e, \dot{e})$.

Тези резултати са установени графически чрез сравняване на полиномната апроксимация на ексцентрицитета $e_{\text{polynomial}}(u)$ (където $u \equiv \ln p$, p е фокалният параметър на елипсата) и точните стойности $e_{\text{exact}}(u)$ получени посредством числено решаване на уравнението. Решенията на уравнението на движение, което е обикновено диференциално уравнение от втори ред, са параметризирани с помощта на граничните условия $e_0 \equiv e(u=0)$ и $\dot{e}_0 \equiv \dot{e}(u=0)$. Коефициентите в разложенията в степенни редове са оценени в явен вид, но поради тяхната сложност ние даваме само техните "размери", разглеждани като дължини на файловете които ги представят. Дискутирани са също проблемите отнасящи се до възможните сингулярности на резултатите.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

INTEGRATION OF THE EQUATION OF THE ARTIFICIAL EARTH'S SATELLITES MOTION WITH SELECTION OF RUNGE-KUTTA-FELBERG SCHEMES OF OPTIMUM PRECISION ORDER

Atanas Atanassov

Solar Terrestrial Influences Laboratory - Bulgarian Academy of Sciences
e-mail:at_m_atanassov@yahoo.com

Abstract

An approach is considered for numerical integration of ordinary differential equations systems of the first order with choice of a computation scheme, ensuring the required local precision. The consideration is made on the basis of schemes of the Runge-Kutta-Felberg type. Criteria are proposed as well as a method for the realization of the choice of an "optimum" scheme. The effectiveness of the presented approach to problems in the field of satellite dynamics is illustrated by results from a numerical experiment. These results refer to a case when a satisfactory global stability of the solution for all treated cases is available. The effectiveness has been evaluated as good, especially when solving multi-variable problems in the sphere of simulation modelling.

Key words: satellite orbit propagation, numerical simulation, Runge Kutta Felberg codes

Introduction

The practice of computing often demands the solution of ordinary differential equation systems (ODES) of the first order:

(1) $$\frac{dy_i}{dx} = \varphi_i(x, y_i)$$

with initial conditions $y_i(x_0) = y_i^0$. . The type of the right side of (1) determines the problem as stiff or non-stiff.

Different groups of methods exist for numerical solution of (1) in cases of non-stiff problems. First, there are ones that are with general character, applicable to a wide class of problems and there are others, which are specialized, oriented to specific problems. The latter have certain advantages, yet they require development for each specific case. An example of the second type of methods is the so-called recurrent power series [1]. The methods with more common application are divided mainly into one-step and multiple-step ones [2].

As far as the order of the equations systems of higher order is subject to lowering, the methods for first-order systems are always applicable. There are also methods, which are applied without preliminary order lowering. As a rule they are more efficient in cases when they can be applied [3, 4] but Nystrom's schemes [5] (which are the most popular) have a more restricted area of application since they require the right part to be independent on the first derivative of the dependent variable y.

In order to minimize the error and the computational expenses for each integration step, used to make the computations, major importance is attributed to the step size for numerical integration and the order of the computational schemes. Approaches exist for determination of optimum step size, which are possible both with the one-step and the multi-step methods. The change of the step in the one-step methods is easier while the multi-step methods require re-computation of the function derivative values in new points and produce complications [6]. The possibility to change the order (line) of the integration method in some multi-step methods [2, 7] attracts attention.

Two basic final approaches exist in the numerical integration of ODES. The first one is connected with integration through optimal step selection. The optimal step size is determined on the basis of error assessment; in general, at smaller error the step increases and vice versa. Optimization of the computational expenses is achieved along with ensuring global stability of the numerical solution at the end of the integration interval.

The second approach requires finding the solution in equidistant values by the independent variable. The integration by a constant step, however, does not always meet the requirement for sufficient local error, connected with the type of the functions on the right side of the equations, hence it can influence the solution stability.

There exists, however, a possibility during integration with optimal step to obtain the solution in desired points on the independent variable on the basis of interpolation. In addition, special methods exist, combining numerical integration and interpolation, with which the solution is obtained in arbitrary points in a natural way with increased computational efficiency [8,9].

Different methods and programs-integrators of common differential equations are developed and their efficiency has been examined. Because of their large number we will mention only a few, having in common with the evident one-step methods of the Runge-Kutta type [10,11,12,13]. Although the efficiency estimations poin to some advantages in behalf of one or another method and computer programs, when solving test problems there is not any certainty as for which method is the most suitable one.

A number of methodological groups exist for numerical solution of (1). The one-step methods of the Runge-Kutta type [1] are characterised by adaptability and easy programme realization. Different computing schemes, corresponding to those methods are known. The advantages of the schemes of the Runge-Kutta-Felberg (RKF) type [2,3] are due to the fact that with minimum additional computations, two solutions are simultaneously obtained with different precision:

$$(2) \qquad y_i = y_i^0 + h\sum_{k=0}^{n} c_k g_k^i + O(h^{p+1}),\ \overline{y_i} = y_i^0 + h\sum_{k=0}^{n+1} \overline{c_k}\, g_k^i + O(h^{p+2})$$

where $$g_0^i = \varphi_i^0(x_0, y_i^0),\ g_k^i = \varphi_i\left(x_0 + a_k.h, y_i^0 + h\sum_{k=0}^{k-1} b_{kj} g_j\right).$$

The difference between the two solutions $\overline{y_i} - y_i$ gives the exact value of the main member of the local error $\overline{O}(h^{p+1})$ for a scheme of the p^{-th} order, on the basis of which the local error can be estimated. The classical RKF schemes were followed by later schemes with enhanced efficiency as well as by

a higher order of the solution precision [16,17,18,19]. In cases of computations made on the basis of a scheme of the p^{th} order, by changing the integration step in definite limits, the necessary local precision is obtained and hence, a certain stability of the solution.

Formulation of the Problem

In the integration region along the independent variable x, the local error is a variable quantity and depends on step h, on the p^{th} order of the integration scheme and on the type of the functions on the right side of (1). Minimization of the local error by means of stepsize control is not the only possible one. Actually, the minimization of the local error by stepsize control is not always suitable. Instead, we can use an integration scheme of the lowest possible order, which provides the necessary local precision in integration with constant step. In this way the computations can turn out to be considerably less than if a scheme is used of an order, providing precision for the entire integration region along an independent variable. The issue for selection of optimal order of the integration schemes is dated far back [20, 21]. This is possible even more in case when parallel integration of several ODES is necessary when the solutions have different character and are obtained with different local precision.

In the classical one-step methods of the Runge-Kutta type, the choice of a scheme with sufficient local precision is a problem, connected with the possibility for evaluation of the local error. In this aspect, the methods of the Runge-Kutta-Felberg type have advantages, which allow to choose a strategy for selection of the optimum integration scheme [22]. RKF schemes are known of 1/2, 2/3, 3/4, 4/5, 5/6 and 7/8 order [14,15]. In the schemes of 1/2 up to 4/5 order, the functions on the right side of (1) are computed in 3 to 6 intermediate points, respectively, and for the schemes of 5/6 and 7/8 order - in 8 and 13, respectively. Special methods are known for integration of ODES with variable order of precision [ ] but instead we shall examine the possibility to use the RKF schemes.

Let us consider the possibility to investigate the efficiency of selection of an optimum scheme, connected with the integration of the equation of the artificial Earth's satellites motion (AES). The most common form of this equation is [5]:

$$(3) \qquad m\frac{d^2\vec{r}}{dt^2} = -G\frac{mM}{r^3}\vec{r} + \vec{f}$$

In (3) $\vec{r}$ is the satellite radius vector, m- its mass, M- the Earth mass, G- the gravitation constant, $\vec{f}$ - the disturbing forces, which determine the motion model and t- the time. The vector equation (3) is usually solved by decreasing the order to a system of two equations:

$$(4) \qquad \begin{aligned} m\frac{d\vec{V}}{dt} &= -G\frac{mM}{r^3}\vec{r} + \vec{f} \\ \frac{d\vec{r}}{dt} &= \vec{V} \end{aligned}$$

In (4) $\vec{V}$ is the velocity vector. The conditions, which specify the numerical integration local error of (4) at a constant step Δt and a scheme of the p^{th} order are different for the separate sections of the trajectory.

A Strategy for Choosing the Integration Scheme

The choice of an integration scheme is not aimed by far at providing a minimum local error, but just a sufficient local error with which the numerical solution will be steady within a given interval, with minimum computation expenses.

In [22,24] we examined the following strategy of choosing an optimum integration scheme for numerical integration based on the following cases:

a) For some functions φ_i the local error $\overline{O^i}$ has a minimum value, bigger than the necessary one; a scheme of a higher order is now used and the computations are repeated.

b) Smaller errors $\overline{O^i} < S^i \cdot O^i_{max}$ are obtained for all functions where S^i are rough estimations of the relations $O^i\left(h^{p+2}\right)/O^i\left(h^{p+1}\right)$; In this case it is reasonable to switch to a scheme of a lower order.

Here it should be kept in mind that the estimation of S^i is an essential problem. The experience shows that the above-stated contradicting requirements can be verified more precisely in the specific case of solving (4). This is made when instead of the local errors $\overline{O^i}$, i=1,..., 6, an evaluation of error $\overline{O} = \sqrt{\overline{O_x}^2 + \overline{O_y}^2 + \overline{O_z}^2}$ is made. Besides, when proceeding with an integration scheme of a higher order, the relative error value should be preserved. Later on, this value can be used as a lower limit, the passing of which should be accompanied by selection of a scheme of a lower order. This solves the problem with the difficulties in estimating. The strategy of choosing a scheme can be presented, as follows:

a') $\overline{O} > \varepsilon.\Delta r$, where $\Delta r = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$, ε is the value of the relative error selected with a view to attaining a global stability of the solution; then it is proceeded with the selection of a scheme of a higher order and when reaching a suitable scheme, the value of $\overline{O}$ is included in $\hat{O}$.

b') $\overline{O} < \hat{O}$ - proceeding with a scheme of a lower order.

The value of $\hat{O}$ is determined on the basis of the product $c_p.\overline{O}$ in which the constants c_p are dtermined empirically and the following values are accepted as suitable: c_2=.05, c_4=.09, c_5=.1, c_7=.2.

Evaluation of the efficiency

The advantage of selecting an optimum scheme for numerical integration of ODES, treated from the computing expenses point of view, is different in each separate case. The stability of the RKF schemes is investigated on the basis of the model equations [3] with regulation of the step length.

It is of major importance to know how the use of schemes of a lower precision order influences the global stability of the solution. Numerical experiments have been carried out for integration of six satellite orbits with main semi-axes and eccentricities, given on Table 1. The RKF scheme of 7/8 order is accepted as a guarantee of the numerical solution's stability with selection of a constant step within definite limits. The efficiency is estimated on the basis of the number of computations on the right side of (4) by the formula

$Q = \frac{n_7 - n_\sim}{n_7}$ where n_7 refers to the guaranteed scheme, and $n_\sim$ - to the case of choosing an optimum scheme.

c') $\bar{O} < \tilde{O}$ - proceeding with a scheme of a lower order.

The value of $\tilde{O}$ is determined on the basis of the product $c_p.\bar{O}$ in which the constants c_p are determined empirically defined and the following values are accepted as suitable: c_2=.05, c_4=.09, c_5=.1, c_7=.2 .

Table 1.

Orbit number	Eccentricity	Main semi-axis [m]
1	.01	8 000 000
2	.02	10 000 000
3	.04	20 000 000
4	.05	30 000 000
5	.07	42 241 085
6	.08	50 000 000

Evaluation of the efficiency

The advantage of selecting an optimum scheme for numerical integration of ODES, treated from the computing expenses point of view, is different in each separate case. The stability of the RKF schemes is investigated on the basis of the model equations [3] with regulation of the step length.

It is of major importance to know how the use of schemes of a lower precision order influences the global stability of the solution. Numerical experiments have been carried out for integration of six satellite orbits with main semi-axes and eccentricities, given on Table 1. The RKF scheme of 7/8 order is accepted as a guarantee of the numerical solution stability with selection of a constant step within definite limits. The efficiency is estimated on the basis of the number of computations on the right side of (4) by the formula

$Q = \frac{n_7 - n_\sim}{n_7}$ where n_7 refers to the guaranteed scheme, and $n_\sim$ - to the case of choosing an optimum scheme.

Fig. 1 presents the efficiency of the computations for each separate orbit at different integration steps as well as the total efficiency. The stability of the numerical solution is estimated on the basis of equation of type (4), which includes different disturbing factors, related to the Earth gravitation field. Fig. 2 presents the differences, obtained in the selection of optimum integration scheme, towards the use of a scheme of the maximum order. The temporal integration interval is one day and night. The computations are made with requirements for local relative error $\varepsilon \leq 2.10^{-7}$. The estimation of the results allows to draw up the following basic conclusions:

- the choice of a scheme, based on error estimation with the RKF methods can result in significant economy of machine time;

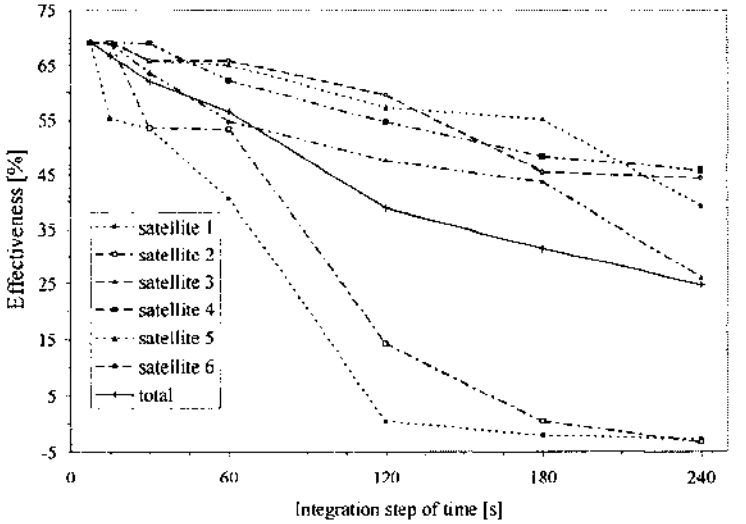


Fig. 1. Average efficiency for one day and night with numerical integration of the equation for artificial earth satellites motion with selection of an optimum scheme

- by decreasing the integration step, the global stability of the solution decreases with using schemes of lower precision order, but it is completely satisfactory for a wide range of problems; when necessary to achieve a better stability for a definite class of orbits, it is necessary to point out a smaller local error;
- when multi-parameter computations are made in the field of immitation modelling, the experimenter can trust the proposed strategy for selection of computation schemes.

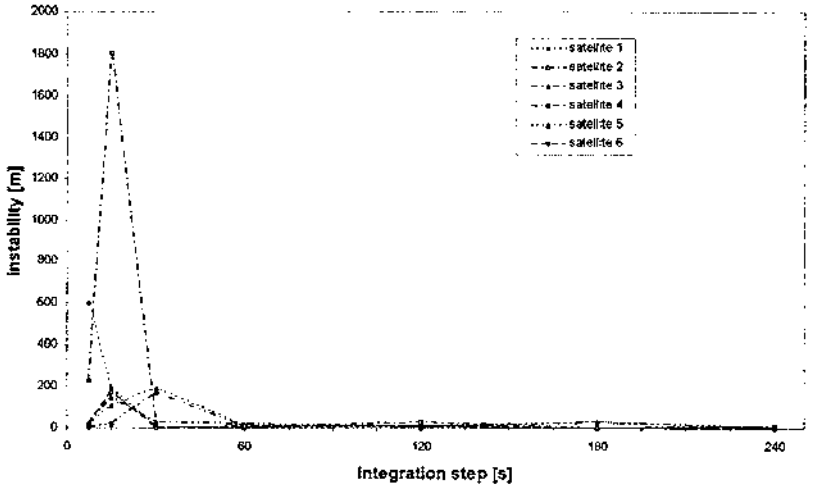


Fig. 2. The difference between the solution, obtained with selection of an optimum scheme and with a scheme of the 7th order, obtained for one day and night at

The numerical experiments were carried out on the basis of programmes using the Fortran programme language [24]. For this purpose an integrator of systems of independent vector differential equations has been developed. It allows the integration of every system to be performed independently on the integration of the remaining ones, with individual choice of a method of appropriate order.

The examined approach for integration of independent differential equations systems is applicable for solution of problems with many objects. The enhancement of the computational efficiency is especially beneficial for problems, whose solution should be made in real time. The examined approach provides additional opportunities, connected with computational parallelization in multi-processor systems as well as in network computations.

References

1. Hadjifotinou K. G. Numerical integration of satellite orbits around an oblate planet., Astron. Astrophys. 354, 328–333 (2000)
2. Крилов, В., В. Бобков, П. Монастырный. Вычислительные методы т.2, Москва, "Наука", 1977, с.21.
3. Бордовицына Т. В. Метод Рунге-Кутта высоких порядков и стабилизирующие преобразования в задачах прогнозирования движения ИСЗ, Космические исследования, т.19, вып.6, 1981г, с. 941-943.
4. Sharp P. W. N-body simulations: the performance of eleven integrators, Report Series 506, Department of Mathematics, University of Auckland, 2003
5. Nystrom, E. J. Uber die numerische Integration von Differentialgleichungen, Acta Soc. Sci. Fennicae 50, (1925), 1-54.
6. Burrage K., P. W. Sharp. *A class of variable-step explicit Nordsieck multistep methods*, SIAM J. Num. Anal., 31 (1994), 1434-1451.
7. Butcher J. C., P. Chartier and Z. Jackiewicz. Experiments with a variable-order type 1 DIMSIM code Numerical Algorithms 22 (1999) 237–261
8. Montenbruck O., E. Gill. State interpolation for on-board navigation systems *Aerosp. Sci. Technol.* 5 (2001) 209–220
9. Shampine L. F. *Interpolation for Runge-Kutta methods*, SIAM J. Numer. Anal., 22 (1985), pp. 1014–1026.
10. Sharp P. W. Numerical Comparisons of Some Explicit Runge-Kutta Pairs of Orders 4 Through 8., ACM Transactions on Mathematical Software, Vol. 17, No. 3, September 1991, 387-409.
11. Brankin R. W., I. Gladwell, L. F. Shampine. RKSUITE: A Suite of Explicit Runge-Kutta Codes., WSSIAA 2(1993) pp. 41-53
12. Hadjifotinou K. G., M. Gousidou-Koutita. Comparison of Numerical Methods for the Integration of Natural Satellite Systems, Celestial Mechanics and Dynamical Astronomy, 70: 99-113, 1998.
13. Sharp P. W. N-body simulations: the performance of eleven integrators, Report Series 506, Department of Mathematics, University of Auckland, 2003
14. Fehlberg E. Klassische Runge-Kutta formeln funfter und siebenter Ordnung mit Schrittweitenkontrolle., Computing, v. 4, p.93-106, 2969.
15. Fehlberg E. Klassische Runge-Kutta formeln vierter und nidrigerer Ordnung mit Schrittweitenkontrolle und ihre Anwendung auf Warmeleitungsprobleme., Computing, v. 6, p. 61-71, 1970.
16. Dormand J. R., P. J. Prince. A family of embedded Runge-Kutta formulae. J.

Comput. Appl. Math. 6(1980), 19-26.
17. Enright W. H., D. J. Higham, B. Owren, P. W. Sharp. *A survey of the explicit Runge-Kutta method*, TR 291/94, Department of Computer Science, University of Toronto, 1994.
18. Papakostas S. N., Ch. Tsitouras and G. Papageorgiou. A general family of Runge-Kutta pairs of orders, 6(5), SIAM J. Numer. Anal. 33(1996), 917-926.
19. Tsitouras Ch. "Optimized explicit Runge-Kutta pair of orders 9(8)", *Appl. Numer. Math.*, 38(2001) 123-134.
20. Petzold L. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations., SIAM J. SCI. STAT. COMPUT., v. 4, No. 1, 1983.
21. Shampine L. F., L. S. Baca. Fixed versus Variable Order Runge-Kutta., ACM Transactions on Mathematical Software, Vol. 12, No. 1, 1986, 1-23.
22. Атанасов Ат. Начин за повишаване на ефективността при числено интегриране на обикновени диференциални уравнения с постоянна стъпка, за нуждите на имитационното моделиране. Сб."Математика и математическо образование", София, БАН, 1987, с. 306.
23. Эскобал П. Методы определения орбит, "Мир", Москва, 1970, с. 44.
24. Атанасов Ат. Программная реализация алгоритма определения оптимальной схемы численного интегрирования систем дифференциальных уравнений космической навигации. Сб. 8-ой конференции Секции № 6, 1986, с. 7.

ИНТЕГРИРАНЕ НА УРАВНЕНИЕТО НА ДВИЖЕНИЕ НА ИЗКУСТВЕНИ СПЪТНИЦИ НА ЗЕМЯТА С ИЗБОР НА СХЕМИ НА РУНГЕ-КУТА-ФЕЛБЕРГ С ОПТИМАЛЕН РЕД НА ТОЧНОСТ

А. Атанасов

Резюме

Обсъжда се подход за числено интегриране на уравнението на движение на изкуствени спътници на Земята с избор на изчислителна схема осигуряваща необходимата локална точност. Разглежданията са направени на основата на изчислителни схеми от типа на Рунге-Кута-Фелберг. Предлага се критерий за избор на "оптимална" схема. Ефективността и устойчивостта на разглеждания подход е илюстрирана с резултати от числени експерименти. Изборът на оптимална схема е подходящ при решаване на многопараметрични задачи в сферата на имитационното моделиране, където е от значение намаляването на изчислителните разходи при достатъчна глобална устойчивот на численото решение.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

A SPATIAL MODEL TO EVALUATE MAN-INDUCED TRANSFORMATION USING GEOINFORMATION TECHNOLOGIES

Eugenia Roumenina, Georgi Jelev, Roumen Nedkov, Vanya Naydenova, Georgi Kanev

Space Research Institute - Bulgarian Academy of Sciences
e-mail: roumenina@space.bas.bg

Abstract

The sustainable use of municipal territory, complying with ecological and social-economic features, is among the top-ranking directions of modern regional policy. A key problem here is the evaluation and control of the territories subject to intensive anthropogenic activity. The paper presents a methodology for performing such type of study. Using conventional-source data and satellite imagery integrated in a data base, a spatial model to evaluate the man-induced transformation of the land of Novi Iskar, Sofia Municipality, was designed. The ranking of land-use categories and the maps of the man-induced transformation distribution index were created after Goffmann's methods adapted for Bulgarian territory by Iliev-Ilieva. As a result of the dynamic urbanization process and the intensive agricultural and industrial-transport activity, the landscape structure of the studied land has experienced material changes. The calculated regional man-induced transformation index (Ual = 554) is close to the index of Sofia Municipality, which is the highest for the country. It is strongly affected by the high values of the index of fields that were utilized for residential and industrial

The study is implemented within the framework of Scientific-research contract NZ – No.1507/05 concluded between SRI-BAS and Scientific Research Fund at the Ministry of Education and Science

purposes. Based on the generated digital terrain model, a series of derivative morphometric cartographic models was created, which reveal the nature and features of the relationships between the relief's plastics and the spatial distribution of the individual land-use categories. The created geodata base provides for express retrieval of thematic information from multi-channel satellite images for the purpose of monitoring the examined territory's man-induced transformation. Obtaining regular unbiased information is of great importance for the Municipality's adequate policy formation and funding.

Keywords: *man-induced transformation, remote sensing, geographic information systems, spatial modeling.*

1. Introduction

The sustainable use of municipal territory, complying with ecological and social-economic features, is among the top-ranking directions of modern regional policy. A key problem here is the evaluation and control of the territories subject to intensive anthropogenic activity. The Town of Novi Iskar is one of the four towns located on the territory of Sofia Municipality. It was proclaimed a town in 1974, as a result of merging the Villages of Kourilo, Koumaritsa, Slavovtsi, and Gnilyane. The Town of Novi Iskar falls within the Municipality's northern industrial area which incorporates as well the industrial settlements of Svetovrachane and Kremikovtzi.

The main objective of the study is the design of a spatial model to evaluate the man-induced transformation of the territory of the Town of Novi Iskar, Sofia Municipality. To achieve it, three interrelated tasks must be solved, namely: 1. To create a geodata base for the territory of the Town of Novi Iskar containing information on land-use; 2. To generate digital terrain model (DTM) and to create a series of derivative morphometric cartographic models, revealing the nature and features of the relationships between the relief's plastics and the spatial distribution of the individual land-use categories. 3. To identify and analyze the man-induced transformation of the examined territory.

2. Methods and information background of the study

In implementing the specified tasks, the basic technologies used were geoinformation technologies (Burrough A., 1996, Lillesand T., Kilfer R., 2000, Тикунов В., 2004), involving geoinformation systems and data processing technologies applied in remote sensing of the Earth. The work methods comprise eight major stages, which are shown in Fig. 1.

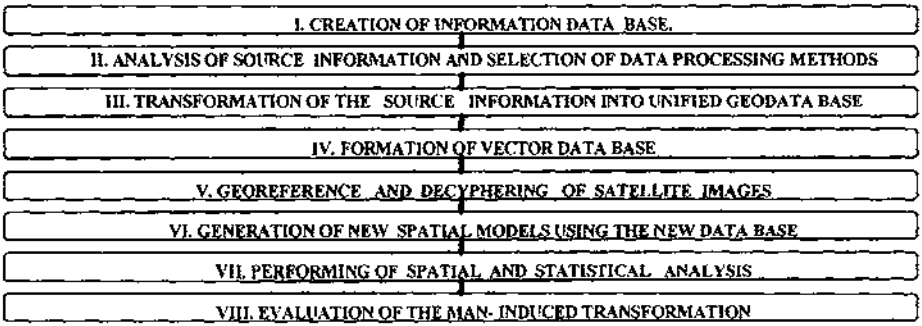


Fig. 1. Work stages

To achieve the objective of the study, the following information sources were used: large-scale topographic maps in scale M 1:10,000; satellite images from Landsat ETM$^+$ and IKONOS taken in 2000 and 2005, accordingly, and data from the National Institute of Statistics. The topographic maps and satellite images were georeferenced to a unified coordinate system (Bulgarian coordinate system, 1970) into a geodata base. By digitizing the topographic maps and visual computer-aided interpretation and photo-revision of the satellite images, vector layers were created. The vector data base was used to generate a digital terrain model and its derivative cartographic models of topographic surface slope and aspect. The performed modelling using GIS tools helps to reveal the relation between the man-induced transformation and the morphographic features of the land of the Town of Novi Iskar. As a result of the performed operations related to information input and preliminary processing, layers of geodata bases were created, required by the subsequent spatial analyses (Table 1). A series of derivative morphometric cartographic models was created, describing the relation between man-induced transformation attributes and relief characteristics. Upon analysis and generalization, the obtained spatial quantitative evaluations were presented in the form of maps (Figs. 2), graphs and tables.

Table 1. Data distribution by layers in the data base

Name of layer	Type of layer
Satellite images from Landsat ETM and IKONOS	Raster
Large-scale topographic maps	Raster
Land boundary	Vector-polygon
Boundary of the quarters' territory	Vector-polygon
Triangulation points	Vector-Points
Ground control points measured by GPS	Vector-Points
Relief isolines	Vector-Line
Forest territories	Vector-Polygon
Natural meadows	Vector-Polygon
Pastures	Vector-Polygon
Perennial plants (vineyards, orchards)	Vector-Polygon
Fields	Vector-Polygon
Water areas, streams, and hydro-melioration equipment (rivers, dams, gullies, canals, etc.)	Vector-Polygon
Transport and infrastructure territories	Vector-Polygon
Built-up lands for residential and industrial purposes	Vector-Polygon
Disturbed lands (mines, quarries, landfills etc.)	Vector-Polygon
Ground measurements 2006–2008	Vector-Points
Digital terrain model	Raster
Slope	Raster
Aspect	Raster
Man-induced transformation	Raster

3. Major results

The spatial analysis and evaluation of man-induced transformation was performed after Goffmann's methods adapted for Bulgarian territory by Iliev-Ilieva (Илиев Ил., М. Илиева 1998). In it, the number of land-use categories is reduced to 10, and each category is assigned an appropriate man-induced transformation rank (r). The values of the man-induced transformation ranks (r) for the respective land-use categories are as follows: Protected territories (protected natural territories, archaeological sites, sanitary-protected areas etc.) – 1; Forest territories – 2; Natural meadows – 3; Pastures – 4; Perennial

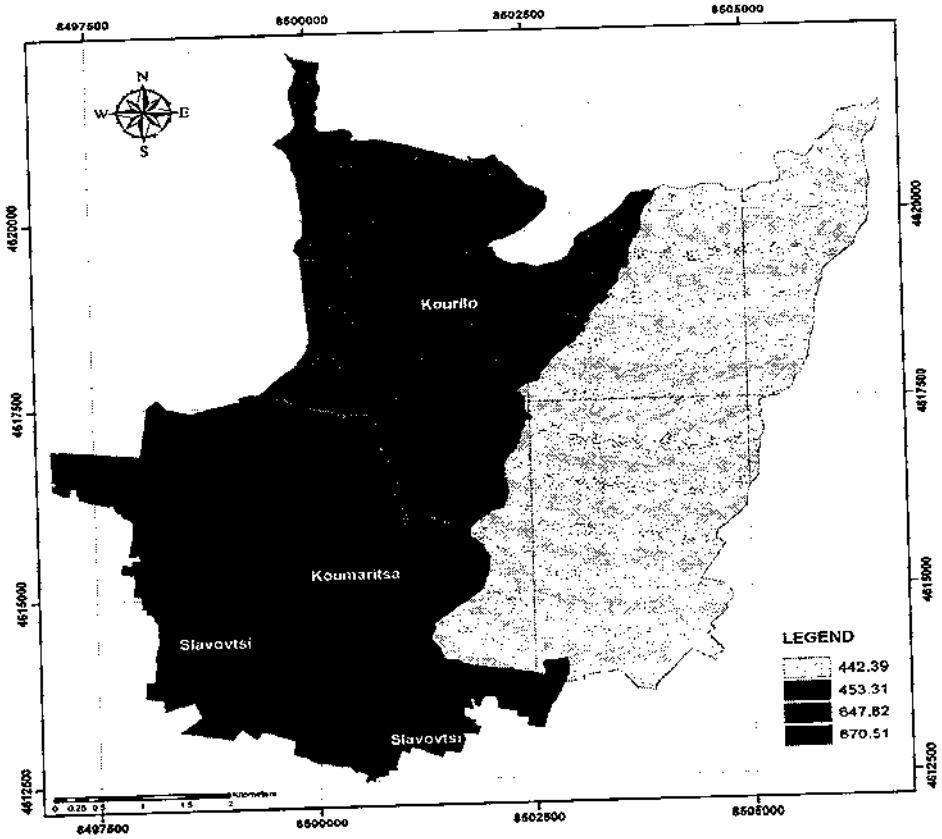


Fig.2. Distribution of man-induced transformation index (Ual) on the territory of the town of Novi Iskar by quarters

plants (vineyards, orchards) – 5; Fields – 6; Water areas (rivers, dams, gullies, canals, etc.) – 7; Transport and infrastructure territories – 8; Built-up lands for residential and industrial purposes – 9; Disturbed lands (mines, quarries, landfills etc.) – 10. The proposed man-induced transformation index (Uam) is equal to the product of the man-induced transformation rank (r) of the respective land-use category and its portion of the overall territory in %. The sum of the man-induced transformation indices of the individual categories represents the local index (Ual).

Within the boundaries of the examined territory, 9 land-use categories were identified, whereas only protected territories (Fig. 2) were not identified. These categories have been formed under the influence of both natural and social-economic factors.

Forest territories

Forest territories are of great ecological importance as a factor maintaining natural equilibrium. The greater portion of mountain lands in the northern part of the region accounts for a greater portion of forest-occupied territories compared to the plane southern part. This land-use category occupies 30.2 % of the region's territory. It is lacking in the Koumaritsa and Slavovtsi quarters. The forests are located mostly in the low mountainous area (altitude of 600 - 1,000 m), along southern slopes with inclination from 10^0 to 30^0 (Table 2). The man-induced transformation index is 60.32, which is close to the average value for Sofia Municipality (58.4). Its relative share ranks third in significance to the formation of Ual.

Natural meadows

The relative portion of natural meadows is 5.52 %, whereas in the Slavovtsi quarter it is less than 1%. They are located mostly in plane territories with altitude of 500 - 600 m (Table 2). Part of the natural meadows is located on the flooded terrace of the Iskar River, which creates marshes at spring water maximum. Another reason for the marshes, particularly in the Gnilyane area, is the high level of underground water.

The man-induced transformation index is 16.56, which is higher than the average for Sofia Municipality (11.7). This relatively high index evidences of available favourable conditions for development of pasture stock-breeding.

Table. 2. Percentage of the area distribution by land-use categories conformity with the relief characteristics on the land of the town of Novi Iskar

Relief characteristics		Land-Use Categories								
		Forest territories	Natural meadows	Pastures	Perennial plants	Fields	Water areas, streams and hidromelio-rative equipment	Transport and infrastructure territories	Built-up lands for residential and industrial purposes	Disturbed lands
Elevation levels	200 m - 600 m	21.3	99.5	66.1	32.9	96.7	99.5	96	95.3	66.7
	600 m - 1000 m	78.6	0.5	33.9	67.1	3.3	0.5	4.0	4.7	33.3
	1000 m - 1600 m	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Slope(°)	0 ° - 1°	9.5	85.5	43.3	14.3	60.4	83.1	56.8	50.4	25.4
	1 ° - 3 °	4.1	8.4	15.2	3.7	24.0	10.2	35.6	25.7	37.4
	3 ° - 7 °	10.0	3.6	18.0	20.4	11.8	5.4	3.5	13.7	33.7
	7° -10°	10.1	2.0	9.9	30.0	3.0	1.0	0.7	5.6	0.6
	10 ° - 15 °	23.7	0.4	9.5	24.0	0.7	0.2	1.1	3.5	0.6
	15 ° - 20 °	20.2	0.1	2.4	5.7	0.1	0.1	0.9	0.9	0.8
	20 ° - 30 °	19.1	0.0	1.4	1.5	0.0	0.0	1.0	0.3	1.4
	>30 °	3.4	0.0	0.2	0.4	0.0	0.0	0.4	0.0	0.1
Aspect	Flat	8.6	75.1	24.6	12.9	36.8	62.4	29.3	13.7	9.5
	North	3.1	0.9	1.8	0.1	0.6	0.1	1.2	0.5	0.6
	Northeast	4.6	6.6	1.4	0.4	1.8	0.9	2.1	3.7	9.4
	East	10.6	3.1	5.8	1.7	2.7	2.5	4.9	13.2	9.5
	Southeast	16.0	0.3	12.3	13.7	16.0	1.3	11.7	31.2	38.0
	South	16.2	2.6	23.4	28.0	21.6	6.0	38.9	13.1	30.6
	Southwest	17.9	7.8	10.3	18.1	10.0	11.8	5.7	10.8	2.3
	West	14.8	3.4	9.4	14.6	8.1	12.4	5.1	10.8	0.1
	Northwest	8.4	0.2	11.0	10.4	2.4	2.7	1.1	3.1	0.0

Pastures

This category occupies 3.44 % off the overall area, whereas most of the pasture territories are located in areas featuring altitude of 500 - 600 m and inclination of 0^0 to 3^0 (Table 2). The man-induced transformation index is 13.76, which is by 24 points less than the average for Sofia Municipality. The relative share in Ual formation is among the lowest ones - 3 %.

Perennial plants

The land-use category of perennial plants on the territory of the Town of Novi Iskar occupies only 1.4 % of the overall area. The greater part of these is located on the territory of the Gnilyane Quarter, at altitude of 600 to 1,000 m, along slopes of southern exposure featuring inclination of 3^0 to 15^0 (Table 2). A look-up into old topographic maps in scale M 1:10,000 issued in 1970 reveals that its territory used to be occupied by vineyards and orchards. The relative share of this category's Uam in the formation of Ual is only 1 %, notwithstanding the value of Uam (6.95) for the examined territory which is nearly threefold greater compared to the respective value for Sofia Municipality (2.5).

Fields

The greatest relative share in the territory's structure is occupied by the field category - 33.94 %. The man-induced transformation index (203.64) for the considered land-use category is substantially higher than the average value for Sofia Municipality (152.4). Its relative share in Ual formation is highest – an overall of 38 % for the whole land of the Novi Iskar region. This tendency is not uniform for the lands of the individual quarters, varying from 30.36 for the Kourilo Quarter to 422.04 for the Slavovtsi Quarter (Table 3). The fields are located in plane areas with altitude of 510 - 600 m.

Currently, the structure of arable land shows that the greatest part of them is occupied by fields planted with grain crops: maize, barley, wheat, and fodder – mainly alfalfa. Vegetable-planted areas rank second.

Water areas, streams, and hydro-ameliorative equipment

Water areas occupy 3.4 % of the overall area, which points to a relatively well organized irrigation system. This parameter is strongly affected by the Iskar River, which flows through the lands of the Quarters of Koumaritsa, Kourilo, and Gnilyane. Notwithstanding the erected protective embankments, part of these quarters' territory was damaged by the floods of June 2005 (Fig. 3). The man-induced transformation index is 23.73, which is close to the average for Sofia Municipality (23.1).

Table 3. Man-induced transformation index (Uam) on the territory of the Town of Novi Iskar

Land-Use Categories	Town of Novi Iskar	Koumaritsa Quarter	Gnilyane Quarter	Slavovtsi Quarter	Kourilo Quarter
Forest territories	60.32		82.68		105
Natural meadows	16.74	22.23	16.56	2.22	19.68
Pastures	13.76	13.32	13.92	14.44	13.52
Perennial plants	6.95	0.5	17.65	0.55	
Fields	203.64	316.8	189.42	422.04	30.36
Water areas, streams, and hydromeliorative equipment	23.73	33.53	22.19	4.27	27.65
Transport and infrastructure territories	40.64	36.24	20.32	137.52	26
Built-up lands for residential and industrial purposes	136.17	202.59	79.65	66.78	197.1
Disturbed lands	18.9	45.3			34

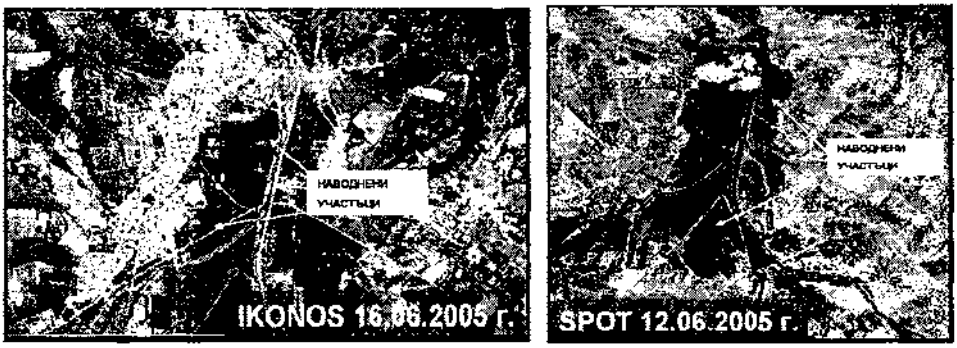


Fig. 3. Satellite images from IKONOS and SPOT on which the areas flooded in June 2005 may be seen

Transport and infrastructure territories

The infrastructure type of land use includes the areas occupied by the transport network elements (inclusive of field and forest roads) outside the populated areas and the pertaining infrastructure. It occupies 5.1 % of the territory's overall area. The man-induced transformation index (40.84) is twice as big as the index of Sofia Municipality (12.0). It is greatly influenced by the lands occupied by the Airport of Dobroslavtsi, part of which is located on the land of the Slavovtsi Quarter. Currently, the airport is not used according to its destination. The relative share of the considered category's Uam in the formation of the local index is 8 %. The transport network on the plane territory is well organized, field roads constituting an important part thereof.

Built-up lands for residential and industrial purposes

The territory's building up is associated with some of the gravest damages during man-induced transformation, occupying 15.1 % of the overall area forming the structure of the examined territory's land use. A number of industrial and service sites and equipments have been erected on the region's territory. The relative share of this category's Uam in the formation of Ual is 26 %, which ranks it second in significance to the formation of the local index, after the field category. The man-induced transformation index is 136.17 which is less than the average for Sofia Municipality (163.8). This tendency is not uniform for the lands of the individual quarters, varying from 66.78 for the Slavovtsi Quarter to 202.59 for the Koumaritsa Quarter (Table 3).

Disturbed lands

The greatest changes in the territory's natural status and regime have occurred in the north-west part of the region as a result of the development of mining and quarry activity. The relatively low stage of development of industrial technologies, the insufficient or lacking depth selection, and last but not least, the deposits' low useful content called for exploitation of large fields. The highest values of the index of the disturbed lands' category are associated with the Quarters of Koumaritsa and Kourilo where Uam is equal to 45.3 and 34, accordingly.

Disturbed lands occupy 1.9 % of the overall area. The land hosts one deserted uranium mine, one deserted open-extraction coal mine and clay-extraction quarries. Despite that these sites feature the highest man-induced transformation rank their share in Ual formation is insignificant – 4 %. The Uam index for this category is 18.9, which is much less than the average for Sofia Municipality (51). This tendency varies with the individual quarters, whereas Gnilyane and Slavovtsi have no disturbed lands (Table 3).

4. Conclusion

As a result of the dynamic urbanization process and the intensive agricultural and industrial-transport activity, the landscape structure of the studied land has experienced material changes. The calculated local man-induced transformation index (Ual = 554) is greater than the value of this index for Sofia Municipality, which is the highest for the country. It is strongly affected by the great values of Ual for fields and built-up lands for residential and industrial purposes. This tendency is characteristic of the Koumaritsa and Slavovtsi Quarters, while the values of Ual for the other two quarters are close to the average value for the country (448.1).

The plane relief in the southern part of the territory, the soil and agro-climatic conditions and resources favour the development of agricultural land use. Despite the relatively favourable agro-climatic conditions, few perennial plants are grown in the region.

The created geodata base provides for express retrieval of thematic information from multi-channel satellite images for the purpose of monitoring the examined territory's man-induced transformation. Obtaining regular unbiased information is of great importance for the Municipality's adequate policy formation and funding.

References

1. Burrough, A. (1996). Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press, Oxford, 210 p.

2. Lillesand T., R. Kilfer (2000). Remote Sensing and Image Interpretation, John Wiley & Sons , 721.

3. Основы Геоинформатики, под ред. Тикунова В. (2004) ACADEMA, Москва, т. 1, 345, т. 2, 477.
4. И л и е в И л., М. И л и е в а, (1998). Оценка на антропогенната преобразуваност на територията на България, Сборник "100 години география в Софийския университет", И-во "Св.Кл.Охридски", С, стр. 523 – 531.
5. Брой на населението по области, общини и населени места, (2005). Национален статистически институт, София.

ПРОСТРАНСТВЕН МОДЕЛ ЗА ОЦЕНКА НА АНТРОПОГЕННАТА ПРЕОБРАЗУВАНОСТ С ИЗПОЛЗВАНЕ НА ГЕОИНФОРМАЦИОННИ ТЕХНОЛОГИИ

Е. Руменина, Г. Желев, Р.Недков, В. Найденова, Г. Кънев

Резюме

Устойчивото използване на територията на общините, съобразено с екологичните и социално-икономически особености, е едно от приоритетните направления на съвременната регионална политика. Един от ключовите проблеми е оценката и контролът на териториите, подложени на интензивна антропогенна дейност. В статията е представена методология за провеждане на такъв тип изследване. С използване на данни, получени от конвенционални източници и спътникови изображения, интегрирани в геобаза данни, е съставен пространствен модел за оценка на антропогенната преобразуваност на землището на гр. Нови Искър, Столична община. Градацията на категориите земеползване и картите за разпределението на индекса на антропогенна преобразуваност са съставени по адаптираната за територията на

България от Илиев–Илиева методика на Гофман. Под въздействието на динамично протичащия урбанистичен процес, както и в резултат на интензивната селскостопанска и промишлено–транспортна дейност ландшафтната структура в изследваното землище е претърпяла съществени изменения. Изчисленият регионален индекс на антропогенна преобразуваност (Ual = 554) е близък до този на Столична община, който е най-високият за страната. Силно влияние върху неговото формиране оказват високите стойности на индекса от ниви и застроени земи от населените места и промишлеността. На базата на генериран цифров модел на релефа са съставени серия от производни морфометрични картографски модели. Те разкриват характера и особеностите на взаимовръзките между пластиката на релефа и пространственото разпределение на отделните категории земеползване. Създадената геобаза данни дава възможност за експресно извличане на тематична информация от многоканални спътникови изображения за извършване на мониторинг на антропогенната преобразуваност на изследваната територия. Получаването на регулярна обективна информация има изключително значение както за формиране на адекватна политика на общината, така и за финансовото й осигуряване.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

RELATIONSHIP BETWEEN LIGHT DISPERSION AND BRIGHTNESS DURING TELESCOPIC DEVICE OBSERVATIONS

Jivko Jekov

Space Research Institute – Bulgarian Academy of Sciences
e-mail: zhekovz@yahoo.com

Abstract

Light diffraction is a very important telescope characteristic which affects the contrast of the image recorded by the observer's eye.

The purpose of the study is to determine to what extent the coefficient of light diffraction affects the brightness of the recorded image.

The subject of the theoretical research are the experimental results obtained during telescope system experiment in the process of observation of remote objects with different brightness of the background at fixed light diffraction coefficients and constant contrast of the object' background.

The received values and the subjection of the contrast of the image to light diffraction coefficient is shown in graphic form. It is found out that when increasing the value of background brightness in constant background contrast in respect to the object, the image contrast sharply decreases. The relation between increase of light diffraction coefficient and the decrease of image brightness can be observed by telescope apparatuses.

Light dispersion is optical device characteristics which affects the contrast of the image recorded by the observer's eye.

Many objects disperse light falling onto them, so the brightness's values

along the various directions appear to be close. According to Lambert's law [3,6], the brightness of a light-dispersing surface is equal in all directions. This assertion may be assumed only as an approximation.

Let σ be a small area with brightness β (Fig. 1) equal in all directions.

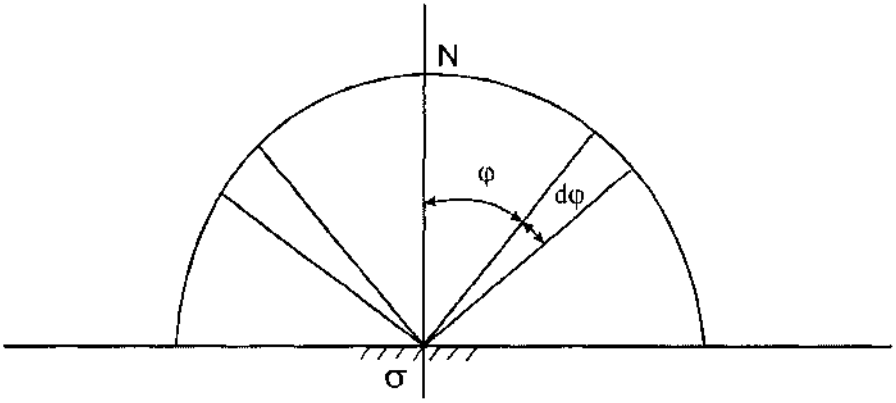


Fig. 1. Definitions

The light flow ψ emitted from area σ along the normal constituent of angle φ is calculated. Isolating the bodily angle dφ located between two ring cones generated by the rotation about normal N of two lines forming angles φ and φ + dφ, produces apparently:

(1) $d_{\varphi} = 2\pi.\sin\varphi\, d\varphi$.

The light intensity within this spatial angle is constant. Therefore, the light flow within the bodily angle dφ will be:

(2) $d\varphi = I_{\varphi} d\varphi = 2\pi\pi B\sigma \sin\varphi\cos\varphi\, d\varphi$.

To determine the light flow ψ emitted by the area within the whole hemisphere, the above expression must be integrated within the limits from 0 to π/2. Then: $\psi = \pi B\sigma$.

(3) $$M = \frac{\varphi}{\sigma} = \pi B \ .$$

The above shows that to brightness $B = 1\,cd/m^2$ corresponds lightness: $M = 3,14\,lm/m^2$.

The surface properties of each diffusely dissipating body differ greatly from those of the ideal light dissipater, i.e., the brightnesses in the various directions are different. To provide numerical characteristics of surface brightness change in various directions, the light dissipation factor for a given surface is used, i.e., the ratio of the brightness of the surface along an arbitrary direction and the brightness of an ideal dissipater, placed under the same illumination conditions. The light dissipation factor is usually denoted by β [1, 4].

The task is to investigate whether the dissipation factor β affects the brightness of the recorded image.

The subject of theoretical research are the results obtained by an experiment with observation telescopic system [5] represented on Table 1 during the observation of remote objects with various background brightnesses ranging between 10^{-2} and 10^{-3} cd/m^2 with given light dissipation factors: $\beta_1 = 0.1$; $\beta_2 = 0.2$; $\beta_3 = 0.3$ and constant contrast of the object's background $K = 0.3$.

Table 1

Light dissipation factor							
	$\beta_\phi = 0,01$	$\beta_\phi = 0,1$	$\beta_\phi = 1$	$\beta_\phi = 10$	$\beta_\phi = 100$	$\beta_\phi = 10^3$	X
$\beta_1 = 0,1$	0,2999	0,2981	0,2431	0,0901	0,0125	0,0013	$\overline{X_1} = 0,1566$
$\beta_2 = 0,2$	0,2986	0,2868	0,2054	0,0517	0,0064	0,0006	$\overline{X_2} = 0,1415$
$\beta_3 = 0,3$	0,2979	0,2806	0,1124	0,0379	0,0043	0,0004	$\overline{X_3} = 0,1239$
							$\overline{X} = 0,1407$

In the last column of Table 1, the obtained data is presented, considered as values of the brightness x for the group of factors β_1, β_2, β_3, i.e., z = 3, where the mean group values are denoted by $\overline{X_1}$, $\overline{X_2}$, $\overline{X_3}$ and the overall mean value $\overline{X}$ for the considered brightnesses n = 6 are calculated using formulae [2]:

(4) $$\overline{X} = \frac{1}{2}\sum_{i=1}^{z} xij \qquad i = 1,2...z$$

(5) $$\overline{\overline{X}} = \frac{1}{zn}\sum_{i=1}^{z}\sum_{i=1}^{n} = \frac{1}{z}\sum_{i=1}^{z}\overline{xi} .$$

The hypothesis H which must be verified suggests that the light dissipation factor β does not affect brightness, while the alternative hypothesis suggests the opposite. To check up the zero hypotheses H, the averaged data from the 18 performed studies must be processed. The data processing includes calculation of the square sums $\varsigma, \varsigma_A, \varsigma_R$ using formulae:

(6) $$\varsigma = \sum_{i=1}^{z}\sum_{i=1}^{n}\left[x_{ij} - \overline{\overline{x}}\right]^2$$

(7) $$\varsigma_A = \sum_{i=1}^{z}\sum_{i=1}^{n}\left[\overline{\overline{x}} - \overline{x_i}\right]^2 = n\sum_{i=1}^{z}(\overline{\overline{x}} - \overline{x_i})^2$$

(8) $$\varsigma_R = \sum_{i=1}^{z}\sum_{i=1}^{n}(x_{ij} - \overline{x_i})^2$$

while the dispersions S^2, S_A^2 è S_R^2 are evaluated using formulae [4]:

(9) $$S^2 = \frac{2}{\nu} = \frac{2}{k\pi - 1}$$

(10) $$S_A^2 = \frac{\varsigma_A}{\nu_A} = \frac{\varsigma_A}{k - 1}$$

(11) $$S_R^2 = \frac{\varsigma_R}{\nu_R} = \frac{\varsigma_R}{k(\pi - 1)}.$$

The obtained values are shown on Table 2.

Table 2

Types of square sums	Square sum	Degree of freedom	Dispersion evaluation
Total	$\varsigma = 0{,}331486$	$\nu 6 = 17$	$S^2 = 0{,}019499$
By factors	$\varsigma_A = 0{,}000537$	$\nu_A = 2$	$S_A^2 = 0{,}000268$
Residual	$\varsigma_R = 0{,}022063$	$\nu_R = 15$	$S_R^2 = 0{,}022063$

The calculation of the disperse ratio F is performed using formula:

(12) $$F = \frac{S_A^2}{S_R^2} = 0{,}0121831.$$

The obtained disperse ratio (12) is compared with the table value F_T at significance level $\alpha = 0.05$ [2] and it is observed that $F > F_T$, which evidences that light dissipation affects image brightness.

Accounting for the fact that the contrast K depends on the object's brightness B_{ob} and the background B_b, K may be determined from:

(13) $$K = \frac{B_{об} - B_ф}{B_ф}$$

and, accounting for the additional brightness ΔB, due to light dissipation, which may be written as:

(14) $$\Delta B = (B_{об} + B_ф)$$

the contrast of the image K` recorded by a visual optic system during observation of a remote object will be equal to:

(15) $$K' = \frac{B_{об} - B_ф}{B_ф + \beta(B_{об} + B_ф)} = \frac{K}{1 + \beta(B_{об} + B_ф} = \frac{K}{1 + \Delta B}.$$

From expression (15) it follows that, with definite object contrast with respect to the surrounding background, the image contrast K will be reduced, while the light dissipation factor increases.

In Fig. 2, the curves for the appropriate dissipation factors are shown. Apart from the image contrast's reduction with the light dissipation factor β's increase, the curves presented in Fig. 2 also reveal that the contrast K` of the recorded image drops abruptly when the background's brightness exceeds (24…30) cd/m^2, i.e., the specified background contrast with respect to the object, which is 0.3, does not provide proper image of the observed remote objects. Therefore, at some given contrast of the object with respect to the surrounding background, the contrast of the recorded image K` is reduced while the light dissipation factor increases. At background brightness within the range from 10^{-2} to 10^3 cd/m^2 it may be shown that, when the value of background brightness increases, while the background contrast with respect to the object K = 0.3, the image's contrast drops abruptly.

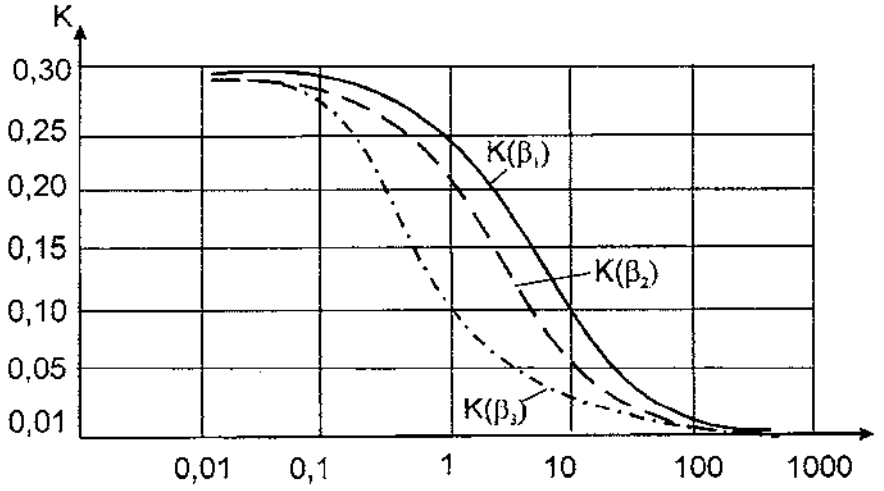


Fig. 2. Dependence of the image contrast on the light dissipation factor

The graphic relationship displays reduction of the image contrast with increase of the light dissipation factor β. Moreover, when the background's

brightness exceeds 24…30 cd/m^2, the contrast of the recorded image drops abruptly.

References

1. Александров А., Ж. Жеков, К. Вълчев. Влияние на светоразсейването върху видимостта на отдалечени обекти посредством визирни оптични уреди, Научна конференция, МТОС, Варна, 1988.
2. Днепровский С. Контрастность приборов, Оптико-механическое производство, Под ред. Фрайбурга, М. Оборонгиз, 1998.
3. Кендалл М., А. Стюарт. Статистические выводи и связи, Наука, М., 1973.
4. Жеков Ж. Проектиране, разчет и конструиране на оптически и електронно-оптически прибори на научни изследвания в областта на космическата физика. Кандидатска дисертация, София, 1983.
5. Жеков Ж. Визир В 3 х 40, Научна сесия ВНВАУ, Шумен, 1983, стр. 115–123.
6. Жеков, Ж. Методи и средства за откриване на отдалечени обекти от космически летателни апарати. Докторска дисертация, София, 2006, 355 с.
7. Lambert J. H. Photometria, sive del mensura et gradibus luminis, colorum et umbrae, Augsburg, 1760.

ЗАВИСИМОСТ МЕЖДУ СВЕТОРАЗСЕЙВАНЕТО И ЯРКОСТТА НА ОБЕКТА ПРИ НАБЛЮДЕНИЕ С ВИЗИРНИ ОПТИЧНИ УРЕДИ

Ж. Жеков

Резюме

Светоразсейването е характеристика на оптичните уреди, която влияе върху контраста на регистрираното изображение от окото на наблюдателя. Обект на теоретично изследване са резултати, получени от експеримент с визирна телескопична система при наблюдение на отдалечени обекти при различна яркост на фона при зададени три различни коефициенти на светоразсейване и постоянен контраст на фона на обекта. Направени са изводи за зависимостите между контраста на изображението, коефициента на светоразсейването, контраста на регистрираното изображение и яркостта.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

DETERMINATION OF THE ANGULAR COORDINATES OF ASTRONOMICAL OBJECTS THROUGH OPTIC-ELECTRONICAL MEASURING SYSTEMS

Jivko Jekov, Garo Mardirossian

Space Research Institute – Bulgarian Academy of Sciences
e-mail: zhekovz@yahoo.com; garo@space.bas.bg

Abstract

Among the methods for determination of the angular coordinates of astronomical objects, particularly topic is the Turner method based on static processing of data information for star number of three and more.

With regard to optical electronic apparatuses used in astronomical practice which high resolution and vision field lower than one degree feature the methods for angular coordinates determination with minimum number of supporting (catalogical) stars become topical. In such conditions it is necessary to re-examine the attitude to orthogonal methods, whose accuracy can be increased through statistical analysis.

The calculation of the ideal coordinates and the equatorial coordinates of an astronomical object is shown as well as the possible errors in their determination.

A priori it is supposed that the orthogonality method is more effective because of the fact that the possibility for appearance of two stars in the vision field is greater than for three ones.

Among the methods for determination of the angular coordinates of astronomical objects, particularly topical is the Turner method based on statical processing of data information for three and more supporting stars number $k \geq 3$ and the method of four constant (orthogonal) ones using two supporting points [1 ,2, 3].

In relation with the widely spread optic-electronical measuring systems in astronomical practice, which have high-resolution ability and relatively a narrow vision field- smaller than one degree [6], the methods for determining the angular coordinates arises with minimum number of points of support (catalog) stars. Become topical in these conditions, the attitude to the orthogonal methods has to be reconsidered, the accuracy of which can be increased by way of statistical arrangement in $k > 2$.

The influence of differential effects on optical systems with a narrow vision field is insignificant.

A method for determination of angular coordinates of astronomical objects, spreading on stars number $k > 2$ and calculation of objects coordinates by the least square method is presented.

The correlation between the ideal and the measured coordinates is expressed by formulas [1,3]:

$$(1) \qquad \begin{aligned} \zeta_i &= ax_i + by_i + c \\ \eta_i &= -bx_i + ay_i + f \end{aligned} \qquad \text{if } i = 1,k,$$

where: x, y – calculated coordinates of supporting stars;
ξ, η – ideal coordinates of supporting stars.

The expression (1) can be also wretten in the following way:

$$(2) \qquad v_{l,i} = x_{1,i}a_1 + x_{2,i}a_2 + 1a_3 + 0a + \xi_{1,i}$$

or in matrix form:

$$(3) \qquad \vec{v} = \overrightarrow{Xa} + \vec{\xi}$$

where: X is a matrix made by rows

$$\vec{x}_{1,i}\{\vec{x}_{1,i},\vec{x}_{2,i},1,0\}$$

(4) and

$$\vec{x}_{2,i}\{\vec{x}_{2,i}-\vec{x}_{1,i},1,0\}$$

and calculating the equations

(5) $$Q = X'X, K = Q^{-1} \ , \ \vec{L} = X'\vec{\xi}, \vec{a} = -K\vec{L} \ ,$$

where: $\vec{a}$ – vector evaluating the constant staff;
Q – a matrix with normal equations as follows:

(6)
$$\begin{matrix} [x_1^2]+[x_2^2] & 0 & [x_1] & [x_2] \\ 0 & [x_1^2]+[x_2^2] & [x_2] & -[x_1] \\ [x_1] & [x_2] & k & 0 \\ [x_2] & -[x_1] & 0 & k \end{matrix}$$

The cell method for matrix calculation [10] with regards to the symmetry and other characters of the matrix Q applying, it allows to find the most suitable form for passing from the calculated coordinates to the matrix K elements, forming and transporting the matrix X, and forming and rotating the matrix Q. The matrix K is as follows:

(7) $$K = Q^{-1} = (X'X)^{-1} = \frac{1}{W},$$

where: $$R = [x_1^2]+[x_2^2];$$

$$W = kR - [x_1^2]-[x_2^2]$$

The elements of the vector $\vec{L}$, have the following values:

$$\vec{L} = X'\vec{\xi} = \begin{vmatrix} [x_1\xi_1]+[x_2\xi_2] \\ [x_2\xi_1]+[x_1\xi_2] \\ [\xi_1] \\ [\xi_2] \end{vmatrix} \tag{8}$$

For ideal coordinates of the astronomical object, the following expression holds:

$$\begin{aligned} \vec{a} &= -KL \ , \\ \xi_1 &= \vec{x}_1, \vec{a} \ , \\ \xi_2 &= \vec{x}_2, \vec{a} \ , \end{aligned} \tag{9}$$

and its equatorial coordinates' estimates at known formula [1]:

$$\alpha = \operatorname{arctg}\left(\frac{\xi_1}{\cos D - \xi_2 \sin D} \right) + A \tag{10}$$

$$\delta = \operatorname{arctg}\left[\frac{(\sin D + \xi_2 \cos D)\cos(\alpha - A)}{\cos D - \xi_2 \sin D} \right], \tag{11}$$

where A and D are the equatorial coordinates in the stuff centre.

The errors in obtaining the coordinates of the object Δ_α and Δ_δ are collected from the error in the reduction Δ_ξ and the error of the calculated coordinates of the object Δ_x.

$$\dot{\Delta}_\alpha = \dot{\Delta}_{\xi_1} + M\dot{\Delta}_{x_1} \qquad \text{and} \qquad \dot{\Delta}_\delta = \dot{\Delta}_{\xi_2} + M\dot{\Delta}_{x2} \tag{12}$$

Here, the scale multiplier: $M = \sqrt{a_1^2 + a_2^2}$ [1].

The reduction error in solving a system of 2k equations of four unknowns by the least square method is determined by dispersion [11]:

$$(13) \qquad \sigma^2(\dot{\Delta}_{\xi_1}) = \frac{|v_1^2|}{2(k-2)} P_1 \ , \ \sigma^2(\dot{\Delta}_{\xi_2}) = \frac{|v_2^2|}{2(k-2)} P_2$$

where the weight coefficients P_1 and P_2 are determined by formulas [11]:

$$(14) \qquad P_1 = \sum \left(\frac{\partial \xi_1}{\partial a_j} \right)^2 K_j + 2 \sum \left(\frac{\partial \xi_1}{\partial a_i} \right) \left(\frac{\partial \xi_1}{\partial a_j} \right) K_i \ ;$$

$$(15) \qquad P_2 = \sum \left(\frac{\partial \xi_2}{\partial a_j} \right)^2 K_j + 2 \sum \left(\frac{\partial \xi_2}{\partial a_i} \right) \left(\frac{\partial \xi_2}{\partial a_j} \right) K_i \ .$$

Having attention to $\frac{\partial \xi}{\partial a} = \vec{x}$, P_1 and P_2 can be presented in square form suitable for the vectors $\vec{x}_1$ and $\vec{x}_2$ and the matrix K [8]:

$$(16) \qquad P_1 = \overrightarrow{x_1} K \overrightarrow{x_1} \quad \text{and} \quad P_2 = x'_2 K x_2$$

The final formula accuracy estimating is the accuracy:

$$(17) \qquad \tilde{\sigma}(\dot{\Delta}_\alpha) = \sqrt{\frac{|v_1^2|}{1(k-2)} P_1 + M^2 \sigma^2(\dot{\Delta}_{x_1})} \cos D$$

$$(18) \qquad \tilde{\sigma}(\dot{\Delta}\delta) = \sqrt{\frac{\left[\nu_2^2\right]}{1(k-2)} P_2 + M^2 \sigma^2 \left(\dot{\Delta}_{x_2}\right)}$$

In conclusion it has to be marked the question for comparison of the effectiveness of that and the other methods can be a topic of special investigation but apriori it is supposed the orthogonality method is more effective because of the fact that the possibility for appearance in the two stars field of vision is bigger than of the three ones.

References

1. Alexandrov, I. Applied astronomy, Moskow, Gostehizdat, 1999 (in Russian).
2. Volinetz, F. Photographical astronomy, Moscow, Mir, 1994 (in Russian).
3. Dache, A. Astronomer Journal, 1998.
4. Jekov, J. An optical system for photometrical objective, Scientifics Session of the *P.Volov* High Military School, Shoumen, 1997 (in Bulgarian).
5. Jekov, J., G. Mardirossian, S. Stoyanov, I. Hristov. Appreciation of the spatial and energetic parameters of the optical devices. Jubilee scientific Session '40 Years of the First Man Flight in Space', *G.Benkovski* High Military School, D. Mitropolia, 2001 (in Bulgarian).
6. Kolmikov, V. The television astronomy, Moscow, Nauka, 1994 (in Russian).
7. Monejerin, V. Excerption AO - the Crimea, 1994, vol. 1 (in Russian).
8. Neil, O. Introduction in statistical optics, Moscow, Mir, 1966 (in Russian)
9. Stoyanov, S., G. Mardirossian, J. Jekov, I. Hristov. Input influences onto the optical-electronic device in the angular coordinates of the remote objects calculation.
10. Fedorenko, O., L. Atroshtenko. Introduction in the matrix theory, Moscow, Nauka, 1996 (in Russian).
11. Chebosharev, A. A least square method with basic theory of probability, Moscow, Geodezizdat, 1998 (in Russian).

ОПРЕДЕЛЯНЕ НА ЪГЛОВИТЕ КООРДИНАТИ НА АСТРОНОМИЧНИ ОБЕКТИ ЧРЕЗ ИЗПОЛЗВАНЕ НА ОПТИКО-ЕЛЕКТРОННИ ИЗМЕРВАТЕЛНИ СИСТЕМИ

Ж. Жеков, Г. Мардиросян

Резюме

Използването в астрономичната практика на оптико-електронни системи с висока проницателна способност и сравнително тясно зрително поле мотивира актуалността на методите за определяне на ъгловите координати на минимален брой опорни (каталожни) звезди.

Предложен е метод за определяне на ъгловите координати на астрономични обекти, разпространени на брой звезди $k > 2$ и изчисляване на координатите на обекти по метода на най-малките квадрати.

В заключение е отбелязано, че проблемът за сравняване на ефективността на този и други методи трябва да бъде предмет на специално изследване, но априорно може да се предположи, че за оптико-електронни системи ортогоналният метод е по-ефективен поради факта, че вероятността за поява в зрителното поле на две звезди е по-голямо, отколкото на три.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

COMPANIES USING SATELLITE AND AIRBORNE REMOTE SYSTEMS FOR OFFSHORE OIL MONITORING

Palmira Panova, Petar Getsov

Space Research Institute - Bulgarian Academy of Sciences
e-mail: ppanova@space.bas.bg;
director@space.bas.bg

Abstract

The paper presents an analysis of remote sensing satellite and airborne oil detection systems and describes the use of their capabilities in exploration programs, monitoring, emergencies by oil and gas companies.

The detection of discharges, legitimate or illegal, can be performed by remote sensing techniques from aircraft and from satellite. Satellite observations complement aerial surveillance due to their wide area coverage with regular revisiting and to some degree is considered to be less weather -dependant.

Furthermore, satellite monitoring is one of the major monitoring components for large sea areas, where other means for surveillance purposes are too expensive or not available. Aerial surveillance aircraft have the capability to carry additional equipment to quantify and classify oil spills. The combination of both may increase the deterrent effect and can assist in optimising the flight activities of the surveillance aircraft.

Two types of remote sensing systems for oil spill detection and monitoring can be defined: space platforms or satellites including onboard sensors for collecting data, and systems with sensors mounted on airborne

platforms. Each type of system include a range of possible instruments onboard which operate sensing the energy in fixed parts of the electromagnetic spectrum, thus providing to determine different features of oil pollution. The satellite systems include the use of visual observations, thermal infrared (TIR) and UV imaging and satellite synthetic aperture radar (SAR). Aircraft remote sensing is based on visual observations, infrared imaging (IR), ultraviolet (UV), hyperspectral imaging, side-looking airborne radar (SLAR), airborne synthetic aperture radar (AIRSAR), down-looking thermal infrared (DTIR) and forward-looking infrared imaging (FLIR) [9]. Radar systems and airborne laser fluorosensor (ALF) are examples of active sensors and IR, UV and microwave radiometers are examples of passive sensors. All sensors must be calibrated and require highly trained personell to operate them and interpret the results. The strenghts and weaknesses of satellite and aircraft remote sensing systems are summarized on Table1.

Table 1

Platforms	Strenghts	Weaknesses
Satellite	Large field of view, spacial coverage of several km.	Long revisiting time and fixed time of overpass and high initial cost of sensors
	Ground infrastructure can be located in area of choice	Difficult to develop algorythms for detecting oil spills; cannot determine the types of oil
	No distance to the coast restrictions for operation	Predictable flight, difficult and expensive to shift once stabilized
		Absolute need of clear skies (optical and infrared sensors)
Aircraft	Equipment can be modified, updated and maintained easily	Restricted spatial coverage
	The devices are less cost by than the space ones	Require a number of resources for continuous coverage
	Flexible and manoeuverable, can fly in different hights and below clouds	Man power
	Highly visible to ships and this insreases the different effects	

Satellite observations complement aerial surveillance due to their wide area coverage with regular revisiting and to some degree are considered to be less weather-dependant. Furthermore, satellite monitoring is a major monitoring component for large sea areas, where other means are too expensive or not available for surveillance purposes. However, aerial surveillance aircraft have the capability to carry additional equipment to quantify and classify oil spills. The combination of both may increase the deterrent effect and can assist in optimising the flight activities of the surveillance aircraft. Fig.1 shows a typical case of a combined satellite and aerial surveillance system [14] .

Remote sensing is useful in several modes of oil spill control, including large area surveillance, site-specific monitoring and tactical assistance in emergencies. It is able to provide essential information to enhance strategic and tactical decision-making, decreasing response costs by facilitating rapid oil recovery and ultimately minimizing impacts for oil companies [10].

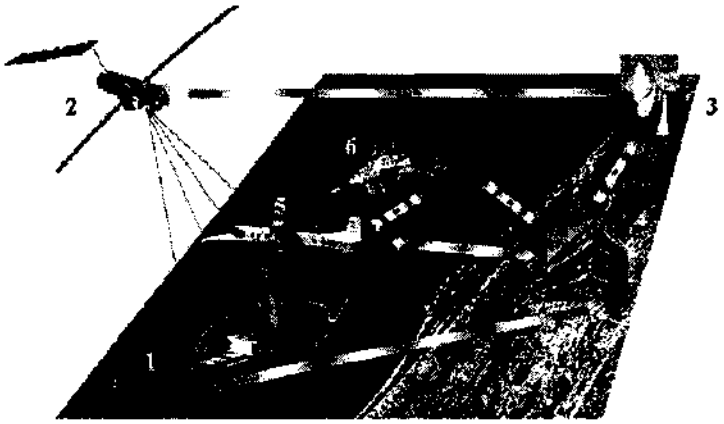


Fig. 1. Interaction diagram of the means of an integrated surveillance system to detect oil spills at sea. The oil spill (1), discharged during a ship disaster is detected by a regular SAR (synthetic aperture radar) satellite overpass (2). The oil can be detected due to the damping effect of oil on capillary waves. After transmission to the ground station (3) and data estimation (4), the aircraft (5) will be alarmed to verify the spill in terms of quantity and type with an advanced set of further remote sensing instruments. In case the quantity or the hazardousness of oil exceeds a certain threshold, the combating vessels will operate to remove the oil from the sea (6). These vessels are co-ordinated and instructed by the surveillance aircraft.

Visual observations of spilled oil from the air, along with still and video photography, are the simplest and most common method of determining the location and extent (scale) of an oil spill. Many devices employing the visible spectrum, including the conventional video camera, digital single-lens-reflex (SLR) cameras and camcorders are now available at reasonable prices. Dedicated remote sensing aircraft often have built-in downward looking cameras linked with a GPS to assign accurate geographic coordinates.

In the visible band, oil has no sharp spectral features. The similarity in the spectral signatures between oil and gas in the visible region of the spectrum explains the problems assosiated with oil slicks identification. For example, the capability of Landsat TM band 7 platform for identification of oil in water does not provide sufficient contrast to distinguish between oil and the background without intensive time processing. Oil slicks form very thin films on the sea surface and the thickness can vary from a tenth of a micron to hundreds of microns. Depending on thickness, oil spills are visualized by satellite optic sensor in colour from silver to black. The visual analysis of the image does not allow to determine oil thickness based on color and is unreliable.

Spills detection by visual observation is limited to favorable sea and atmospheric conditions and is inoperable in rain, fog, or darkness; visual observations are restricted to spill documentation because there is no mechanism for positive oil detection. Oil can be difficult to see in high seas and among debris or weeds where it can blend in to dark backgrounds, such as water, soil, or shorelines. In addition, the sun angle, glare, sea state and satellite and airborne camera view angle can affect the appearance of oil slicks [8] .

Infrared sensors, which detect infrared radiation levels given off surfaces, have been developed into relatively inexpensive sensors for ship-board and aerial observation of oil slicks.

There are airborne downward-looking thermal IR and forward-looking IR (FLIR). Thick oil (>0.003mm) appears hot or white in infrared data, middle oil thicknesses appears cool and black, and thin oil (0.0001mm) or sheens are not detected. Oil becomes 'visible' on thermal channels during daily images while at night the opposite is true- the oil body will lose heat faster than the surrounding water and will be seen as a cooler area on the image. These sensors are capable of detecting thicker parts of a slick only (>100 μm),

so they are useful for guiding response to the thicker oil parts. Infrared devices cannot detect emulsions and could be combined with an ultraviolet sensor for complete imaging of both the thick and thin portions of a slick [8].

Tests of a mid-band IR system MIR (3.4 to 5.4 μm) over oil spill showed no detection in this range, however, ship scars were visible. Studies in the thermal infrared TIR (8 to 14 μm) show that there is no spectral structure in this region. Oil detection in the infrared is not positive as several false targets can interfere, including weeds, shoreline, and oceanic fronts. Infrared is reasonably inexpensive, however, and is currently the prime tool used by the spill remote sensor operator.

Since thermal imagers, also known as forward looking IR (FLIR) devices, became commercially available, they are increasingly installed on smaller and mid-size aircraft. In September 2003, a StarSAFIRE thermal imaging system with a newly developed covert action laser illuminator (CALI) started operation as part of the second German surveillance aircraft, enabling the operator to read the ships name in the absence of daylight or investigate areas of specific interest. The laser is illuminating the NIR sensitivity of the inbuilt CCD camera which normally use an IR-cut filter during daytime, blocking unwanted NIR radiation from its detector and during night time and in combination with the CALI, this NIR cut filter is removed and the laser illumination is available up to a distance of 1500 m (Fig. 2).

This thermal imaging capability brings an important step towards night operation and polluter identification. It is also a sensor with a wide range of application, like SAR or border patrol, making it a perfect device for multi-role aircraft, often only equipped with basic maritime surveillance sensors [1] A disadvantage of any type of infrared detector, however, is that they require cooling to avoid thermal noise, which would overwhelm any useful signal [15].

Oil slicks have been detected using several types of optical and thermal satellite imagery but in all cases the oil position should be known in advance. Most optical satellites provide data from the visible part of the spectrum. Satellite sensors like Landsat TM (with spatial resolution of 120 m) and NOAA AVHRR provide thermal data. The SPOT platform is able to identify the sheen area of the slick due to the wider spectral bands it employs. Optical and thermal satellite imagery offers little potential for oil spill detection while SAR satellites are nowadays the most useful space system for this purpose.

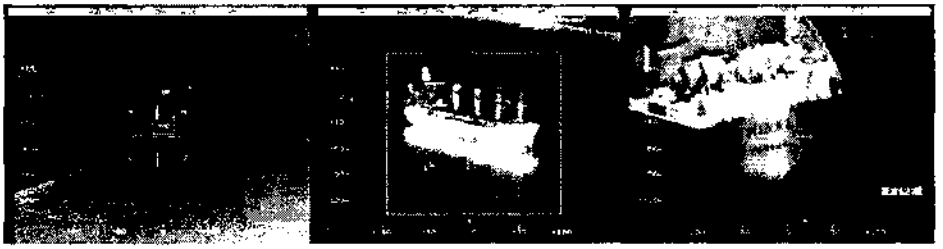


Fig.2. Helicopter approach to a non-moving ship on the Columbia River near Portland, Oregon. ***Left:*** *Approaching the vessel using the thermal camera and the autotracker.* ***Middle:*** *Close-up to medium field of view, still observing the IR image and using the autotracker.* ***Right:*** *Laser-illuminated stern of the ship, supplied by the CCD camera with disabled NIR-cut filter.*

At the end of the Persian Gulf War in the spring of 1991, 732 oil wells were set ablaze in Kuwait. The only source of objective information during the military conflict were the satellite images of the territory of Iraq which provided a means to locate the sources, monitor oil fires during military operations and present visually smoke plumes. Some of those wells were burning for seven, eight, nine months, the oil field fires provided air pollutant of greatest concern, environmental land and marine damage, drilling oil companies losses, Coalition military forces and people health damages. Troop reports of battlefield operations in conditions of heavy smoke and petroleum rain verified the severity of the actual exposures to the oil field fires. Equally important as a toxic exposure was the oil rain from the oil fires [13] .

The total area of Kuwait is covered by four TM scenes of approximately 185 km x 177 km . In Fig.3, oil fires and offshore oil slick are taken from the Landsat-5 Thematic Mapper, the fires appear as red dots. The black smoke plume extends south along the Persian Gulf coast as a strip about 25 kilometers wide in areas near the fires. Unburned pools of oil on the ground and oil offshore in the Persian Gulf reflect sunlight, the same way as water does, and appear as white or light toned features. Oil slicks drifted south toward the Arab Emirate States [16].

This scene of the Persian Gulf was taken hours before the Desert Storm operation began on March 11, 1991 (Fig. 4). A week later, Iraqi soldiers began blowing up Kuwaiti oil wells in response to the multinational offensive. The following image shows the dramatic development of the environmental

catastrophe and the capability of the visible and IR sensors of NOAA Advanced Very High Resolution Radiometer (AVHRR) for early detection and monitoring of oil spills. Thick and thin oil layers and the boundary between water and oil were possible to detect by the IR channel, but the oil spills could not have a significant different temperature signature from the surrounding water at night. Oil spills could be detected in the visible images only under highly favourable lighting and sea conditions [2] .

Fig.3. Color infrared view of the Kuwait oil fires and offshore oil slick taken from the LandSat-5 TM

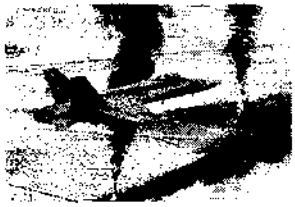


Fig.4. Scene taken fron NOAA-11 AVHRR multispectral color image of the Persian Gulf.

GIS-based risk management system uses the latest spatial information technology to store data required for oil spill risk assessment, response, planning, training and risk management. The integration of remote sensing with GIS techniques offers an effective tool for analysis in the risk management.

Exploration companies are always looking for ways to efficiently and cost-effectively monitor assets, make better-informed decisions and meet environmental guidelines. Oil seeps and oil-impacted soils are often too subtle to detect using multispectral satellite-based sensors because they are diluted on the surface when mixed with other materials and are of limited surface area. By contrast, airborne hyperspectral sensors have sufficient spectral resolution to identify different surface materials based solely on spectral signatures. On land, hyperspectral data has been used to delineate the extent of an oil well blowout.

Most past and current hyperspectral sensors have been airborne, with two recent exceptions: NASA's Hyperion sensor on the EO-1 satellite, and the U.S. Air Force Research Lab's FTHSI sensor on the MightySat II satellite. Several new space-based hyperspectral sensors have been proposed recently. AIG during 1995 using the AVIRIS instrument as a first attempt to broaden the use of hyperspectral data. A hyperspectral AIG/HyVista Group Shoot has been organised in the USA during 1999 using the commercial HyMap sensor [9] .

Oil and gas explorers have remained interested in the technology for many years and companies, such as Shell and ChevronTexaco are known users. Recently, effort towards detection and mapping of offshore hydrocarbon seepage has been made, especially at well-known seeps, such as those off of Santa Barbara, CA, USA. The HyMap scanner provides 126 spectral channels spanning the wavelength range from the 0.4 to 2.5mm (visible to shortwave infrared) spectral region over a 512-pixel swath. HyMap is mounted in a gyro-stabilised platform Zeiss-Jena SM2000 augmented with a Boeing C-MIGITS GPS/INS. Fig.5 (Hy Map) shows seamless data product mosaics of the 4 flight lines imaged some of the world's largest natural oil seeps located just offshore of Santa Barbara, CA covering the area of offshore production platform Holly, with different seep specific spectral components highlighted in the images [9].

A joint research project was organized by Geosat and sponsored by Chevron, Exxon and Shell in 1999. The goal of the project was to determine the viability of hyperspectral technology for detecting oil seeps (Fig. 6) [11] .

The two basic types of microwave radar that can be used to detect oil spills and for environmental remote sensing in general are Side-Looking Airborne Radar (SLAR) and Synthetic Aperture Radar (SAR).

The Side-Looking Airborne Radar (SLAR) is the primary sensor for long-range detection of oil pollution on the sea surface. SLAR sends out short pulses in the X-band perpendicular to the flight direction to left and right side of the aircraft and receives their reflection from small gravity and capillary waves up to a distance of typically 30 km, depending on wind conditions and aircraft altitude [8] .

An airborne surveillance system has been used to monitor the German territorial waters in the North Sea and Baltic Sea for oil discharges and marine pollution. A SLAR unit will cost between $700,000 to $1,000,000.

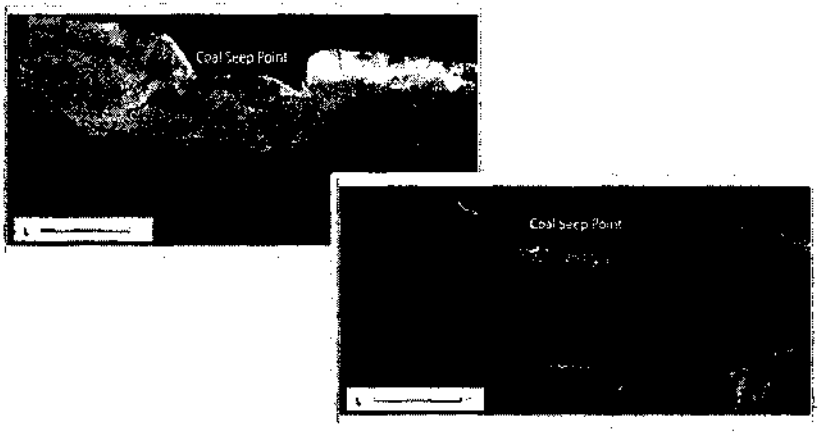


Fig.5. ***Left up*** *- 'False' colour composite mosaic of the survey area highlighting different water constituents using spectral ratios in the visible and near infrared spectral region. The coastal turbid water maps in yellow while darker currents map in red/orange. Open waters are in blue colors with darker tones representing offshore currents. Darkest blue tones represent slick areas.* ***Right down-*** *Spectral component map of slick areas. The land areas, open water and kelp beds have been masked out deliberately and the remaining data spectrally processed to highlight variation.*

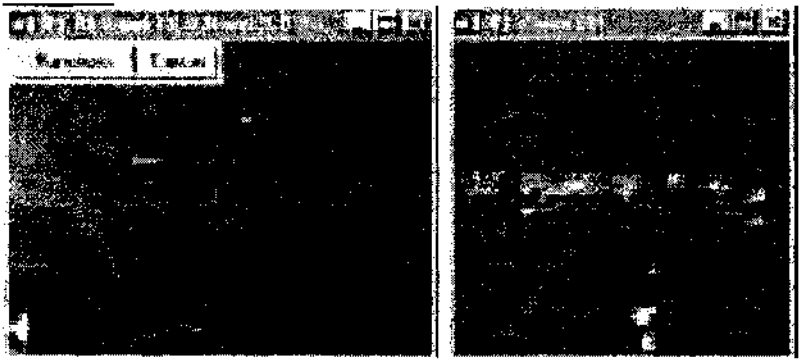


Fig.6 . A wharf where an oil spill was documented in 1992. Hyperspectral imaging detects oil-impacted surfaces (red) along the edge of the wharf.

Satellite Synthetic Aperture Radar (SAR) is shown to be an important tool for oil spill remote sensing because it is the only sensor that can be used over large areas and that can 'see' at night and through clouds or fog. This "all-weather" operation and the wide swath width provided by the available SAR satellites have been the major reasons why satellite-based SAR has been most commonly promoted for operational detection of oil slicks.

There are presently three SAR satellites in orbit with global coverage: RADARSAT, ERS-2 and ENVISAT. These provide revisit times for most places on the globe that are impractically long and irregular for operational sensing of a given spill. RADARSAT has a repeat cycle of 24 days for coverage of a given area of latitude, while ENVISAT and ERS-1 have revisit frequencies of 35 days and have the ability to image surface oil seeps remotely with wide swath coverage (typically 100 x 100km scenes for ERS and 165 x 165 km for Radarsat Wide1) [8].

Monitoring illegal oil discharges is thus an important component in ensuring compliance with marine protection legislation and general protection of the coastal environments. Fast delivery SAR products are proving to be of great value in the optimisation of airborne surveillance resources, due to the large area they can image at any one time. Size, location and dispersement of the oil spill can be determined using this type of imagery.

The Arabian Gulf region is the largest offshore oil development area in the world producing over 27% of the world's oil in 2000. The area also holds 65% of the world's oil reserves. The study area has one of the busiest and most important tanker shipping lanes in the world; more than 40 % of the world's total oil transportation passes through the region. The oil sludge, released by tankers cruising in the Arabian Gulf is estimated to be around 8 million metric tons per year, representing 60 % of the total pollution in the area. The Japan Oil Development Co., Ltd. (JODCO) with the collaboration of the Japanese Information Centre for Petroleum Exploration and Production (ICEP) have supported the Satellite Image Processing project of the offshore waters of the United Arabian Emirates (UAE) [4].

A total of more than 300 frames of ERS-1/2, Landsat-7 ETM, JERS-1 SAR, Terra ASTER, RADARSAT, and Space Shuttle images derived from different platforms that covered most of the offshore water of the UAE have been used (required within the period 1980's and early 2001). Attention has been focused on the area of the offshore from Abu Dhabi to Ajman with analysis using seven ERS-1/2 C-band SAR images including 2 tandem mode

24-hour interval image observed on 24 April, 29 May and 30 May 1996 (**Fig.7**). Observed slicks and bright spots in the 3 images acquired on different dates are confirmed as leakage oil slicks from same oil production platforms. In addition, known well location correspond to the leaking pints as well. Oil platforms and ships are observed as brilliant punctual reflectors in the background generally indistinguishable unless the well locations are superimposed on the images. This detection depends on the sea conditions, incidence angle increasing provide more contrast on the SAR image among the ship and the sea, it is possible to detect the ship trace depending on the size and speed of the ship. A ship can be distinguished by the presence of short, faint lines suggesting that the slicks are ephemeral surface events from a moving source, probably oil spills from a passing vessel moving down from north to south [4].

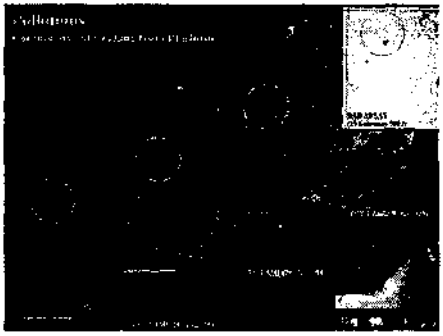


Fig.7. ERS-1/2 SAR images including 2 tandem mode 24-hour interval image observed on 24 April,29 and 30 May1996.

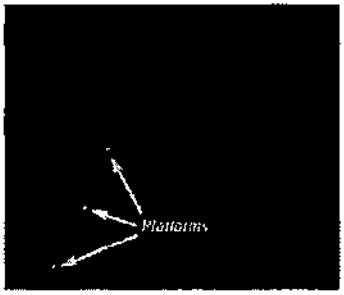


Fig.8. RADARSAT data © Canadian Space Agency 1997. Received by the Canada Centre for Remote Sensing

Three oil drilling platforms can be seen in this RADARSAT image **on** Fig.8 off the northwest coast of Australia. The dark trails in the water in the east and northeast of the scene may indicate surface oil pollution from the production platforms [3].

Influx of oil from tankers and offshore oil operations are major causes of pollution in the marine environment. According to statistics of the US Coat Guard (1990), sources of oil in the sea are classified into 6 categories. By far the highest contributor to oil pollution in the ocean, about 52%, result from a mix of materials and waste which make up urban runoff and discharge from land-based industrial plants. Another 19% of the oil in the sea is directly

attributable to world oil industry. Two per cent of this occurs in spills from rigs and platforms during the exploration and production phases, and only 5 % of oil pollution in the oceans is attributable to accidents involving oil tankers. The remaining 13 % of hydrocarbons in the oceans is absorbed from the atmosphere by particle settlement and rain-wash.

According to the European Space Agency (1998), 45% of the oil pollution comes from operative discharges from ships. When taking into account how frequent such spillages occur, controlled regular oil spills can be a much greater threat to the marine environment and the ecosystem than larger oil spill accidents like the Prestige tanker (carrying N77,000 ton of fuel oil accident at Galice, northwest coast of Spain in 2002) (**Fig.9**) and the Sea Empress supertanker (**Fig.10**) [5] .

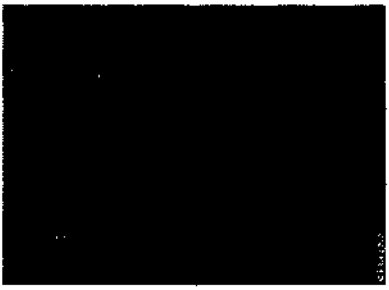


Fig.9. A satellite image of November 20, 2004. This SAR image shows tanker, Prestige, 100 km off the Spanish coast. The ESA 'ENVISAT ASAR was operating in its wide-swath mode covering an area of approximately 400 km by 400 km.

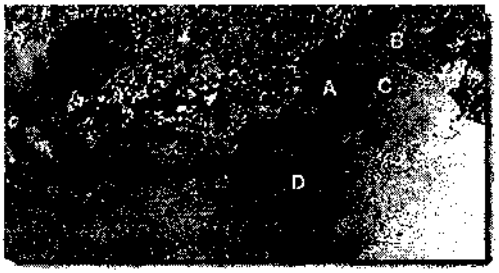


Fig.10. The Sea Empress, a 147,000 ton supertanker accident the south of Wales February 15th, 1996. Seven days later, RADARSAT captured this image, clearly delineating the remaining oil slick. The spill appears on the image in black tones.

In June 2004, the Russian oil company "Lukoil" launched the exploitation of D-6 oilfield on a platform raised on the Russian Baltic shelf. Obviously, the least accident on the oil platform or pipelines connecting platform with coast terminal is capable of causing an ecological disaster in the region. In order to control the situation and to take timely measures and operationally predict the propagation of pollution in the case of disaster, experimental monitoring of the area using satellite remote sensing has been established.

The ESA's ERS-2 and ENVISAT radars conduct remote sensing of the sea surface and send back medium or high resolution images in emergency situations (after a large oil spill has been detected), wide swath (from 100 to 500 km) and keen sensitivity to sea surface roughness.The D-6 platform's location point will appear on the images regularly, 19 times a month, on the average. Most of these spills are produced by ships. No spills caused by leakage from oil exploitation installations have been detected. This is an evidence of the platform's ecological safety guaranteed by modern technologies used in its construction and operation.

RADARSAT International (RSI) has signed a multi-year, multi-satellite oil spill and oil seep monitoring contract with the Marine North East Region (MNE) of PEMEX Exploration and Production (PEP) of Mexico in establishing a maritime surveillance strategy for the Gulf of Mexico, including offshore oil seep exploration and environmental monitoring. PEP produces 3 million barrels of oil and 4,500 billion cubic meters of gas daily, ranking fourth in international production capacity. It will use RADARSAT-1 and ENVISAT imagery as well as RADARSAT-2 imagery when it becomes available. The contract includes imagery acquisition, processing, interpretation and services designed for oil spill and seep detection. Using this multisatellite program, PEP-MNE will be able to maximize imaging opportunities over their area of interest [17].

It is important to recognize that potentially damaging discharges of crude oil or petroleum products can and do occur at every point in the oil production and transportation system. The costs resulting from a spill are numerous and include economic, social, recreational, and ecological losses. The cost of recovering or eliminating oil offshore is typically 10 to 100 times less than removing the same oil from shorelines. On March 24, 1989, the tanker Exxon Valdez grounded on Bligh Reef in Prince William Sound carrying 53 million gallons of crude oil. Federal agency costs for the EXXON VALDEZ 1989 cleanup season alone were $110 million. EXXON's response costs exceeded $2 billion. Union Oil of California spent $13 million to clean up the oil and settle claims resulting from a 6,300 gallon oil spill from a near shore pipeline in south-central California in 1992 [19]. The conclusion is that the potential ecological disaster in the area of oil exploration is very dangerous and the regular monitoring of oilfields based on satellite SAR is very effective means of oil spills primary detection and tracking [6,12].

A number of limitations with Satellite SAR for oil spill detection have been recognised for some time. SAR systems rely on the detection of surface

roughness and wind speed. Detection is difficult or impossible for some oil types. SAR images are also known to return many "false positives" for oil slicks caused by natural phenomena which generate patches of similar appearance. Automatic analysis of SAR images is not applied routinely yet. Several algorithms based on application of different approaches are suggested, realized and tested.The major limitation in the use of satellite remote sensing in the monitoring of oil slicks is the relatively 'poor' temporal resolution of the higher resolution near polar orbiting satellites. SPOT and RADARSAT have the ability to aim its sensor toward a designated area (at low incidence angles RADARSAT cannot differentiate slicks from sea water but the images will require extensive geometrical correction as the images are off nadir). A single satellite cannot provide the necessary cover of the area. Therefore SPOT should be used in conjunction with other satellites, such as Landsat. At present the two SPOT satellites have the ability to provide frequent data coverage of the higher latitudes. The same is true for the current SAR satellites, using the ERS and RADARSAT platforms in tandem it will be possible obtain greater number of slick images for a given period of time.

SAR is the preffered radar technology and a unit cost between $2,000,000 to $4,000,000.

The line scanning microwave radiometer (MWR) enables quantitative assessments of detected oil slicks by analysing the radiant emission from the sea surface and oil slicks at two or three frequencies for avoidance of interference (18.7, 36.5 and 89 GHz with geometrical resolution respectively 22 m, 11 m and 5 m) and allow to determine the thickness and volume for thicker and thin layers. This detection method has not been very successful in the field, however, as several environmental and oil-specific parameters must be known. In addition, the signal return is dependent on oil thickness but in a cyclical fashion [8] .

The microwave scatterometer is a device that measures the scattering of microwave or radar energy by a target. The main disadvantages include the lack of discrimination for oil and the lack of imaging capability.

Airborne Laser Fluorosensor (ALF) [9] seepage detection system is a seep detection system that uses a sophisticated, solid state laser to generate UV light which is pulsed from a low flying aircraft. The laser induces fluorescence in any fresh hydrocarbons on the sea surface. Due to the

differences in the composition, the fluorescence spectra of different oils show variations with respect to the spectral form and the intensity of the fluorescence observed. Flying height is typically 80 m (to 600 m) and flying speed is typically 270 km/h. A surface swath width is 100 m at 300 m flight altitude conical scanner for two-dimential mapping of the sea surface.

The ALF acquisition system performs using excimer laser at 308 nm for the analysis of oil and organic pollutant (MkII system, Barringer's Fluoroscan) and the newest MkIII with NdYAG laser at 266 nm with 176 recorded channels and records fluororsensor data at the 50 Hz acquisition rate.

ALF detects oil sea surface films that may be too thin for satellite or other airborne methods and detects oil spills below the sea surface; estimates the oil volume - quantity and type of oil, identification of the oil through its spectral form; can be used during day and night if the visibility is sufficient and within certain limits is practically independent of the sea state. ALF gives information about hydrocarbon source, charge rates, trap integrity and oil degradation, helping reduce risk in further exploration of the area.

Airborne laser fluorosensor (ALF) survey was flown by World Geoscience in the Timor Sea, Australia in 1997 for BHP Petroleum and joint venture partners using the ALF MkIII system to detect natural oil seepage over the permit in an effort to refine the petroleum prospectivity assessment. The survey covers a triangular area over the northern two thirds of the permit and extends about 60km north to south and nearly 50km east to west, a total of 285 fluorescence anomalies (fluors) are picked out of the 1,751,550 recorded spectra in the final interpretation [9] .

Conclusions

Remote sensing is a critical element for an effective response to marine oil spills. Remote sensing data can provide information to enhance strategic and tactical decision-making, decreasing response costs by rapidly determining oil exact location and ultimately minimizing impacts. Spills detection by visual observation, despite of the low price and common use, is limited to favorable sea and atmospheric conditions, there is no mechanism for positive oil detection. The infrared devices are commercially available at affordable prices, for ship-board and aerial observation of oil slicks. There are airborne

downward-looking thermal IR and Forward-Looking Infrared (FLIR). These sensors can 'see' the oil slick

during day and night, detecting thicker parts of a slick only. This thermal imaging capability brings an important step towards night operation and polluter identification. Satellite sensors like Landsat TM and NOAA AVHRR provide thermal data and oil slicks can be detected but in all cases the oil position should be known in advance. Optical and thermal satellite imagery offers little potential for oil spill detection while SAR satellites are nowadays the most useful space system for this purpose. The Side-Looking Airborne Radar (SLAR) is the primary sensor for long-range detection of oil pollution on the sea surface. Fast delivery Satellite Aperture Radar (SAR) products are proving to be of great value in the optimisation of airborne surveillance resources, due to the large area they can image at any one time. Regular monitoring of oilfields, oil influxes from tankers and offshore oil operations, illegal oil discharges, operative discharges from ships based on satellite SAR is very effective for companies means of oil spills primary detection and tracking and oil spills incidents prevention. As oil becomes harder to find, pursuit of fractured reservoirs and subtle signs of hydrocarbons at the surface and sea surface will receive increased attention. Airborne ALF system and hyperspectral hydrocarbron mapping are the insights of detection oil seeps source on the sea surface because of their high effectiveness and should be used in conjuction with GPS and aircraft inertial systems which allow pinpointing the oil's location and other airborne geo-data sets.

References

1. Brown C., R. Marois, M. Fingas. Airborne oil spill sensor testing: progress and recent developments, Canada, USA, 1997
2. Brekke C., A. Solberg. Oil spill detection by satellite remote sensing, Norwegian Defence Resesarch Establishement, 2004
3. Coastal zone monitoring- Application profile, Radarsat International, Canada, 1999
4. Harahsheha H., T. Nishidaid et all. Operational satellite monitoring and detection for oil spill in offshore of UAE, Global Scan technologies UAE, Japan Oil Development Co., Ltd Japan, 2003
5. Inggs M.R., R.T. Lord. Applications of Satellite Imaging Radar, University of Cape Town, 2001

6. Litovchenko K. Operational monitoring of oil spills in Baltic sea using ENVISAT ASAR, Russian Institute of Space Device Engineering, 2005
7. Moutaz Dalati. Detection of Oil Spills along a part of the Eastern Coast of the Mediterranean Sea, General Organization for Remote Sensing, Syria
8. Panova P., P. Getsov. The airborne remote systems for offshore oil seepage detection, International Conference 2005,Varna, Bulgaria
9. Panova P. Offshore hydrocarbon seep detection with ALF and airborne hyperspectral seep mapping possibilities, International Conference RAST 2005, Istanbul, Turkey (In english)
10. Remote Sensing and Surveillance of Oil Spills, MMS, Offshore Mineral Management, USA 2005
11. RSInc. RSI's ENVI Software Plays Key Role in Oil Seep Detection, Australia
12. Satellites on ecology safety watch, Oil of Russia, Journal N 2, 2005
13. Stead C. Oil fires, petroleum and gulf war illness, USA, 1999
14. Trieschmann O. European Group of Experts on Satellite Monitoring of Sea-based Oil Pollution, Federal Institute of Hydrology, Germany, 2004
15. http://www.geog.ucl.ac.uk/~salmond/essay.html
16. http://eospso.gsfc.nasa.gov/eos_homepage/for_educators/eos_edu_pack/p16.php
17. http://www.spacenewsfeed.co.uk/2001/28October2001.html
18. http://www.uscg.mil/hq/g-m/nmc/gendoc/coop/coop.htm

ИЗПОЛЗВАНЕ НА СПЪТНИКОВИ И САМОЛЕТНИ СИСТЕМИ ЗА ОТКРИВАНЕ НА НЕФТ В МОРСКИТЕ ЗОНИ

П. Панова, П. Гецов

Резюме

Докладът представя анализ на спътникови и самолетни системи за дистанционно търсене на нефт и използване на техните възможности от нефтени и газови компании за проучвателни програми и при

аварийни изтичания и разливи, както и при други извънредни ситуации.

Тези системи взаимно се допълват и спомагат за комплексно наблюдение с различна разделителна пространственна способност и повторяемост на получаваната информация.

Мониторингът от спътници е един от основните компоненти за наблюдение на обширни морски зони, което за другите типове платформи може да се окаже прекалено скъпо, особено за проучвателни цели.

При самолетите пък, на борда може да се монтира допълнително оборудване позволяващо по прецизно откриване на нефт и определяне на количеството и вида на нефтените петна. Така че, само съвместното им използване може да спомогне за постигане на най–добри резултати при големия спектър от задачи, свързани с мониторинга на обширните водни пространства.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

A METHOD FOR INVESTIGATION AND EVALUATION OF THE RELIABILITY OF ELECTRONIC DEVICES DURING THE PROCESS OF DEVELOPMENT AND OPERATION

Nikolay Petrov[1], Boycho Boychev[2]

[1]Thracean University – Stara Zagora
[2]Space Research Institute – Bulgarian Academy of Sciences
e-mail: boytchev@space.bas.bg

Abstract

The requirement for evaluation of the reliability of electronic devices (ED) for every day and special use is connected with new methods for design and construction. This applies most of all to aerospace equioment and sistems. They must satisfy a complex of technological requirements. The most important of them must be restrictions by size, weight, power consumption, high stability of output parameters and high reliability in the process of usage. It should be taken into account that process in question takes place in the condition of a wide range of change in temperate, humidity, pressure, vibrations, in the presence of active disturbances, radiation and random variances of the nominal of the elements of ED, provoked by change in production technology. This paper examines an engineering method for research of the local sub domains for stable work of the circuits in ED. The method is continuation of the method of boundary tests, created in 1968 and suggested in[1].

1. Introduction

Satisfying the requirements for ED is a complex and multisides problem of the optimization and modelling theory. Three basic groups of methods for optimization of ED work areas with respect to decreasing

probability for gradual failures are described in technical literature : the method of statistical tests (normal and speeded) [1; 2], the method of optimization with analysis of intermediate results [2] and the "Monte-Carlo" statistical modelling method [3]. Among the most effective methods from the second group are the Gauss – Zeidel method, the relaxation method, and the gradient method [3]. From the point of view of engineering design, the general disadvantage of the mentioned methods is their complexity, labour-consumption and poor clearness. The treatment of the problem consists in dividing the multidimensional domain of the circuit's stable work into local sub-domains and optimizing the last domains with sufficiently high probability for stable work ($P \geq 0,95$). Thus, the clearness of the analysis of the local sub-domains (two-dimensional or three-dimensional) give the possibility to apply the heuristic approach during the design process.

2. Solution of the problem

The solving of the formulated problem after the proposed method takes place in 4 steps described below.

2.1. Detrmination of the local sub domains of the ED cicuit's stable work

Thus can be done using the system of equations:

(1) $$V_B(t) \leq W_B(t),\ (B = 1,2,.....,k),$$

where: $V_B(t)$ and $W_B(t)$ are generalized functions of the ED parameters $q_1, q_2, ..., q_n$, which include the parameters of the electronic elements, semiconductor depending on one devices and power supply sources common independent variable, the time t.

The functions $V_B(t)$and $W_B(t)$ also depend on the characteristics of the disturbing signals and the exploitation conditions. Conditions (1) are obtained from the analysis of ED's physical model. When the system of equations (1) is functionally incomplete with respect to the parameters $q_i\ (i = 1,2,...,n)$ it is completed based on physical considerations by introducing new conditions or adopting part of the parameters q_i, as determinative. The stable work domain of ED, determinate by system (1)

is multidimensional with respect to the parameters q_i. The analysis of such a multidimensional domain is not very effective in obtaining engineering formulas to determine the optimal values of the parameters q_i. Therefore, it is advisable to divide the steady work domain of ED into local sub-domains.

The nature of the division is in solving the system of equations (1) in such a way as to obtain two relations of the type:

$$(2) \qquad q_m \le F_1(q_r),\ q_m \ge \Phi_1(q_r),\ (m,r \le n,\ m \ne r),$$

where: q_m, q_r are two of the parameters, that are searched for.

Conditions (2) determinate the local sub domain (q_m, q_r) of stable work of ED. Moreover, we must take into account that in relation (2) more than two parameters q_i can be included . In this case, the boundary of the local sub domain depends parametrically on the additional parameters q_i. Taking into consideration the values (q_m, q_r) in the initial system (1), we obtain a second local sub domain.

$$(3) \qquad \frac{q_p}{\overline{q}_m} \le F_2\left(q_s, \overline{q}_m, \overline{q}_r\right), \frac{q_p}{\overline{q}_m} \ge \Phi_2\left(q_s, \overline{q}_m, \overline{q}_r\right), \ldots \left(p,s \le n,\ p \ne s\right),$$

from the analysis of which we defines the mean values $\overline{q}_m, \overline{q}_r$ of another two parameters of the ED.

Using the values $\overline{q}_m, \overline{q}_r, \overline{q}_p, \overline{q}_s$ obtained from the initial system (1), we determine the next local sub-domain and etc. until we obtain the mean values of all parameters q_i. The values $\overline{q}_i$ must be chosen taking into account the possible dispersion of the quantities q_i and the coefficients, which define the functions $F_1, F_2, \ldots, \Phi_1, \Phi_2, \ldots$.

2.2. Stochastic analysis of the local sub domain of stable work of the ED

When taking into account the dispersion of the quantities q_i and the coefficients which determine the functions $F_1, F_2, \ldots, \Phi_1, \Phi_2, \ldots$ each of the

relations of a type (2), (3) having the following general appearance:

(4) $\mu_j(t) \le \nu_j(t),\ (j = 1,2,3,....,l),\ (l \le n/2)$

must be replaced by an additionally specified relation of the following type:

(5) $\bar{\mu}_j(t) \le \bar{\nu}_j(t).\left[1 + Z_j(t)\right],$

where the dimensionless random quantity $Z_j(t)$ determines the influence of the destabilizing factorswith time t, on the output characteristic $Z(t)$ of the ED, i.e. its relative error.

In the general case, the dimensionless random quantity $Z_j(t)$ is determinated by the equation:

(6) $Z_j(t) = \frac{\Delta Z}{\bar{Z}},$

where: ΔZ is the displacement of the output characteristics $Z(t)$ of the ED from its nominal value $\bar{Z}(t)$ caused by the influence of the destabilizating factors with time t.

We assume, that the displacement of the basic perameters q_i of the ED are small (this condition is satisfied with correct by true design and normal exploitation of the ED), i.e.:

(7) $\Delta q_i << \bar{q}_i,$

and $\Delta\nu_j(t) << \bar{\nu}_j(t),\ \Delta\mu_j << \bar{\mu}_j$ we represent the relative error $Z_j(t)$ of the functions $L_j = \nu_j/\mu_j$ in the form:

(8) $Z_j(t) = \frac{\Delta L_j}{\bar{L}_j} \approx \frac{\Delta\nu_j}{\bar{\nu}_j} - \frac{\Delta\mu_j}{\bar{\mu}_j}.$

Taking into account relation (6), we represent $Z_j(t)$ in the following form:

$$(9) \qquad Z_j(t) = \sum_{i=1}^{n} a_{ij} \cdot (\Delta q_i / \overline{q}_i),$$

where: $a_{ij} \approx \dfrac{\partial L_j(q_1, q_2, \ldots, q_n)}{\partial q_i} \cdot \dfrac{\overline{q}_i}{\overline{L}_j(\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_n)}$, and the derivatives $\partial L_j / \partial q_i$ are determinated for the points $(\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_n)$.

Equation (9) is the error equation of the error of the output characteristics of the ED which determines the relation of the relative displacements $Z_j(t)$ at the boundaries of the determined local sub-domains with the relative displacements $\Delta q_i / \overline{q}_i$ of the parameters of the ED.

Based on the theoretical investigations made in [4], equations (9) can be transformed in the following form:

$$(10) \qquad \frac{\Delta Z}{\overline{Z}} = \sum_{i=1}^{n} \left(\frac{\Delta Z}{\overline{Z}} \right)_{q_i} \cong \sum_{i=1}^{n} a_{i1} \cdot \left(1 + m_i \cdot \frac{\Delta q_i}{\overline{q}_i} \right)^{r_i} \cdot \frac{\Delta q_i}{\overline{q}_i}.$$

Relation (10) is an expanded interpretation of the equation of the "small" relative errors from (9), which is characterized by the existence of a "partial" (by „ i ") influence coefficient, defined by the multiplier, accounting for the non-linear terms of the second and higher order.

2.3. A priory determination of the parameters of the partial series from equation (10)

This determination is based on the use of the influence of the artificial by formed change of the *i*-th **basic parameter** (BP) Δq_i, on **the output characteristics** (OC) $(\Delta Z)_{q_i}$ at constant values of the other BPs.

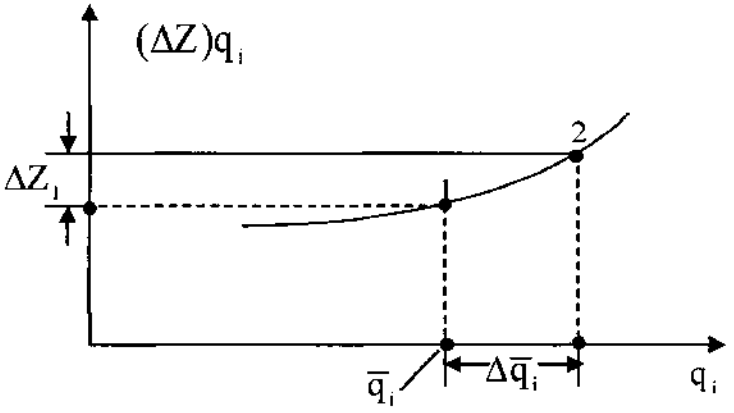


Fig. 1. A priory determination of the parameters of the partial series from (10)

That is accomplished by using the following initial data and procedures:

1. For BP q_i, two a priory points are used with coordinates - p.1 $\left[\overline{q}_i, (\Delta Z)_{q_i}\right]$, p.2 $\left[(\overline{q}_i + \Delta\overline{q}_i), \Delta\overline{Z}_1\right]$, which are shown in the Fig. 1, where $\overline{q}_i, \overline{Z}$ are the known nominal of the i-th BP and the OC; $\Delta\overline{q}_i, \Delta\overline{Z}_1$ are the given changes of the i- th BP and the measured change of ΔZ.

2. For p.1 we determine according by the derivative $(\partial Z/\partial q_i)$ and the value of the coefficient $\overline{a}_{i1} = \left[(\partial Z/\partial q_i)_{\overline{q}_i}\right].(\overline{q}_i/\overline{Z})$.

3. Along the coordinates of p.1 (the end of the a priory range, beginning of the a posteriori range of prediction) $(\Delta Z_1, \Delta\overline{q}_i)$ we determine the parameter $\overline{m}_i$ according to:

$$\overline{m}_i = \frac{\Delta\overline{Z}_1/\Delta\overline{q}_i}{(\partial Z/\partial q_i)_{\overline{q}_i}}. \quad (11)$$

4. From relation (9) presented in suitable a priory form:

$\frac{\Delta\bar{Z}_1}{\bar{Z}} = \bar{a}_{i1} \cdot \frac{\Delta\bar{q}_i}{\bar{q}_i} \cdot \left(1 + \bar{m}_i \cdot \frac{\Delta\bar{q}_i}{\bar{q}_i}\right)^{r_i}$, we obtain after appropriate transformations:

$$(12) \qquad \bar{r}_i = \frac{\lg \frac{(\Delta\bar{Z}_1/\bar{Z})\bar{q}_i}{\bar{a}_{i1}.\Delta\bar{q}_i}}{\lg[1 + \bar{m}_i.(\Delta\bar{q}_i/\bar{q}_i)]}.$$

2.4. Prediction of the relative displacement (error) of the output characteristics of the ED in the case of its complex dependence on individual BPs

During stable work regimes of the ED, the following is valid for formulas (7) to (10):

$(\Delta q_i/\bar{q}_i) << 1$, $(\Delta Z/\bar{Z}) << (a_{i1}.\Delta q_i/\bar{q}_i)$, $|r_i| < 1$. From these conditions we can transform (10) in the following form:

$$(13) \qquad \frac{\Delta Z}{\bar{Z}} = \sum_{i=1}^{n}\left(\frac{\Delta Z}{\bar{Z}}\right)_{q_i} \cong \sum_{i=1}^{n} a_{i1}.\left(1 + m_i.\bar{r}_i \cdot \frac{\Delta q_i}{\bar{q}_i}\right).\frac{\Delta q_i}{\bar{q}_i}.$$

Equation (13) is "*expanded equation of the accidental relative error*", which gives the probability for predicting the relative error ξ_{OC} of the OC of the ED with complex change of the BP$_i$ $(i = 1,...,n)$ which determine it out side the range $\Delta\bar{q}_i$ for $|\Delta q_i| > |\Delta\bar{q}_i|$ according to:

$$(14) \qquad \xi_{OC} = \frac{\Delta Z}{\bar{Z}} \cong \sum_{i=1}^{n} \bar{a}_{i1} \cdot \left(1 + \bar{m}_i.\bar{r}_i \cdot \frac{\Delta q_i}{\bar{q}_i}\right) \cdot \frac{\Delta q_i}{\bar{q}_i}.$$

As a matter of principle, equation (14) is a relation of relative accidental errors, which depends on a great number of non-dominating factors with normal distribution. It is important to note, that in the common case the distribution of the process of change in ξ_{OC}, as a result of its non-linear character is a multiplication of Veybul distributions [4].

With $\left[\overline{m}_i.\overline{r}_i.(\Delta q_i/\overline{q}_i)<<1\right]$ and in the general case condition $(x/x^2)\cong M(x)/M(x^2)\cong\sigma_X/\sigma_X^2$ where is satisfied for the mathematical expectation $M(\Delta q_i/\overline{q}_i)$ and the dispersion $D(\Delta q_i/\overline{q}_i)$ of the individual BP_i, the mathematical expectation and the dispersion of the OC of ED $(\Delta Z/\overline{Z})$ can be determined, according to:

$$M\left(\frac{\Delta Z}{\overline{Z}}\right)\cong\sum_{i=1}^{n}\overline{a}_{i1}.\left[1+\overline{m}_i.\overline{r}_i.M\left(\frac{\Delta q_i}{\overline{q}_i}\right)\right].M\left(\frac{\Delta q_i}{\overline{q}_i}\right), \tag{15}$$

$$D\left(\frac{\Delta Z}{\overline{Z}}\right)\cong\sum_{i=1}^{n}\overline{a}_{i1}^2.\left[1+\overline{m}_i^2.\overline{r}_i^2.D\left(\frac{\Delta q_i}{\overline{q}_i}\right)\right].D\left(\frac{\Delta q_i}{\overline{q}_i}\right). \tag{16}$$

3. Case study of an electronic circuit

As a subject of the case study, the circuit of a generator based on the integrated circuit NE 555 - electronic timer [5], is chosen. The circuit and its elements are shown in Fig. 2.

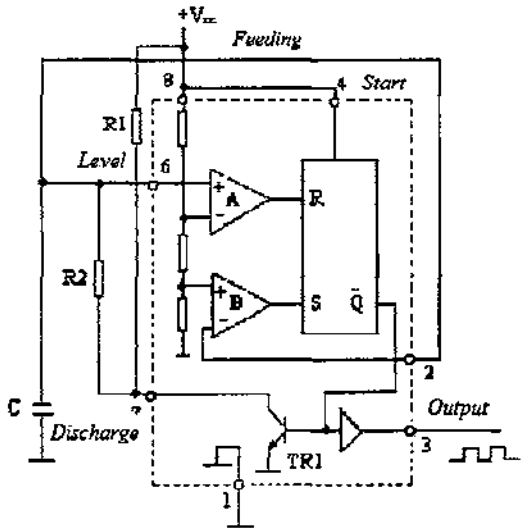


Fig. 2. Electronic circuit of a generator based on the timer NE 555

The investigation was carried out on the basis of equations (15) and (16), which were transformed in the following form:

$$(17)\quad M\left(\frac{\Delta f}{\bar{f}}\right) \cong \sum_{i=1}^{n} a_{i1}.\left[1+\bar{m}_i.\bar{r}_i.M\left(\frac{\Delta C}{\bar{C}}\right)\right].M\left(\frac{\Delta C}{\bar{C}}\right),$$

$$(18)\quad D\left(\frac{\Delta f}{\bar{f}}\right) \cong \sum_{i=1}^{n} a_{i1}.\left[1+\bar{m}_i.\bar{r}_i.D\left(\frac{\Delta C}{\bar{C}}\right)\right].D\left(\frac{\Delta C}{\bar{C}}\right),$$

where: $\bar{f}$ - the mean (nominal) frequency of the generated electrical oscillation at the output of the timer NE 555 $(\bar{f}=1\,kHz)$; Δf - the displacement of the timer frequency from $\bar{f}$; $\bar{C}$ - the mean (nominal) value of the capacitor C from Fig. 2 $(\bar{C}=1\,nF)$; ΔC - the displacement of the value of capacitor C from the its mean (nominal) value $\bar{C}$; $R_1=5\,k\Omega$; $R_2=470\,k\Omega$.

In Fig. 3 and Fig. 4 the graphic simulation investigations of equations (17) and (18) using the "*Mathematics*" program are shown, investigating the functional relations:

$$(19)\quad M\left(\Delta f/\bar{f}\right)=F\left[M\left(\Delta C/\bar{C}\right)\right],$$

$$(20)\quad D\left(\Delta f/\bar{f}\right)=F\left[D\left(\Delta C/\bar{C}\right)\right].$$

Functional relations (19) and (20) are investigated under the conditions of the experiments with the timer from Fig. 2, whereas with a change of the capacitor value C by $\Delta C=100\,pF$ we have $\Delta f=10\,Hz$. Under this condition, from equations (9), (10) and (11), the following relations for mathematical expectation M and dispersion D were obtained:

$$(21)\quad M\left(\frac{\Delta f}{10^3}\right)=\left[1+10^{20}.\Delta C.\lg(\Delta f)\right].M\left(\frac{\Delta C}{10^{-9}}\right),$$

$$(22)\quad D\left(\frac{\Delta f}{10^3}\right)=10^{12}.\left[1+10^{40}.\lg^2(\Delta f)\right].D\left(\frac{\Delta C}{10^{-9}}\right),$$

where M is mathematical expectation and D is dispersion.

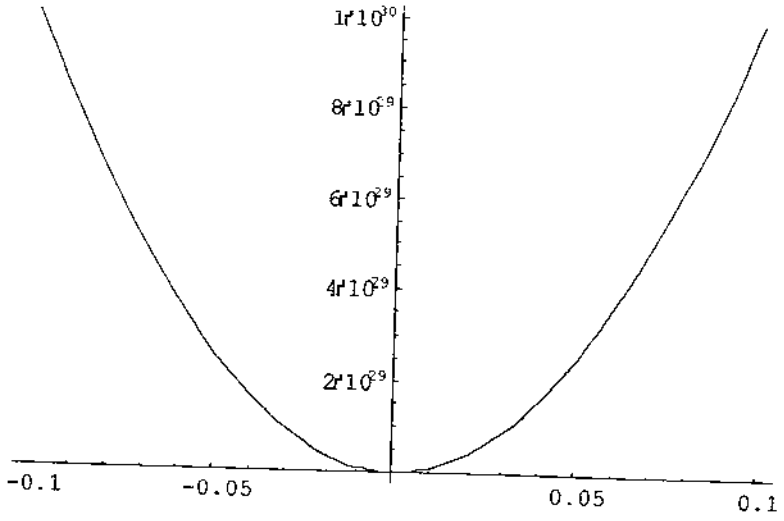


Fig.3. Graphical representation of the function

$$M\left(\frac{\Delta f}{10^3}\right)=F\left[M\left(\frac{\Delta C}{10^{-9}}\right)\right]$$

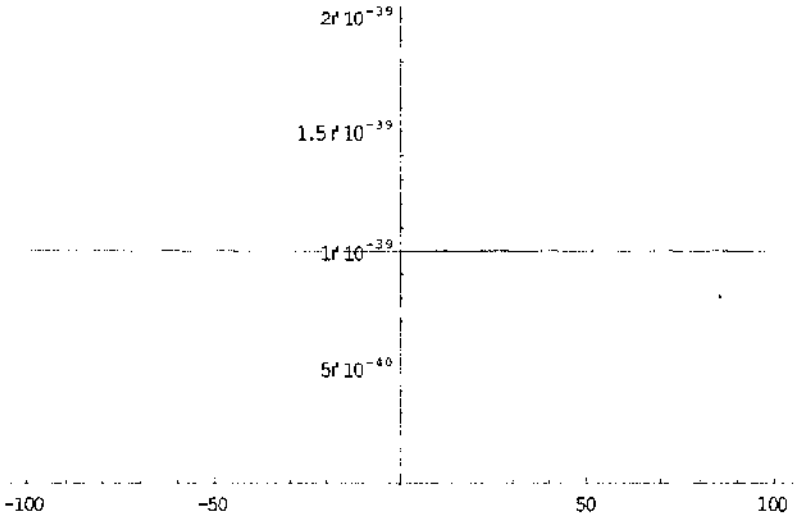


Fig.4. Graphical representation of the function

$$D\left(\frac{\Delta f}{10^{3}}\right)=F\left[D\left(\frac{\Delta C}{10^{-9}}\right)\right]$$

4. Conclusions

On the basis of the results from this investigation, we can formulate the following conclusions:

1. Equations for mathematical expectation and dispersion of the output characteristics of electronic devices for every day and special use have been derived. From the analysis of these equations the technological conclusion for the high reliability to external influence on the test circuit in the connection regime recommended by the producer has been made.

2. On this basis simulation case investigation of an electronic circuit (electronic timer) has been carried out. The simulation investigation evidences of the good symmetry of function $M\left(\Delta f/\bar{f}\right)$ and stability of function $D\left(\Delta f/\bar{f}\right)$.

3. Based on the type of the graph of function $M\left(\Delta f/\bar{f}\right)$, it can be ascertained that the probability for generation of the nominal specified by the producer frequency of the output pulses is maximum in the domain of the solution of this function. The uncertainty of the dispersion of the generated frequency is uniform and symmetrical at deviation from the nominal value. All investigation have been made only for a change of the capacity of capacitor C from the connection shown in diagram Fig.2, whereas the influence of resistors R_1 and R_2 has not been accounted for.

References

1. Фомин А. В., Ю. Т. Обичкин. Надеждность полупроводниковых радиоустройств летательных аппаратов. Изд. "Машиностроение", Москва, 1968.
2. Маслов А. Я., В. Ю. Татарский. Повышение надеждности радиоелектронной аппаратурый. Изд. "Сов. Радио", Москва, 1972.
3. Гиндев Е. Г. Надеждност на сложни системи. Институт по специална оптика и радиоелектроника. София, 1976.
4. Петров Н. И. Експлоатационна надеждност на рискови технически системи. Изд. къща "Учков", Ямбол, ISBN 954-9978-26-5, 2002.
5. Tooley M. Practical Digital Electronics Handbook. PC Publishing, London, 1988.

МЕТОД ЗА ИЗСЛЕДВАНЕ И ОЦЕНКА НА НАДЕЖДНОСТТА НА ЕЛЕКТРОННИ АПАРАТУРИ В ПРОЦЕСА НА РАЗРАБОТКА И ЕКСПЛОАТАЦИЯ

Н. Петров, Б. Бойчев

Резюме

Изискването за оценка на надеждността на електронните апаратури (ЕА) с битово и специално предназначение изисква принципно нови методи за проектиране и конструиране. Това е особено актуално за аерокосмическите апарати и системи. Те трябва да удовлетворяват комплекс

от тактико-технически и технологични изисквания. Най-важните от тях следва да бъдат ограничения по отношение на габарити, тегло, консумирана и разсейвана мощност, висока стабилност на изходните параметри и висока надеждност в процеса на експлоатация. Трябва да се има предвид, че въпросният процес се извършва в условията на широк диапазон на изменение на температурата, влажността, налягането, вибрациите, наличието на активни смущения, радиация и неизбежни случайни вариации на номиналите на елементите на ЕА предизвикани от изменението на технологията на производство. В настоящата статия се разглежда инженерен метод за изследване на локални подобласти за устойчива работа на схемите в ЕА. Методът се явява продължение на метода на граничните изпитания създаден през 1968 г. и предложен в [1].

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

PULSE EXPANSION OF A THERMOELASTIC VISCOPLASTIC CYLINDRICAL SHELL BY INTERNAL PRESSURE

Viktor Baranov[1], Ivan Getsov[2]

[1]Tula State University, Russia
[2]Vazov Engineering Plants, Sopot, Bulgaria.
e-mail: igecov@abv.bg

Abstract

A conjugate problem of predicting the kinematic and thermal parameters as well as the parameters of the stressed-strained state of a cylindrical shell made of thermoelastic viscoplastic material and loaded by an internal pressure pulse has been solved in terms of waves.

1. Introduction

The subject of this research is a thin-wall cylindrical shell made of elastic viscoplastic material of the Kristesku-Malvern-Sokolovski type [1], with the latter's indicial equation describing the properties typical for most metals – strain and dynamic hardening, as well as thermal weakening in the process of loading, with the latter being caused by dissipation of inelastic deformation energy into thermal energy, having taken into account the heat transfer attending the process. The outer and inner radii of the shell have been designated R_0 and r_0, accordingly. At a time moment $t = 0$, a radial pressure pulse $p = p(t)$, with $p(0) \neq 0$, has been applied to the shell's internal surface. The pulse's amplitude parameters are such, so as to take the shell out of the elastic state.

2. Problem positing

The solution has been carried out in a cylindrical coordinate system *(R, θ, z)* within the framework of the hypothesis of planar stressed state. The stress tensor's components are σ_{RR} and $\sigma_{\theta\theta}$ which are different from zero (hereinafter referred to as σ_R and σ_θ, accordingly), and the strain tensor's components are ε_{RR}, $\varepsilon_{\theta\theta}$ and ε_{ZZ} (ε_R, ε_θ, and ε_Z, accordingly).

The equation of motion of a circular element of the shell with radius $R \in [r_0; R_0]$ and thickness dR has been expressed as follows:

$$\rho \frac{\partial V_R}{\partial t} = \frac{\partial \sigma_R}{\partial R} + \frac{\sigma_R - \sigma_\theta}{R} \tag{1}$$

where V_R –radial velosity; - material density

The relative strains ε_R and ε_θ and velocity V_R are interrelated by conditions for continuity:

$$\frac{\partial \varepsilon_R}{\partial t} = \frac{\partial V_R}{\partial R}; \frac{\partial \varepsilon_\theta}{\partial t} = \frac{V_R}{R} \tag{2}$$

The constitutive equations for the shell material have been assumed in the form [2]:

$$\frac{\partial \varepsilon_R}{\partial t} = \frac{1}{E\left(\varepsilon_i^p, T\right)}\left(\frac{\partial \sigma_R}{\partial t} - \mu \frac{\partial \sigma_\theta}{\partial t}\right) + \left[\frac{\sqrt{J_2} - \sigma\left(\varepsilon_i\right)}{\beta\left(\varepsilon_i\right)}\right]^{\alpha\left(\varepsilon_i\right)} *$$

$$* \frac{2\sigma_R - \sigma_\theta}{6\sqrt{J_2}}\left[1 - \frac{1}{E^2\left(\varepsilon_i^p, T\right)} \frac{\partial E\left(\varepsilon_i^p, T\right)}{\partial \varepsilon_i^p}\left(\sigma_R - \mu\sigma_\theta\right)\right];$$

$$\frac{\partial \varepsilon_\theta}{\partial t} = \frac{1}{E\left(\varepsilon_i^p, T\right)}\left(\frac{\partial \sigma_\theta}{\partial t} - \mu \frac{\partial \sigma_R}{\partial t}\right) + \left[\frac{\sqrt{J_2} - \sigma\left(\varepsilon_i\right)}{\beta\left(\varepsilon_i\right)}\right]^{\alpha\left(\varepsilon_i\right)} *$$

$$* \frac{2\sigma_\theta - \sigma_R}{6\sqrt{J_2}}\left[1 - \frac{1}{E^2\left(\varepsilon_i^p, T\right)} \frac{\partial E\left(\varepsilon_i^p, T\right)}{\partial \varepsilon_i^p}\left(\sigma_\theta - \mu\sigma_R\right)\right]; \quad (3)$$

$$\frac{\partial \varepsilon_z}{\partial t} = -\frac{\mu}{E\left(\varepsilon_i^p, T\right)}\left(\sigma_R + \sigma_\theta\right) - \left[\frac{\sqrt{J_2} - \sigma\left(\varepsilon_i\right)}{\beta\left(\varepsilon_i\right)}\right]^{\alpha\left(\varepsilon_i\right)} *$$

$$* \frac{\sigma_R + \sigma_\theta}{6\sqrt{J_2}}\left[1 - \frac{1}{E^2\left(\varepsilon_i^p, T\right)} \frac{\partial E\left(\varepsilon_i^p, T\right)}{\partial \varepsilon_i^p}\left(\sigma_\theta + \sigma_R\right)\right],$$

where I_2 – invariant of the stress deviator tensor;

ε_i – strain intensity;
$\varepsilon^p{}_i$ – reduced plastic deformation;
μ – Poisson's ratio;

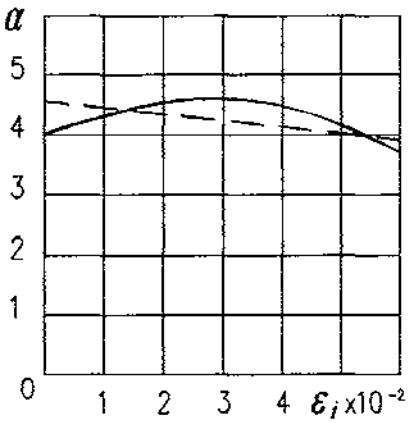


Fig. 1. Approximation of relation $\alpha = \alpha(\varepsilon_i)$ *for steel 3*

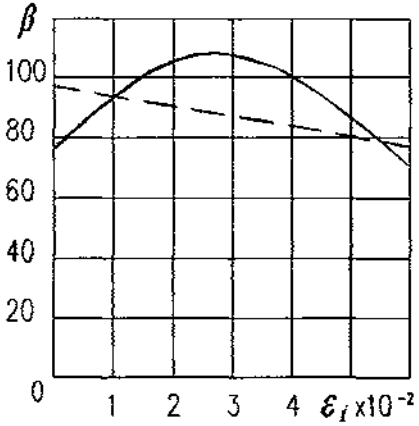


Fig. 2. Approximation of relation $\beta = \beta(\varepsilon_i)$ *for steel 3*

The approximating functions $\alpha = \alpha(\varepsilon_i)$, $\beta = \beta(\varepsilon_i)$ and $\sigma = \sigma(\varepsilon_i)$ have been expressed with reference to steel 3 in the form [2]:

(4) $$\alpha = 4{,}5 - 7\varepsilon_i \; ; \quad \beta = 98{,}1 - 232\varepsilon_i ; \quad \sigma = 241 + 2196\varepsilon_i \text{ , [MPa].}$$

Figures 1..., 3 illustrate the satisfactory matching between the results from the experiment (solid lines) and theoretical (design) relations (4) (dash lines). The relation between the modulus of elasticity and plastic deformation value has been expressed by means of the exponential function in the form [3]:

The boundary conditions of the problem have been assumed in the following form

(5) $$\sigma_R(r_0,t)=\sigma_{Ro}H(t);\sigma_R(R_0,t)=0,$$

and the initial conditions –

(6) $$\sigma_R(r_0,0)=\sigma_\theta(r_0,0)=V_R(r_0,0)=\varepsilon_R(r_0,0)=\varepsilon_\theta(r_0,0)=\varepsilon_z(r_0,0)=0$$

where $H(t)$ – Heaviside single function.

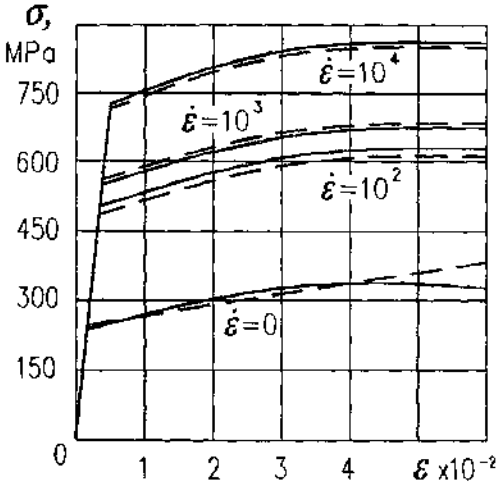


Fig. 3. Approximation of dynamic load diagrams for steel 3 by means of constitutive equations (3), (4)

The system of equations (1),…,(3) is quasi-linear of the hyperbolic type. It has been solved numerically by the characteristics method [4] with the subsequent finite-difference approximation of the differential relations between the sought functions along the characteristic directions.

The latter are in the following form:

- along the divisible characteristics $dR = 0$:

$$d\varepsilon_\theta = \frac{V_R}{R}dt;$$

$$d\varepsilon_\theta = \frac{d\sigma_\theta - \mu d\sigma_R}{E} + \left[\frac{\sqrt{J_2} - \sigma(\varepsilon_i)}{\beta(\varepsilon_i)}\right]^{\alpha(\varepsilon_i)} *$$

$$* \frac{2\sigma_\theta - \sigma_R}{6\sqrt{J_2}} \left[1 - \frac{1}{E^2}\frac{\partial E}{\partial \varepsilon_i^p}\left(\sigma_\theta - \mu\sigma_R\right)\right] dt;$$

$$d\varepsilon_R = \frac{d\sigma_R - \mu d\sigma_\theta}{E} + \left[\frac{\sqrt{J_2} - \sigma(\varepsilon_i)}{\beta(\varepsilon_i)}\right]^{\alpha(\varepsilon_i)} *$$

$$* \frac{2\sigma_R - \sigma_{\theta R}}{6\sqrt{J_2}} \left[1 - \frac{1}{E^2}\frac{\partial E}{\partial \varepsilon_i^p}\left(\sigma_R - \mu\sigma_\theta\right)\right] dt;$$

$$d\varepsilon_Z = -\mu\frac{d\sigma_\theta + \mu d\sigma_R}{E} - \left[\frac{\sqrt{J_2} - \sigma(\varepsilon_i)}{\beta(\varepsilon_i)}\right]^{\alpha(\varepsilon_i)} *$$

$$* \frac{2\sigma_R + \sigma_\theta}{6\sqrt{J_2}} \left[1 + \frac{1}{E^2}\frac{\partial E}{\partial \varepsilon_i^p}\left(\sigma_R - \sigma_\theta\right)\right] dt; \qquad (7)$$

- along the characteristics $dR = \pm\sqrt{\frac{E}{\rho(1-\mu^2)}}dt$:

$$\mp dV_R + \sqrt{\frac{1-\mu^2}{E\rho}}d\sigma_R + \frac{\sigma_R - \sigma_\theta}{R\rho}dt \mp \sqrt{\frac{E}{\rho(1-\mu^2)}}\mu\frac{V_R}{R}dt +$$

(8) $$+\sqrt{\frac{E}{\rho(1-\mu^2)}}\left[\frac{\sqrt{J_2}-\sigma(\varepsilon_i)}{\beta(\varepsilon_i)}\right]^{\alpha(\varepsilon_i)}(A+B)dt=0.$$

where

$$A=\frac{2\sigma_R-\sigma_\theta}{6\sqrt{J_2}}\left[1-\frac{1}{E^2}\frac{\partial E}{\partial \varepsilon_i^p}\left(\sigma_R-\mu\sigma_\theta\right)\right];$$

$$B=\mu\frac{2\sigma_\theta-\sigma_R}{6J_2}\left[1-\frac{1}{E^2}\frac{\partial E}{\partial \varepsilon_i^p}\left(\sigma_\theta-\mu\sigma_R\right)\right]$$

Integration of the system of six ordinary differential equations (7), (8) along the curvilinear characteristic directions has been carried out numerically. Moreover, at each integration step, a cycle of iterations to ε^p_i has been organized and temperature T has risen as a result of the dissipation of inelastic deformation energy into thermal energy. Furthermore, heat redistribution in the shell due to heat transfer has been taken into consideration by means of the heat flow equation in the form [5]:

(9) $$\frac{\partial T}{\partial t}+\frac{a}{\omega^2}\frac{\partial^2 T}{\partial t^2}=a\left(\frac{\partial^2 T}{\partial R^2}+\frac{1}{R}\frac{\partial T}{\partial R}\right)$$

in which the term $\frac{a}{\omega^2}\frac{\partial^2 T}{\partial t^2}$ takes into account the finiteness of the rate of heat transfer; a – coefficient of temperature conductivity. In calculations, the rate of heat flow has been assumed to be ω= 1800 m/s (divisible by the velocity of propagation of radial pressure wave in the shell material which for steel 3 is equal to c = 5400 m/s).

It is convenient to solve equation (9) by the characteristics method, too. The characteristic directions and differential relations between the sought functions

$$v = \frac{\acute{I}\grave{D}}{\acute{I}\eth} \qquad u = \frac{\acute{I}\grave{D}}{\acute{I}\grave{B}}$$

along them are in the following form:

$$dR = \pm\omega dt;$$

$$\pm du - \frac{1}{\omega} dV \pm \frac{\omega^2}{a} udt + \omega \frac{u}{R} dt = 0. \tag{10}$$

In Fig. 4, it is plotted in the phase plane R, t, the combined grid of characteristics for a conjugate solution of the system of equations (7), (8) and (10). The selection of a numerical integration step for equations (10) with relation to the integration step of equations (7), (8) is secondary, assuming that the characteristics lines (10) pass through the node points of the characteristics of equations (7), (8). At the initial iteration step in the current node point, the characteristics of the stressed-strained and kinematic state have been determined, as well as the temperature caused by inelastic deformation of the material, the temperature being driven to the given point as a result of heat transfer. The temperatures have been then summarized, and according to the summary temperature the values of the six sought functions σ_R, σ_θ, V_R, ε_R, ε_θ and ε_Z have been defined more precisely, then the values of temperature T and modulus of elasticity E have been defined more precisely – and the iteration cycle has been repeated until the required precision is achieved. The heat interaction of the shell with the environment has been assumed to be lacking.

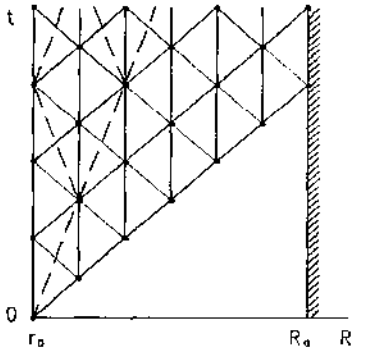


Fig. 4. Combined grid of characteristics in a phase plane of independent variables

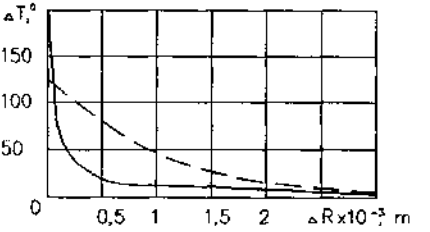


Fig. 5. Effect of heat transfer on temperature distribution in the shell at a fixed time moment

3. Conclusion and deductions

Some results of the calculations are illustrated in Fig. 5,..., 8. In Fig. 5, snapshots of temperature distribution through the shell thickness at a fixed time moment of $t = 3.3 \times 10^{-6}$ s have been plotted. The problem has been solved for boundary conditions (5) at $\sigma_{R0} = 350$ MPa, $r_0 = 0.05$ m, $R_0 = 0.06$ m. The effect of heat transfer in the process of loading has been analyzed. The solid line in the figure has been plotted without taking into account the heat transfer ($a = 0$), and the dash one – with taking into account the heat transfer at $a = 0.024$ m^2/s. Apparently, when taking into account the heat transfer in the shell dynamic loading, it affects materially the current temperature distribution, moreover, the loaded shell surface heating decreases.

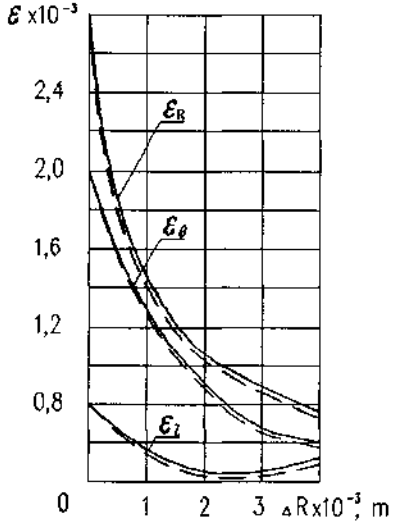


Fig. 6. Snapshots of deformation components for $t = 1.5 \times 10^{-5}$ s

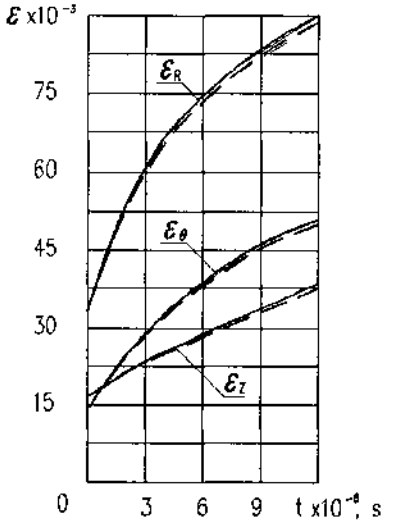


Fig. 7. Variation of deformation components on boundary surface

The variation of deformations and radial stress through the shell thickness and in time is shown in Fig. 6,...,8. Snapshots showing deformation components for $t = 1.5 \times 10^{-5}$ s are shown in Fig. 6, the variation of deformations in time on the loaded surface $R = r_0$ is shown in Fig. 7, and the variation of radial stresses through the shell thickness at a time moment $t = 1.5 \times 10^{-5}$ s is shown in Fig. 8. All curves have been plotted at $a = 0$, the impact of temperature and reduced plastic deformation on the variation of the material's modulus of elasticity has been analyzed. The solid lines in the figures correspond to $E = const$, and the dash ones – to the variable modulus of elasticity. Apparently, when taking into account the shell material heating, it does not affect materially the variation of the stress-strained state parameters. Obviously, the picture will change materially in the cases, when loading of the shell's inner surface will be attended by its pulse heating by an external independent source, which is typical, for example, for the cases of gun firing or loading the shell's inner surface by products of detonation of high explosives. In these cases it should be expected (as confirmed by the carried out calculations, too) that, as a result

of the intensive heating of the shell's inner surface, the inner layers of its material will be in a compressed state, including the layers along the tangential components of stress and strain, thus increasing artificially the safety factor of the shell and changing the time-sequence of appearance of microdefects and destruction through the shell space in the process of its expansion: destruction begins not from its inner surface, as it follows from the solution of the well-known Lame's problem, but in the middle layers, in the area where tangential stresses change their sign.

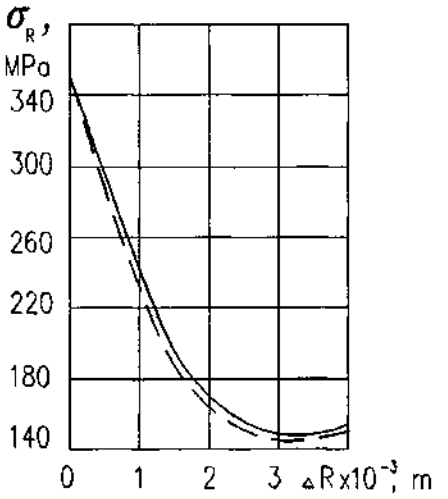


Fig. 8. Snapshots showing profiles of radial stresses waves in the shell for $t = 1.5 \times 10^{-5}$ s

Tooling has been thereby developed that allows to predict the behavior of cylindrical shells under conditions of pulse loading on the inner surface as well as to solve a series of problems of practical importance.

References

1. P e z h i n a P. Basic Problems of Viscoplasticity, M., Mir, 1968, p.176.
2. B a r a n o v V., C h r i s t o v Ch., G e c o w I. Radial Wave Expansion of the Thick Wall Homogeneous Cylindrical Shell, Journal of Materials Science and Technology, Sofia, Bulgarian Academy of Science, 2002, Vol.10, № 1, p.17-24.

3. B a r a n o v V. L. Longitudinal Waves in Elastic Viscoplastic Rods with Variable Modulus of Elasticity, Works on Mechanics of Deformable Rigid Body, Tula, 1981, p.142-150.
4. B e r y o z i n I. S. , Z h i d k o v N. P. Methods of Calculation, Vol.2, Moscow: State Publishing Office for Physics & Mathematics Publications, 1962, p. 620.
5. L y k o v A. V. Theory of Heat Conduction., Moscow: State Publishing Office for Technical & Theoretical Publications. 1952, p. 392.

ИМПУЛСНО РАЗШИРЕНИЕ НА ЦИЛИНДРИЧНА ОБЛИЦОВКА ОТ ТЕРМОЕЛАСТИЧЕН - ВЯЗКОПЛАСТИЧЕН МАТЕРИАЛ ПОД ДЕЙСТВИЕТО НА ВЪТРЕШНО НАЛЯГАНЕ

В.Баранов, И.Гецов

Резюме

Решава се вълнова задача за определяне на кинематичните параметри и параметрите на напрегнато - деформираното състояние на цилиндрицната облицовка от термоелестичен-вязкопластичен материал, натоварена от импулс на вътрешно налягане.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

A METHOD OF PHASE-MANIPULATED COMPLEMENTARY SIGNALS APPLIED IN SPACECRAFT-BASED RADARS

Borislav Bedzhev, Zhaneta Tasheva, Rosen Bogdanov

National Military University
e-mail: bedzhev@mail.pv-ma.bg

Abstract

Radar imagery, realized by means of synthetic aperture radars (SARs) is very important in exploring planet, satellite and comet surfaces. The most valuable feature of the autocorrelation function (ACF) of the SAR signals is the level of their side lobes, because they determine the dynamic range of the image and the possibility to discover small objects. With regard to this, our paper suggests a method for applying in spacecraft-based SARs the so named generalized complementary signals, whose ACF does not have any side-lobes. It uses the polarization features of electromagnetic waves.

1. Introduction

Radar imagery is very important in exploring planet, satellite and comet surfaces [1, 2]. It may be sketched as follows. The transmitter of the spacecraft-based radar sends electromagnetic signals.

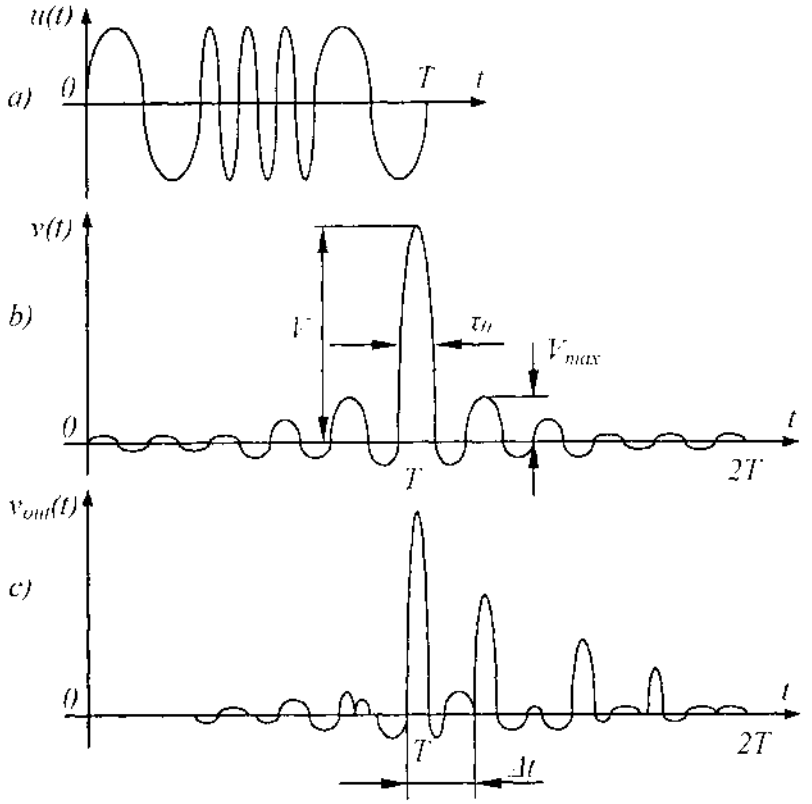


Fig.1. Processing of radar signals

The examined objects reflect the signals, producing so named echo-signals. They are the input to the radar receiver. Usually, in order to maximize the "signal/noise" ratio, the receiver is constructed as a filter, matched to the sent signals. In this case, the receiver's output is the autocorrelation function (*ACF*) of the sent signals. This is clarified in Fig. 1, where a radar signal is shown (Fig. 1a). The duration of the signal is T, but it is separated in n sub-signals (or "elementary signals") with duration τ_0 (i.e. $n = T / \tau_0$) and different carrier frequency. This technique is named "discrete frequency shift-keying" (*DFSK*).

It allows obtaining a different echo-signal from every "reflected point" of the object. Usually, the receiver's output, produced by a single point echo, is characterized by a main peak V and a sequence of side-lobes with maximal amplitude V_{max}, as shown in Fig. 1b. At the end, the radar receiver output signals are sampled and processed to extract the object image [3].

In general, the above-described technique of using complex radar signals provides for both large performance range (provided by the aggregated power of elementary signals) and high distance resolution (defined by τ_0) of the spacecraft-based synthetic aperture radars (*SAR*s). Unfortunately, real objects comprise more than one reflected points. As a result, the echo-signals of all reflected points interfere, as shown in Fig.1c. In this situation, it is hard to obtain a detailed object image, because the side-lobes of the more powerful signals mask the main peaks of the weak signals.

With regard to this, our paper aims to suggest a method for applying the so named generalized complementary signals, whose *ACF* does not have any side-lobes. It uses the polarization features of electromagnetic waves.

2. A Method of Phase-Manipulated Complementary Signals Applied in Spacecraft-Based Radars

It is known [4] that discrete phase- and frequency-modulated signals may be presented as the real part of the complex-valued function:

$$(1) \qquad V(t) = \sum_{j=1}^{n} \{U_j . \exp(i\theta_j) . \exp[2\pi i(f_0 + f_j)t]\} . u_0(t - j\tau_0),$$

where $i = \sqrt{-1}$; U_j is the amplitude of the j[th] elementary pulse $j = 1,2,...,n$; f_0 is the carrier frequency; $\{f_1, f_2, ..., f_n\}$ are real time functions, which express the frequency modulation; $\{\theta_j;\ 0 \le \theta_j < 2\pi;\ j = 1,2,...n\}$ is the set of numbers, describing the phase modulation and:

(2) $$u_0(t)=\begin{cases} 1, \ if \ \ 0\le t\le \tau_0 \\ 0, \ if \ \ t<0, \ or \ \ t>\tau_0 \end{cases}.$$

To maximize transmitter efficiency and to simplify the practical accomplishment of the process of signal receiving, the so-named uniform signals with $\tau_0 = const;\ U_j = const;\ j = 1,2,...,n;; \theta_j \in \{(2\pi l)/m;\ l = 0,1,...,m-1\}$ are most widely applied.

In this case and if only discrete phase shift keying (*DPSK*) is applied, the signal is named "discrete phase-manipulated (*PM*) signal". It can be described precisely by the sequence $\{\zeta(j)\}_{j=1}^{n}$ of normalized complex amplitudes of elementary signals [4]:

(3) $$\zeta(j)=\exp(i.\theta_j);\ \ \zeta(j)\in\{\exp(2\pi il/m); l=0,1,...,m-1\}.$$

As mentioned above, the signals whose *ACF* has close-to-zero level of the side-lobes, are the most attractive for implementation in spacecraft based *SAR*s. With this regard, in the rest part of the paper our attention shall be focused on the so-named generalized complementary signals, whose *ACF* is free of any side-lobes.

It is known that a single radar signal does not have non-periodical *ACF* with zero level of the side-lobes. Moreover, the classes of single uniform discrete radar signals with small level of their side-lobes seem to be very rare. For this reason, Golay introduced [5] the so-named complementary series (or signals (*CS*s)). They are a pair of two uniform binary phase-manipulated signals, whose aggregated non-periodical *ACF* is similar to a delta pulse.

It is necessary to emphasize that Golay's definition of *CS*s is not useful in some important cases. This situation has motivated some theoreticians to extend the classical definition as follows [6, 7, 8].

Definition 1: *The set of p sequences (PM signals), whose elements are complex numbers, belonging to the multiplicative group of the m-th ($m > 2$) roots of unity:*

$$\{A_1 = \{\xi_1(j)\}_{j=1}^{n_1}; A_2 = \{\xi_2(j)\}_{j=1}^{n_2}; ...; A_p = \{\xi_p(j)\}_{j=1}^{n_p}\};$$
(4)
$$\xi_k(j) \in \{\exp(2\pi i l / m_k); l = 0,1,...,m_k - 1\}; k = 1,2,...,p.$$

are ***a set of generalized complementary signals*** *(GCCs) if and only if their aggregated ACF has ideal shape, similar to a delta pulse:*

$$R_c(r) = \sum_{k=1}^{p} R_{A_k}(r) = \begin{cases} n = n_1 + n_2 + ... + n_p; & if \quad r = 0; \\ 0; & if \quad r = 1,2,..., \max\{n_k\}. \end{cases} \tag{5}$$

In (5) the non-periodical *ACF* $R_{A_k}(r)$ are defined by the well known formula [4]:

$$R_\zeta(r) = \begin{cases} \sum_{j=1}^{n-|r|} \zeta(j)\zeta^*(j+|r|), & -(n-1) \le r \le 0 \\ \sum_{j=1}^{n-r} \zeta^*(j)\zeta(j+k), & 0 \le r \le n-1. \end{cases} \tag{6}$$

Consequently, Golay's codes are a particular case of the *GCCs*, when $p = 2$, $m = 2$. The *CCs* and *GCCs* are unique among all *PM* signals for the following features:

- their aggregated *ACF* has ideal shape, similar to a delta pulse;
- if a pair of *GCCs*, consisting *of n* elements, is known, then it is easy to create an infinite set of pairs with unlimited code-length.

It should be emphasize that most types of uniform *PM* signals with close to ideal *ACF* have limited code-length. For instance, Barker codes exist only for $n \le 13$, if n is an odd integer.

With regard to the *GCCs*' positive features, they are studied very intensively and a quick reference revealed more than 200 conference reports and magazine articles related to this theme during the past ten years.

The natural question, which arises from Definition 1, is "How can the *GCC*s be implemented in a real communication system?". The most obvious answer is to use p different frequency carriers $f_k, k = 1,2,...,p$, phase manipulated according to the sequences $A_k, k = 1,2,...,p$. Unfortunately, this is not the best approach when the communication system is a spacecraft-based *SAR*. This conclusion will be clarified by the following example. Let us assume that a spacecraft-based *SAR* exploits *GCC*s with $p = 2$ and the transmitter radiates simultaneously two uniform *PM* signals with carriers f_1, f_2, manipulated according to the sequences A_1, A_2. As a result of the so-named Doppler effect, the carriers f_{ek} of the echo-signals will be:

$$(7) \qquad f_{ek} = f_k \frac{1 - V_R / c}{1 + V_R / c} \approx f_k (1 - 2V_R / c), \; k = 1,2,$$

where V_R is the radial velocity of the spacecraft relatively to the object and c is the velocity of the electromagnetic waves propagation. If the explored object is on the earth's surface, then V_R must be at least 27 360 km/h. Then the difference $\Delta f = |f_{e1} - f_{e2}|$ will be too significant and it may lead to irreparable phase distortions between the components of the *GCC*s.

For the above reason, in the rest part of our report we shall prove a more appropriate approach to *GCC*s implementation in spacecraft-based *SAR*s. Namely, we propose the two uniform *PM* signals, composing a pair of *GCC*s, to be transmitted simultaneously on one frequency carrier f_0 but by means of different types of polarization. Let the horizontal and the vertical polarized *PM* signals be described by the sequences $A_1 = \{\mu(j)\}_{j=1}^{n}$ and $A_2 = \{\eta(j)\}_{j=1}^{n}$, respectively. Then the signals, reflected by an object

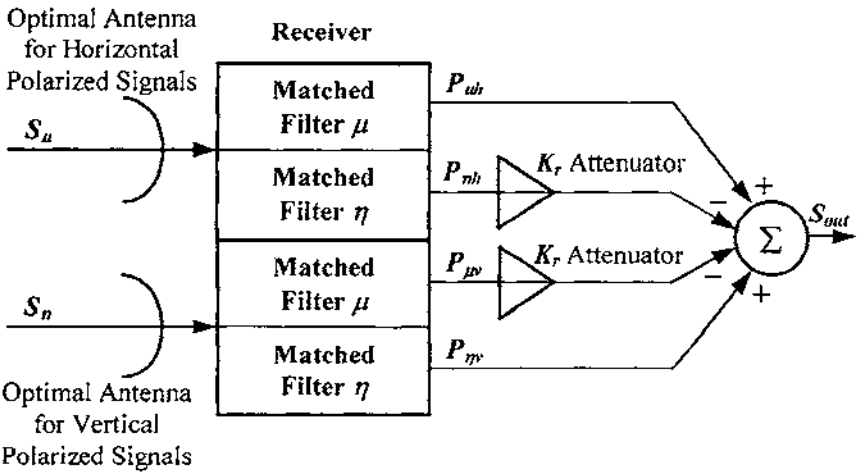


Fig. 2. Method of GCCs applied in spacecraft-based SARs

point, will be S_μ, S_η respectively. The reflected signals are connected with the transmitted signals by the following matrix equation:

$$(8) \qquad \begin{Vmatrix} S_\mu \\ S_\eta \end{Vmatrix} = \begin{Vmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{Vmatrix} \begin{Vmatrix} A_1 \\ A_2 \end{Vmatrix}; \quad \|D\| = \begin{Vmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{Vmatrix},$$

where the complex valued matrix $\|D\|$ is the so-named "polarization matrix of the target scattering". Its entries depend on the physical features of the object, its orientation and position relatively to the radar and the carrier frequency of transmitted signals. When these parameters are constant, then the entries of the matrix $\|D\|$ are also constants and more over, $D_{12} = D_{21}$. Accounting that the size of a single reflected point is small, it may be concluded that $D_{11} \approx D_{22}$. Consequently, the matrix can be presented in the form:

(9) $$\|D\| = \begin{Vmatrix} k_1 & k_1.k_2 \\ k_1.k_2 & k_1 \end{Vmatrix},$$

where $D_{11} = D_{22} = k_1$; $(D_{12}/k_1) = (D_{21}/k_1) = k_2$.

As stated above, the main obstacle for *GCC*s practical implementation by means of polarized electromagnetic waves is the fact that each echo-wave comprises both horizontal and vertical polarized components. For this reason, we propose the method of signal processing, shown in Fig. 2. We shall explain it using the following notation. Let $\{\zeta(j)\}_{j=1}^{n}$ be the sequence of normalized complex amplitudes of elementary signals, composing an arbitrary complex signal. As mentioned, the result of processing this signal by its matched filter will be the *ACF* of the signal. It may be presented by the following polynomial:

(10) $$P(x) = F(x).F^*(x^{-1}).$$

Here:

(11) $$F_\zeta(x) = \zeta(n).x^{n-1} + \zeta(n-1).x^{n-2} + ... + \zeta(2).x + \zeta(1),$$

is the so-named "Hall polynomial", corresponding to the sequence $\{\zeta(j)\}_{j=1}^{n}$, $F_\zeta^*(x^{-1})$ is the polynomial:

(12) $$F_\zeta^*(x^{-1}) = \zeta^*(n).x^{-(n-1)} + \zeta^*(n-1).x^{-(n-2)} + ... + \zeta^*(1),$$

the coefficients of the polynomial $P(x)$ are:

$$p_k = R_\zeta(k);\ k = -(n-1), -(n-2), ... -1, 0, 1, ..., n-2, n-1,$$

and the values $R_\zeta(k)$ of the *ACF* are computed according to (6).

Accounting for the above notation, (8) and (9), the outputs of the matched filters shown in Fig. 2 may be expressed by the following polynomials:

$$P_{\mu h}(x) = k_1[F_\mu(x) + k_2.F_\eta(x)].F_\mu^*(x^{-1});$$

$$P_{\eta h}(x) = k_1[F_\mu(x) + k_2.F_\eta(x)].[k_r F_\eta^*(x^{-1})];$$

(13) $$P_{\eta v}(x) = k_1[k_2.F_\mu(x) + F_\eta(x)].F_\eta^*(x^{-1});$$

$$P_{\mu v}(x) = k_1[k_2.F_\mu(x) + F_\eta(x)].[k_r F_\mu^*(x^{-1})].$$

In (13), k_r is a special coefficient, brought into the scheme by means of two directed attenuators. Now it is easy to see, that if the attenuators in Fig.2 are regulated to obtain $k_r = k_2$, then:

$$S_{out}(x) = P_{\mu h}(x) - P_{\eta h}(x) + P_{\eta v}(x) - P_{\mu v}(x) =$$

$$= k_1\{[F_\mu(x)F_\mu^*(x^{-1}) + F_\eta(x)F_\mu^*(x^{-1})] -$$

(14)

$$- k_2^2.[F_\mu(x)F_\mu^*(x^{-1}) + F_\eta(x)F_\mu^*(x^{-1})]\} = k_1.2n.(1 - k_2^2).$$

Formula (14) shows that the method of *GCC*s use, shown in Fig. 2, preserves the cancellation of the *ACF* side-lobes, despite the harmful presence of cross-reflected signals. The signal power losses depend on the relative coefficient of the cross-polarized reflection k_2.

4. Conclusions

The method of *GCC*s applied in spacecraft-based radars, presented above, preserves the positive features of the *GCC*s, especially the cancellation of the *ACF* side-lobes, despite the harmful presence of cross-reflected signals. This result is achieved by small losses of signal power, because mostly $k_2 \le 20\%$ and hence $(1 - k_2^2) \ge 96\%$.

References

1. Гецов П., Ж. Жеков, Г. Мардиросян, И. Христов, Ефективност на визирни оптични системи при наблюдение на отдалечени обекти при различна яркост на фона, Сб. научни трудове от НС на ВВУАПВО, ч. 2, Шумен, 1997г., с. 243-249.

2. Жеков Ж., Г. Мардиросян, И. Христов, Д. Иванова. Абсорбционый ультрафиолетовый озонометр, Бълг. геофиз. сп. 1998, № 3 – 4, с. 50 – 54
3. Лазаров, А.Д. Радиовълни, антеннофидерни и микровълнови устройства, БСУ, 2003.
4. Варакин Л. Е. Системы связи с шумоподобными сигналами - М.: Радио и связь, 1985. – 384 с.
5. Golay M. Y. E. Complementary series, IRE Trans. on Information Theory, 1961, vol. IT – 7, №2, pp. 82 – 87.
6. Tseng C. C., S. L. Liu. Complementary sets of sequences, IEEE Trans. on Information Theory, 1972, vol. IT- 18, №5, p. 644 – 652.
7. Игнатов В. В., С. А. Добровольский, А. Ю. Гужва. Матричные системы сигналов для использования в системах CDMA, Электросвязь, 2003, № 9, с. 41-42.
8. Frank L.R. Polyphase Complementary Codes, IEEE Trans. on Inform. Theory, 1980, vol. IT-26, № 6, pp. 641 – 647.

МЕТОД ЗА ПРИЛАГАНЕ НА ФАЗОВО МАНИПУЛИРАНИ КОМПЛЕМЕНТАРНИ СИГНАЛИ В КОСМИЧЕСКИТЕ РАДИОЛОКАЦИОННИ СИСТЕМИ

Б. Беджев, Ж. Ташева, Р. Богданов

Резюме

Получаването на радарни изображения на повърхността на планетите, спътниците и кометите е важен момент при тяхното изучаване. В този процес най-важното свойство на автокорелационната функция (АКФ) на радиолокационните сигнали е нивото на страничните листи на АКФ, защото то определя динамичния диапазон на изображението и възможността за откриване на малоразмерни обекти. По тази причина в статията се обосновава метод за използване на така наречените обобщени комплементарни сигнали (ОКС), чиято сумарна АКФ няма странични листи. Методът се характеризира с това, че запазва ценните свойства на ОКС въпреки ефекта на кръстосано поляризционно отражение на радарните сигнали. Този положителен резултат се постига технически просто и с минимални загуби на енергия на ехо-сигналите.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

SOME STATISTICAL PROPERTIES OF SHRINKING–MULTIPLEXING GENERATOR

Zhaneta Tasheva

National Military University
e-mail: tashevi86@yahoo.com

Abstract

The binary additive stream ciphers are one of the cryptographic systems that are often used in aerospace and communication applications where high speed and low delay are requirements.

The need for software-oriented stream ciphers has lead to several alternative proposals in the last few years. A new Pseudo Random Number Generator (PRNG) named Shrinking–Multiplexing Generator (SMG) is described in this paper. It uses p-1 parallel working Linear Feedback Shift Registers (LFSRs) with large period and one p-adic Feedback with Carry Shift Registers (FCSR), which control nonlinearity in the generator.

To be suitable for use in cryptographic systems, PRNG must have large period, large linear complexity and good statistical properties. The results from statistical analysis of SMG are given in the paper. The sequence generated by SMG is uniform, scalable, uncompressible, with large period; consistent and unpredictable are shown by analysis of proposed PRNG. This gives us reason to say that the SMG is suitable for particular cryptographic application.

I. Introduction

The binary additive stream ciphers are one of the cryptographic systems that are often used in applications where high speed and low delay are requirements like modern aerospace and communication information systems.

In the binary additive stream cheaper, the keystream, the plaintext, and the ciphertext are sequences of binary digits. The keystream is generated by a keystream generator, which takes a secret key as a seed, and produces a long pseudorandom sequence. The ciphertext is generated by bitwise modulo 2 additions of the keystream and the plaintext.

The goal of stream cipher design is to efficiently produce pseudorandom sequences that in all senses are "indistinguishable" from truly random sequences. To do this, the pseudorandom sequence must have various properties, such as high linear span, high pair-wise Hamming distance, low cross-correlation value, large linear complexity and good statistical properties.

The need for software-oriented stream ciphers has lead to several alternative proposals in the last few years. To satisfy this need we propose a new Pseudo Random Number Generator (*PRNG*) and give its statistical analysis.

The paper is organized as follows. A new Shrinking–Multiplexing Generator (*SMG*) is described in Section II. The statistical analysis of the proposed *SMG* is given in Section III. Finally, some future works and conclusions are presented.

II. The Shrinking–Multiplexing Generator

In this section we give a brief review of the architecture and mathematical background of the new Shrinking–Multiplexing Generator (*SMG*) proposed by us in [8].

We generalized the architecture of the Shrinking generator [1] by using a p–adic control *FCSR R* (Fig. 1) instead of the control *LFSR* R_1. Thise increases a number of used slave *LFSRs* from 1 in Shrinking generator to $p-1$ in *SMG*. The control register *LFSR* R_1 is used to select a portion of the output sequences of a used *LFSRs* $R_1 \div R_{p-1}$.

Definition 1: A *Shrinking–Multiplexing Generator* (*SMG*) (Fig. 1) consists of a control *FCSR* of length L and of $p-1$ slaved *LFSRs* of length L_1, L_2, ..., L_{p-1}, each capable to produce one bit in its output; and a clock which controls the movement of data in registers. The algorithm of *SMG* with control p-adic *FCSR* consists of the following steps:

1. All *LFSRs* from R_1 to R_{p-1} and *FCSR R* are clocked.

2. If the p-adic output $b_1 = j$ of control register R is not equal to 0, the output bit of register R_j forms part of the keystream.

3. Otherwise, if the output $b_i = 0$ of control register R is equal to 0, the output bits of all slave registers R_1 to R_{p-1} discards.

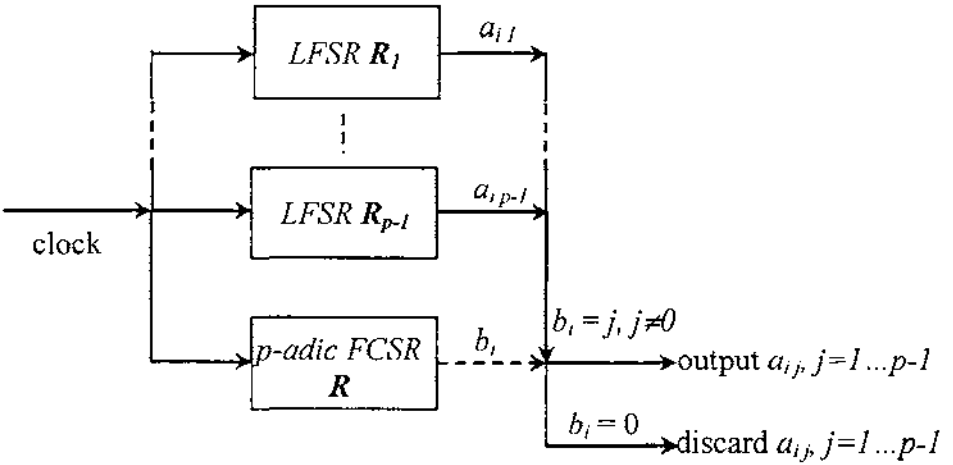


Fig 1. The Shrinking–Multiplexing Generator

Therefore, the produced keystream is a *shrunken* version of the output sequences of *LFSRs* R_1 to R_{p-1} when there is zero output of control *FCSR*, and a *mixed* version of the output sequences of *LFSRs* R_1 to R_{p-1} when output is not equal to zero.

The proposed *SMG* uses the generalization of 2-adic *FCSRs* [3] with stage contents and feedback coefficients in $\mathbf{Z}/(p)$ where p is a prime number, not necessarily 2.

Definition 2: A *p-adic feedback with carry shift register* (*FCSR*) with Galois architecture of length L (Fig. 3) consists of L *stages* (or *delay elements*) numbered *0, 1, ..., L-1*, each capable to store one p-adic (0, 1, ..., p-1) number and having one input and one output; and a clock which controls the movement of data. During each clock cycle the following operations are performed:

1. The content of stage 0 is output and forms part of the *output sequence*;
2. The sum modulo p after stage i is passed to stage i - *1* for each i, $1 \leq i \leq L-1$;

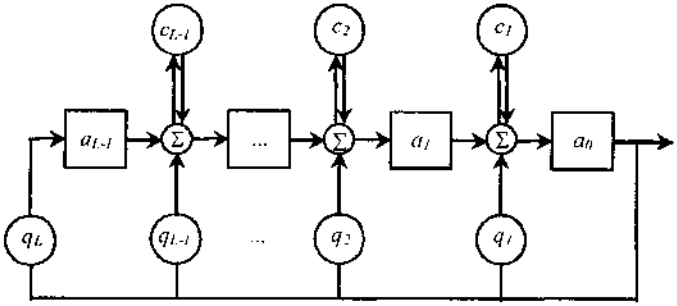


Fig. 2. Galois FCSR

3. The output of the last stage 0 is introduced into each of the tapped cells simultaneously, where it is fully added (with carry) to the contents of the preceding stages. The $q_1, q_2, \ldots, q_L$ are the *feedback multipliers* and the cells denoted by $c_1, c_2, \ldots, c_{L-1}$ are the *memory* (or "*carry*") bits. If

(1) $$q = -1 + q_1 p + q_2 p^2 + \ldots + q_L p^L$$

is the base p expansion of a positive integer:

(2) $$q \equiv -1 \pmod{p},$$

then q is a connection integer for a *FCSR* with feedback coefficients $q_1, q_2, \ldots, q_L$ in $\mathbf{Z}/(p)$.

With each clock cycle, the integer sums:

(3) $$\sigma_j = a_j + a_0 q_j + c_j$$

is accumulated.

At the next clock cycle this sum modulo p

(4) $$a'_{j-1} = \sigma_n \pmod{p}$$

is passed on to the next stage in the register, and the new memory values are:

(5) $$c'_j = \sigma_n \ (\mathrm{div}\, p).$$

The proposed *SMG* uses the *LFSRs* with stage contents and feedback coefficients in Extended Galois Field $GF(p^r)$. There are two architectures for *LFSRs*, the Galois and Fibonacci. We use the Galois one and next we give its definition [3].

Definition 3: A *linear feedback shift register* (*LFSR*) with Galois architecture of length L (Fig. 2) consists of L *stages* (or *delay elements*) numbered *0, 1, ..., L- 1*, each capable to store one bit and having one input and one output; and a clock which controls the movement of data. During each clock cycle the following operations are performed:

1. The content of stage 0 is output and forms part of the *output sequence*;
2. The content of stage i is moved to stage i - *1* for each i, $1 \le i \le L"1$;
3. The output of the last stage 0 is introduced into each of the tapped cells simultaneously, where it is added (modulo 2) to the contents of the preceding stages. If $q_1, q_2, ..., q_L$ are the *feedback multipliers* then the recurrence equations are as follows:

$$(6) \qquad a'_i = a_{i+1} + q_{i+1} a_0, \quad for\ 0 \le i \le L-2$$

$$a'_{L-1} = q_L a_0$$

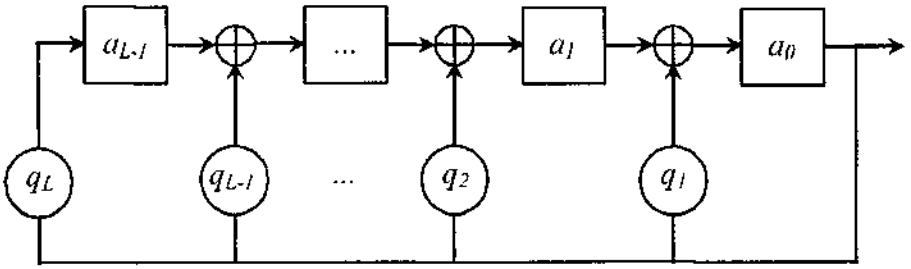


Fig. 3. Galois LFSR

Suppose a Galois *LFSR* with connection polynomial *q(X)*

$$(7) \qquad q(X) = -1 + \sum_{i=1}^{L} q_i X^i$$

has initial loading $(a_0, a_1, ..., a_{L"1})$. Then the output sequence $\boldsymbol{b} = (b_0, b_1, ...)$ of the *LFSR* is the coefficient sequence of the function

$$(8) \qquad B(X) = -\frac{h(X)}{q(X)},$$

where

(9) $h(x) = a_0 + a_1 X + \ldots + a_{L-1} X^{L-1}$.

Conversely, if $\boldsymbol{b}$ is any strictly periodic sequence and let $B(X) = -\dfrac{h(X)}{q(X)}$ be its generating function, then $q(X)$ is a connection polynomial of a Galois *LFSR* which generates the sequence $\boldsymbol{b}$, and $h(X)$ determines the initial loading by (4).

III. Statistical Properties of Shrinking–Multiplexing Generator

Generators suitable for use in cryptographic applications need to meet stronger requirements than for other applications. In particular, their outputs must be unpredictable, i.e. random. To test the statistical properties of *SMG* we have used some procedures proposed by the National Institute of Standards and Technology (*NIST*) to be useful in detecting deviations of a binary sequence from randomness. These tests show as a first step whether or not a generator is suitable for a particular cryptographic application.

The *NIST* statistical tests [5, 6] are formulated to test a specific null hypothesis (H_0): "The sequence being tested is random". Associated with this null hypothesis the alternative hypothesis (H_a) is that the sequence is not random. For each applied test a decision is derived that accepts or rejects the null hypothesis, i.e., whether the generator is (or is not) producing random values, based on the sequence that was produced.

For each test a statistical *P-value* is computed, which is a function of the data. For these NIST tests, each *P-value* [5] is the probability that a perfect random number generator would have produced a sequence less random than the sequence that was tested, given the kind of non-randomness assessed by the test. If a *P-value* for a test is determined to be equal to 1, then the sequence appears to have perfect randomness. A *P-value* of zero indicates that the sequence appears to be completely nonrandom. A significance level (α) can be chosen for the tests. If *P-value* $\geq \alpha$, then the null hypothesis is accepted; i.e., the sequence appears to be random. If *P-value* $< \alpha$, then the null hypothesis is rejected; i.e., the sequence appears to be non-random. The parameter α denotes the probability of the Type I error. Typically, α is chosen in the range [0.001, 0.01]. In our test we chose $\alpha = 0.01$. This indicates that one would

expect 1 sequence in 100 sequences to be rejected. A $P\text{-}value \geq 0.01$ means that the sequence would be considered to be random with a confidence of 99 %. A $P\text{-}value < 0.01$ means that the conclusion was that the sequence is non-random with a confidence of 99 %.

Two types of errors are supposed by statistical hypothesis testing (Table 1. [5]). The conclusion, Type I error, occurs when the data is, in truth, random, and the test rejects the null hypothesis H_0 (the data is random). The Type II error occurs if the data is, in truth, non-random, and a conclusion to accept the null hypothesis H_0 is made.

Table 1. Possible statistical test outcomes

TRUE SITUATION	**CONCLUSION**	
	Accept H_0	***Accept Ha (reject H_0)***
Data is random (H_0 is true)	No error	Type I error
Data is not random (H_a is true)	Type II error	No error

To determine the randomness of arbitrarily long binary sequences produced by *SMG* we used the *NIST* Test Suite statistical package. It consists of 16 functional tests. Some of them are decomposable into a variety of subtests. Here we describe briefly the purpose of the tests and the characteristics that they detect. More information about the mathematical background of the tests the reader may be found in [2, 5, 6]. The 16 tests are:

1. ***The Frequency (Monobit) Test.*** The purpose of this test is to determine whether the number of ones and zeros in a sequence is approximately the same as would be expected for a truly random sequence. It detects too many zeroes or ones in the sequence.

2. ***Frequency Test within a Block.*** The purpose of this test is to determine whether the frequency of ones in an M-bit block is approximately $M/2$, as would be expected under an assumption of randomness. The test detects too many zeroes or ones at the beginning of the sequence.

3. ***The Runs Test.*** The purpose of the runs test is to determine whether the number of runs of ones and zeros of various lengths is as expected for a random sequence. A run of length k consists of exactly k identical bits and is bound before and after by a bit of the opposite value. In particular, this test determines whether the oscillation between such zeros and ones is too fast or too slow.

4. ***Test for the Longest-Run-of-Ones in a Block.*** It determines whether the length of the longest run of ones within the tested sequence is consistent with the length of the longest run of ones that would be expected in a random sequence. The small *P-values* indicate that the tested sequence has clusters of ones.

5. ***The Binary Matrix Rank Test.*** Test checks for linear dependence among fixed length substrings of the original sequence. The small *P-values* indicate a deviation of the rank distribution from that corresponding to a random sequence.

6. ***The Discrete Fourier Transform (Spectral) Test.*** The focus of this test is the peak heights in the Discrete Fourier Transform of the sequence. The test detects periodic features in the tested sequence.

7. ***The Non-Overlapping Template Matching Test.*** The purpose of this test is to detect generators that produce too many occurrences of a given non-periodic (aperiodic) pattern. If the *P-value* is very small (< 0.01), then the sequence has irregular occurrences of the possible template patterns. The test consists of 148 subtests for different aperiodic patterns.

8. ***The Overlapping Template Matching Test.*** Unlike the Non-Overlapping Template Matching Test, the focus of this test is the number of occurrences of prespecified target strings.

9. ***Maurer's "Universal Statistical" Test.*** The purpose of the test is to detect whether or not the sequence can be significantly compressed without loss of information. A significantly compressible sequence is considered to be non-random.

10. ***The Lempel-Ziv Compression Test.*** The test determines how far the tested sequence can be compressed, using the number of cumulatively distinct patterns (words) in the sequence. The sequence is considered to be non-random if it can be significantly compressed.

11. ***The Linear Complexity Test.*** The purpose of this test is to determine whether or not the sequence can be produced by some *LFSR*. A *LFSR* that is too short implies non-randomness.

12. ***The Serial Test.*** The purpose of this test is to determine whether the number of occurrences of the 2^m m-bit overlapping patterns is approximately the same as would be expected for a random sequence. Random sequences have uniformity; that is, every m-bit pattern has the same chance of appearing as every other m-bit pattern.

13. ***The Approximate Entropy Test.*** This test compares the frequency of overlapping blocks of two consecutive/adjacent lengths (m and m+1) against the expected result for a random sequence. It detects non-uniform distribution of m-length words.

14. ***The Cumulative Sums (Cusums) Test.*** It calculates the maximal excursion (from zero) of the random walk defined by the cumulative sum of adjusted (-1, +1) digits in the sequence. For a random sequence, the excursions of the random walk should be near zero. For certain types of non-random sequences, the excursions of this random walk from zero will be large. Test consists of two subtests.

15. ***The Random Excursions Test.*** The focus of this test is the number of cycles having exactly K visits in a cumulative sum random walk. The purpose of this test is to determine if the number of visits to a particular state within a cycle deviates from what one would expect for a random sequence. This test is actually a series of eight tests (and conclusions), one for each of the states: -4, -3, -2, -1 and +1, +2, +3, +4.

16. ***The Random Excursions Variant Test.*** The test calculates total number of times that a particular state is visited in a cumulative sum random walk. It is actually a series of eighteen tests, one for each of the states: -9, -8, ..., -1 and +1, +2, ..., +9.

We test the *SMG* with 3-adic control register *FCSR* and 2 slave *LFSRs*. We test 3 different *SMGs* with connection tabs given on Table 2. For each generator we make 100 different tests with sequence of 1 000 000 bits, which is generated with a variety of seeds (different initial states). In all, this resulted in 300.189 = 56 700 *P-values*.

The results from all 56 700 tests are given on Table 3. On the table are given the average values for 100 *P-values* and 100 proportions for the 3 different *SMG* are tested. The number of tests is the same as those reviewed above. The letter "A" after the number of test indicates that the row gives the average value from all subtests included in this statistical test. The values of the cumulative-sums test (number 3.A) are the average from two subtests. The given values from the random-excursions test (number 12.A) are the average

from the eight subtests, the values from the random-excursions-variant test (number 13.A) are the average from the eighteen subtests, the values from nonperiodic-templates are the average (number 11.A) from 148 subtests, and the serial test (number 14.A) are the average from two subtests.

Table 2. Connections for tested SMG

№	*SMG*	Connection Integer *q*	Connection Polynomialsfor *R1* and *R2*
1.	*3-adic FCSR, L = 16;* *LFSR R_1, L1=131;* *LFSR R_2, L2=166.*	100 000 073	$x^{131} + x^8 + x^3 + x^2$; $x^{166} + x^{10} + x^3 + x^2$
2.	3-adic *FCSR, L = 16;* *LFSR R_1, L1 = 176;* *LFSR R_2, L2 = 200.*	100 000 037	$X^{176} + x^7 + x^5 + x^4 + x^3 + x^2 + 1$; $X^{200} + x^5 + x^3 + x^2 + 1$
3.	3-adic *FCSR, L = 14;* *LFSR R_1, L1 = 34;* *LFSR R_2, L2 = 32.*	10 000 229	$x^{34} + x^7 + x^6 + x^5 + x^2 + x + 1$; $x^{32} + x^7 + x^5 + x^3 + x^2 + 1$

To determine how well the empirical results match their theoretical counterparts we assess the goodness of fit of the distribution of *P-values* to a uniform distribution.

We build the histogram of *P-values* (fig. 4) among sub-intervals determined by dividing the unit interval by ten. To make a histogram all 567 average *P-values* from all statistical tests are used. As one can see from Fig. 4, the *P-values* are distributed approximately uniformly in all ten subintervals.

Table 3. Results from statistical analysis of SMG

№ of Test	Average from 100 tests of 1st *SMG*		Average from 100 tests of 2nd *SMG*		Average from 100 tests of 3th *SMG*		Average	
	P-value	Proportion	P-value	Proportion	P-value	Proportion	P-value	Proportion
1.	0,935716	1,00000	0,026948	1,0000	0,554420	0,9700	**0,505695**	**0,990000**
2.	0,236810	1,00000	0,437274	0,9900	0,129620	1,0000	**0,267901**	**0,996667**
3.A	0,357852	0,98500	0,505598	0,9900	0,586419	0,9750	**0,483290**	**0,983333**
4.	0,935716	0,99000	0,249284	0,9800	0,162606	0,9800	**0,449202**	**0,983333**
5.	0,834308	0,98000	0,867692	0,9900	0,699313	0,9700	**0,800438**	**0,980000**
6.	0,085587	1,00000	0,249284	0,9900	0,055361	1,0000	**0,130077**	**0,996667**
7.	0,779188	0,99000	0,032923	1,0000	0,935716	1,0000	**0,582609**	**0,996667**
8.	0,304126	1,00000	0,955835	0,9800	0,816537	0,9900	**0,692166**	**0,990000**
9.	0,964295	0,96000	0,162606	1,0000	0,236810	0,9600	**0,454570**	**0,973333**
10.	0,759756	0,99000	0,514124	0,9900	0,042808	0,9900	**0,438896**	**0,990000**
11.A	0,480569	0,99162	0,503849	0,9895	0,526539	0,9897	**0,503652**	**0,990292**
12.A	0,389515	0,97759	0,419667	0,9881	0,552950	0,9839	**0,454044**	**0,983183**
13.A	0,276655	0,99586	0,502274	0,9921	0,469981	0,9866	**0,416304**	**0,991487**
14.A	0,679079	1,00000	0,513301	1,0000	0,726828	0,9900	**0,639736**	**0,996667**
15.	0,657933	0,99000	0,075719	1,0000	0,401199	0,9800	**0,378284**	**0,990000**
16.	0,262249	0,97000	0,834308	0,9700	0,935716	1,0000	**0,677424**	**0,980000**

Then, the Goodness-of-Fit Distributional Test is made on the *P-values* obtained for an arbitrary statistical test using a distribution of χ^2. This is accomplished by computing:

$$(10) \qquad \chi^2 = \sum_{i=1}^{10} \frac{\left(F_i - \frac{s}{10}\right)^2}{\frac{s}{10}},$$

where F_i is the number of *P-values* in subinterval i and s is the sample size.

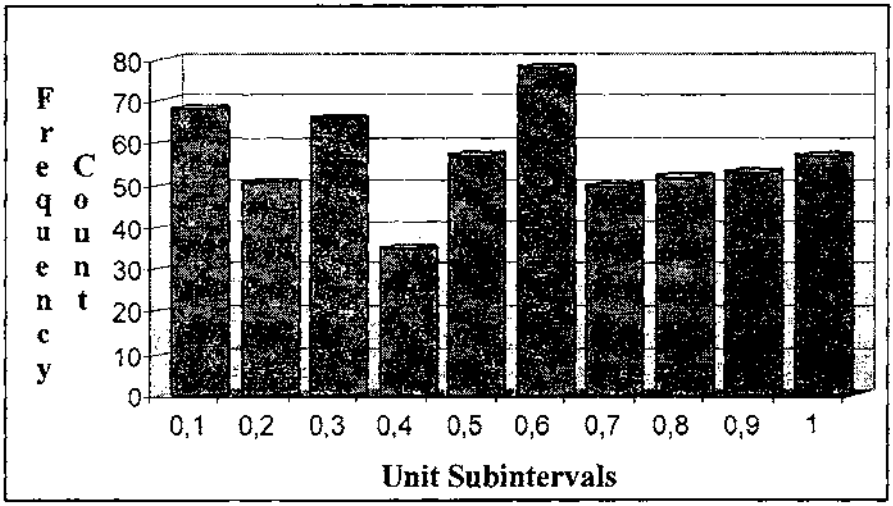


Fig. 4. Histogram of P-values

Thenthe *P-value* of the *P-values* is computed:

$$(11) \qquad P\text{-}value = Q\left(\frac{9}{2}, \frac{\chi^2}{2}\right),$$

where $Q(a,x)$is an Incomplete Gamma Function:

$$(12) \qquad Q(a,x) \equiv \frac{\Gamma(a,x)}{\Gamma(a)} \equiv \frac{1}{\Gamma(a)} \int_x^{\infty} e^{-t} t^{a-1} dt,$$

where $\Gamma(a)$ is a Gamma Function:

$$(13) \qquad \Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt.$$

Using the histogram of all 567 P-values in Fig. 4 and equations (10) – (13) *P–value = Q(4.5, 11.19) = 0.0055* are found. Because *P-value* ≥ 0.0001 [5], the sequence generated by *SMG* can be considered *uniformly distributed.*

To determine the proportion of sequences passing a statistical test we compute the proportion of sequences that had *P-values* $\geq$ *0.01*. The range of acceptable proportions is determined using the confidence interval defined as:

$$(14) \qquad \hat{p} \pm 3\sqrt{\frac{\hat{p}(1-\hat{p})}{m}},$$

where $\hat{p} = 1 - \alpha$, and m is the sample size. If the proportion falls outside this interval, then there is evidence that the data is nonrandom. Note that with m = 300 the confidence interval is 0.9728 ÷ 1.0. These bounds are given in Fig. 5 as dashed lines. As one can see from the figure, none of the proportions falls outside these thresholds.

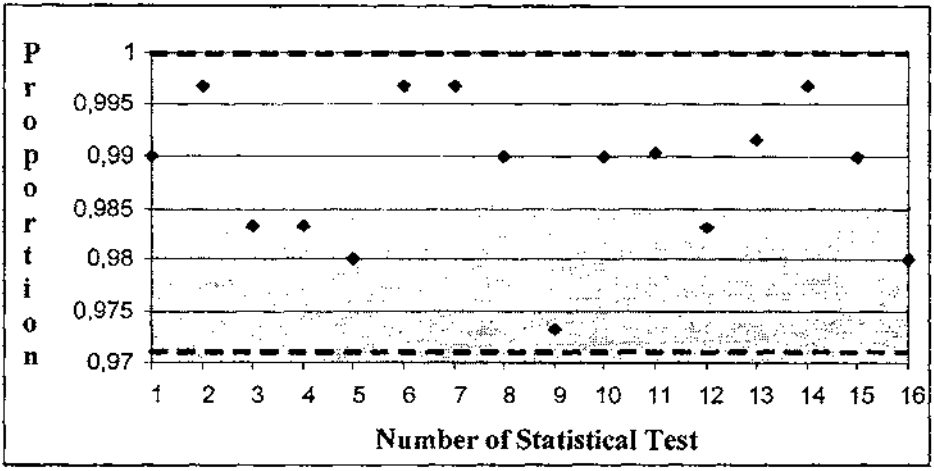


Fig. 5. Proportion of sequences passing a statistical test

The analysis of the empirical results from all 16 *NIST* statistical test shows:

1. The sequences generated by *SMG* are ***uniform***. This means that the probability of occurrence of a zero or one is equal, i.e. exactly 1/2. The expected number of zeros (or ones) is *n/2*, where *n* = the sequence length. This property is proved by follow tests and results:

1.1. The number of ones and zeros in a sequence and in sub-blocks are

approximately the same as in the truly random sequence is shown by 1st and 2nd *NIST* statistical tests. The average numbers for all 300 tested sequences are shown on Table 4.

Table 4. Number of Zeroes and Ones in SMG sequences

***SMG* Number**	**Average from 100 x 1 000 000 sequences**	
	Number of Zeroes	**Number of Ones**
SMG 1	499952	500048
SMG 2	500017	499983
SMG 3	500089	499911
Average from all *GSM*	**500019**	**499981**

1.2. The number of runs of ones and zeros of various lengths is as expected for a random sequence which is proved by 3rd test.

1.3. The results from 4th and 5th tests demonstrate that the generated sequences have not clusters of ones and there is not linear dependence among fixed length substrings of the original sequence.

1.4. The serial test (number 12) shows that every m-bit pattern has the same chance of appearing as every other m-bit pattern.

1.5. The excursions of the random walk is near zero which is given by test 14.

2. The sequences generated by *SMG* are ***scalable***, i.e. any extracted subsequence of sequence generated by *SMG* pass any test for randomness. This property is demonstrated by follow tests and results:

2.1. The approximate entropy test did not detect non-uniform distribution of m-length words.

2.2. The generated sequences have not irregular occurrences of the possible aperiodic template patterns which is shown by tests 7 and 8.

2.3. The number of visits to a particular state within a cycle did not deviate from the number expected for a random sequence.

3. The sequences generated by *SMG* are ***uncompressible***. This property is argued by follow tests and results which is by tests 15 and 16.

3.1. The sequence cannot be significantly compressed without loss of information which is shown by universal statistical test (number 9).

3.2. The sequence cannot be significantly compressed with Lempel-Zip algorithm (test 10).

4. The sequences generated by *SMG* have a ***large period***. This property is shown by follow results:

4.1. The spectral tests (number 9) did not detect periodic features in the tested sequence.

4.2. The linear complexity test (number 11) proved that the generated sequences cannot be produced by some too short *LFSRs* with length up to 16.

IV. Future works and conclusions

The empirical results from the above mentioned *NIST* statistical tests gives us reason to say that the binary sequences generated by *SMG* pseudorandom number generator have properties like ***uniformity, scalability, large period, incompressibility*** and ***consistency***, i.e. the behavior of a *SMG* is consistent across seeds. These properties show that the Shrinking–Multiplexing Generator works ***unpredictably*** like a true random number generator.

The proposed *SMG* is characterized with fast software implementation. It can be used in cryptographic applications in modern aerospace and communication information systems where high speed and low delay are requirements.

The proposed idea of using a p-adic control *FCSR* in *PRNG* allows generalizing all well-known clock controlled generators like p-adic Combiner Generator [7], Summation-Shrinking Generator [9], Alternating-Shrinking Generator and so on.

References

1. Coppersmith D., H. Karwczayk, Y. Mansou. "The Shrinking Generator", Crypto'93, http://imailab-www.iis.u-tocio.ac.jp/limit/Papers/Crypto_Eurocrypt/HTML/PDF/C93/22.pdf.

2. Getsov P., G. Mardirossian, S. Stoyanov, J. Jekov, P. Panova. "A Posibility of Storm and Hall Prediction using Data about Atmospheric Ozone Variations", Proceedings of International conference on Recent Advances in Spase Technologies – RAST, Istambul, 2003, pp.295-298.

3. Goresky M., A. Klapper. "Fibonacci and Galois Representations of Feedback-With-Carry Shift Registers", IEEE Trans. Inform. Theory, vol. 48, pp. 2826"2836, November 2002.
4. Menezes A., P. Oorschot, S. Vanstone. Handbook of Applied Cryptography, CRC Press, p. 780, 1997, www.cacr.math.uwaterloo.ca/hac.
5. Rukhin A., J. Soto, J. Nechvatal, M. Smid, El. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray, S. Vo. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, NIST Special Publication 800-22 (with revisions dated May 15, 2001), p. 162, 2001.
6. Soto J.. Statistical Testing of Random Number Generators, NIST Special Publication (301) 975-4641, p. 12, 2001.
7. Stoyanov B., B. Bedzhev. Algorithm for *p*-adic Combiner Generator Synthesis, XXXIX International Scientific Conference on Information, Communication and Energy Systems and Technology, ICEST 2004, 16-19 June 2004, Bitola, Macedonia, under printing.
8. Tasheva Zh., B. Bedzhev, V. Mutkov. "A Shrinking Data Encryption Algorithm with p-adic Feedback with Carry Shift Register", Conference Proceedings of XII International Symposium on Theoretical Electrical Engineering, ISTET'03, July 6-9, 2003, Warsaw, Poland, Volume II, pp. 397 – 400, 2003.
9. Tasheva Zh., B. Bedzhev, B. Stoyanov. Summation-Shrinking Generator, International Conference "Information Technology and Sequrity ITS – 2004", June 22-26, 2004, Partenit, Crimea, Ukraine, under printing.

НЯКОИ СТАТИСТИЧЕСКИ СВОЙСТВА НА СВИВАЩИЯ-МУЛТИПЛЕКСИРАЩ ГЕНЕРАТОР

Ж. Ташева

Резюме

Двоичните поточни шифри са едни от криптографските приложения, които често се използват в космическите и комуникационни системи, където се изискват високи скорости и малки закъснения.

Необходимостта от високоскоростни софтуерни поточни шифри през последните години доведе до предлагането на няколко нови алтернативни решения. В статията накратко е представена архитектурата

и принципа на работа на нов псевдослучаен генератор (*PRNG*), наречен Свиващ-Мултиплексиращ Генератор (Shrinking–Multiplexing Generator - *SMG*). Той е изграден от p - 1 паралелни линейни преместващи регистри с обратни връзки (*LFSRs*) с голям период и един преместващ регистър с пренос и обратни връзки (*FCSR*), който управлява нелинейността в работата на *SMG*.

За да бъде приложим в криптографските системи, всеки *PRNG* трябва да има големи период и линейна комплексност, както и добри статистически свойства. В статията са представени резултатите от направения статистически анализ на 300 различни последователности с дължина 1 000 000 бита, генерирани при различни ключове на *SMG* генератора. За анализа е използван предложения от Националния институт за стандарти и технологии комплект от 16 различни функционални статистически тестове. Анализът потвърждава, че генерираните от *SMG* последователности са еднообразни, независещи от използваното ядро на генератора, с голям период и не могат да се компресират. Всички тези свойства се отнасят и за всички подредици на генерираната последователност. Тези резултати ни дават основание да твърдим, че генерираната от *SMG* последователност е непредсказуема и се характеризира със свойствата на истински случаен генератор.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

APPLICATION OF BPN AND LVQ NEURAL NETWORKS FOR RADIOHOLOGRAPHY IMAGE RECOGNITION

Milena Kostova, Valeriy Dzhurov

Angel Kunchev University of Rousse
e-mail: mpk@mail.bg; vdjurov@yahoo.com

Abstract

Informative indication is taken from a contour obtained upon filtration of a radioholography image. Recognition is performed by classification of the types of aircrafts. BPN and LVQ neural networks are used by making comparative analysis of the two approaches.

1. Introduction

One of the important directions of the contemporary radiolocation is the creating of methods for having the whole information that is contained in the radiolocationable signals and noises. Radiolocationable recognition is a part of this direction. It includes definition of the radiolocationable characteristics of different objects, a choice of informative indication and decisive rules for recognition [1]. There are many ground, overwater, air and other targets called 'lying' targets, which vary greatly by their parameters and characters. This poses serious challenges and calls for upgrading and following development of the theory and practice of radiolocationable recognition.

Using radioholography approaches to synthesize radiolocationable images provides the opportunity to estimate the shape and the size of the radiolocationable target, as well as the typical indication of the object, on which radiolocationable recognition can be performed [6].

The interest of the specialists working in the sphere called 'inteligent' modeling to the application of shapes of artifical inteligence (AI) in recognition systems has suddenly grown up lately. The great variety of types of neural networks, as well as of radiolocationable recognition systems requires profound analysis and experimental work which purpose is achieved best when the shapes of AI are selected. [2]

2. Choice of informative indication

2.1. Methods for obtaining a contour

The contour of the object is reputed to be the most informative in object recognation. Methods for having a contour are separated in two groups:

- gradient methods – based on the mark of the first or the second derivative;
- methods, based on a functional approximation and on the analysis of the spectrum of the ingoing image in the sphere of the spatial frequency.

Spatial element with x and y of the function f(x.y) is done in gradient methods for separation of edge and limits. The module of the gradient value is defined for this function, which forms gradient image J(x, y):

$$(1) \qquad J(x,y) = \|\nabla f(x,y)\| = \sqrt{J_x + J_y} = \sqrt{\left(\frac{\partial f(x,y)}{\partial x}\right)^2 + \left(\frac{\partial f(x,y)}{\partial y}\right)^2}$$

From gradient image we pass to its double presentation to the following formula:

$$(2) \qquad b(x,y) = \begin{cases} 1, & \text{ако} \quad J(x,y) \geq \vartheta \\ 0, & \text{ако} \quad J(x,y) < \vartheta \end{cases}$$

where ϑ is a threshold.

Usually, the value of the threshold is defined by analyzing the histogram of the gradient presentation:

$$(3) \qquad J(x,y), \quad x = 1,2,\ldots,N, \quad y = 1,2,\ldots,M$$

The image is stored in the memory of the computer by its discreet (figure) presentation:

$$g(i,j), \quad i \in 1,2,\ldots,N, \quad j \in 1,2,\ldots,M \,.(4)$$

The marks of the derivative for the function g(i, j) are accomplished by figure differences, performing in determinate scheme (single, double), called differential (gradient) operators.

Such gradient methods are Roberts, Sobel, Canny, LoG, Prewitt and zerocross's.[2]

Another group includes:

- method, based on functional approximation;
- method, based on the analysis of the spectrum of the ingoing image in the sphere of the spatial frequency.

The methods of this group require a lot of calculations, which makes difficult their use.

Methods in pawn in MATLAB7.0 are used for receiving the contour of the image[3]. The method for the contour's assignment is chosen on the basis of experiments with tentative objects.

A selection of n = 30 images is chosen to choose a method for obtaining the contour. The selection contains images of aircrafts which are accepted as models.

A contour of the object (aircraft) and the number of the points lying on it in each method is defined for each image of the selection. The results from the experiment are shown on Table 1. From the received results we can see that when Roberts's method is used, the greatest number of points of the contour is given, therefore it is the most informative. Another priority of this method is that the contour hammers out but it does not rend. That is why Roberts's method is chosen to obtain the contour of the image.

Obtaining the contour and taking the coordinates of its points down is done accordingly by MATLAB7.0, the functions *edge* and *find* [3].

2.2. Finding indication from a contour

2.2.1.General characteristic of the indications of the objects and vector formation from indications.

Composing of a basis of information related with the characteristics of the investigated objects is necessary to make concrete classification of flying objects in the described groups.

These are morphology characteristics describing the shape and the special geometric features of the investigated objects or their silhouette projections[2].

The contour curvature of the object silhouette, its area,

circumference, fuselage axis, width of the aircraft's wings, symmetry of the contour and others treat to this group.

The properties of the objects related with the shape of the typical representative plane for the different groups, the location of the wings in relation to the fuselage axis and others can be estimated based on the analysis of the morphology characteristics. The analyses of the investigated objects are based mostly on the morphology characteristics of the objects silhouette images in many investigations connected with the creating of automation methods and instruments for aircrafts recognition[5]. In general, the morphological properties of the silhouette images have great potential. However, the results of the published investigations show that their potential of aircrafts in 2D-images is not used enough. This refers mostly to the approaches that ensure invariant for the scale, rotation and translation of the object in the frame. Another recognition approach must be searched according to this base.

The following vector from indications has been formed to recognize separate types of aircrafts (specific 'Stealth', military, vehicular) in a contour synthesized after a cultivation of radioholographie image:[6]

1. an attitude A of the width to the length of the object where the width is the distance between the most distant points of the wings and the length is the distance between the most distant points of the fuselage axis;

2. an incidence α of the right which connects the mass centre and the most distant point of one wing in relation to the fuselage axis.

3. the position of the mass centre in relation to the geometric environment of the fuselage axis. The difference 'geometric environment-mass centre'- L_r is calculated.

4.width of the aircraft's wing – L_k.

The indications for 'F22' are shown in Fig.1.

In this way the introduced indications are invariant in respect of the scaling and translation of the object in the frame.

The typical values calculated for the relevant indications of the separate types of aircrafts are shown on Table 2.

Table 1.Number of points from the contour of the image, obtained in the different methods.

Number of the image	Method of Roberts's	Method of Sobel's	Method of Prewitt's	Method of Canny's	Method of zerocross-'s	Method of LoG's
1	669	480	508	478	480	480
2	599	436	448	438	437	437
3	625	464	480	458	458	458
4	604	449	469	451	450	450
5	648	467	491	461	463	463
6	650	466	516	466	470	470
7	622	460	478	462	463	463
8	608	432	445	434	435	435
9	638	483	510	478	480	480
10	616	449	502	449	449	449
11	578	430	448	431	429	429
12	682	489	521	486	490	490
13	618	459	497	456	459	459
14	672	492	533	492	493	493
15	603	443	477	440	443	443
16	671	497	540	495	499	499
17	615	469	512	466	469	469
18	608	452	473	447	448	448
19	603	438	464	436	440	440
20	620	429	456	429	435	435
21	577	416	433	414	413	413
22	554	412	416	414	414	414
23	554	408	414	409	408	408
24	604	449	485	448	450	450
25	658	479	499	478	478	478
26	589	436	455	437	436	436
27	629	459	497	458	460	460
28	608	450	482	448	447	447
29	588	431	467	432	432	432
30	663	503	520	497	495	495

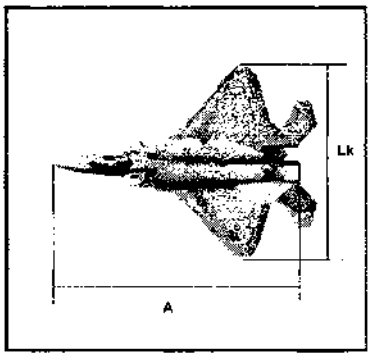


Fig. 1. A diagram of aircraft F22 and its geometric size

The following classification and the relevant recognition of the aircrafts can be made by Table 2 according to the enunciated indications and received costs for the concrete type of aircraft.

1. The aircrafts can be classified in three groups according to the indication attitude: Specific, Military and Vehicular.

2. The second group of aircrafts can be separated according to the indication position of the mass centre in relation to the geometric environment of the fuselage axis on Bombers, Destroyers and UAV.

The following classification and relevant recognation of the aircrafts can be made by table 2 according to the enunciated indications and received costs for the concrete type of aircraft:

Because of the fact that UAV have a typical cost of the attitude A that is notable from the cost for the other types of aircrafts, this group could be separated and classified correctly in attitude.

3. The vehicular aircrafts could be recognized by the indication **A** as a type *An* or as a type *Boeing*. The number of the aircraft's engines can be defined by the indication width of the wings because those with more engines 4, 6 have greater width of the wings.

For their obteining the program environment MATLAB7.0 is used [3]. The program for assignment and calculating the relevant indications conditions the received images in the following way:

- the obtained image passes through a cultivation that separates the contour of the object.
- calculates the coordinates of the mass centre.
- the coordinates of the last points of the wings and the fuselage axis (conditionally represented as head and tail) are extracted.
- calculates the costs of the indications.

2.2.2. Estimation of the information of the enunciated indications

Interval marks are used for estimation of the information of the introduced indications [1]. A confidential interval is made for the middle cost and Student's distribution is applied:

$$(5) \qquad T = \frac{\overline{X} - E[X]}{s}\sqrt{n}$$

where $\overline{X}$ and s are the marks of the middle cost and the middlesquare diversion of the costs of the relevant indication accordingly;

n - volume of the extract data;

$E[X]$ - middle cost of the indication .

The confidential interval usually has symmetrical frontiers according to the point mark. The cost of these frontiers depends on the stages of familiarity $k = n - 1$ and the wished confidential probability γ. The confidential interval for the middle cost is in the equation:

$$(6) \qquad I_{\gamma;k}^{E[X]} = \left(\overline{X} - t_{\gamma;k}\frac{s}{\sqrt{n}}; \overline{X} + t_{\gamma;k}\frac{s}{\sqrt{n}}\right)$$

where $t_{\gamma;k}$ is Student's criterion.

The conterminous relative mistake $\Delta_{\gamma;k}$ that is allowed in an exchange of the middle cost and its mark is calculated:

$$(7) \qquad \Delta_{\gamma;k} \approx t_{\gamma;k}\frac{s}{\sqrt{n}\overline{X}}.100,\%$$

The confidential intervals of the indications, estimated at (6) for particular types of aircraft are presented on Table 3, 4, 5 accordingly.

Table 2. Values of the indications for the particular types of aircraft

Type of aircraft	Xw 1	Yw 1	Xw 2	Yw 2	Xn ax	Yn ax	Xopt	Yopt	Xc	Yc	A= w/le.	Alfa	N= Xc- Xmd
F22	328	92	328	365	73	232	411	237	274	239	0.807	66.83	-32
F117	72	125	250	128	161	351	161	116	161	203	0.755	41.25	-30.5
Military aircraft Exterminators													
Mirag 2000 EU	363	275	363	109	191	123	191	409	191	304	0.58	54.94	-38
Su - 34 RU	275	145	275	393	8	270	387	270	237	270	0.654	72	-39.5
Jaguar EU	258	213	258	32	17	123	337	123	224	123	0.566	20.71	-47
F16 USA	328	73	328	276	119	176	418	176	307	176	0.679	78	-38.5
Bomber													
Tu 160 RU	339	102	339	242	104	174	392	174	272	174	0.486	45.447	-24
Buccaneer EU	236	41	236	199	86	121	330	121	217	121	0.647	76.35	-9
Reconnaissance													
UAV											1.4706	90	-
											1,559	90	-
											1,63	90	-
											1,71	90	-
											1.9238	90	-

Type of aircraft	Xw 1	Yw 1	Xw 2	Yw 2	Xn ax	Yn ax	Xopt	Yopt	Xc	Yc	Le.	W	A=W/le	N= Xc- Xmd
Transports Boeing - USA														
B 7474 engines											70.66	64.44	0.912	-
B 7572 engines											47.3	45	0.9514	-
B 7672 engines											48.5	47.6	0.9814	-
An - RU														
An - 722 engines											28.07	31.89	1.1361	-
An - 224 engines											57.90	64.40	1.1123	-
An - 1244 engines											69.1	73.30	1.0608	-
An - 2256 engines											84.00	88.40	1.0524	-

Table 3. Confidential intervals and border relative error for indication A

Type of aircraft	Down border	Upper border	*, % **
Stealth	0.7575	0.8077	
Military			
Exterminators	0.56	0.6789	
Bombers	0.486	0.647	
UAV	1.4706	1.9238	
Transports			
Boeing	0.912	0.9814	

*= $\Delta_{\gamma,\kappa}$; **= γ=0,95

Table.4. Confidential intervals and border relative error for indication α.

Type of aircraft	Down border	Upper border	*, % **	*, % ***
Other types of military aircraft	78,9316	85,9434	4,2529	7,8103
UAV aircrafts	90			

*= $\Delta_{\gamma\kappa}$; **= $_{\gamma=0,95}$; ***= $_{\gamma=0,99}$

Table 5. Confidential intervals and border relative error for indication L_r

Type of aircraft	Down border	Upper border	*, % **	*, % ***
Exterminators	34.915	46.585		
Bombers	9	24		

*= $\Delta_{\gamma\kappa}$; **= $_{\gamma=0,95}$; ***= $_{\gamma=0,99}$

Table 6. Confidential intervals and border relative error for indication L_K

	Boeing	
	2 engines	4 engines
Lк	<56.02	>56.02

Table 7. Confidential intervals and border relative error for indication L_K

	An		
	2 engines	4 engines	6 engines
Lk	<48.145	80.85<Lk-<48.145	>80.85

They are obtained using the program environment MATLAB 7.0 [3].

3. Using neural networks for objects recognition by classification.
Comparative analysis of two types of neural networks - BPN and LVQ is made.

3.1. Presenting of the entrance data for the neural networks

The application of the neural networks requires special preparation of the data, at which it will accomplish aircrafts recognition.[4] These indications are defined in p. 2.2.1. When working with real numbers, the error teaching of the neural network is relatively big. That is why, the entrance data are coded with 0 and 1.

The entrance data are coded as follow:

Indication attitude A: This indication is changed in 5 intervals, characteristic for three classes of aircraft – Stealth, Military and Transport, presented in p.2.2.2. The coding of the entrance is the following:

1 0 0 0 0
0 1 0 0 0
0 0 1 0 0
0 0 0 1 0
0 0 0 0 1,

To the first column corresponds first class (Stealth);
To the second and third – second class (Military);
To the third and fourth – third class (Transports).
The coding of the wished exit (target) is the following

1 0 0 0 0
0 1 1 0 0
0 0 0 1 1,

where to first class corresponds coding 0 0 1;
to second class - coding 0 1 0;
to third class - coding 1 0 0;

➢ Indications N

The coding depends on the intervals:

Exterminators – 34,915 - 46,585 - [0 0]

Bombers – 9 – 24 - [1 1]

Respectively, there are two classes at the end of the neural network

[1 0

0 1].

➢ Width of the wings L_k.

If the attitude is in the following borders:

- from 0.8701 – 1.0059 – coding is 1;
- from 1.035 – 1.1458 coding is 0.

Respectively the neural network has two exits:

- 0 0 – Boeing;
- 1 1 – Àn.

3.2. BPN neural network

3.2.1. Creating a neural network

Diffusive neural network with converse spreading of the error (BPN) can be created in two ways – by program code or by using dialogue windows.

➢ Creating of the network by dialogue windows.

It is using *Neural Networks Toolbox* of the program package *Matlab 7.0*.

➢ Creating by program code:

The function, which creates diffusive neural network with converse spreading of the error is *newff*. The first step in creating a "*feadforward*" network is initializing the network object. The function *newff* creates "*feadforward*" network by the code, written as follows:

net=newff([],[],{''});

Therefore, the code for the data from the table has the following aspect:

net=newff([0 1; 0 1; 0 1],[3,8,3],{'tansig', 'tansig'});

This command creates network object, and initializes the weights and the diversions of the network. Thus the network is prepared for training. Before starting training of "feadforward" network, the weights and the diversions have to be initialized, which is done by the function *newff*.

3.2.2. Algorithm for training of the network

The presented algorithm uses the gradient method, which decreases the weights. [7] The gradient method is called backpropagation, which is a training algorithm, in which the weights are moved in the direction of the negative gradient. There are many varieties of the backpropagation algorithm: the most simple implcmentation of backpropagation is the renovation of the networks weights and diversions in the direction in which the function decreases – negative gradient.

A training multitude of last-number vector couples is used, presenting input image and wanted image. The input image is presented on the entrance layer of HM. The information is spread on the initial layer of the network and then concrete image appears.

The difference between this image and the wanted image represents the operation error of HM. The value of the error can be estimated like middlesquare. If the value of the error is lower than a given threshold, then a next couple is chosen. Otherwise, the vector of the error in the initial layer is used for estimation of the initial errors of the neural from the nearest layer. These errors spread to the previous layer and so on. The errors from each layer are used for correction of the weights of the connections to this layer. The procedure is repeated for all couples from the trained multitude of (input-wanted image).

One round of all couples from the trained multitude is called cycle or epoch. In other words, the meaning of every epoch is the presentation of each couple input-wanted image from the trained multitude only once. The process is repeated for as much epochs, as required that the costs of the error become less or equal to a given threshold (the operation accuracy of the neural network).

3.2.3. Simulation of the neural network.

The function **sim** imitates network. **Sim** takes network input *p* and network object *net*, and returns network output *a*. Below is presented the general appearance of the function **sim**:

```
p={[    ]};
y=sim(net,p).
```

3.3. LVQ

3.3.1. Creating the neural network.

LVQ neural network analogy of BPN network can be created in two ways – by program code or by using dialogue windows.

3.3.2.Creating LVQ by program code

The LVQ network could be created by the newlvq function.

net= newlvq (PR,S1,PC,LR,LF)

where:

- PR is a Rõ 2 matrix of minimal and maximal values for R input elements.
- S^1 is the number of the first layer of hidden neurons.
- PC is a S^2 element vector with the percentage for the typical class
- LR is the precision of education(by understanding 0.01).
- LF is the educating function (by understanding learnlv1).

After calling the *newlvq,* a net with 10 neurons in the first and 15 neurons in the second layer will be created with its arguments, The first layer of weights are initiated in the center of the input range with a *midpoint* function. The prime values of the first layer of the weight matrix are checked with:

net.IW{1,1}

ans =

.....

The weights of the second layer are checked with:

net.LW{2,1}

ans =

.....

In this way, the competitive neuron and the neurons in the line layer it is connected to can be determined.

All other weights between the competitive and the line neurons are equal to 0.

3.3.3. Algorythm of education.

The educating function for the *LVQ* net is *learnlv*, and it is used by understanding. The education in the competitive layer is based on a pair of input and target neurons.

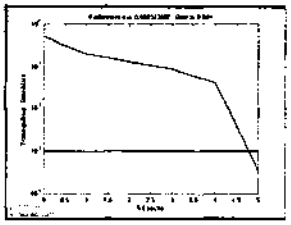
The simulation of the *LVQ* neuron network is done in a way similar to

the simulation of the *BPN* neuron network, and this procedure is described in point 3.2.3.

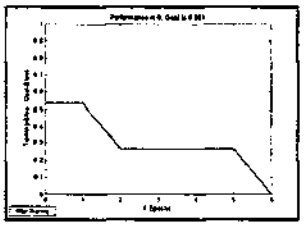
3.4. Comparison of *BPN* and *LVQ* neuron networks according to the number of the training epochs.

200, 500 and 1000 epochs have been offered. A comparative analysis of the training errors of the two networks with reference data has been made.

200 epochs.



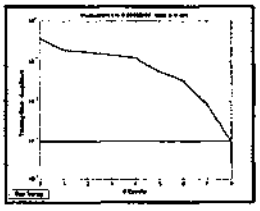
a)



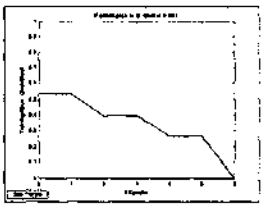
b)

Fig.2. Training of the neuron nets for 200 epochs – a) BPN neuron network, b) LVQ neuron network.

500 epochs.



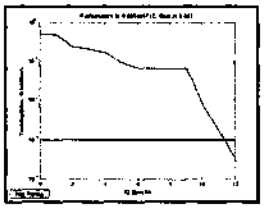
a)



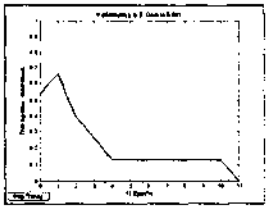
b)

Fig.3. Training of the neuron networks for 500 epochs – a) BPN neuron network, b) LVQ neuron network.

1000 epochs.



a)



b)

Fig.4. Training of the neuron networks for 1000 epochs – a) BPN neuron network, b) LVQ neuron network.

The desired level of mean square level of erroneous answers is 0.001. The average values of the reached level for both networks is presented on Table 8. The number of the epochs needed for reaching the level is shown, too.

Table 8. Values of the average level of errors and number of epochs for training BPN and LVQ neuron nets.

Number of epochs	BPN		LVQ	
	Average level of errors	Number of epochs needed for reaching the level	Average level of errors	Number of epochs needed for reaching the level
200	0.000343	5	0	6
500	0.000964	8	0	6
1000	0.0003357	12	0	12

Conclusion: We see from the results, that the LVQ neuron network reaches the wanted level of precision 0.001 unlike the BPN neuron net. The LVQ network gets trained for 12 epochs when 1000 are given, while if a smaller number of epochs is given (200, 500) it gets trained faster. Therefore, when experimenting with real objects it is better to use an LVQ neuron network with 500 epochs of training.

References

1. В е н к о в, *П. Г.* Анализ и разпознаване на изображения и сцени, София, Технически университет, 1996 г.
2. Го ч е в, Г. Компютърно зрение и невронни мрежи, София, Технически университет, 1998 г.
3. Дь я к о н о в, В., В. К р у г л о в. Математические пакеты расширения Matlab. Специальный справочник, Санкт Петербург, Питер, 2001 г.
4. К и р о в а, Т. Невронни мрежи (Основни архитектури и обучаващи алгоритми), София, Софтех, 1995 г.
5. http://spacecom.gvc.nasa.govienscon f
6. http://pvo.gun.ru
7. Lu g e r, G. F., W. A. S t u b B l e f i e l d. Artificial Intelligence – Structures and Strategies for Complex Problem Solving (Chapter 14, Machine Learning: Connectionist), Third Edition, USA, Addison Wesley Longman, Inc., 1998 г.

ПРИЛОЖЕНИЕ НА BPN И LVQ НЕВРОННИ МРЕЖИ ПРИ РАЗПОЗНАВАНЕ НА РАДИОХОЛОГРАФСКО ИЗОБРАЖЕНИЕ

М. Костова, В. Джуров

Резюме

От контур получен след филтрация на радиохологафско изображение са снети информативни признаци. Извършено е разпознаване чрез класификация на типове самолети. Използвани са BPN и LVQ невронни мрежи като е направен сравнителен анализ на двата подхода.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

APPLICATION OF FUZZY LOGIC IN THE QUALITY OF A FUZZY IN IDENTIFYING A DYNAMIC TARGET

Milena Kostova

Angel Kunchev University of Rousse
e-mail: mpk@mail.bg;

Abstract

A stage of an "intelligent" system for identifying a flying target (airplane), making a classification according to features taken down from a contour via a neural network is treated in this paper. Aiming at correct and true coding of the input data of the neural network for all the intervals of the values of the features, a FUZZY system, which uses the Mamdani algorithm, is synthesized in its quality of a fuzzy inference.

Key words: *features, intervals, coding, fuzzy inference, fuzzy logic*

1. Introduction

An important factor for a correct classification at solving problems connected with identification is the preceding preparation, ensuring optimal setting of the intervals of the features. A particular radio-location problem is treated, connected with identification via a classification of dynamic targets (airplanes) according to features taken down from a contour drawn at processing a radioholographic image. [3] It is possible, due to the great variety of flying devices, planes with values of a certain feature close to the

limits of a respective interval but beyond them or getting into an interval characteristic of another group of planes not to be classified or to be classified incorrectly, i.e. a correct coding cannot be realized. This problem is solved by applying fuzzy logic in the quality of fuzzy inference. The fuzzy system guarantees a correct coding in cases where there are uncovered intervals or coinciding (overlapping) intervals. In the rest of the cases it reacts as a fuzzy inference.

2.Theoretical part

Definition: *Fuzzy logic inference* means reaching a conclusion in the form of a fuzzy multitude, corresponding to respective values of the input, using a fuzzy data base of knowledge and fuzzy operations.

The fuzzy inference is on the grounds of Zadeh's compositional rule. [1]

Definition: Lotfi *Zadeh's compositional rule of the conclusion is formulated as follows:*

if the fuzzy ratio $\tilde{R}$ between the input variable (x) and the output variable (y) are known, then when the fuzzy value of the input variable $x = \tilde{A}$, the fuzzy value of the output variable is defined by the expression:

$$y = \tilde{A} \circ \tilde{R} \tag{1}$$

where o -is the maxmin. composition.

The fuzzy logic inference according to *Mamdani* 's algorithm is as follows [1]:

$$\bigcup_{p=1}^{k_j}[\bigcap_{i=1}^{n} X_i = a_{i,jp}.w_{jp}] \to y = d_j \tag{2}$$

$$j = 1 \div m$$

Let us assume that $\mu_{jp}(x_j)$ is the membership function of the input x_i Fuzzy term $a_{i,jp}$, i.e.

$$a_{i,jp} = \int_{\underline{x}_i}^{\tilde{x}_i} \mu_{jp}(x_i)/x_i, \quad x_i \in [\underline{x}_i, \overline{x}_i] \tag{3}$$

$\mu d_j(y)$ -is membership function of the output y of the fuzzy term d_j, i.e. $d_j = \int_{\underline{y}}^{\bar{y}} \mu d_j(y)/y$, ,

(4) $y \in [\underline{y}, \bar{y}]$

The degree of membership of the input vector $X^* = (x_1^*, x_2^*, ..., x_n^*)$ of the fuzzy term d_j is defined as follows:

(5) $$\mu_{d_j}(X^*) = \bigvee_{p=1,k_j} w_{jp} \cdot \bigwedge_{i=1,n} \left[\mu_{jp}(x_i^*) \right], \; j = \overline{1,m}$$

where $\vee(\wedge)$ -are the logic operations OR (AND). This results into a fuzzy multitude $\tilde{y}$, corresponding to the input X^*:

(6) $$\tilde{y} = \frac{\mu_{d_1}(X^*)}{d_1} + \frac{\mu_{d_2}(X^*)}{d_2} + ... + \frac{\mu_{d_m}(X^*)}{d_m}.$$

For a transition from a sub-multitude of the universal multitude of the fuzzy terms $\{d_1, d_2, ...d_m\}$ to the fuzzy multitude of an interval $y \in [\underline{y}, \bar{y}]$ it is necessary: to "cut" (agg) the function of belonging $\mu d_j(y)$ at level $\mu d_j(X^*)$, and to join the resulting fuzzy multitudes, which can be written down as follows:

(7) $$\tilde{y} = \underset{j=\overline{1,m}}{agg} \left(\int_{\underline{y}}^{\bar{y}} \min\left(\mu_{d_j}(X^*), \mu_{d_j}(y) \right) / y \right)$$

The crisp (non-fuzzy) value of the output y, corresponding to the input vector X^* is determined as a result of a defuzzification of the fuzzy multitude $\tilde{y}$. Defuzzification is applied mostly in accordance with the method of the weight centre.

$$y = \frac{\int_{\underline{y}}^{\overline{y}} y \cdot \mu_{\tilde{y}}(y)dy}{\int_{\underline{y}}^{\overline{y}} \mu_{\tilde{y}}(y)dy},$$

3. Methodology for creating a Fuzzy system

3.1. General characteristic

One of the effective approaches for choosing an optimal decision to which interval to consider the values of a given feature is the so-called approach for the logic conclusion in the conditions of obscurity and inconclusiveness (fuzzy logic inference).

The methods of the fuzzy multitudes are applied in the quality of a formal apparatus. To be more exact, fuzzy relations are introduced about quality values of the area factors and the target function (the criterion for optimality). A fuzzy relation is characterized by a membership function, which is a subjective measure of the degree of fulfillment (truthfulness) of the factor-criteria ratio. Using the Bellman-Zadeh compositional rule, the fuzzy ratio is applied for calculating the value of the conclusion (the so-called composition -based inferencing).

A suitable apparatus for formal description of the conclusion are the multi-valued logic probabilities and respectively the fuzzy multi-valued logic functions. They are based on the multi-valued logic (k-valued logic), $k \geq 2$, which is a generalization of the two-valued logic.

The functions of the ê-valued logic $f(x_1, ... x_i ..., x_n)$, x_i, $i = 1 \div n$, where for each x_i there are k number of logic truth values, can be presented in a table or analytically.

3.2.Application of a fuzzy inference for a given localizing task.

3.2.1.Formulating the task:

Digital radioholographic images of ten types of airplanes are treated: F16, An 124, McDowell, B52, Bucaneer, F117, Jaguar, Mig 29, Miraj 2000, Su 34. A filtration of the noisy images is done using a CNN neural network. *Candy* software is used to achieve the filtration via a CNN neural network.[2]

The method used for taking down a contour from the filtered radio-holographic image of the target is Robert's method because it provides the biggest number of points of the contour, therefore it is the most informative: the contour only gets thinner but is not broken.

For identifying particular types of planes (specific "Stealth"; military and transport) the following vector of features is synthesized from the produced contour:

1. Ratio A of the width to the length of the target, where the width is the distance between the endmost points of the wings and the length - the distance between the endmost points of the fuselage axis:

2. Slope α of the line linking the centroid (the mass centre) and the endmost point of one of the wings, toward the fuselage axis:

3. The position of the mass centre in relation to the geometric centre of the fuselage axis The subtraction "geometric centre – mass centre" - L_r is calculated.

4. Width of the wings of the plane – L_k.

The characteristic values are calculated for the respective features of the given types of planes.

The following classification of planes is done according to the formulated features and the values calculated for the particular type of plane:

- According to the feature 'relation', the planes can be classified in three groups: specific, military and transport.

- According to the feature 'position of the mass centre in relation to the geometric centre of the fuselage axis', the second group of planes can be divided into bombers, exterminators, and unmanned.

Due to the fact that the unmanned planes have characteristic value of the ratio A, different from the value of the other types of planes, this group of planes could be classified correctly as a separate one according to this ratio.

- According to the feature 'A', the transport planes can be identified either as type *An* or as type *Boing*. According to the feature width of the wings, the number of the plane engines could be identified as those with more engines, e.g. 4, 6, have a larger width of the wings.

The confidence intervals of the features of the separate types of planes, shown in tables 1, 2, 3, 4, 5, are given by means of statistic methods.

Table 1.
Confidence intervals and limiting relative error for the feature 'A'.

Type of plane	Lower limit	Upper limit	*, % **
Stealth	0.7575	0.8077	
Military			
Fighters	0.56	0.6789	
Bombers	0.486	0.647	
Unmanned	1.4706	1.9238	
Transport			
Boing	0.912	0.9814	
An	1.0524	1.1361	

*$= \Delta_{\gamma;k}$; **$= \gamma = 0{,}95$

Table 2. Confidence intervals and limiting relative error for the feature α :

Type of plane	Lower limit	Upper limit	*,% **	*, % ***
Other types of planes	78,9316	85,9434	4,2529	7,8103
Unmanne-d planes	90			

*= $\Delta_{\gamma;k}$; **= $\gamma = 0{,}95$; ***= $\gamma = 0{,}99$

Table 3.
Confidence intervals and limiting relative error for the feature L_{γ}

Type of plane	Lower limit	Upper limit	*,% **	*, % ***
Fighter	34.915	46.585		
Bomber	9	24		

*= $\Delta_{\gamma;k}$; **= $\gamma = 0{,}95$; ***= $\gamma = 0{,}99$

Table 4.
Confidence intervals and limiting relative error for the feature L_{ε}

	Boeing		
	2 engines	4 engines	
Lk	<56.02	>56.02	

Table 5.
Confidence intervals and limiting relative error for the feature $L_{\tilde{e}}$

	An		
	2 engines	4 engines	6 engines
Lk	<48.145	80.85<Lk<-48.145	>80.85

The *Matlab* programme medium is used for getting these intervals [3].

On Table 1 it is seen that the intervals of the feature ratio A, determined for each type of plane have both overlapping and not overlapping plots. For example:

1. When the ratio A is in the interval from 0.677 to 0.748, a concrete decision cannot be imposed how to code the input of the neural network – it could go into group 2 or into group 1.

2. When the ratio A is in the interval from 0.8056 to 0.8904 – it could go into group 3 or into group 1.

3. When the ratio A is in the interval from.1324 to 1.2516 – it could go into group 3 or into group 2.

Because of this, a great deal of planes, which have values of a certain feature close to the limits of a respective interval but beyond them or get into an interval characteristic for another group of planes, cannot be classified or can be classified incorrectly, as a correct coding cannot be realized.

In order to eliminate this problem a second feature is introduced: the angle α, defined in (10) and Fuzzy Logic is used with the aim of covering a larger interval of the values of the feature.

3.2.2. Getting optimized values for the feature 'A'.

The input variables for the Fuzzy system are the two features: the ratio A and the slope α. The membership functions of A are set by 9 linguistic

variables in the range from 0.4617 to 2.016. The limits are chosen in accordance with the following:

- the lowest value of the ratio is 0.486 which corresponds to a plane type "Military". An admissible deviation from this value is set in the limits of 5% -- A=0.4617.

- the highest value of the ratio is 1.92 (type of plane "Military"). An admissible deviation of 5 % from this value is set – A=2.016.

The linguistic variables and their intervals of operation are shown on Table7.

The limiting values of the feature A for the three groups of planes are shown on Table 6 in ascending order.

Table 6. Limiting values of A for the types of planes:

Type of plane	Limiting values of A	
	Min	Max
Military	0.486	0.6789
Stealth	0.7575	0.8077
Military	1.4706	1.9238
Transport	0.912	1.1361

For the linguistic variables A_3, A_5, A_7 and A_9 as a mean value is assumed the upper limit of each interval for the preceding type of plane enlarged by 5%. The linguistic variables A_2, A_4, A_6, and A_8 cover the interval for the respective type of plane.

The *Matlab 7.0* programme medium is used for modelling and simulation of fuzzy logic.[5] The ratio A as an input variable in fuzzy logic is shown visually in Fig.1

Table 7. Linguistic variables for A and their intervals

A	Intervals		
	Min	Mean	Max
A1	0.4617	0.486	0.5103
A2	0.486	0.599	0.7128
A3	0.677	0.7128	0.748
A4	0.7128	0.7804	0.848
A5	0.8056	0.848	0.8904
A6	0.848	1.02	1.192
A7	1.1324	1.192	1.2516
A8	1.192	1.556	1.92
A9	1.824	1.92	2.016

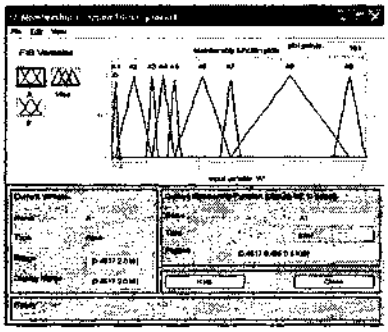


Fig.1. Fuzzy logic according to feature A.

The membership functions of the other input variable - the slope α are represented by 8 linguistic variables in the range 19.6745 ÷ 94.5. For each type of plane the values of α are shown on Table 8. The admissible intervals of the linguistic variables are shown on Table 9.

Table 8. Limiting values of α for each type of plane

Type of plane	*, °
Jaguar - Military	20.71
Stealth - Specific	41.25
Transport	54
Mirag 2000 - Military	60
Stealth - specific	66
Su - 34 - Military	72
F16 - Military	78
UAV - Military	90

*= α

This angle varies in small limits depending on the position of the plane in the frame. Due to that fact, overlapping of the limiting values is impossible to be set for the types of planes. Therefore, the intervals of that input variable are formed by the limiting value and an admissible deviation from it, ± 5 %.

Table 9. Linguistic variables for α and their intervals.

F	Intervals		
	Min	Mean	Max
F1	19.6745	20.71	21.7455
F2	39.1875	41.25	43.3125
F3	51.3	54	56.7
F4	62.7	66	69.3
F5	68.4	72	75.6
F6	74.1	78	81.9
F7	85.5	90	94.5
F8	57	60	63

Visually the relation α, as an input variable in fuzzy logic, is shown in Fig.2.

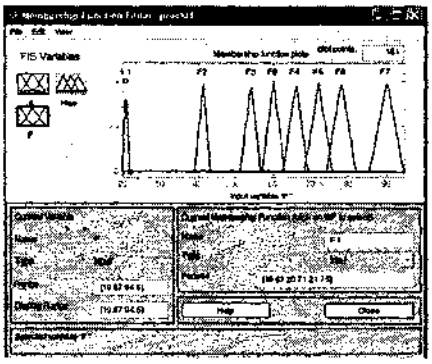


Fig.2. Fuzzy logic according to feature 'α'.

The output variable is determined by the discrete levels 1, 2, and 3, which correspond to the respective group of planes: Stealth, Military and Transport. It is shown in Fig.3

Triangular membership functions to the linguistic values of the variables are chosen. The transforming of the input variable into output ones is done by the algorithm of *Mamdani* [4]. Thirty-nine rules are introduced for control which are shown on Table 10.

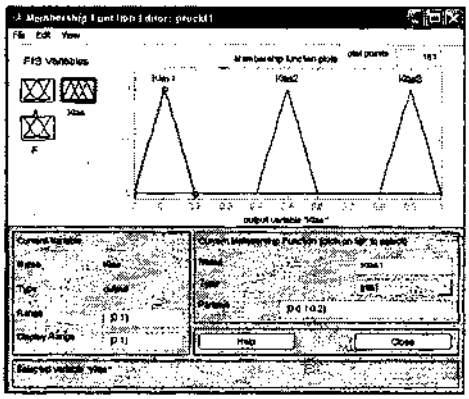


Fig. 3. Fuzzy logic for the output variable

These rules are introduced in the *Fuzzi Logic editor of Matlab* and can be visualized as follows :

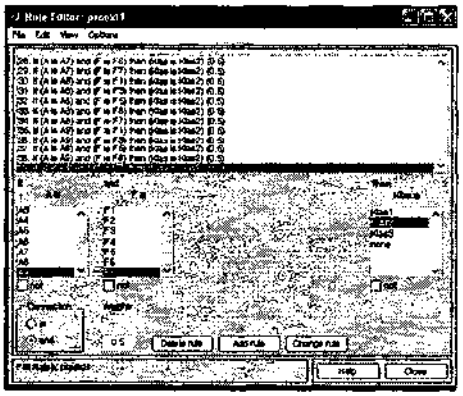


Fig.4. Fuzzy logic for introducing rules.

4.Computer simulation

The result from the operation of the *Fuzzy* system for the following input data is:

1. When A=0.69 and α=21° - the expected result is coding of the input into Group 2. The result from the *Fuzzy* system is shown in Fig. 5.

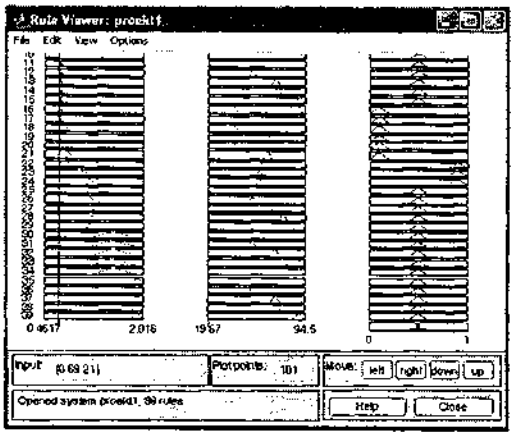


Fig.5. Result from the operation of Fuzzy logic

2. When A=0.73 and α=39° - the expected result is coding of the input into Group 1. The result from the *Fuzzy* system is shown in Fig. 6.

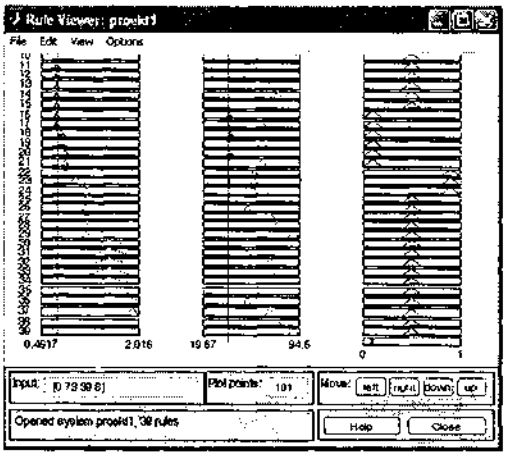


Fig.6. Result from the operation of Fuzzy logic.

3. When A=0.81 and α=63.1° - the expected result is coding of the input into Group 1. The result from the *Fuzzy* system is Group 1.

4. When A=0.856 and =59.7° - the expected result is coding of the input into Group 3. The result from the *Fuzzy* system is shown in Fig. 7.

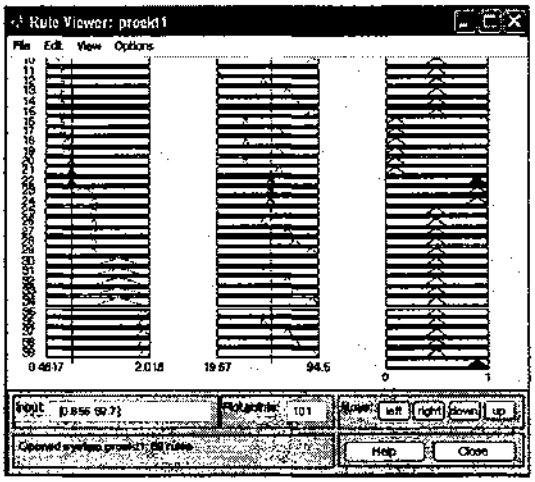


Fig.7. Result from the operation of Fuzzy logic

5. When A=1.17 and α=62.4° - the expected result is coding of the input into Group 3. The result from the *Fuzzy* system is Group 3.

6. When A=1.21 and =73.8° - the expected result is coding of the input into Group 2. The result from the *Fuzzy* system is Group 2.

Table 10. Rules for getting input data for a fuzzy conclusion

A	F							
	F1	F2	F3	F4	F5	F6	F7	F8
A1	Klas2	-	Klas2	-	Klas2	Klas2	Klas2	-
A2	Klas2	-	Klas2	-	Klas2	Klas2	Klas2	-
A3	Klas2	Klas1	Klas2	Klas1	Klas2	Klas2	Klas2	-
A4	-	Klas1	-	Klas1	-	-	-	-
A5	-	Klas1	-	Klas1	-	-	-	Klas3
A6	-	-	-	-	-	-	-	Klas3
A7	Klas2	-	Klas2	-	Klas2	Klas2	Klas2	Klas3
A8	Klas2	-	Klas2	-	Klas2	Klas2	Klas2	-
A9	Klas2	-	Klas2	-	Klas2	Klas2	Klas2	-

Conclusion:

The use of Fuzzy Inference as an element of an intelligent classificatory on the basis of a neural network provides correct and true coding of the input data for all the intervals.

References

1. Erich Peter Klement, Wolfgang Slany. "Fuzzy logic in artificial intelligence", 1996, ISBN: 0-8247-2287-6
2. http://lab.analogic.sztaki.hu
3. Lazarov, A.. D., Ch.M. Minchev. Algorithm for ISAR Target Recognition and Neural Network Architecture Implementation, In proceedings of ISPC 2003, Dallas, Texas, USA, March 31-April 3, 2003. Lazarov, A.. D., Ch.M. Minchev. Algorithm for ISAR Target Recognition and Neural Network Architecture
4. Гочев, Г. Компютърно зрение и невронни мрежи.Трудове на ТУ - София, 1995.

5. Дьяконов, В., В. Круглов. Математические Пакеты Расширения Matlab. Специальный справочник, Санкт Петербург, Питер, 2001 г.

ПРИЛОЖЕНИЕ НА FUZZY ЛОГИКА В КАЧЕСТВОТО НА РАЗМИТ ИЗВОД ЗА ОПТИМИЗИРАНЕ НА ВХОДНИ ДАННИ ПРИ РАЗПОЗНАВАНЕ НА ДИНАМИЧЕН ОБЕКТ

М. Костова

Резюме

Разглежда се етап от "интелигентна" система за разпознаване на летящ обект (самолет), извършваща класификация по признаци снети от контур.чрез невронна мрежа. С цел коректно и правилно кодиране на входните данни на невронната мрежа за всички интервали от стойности на признаците е синтезирана FUZZY система в качеството на размит извод, която използва алгоритъма на *Mamdani*.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

EGNOS DATA COLLECTION AND EVALUATION IN THE EASTERN AND SOUTERN EUROPE REGION – FIRST RESULTS

Boris Vassilev, Tsvetan Stoyanov

Technical University of Sofia
e-mail: tsstoianov@tu-sofia.bg

Abstract

This paper presents a summary of the activities of the Sofia branch of the Data Collection Network that Eurocontrol set up to perform data collections and analysis with the ESTB signal. The actual performance achieved at Sofia location is computed and checked against the RNP requirements, and any anomalies identified are analysed in detail to identify the cause.

1. Introduction

The existing satellite navigation systems GPS and GLONASS, as single systems, do not satisfy a number of user requirements, first of all for safety critical applications, in particular for precision approaches. The European Satellite-Based Augmentation System (SBAS) covering Europe is called the European Geostationary Navigation Overlay Service (EGNOS). EGNOS is the first phase of the European Union's policy on a global navigation satellite system (GNSS 1). EGNOS is being developed by the European Space Agency (ESA) in co-operation with the European Union (EU) and Eurocontrol.

The system provides additional signals to users of satellite navigation services, broadcast through geostationary satellites guaranteeing the integrity of GPS so that it can be used in support of life safety services such a civil aviation.

ESA is performing an extensive EGNOS verification campaign but this is focusing on the signal-in-space (SIS) as seen by a network of independent reference stations. Operational validation includes all activities that will demonstrate that EGNOS is ready to be used to support the flight operations for which it is intended. Within the frame of activities in preparation for the EGNOS operational validation, Eurocontrol has established a standardized data collection environment to perform regular EGNOS System Test Bed (ESTB) performance monitoring. ESTB is a prototype of EGNOS will a limited number of monitor stations. Monitor station have been set up at six different universities geographical by distributed around Europe. The ESTB has been providing a SIS since February 2000 to support the system development as well as to give potential users the opportunity to gain experience with EGNOS-like signals. Currently, this data collection network is using for daily records and first glance analyses and weekly and monthly data collection and evaluation, as well for analysis and assessment of the SBAS system performance. The actual performance achieved at each location is computed and can be checked against the PRN requirements, and in addition all revealed anomalies are analyzed in detail to identify the cause and the probability of re-occurrence.

2. Data Collection Activities in Sofia

The authors of this paper as representatives of the Technical University of Sofia are a part of this monitoring network since November 2003. A standardized data collection environment for the ESTB has been set up at the Department of Aeronautics. The precise position of the receiver (antenna phase center) in WGS-84 co-ordinates is:

[x, y, z] = [4313692.39, 1862642.84, 4299661.79] (in meters), and

[longitude, latitude, height (in meters)] = [23.3545800101, 42.652809391, 660.1147].

The used receiver is a NOVATEL OEM-3 "Milenium" with a Novatel Pinweel 600 S/N antenna. This L1/L2 receiver has 12 channels (11 GPS+1 SBAS). The data is logged on a personal computer using the SLOG software from Novatel. The broadcast signal complies with RTCA's Do229A(B) Minimum Operational Performance Standards (MOPS) and is broadcast through the geostationary satellite Inmarsat IOR (PRN 131).

The data is collected at 1Hz frequency and processed with Pegasus*Plus v3.x.x software. Pegasus is software prototype capable of processing receiver-

native data from a limited set of SBAS receivers and computing the position and integrity solution in accordance with the RTCA MOPS Do229 standards. It consists of three major software components. The CONVERTOR program translates receiver-native GNSS data into generic format. The WinGPSAll program uses the output of the CONVERTOR to determine a GNSS navigation solution. The ALGORITHMS programs uses the output of the CONVERTOR and the WinGPSAll programs to analyse the constellation, to determine predictive integrity monitoring qualifiers and to perform integrity monitoring using Receiver and Aircraft Autonomous Integrity Monitoring Algorithms. MATHLAB™ files are provided as a support for the user to automate and standardise the evaluation of the performance of the ESTB.

3. The ESTB Signal in Space

The raw navigation message of the SBAS contains 500 bits, transmitted in each second. It is 1/2 encoded with a Forward Error Correcting (FEC) Code. Thus, the baseline data rate of the SBAS SIS will be 250 bits per second. The block format for the 250 bits includes Preamble (8 bits), Message type identifier (6 bits), Binary message (212 bits) and Parity (24 bits). The message type identifier is a binary coded integer value (range 0-63), thus resulting in 64 different possible message types (MT) for the SBAS SIS.

4. Pegasus Data Processing

The real data collection and evaluation, together with the theoretical analysis, modeling and simulation is a part of the validation process of the EGNOS system. This process requires adequate demonstration of the accuracy, integrity, availability and continuity of the positioning service provided.

The measured pseudo range is corrected using the ESTB correction parameters. The corrected pseudo range will be [1,5]:

$$(1) \qquad \rho = \rho_{meas} + RC_{fast} - RC_{iono} + RC_{tropo} + RC_{clock}$$

with ρ_{meas} - measured pseudo range; RC_{fast}, RC_{iono}, RC_{tropo} and RC_{clock} - fast, ionospheric, tropospheric and satellite clock corrections.

The position solution is then calculated by means of a weighted least square algorithm [1,5]:

(2) $$x = \left(H^T W H\right)^{-1} H^T W R ,$$

with R – pseudo ranges; H – line-of-side matrix, W – weighting matrix. A column of the matrix H for a particular satellite is:

(3) $$H_i = \left[-\cos(E)\cos(A) \quad -\cos(E)\sin(A) \quad -\sin(E) \quad 1\right] ,$$

with E and A – elevation and azimuth of the satellite.

The weighting of the last squares is achieved by a matrix, which contains on its main diagonal a model of the pseudo range error after its correction by the transmitted ESTB parameters:

(4) $$W^{-1} = \mathrm{diag}\left(\sigma_i^2\right) .$$

Since the actual variance of the pseudo range measurement can not be observed in real-time, the variance for an individual satellite is modelled based on the model parameters supplied in the following equation:

(5) $$\sigma_i^2 = \sigma_{ifIt}^2 + \sigma_{iUIRE}^2 + \sigma_{iair}^2 + \sigma_{itropo}^2 ,$$

with σ_{ifIt}^2, σ_{iUIRE}^2 and σ_{itropo}^2 - variance of the residual error after application of fast and slow, ionospheric and tropospheric corrections, σ_{iair}^2 - variance of the contributions of the receiver to the residual error.

The contribution of the fast and slow corrections to variance of range measurement is determined mainly by the actual residual variance of each range correction and by taking the degradation of the variance with respect to time into account.

(6)

$$\sigma_{ifIt}^2 = \left\{ \begin{array}{l} \left(\sigma_{UDRE} + \delta\sigma_{UDRE} + \varepsilon_{fc} + \varepsilon_{rrc} + \varepsilon_{ltc} + \varepsilon_{er}\right)^2; \quad RSS_{UDRE} = 0 \\ \left(\sigma_{UDRE} + \delta\sigma_{UDRE}\right)^2 + \varepsilon_{fc}^{\ 2} + \varepsilon_{rrc}^{\ 2} + \varepsilon_{ltc}^{\ 2} + \varepsilon_{er}^{\ 2}; \quad RSS_{UDRE} = 1 \end{array} \right\}$$

with σ^2_{UDRE} - variance of User Differential Range Error (MT2 – MT5, MT6); $\delta\sigma^2_{UDRE}$ - increment for the variance of the UDRE (MT 27); ε_{fc}, ε_{rrc}, ε_{er}, ε_{ltc} - degradation parameters for fast correction data (MT 7), range rate correction data (MT 10), long term correction data (MT 10) and flight phases en-route through non-precision approach (MT 10); RSS_{UDRE} - root-sum-square flag for UDRE (MT 10).

The user Ionospheric Range Error Estimate is calculated by:

$$\sigma^2_{UIRE} = \left\{1 - \left[\frac{R_e \cos(E)}{R_e + h}\right]^2\right\}\sigma^2_{UIVE}, \quad (7)$$

with R_e – Earth radius (assumed to be 6378 km); E – elevation of satellite.

To determine an upper boundary of the vertical error for the location of the ionospheric pierce point, it is necessary to use the four or three point interpolation scheme [1,5]:

$$\sigma^2_{UIVE} = \sum_{i=1}^{4} W_i(x,y)\sigma^2_{ionogrid}, \quad \text{or} \quad (8)$$

$$\sigma^2_{UIVE} = \sum_{i=1}^{3} W_i(x,y)\sigma^2_{ionogrid}, \quad (9)$$

with $W_i(x,y)$ – weighting function; $\sigma^2_{ionogrid}$ - grid ionospheric vertical error boundary with degradation over time (MT 18 and MT 26).

The model for the residual error for the tropospheric delay estimate for a particular satellite is given as [1,5]:

$$\sigma^2_{itropo} = 0.12\left[\frac{1.001}{0.002001 + \sin^2(E_i)}\right]^2. \quad (10)$$

The standard deviation receiver noise for an SBAS satellite, including the multipath is:

$$(11) \qquad \sigma^2_{iair} = \sigma^2_{inoise} + \sigma^2_{imp} .$$

The variance of a normal distribution that models the residual miltipath error of an airborne subsystem can be obtained to [1,4]:

$$(12) \qquad \sigma^2_{imp} = 0.2 \exp\left(-\frac{E_i}{75}\right) .$$

As minimum requirements representing the worst case signal reception condition are used [1,4] $\sigma^2_{inoise} = 1.8$ m.

For certification of GNSS based navigation systems for aviation, it is necessary to guarantee that the user is informed of his position with sufficient integrity. The probability that the navigation system supplies the so-called hazardously misleading information (HMI) should be proven to remain extremely small. The integrity is specified in terms of the horizontal and vertical protection level (HPL and VPL), which is related to the probability that the alert limit may be exceeded. The SBAS protection levels are functions of the satellite constellation and the estimated SBAS performances. Thus, using SBAS correlation data, the protection levels can be determined without using actual pseudo range measurements. However, based on the pseudo range error model, the HPL and VPL provide an estimation of the upper boundary of the horizontal and vertical position error:

$$(13) \qquad HPL = k_h \sqrt{\left(H^T W H\right)^{-1}_{11} + \left(H^T W H\right)^{-1}_{22}} ;$$

$$(14) \qquad VPL = k_v \sqrt{\left(H^T W H\right)^{-1}_{33}} ,$$

with $k_h = 6.0$ (integrity risk: 2.10^{-9}) and $k_v = 5.33$ (integrity risk: 2.10)– horizontal and vertical level of integrity [3].

5. ESTB Results in Sofia

The receiver-native data obtained by the Sofia set of the SBAS receiver is processed by the PEGASUS software, and position and integrity solution

are computed in conformity with the above mentioned expressions. Daily First Glance reports and weekly reports are generated. These reports summarize the results obtained after applying the proposed algorithms on the measurements. In this way are generated performance values that can be checked against the PRN requirements. Summary of the performance obtained during the ESTB campaign in Sofia is presented in Fig. 1, 2 and 3.

Fig. 1 presents the position errors (HPE and VPE) of the navigation solution with respect to the precisely surveyed antenna location. They are obtained when applying all available differential corrections (fast, slow and ionospheric corrections) from the ESTB. The accuracy requirements are derived from ICAO's GNSS Standards and Recommended Practices (SARPS). **The obtained results prove that the horizontal and vertical requirements are fulfilled for all categories (NPA, APVI, APV II and CAT I).**

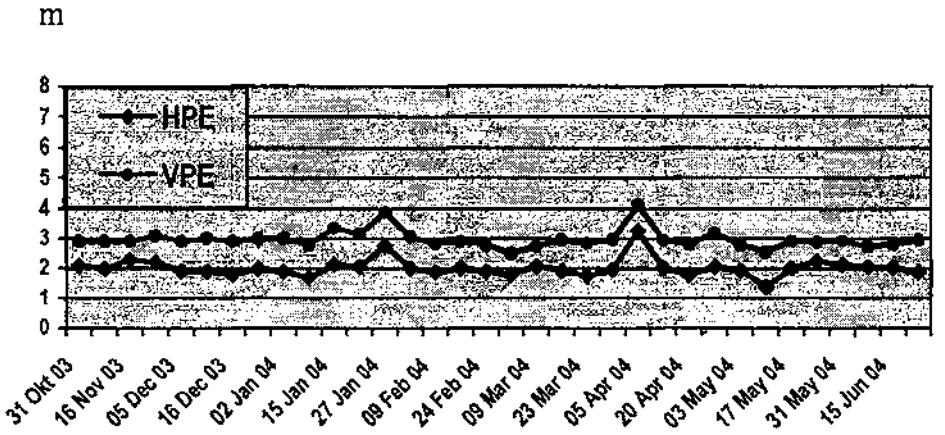


Fig. 1. Statistic data of the position errors

The computed protection levels (HPL and VPL), which represent the upper boundary of its position error are presented in Fig. 2.

The alarm limits against which a user has to compare its protection levels are defined in the ICAO's GNSS SARPS. We can see that **the**

horizontal integrity requirements for all categories are fulfilled (with one exception on January 27). The vertical integrity requirements are fulfilled for categories NPA and APV-I.

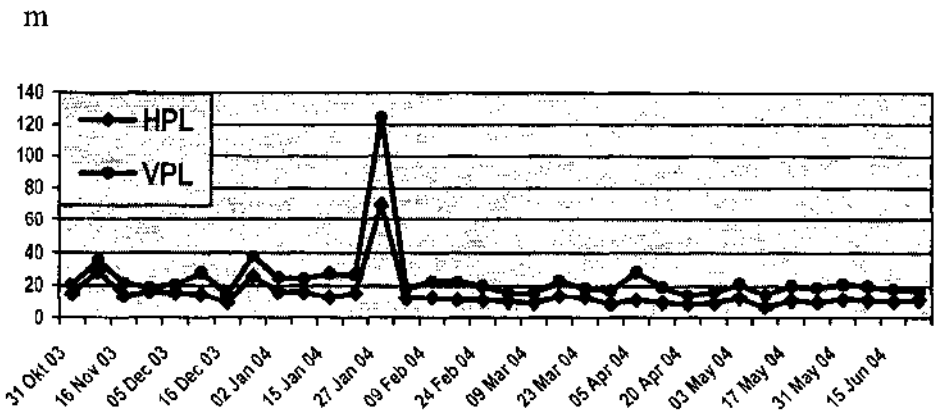


Fig. 2. Statistic data of the protection levels

The availability is defined as the percentage of time during which the system fulfils the accuracy, integrity and continuity requirements for the intended operation. The summary results obtained during the ESTB campaign in Sofia are presented in Fig. 3.

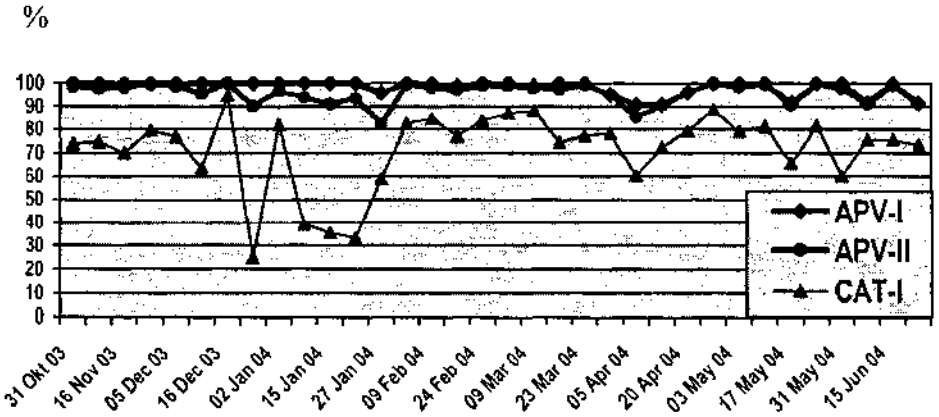


Fig. 3. Statistic data of the availability

It is important to highlight that the system was improving from the start (October 2003) to the end (July 2004) of the ESTB campaign in Sofia.

6. Anomaly Investigation

If any anomalies are encountered a detailed analysis will have to be done on possible measured integrity failures, discontinuity of service or other anomalies, to assess whether those are really related to system malfunctions or are caused by the data collection and valuation environment. The anomaly investigation aims to identify whether they were related to local effects such as multipath, antenna/receiver effects, SIS malfunctions, or other unexpected effects like ionospheric storms, GPS satellite clock malfunctions, etc.

Anomaly 1: Big jumps in XPL (sometimes reaching to several hundred meters). XPE is normal or little bigger than normal (Fig. 4).

The analysis shows that the possible reason for the integrity failures appearing is directly related to the contents of the GEO signal messages. As presented in Fig. 5 and 6, the fast corrections (MT 2 and MT 3 messages) are absent during the periods of the first and second jumps. **The cause of the third jump is the interruption of the broadcasting of all types of messages by PRN 131, except MT 0** (Fig. 7).

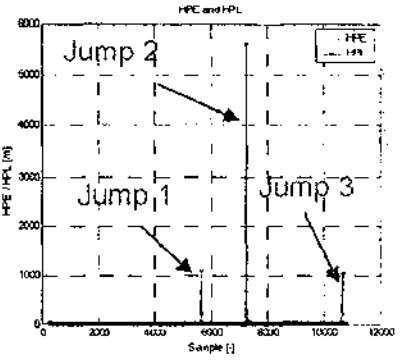


Fig. 4.

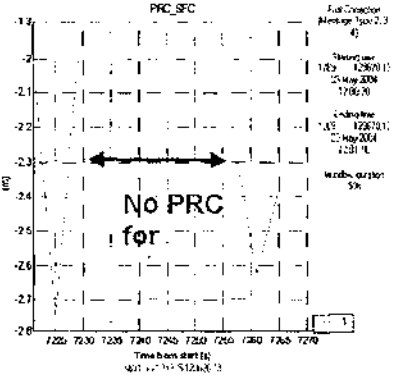


Fig. 5.

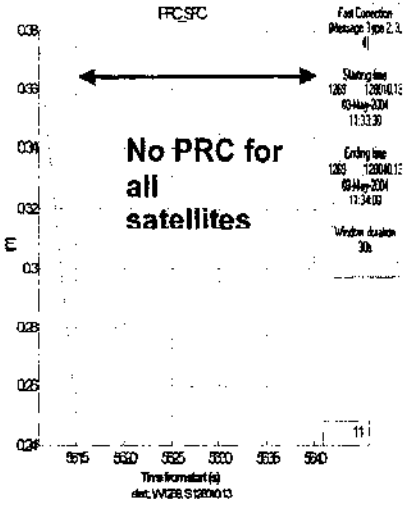


Fig. 6

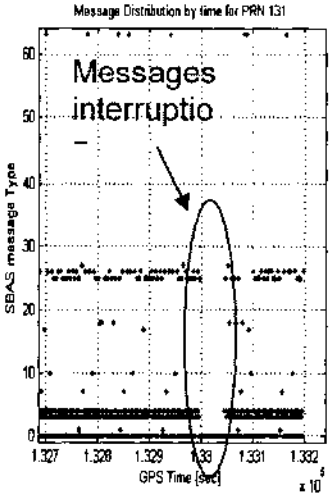


Fig. 7

Anomaly 2: No position data during a period of 7648 seconds (for GPS time from 173591 to 181239) and big peaks in XPL (sometimes reaching to several hundred meters) (Fig.8). The possible reason is the interruption of the broadcasting of all types of messages by PRN 131, including MT 0 (Fig. 9). **This causes the interruption of the ionospheric and fast corrections** (Fig. 10 and 11).

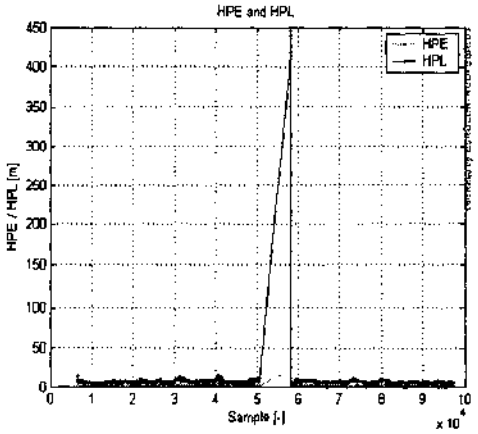


Fig. 8

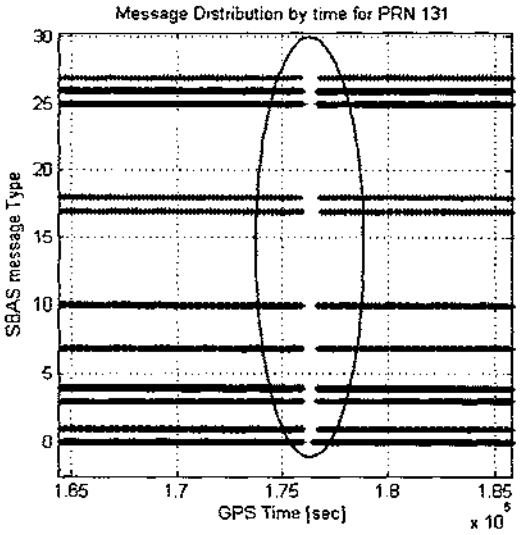

Message Distribution by time for PRN 131
SBAS message Type
GPS Time [sec]
x 10^5

Fig. 9

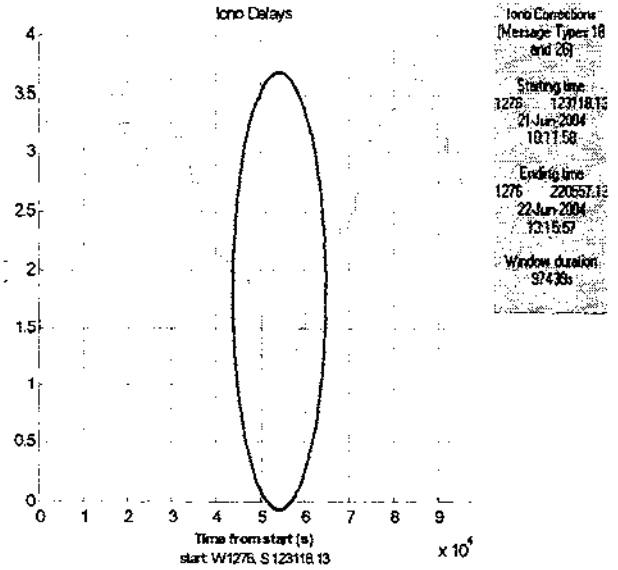

Iono Delays
Time from start (s)
start W1275, S123118.13
x 10^4
Iono Corrections
(Message Types 18 and 26)
Starting time
1275 123118.13
21-Jun-2004
10:11:58
Ending time
1275 220557.13
22-Jun-2004
13:15:57
Window duration
97439s

Fig. 10

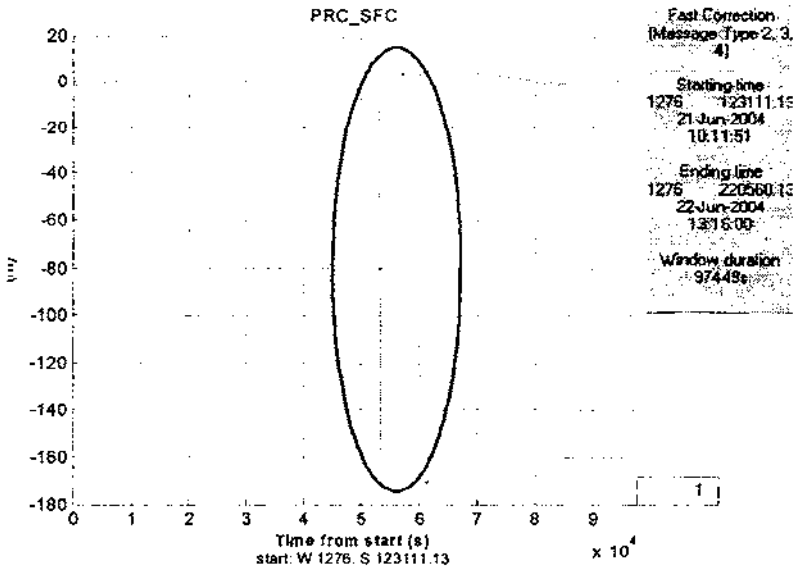


Fig. 11

Anomaly 3: A jump in the XPE with magnitude of several hundred meters and duration about 10 s. The XPL did not cover this jump (XPE>XPL>XAL for APV II and CAT I) and therefore HMI were created (Fig.12). The analysis shows that the possible causes of this event are the problems with the fast corrections. The MT 2 is transmitted at the start of the period and then repeated 4 times in the next epochs. This fact shows that there had been an "alarm situation" and that the ESTB had transmitted that information to the users (according to the standard requirements [1,2]). Since the CONVERTOR was running in SBAS mode 0/2, it interpreted an incoming MT0 with non-zero bits as MT 2. We discovered that indeed the MT0 transmissions contained bits set to"1". Thus, the explanation for XPE jump and HMI creation can be found in the ESTB SIS.

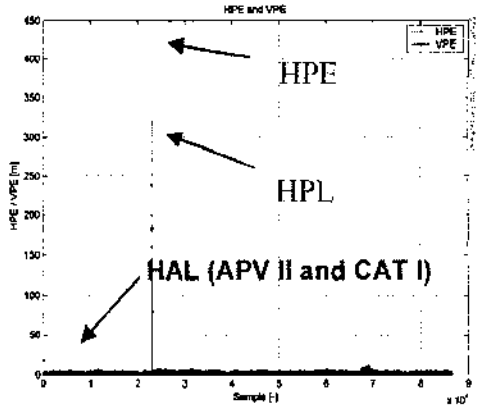


Fig. 12

7. Conclusion

The ESTB, a prototype of EGNOS, broadcasts ranging information, pseudo range corrections and integrity information for GPS. In order to gain experience with the Eastern and Southern Europe Region implementation of the SBAS system, confidence in the performance of that system has to be established. The good experimental results and their successful analysis demonstrate the working capacity and effectiveness of the ESTB monitoring station in Sofia. These results raise a lot of hope for the real EGNOS validation activity, which started in September 2004.

References

1. RTCA Do-229B, "MOPS for Global Positioning System/WAAS Airborne Equipment", October 6, 1999.
2. RTCA Do-229C, "MOPS for Global Positioning System/WAAS Airborne Equipment", November 28, 2001.
3. Soley S., A. van den Berg, R. Kremers, J. Sanz, Ch. Macabiau, A. Fonseca. "The Data Collection Network: EGNOS related", Eurocontrol EATMP Navigation Domain, April 2003.
4. Butzmuhlen C., R. Stolz, R. Farrnworth, E. Breeuwer. "PEGASUS – Prototype Development for EGNOS Data Evaluation – First User Experiences with the ESTB", Proceeding of the ION National Technical Meeting 2001.
5. PEGASUS*PLUS – Technical Notes on SBAS, GNSS Tools Team, September

СЪБИРАНЕ И ОБРАБОТКА НА ДАННИ ЗА СИСТЕМАТА EGNOS ЗА РЕГИОНА НА ИЗТОЧНА И ЮЖНА ЕВРОПА – ПЪРВИ РЕЗУЛТАТИ

Б. Василев, Цв. Стоянов

Резюме

В статията са представени първите резултати от дейността на софийския клон на мрежата, създадена от EUROCONTROL, за събиране и анализ на данни за европейската спътникова диференциална навигационна система EGNOS, намираща се в процес на валидация. Приведени са характеристиките на действащата на територията на ТУ-София станция, хардуерните и софтуерните и средства за приемане и обработка на навигационната информация. Представените резултати са получени с помощта на системата ESTB, представляваща тестови прототип на EGNOS.

В работата са изведени математическите зависимости, с помощта на които се определят основните характеристики на системата. На базата на приетите от GPS спътниците сигнали и сигналите от ретранслираща геостационарен спътник IOR (PRN 131) от системата INMARSAT, носещи навигационни съобщения с коригиращите поправки, са изчислени характеристиките на системата (accuracy, availability, integrity и continuity) за периода на действие на системата ESTB. Идентифицирани са и са анализирани детайлно най-характерните аномалии в получените резултати, като е търсена причината за появата им.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

A MODEL FOR USING MULTIPLE COMMUNICATION CHANNELS IN SIGNAL-SECURITY SYSTEMS*

Velin Kralev

Neofit Rilski South-West University
e-mail: vickson_kralev@abv.bg

Abstract

The paper describes the operation principles of signal-security systems using wireless channel for transmission of messages. The hardware and software shortcomings are described. The message transmission frame and the restrictions on objects' numbering imposed by it are discussed. A model for using multiple communication channels based on the available hardware is proposed. A method for logical extending of the message transmission frame is described, as well as the functionality of applied module, supporting multiple communication channels. New trends for development of signal-security systems are presented.

Key words: *signal-security system; communication channel; message frame.*

1. Introduction

Companies dealing with security, design and build security systems to restrict access to certain objects, watch the movement of people through the pertaining area of a certain object, signalling the violation of established restrictions and documenting violations. Each of these purposes is accomplished by an individual subsystem. All such subsystems, interacting

*Докладът е изнесен на Научната конференция с международно участие SENS 2006

between themselves and exchanging information, constitute an object's complete security system, named signal-security system [1].

Centralized security is implemented by monitoring centres, comprised of hardware and software. The hardware acquires the incoming signals and the software processes, analyzes, and stores the received messages, indicating the changes in the objects' status.

The communication between the different objects and the monitoring centre can be realized by means of different independent message transmission systems: wireless security system, telephone line security system and GSM transmitter security system [2].

2. Operation principles of signal-security systems

Each signal-security system has its message transmission technology. This technology includes installed hardware, both in the security objects and the monitoring centre.

The system software provides information about all registered objects, whereas the received signals are analyzed, printed and saved in physical files in the form of messages [3].

Wireless security system. The wireless security is realized through cellular digital wireless network using two wireless frequencies. The transmitters, located in the security objects, broadcast signals in one wireless frequency, in contrast to repeaters and receivers which communicate between themselves in another wireless frequency. Usually, each monitoring centre has an individual receiver station [2].

Each security object delivers a test signal over a certain span of time. This signal indicated the status of the security system.

Fig. 1 presents the connection circuitry between the hardware and the security object.

The designation of the inputs and outputs of the security station are: **AC** – alternating current, **AUX** – direct current output for detectors' power-supply, **GRND** – grounding, **SIRN** – siren output, $\mathbf{PGM_0}$, $\mathbf{PGM_1}$, ..., $\mathbf{PGM_n}$ – programmable outputs, $\mathbf{Z_1}$, **Z2**, ..., $\mathbf{Z_n}$ – inputs for zone detectors connected in series, **COM** – grounding, **R, G, Y, B** – keypad interface.

The outputs of the security station – $\mathbf{PGM_0}$, $\mathbf{PGM_1}$, ..., $\mathbf{PGM_n}$ are programmable. Each of them ($\mathbf{PGM_i}$) can be programmed to transmit two events, one when there is voltage (i.e. "high potential") and second, when there is no voltage (i.e. "low potential").

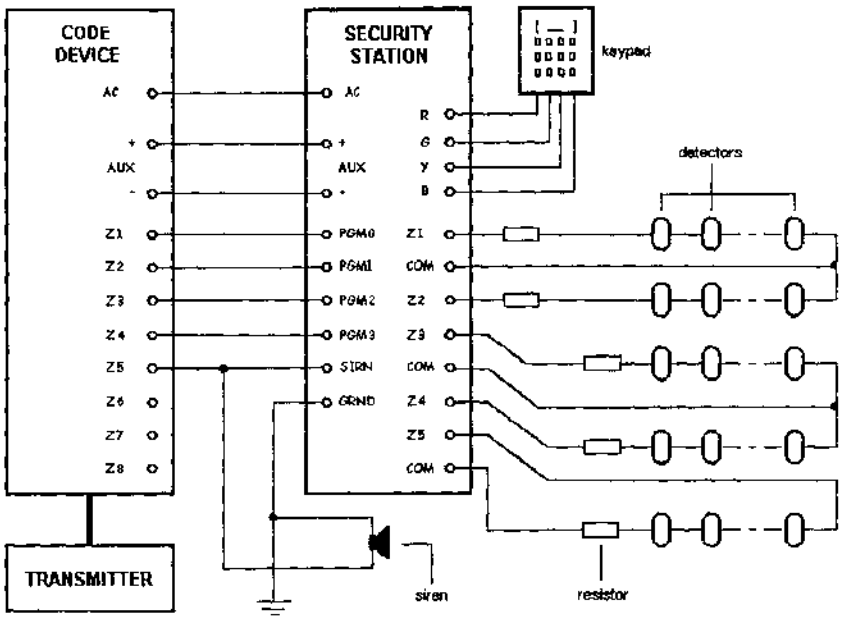


Fig. 1. Hardware connection circuitry

A coder and a transmitter are connected to the security station. They can be assembled in one module, or can be autonomous. When starting, the coder generates a message about the current status of all zones. The coder has its own control, comprising a central processor, an operating memory, registers, and message coding software. In the transmitter, the coded message is modulated by a higher frequency and is sent in the air.

The sent messages from the security objects are received in the monitoring centres. The receiver station demodulates (separates the low frequency from the high one), after which sends a message to the decoder. In the decoder, the received message is transformed into standard RS232 asynchronous signal and then is transferred bit by bit to a communication port of an application server. After analyzing the message by the application server, it is transferred to the database server where it is stored. Fig. 2 presents the schematic diagram of the described process.

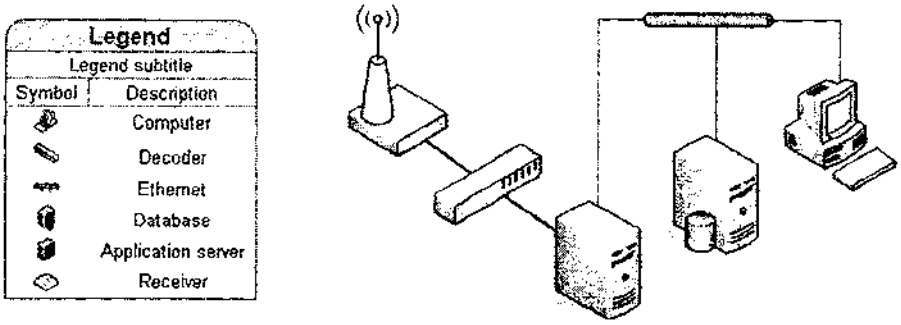


Fig. 2. Receiving, transforming, and storing a message

The frame of the message is fixed (i.e. with constant size) and its length is 24 bytes (24*8 = 192 bits). Fig. 3 presents an example frame of a message.

1	1	0	7	0	1	7	R	E	S	T	O	R	A	L		Z	O	N	E		1	#13
object number				message code			message text															end

Fig. 3. Example frame of a message

For each object, only 4 (four) symbols (0..9) are used, which sets the restriction for a total number of objects not exceeding 10,000. The paper suggests a solution for this problem.

3. A model for using multiple communication channels in signal-security systems

In the security station, the number of objects can be programmed and it may vary between 0 and 9999. The transmitter may be set up to broadcast signals of a second wireless frequency, thus creating a system of objects transmitting messages over a new transmission network. We shall name this transmission network "communication channel".

During this implementation it is possible that, in the monitoring centre, two identical messages are transmitted from two receivers to the decoder, the

difference between them being the frequency by which they are sent and the receivers that receive them.

The described problem can be solved in two ways:

Hardware solution to the problem. This solution requires reconstruction or replacement of a part of the hardware which codes and decodes the messages, received from the security objects. The purpose of the reconstruction is to extend the physical frame of the sent messages so that the object numbers be coded in more than 4 (four) symbols. This solution has the following disadvantages: the reconstruction will affect a great part of the hardware (coders, decoders or/and transmitters and receivers); it will take time and money. Taking into consideration these disadvantages, the conclusion may be drawn that it is necessary to find another, more effective solution in terms of time, money and use of the available devices.

Software solution to the problem. This article suggests another possible solution to the above-described problem. The idea is to use the possibility of building different communication channels, which is possible with the available hardware. The setting of the receivers and transmitters to another frequency does not resolve the problem with the length of the messages' frame, but only creates a new communication channel, which nevertheless provides to extend the frame of the message at its receiving. This solution to the problem has the following advantages: the available hardware can be used without reconstruction; a new communication channel requires only a program setting of a definite number of receivers and transmitters, using the available hardware, as well as new ones. A prerequisite for this solution to the problem is the design of an applied module providing the required functionality.

4. Requirements to the applied module, complying with the functionality of the proposed model

The applied module must provide for the extending the messages' frame, when they are received, but before being analyzed and saved. In this way, the maximal possible object number is increased and the problem of receiving two identical messages from two different objects is resolved. This can be achieved by adding to the number of the object in the message the number of the communication channel that has received the message. The applied module should provide for registration of the communication channels, security objects, messages from objects and hardware (including coders, decoders, transmitters,

receivers, and communication ports). Fig. 4 presents the frame of a message, that has been extended by the applied module.

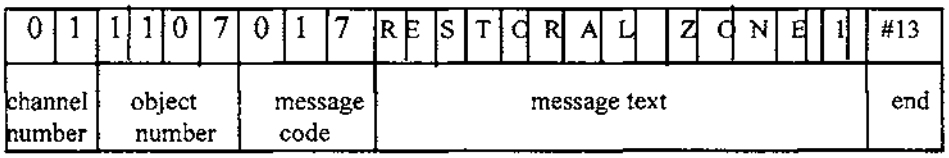


Fig. 4. Frame of the message after the number of the channel has been added to it

5. Conclusion

The size of the paper does not allow to describe all aspects of signal-security systems, but the proposed ideas might be used as a basis for further development of these systems, by building an integrated information system, oriented to them. This system should provide the for registration of hardware and messages, computer processing of documents, automated dispatching (GIS and GPS), Web- and WAP-based information system for customers, and data and applied modules security system.

References

1. http://www.elimpex.org
2. http://www.3ssot.com
3. http://www.sot.bg

МОДЕЛ ЗА ИЗПОЛЗВАНЕ НА МНОЖЕСТВО КОМУНИКАЦИОННИ КАНАЛИ В СИГНАЛНО-ОХРАНИТЕЛНИТЕ СИСТЕМИ

В. Кралев

Резюме

В статията са разгледани принципите на действие на сигнално-охранителните системи, използващи радиоканал за пренос на съобщения. Описани са хардуерните и софтуерните недостатъци на апаратното и програмно осигуряване. Разгледан е кадъра за пренос на съобщения и наложените от него ограничения за номериране на обекти. Предложен е модел за използване на множество комуникационни канали, с помощта на съществуващото апаратно осигуряване. Описан е начин за логическо разширяване на кадъра за пренос на съобщения. Описана е функционалността на приложен модул, поддържащ множество комуникационни канали. Представени са тенденциите за развитие на сигнално-охранителните системи.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

УПРАВЛЕНИЕ НА КАЧЕСТВОТО В ПРОЕКТИ ПО АКВИЗИЦИЯ - СТАНДАРТИ И ПРАКТИКИ*

Бранимир Жеков

Военна академия "Г.С. Ракавски"
e-mail: bzhekov@yahoo.com

Резюме

В доклада са посочени принципите за изграждане на съюзните публикации на НАТО за осигуряване на качеството AQAP's "2000", като е отделено специално внимание на тези, които определят изискванията към различните модели системи за управление на качеството. Дадени са критериите за избор на подходящ модел за включване като изискване в договора за реализация на проекта. Разгледани са концепциите на интегрирания системен подход към качеството през целия жизнен цикъл на отбранителните продукти. Посочено е съдържанието на процеса на управление на риска при реализирането на проекти по аквизиция и основните задачи, които следва да се решават в рамките на този процес.

Ключови думи: *качество, система за управление на качеството, отбранителни продукти, аквизиция, управление на риска.*

Качеството при управление на проекти

Политиката за модернизация и снабдяване на въоръжените сили на страните-членки на НАТО намира своята практическа реализация чрез точно дефинирани и всестранно обосновани инвестиционни национални и международни проекти, наричани проекти по аквизиция (придобиване).

*Докладът е изнесен на Научната конференция с международно участие SENS 2006

При това от особено значение е фактът, че качеството на постъпващите на въоръжение отбранителни продукти и системи определя в значителна степен очакваното удовлетворяване на съответните оперативни способности на войските. Финансовите разходи и разходите на време, свързани с разработването на тези проекти, изискват предприемане на конкретни мерки за гарантиране на тяхното качество, още преди да бъдат доставени от съответните контрагенти.

В отговор на важността на проблема "качество", в Международната организация по стандартизация ISO са създадени редица документи. Основен документ по проблемите на качеството е стандартът ISO 10006:2003 "Управление на качеството. Указания за качество при управление на проектите" [4].

Ще се спрем на няколко определения, съгласно ISO 10006:2003, които имат отношение към разглежданите проблеми в настоящата работа [1].

Проект е уникален процес, който се състои от съвкупност от координирани и управлявани дейности с дата на започване и на завършване, предприемани за постигане на цел, съответстваща на конкретни изисквания, включващ ограничения по срокове, цени и ресурси.

Трябва да се отбележи, че един отделен проект може да бъде част от структурата на по-голям проект, в някои проекти целите се усъвършенстват и характеристиките на продукта се определят в процеса на развитие на проекта, резултатът от проекта може да бъде една или няколко единици от продукта.

Качество - степента, до която съвкупност от присъщи характеристики изпълнява изисквания. Забележки към определението:

- терминът "качество" може да се използува с прилагателни като лошо, добро или отлично;
- "присъщи" за разлика от "приписани", означава съществуващи в нещо, особено като постоянни характеристики.

Изискване - потребност или очакване, което е обявено, обикновено се подразбира или се явява задължително.

Процес - съвкупност от взаимно свързани и взаимодействащи си дейности, които преобразуват входовете в изходи. Забележки към определението:

- за да се осъществи даден процес, освен необходимите дейности трябва да са налични и съответните ресурси - персонал, финанси,

съоръжения, инсталации, техники и методи;

- всеки процес има вход. На изхода са резултатите от процеса. Това са продуктите, материални или нематериални. Самият процес е едно преобразуване, което добавя стойност. Входовете на един процес обикновено са изходи на други процеси.

Продукт на проекта - това, което е дефинирано в обхвата на проекта и се доставя на потребителя.

Оценка на хода на проекта - оценка на резултатите от дейностите по проекта, провеждана през процесите на проекта, в подходящи пунктове от жизнения му цикъл, на база на определени критерии за процесите и продукта на проекта.

Може да се изисква ревизия на плана на проекта в резултат от оценката на хода на изпълнението на проекта.

Заинтересован от проекта - отделна личност или група индивиди с общи интереси в работата на организацията - доставчик и средата, в която тя работи. Забележки към определението:

- организацията - доставчик е организацията по проекта;
- заинтересовани от проекта могат да са клиент, потребител, собственик, партньор, финансиращ, подизпълнител, общественост, вътрешен персонал.

Процесите на проекта основно се групират в две категории [4]:

- процеси, свързани с управлението на проекта;
- процеси, свързани с продукта в резултат от проекта (касаещи единствено продукта от проекта, като планиране, разработване, производство и проверка).

Процесите, свързани с управлението на проекта, могат да бъдат обособени в следните групи (Фиг.1):

Първата група е така наречения *стратегически процес*. Чрез него се обосновава и определя насоката на проекта.

За стратегическия процес е характерно:

- задоволяване на потребностите на потребителя и други заинтересовани е от първостепенна важност;
- проектът се изпълнява като комплекс от планирани и взаимно зависими процеси;

➢ за да се постигнат целите на проекта е необходимо да се постави фокуса върху качеството на процесите и продуктите на проекта;

➢ ръководството е отговорно за създаване на подходяща среда за постигане целите по качеството;

➢ ръководството е отговорно за непрекъснатото подобряване.

Втората група *обхваща управлението на взаимните зависимости между другите процеси, преди всичко във времето на проекта - начало на проекта, управление на взаимодействието, управление на измененията, прекратяване.*

Фиг.1. Схематично представяне на десетте групи процеси, свързани с управлението на проекта

Принципите за управление на измененията се основават на международния стандарт ISO 10007:2003 "Системи за управление на качеството – Указания за управление на конфигурацията", STANAG 4159, издание 1 от 1999 г. „Политиката и процедурите на НАТО за общо управление на конфигурацията на материалите за многонационалните съвместни проекти", STANAG 4427, издание 1 от 1999 г. „Въвеждане на съюзните публикации по общо управление на конфигурацията и на съюзните публикации на НАТО за управление на конфигурацията ACMP's. Тази група процеси (Фиг.2) обхваща: начало на проект и разработване на план на проекта, управление на взаимодействието, управление на измененията, прекратяване.

Останалите осем групи са процеси, свързани с *обхвата, времето разходите, ресурсите, персонала, комуникацията, риска и закупуването.* От тях накратко ще разгледаме само процесите, свързани с времето [8].

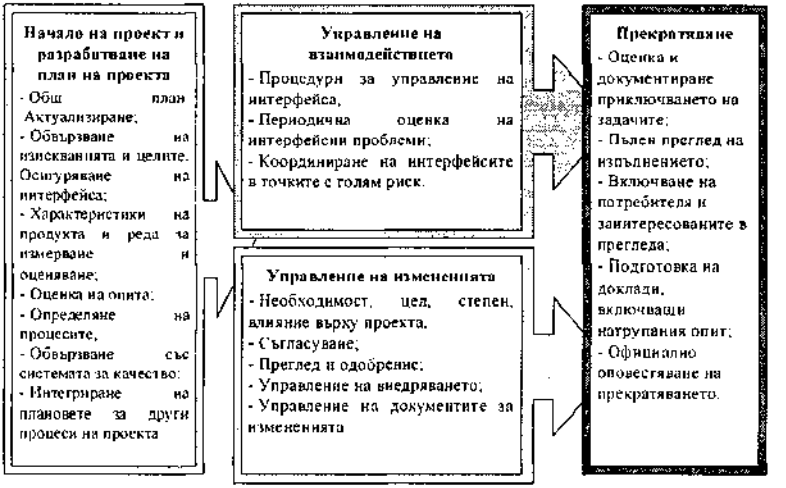
Основната цел на процесите, свързани с времето е да се определят зависимостите и продължителността на дейностите и да се гарантира навременното завършване на проекта. Процесите, свързани с времето са:

- *планиране на зависимите дейности* – определяне на вътрешните зависимости и логическите взаимодействия и зависимости между дейностите на проекта;
- *оценка на продължителността* – оценяване на продължителността на всяка дейност във връзка със специфичните условия и с необходимите ресурси;
- *разработване на програма* – взаимно обвързва времевите цели на проекта, взаимно зависимите дейности и тяхната продължителност като рамка за разработване на общи и подробни програми;
- *управление на програмата* – управление на реализацията на дейностите на проекта в съответствие с програмата или предприемане на съответни дейности за компенсиране на закъсненията.

Различни методи за управление на проекта (програмата) са намерили практическо приложение, като по-известни са:

- *график на Гант (лентов график):*

Графикът на Гант не представлява голяма статистическа ценност, но се използва като средство за оперативно въздействие върху хода на работата с цел осигуряване на нейното своевременно завършване.



Фиг.2 Взаимно зависими процеси на управление

➢ *метод на критичния път и метод ПЕРТ.*

Мрежовият график, създаден по метода ПЕРТ (метод за оценка и преглед на проектите), съдържа информация като в графика на Гант, но освен това показва връзката между събитията и определя какви операции е необходимо да бъдат изпълнени, за да настъпи едно или друго събитие. От графика се определя и критичния път.

Интегриран системен подход към качеството

В основата на политиката по качеството на НАТО е поставено разбирането, че управлението на качеството е непрекъснат, интегриран процес, който е основан на концепциите, свързани с фазите на жизнения цикъл, с процесите и участниците в него и със системата за управление.

Интегрираният системен подход е насочен както към управленските, така и към техническите елементи и се основава на следното:

➢ работата на организацията през различните фази от жизнения цикъл се осъществява чрез мрежа от взаимно свързани процеси, които

трябва да се идентифицират, анализират, управляват и усъвършенстват;

- отделните елементи са интегрирани в система;
- интересите на всички участници в жизнения цикъл се вземат под внимание;
- участниците в жизнения цикъл използват обща структура и терминология за създаване и управление на системата по качеството;
- принципите за управление на качеството и свързаните с това дейности се прилагат както към продукта, така и към всички процеси на жизнения цикъл.

Основани на договорните изисквания (включително на посочения в договора модел на система за управление на качеството в съответствие с някоя от публикациите на НАТО за осигуряване на качеството от серията AQAP's „2000"), дейностите по управление на качеството трябва да се прилагат за всички процеси.

Участниците директно включени в процесите на жизнения цикъл или в асоциирани към тях дейности са: потребител, купувач, собственик, доставчик и персонал, извършващ Държавна гаранция за качество (ДГК) [6].

За да постигнат максимална ефективност, организациите на участниците в жизнения цикъл трябва да изградят, документират, поддържат и развиват ефективни и ефикасни системи за управление на качеството [2, 3]. Системата за управление на качеството е тази част (подсистема) от системата за управление на организацията, която установява политиката по качество и целите по качество, и фокусира вниманието върху постигането на тези цели.

В Анекс С на AQAP 2000 са дадени някои широко използвани методи за оценка и подобряване - осемте принципа за управление на качеството от ISO 9004:2000 (ориентация към клиента, лидерство, съпричастност на хората, процесен подход, системен подход в управлението, постоянно подобрение, фактологичен подход при изработване на решения, взаимноизгодни отношения с поддоставчиците), наградите за качество Malcolm Baldrige, модел EFQM.

Съюзни публикации на НАТО AQAP's "2000"

Една от най-важните цели на стандартизацията, в това число и на стандартизацията в НАТО, е постигане на високо качество

(съответствие с предназначението, надеждност, безопасност, ергономичност, съвместимост, взаимозаменяемост и др.) при реализирането на продуктите и услугите, удовлетворяващи в най-висша степен потребностите от тях. Стандартизацията в НАТО повишава ефективността при използването на наличните отбранителни ресурси, като това включва, освен всичко друго, задълбочаване на сътрудничеството и елиминиране на излишното дублиране сред страните на Алианса при изследванията, разработването, производството, снабдяването и поддръжката на продуктите, предназначени за отбрана.

AQAP's "2000" [5] се базира на третата версия на фамилията ISO 9000:2000. ACMP's имат за основа международния стандарт за управление на конфигурацията ISO 10007, а ARMP-1 "Изисквания на НАТО за безотказност и ремонтопригодност" от серията ARMP's има за основа стандартите SAE JA 1000 (стандарт за "Програма по безотказност") и JA 1010 (стандарт за „Програма по ремонтопригодност"), заедно с техните ръководства SAE JA 1000-1 и JA 1010-1.

Използването на съюзните публикации на НАТО за осигуряване на качеството AQAP's "2000" [7] се основава на стандартизационното споразумение STANAG 4107 (издание 6 от 1997 г.), което определя и принципите за взаимното признаване на държавната гаранция за качеството (ДГК) на продуктите, предназначени за отбрана. Според него съществуват два вида документи AQAP's:

- *договорен тип* - AQAP 2110, -2120, -2130, -2131, -160, -150, -2105, които са във вид на "техническа спецификация", предназначена за договорно използване;
- *ръководен тип* – AQAP-2000, -2009, -159, 169, -2070, -2050, които осигуряват общото ръководство за оценка на използването на AQAP's от договорен тип, както и ръководство за личния състав на органа за ДГК при разработването на плановете за ДГК, посредством комплект от стандартизирани процедури.

Принципите и критериите за прилагане на договорните изисквания за осигуряване на качеството могат да бъдат обобщени както следва:

- съюзните публикации по осигуряване на качеството AQAP-2110, -2120, -2130, -2131, -160 са предназначени за изготвяне на предписание към договорните изисквания за управление на качеството;
- приложимостта или неприложимостта на някои от тези AQAP's зависи от вида и сложността на доставения продукт;

➢ подборът и приложението на подходящи изисквания за управление на качеството трябва да следва "процес на елиминиране", който да започва с решение относно необходимостта от прилагане на AQAP's. В договора не трябва да бъдат включени изисквания за прилагане на AQAP's когато:

- не е необходимо включване на специфични изисквания в договора за управление на качеството и не е необходимо да бъдат изпълнени действия по осигуряване на ДГК;
- характеристиките на заявения продукт са такива, че годността му би могла да бъде подходящо определена след получаването.

В случай, че прилагането на AQAP's е необходимо, процесът продължава с вземане на решение дали AQAP-2131 е достатъчен. В случай, че AQAP- 2131 е достатъчен и удовлетворява Националния орган по гарантиране на качеството, то този документ трябва да бъде договорно определен.

➢ аналогичен процес на елиминиране трябва да бъде извършен и по отношение на AQAP-2130. В случай, че AQAP-2130 е достатъчен, AQAP-2120 не трябва да се избира като изискване. AQAP-2120 трябва да бъде избран само, ако е сигурно, че AQAP-2130 няма да бъде ефективен. AQAP-2110 трябва да бъде включен само, ако договорът включва изискване за проектиране и разработване на продукти;

➢ когато разработването на софтуер е част от изискванията на даден договор, AQAP-160 трябва да бъде приложен в комбинация с AQAP-2110.

Заключение

Важна роля за успешното реализиране на проектите по аквизиция има всеобхватното и пълноценно прилагане на политиката по качеството на НАТО, базираща се на разбирането, че управлението на качеството е непрекъснат процес, протичащ през целия жизнен цикъл на отбранителните продукти и системи. Нейната практическа реализация води до провокиране на предизвикателства за отбранителната промишленост, акцентиращи върху дейностите по планиране, управление, осигуряване и подобряване на качеството, както на процесите, свързани с проектирането, разработването и производството на самите продукти, така и на процесите по управление на проектите при тяхната реализация.

Литература

1. ISO 9000:2000 "Системи за управление на качеството. Основи и речник."
2. ISO 9001:2000 "Системи за управление на качеството. Изисквания."
3. ISO 9004:2000 "Системи за управление на качеството. Ръководство за непрекъснато подобрение."
4. ISO 10006:2003 "Управление на качеството. Указания за качество при управление на проектите"
5. AQAP 2000 "Политиката на НАТО за интегриран системен подход към качеството през жизнения цикъл.", 2003 г.
6. AQAP 2070 "Ръководство на НАТО за предоставяне на пълномощия за държавна гаранция за качеството", 2004 г.
7. Съюзни публикации на НАТО за осигуряване на качеството от серията AQAP's "2000".
8. Георгиев В., Е. Тимева. Управление на проекти. Същност, съдържание, процеси и взаимодействие. С., Военно издателство, 2006.

QUALITY MANAGEMENT OF ACQUISITION PROJECTS - STANDARDS AND PRACTICES

B. Zhekov

Abstract

In the paper the principles of construction of allied publications of NATO for maintenance of quality AQAP's "2000" are considered and the special attention is allocated by what define requirements to different models of control systems of quality. Criteria of a choice of suitable model which is necessary for including as the requirement in the contract for realization of the project are given. Concepts of the integrated system approach at all stages of life cycle of defensive products are considered. The contents of managerial process by risk during realization of acquisition projects and the primary goals which should be solved within the limits of this process is specified.

Keywords: quality, control system of quality, defensive products, acquisition, management of risk

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

ЕДИН ПОДХОД ЗА ОЦЕНКА ЕФЕКТИВНОСТТА НА ИНФОРМАЦИОННА ЗАЩИТА*

Бранимир Жеков

Военна академия "Г.С. Ракавски"
e-mail: bzhekov@yahoo.com

Резюме

В доклада е представен кратък анализ на възможността за прилагане на системния подход като методология на анализа и синтеза на системи за информационна защита. Направена е оценка на приложимостта на съществуващата нормативна и методологическа база за оценка на ефективността на системи за информационна защита. Предложен е вероятностен подход за оценка на ефективността на системата за информационна защита, като са дефинирани съответни показатели на ефективността на системата и са дадени критерии за нейната оценка.

Ключови думи: *сигурност, информационна защита, системен подход, целева функция, качество, ефективност, показатели, критерии.*

Стремителното развитие и повсеместното внедряване на информационни и телекомуникационни технологии, през последното десетилетие, стана нов етап в икономическия и научно-техническия прогрес на човешката цивилизация, като при това все по-ясно се обозначава устойчивата тенденция към формиране на информационното общество.

*Докладът е изнесен на Научната конференция с международно участие SENS 2006

Ръстът на информационния фактор в съвременния свят, масовото създаване, внедряване и експлоатация на информационни системи са предизвикали и продължават да предизвикват възникване на доста значително множество от проблеми в сферата на сигурността на личността, обществото и държавата.

Няма съмнение, че за защита на критически важните информационни системи се разработват и на практика са в наличност множество международни, национални, отраслови, фирмени и други стандартизационни, нормативно-технически и методически документи. Въпреки това обаче, няма еднозначен и категоричен отговор на най важния въпрос – доколко предлаганото или реализираното решение е добро, каква е неговата планирана или реална ефективност. За съществуващото положение в областта на информационните системи, и по конкретно в областта на информационната сигурност, съществуват редица причини [6]:

- пренебрегване на системния подход като методология за анализ и синтез на системи за информационна защита;
- липсата на механизми за пълно и достоверно потвърждаване на качеството на системите за информационна защита;
- недостатъци на нормативната и методическата база по информационна сигурност, преди всичко в областта на показателите и критериите.

Системната парадигма

Реализацията на комплекса от мероприятия и мерки по осигуряване на информационна защита преди всичко е насочена към постигане на определена цел, т. е. тя има целево предназначение, което на формално ниво може да бъде представено посредством конкретна целева функция. При това може да се очаква, че желаемият резултат ще бъде постигнат в толкова по-голяма степен, колкото по-точно, по-еднозначно и по-конкретно е дефинирана самата функция, колкото по-задълбочено и по-прецизно са зададени и уточнени параметрите за нейното постигане, и колкото по-пълно са посочени съществуващите ресурсни, пространствено-времеви и други ограничения.

Ако целевата функция е с ниска степен на сложност и съществува сравнително висока вероятност за нейното постигане, т. е. тя може да бъде зададена чрез скаларен показател, то като правило нейната

реализация се осъществява чрез използване на ограничен и сравнително несложен по състав и структура комплекс от мероприятия и мерки.

При разширяване на кръга от проблеми, които следва да бъдат решени за постигане на цялостна и комплексна информационна защита, целевата функция придобива многомерен, векторен характер. При това коефициентът на значимост на отделно взетите елементи се намалява, а се повишава ролята и значението на общосистемните задачи – определяне на оптималната структура и режимите на функциониране, организация на взаимодействието между отделните елементи и подсистеми, отчитане влиянието на външната среда и др.

Целенасоченото обединяване на крайното множество от взаимодействащи и взаимозаменяеми функционални елементи на информационната защита и отношенията между тях формира тотално цяло с входни и изходни интерфейси, т.е. образува се системата на информационна защита. Отличително качествено свойство на агрегирането на отделните елементи на информационна защита в едно цяло (в система) е емергентността, което в случая изпълнява ролята на системообразуващ фактор, т. е. в новосъздаденото цяло се създават условия за възникване на т. нар. синергичен ефект. Същността на този ефект се свежда до появата на специфични, качествено нови свойства в цялото (системата), първоначално неприсъщи за нито една от неговите (нейните) съставни елементи. Трябва да се има предвид, че при агрегирането на елементите първостепенно значение имат само онези техни свойства, които определят взаимодействието помежду им и оказват влияние на системата като цяло, а също така и на изпълнението на дефинираната целева функция. При това посочените в [1] основни принципи за изграждане на съвременните сложни структурно-функционални построения могат да бъдат приложени и към системата за информационна защита, съответно в морфологичен, функционален и информационен смисъл:

- *съгласуваност* – структурно-функционална стандартизация, унификация, причинно-следствена и пространствено-времева свързаност;
- *ортогоналност* – модулност, агрегативност и автономност;
- *съответствие* – максимизация на критериалната ефективност по пълнота, достоверност, точност, надеждност;
- *икономичност* – ефект, ефективност, целесъобразност и

полезност;

- *прозрачност* – обща наблюдаемост и управляемост;
- *общност* – циклична завършеност, затвореност, устойчивост и сходимост;
- *откритост* – максимална потребителска и йерархично-управленска достъпност;
- *пълнота* – максимизация на функционалната ефективност при структурно-параметрична ограниченост;
- *реверсивна каузалност* – интелигентни положителни обратни връзки на управлението, надстроени над функционалната хомеостаза на системата.

От приведените по-горе принципи, принципът "икономичност" в смисъл ефект, ефективност, целесъобразност и полезност на системата за информационна защита представлява определен интерес от теоретична и практична гледна точка, поради което е обект на по-нататъшните съждения.

Като се има предвид, че в основата на реализацията на системния подход лежи циклично-реверсивното приложение на методите на анализа и синтеза [2], то резултативно решение при проектирането и изследването на системата за информационна защита не може да бъде получено само чрез логически съждения за нейното поведение при различни ситуации и условия. Преобладаваща част от проблемите, имащи конкретна практическа насоченост, изискват количествени, а не качествени оценки на съответните характеристики. Необходими са конкретни данни, които да разкриват свойствата на системата. В този смисъл обикновено възниква остра необходимост от количествена мярка за качеството на съответствие.

За оценка на качеството на съответствие на системата за информационна защита в редица случаи е целесъобразно използването на едно от основните свойства на системата, а именно ефективността на нейното функциониране, под която съгласно [3], се разбира някаква количествена характеристика на качеството и обема на изпълняваната от системата работа по информационната защита, или с други думи степента на съответствие на резултатите от защитата на информацията на поставената цел. Като количествена характеристика на качеството на съответствие, ефективността е непосредствено свързана с други системни свойства, като качество, надеждност, устойчивост, управляемост,

шумозащитеност. Поради това количествената оценка на ефективността позволява да се измерват и обективно да се анализират основните свойства на системата за информационна защита на всички етапи от нейния жизнен цикъл.

Акредитация, сертификация, качество

Информационните системи, в които се създава, обработва, съхранява и пренася класифицирана информация, подлежат на задължителна акредитация в съответствие с действащите законови и нормативни документи. При това обаче възниква въпросът доколко сертификацията на съответствие на изискванията е най-добрият инструмент, с каква степен на достоверност се потвърждава това съответствие и дава ли тя необходимите гаранции. И ако се допусне, че степента на достоверност все пак е известна, означава ли това, че този термин е еквивалентен на вероятностно-статистическото разбиране на това понятие. Нещо повече, контролът и отговорността за достоверността на резултатите са непосредствено възложени на съответните сертифициращи органи (лаборатории), т. е. те сами извършват измерванията и сами се контролират. При една такава ситуация нормативното изискване за осигуряване на определена достоверност на резултатите от изпитванията на отделните средства, и още повече на цялата система за информационна защита, може би ще остане трудно осъществима декларация.

Неопределеността на постигнатия резултат се засилва и от факта, че много често заявителите на системи за информационна защита не са запознати в достатъчна степен с функционалните възможности на отделните средства и със степента на тяхното влияние върху общото ниво на информационна сигурност и извършват необосновани разходи. В резултат на това, заявителя не винаги получава това, от което реално има нужда и не е в състояние обективно да провери и оцени качеството и ефективността на предложеното решение.

Както беше посочено по-горе, затрудненията при обективното потвърждаване на ефективността на информационната защита произтичат както от съществуващата нормативна база, така и от недостатъчно разработената система от показатели за информационна сигурност [4].

Не удовлетворява напълно потребностите и съществуващата система от критерии за информационна сигурност, в това число и такива, като ефективност на системата за информационна защита. Към сериозните проблеми следва да бъде отнесено и честото пренебрегване на стохастичната природа на събитията и явленията, които възникват в процеса на защита на информацията, абстрахиране от тяхното икономическо съдържание в нормативен, методически и приложен аспект [5].

По този начин, съществуващите национални и международни стандарти и документи на тяхна основа не дават отговор на редица ключови въпроси [6]:

- Как да се създаде информационна система, която да притежава информационна защита със зададено, измеримо, обективно проверяемо ниво?
- Как на практика да се организира режима на информационна защита и той да се поддържа в условията на постоянно изменящата се външна среда и структура на самата система?
- Какво е реалното ниво на сигурност и колко е ефективна системата за информационна защита?

Показатели и критерии

Тъй като качеството на който и да е обект, в това число и системата за информационна защита, съгласно [7], се проявява само в процеса на неговото използване по предназначение, т.е. при неговото целево функциониране, то най обективната оценка на този обект е оценката по ефективността на неговото използване.

Проектирането, организацията и използването на системата за информационна защита фактически са свързани с неизвестни събития в бъдещето и поради това винаги съдържат елементи на неопределеност. С реализацията на проекта нивото на тази неопределеност се намалява, но никога ефективността на системата за информационна защита не може да бъде адекватно изразена и описана чрез детерминирани показатели. Процедурите по изпитване, сертифициране или лицензиране не отстраняват напълно неопределеността на свойствата на системата за информационна защита или на отделните нейни елементи и не отчитат случайния характер на атаките. Ето защо обективна характеристика на

качеството на системата за информационна защита – степента на нейната адаптируемост за достигане на зададеното ниво на информационна сигурност в условията на реално въздействие на случайни фактори, може да служи вероятността, характеризираща степента на възможностите конкретната система за информационна защита, при зададен комплекс от условия. Съгласно общата теория на системите такава вероятност се нарича вероятност за достигане на целта на операцията или вероятността за изпълнение на поставените задачи пред системата. Дадената вероятност следва да бъде поставена в основата на комплекса от показатели и критерии за оценка на ефективността на системата за информационна защита. При това в качеството на критерии за оценка могат да се използват понятията годност и оптималност, като годност означава изпълнение на всички изисквания, поставени към системата за информационна защита, а оптималност – достигане на една от характеристиките на оптимално значение при спазване на ограниченията и условията на другите свойства на системата. При избора на конкретен критерий е необходимо той да бъде съгласуван с целта, поставена пред системата за информационна защита. В Табл.1 са дадени възможни показатели за ефективност, а в Табл.2 – възможни критерии за ефективност на система за информационна защита [6].

Таблица 1

№	Изисквания към системата	Показатели за ефективност
1.	Поява или отсъствие на събитие	Вероятност на събитието
2.	Достигане на зададени характеристики	Вероятност за достигане на резултат не по-нисък от зададено ниво
3.	Отклонение от зададени характеристики	Средно квадратично отклонение на резултата от зададена стойност
4.	Осигуряване на гарантирано ниво на характеристики	Квантил на зададено ниво на гаранция
5.	Не са предявени	• Математическо очакване на резултат • Дисперсия на резултат • Среден риск

Таблица 2

№	Концепция за ефективност на системата	Критерии за ефективност
1.	Годност	• Приемлив резултат • Допустима гаранция • Допустим гарантиран резултат
2.	Оптималност	• Най-добър резултат • Най-добър среден резултат • Най-голяма вероятност за гарантиране на резултат • Най-голям гарантиран резултат

В съвременните нормативни документи по информационна защита, както е известно, основно се използва квалификационният подход. Значително по-конструктивни на практика са вероятностните методи, намерили широко разпространение в други области на науката и практиката. Съгласно тези методи, нивата за гарантиране на сигурността на системата за информационна защита се трансформират в доверителни вероятности на съответните оценки на показателите. За решаването на тази задача може да се използва теорията на статистическите решения [7], позволяваща да се определят оптимални нива на гарантиране на сигурността.

Преди всичко, оценката на оптималното ниво на гарантиране на сигурността в определена степен зависи от щетата, свързана с грешката при избора на конкретното значение на показателя на ефективността. Освен това, за получаване на числени оценки на риска е необходимо да бъде известно разпределението на множество случайни величини. В редица практически случаи такива оценки е възможно да бъдат получени по резултатите от активните одити на системите за информационна защита или с помощта на симулационно моделиране.

Симулационно моделиране

Както е известно, целта на програмната симулация е чрез проиграване на създадените за целта симулационни модели да се изследва поведението на обекти при различни параметри на въздействуващите

фактори и анализиране ефекта, който тази промяна оказва върху крайния резултат, при характерните за това предимства – изследването се извършва в лабораторни условия, лисват проблеми, свързани с предизвикване на възможни реални повреди, наличие на определена икономическа ефективност.

Симулационото моделиране освен за оценка на разпределението на случайни величини, необходими за прилагане на теорията на статистическите решения, може да намери полезно практическо приложение, съгласно [8], за изследване на ефективността на информационната защита и по-конкретно при анализ на риска от потенциални атаки, оценка на вероятните щети, оценка на надеждността на топологията, симулиране на атаки и др.

Съществена роля при използване на симулациите в областта на информационната защита имат инструментите за моделиране на мрежи. Техен основен представител е продукта Opnet (Optimized Network Engineering Tool) на фирма OPNET Technologies Inc. и по конкретно неговия модул, предназначен за анализ на мрежовата сигурност, NetDoctor.

Независимо от сложността на мрежовите симулатори, симулацията остава един добър подход за анализ на процесите в реалните мрежови конфигурации и за оценка на ефективността на тяхната защита.

Заключение

Прилагането на системния подход при получаване на количествени оценки на ефективността на информационната защита създава условия за конкретна обективност и значимост на резултатите. Вероятностният характер на процесите, голямата динамика на технологичните иновации и разнородните заплахи предопределят симулационното моделиране в областта на информационната защита като един перспективен и ценен подход.

Литература

1. B l a a u w, G. Coputer architecture. Electronische Rechenanlagen. 14, №4, 1972.
2. С е м е р д ж и е в Ц. Инструменти за стратегическо ръководство "C⁴I". С., Софттрейд, София, 2001.
3. У ш а к о в И. А. Надеждность технических систем. М., Радио и связь, 1985.
4. ISO/MEC 15408-99 " Criteria of an estimation of safety of information technologies."

5. Баутов А. Стандарты и оценка эффективности защиты информации. Доклад на Третьей Всероссийской практической конференции "Стандарты в проектах современных информационных систем", М., 2003.
6. Баутов А. Эффективность защиты информации. Открытые системы №7, 2003.
7. Петухов Г. Основы теории эффективности целенаправленных процессов. Часть I, Методология, методы, модели. М., МО, 1989.
8. Пугачев В. Теория вероятностей и математическая статистика. М., Наука, 1979.
9. Saunders, J. *Simulation Approaches in Information Security Education.* Presented at the National Colloquium on Information Systems Security Education, Seattle, WA, 2002.

ONE APPROACH FOR AN ESTIMATION OF EFFICIENCY OF INFORMATION PROTECTION

B. Zhekov

Abstract

In the report the short analysis of an opportunity of the appendix of the system approach as methodology of the analysis and synthesis of systems of information protection is presented. The analysis of applicability of existing normative and methodological maintenance for an estimation of efficiency of systems of information protection is made. The likelihood approach for an estimation of a system effectiveness of information protection is offered, corresponding parameters of system effectiveness are certain and criteria of its estimation are given.

Keywords: safety, information protection, system approach, criterion function, quality, efficiency, parameters, criteria.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

СЪВРЕМЕННИ СРЕДСТВА ЗА ЗАЩИТА И ЛИКВИДИРАНЕ НА ПОСЛЕДСТВИЯТА ПРИ ТЕРОРИСТИЧНИ АТАКИ С ТОКСИЧНИ ХИМИЧЕСКИ ВЕЩЕСТВА*

Иван Попов[1], Георги Попов[2]

[1]Военна академия "Г. С. Раковски"- Институт за перспективни изследвания за отбраната
[2]Кингстън Енвайрънментал Сървисиз, Канзас Сити, САЩ

Резюме

Конвенционалната стратегия на държавите и техните армии, отнасяща се до воденето на "химическа война", претърпя особена преоценка след 11 септември, когато тероризмът започна да се разглежда като съществена заплаха за националната сигурност. От самото си създаване химическото оръжие се е считало и се е разглеждало като оръжие на слабите. В момента американските специалисти го оценяват като най-мощното и вероятно за използуване при очакваните терористични атаки. Един от най-съществените елементи на химическа защита са средствата за защита на дихателните органи и кожата. В доклада са разгледани съвременните средства за индивидуална противохимическа защита на дихателните органи и кожата като ефективни средства за ликвидиране на последствията в огнища на поражение с токсични химични вещества, употребени от терористи. Разгледани са основните характеристики на средствата за защита и възможностите за използването им в различни ситуации. Описани са способите и средствата за оказване на първа помощ на поразените и ликвидиране на последствията в районите, заразени

*Докладът е изнесен на Научната конференция с международно участие SENS 2006

с токсични вещества. Оценени са предимствата и недостатъците на съвременните средства за специална и санитарна обработка. Посочени са перспективите на развитие и усъвършенстване на средствата за защита, специална и санитарна обработка.

Десетилетия наред се водят конструктивни преговори за забрана и унищожаване на радиологичните, химическите и биологическите (РХБ) оръжия. Подписаните и ратифицирани конвенции и договори недвусмислено показват, че човечеството се приближава към пълната забрана и ликвидиране на натрупаните запаси от тези оръжия. Независимо от това, събитията през последните години (ядрени аварии, терористични актове, скрито производство на химични и биологични оръжия) показва, че вероятността войсковите формирования, участващи в разнообразни мисии и операции да срещнат РХБ агенти е висока. Това налага своевременно изучаване на поразяващото въздействие на РХБ агентите, ефективните средства за защита от тях, както и способите и средствата за тяхното отстраняване и обезвреждане от заразените обекти.

Важен дял в арсенала на терористичните организации заемат химическите оръжия. Химическото оръжие дълго време е разглеждано като "атомната бомба на бедните" поради ниската си себестойност и лесното му производство. Химичните агенти лесно се разпръскват и заразяват широки области, а откриването им е трудно [1,2].

Терористите предвиждат тежки поражения от употребата на химични агенти, разчитайки и на фактора на шока.

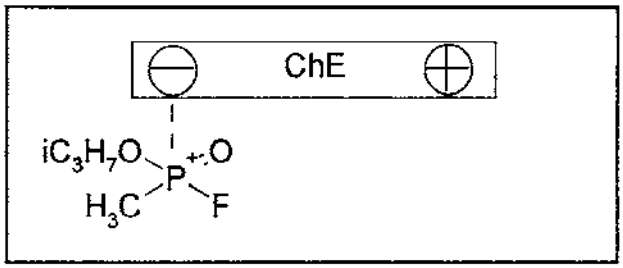
Основните изисквания за висока ефективност при употребата на токсични химични агенти са да притежават висока токсичност, да са сравнително устойчиви към водата, кислорода във въздуха и светлината и да притежават физични и химични свойства, осигуряващи въвеждане на токсичните вещества в приземния слой на атмосферата в парообразно или аерозолно състояние.

Най-пълно отговарят на тези изисквания фосфорорганичните бойни отровни вещества от типа на зарина и V-газовете.

Флуоранхидирида на изопропиловия естер на метилфосфоновата киселина - зарин е пригоден за използване целогодишно. Температурата му на кипене е 151С, а на топене - 53°С. Има висока летливост ($C_{max}^{20^{\circ}C} = 13\,mg/l$). Смъртоносната му концентрация при експозиция 1 min с 0,1 mg/l.

По характер на токсично действие заринът се отнася към отровните вещества с нервнопаралитично и миотично действие. Той е ферментна отрова, предизвикваща блокиране на хлоинестеразата, което води до тежки поражения на нервната система и целия организъм [2]:

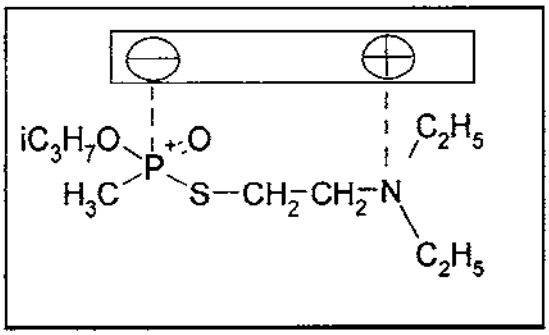
(1)



Токсичното му действие се проявява при всеки способ на попадане в организма и предизвиква поражения както в парообразно, така и в капкотечно състояние. Малки концентрации на зарина предизвикват миоза и стягане в гърдите. Смъртоносната доза за човека при инхалация LCt100 е около 0,1 mg.min/l, а кожнообриваната му доза е 7-9 mg/kg. Със зарин може да се заразяват големи площи за продължително време, а високата му летливост осигурява разпространение на парите му на големи разстояния (до 10-12 km) [2]. Зарин бе използуван от терористи в токийското метро през 1995 г.

Фосфорилтиохолините (V-газовете) притежават по-висока токсичност от зарина. Причина за това е наличието в молекулата на диалкиламинови естери с наличие на анионен център, който способства за по-бързо и по-трайно блокиране на ензима холинестераза.

(2)

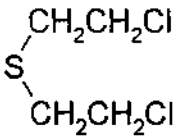


Например за нанасяне на смъртоносни поражения на човека върху незащитена кожа са необходими 0,005 mg, докато за зарин тази стойност е 0,5 mg. Основното бойно състояние на V-газовете е аерозолно поради незначителната им летливост и високата температура на кипене (около 280С). Предвидено е да се използват чрез снаряди, ракети, авиобомби, контейнери и др. На практика се цели да се създаде такава аерозолна концентрация, способна да предизвика най-бърз смъртоносен ефект, дори при едно вдишване.

Ефективно средство за оказване първа помощ на поразените с нервно-паралитични отровни вещества са шприц-ампулите в индивидуалния защитен пакет. Антидотът в ампулите осигурява бързо и ефикасно деблокиране (реактивиране) на ензима холинестераза и възстановяване нормалната функционална дейност на организма.

Макар и с по-малка токсичност, кожнообривните отровни вещества също се предвиждат в арсенала на терористите.

Основен представител на този клас бойни отровни вещества е $\beta\ \beta'$ -дихлордиетилсулфидът с тривиално название

иприт - 

Ипритът е общоотровно вещество с ярко изразено кожнообривно и задушливо действие. Химически чистият иприт представлява безцветна течност с температура на замръзване 14,4°С и температура на кипене 217°С. Максималната му летливост $C_{max}^{20^{\circ}C} = 0{,}625\,mg/l$, а относителната му маса е 1,27. Създава устойчиви замърсявания – през лятото до няколко денонощия, а през зимата до няколко седмици. За предотвратяване на по-тежки последици от поразяващото действие на иприта от особена важност е бързото отстраняване на капките на иприта и своевременната обработка на заразеното място с универсалния дегазиращ разтвор от индивидуалния защитен пакет.

Индивидуални средства за противохимическа защита на личния състав са средствата за защита на дихателните органи и кожата.

Средствата за защита на дихателните органи са един от най-съществените елементи на противохимическата защита.

Независимо от съществуващото разнообразие на средства за защита на дихателните органи основни си остават трите типа, с които най-масово са снабдени армиите - филтриращи полумаски, филтриращи противогази и изолиращи противогази [1]. Армиите от различните страни използват основно филтриращи полумаски от клас Р3. Маските от този клас имат много-добри защитни свойства по отношение на финни токсични и бактериални аерозоли, радиоактивни аерозоли и течни и твърди аерозоли на съвременните дразнещи агенти. Те трябва да отговарят на новия стандарт EN 149:2001, който регламентира минималните изисквания към тези средства [4].

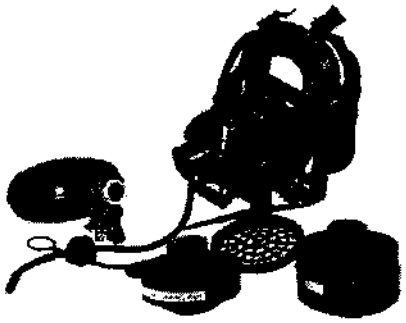
Всички армии в света са снабдени на 100 % с филтриращи противогази и респиратори.

С помощта на филтриращите противогази се решават всички основни задачи за защита на дихателните органи в условията на военни действия и терористични атаки. Те са предназначени за защита на дихателните органи, лицето и очите от въздействието на БОВ, радиоактивни и бактериални аерозоли.

Основни характеристики на приетият на снабдяване в БА филтриращ противогаз ПФ-90, производство на „Зебра" АД са:

- съпротивление при вдишване при непрекъснат въздушен поток 30 л./min – максимум 0,3 μbar;
- съпротивление при издишване при непрекъснат въздушен поток 30 л./min – максимум 0,8 μbar;
- коефициент на просмукване на маслена мъгла /при концентрация на маслената мъгла 2500 мг/м3 – до 0,0001 %;
- период на защита на кожата на лицето от въздействието на капкотечни отровни вещества – минимум 8 часа;
- общо поле на зрение – минимум 70 %;
- бързо и удобно поставяне и сваляне посредством 6 – точков гумен закрепващ елемент.

При извършване на дейности в райони с недостиг на кислород или комбинирано заразяване и с промишлени отровни вещества се използват въздушно-дихателни апарати.



Фиг. 1. Филтриращ противогаз ПФ-90

Фиг. 2. Въздушно-дихателен апарат „Дрегер"

Тенденциите в усъвършенстването на въздушно-дихателните апарати са:

- олекотяване на бутилките и увеличаване продължителността на използване;

- оптимизиране състава на подавания въздух (кислород, влага и др.);

- внедряване на алармени системи.

Ефективни средства за защита на кожата от радионуклеиди, бойни отровни вещества и биологични агенти са филтриращите защитни облекла. С най-добри характеристики са облеклата на фирма „Блюхер" със защитна материя "Saratoga":

Основни характеристики:

- защита от пари и капки на бойни отровни вещества – не по-малко от 12 h;

- устойчивост срещу топлинна радиация – импулс от 60 J/cm² – за 66 s.;

- пропускливост на въздух – минимум 167 mm/sek, максимум – 915 mm/s.

Фиг.3. Филтриращи защитни облекла, производство на фирма „Блюхер"

При дейности с високи нива на риск (недостиг на кислород, пожари, свързани с отделяне на разнообразни токсични вещества и др.) се използват изолиращи костюми за многократна употреба. С най-добри характеристики са облеклата на фирма „Дрегер":

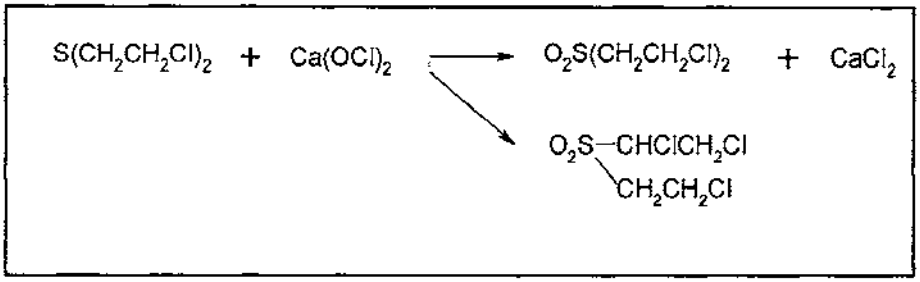
Фиг. 4. Изолиращи защитни облекла, производство на фирма „Дрегер"

Важен дял при ликвидирането на последствията от употребата на радиологични, химични и биологични агенти заема деконтаминацията.

Токсичните химични агенти могат да бъдат отмити и отстранени, изсушени, обезвредени посредством химически активни рецептури и сорбиращи субстанции или посредством термична обработка.

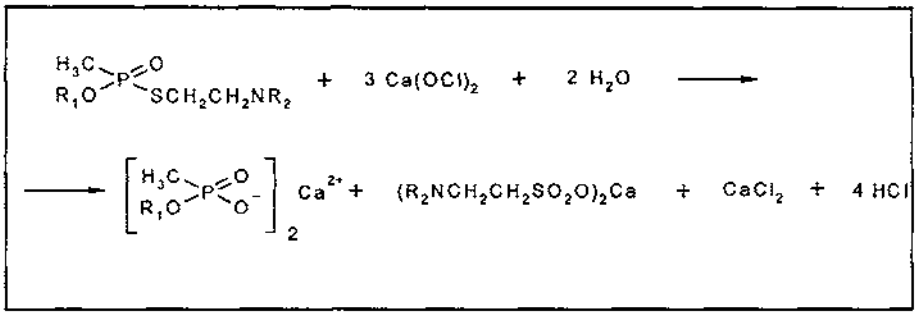
Водните хипохлоритни рецептури с понижено pH на средата (до 8,5-9) притежават висока дегазираща ефективност по отношение на видове бойни отровни вещества и са намерили приложение като универсални рецептури за дегазация на въоръжение и техника. Високият им окислителен потенциал осигурява ефективно обезвреждане на БОВ от типа на иприт (Схема 3).

(3)



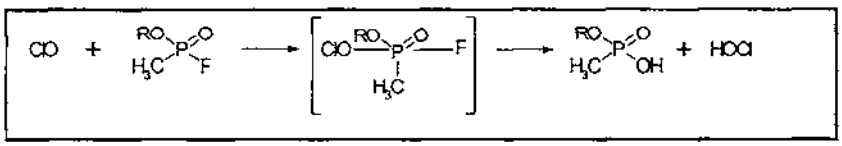
и V-газ (Схема 4),

(4)



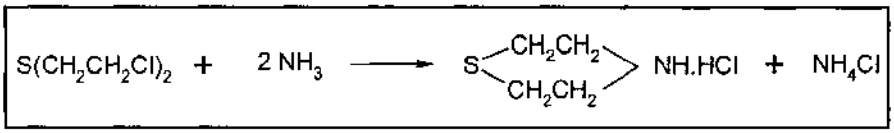
Такова pH гарантира и бърза дегазация на отровни вещества от типа на зарина (зомана), благодарение на каталитичното действие на хипохлоритния йон (OCl-), с което рецептурата се превръща в полидегазираща (Схема 5).

(5)



Станциите за ликвидиране на последствията от типа на „Саниджет" осигуряват още по-ефективно и бързо разграждане на токсичните вещества до нетоксични продукти благодарение на високата температура на подавания за обработка хипохлоритен разтвор.

Високоефективна е и дегазацията на текстилни материали по пароамонячния способ при температура около 100°C. Получените продукти в резултат на реакцията са нетоксични (Схема 6).

(6) 

Хипохлоритните рецептури обезвреждат ефективно и бактериологични агенти.

Своевременната обработка на заразените обекти е гаранция за бързо отстраняване и обезвреждане на токсични химически агенти и свеждане до минимум на тяхното вредно въздействие.

Литература

1. Dr. Norbert Gass. Degradation in operational capabilities due to decontamination of force unit, Canada, 2004.
2. Франке З. „Химия на отровните вещества", Москва, 1976.
3. Симеонов А., Е. Христов и др. Гражданска отбрана, София, 1986.
4. Stanag 2352 NBC (издание 4) – снаряжение за ядрена, химическа и биологична защита.

NEW DEVICES FOR PROTECTION AND CONSEQUENCE MANAGEMENT IN CASE OF TERRORIST ACTS INVOLVING TOXIC CHEMICAL SUBSTANCES

I. Popov, G. Popov

Abstract

New means for individual chemical protection of respiratory tract and skin as effective devices for emergency response in case of use of toxic chemical substances by terrorists are presented. Methods and ways for providing of first aid to injured persons and consequence management in contaminated areas are defined. Perspectives for development and improvement of the protection devices and special treatment means are indicated.

Bulgarian Academy of Sciences. Space Research Institute.
Aerospace Research in Bulgaria. 21, 2007, Sofia

СЪВРЕМЕННИ СИСТЕМИ ЗА АЕРОЗОЛНА МАСКИРОВКА НА БРОНЕТАНКОВА ТЕХНИКА*

Иван Попов[1], Георги Попов[2]

[1]Военна академия "Г. С. Раковски"- Институт за перспективни изследвания за отбраната
[2]Кингстън Енвайрънментал Сървисиз, Канзас Сити, САЩ

Резюме

Представени са нови устройства за въздействие върху бронетранспортьори и системи за аерозолно противодействие. Описани са методите за събиране на ориентировъчни данни и автоматично генериране на аерозолен екран. Определени са свойствата на генерираните аерозолни екрани и режимите на защита спрямо боеприпасите с точно насочване. Направена е оценка на предимствата и недостатъците на новите системи за аерозолен камуфлаж на бронетранспортьори.

Непрекъснатото развитие и усъвършенстване на оптико-електронните средства за наблюдение и разузнаване, както и оръжията с дистанционно насочване и управление, наложи търсене и внедряване на нови съвременни способи и средства за защита. Ефективно средство за противодействие на съвременните високоточни оръжия за защита на

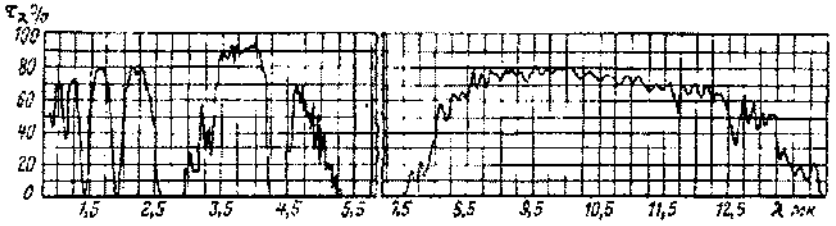
*Докладът е изнесен на Научната конференция с международно участие SENS 2006

бронетанковата техника са системите за аерозолна маскировка. В доклада са разгледани съвременните средства за поразяване на бронетанкова техника и системи за аерозолно противодействие. Разгледано е усъвършенстването на системите в различни страни. Описани са способите за пеленгация и автоматизирано създаване на аерозолните завеси. Разяснени са оптичните свойства на създадените аерозолни дисперсни системи и способността им да осигуряват ефективна маскировка в оптическия и инфрачервения диапазон. Обоснована е зависимостта на защитата на бронетанковата техника от съразмерността на аерозолните частици и дължината на вълната. Указани са способите за защита от управляеми боеприпаси. Оценени са предимствата и недостатъците на съвременните системи за аерозолна защита на бронетанкова техника и перспективите на развитие и усъвършенстване.

Важен дял в комплексните инженерно-технически мероприятия за осигуряване бойните действия на войските заема аерозолната маскировка. Тя намалява вероятността за откриване и опознаване на обектите, а в отделни случаи осигурява и пълното им скриване от разузнаването [1].

Появата на новите съвременни високоточни оръжия (ВТО), обединени в единни системи за разузнаване, целеуказване и насочване на боеприпасите в последната фаза от полета към целта, изведе на преден план ролята на аерозолната маскировка като едно от най- ефективните средства за противодействие. Особено голямо внимание в много страни през последните години се отделя на разработката и внедряването на нови аерозолни състави и средства за защита на бронетанковата техника.

Едно от основните изисквания, които се предявяват към съвременните аерозолни средства е ефективно противодействие на квантово електронните средства за разузнаване и управление на оръжията, работещи в оптическия и микровълновия диапазон. Голяма част от тези средства работят в т.н. "прозорци на атмосферата", които се намират в обхватите 0.7-1.5 мкм., 2 - 2.3 мкм., 3 –5 мкм., и 8 - 14 мкм.[фиг. 1]. Най-предпочитан диапазон е 8-14 мкм., тъй като разсейването, отразяването и поглъщането на различните по честота вълни, присъщи на обикновения слой на атмосферата, в този диапазон са най–ниски [1]. Още по-сложно е аерозолното екраниране на радиолокационни средства, работещи в милиметровия и ултракъсия (микровълнов) диапазон.



Фиг.1. Пропускане на атмосферата в областта 0.61-15 мкм

Аерозолите притежават ясно изразени оптични свойства, благодарение на които могат да се използват за маскировка. Тези свойства се дължат главно на три основни явления [2]:

- разсейване на светлината в аерозолния облак (Тиндалов ефект);
- поглъщане на светлината от частиците на дисперсната фаза;
- отразяване на светлината на границата между димния облак и чистия въздух.

Разсейването на светлината е резултат на взаимодействието на електромагнитните вълни с електронната обвивка на атомите, изграждащи частиците на дисперсната фаза .

Падащите вълни предизвикват периодични колебания в електронната обвивка, в резултат на което тя изпуска вторични импулси, които представляват разсеяна светлина. Тя се получава на границата между димната частица и въздуха и се дължи на явленията отразяване, пречупване, дифракция и др. [2]. Отразяването и пречупването на светлината на границата на аерозолните частици става в случаите, когато размерите на частиците са по-голями от дължината на вълната на падащата светлина. Интензивността на разсеяната светлина, получена в резултат на отразяването и пречупването, се дава със зависимостта [3]:

$$(1) \qquad I_s = k\frac{C}{r},$$

където:
Is - интензивност на разсеяната светлина;
k - коефициент на пропорционалност;

C - масова концентрация на аерозолните частици, g/m3;
r - радиус на частиците, m.

Когато размерите на частиците на аерозола са значително по-малки от дължината на вълната на падащата светлина, интензивността на разсеяната светлина нараства по зависимостта:

(2) $Is = k\,C\,r\,3$.

Най- голямо разсейване, дължащо се на явлението дифракция, се наблюдава при частици с радиус близък до дължината на вълната на преминаващата светлина. Дифракцията е основно явление, което довежда до разсейване на светлината в аерозолите. Интензивността на разсеяната светлина значително се увеличава при намаляване на дължината на вълната. Съгласно закона на Релей, интензивността на разсеяната светлина за каквато и да е дисперсна система се изразява с формулата [3,4]:

(3) $$I_s = I_0 \frac{k.C.r}{\lambda^4},$$

където:
I_0 е интензивност на падащата светлина;
l - дължина на вълната.

Разсейването на светлината от аерозолните частици лежи в основата на нефелометричните определения на структурата, големината и концентрацията на аерозолите. Поглъщането на светлината в различните аерозоли става по различен начин. Тук освен размерът на частиците оказва влияние и химичният състав на дисперсната фаза. Количествена зависимост между погълнатата от аерозола светлина, неговата концентрация и дебелината на поглъщащия слой дава законът на Бугер - Ламберт-Беер [1,2]:

(4) $I = I_0 . e - k.d$,

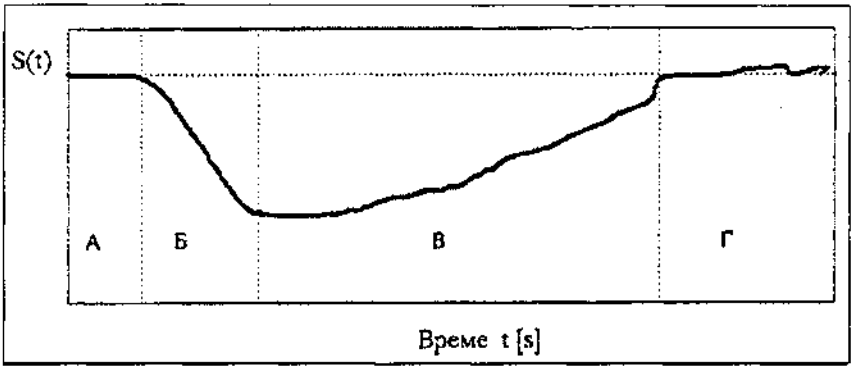
където I_0 е интензивност на светлинния поток, постъпващ в поглъщащия аерозолен слой;

I - интензивност на светлинния лъч, преминал през поглъщащия аерозолен слой;

e - основа на натуралните логаритми;

d - дебелина на аерозолния слой.

Оптическата плътност (прозрачност) на аерозолните дисперсни системи се определя със специална спектроскопична апаратура, която позволява измерването на прозрачността на моделираната в камерата аерозолна система едновременно в пет спектрални канала. Измерването става в относителни единици, като сигналът S (t) в i- тия канал има следния времеви ход[5,6] (фиг. 2).



Фиг. 2. Времеви ход на прозрачността

Този сигнал лесно може да бъде преизчислен в прозрачност. За целта се намира средната стойност S на S(t) в интервалите (А) и (Г). През тези си интервали сигналът S(t) се регистрира преди запалването (изпарението) на аерозолния състав в камерата и след започване на вентилацията. Получената средна стойност S съответства на прозрачност T(t) = 1, т.е. на отсъствие на отслабване на оптично лъчение вътре в камерата. Сигналът през интервала (В) се преизчислява в прозрачност съгласно зависимостта:

$$T(t) = \frac{S(t)}{\bar{S}} \qquad (5)$$

Това е временния ход на прозрачността на аерозолната дисперсна система, създадена след пълното изгаряне на сместа през интервала (Б). Съгласно Закона на Бугер

(6) $T(\lambda;t) = \exp[-k(\lambda i)C(t).l],$

където l е геометрична дължина на пътя на лъчението през камерата (m); k((li) – коефициент на екстинкция на аерозола за i- тия спектрален канал ($g - 1m^2$).

Масовият коефициент е оптичната характеристика, която подлежи на оценяване на спектроскопичните измервания. На практика при микрофизичните измервания се оценява не C(t), а средната стойност:

(7) $$C = \frac{1}{t}\int_0^{\Delta t} C(t)dt$$

където [0, Dt] е подинтервал на B, през който през мембранен филтър се извършва пробоотбор на аерозоли в камерата.

Това налага да се направи съответното усредняване на T(t). Сигналът S(t) се регистрира в цифров вид и подобно усредняване не представлява проблем. Много често в интервала [0, Dt] прозрачността се изменя по почти линеен закон, чиито параметри T_0 и T_1 могат да се определят по метода на най-малките квадрати[5,6,7], т.е.:

(8) $$T(t) = T_0 + \frac{T_1 - T_0}{\Delta t}.t, T_1 > T_0$$

Използваните до скоро димообразуващи състави, известни като "класически ",имат размери на частиците от 0,3 до 1.2 мкм и осигуряват ефективна маскировка на обектите само във видимия и близкия инфрачервен (оптически) диапазон в обхвата 0 .4 – 1.5 мкм [2,3].

Новите изисквания за повишаване обхвата на маскировка са поставили пред разработчиците и производителите на средства за аерозолна маскировка сложни проблеми за повишаване абсорбционната, разсейващата и отразяващата способност на аерозолната дисперсна фаза.

При това в много от случаите остро е стоял проблемът за бързо създаване на аерозолната завеса. За привеждане на аерозолните състави в съответствие с новите изисквания са се очертали няколко тенденции.

На първо място продължават да се усъвършенстват аерозолните състави с ефективна маскираща способност в оптическия диапазон.

Втората тенденция е модифициране на традиционните състави чрез увеличаване маскиращата им способност и в инфрачервения диапазон.

Особена актуалност през последните години придоби третата тенденция за разработване на нови състави с ефективна маскираща способност в оптическия, инфрачервения и микровълновия диапазон.

Успоредно с развитието и усъвършенстването на аерозолните състави се разработват и усъвършенстват и средствата за генерация на аерозоли.

Особено голямо внимание в много страни през последните години се отделя на разработката и внедряването на нови аерозолни състави и средства за защита на бронетанковата техника. Като основно средство за индивидуална аерозолна защита на танкове и БМП военните специалисти считат мортирните гранатохвърлачни димни установки. Основни производители са САЩ, Англия, Франция и Германия. Установките са предназначени за изстрелване на димни гранати и са интегрирани със система от датчици за регистриране на лазерна подсветка за автоматизирана защита. Приети са на снабдяване 12-стволни мортирни установки за танкове и 8-стволни за бронирани машини от типа на БМП, БТР и др.

Типовата гранатометна система монтирана на танкове в САЩ се състои от дванадесет 66 мм. димни гранати, разположени по шест броя от всяка страна на куполата. Основно димообразуващо вещество е червен фосфор. За 2-3 сек. на разстояние 25-30 м. може да бъде поставена димна завеса с височина 13 м. и широчина 38 м. (защитаван сектор –110 градуса), ефективна в продължение на 2-3 мин.

Изискването за маскировка в по-широк спектрален диапазон е реализирано в създадената 66 мм. граната на САЩ, която е предназначена за създаване на смущения в средствата за насочване, работещи в оптическия и средния инфрачервен диапазон. Тя може да се изстрелва от болшинството приети на въоръжение гранатохвърлачни установки М-239, М-250, М-258 и др., монтирани на танкове и бронетранспортьори.

Две нови димови гранатохвъргачни установки са разработени в Англия. Едната от тях, "VIRSS", се състои от 12 огневи контейнера, във всеки от които са разположени по 20 димни гранати. Аерозолната завеса се образува и поддържа чрез последователен запуск (регулира се автоматично) на всичките 240 гранати от комплекта. Взривяването на гранатите създава област на висока температура, благодарение на което се осигурява маскировка на бронетанковата техника във видимия и инфрачервения диапазон в продължение на около 2 минути.

В другата установка, MBSMR-3, изстрелването на гранатите (общо 12 броя) се извършва едновременно.След това всяка граната изхвърля по два земни подбоеприпаса и един въздушен. Последният от своя страна се разделя на 6 елемента със сферична форма. В продължение на 3 сек. на разстояние 15-20 м от машината в сектор около 110 градуса се образува димна завеса с височина 5 м и широчина до 40 м, която има ефективна маскираща способност в средния инфрачервен диапазон 3-5 и 8-14 мкм в течение на 35-40 сек, а в оптическия диапазон прикритието на обекта продължава 60-80 сек. Установката MBSMR-3 е съвместима с много приети на въоръжение в армиите от НАТО 66 мм гранатохвъргачни установки и освен това позволява на екипажа веднага след залпа да извърши маньовър – да смени местоположението си. За установката VIRSS, маньовърът е възможен само след изстрелване на всички гранати.

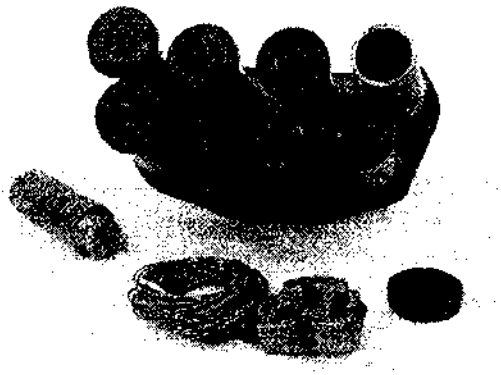
Френската универсална гранатохвъргачна установка GALIX се състои от 8 пускови тръби, които са ориентирани по чифтно в различни направления. Количеството на гранатите в един залп е 4, 6 или 8 броя. В комплекта на установката са включени следните видове боеприпаси: димни гранати, топлинни ловушки, създаващи смущения в инфрачервените разузнавателни средства в продължение на 10 сек, противопехотни гранати с повишена защита на предния сектор на машината, противопехотни боеприпаси, всеки от които включва две осколочни и една фугасна граната, сълзотворни и осветителни гранати.

Някои американски танкове са снабдени със система за димопуск VEESS. В качество на димообразуващо вещество в нея се използва дизелово гориво, което се впръсква в потока горещи газове на двигателя. Системата осигурява създаване на димна завеса на височина 10 м, широчина 8 м за 5 сек. Продължителността на действие на завесата се определя от запаса на гориво. Тази система се използва в съчетание

с бордовата гранатохвъргачна установка. Конкретният способ на действие се избира от екипажа на машината в зависимост от скоростта и направлението на вятъра, интензивността на слънчевата светлина и характера на действие на противника.

За самозащита на бронетанковата техника в Русия е приета на снабдяване гранатохвъргачна установка " Туча" [фиг. 2]. Максималната далекобойност на изстрелваните гранати 3Д-6 достига 300 – 400 м, времето за разгаряне на димната смес е около 15 сек, а времето за интензивно димообразуване от 1 до 1.5 мин. Броят на мортирите за изстрелване на гранатите е съобразен с характеристиките на бронетанковата техника. На БТР се монтират 4 мортири, МТЛБ (БМП),-6, леки танкове-8 и тежки танкове-12. В зависимост от броя на изстрелваните димни гранати широчината на заслепяващата за противниковите огневи средства полоса може да достигне до 140 м.

Аерозолният състав на гранатите е металхлориден, с размери на аерозолните частици на дисперсната фаза в интервала от 0.1 до 1.7 микрона. Съставът се отнася към "класическите" и осигурява ефективна маскираща защита само в оптическия диапазон [3]. Горенето на състава е бавно (за около 1.5 мин), при температура около 900 градуса и е съпроводено с отделяне на метални хлориди. Те от своя страна поглъщат влагата от въздуха и образуват капчици разтвор:



Фиг.3. Система "Туча"

Затова, образуващият се при горенето дим първоначало има сив цвят, който впоследствие преминава в сламено жълт.

По схващания на руските военни специалисти, гранатите ще се изстрелват при достигане на 2 – 3 км от предния край на противника за извършване на скрит маньовър на бойното поле.

Системата "Туча" е усвоена по лиценз и в нашата страна. Подобна система "Телур" е разработена и приета на снабдяване в Полша. През последните години полските специалисти са разработили за ситемата нови димни гранати, задействащи се по взривния способ [4]. В съчетание с датчици за регистрация на лазерна подсветка, новите димни гранати осгуряват за около 15 сек ефективна защита в средния инфрачервен диапазон и около 2 мин. в оптическия диапазон.

В съответствие с новите постановки и изисквания руските военни специалисти са разработили през 2001 г. нова танкова гранатохвъргачна система за защита от високоточните оръжия ТШУ-1С ("Штора-1").Системата защитава обектите на бронетанковата техника от високоточните оръжия, използуващи при своята работа лазерно излъчване. Със системата се решават автономно следните по важни задачи:

- определя се видът и източникът на лазерно излъчване;
- бронетанковият обект се защитава автоматично от източника на лазерна подсветка чрез поставяне на аерозолна завеса;
- оповестяват се членовете на екипажа за регистрирано лазерно облъчване;
- създава се димна завеса едновременно от всички гранатомети в екстремални условия.

Системата "Штора –1", съвместно с електрооборудването на обекта, осигурява автоматична защита на бронетанковата техника и понижава прицелния огън на управляемите противотанкови системи с вероятност 0.8-0.9. Внедряването на системата осигурява надеждна защита и значително повишаване на тактико-техническите характеристики и живучестта на бронетанковата техника.

Съвременната бронетанкова техника на страните от бившия "Варшавски договор" е оборудвана със собствена система за димопуск, аерозолният състав на която е дизелово гориво. Системата осигурява създаване на аерозолни екрани за самозащита при маньовър на бронетанковата техника.

Бронетанковата техника на Българката армия е комплектована с гранатохвъргачни установки "Туча" и системи за димопуск с дизелово гориво.

Новите изисквания за оперативна и техническа съвместимост с армиите от НАТО налагат преразглеждане на съществуващите системи и замяната им с по-високоефективни.

Литература

1. Козелкин В., И. Усольцев. "Основы инфракрасной техники", Москва, 1974.
2. Шидловский А. "Основы пиротехники", Москва, 1976.
3. Огнеметно-запалителни и димни средства. ДВИ, София,1975.
4. Полски патент 119927/С06Д3/00/1987.
5. Цанев В., Б. О. Тодоров и др. "Камера и спектроскопична апаратура за изследване оптичните свойства на маскиращи димове", ЮНС ,ВНТИ, 1997.
6. Цанев В. И., Г. К. Попов. "Предварителна обработка на данни от изследванията на оптичните свойства на маскиращи димове", ЮНС, ВНТИ, 1997.
7. Гаврилов В. А. " Видимость в атмосфере", Ленинград, 1966.

NEW SYSTEMS FOR AEROSOL CAMOUFLAGE OF ARMOURED VEHICLES

I. Popov, G. Popov

Abstract

New devices for impact on armoured vehicles and systems for aerosol counteraction are presented. Methods for taking the bearings and automatic generation of aerosol screen are described. Properties of generated aerosol screens and modes for protection against Precision Guided Munitions (PGM) are determined. Advantages and disadvantages of the new systems for aerosol camouflage of armoured vehicles are assessed.